

# A combinatorial indexing strategy for low-cost epigenomic profiling of plant single cells

Xiaoyu Tu<sup>1,2,5</sup>, Alexandre P. Marand<sup>3,5</sup>, Robert J. Schmitz<sup>3,\*</sup> and Silin Zhong<sup>4,\*</sup>

<sup>1</sup>Joint Center for Single Cell Biology, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup>Shanghai Collaborative Innovation Center of Agri-Seeds, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>3</sup>Department of Genetics, University of Georgia, Athens, GA 30602, USA

<sup>4</sup>State Key Laboratory of Agrobiotechnology, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, China

<sup>5</sup>These authors contributed equally to this article.

\*Correspondence: Silin Zhong ([silin.zhong@cuhk.edu.hk](mailto:silin.zhong@cuhk.edu.hk)), Robert J. Schmitz ([schmitz@uga.edu](mailto:schmitz@uga.edu))

<https://doi.org/10.1016/j.xplc.2022.100308>

## ABSTRACT

Understanding how *cis*-regulatory elements facilitate gene expression is a key question in biology. Recent advances in single-cell genomics have led to the discovery of cell-specific chromatin landscapes that underlie transcription programs in animal models. However, the high equipment and reagent costs of commercial systems limit their applications for many laboratories. In this study, we developed a combinatorial index and dual PCR barcode strategy to profile the *Arabidopsis thaliana* root single-cell epigenome without any specialized equipment. We generated chromatin accessibility profiles for 13 576 root nuclei with an average of 12 784 unique Tn5 integrations per cell. Integration of the single-cell assay for transposase-accessible chromatin sequencing and RNA sequencing data sets enabled the identification of 24 cell clusters with unique transcription, chromatin, and *cis*-regulatory signatures. Comparison with single-cell data generated using the commercial microfluidic platform from 10X Genomics revealed that this low-cost combinatorial index method is capable of unbiased identification of cell-type-specific chromatin accessibility. We anticipate that, by removing cost, instrumentation, and other technical obstacles, this method will be a valuable tool for routine investigation of single-cell epigenomes and provide new insights into plant growth and development and plant interactions with the environment.

**Key words:** combinatorial indexing, single-cell ATAC-seq, epigenome

Tu X., Marand A.P., Schmitz R.J., and Zhong S. (2022). A combinatorial indexing strategy for low-cost epigenomic profiling of plant single cells. Plant Comm. **3**, 100308.

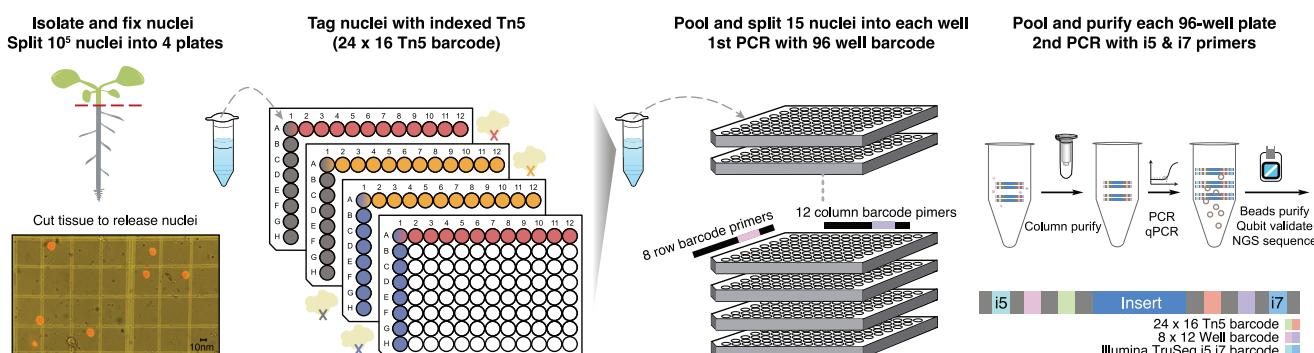
## INTRODUCTION

In multicellular organisms, each cell shares an identical set of genetic instructions despite having highly specialized structures and functions. Cell-specific physiology caused by dynamic gene expression characteristics is often the result of the binding and activity of cell-specific transcription factors (TFs) to proximal and distal *cis*-acting regulatory elements (CREs) in accessible chromatin regions (ACRs) associated with a given gene (Thurman et al., 2012; Kundaje et al., 2015). Chromatin accessibility profiling methods such as DNase I hypersensitive site sequencing and assay for transposase-accessible chromatin sequencing (ATAC-seq) have been developed to measure chromatin accessibility as a generalized proxy for regulatory DNA across numerous plant species (Crawford et al., 2006; Buenrostro et al., 2013; Moore et al., 2020; Minnoye et al., 2021). However, these assays can only capture the average chromatin accessibility signal across a population of cells, masking cell-specific and rare events within a given tissue.

The recent development of droplet microfluidics device-based single-cell systems has enabled researchers to co-encapsulate individual cells or nuclei with barcode beads in nanoliter water-in-oil emulsions, making it possible to perform sequencing at the single-cell level (Lareau et al., 2019; Satpathy et al., 2019). For example, single-cell ATAC-seq (scATAC-seq) assays based on the 10X Genomics Chromium system have been used successfully to study cell-specific chromatin accessibility in plants (Dorrity et al., 2021; Farmer et al., 2021; Marand et al., 2021). However, these commercial single-cell assays often require specialized equipment and trained personnel and are very expensive to purchase and maintain. An alternative low-cost scATAC-seq method based on combinatorial indexing (sci-ATAC-seq) has been developed for single-cell epigenome profiling in

---

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.



**Figure 1.** Overview of sci-ATAC-seq.

Shown is a schematic of the sci-ATAC-seq experimental workflow. Tagmentation of permeabilized nuclei was carried out using 384 barcoded Tn5 in four 96-well plates. After the first pooling, ~15 nuclei were split into each well, and PCR was performed to introduce the second set of barcodes. Finally, PCR products from the same plate were pooled and purified for the final amplification with the third set of barcodes.

mammalian cells (Cusanovich et al., 2015; Fan et al., 2015; Rosenberg et al., 2018). By circumventing expensive microfluidics systems, this technique has already found widespread use in animal studies but remains to be adopted for plants.

It is clear to us that an easy-to-use, low-cost, sensitive, and high-throughput epigenomics profiling method is urgently needed for plant scientists to decode the full repertoire of *cis*-regulatory diversity and chromatin dynamics in different tissues throughout the life cycle and in response to environmental stimuli. To meet this challenge, we developed an sci-ATAC-seq protocol for plant tissues based on combinatorial indexing and dual PCR barcoding. It does not require any specialized equipment and has a low library preparation cost of ~\$100 per 5000 cells, whereas a 10X Chromium microfluidics system could cost over \$80,000, with an individual assay costing ~\$2000 per 5000 cells.

We applied this sci-ATAC-seq protocol to *Arabidopsis thaliana* root tissues and obtained high-quality chromatin accessibility data for 13 576 cells from two biological replicates. We benchmarked our data to previously generated 10X Genomics scATAC-seq data from the same tissue. The results showed that sci-ATAC-seq could readily capture known cell-type-specific chromatin accessibility profiles, was similarly correlated with nuclear transcription, and was characterized by lower background signal, organellar contamination, and doublet rates. A step-by-step protocol with preparation time is provided in the [Supplemental information](#). We anticipate that this technological breakthrough will empower plant researchers to overcome the difficulty and cost barriers of single-cell epigenome studies and enable routine single-cell epigenomics profiling at a larger scale.

