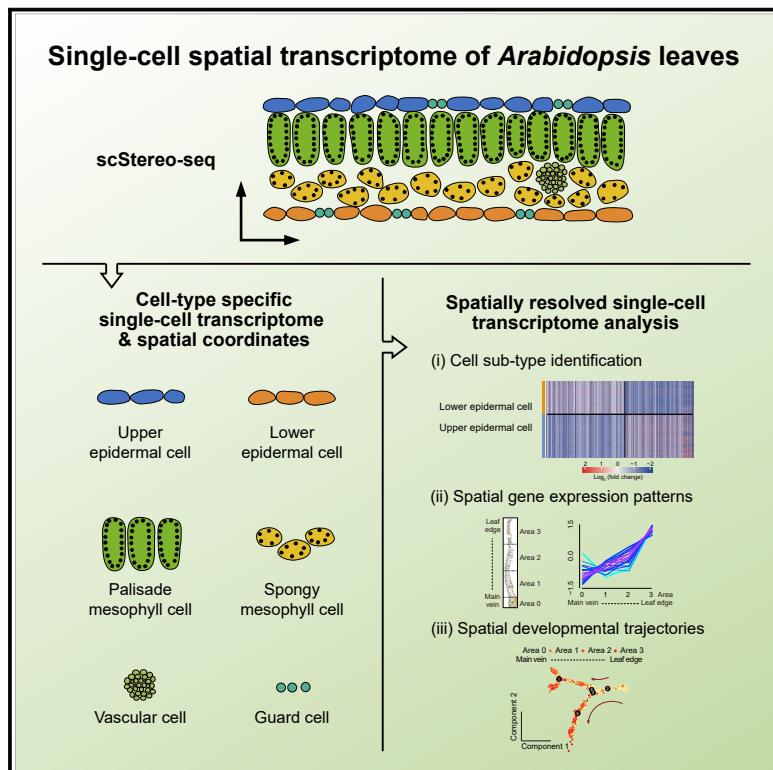


Developmental Cell

The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in *Arabidopsis* leaves

Graphical abstract



Authors

Keke Xia, Hai-Xi Sun, Jie Li, ..., Jian Wang, Ying Gu, Xun Xu

Correspondence

wangjian@genomics.cn (J.W.), guying@genomics.cn (Y.G.), xuxun@genomics.cn (X.X.)

In brief

Xia et al. establish the single-cell spatial transcriptome technique (scStereo-seq) in *Arabidopsis* leaves and discover subtle but significant transcriptional differences between cell subtypes, demonstrating that scStereo-seq is a powerful tool to integrate single-cell location and transcriptome information for plant biology study.

Highlights

- Establishing an *in situ* single-cell spatial transcriptome method in plant
- Identification of cell subtypes and subtle transcriptional differences
- Cell-type-specific gene expression gradients from the main vein to the leaf edge
- scStereo-seq enables the construction of spatially resolved developmental trajectories



Resource

The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in *Arabidopsis* leaves

Keke Xia,^{1,8} Hai-Xi Sun,^{1,2,8} Jie Li,^{1,2,8} Jiming Li,^{1,8} Yu Zhao,^{1,8} Lichuan Chen,¹ Chao Qin,¹ Ruiying Chen,^{1,3} Zhiyong Chen,¹ Guangyu Liu,¹ Ruilian Yin,^{1,2} Bangbang Mu,¹ Xiaojuan Wang,¹ Mengyuan Xu,¹ Xinyue Li,¹ Peisi Yuan,¹ Yixin Qiao,¹ Shijie Hao,^{1,2} Jing Wang,¹ Qing Xie,¹ Jiangshan Xu,^{1,2} Shiping Liu,¹ Yuxiang Li,¹ Ao Chen,¹ Longqi Liu,^{1,4} Ye Yin,⁵ Huanming Yang,^{1,6} Jian Wang,^{1,6,*} Ying Gu,^{1,7,*} and Xun Xu^{1,7,9,*}

¹BGI-Shenzhen, Shenzhen 518083, Guangdong, China

²College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

³College of Informatics, Huazhong Agricultural University, Wuhan 430070, Hubei, China

⁴Shenzhen Key Laboratory of Single-Cell Omics, BGI-Shenzhen, Shenzhen 518120, Guangdong, China

⁵BGI Genomics, BGI-Shenzhen, Shenzhen 518083, Guangdong, China

⁶James D. Watson Institute of Genome Sciences, Hangzhou 310058, Zhejiang, China

⁷Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen 518120, Guangdong, China

⁸These authors contributed equally

⁹Lead contact

*Correspondence: wangjian@genomics.cn (J.W.), guying@genomics.cn (Y.G.), xuxun@genomics.cn (X.X.)

<https://doi.org/10.1016/j.devcel.2022.04.011>

SUMMARY

Understanding the complex functions of plant leaves requires a thorough characterization of discrete cell features. Although single-cell gene expression profiling technologies have been developed, their application in characterizing cell subtypes has not been achieved yet. Here, we present scStereo-seq (single-cell spatial enhanced resolution omics sequencing) that enabled us to show the *bona fide* single-cell spatial transcriptome profiles of *Arabidopsis* leaves. Subtle but significant transcriptomic differences between upper and lower epidermal cells have been successfully distinguished. Furthermore, we discovered cell-type-specific gene expression gradients from the main vein to the leaf edge, which led to the finding of distinct spatial developmental trajectories of vascular cells and guard cells. Our study showcases the importance of physical locations of individual cells for exerting complex biological functions in plants and demonstrates that scStereo-seq is a powerful tool to integrate single-cell location and transcriptome information for plant biology study.

INTRODUCTION

Leaves are one of the most important organs in plants. A mature leaf is mainly composed of epidermis, mesophyll, and vasculature. In *Arabidopsis*, the epidermis contains guard cells and trichomes that are embedded in the single-cell layer of pavement cells. Between the upper and lower epidermal cells are the palisade and spongy mesophyll, which are the main sites in the leaf for photosynthesis. Embedded in the mesophyll cells are the vascular cells, which are responsible for substance transmission and exchange with other types of cells (Svozil et al., 2015; Tsukaya, 2013). Since individual cell types have specialized functions during leaf development and response to environmental stimuli (Berkowitz et al., 2021), tools for systematic research are needed to understand how distinct types of cells are regulated and how they communicate with each other to build the complex functional network in plant leaves.

Recently, several studies in leaf cell biology using single-cell RNA sequencing (scRNA-seq) technology have greatly increased our understanding of stomata and vasculature development, showing flexible stomatal lineage cell states existing along a continuum and phloem parenchyma cells enriched for transporters (Kim et al., 2021; Lopez-Anido et al., 2021). On the other hand, however, the application of scRNA-seq in plant leaf studies faces several challenges (Gurazada et al., 2021). First, protoplasting is technically difficult for leaves and other organs of many plants. It may artificially cause changes in the expression of hundreds of genes and influence the following transcriptome profiling (Tian et al., 2020). Second, many cell types are resistant to protoplasting, which may yield biased proportions of cell types (Bezrutczyk et al., 2021). Third, tissue dissociation leads to the loss of spatial information about cells, which is not only important for analyzing cell-cell and cell-environment interaction and assigning a specific function for cells at the right physical location but is also crucial for cell-type



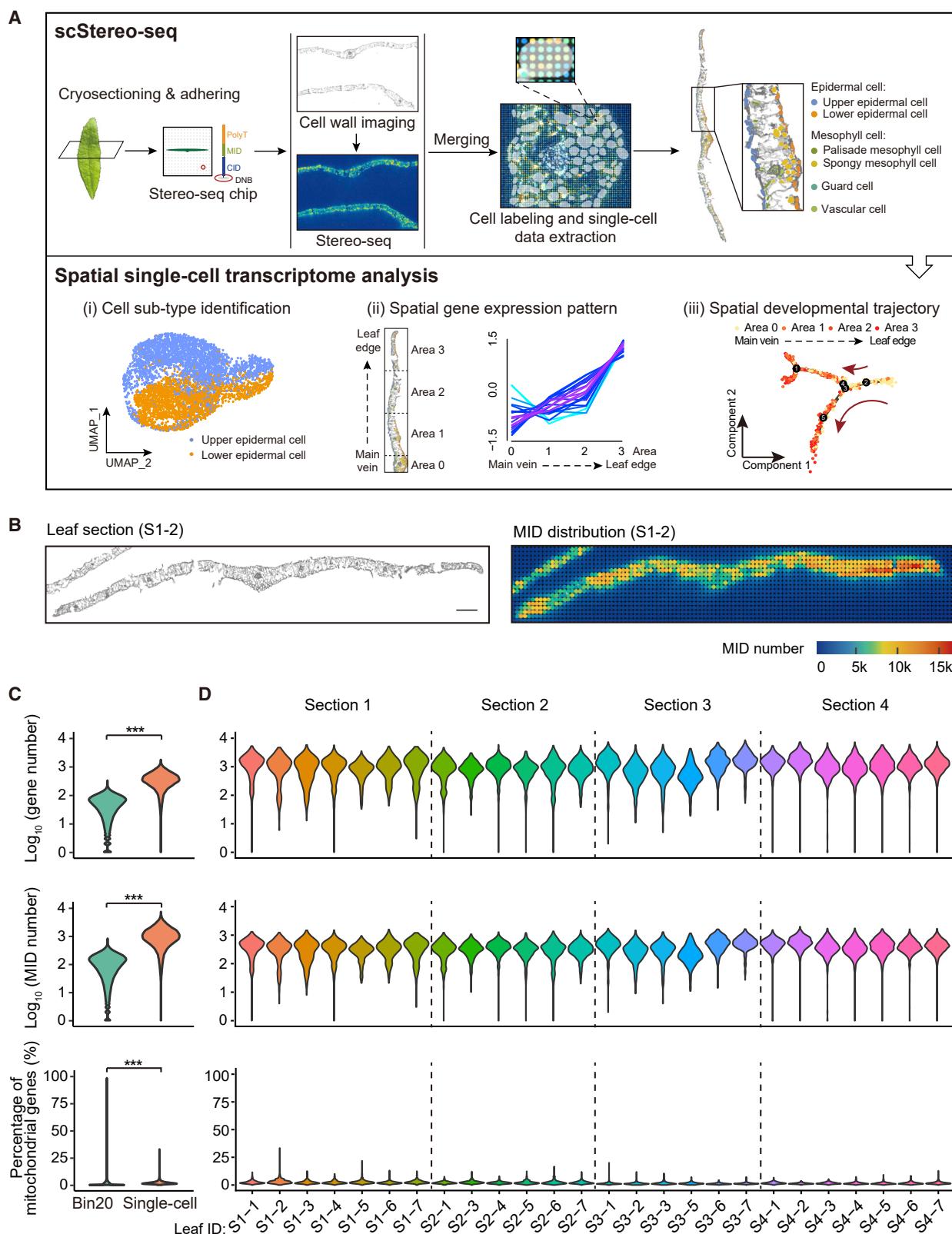


Figure 1. Single-cell level of spatially resolved transcriptome in *Arabidopsis thaliana* caulin leaves

