ANNUAL REVIEWS

Annual Review of Plant Biology

Advances and Opportunities in Single-Cell Transcriptomics for Plant Research

Carolin Seyfferth,[1,2,*] Jim Renema,[1,2,*] Jos R. Wendrich,[1,2] Thomas Eekhout,[1,2] Ruth Seurinck,[3,4] Niels Vandamme,[3,4] Bernhard Blob,[5,6] Yvan Saeys,[3,4] Yrjo Helariutta,[5,6] Kenneth D. Birnbaum,[7] and Bert De Rybel[1,2]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium; email: bert.derybel@psb.vib-ugent.be

[2]VIB Center for Plant Systems Biology, 9052 Ghent, Belgium

[3]Data Mining and Modeling for Biomedicine, VIB Center for Inflammation Research, 9052 Ghent, Belgium

[4]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 9000 Ghent, Belgium

[5]The Sainsbury Laboratory, University of Cambridge, Cambridge CB2 1LR, United Kingdom

[6]Viikki Plant Science Centre, HiLIFE/Organismal and Evolutionary Biology Research Program, Institute of Biotechnology, Faculty of Biological and Environmental Sciences, University of Helsinki, 00014 Helsinki, Finland

[7]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA; email: ken.birnbaum@nyu.edu

Annu. Rev. Plant Biol. 2021. 72:847–66

First published as a Review in Advance on March 17, 2021

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

https://doi.org/10.1146/annurev-arplant-081720-010120

*These authors contributed equally.

Keywords

scRNA-seq, plant transcriptomics, protoplast, cell atlas, cell trajectory

Abstract

Single-cell approaches are quickly changing our view on biological systems by increasing the spatiotemporal resolution of our analyses to the level of the individual cell. The field of plant biology has fully embraced single-cell transcriptomics and is rapidly expanding the portfolio of available technologies and applications. In this review, we give an overview of the main advances in plant single-cell transcriptomics over the past few years and provide the reader with an accessible guideline covering all steps, from sample preparation to data analysis. We end by offering a glimpse of how these technologies will shape and accelerate plant-specific research in the near future.

Contents

1. MILESTONES IN THE DEVELOPMENT OF PLANT SINGLE-CELL RNA-SEQUENCING

For centuries, biologists have been striving to precisely capture, characterize, and classify cells within multicellular organisms. The identity of a cell has long been primarily determined according to morphological features, spatial information within the tissue or organ, cell-to-cell communication, or unique cellular content such as specific organelles. However, with the onset of molecular biology, it became clear that the identity of individual cells can be characterized by unique spatiotemporal gene expression profiles. As such, the transcriptome can be thought of as a unique molecular fingerprint in which gene activity is a readout for cell function, developmental stage, or disturbances caused by internal and environmental stimuli. This concept has intriguing implications, as the transcriptome of a single cell would theoretically contain all required information to determine its origin, identity, and developmental stage within a complex multicellular tissue, organ, or even the full organism. Harnessing this highly complex information at a sufficient resolution has, however, been a major technical hurdle, even in well-studied model species such as *Arabidopsis thaliana*. Over the past two decades, stepwise technical advances in next generation sequencing (NGS) and microfluidics technology have been instrumental in mapping the transcriptome at cellular resolution (**Figure 1**).

Microfluidics: manipulation of liquids down to the picolitre range, used in single-cell approaches to generate droplets

Bulk RNA-sequencing (bulk RNA-seq): the common name for RNA-sequencing of pooled cells, tissues, or organisms

Fluorescence-activated cell sorting (FACS): a flow cytometry–based method to separate and sort cells based on parameters such as size or fluorescence

A first major step moving beyond profiling individual genes in a complex sample came with the onset of mapping a large subset of the *Arabidopsis* transcriptome via microarray (89) and, later, the full transcriptome via bulk RNA-sequencing (bulk RNA-seq) technologies (46) (**Figure 1**). Although the sample itself usually was a convoluted mixture of different cell types, tissues, or entire organisms, RNA-seq revealed molecular mechanisms and pathways underlying developmental processes, adaptation, stress, and immune responses at a whole-genome level. Pioneering work in the plant field combining cell-type-specific reporters and fluorescence-activated cell sorting (FACS) allowed the deconvolution of the complexity of the *Arabidopsis* root into radial cell identities and longitudinal zones (9). This resulted in high-resolution gene expression maps of an organ at whole-genome resolution (9, 10, 20, 29, 44). Together with subsequent studies mapping

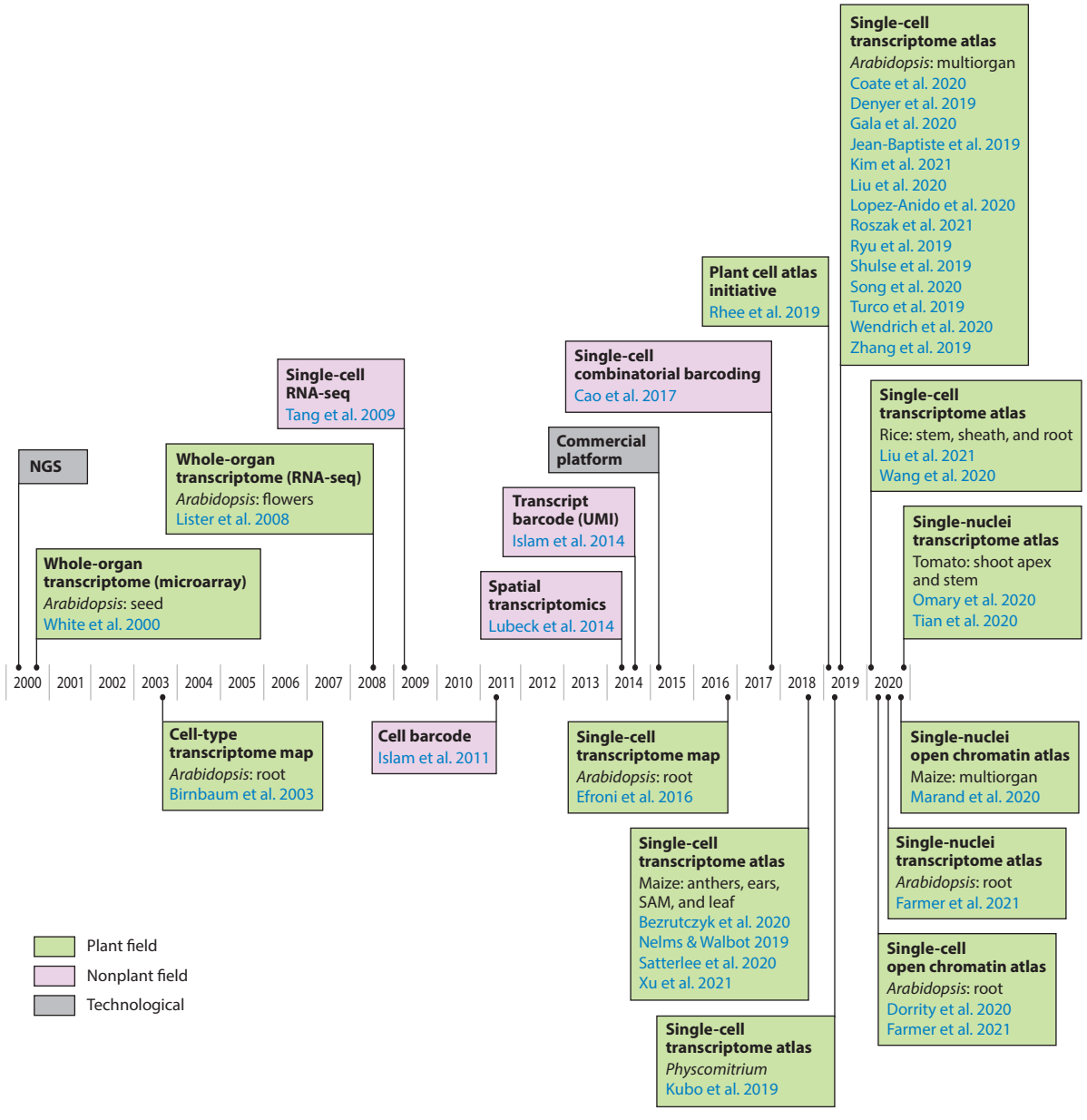

Figure 1

Timeline of key events in plant single-cell transcriptomics. Milestones related to the development and establishment of single-cell transcriptomics in the plant field are illustrated by green boxes. General technological developments are shown in gray, and key developments outside of the plant field are indicated in pink. Each dot marks the appearance of the first publication related to each box. Abbreviations: NGS, next generation sequencing; RNA-seq, RNA sequencing; SAM, shoot apical meristem; UMI, unique molecular identifier.

cellular responses to different environmental factors, such as nitrogen (29), salt, and iron (20), these data sets became heavily used resources to trace spatiotemporal gene expression profiles in roots. The requirement for a priori knowledge in the construction of tissue-specific reporter genes has, however, hampered the use of this technology beyond well-characterized model species.

