THE DIFFERENCE IS BREAKTHROUGH DATA EVERY STEP OF THE WAY

MOVE FROM TRANSLATIONAL RESEARCH TO CLINICAL APPLICATIONS FASTER WITH BD AS YOUR TRUSTED WORKFLOW SOLUTIONS PARTNER. BD is your partner in immuno-oncology from discovery research to clinical applications, with a range of high-performance solutions designed to give you high-quality, reliable research data for even the most complex experiments. More solutions. More answers. More data you can trust. Discover more with BD at your side. **Discover the new BD.**

Learn how you can advance your research >

BD
Advancing the world of health

Status and Potential of Single-cell Transcriptomics for Understanding Plant Development and Functional Biology

Running Title : Plant functional and development Biology in the era of single-cell Omics

Muhammad Munir Iqbal[1, 2†], Bhavna Hurgobin[3†], Andrea Lisa Holme[4], Rudi Appels[5, 6], Parwinder Kaur[1*]

[1] UWA School of Agriculture and Environment, Faculty of Science, The University of Western Australia, 35 Stirling Hwy, Perth, 6009, WA, Australia

[2] Genome Innovation Hub, Telethon Kids Institute, Perth Children Hospital, Nedlands, 6009, WA, Australia

[3] School of Life Sciences, La Trobe University, Bundoora, Victoria 3086, Australia

[4] Iain Fraser Cytometry Centre, IFCC Institute of Medical Sciences (IMS), School of Medicine, Medical Sciences and Nutrition University of Aberdeen Forester hill, Aberdeen AB25 2ZD, United Kingdom

[5]School of BioSciences, The University of Melbourne, Victoria 3010, Australia

[6] School of Applied Biology, La Trobe University, Bundoora, Victoria 3086, Australia

† Equal first authors

*Correspondence to: Assoc. Professor Parwinder Kaur,

UWA School of Agriculture and Environment, Faculty of Science

The University of Western Australia, Perth, WA 6009

Email: Parwinder.kaur@uwa.edu.au

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/cytoa.24196

Abstract

The advent of modern "omics" technologies (genomics, transcriptomics, proteomics, and metabolomics) are attributed to innovative breakthroughs in genome sequencing, bioinformatics, and analytic tools. An organism's biological structure and function is the result of the concerted action of single cells in different tissues. Single cell genomics has emerged as a ground-breaking technology that has greatly enhanced our understanding of the complexity of gene expression at a microscopic resolution and holds the potential to revolutionize the way we characterize complex cell assemblies and study their spatial organization, dynamics, clonal distribution, pathways, function, and networking.

Mammalian systems have benefitted immensely from these approaches to dissect complex systems such as cancer, immunological disorders, epigenetic controls of diseases and understanding of developmental biology. However, the applications of single-cell omics in plant research are just starting. The potential to decipher the fundamentals of developmental and functional biology of large and complex plant species at the single-cell resolution are now becoming important drivers of research. In this review, we present the status, challenges and potential of one important and most commonly used single-cell omics technique in plants, namely single cell transcriptomics.

Key terms

Single-cell omics, transcriptomics, scRNA-Seq, functional biology; development biology; plant; cytometry

Introduction

Plants are known to be structurally flexible. Plants possess small, self-replicating chloroplast DNA molecules varying in size from 120 to 220 kb with highly conserved gene content across species [1]. Plant organ development is generally attributed to specific areas of growth, commonly known as 'meristems'[2-4]. Meristems are located in the high growth activity regions of most plants and contain undifferentiated cells. These undifferentiated cells are solely responsible for development and functional trajectories [2, 3, 5]. Although cells arise from the same zygote, they proliferate and differentiate to multiple types and different organs. This cellular heterogeneity is a virtue in terms of responding to different physiological stimuli affecting gene expression profiles, functionalities and developmental patterns [6]. Traditionally, spectrophotometry, low-throughput phenotyping and bulk genotyping methodologies have played an important role in improving our understanding of the development and functional biology of key traits governed by different tissue specific cells in plants. These methodologies have helped in generating important information such as genome size estimation, level of ploidy, and nuclear replication state [7-9].

The fluorescence microscopy techniques such as fluorescence in situ hybridization (FISH) has opened the ways to study the conformation of chromosomes by which longer strands of DNA fit in a nucleus of size between 1-10 µm in individual cells [10, 11]. However, the availability of limited number of fluorescent markers available to conduct this analysis and its low resolution even under the microscopes of good resolution has limited its potential. The biochemical genomic conformation capture techniques such as 3C and Hi-C mitigated the limitations of fluorescence microscopy techniques by enabling the genome

wide profiling of chromatin interactions based on the average analysis of cell populations [10, 12].

Despite the promise of these methodologies, they lack the ability to identify the developmental trajectories and functionalities of certain 'rare' cellular types governing a plant's key developmental and functional decisions. Hence, their masked effect by major cell types can result in an ambiguous and fragmented understanding of cell fate decision [13]. Moreover, these technologies lack the ability to decipher the heterogeneity among small cell populations.[14].

Plant systems have unique cell models in different tissues and organs and these cellular models are highly specie specific too, hence different gene expression emerge from each of those models. For instance, germline cells, trichomes cells, fibre cells, root hair cells, floral cells and stomatal guard cells are all different from each other and exhibit different gene expression pattern [15]. Consequently, a bulk gene expression profile from one tissue would not provide a clear insight to the development decisions governing these and other key plant traits. This has warranted the development of new techniques which have the capability to accurately measure the genotypic and phenotypic profiles of single cells in particular tissues or organs.

In the last two decades, it has become feasible to scan plant genomes at the DNA, RNA and protein expression levels of resolution via technologies collectively referred to as multi-omics [16]. This revolution has impacted the ways in which different plant populations are mined for high throughput genome wide association studies and select haplotypes for high resolution and stable genomic regions in order to select for desirable traits for breeding future crop varieties [17]. This rapid technological advancement has also

enabled researchers to study the biological processes responsible for development and the function of single cells. These single-cell analysis methodologies have also progressed to enable us to read and measure the genetic architecture and spatial organisation at a single-cell scale. These single-cell omics (sc-Omics) techniques allow us to study the genome, transcriptome, proteome and metabolome of organisms at single-cell level using genomic readouts. The latter can be inherent segments of RNA and DNA molecules of interest or a small surrogate sequence known as 'barcode' to identify different cells. More specifically, sc-Omics allows us to measure cells at the DNA-level (scDNA-Seq) [18], RNA-level (scRNA-Seq) [19], DNA conformation level (scATAC-Seq) [20] and protein-level (CITE-Seq [21], Abseq [22]). Different single-cell analysis techniques that have evolved under the umbrella of sc-Omics [23-25] are mainly based on droplet-based infrastructure or well-based infrastructure to capture single cells into a single reaction chamber to then be individually sequenced. The former has more high-throughput cell capture capacity than the latter but yields lower sequencing depths of subject material [26, 27].

