

## TECHNICAL ADVANCE

# Single-nucleus RNA sequencing of plant tissues using a nanowell-based system

Daniele Y. Sunaga-Franze<sup>1,†</sup> , Jose M. Muino<sup>2,†</sup> , Caroline Braeuning<sup>1,†</sup> , Xiaocai Xu<sup>3,†</sup> , Minglei Zong<sup>3</sup> , Cezary Smacznik<sup>3</sup> , Wenhao Yan<sup>3</sup> , Cornelius Fischer<sup>1</sup> , Ramon Vidal<sup>1</sup> , Magdalena Kliem<sup>1</sup> , Kerstin Kaufmann<sup>3,\*</sup>  and Sascha Sauer<sup>1,\*</sup> 

<sup>1</sup>Genomics Platforms, Max Delbrück Center for Molecular Medicine in the Helmholtz Association/Berlin Institute of Health, Berlin, Germany,

<sup>2</sup>Systems Biology of Gene Regulation, Institute of Biology, Humboldt-Universität zu Berlin, Berlin, Germany, and

<sup>3</sup>Plant Cell and Molecular Biology, Institute of Biology, Humboldt-Universität zu Berlin, Berlin, Germany

Received 29 January 2021; revised 16 July 2021; accepted 2 August 2021; published online 13 August 2021.

\*For correspondence (e-mail kerstin.kaufmann@hu-berlin.de; sascha\_sauer\_2020@gmx.de).

†These authors contributed equally to this work.

## SUMMARY

Single-cell genomics provides unprecedented potential for research on plant development and environmental responses. Here, we introduce a generic procedure for plant nucleus isolation combined with nanowell-based library preparation. Our method enables the transcriptome analysis of thousands of individual plant nuclei. It serves as an alternative to the use of protoplast isolation, which is currently a standard methodology for plant single-cell genomics, although it can be challenging for some plant tissues. We show the applicability of our nucleus isolation method by using different plant materials from different species. The potential of our single-nucleus RNA sequencing method is shown through the characterization of transcriptomes of seedlings and developing flowers from *Arabidopsis thaliana*. We evaluated the transcriptome dynamics during the early stages of anther development, identified stage-specific activities of transcription factors regulating this process, and predicted potential target genes of these transcription factors. Our nucleus isolation procedure can be applied in different plant species and tissues, thus expanding the toolkit for plant single-cell genomics experiments.

**Keywords:** single-nucleus isolation, snRNA-seq, *Arabidopsis thaliana* seedlings, flower development, anther, technical advance.

## INTRODUCTION

The fundamental units of life, cells, can vary tremendously within an organism. The analysis of specialized cells and their interactions is essential to develop a comprehensive understanding of the functions of tissues and biological systems in general. Major biological processes such as growth, development, and physiology ultimately gain plasticity from heterogeneity in cellular gene expression (Fischer et al., 2019).

Without precise transcriptional maps of different cell populations, we cannot accurately describe all their functions and the underlying molecular networks that drive their activities. Recent advances in single-cell RNA sequencing (scRNA-seq) and in particular single-nucleus RNA sequencing (snRNA-seq) have put the establishment

of comprehensive, high-resolution reference transcriptome maps of mammalian cells and tissues on the agenda of international consortia such as the Human Cell Atlas (Regev et al., 2017).

Similar efforts are made by the Plant Cell Atlas consortium (Rhee et al., 2019). Plant tissues and cells pose specific challenges compared to mammalian systems (Efroni and Birnbaum, 2016). Plant cells are immobilized in a rigid cell wall matrix, which must be removed to isolate single cells. Additional technical demands include size variability of plant cells and the presence of plastids and vacuoles. Consequently, these characteristics require considerably different operational procedures compared with mammalian tissues. Recently, plant scRNA-seq studies using protoplast isolation (PI) have been published (e.g., Denyer et al., 2019; Efroni et al., 2015, 2016; Jean-Baptiste et al.,

2019; Nelms and Walbot, 2018; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019). This procedure enables the identification and classification of plant cell types. However, it is known that enzymatic digestion of plant cell walls can introduce artifacts at the transcriptome level, limiting the applicability of this approach (Jean-Baptiste et al., 2019; Shulse et al., 2019). Approaches based on nucleus isolation offer an alternative approach (Thibivilliers et al., 2020).

Several methodologies are available for the generation of RNA-seq libraries from isolated cells or nuclei. Two of the most popular are droplet-based (e.g., 10x Chromium) and nanowell-based (e.g., Takara ICELL8) systems. Droplet-based methods are popular because of their scalability. They enable rapid processing of thousands of cells simultaneously. Particularly in the Chromium system, gel beads are supplied with a unique barcoded oligonucleotide. Cells or nuclei are encapsulated together with these beads and lysed, and the gel bead releases the barcoded oligonucleotide for reverse transcription (RT). RT is performed inside droplets and the product is transferred to a tube where amplification of cDNA occurs. One disadvantage is that, in some events, more than one cell or nucleus enters the same capsule, producing a mixed cDNA population (Lareau et al., 2020). Nanowell-based systems trap the isolated cells or nuclei in wells where the cDNA is produced. In particular, the ICELL8 system consists of a chip with more than 5000 nanowells containing barcoded oligonucleotides attached to their surface. Each cell or nucleus is deposited in a nanowell and its quality and number are checked by microscopy, which reduces the probability of obtaining transcriptomes from more than one cell/nucleus. One of the main disadvantages of nanowell systems is their more limited scalability, as each chip has a fixed number of nanowells. However, it is crucial to continue enriching and improving the repertoire of single-cell omics methodologies available for the plant research community.

Here, we introduce a single-nucleus sequencing protocol using the nanowell-based ICELL8 system by studying the dynamics of *Arabidopsis* transcriptomes during flower development. Working with nuclei has the advantage of eliminating organelles and vacuoles, as well as secondary metabolites localized in the cytoplasm that can interact with RNA and negatively affect next-generation sequencing (NGS) library preparation.

## RESULTS AND DISCUSSION

### Nucleus isolation and snRNA-seq library preparation

While PI-based methods have been shown to readily identify plant cell types, it is also known that they can lead to changes in gene expression and different cell types may be affected in different degrees (Jean-Baptiste et al., 2019; Shulse et al., 2019; Thibivilliers et al., 2020). To address