(A) Schematic representation of the single-cell Stereo-seq procedure. *Arabidopsis thaliana* caulin leaves are cryo-sectioned and positioned on top of the chip surface with DNA nanoball (DNB) docked in a grid-patterned array of spots, and the capture probes contain CID (coordinate identity) bar code, MIDs (molecular

(legend continued on next page)

identification, especially for cell types with no known marker genes. Because of these limitations of scRNA-seq for plant studies, the development of gene expression profiling technologies with transcript location information is in great need.

In this regard, several remarkable high-throughput methodologies of *in situ* gene expression profiling have been developed recently (Chen et al., 2021; Cho et al., 2021; Eng et al., 2019; Liu et al., 2020a; Stickels et al., 2021; Wang et al., 2018), such as Slide-seq (Stickels et al., 2021), DBiT-seq (deterministic bar coding in tissue for spatial omics sequencing) (Liu et al., 2020a), and 10x Genomics Visium platform (Bergensträhle et al., 2020). Most of these methods have been successfully applied in mammalian studies, leading to many novel insights into development and disease progression (Choi et al., 2021). Up to now, however, only one method has been applied in plant systems, and the technical difficulties of plant tissue preparation remain a challenge (Giacomello et al., 2017). Moreover, although the spatial transcriptome profiling methods were able to detect a comparable number of genes to those detected by typical scRNA-seq technologies, currently, the spatial resolution of detection that most approaches can achieve is still at the tissue/domain level rather than the single-cell level. Among all *in situ* gene expression methodologies, Stereo-seq (spatial enhanced resolution omics sequencing) is so far the only sequencing-based spatially resolved transcriptomics technology with subcellular resolution (Chen et al., 2021), based on which we set out to accomplish *in situ* single-cell transcriptome analysis of plant tissues.

In this study, by combining plant cell wall staining with high-resolution spatial transcriptomics using Stereo-seq, we present *in situ* single-cell transcriptome profiling in mature *Arabidopsis*-sleaves and therefore named this new approach single-cell Stereo-seq (scStereo-seq). It not only enables us to identify the main leaf cell types but also allows us to distinguish cell subtypes with similar transcriptome profiling, including the upper and lower epidermal cell, palisade mesophyll cell, and spongy mesophyll cell. Furthermore, we show the existence of cell-type-specific spatial gene expression gradients from the leaf's main vein to the leaf edge. We reconstructed those gradients to show the developmental trajectories of specific cell types according to their spatial distribution.

RESULTS

Establishing single-cell stereo-seq method with *Arabidopsis* leaves

We first applied Stereo-seq using leaf samples of *Arabidopsis thaliana* to obtain *in situ* transcriptome profiling. Cauline leaves were cryo-sectioned and positioned on top of four separate

chips (or sections) with seven leaf samples on each chip. On the surface of the chip, a DNA nanoball (DNB) is docked in a grid-patterned array of spots. Each spot is 220 nm in diameter and the center-to-center distance between neighboring spots is 500 nm (Figure 1A). The DNB contains random barcoded sequences, the coordinate identity (CID), molecular identifiers (MIDs), and polyT sequence-containing oligonucleotides designed to capture mRNAs. After cell wall staining and imaging, the chips were used for Stereo-seq library construction and data acquisition. In short, mRNA was released from tissue cells through permeabilization and was captured by polyT in the DNB. The released mRNA was then reverse-transcribed and amplified into cDNA, which was used for PCR amplification and library sequencing. The sequencing data were visualized in the STOmics visualization system (<https://stereomap.cngb.org/>) and processed using a series of Stereo-seq exclusive tools, including SAW (<https://github.com/BGIResearch/SAW>) and stereopy (<https://github.com/BGIResearch/stereopy>). In total, we selected data from 26 leaf samples with good morphology in these four chips for further analyses.

To assess whether the detected transcript signal was stringently confined in leaf areas, we assessed the MID distribution on the chip. As expected, the MID distribution was nicely overlapped with the leaf areas, and no transcript signal diffusion was observed (Figures 1B and S1). Next, we performed fluorescent brightener (FB) 28 staining of the plant cell walls on the chip to visualize cell-cell boundary for establishing scStereo-seq. After merging the stained image with the expression data generated by the STOmics visualization system, we then extracted transcriptome information from individually sampled cells based on the coordinate information of the cell wall boundary (hereafter referred to as the single-cell method) and of clustered bins (bin 20, 20 × 20 DNB bins, referred to as the Bin20 method) (Chen et al., 2021). To enhance the accuracy of cell capture and data interpretation, we only selected cells that had clear cell boundaries and could be classified into specific cell (sub) types based on histological features. Based on these criteria, we initially captured 17,406 cells in total. After performing quality control (see STAR Methods), 13,950 high-quality single cells with a total of 19,720 genes detected were kept (Table S1). To investigate the differences between the single-cell method and the Bin20 method, we checked the gene number, MID number, and the percentage of mitochondrial genes in each cell. As shown in Figure 1C, we observed a magnitude of higher gene and MID numbers for the single-cell method in comparison with the Bin20 method. This result indicates that the single-cell method would provide much more transcript information for defined cells than that of a defined bin area, which is essential for accurately identifying cell features in the following analyses. We next

identifiers), and PolyT Oligos to enable the recordation of the spatial coordinates, the identification of unique transcripts per gene, and the capture of mRNAs. After cell wall staining and imaging, the same section is sequenced with Stereo-seq. Through the combined high-resolution image and MIDs, a single-cell level of MID distribution is achieved. A robust extraction method is built to be used in extracting single cells and in the identification of major cell types in cauline leaves. Using spatial single-cell data, several cell subtypes are distinguished (i). Next, the leaf is divided into four distinct parts, and spatial gene expression pattern (ii) and spatial developmental trajectory (iii) are determined.

(B) *Arabidopsis* cauline leaf (leaf #2 of Section #1, S1–2) and MID distribution in the leaf. The leaf samples were stained with FB28 and the picture was transformed into black and white. The color bar represents the number of MIDs. The scale bar represents 500 μm.

(C and D) Violin plots represent the number of genes and MIDs per cell/bin and the percentage of mitochondrial genes per cell/bin of sequenced cells/bins in all leaf samples (C) and in separate leaf samples (D). Asterisks indicate statistically significant differences: (**) p < 0.001.

expanded our data quality examination by calculating the gene number, the MID number, and the percentage of mitochondrial genes in each cell across all 26 leaf samples. The majority of cells had similar gene numbers and MID numbers among all leaf samples, and the gene number was positively correlated with the MID number in each cell (Figures 1D and S2), thus strongly indicating robust and stable RNA capture efficiency of our single-cell method in different sample areas. Taken together, using the newly developed scStereo-seq protocol, we obtained high-quality and spatially resolved single-cell transcriptome profiles for more than 10,000 cells from *Arabidopsis* caulin leaves.

Identifying cell-type-specific transcriptomes at bona fide single-cell level

Based on histological features and their spatial information in the leaves, the 13,950 cells were classified into 5,139 epidermal cells (2,898 upper and 2,241 lower epidermal cells), 7,490 mesophyll cells (4,254 palisade and 3,236 spongy mesophyll cells), 377 guard cells, and 944 vascular cells (Figure 2A; Table S1). The comparable gene number, UMI number, and percentage of mitochondrial genes per cell were observed for all four major cell types (Figure 2B), implying that our approach was able to capture transcripts of major cell types in *Arabidopsis* leaves with similar efficiency.

To confirm the reliability of the cell-type assignment above, we compared the expression level of marker genes of each cell type, and also performed principal component analysis (PCA) and differentially expressed gene (DEGs) analyses among the four major cell types. The results showed that each major cell type expressed a unique set of marker genes for corresponding cell types (the top-five up-regulated genes were listed) (Figure 2C; Table S2). Meanwhile, the percentage of cells expressing cell-type markers of the single-cell method was significantly higher than those using the Bin20 method (Figures 2D, S3A, and S3B), demonstrating that the single-cell method captured the transcriptome from one cell at a higher possibility than the Bin20 method. Moreover, PCA and differential expression analyses using the single-cell transcriptome revealed that the same cell types from different leaf samples were clustered together, while different cell types from different leaf samples were separately grouped (Figures 2E, 2F, S3C, and S3F). Altogether, our results demonstrated that scStereo-seq can reliably capture individual cells with both transcriptome and spatial information.