The limitations posed by fluorescence microscopy techniques such as FISH were resolved by the single-cell transcriptomics techniques such as Spatial transcriptomics [28]. Spatial transcriptomics is another allied single-cell technique to study the gene expression among different cellular populations of a tissues [28]. This technique employs the tissue sections on the arrays with known probes or barcodes thus fixing the cells with different probes in order to help differentiate them for downstream gene expression analysis. The size limit of array probe is a major drawback of this technique where a large single cell could attach to more than one probe thus making it difficult to associate a certain gene expression profile to it [29]. Other methods of single-cell capture for downstream analysis include Fluorescence-activated cell-sorting (FACS), Magnetic-activated cell sorting (MACS), Laser

capture microdissection (LCM) and manual/automated cell picking [30-33]. The major drawbacks of these techniques as they are currently used include their dependence on already known fluorescence labelled markers or barcode printing in live cells prior to sorting, surface antigens or time-consuming cell picking under microscope. Hence, the high-throughput single-cell capture by microfluidics or well-based infrastructure make them ideal techniques for single cell analysis [34]. The amplification-enabled readouts of individual cells provide extra edge to capture 'minor' cell types and analyse their genotypic profiles in greater details which can be further enhanced by combining with upstream isolation methods such as FACS. MACS, LCM. [34].

sc-Omics has opened ways to study the detailed genetic architecture of an organism at the single-cell level in detail. This information is then translated to infer an accurate understanding of genetic control of biological processes happening in an organism [35]. So far, mammalian systems have been well studied using sc-Omics approaches [36, 37], while plants lag behind mainly because the processing of tissue with cell walls has been a challenge for the application of single-cell technologies, often resulting in low capture rates [38]. A few successes have been achieved in successful protoplast isolations in plants such as moss, to load them on commercially available droplet-based platform such as 10x Genomics [39]. The relatively untapped plant systems have the potential to efficiently utilise the modern sc-Omics techniques to enhance our broader understanding of cellular-level plant development and functional trajectories [36]. Single-cell based omics, in particular single-cell transcriptomics has the capability to provide a detailed insight into cell responses to different environmental cues, such as biotic and abiotic stresses and provides the strong basis for understanding of genetic control of more complex traits [40, 41]. In this short

review, we cover the status of current use and potential of single-cell transcriptomics in regard to the development and functional biology of plants.

Single-cell transcriptomics and plant development biology

Like many other species, crops have evolved to very different states compared to their wild progenitors either through unintended selection or selected breeding to better fit in a range of environments [42]. During this process, the cell, being the basic unit of life, has also evolved to multicellular states in order to support plant development trajectories in response to different environmental stimuli [43].

Most available single-cell studies in plants have utilised scRNA-Seq to identify different cell types. As is principally correct for sc-Omics, different barcodes (both synthetic and plant genome derived) are employed to tag and track cell differentiation along the course of development in plant systems [36]. These studies provided a proof of concept to identify intermediate differentiation cell states by providing a link to track cell cycle state from meristematic phase to a fully differentiated phase based on plant development trajectories [44].

A recent single-cell transcriptomic study by Shulse et al. [45] involving ~4,000 root xylem cells identified several novel candidate marker genes which have no or very little information linked to their functional activity, highlighting the power of this technology. This study also revealed that genes with already well-established roles were a small subset of overall number of genes found while many new marker genes with unknown functionality were discovered. These findings not only challenged the previously thought genetic control of cellular pathways but also opened up the avenues to study the transcriptional role of newly found genes in root cells differentiation. The droplet-based cell profiling approach

used in this study was further extended to >12,000 root cells, which revealed the representation of all root tissues, their distinct developmental stages and specific marker genes associated with those cell populations [45].

In another study by Turco et al. [46], a Drop-seq based isolation of single-cells from a bulk of ten plates with 200 roots per plate revealed a novel regulatory network of four targets controlled by a transcription factor called *VASCULAR-RELATED NAC DOMAIN7 (VND7)*, which has a well-established role in terminal differentiation of xylem cells. This study further revealed that two other xylem cell differentiation transcription factors, MYB46 and MYB83, do not respond to VND7 in a switch like fashion as it was previously thought [46]. Furthermore, the cell coverage used in the study was comparable to the previously conducted microscopy-based analysis with an extra advantage of finding cellular quiescence centres, which were absent in earlier studies [46, 47]. This important finding not only cleared the ambiguity in the understanding of an important biological pathway but also provided an alternate pathway where VND7 was shown to work in a cycle by governing through four targets. In another study by Ryu et al. [48], a representation of all major tissues along with distinct and rare cell populations was found using ~10,000 Arabidopsis root protoplasts. The development trajectories of hair and non-hair root cells were identified from meristematic to completely differentiated state. A proof of concept for gene expression at the single-cell level was also presented in this study by conducting a comparative analysis on root epidermal mutants and wild type cells [48]. This is an significant addition towards the phase-2 of Plant Cell Atlas development which addresses building out the initial maps and networks [49].

The dynamics of gene expression in Arabidopsis root cells was studied by Jean-Baptise et al. [50], where they analysed 3,121 cells and found 6,152 unique molecular identifiers (UMIs) on average per cell, which corresponded to 2,445 both known and novel genes per cell. These UMIs, which are random nucleotide sequences that tag individual mRNA molecules, discriminate between original mRNA molecules and those that arise from duplicate polymerase amplifications [51]. This study also found the transcription factors associated with early and late cell life states. Recently another study by Zhang et al. [52] involving ~15,000 Arabidopsis root cells showed that the cells could be distributed into 24 distinct cell clusters, each with its own cluster-specific markers. The clusters were found to be involved in a wide range of functions from cell differentiation and hormone response to ion assimilation in root cells [52]. Similar studies on root phloem are missing yet, perhaps due to very low abundance of root phloem cells, hence difficult access to enough cDNA [43].

The aforementioned studies have employed single cells from root tissues to investigate key developmental processes in plants. This was important for trialling these modern single cell techniques in plants but their application to single cells from other tissue types in plants is also warranted as several key developmental decisions are being taken in plant organs other than roots and could better be studied using cells from those tissues and organs. A recent study by Nelms and Walbot [53] employed anther cells to investigate the previously poorly understood mechanisms of meiotic entry in plants. This study provided very useful insights about the change in gene expression at mid-leptotene stage and cell morphology just prior and at the start of meiosis in maize plants. The increase in gene expression related to membrane bound organelles and mitochondria and degradation of ribosomal templates suggests the preparation of switch towards functions encoded by haploid genome.

Single-cell transcriptomics and plant functional biology

Single-cell studies in relation to plant functional mechanisms such as response to different environment stimuli are scarce and most single-cell studies to date have focused on cell differentiation. However, the fact that each cell also contains the foot print of whole organism in the nucleus also warrants its use for studying the biological functions happening in an organism [16]. This is especially important in relation to complex multigenic traits, which otherwise need a lot of time and resources for their accurate measurement at both genotypic and phenotypic levels. The examples of such traits include plant responses to various biotic and abiotic stresses where an interaction of more than one biological pathway is involved in enabling a plant to cope a certain stress.