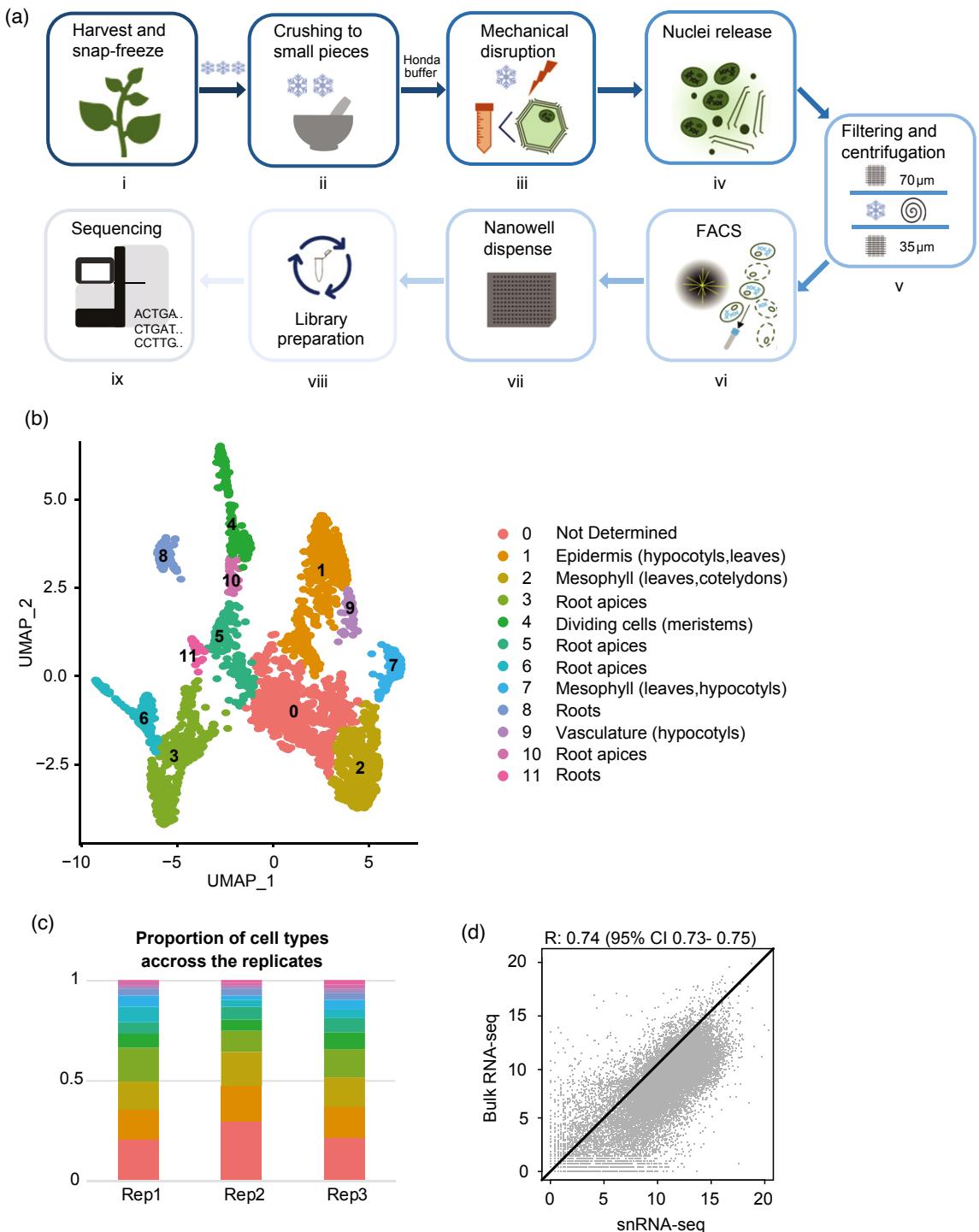
this issue, PI-responsive genes can be identified through an independent bulk RNA-seq experiment and subsequently eliminating them from the scRNA-seq analysis (Denyer et al., 2019). However, we show that the impact of PI cannot be completely eliminated in this way (Figure S1).

Here, we propose an snRNA-seq strategy for the transcriptome sequencing of individual plant nuclei (Figure 1a); full protocol in Experimental Procedures). The key step of our plant nucleus sequencing procedure consists of gentle but efficient isolation of plant nuclei. Snap-frozen *Arabidopsis* tissue is gently physically dissociated by pestle and transferred to Honda buffer for cell lysis (Moreno-Romero et al., 2017). Cell walls and cell membranes are mechanically disrupted using a gentleMACS Dissociator, keeping the nuclei largely intact, as observed by DAPI staining (Figure S2a). Released intact *Arabidopsis* nuclei are collected using fluorescence-activated cell sorting (FACS; Figure S2b). A clear separation between nuclei and debris was obtained (Figure S2c). To show the applicability of this method to different plant species/tissues, we performed nucleus isolation in *Arabidopsis thaliana* (seedlings and flowers), *Petunia hybrida* (flowers), *Antirrhinum majus* (flowers), and *Solanum lycopersicum* (flowers and leaves) (Figure S2a,b). The RNA that was isolated from these nuclei was of high quality, as observed by electrophoresis for *Arabidopsis* (Figure S2d).

The next step consists of generating high-quality cDNA libraries from the isolated nuclei. There are a number of different library preparation protocols and sequencing procedures that can be combined (Cao et al., 2017; Denyer et al., 2019). We opted for the Takara ICELL8 system, a sensitive nanowell-based approach that includes standardized lysis of nuclei by detergents and a freeze-thaw cycle (Goldstein et al., 2017). One of the main advantages of this system is that it allows for manual selection of single-nucleus-containing wells, as well as the visual inspection and selection of intact nuclei, thereby introducing additional quality control. Using SMARTer ICELL8 3' chemistry, we prepared DNA libraries for short paired-end sequencing using fresh, snap-frozen *Arabidopsis* seedlings.

### snRNA-seq performance in *Arabidopsis* seedlings

To establish the method, we used 10-day-old *A. thaliana* seedlings. Seedlings feature diverse plant structures, including the primary root, hypocotyl, and cotyledons. This allowed us to evaluate the ability of the method to recover the transcriptomes of diverse tissue types and to know if we could apply this method to each of these tissue types in one single experiment. A total of 3348 nuclei were obtained from three biological replicates, with an average of 2802 expressed genes per nucleus and more than 90% of their reads mapped to the reference genome. More than 92% of the nuclei were found with less than 5% mitochondrial reads (Figure S3), indicating a very low level of



**Figure 1.** Single-nucleus RNA sequencing. (a) Schematic overview of snRNA-seq experimental strategy. (b) UMAP plot and clustering analysis of *Arabidopsis* seedling samples (three biological replicates, 12 clusters, 2871 nuclei in total). (c) Barplot showing a similar proportion of nuclei per cluster across the three replicates (the color code used to identify clusters is the same in panels (b) and (c)). (d) Correlation ( $R = 0.74$ ) between snRNA-seq and bulk RNA-seq (both with three biological replicates), indicating that snRNA-seq is able to recover similar transcriptomes as bulk RNA-seq

damaged nuclei. See Data S1 for more detailed statistics of the read mapping step. A Pearson correlation coefficient of 0.9 was observed among the biological replicates,

indicating high reproducibility of the method (Figure S4a,b). A good reproducibility was also observed between snRNA-seq and bulk RNA-seq (Pearson correlation of 0.74,