Moving from pooled cells to profiling single-cell transcriptomes required retrieving reliable gene expression values from the very low amounts of RNA present in a single cell (11). Early technological advances that enabled an exponential amplification of messenger RNA (mRNA) from a single cell (43, 80) provided the basis for the first whole-genome single-cell RNA-seq studies in plants, in which single *Arabidopsis* root cells carrying a fluorescent tag were sorted into wells using FACS, followed by full-transcriptome sequencing (24). Although limited in practice to a few hundred cells, these studies illustrated the immense potential of single-cell transcriptomics in plants.

The transition from plate-based to nanoliter-sized droplet-based assays using microfluidics (79) allowed an increase in the throughput from hundreds to thousands of single cells per experiment at affordable costs. A cell barcode system was introduced in which randomly designed short nucleotide sequences serve as unique labels that enable sequencing reads obtained in a pool of mRNAs from single cells to be assigned back to their original cell (37). This barcoding principle was also implemented to tag each captured individual transcript within a cell with unique molecular identifiers (UMIs) in order to correct for polymerase chain reaction (PCR)-induced replication biases (38). The adoption of droplet-based methodologies into accessible commercial platforms fully embedded this technology in modern plant research (7, 18, 19, 21, 26, 28, 39, 41, 47–49, 54, 68, 71–73, 81, 83, 87, 88, 91, 94). These studies illustrate how single-cell transcriptomes give sufficient information to cluster cells according to their identity and response to stimuli. Moreover, cells can be ordered along a pseudotime axis, allowing the reconstruction of cell trajectories within multiple cell types (7, 18, 19, 28, 39, 42, 47–49, 54, 60, 62, 67, 68, 71–73, 81, 83, 88) and gene regulatory networks (GRNs) that control them (19, 81). These networks will be further refined by the addition of single-nuclei assay for transposase-accessible chromatin (snATAC)-seq data (21, 26, 54), allowing the exploration of chromatin accessibility at the level of individual cells.

In conclusion, plant research has fully embraced single-cell transcriptomics, and its application to a variety of plant species is ongoing. Pioneering studies in rice, maize, and tomato (7, 47, 54, 60, 62, 71, 81, 87, 91) indicate the utility of single-cell transcriptomics in crop species and thus lead the way for its integration into applied plant research. Although the field of single-cell technologies progresses at a very fast pace, in this review we provide the reader with an accessible guideline encompassing all steps, from sample preparation to data analysis. We furthermore illustrate how single-cell technologies will shape and accelerate plant-specific research in the near future.

Barcode: a short nucleotide sequence used as a unique cell- or sample-specific recognition tag

Unique molecular identifier (UMI): a short oligonucleotide sequence unique to each transcript from an individual cell that enables quantification and correction of polymerase chain reaction amplification biases

Cell trajectories: biological processes that cause heterogeneity among cells

Gene regulatory network (GRN): a network of molecular regulators that regulates molecular processes through transcriptional regulation

Single nuclei assay for transposase-accessible chromatin-sequencing (snATAC-seq): retrieves information on the open chromatin status of individual nuclei

Compart-mentalization: isolating a single cell into a compartment (e.g., a droplet, tube, or well)

2. AN OVERVIEW OF THE SINGLE-CELL RNA–SEQUENCING WORKFLOW

The generation and analysis of a single-cell transcriptome data set from a plant sample require a multifaceted workflow, comprising (*a*) the generation of individual cells and compartmentalization; (*b*) library construction and sequencing; and (*c*) demultiplexing, analysis, integration, and validation of the data (**Figure 2**). This section summarizes the state-of-the-art process used to generate single-cell transcriptomes from plant samples and highlights the most important pitfalls. We end this section by describing more advanced analysis methods to retrieve meaningful biological information from the obtained data set.

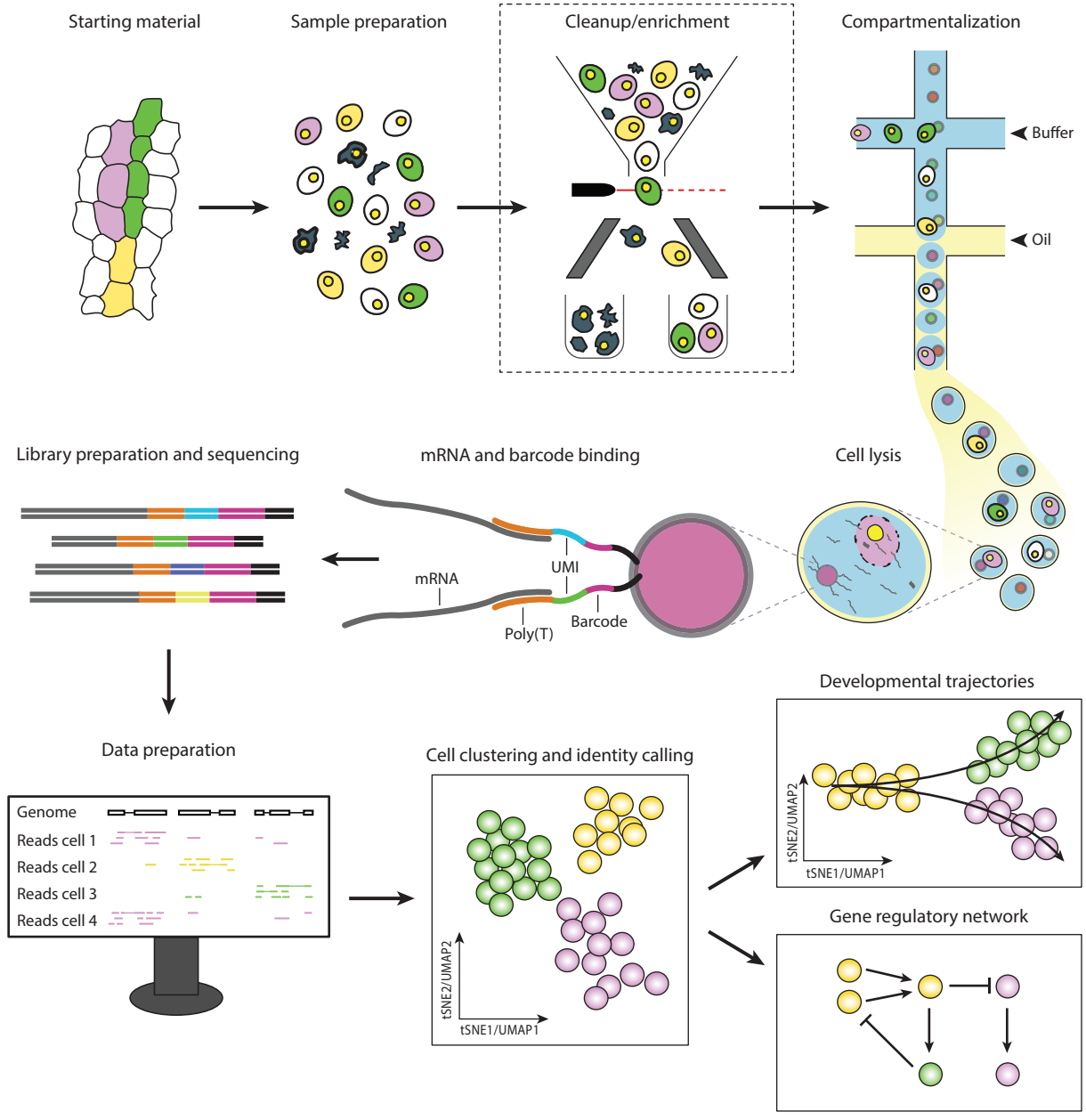

Figure 2

Schematic overview of a typical droplet-based plant scRNA-seq experiment. Individual cells are isolated from the tissue, which contains different cell identities (*colored cells*), by enzymatic digestion. As an optional step (*dashed line*), specific cell populations can be enriched using FACS. Next, cells are compartmentalized together with barcoded beads, buffers, and enzymes required for library preparation into droplets. After droplet formation, cells are lysed, and the released mRNAs are bound to the bead; this is followed by library preparation and next generation sequencing. Next, raw sequence reads are filtered and normalized, after which the reads are mapped against a reference genome. Dimensionality reduction clustering is performed before assigning cell identities. The resulting data set can then be used to, e.g., predict gene regulatory networks or perform trajectory inference analysis. Abbreviations: FACS, fluorescence-activated cell sorting; mRNA, messenger RNA; poly(T), polythymidylic acid tail; scRNA-seq, single-cell RNA-sequencing; tSNE, t-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection; UMI, unique molecular identifier.