We further validated the reliability of the four cell types in the following two aspects. On one hand, we detected the expression of the previously reported canonical marker genes in each cell type. We found that these genes were highly expressed in their corresponding cell types and lowly expressed in other cell types (Figures 3A and S4A, the upper panel) (Kim et al., 2021; Liu et al., 2020b; Zhang et al., 2021). For example, *FAMA* and *SWEET11* were specifically expressed in guard cells and vascular cells, respectively (Figures 3A and S4A). In addition, these marker genes were specifically expressed in the right loci where the corresponding cells reside (Figures 3A and S4A, the lower panel). On the other hand, we investigated the biological processes enriched in each cell type and found biological processes well-matched to the specialized functions of corresponding cell types (Figure 3B). For instance, vascular cells were enriched in

transportation-related pathways, while mesophyll cells were enriched in photosynthesis-related pathways and sucrose metabolic processes (Figure 3B) (Ruan, 2014). These results demonstrated the reliability of cell-type identification using the single-cell method of scStereo-seq.

Identification of cell subtypes and subtle transcriptional differences by scStereo-seq

For *Arabidopsis* leaves, epidermis can be subdivided into upper epidermal cells and lower epidermal cells, and mesophyll cells can also be subdivided into spongy mesophyll cells and palisade mesophyll cells. However, because of highly similar transcriptomes between cell subtypes, previous scRNA-seq data and analysis alone were not able to distinguish these cell subtypes (Kim et al., 2021; Liu et al., 2020b; Zhang et al., 2021). Similar to previous studies, using the single-cell transcriptome data of scStereo-seq alone without the spatial histological information of each cell, we cannot distinguish two cell subtypes (Figures 3C and S4B). However, by combining the single-cell method with the histological spatial information, we could successfully distinguish upper and lower epidermal cell types, as well as spongy and palisade mesophyll cell types (Figures 3C and S4B). Moreover, we found that accurate cell boundary assignment was also essential for cell subtype assignment, as data from the Bin20 method could not clearly distinguish cell subtypes (Figures S4E and S4F). This was possibly because the defined area in Bin20 method comprises a mixture of different cells, which may significantly dilute the cell-type-specific transcriptome features (Figures 1C and 2D). In sum, we demonstrated that both accurate cell boundary demarcation and the exact locations of identified cells were required for cell subtype classification in plant leaves and, therefore, highlighted the advantages of scStereo-seq technology for studying complex plant tissues compared with other methods.

Various subtypes of cells in leaves may infer different functions, which have not been extensively studied so far. Based on the DEG analysis results, we captured cell-type-specific clusters of genes (Figures 3D and S4C; Table S3). The following Gene Ontology (GO) term analysis of DEGs for upper versus lower epidermal cells revealed that biological processes associated with photosynthesis (Berkowitz et al., 2021), cuticle development, and fatty acid biosynthesis were significantly enriched in upper epidermal cells, while biological terms related to immune response, response to absence of light, and regulation of cell death were significantly enriched in lower epidermal cells (Figure 3E). *WAX2*, which has a metabolic function associated with both cuticle membrane and wax synthesis (Chen et al., 2003), was highly expressed in upper epidermal cells, whereas *DIN6*, which is involved in nitrogen metabolism repressed by light (Lam et al., 1998), was highly expressed in the lower epidermis (Figure 3D). Since guard cells were excluded from the epidermal cell in our dataset, this result indicates that epidermal pavement cells have photosynthetic capacity, consistent with the finding that pavement cells of *Arabidopsis* have chloroplasts (Barton et al., 2016; Berkowitz et al., 2021). For mesophyll cell subtypes, we observed that photosynthesis-related GO terms were significantly enriched in the palisade mesophyll, while stress-related GO terms were enriched in spongy mesophyll cells, which is suggestive of functional distinction as well (Figure S4D). Taken

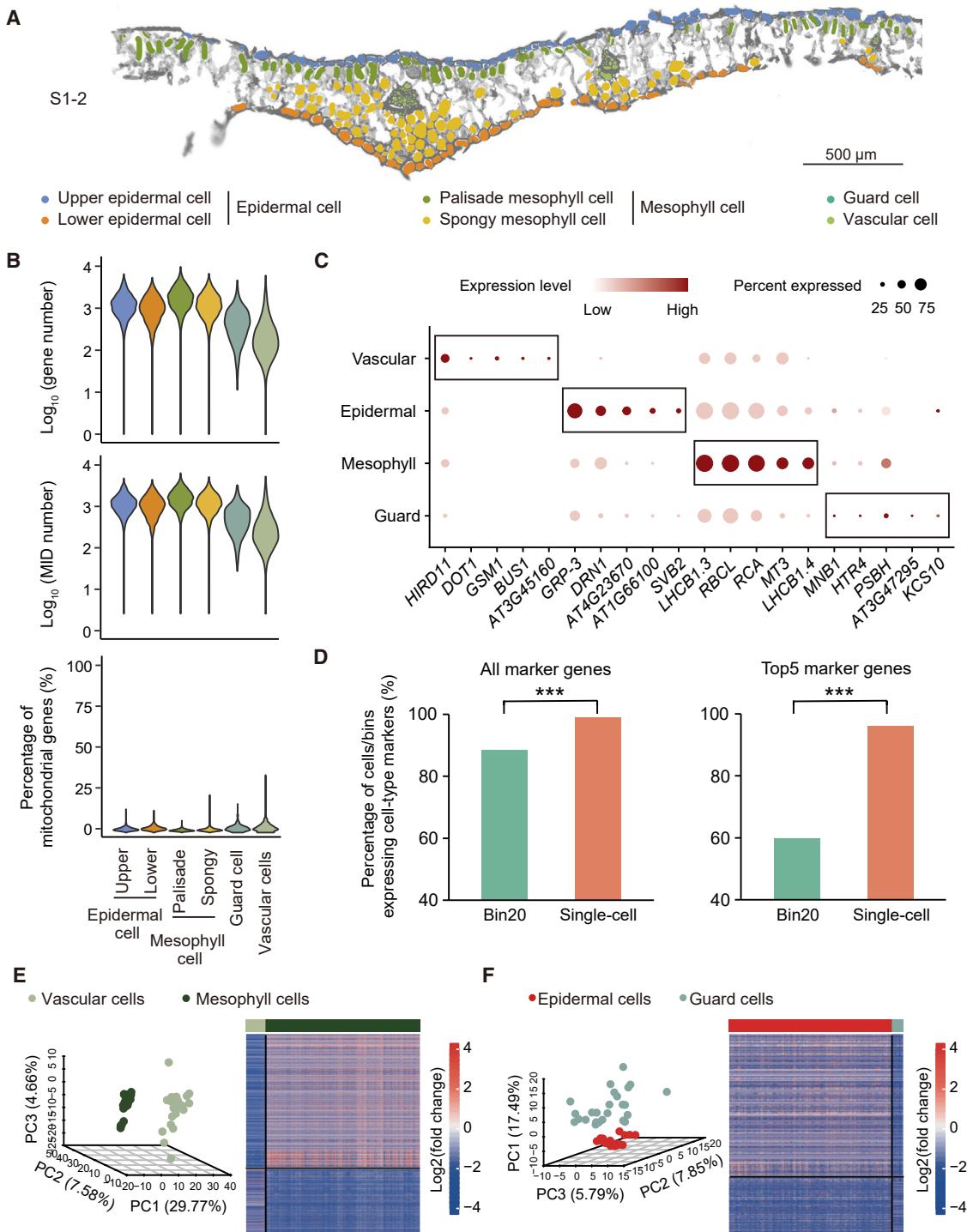


Figure 2. Technical reproducibility and transcriptome diversity of different cell types

(A) Distribution of different cell types in a leaf sample. The scale bar represents 500 μm .

(B) Violin plots show the gene number, the MID number, and the percentage of mitochondrial genes per cell in different cell types.

(C) Dot plots showing the top-five up-regulated marker genes of each cell type. The circle size indicates the relative percentage of cells expressing the marker genes, while color represents the relative expression level.

(D) Percentage of cells/bins expressing cell-type markers. Asterisks indicate statistically significant differences: (***) $p < 0.001$.

(E and F) PCA plot between vascular cells and mesophyll cells in different leaf samples (left) and heatmap showing DEGs between vascular cells and mesophyll cells (right) (E). PCA plot between epidermal cells and guard cells in different leaf samples (left) and heatmap showing DEGs between vascular cells and mesophyll cells (right) (F). Blue and red represent \log_2 -transformed fold change < 0 and > 0, respectively.