Figure 1d), even though snRNA-seq data comprise the nuclear transcriptome while bulk RNA-seq data comprise the nuclear and cytoplasmic transcriptome. This indicates that the method is able to recover the general transcript abundance present in the bulk RNA-seq data. To further characterize any bias introduced by our method, we performed gene ontology enrichment analysis on the genes differentially expressed between snRNA-seq and bulk RNA-seq data (Figure S5). No stress-related terms were found enriched; however, we detected the enrichment of terms related with plastid-localized proteins. Because a large fraction of genes encoding plastid-localized proteins are intron-less, we speculate that this difference in expression may be produced by the difference in splicing efficiency: genes with no or few introns are transported to the cytoplasm and depleted from the nucleus faster than genes with more introns (Boeri et al., 2011). Indeed, we observed a strong link between the number of introns and differential abundance in nuclei versus bulk RNA-seq data (Figure S5b,c). The integrated analysis of the three seedling datasets by Seurat revealed 12 major clusters (Figure 1b). A similar proportion of nuclei from each annotated cluster was observed across the three replicates, again indicating the good reproducibility of the method (Figure 1c). To annotate the major tissue types enriched in each cluster, we first obtained the top 20 marker genes of each cluster (Table S2). Then, the expression of these marker genes was characterized using a set of plant organ-specific bulk RNA-seq datasets (Figure S6) and by plotting the expression of the identified marker genes (see Figure S7 for some examples). For example, cluster 11 shows the highest signal in Transcriptome Variation Analysis Database (TraVaDB) root samples, and was therefore labeled as 'roots'. Since seedlings comprise a large diversity of cell and tissue types, for many of which no cell-type-specific transcriptome data are available, we did not pursue a comprehensive annotation of this dataset. Instead, considering that the majority of published plant scRNA-seq experiments are focused on roots, we investigated the ability of our method to recover the main root cell types from the seedling dataset. A subset of 980 nuclei identified as 'root' in our seedling dataset (Figure 1b) were further analyzed. Fourteen clusters were identified (Figure S8a). A clear overlap was found between the top 20 marker genes of each cluster (Table S2) and the top 500 markers identified in a previous scRNA-seq study of root (Denyer et al., 2019, Figure S8b). For example, 55% of the markers from cluster 8 overlapped with the markers of cluster '10-Trichoblast' from Denyer's study. No overlap bigger than 10% was found with any other cluster, leading us to infer cluster 8 is a cluster of trichoblasts. This result shows that, despite the small number of nuclei (980) compared to 4727 cells from the comprehensive root atlas (Denyer et al., 2019), we were able to

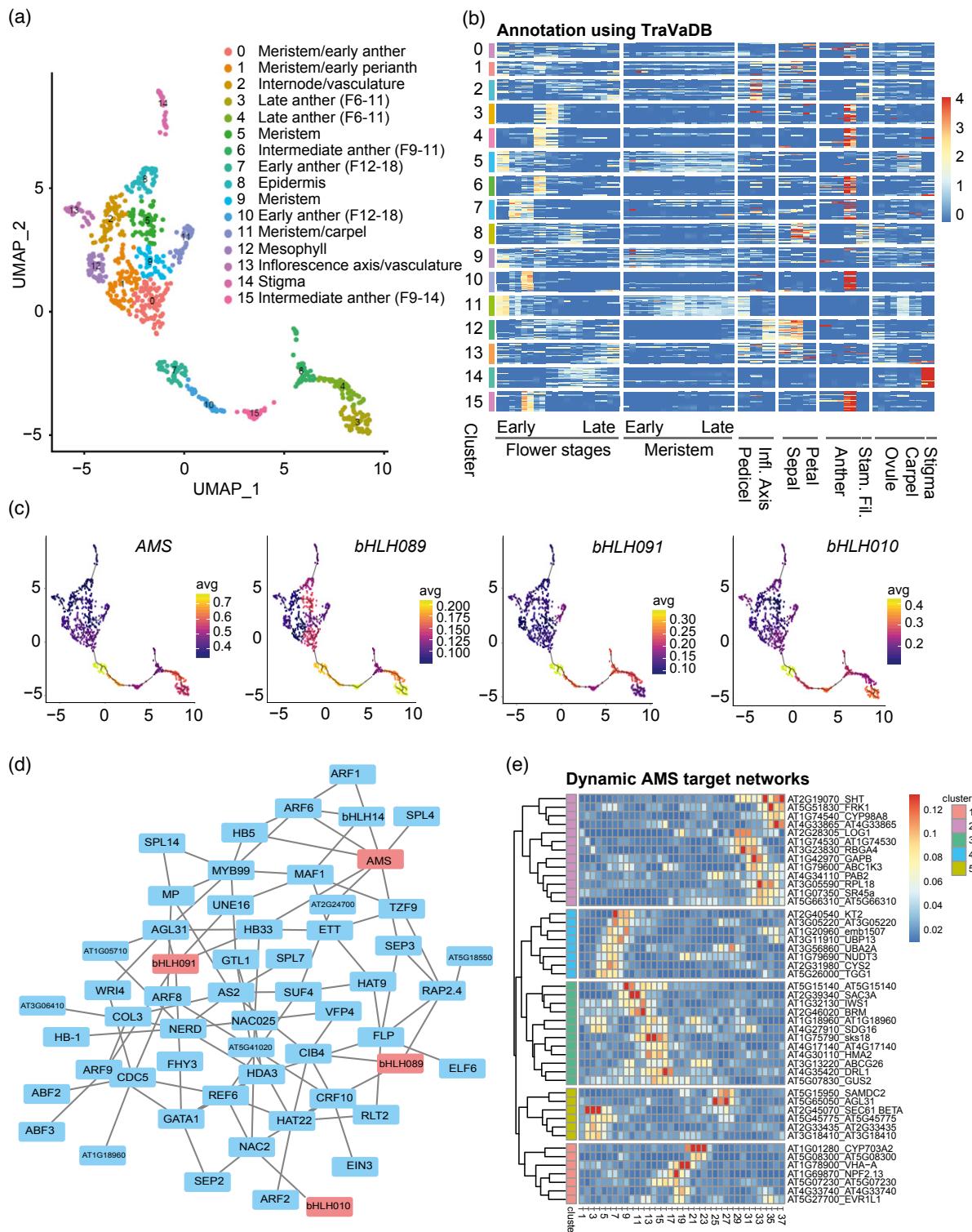
recover the main root cell types and this approach can be used to investigate subsets of cells within complex samples like seedlings.

### **Similarity between snRNA-seq data generated from fixed and unfixed plant material**

To allow for more technical flexibility in our method, i.e., the possibility to simplify the storage of plant samples and maintain *in situ* expression states (Alles et al., 2017), we fixed seedlings using methanol directly after harvesting and performed snRNA-seq as described before (see Data S1 and Figure S3 for the read statistics). After quality control filtering, we obtained a similar number of nuclei (850) and an average number of expressed genes (2292) when using methanol fixation compared to no fixation (1116 nuclei and 2802 genes). A similar nucleus distribution was also observed between fixed and non-fixed samples (Figure S9a). Additionally, an expression correlation of 0.88 and a *P*-value of  $<2.2\text{e-}16$  was observed among both groups of samples (Figure S9b,c), indicating that fixation of the material does not introduce major differences in the number of nuclei and obtained cell types. A good reproducibility was also observed between snRNA-seq of fixed samples and bulk RNA-seq (Pearson correlation of 0.80, Figure S9d).

