

Advances and opportunities of single-cell transcriptomics for plant research

Authors

Carolin Seyfferth^{1,2,*}, Jim Renema^{1,2,*}, Jos R. Wendrich^{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,#}

Affiliations

¹ Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium

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

³ Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium

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

⁵ The Sainsbury Laboratory, University of Cambridge, Cambridge, UK.

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

⁷ Center for Genomics and Systems Biology, Department of Biology, New York University, New York

* These authors contributed equally

Correspondence to: Bert De Rybel (bert.derybel@psb.vib-ugent.be) and Kenneth D. Birnbaum (ken.birnbaum@nyu.edu)

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 single 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 on how single cell technologies will shape and accelerate plant-specific research in the near future.

Summary points

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

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, behavior towards other cells, or a 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 read-out 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 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**).

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 (69) and later the full transcriptome via **bulk RNA-sequencing** (RNA-seq) technologies (37) (**Figure 1**). Although the sample itself usually was a convoluted mixture of different cell types, tissues, or entire organisms; RNA-seq allowed to reveal 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 to deconvolute the complexity of the *Arabidopsis* root into radial cell identities and longitudinal zones (6). This resulted in high resolution 3D gene expression maps of an organ at whole genome resolution (6, 7, 15, 23, 35). Together with subsequent studies mapping cellular responses to different environmental factors such as nitrogen (23), salt and iron (15), these data sets became heavily used resources to trace spatiotemporal gene expression profiles in roots. The requirement for *a priori* knowledge to construct tissue specific reporter genes has however hampered the use of this technology beyond well-characterized model species.

The step from obtaining gene expression values of pooled sorted cells to actual sequencing of single cells required retrieving reliable gene expression values from the very low amounts of

RNA present in a single cell (8). Performing an exponential amplification of mRNA from a single cell (34, 64) enabled the generation of 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 (19). Although limited in practice to a few hundreds of 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* (63) allowed to increase 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 the assignment of sequencing reads obtained from a pool of single cells back to its original cell (30). This barcoding principle was also implemented to tag each individual transcript within a cell in the form of *Unique Molecular Identifiers* (UMIs) in order to correct for Polymerase Chain Reaction (PCR)-induced replication biases (31). The adoption of droplet-based methodologies into accessible commercial platforms fully embedded this technology in modern plant research (14, 16, 72, 21, 32, 38, 55, 58, 59, 66, 68). 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 *pseudo-time* axis allowing the reconstruction of *cell trajectories* within multiple cell types (14, 32, 33, 38, 49, 55, 58, 59, 66) and *Gene Regulatory Networks* (GRNs) that control them (14). These networks will be further refined by the addition of *single nuclei Assay for Transposase-Accessible Chromatin* (snATAC)-Seq data (16, 21), allowing to also explore 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 and maize (49, 68) 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 very high pace, in this review we hope to provide the reader with an accessible guideline encompassing all steps from sample preparation to data analysis. We furthermore aim to illustrate how single cell technologies will shape and accelerate plant-specific research in the near future.

2. An overview of the scRNA-seq workflow

The generation and analysis of a single cell transcriptome data set from a plant sample requires a multifaceted workflow, comprising 1) the generation of single cells and *compartmentalization*; 2) library construction and sequencing; and 3) demultiplexing, analysis, integration, and validation of the data (**Figure 2**). This chapter summarizes the state-of-the-art for each of the steps 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 biological meaningful information from the obtained data set.

2.1. The generation of single cells in plants

Cell isolation methods

Similar to any other research field, the first step in plant-based single cell experiments is the isolation of single cells. Given plant cells are characterized by rigid cell walls, this first step can be specifically challenging depending on the tissue type or plant species. Enzymatic cell wall digestion and manual isolation remain the two principal ways to enable 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, and rice (6, 7, 66, 68, 72, 14, 18, 19, 32, 38, 49, 58, 59) (**Figure 2**). Depending on the cell wall composition, the enzymatic solution is composed of cellulases, hemicellulases and pectinases. Because of the variable composition of plant cell walls in tissues, organs, and species, not all cells are digested with the same efficiency. As just one example, differentiated cells develop thick cell walls which in some cases incorporate the very recalcitrant lignin polymer (48). As such, a careful optimization of the enzymatic solution and the conditions is advisable for every new tissue type and plant species. Validation of the obtained protoplast is required to determine the composition of retrieved cell types; this could be performed by quantitative real-time PCR analysis to determine the relative expression levels of specific reporter genes. Additionally, if fluorescent reporter lines are available, the relative abundance of protoplast expressing Fluorescent Proteins (FPs) can be determined by *in vivo* microscopic observations (33). Additional bulk RNA-seq data from undigested tissues is routinely used to filter the effect of tissue digestion on the endogenous gene expression (14, 59, 68).