## RESULTS

### Implementation of sci-ATAC-seq to profile single-nucleus chromatin accessibility

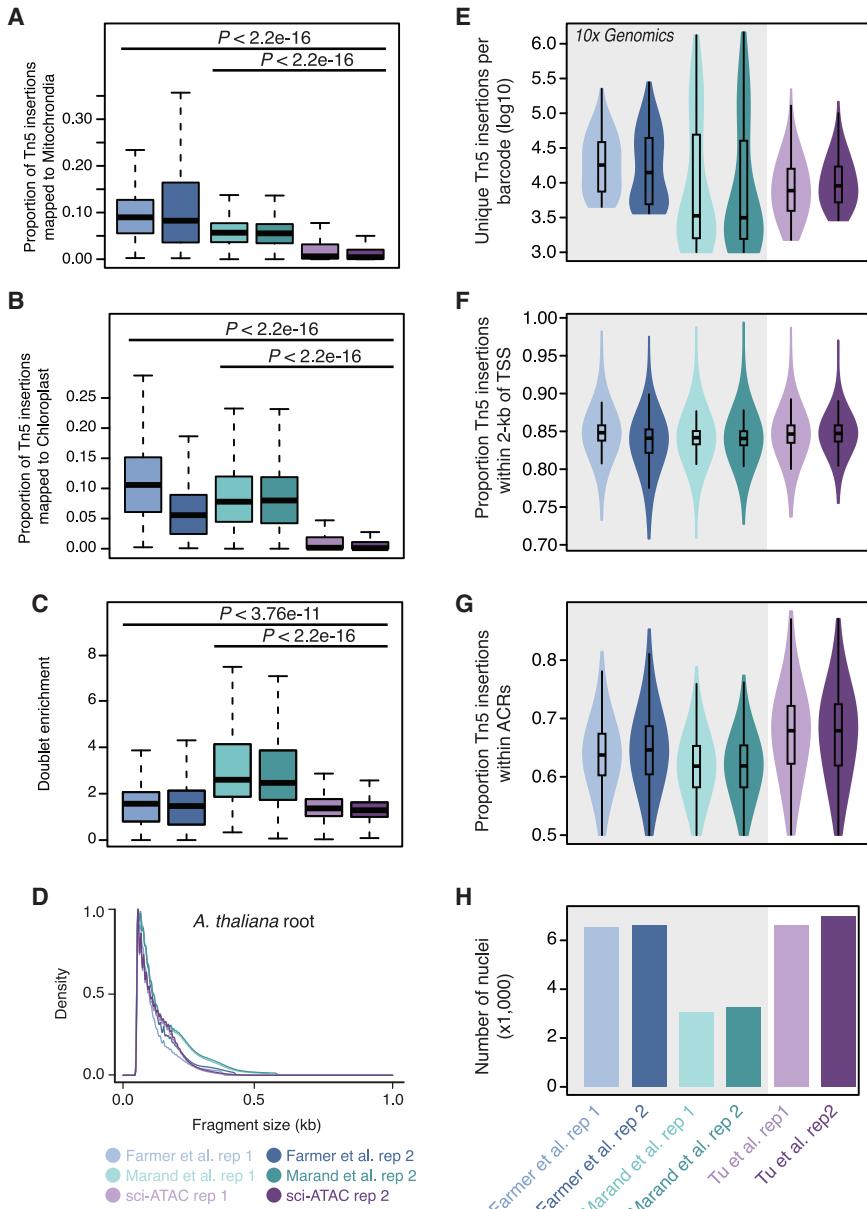
Low-cost combinatorial index-based sci-ATAC-seq is used routinely for animal cells and tissues. One of the key technological challenges presented by plant tissues is the clumping of broken nuclei because detergents, such as Triton X-100 and Tween 20, which are used to lyse plant organelles, also disrupt the outer nuclear membrane. We observed previously that plant nuclei after

formaldehyde fixation are highly resistant to clumping and remain intact after heat shock, SDS denaturation, centrifugation, and overnight incubation steps during *in situ* Hi-C library preparation (Dong et al., 2017). Because it has been reported that ATAC-seq can be performed on formaldehyde-fixed animal tissues (Chen et al., 2016), we introduced a short fixation step to the nuclei isolation process before lysing the plant organelles with detergent to preserve the nuclear membrane and prevent clumping during single-cell library preparation. The ability to use detergent washes to lyse organelles without affecting nuclear integrity enabled us to reduce doublet rates and remove organelle DNA contamination without the need for cytometrical sorting.

To demonstrate the utility of sci-ATAC-seq for resolving heterogeneous chromatin accessibility signals from complex plant tissues, we applied it to whole-root tissues from 2-week-old *A. thaliana* plants ([Figure 1; Supplemental information](#)). First, isolated nuclei were split into four 96-well plates with 24 × 16 barcoded Tn5 transposons to integrate 384 adapter combinations into the ACRs of individual nuclei. The tagmentation reaction was then quenched, and tagged nuclei were pooled and redistributed to multiple 96-well plates with approximately 15 nuclei per well. Plate PCR was then performed using 12 × 8 sets of unique dual-indexed primers complementary to the transposase-introduced adapters, which added a second round of well-specific barcodes to the Tn5-tagged DNA. Subsequently, PCR products from individual wells in each plate were combined and purified, and a final round of PCR was performed to add a third barcode to each molecule to label the sample plate. Using this method, a single 96-well plate can generate sci-ATAC-seq profiles for ~1500 cells, and one could easily perform 10–20 plate PCRs in a few days to obtain 15 000–30 000 chromatin accessibility profiles, making this method highly scalable.

### Comparison of sci-ATAC-seq and scATAC-seq

To evaluate the efficacy of our method, we compared our data with two recent *A. thaliana* root scATAC-seq datasets generated using the 10X Genomics Chromium system, which have similar nuclei counts and identified cell types (Farmer et al., 2021; Marand et al., 2021). We uniformly processed all six libraries to demultiplex reads, assign cell barcodes, align fragments to the TAIR10 reference genome, and remove duplicated fragments, and we used the R package Socrates to analyze all three datasets



**Figure 2. Comparison of sci-ATAC-seq quality control metrics with those of scATAC-seq.**

(A) Proportion of Tn5 insertions derived from the mitochondrial genome.

(B) Proportion of Tn5 insertions derived from the chloroplast genome.

(C) Distributions of doublet enrichment scores.

(D) Fragment size distributions by dataset and replicate.

(E) Distribution of unique Tn5 integration sites per nucleus.

(F) Distributions of the proportion of Tn5 integration sites within the promoter regions 2-kb upstream of gene TSSs.

(G) Distributions of the proportion of Tn5 integration sites within ACRs per nucleus.

(H) Number of nuclei that passed quality control thresholds.

previously published 10X scATAC-seq data (Figure 2H), with a high proportion of recovery at ~90% (Supplemental Figure 1).

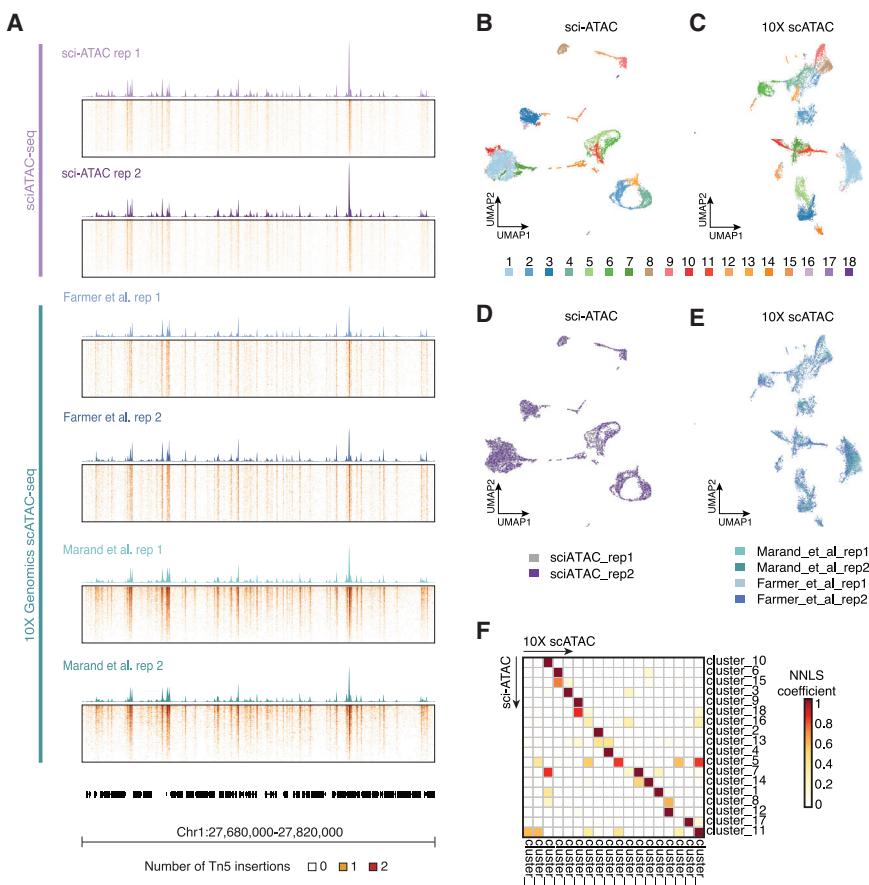
Next, we processed our sci-ATAC-seq and 10X scATAC-seq data sets in parallel to evaluate whether the differences in data quality could extend to nucleus clustering (Figure 3). Using identical parameters for normalization, graph-based clustering, and uniform manifold approximation and projection (UMAP) visualization in Socrates, we identified 18 clusters in the sci-ATAC-seq and 10X scATAC-seq datasets (Figure 3B and 3C). Qualitative assessment of the UMAP embedding indicated that the sci-ATAC-seq replicates were highly reproducible (Figure 3D). A similar degree of reproducibility was observed for the 10X scATAC-seq data, albeit after first removing lab-specific batch effects (Figure 3E).

