

# Single-cell RNA-seq describes the transcriptome landscape and identifies critical transcription factors in the leaf blade of the allotetraploid peanut (*Arachis hypogaea* L.)

Hao Liu<sup>1,†</sup> , Dongxiu Hu<sup>1,†</sup>, Puxuan Du<sup>1,†</sup>, Liping Wang<sup>1</sup>, Xuanqiang Liang<sup>1</sup>, Haifen Li<sup>1</sup>, Qing Lu<sup>1</sup>, Shaoxiong Li<sup>1</sup>, Haiyan Liu<sup>1</sup>, Xiaoping Chen<sup>1,\*</sup>, Rajeev K Varshney<sup>2,3,\*</sup>  and Yanbin Hong<sup>1,\*</sup>

<sup>1</sup>Guangdong Provincial Key Laboratory of Crop Genetic Improvement, South China Peanut Sub-Center of National Center of Oilseed Crops Improvement, Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong Province, China

<sup>2</sup>Center of Excellence in Genomics & Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Telangana, India

<sup>3</sup>State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia

Received 16 February 2021;

revised 22 June 2021;

accepted 23 June 2021.

\*Correspondence (Tel +86 138 26214570; fax +020 87511794; email

hongyanbin@gdaas.cn (Y.H.); Tel +86 150 11858620; fax +86 208 5514269; email

chenxiaoping@gdaas.cn (X.C.); Tel +91 8455 683305; fax +91 8455 683071; email R.K.Varshney@CGIAR.ORG (R.K.V.))

† These authors contributed equally to this article.

## Summary

Single-cell RNA-seq (scRNA-seq) has been highlighted as a powerful tool for the description of human cell transcriptome, but the technology has not been broadly applied in plant cells. Herein, we describe the successful development of a robust protoplast cell isolation system in the peanut leaf. A total of 6,815 single cells were divided into eight cell clusters based on reported marker genes by applying scRNA-seq. Further, a pseudo-time analysis was used to describe the developmental trajectory and interaction network of transcription factors (TFs) of distinct cell types during leaf growth. The trajectory enabled re-investigation of the primordium-driven development processes of the mesophyll and epidermis. These results suggest that palisade cells likely differentiate into spongy cells, while the epidermal cells originated earlier than the primordium. Subsequently, the developed method integrated multiple technologies to efficiently validate the scRNA-seq result in a homogenous cell population. The expression levels of several TFs were strongly correlated with epidermal ontogeny in accordance with obtained scRNA-seq values. Additionally, peanut *AHL23* (*AT-HOOK MOTIF NUCLEAR LOCALIZED PROTEIN 23*), which is localized in nucleus, promoted leaf growth when ectopically expressed in *Arabidopsis* by modulating the phytohormone pathway. Together, our study displays that application of scRNA-seq can provide new hypotheses regarding cell differentiation in the leaf blade of *Arachis hypogaea*. We believe that this approach will enable significant advances in the functional study of leaf blade cells in the allotetraploid peanut and other plant species.

**Keywords:** scRNA-seq, plant single-cell, leaf cell, leaf development, peanut improvement.

## Introduction

Leaf is an integral aerobic component of a plant and is vital for its ability to perform photosynthesis. Plants evolved to have diversity of leaf shapes to adapt to the environment, in which they are grown, with an increased density of chlorophyll to maximize the ability to absorb sunlight. This process allows the plant to convert light into chemical energy to fuel the continuous production of carbon substrates required for plant vegetative and reproductive growth (Song *et al.*, 2015). Owing to their size and cell density, leaves have been identified as the perfect source for mRNA extraction for transcriptomics research. While current methods of physiology and molecular biology research have provided great insights into leaf growth and behaviour, they have failed to capture the heterogeneity of function that individual cells exhibit in leaf development. Elucidation of these behaviours may expand our understanding of leaf functionality.