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 to obtain the content of specific or rare cells from a broad spectrum of plant species and tissues (36, 39), and were recently also adopted to single cell experiments (33). This approach was used to obtain scRNA-seq data from *Physcomitrella patens* gametophytes (**Figure 1**) (33) 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 are isolated and removed using a laser. LCM has been applied in hard stems (13) and even in fleshy tomato fruit bodies (43), indicating its efficacy in a wide range of tissue types.

Although each of these manual approaches has its specific merits, a main advantage for both is that they can be applied to any tissue or plant species without the need of 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 are fixed or directly processed, the influence on gene expression is low as seen in scRNA-seq data from *P. patens* (33). These manual approaches are however laborious and come with relatively low throughput.

Compartmentalization of cells

Once a cell mixture is obtained, individual cells need to be separated and compartmentalized into tubes (33, 49), wells (19), or droplets (14, 16, 32, 55, 58, 59, 66, 68, 72) 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 sufficient attention, as the presence of multiple cells in a compartment (***multiplets***) or the inclusion of impurities such as cell debris and free mRNA will lead to data loss in downstream processing. FACS has been used to distribute cells in plate-based scRNA-seq approaches (19), 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 score multiple characteristics of a single cell (e.g. size and/or fluorescence), which can reduce impurities. In this aspect, the implementation of FACS devices equipped with image stream analysis will be of high interest (26) and can allow more precise cell quantification to reduce

the chance of multiplets formation. However, higher purity and cell enrichment comes at the cost of increased processing time and additional processing steps. The final choice on the cell isolation and compartmentalization strategy is highly dependent on the tissue, species, the usage of commercial platform, and the envisaged application. In any case, careful consideration is required before initiating a single cell experiment.

2.2 Single cell transcriptomics technologies used in plant research

Its high-throughput and low-cost cell processing allowed droplet-based single cell methods (e.g. Drop-seq (42) or the Chromium 10X platform) to dominate the plant single cell transcriptome landscape including applications in Arabidopsis roots (14, 16, 21, 32, 55, 58, 72) and cotyledons (38); and rice stem and sheaths (68) (**Figure 1**). Given the prominent role of droplet-based protocols in the plant field, a typical single cell experiment using this set-up is described below and depicted in **Figure 2**.

The general principle underlying droplet-based protocols 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, a transcript-specific barcode (as UMIs), and poly(T) sequence (**Figure 2**). The oil droplets contain all buffers and enzymes required for 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 information on genetic variation, gene expression, or relative abundance of transcript isoforms is not obtained (74). The STAMP containing droplets are combined to facilitate further amplification using PCR and library preparation as a single sample, followed by RNA-Seq. This set-up 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 one single experiment.

A different 3'-based method, called Cell Expression by Linear amplification and sequencing (CEL-seq2) (27) has been applied to precisely define cellular transition states during maize anther development (49). 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 (27).

For scRNA-seq experiments where high (or full-length) transcript coverage is of the essence, the SMART-Seq protocol has been applied in a plate-based set-up (19) (**Figure 1**). Compared to 3'-based methods, SMART-Seq protocols have been associated with a higher number of sequenced transcripts and genes detected per cell (74). Additionally, the full-length transcript coverage provides information on expression balances between transcript isoforms. 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 e.g. rare cell types such as regenerating root meristem cells in *Arabidopsis* (19).

2.3. Data analysis

The analytical workflow of single cell transcriptomes is a multistep process. Several software packages, including *Cell Ranger* (73) and *Seurat* (10), combine pipelines to perform quality control, normalization, ***dimensionality reduction*** of the raw data and visualization. In this section, we will describe the different steps during data analysis as currently performed in the plant field and highlight some of the potential options and pitfalls.

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 allows to reduce background noise and PCR amplification bias (31). The barcode and UMI sequences are removed before alignment to the reference genome/transcriptome. A quality assessment can be performed 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 with low mapping ratios to endogenous RNA while 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 low count depth, low number of genes, and high amount of counts for 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 potential multiplets. A low count depth can indicate either a cell with low quality RNA or an empty droplet. In either case, barcodes should be removed before the normalization to reduce the risk of influencing the quantification of gene expression, cell clustering, and the identification of cell **marker genes**. A subsequent normalization step is used to adjust variation between cells of the same sample that are caused by differences in cell lysis efficiency, sequencing depth, or efficiency differences during library preparations. 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 an absolute requirement. Normalization by scaling counts can be expanded with an imputation algorithm that tries to infer gene expression counts to correct for excess zero counts or drop-out. Single-cell data has a higher dropout rate than bulk RNA-seq, where zeros can reflect technical limitations instead of the true absence of expression. Once normalization is performed and gene expression has been quantified, a subset of genes with either high biological variability (**Highly Variable Genes** (HVGs) (8, 12)) or highest variation in average expression level across all cells (highest expressed genes (17)) are used to determine cell clusters.

