Single-nuclei RNA-sequencing of plants

Daniele Y. Sunaga-Franze[1,#], Jose M. Muino[2,#], Caroline Braeuning[1,#], Xiaocai Xu[3,#], Minglei Zong[3], Cezary Smaczniak[3], Wenhao Yan[3], Cornelius Fischer[1], Ramon Vidal[1], Magdalena Kliem[1], Kerstin Kaufmann[3,*], Sascha Sauer[1,*]

[1] Max Delbrück Center for Molecular Medicine/Berlin Institute of Health, Berlin, Germany
[2] Humboldt-Universität zu Berlin, Institute of Biology, Systems Biology of Gene Regulation, Berlin, Germany
[3] Humboldt-Universität zu Berlin, Institute of Biology, Plant Cell and Molecular Biology, Berlin, Germany

*Correspondence to: kerstin.kaufmann@hu-berlin.de, sascha.sauer@mdc-berlin.de

Abstract

Single-cell genomics has slowly advanced in plant research. Here, we introduce a generic procedure for plant nuclei isolation and nanowell-based library preparation for short-read sequencing. This plant-nuclei sequencing (PN-seq) method allows for analyzing several thousands of genes in thousands of individual plant cells. In this way, we expand the toolset for single-cell genomics in the field of plant biology to generate plant transcriptome atlases in development and environmental response.

Introduction

The fundamental units of life, the cells, can vary tremendously within an organism. The analysis of specialized cells and their interactions is essential for a comprehensive understanding of the function of tissues and biological systems in general. Major biological roles such as growth, development and physiology ultimately gain plasticity from heterogeneity in cellular gene expression [1].

Without precise transcriptional maps of different cell populations, we cannot accurately describe all their functions and underlying molecular networks that drive their activities. Recent advances in single-cell (sc) and in particular single-nucleus (sn) RNA-sequencing have put comprehensive, high-resolution reference transcriptome maps of mammalian cells and tissues on the agenda of international consortia such as the Human Cell Atlas [2].

Similar efforts have been made by the Plant Cell Atlas [3]. Plant tissues and plant cells pose specific challenges compared to mammalian systems [4]. Plant cells are immobilized in a rigid cell wall matrix, which is required to be removed for isolating 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 single-cell RNA-sequencing studies using protoplast isolation (PI) have been published [5,6,7,8,9]. However, it is known that enzymatic digestion of plant cell walls is an important stressor for the plant and thus can introduce artifacts at the transcriptome level. PI-response genes can be identified through an independent bulk RNA-seq experiment to eliminate at least the most strongly affected genes from a scRNA-seq [9], as is shown in the Supplementary Fig. 1. In summary, there is an urgent need for alternative efficient single-cell genomics methods tailored for plant research.

Here, we introduce a single-nucleus sequencing protocol for plants (Fig. 1a; full protocol in Supplementary Materials and Methods). Working with nuclei has the advantage to eliminate organelles and vacuoles, as well as secondary metabolites localized in the cytoplasm that can

interact with RNA. SnRNA-seq experiments have specific challenges, such as lower RNA yield, that need to be overcome by optimized experimental procedures and data analysis strategies [10,11,12,13]. As yet, there is no report of snRNA-seq methodology in plants.

Results and Discussion

Preparation of plant tissue and nuclei

Here we propose a single-nucleus sequencing strategy to detect nucleic acids derived from individual plant cells. The key step of our plant-nuclei sequencing (PN-seq) procedure consists of gentle but efficient isolation of plant nuclei. Plant tissue was frozen, gently physically dissociated and transferred to sucrose-rich, protease and RNAse inhibitor containing Honda buffer to support cell lysis. Cell walls and cell membranes were mechanically disrupted using a gentleMACS Dissociator, keeping the nuclei largely intact as detected by microscopy (Supplementary Fig. 2a, Materials and Methods). Finally, released intact plant nuclei were collected using Fluorescence Activated Cell Sorting (FACS), as demonstrated for a variety of samples including complex seedlings of the model plant *Arabidopsis thaliana*, as well as flowers of *Arabidopsis thaliana*, *Petunia hybrida* and *Antirrhinum majus* (snapdragon) and flowers and leaves from *Solanum lycopersicum* (tomato) (Fig. 1b and Supplementary Fig. 2b and c). The RNA that was isolated from nuclei was of higher quality than RNA that was conventionally purified from plant tissue, as observed by chromatrography analysis (Supplementary Fig. 2d).

Plant-nuclei (PN) sequencing of *Arabidopsis thaliana* seedlings

The next step consists of generating high quality cDNA libraries from the isolated nuclei. In principle, a number of library preparation and sequencing procedures can be combined. Of note, nuclei from *Arabidopsis thaliana* (~2 µm diameter) are much smaller than typical human or mouse nuclei (~10 µm diameter), and may contain only a fraction of the average mammalian nuclei RNA amount. We thus opted for a sensitive nanowell-based approach that includes lysis of nuclei by detergents and a freeze-thaw-cycle. In microdroplet-based single-cell RNA-sequencing methods such as the popular commercial 10x Genomics Chromium procedures, only a relatively mild lysis by detergents can be applied, since reverse transcription (RT) reactions take place in the same environment. Moreover, nanowells allow for selection of single-nucleus-containing wells and exclusion of no-nucleus- and multiple-nuclei-wells, thereby introducing additional quality control (Supplementary Materials and Methods). Using SMARTer ICELL8 3' chemistry, we prepared DNA libraries for short paired-end sequencing.

107 Raw sequencing data were preprocessed with ICELL8 mappa analysis pipeline and the R
108 package Seurat v3 was used for downstream analysis [14]. Global properties of single-nuclei
109 RNA-sequencing libraries including the number of sequenced, barcoded and mapped reads
110 were summarized using hanta software from ICELL8 mappa analysis pipeline (Supplementary
111 Data 1).
112 To validate our method, we set up our protocol using pools of *Arabidopsis thaliana* seedlings
113 (3 biological replicates), which feature diverse plant structures comprising the radicle
114 (embryonic root), the hypocotyl (embryonic shoot), and the cotyledons (seed leaves). On
115 average, we obtained 1,116 nuclei per replicate and 2,802 genes per nucleus at ~220,000 reads
116 per nucleus from these complex samples (Supplementary Data 1, Supplementary Fig. 3).
117 Previous studies using PI of much less complex roots and droplet-based scRNA-seq obtained
118 on average of 2,300 cells and the median number of 4,300 genes per cell [9]. As expected, the
119 lower number of expressed genes per nuclei found by PN-seq can be explained by the fact that
120 plant nuclei were small (~2 µm diameter) and thus contained less RNA. The ICELL8 system
121 provides power for unbiased isolation of up to 1,800 single cells on a single chip, which can be
122 upscaled by denser and/or bigger nanowell formats. As shown below, we achieved a reasonable
123 number to perform further biology analysis, indicating the potential of PN-seq as a broadly
124 applicable method.
125 Next, we analyzed the reproducibility of PN-seq. Correlation between the replicates was
126 assessed using MA plots and Pearson's correlation, which ranged between 0.90 and 0.91
127 (Supplementary Fig. 4a-b). In silico pooling of our PN-seq data and subsequent correlation with
128 gene expression data derived from conventional bulk RNA-sequencing resulted in Pearson
129 correlation of 0.74 (Supplementary Fig. 4c, Supplementary Materials and Methods). This
130 correlation is consistent with correlation coefficients found in previous publications (ranging
131 from 0.7 to 0.85) [15] and therefore we found good agreement between both experiments. As
132 shown in Supplementary Fig. 5, gene expression differences between both experiments were
133 similar across all the chromosomes. No bias versus expression of specific sets of genes was
134 observed.
135
136 Main organs and cell types of seedlings
137 The three seedling replicates were subsequently assessed using Seurat integration analysis,
138 which initially revealed 13 distinct nuclei clusters (n=2,871) (Supplementary Fig. 6a,
139 Supplementary Materials and Methods). Similar distribution of nuclei across clusters was
140 observed, indicating that highly similar nuclei populations were recovered in each biological