2.1. The Generation of Single Cells in Plants

Protoplast: a plant cell from which the cell wall has been removed by enzymatic digestion

Multiplets: a droplet containing two or more cells

Similar to other research fields, the first step in plant-based single-cell experiments is the isolation of single cells. Several methods are available for isolation and separation of plant single cells.

2.1.1. Cell isolation methods. Given that plant cells are characterized by rigid cell walls, this first step can be especially challenging depending on the tissue type or plant species. Enzymatic cell wall digestion and manual isolation remain the two principal ways to enable the dissociation of single cells from the tissue context in plants.

Enzymatic digestion to release single cells from their tissue context results in a suspension of cells without cell walls (called protoplasts) and has been the most common way to isolate the individual cells for recent single-cell transcriptome experiments in *Arabidopsis*, maize, tomato, and rice (7, 9, 10, 18, 19, 23, 24, 28, 39, 41, 47–49, 60, 62, 67, 71–73, 83, 87, 88, 91, 94) (**Figure 2**). Careful optimization and validation are required in both biological and technical aspects to retrieve all cell types of interest from the isolated tissue.

Manual isolation methods provide an alternative to enzymatic dissociation of cell walls. These include microcapillary-based approaches to extract the content of individual cells and microdissection-based techniques to isolate a region or cell of interest. Microcapillary-based extractions have been used extensively in plant research, as they allow the acquisition of the content of specific or rare cells from a broad spectrum of plant species and tissues (45, 50), and were recently also adapted to single-cell experiments (17, 42, 75). This approach was used to obtain single-cell RNA-sequencing (scRNA-seq) data from *Physcomitrium patens* [formerly known as *Physcomitrella patens* (64)] leaves (42) (**Figure 1**) in combination with a nuclear fluorescent protein, allowing a controlled disruption of cells and subsequent uptake of the nucleus and surrounding cytoplasm. In laser capture microdissection (LCM), tissues are first fixed by embedding or freezing, sectioned, and placed on a slide before individual cells or a specific region is isolated and removed using a laser. Although LCM has not been used to generate single-cell transcriptome data yet, this method has been applied in hard stems (15) and fleshy tomato fruit bodies (55), indicating its efficacy in a wide range of tissue types.

Although each of these manual approaches has its specific merits, a main advantage of both is that they can be applied to any tissue or plant species without the need for prior knowledge or molecular tools as long as the cells of interest can be differentiated according to, e.g., size or anatomic features. Moreover, because cells do not undergo enzymatic digestion but are fixed or directly processed, the influence of preparation on gene expression is considered low, as seen in scRNA-seq data from *P. patens* (42). However, these manual approaches are laborious and come with relatively low throughput.

2.1.2. Compartmentalization of cells. Once a cell mixture is obtained, individual cells need to be separated and compartmentalized into tubes (42, 60, 75), wells (24, 49, 62, 67, 71), or droplets (7, 18, 19, 21, 28, 39, 41, 47, 68, 71, 72, 73, 83, 87, 88, 91, 94) for further processing. Each tube, well, or droplet contains the necessary components to perform the amplification of transcripts and library preparation. The technique used for cell compartmentalization merits high precision, as the presence of multiple cells in a compartment (multiplets) or the inclusion of impurities such as cell debris and free mRNA leads to data loss in downstream processing. FACS has been used to distribute cells in some plate-based scRNA-seq approaches (24, 49, 62, 67), allowing simultaneous enrichment for specific cell populations of interest using fluorescent reporter lines. Independent of the compartmentalization into wells or droplets, a FACS-based sorting step can be useful to obtain multiple characteristics of a single cell (e.g., size and/or fluorescence), which can help isolate the

targeted cell population (24, 49, 62, 67) and/or reduce undesired cell debris (62, 67, 88). In this aspect, the implementation of FACS devices equipped with image stream analysis is of high interest (33) and can allow more precise cell quantification to reduce the chance of multiple cells being sorted into a single well. However, cell sorting comes at the cost of additional processing steps and increased processing time. The final choice of cell isolation and compartmentalization strategy is highly dependent on the tissue, species, commercial platform usage, and envisaged application. In any case, careful consideration is required before initiating a single-cell experiment.

Single-cell transcriptomes attached to microparticles (STAMPs): the combination of barcoded beads with attached mRNA transcripts from a single cell

2.2. Single-Cell Transcriptomics Technologies Used in Plant Research

High-throughput and low-cost cell processing allowed droplet-based single-cell methods [e.g., Drop-seq (53) or the Chromium Controller from 10X Genomics] to dominate the plant single-cell transcriptome landscape, including applications in *Arabidopsis* (18, 19, 21, 26, 28, 39, 41, 47–49, 68, 72, 88, 94), rice (47, 87), and maize (6, 71, 91) (**Figure 1**). Given the prominent role of droplet-based methods in the plant field, a typical single-cell experiment using this setup is described below and depicted in **Figure 2**.

The general principle underlying droplet-based methods is the encapsulation of an individual cell and a barcoded bead in a nanoliter-sized droplet. The bead is coated with DNA oligonucleotide probes that encode a cell-specific barcode and transcript-specific barcodes (UMIs) equipped with a poly(T) sequence (**Figure 2**). The droplets contain all buffers and enzymes required for complementary DNA (cDNA) synthesis. Upon encapsulation in the droplets, cells are lysed, which allows the poly(A)-tailed mRNA to bind to the beads. Reverse transcription is used to generate single-cell transcriptomes attached to microparticles (STAMPs). Because transcript coverage mainly occurs at the 3′ end, processing is fast but contains limited information about genetic variation and relative abundance of transcript isoforms is not obtained (96). After mRNAs aggregated into a STAMP, the individual droplets are combined into one sample in order to facilitate further PCR-based amplification and library preparation, followed by RNA-Seq. This setup allows droplet-based scRNA-seq methods to process thousands of cells while providing sufficient resolution and sequencing depth to generate atlases of whole organs in a single experiment.

A different 3′-based method called cell expression by linear amplification and sequencing (CEL-seq2) (34) has been applied to generate a transcriptome map of the transcriptional landscape of the maize shoot apical meristem (71) and to precisely define cellular transition states during maize anther development (60). Similar to droplet-based methods, CEL-seq2 uses primers that are equipped with a cell barcode and UMI, while the major difference is that the mRNA amplification in this plate-based approach is done via in vitro transcription instead of PCR, which can reduce amplification bias (34).

For scRNA-seq experiments where high (or full-length) transcript coverage is of the essence, the SMART-Seq protocols have been applied in a plate-based setup (24) (**Figure 1**). SMART-Seq combines reverse transcription with a template-switching mechanism at the 5′ end of the RNA template to generate a full-length cDNA (96). Additionally, the full-length transcript coverage can provide information on expression levels of transcript isoforms (31). Although the extended processing time associated with full-length transcript amplification makes this approach less suitable for high-throughput single-cell analysis, it is highly suitable for experiments focusing on specific or rare cell types (24, 49, 67). The SMART-Seq protocol has been utilized to study transcriptional changes that govern all steps in the development of protophloem sieve elements, a rare cell subgroup that is most of the time not captured in the scRNA-Seq data set of the *Arabidopsis* root (67).