### **snRNA-seq performance in *Arabidopsis* inflorescences**

To evaluate the performance of snRNA-seq to study cell differentiation, we applied snRNA-seq to *A. thaliana* inflorescences, which cover all stages of flower development prior to anthesis. After quality control filtering, we obtained transcriptomes of 856 nuclei with an average of 2967 expressed genes per nucleus (Figure S10a). The analysis identified 16 clusters corresponding to distinct organs and developmental stages (Figure 2a; Figure S10c). To annotate these clusters with particular cell, tissue, and organ types, we first identified specific marker genes of each cluster (Table S2) and then plotted their expression profiles in the different floral organs and developmental stages obtained from TraVaDB (Figure 2b) as well as in tissue-specific shoot apical meristem RNA-seq data (Yadav et al., 2014, Figure S11). Last, we correlated the gene expression of each cluster with each TraVaDB sample and indicated these values in the UMAP plot (Figure S10c). A major proportion of clusters (37% of the nuclei) were annotated as differentiating anthers at different developmental stages (clusters 0, 3, 4, 6, 7, 10, 15). This can be explained by the fact that anthers comprise a large fraction of tissues (Gómez et al., 2015; Smyth et al., 1990) in developing flowers. Furthermore, anther/pollen has very specific gene expression profiles (Gómez et al., 2015; Smyth et al., 1990), which may facilitate the computational identification of the clusters. Our data captured gene expression dynamics during anther/pollen development from undifferentiated stem cells (cluster 0;



**Figure 2.** Anther development at single-nucleus resolution. (a) UMAP plot and clustering of the snRNA-seq data from *Arabidopsis* flowers before anthesis. (b) Heatmap showing the expression of the top 20 significant marker genes for each cluster. The flower stages considered ranged from F19<sup>+</sup> to F1 and meristem stages considered ranged from M5 to M10, based on the definitions from <http://travadb.org/samples>. (c) Average expression of known representative anther TF genes *AMS*, *bHLH089*, *bHLH091*, and *bHLH010* plotted in the UMAP coordinates. (d) Gene network estimated from cluster 15 (early anther) using GENIE3 (only TFs with more than three targets are shown). (e) Heatmap showing the strength of the interaction between *AMS* and its targets obtained by GENIE3 at different developmental stages. Cells belonging to an anther cluster were ordered by their developmental stage predicted by Monocle 3 pseudotime analysis (Figure S13c) and GRNs were predicted independently for overlapping sets of 50 cells ordered by pseudotime; T1 is the first 50 cells (cluster 0, meristem/early anthers), and T37 is the latest stage (cluster 3, late anther).

Figure 2a) to late anther stages close to organ maturity, prior to anthesis (cluster 3; Figure 2a). This led us to use Monocle 3 to estimate the pseudotime of each anther cell (Figure S13c). When we plotted the average pseudotime of the cells of each anther cluster against the developmental time of each cluster obtained with TraVaDB annotation (Figure S13d), we observed a strong concordance with anther developmental stages. This concordance indicates that the estimated pseudotime of each cell can be used as a proxy of its developmental stage and consequently to study transcriptional dynamics of anther differentiation.

### Gene regulatory trajectories of anther development

We used GENIE3 (Huynh-Thu et al., 2010) to show the utilization of the snRNA-seq data to infer the dynamics of gene regulatory networks (GRNs) during plant development. We reconstructed the GRNs for all clusters that were identified as anthers and estimated the strength of interactions between known transcription factors (TFs) versus all expressed genes. For example, Figure 2d shows the GRN for cluster 15, representing an early anther stage. In our analysis, one of the main master TF genes (with most interactions) was *ABORTED MICROSPORES* (*AMS*), which encodes a known regulator of anther development. When the regulatory dynamics of this TF was investigated in more depth using our data, we found that predicted targets of *AMS* and the related TF genes *bHLH089*, *bHLH091*, and *bHLH010* (Xu et al., 2010; Zhu et al., 2015) were expressed in a highly dynamic manner (Figure 2c,d). *AMS* target genes at early stages were functionally enriched in chromatin remodeling (e.g., *BRAHMA* and *SET DOMAIN PROTEIN 16*) and pollen development (*DIHYDROFLAVONOL 4-REDUCTASE-LIKE1* and *ATP-BINDING CASSETTE G26*) (Figure 2e). Late targets included metabolic enzymes as well as genes associated with RNA regulatory processes. Newly identified marker genes covered the full anther developmental trajectory and are candidates for further mechanistic analyses.

### Validation of cell type marker genes

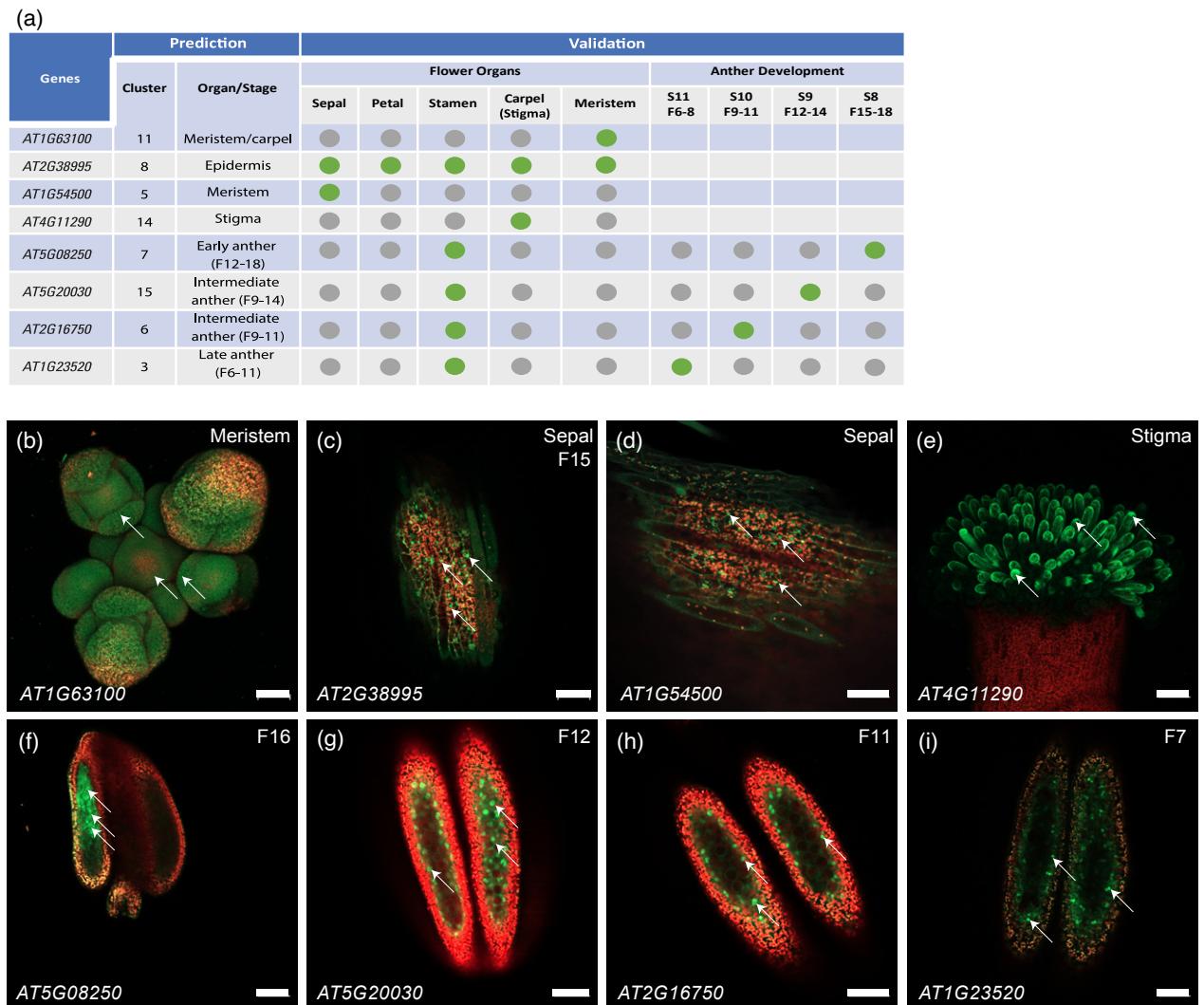
To validate the clustering analysis and dynamic anther transcriptome trajectory, we assessed the expression patterns of genes using promoter::NLS-GFP reporter lines (NLS stands for ‘nuclear localization signal’). We selected 10 previously uncharacterized genes predicted to be specifically or preferentially expressed in one of the clusters: *AT1G63100*, *AT3G51740*, *AT1G54500*, *AT4G11290*, *AT5G08250*, *AT5G20030*, *AT2G16750*, *AT1G23520*, *AT3G05570*, and *AT2G38995*. Eight out of the 10 selected genes showed a specific expression in line with the predictions (Figure 3, Figure S14). Specific expression in the floral meristem and young organ primordia was observed for the genes *AT1G63100* and *AT3G51740* from cluster 11 (Figure 3b,c, Figure S15e,h). The gene *AT4G11290*, from cluster 14, showed highly specific expression in the stigma (Figure 3e,