To identify clusters with matching chromatin accessibility profiles between the datasets, we performed reciprocal non-negative least squares (NNLS) regression for each sci-ATAC-seq cluster using all 10X scATAC-seq clusters as the explanatory variable and vice versa,

multiplying the resulting coefficients (Methods). NNLS analysis indicated that the cluster-aggregate chromatin accessibility profiles were well matched between sci-ATAC-seq and 10X scATAC-seq (Figure 3F).

Overall, our analysis showed that a sci-ATAC-seq experiment with only 10 96-well-plates was sufficient to generate high-quality epigenome profiles for 13 576 nuclei, with an average of 12 784 unique Tn5 integrations per cell. Analysis of aggregated and single-cell signals indicated that sci-ATAC-seq produced chromatin accessibility profiles equivalent to those generated by the 10X Genomics Chromium system, with similar quality and reproducibility. Our method provides an efficient and

(Methods; Supplemental Figure 1). We first evaluated several common quality metrics, including organelle DNA contamination, doublet rate, fragment size, unique Tn5 insertions per cell, and proportion of insertions within promoters and ACRs (Figure 2). Our sci-ATAC-seq data showed significantly lower organelle DNA contamination (Wilcoxon rank-sum test,  $P < 2.2e^{-16}$ ) and doublet rates (Wilcoxon rank-sum test,  $P < 3.76e^{-11}$ ) compared with the two published datasets (Figure 2A–C). All libraries were enriched in subnucleosomal fragments, and the number of insertions per cell was comparable between the sci-ATAC-seq and 10X scATAC-seq data (Figure 2D and 2E). We also observed more Tn5 insertions mapping to gene promoter regions and within ACRs in the sci-ATAC-seq data (Figure 2F and 2G). Importantly, these gains in data quality did not come at the cost of reduced throughput. The total number of nuclei that passed quality control in the sci-ATAC-seq libraries was similar to that in



**Figure 3. Comparison of cluster separation and replicate overlap of sci-ATAC-seq and scATAC-seq data.**

(A) Pseudobulk aggregates and single-cell accessibility profiles from 1000 random nuclei across all datasets used in this study.

(B and C) sci-ATAC-seq (B) and 10X scATAC-seq (C) UMAP colored by Louvain clusters.

(D and E) sci-ATAC-seq (D) and 10X scATAC-seq (E) UMAP colored by biological replicate.

(F) Non-negative least squares (NLLS) regression coefficients (described by Domcke et al., 2020) between sci-ATAC-seq and 10X scATAC-seq datasets.

low-cost alternative for high-quality and high-resolution profiling of plant single-cell epigenomes.

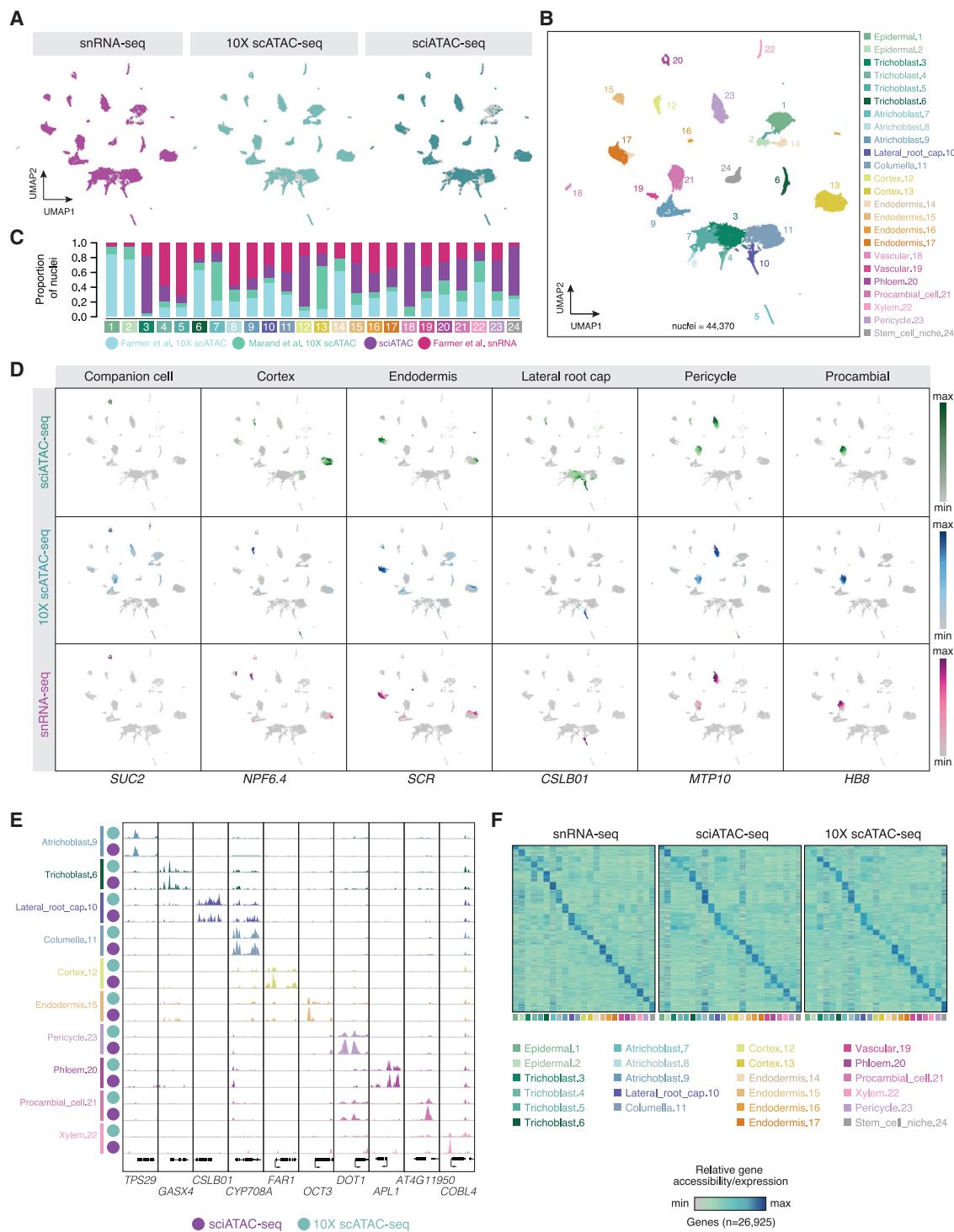
#### Integrative analysis of single-cell ATAC-seq and RNA-seq revealed cellular heterogeneity in the *A. thaliana* root

Next, we sought to resolve the chromatin landscapes of individual cell types within the *A. thaliana* root. We generated a chromatin accessibility and nuclear transcription unified co-embedding by integrating three single-cell ATAC-seq datasets and one single-nucleus RNA-seq (snRNA-seq) dataset from *A. thaliana* roots using a combination of integrative non-negative matrix factorization (INMF) and UMAP. The co-embedding discriminated ~1 billion pseudobulk Tn5 integration sites into 24 clusters based on shared cell states, ranging in size from 208 to 8755 nuclei (Figure 4A–4C). Automated annotation of cell identities using established marker genes recovered nearly all expected cell types of the *A. thaliana* root (Supplemental Table 1).