141 replicate (Fig. 1c-d). The Wilcoxon rank sum test was applied to identify the significant marker
142 genes of the clusters (Supplementary Table 1). Annotation of the initial 13 clusters was done
143 based on the TraVaDB (Transcriptome Variation Analysis Database - http://travadb.org [17],
144 Supplementary Materials and Methods), a plant gene expression resource, which resulted in 10
145 cluster labels that could be roughly classified into expected main basic organ types of seedlings:
146 Leaves/Cotyledons (n=643 nuclei), Shoot meristems (n=180 nuclei), Hypocotyls (n=393
147 nuclei), Root apices (n=192 nuclei), Vasculature (Leaves/Roots) (n=342 nuclei), Roots (n=267
148 nuclei), Leaves (n=152 nuclei), Mature roots (n=27 nuclei), Roots/Hypocotyls (n=136 nuclei)
149 and non-determined nuclei (n=539 nuclei) (Fig. 1e). The heatmap showing the expression of
150 marker genes recovered from TraVaDB is displayed in Supplementary Fig. 6b. The entire
151 annotation process is illustrated in Supplementary Fig. 6c. The clusters contain similar numbers
152 of nuclei from each replicate, corroborating again the reproducibility of the method (Fig. 1f).
153 Next, we performed in depth analysis of root nuclei using 964 nuclei derived from root tissue
154 of the seedlings (Root apices = 192 nuclei, Vasculature (Leaves/Roots) = 342 nuclei, Roots =
155 267 nuclei), Mature roots = 27 nuclei, Roots/Hypocotyls = 136 nuclei). The 964 nuclei were
156 reorganized into 15 sub-clusters. The marker genes of the predicted 15 sub-clusters were
157 compared to the list of markers from recently published atlas of the *Arabidopsis* root [5,9]
158 (Supplementary Table 1, Supplementary Materials and Methods). PN-seq faithfully recovered
159 major root cell types from our complex seedling dataset: mature, cortex/endodermis, stele,
160 trichoblast, atrichoblast, endodermis and xylem (Supplementary Fig. 7). When looking at the
161 expression of those genes in the RNA-seq based TraVa dataset, we notably observed the
162 clusters 0, 1, 4, 10, 11, 12 and 14, in particular cluster 10 (trichoblast), enriched with multiple
163 marker genes known from flowers, indicating i) the higher complexity of cell and organ type
164 of inflorescence data compared to roots, ii) the still poor spatiotemporal resolution currently
165 available, as TravA dataset is not cell-type specific and iii) the basic principle of the biology in
166 which a gene may play role in multiple biological processes (Supplementary Table 1).
167
168 High similarity between fixed and unfixed material
169 In order to allow for more technical flexibility, i.e. the possibility to simplify the harvest and
170 storage of plant samples, we fixed seedlings using methanol prior to our workflow
171 (Supplementary Materials and Methods). High similarity across the samples from fixed and
172 unfixed procedures was observed (Supplementary Fig. 8a-c) and the output was similar to the
173 unfixed procedure: 850 nuclei and 2,292 genes (mean) per nucleus (Supplementary Fig. 3),
174 implying that fixation of the material does not introduce major differences in the results. The

175 option of PN-seq to process frozen or methanol-fixed materials offers an additional advantage
176 over protoplast-based procedures, since the latter requires immediate processing of fresh,
177 unfrozen plant material.
178
179 Developmental flower stages covered by *Arabidopsis thaliana* inflorescences
180 Next, to study cell differentiation in plants, we applied PN-seq to *Arabidopsis thaliana*
181 inflorescences covering all developmental stages prior to anthesis. After quality control
182 filtering, we obtained transcriptomes of 856 nuclei with an average number of 2,967 expressed
183 genes per nucleus (Supplementary Fig. 9a), and with 14,690 genes expressed in at least five
184 nuclei. The analysis identified 15 clusters corresponding to distinct organs and developmental
185 stages (Fig. 2a; Supplementary Fig. 9b). In order to annotate these clusters with particular cell
186 types, we first identified specific marker genes of each cluster (Supplementary Table 1), then
187 plotted their expression profiles in the different floral organs and developmental stages obtained
188 from TraVaDB [17] (Fig. 2b). Last, we correlated the gene expression of each cluster with each
189 TraVaDB sample and indicated these values in the UMAP plot (Supplementary Fig. 9c). A
190 major proportion of clusters (37% of the nuclei population) were annotated as differentiating
191 anthers at different developmental stages (clusters 3, 4, 6, 7, 10, 15). This can be explained by
192 the fact that six anthers per developing flower comprise a large fraction of floral tissues [18,19].
193 Furthermore, anthers/pollen express unique genes [18,19] which facilitates the bioinformatic
194 identification of the clusters. Our data captured the developmental gene expression profiles
195 during anther/pollen development from undifferentiated stem cells (cluster 0; Fig. 2) to late
196 anther stages close to organ maturity, prior to anthesis (cluster 3; Fig. 2). Pseudotime analysis
197 using Monocle 3 showed a strong concordance with anther developmental stages
198 (Supplementary Fig. 10).
199
200 Regulatory link during anther and pollen development
201 Next, we analyzed how the gene regulatory network dynamically changes during anther and
202 pollen development. Gene regulatory networks (GRNs) were inferred from transcriptome data
203 using GENIE3 [20] to estimate the strength of interaction between known transcription factors
204 (TFs) versus all the expressed genes for each cluster independently (Supplementary Fig. 11).
205 One of the most connected TFs representing anthers was *ABORTED MICROSPORES* (*AMS*).
206 *AMS* and the related TF genes *bHLH089*, *bHLH090* and *bHLH010* [21,22] were expressed in a
207 highly dynamic manner (Fig. 2c, 2d). AMS target genes at early stages were functionally
208 enriched in chromatin remodeling (e.g. *BRAHMA*; *SET DOMAIN PROTEIN 16*) and pollen
209 development (*DIHYDROFLAVONOL 4-REDUCTASE-LIKE1; ATP-BINDING CASSETTE*

210 *G26*) (Fig. 2e). Late targets included metabolic enzymes as well as genes associated with RNA-
211 regulatory processes. Newly identified marker genes covered the full anther developmental
212 trajectory (Supplementary Fig. 12) and are candidates for further mechanistic analyses.
213 To further validate our clustering analysis, we assessed the expression patterns of genes using
214 promoter: GFP -NLS reporter lines. We selecteded genes with significant specificity and
215 unknown function (Supplementary Fig. 13). In general, all genes showed expression in line
216 with predictions. Specific expression in the floral meristem was observed for genes
217 (AT1G63100, AT3G51740) from clusters 11, while genes AT4G11290 selected from cluster
218 14 showed highly specific stigma tissue expression. AT2G38995 was expressed in the sepals
219 and petals, as expected for a marker from cluster 8, and it also showed slight expression in
220 anthers. Genes AT5G20030, AT1G23520 and AT2G16750 were expressed in anthers and
221 showed stage specificities that correlated with our analysis. Gene AT5G20030 from cluster 15,
222 which is early anther cluster, showed expression in young anthers from flower 9 to flower 14.
223 Finally, genes AT1G23520 and AT2G16750 were found in clusters of older anthers and indeed
224 showed expression in old flowers (from 6-8 and 4-5), respectively.
225 In conclusion, PN-seq allowed for efficiently building transcriptome maps of plant samples and
226 for studying at the level of individual cells dynamic GRNs during development, and revealed
227 cell-type and stage-specific TF target pathways in an unprecedented manner.
228
229 Conclusion
230 Althought it is known that protoplast isolation (PI) procedure can significantly affect the plant
231 transcriptome, it has been the basic choice for plant single-cell sequencing and had been mostly
232 applied to roots samples [7,8,9,23,24]. Here, we introduced PN-seq that can be applied to analyze
233 nucleic acids in bulk or in individual cells. Our new PN-seq methodology - based on efficient
234 isolation of nuclei - is directly and easily applicable to a broad range of different plant tissues
235 such as complex seedlings, flowers and leaves, and thus provides a versatile tool for multiple
236 plant studies. In principle, various library preparation and sequencing methods can be combined
237 with our nuclei isolation procedure.
238 Nanowell-based library preparation offered the possibility of visual quality control of
239 individual nuclei, achieved high numbers of several thousand genes per cell and more than
240 thousand nuclei per run to sensitively detect plant cell (sub)types. The number of nuclei can
241 potentially be upscaled by using denser and/or larger nanowell-formats to further increase the
242 number of nuclei for sequence analysis. The here applied nanowell-based approach resulting in
243 deep cellular transcriptome data was of particular advantage to identify co-regulated genes and