2.3. Data Analysis

Dimensionality reduction: the process of reducing the number of dimensions in a high-dimensional space using visualization

Highly variable genes (HVGs): genes with high contributions to cell-to-cell variation within a homogeneous cell population

The analytical workflow of single-cell transcriptomes is a multistep process. Several software packages, including Cell Ranger (95), Seurat (13), and COPILOT (72), combine pipelines to perform quality control, normalization, dimensionality reduction of the raw data, and visualization. In this section, we describe the different steps during data analysis as currently performed in the plant field and highlight some of the potential options and pitfalls.

2.3.1. Demultiplexing, quality control, and normalization. After sequencing and the initial quality control used for the detection and removal of low-quality bases (e.g., with FastQC), the sequence files are demultiplexed to assign sequences back to their cells of origin using the unique cell barcodes. Next, another round of demultiplexing is performed using UMIs to quantify the number of mRNAs captured within a single cell. The usage of UMIs largely improves expression quantification across thousands of analyzed cells, as it identifies PCR duplicates and reduces background noise and PCR amplification bias (38). The non-mRNA sequences (e.g., the barcode and UMI) are removed before alignment to the reference genome/transcriptome. Several steps can be taken at this point to evaluate the quality of cell compartmentalization and library preparation. For example, a comparison between the mapping efficiency of endogenous RNA to spike-ins (control transcripts with known sequence and user-defined concentration) gives an estimate about the quality of the library used for sequencing. RNA degradation or inefficient cell lysis can lead to a low-quality library, which would be detectable by low mapping ratios to endogenous RNA but sufficient mapping to spike-ins. An additional quality check is done by calculating the number of counts per barcode (summarized as count depth), the number of genes per barcode, and the ratio of counts from mitochondrial and chloroplast genes per barcode. A combination of a low count depth, low number of genes, and high number of counts from mitochondrial or chloroplast genes per barcode can indicate a low-quality cell, while an unexpectedly high number of counts and genes per barcode can indicate a potential mixture of two or more cells (multiplets). A low count depth can indicate either a cell with low-quality RNA or an empty droplet. In either case, low-quality cells or multiplets should be removed before normalization. This normalization step is used to adjust variation between cells of the same sample that is caused by differences in cell lysis, library preparation efficiency, or sequencing depth. These variations occur because each count per barcode is the result of an mRNA molecule that underwent capturing, reverse transcription, and sequencing. Thus, even slight differences in efficiency within these steps can cause a high variability between cells, making normalization a critical step (85). Once normalization is performed and gene expression has been quantified, a subset of genes with the highest variation in average expression level across all cells [highly variable genes (HVGs)] (11, 16, 22) is used to determine cell clusters.

2.3.2. Dimensionality reduction and clustering. Depending on the plant species, sequence coverage, and depth, scRNA-seq data sets can contain the expression values of, for example, 25,000 genes in *Arabidopsis* (2) or 40,000 in maize (40) for each of the thousands of cells. To cope with the computational and statistical challenges that come with comparing gene expression profiles among thousands—and up to millions—of cells, HVGs are identified (19, 26, 39, 47, 48, 60, 67, 68, 72, 73, 83, 87, 91, 94) and uninformative genes are removed (referred to as feature selection). Dimensionality reduction tools are used to visualize the high-dimensional data set in only two or three dimensions. Principle component analysis (PCA) has been used to reduce dimensionality and determine cell clusters in plant scRNA-seq data sets (19, 26, 47, 48, 60, 72,

83). PCA determines variables that contribute to the overall variance of expression among all individual cells and compares them to establish which variables contribute the most to the overall variance. PCA has been applied in combination with other dimensionality reduction tools, e.g., t-distributed stochastic neighbor embedding (t-SNE) (86), to visualize cell clusters in plant scRNA-seq studies (18, 19, 26, 48, 67, 72, 81, 83, 88). t-SNE plots visualize how many clusters were identified and how many cells belong to each cluster. However, the positions and distances between clusters on the t-SNE plot hold little information. By contrast, uniform manifold approximation and projection (UMAP) (57) allows comparison of both cell clusters and cells within these clusters based on distances between them. Given that UMAP also allows the visualization of more than two dimensions and has a rapid runtime performance (5), it is quickly becoming the preferred method in the plant field (7, 21, 26, 28, 39, 41, 47–49, 54, 62, 71, 72, 81, 87, 91, 94).

Several algorithms are available to identify and characterize cell populations by unsupervised clustering (22). Plant scRNA-seq studies mainly used the Louvain clustering algorithm (21, 28, 39, 54, 67, 83) based on shared nearest neighbor networks (19, 21, 26, 39, 47, 48, 68, 72, 73, 83, 91, 94), which calculates the nearest neighbors of each cell and clusters those with the highest overlap. However, comparative studies indicate that clustering outputs can differ significantly between methods (27), which can, for example, affect connections between cells and the total number of clusters.

t-Distributed stochastic neighbor embedding (tSNE): a dimension reduction and projection algorithm

Uniform manifold approximation and projection (UMAP): a dimension reduction and projection algorithm

Marker gene: a gene with a specific expression pattern restricted to a cell identity, subpopulation, or cells, preferably across multiple data sets

2.3.3. Cell identity assignment and validation. The cell identity of each cluster can be defined by overlaying the expression of specific marker genes, requiring previous knowledge of the gene expression in the tissue or species of interest. For the root scRNA-seq data sets in *Arabidopsis*, this can be done via correlation analyses of marker gene expression in specific clusters and validated data sets (10, 44) and by computing an index of cell identity score (23) for each cell (19, 72, 73, 83). The process of cell clustering and identity calling is, however, an iterative one, as changing cluster parameters has a great impact on clustering output and the ability to call subcell types.

Once clustering and annotation have been performed, a typical first step in validating newly generated data sets is to investigate cell cluster–specific genes for known marker genes (besides those used for initial clustering) or other indications that match the expected biological function, such as metabolic pathways. Typically, marker gene expression is not only specific to a cell cluster but also expressed in most cells of that cluster. Variations of marker gene expression among cells of the same cluster can have multiple causes. For example, the gene in question might only be expressed in a specific developmental stage or cellular subtype. In addition, the technical noise of single-cell techniques could lead to false negatives for lowly expressed genes. Comparing marker gene expression of genes associated with proliferation (e.g., cell cycle–related genes) can be used to fine-tune the composition of a cell cluster (21, 28, 49, 60, 71, 88). Alternatively, plants show distinct changes in cell ploidy levels throughout their development, and sets of marker genes have been associated with the switch from the mitotic cell cycle to the endocycle (8). These endocycle markers can be used to validate the progression of cell lineages in time (21, 28, 49, 60, 71, 88). It is worth noting that some developmental processes (e.g., cells with different identities undergoing division) might even cause dominant transcriptional changes that could mask the actual cell identities, causing cells from different cell lineages to cluster together. Additionally, marker gene expression might be affected by biotic or abiotic stimuli, which need to be taken under consideration during the experimental and project design. While marker genes for root cell identities in *Arabidopsis* have been extensively studied and validated, fewer marker genes are available for other organs and even fewer for other species. Automated cluster annotation methods have facilitated cell identity annotation in the mammalian field (1), offering important alternatives to the manual annotation currently done in the plant field.

Although the options mentioned above should be explored for each data set generated and offer a fast way to provide an initial validation of the data, it is important to stress that there is no substitute for in vivo validation. As the predictive power of a large-scale resource defines its usefulness for the community, validating that predicted cell clusters correspond to spatial or temporal cell identities is a key step in ensuring the utility of the data. This can be done by validating expression patterns with transgenic reporter lines (19, 41, 49, 62, 67, 88) or by in situ hybridization (7, 47, 54, 67, 71, 91). Ultimately, functional relevance is a key step in validating the utility of the profiles using genetically perturbated mutants (28, 41, 62, 68, 72, 83, 91). Once validated, scRNA-seq data sets can be mined to extract additional transcriptional signatures that give relevant information on biological processes within individual cells.