Dimensionality reduction and clustering

Depending on the plant species, sequence coverage and depth; scRNA-seq data sets can contain the expression values of 25,000 (Arabidopsis) to 50,000 (maize) genes for each of the thousands of cells. To cope with the computational and statistical challenges that come with comparing gene expression profiles among thousands - up to millions - of cells, HVGs are identified (14, 21, 72, 32, 38, 49, 55, 58, 59, 66, 68) 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 set (14, 21, 38, 49, 58, 66). PCA determines variables that contribute to the overall variance of expression among all individual cells and compares them to estimate the variable that contributes 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) (41), to visualize cell clusters in plant scRNA-seq studies (14, 21, 38, 58, 66). t-SNE is specifically useful to highlight

differences between cell clusters, but not within cells of the same cluster. ***Uniform Manifold Approximation and Projection*** (UMAP) (45) on the other hand, allows comparison of both cell clusters and cells within these clusters based on distances between them. Given that UMAP also allows to visualize more than two dimensions and has a high run time performance (3), it is quickly becoming the preferred method in the plant field (16, 21, 32, 38, 58, 68, 72).

Several algorithms are available to identify and characterize cell populations by unsupervised clustering (17). Plant scRNA-seq studies used Louvain clustering algorithm based on shared nearest neighbor networks (14, 16, 21, 32, 38, 55, 58, 59, 66, 72), which calculates the nearest neighbors of each cell and clusters those with the highest overlap. However, comparative studies indicate very method-specific clustering outputs (22).

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 (7, 35) and by computing an index of cell identity score (18) for each cell (58, 59, 66). 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 sub-cell types.

Once clustering and annotation has been performed, a 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 e.g. metabolic pathways. Marker gene expression should not only be specific to a cell cluster but should also be expressed in most cells of that cluster. Variations of marker gene expression among cells of the same cluster can have multiple reasons, one being variations related to the developmental stage that individual cells are in. Comparing marker gene expression of genes associated with proliferation (e.g. such as cell cycle-related genes) can be used to fine-tune the composition of a cell cluster (49). Alternatively, plants have the unique characteristic of showing distinct changes in cell ploidy levels through development and sets of marker genes have been associated with this (5). These ploidy changes can be used to validate the progression of cell lineages in time (16). 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 less so for other species. Automated cluster annotation methods have facilitated cell identity annotation in 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, an extensive validation of predicted cell clusters, cell identities, or developmental stages must be performed to ensure the data can be used for downstream analysis. This can be done by validating expression patterns with transgenic reporter lines (14), *in situ* hybridization, or by analyzing genetically perturbed mutants (55, 58, 66). Once validated, scRNA-seq data sets can be mined to extract additional transcriptional signatures that give relevant information on biological processes within individual cells.

Data set integration

Conceptually, scRNA-seq data sets provide power by numbers. Increasing the number of cells analyzed can be achieved by increasing the sample size or by integrating 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 (28). 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 their requirement of having the exact same amount of cell types present across all batches (e.g. MetaSparseKmeans (29)) or according to their ability to handle cell identities or states that are not present across all samples (Mutual-Nearest-Neighbor (MNN) (25)). Wang and collaborators (68) for example 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. Shahan and collaborators (58) used a customized tool (*Cell preprOcessing PIpeline kaLlistO busTools (COPILLOT)*) to align and combine in-house generated and publicly available scRNA-seq data sets from *Arabidopsis* roots.

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 all generated data can be incorporated into a comprehensive atlas (see **plant cell atlas** (53)).

2.4. Beyond cell clustering: identification of single cell transcriptome signatures

The initial output of scRNA-seq data sets are ***Differentially Expressed Genes*** (DEGs) for e.g. specific tissue types or treatments. This already provides a high spatial resolution on 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 analyzing cellular state transitions and developmental trajectories due to the high temporal resolution. The spatiotemporal resolution and power by numbers also provides opportunities to enhance GRN analysis.

Trajectory inference methods

The dynamic processes that occur within cells can be uncovered using ***Trajectory inference*** methods. It can highlight transitions between different cell states or even identify branching points in the developmental trajectory of a cell type lineage. Pseudo-time analysis can be used to align cells according to their incremental changes in transcriptome 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 point can provide some robustness in this prediction.

The resulting trajectories can be simple (e.g. cyclic or linear), bifurcated, or even containing disconnected events (56). In all cases, however, biological validation is required to ensure the predicted trajectories match biological data. 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 increases the temporal resolution of transcriptome dynamics (65) and has already been used intensively to define developmental trajectories in the Arabidopsis root meristem cell types (14, 16, 55, 59, 68, 72). Other tools including *CytoTRACE* (24) and *scVelo* (4) 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 (58).