244 decipher gene networks underlying biological processes of interest. Along with the ever
245 growing arsenal of nucleic acids sequencing technologies and plant genomics reference
246 databases, single-nuclei genomics procedures are expected to become valuable tools to build
247 maps of all plant cells of developing and adult tissues, and to measure cell-type specific
248 differences in environmental responses to gain novel mechanistic insights into plant growth and
249 physiology [3].
250
251 **Materials and Methods**
252 Preparation of plant tissues
253 One gram of *Arabidopsis thaliana* (Col-0) seedlings or 10 inflorescences were collected and
254 snap-frozen in liquid nitrogen. The same procedure was applied for the following samples: 10
255 unopened buds of *Petunia hybrida* (W115), 8 unopened buds of *Antirrhinum majus*, 20 fully
256 developed flowers and 1.3 g leaves of *Solanum lycopersicum*. A step-by-step protocol for the
257 preparation of plant tissues, nuclei and single-nucleus libraries as well as the steps for the data
258 pre-processing analysis can be found at Protocol Exchange.
259
260 Preparation of nuclei
261 Frozen tissue was carefully crushed to small pieces in liquid nitrogen using a mortar and a pestle
262 and transferred to a gentleMACS M tube that was filled with 5 ml of Honda buffer (2.5 % Ficoll
263 400, 5 % Dextran T40, 0.4 M sucrose, 10 mM $MgCl_2$, 1µM DTT, 0.5% Triton X-100, 1
264 tablet/50mL cOmplete Protease Inhibitor Cocktail, 0.4 U/µl RiboLock, 25 mM Tris-HCl, pH
265 7.4). The M tubes were put onto a gentleMACS Dissociator and a specific program
266 (Supplementary Table 3) was run at 4°C to disrupt the tissue and to release nuclei. The resulting
267 suspension was filtered through a 70 µm strainer and centrifuged at 1000 g for 6 min at 4°C.
268 The pellet was resuspended carefully in 500 µl Honda buffer, filtered through a 35 µm strainer
269 and stained with 3x staining buffer (12 µM DAPI, 0.4 U/µl Ambion RNase Inhibitor, 0.2 U/µl
270 SUPERaseIn RNase Inhibitor in PBS). Nuclei were sorted by gating on the DAPI peaks using
271 a BD FACS Aria III (200,000 – 400,000 events) into a small volume of landing buffer (4%
272 BSA in PBS, 2 U/µl Ambion RNase Inhibitor, 1 U/µl SUPERaseIn RNase Inhibitor). Sorted
273 nuclei were additionally stained with NucBlue from the Invitrogen Ready Probes Cell Viability
274 Imaging Kit (Blue/Red), then counted and checked for integrity in Neubauer counting
275 chambers. Quality of RNA derived from sorted nuclei was analyzed by Agilent TapeStation
276 using RNA ScreenTape or alternatively by Agilent's Bioanalyser 2100 system.
277

278 Preparation of single-nucleus libraries using SMARTer ICELL8 Single-Cell System
279 The NucBlue and DAPI co-stained single-nuclei suspension (60 cells/µL) was distributed to
280 eight wells of a 384-well source plate (Cat. No. 640018, Takara) and then dispensed into a
281 barcoded SMARTer ICELL8 3' DE Chip (Cat. No. 640143, Takara) by an ICELL8
282 MultiSample NanoDispenser (MSND, Takara). Chips were sealed and centrifuged at 500 g for
283 5 min at 4°C. Nanowells were imaged using the ICELL8 Imaging Station (Takara). After
284 imaging, the chip was placed in a pre-cooled freezing chamber, and stored at −80 °C for at least
285 2 h. The CellSelect software was used to support identification of nanowells that contained a
286 single nucleus. One chip yielded on average between 800 - 1200 nanowells with single nuclei.
287 These nanowells were selected for subsequent targeted deposition of 50 nL/nanowell RT-PCR
288 reaction mix from the SMARTer ICELL8 3' DE Reagent Kit (Cat. No. 640167, Takara) using
289 the MSND. After RT and amplification in a Chip Cycler, barcoded cDNA products from
290 nanowells were pooled by means of the SMARTer ICELL8 Collection Kit (Cat. No. 640048,
291 Takara). cDNA was concentrated using the Zymo DNA Clean & Concentrator kit (Cat. No.
292 D4013, Zymo Research) and purified with AMPure XP beads. Afterwards, cDNA was used to
293 construct Nextera XT (Illumina) DNA libraries followed by AMPure XP bead purification.
294 Qubit dsDNA HS Assay Kit, KAPA Library Quantification Kit for Illumina Platforms and
295 Agilent High Sensitivity D1000 ScreenTape Assay were used for library quantification and
296 quality assessment. Strand-specific RNA libraries for sequencing were prepared with TruSeq
297 Cluster Kit v3 and sequenced on an Illumina HiSeq 4000 instrument (PE100 run).
298
299 Preparation of bulk libraries
300 Five 10-days-old *Arabidopsis thaliana* seedlings were collected into 1.5 ml screw cap tubes
301 with 5 glass beads, precooled in liquid nitrogen. Samples were homogenized by adding one half
302 of TRI-Reagent (Sigma-Aldrich, 1 ml per 100 mg) to each sample following sample disruption
303 by using the Precellys 24 Lysis & Homogenization instrument for 30 sec and 4000 rpm. After
304 homogenization, total RNA was extracted by adding 2nd half of TRI-Reagent and the protocol
305 was proceeded according to the manufacturer. To remove any co-precipitated DNA, a DNase-
306 I digest was performed by using 1U DNase-I (NEB) in a total volume of 100 µl. Total RNA
307 was cleaned-up by LiCl-precipitation using 10 µl 8 M LiCl and 3 vol 100% Ethanol pa
308 incubating at -20 °C overnight. Following a spin down at 4 °C, 13,000 rpm for 30 min and 2
309 washing steps with 70% Ethanol pa. The RNA pellet was dried on ice for 1 h and resuspended
310 in 40 µl DEPC-H2O incubating at 56 °C for 5 min. Quality of total RNA was analyzed by
311 Agilent TapeStation using RNA ScreenTape or alternatively by Agilent's Bioanalyser 2100