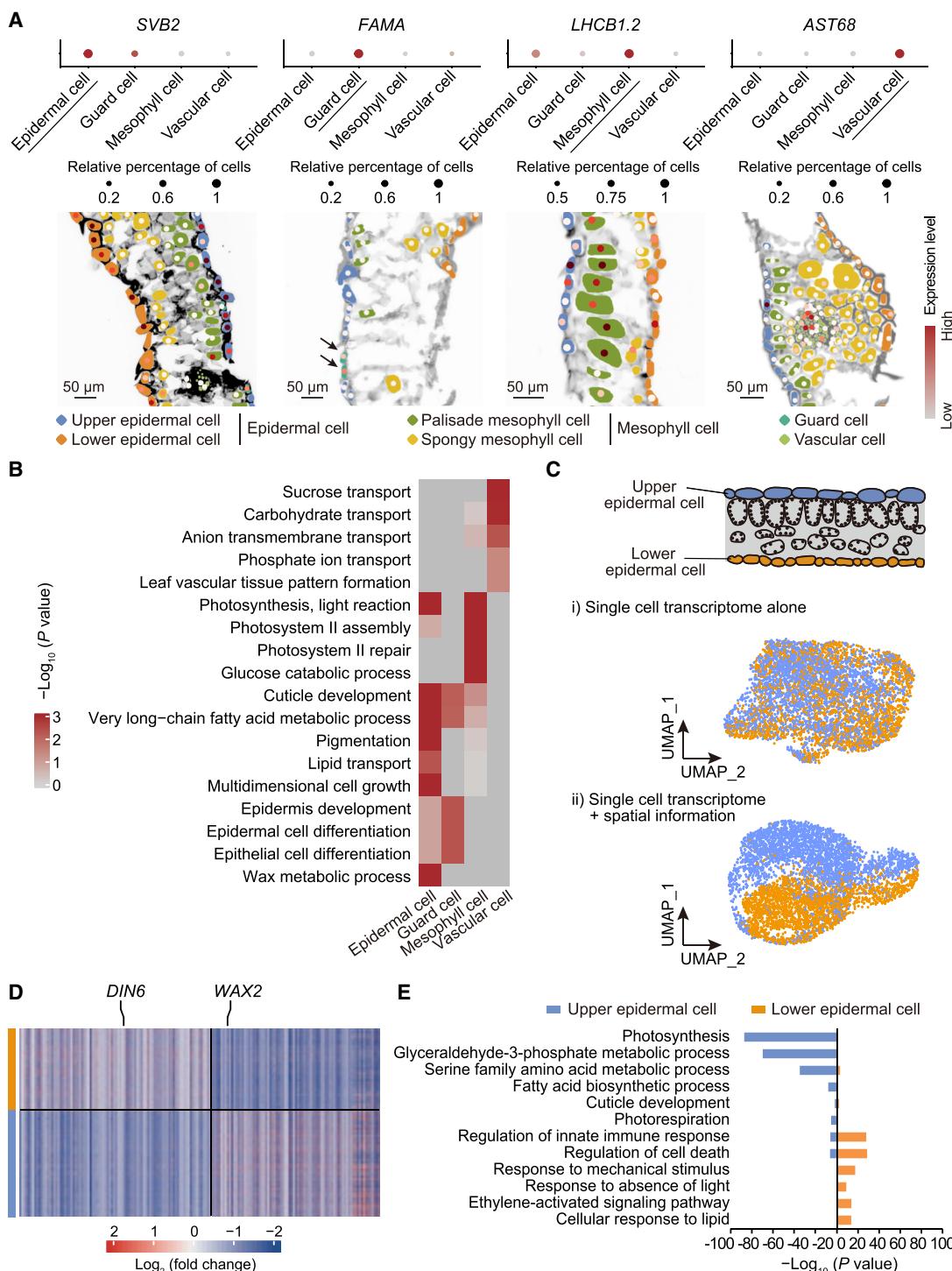


Figure 3. Validation and investigation of transcriptome characteristics of cell types and cell subtypes

(A) Validation of cell types using cell-type-specific marker genes (top panel). Representative images showing expression levels of marker genes in the corresponding cells in leaf samples (lower panel).

(B) Enriched GO terms for four different cell types.

(C) scStereo-seq allows separation between upper and lower epidermal cells. Distribution of upper and lower epidermal cells (i) without spatial information (using variable genes of all epidermal cells) and (ii) with spatial information (using DEGs between upper and lower epidermal cells). Blue dots represent upper epidermal cells and orange dots represent lower epidermal cells.

(D) DEGs between upper and lower epidermal cells. Blue and red represent log₂-transformed fold change < 0 and > 0, respectively.

(E) GO enrichment analysis of DEGs between upper and lower epidermal cells.

together, by distinguishing cell subtypes we revealed the subtle transcriptional differences between upper and lower epidermal cells and spongy and palisade mesophyll cells. This implies distinct functions of cell subtypes and warrants further efforts to explore the molecular mechanisms behind cell-type specification during leaf development, as well as potentially unique roles of cell subtypes under physiological and pathological conditions.

Photosynthesis-related genes exhibit expression gradient from the main vein to the leaf edge

Besides functional difference between cell subtypes, genes specifically expressed in certain areas of leaves play key roles in various plant biological processes, such as leaf development and plant photomorphogenesis (Martinez et al., 2021; Tian et al., 2019; Wang et al., 2020). We took advantage of spatial information from the sampled cell. The main vein area was designated as area 0, and the area on either side of the main vein was divided into three areas, which were designated as areas 1, 2, and 3 (Figure 4A). The data from the same designated areas of each leaf sample were merged. We then determined gene expression patterns in the designated areas of the leaves.

First, we performed gene expression pattern analysis along the leaf for all cells. Interestingly, three patterns of spatial gene expression gradients from the main vein to the leaf edge were observed (Figure 4B). For genes in pattern 1, their expression levels were gradually increased from area 0 to 3, while genes from pattern 3 showed the opposite trend. Genes in pattern 2 showed the highest expression levels in area 1. We then examined the function of genes in different groups. The enriched GO terms for the pattern 1 gene-set were related to photosynthesis and stomatal closure (Figure 4C), suggesting that the photosynthetic efficiency gradually increases from the main vein to the leaf edge. For the list of genes in pattern 3, enriched GO terms are associated with responses to cadmium ion, fluid transport, drug transport, and water-soluble vitamin metabolic processes, as well as a response to zinc ion, well reflecting the main roles of the vein in substance transportation (Figure 4C).

Given that the photosynthesis process mainly takes place in mesophyll cells, we next analyzed data from mesophyll cells to deduce the photosynthetic efficiency in different leaf regions. Gene expression patterns similar to those for all-cell analysis were observed (Figure 4E). Consistently, the enriched GO terms for pattern 1 genes were related to photosynthesis (Figure 4F). Interestingly, genes in patterns 2 and 3 were enriched in GO terms that are associated with fluid transport, hyperosmotic response, and response to mechanical stimulus, suggesting frequent substance exchanges might happen between mesophyll cells and vascular cells in area 0. Representative genes involved in photosynthesis, including *LHCB6*, *LHCB1.2*, and *LHCB1.3* (Voitsekhovskaja and Tyutereva, 2015), showed increased expression from the main vein to the leaf edge in both the all-cell group and mesophyll cell group, but transportation-related genes, such as *ROG1*, *ERD15* (Feng et al., 2022), and *GAPC2*, showed decreased expression (Figures 4D and 4G). Taken together, these results show that distinct gene expression patterns could reflect functional differences in different parts of the leaves. The spatial gene expression patterns along the medio-lateral axis of a leaf provide important

clues for understanding the complex functions of plant leaves from the spatial perspective.

Vascular cells and guard cells exhibit distinct spatially resolved developmental trajectories

It is known that the differentiation at the margin of a leaf is decelerated relative to the more medial regions (Martinez et al., 2021). However, it is unknown whether all types of cells in leaves follow the same development pattern. To tackle these questions, we performed pseudotime analysis on vascular cells and guard cells and compared their developmental stages in the four areas as described in the previous section (Figure 4A). The result showed three major trajectories, and vascular cells from different leaf areas were separately located along the pseudo time path (Figures 5A and 5B). Surprisingly, compared with vascular cells in areas 1, 2, and 3, a significantly large number of vascular cells from area 0 were distributed at a relatively early pseudo time stage (Figures 5A and 5B), while vascular cells from areas 1, 2, and 3 were mainly distributed on other branches (Figures 5A and 5B), suggesting that vascular cells from area 0 were, in general, less differentiated than the cells from areas 1, 2, and 3. Moreover, to further investigate the functions of cells in different developmental stages, we plotted the expression changes of DEGs along the pseudo time, which were classified into three distinct clusters (Figure 5C). For cluster 3, genes were highly expressed at the early stage of the pseudotime axis and enriched in GO terms, which is related to a response to oxidative stress and meristem development, suggesting a higher percentage of pluripotent vascular cells in the main vein area. Genes in cluster 2 were highly expressed at the late stage of the pseudo time axis, corresponding to cells in area 3. Their functions mainly comprised photosynthesis and transportation via GO analysis. Genes in cluster 1 were mainly upregulated in the middle stage of the pseudo time axis and were related to defense and photosynthesis.

Because epidermal and guard cells differentiate from the same basal epidermal cells (Glover et al., 2016), we next investigated the spatial developmental trajectory of epidermal cells and guard cells together. As shown in Figure 5D, epidermal and guard cells were located on different branches with two branches mainly containing epidermal cells, designated as cell fate 1. Another branch that predominantly contained guard cells was designated as cell fate 2 (Figure 5E). However, when adding spatial information to the development trajectory, we did not find a clear spatial distribution pattern for cells from pre-branch and fates 1 and 2, with no obvious difference in the spatial distribution of these cells from area 0 to 3 (Figures S5A and S5B). In addition, we assessed the gene expression pattern of DEGs along the pseudo time from pre-branch to cell fate 2 and from pre-branch to cell fate 1. Four expression clusters were observed (Figure 5F). As expected, GO terms associated with response to stress and stomatal movement were significantly enriched in cell fate 2, which had the highest number of guard cells among all branches (Figure 5F). GO terms related to response to mechanical stimulus and response to desiccation and succinate transmembrane transport were enriched in cell fate 1 (Figure 5F). These data suggest that the development of some vascular cells, but not guard cells, at the margin of the leaf was accelerated relative to the more medial regions. The