The improved resolution and coverage of the combined data enabled us to distinguish putative cell identities by evaluating differential chromatin accessibility profiles of known cell-type-specific marker genes (Figure 4D). For example, a known companion cell-specific SUCROSE-PROTON SYMPORTER 2 (*SUC2*; AT1G22710) gene was found to be highly expressed in

phloem cells (cluster 20) (Rodriguez-Villalon et al., 2014), concordant with the elevated promoter and gene body chromatin accessibility detected in our data. Gene body chromatin accessibility for NITRATE PEPTIDE TRANSPORTER 6.4 (*NPF6.4*; AT3G21670), a cortical parenchyma marker gene (Farmer et al., 2021), was enriched in predicted cortex clusters (clusters 12 and 13). The SCARECROW (*SCR*; AT3G54220) gene, which encodes a putative TF that is first expressed in quiescent center precursor cells during embryogenesis (Di Laurenzio et al., 1996) and extends to the initial cells for the ground tissue (cortex and endodermis) and the endodermis (Wysocka-Diller et al., 2000), showed enriched chromatin accessibility in the cortex (cluster 13) and endodermis (cluster 17) annotated clusters in our sci-ATAC-seq profiles.

To better annotate the 24 identified cell populations, we also generated pseudobulk chromatin accessibility maps for each cell type cluster and examined the ACRs neighboring known cell-type-specific marker genes (Figure 4E). For instance, enrichment of chromatin accessibility in TERPENE SYNTHASE 29 (*TPS29*; AT1G31950) from cluster 9 and XYLAN GLYCURONOSYLTRANSFERASE 4 (*GUX4*; AT1G54940) from cluster 6 suggested that the two clusters correspond to two cell types of the root epidermis, trichoblast, and atrichoblast, respectively. We observed cluster 10 with specific chromatin accessibility at the CELLULOSE SYNTHASE-LIKE PROTEIN B1 (*CslB1*; AT2G32610) gene, a known early non-hair/lateral root cap marker (Jean-Baptiste et al., 2019). Cluster 11 demonstrated chromatin accessibility at cis elements neighboring the CYTOCHROME P450 CYP708 A2 (*AT5G48000*) gene, which is associated with columella-specific expression (Brady et al., 2007). Cluster 12 showed significant ATAC-seq signal enrichment at FATTY ACID REDUCTASE 1 (*FAR1*; AT5G22500), a predicted mobile RNA associated with the suberization of endodermal cells (Domergue et al., 2010). Cluster 15 had enhanced chromatin accessibility at ORGANIC CATION TRANSPORTER 3 (*OCT3*; AT1G16390), a known endodermis marker gene (Jean-Baptiste et al., 2019). The DEFECTIVELY ORGANIZED TRIBUTARIES 1 (*DOT1*; AT2G36120) gene, involved in vascular patterning,



**Figure 4. Chromatin accessibility and nuclear transcription hallmarks of *A. thaliana* root cells.**

(A) Comparison of snRNA-seq, 10X Genomics scATAC-seq, and sci-ATAC-seq from *A. thaliana* root nucleus coordinates on an integrated UMAP embedding. Gray dots indicate nuclei from a mutually exclusive dataset.

(B) Cell type identities of nuclei from the integrated UMAP embedding.

(C) Proportion of nuclei derived from published 10X Genomics scATAC-seq and snRNA-seq data, as well as our sci-ATAC-seq, and sets for each cell type cluster.

(D) Gene accessibility and RNA abundance distributions for six cell-type-specific marker genes among sci-ATAC-seq, 10X Genomics scATAC-seq, and snRNA-seq datasets.

(legend continued on next page)

showed an ATAC-seq signal in cluster 23 and, to a lesser extent, in cluster 21, whereas cluster 20 showed enrichment of chromatin accessibility in *ALTERED PHLOEM DEVELOPMENT* (*APL*; AT1G79430), the marker gene known to appear earliest during companion cell and phloem sieve element specification (Jean-Baptiste et al., 2019; Shulse et al., 2019). Cluster 21 had ACRs neighboring a putative transmembrane protein that is preferentially expressed in procambial cells (AT4G11950) (Brady et al., 2007), whereas *COBRA-LIKE 4* (*COBL4*; AT5G15630), which showed a signal in cluster 22, is a marker gene for xylem (Denyer et al., 2019).

To evaluate the integrated embedding-based cell type annotations in the independent clustering analyses of sci-ATAC-seq and 10X scATAC-seq, we plotted the inferred cell identities from the integrated analysis on the UMAP embeddings for each scATAC-seq data set clustered in isolation. Surprisingly, we observed greater adjusted Rand index scores (a measure of similarity between clustering) for sci-ATAC-seq clusters relative to 10X scATAC-seq, indicating that the sci-ATAC-seq data faithfully grouped by cell identity independent of external data (Supplemental Figure 3). In addition to annotating cell types based on aggregate chromatin profiles, we also sought to evaluate cell type variation among modalities. We subset nuclei by snRNA-seq, 10X scATAC-seq, and sci-ATAC-seq and generated heatmaps of gene expression or chromatin accessibility across 26 925 protein-coding genes for each cell type (Figure 4F). Matched cell type comparisons between single-cell RNA-seq and ATAC-seq revealed that sci-ATAC-seq chromatin accessibility profiles (Pearson's correlation coefficient = 0.26) were comparable predictors of transcript abundance relative to 10X Genomics scATAC-seq (Pearson's correlation coefficient = 0.22), although the overall correlation with RNA abundance was relatively low for sci-ATAC-seq and 10X scATAC-seq (Supplemental Figure 4). Our analysis demonstrated that sci-ATAC-seq could robustly recapitulate cell-type-specific open chromatin identified by 10X scATAC-seq at a fraction of the cost.

### Regulatory consequences of cell-specific ACRs

A major advantage of single-cell methods over bulk profiling is the ability to enrich signals that originate from rare cell types. To empirically test this ability, we identified ACRs at the bulk and cell-type scales using sci-ATAC-seq, 10X Genomics scATAC-seq, and their union. Identification of ACRs using data partitioned by cell type resulted in 11 850 more ACRs than bulk on average, demonstrating a 32% increase in sensitivity (Figure 5A). As a specific example, we selected a random ACR on chromosome 1 (Chr1:25450953–25451274) detected after cell type partitioning that was undetected in the bulk data (Figure 5B). This ACR was highly enriched (250-fold) in the phloem (accessible in 5% of phloem cells) despite being detected in less than 0.02% of total nuclei, suggesting that a substantial portion of cell-type-specific ACRs are masked by bulk approaches (Figure 5B). In addition, we observed that ACRs identified after cell type partitioning were more concordant across datasets compared with bulk-

### Combinatorial index single-cell ATAC-seq

scale profiling (Figure 5C). Thus, interrogation of chromatin accessibility from cell-type resolved data significantly improves the sensitivity for the identification of rare and cell-type-specific CREs.

Next, we compared the chromatin accessibility profiles generated by sci-ATAC-seq and 10X scATAC-seq in different cell types. As expected, distributions of chromatin accessibility across cell types were highly concordant between the two different technologies (Figure 5D). To quantify potential differences, we performed differential accessibility analysis for sci-ATAC-seq and 10X scATAC-seq independently. The effect of cell type on ACR accessibility status was well correlated between technologies, indicating that the magnitude and direction of differences in chromatin accessibility across cell types were consistent between sci-ATAC-seq and 10X scATAC-seq (Figure 5E). By applying uniform thresholds (false discovery rate [FDR] <0.1 and beta coefficient >1) for each technology, we identified significant changes for ~25% of ACRs (12 363 of 50 394). Approximately 46% of cell-type-specific ACRs (5498 of 11 850) were not identified in the bulk ACR set, highlighting the usefulness of single-cell analysis for finding ACRs in rare cell types. Cell-type-specific ACRs were, on average, nearly 2.5 times (895 bp versus 375 bp) farther away from the nearest transcription start site (TSS) compared with non-specific ACRs, suggesting that even in the small *A. thaliana* genome, distal CREs may be important contributors to cell identity (Figure 5F).