312 system. Concentration was measured by a Qubit RNA BR Assay Kit (Thermo Fisher
313 Scientific). One µg of total-RNA was used for RNA library preparation with Illumina TruSeq®
314 Stranded mRNA Library Prep, following the protocol according to the manufacturer. Quality
315 and fragment peak size were checked by Agilent TapeStation using D1000 ScreenTape or
316 alternatively by Agilent's Bioanalyser 2100 system. Concentration was measured by the Qubit
317 dsDNA BR Assay Kit (Thermo Fisher Scientific). Three replicates, composed by five seedlings
318 each, were used separately through the whole procedure. Strand-specific RNA libraries were
319 prepared using TruSeq Stranded mRNA library preparation procedure and the three replicates
320 were sequenced on an Illumina NextSeq 500 instrument (PE75 run).
321
322 Data pre-processing
323 The overall data analysis workflow is shown in Supplementary Fig. 14. Raw sequencing
324 files (bcl) were demultiplexed and fastq files were generated using Illumina bcl2fastq software
325 (v2.20.0). The command-line version of ICELL8 mappa analysis pipeline (demuxer and
326 analyzer v0.92) was used for the data pre-processing. Mappa_demuxer assigned the reads to
327 the cell barcodes present in the well-list file. Read trimming, genome alignment (*Arabidopsis*
328 *thaliana* reference genome: TAIR10), counting and summarization were performed by
329 mappa_analyzer with the default parameters. A report containing the experimental overview
330 and read statistics for each PN-seq library was created using hanta software from the ICELL8
331 mappa analysis pipeline (Supplementary Data 1). The gene matrix generated by
332 mappa_analyzer was used as input for Seurat v3.
333
334 Quality control and data analysis
335 The downstream analysis started by removing the negative and positive controls included in all
336 Takara Bio's NGS kits. For the seedling samples, R package Seurat v3 was used to filter viable
337 nuclei, removing genes detected in less than 3 nuclei, nuclei with less than 200 genes, nuclei
338 with more than 5% of mitochondria and nuclei with more than 5% of chloroplasts. Seurat
339 *SCTransform* normalization method was performed for each one of the seedling replicates
340 separately. Data from three seedling replicates were integrated using *PrepSCTIntegration*,
341 *FindIntegrationAnchors* and *IntegrateData* functions. After running the *RunPCA* (default
342 parameters), we performed UMAP embedding using *runUMAP* with *dims*=1:20. Clustering
343 analysis was performed using *FindNeighbors* (default parameters) and *FindClusters* function
344 with *resolution*=0.5. Differentially expressed genes were found using *FindAllMarkers* function
345 and "wilcox" test. The sub-clustering analysis of root was performed using the *subset* function

346 and the seedling clusters containing root cells (clusters: 3, 4, 6, 7, 9, 11 and 12; Supplementary
347 Fig. 6a). *SCTransform* and *RunPCA* were re-run after sub-setting the data and subsequently
348 *FindAllMarkers* to find the differentially expressed genes across the sub-clusters, with the
349 "wilcox" test and using the RNA assay (normalized counts). The annotation of the clusters was
350 based on the top 20 markers of each cluster.
351 For the flower PN-seq dataset (900 nuclei), only genes encoded in the nucleus were used
352 (32,548 genes). Nuclei with i) less than 10,000 reads, ii) less than 500 genes containing 10
353 reads, or iii) at least one gene covering more than 10% of the reads of a particular nucleus were
354 filtered out. In addition, genes with less than 10 reads in at least 15 nuclei were also filtered
355 out. The filtering step resulted in a dataset containing 856 nuclei and 14,690 genes. Seurat v3
356 *SCTransform* normalization was applied to the filtered data using all genes as *variable.features*,
357 and with parameters: *method*="nb", and *min_cells*=5. We used the *JackStraw* function in Seurat
358 to estimate the optimal number of PCAs to be used in the analysis (Supplementary Fig. 8b).
359 After calculating the first 12 PCAs with *RunPCA,* we performed UMAP embedding using
360 *runUMAP* with parameters *n.neighbors*=10, *min.dist*=.1, *metric*="correlation", and
361 *umap.method*="umap-learn". Clustering was done with *FindNeighbors* (default parameters),
362 and *FindClusters* function using the SLM algorithm, *resolution*=1.15, and *n.iter*=100. Markers
363 genes were found with the function *FindAllMArkers*, using the "wilcox" test and *min.pct*=0.25.
364 The annotation of the clusters was based on the top 20 markers of each cluster.
365
366 Annotation
367 Annotation of the seedling and flower clusters was based on TraVaDB (Transcriptome
368 Variation Analysis Database, http:/travadb.org). TraVaDB is an open-access database based
369 on RNA-Seq data, which includes 79 samples, each with at least two biological replicates,
370 corresponding to different developmental stages and parts of roots, leaves, flowers, seeds,
371 siliques and stems. The top 20 differentially expressed genes of each cluster was used as input
372 for the analysis with TraVaDB [17]. The complete TraVaDB was downloaded. The heatmaps
373 showing the tissue types in which genes were found expressed were created using a R script.
374 For the annotation of the root clusters, we developed a function in R (available at
375 https://github.com/ramonvidal/punyplatypus). This function processes the output of Seurat
376 *FindAllMarkers* and predicts cell-type(s) for each cluster based on the match between the genes
377 from the input list and the marker genes from a reference list containing one or more single-
378 cell experiments. An adjusted p-value (by Bonferroni correction for multiple hypothesis testing)
379 and a PPV (positive predictive value) are calculated for each predicted cell-type. The smaller

380 the p-value the bigger the evidence the genes is a cell-type specific marker. The PPV describes
381 the performance of the prediction. It represents the proportion of positive results that are truly
382 positive. High PPV can be interpreted as indicating the accuracy of the prediction. Only
383 differentially expressed genes with adjusted p-value <= 0.05 were used as input for the
384 punyplatypus. The reference list was created using marker genes from recently published
385 single-cell RNA expression data of the *Arabidopsis* root [5,9]. They are listed in Supplementary
386 Table 1.
387
388 Reproducibility and correlation
389 To assess technical reproducibility of our protocol, we used MA plots to evaluate the variability
390 in each pair of seedling replicates. We compared three replicates against one another, resulting
391 in three comparisons. Pearson's correlation coefficient was calculated across seedling
392 replicates. The consistency between bulk and PN-seq experiments was investigated throught
393 the comparison between the log2 mean expression of genes detected in both experiments.
394 Expression of bulk RNA-seq data was quantified with RSEM[25]. They are listed in
395 Supplementary Table 3.
396
397 Network analysis
398 GENIE3 was used to infer gene networks starting from the normalized expression data obtained
399 from Seurat for each cluster independently, using the parameters *nTrees*=1000, and using as
400 regulators the list of DNA binding proteins obtained from TAIR (www.arabidopsis.org). Genes
401 expressed in less than 33% of the nuclei in a particular cluster were removed. Only the top
402 10,000 interactions were kept. Gene regulators with less than 10 predicted targets were also
403 removed. Dynamics of the gene network through anther development were obtained by the
404 following approach: First, all nuclei were ordered by their estimated developmental pseudotime
405 using Monocle 3 [26] and cluster 0 (meristem/Early anther) as root cluster. Next, gene networks
406 were estimated with GENIE, as described before, using groups of non-overlapping sets of 50
407 nuclei that were previously ordered by its developmental pseudotime.
408
409 Generation and Confocal Imaging of Reporter Lines
410 To validate expression specificity of the marker genes from our single cell PN-seq approach,
411 promoter:NLS-GFP (nuclear locatisation signal-green fluorescent protein) reporter lines were
412 generated.The marker genes for validation were chosen from the pool of cluster specific marker
413 genes (p<0.05) that were not previously characterized in the literature (unknown marker genes).