Gene Regulatory Network analysis

Single cell data sets offer unprecedented temporal and spatial resolution to reconstruct networks of transcriptional regulators (as GRNs) that control cell type-specific processes and cell stage transitions. Although harnessing the full spatiotemporal resolution at cellular level of scRNA-seq data is still challenging, constructing GRNs between groups of cells along the developmental trajectory is already feasible (14). For example, analyzing the dynamics of transcription factors expression during trichoblast cells differentiation in Arabidopsis revealed novel players and so far undescribed feedback mechanisms (14). Elaborating these comparisons to include transcription factors and their respective target genes can shed further light into cell-to-cell communication to orchestrate cell proliferation and differentiation.

3. Emerging opportunities for single cell technology in plant research

3.1. Digging deeper and faster

Plant research has embraced single cell approaches in the past years and this has allowed to study developmental processes and cellular responses at unprecedented resolution (14, 16, 66, 68, 72, 21, 32, 33, 38, 49, 55, 58, 59). 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 we will soon also 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 e.g. the Arabidopsis root 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) (14, 21, 32, 58, 66, 72). There should however be sufficient information in existing data sets to perform a more detailed classification of subpopulations (e.g. into proto- and metaxylem, sieve element, metaphloem, companion cells etc.) (52, 58, 66, 72) and perhaps even trajectory bifurcation analysis to validate cell lineage tracing studies. This suggests that currently a main limitation is the lack of reliable marker genes. Moving from general profiling of all tissue types into the generation of more dedicated data sets comprising

one specific tissue type (66) 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 (46) offers an alternative to overcome marker-based cell selection or analysis of overexpression mutants. Here, a FP is constitutively expressed in all cells and two-photon irradiation is used to locally excite it before tissue dissociation. Cells in which the FP is stimulated are separated from non-stimulated cells using FACS, and afterwards sequenced. The concept of stimulating 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 (50). As the simplest example, the amount of cells along a trajectory from state 1 to state 2 can predict whether this transition is slow (gradual change typed by a homogenous distribution from state 1 and state 2; **Figure 3A**) or very fast (switch-like behavior characterized by few cells between state 1 and state 2; **Figure 3B**). This by itself provides important biological insights into the **developmental process of transitions into certain cell identities**. The interpretation of developmental trajectories in scRNA-seq data sets of lines overexpressing xylem specific transcription factors has revealed a bi-stable hysteric switch behavior of xylem vessel identity in Arabidopsis (66).

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 (54) or single-cell combinatorial indexing (sci)-RNA sequencing (11), allow ultra-high-throughput processing of up to millions of cells using multiple rounds of random pooling and barcoding. Although coming at the expense of data depth per cell, analysis of such high cell numbers allows to enrich and identify 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 (57). Identifying and studying such **exceptional responder cells** in plants (**Figure 3C**) could hypothetically reveal e.g. 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 **multi-omics** analyses (**Figure 3D**). Upcoming new methods can capture transcriptome and epigenome information from individual cells simultaneously, e.g. Assay for Single-cell Transcriptome and Accessibility Regions (ASTAR-seq (70)), which can for example provide unique opportunities to understand acquired disease resistance or responses to other internal and external stimuli at cellular level.

3.2. Single cell transcriptomics for all

Although the potential of single cell transcriptomics has been illustrated in the model plant *Arabidopsis* (14, 16, 21, 32, 38, 55, 58, 59, 66, 72), the moss *P. patens* (33), and economically relevant crop species such as maize (49) and rice (68), there are clear challenges to implement 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 efficient plant transformation protocols. **Spatial transcriptomics** approaches offer spatially resolved transcriptomics without having the need to generate single cells or having sets of tissue-specific marker genes to annotate cell identities (**Figure 3E**) (40, 47). The combination of cellular resolution via single cell transcriptomics with spatially resolved transcriptomics is specifically attractive as it offers opportunities to reveal cellular communication mechanisms such as ligand-receptor pairs (20) (**Figure 3E**) and could even trace e.g. signal amplification across cell types when exposed to chemicals, fertilizers, or pathogens.

Given that plant cells are differentially but highly responsive to their environment and this is changing over time, the use of scRNA-seq in high-resolution phenotyping and breeding applications (including yield, abiotic/biotic stress resistance, 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 (44). Other

techniques use oligo-tagged antibodies to label cells from the same sample before pooling samples together (61, 62). Pooling of 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 Drop-seq (C-Drop-seq) (67). 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 (21). Nuclei enrichment has been used widely in plants to obtain cell type-specific transcriptomes (reviewed in (2)). Fluorescence-Activated Nuclei Sorting for example was used to capture transcriptomes from rare cells, such as phloem companion cells (71) or embryonic cells (60). Recently, Arabidopsis root nuclei were used to profile the open chromatin regions using snATAC-seq (16, 21). In theory, the open chromatin state of a cell could act as a read-out for transcriptional activity (51). Indeed, snRNA-seq and snATAC-seq results correlated for most cell types, but snRNA-seq identified a novel cell cluster which was not captured in scRNA-seq data sets (21). Although the identity of this cluster is yet to be assigned, it suggests that these nuclei might come from cells that are not captured via enzymatic digestion (21).