Figure S15g), while *AT2G38995* from cluster 8 showed epidermis-specific expression and was preferentially expressed in sepals rather than other organs in young flowers (such as flower 15 and flower 16, TraVaDB nomenclature) (Figure 3c, Figure S15f). Additionally, its expression spreads to epidermis of petals, stamens, and carpels in old flowers (such as flower 4 and flower 5) (Figure S15f). The genes *AT5G20030*, *AT5G08250*, *AT1G23520* and *AT2G16750* were found expressed in anthers and showed stage-specific expression as predicted by our analysis shown in Figure 2 (Figure 3f–i, Figure S15a–d). Gene *AT5G08250* from cluster 7, the first cluster of anther lineage, showed very strong expression in young anthers from flower 16 to flower 18 (TraVaDB nomenclature; Figure 3f, Figure S15a); *AT5G20030* from cluster 15, which is an ‘intermediate anther’ cluster, showed a peak in expression in flower 12 (Figure 3g, Figure S15b). *AT2G16750*, from cluster 6, was found strongly expressed in older anthers in flower 10 and flower 11 (Figure 3h, Figure S15c). Finally, *AT1G23520* from cluster 3, the last cluster of anther lineage, was found expressed in old anthers in flower 6 to flower 8 (Figure 3i, Figure S15d). On the other hand, *AT1G54500* was expressed in sepal primordia and developing sepals (Figure 3d), indicating that it is not specific to meristems as predicted for cluster 5. *AT3G05570*, from cluster 6, was not found expressed in any specific tissue.

### Conclusions

Although the PI procedure may affect the plant transcriptome (Figure S1), it has been the main choice for plant single-cell sequencing and has been mostly applied to root samples (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019). Here, we introduced a procedure for plant nuclei isolation that is applicable to different plant tissues and organs such as flowers and leaves, and thus provides a versatile tool for plant single-cell omics. Working with nuclei has the overall advantages of (i) eliminating the dissociation-induced transcriptional responses, (ii) the compatibility with frozen samples, and (iii) the possibility to carry out RNA-seq from individual cells to study cell types, like neurons, in which it is very difficult to recover intact cells (Bakken et al., 2018; Grindberg et al., 2013; Wu et al., 2019). More specifically, for plants, the advantages of using nuclei include (i) the elimination of the need to lyse the cell wall and (ii) the elimination of organelles and vacuoles. As disadvantage, working with nuclei decreases the amount of RNA per individual cell and consequently reduces the sensitivity of transcript detection. Additionally, snRNA-seq cannot be used to study transcriptomes of cells without nucleus (i.e., dividing cells without nuclear membrane or some specific vascular tissue types), nor the translocation of transcripts from one cell to another.

Reporter gene analysis based on promoter fusion to a reporter gene (e.g., *GFP*) measures promoter activity



**Figure 3.** Validation of cluster-specific marker genes with nucleus-targeted GFP reporter lines. (a) Summary of expression specificity validation of selected marker genes. Green dots indicate positive and gray dots indicate negative GFP signals at particular developmental stages or flower organs. Flower numbers (F6–F18) are based on TraVaDB nomenclature; flower developmental stages (S8–S11) are based on Smyth et al., 1990. (b) AT1G63100, Meristem. (c) AT2G38995, Sepal. (d) AT1G54500, Sepal. (e) AT4G11290, Stigma. (f) AT5G08250, anther from flower 16. (g) AT5G20030, anther from flower 12. (h) AT2G16750, anther from flower 11. (i) AT1G23520, anther from flower 7. More confocal images are shown in Figure S15. White arrowheads indicate exemplary GFP signals. Scale bars, 50 µm.

without considering post-transcriptional regulatory processes such as nuclear export, splicing, transcript degradation, or translational efficiency. Besides, it can be challenging to select a genomic region that contains all regulatory elements for correct spatiotemporal expression of a gene. RNA fluorescence *in situ* hybridization experiments could be performed to further validate the results.

Nanowell-based library preparation offers the possibility of visual quality control of individual nuclei, to achieve high numbers (several thousand) of genes per cell and more than a thousand nuclei per run to sensitively detect plant cell (sub-)types. The number of nuclei can potentially be upscaled by using denser and/or larger nanowell

formats to further increase the number of nuclei for sequence analysis. The nanowell-based approach resulting in deep cellular transcriptome data is of particular advantage to identify co-regulated genes and decipher gene networks underlying biological processes of interest. Along with the ever-growing range of nucleic acid sequencing technologies and plant genomics reference databases, single-nucleus genomics procedures are expected to become valuable tools to build maps of all plant cells of developing and adult tissues and to measure cell-type-specific differences in environmental responses to gain novel mechanistic insights into plant growth and physiology (Rhee et al., 2019).

## EXPERIMENTAL PROCEDURES

### Preparation of plant tissues

Plants were grown under long-day conditions (16 h light, 8 h dark) at 22°C in a growth chamber. One gram of 10-day-old *A. thaliana* (Col-0) seedlings growing on ½ MS medium with 1% sucrose or 10 inflorescences from *A. thaliana* growing on soil were collected and snap-frozen in liquid nitrogen at day time. The same procedure was applied for the following samples to test the nucleus isolation pipeline: 10 unopened buds of *P. hybrida* (W115), eight unopened buds of *A. majus*, and 20 fully developed flowers and 1.3 g of leaves of *S. lycopersicum*.

The biological replicates of *A. thaliana* seedlings for snRNA-seq were grown and collected independently on different dates. The three biological replicates of *A. thaliana* seedlings for bulk RNA-seq were grown on different plates and harvested in parallel at once.