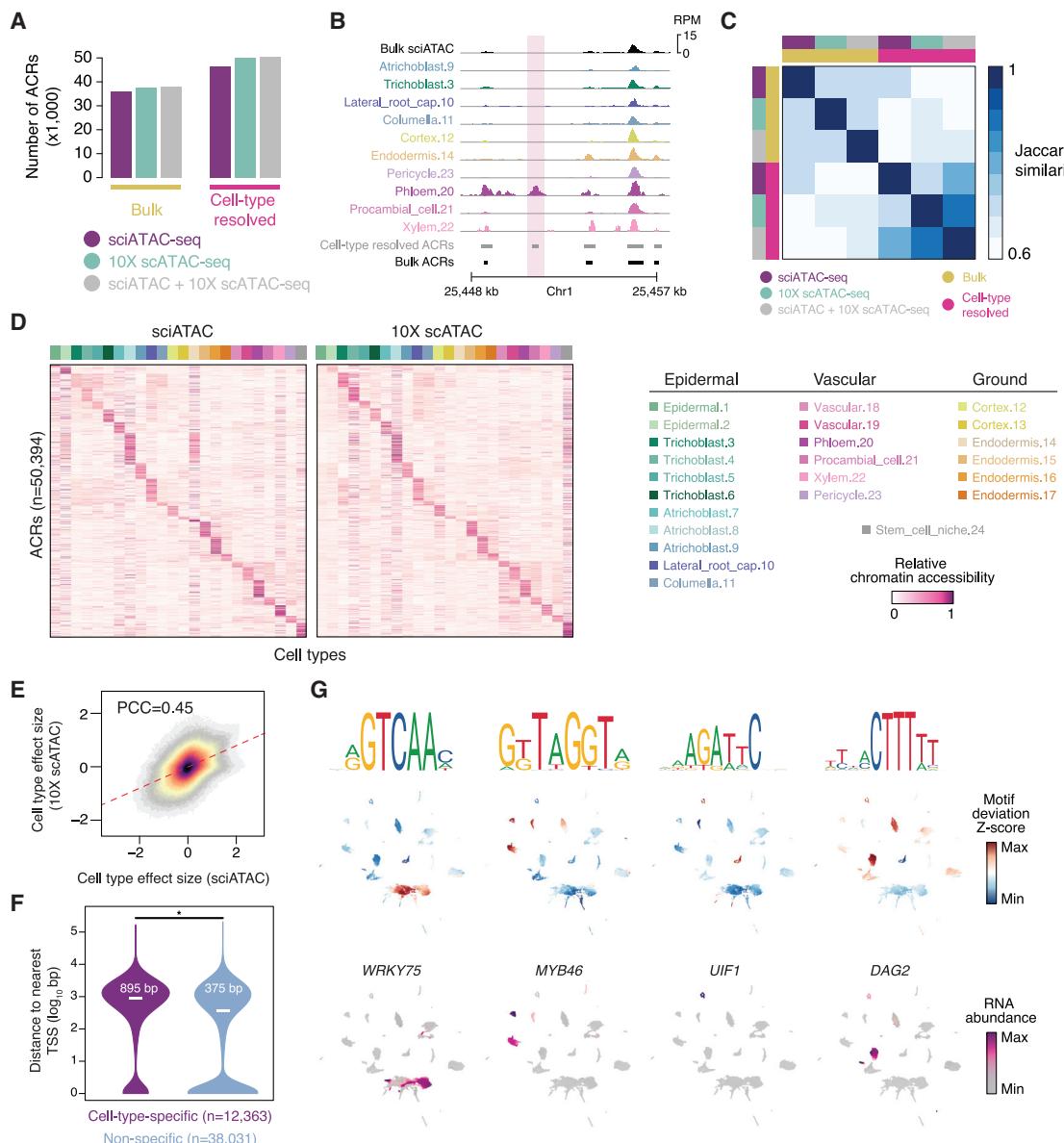
A long-standing challenge in developmental biology is defining the sets of TFs that are involved in generating and maintaining cell type diversity from an invariant genome. To identify transcriptional regulatory sequences underlying accessible chromatin in distinct cell types, we performed global motif enrichment analysis across all ACRs specific to each cell type. The observed enrichment of motif sequences in ACRs relative to background (motif deviation scores) for known regulators was consistent with the predicted cell identities and often correlated with the RNA abundance of the cognate TF (Figure 5G). For example, we found enrichment of *WRKY* motifs in root epidermal progenitors (Marand et al., 2021); *MYB* motifs in the endodermis, cortex, and pericycle (Cohen et al., 2020); *ULT1 interacting factor (UIF)* motifs in the stem cell niche and phloem cells (Moreau et al., 2016); and *DOF AFFECTING GERMINATION (DAG)* motifs in the vascular system (Gualberti et al., 2002). Our findings highlight the ability of scATAC-seq to uncover TF target sequences associated with lineage specification that are important for plant growth and development.

### DISCUSSION

Cellular heterogeneity is a major obstacle to the dissection of gene-regulatory programs that drive plant development. Although great progress has been made with single-cell-based analysis of chromatin status in plants using commercial platforms, there are trade-offs between data quality, throughput, and cost (Marand and Schmitz, 2022). Our study expands the potential application of combinatorial indexing in plant

(E) 10X Genomics scATAC-seq and sci-ATAC-seq pseudobulk cell type Tn5 integration site coverage at 10 cell-type-specific marker gene loci. Track heights denote read coverage scaled per million.

(F) Comparison of relative snRNA-seq gene expression (left) with sci-ATAC-seq (center) and 10X Genomics scATAC-seq (right) gene accessibility across *A. thaliana* root cell types and ~27 000 protein-coding genes.

**Figure 5. Cell-type-specific ACRs and TF regulatory signatures.**

- (A) The number of ACRs identified in bulk (all nuclei) compared with ACR counts identified after merging ACRs from distinct cell types for sci-ATAC-seq, 10X Genomics scATAC-seq, and sci-ATAC-seq + 10X Genomics scATAC-seq datasets.
- (B) Whole-root bulk and pseudobulk cell type Tn5 integration site coverage from sci-ATAC-seq. Track heights denote read coverage scaled per million. The bottom two tracks represent ACRs from cell type and whole-root bulk resolutions, respectively.
- (C) Pairwise Jaccard similarity (fraction of overlapping ACRs versus the union of ACR sets) for ACR calls among various scATAC-seq technologies and resolutions.
- (D) Relative chromatin accessibility across sci-ATAC-seq-based and 10X Genomics scATAC-seq-based cell types and 50 394 ACRs.
- (E) Comparison of beta values from likelihood ratio tests of the contribution of cell type membership toward ACR chromatin accessibility status between sci-ATAC-seq and 10X Genomics scATAC-seq.
- (F) Distributions of ACR distance to the nearest TSS for cell-type-specific and non-specific ACRs. \* $P < 2.2e^{-16}$ , Wilcoxon rank-sum test between the distance to TSS distributions (in base pairs) for cell-type-specific and non-specific ACRs.
- (G) Comparison of matched motif deviation and imputed TF RNA abundance levels for four TFs with cell-type-specific gene expression patterns.

single-cell sequencing and demonstrates its usefulness for single-cell chromatin accessibility profiling in *A. thaliana* root tissue. Compared with scATAC-seq generated using microfluidics systems such as 10X Chromium, our sci-ATAC-seq strategy increases yield and efficiency, does not require cell sorting or commercial reagents, and reduces the cost to  $\sim \$0.01$  per nucleus

(Supplemental information; Supplemental Tables 1–4). A relaxed sci-ATAC-seq workflow for 10–20 plates can be completed in one week to obtain chromatin accessibility information from 15 000–30 000 nuclei. In addition, the ability to safely freeze and store tagged nuclei in 96-well plates during the protocol makes this method highly modular and improves scalability.

# Plant Communications

Our analysis demonstrated that sci-ATAC-seq maps provide information-rich measurements for individual nuclei, enabling identification and annotation of cell types. Moreover, sci-ATAC-seq is able to reveal key gene-regulatory features, such as genomic regions with cell-type-specific chromatin accessibility and underlying TF binding motifs. Throughput may potentially be increased to millions of cells by coupling this approach to a droplet-based microfluidics platform for PCR instead of using 96-well plates. It has been modified to jointly profile chromatin accessibility in animal cells with RNA abundance or other epigenomic features, such as TF binding and histone modifications (Luo et al., 2017; Cao et al., 2018; Zhu et al., 2021). We hope that by removing cost, instrument, and other technical obstacles, this sci-ATAC-seq method can be readily adopted by more plant research laboratories and facilitate new insights into the complex molecular mechanisms underlying cell-specific gene expression, cell fate determination, and plant growth and development.

## METHODS

sci-ATAC-seq

A detailed step-by-step sci-ATAC-seq protocol with reagent and equipment lists is included in the [supplemental information](#). Tn5 expression and purification were performed as described previously ([Tu et al., 2020](#)), and the plasmids were obtained from Addgene (accession number 127916). All computer codes have been deposited in GitHub.

## Nucleus isolation

Seeds of *A. thaliana* were surface sterilized and then sown on half-strength Murashige and Skoog medium plates containing 1% agar without added sucrose. Plates were placed vertically in a 22°C plant growth chamber with a photoperiod of 18 h light, 6 h dark for 2 weeks.