3.3. Standardization and community building

Single cell RNA-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 in the amount of technical and biological replicates required per experiment and suggest means to properly validate the data via appropriate controls. In case of previously uncharacterized tissue types or plant species, bulk RNA-seq data of undigested tissues might be included to provide useful controls to separate gene expression changes that are triggered during tissue dissociation (14, 16, 21, 55, 59, 68). Additionally, a quantitative indication of tissue digestion efficiency across cell types and developmental stages would also be beneficial during analysis (32, 58, 59, 66, 68). 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 planet, guidelines for batch effect correction will become very important. Moreover, when comparing different tissues and/or different species, unsupervised **multi-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 (10). 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 3F**). This will however require well-defined and reliable GO-term annotations in both model and non-model plant species. In addition, machine learning algorithms could also be used to reduce noise between cells of the same cell type, that can for example be triggered by cell cycle- and cell stress-related genes (9). This strategy has already been utilized to remove cell cycle related genes, which could overwrite other transcriptional responses (49). These and other initiatives will allow us to build integrated virtual plant databases, such as the **plant cell atlas** (53). This not only serves as data repository but can also act as a communications hub to find consensus on experimental design, data correction and statistical analysis.

Glossary Terms

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

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

Compartmentalization: Isolating a single cell into a compartment (droplet, tube, or well)

Cell trajectory: A biological process that causes heterogeneity among cells

Differentially Expressed Gene (DEG): A gene with significant differences in expression levels between two experimental conditions

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

Fluorescence-Activated Cell Sorting (FACS): Flow cytometry-based method to separate and sort cells based on parameters such as size or fluorescence

Gene Regulatory Network (GRN): Network of molecular regulators that is predicted to regulate each other's expression levels

Highly Variable Genes (HVG): Genes with high contribution to cell-to-cell variation within a homogeneous cell population

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

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

Multiplets: A droplet containing two or more cells

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

Pseudo-time: Inference of time or progression through, for example, development

Single nuclei Assay for Transposase-Accessible Chromatin-Seq (snATAC-seq): Retrieving information on the open chromatin status of individual nuclei

Single-cell Transcriptomes Attached to MicroParticles (STAMPs): The combination of barcoded-beads with attached mRNA transcripts from a single cell

t-distributed Stochastic Neighbor Embedding (tSNE): Dimension reduction and projection algorithm

Trajectory inference: Process of ordering cells along pseudo-time to identify cellular transitions and developmental trajectories

Uniform Manifold Approximation and Projection (UMAP): Dimension reduction and projection algorithm

Unique Molecular Identifier (UMI): Small oligonucleotide sequence unique to each transcript from an individual cell that enables quantification and correction of PCR amplification biases

Acknowledgments: We would like to acknowledge 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.

Declaration of interests: The authors declare no conflict of interest related to this work.