### Preparation of nuclei

Frozen tissue was carefully crushed to small pieces in liquid nitrogen using a mortar and pestle and transferred to a gentleMACS M tube (Miltenyi Biotec, <https://www.miltenyibiotec.com>) that was filled with 5 ml of Honda buffer (2.5% Ficoll 400, 5% Dextran T40, 0.4 M sucrose, 10 mM MgCl<sub>2</sub>, 1 µM DTT, 0.5% Triton X-100, 1 tablet/50 ml cComplete Protease Inhibitor Cocktail (Roche, <https://www.roche.com>), 0.4 U µl<sup>-1</sup> RiboLock (Thermo Fisher Scientific, <https://www.thermofisher.com/>), 25 mM Tris-HCl, pH 7.4). This buffer composition enables efficient lysis of cell membranes while keeping the nuclear membranes intact (Moreno-Romero et al., 2017). The M tubes were placed onto gentleMACS Dissociator (Miltenyi Biotec, <https://www.miltenyibiotec.com>) and a specific program (Table S1) was run at 4°C to disrupt the tissue and to release nuclei. The resulting suspension was filtered through a 70-µm strainer and centrifuged at 1000 **g** for 6 min at 4°C. The pellet was resuspended carefully in 500 µl Honda buffer, filtered through a 35-µm strainer, and stained with 3× staining buffer (12 µM DAPI, 0.4 U µl<sup>-1</sup> Ambion RNase Inhibitor (Thermo Fisher Scientific, <https://www.thermofisher.com>), 0.2 U µl<sup>-1</sup> SUPERaseIn RNase Inhibitor (Thermo Fisher Scientific, <https://www.thermofisher.com>) in PBS). Nuclei were sorted by gating on the DAPI peaks using a BD FACS Aria III (BD Biosciences, <https://www.bdbiosciences.com>) (200 000–400 000 events) into a small volume of landing buffer (4% BSA in PBS, 2 U µl<sup>-1</sup> Ambion RNase Inhibitor, 1 U µl<sup>-1</sup> SUPERaseIn RNase Inhibitor). Sorted nuclei were additionally stained with NucBlue from the Invitrogen Ready Probes Cell Viability Imaging Kit (Thermo Fisher Scientific, <https://www.thermofisher.com>, Blue/Red) and then counted and checked for integrity in Neubauer counting chambers. Quality of RNA derived from sorted nuclei was analyzed by an Agilent TapeStation using RNA ScreenTape (Agilent, <https://www.agilent.com>) or alternatively by Agilent's Bioanalyzer 2100 system (Agilent, <https://www.agilent.com>). The seedling nuclei for the three biological replicates for snRNA-seq were isolated and FACS-sorted independently at different dates.

The procedures took approximately 6 min for gentleMACS dissociation, 6 min for centrifugation, approximately 10 min for FACS sorting, and approximately 10 min for nucleus counting, so together with other operations, the whole procedure of nucleus isolation should be completed within 40 min.

### Preparation of single-nucleus libraries using SMARTer ICELL8 single-cell system

The suspension of NucBlue and DAPI co-stained single nuclei (60 nuclei µl<sup>-1</sup>) was distributed to eight wells of a 384-well source

plate (Takara, <https://www.takarabio.com>) and then dispensed into a barcoded SMARTer ICELL8 3' DE Chip (Takara, <https://www.takarabio.com>) by an ICELL8 MultiSample NanoDispenser (MSND, Takara, <https://www.takarabio.com>). Chips were sealed and centrifuged at 500 **g** for 5 min at 4°C. Nanowells were imaged using an ICELL8 Imaging Station (Takara, <https://www.takarabio.com>). After imaging, the chip was placed in a pre-cooled freezing chamber which was stored at -80°C for at least 2 h. CellSelect software was used to support the identification of nanowells that contained a single nucleus. One chip yielded on average between 800 and 1200 nanowells with single nuclei. These nanowells were selected for subsequent targeted deposition of 50 nl RT-PCR reaction mix from the SMARTer ICELL8 3' DE Reagent Kit (Takara, <https://www.takarabio.com>) per nanowell using the MSND. After RT and amplification in a Chip Cycler, barcoded cDNA products from nanowells were pooled by means of the SMARTer ICELL8 Collection Kit (Takara, <https://www.takarabio.com>). cDNA was concentrated using the Zymo DNA Clean & Concentrator kit (Zymo Research, <https://www.zymoresearch.com>) and purified with AMPure XP beads (Beckman Coulter, <https://www.beckman.de>). Afterwards, cDNA was used to construct Nextera XT (Illumina, <https://www.illumina.com>) DNA libraries followed by AMPure XP bead purification. A Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, <https://www.thermofisher.com>), a KAPA Library Quantification Kit (Kapa Biosystems, Roche, <https://www.roche.com>) for Illumina Platforms, and the Agilent High Sensitivity D1000 ScreenTape Assay (Agilent, <https://www.agilent.com>) were used for library quantification and quality assessment. Strand-specific RNA libraries for sequencing were prepared with TruSeq Cluster Kit v3 (Illumina, <https://www.illumina.com>) and sequenced on an Illumina HiSeq 4000 instrument (Illumina, <https://www.illumina.com>, PE100 run).

### Preparation of bulk RNA-seq libraries

Five 10-day-old *A. thaliana* seedlings were collected into 1.5-ml screw-cap tubes with five glass beads, pre-cooled in liquid nitrogen. Samples were homogenized by adding one half of the TRI-Reagent (Sigma-Aldrich, <https://www.sigmaaldrich.com>, 1 ml per 100 mg) to each sample following sample disruption using a Pre-celllys 24 Tissue Homogenizer (Bertin, <https://www.bertin-instruments.com>) instrument for 30 sec at 4000 rpm. After homogenization, total RNA was extracted by adding the second half of the TRI-Reagent and the manufacturer's protocol was followed. To remove any co-precipitated DNA, DNase-I digestion was performed using 1 U DNase-I (NEB, <https://international.neb.com>) in a total volume of 100 µl. Total RNA was cleaned up by LiCl precipitation using 10 µl 8 M LiCl and 3 volumes 100% ethanol and overnight incubation at -20°C, followed by centrifugation at 4°C and 17 900 **g** for 30 min and two washing steps with 70% ethanol. The RNA pellet was dried on ice for 1 h, resuspended in 40 µl DEPC water, and incubated at 56°C for 5 min. Quality of total RNA was analyzed by an Agilent TapeStation using RNA ScreenTape (Agilent, <https://www.agilent.com>). Concentration was measured by a Qubit RNA BR Assay Kit (Thermo Fisher Scientific, <https://www.thermofisher.com>). One microgram of total RNA was used for RNA library preparation with the TruSeq® Stranded mRNA Library Prep kit (Illumina, <https://www.illumina.com>), following the manufacturer's protocol. Quality and fragment peak size were checked by a TapeStation using D1000 ScreenTape (Agilent, <https://www.agilent.com>). Concentration was measured by the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, <https://www.thermofisher.com>). Three replicates, composed of five seedlings each, were used separately throughout the whole procedure. Strand-specific RNA libraries were prepared using the TruSeq Stranded mRNA library

preparation procedure and the three replicates were sequenced on an Illumina NextSeq 500 instrument (Illumina, <https://www.illumina.com>, PE75 run). The three biological replicates of bulk RNA-seq libraries were generated independently in parallel at once.

### Data pre-processing

Raw sequencing files (bcl) were demultiplexed and fastq files were generated using Illumina bcl2fastq software (v2.20.0). The command-line version of the ICELL8 mappa analysis pipeline (demuxer and analyzer v0.92 with STAR v2.7.2a) was used for data pre-processing and read mapping. Mappa\_demuxer assigned the reads to the cell barcodes present in a pre-defined list of barcode sequences. Read trimming, genome alignment (*A. thaliana* reference genome: TAIR10), counting, and summarization were performed by mappa\_analyzer with default parameters. mappa\_analyzer generates a gene matrix containing the gene counts for each barcode, with the genes in the rows and barcodes/cells in the columns. Each entry in the matrix is an expression value representing the expression of a gene (i) for a barcode/cell (j). An html report containing the read statistics for each snRNA-seq library, including the number of reads per cell, reads mapped to the reference genome, and numbers of exons, introns, and intergenic, ribosomal, and mitochondrial reads, was created using hanta software from the ICELL8 mappa analysis pipeline (Data S1).