To isolate intact nuclei, roots from ~40 *Arabidopsis* seedlings were first immersed in 1× phosphate buffered saline (PBS) with 1% formaldehyde and 1.5 mM phenylmethanesulfonylfluoride or phenylmethylsulfonyl fluoride (PMSF). The solution was placed under vacuum twice for 5 min, and the roots were washed with distilled water. The fixed tissues were then transferred to a Petri dish with 10 ml pre-chilled cutting buffer and chopped into small pieces to release the nuclei. The solution was then filtered through a 40-μm cell strainer, and nuclei were collected by centrifugation at 500 g and 4°C for 15 min. The pellet was then resuspended with cutting buffer without formaldehyde and filtered through a 10-μm nylon filter. The nuclei were then pelleted again by centrifugation and washed once more with cutting buffer. Before the final centrifugation, a small fraction of the solution was removed and stained with SYBR Green to determine nucleus concentration. We recommend using  $2-5 \times 10^5$  nuclei for each sci-ATAC-seq experiment because using too many nuclei increases clumping during the centrifugation steps. The nucleus pellet was then washed once with 5 ml of fragmentation buffer without dimethylformamide (DMF) and resuspended in 5 ml of fragmentation buffer before addition to the four PCR plates with uniquely indexed Tn5 in each well.

## Tn5 fragmentation and split-pool-split

The sci-ATAC-seq technique depends on the Tn5 transposon to introduce barcoded adapters into individual nuclei for subsequent demultiplexing of single-cell data from the bulk sequencing reads. For such transposon-based ATAC-seq, it is often necessary to experimentally determine an optimized Tn5 concentration to avoid over- or undertagmentation (Orchard et al., 2020). This is particularly challenging for sci-ATAC-seq because nuclei would have to be tagged by hundreds of differentially barcoded transposons. We used the hyperstable TS-Tn5 transposase in which Tn5 is fused to the *Escherichia coli* chaperone elongation

## Combinatorial index single-cell ATAC-seq

factor (Tu et al., 2020). TS-Tn5 has been shown to limit transposition into inaccessible chromatin because of its increased protein size, simultaneously simplifying the assay and improving the signal-to-noise ratio. We observed that read pileup around genes from our method appeared more discrete than that from 10X Genomics scATAC-seq with normal Tn5, consistent with the hypothesized reduction in background integrations afforded by the TS-Tn5 fragmentation enzyme (Figure 2F and 2G; Supplemental Figure 2).

To obtain a combination of 384 indexed transposons, 24 forward Tn5-ME-A and 16 reverse Tn5-ME-B adapters were used to assemble the transposons at 25°C for 1 h. First, 1.5 µl each of A and B transposons were distributed to four 96-well plates. Each well contained a unique combination of A and B indexed Tn5. The nuclei in 12 µl fragmentation buffer were added to each well of the four 96-well plates. The plates were then sealed, and the fragmentation reaction was performed for 45–60 min at 37°C with occasional shaking. The reaction was stopped by flushing each well with 100 µl wash buffer supplemented with ethylenediaminetetraacetic acid (EDTA) to quench the Mg<sup>2+</sup>. All nuclei were transferred to a reservoir and then to a 14-ml round-bottom polystyrene tube. Nuclei were then pelleted and washed with wash buffer. They were resuspended to a concentration of ~10 nuclei/ml, and 1.5 µl of nucleus solution was then pipetted into each well of 96-well plates. Finally, 3 µl of lysis buffer containing SDS was added to each well to lyse the nuclei. The plates were then sealed and stored at –20°C.

## Library amplification

To perform plate PCR with unique row and column barcode primers, the stored nucleus plates were briefly centrifuged and heated to 50°C in a PCR machine for 5 min before addition of 7.5 µl row PCR master mix containing 2% Tween 20 and 7.5 µl column PCR master mix containing DNA polymerase to reconstitute a 20-µl PCR reaction in each well. The detergent in the row PCR master mix is required to quench the SDS in the lysis buffer. The plate was then sealed, and the tagged DNA was PCR amplified to acquire the second barcode that labeled the DNA from each well. The 96 PCR reactions were pooled and purified using a QIAGEN MinElute Kit on a vacuum manifold. The purified PCR products were re-amplified using modified Illumina TruSeq PCR primers ([Supplemental information](#) and [Supplemental Tables 1–4](#)), which gave the tagged DNA a third barcode to label each plate. Libraries prepared from 10 plates (5 for each biological replicate) were sent for Illumina HiSeq X sequencing using the manufacturer’s conventional sequencing reagent and primers. When a HiSeq X lane is used for sequencing the sci-ATAC-seq libraries, 5%–10% spike-in library (e.g., PhiX control from Illumina) must be added to the lane to balance the nucleotide distribution at the beginning of the forward and reverse reads. Spike-in control is not required when the sci-ATAC-seq libraries are sequenced on the latest NovaSeq system, where libraries of different formats are mixed and sequenced in one lane.

## Data processing

Cutadapt (version 2.10) was used to demultiplex the inline barcode of the sci-ATAC-seq data. The 10X Genomics scATAC-seq data were processed as described previously (Marand et al., 2021). Barcode index sequences were appended to the read names of paired-end reads using the extract command of UMItools v.1.0.1 with non-default parameters (`-bc-pattern = NNNNNNNNNNNNNNNNNNNNNNNNNNNNN`) and aligned to the *A. thaliana* TAIR10 reference genome with BWA-MEM v.0.7.17. Aligned reads were filtered with SAMtools v.1.6.0 to remove low-quality alignments (`samtools view -q 30`) and improperly paired reads (`samtools view -f 3`), the Picard Tools v.2.21.6 `MarkDuplicates` function to remove duplicate reads, and a simple PERL script to remove multi-mapped reads (any read pair associated with the XA tag appended by BWA-MEM). The remaining alignments were converted to single-base-resolution Tn5 integration sites by adjusting the start position of forward and reverse strand alignments by +4 and -5, respectively. Only unique Tn5 insertion sites per barcode were retained.

### Nucleus identification and quality control

Nucleus identification and quality control steps were implemented with the R package Socrates v.0.0.1. In brief, single-base-resolution Tn5 insertion sites for each data set were loaded into Socrates with the TAIR10 GFF gene annotation and the TAIR10 genome index. Bulk-scale ACRs were identified in Socrates using callACRs (genome size = 9.5e7, shift = -50, extsize = 100, FDR = 0.05) to estimate the fraction of Tn5 integration sites within ACRs for each nucleus. Metadata for each nucleus were collected using the buildMetaData function with a TSS window size of 2 kb (tss.window = 2000). Thresholds for the minimum number of unique Tn5 insertion sites, the fraction of Tn5 insertion sites within 2 kb of TSSs, and the fraction of Tn5 insertion sites within ACRs were identified manually for each independent dataset (minimum of 1000 Tn5 insertion sites per nucleus, 70% of Tn5 insertions within 2 kb, and 50% of Tn5 insertions within ACRs across all datasets). Doublet enrichment scores were estimated in Socrates using an approach described previously ([Granja et al., 2021](#)).

### snRNA-seq data processing

Raw snRNA-seq data from *A. thaliana* roots were acquired from NCBI GEO: GSE155304. Raw reads were aligned to the TAIR10 reference genome and processed with Cell Ranger v.4.0.0. snRNA-seq quality control and nucleus identification were performed as described previously ([Marand et al., 2021](#)).

### snRNA-seq, scATAC-seq, and sci-ATAC-seq data integration

UMI and Tn5 insertion site count matrices were generated from gene body and gene body +500 bp upstream TSSs for snRNA-seq and scATAC-seq datasets, respectively, and loaded into R using LIGER. Each dataset was normalized independently using the normalize function. Highly variable genes were selected based on the substructure observed in snRNA-seq with the function selectGenes (datasets.use = "snRNA") and used to subset and scale all datasets via the function scaleNotCenter. A joint embedding containing snRNA-seq, scATAC-seq, and sci-ATAC-seq nuclei was generated using iNMF with the function optimizeALS (k = 15, lambda = 5). The metagene factors estimated by iNMF were then quantile normalized with the function quantile\_norm (do.center = F, ref\_dataset = "snRNA") using the snRNA-seq nuclei as the reference to fully integrate all datasets into a shared embedding. Groups of nuclei with similar patterns of gene expression and chromatin accessibility were identified with the function louvainCluster with non-default parameters (resolution = 0.25, k = 35, eps = 0, prune = 1/10). Nucleus–nucleus relationships were visualized by reducing the dimensions of the iNMF loadings with runUMAP (n\_neighbors = 35, min\_dist = 0.01). RNA abundances were predicted for each scATAC-seq and sci-ATAC-seq nucleus by averaging the expression levels of the 20 nearest snRNA-seq nucleus neighbors using the function imputeKNN.