Figures

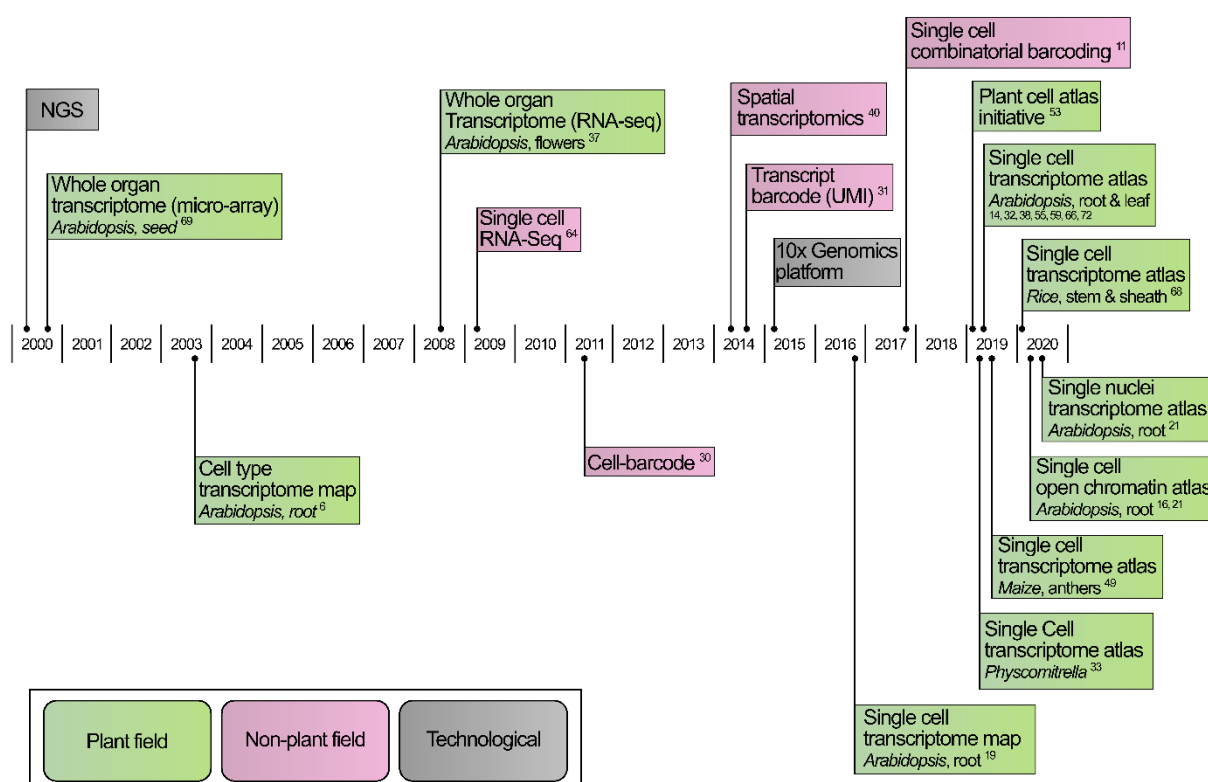


Figure 1. Timeline of key events in plant single cell transcriptomics. Selection of milestones related to the development and establishment of single cell transcriptomics in the plant field are illustrated by green boxes on the timeline. General technological developments are shown in grey and key development outside of the plant field are indicated in pink.