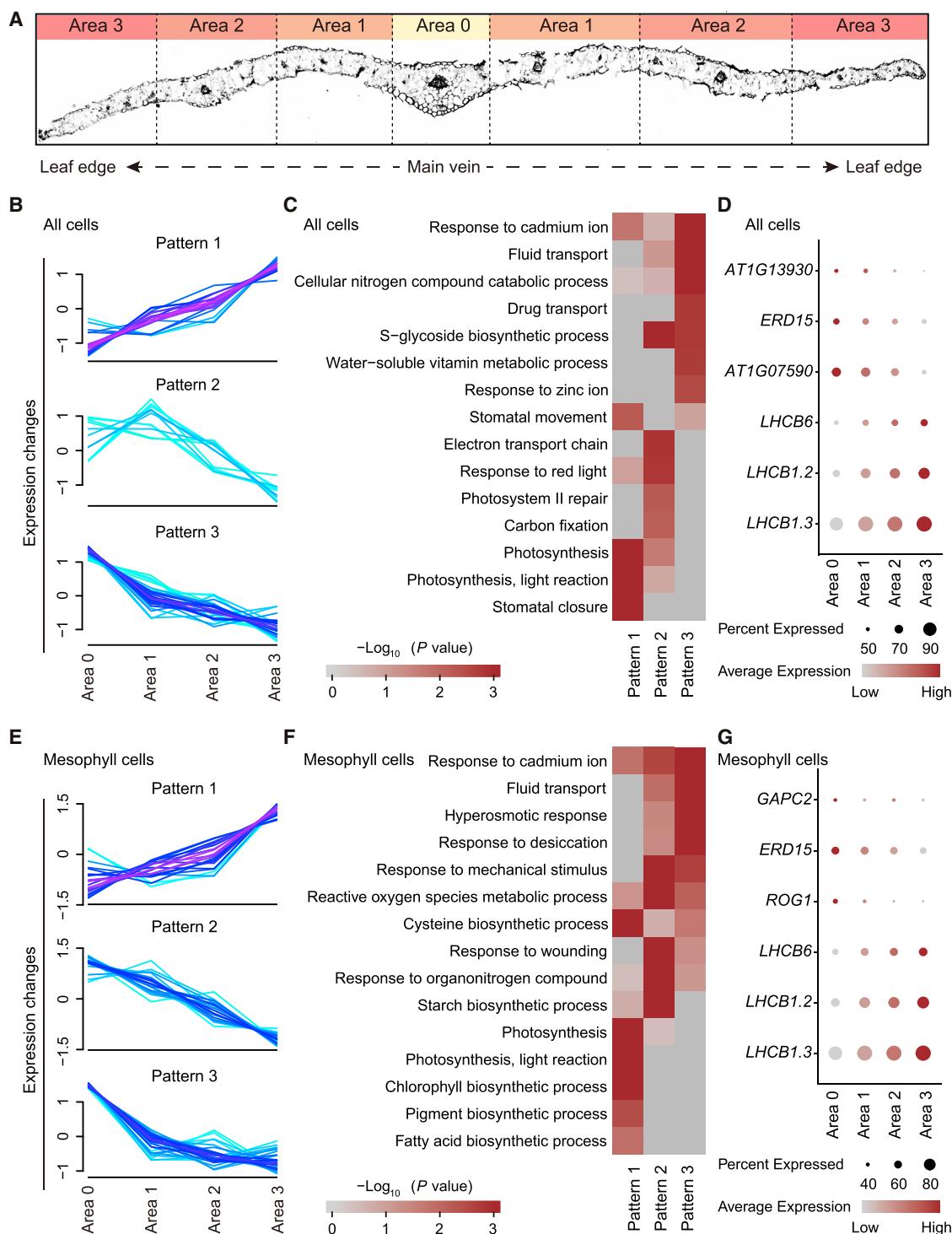


Figure 4. Gene expression patterns along the cross-section of *Arabidopsis* leaves

- (A) Four unique areas (areas 0, 1, 2, and 3) from main vein to leaf edge.
- (B) Gene expression patterns for all cells.
- (C) GO enrichment analysis for genes present in each pattern.
- (D) Dot plot showing expression levels of representative transportation and photosynthesis-related genes in all cells.
- (E) Gene expression patterns for mesophyll cells.
- (F) GO enrichment analysis for genes present in each pattern.
- (G) Dot plot showing expression levels of representative transportation and photosynthesis-related genes in mesophyll cells. For dot plots, the circle size indicates the percentage of cells expressing the marker genes, while color represents expression value. For pattern analysis, only genes with membership > 0.43 are shown.

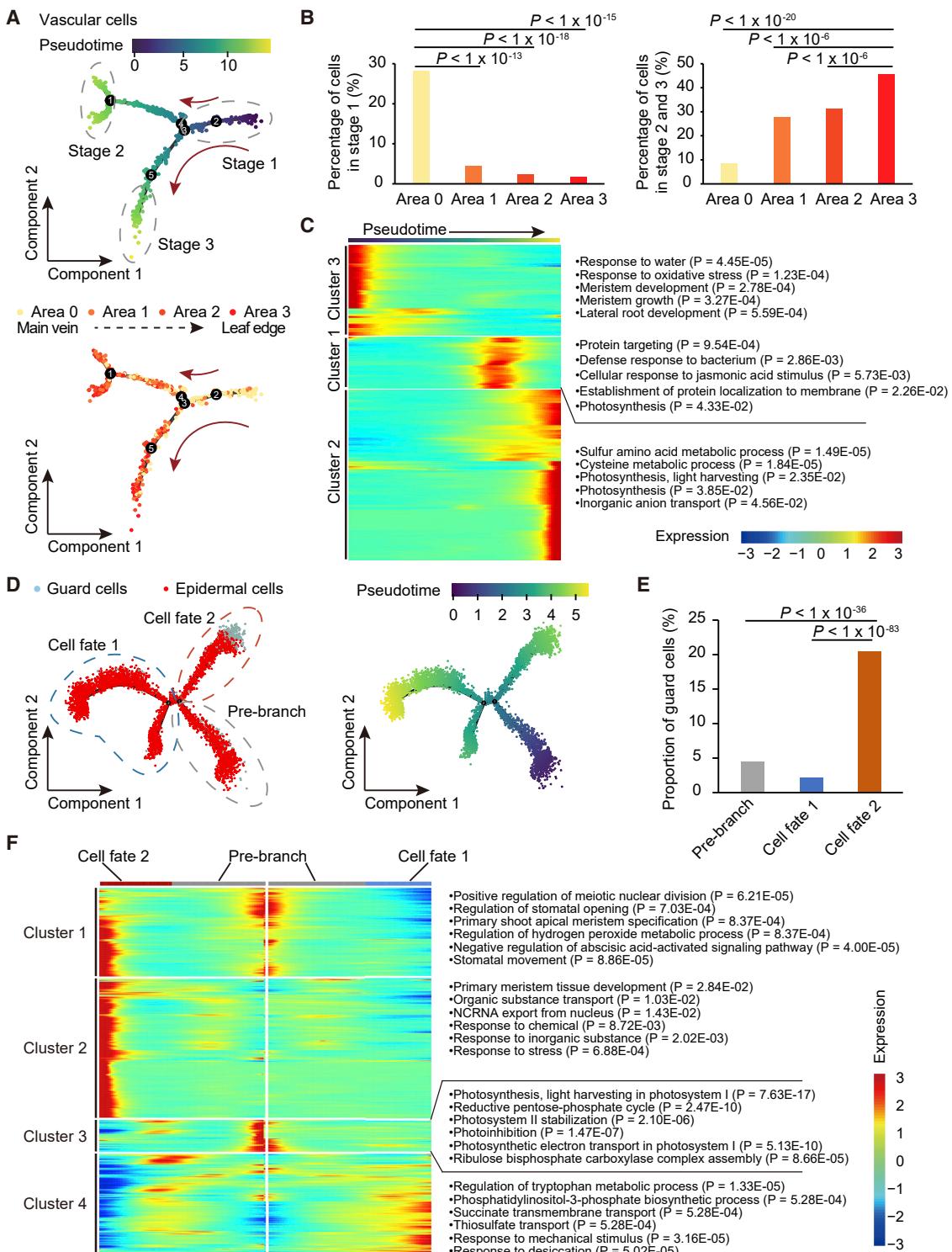


Figure 5. Spatially resolved developmental trajectories of vascular cells and guard cells

(A) Distribution of vascular cells on the pseudotime trajectory (top); distribution of vascular cells from each area on the pseudotime trajectory (bottom).

(B) Percentage of cells for each area in stage 1 (left); percentage of cells for each area in stage 2 and stage 3.

(C) Clustering of differentially expressed genes along a pseudotime progression of vascular cells (left); top-five enriched GO terms for each cluster (right).

(D) Distribution of guard cells and epidermal cells on the pseudotime trajectory branches, including pre-branch, cell fate 1, and cell fate 2.

(E) Proportion of guard cells in each branch.

(F) Clustering of differentially expressed genes along a pseudotime progression of epidermal cells and guard cells (left); representative enriched GO terms for each cluster (right).

development stages with spatial information revealed here could provide important references for future studies on the special development stages of vascular cells.

DISCUSSION

Leaves are the main photosynthetic organs and are also involved in responses to biotic and abiotic stresses (Maugarny-Calès and Laufs, 2018). Different cell types in plant leaves exhibit various molecular signatures and perform specific functions (Berkowitz et al., 2021). Cell-type-specific characterizations are essential for a better understanding of the developmental mechanisms of leaves and their responses to environmental stimuli. With the advantages of scStereo-seq that was developed in this study, we successfully revealed the complex cell-type-specific and spatial-temporal gene expression features of *Arabidopsis* leaves.