Differentially expressed gene (DEG): a gene with significant differences in expression levels between two experimental conditions

2.3.4. Data set integration. Conceptually, scRNA-seq data sets provide statistical power due to the high number of data points, often referred to as power by numbers. As more cells are collected, the sensitivity increases to detect a wider range of cell identities, including both rare cell types and specific cell states of common cell types. Increasing the number of analyzed cells can be achieved by either increasing the number of sampled cells or by integrating multiple existing data sets. The integration of data from different experiments or batches requires additional corrections and quality controls. Technical variance associated with protocols, equipment, staff, and time of processing can interfere with the detection of the biological signal of interest and lead to false discoveries when combining samples (35). Integration methods correct for batch effects between cells from different samples by transforming the data to recreate cell identities in low-dimensional space that are no longer dominated by batch effects. Available methods can be differentiated based on whether they require the exact same amount of cell types present across all batches [e.g., MetaSparseKmeans (36)] or whether they can process cell identities or states that are not present across all samples [e.g., mutual nearest neighbor (MNN) (32)]. Shahan and collaborators (72), for example, used the canonical correlation analysis (CCA) that is implemented in the Seurat toolkit (13) to integrate in-house-generated and publicly available scRNA-seq data sets from *Arabidopsis* roots. Wang and collaborators (87) used MNN to correct for batch effects between rice samples that were exposed to different environmental conditions enabling the comparison of cell type assignment independent of the environmental conditions.

Given that batch effects can only be accounted and corrected for in the analysis pipeline if they are controlled for through an appropriate experimental design, an upcoming challenge for the plant community will be to join forces and align protocols and experimental designs to ensure that all generated data can be incorporated into a comprehensive plant cell atlas (for more about the Plant Cell Atlas initiative, see 65).

2.4. Beyond Cell Clustering: Identification of Single-Cell Transcriptome Signatures

The initial outputs of a scRNA-seq data set are differentially expressed genes (DEGs) between cell types or subsets of cells. This already provides a high spatial resolution for the involvement of unknown regulators and differential responses of specific tissue types to a general stimulus such as abiotic stresses or treatment with phytohormones. However, scRNA-seq data also allow characterizing cellular state transitions and developmental trajectories due to the high temporal resolution. Additionally, the combination of spatiotemporal resolution and high cell numbers provides opportunities to enhance GRN analysis.

2.4.1. Trajectory inference methods. The dynamic processes that occur within cells can be uncovered using trajectory inference methods. A trajectory analysis can highlight transitions between different cell states or even identify branching points in the developmental trajectory of a cell-type lineage. Pseudotime analysis can be used to align cells according to their incremental transcriptional changes within a temporal order along a trajectory that corresponds to a biological process. Running different algorithms in an iterative way and testing different cells as starting points can provide some robustness in this prediction.

The resulting trajectories can be simple (e.g., cyclic or linear) or bifurcated or can even contain disconnected events (69). However, in all cases, validation is required to ensure the predicted trajectories match detectable biological events. Genes specifically expressed at the branching points or transitions between cell states can contribute to understanding these cell lineage–specific processes. The *Arabidopsis* root tip is uniquely fit to trace developmental trajectories, as cells are orderly aligned in cell files and cell lineages are well defined. Monocle is an unsupervised algorithm that generates a pseudotemporal resolution of transcriptome dynamics (82) and has already been used intensively to define developmental trajectories in *Arabidopsis* (19, 21, 28, 67, 68, 73, 87, 94), tomato (81), rice (47), and maize (71). Other tools, including CytoTRACE (30, 72), Slingshot (62, 78, 88), GNG (69, 88), and scVelo (6, 49, 72), have also been used for the same application. By focusing on the trajectories of cortex and endodermis cells, which both originate from the same stem cell, potential early regulators in ground tissue development could be identified (72).

Trajectory inference: the process of ordering cells along pseudotime to identify cellular transitions and developmental trajectories

Pseudotime: inference of time or progression through, for example, development

2.4.2. Gene regulatory network analysis. Single-cell data sets offer unprecedented temporal and spatial resolution to reconstruct GRNs with transcriptional regulators that control cell-type-specific processes and cell stage transitions. Although harnessing the full spatiotemporal resolution at the cellular level of scRNA-seq data is still challenging, constructing GRNs between groups of cells along the developmental trajectory is already feasible (19, 67). For example, analyzing the dynamics of transcription factor expression during trichoblast cell differentiation in *Arabidopsis* revealed novel players and currently undescribed feedback mechanisms (19). Expanding these comparisons to include transcription factors and their respective target genes can shed further light on how cell-to-cell communication orchestrates cell proliferation and differentiation.

3. EMERGING OPPORTUNITIES FOR SINGLE-CELL TECHNOLOGY IN PLANT RESEARCH

3.1. Digging Deeper and Faster

Plant research has embraced high-throughput single-cell transcriptomic approaches in recent years, and this has allowed the study of developmental processes and cellular responses at unprecedented resolutions (7, 18, 19, 21, 26, 28, 39, 41, 42, 47–49, 54, 60, 62, 67, 68, 71–73, 75, 81, 83, 87, 88, 91, 94). However, we are only observing the tip of the proverbial iceberg, as we will not only see the amount and type of information increase exponentially over the next few years, but also soon be able to extract much more detailed information from existing data.

A prime example is the identification of specific or novel subpopulations of cells. Current resolution of the *Arabidopsis* root, for example, allows precise identification of developmental trajectories of general cell identities (e.g., vascular tissues) and some subdivision into cell types (e.g., xylem and phloem) (19, 26, 28, 39, 67, 72, 83, 88, 94). Using recently defined marker genes for such subpopulations gives sufficient information to perform a more detailed classification of subpopulations (e.g., into proto- and metaxylem, sieve element, and companion cells) and even trajectory bifurcation analysis to validate cell lineage tracing studies (67, 88). Moving from general profiling

of all tissue types to the generation of more dedicated data sets comprising one specific tissue type (41, 49, 67, 83) will provide the required resolution and identification of validated marker genes and allow direct detection of these identities in all existing data sets. NICHE-seq (58) offers an alternative to overcome marker-based cell selection or analysis of overexpression mutants. Here, a photoactivatable fluorescent protein is constitutively expressed in all cells, and two-photon irradiation is used to locally excite it before tissue dissociation. Cells in which the fluorescent protein is stimulated are separated from nonstimulated cells using FACS and sequenced afterwards. The ability to stimulate cells of interest within a tissue niche has the potential to capture cell states and to recreate spatial information in any tissue of interest.

Besides providing temporal alignment of cell types along a developmental trajectory, trajectory inference methods can offer additional information on the stability and speed of cell state transitions (61). In the simplest example, the number of cells along a trajectory can predict whether this transition is slow (gradual change typed by a homogenous distribution from start to end) (**Figure 3*a***) or very fast (switch-like behavior characterized by few cells between the start and end) (**Figure 3*b***). This by itself provides important biological insights into the developmental process of transitions into certain cell identities. For example, the interpretation of developmental trajectories in scRNA-seq data sets of lines overexpressing xylem-specific transcription factors has revealed a bistable hysteric switch behavior of xylem vessel identity in *Arabidopsis* (83).

Obtaining more reliable developmental trajectories to pinpoint rare cells undergoing this type of switch-like behavior can be assisted by scaling up the number of processed cells. Single-cell combinatorial barcoding protocols, such as split-pool ligation-based transcriptome (SPLiT)-seq (66) or single-cell combinatorial indexing (sci)-RNA sequencing (14), allow ultrahigh-throughput processing of up to millions of cells using multiple rounds of random pooling and barcoding. Analysis of such high cell numbers, although they come at the expense of data depth per cell, allows enrichment and identification of very rare or transient subpopulations. This could become very relevant, as groups of cells within a cell type might respond differently to an endogenous or environmental stimulus. Expression mosaicism of transgenes in plant calli, in which transgene expression is only observed in a fraction of cells of the same identity, is an ideal example for heterogeneity within cell types (70). Identifying and studying such exceptional responder cells in plants (**Figure 3*c***) could hypothetically reveal, for example, gene regulatory elements that cause their resistance towards biotic and abiotic stimuli. These rare cell populations and transient cell states might challenge current definitions of cell identity but, at the same time, offer exciting new opportunities for biotechnological applications.