The study by Jean-Baptise et al. [50] also tried to deconvolute the dynamics of the functional control of plant traits at the single-cell level by studying the effect of heat stress on Arabidopsis seedlings. While assessing gene expression heterogeneity among different cell types, they found overall canonical heat shock gene expression while significant change in the expression of other genes such as root hair and non-hair controlling genes was also observed in response to heat stress[50]. The above mentioned study by Shulse et al. [45] also studied the functional response to environmental stimuli by applying with and without sucrose treatments on root development. They found that sucrose application does not substantially changes the cell identity but alter the cell proportion and gene expression in different tissue types. Another study by Hossain et al. [54] identified differentially methylated regions in soybean root hairs, which are single-cell extensions of roots, and multicellular stripped roots, under control and heat stress treatments. While both root hairs and multicellular stripped roots exhibited a hypomethylation pattern under heat stress, root

hairs were found to be more hypermethylated than multicellular stripped roots under control conditions.

Another example illustrating the use of scRNA-Seq in the context of plant functional biology is the study by Gould et al. [55], where circadian rhythm activity was investigated at the single-cell level in Arabidopsis. This study employed confocal microscopy to study single-cells tracked with the help of certain fluorescence markers. The results, which were based on spatial waves of gene expression of plant circadian clock, revealed the presence of multiple coordination points as opposed to the centralised clock present in mammals [55, 56]. These robust, but desynchronised single-cell oscillations found in this study, suggest that a much more synchronised mechanism of circadian clock control in plants could be achieved by employing scRNA-seq on the same treatments. Such an approach could potentially uncover many rarer cell types and quiescent centres.

The plant germlines also have enormous role in the functional biology of plants. Understanding the mechanism of pollination in plants provides a better control over pollen sterility and its appropriate use during various stages of research and trait improvement such as hybrid seed development [57]. In a recent attempt to combine fluorescence microscopy and single-cell transcriptomics on plant sperm cells, Misra et al. [58] found the expression of 1900 genes per cell compared to 7000 genes in a bulk of 50 single-cells. These results show the potential of utilising single-cell transcriptomic methodologies to produce high quality gene expression data from such low cell numbers for future targeted genomics [58]. The optimisation of threshold in algorithms to detect gene expression from such low number of cells improves the study of gene expression levels compared to bulk populations.

Bioinformatics analysis of single-cell transcriptomics data

In order to obtain important biological insights into experiments at the single-cell level, bioinformatics approaches must be employed to analyse the wide array of data types generated. To this end, over 600 scRNA-Seq bioinformatics tools have been developed; these include tools that have either been created solely for the purpose of scRNA-Seq data analysis or have been modified from existing bulk RNA-Seq analysis tools [59]. In an effort to make these resources more accessible to the scientific community, the Oshlack Lab [59] created an online database with a curated list of relevant software packages (https://www.scrna-tools.org/tools). Each package is categorised according to its function, contains a brief description of its intended use and a link to the source code is also provided. Several reasons have necessitated the creation of bioinformatics tools that can specifically handle scRNA-Seq data analysis from mammalian and plant systems alike, a major one being the amplification biases introduced into the minute amounts of starting cellular RNA material. If not handled properly, lowly-expressed transcripts may erroneously appear to be significantly expressed and incorrectly estimate cell-to-cell variability [60]. Another important source of technical noise is batch effects, which can be introduced when cells from one population are captured and sequenced separately from another population [60]. While careful experimental designs can mitigate these effects, unwanted sources of variation can remain in the data and must be removed using computational approaches. Yet another source of technical noise, which is specific to droplet-based scRNA-Seq data, is the presence of empty droplets, i.e. droplets that are devoid of cells but contain ambient RNA that have leaked from damaged cells [61]. While the resulting transcripts can still have UMIs assigned to their corresponding barcodes, their cellular identity is lost. Such transcripts must be filtered out from the data as they can lead to misleading biological conclusions. Additionally, the sparse nature of scRNA-Seq data means that resulting measurements can

harbour a large number of zeroes stemming from genes that have no mapped reads or no UMIs associated to them; this can occur due to technical reasons such as the scRNA-Seq platform or biological reasons such as low gene expression levels [62]. This explains why a sequencing coverage of 30X would be required in single-cell sequencing to achieve a genome coverage of at least 90%. In contrast, a sequencing depth of only 4X would be required at the bulk-level to achieve a similar coverage [36]. While existing scRNA-Seq bioinformatics tools have in-built statistical models that take into account the sparse nature of scRNA-Seq data, data imputation, i.e. the replacement of missing values with estimated values [63] may also be needed. Last but not least, cell-to-cell heterogeneity such as differences in cell size and cell cycle state can act as important confounding factors during cell type identification [64] and must be addressed during data analysis.

Bioinformatics tools that have been designed to analyse mammalian single-cell data can also be used to analyse plant single-cell data, especially scRNA-Seq as demonstrated in several studies [45, 48, 50, 52]. A summary of these tools as well as an overview of a typical scRNA-Seq data analysis workflow in Table 1. and Figure 1., respectively. The first stage of any scRNA-Seq data analysis is to determine the expression of each gene in each cell. This is achieved via read alignment of the scRNA-Seq data against an appropriate reference. As with any type of Next-Generation Sequencing (NGS) data analysis, it is important to assess the quality of the raw data using FastQC [65] or similar tools prior to read alignment. Adapter sequences and low-quality bases must also be removed if present using tools such as Trimmomatic [66], TrimGalore [67] among others. Read alignment and gene quantification at the cellular level can then be performed using standalone tools such as CellRanger [68] and scPipe [69], or a combination of tools such as Hisat2 [70] and HTSeq

[71]. Once the gene-cell matrix is obtained, it can be further processed to address the biological question at hand, and there are several tools that are available for this purpose.

A very popular tool for scRNA-Seq data analysis is Seurat [72]. This R package, which is available on the Comprehensive R Archive Network (CRAN) [https://cran.r-project.org/], employs machine learning techniques to identify positive and negative markers that define cell clusters based on gene expression values within these clusters. Seurat uses as input a gene-cell matrix and employs several quality control metrics such as unique gene counts and percentage of mitochondrial sequence present per cell to filter out low quality cells. Then, data normalisation, which is an important step to correct for the inherent noise and variability of scRNA-Seq datasets [73], is carried out. This is followed by the identification of highly variable genes, linear dimensional reduction of the data via Principal Component Analysis (PCA), unsupervised clustering of cells, identification of differentially expressed markers that define clusters and assignment of cell type identity to markers. Seurat also allows non-linear dimensional reduction via t-Distributed Stochastic Neighbor Embedding (tSNE) and Uniform Manifold Approximation and Projection (UMAP). These are valuable options for visualising and exploring scRNA-Seq datasets [74]. Additionally, the latest releases of Seurat (v2 and v3) an alignment procedure that allows the integration of scRNA-seq datasets such that heterogeneous tissues across different conditions can be compared, and measurements produced by different protocols and research environments can be integrated [75, 76].