Currently, scRNA-seq is frequently used to analyse tissue cells of various types, large quantities of data are generated but the specific functionalities of individual cells are often not considered. Recently invented cell isolation technologies, including the

utilization of glass microcapillaries, flow cytometric sorting, laser capture microdissection and enzyme-isolated protoplasts, have been applied to plant cells. The aforementioned approaches to plant single-cell RNA-sequencing (scRNA-seq) have allowed compilation of gene expression atlases for single cells isolated from the *Arabidopsis* root (Liu *et al.*, 2020b). These atlases presented a widespread heterogeneous and monoallelic transcriptome landscape that differed greatly when compared to the results of bulk RNA-seq (Denyer *et al.*, 2019; Ryu *et al.*, 2019). Evidence presented in a recent study enabled reinterpretation of the development procedure of root cell lineage by adopting scRNA-seq, the root cell trajectory presented is different to the traditional theory supported by integrating morphological observation with bulk RNA-seq (Jean-Baptiste *et al.*, 2019; Liu *et al.*, 2021b; Shulse *et al.*, 2019). In recent times, scRNA-seq has been utilized to dissect the transcriptome profile of stomatal lineage cell development in *Arabidopsis* (Liu *et al.*, 2020b), to decipher the transcriptional landscape of maize stem-cell reorganization at single-cell resolution (Satterlee *et al.*, 2020) and also in functional analysis and gene discovery in the development of ear in maize (Xu *et al.*, 2021). Additionally, scRNA-seq has revealed the

mechanisms of meiosis in the another of maize, which provided a greater insight into plant reproductive cell lineage and development (Nelms and Walbot, 2019). In the case of understanding of leaf blade development, the scRNA-seq study has not provided very satisfactory results (Kim et al., 2021). The use of scRNA-seq on non-model or allotetraploid plant cells has remain limited due to the absence of information on cellular compositions and lack of suitable techniques.

Peanut or groundnut (*Arachis hypogaea* L.) is an allopolyploid leguminous crop that has a genome consisting of two diploid sub-genomes, *Arachis ipaensis* and *Arachis duranensis* (Bertioli et al., 2019; Chen et al., 2019;). It is cultivated worldwide for edible oil and easy source of protein (Liu et al., 2020a). The process of underground pod development requires a large amount of energy which the peanut plant generates via photosynthesis (Chen et al., 2016). Therefore, leaves play an integral role in the overall survival and functional development of the plant (Kumar et al., 2019). In view of above, this study presents the first scRNA-seq expression atlas of the leaf blade of peanut at single-cell level. We report precise spatiotemporal information, critical marker gene transcripts and define features of major leaf cell types that will allow for improvements in peanut cultivation.

## Results

### Peanut leaf protoplast cell isolation and application of scRNA-seq

Initially, peanut seedlings of the variety Hanghua2hao were grown in a dark setting for one week before the etiolated leaf blade was plucked from the leaf stalk and cut into small strips. The leaf strips were then transferred into 30 ml of cellulase and pectinase enzyme solution and shaken slightly for two hours to isolate protoplasts. Cell viability was assessed by staining with trypan blue. Consequently, more than 1280 protoplast cells/ $\mu$ L were obtained. Microscopic observation identified the percentage of live cells to be over 94%. Thus, the cell counts and quality was consistent with the standard for next-generation scRNA-seq (Figure 1a). Analysis was then performed using the 10x Genomics scRNA-seq platform. A total of  $2 \times 10^4$  cells were loaded to examine gene expression profiles against the previous published transcriptome information (Bertioli et al., 2019).