Although single-cell transcriptomic approaches are well established in the plant field, cell-specific responses and consequent phenotypic changes are the result of the combined effect of the transcriptome, epigenome, metabolome, and proteome. The coming years will be characterized by initiatives to start single-cell multiomics analyses (**Figure 3*d***). Upcoming new methods can capture transcriptome and epigenome information from individual cells simultaneously, such as assay for single-cell transcriptome and accessibility regions (ASTAR-seq) (90) or simultaneous high-throughput ATAC and RNA expression with sequencing (SHARE-seq) (52), which can, for example, provide unique opportunities to understand acquired disease resistance or responses to other internal and external stimuli at the cellular level.

3.2. Single-Cell Transcriptomics for All

Although the potential of single-cell transcriptomics has been illustrated in the model plant *Arabidopsis* (18, 19, 21, 26, 28, 39, 41, 47–49, 67, 68, 72, 73, 75, 83, 88, 94), the moss *P. patens* (42), and economically relevant crop species such as maize (7, 54, 60, 71, 91), tomato (62, 81), and rice (47,

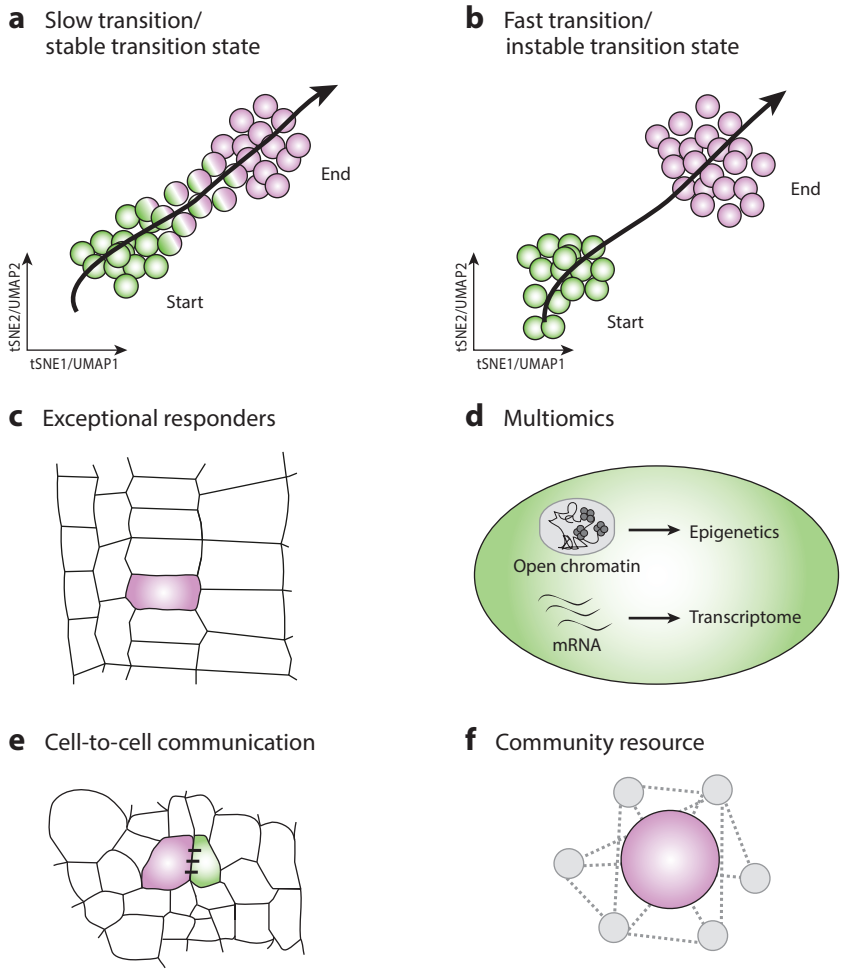

Figure 3

Application of single-cell technologies in plant cell research. (*a–b*) Enrichment of subpopulations can be used to identify and characterize changes along a developmental trajectory. Slow transitions between the start and end of a trajectory suggest gradual changes in the transcriptome, while fast transitions, without intermediate stages, can suggest a switch-like behavior. (*c*) Ultra high-throughput analysis of ten-thousand to hundreds of thousands of single cells could reveal heterogeneity within cells of the same cell type. The transcriptomes of such exceptional responders carry useful information to understand phenotypic plasticity. (*d*) Multiomics single-cell approaches can be used to correlate cell-specific transcriptome profiles with gene regulatory elements or other cellular information (metabolome, proteome, etc.). (*e*) Combining single-cell transcriptomic profiles with spatial information can reveal cell-to-cell communication signals as seen in ligand-receptor-mediated pathways. (*f*) Data depository and integration initiatives, like the Plant Cell Atlas (65), aim to unify experimental conditions and sample processing to allow a standardized analysis and integration of scRNA-seq data sets as valuable community resources. Abbreviations: mRNA, messenger RNA; tSNE, t-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection.

87), there are clear challenges to implementing the technology in genetically poorly characterized and emerging model plant species. In many cases, a set of well-defined marker genes to annotate cell clusters is simply not available, and generating fluorescent reporter lines for dozens of cluster-specific genes might be technically challenging due to long generation times or low-efficiency in plant transformation. Spatial transcriptomics approaches offer spatially resolved single-cell transcriptomes within whole tissue sections, which can be used to circumvent the need to generate single cells or have sets of tissue-specific marker genes to annotate cell identities (51, 59) (**Figure 3*e***). Furthermore, keeping cells within the tissue can help to maintain spatial responses, including, for example, transient transcriptional responses between cells induced by internal [e.g., ligands (25) (**Figure 3*e***)] or external cues such as chemicals, fertilizers, or pathogens.

Given that plant cells are differentially but highly responsive to their environment and this changes over development, the use of scRNA-seq in high-resolution phenotyping and breeding applications (including yield, abiotic/biotic stress resistance, and plant immunity) has immense potential, but requires processing of many samples. Sample multiplexing can be useful to test the variety of individual cell sensitivities and responses during chemical screens or chemical concentration tests. Multiplexing using lipid-tagged indices for single-cell and single-nucleus RNA-seq (MULTI-Seq) uses lipid- or cholesterol-conjugated oligonucleotides that bind to plasma membranes to provide cells from each sample with an additional barcode (56). Other techniques use oligo-tagged antibodies to label cells from the same sample before pooling samples together (76, 77). Pooling samples prior to library preparation and sequencing lowers costs by an order of magnitude. The associated loss of sequence coverage might not be a problem when studying specific processes or for breeding purposes, as it can be sufficient to evaluate a subset of genes by enriching for a specific panel of gene expression using, for example, Constellation DropSeq (C-DropSeq) (84). Although targeted gene panels reduce the overall information obtained from a single-cell transcriptome, they offer a higher sensitivity.

Many of the key cellular responses to environmental stimuli occur in differentiated cells, which are difficult to liberate using enzymatic digestion. Additionally, some plant species might be specifically recalcitrant to tissue dissociation. In these cases, profiling nuclei in single-nucleus RNA-seq (snRNA-seq) could offer a solution (26, 54, 81). Nuclei enrichment has been used widely in plants to obtain cell-type-specific transcriptomes (for a review, see 3). Fluorescence-activated nuclei sorting, for example, was used to capture transcriptomes from rare cells, such as phloem companion cells (93) or embryonic cells (74). Recently, *Arabidopsis* root nuclei were used to profile the open chromatin regions using snATAC-seq (21, 26). In theory, the open chromatin state of a cell could act as a readout for transcriptional activity (63). Indeed, snRNA-seq and snATAC-seq results correlated for most cell types, but snRNA-seq identified a novel cell cluster that was not captured in scRNA-seq data sets (26). Although the identity of this cluster has yet to be assigned, it suggests that these nuclei might come from cells that are not captured via enzymatic digestion (26).

3.3. Standardization and Community Building

scRNA-seq data sets contain a wealth of information that is of use to a broad range of plant scientists. However, to ensure that these large, high-dimensional data sets can be properly mined and used by the community, we need to implement standardized procedures and protocols for sample preparation, quality control, data analysis, and validation. Establishing detailed experimental design guidelines could provide a standard for the amount of technical and biological replicates required per experiment and suggest means to properly validate the data via appropriate controls. In the case of previously uncharacterized tissue types or plant species, bulk RNA-seq data of undigested tissues can be profiled to provide useful controls to separate gene expression changes that

are triggered during tissue dissociation (7, 19, 21, 26, 41, 47, 67, 68, 73, 81, 87, 91). Additionally, a quantitative indication of tissue digestion efficiency across cell types and developmental stages would also be beneficial during analysis to determine potential biases in the profiled cell population (39, 72, 73, 83, 87, 88). A set of experimental and analysis guidelines could be formulated based on dedicated benchmark studies, which are not available in the plant field so far.