414 The genomic promoter region upstream of the ATG and until the closest neighboring gene was
415 amplified by PCR and introduced into the entry vector pCR8:GW:TOPO by TA cloning
416 (primers used for PCR are listed in Supplemental Table 4. Afterwards, the LR reactions were
417 performed with the binary vector pGREEN:GW:NLS-GFP (Smaczniak et al. 2017) to generate
418 GFP transcriptional fusions to a nuclear-localization signal peptide. All reporter constructs were
419 transformed into the Col-0 Arabidopsis background, and multiple independent lines per
420 construct were analyzed under a Zeiss LSM800 laser-scanning confocal microscope. Different
421 floral organs were dissected and screened for the GFP signal by confocal microscopy under
422 20× and 63× magnification objectives. Auto-fluorescence from chlorophyll was collected to
423 give an outline of the flower organs. A 488-nm laser was used to excite GFP and chlorophyll
424 and emissions were captured using PMTs set at 410–530 nm and 650– 700nm. Z-stack screens
425 were performed for the floral meristem and stigma tissues to give a 3D structure visualization.
426

427 **Data availability**

428 All relevant data have been deposited in EBI ArrayExpress, accession number E-MTAB-9174.

429 **Code availability**

430 R function developed for the annotation of the root clusters is available on Github:
431 https://github.com/ramonvidal/punyplatypus).
432

433 **Acknowledgements**

434 The study was supported by the European Commission (FP-7, grant agreement no. 262055,
435 ESGI, and Horizon 2020, grant agreement no. 824110, EASI-Genomics) (S.S.), and by DFG
436 grant KA 2720/5-1 (X.X, K.K). We thank Solenne Bourdier and Dijun Chen for support.
437

438 **References**

439 1. Fischer, C. *et al.* Signals trigger state-specific transcriptional programs to support
440 diversity and homeostasis in immune cells. *Sci. Signal.* **12**, 1–17 (2019).
441 2. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, e27041 (2017).
442 3. Rhee, S. Y., Birnbaum, K. D. & Ehrhardt, D. W. Towards Building a Plant Cell Atlas.
443 *Trends Plant Sci.* **24**, 303–310 (2019).
444 4. Efroni, I. & Birnbaum, K. D. The potential of single-cell profiling in plants. *Genome*
445 *Biol.* **17**, 65 (2016).
446 5. Efroni, I., Ip, P.-L., Nawy, T., Mello, A. & Birnbaum, K. D. Quantification of cell
447 identity from single-cell gene expression profiles. *Genome Biol.* **16**, 9 (2015).
448 6. Efroni, I. *et al.* Root Regeneration Triggers an Embryo-like Sequence Guided by
449 Hormonal Interactions. *Cell* **165**, 1721–1733 (2016).
450 7. Zhang, T.-Q., Xu, Z.-G., Shang, G.-D. & Wang, J.-W. A Single-Cell RNA Sequencing

Profiles the Developmental Landscape of Arabidopsis Root. *Mol. Plant* **12**, 648–660 (2019).

8. Jean-Baptiste, K. *et al.* Dynamics of gene expression in single root cells of arabidopsis thaliana. *Plant Cell* **31**, 993–1011 (2019).
9. Denyer, T. *et al.* Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Resource Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using. *Dev. Cell* **48**, 840-852.e5 (2019).
10. Lake, B. B. *et al.* A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* **7**, 1–8 (2017).
11. Macaulay, I. C., Ponting, C. P. & Voet, T. Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends Genet.* **33**, 155–168 (2017).
12. Selewa, A. *et al.* Systematic Comparison of High-throughput Single-Cell and Single-Nucleus Transcriptomes during Cardiomyocyte Differentiation. *Sci. Rep.* **10**, 1–13 (2020).
13. Ding, J. *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
14. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
15. Beisang, D. J. *et al.* Single-cell RNA sequencing reveals that lung mesenchymal progenitor cells in IPF exhibit pathological features early in their differentiation trajectory. *Sci. Rep.* **10**, 1–12 (2020).
16. Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740–742 (2014).
17. Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D. & Penin, A. A. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant J.* **88**, 1058–1070 (2016).
18. Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early flower development in Arabidopsis. *Plant Cell* **2**, 755–767 (1990).
19. Gómez, J. F., Talle, B. & Wilson, Z. A. Anther and pollen development: A conserved developmental pathway. *J. Integr. Plant Biol.* **57**, 876–891 (2015).
20. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, 1–10 (2010).
21. Xu, J. *et al.* The ABORTED MICROSPORES regulatory network is required for postmeiotic male reproductive development in Arabidopsis thaliana. *Plant Cell* **22**, 91–107 (2010).
22. Zhu, E. *et al.* The DYT1-interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for Arabidopsis anther development and transcriptome. *Plant J.* **83**, 976–990 (2015).
23. Shulse, C. N. *et al.* Profiling of Plant Cell Types Resource Profiling of Plant Cell Types. *CellReports* **27**, 2241-2247.e4 (2019).
24. Ryu, K. H., Huang, L., Kang, M. & Schiefelbein, J. Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells 1 [OPEN]. **179**, 1444–1456 (2019).
25. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *Bioinforma. Impact Accurate Quantif. Proteomic Genet. Anal. Res.* 41–74 (2014) doi:10.1201/b16589.
26. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

Acknowledgements

The study was supported by the European Commission (FP-7, grant agreement no. 262055, ESGI, and Horizon 2020, grant agreement no. 824110, EASI-Genomics) (S.S.), and by DFG grant KA 2720/5-1 (X.X, K.K). We thank Solenne Bourdier and Dijun Chen for support.

Author information

Conception and design: S.S. and K.K.; Protocol optimization to obtain plant nuclei: W.Y.,X.X.,C.S., C.B. C.F., and S.S.; Experimental work: X.X. and C.B.; FACS analysis: C.B.; Library preparation for single-cell and bulk RNA-sequencing: C.B. and C.S.; Nuclei Isolation and selection by ICELL8 Imaging System: X.X.,C.B. and C.S.; Validation with transgenic lines and confocal imaging: X.X. and M.Z.; Provision of tools and materials: K.K., M.K. and S.S. Data analysis: D.Y.S.F. and J.M.M.; R function development: R.V.; Data interpretation: D.Y.S.F., J.M.M., C.S., K.K. and S.S; Drafting of the manuscript: S.S. and D.Y.S.F.; Manuscript writing, review and edition: all authors. Funding and supervision: K.K. and S.S.

Ethics declarations

The authors declare no competing interests.

Figures

Figure 1: Single-nucleus RNA-sequencing. a) Schematic overview of the here-applied experimental strategy consisting of i) harvest and snap-freeze of plant material, ii) crushing to small pieces in liquid nitrogen, iii) mechanic disruption in Honda buffer (with a gentleMACS Dissociator), iv) nuclei release, v) filtering and centrifugation, vi) FACS, vii) nanowell-based single-nucleus preparation, viii) library preparation and finally ix) sequencing. After nuclei isolation (FACS) alternative experimental approaches are conceivable to produce sequence data. b) FACS histogram plots of DAPI fluorescent nuclei from different plants and different tissue types: *Arabidopsis thaliana* seedlings and flowers, *Petunia hybrid* flowers, *Antirrhinum majus* flowers (snapdragon), *Solanum lycopersicum* (tomato) flowers and leaves are shown after conventional gating for rough debris exclusion and doublet discrimination. The grey filled sections represent the gate that was set for sorting. The different tissue types produce different amount of nucleus-like sized, low DAPI-fluorescent debris that can only be separated from intact nuclei by gating the high DAPI-fluorescent peaks (Suppl. Fig. 2). c-d) UMAP plots showing the reproducibility among three independent biological replicates of *Arabidopsis thaliana* seedlings. e) UMAP plot of 2,871 nuclei showing the single-nuclei cluster by identity.