Though the plant cell wall is considered to bring difficulties in cell isolation for most single-cell plant studies, our study innovatively used it for accurate single-cell extraction by combining plant cell wall staining with Stereo-seq to obtain well-displayed cell-cell boundaries. The single-cell method displayed a great improvement in data quality compared with the Bin20 method. This method enabled us to generate a *bona fide* single-cell spatial transcriptome profile in plants and to clearly distinguish cell types and subtypes. Although laser capture micro dissection (Martinez et al., 2021), MERFISH (multiplexed error-robust FISH) (Xia et al., 2019) and 1cell-DGE (single-cell digital gene expression) (Kubo et al., 2019) can also perform single-cell profiling with spatial information in plants, these methods suffer from the disadvantage of low throughput and require special equipment that is not widely available for most labs. In contrast, transcriptome signals of cells from 26 leaf samples were easily captured on chips at one time using our method, and the high-throughput *in situ* single-cell transcriptome profile method allowed us to conveniently capture cells in batches without the need for special equipment. Although the classification of epidermal, mesophyll, vascular, and guard cell types has been achieved using single-cell transcriptomes by scRNA-seq in previous studies (Liu et al., 2020b; Lopez-Anido et al., 2021), cell subtypes such as the upper and lower epidermal cell and spongy and palisade mesophyll cell have not been well-differentiated and investigated so far. This is likely because subtypes of cells have highly similar transcriptome profiling and lack specific marker genes, thus requiring more features for subtype identification and comparison. Indeed, by combining spatial information with single-cell transcriptomes, we successfully identified these cell subtypes and obtained their cell-type-specific profiles. These are of course measured with other sequencing methods that do not provide subcellular information, but their inference is relatively facile (Kim et al., 2021). In addition, validation of single-cell sampling is challenging, and cell-type-specific genes might be used as legitimate references for cell-type identification in scRNA-seq studies. Thus, scStereo-seq can be easily extended to a variety of model or non-model plant species, avoiding the need for well-established marker genes for highly similar cell types.

Furthermore, for high-throughput single-cell transcriptome analysis in *Arabidopsis* leaves, a large number of single proto-

plast cells are needed for capturing rare cell types (Kim et al., 2021; Lopez-Anido et al., 2021; Zhang et al., 2021). In contrast, the scStereo-seq procedure does not need protoplast preparation; thus, in principle, more accurate single-cell transcriptome data without biased cell capture can be obtained (Rich-Griffin et al., 2020; Shaw et al., 2021) and may better reflect the real transcriptional characteristics of all cell types in leaves.

We found the gene expression gradient from the main vein to the leaf edge in specific cells, such as *LHCB6*, *LHCB1.2*, and *LHCB1.3* in mesophyll cells, which are important components of photosynthesis (Voitsekhovskaja and Tyutereva, 2015). The low level or absence of photosynthesis-related gene transcription in the main vein region (area 0) is consistent with the main function of leaf vasculature in solute translocation rather than in photosynthesis, reflecting the reliability of our data. The expression levels of *LHCB6*, *LHCB1.2*, and *LHCB1.3* were significantly higher in the leaf edge, indicating that these areas might have a higher photosynthetic capacity or have accumulated more proteins required for photosynthesis in comparison with other leaf areas close to the main vein. Such a unique expression pattern implies a quantitative and spatial difference of functions for the same types of cell, the regulatory mechanisms and significance of which requires future investigation. Meanwhile, we reconstructed those gene expression gradients to show the developmental trajectories of vascular cells and guard cells according to their spatial distribution.

While the scStereo-seq technology enabled us for the first time to reveal the gene expression profiling of *Arabidopsis* leaves at single-cell resolution and location information together, several improvements are expected in future efforts. First, the current single-cell labeling method is low throughput and time consuming. Cell boundary determination algorithms for automatic and high-throughput single-cell capture with scStereo-seq data are being developed, which will be pivotal for large tissue studies in the future. Second, because of the high water-content in plants, cryo-sectioning is challenging for many plants or tissues. Possibly due to this technical issue, we identified fewer cell types than that of the previous study (Kim et al., 2021). In this regard, protocol improvements in sectioning to obtain high-quality cell wall staining and RNA content will enable a more complete global picture of all cell types and their gene expression profiling in a given tissue. Third, further improvements in the resolution of the stereo-seq chips could facilitate identifying tiny cells in certain tissues, such as the plant shoot apex. Lastly, the number of genes detected in each guard cell and vascular cell was relatively less compared with those detected in each palisade and spongy mesophyll cell. It could be that a gene was randomly detected in a subset of cells, and it might have an effect on differential expression analysis. To circumvent this issue, we only chose genes that were expressed in more than 25% of cells in a defined cell-type when performing DEG analysis (see STAR Methods) (Figures 2E and 2F). Although we could partially solve the issue using specified parameters, we are also optimizing the experimental procedure to improve gene capture efficiency.

With continued upgrading as expected above and due to the simplicity of its application and the fewer technical requirement compared with MERFISH (Xia et al., 2019) and FISSEQ (fluorescent *in situ* RNA sequencing) (Lee et al., 2014), scStereo-seq

has great potential to advance future studies in plant biology and other areas of research. For example, we would be able to analyze transcriptomes at the subcellular level, which will be important to understand RNA metabolism and gene posttranscriptional regulation in different compartments of a cell. scStereo-seq can also be applied to analyze how different cell types respond to pathogen infection in plant leaves and roots, and how response signals are transmitted to distant cells and tissues. Moreover, scStereo-seq can be used to analyze the leaf differences between C3 and C4 plants. By comparing transcriptome differences among bundle sheath cells, other vascular cells, and the surrounding mesophyll cells in C3 and C4 plants and constructing developmental trajectories of bundle sheath cells, we could provide insights for transforming C3 plants into C4 plants. In addition, scStereo-seq is also an ideal technology to facilitate building a plant cell atlas (Rhee et al., 2019). In sum, the scStereo-seq technology developed in this study offers a powerful spatially resolved single-cell transcriptomic strategy for a systematic study of plant biology, and it will enable us to understand plants at an unprecedented resolution.

Limitations of the study

While we were able to distinguish cell subtypes, including lower epidermal cell and upper epidermal cell and spongy mesophyll cell and palisade mesophyll cell, we did not manage to distinguish cells belonging to the vasculature, such as xylem and phloem. Further optimization of the experimental procedure to obtain different cell types of the vasculature is underway. In addition, we only obtained spatiotemporal transcriptomic data from the main vein to the leaf edge, and further spatiotemporal transcriptome investigation along the main vein would enable construction of a 3D spatiotemporal transcriptome map of *Arabidopsis* leaves, which would reveal new insights into the development of *Arabidopsis* leaves.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Plant growth condition and tissue collection
 - Stereo-seq chip structure
 - Cryo sectioning, fixation, staining and imaging
 - Permeabilization, reverse transcription, tissue removal and cDNA release, Library preparation and sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Raw Stereo-seq data processing
 - Single-cell data acquisition
 - Quality control of single-cell Stereo-seq data
 - PCA analysis
 - DEG analysis
 - Expression pattern analysis

- Trajectory analysis
- GO enrichment analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.devel.2022.04.011>.

ACKNOWLEDGMENTS

This work was supported by the grant of Top Ten Fundamental Research Institutes of Shenzhen, Guangdong Provincial Key Laboratory of Genome Read and Write (No. 2017B030301011) and Shenzhen Key Laboratory of Single-Cell Omics (No. ZDSYS20190902093613831). We also sincerely thank the China National GenBank (CNGB) for the support they provided. This work is part of the SpatioTemporal Omics Consortium (STOC) paper package. A list of STOC members is available at www.sto-consortium.org.

AUTHOR CONTRIBUTIONS

X.X., Y.G., J. Wang., K.X., and H.-X.S. designed and supervised the study. K.X., H.-X.S., and Y.Z. designed the experiment. Y.Z., L.C., G.L., Z.C., R.Y., B.M., and X.W. performed the library preparation and sequencing. J. Li., J.Li., C.Q., R.C., L.C., S.H., and H.-X.S. performed bioinformatics analysis. M.X., X.L., P.Y., Y.Q., and J.X. provided technical support. J.W., Q.X., S.L., Y.L., A.C., L.L., Y.Y., and H.Y. gave the relevant advice. K.X., H.-X.S., J.-M.L., J. Li., G.L., Z.C., and R.C. wrote the manuscript. K.X., H.-X.S., X.X., and Y.G. participated in the manuscript editing and discussion. All authors edited and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare the following competing interests: the chip, procedure, and applications of Stereo-seq are covered in pending patents. Ye Yin is an employee and shareholder of BGI Genomics.

INCLUSION AND DIVERSITY

The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: August 10, 2021

Revised: January 27, 2022

Accepted: April 6, 2022

Published: May 4, 2022

REFERENCES

- Barton, K.A., Schattat, M.H., Jakob, T., Hause, G., Wilhelm, C., McKenna, J.F., Máthé, C., Runions, J., Van Damme, D., and Mathur, J. (2016). Epidermal pavement cells of *Arabidopsis* Have chloroplasts. *Plant Physiol.* **171**, 723–726.
- Bergensträhle, J., Larsson, L., and Lundeberg, J. (2020). Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 482.
- Berkowitz, O., Xu, Y., Liew, L.C., Wang, Y., Zhu, Y., Hurgobin, B., Lewsey, M.G., and Whelan, J. (2021). RNA-seq analysis of laser micro dissected *Arabidopsis thaliana* leaf epidermis, mesophyll and vasculature defines tissue-specific transcriptional responses to multiple stress treatments. *Plant J.* **107**, 938–955.
- Bezrutczyk, M., Zöllner, N.R., Kruse, C.P.S., Hartwig, T., Lautwein, T., Köhrer, K., Frommer, W.B., and Kim, J.Y. (2021). Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. *PlantCell* **33**, 531–547.
- Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Yang, J., Li, W., Xu, J., Hao, S., et al. (2021). Large field of view-spatially resolved transcriptomics at nanoscale resolution. Preprint at bioRxiv. <https://doi.org/10.1101/2021.01.17.427004>.