To allow comparisons between the quickly increasing number of experiments performed by many different research groups across the field, guidelines for batch effect correction will become very important. Moreover, when comparing different tissues and/or different species, unsupervised multiple data set integration algorithms will need to be developed to compensate for the complete absence or variable relative composition of individual cell types across samples (13). In such cases, gene panels related to biological processes or functions could be used in machine learning algorithms to automatically define cell identities across tissues and species (**Figure 3*f***). A recent machine learning approach identified several novel marker genes for *Arabidopsis* root cell types (92), showing its potential in capturing detailed clustering signatures. This will, however, require well-defined and reliable Gene Ontology (GO)-term annotations in both model and non-model plant species. Computational simulations indicate that heterogeneity in the cell cycle can even override cell identity transcriptomes (4). To circumvent this problem, removal of variation contributed by the cell cycle has lately also been implemented in plant single-cell transcriptome data sets (49, 71). Machine learning approaches can be used to give an unbiased estimation of parameters that trigger variation among cells and data sets (12). These and other initiatives will allow us to build integrated virtual plant databases, such as the Plant Cell Atlas (65).

SUMMARY POINTS

1. Single-cell approaches are established in the plant field, and the portfolio of technologies is rapidly expanding.
2. Efficient cell isolation is a bottleneck for some tissues and plant species.
3. Single-cell transcriptomics offers solutions to solve longstanding questions in fundamental and applied plant research.
4. Unified experimental design and analysis methods are crucial to allow efficient data integration and comparison.
5. Initiatives such as the Plant Cell Atlas will assist in uniting data sets and research tools for the entire community.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

AUTHOR CONTRIBUTIONS

C.S., J.R., J.R.W., B.B., Y.H., K.D.B., and B.D.R. contributed to Section 1; C.S., J.R., J.R.W., T.E., R.S., N.V., B.B., Y.S., Y.H., K.D.B., and B.D.R. contributed to Section 2; and C.S., J.R., J.R.W., K.D.B., and B.D.R. contributed to Section 3. C.S., J.R., and B.D.R. generated the figures.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Camilla Ferrari and Max Minne (VIB Center for Plant Systems Biology) for critical reading of the manuscript. This work was supported by funding from the European Research Council (ERC Starting Grant TORPEDO; 714055) and the Research Foundation—Flanders (FWO; Odysseus II G0D0515N) to B.D.R.

LITERATURE CITED

1. Abdelaal T, Michielsen L, Cats D, Hoogduin D, Mei H, et al. 2019. A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol*. 20(1):194
2. Arabidopsis Genome Initiat. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6814):796–815
3. Bailey-Serres J. 2013. Microgenomics: genome-scale, cell-specific monitoring of multiple gene regulation tiers. *Annu. Rev. Plant Biol.* 64:293–325
4. Barron M, Li J. 2016. Identifying and removing the cell-cycle effect from single-cell RNA-Sequencing data. *Sci. Rep.* 6(1):33892
5. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, et al. 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37(1):38–44
6. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* 38:1408–14
7. Bezrutczyk M, Zöllner NR, Kruse CPS, Hartwig T, Köhrer K, et al. 2020. Phloem loading via the abaxial bundle sheath cells in maize leaves. bioRxiv 2020.09.06.284943. **https://doi.org/10.1101/2020.09.06.284943**
8. Bhosale R, Boudolf V, Cuevas F, Lu R, Eekhout T, et al. 2018. A spatiotemporal DNA endoploidy map of the Arabidopsis root reveals roles for the endocycle in root development and stress adaptation. *Plant Cell* 30(10):2330–51
9. Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, et al. 2003. A gene expression map of the *Arabidopsis* root. *Science* 302(5652):1956–60
10. Brady SM, Orlando DA, Lee J-Y, Wang JY, Koch J, et al. 2007. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318(5851):801–6
11. Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, et al. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10(11):1093–95
12. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, et al. 2015. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33(2):155–60
13. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36(5):411–20
14. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357(6352):661–67
15. Chen H, Wang JP, Liu H, Li H, Lin Y-CJ, et al. 2019. Hierarchical transcription factor and chromatin binding network for wood formation in *Populus trichocarpa*. *Plant Cell* 31(3):602–26
16. Chen H-IH, Jin Y, Huang Y, Chen Y. 2016. Detection of high variability in gene expression from single-cell RNA-seq profiling. *BMC Genom*. 17(S7):508
17. Chen S, Lake BB, Zhang K. 2019. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* 37(12):1452–57
18. Coate JE, Farmer AD, Schiefelbein JW, Doyle JJ. 2020. Expression partitioning of duplicate genes at single cell resolution in *Arabidopsis* roots. *Front. Genet.* 11:596150
19. Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MCP. 2019. Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* 48(6):840–852.e5
20. Dinneny JR, Long TA, Wang JY, Jung JW, Mace D, et al. 2008. Cell identity mediates the response of *Arabidopsis* roots to abiotic stress. *Science* 320(5878):942–45

21. Dorrity MW, Alexandre C, Hamm M, Vigil A-L, Fields S, et al. 2020. The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. bioRxiv 2020.07.17.204792. **https://doi.org/10.1101/2020.07.17.204792**
22. Duò A, Robinson MD, Soneson C. 2018. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Research* 7:1141
23. Efroni I, Ip P-L, Nawy T, Mello A, Birnbaum KD. 2015. Quantification of cell identity from single-cell gene expression profiles. *Genome Biol.* 16(1):9
24. Efroni I, Mello A, Nawy T, Ip P-L, Rahni R, et al. 2016. Root regeneration triggers an embryo-like sequence guided by hormonal interactions. *Cell* 165(7):1721–33
25. Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, et al. 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 568(7751):235–39
26. Farmer A, Thibivilliers S, Ryu KH, Schiefelbein J, Libault M. 2021. Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in *Arabidopsis* roots at the single-cell level. *Mol. Plant* 14(3):372–83
27. Freytag S, Tian L, Lönnstedt I, Ng M, Bahlo M. 2018. Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data [version 2; peer review: 3 approved]. *F1000Research* 7:1297
28. Gala HP, Lanctot A, Jean-Baptiste K, Guiziou S, Chu JC, et al. 2020. A single cell view of the transcriptome during lateral root initiation in *Arabidopsis thaliana*. bioRxiv 2020.10.02.324327. **https://doi.org/10.1101/2020.10.02.324327**
29. Gifford ML, Dean A, Gutierrez RA, Coruzzi GM, Birnbaum KD. 2008. Cell-specific nitrogen responses mediate developmental plasticity. *PNAS* 105(2):803–8
30. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, et al. 2020. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 367(6476):405–11
31. Hagemann-Jensen M, Ziegenhain C, Chen P, Ramsköld D, Hendriks G-J, et al. 2020. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* 38(6):708–14
32. Haghverdi L, Buettner F, Theis FJ. 2015. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* 31(18):2989–98
33. Han Y, Gu Y, Zhang AC, Lo Y-H. 2016. Review: imaging technologies for flow cytometry. *Lab Chip* 16(24):4639–47
34. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, et al. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* 17:77
35. Hicks SC, Townes FW, Teng M, Irizarry RA. 2018. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* 19(4):562–78
36. Huo Z, Ding Y, Liu S, Oesterreich S, Tseng G. 2016. Meta-analytic framework for sparse *K*-means to identify disease subtypes in multiple transcriptomic studies. *J. Am. Stat. Assoc.* 111(513):27–42
37. Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, et al. 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 21(7):1160–67
38. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, et al. 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11(2):163–66
39. Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, et al. 2019. Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* 31(5):993–1011
40. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, et al. 2017. Improved maize reference genome with single-molecule technologies. *Nature* 546(7659):524–27
41. Kim J-Y, Symeonidi E, Pang TY, Denyer T, Weidauer D, et al. 2021. Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* koaa060. **https://doi.org/10.1093/plcell/koaa060**
42. Kubo M, Nishiyama T, Tamada Y, Sano R, Ishikawa M, et al. 2019. Single-cell transcriptome analysis of Physcomitrella leaf cells during reprogramming using microcapillary manipulation. *Nucleic Acids Res.* 47(9):4539–53
43. Kurimoto K, Yabuta Y, Ohinata Y, Saitou M. 2007. Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat. Protoc.* 2(3):739–52
44. Li S, Yamada M, Han X, Ohler U, Benfey PN. 2016. High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. *Dev. Cell* 39(4):508–22