535 f) Barplot showing the proportion of nuclei corresponding to the identified cluster across the
536 biological replicates.
537
538 Figure 2: Anther development at single-nuclei resolution. a) UMAP plot of the flower PN-seq
539 data. b) Heatmap showing the expression of the top 20 significant markers genes for each
540 cluster. c) Gene expression of known anther regulators *AMS*, *bHLH089*, *bHLH090* and
541 *bHLH010* plotted in the UMAP coordinates. d) Gene network estimated from cluster 15 (early
542 anther) using GENIE3 (Supplementary Materials and Methods); only top 5,000 interactions
543 were used, and only TFs with more than 3 targets are shown. e) Heatmap showing the strength
544 of the interaction between AMS and its target obtained by GENIE3 across overlapping sets of
545 50 cells from anther clusters ordered by pseudotime; T1 is the first 50 cells (cluster 0,
546 meristem/early anthers), and T37 is the latest stage (cluster 3, late anther).
547
548 **Supplementary Information**
549 Supplementary Figures
550 Suppl. Fig. 1: Effect of protoplast isolation (PI) on root scRNA-seq. a) Expression correlation
551 of each cell from scRNA-seq with bulk RNA-seq sample using PI and with bulk RNA-seq
552 sample without using PI (y and x-axis, respectively; Denyer *et al.* data [9]). Cells have different
553 response to protoplasting, with some groups of cells being more sensitive to protoplasting
554 (higher correlation with PI data than with data from intact tissues) than others. b) Re-analysis
555 of scRNA-seq full data and c) re-analysis of scRNA-seq data removing the top 6,000 PI-
556 responsive genes (Denyer *et al.* data [9]) from the clustering step. In b) and c), the left UMAP
557 plots show the cell clusters of scRNA-seq data; the UMAP plots in the center show the
558 difference between the correlation of each cell from scRNA-seq with bulk RNA-seq PI and
559 non-PI samples. Thus, positive correlation numbers indicates cells with stronger similarity to
560 the transcriptome of PI samples. The violin plots in the right show the difference in the
561 correlation of cells between PI and non-PI, per cluster. When no PI-responsive genes were
562 removed (b) we observed several clusters containing cells with strong response to PI, with the
563 most extreme cluster with up to 55% of the top 20 marker genes being PI-responsive genes.
564 This effect largely persisted when PI-responsive genes were excluded from the primary scRNA-
565 seq analysis. After the exclusion of PI-responsive genes from the clustering step, but still using
566 them to identify markers, we observed the most extreme cluster with up to 46% of the top 20
567 marker genes being PI-responsive genes (c). These results highlight a need for alternative
568 methods beyond PI for plant single-cell genomics.

Suppl. Fig. 2: Generic single-nuclei isolation procedure. a) Microscopy analysis. Sections from disposable Neubauer counting chambers with DAPI stained nuclei after FACS from *Arabidopsis* thaliana seedlings and flowers, *Petunia hybrid* flowers, *Antirrhinum majus* flowers (snapdragon), *Solanum lycopersicum* (tomato) flowers and leaves. The brightfield images are overlain with the blue fluorescence images. Images of *Arabidopsis thaliana* samples were captured with a DMi8 microscope by Leica and the others by a BZ-X700 Series microscope by Keyence. The images show that FACS yields clean, debris-free nuclei suspensions irrespective of the initial amount of debris. b) Contour plots of flow cytometry experiments. c) Gating strategy used for flow cytometry, exemplified for *Arabidopsis thaliana* flower (inflorescence) samples. d) Quality control of RNA nuclei samples through chromatrography-based analysis of RNA derived from pooled nuclei of *Arabidopsis thaliana* seedlings and RNA derived from conventional purification (RIN = RNA Integrity Number). B1 corresponds to bulk RNA from tissue and C1 corresponds to RNA from sorted nuclei.

Suppl. Fig. 3: Summary of PN-seq seedling datasets. Violin plots showing the total number of detected genes (nFeature), reads counts (nCount), proportion of mitochondria (percent.mt) and chloroplast (percent.ch) contamination per nucleus for each replicate.

Suppl. Fig. 4: Correlation among PN-seq replicates and between PN-seq and bulk RNA-seq libraries. a) MA-plot showing the differences between samples: Replicate 1 versus Replicate 2, Replicate 1 versus Replicate 3 and Replicate 2 versus Replicate 3, plotted against the average gene count value (A). The red line shows the average differences. b) Scatterplot of gene expression obtained by different PN-seq replicates. c) Scatterplot of gene expression obtained by pooling the three PN-seq replicates against the bulk RNA-seq. Read counts from both datasets, PN-seq and bulk RNA-seq, were log2 transformed.

Suppl. Fig. 5: Comparison of bulk-RNA and PN-seq data indicates no expression bias of specific gene groups. Expression of top 25% genes with highest variance between bulk (red boxes) and PN-seq (green boxes) experiments. *A. thaliana* chromosomes are shown on the x-axis and the log2 mean expression of the genes in 3 replicates are shown on the y-axis. The thick line in the middle of the box represents 50% of observations, with the lower end of the box at 25% and the upper end of the box at 75%.

603 Suppl. Fig. 6: Annotation of seedling cell-clusters using TraVaDB dataset. a) UMAP plot of
604 2,871 nuclei organized in 13 clusters before the annotation. b) Heatmaps of the 13 clusters
605 showing the expression level of the top 20 differentially expressed genes from each cluster in
606 the TraVaDB dataset. c) Illustration of the cluster annotation process.
607
608 Suppl. Fig. 7: Single-nuclei transcriptome analysis of a subset of root cells derived from
609 seedling data. a) UMAP of 15 clusters (n=964 nuclei). Eight clusters were faithfully annotated
610 using punyplatypus function: cluster 0 - mature (p-value=7.785874e-04, PPV=0.55), cluster 1
611 - endodermis (p-value=1.550274e-13, PPV=0.86), cluster 1 - cortex (p-value=1.689028e-10,
612 PPV=0.71), cluster 4 - stele (p-value=8.827076e-12, PPV=0.90), cluster 9 - mature (p-
613 value=5.837800e-10, PPV=1.0, cluster 10 - trichoblast (p-value=1.655356e-92, PPV=0.90),
614 cluster 11 - trichoblast (p-value=3.641153e-11, PPV=0.47), cluster 12 - endodermis (p-
615 value=2.486261e-05, PPV=0.54) and cluster 14 - xylem (p-value=8.777430e-37, PPV=0.94).
616 Punyplatypus calculates a p-value and PPV per cell type, which indicate the performance of
617 annotation. The smaller the p-value the bigger the evidence that the genes are cell-type specific
618 markers. High PPV can be interpreted as indicating the proportion of genes in a cluster found
619 annotated with the same cell type in the reference list. b) Violin plots showing expression of
620 marker genes per annotated cluster. c) Heatmap corroborating the annotation by punyplatypus.
621 It shows the top 100 markers of Denyer *et al.*[9] among the top 1000 expressed genes per nuclei
622 of PN-seq. Almost all cell types could be recovered using a subset of root nuclei from the
623 complex seedling data.
624 Suppl. Fig. 8: Correlation between unfixed and fixed seedling samples. a) UMAP plot showing
625 similar cell distribution of unfixed and fixed seedling samples. b) UMAP plot showing the
626 overlapping between cells from unfixed and fixed seedling samples. c) Pearson's correlation
627 coefficient across unfixed and fixed seedling samples. The thick line in the middle of the box
628 represents 50% of observations (log2 of read counts), with the lower end of the box at 25% and
629 the upper end of the box at 75%. Bars extend to the lowest and highest values that are not
630 outliers. The correlation found over all samples was 0.90.
631
632 Suppl. Fig. 9: Single-nuclei transcriptome analysis of *Arabidopsis thaliana* flower
633 development. a) Number of genes per nuclei (nFeature) and number of reads per gene (nCount).
634 b) JACKSTRAW plot to identify the optimal number of PCAs for the analysis of the
635 inflorescence dataset. c) Annotation of clusters based on correlation: the average gene
636 expression of each cluster was Spearman correlated against each one of the TraVaDB