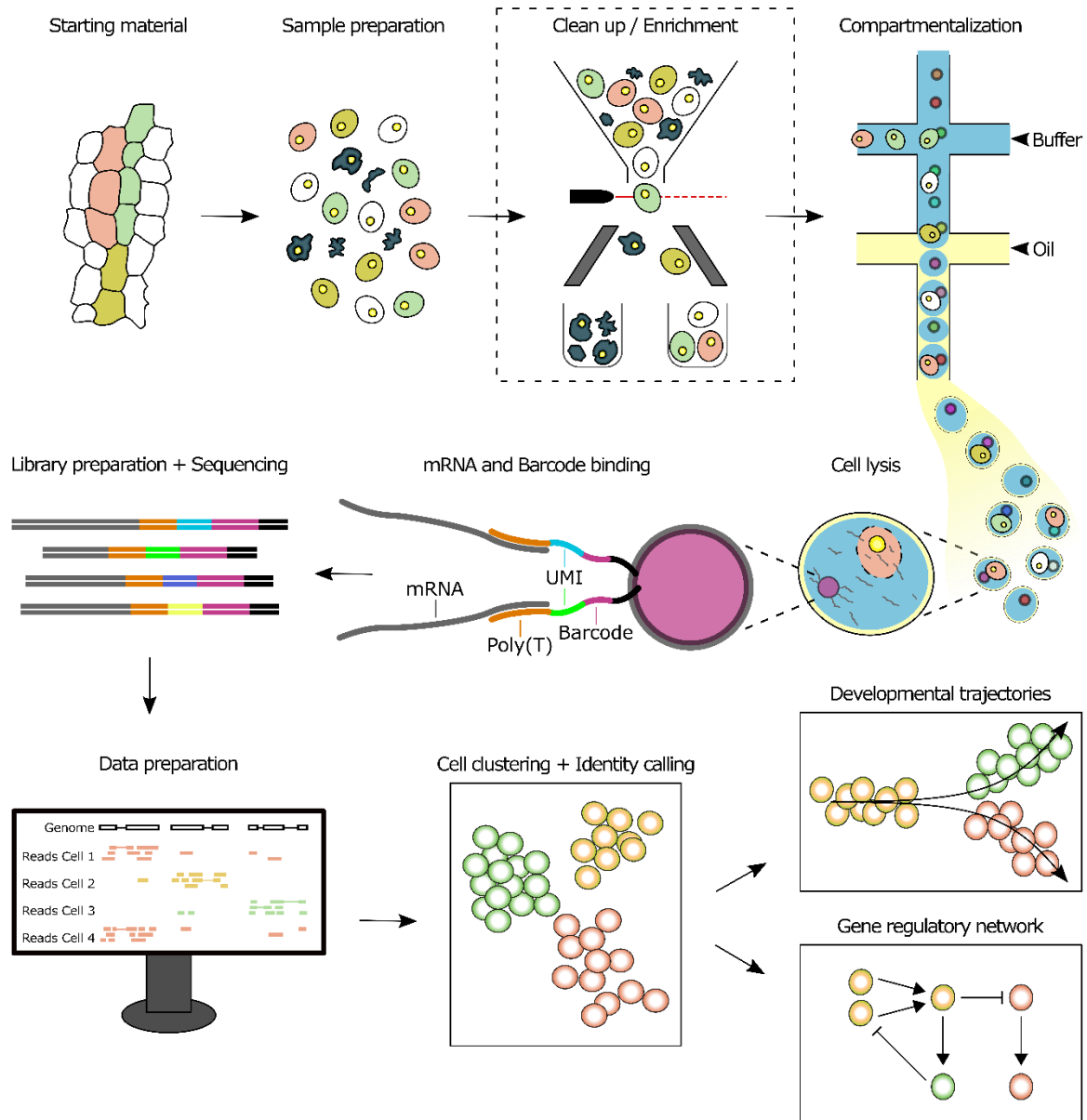


Figure 2. Schematic overview of a typical droplet-based plant scRNA-seq experiment.

Individual cells are isolated from the tissue containing different cell identities (illustrated by differently 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 oil droplets. After droplet formation, cells are lysed, and the released mRNAs are bound to the bead; followed by library preparation and next generation sequencing. The data is next 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.

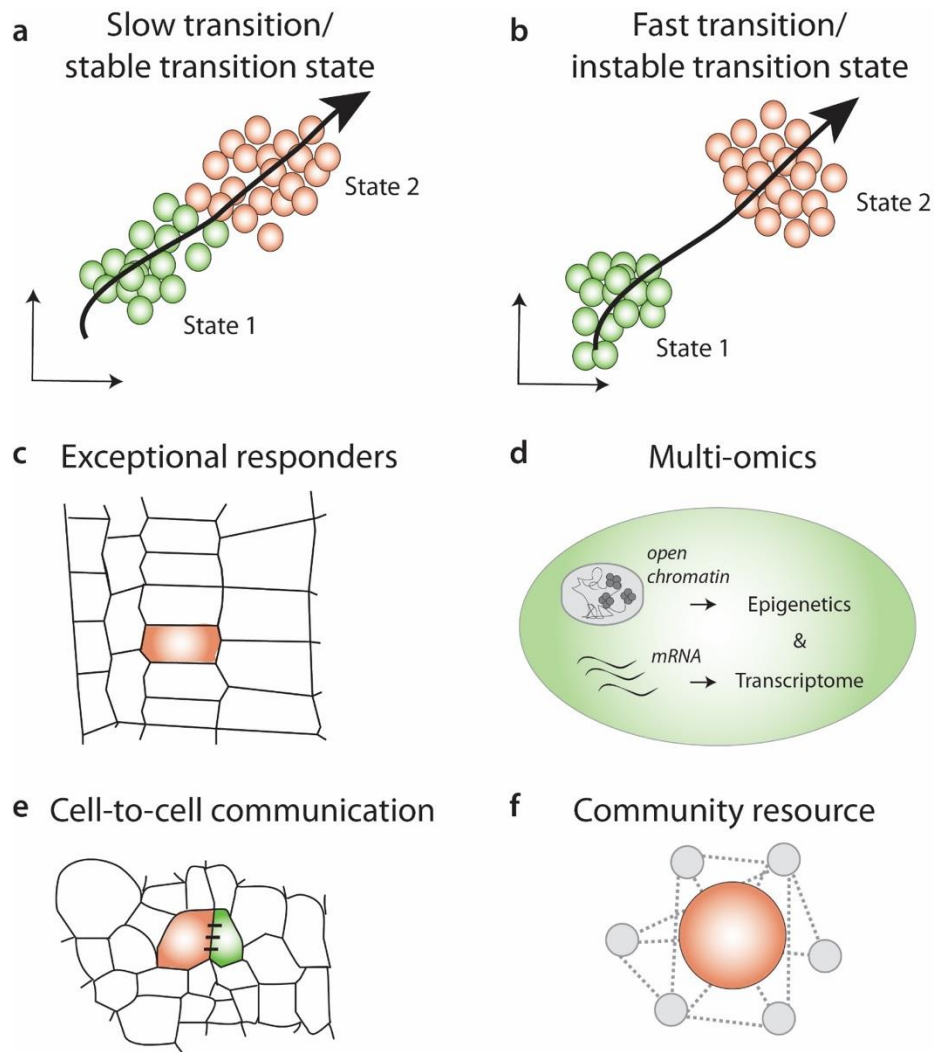


Figure 3. Application of single cell technologies in plant cell research. (a-b) Enrichment of subpopulation can be used to identify and characterize cell state changes along a developmental trajectory. Slow transitions between two states suggest gradual changes in the transcriptome, while fast cell state transitions, without intermediate stages, can suggest a switch-like behavior in the cell states. (c) Ultra high-throughput analysis of ten-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) Multi-omics 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, aim to unify experimental conditions and sample processing to allow a standardized analysis and integration of scRNA-seq data sets as valuable community resources.