Another important application of scRNA-Seq in the context of plant research is single-cell trajectory analysis. Plant cells are always changing from one functional state to another during development. Cells in different states express different sets of genes, and as

the cells move between states, they undergo transcriptional re-configuration where some genes are silenced, and others are newly activated [77]. While it is very difficult to experimentally purify cells that belong to different transient stages via bioinformatics analysis [40, 77], a popular tool that has been designed for this purpose is Monocle [77]. This R package, which is available on Bioconductor [http://bioconductor.org/], uses a machine learning technique called reversed graph embedding to construct single-cell trajectories using pseudotime. In other words, Monocle learns the sequence of gene expression changes that each cell would undergo during any given biological process and places the cells at their proper positions on this learned trajectory. A simplistic Monocle workflow consists of three steps. First, the algorithm identifies genes that define a cell's progress. These are genes whose expression levels change as a function of progress through the biological process being studied. Then, data dimensional reduction techniques such as tSNE are applied to the data. Lastly, the cells are ordered along their trajectories. The trajectories may also include branches that occur as a result of cells being subjected to gene expression programs underlying alternative cell fates. The latest version of Monocle (v3) not only supports the analysis of millions of cells at a time but also offers some new and powerful features. These include an optimised workflow for more accurate developmental trajectory analyses, and an updated statistical test to survey genes that have trajectory-dependent expression. Monocle v3 also allows UMAP visualisation as well as a 3D interface to view trajectories and gene expression [https://cole-trapnell-lab.github.io/monocle3/].

Seurat and Monocle are well-suited for the analysis of plant scRNA-Seq data as demonstrated in several studies [45, 48, 50, 52], and the contribution of data visualisation in this regards cannot be overlooked. Both tools mentioned here use tSNE and UMAP as a means to visualise clusters in single-cell data. Recently, the application of an eFP (electronic

fluorescent pictographis)-Seq Browser to survey scRNA-Seq data has also been demonstrated [48]. The eFP-Seq Browser was originally designed to explore microarray data [78] and has since been applied to several bulk RNA-Seq studies of major crops such as wheat, maize, rice, potato, soybean, barley and tomato [http://bar.utoronto.ca/]. This tool shows the read depth coverage of candidate genes, coupled with eFP images that allow visual assessment of expression levels of these genes across samples. When applied to scRNA-Seq data, the user is able to see the expression levels of their gene of interest in 3 different modes: absolute, relative and compare. The 'absolute' mode compares the expression level of the candidate gene in each cell of various tissue types to the highest expression level that has been recorded for the same gene. The 'relative' mode shows the ratio of a cell's expression level in different tissue types relative to an appropriate control. As for the 'compare' mode, it compares the expression levels of 2 genes across cells of different tissue types. An example of scRNA-Seq data visualisation via an eFP browser is available for the data presented by Ryu et al. [48]. To reiterate, this study assessed the expression levels of ~10,000 Arabidopsis root protoplasts in all major tissues.

While all plant-related single-cell studies conducted to date have focused on the model plant species *A. thaliana,* they are also amenable to crop species. The identification of cell types that respond most strongly to abiotic stresses such as heat, drought, and nutrient starvation will allow us to genetically manipulate relevant cell types to generate stress-tolerant crops without pleiotropically affecting plant fitness and yield [50]. Additionally, a better understanding of the underlying biological processes will allow scRNA-seq to be combined with important modalities such as CRISPR/Cas9-based genetic screens, for instance. CRISPR/Cas9 has been widely used to systematically characterise gene functions and if combined with scRNA-Seq will act as a powerful tool for the high-

throughput dissection of gene functions by allowing sequential knockout of candidate genes and studying the resulting effects [36]. A potential application of such multimodal technique could be during the assessment of expression levels of plant regulatory factors that can initiate the formation of somatic embryos from plant tissue cultures to accelerate the breeding cycle [79].

Future perspectives

Sufficient numbers of viable cells are a prerequisite to conducting any single-cell based assay and obtaining accurate information. In order to run plant cells through a single-cell RNA-seq analysis, they must be stripped of their cell walls using a cocktail of enzymes and although this most likely affects the internal cell biology of the cells, the data to date suggests this may not prevent the recognition of different cells within a tissue [45]. Using Arabidopsis root tissue Shulse et al. [45] showed that single-cell RNA-Seq can be used to profile developmental processes and alterations induced by external factors. The future comparative studies will establish the boundaries of interpretation that may exist around the use of protoplasts to study plant tissues. However, the smooth and efficient isolation of viable plant single-cells remains a challenge. Optimisation of already available cell isolation techniques, especially the enzymatic treatment time to isolate protoplast from different origins, could however circumvent this problem. The applications of commercially available cell suspension protocols especially through 10x Genomics could be extended and modified for plant species other than Moss for even single-cell capture [39]. The optimisation of already available methods such as serial dilution, micro-manipulation, FACS and optical tweezers would further enhance our capability to isolate large number of viable individual cells [80-83]. The time and resource allocation for purpose are justified now more than ever

given the availability of such revolutionary techniques relying on presence of viable single-cell.

The in-depth bioinformatic analysis of single cells is dependent on the availability of large number of viable single cells. As mentioned above, single-cell genomics techniques require at least 30x sequencing depth compared to 4x bulk sequencing, thus increasing the sequencing costs substantially. To cope with the cost issues, 5´ or 3´ sequencing with UMI tagging was introduced [84, 85]. This, together with the application of bioinformatics tools which can handle the inherent noise and sparsity of scRNA-Seq data, as previously discussed, have been key to addressing the issues related to the accuracy of read assignment and PCR bias under shallow sequencing. However, allele-specific expression profiling still remains a challenge under shallow sequencing depths [86].The development of novel bioinformatics pipelines to process such datasets will open avenues for obtaining useful biological insights from single-cell data. Additionally, the decrease in sequencing costs, coupled with the increased throughput of sequencing data are likely to create a data deluge which will need to be handled efficiently. Existing and novel bioinformatics pipelines will need automation and adaptation to run in high-performance computing environments to significantly reduce the time required to undertake computationally intensive steps. Nonetheless, the large volumes of data to be analysed also offer the promise of uncovering many new aspects regarding the biology of many model and non-model plant species, and the application of unsupervised machine learning (ML) techniques in this regard will be of utmost importance [49]. Further, innovative ways of visualising plant single-cell data will also be required. For instance, the application of semantic zooming will increase the resolution at which images can be visualised, and this is bound to offer important insights at the cellular level [49]. Coupling scRNA-Seq data with such an imaging technique would allow

the more accurate spatial and temporal reconstruction of various cellular types and states in any tissue [87]. Single-cell Omics based data analysis uses many of the approaches applied in genomics data analysis and is learning from genomics field. SeqGeq 'seek-geek' by FlowJo or Seurat R package allows overlaying sequencing data onto the single cell sorted into a specific well and sequenced [88]. BD Abseq is another approach which enables the parallel detection of protein and mRNA in a single-cell and is yet to be utilised in plants to study cell surface protein expression along with cell specific mRNA expression profiles [89].