45. Lieckfeldt E, Simon-Rosin U, Kose F, Zoeller D, Schliep M, Fisahn J. 2008. Gene expression profiling of single epidermal, basal and trichome cells of *Arabidopsis thaliana*. *J. Plant Physiol.* 165(14):1530–44
46. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133(3):523–36
47. Liu Q, Liang Z, Feng D, Jiang S, Wang Y, et al. 2021. Transcriptional landscape of rice roots at the single-cell resolution. *Mol. Plant* 14:384–94
48. Liu Z, Zhou Y, Guo J, Li J, Tian Z, et al. 2020. Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA Sequencing. *Mol. Plant* 13(8):1178–93
49. Lopez-Anido CB, Vatén A, Smoot NK, Sharma N, Guo V, et al. 2020. Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. bioRxiv. 2020.09.08.288498. **https://doi.org/10.1101/2020.09.08.288498**
50. Lu C, Koroleva OA, Farrar JF, Gallagher J, Pollock CJ, Tomos AD. 2002. Rubisco small subunit, chlorophyll *a/b*-binding protein and sucrose:fructan-6-fructosyl transferase gene expression and sugar status in single barley leaf cells in situ. Cell type specificity and induction by light. *Plant Physiol.* 130(3):1335–48
51. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. 2014. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* 11(4):360–61
52. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, et al. 2020. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* 183(4):1103–1116.e20
53. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161(5):1202–14
54. Marand AP, Chen Z, Gallavotti A, Schmitz RJ. 2020. A *cis*-regulatory atlas in maize at single-cell resolution. bioRxiv 2020.09.27.315499. **https://doi.org/10.1101/2020.09.27.315499**
55. Martin LBB, Nicolas P, Matas AJ, Shinozaki Y, Catalá C, Rose JKC. 2016. Laser microdissection of tomato fruit cell and tissue types for transcriptome profiling. *Nat. Protoc.* 11(12):2376–88
56. McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, et al. 2019. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* 16(7):619–26
57. McInnes L, Healy J, Melville J. 2018. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. arXiv:1802.03426 [stat.ML]
58. Medaglia C, Giladi A, Stoler-Barak L, De Giovanni M, Salame TM, et al. 2017. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* 358(6370):1622–26
59. Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. 2016. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *PNAS* 113(39):11046–51
60. Nelms B, Walbot V. 2019. Defining the developmental program leading to meiosis in maize. *Science* 364(6435):52–56
61. Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, et al. 2016. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* 537(7622):698–702
62. Omary M, Gil-Yarom N, Yahav C, Steiner E, Efroni I. 2020. A conserved superlocus regulates above- and belowground root initiation. bioRxiv 2020.11.11.377937. **https://doi.org/10.1101/2020.11.11.377937**
63. Preissl S, Fang R, Huang H, Zhao Y, Raviram R, et al. 2018. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* 21(3):432–39
64. Rensing SA, Goffinet B, Meyberg R, Wu SZ, Bezanilla M. 2020. The moss *Physcomitrium* (*Physcomitrella*) *patens*: a model organism for non-seed plants. *Plant Cell* 32(5):1361–76
65. Rhee SY, Birnbaum KD, Ehrhardt DW. 2019. Towards building a Plant Cell Atlas. *Trends Plant Sci.* 24(4):303–10
66. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, et al. 2018. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 360(6385):176–82
67. Roszak P, Heo J-O, Blob B, Toyokura K, de Luis Balaguer MA, et al. 2021. Analysis of phloem trajectory links tissue maturation to cell specialization. bioRxiv 2021.01.18.427084. **https://doi.org/10.1101/2021.01.18.427084**
68. Ryu KH, Huang L, Kang HM, Schiefelbein J. 2019. Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* 179(4):1444–56

69. Saelens W, Cannoodt R, Todorov H, Saeys Y. 2019. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* 37(5):547–54
70. Saika H, Nonaka S, Osakabe K, Toki S. 2012. Sequential monitoring of transgene expression following agrobacterium-mediated transformation of rice. *Plant Cell Physiol.* 53(11):1974–83
71. Satterlee JW, Strable J, Scanlon MJ. 2020. Plant stem-cell organization and differentiation at single-cell resolution. *PNAS* 117(52):33689–99
72. Shahan R, Hsu C-W, Nolan TM, Cole BJ, Taylor IW, et al. 2020. A single cell *Arabidopsis* root atlas reveals developmental trajectories in wild type and cell identity mutants. bioRxiv 2020.06.29.178863. **https://doi.org/10.1101/2020.06.29.178863**
73. Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, et al. 2019. High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* 27(7):2241–47.e4
74. Slane D, Kong J, Berendzen KW, Kilian J, Henschen A, et al. 2014. Cell type-specific transcriptome analysis in the early *Arabidopsis thaliana* embryo. *Development* 141(24):4831–40
75. Song Q, Ando A, Jiang N, Ikeda Y, Chen ZJ. 2020. Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in *Arabidopsis* female gametophytes. *Genome Biol.* 21:178
76. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, et al. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14(9):865–68
77. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, et al. 2018. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* 19:224
78. Street K, Risso D, Fletcher RB, Das D, Ngai J, et al. 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom.* 19:477
79. Streets AM, Zhang X, Cao C, Pang Y, Wu X, et al. 2014. Microfluidic single-cell whole-transcriptome sequencing. *PNAS* 111(19):7048–53
80. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6(5):377–82
81. Tian C, Du Q, Xu M, Du F, Jiao Y. 2020. Single-nucleus RNA-seq resolves spatiotemporal developmental trajectories in the tomato shoot apex. bioRxiv 2020.09.20.305029. **https://doi.org/10.1101/2020.09.20.305029**
82. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, et al. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32(4):381–86
83. Turco GM, Rodriguez-Medina J, Siebert S, Han D, Valderrama-Gómez MÁ, et al. 2019. Molecular mechanisms driving switch behavior in xylem cell differentiation. *Cell Rep.* 28(2):342–51.e4
84. Vallejo AF, Davies J, Grover A, Tsai C-H, Jepras R, et al. 2019. Resolving cellular systems by ultra-sensitive and economical single-cell transcriptome filtering. bioRxiv 800631. **https://doi.org/10.1101/800631**
85. Vallejos CA, Risso D, Scialdone A, Dudoit S, Marioni JC. 2017. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat. Methods* 14(6):565–71
86. van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9:2579–605
87. Wang Y, Huan Q, Chu X, Li K, Qian W. 2020. Single-cell transcriptome analyses recapitulate the cellular and developmental responses to abiotic stresses in rice. bioRxiv 2020.01.30.926329. **https://doi.org/10.1101/2020.01.30.926329**
88. Wendrich JR, Yang B, Vandamme N, Verstaen K, Smet W, et al. 2020. Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* 370(6518):eaay4970
89. White JA, Todd J, Newman T, Focks N, Girke T, et al. 2000. A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol.* 124(4):1582–94
90. Xing QR, El Farran CA, Zeng YY, Yi Y, Warrier T, et al. 2020. Parallel bimodal single-cell sequencing of transcriptome and chromatin accessibility. *Genome Res.* 30(7):1027–39
91. Xu X, Crow M, Rice BR, Li F, Harris B, et al. 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell* 56:557–68.e6
92. Yan H, Song Q, Lee J, Schiefelbein J, Li S. 2020. Identification of cell-type marker genes from plant single-cell RNA-seq data using machine learning. bioRxiv 2020.11.22.393165. **https://doi.org/10.1101/2020.11.22.393165**

93. Zhang C, Barthelson RA, Lambert GM, Galbraith DW. 2008. Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. *Plant Physiol.* 147(1):30–40
94. Zhang T-Q, Xu Z-G, Shang G-D, Wang J-W. 2019. A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant.* 12(5):648–60
95. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049
96. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, et al. 2017. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65(4):631–643.e4