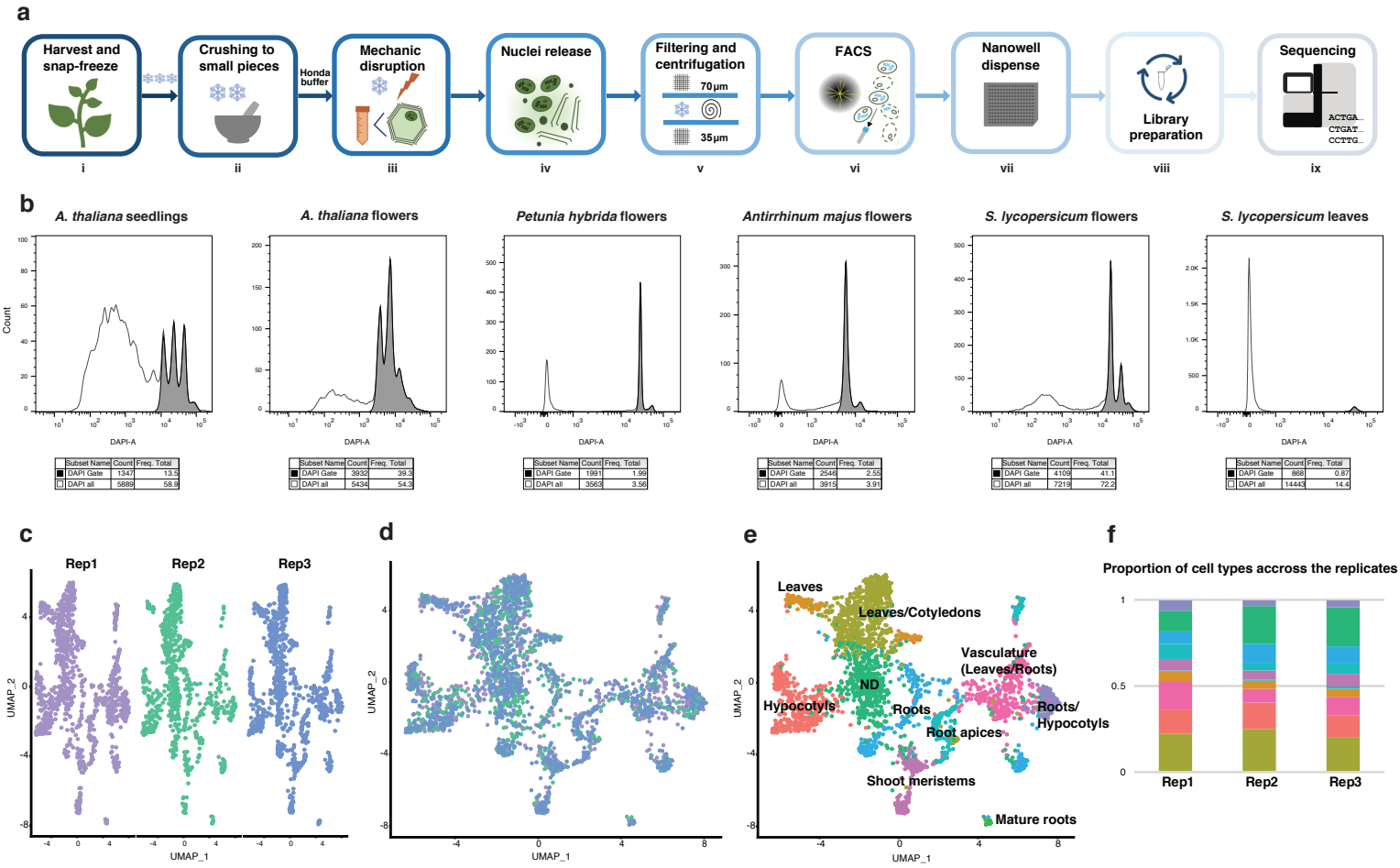
637 transcriptome dataset considered. The two labels plotted on top of each cluster indicate the two
638 TraVa samples with highest correlation.
639
640 Suppl. Fig. 10: Temporal trajectories in the floral PN-seq dataset. a) Annotation of clusters
641 based on correlation with the "stages" samples in the TraVa dataset. b and c) Pseudotime
642 analysis using Monocle 3. d) Flower developmental stages recovered from the TraVa dataset.
643
644 Suppl. Fig. 11: Main master regulators in flower cells. GENIE3 was used to estimate the gene
645 network of each cluster. The heatmap shows the number of predicted target genes for each TF
646 among the top 10,000 strongest interaction in the network. Only the top 4 TFs with most
647 predicted target genes per cluster are shown.
648
649 Suppl. Fig. 12: Novel marker genes covering the developmental trajectory of anther/pollen
650 development.
651
652 Suppl. Fig. 13: a) Summary of functional validation. Clearly visually validated genes are
653 indicated by green dots, whereas grey dots indicate negative results. b) Validation for cluster-
654 specific genes with transcriptional reporter lines. Expression patterns of reporter lines for the
655 following genes reveal the predicted floral organ specificities: (a) AT1G63100, floral meristem;
656 (b) AT3G51740, floral meristem; (c) AT4G11290, stigma; (d) AT2G38995, sepal; (e)
657 AT2G38995, petal; (f) AT2G38995, anther. Expression patterns of reporter lines for the
658 following genes reveal the predicted stage specificities for anther development: (g)
659 AT5G20030, flower 11-13; (h) AT1G23520, flower 6-8; (i) AT2G16750, flower 4-5. White
660 arrowheads indicate GFP signals in nuclei;. Scale bars, 50 µm.
661
662 Suppl. Fig. 14: PN-seq data analysis pipeline. Blue boxes represent the main analysis steps, the
663 software used in each step is shown in italic and the steps using *mappa* analysis pipeline are
664 shown in the dashed gray box. The in/output files from each step of analysis are represented
665 with file icons.
666
667 Supplementary Tables:
668 Suppl. Table 1: List of marker genes in *Arabidopsis thaliana* seedlings, roots and flowers.
669 Suppl. Table 2: Program steps for gentle tissue disruption and nuclei release on the
670 gentleMACS Dissociator and the instrument specific commands. The instrument specific tubes

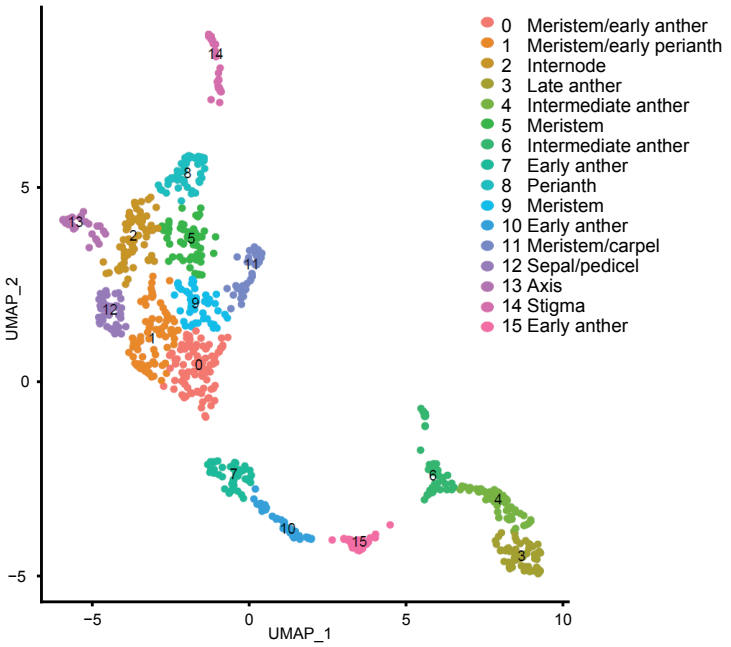
have got a stator and rotor. Latter can be moved at certain a certain speed (rpm = rounds per minute) for a certain time (time in s) in several rounds (loops).