The progress of all above-mentioned droplet- based single-cell approaches in plants has been impeded because of their large cell size. Majority of available sc-Omics studies on plants so far have focused on single-cells extracted from roots perhaps due to their size, which aligns with the available single-cell encapsulation and capture infrastructure [90]. The differentiated cell size from the majority of plant tissues is too big to meet the requirements of available droplet or well-based platforms, which have the capacity of fit a maximum of 50 um cell size [91]. This warrants the development of plant specific chips and panels along with plant based unique barcodes to capture a wide range of plant cell types. The other major reason of extensive use of root cells in previous studies may be the availability of 'relatively' small number and types of cells in the roots compared to other tissues [48], which makes them ideal to be used for proof of concept studies in plants. Moreover, major costs of single cell studies arise from the sequencing part of experiment, where a trade-off comes into play between number of cells per analysis and read depth per cell. Now that these single cell techniques have been successfully applied in plants at proof of concept level, deconvolution of key biological processes would warrant utilisation of single cells from tissues other than roots.

Challenges in finding the true developmental trajectories as cells are lost during tissue dissociation for cells isolation need to be considered as available bioinformatics approaches rely on development stage as dominant signals in single cell profiles [40]. The traditional flow cytometry's prerequisite of prior knowledge of morphological and cellular markers could become handy here in order to resolve those developmental trajectories profile of different plant tissues and organs. Polyploidy is thought to induce bigger cell sizes to accommodate multiplying chromatin material or increase number of cells. Both could potentially pose issues in the single cell analysis [92]. Larger size of cells may require special infrastructure for different plant cell types to capture single cells in droplets or wells depending on the methods used. Larger numbers of cells required for the analyses of plant tissues als increase the costs of sequencing by increasing the threshold of sufficient number of cells to be captured to represent certain tissues or organs. Single nuclei analysis methods mainly in human and mouse have recently attained the attention of scientists to study the genome wide chromatin accessibility profiles in single-cells. These methods would open the ways to map different development and functional pathways with chromatin accessible regions.

The portability and long read sequencing capability of nanopore embedding flow cytometry in existing nanopore infrastructure could potentially open the ways for plant single-cells sequencing of variety of cell types [93]. PacBio platforms especially the Real-Time (SMRT) solutions are another option to capture and read variety of plant cells at single-cell resolution [94]. The calibration of reagent volumes used in the existing single-cell sequencing platforms, sample digestion artefacts and resultant debris clean up shall further enhance the quality of output data from single-cells originating from different plant tissues. Furthermore, the study of phloem cells differentiation and their role in plant developmental

and functional biology still remains unanswered. The employment of sc-Omics techniques to study single cells from shoot apical meristems could further enhance our understanding of cell signalling among different tissue types along the lengths of whole plants.

Finally, all published single-cell transcriptomic studies in plants thus far agreed on the fact that high-throughput scRNA-seq enables the identification and characterisation of rare cell types and quiescence centres present in a large heterogeneous cell population which were unknown or uncharacterised previously. We anticipate, based on the initial findings in plants published to date, that these technologies have immense potential to disrupt our traditional ways of studying hormone profiling, metabolomics, and protein-interactions. This would be possible by complementation of different single-cell techniques such as single-cell RNA-seq, ATAC-seq, ChIPseq and DNA methylation protocols in plants in similar ways as this model has been successfully implemented in mammalian system [95-98]. The challenge of demonstrating single-cell data in simple and informative ways could be converted into an opportunity by combining different data sets and algorithms to make a universal Plant Cell Atlas [49, 99] similar to Human Cell Atlas Project.

Acknowledgments

The authors would like to extend sincere thanks to Robert Salomon for inspiring to write this manuscript. Resources were provided by The University of Western Australia.

Literature cited

1. Palmer JD: **Comparison of chloroplast and mitochondrial genome evolution in plants**. In: *Cell organelles.* Springer; 1992: 99-133.
2. Quiroz-Figueroa FR, Rojas-Herrera R, Galaz-Avalos RM, Loyola-Vargas VM: **Embryo production through somatic embryogenesis can be used to study cell differentiation in plants**. *Plant Cell, Tissue and Organ Culture* 2006, **86**(3):285.
3. Buvat R: **Ontogeny, cell differentiation, and structure of vascular plants**: Springer Science & Business Media; 2012.
4. Fletcher JC: **Shoot and floral meristem maintenance in Arabidopsis**. *Annual review of plant biology* 2002, **53**(1):45-66.
5. Doležel J, Greilhuber J, Suda J: **Flow cytometry with plants: an overview**. *Flow cytometry with plant cells: analysis of genes, chromosomes and genomes* 2007:41-65.
6. Song Y, Xu X, Wang W, Tian T, Zhu Z, Yang C: **Single cell transcriptomics: moving towards multi-omics**. *Analyst* 2019, **144**(10):3172-3189.
7. Bourge M, Brown SC, Siljak-Yakovlev S: **Flow cytometry as tool in plant sciences, with emphasis on genome size and ploidy level assessment**. *Genetics & Applications* 2018, **2**(2):1-12.
8. Kalia A, Sharma SP: **Single-Cell Omics in Crop Plants: Opportunities and Challenges**. In: *Single-Cell Omics.* Elsevier; 2019: 341-355.
9. Pieruschka R, Poorter H: **Phenotyping plants: genes, phenes and machines**. *Functional Plant Biology* 2012, **39**(11):813-820.
10. Fraser J, Williamson I, Bickmore WA, Dostie J: **An overview of genome organization and how we got there: from FISH to Hi-C**. *Microbiol Mol Biol Rev* 2015, **79**(3):347-372.
11. Smith H: **The molecular biology of plant cells/edited by H. Smith**. *Botanical monograph; v 14*.
12. Dekker J, Rippe K, Dekker M, Kleckner N: **Capturing chromosome conformation**. *science* 2002, **295**(5558):1306-1311.
13. Ye F, Huang W, Guo G: **Studying hematopoiesis using single-cell technologies**. *Journal of hematology & oncology* 2017, **10**(1):27.
14. Tang X, Huang Y, Lei J, Luo H, Zhu X: **The single-cell sequencing: new developments and medical applications**. *Cell & Bioscience* 2019, **9**(1):53.
15. Libault M, Pingault L, Zogli P, Schiefelbein J: **Plant systems biology at the single-cell level**. *Trends in plant science* 2017, **22**(11):949-960.
16. de Oliveira Dal'Molin CG, Nielsen LK: **Plant genome-scale reconstruction: from single cell to multi-tissue modelling and omics analyses**. *Current opinion in biotechnology* 2018, **49**:42-48.
17. Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS, Cairns JE: **Translating high-throughput phenotyping into genetic gain**. *Trends in plant science* 2018, **23**(5):451-466.
18. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, Tsai P-C, Casasent A, Waters J, Zhang H: **Punctuated copy number evolution and clonal stasis in triple-negative breast cancer**. *Nature genetics* 2016, **48**(10):1119.

19. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A: **mRNA-Seq whole-transcriptome analysis of a single cell**. *Nature methods* 2009, **6**(5):377.
20. Chen X, Miragaia RJ, Natarajan KN, Teichmann SA: **A rapid and robust method for single cell chromatin accessibility profiling**. *Nature Communications* 2018, **9**(1):5345.
21. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P: **Simultaneous epitope and transcriptome measurement in single cells**. *Nature methods* 2017, **14**(9):865.
22. Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR: **Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding**. *Scientific reports* 2017, **7**:44447.
23. Wang D, Bodovitz S: **Single cell analysis: the new frontier in 'omics'**. *Trends in biotechnology* 2010, **28**(6):281-290.
24. Chappell L, Russell AJ, Voet T: **Single-cell (multi) omics technologies**. *Annual review of genomics and human genetics* 2018, **19**:15-41.
25. Galli E, Friebel E, Ingelfinger F, Unger S, Núñez NG, Becher B: **The end of omics? High dimensional single cell analysis in precision medicine**. *European journal of immunology* 2019, **49**(2):212-220.
26. Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z, Huang Y, Wang J: **Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems**. *Molecular cell* 2019, **73**(1):130-142. e135.
27. Zhang L, Vertes A: **Single-Cell Mass Spectrometry Approaches to Explore Cellular Heterogeneity**. *Angewandte Chemie International Edition* 2018, **57**(17):4466-4477.
28. Burgess DJ: **Spatial transcriptomics coming of age**. *Nature Reviews Genetics* 2019, **20**(6):317-317.
29. Crosetto N, Bienko M, Van Oudenaarden A: **Spatially resolved transcriptomics and beyond**. *Nature Reviews Genetics* 2015, **16**(1):57-66.
30. Herzenberg LA, Parks D, Sahaf B, Perez O, Roederer M, Herzenberg LA: **The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford**. *Clinical chemistry* 2002, **48**(10):1819-1827.
31. Schmitz B, Radbruch A, Kümmel T, Wickenhauser C, Korb H, Hansmann M, Thiele J, Fischer R: **Magnetic activated cell sorting (MACS)—a new immunomagnetic method for megakaryocytic cell isolation: comparison of different separation techniques**. *European journal of haematology* 1994, **52**(5):267-275.
32. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA: **Laser capture microdissection**. *Science* 1996, **274**(5289):998-1001.
33. Brehm-Stecher BF, Johnson EA: **Single-cell microbiology: tools, technologies, and applications**. *Microbiol Mol Biol Rev* 2004, **68**(3):538-559.
34. Xu X, Wang J, Wu L, Guo J, Song Y, Tian T, Wang W, Zhu Z, Yang C: **Microfluidic Single-Cell Omics Analysis**. *Small* 2019.
35. Giladi A, Amit I: **Single-cell genomics: a stepping stone for future immunology discoveries**. *Cell* 2018, **172**(1-2):14-21.
36. Yuan Y, Lee H, Hu H, Scheben A, Edwards D: **Single-cell genomic analysis in plants**. *Genes* 2018, **9**(1):50.
37. Li L, Guo F, Gao Y, Ren Y, Yuan P, Yan L, Li R, Lian Y, Li J, Hu B: **Single-cell multi-omics sequencing of human early embryos**. *Nature cell biology* 2018, **20**(7):847.

38. Jia X, Zhang X, Qu J, Han R: **Optimization conditions of wheat mesophyll protoplast isolation**. *Agricultural Sciences* 2016, **7**(12):850-858.
39. Genomics x: **Moss Protoplast Suspension for Single Cell RNA Sequencing. In: Sample Preparation Demonstrated Protocol** 2017.
40. Efroni I, Birnbaum KD: **The potential of single-cell profiling in plants**. *Genome biology* 2016, **17**(1):65.
41. Bennett M, Leitch I: **Plant genome size research: a field in focus**. *Annals of botany* 2005, **95**(1):1-6.
42. Zohary D, Hopf M: **Domestication of plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley**: Oxford University Press; 2000.
43. Rodriguez-Villalon A, Brady SM: **Single cell RNA sequencing and its promise in reconstructing plant vascular cell lineages**. *Current opinion in plant biology* 2019, **48**:47-56.
44. Herring CA, Banerjee A, McKinley ET, Simmons AJ, Ping J, Roland JT, Franklin JL, Liu Q, Gerdes MJ, Coffey RJ: **Unsupervised trajectory analysis of single-cell RNA-seq and imaging data reveals alternative tuft cell origins in the gut**. *Cell systems* 2018, **6**(1):37-51. e39.
45. Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, Gouran M, Turco GM, Zhu Y, O'Malley RC, Brady SM: **High-throughput single-cell transcriptome profiling of plant cell types**. *Cell reports* 2019, **27**(7):2241-2247. e2244.
46. Turco GM, Rodriguez-Medina J, Siebert S, Han D, Vahldick H, Shulse CN, Cole BJ, Juliano C, Dickel DE, Savageau MA: **Molecular Mechanisms Driving Bistable Switch Behavior in Xylem Cell Differentiation**. *bioRxiv* 2019:543983.
47. Demura T, Tashiro G, Horiguchi G, Kishimoto N, Kubo M, Matsuoka N, Minami A, Nagata-Hiwatashi M, Nakamura K, Okamura Y *et al*: **Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells**. *Proceedings of the National Academy of Sciences* 2002, **99**(24):15794-15799.
48. Ryu KH, Huang L, Kang HM, Schiefelbein J: **Single-cell RNA sequencing resolves molecular relationships among individual plant cells**. *Plant physiology* 2019, **179**(4):1444-1456.
49. Rhee SY, Birnbaum KD, Ehrhardt DW: **Towards building a plant cell atlas**. *Trends in plant science* 2019.
50. Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus J: **Dynamics of gene expression in single root cells of A. thaliana**. *The Plant Cell* 2019:tpc. 00785.02018.
51. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W: **Comparative analysis of single-cell RNA sequencing methods**. *Molecular cell* 2017, **65**(4):631-643. e634.
52. Zhang T-Q, Xu Z-G, Shang G-D, Wang J-W: **A Single-Cell RNA Sequencing Profiles the Developmental Landscape of Arabidopsis Root**. *Molecular plant* 2019, **12**(5):648-660.
53. Nelms B, Walbot V: **Defining the developmental program leading to meiosis in maize**. *Science* 2019, **364**(6435):52-56.