### Quality control and data analysis

The analysis started by removing reads with barcodes representing the negative and positive controls included in all Takara Bio NGS kits. For the seedling samples, only genes encoded in the nucleus were used. Seurat v3 (Butler et al., 2018; Stuart et al., 2019) was used to filter viable nuclei by removing (i) genes detected in less than three nuclei and (ii) nuclei with less than 200 genes. The Seurat *SCTransform* normalization method was performed for each one of the seedling replicates separately. Data from three seedling replicates were integrated using *PrepSCTIntegration*, *FindIntegrationAnchors*, and *IntegrateData* functions. After running *RunPCA* (default parameters), we performed UMAP embedding using *runUMAP* with the following parameters: *dim*=1:12, *n.neighbors*=10, *min.dist*=.1, *metric*="correlation". Clustering was done with the *FindNeighbors* (default parameters) and *FindClusters* functions using the smart local moving (SLM) algorithm, *resolution*=0.1, and *n.iter*=100. Differentially expressed genes were found using the *FindAllMarkers* function and the 'wilcox' test with *logfc.threshold*=0.25 and *min.pct*=0.25. The sub-clustering analysis of roots was performed using the *subset* function and the seedling clusters containing root cells (clusters 3, 5, 6, 8, 10, and 11; Figure 1b). *SCTransform*, *RunUMAP*, *FindNeighbors*, and *FindClusters* with *dims*=1:10 and *resolution*=1 (other parameters as previously indicated) were re-run after sub-setting the data and *FindAllMarkers* was applied to the RNA assay (normalized counts) to find the differentially expressed genes across the sub-clusters with the 'wilcox' test, with *logfc.threshold*=0.25. For the analysis of fixed seedling samples, we used the same Seurat parameters used for the unfixed samples.

For the analysis of flower snRNA-seq samples (900 nuclei), only nuclear genes were used. Nuclei with (i) less than 10 000 reads, (ii) less than 500 genes containing 10 reads, or (iii) at least one gene covering more than 10% of the reads of a particular nucleus were filtered out. In addition, genes with less than 10 reads in at least 15 nuclei were also removed. The filtering step resulted in a dataset containing 856 nuclei and 14 690 genes. Seurat *SCTransform* normalization was applied to the filtered data using all genes

as *variable.features* with *method*='nb' and *min\_cells*=5. We used the *JackStraw* function in Seurat to estimate the optimal number of PCAs to be used in the analysis (Figure S10b). After calculating the first 12 PCAs with *RunPCA*, we performed UMAP embedding using *runUMAP* with parameters *n.neighbors*=10, *min.dist*=.1, *metric*="correlation", and *umap.method*="umap-learn". Clustering was done with *FindNeighbors* (default parameters) and *FindClusters* function using the SLM algorithm, *resolution*=1.15, and *n.iter*=100. Markers genes were found with the function *FindAllMarkers* using the 'wilcox' test and *min.pct*=0.25.

### Annotation

Annotation of the seedling and flower clusters was performed by visualizing the expression of the top 20 marker genes of each cluster on tissue- and stage-specific transcriptomes of TraVaDB (<http://travadb.org>, Klepikova et al., 2016). For the annotation of the flower clusters, the floral-meristem-specific expression datasets from (Tian et al., 2019) and from (Yadav et al., 2014) were also used.

### Reproducibility and correlation

To assess the reproducibility of our method, we compared the pooled number of reads overlapping each gene of each seedling replicate against one another in  $\log_2$  space. The same was done to verify the similarity between unfixed and fixed seedling datasets.

The correlation between bulk and snRNA-seq datasets (with three biological replicates in each dataset) was investigated by comparing the average number of reads overlapping each gene in the snRNA-seq against bulk RNA-seq datasets. Expression of bulk RNA-seq data was quantified with RNA-seq with RSEM tool (Li and Dewey, 2014).

### Network analysis

GENIE3 (Huynh-Thu et al., 2010) was used to infer gene networks starting from the normalized expression data obtained from Seurat for each cluster independently with parameter *nTrees*=1000 and using the list of DNA-binding proteins obtained from TAIR ([www.arabidopsis.org](http://www.arabidopsis.org)) as regulators. Genes expressed in less than 33% of the nuclei in a particular cluster were removed. Only the top 10 000 interactions were kept. DNA-binding proteins with less than 10 predicted targets were also removed. Dynamics of the gene network through anther development were obtained by the following approach. First, all nuclei were ordered by their estimated developmental pseudotime using Monocle 3 (Trapnell et al., 2014) and using cluster 0 (meristem/early anther) as root cluster. Second, gene networks were estimated with GENIE, as described previously, using groups of non-overlapping sets of 50 nuclei that were previously ordered by their developmental pseudotime.

### Generation and confocal imaging of reporter lines

To validate expression specificity of the marker genes from our snRNA-seq approach, promoter::NLS-GFP reporter lines were generated. The marker genes for validation were chosen from the pool of cluster-specific marker genes ( $P < 0.05$ ) that were not previously characterized in the literature (unknown marker genes). The genomic promoter region upstream of the ATG and until the closest neighboring gene was amplified by PCR and introduced into the entry vector pCR8:GW:TOPO by TA cloning (primers used for PCR are listed in Table S3). The LR reactions were performed with the binary vector pGREEN:GW:NLS-GFP (Smacniak et al., 2017) to generate GFP transcriptional fusions to an NLS peptide.

All reporter constructs were transformed into the *A. thaliana* Col-0 background, and multiple independent lines per construct were analyzed under a Zeiss LSM800 laser-scanning confocal microscope. Different floral organs were dissected and screened for the GFP signal by confocal microscopy under 20 $\times$  and 63 $\times$  magnification objectives. Auto-fluorescence from chlorophyll was observed to give an outline of the flower organs. A 488-nm laser was used to excite GFP and chlorophyll and emissions were captured using photomultiplier tubes set at 410–530 and 650–700 nm. Z-stack screens were performed for the floral meristem and stigma tissues to give a 3D structure visualization.

## ACKNOWLEDGMENTS

We thank Solenne Bourdier and Dijun Chen for their support. The work was supported by the European Commission (FP-7, grant agreement no. 262055 to SS, ESGI and European Union's Horizon 2020 research and innovation programme, grant agreement no. 824110 to SS, EASI-Genomics) and DFG (grant no. KA 2720/5-1 to XX and KK). Open Access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Table S1.** Program steps for gentle tissue disruption and nucleus release on the gentleMACS Dissociator and the instrument-specific commands. The instrument-specific tubes have a stator and rotor. The latter can be moved at a certain speed (*rpm* = revolutions per minute) for a certain time (time in s) in several rounds (*loops*).