### Independent clustering analysis for sci-ATAC-seq and 10X scATAC-seq

Independent analysis of sci-ATAC-seq and 10X scATAC-seq data was performed using the R package Socrates ([Marand et al., 2021](#)). In brief, a sparse nuclei by 500-bp window matrix scored by the presence/absence of Tn5 insertions was filtered to remove nuclei with fewer than 100 accessible windows and windows in which less than 0.1% of nuclei had a Tn5 insertion. The filtered matrices were normalized using a quasi-binomial logistic regression regularized model with the default settings of the function regModel. The dimensionality of the normalized accessibility scores was reduced using the function reduceDims (n.pcs = 50, cor.max = 0.8). The reduced embedding was visualized as a UMAP embedding using projectUMAP (k.near = 35, m.dist = 0.05). Finally, we identified graph-based clusters using the Louvain neighborhood clustering algorithm with callClusters (res = 0.6, threshold = 5, e.thresh = 3, k.near = 35). Because of lab-specific batch effects in the 10X scATAC-seq data, we used the R package Harmony ([Korsunsky et al., 2019](#)) to remove batch effects using a covariate for lab of origin prior to graph-based clustering with non-default parameters (do\_pca = F, max.iter.harmony = 30, theta =

2, lambda = 0.1). Adjusted Rand index scores were determined using the R package mclust. Reciprocal NNLS between sci-ATAC-seq clusters and 10X scATAC-seq clusters was performed as described previously ([Domcke et al., 2020](#)) using the R package nnls.

### Cell type annotation

To enable cell type annotations for each individual nucleus, we devised an annotation strategy that estimates the relative enrichment of cell-type-specific markers against permuted sets of randomly selected cell-type-specific genes. Specifically, we collected a list of known cell-type-specific marker genes for various cell types in the root of *A. thaliana* and estimated Z scores of gene expression for these genes across nuclei specifically using the snRNA-seq dataset. For each cell type, we estimated the average Z score from X markers specific to the given cell type. Then we derived 1000 random sets of X cell-type-specific markers, excluding the cell type of interest, estimating the average Z score of each randomized set. The relative cell-type-specific enrichment was then estimated as follows Equation:

$$\text{cell-type enrichment} = \frac{u_i - u_r}{\sqrt{\sum_{k=1}^{1000} (x_k - u_r)^2 / 1000}}$$

where  $i$  denotes the average Z score for the cell type of interest,  $r$  is the average Z score from the randomized permuted gene sets, and  $k$  is the index for the vector of permuted gene sets. Because of the high degree of sparsity and dropouts in single-cell experiments, we smoothed the cell type enrichment scores using a diffusion-based weighted affinity matrix derived from the iNMF loadings, 25 nearest neighbors, and three diffusion time steps using an approach described previously for imputing single-cell gene expression values ([Welch et al., 2019](#)). In brief, we first computed a nucleus–nucleus distance matrix using the iNMF loadings and transformed distances into nucleus–nucleus affinities with a Gaussian kernel function. The affinity matrix was then symmetrized to reduce noise and the effect of outliers by addition of the transposed affinity matrix and rescaling of the rows to sum to one, resulting in a Markov transition matrix ( $M_{i,j}$ ) that represents the probability of a transition between nucleus  $i$  and  $j$ . To remove sources of technical noise, including dropouts and undersequencing, we used “diffusion time,”  $t$ , to raise  $M$  to the power  $t$ . Smoothed cell type enrichment scores were estimated by matrix multiplication of the Markov transition matrix with the original nucleus by cell type matrix. The cell type enrichment scores were converted into a probability distribution by rescaling the sum across all possible cell types. Nuclei were labeled with the cell type with the greatest probability ([Welch et al., 2019](#)).

To extend snRNA-seq cell type annotations to scATAC-seq and sci-ATAC-seq within the same embedding, we trained a weighted knn model on the snRNA-seq cell type annotations using the iNMF loadings with the R function kknn and assigned cell type annotations based on the highest probability. From the full embedding, clusters were labeled as the cell type with the greatest frequency among nuclei. Automated cell type annotations were assessed and revised as necessary by computing differential gene expression and accessibility and visualizing normalized gene expression and accessibility values on the reduced UMAP embedding as described previously.

### ACR identification

ACRs were first identified using the pseudobulk Tn5 integration sites (ignoring barcode information) for droplet-based scATAC-seq and sci-ATAC-seq independently with MACS2 v.2.2.7.1 (-nomodel -keep-dup all --extsize 100 --shift -50 --qvalue 0.05). ACRs for scATAC-seq and sci-ATAC-seq were then filtered using an empirical FDR method ( $FDR < 0.01$ ) as described previously ([Hufford et al., 2021](#)). Following nucleus clustering and annotation, we performed an additional round of ACR identification (using the same parameters and filtering steps described above) for each cluster using all Tn5 integration sites regardless of

technology (scATAC-seq + sci-ATAC-seq, hereafter referred to as “combined scATAC-seq”) as well as scATAC-seq technology in isolation by aggregating Tn5 integration sites for barcodes associated with the focal cluster, generating ACR calls from each cluster for the combined scATAC-seq and scATAC-seq and sci-ATAC-seq in isolation. Cluster-based ACRs were merged into single BED files for the combined dataset, scATAC-seq, and sci-ATAC-seq independently using the BEDtools merge command. Concordance among bulk, cell-type-resolved, and scATAC-seq technologies was performed using the BEDtools jaccard function. Differential chromatin accessibility was determined as described previously, except that all nuclei were included in the per-ACR and per-cluster models (replacing the randomly selected nuclei as a reference set). The differential accessibility logistic regression models were determined independently for scATAC-seq and sci-ATAC-seq nuclei. The beta values (effect sizes) from the likelihood ratio tests (between a model including cell type membership and nucleus read depth and a model with only nucleus read depth) were extracted for each ACR and cell type for scATAC-seq and sci-ATAC-seq datasets. Thus, the effect sizes represent the contribution of cell type identity to the relative chromatin accessibility of a locus conditioned by the cognate scATAC-seq technology. Finally, we defined differential accessibility as ACRs with FDR <0.1 and beta >1 in both datasets, resulting in a total of 12 363 differential ACRs and an average of 515 per cell type.

#### TF motif analysis

To determine the relative enrichment of various TF motifs within ACRs for each nucleus, we first constructed a sparse binarized matrix of ACRs by nucleus by scoring the presence/absence of Tn5 integration sites for each nucleus using the combined scATAC-seq ACRs with differential chromatin accessibility (FDR <0.05 and log<sub>2</sub> foldchange > 2). The sparse ACR by nucleus matrix was filtered to retain ACRs that were accessible in at least 100 nuclei, followed by removal of nuclei with fewer than 100 accessible ACRs and a fraction of Tn5 insertions within the remaining ACRs less than 0.1. JASPAR 2020 motif occurrences within differentially accessible ACRs were collected using the function matchMotifs from the R package chromVAR based on significant motif matches (default  $P = 5e^{-5}$ ) to the underlying ACR sequences. Motif deviation scores for each nucleus were then estimated using the deviationScores function after normalizing by GC bias. To visualize motif deviations and remove the effects of variable sequencing coverage and point density on the UMAP embedding, motif deviations were smoothed using the diffusion-based affinity matrix constructed from the multiple modality integrated iNMF loadings and subsequently transformed to Z scores.

#### ACCESSION NUMBERS

The sequencing data have been deposited in the NCBI SRA under accession code PRJNA758591. The R source code and package used throughout the analysis can be found in the following GitHub repository: [https://github.com/plantformatics/plant\\_sciATAC](https://github.com/plantformatics/plant_sciATAC).

#### SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

#### FUNDING

This study was funded with support from the NSFC for Young Scientists (32100438) and the China Postdoctoral Science Foundation (2020M672858 and 2021T140677) (to X.T.); Hong Kong GRF-14104119, GRF-14109420, and AoE/M-403/16 and funding from the State Key Laboratory of Agrobiotechnology (to S.Z.); the NSF (IOS-1856627) and the UGA Office of Research (to R.J.S.); and an NSF postdoctoral fellowship in biology (DBI-1905869 to A.P.M.).