54. Hossain MS, Kawakatsu T, Kim KD, Zhang N, Nguyen CT, Khan SM, Batek JM, Joshi T, Schmutz J, Grimwood J *et al*: **Divergent cytosine DNA methylation patterns in single-cell, soybean root hairs**. *New Phytologist* 2017, **214**(2):808-819.
55. Gould PD, Domijan M, Greenwood M, Tokuda IT, Rees H, Kozma-Bognar L, Hall AJ, Locke JC: **Coordination of robust single cell rhythms in the Arabidopsis circadian clock via spatial waves of gene expression**. *Elife* 2018, **7**:e31700.
56. Reppert SM, Weaver DR: **Coordination of circadian timing in mammals**. *Nature* 2002, **418**(6901):935.
57. Budar F, Pelletier G: **Male sterility in plants: occurrence, determinism, significance and use**. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie* 2001, **324**(6):543-550.
58. Misra CS, Santos MR, Rafael-Fernandes M, Martins NP, Monteiro M, Becker JD: **Transcriptomics of Arabidopsis sperm cells at single-cell resolution**. *Plant reproduction* 2019, **32**(1):29-38.
59. Zappia L, Phipson B, Oshlack A: **Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database**. *PLoS computational biology* 2018, **14**(6):e1006245.
60. Yuan G-C, Cai L, Elowitz M, Enver T, Fan G, Guo G, Irizarry R, Kharchenko P, Kim J, Orkin S: **Challenges and emerging directions in single-cell analysis**. *Genome biology* 2017, **18**(1):84.
61. Lun AT, Riesenfeld S, Andrews T, Gomes T, Marioni JC: **EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data**. *Genome biology* 2019, **20**(1):63.
62. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Campbell KR, Beerenwinkel N, Mahfouz A: **Eleven grand challenges in single-cell data science**. *Genome biology* 2020, **21**(1):1-35.
63. Peng T, Zhu Q, Yin P, Tan K: **SCRABBLE: single-cell RNA-seq imputation constrained by bulk RNA-seq data**. *Genome biology* 2019, **20**(1):88.
64. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O: **Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells**. *Nature biotechnology* 2015, **33**(2):155.
65. Andrews S: **FastQC: a quality control tool for high throughput sequence data. 2010**. In.; 2016.
66. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
67. Krueger F: **Trim galore**. *A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files* 2015.
68. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J: **Massively parallel digital transcriptional profiling of single cells**. *Nature communications* 2017, **8**:14049.
69. Tian L, Su S, Dong X, Amann-Zalcenstein D, Biben C, Seidi A, Hilton DJ, Naik SH, Ritchie ME: **scPipe: a flexible R/Bioconductor preprocessing pipeline for single-cell RNA-sequencing data**. *PLoS computational biology* 2018, **14**(8):e1006361.
70. Kim D, Langmead B, Salzberg S: **HISAT2: graph-based alignment of next-generation sequencing reads to a population of genomes**. In.; 2017.
71. Anders S, Pyl PT, Huber W: **HTSeq—a Python framework to work with high-throughput sequencing data**. *Bioinformatics* 2015, **31**(2):166-169.

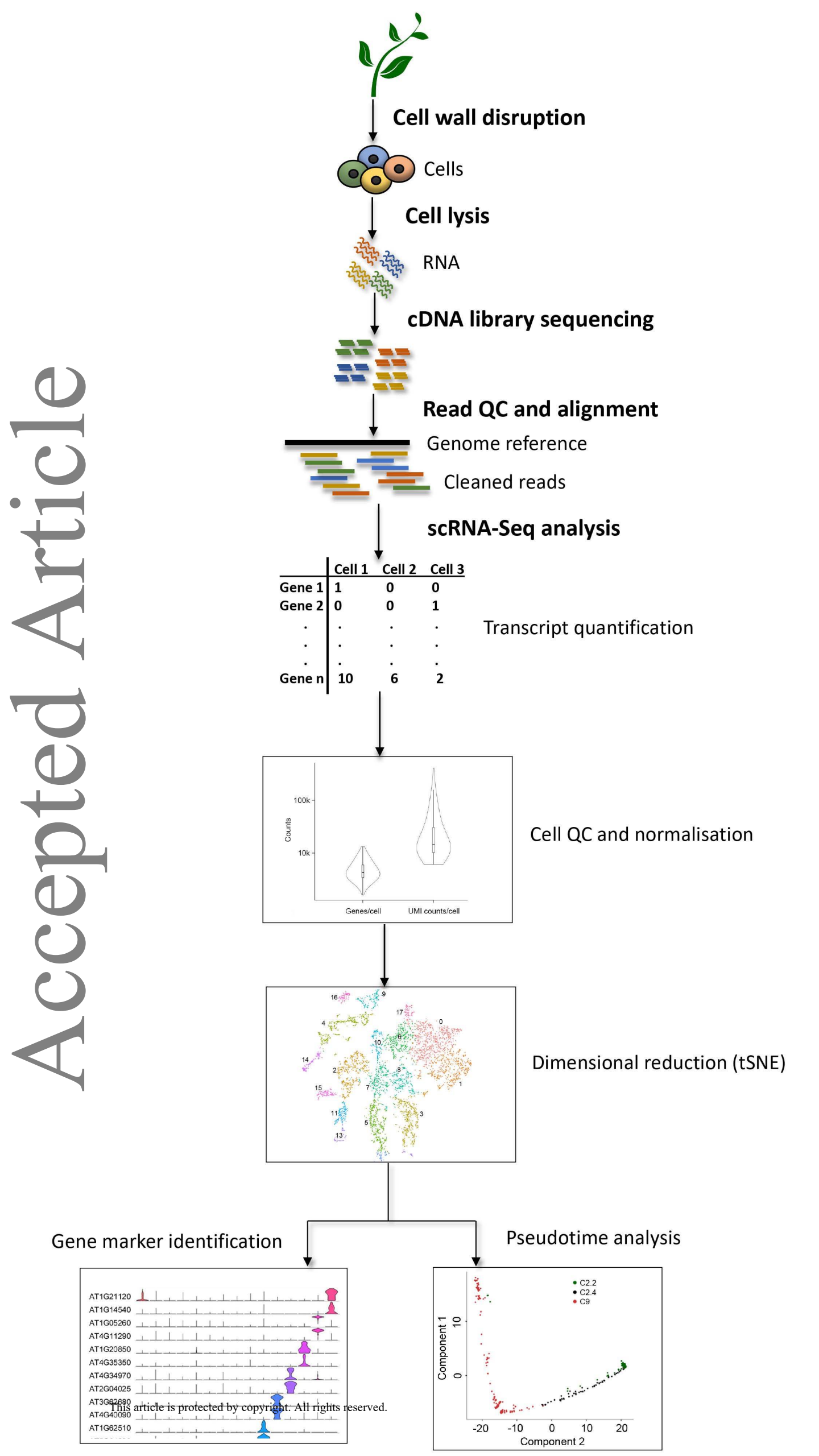
72. Satija R, Farrell JA, Gennert D, Schier AF, Regev A: **Spatial reconstruction of single-cell gene expression data**. *Nature biotechnology* 2015, **33**(5):495.
73. Vallejos CA, Risso D, Scialdone A, Dudoit S, Marioni JC: **Normalizing single-cell RNA sequencing data: challenges and opportunities**. *Nature methods* 2017, **14**(6):565.
74. Lafzi A, Moutinho C, Picelli S, Heyn H: **Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies**. *Nature protocols* 2018:1.
75. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: **Integrating single-cell transcriptomic data across different conditions, technologies, and species**. *Nature biotechnology* 2018, **36**(5):411.
76. Stuart T, Satija R: **Integrative single-cell analysis**. *Nature Reviews Genetics* 2019:1.
77. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells**. *Nature biotechnology* 2014, **32**(4):381.
78. Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ: **An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets**. *PloS one* 2007, **2**(8).
79. Altpeter F, Springer NM, Bartley LE, Blechl AE, Brutnell TP, Citovsky V, Conrad LJ, Gelvin SB, Jackson DP, Kausch AP: **Advancing crop transformation in the era of genome editing**. *The Plant Cell* 2016, **28**(7):1510-1520.
80. Ham RG: **Clonal growth of mammalian cells in a chemically defined, synthetic medium**. *Proceedings of the National Academy of Sciences of the United States of America* 1965, **53**(2):288.
81. Spangrude GJ, Heimfeld S, Weissman IL: **Purification and characterization of mouse hematopoietic stem cells**. *Science* 1988, **241**(4861):58-62.
82. Landry ZC, Giovanonni SJ, Quake SR, Blainey PC: **Optofluidic cell selection from complex microbial communities for single-genome analysis**. In: *Methods in enzymology.* vol. 531: Elsevier; 2013: 61-90.
83. Shapiro E, Biezuner T, Linnarsson S: **Single-cell sequencing-based technologies will revolutionize whole-organism science**. *Nature Reviews Genetics* 2013, **14**(9):618.
84. Hashimshony T, Wagner F, Sher N, Yanai I: **CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification**. *Cell reports* 2012, **2**(3):666-673.
85. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM: **Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets**. *Cell* 2015, **161**(5):1202-1214.
86. Hwang B, Lee JH, Bang D: **Single-cell RNA sequencing technologies and bioinformatics pipelines**. *Experimental & molecular medicine* 2018, **50**(8):96.
87. Rich-Griffin C, Stechemesser A, Finch J, Lucas E, Ott S, Schäfer P: **Single-Cell Transcriptomics: A High-Resolution Avenue for Plant Functional Genomics**. *Trends in plant science* 2020, **25**(2):186-197.
88. Vazquez J, Ong IM, Stanic AK: **Single-Cell Technologies in Reproductive Immunology**. *American Journal of Reproductive Immunology* 2019:e13157.
89. Shum EY, Walczak EM, Chang C, Fan HC: **Quantitation of mRNA Transcripts and Proteins Using the BD Rhapsody™ Single-Cell Analysis System**. In: *Single Molecule and Single Cell Sequencing.* Springer; 2019: 63-79.