- Chen, F.Z., You, L.J., Yang, F., Wang, L.N., Guo, X.Q., Gao, F., Hua, C., Tan, C., Fang, L., Shan, R.Q., and Wei, X.F. (2020). CNGBdb: China national Gene Bank DataBase. *Yi Chuan* 42, 799–809.
- Chen, X., Goodwin, S.M., Boroff, V.L., Liu, X., and Jenks, M.A. (2003). Cloning and characterization of the WAX2 gene of *Arabidopsis* involved in cuticle membrane and wax production. *PlantCell* 15, 1170–1185.
- Cho, C.-S., Xi, J., Park, S.-R., Hsu, J.-E., Kim, M., Jun, G., Kang, H.-M., and Lee, J.H. (2021). Seq-Scope: submicrometer-resolution spatial transcriptomics for single cell and subcellular studies. *bioRxiv*.
- Choi, H., Lee, E.J., Shin, J.S., Kim, H., Bae, S., Choi, Y., and Lee, D.S. (2021). Spatiotemporal characterization of glial cell activation in an Alzheimer's disease model by spatially resolved transcriptome. *bioRxiv*.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics Oxf. Engl.* 29, 15–21.
- Eng, C.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.C., and Cai, L. (2019). Transcriptome-scale super-resolved imaging in tissues by RNaseqFISH. *Nature* 568, 235–239.
- Feng, G., Zhong, Y., and Zou, W. (2022). Lipid transporter LSR1 positively regulates leaf senescence in *Arabidopsis*. *Plant Signal. Behav.* 17, 2007328.
- Giacomello, S., Salmén, F., Terebieniec, B.K., Vickovic, S., Navarro, J.F., Alexeyenko, A., Reimegård, J., McKee, L.S., Mannapperuma, C., Bulone, V., et al. (2017). Spatially resolved transcriptome profiling in model plant species. *Nat. Plants* 3, 17061.
- Glover, B.J., Airoldi, C.A., and Moyroud, E. (2016). Epidermis: outer cell layer of the plant. In eLS (John Wiley & Sons), pp. 1–7.
- Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., Hua, C., Yang, F., Liu, W., Peng, C., et al. (2020). CNSA: a data repository for archiving omics data. Database (Oxford) 2020.
- Gurazada, S.G.R., Cox, K.L., Czymbmek, K.J., and Meyers, B.C. (2021). Space: the final frontier - achieving single-cell, spatially resolved transcriptomics in plants. *Emerg. Top. Life Sci.* 5, 179–188.
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29.
- Kim, J.Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M., Zöllner, N., Hartwig, T., Wudick, M.M., et al. (2021). Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* 33, 511–530.
- Kubo, M., Nishiyama, T., Tamada, Y., Sano, R., Ishikawa, M., Murata, T., Imai, A., Lang, D., Demura, T., Reski, R., and Hasebe, M. (2019). Single-cell transcriptome analysis of *Physcomitrella* leaf cells during reprogramming using microcapillary manipulation. *Nucleic Acids Res.* 47, 4539–4553.
- Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* 2, 5–7.
- Lam, H.M., Hsieh, M.H., and Coruzzi, G. (1998). Reciprocal regulation of distinct asparagine synthetase genes by light and metabolites in *Arabidopsis thaliana*. *Plant J.* 16, 345–353.
- Lee, J.H., Daugherty, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science* 343, 1360–1363.
- Liu, Y., Yang, M., Deng, Y., Su, G., Enninfel, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020a). High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* 183, 1665–1681.e18.
- Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J., Wu, R., Zhang, B., Hu, Y., et al. (2020b). Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol. Plant* 13, 1178–1193.
- Lopez-Anido, C.B., Vatén, A., Smoot, N.K., Sharma, N., Guo, V., Gong, Y., Anleu Gil, M.X., Weimer, A.K., and Bergmann, D.C. (2021). Single-cell resolution of lineage trajectories in the *Arabidopsis* stomatal lineage and developing leaf. *Dev. Cell* 56, 1043–1055.e4.
- Martinez, C.C., Li, S., Woodhouse, M.R., Sugimoto, K., and Sinha, N.R. (2021). Spatial transcriptional signatures define margin morphogenesis along the proximal-distal and medio-lateral axes in tomato (*Solanum Lycopersicum*) leaves. *Plant Cell* 33, 44–65.
- Maugarny-Calès, A., and Laufs, P. (2018). Getting leaves into shape: a molecular, cellular, environmental and evolutionary view. *Development* 145.
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982.
- Rhee, S.Y., Birnbaum, K.D., and Ehrhardt, D.W. (2019). Towards building a plant cell atlas. *Trends Plant Sci.* 24, 303–310.
- Rich-Griffin, C., Stechemesser, A., Finch, J., Lucas, E., Ott, S., and Schäfer, P. (2020). Single-cell transcriptomics: A high-resolution a venue for plant functional genomics. *Trends Plant Sci.* 25, 186–197.
- Ruan, Y.L. (2014). Sucrose metabolism: gateway to diverse carbon use and sugar signaling. *Annu. Rev. Plant Biol.* 65, 33–67.
- Shaw, R., Tian, X., and Xu, J. (2021). Single-cell transcriptome analysis in plants: advances and challenges. *Mol. Plant* 14, 115–126.
- Stickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J., Arlotta, P., Macosko, E.Z., and Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* 39, 313–319.
- Svozil, J., Gruissem, W., and Baerenfaller, K. (2015). Proteasome targeting of proteins in *Arabidopsis* leaf mesophyll, epidermal and vascular tissues. *Front. Plant Sci.* 6, 376.
- Tian, C., Du, Q., Xu, M., Du, F., and Jiao, Y. (2020). Single-nucleus RNA-seq resolves spatiotemporal developmental trajectories in the tomato shoot apex. *bioRxiv*.
- Tian, C., Wang, Y., Yu, H., He, J., Wang, J., Shi, B., Du, Q., Provart, N.J., Meyerowitz, E.M., and Jiao, Y. (2019). A gene expression map of shoot domains reveals regulatory mechanisms. *Nat. Commun.* 10, 141.
- Tsukaya, H. (2013). Leaf development. *Arabidopsis Book* 11, e0163.
- Voitsekhovskaja, O.V., and Tyutereva, E.V. (2015). Chlorophyll b in angiosperms: functions in photosynthesis, signaling and ontogenetic regulation. *J. Plant Physiol.* 189, 51–64.
- Wang, J., Sun, N., Zhang, F., Yu, R., Chen, H., Deng, X.W., and Wei, N. (2020). SAUR17 and SAUR50 differentially regulate PP2C-D1 during apical hook development and cotyledon opening in *Arabidopsis*. *Plant Cell* 32, 3792–3811.
- Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361.
- Xia, C., Fan, J., Emanuel, G., Hao, J., and Zhuang, X. (2019). Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. USA* 116, 19490–19499.
- Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16, 284–287.
- Zhang, T.Q., Chen, Y., and Wang, J.W. (2021). A single-cell analysis of the *Arabidopsis* vegetative shoot apex. *Dev. Cell* 56, 1056–1074.e8.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Cauline leaves of <i>Arabidopsis thaliana</i>	This study	N/A
Chemicals, peptides, and recombinant proteins		
OCT	Leica	4583
FluorescentBrightener 28	SIGMA	F3543-5G
Pepsin	Sigma	P7000
HCl	Dongjiang	N/A
0.1x SSC buffer	Thermo	AM9770
RNase inhibitor	NEB	M0314L
SuperScript II	Invitrogen	18064-014
Tissue Removal buffer	BGI	N/A
Exonuclease I	NEB	M0293L
KAPA HiFi Hotstart Ready Mix	Roche	KK2602
Qubit dsDNA Assay Kit	Thermo	Q32854
SDS buffer	BBI	A600485-0500
1x KAPA HiFi Hotstart Ready Mix	BGI	N/A
Stereo-Library-F primer	BGI	N/A
Stereo-Library-R primer	BGI	N/A
Nuclease-free H ₂ O	Ambion	AM9937
Ampure XP Beads	Vazyme	N411-03
Ethanol	XIOLONGSCIENTIFIC	171225-02
Deposited data		
Raw sequencing data	This study	CNGB: CNP0002618
STOmics database to view these data	This study	https://db.cngb.org/stomics/datasets/STDS0000104
TAIR10 <i>Arabidopsis</i> annotation	TAIR	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/
Experimental models: Organisms/strains		
<i>Arabidopsis</i> : Col-0	N/A	N/A
Software and algorithms		
STAR v2.7.9a	Dobin et al., 2013	https://github.com/alexdobin/STAR
SAW v02.1.0		https://github.com/BGIResearch/SAW
Stereopy v0.2.2		https://github.com/BGIResearch/stereopy
Seurat v4.0.0	Hao et al., 2021	https://github.com/satijalab/seurat
Mfuzz v2.50.0	Kumar et al., 2007	https://bioconductor.org/packages/release/bioc/html/Mfuzz.html
Monocle v2.18.0	Qiu et al., 2017	https://bioconductor.org/packages/release/bioc/html/monocle.html
clusterProfiler	Yu et al., 2012	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R	R Core	https://www.r-project.org/
Python	Python Software Foundation	https://www.python.org/
Other		
Stereo-seq capture chips	BGI	N/A
DNBSEQ-T10 sequencer	BGI	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Xun Xu (xuxun@genomics.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The high-throughput sequencing data that support the findings of this study have been deposited into CNGB Sequence Archive (Guo et al., 2020) of CNGBdb (Chen et al., 2020) with accession number CNP0002618. The analyses results have also been uploaded to STOMICS DB website (<https://db.cngb.org/stomics/datasets/STDS0000104>). The scripts have been uploaded to GitHub (<https://github.com/Jie57Li/Arabidopsis-scStereo-seq-data-process>).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The *Arabidopsis thaliana* plants were used as the experimental model in the study. The Columbia-0 (Col-0) ecotype was used as wild type.