#### AUTHOR CONTRIBUTIONS

R.J.S. and S.Z. designed and supervised the project. X.T. and S.Z. conducted the experiments. A.P.M. performed computational analyses. X.T. and A.P.M. wrote the manuscript. All the authors reviewed and contributed to the paper.

#### ACKNOWLEDGMENTS

R.J.S. is a co-founder of REquest Genomics, a company that provides epigenomic services.

Received: November 7, 2021

Revised: February 22, 2022

Accepted: February 28, 2022

Published: March 2, 2022

#### REFERENCES

- Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N.** (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**:801–806. <https://doi.org/10.1126/science.1146265>.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J.** (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**:1213–1218. <https://doi.org/10.1038/nmeth.2688>.
- Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al.** (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**:1380–1385. <https://doi.org/10.1126/science.aau0730>.
- Chen, X., Shen, Y., Draper, W., Buenrostro, J.D., Litzenburger, U., Cho, S.W., Satpathy, A.T., Carter, A.C., Ghosh, R.P., East-Seletsky, A., et al.** (2016). ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**:1013–1020. <https://doi.org/10.1038/nmeth.4031>.
- Cohen, H., Fedyuk, V., Wang, C., Wu, S., and Aharoni, A.** (2020). SUBERMAN regulates developmental suberization of the *Arabidopsis* root endodermis. *Plant J.* **102**:431–447. <https://doi.org/10.1111/tpj.14711>.
- Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., et al.** (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* **16**:123–131. <https://doi.org/10.1101/gr.407410>.
- Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J.** (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**:910–914. <https://doi.org/10.1126/science.aab1601>.
- Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P.** (2019). Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* **48**:840–852.e5. <https://doi.org/10.1016/j.devcel.2019.02.022>.
- Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A., and Benfey, P.N.** (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* **86**:423–433. [https://doi.org/10.1016/S0092-8674\(00\)80115-4](https://doi.org/10.1016/S0092-8674(00)80115-4).
- Domcke, S., Hill, A.J., Daza, R.M., et al.** (2020). A human cell atlas of fetal chromatin accessibility. *Science* **370**:eaba7612. <https://doi.org/10.1126/science.aba7612>.
- Domergue, F., Vishwanath, S.J., Joubès, J., Ono, J., Lee, J.A., Bourdon, M., Alhattab, R., Lowe, C., Pascal, S., Lessire, R., et al.** (2010). Three *Arabidopsis* fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty alcohols associated with suberin deposition. *Plant Physiol.* **153**:1539–1554. <https://doi.org/10.1104/pp.110.158238>.

## Combinatorial index single-cell ATAC-seq

- Dong, P., Tu, X., Chu, P.-Y., Lü, P., Zhu, N., Grierson, D., Du, B., Li, P., and Zhong, S. (2017). 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* **10**:1497–1509. <https://doi.org/10.1016/j.molp.2017.11.005>.
- Dorrity, M.W., Alexandre, C.M., Hamm, M.O., Vigil, A.-L., Fields, S., Queitsch, C., and Cuperus, J.T. (2021). The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *Nat. Commun.* **12**:3334. <https://doi.org/10.1038/s41467-021-23675-y>.
- Farmer, A., Thibivilliers, S., Ryu, K.H., Schiefelbein, J., and Libault, M. (2021). Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in *Arabidopsis* roots at the single-cell level. *Mol. Plant* **14**:372–383. <https://doi.org/10.1016/j.molp.2021.01.001>.
- Fan, C., Fu, G., and Fodor, S. (2015). Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**:1258367.
- Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**:403–411. <https://doi.org/10.1038/s41588-021-00790-6>.
- Gualberti, G., Papi, M., Bellucci, L., Ricci, I., Bouchez, D., Camilleri, C., Costantino, P., and Vittorioso, P. (2002). Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of *Arabidopsis* seeds. *Plant Cell* **14**:1253–1263. <https://doi.org/10.1105/tpc.010491>.
- Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., et al. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**:655–662. <https://doi.org/10.1126/science.abg5289>.
- Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorrity, M.W., Saunders, L., Bubb, K.L., Trapnell, C., Fields, S., Queitsch, C., and Cuperus, J.T. (2019). Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* **31**:993–1011. <https://doi.org/10.1105/tpc.18.00785>.
- Korsunsky, I., Millard, N., Fan, J., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**:1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>.
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* **518**:317–330. <https://doi.org/10.1038/nature14248>.
- Lareau, C.A., Duarte, F.M., Chew, J.G., Kartha, V.K., Burkett, Z.D., Kohlway, A.S., Pokholok, D., Aryee, M.J., Steemers, F.J., Lebofsky, R., et al. (2019). Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **37**:916–924. <https://doi.org/10.1038/s41587-019-0147-6>.
- Luo, C., Keown, C.L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J.R., Sandoval, J.P., et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**:600–604. <https://doi.org/10.1126/science.aan3351>.
- Marand, A.P., Chen, Z., Gallavotti, A., and Schmitz, R.J. (2021). A cis-regulatory atlas in maize at single-cell resolution. *Cell* **184**:3041–3055.e21. <https://doi.org/10.1016/j.cell.2021.04.014>.
- Marand, A.P., and Schmitz, R.J. (2022). Single-cell analysis of cis-regulatory elements. *Curr. Opin. Plant Biol.* **65**:102094. <https://doi.org/10.1016/j.pbi.2021.102094>.

## Plant Communications

- Minnoye, L., Marinov, G.K., Krausgruber, T., Pan, L., Marand, A.P., Secchia, S., Greenleaf, W.J., Furlong, E.E.M., Zhao, K., Schmitz, R.J., et al. (2021). Chromatin accessibility profiling methods. *Nat. Rev. Methods Primers* **1**:1–24. <https://doi.org/10.1038/s43586-020-00008-9>.
- Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shores, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**:699–710. <https://doi.org/10.1038/s41586-020-2493-4>.
- Moreau, F., Thévenon, E., Blanvillain, R., Lopez-Vidriero, I., Franco-Zorrilla, J.M., Dumas, R., Parcy, F., Morel, P., Trehin, C., and Carles, C.C. (2016). The Myb-domain protein ULTRAPETALA1 INTERACTING FACTOR 1 controls floral meristem activities in *Arabidopsis*. *Development* **143**:1108–1119. <https://doi.org/10.1242/dev.127365>.
- Orchard, P., Kyono, Y., Hensley, J., Kitzman, J.O., and Parker, S.C.J. (2020). Quantification, dynamic visualization, and validation of bias in ATAC-seq data with ataqv. *Cell Syst.* **10**:298–306.e4. <https://doi.org/10.1016/j.cels.2020.02.009>.
- Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**:176–182. <https://doi.org/10.1126/science.aam8999>.
- Rodriguez-Villalon, A., Gujas, B., Kang, Y.H., Breda, A.S., Cattaneo, P., Depuydt, S., and Hardtke, C.S. (2014). Molecular genetic framework for protophloem formation. *PNAS* **111**:11551–11556. <https://doi.org/10.1073/pnas.1407337111>.
- Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**:925–936. <https://doi.org/10.1038/s41587-019-0206-z>.
- Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M., Zhu, Y., O’Malley, R.C., Brady, S.M., et al. (2019). High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* **27**:2241–2247.e4. <https://doi.org/10.1016/j.celrep.2019.04.054>.
- Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* **489**:75–82. <https://doi.org/10.1038/nature11232>.
- Tu, X., Mejia-Guerra, M.K., Valdes Franco, J.A., Tzeng, D., Chu, P.-Y., Shen, W., Wei, Y., Dai, X., Li, P., Buckler, E.S., et al. (2020). Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nat. Commun.* **11**:5089. <https://doi.org/10.1038/s41467-020-18832-8>.
- Welch, J.D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E.Z. (2019). Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**:1873–1887.e17. <https://doi.org/10.1016/j.cell.2019.05.006>.
- Wysocka-Diller, J.W., Helariutta, Y., Fukaki, H., Malamy, J.E., and Benfey, P.N. (2000). Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot. *Development* **127**:595–603. <https://doi.org/10.1242/dev.127.3.595>.
- Zhu, C., Zhang, Y., Li, Y.E., Lucero, J., Behrens, M.M., and Ren, B. (2021). Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat. Methods* **18**:283–292. <https://doi.org/10.1038/s41592-021-01060-3>.