90. Carey TR, Cotner KL, Li B, Sohn LL: **Developments in label-free microfluidic methods for single-cell analysis and sorting**. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* 2019, **11**(1):e1529.
91. Valihrach L, P.Androvic, Kubista M: **Platforms for Single-Cell Collection and Analysis**. *International Journal of Molecular Sciences* 2018, **19**(807).
92. Comai L: **The advantages and disadvantages of being polyploid**. *Nature reviews genetics* 2005, **6**(11):836-846.
93. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT: **Nanopore sequencing and assembly of a human genome with ultra-long reads**. *Nature biotechnology* 2018, **36**(4):338.
94. Schloss PD, Jenior ML, Koumpouras CC, Westcott SL, Highlander SK: **Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system**. *PeerJ* 2016, **4**:e1869.
95. Clark SJ, Argelaguet R, Kapourani C-A, Stubbs TM, Lee HJ, Alda-Catalinas C, Krueger F, Sanguinetti G, Kelsey G, Marioni JC: **scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells**. *Nature communications* 2018, **9**(1):781.
96. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, Bernstein BE: **Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state**. *Nature biotechnology* 2015, **33**(11):1165.
97. Hui T, Cao Q, Wegrzyn-Woltosz J, O'Neill K, Hammond CA, Knapp DJ, Laks E, Moksa M, Aparicio S, Eaves CJ: **High-resolution single-cell DNA methylation measurements reveal epigenetically distinct hematopoietic stem cell subpopulations**. *Stem cell reports* 2018, **11**(2):578-592.
98. Lee HJ, Smallwood SA: **Genome-wide analysis of DNA methylation in single cells using a post-bisulfite adapter tagging approach**. In: *Next Generation Sequencing.* Springer; 2018: 87-95.
99. Kaur P, Sahab S, Dickel D, Appels R, Datla R, Xiang D, Gao P, Hayden M, Tibbits J: **The Wheat Cell Atlas (WCA) for a New View of the Wheat Plant- IWGSC - Towards a Functionally Annotated Wheat Genome Reference Sequence**. In: *PAG ASIA, Plant and Animal Genome Conference, June 6-8, 2019: 2019; Shenzhen, China*.
100. Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MC: **Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing**. *Developmental cell* 2019, **48**(6):840-852. e845.

Figure legends and Table titles

Table 1. List of bioinformatics tools used for scRNA-Seq data analysis.

Figure 1. Overview of bioinformatics analysis of scRNA-Seq data. This includes read quality control, read alignment, transcript quantification of single cells, data normalisation, dimensional reduction and downstream applications such as gene marker identification and pseudotime analysis. Panels representing dimensional reduction (tSNE), gene marker identification and pseudotime analysis were taken from [100].



Step	Tool	Main function	Notes	References	Citations in plant scRNA-Seq studies
1	FastQC	Quality assessment of raw sequencing data		[64]	
2	Trimmomatic	Removal of low-quality bases and adapters from raw sequencing data		[65]	
	TrimGalore			[66]	
3	scPipe	Alignment of sequencing data against a reference genome assembly	scPipe processes scRNA-Seq data from UMI and non-UMI protocols. It also outputs a gene-cell matrix on which it performs quality control. Cell clustering and visualisation are also possible via an R shiny app.	[68]	
	CellRanger		CellRanger processes Chromium single cell 3' RNA-Seq data. It also performs clustering and gene expression analysis.	[67]	[44, 47, 51]
	Hisat2			[69]	
	STAR			[99]	[44, 47]
	HTSeq			[70]	
4	ascend	Cell-matrix quality control: filtering, normalisation, dimensionality reduction, clustering and visualisation	ascend processes Chromium, dropSeq and InDrop data.	https://github.com/powellgenomicslab/ascend	
	dropEST		dropEST processes droplet-based scRNA-Seq data.	[100]	
	DropletUtils		DropletUtils processes droplet-based scRNA-Seq data.	https://github.com/MarioniLab/DropletUtils	
	Scanpy		Scanpy is a highly-scablable Python program that can process up to 1 million cells.	[101]	
	Scater			[102]	

	Seurat		Seurat also has an alignment procedure that allows integration of scRNA-seq datasets so that heterogeneous tissues across different conditions can be compared, and measurements produced by different protocols and research environments can be integrated.	[71]	[47, 51]
5	ascend	Downstream analyses	Differential expression analysis.		
	Seurat		Differential expression analysis and marker gene discovery.	[71]	[47, 51]
	Scanpy		Differential expression analysis, trajectory inference and simulation of gene networks.	[101]	
	Monocle		Differential expression analysis and trajectory inference.	[76]	[44, 49, 51]
	RaceID		Identification of rare cell types.	https://github.com/dgrun/RaceID	