METHOD DETAILS

Plant growth condition and tissue collection

Arabidopsis thaliana (Col-0) seeds were sown and incubated at 25 °C in the photoperiod of 16 hours light / 8 hours dark after surface sterilization with 8% sodium hypochlorite in 0.1% triton X-100 solution. Fresh leaves were collected from six-week-old seedlings. Then tissues were embedded in pre-cooled OCT (Sakura) and stored at -80 °C until processed.

Stereo-seq chip structure

The Stereo-seq capture chips (Chen et al., 2021) were used in this study. To generate the DNA nanoball (DNB) array for *in situ* RNA capture, we first synthesized random 25-nt CID (coordinate identity) -containing oligos, circularized with T4 ligase and splint oligos. DNBS were then generated by rolling circle amplification and were loaded onto the patterned chips (65 mm × 65 mm). The chip surface consists of DNA nanoball (DNB) containing random barcoded sequences, the coordinate identity (CID), molecular identifiers (MID) and polyT sequence-containing oligonucleotides. The DNBS are docked in a grid-patterned array of spots, each spot being approximately 220 nm in diameter and with a center-to-center distance of 500 nm. Next, to determine the distinct DNB-CID sequences at each spatial location, single-end sequencing was performed using sequencing primers in a MGIDNBSEQ-Tx sequencer with sequencing strategy SE25. Finally, MID and polyT-containing oligos were hybridized and ligated to the DNB on the chip. This procedure produces capture probes containing a 25 bp CID barcode, a 10 bp MID and a 22 bp polyT ready for *in situ* capture.

CID sequences together with their corresponding coordinates for each DNB were determined using a base calling method according to manufacturer's instruction of DNBSEQ sequencer. After sequencing, the capture chip was split into smaller size chips (10 mm × 10 mm). At this stage, all duplicated CID that corresponded to non-adjacent spots were filtered out.

Cryo sectioning, fixation, staining and imaging

The pre-frozen leaf tissues in OCT were transversely sectioned at 10 µm thickness using a Leika CM1950 cryostat. Tissue sections were adhered to the Stereo-seq chip surface and incubated at 37 °C for 3 minutes. Then, tissues were fixed in methanol and incubated at -20 °C for 30 minutes. The same tissue sections were stained with Fluorescent Brightener 28 (FB) for Stereo-seq, while tissue sections adjacent to those were adhered to glass slides for histological examination using the same staining method. Imaging for both procedures was performed with a moticfluorescence microscope.

Permeabilization, reverse transcription, tissue removal and cDNA release, Library preparation and sequencing

These processes were performed according to the previously reported Stereo-seq method (Chen et al., 2021). Tissue patches on the chip were permeabilized using 0.1% pepsin (Sigma, P7000) in 0.01 M HCl buffer (pH = 2), incubated at 37 °C for 10 minutes and then washed with 0.1x SSC buffer (Thermo, AM9770) supplemented with 0.05 U/µLRNase inhibitor (NEB, M0314L) after FB in tissues was rinsed off with 80% ethanol. Released RNA from permeabilized tissues was captured by the DNB and reverse transcribed overnight at 42 °C using SuperScript II (Invitrogen, 18064-014, 10 U/µL reverse transcriptase, 1 mM dNTPs, 1 M betaine solution PCR reagent, 7.5 mM MgCl₂, 5 mM DTT, 2 U/µLRNase inhibitor, 2.5 µM Stereo-TSO and 1x First-Strand buffer).

After *in situ* reverse transcription, tissue patches were washed twice with 0.1x SSC buffer and digested with Tissue Removal buffer (10 mM Tris-HCl, 25 mM EDTA, 100 mM NaCl, 0.5% SDS) at 37 °C for 30 minutes. cDNA-containing chips were then subjected to Exonuclease I (NEB, M0293L) treatment for 3 hours at 37 °C and were finally washed once with 0.1x SSC buffer. The resulting cDNAs

released from chips were amplified with KAPA HiFi Hotstart Ready Mix (Roche, KK2602) with 0.8 μ M cDNA-PCR primer. PCR reactions were conducted as: first incubation at 95 °C for 5 minutes, 15 cycles at 98 °C for 20 seconds, 58 °C for 20 seconds, 72 °C for 3 minutes and a final incubation at 72 °C for 5 minutes.

The concentrations of the resulting PCR products were quantified by Qubit dsDNA Assay Kit (Thermo, Q32854). A total of 20 ng of DNA were then fragmented with in-house Tn5 transposase at 55 °C for 10 minutes. Then, the reactions were stopped by the addition of 0.02% SDS buffer and gently mixing at 37 °C for 5 minutes. Fragmentation products were amplified as described below: 25 μ L of fragmentation product, 1x KAPA HiFi Hotstart Ready Mix and 0.3 μ M Stereo-Library-F primer, 0.3 μ M Stereo-Library-R primer in a total volume of 100 μ L with the addition of nuclease-free H₂O. The reaction was then run as: 1 cycle of 95 °C for 5 minutes, 18 cycles of (98 °C 20 seconds, 58 °C 20 seconds and 72 °C 30 seconds) and 1 cycle of 72 °C for 5 minutes. PCR products were purified using the Ampure XP Beads (Vazyme, N411-03) (0.6x and 0.8x) for DNB generation and finally sequenced (paired-end 50 bp) on a DNBSEQ-T10 sequencer.

QUANTIFICATION AND STATISTICAL ANALYSIS

Raw Stereo-seq data processing

Fastq files were generated using a DNBSEQ-T10 sequencer, and the raw data was processed according to the Stereo-seq method (Chen et al., 2021). CID and MID are contained in the forward reads (CID: 1-25 bp, MID: 26-35 bp) while the reverse reads consist of the cDNA sequences. CID sequences on the forward reads were first mapped to the designed coordinates of the *in situ* captured chip, allowing 1 base mismatch to correct for sequencing and PCR errors. Reads with MID containing either N bases or more than 2 bases with quality scores lower than 10 were filtered out. CID and MID associated with each read were appended to each read header. The remaining reads were then aligned to *Arabidopsis* genome (TAIR10) using STAR (Dobin et al., 2013) and mapped reads with MAPQ 10 were counted and annotated to their corresponding genes using an in-house script (available at <https://github.com/BGIResearch/SAW>). MID with the same CID and the same gene locus were collapsed, allowing 1 mismatch to correct for sequencing and PCR errors. Finally, this information was used to generate a CID-containing expression profile matrix.

Single-cell data acquisition

To obtain single-cell level transcriptome data, we built a single-cell data batch extraction process: (1) Cell extraction and classification based on cell boundary information in cell wall stained images; (2) Image alignment between stained images and dot plots generated from expression profile matrix; (3) Extracting the expression data of each labeled cell regions in batches. With data of all identified cells extracted, an information file was generated with bin1 size. We obtained the X-Y coordinates and overall MID count for every single cell using Stereopy (<https://github.com/BGIResearch/stereopy>) and processed the data to be suitable for subsequent analyses at the same time.

Quality control of single-cell Stereo-seq data

Processing of the MID count matrix obtained from Stereo-seq data was implemented using the R package Seurat (v4.0.0) (Hao et al., 2021). To remove low-quality cells, we filtered out cells with gene numbers no more than 200 or higher than 5000 for mesophyll cells and epidermal cells, while for vascular cells and guard cells which had lower RNA capture due to their smaller cell areas, we filtered out cells with gene numbers no more than 50 or higher than 5000. Furthermore, we discarded low-quality cells with a high percentage of mitochondrial genes (>10%) to avoid perforated cells which lose cytoplasmic RNAs (Rich-Griffin et al., 2020). After filtering, the remaining 13,950 single cells with 17,023 genes were included in the downstream analyses.

PCA analysis

The base R function “prcomp” was used to perform Principal Component Analysis (PCA) with the mean matrix for each cell type in each leaf sample. We obtained three-dimensional information of three principal components in our data set and visualized the spatial relationship between the different cell types of each leaf sample.

DEG analysis

Genes differentially expressed across cell types or cell subtypes were identified by comparing average transcript levels in cells of a given cluster to that of cells in all other clusters using the Seurat package Wilcoxon Rank Sum test. For cell types, the default parameters were used. For cell subtypes, the following cutoffs were applied: logfc.threshold = 0.2, min.pct = 0.25.

Expression pattern analysis

We divided each leaf into seven parts, specifically, the main vein and the surrounding three leaf parts of equal length. The symmetrical leaf parts according to main vein were merged, which remains four unique areas in a leaf. The four parts of a single leaf were designated as Area 0, 1, 2 and 3, from main vein to leaf edge and were used for expression pattern analysis. The pattern analysis was performed utilizing the R package Mfuzz (v2.50.0) with the mean expression of genes in four leaf areas (Kumar and E Futschik, 2007).

Trajectory analysis

We used the R package Monocle (v2.18.0) ([Qiu et al., 2017](#)) to carry out pseudotime analyses for vascular cell development and epidermal cell differentiation. The “differentialGeneTest” function was used to calculate the differentially expressed genes in different areas. The genes with high variation were selected using the “dispersionTable” function. The overlap genes between the differentially expressed genes in different areas and the genes with high variations were used for subsequent analysis. Then, we reduced our data to two components using the method “DDRTree”. Cells were ordered along the developmental paths and visualized in two-dimensional space. Heatmaps were used to demonstrate the gene expression that differs in cells. The data used in plotting the heat maps were subsequently used for GO biological process analysis.

GO enrichment analysis

GO enrichment analysis was performed using R package clusterProfiler([Yu et al., 2012](#)) with TAIR10 annotation as the background. The smaller the *P*-value is, the more the GO term is significantly enriched.