References

1. Abdelaal T, Michielsens 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. Bailey-Serres J. 2013. Microgenomics: Genome-Scale, Cell-Specific Monitoring of Multiple Gene Regulation Tiers. *Annu. Rev. Plant Biol.* 64(1):293–325
3. 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
4. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.*
5. 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
6. 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
7. 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* (80-.). 318(5851):801–6
8. 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
9. 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
10. 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
11. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* (80-.). 357(6352):661–67
12. Chen H-IH, Jin Y, Huang Y, Chen Y. 2016. Detection of high variability in gene expression from single-cell RNA-seq profiling. *BMC Genomics.* 17(S7):508
13. 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*.

14. 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
15. Dinnyen JR, Long TA, Wang JY, Jung JW, Mace D, et al. 2008. Cell Identity Mediates the Response of Arabidopsis Roots to Abiotic Stress. *Science (80-.)*. 320(5878):942–45
16. 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*
17. Duò A, Robinson MD, Soneson C. 2018. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Research*. 7:1141
18. 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
19. 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
20. Eng C-HL, Lawson M, Zhu Q, Dries R, Koulina N, et al. 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*. 568(7751):235–39
21. Farmer A, Thibivilliers S, Ryu KH, John S, Marc L. 2020. The impact of chromatin remodeling on gene expression at the single cell level in Arabidopsis thaliana. *bioRxiv*
22. 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(0):1297
23. Gifford ML, Dean A, Gutierrez RA, Coruzzi GM, Birnbaum KD. 2008. Cell-specific nitrogen responses mediate developmental plasticity. *Proc. Natl. Acad. Sci*. 105(2):803–8
24. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, et al. 2020. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science (80-.)*. 367(6476):405–11
25. Haghverdi L, Büttner F, Theis FJ. 2015. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*. 31(18):2989–98
26. Han Y, Gu Y, Zhang AC, Lo Y-H. 2016. Review: imaging technologies for flow cytometry. *Lab Chip*. 16(24):4639–47
27. Hashimshony T, Senderovich N, Avital G, Klochender A, de Leeuw Y, et al. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol*. 17:77

28. 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
29. 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
30. 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
31. 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
32. 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
33. 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
34. 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
35. 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
36. 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
37. 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
38. Liu Z, Zhou Y, Guo J, Jiao Li ZT, Zhu Z, et al. 2020. Global Dynamic Molecular Profiling of Stomatal Lineage Cell Development by Single-Cell RNA Sequencing. *Mol. Plant*. 13(8):1178–93
39. Lu C, Koroleva OA, Farrar JF, Gallagher J, Pollock CJ, Tomos AD. 2002. Rubisco Small Subunit, Chlorophylla/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
40. 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
 41. Maaten L Van Der, Hinton G. 2008. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* 9(Nov):2579–2605
 42. 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
 43. 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
 44. 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
 45. McInnes L, Healy J, Melville J. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*
 46. 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 (80-.).* 358(6370):1622–26
 47. 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. *Proc. Natl. Acad. Sci. U. S. A.* 113(39):11046–51
 48. Naseer S, Lee Y, Lapierre C, Franke R, Nawrath C, Geldner N. 2012. Casparian strip diffusion barrier in Arabidopsis is made of a lignin polymer without suberin. *Proc. Natl. Acad. Sci.* 109(25):10101–6
 49. Nelms B, Walbot V. 2019. Defining the developmental program leading to meiosis in maize. *Science (80-.).* 364(6435):52–56
 50. 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
 51. 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
 52. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, et al. 2017. The Human Cell

Atlas. *Elife*. 6:e27041

53. Rhee SY, Birnbaum KD, Ehrhardt DW. 2019. Towards Building a Plant Cell Atlas. *Trends Plant Sci*. 24(4):303–10
54. 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* (80-.). 360(6385):176–82
55. 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
56. Saelens W, Cannoodt R, Todorov H, Saeys Y. 2019. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol*. 37(5):547–54
57. 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
58. 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*
59. 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):2241–2247.e4
60. Slane D, Kong J, Berendzen KW, Kilian J, Henschen A, et al. 2014. Cell type-specific transcriptome analysis in the early arabidopsis thaliana embryo. *Dev*. 141(24):4831–40
61. 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
62. 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(1):224
63. Streets AM, Zhang X, Cao C, Pang Y, Wu X, et al. 2014. Microfluidic single-cell whole-transcriptome sequencing. *Proc. Natl. Acad. Sci. U. S. A*. 111(19):7048–53
64. 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
65. 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

66. 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):342-351.e4
67. 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*
68. 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*
69. 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
70. Xing QR, Farran CA El, Zeng YY, Yi Y, Warriar T, et al. 2020. Parallel bimodal single-cell sequencing of transcriptome and chromatin accessibility. *Genome Res.* 30(7):1027–39
71. 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
72. 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
73. 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
74. 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