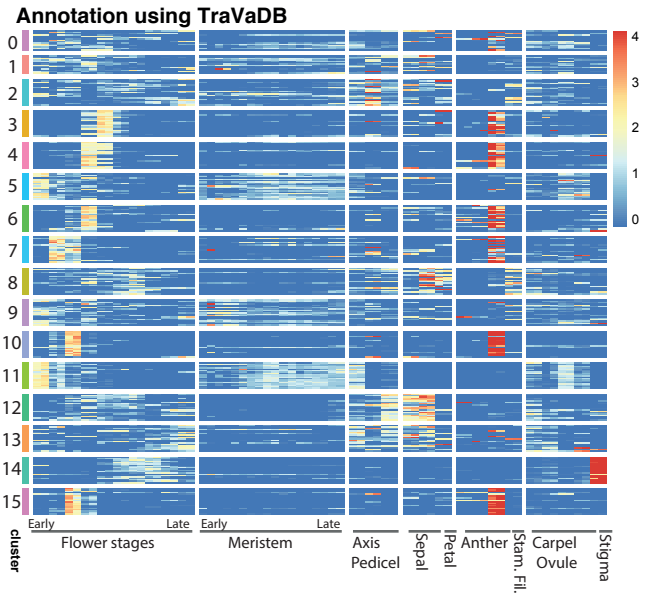
Suppl. Table 3: Sequencing data metrics and read counts of PN-seq and bulk RNA-seq libraries.

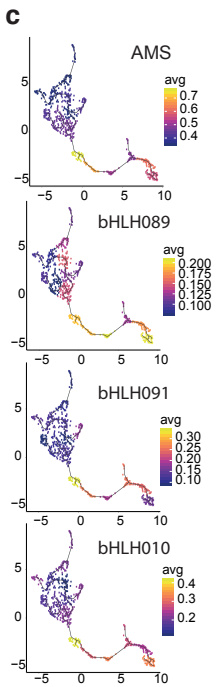
Supplementary Data:

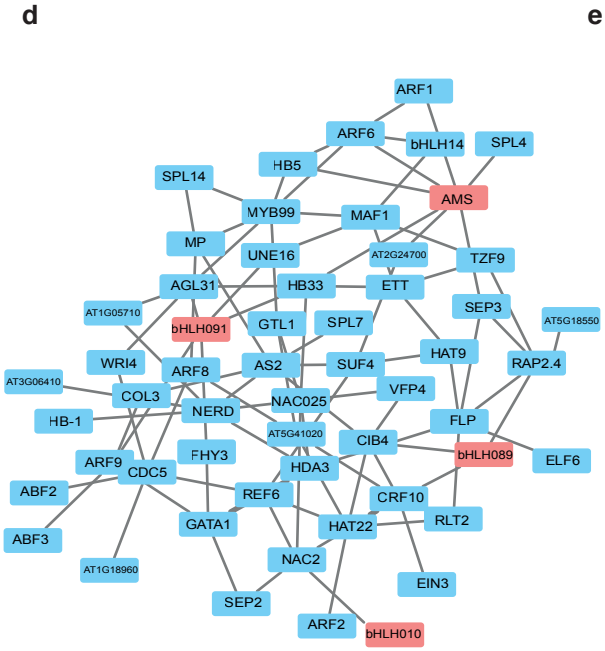
Suppl. Data 1: Summary and read statistics of the PN-seq data (adapted from the html reports generated by ICELL8 hanta software)

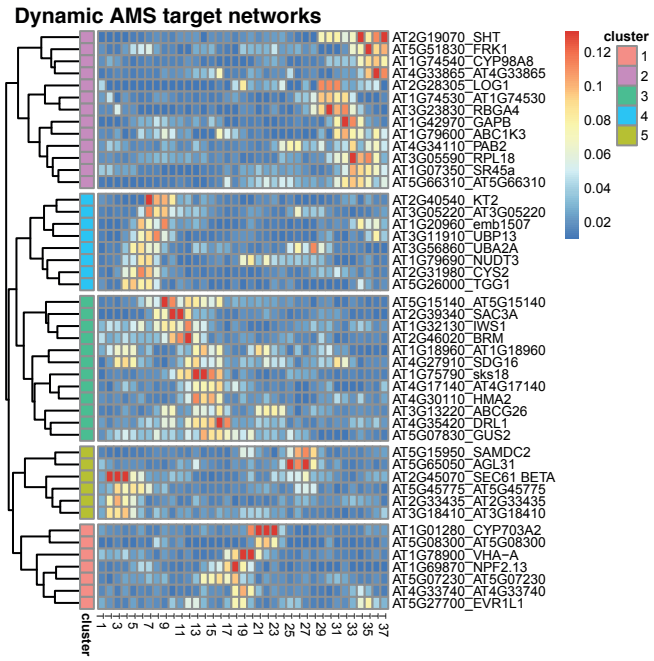

a
Harvest and snap-freeze
Crushing to small pieces
Honda buffer
Mechanic disruption
Nuclei release
Filtering and centrifugation
70µm
35µm
FACS
Nanowell dispense
Library preparation
Sequencing
ACTGA...
CTGAT...
CCTTG...
i
ii
iii
iv
v
vi
vii
viii
ix
b
A. thaliana seedlings
A. thaliana flowers
Petunia hybrida flowers
Antirrhinum majus flowers
S. lycopersicum flowers
S. lycopersicum leaves
Count
DAPI-A
Subset Name | Count | Freq. Total
DAPI Gate | 1347 | 13.5
DAPI all | 5889 | 58.9
DAPI Gate | 3932 | 39.3
DAPI all | 5434 | 54.3
DAPI Gate | 1991 | 1.99
DAPI all | 3563 | 3.56
DAPI Gate | 2546 | 2.55
DAPI all | 3915 | 3.91
DAPI Gate | 4109 | 41.1
DAPI all | 7219 | 72.2
DAPI Gate | 868 | 0.87
DAPI all | 14443 | 14.4
c
Rep1
Rep2
Rep3
UMAP_2
UMAP_1
d
UMAP_2
UMAP_1
e
Leaves
Leaves/Cotyledons
Vasculature (Leaves/Roots)
ND
Hypocotyls
Roots
Root apices
Roots/ Hypocotyls
Shoot meristems
Mature roots
UMAP_2
UMAP_1
f
Proportion of cell types accross the replicates
Rep1
Rep2
Rep3


a
0 Meristem/early anther
1 Meristem/early perianth
2 Internode
3 Late anther
4 Intermediate anther
5 Meristem
6 Intermediate anther
7 Early anther
8 Perianth
9 Meristem
10 Early anther
11 Meristem/carpel
12 Sepal/pedicel
13 Axis
14 Stigma
15 Early anther
UMAP_2
UMAP_1


b
Annotation using TraVaDB
cluster
Early
Late
Flower stages
Meristem
Axis
Pedicel
Sepal
Petal
Anther
Stam. Fil.
Carpel
Ovule
Stigma


c
AMS
bHLH089
bHLH091
bHLH010
avg


d
ARF1
ARF6
bHLH14
SPL4
HB5
SPL14
AMS
MYB99
MAF1
MP
UNE16
AT2G24700
TZF9
AGL31
HB33
ETT
SEP3
AT1G05710
bHLH091
GTL1
SPL7
AT5G18550
WRI4
ARF8
AS2
SUF4
HAT9
RAP2.4
AT3G06410
COL3
NAC025
VFP4
HB-1
NERD
FLP
AT5G41020
CIB4
bHLH089
ELF6
ARF9
FHY3
CDC5
HDA3
ABF2
REF6
CRF10
GATA1
HAT22
RLT2
ABF3
AT1G18960
NAC2
EIN3
SEP2
ARF2
bHLH010


e
Dynamic AMS target networks
cluster