**Table S2.** List of marker genes in *Arabidopsis thaliana* seedlings and flowers.

**Table S3.** List of primers used for PCR.

**Data S1.** Summary and read statistics of the snRNA-seq and bulk RNA-seq data.

**Figure S1.** Effect of protoplast isolation (PI) on root scRNA-seq experiments.

**Figure S2.** Quality of isolated nuclei from different plant tissues/species.

**Figure S3.** Summary of snRNA-seq seedling datasets.

**Figure S4.** Reproducibility of snRNA-seq.

**Figure S5.** Expression bias of snRNA-seq compared to cytoplasmic bulk RNA-seq.

**Figure S6.** Annotation of seedling clusters using TraVaDB and shoot apical meristem domains.

**Figure S7.** Gene expression of representative seedling marker genes.

**Figure S8.** Analysis of a subset of root nuclei derived from snRNA-seq seedling dataset.

**Figure S9.** Correlation between fixed and unfixed seedling replicates and between fixed and bulk RNA-seq replicates.

**Figure S10.** Single-nucleus transcriptome analysis of *A. thaliana* flower development.

**Figure S11.** Annotation of flower clusters using the cell-specific shoot apical meristem population bulk RNA-seq dataset.

**Figure S12.** Expression of selected marker genes of flower with known function.

**Figure S13.** Temporal trajectory of the floral snRNA-seq dataset.

**Figure S14.** Gene expression of the validated marker genes of flowers illustrated in Figure 3.

**Figure S15.** Validation of cluster-specific marker genes of flowers with transcriptional reporter lines.

## REFERENCES

- Alles, J., Karaikos, N., Praktiknjo, S.D., Grosswendt, S., Wahle, P., Ruffault, P.-L. *et al.* (2017) Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biology*, **15**, 1–14.
- Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B. *et al.* (2018) Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One*, **13**, 1–24.
- Boeri, M., Verri, C., Conte, D., Roz, L., Modena, P., Facchinetto, F. *et al.* (2011) MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 3713–3718.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, **36**, 411–420.
- Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R. *et al.* (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*, **357**, 661–667.
- Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K. & Timmermans, M.C.P. (2019) Spatiotemporal developmental trajectories in the arabidopsis root revealed using high-throughput resource spatiotemporal developmental trajectories in the arabidopsis root revealed using. *Developmental Cell*, **48**, 840–852.e5. <https://doi.org/10.1016/j.devcel.2019.02.022>
- Efroni, I. & Birnbaum, K.D. (2016) The potential of single-cell profiling in plants. *Genome Biology*, **17**, 65.
- Efroni, I., Ip, P.-L., Nawy, T., Mello, A. & Birnbaum, K.D. (2015) Quantification of cell identity from single-cell gene expression profiles. *Genome Biology*, **16**, 9.
- Efroni, I., Mello, A., Nawy, T., Ip, P.-L., Rahni, R., DelRose, N. *et al.* (2016) Root regeneration triggers an embryo-like sequence guided by hormonal interactions. *Cell*, **165**, 1721–1733.
- Fischer, C., Metsger, M., Bauch, S., Vidal, R., Böttcher, M., Grote, P. *et al.* (2019) Signals trigger state-specific transcriptional programs to support diversity and homeostasis in immune cells. *Science Signalling*, **12**, 1–17.
- Goldstein, L.D., Chen, Y.-J., Dunne, J., Mir, A., Hubschle, H., Guillory, J. *et al.* (2017) Massively parallel nanowell-based single-cell gene expression profiling. *BMC Genomics*, **18**, 1–10.
- Gómez, J.F., Talle, B. & Wilson, Z.A. (2015) Anther and pollen development: a conserved developmental pathway. *Journal of Integrative Plant Biology*, **57**, 876–891.
- Grindberg, R.V., Yee-Greenbaum, J.L., McConnell, M.J., Novotny, M., O'Shaughnessy, A.L., Lambert, G.M. *et al.* (2013) RNA-sequencing from single nuclei. *Proceedings of the National Academy of Science of United States of America*, **110**, 19802–19807.
- Huynh-Thu, V.A., Irthrum, A., Wehenkel, L. & Geurts, P. (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, **5**, 1–10.
- Jean-Baptiste, K., McFainlie-Figueroa, J.L., Alexandre, C.M., Dorrrity, M.W., Saunders, L., Bubb, K.L. *et al.* (2019) Dynamics of gene expression in single root cells of arabidopsis thaliana. *The Plant Cell*, **31**, 993–1011.
- Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D. & Penin, A.A. (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *The Plant Journal*, **88**, 1058–1070.
- Lareau, C.A., Ma, S., Duarte, F.M. & Buenrostro, J.D. (2020) Inference and effects of barcode multiplets in droplet-based single-cell assays. *Nature Communications*, **11**, 1–9.
- Li, B. & Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
- Moreno-Romero, J., Santos-González, J., Hennig, L. & Köhler, C. (2017) Applying the INTACT method to purify endosperm nuclei and to

- generate parental-specific epigenome profiles. *Nature Protocols*, **12**, 238–254.
- Nelms, B. & Walbot, V.** (2018) Defining the developmental program leading to meiosis in maize. *bioRxiv*, **56**, 52–56.
- Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E. et al.** (2017) The human cell atlas. *Elife*, **6**, e27041.
- Rhee, S.Y., Birnbaum, K.D. & Ehrhardt, D.W.** (2019) Towards building a plant cell atlas. *Trends in Plant Science*, **24**, 303–310.
- Ryu, K.H., Huang, L., Kang, H.M. & Schiefelbein, J.** (2019) Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiology*, **179**, 1444–1456.
- Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M. et al.** (2019) High-throughput single-cell transcriptome profiling of plant cell types. *Cell Reports*, **27**, 2241–2247.e4.
- Smaczniak, C., Muñoz, J.M., Chen, D., Angenent, G.C. & Kaufmann, K.** (2017) Differences in DNA binding specificity of floral homeotic protein complexes predict organ-specific target genes. *The Plant Cell*, **29**, 1822–1835.
- Smyth, D.R., Bowman, J.L. & Meyerowitz, E.M.** (1990) Early flower development in *Arabidopsis*. *The Plant Cell*, **2**, 755–767.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M. et al.** (2019) Comprehensive integration of single-cell data. *Cell*, **177**, 1888–1902.e21.
- Thibivilliers, S., Anderson, D. & Libault, M.** (2020) Isolation of plant root nuclei for single cell RNA sequencing. *Current Protocols in Plant Biology*, **5**, 1–12.
- Tian, C., Wang, Y., Yu, H., He, J., Wang, J., Shi, B. et al.** (2019) A gene expression map of shoot domains reveals regulatory mechanisms. *Nature Communications*, **10**, 1–12.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M. et al.** (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, **32**, 381–386.
- Wu, H., Kirita, Y., Donnelly, E.L. & Humphreys, B.D.** (2019) Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: rare cell types and novel cell states revealed in fibrosis. *Journal of the American Society of Nephrology*, **30**, 23–32.
- Xu, J., Yang, C., Yuan, Z., Zhang, D., Gondwe, M.Y., Ding, Z. et al.** (2010) The ABORTED MICROSPORES regulatory network is required for post-meiotic male reproductive development in *Arabidopsis thaliana*. *The Plant Cell*, **22**, 91–107.
- Yadav, R.K., Tavakkoli, M., Xie, M., Girke, T. & Venugopala, R.G.** (2014) A high-resolution gene expression map of the *arabidopsis* shoot meristem stem cell niche. *Development*, **141**, 2735–2744.
- 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. *Molecular Plant*, **12**, 648–660.
- Zhu, E., You, C., Wang, S., Cui, J., Niu, B., Wang, Y. et al.** (2015) The DYT1-interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for *Arabidopsis* anther development and transcriptome. *The Plant Journal*, **83**, 976–990.