### Identification of major cell clusters in peanut leaf

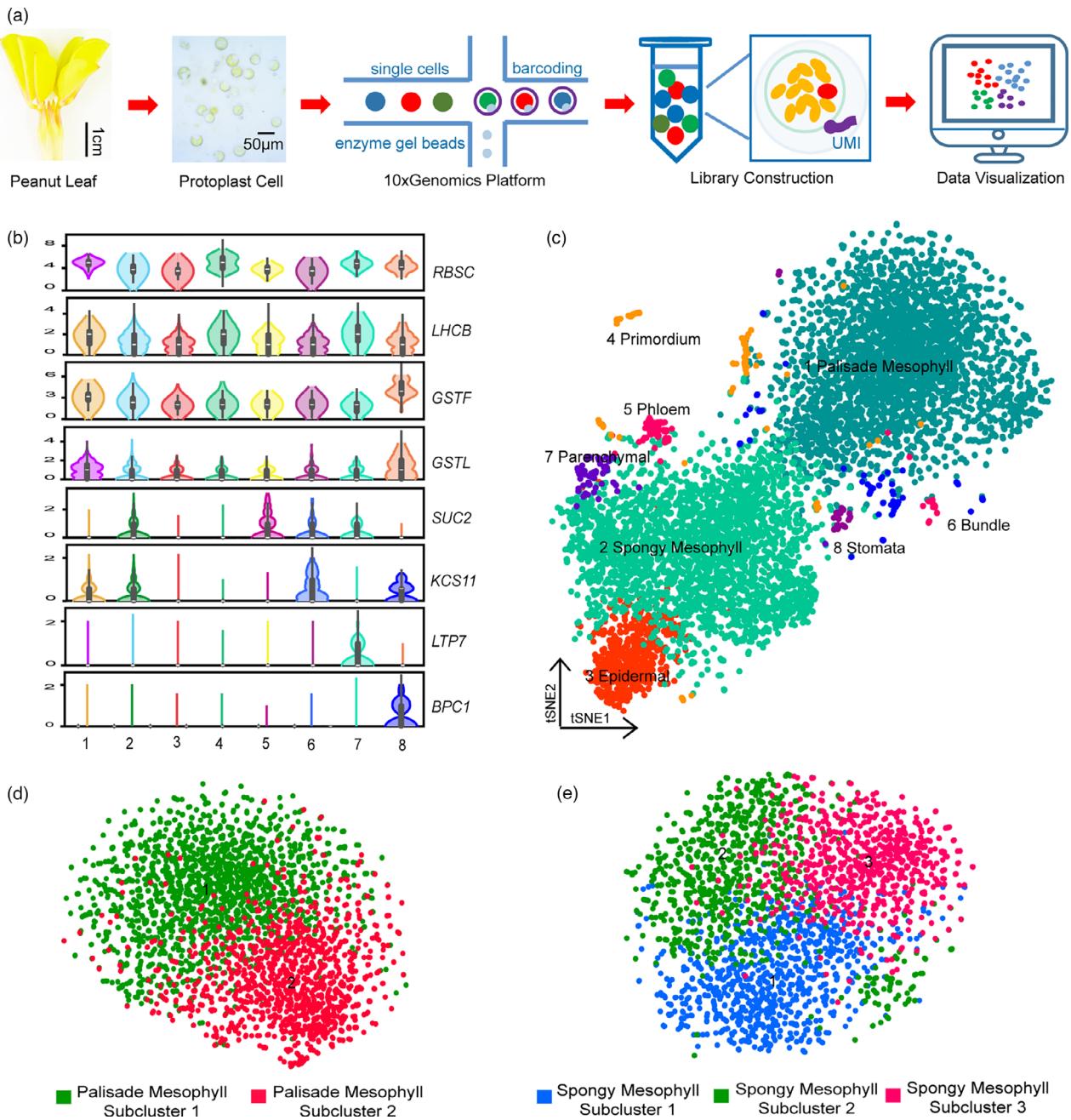
Consequently, the transcriptome profiles of 6,815 single cells were obtained and characterized by analysis of the total UMI (Unique Molecular Index) counts. Data quality control indicated that a median of 3,070 filtered UMI and 1,807 genes (Figures S1–S3, Table S1–S2) were distributed in each cell. Forty-six thousand and four genes (Table S3) were identified following assessment using the improved filter parameter and, subsequently, their expression abundances were anchored into a circular map that consisted of 20 peanut chromosomes to describe the single-cell gene expression pattern of the analysed leaf cells (Figure S4). In order to visualize populations of distinct cell types following low quality filtering of the cells, the original normalized transcriptome data were processed using the canonical correspondence analysis function of Seurat (R package) and by Z-score normalization. The processed data were subsequently stratified using reducing dimensionality through principal component analysis, which resulted in 6,815 single cells clustered by t-distributed stochastic neighbourhood embedding (t-SNE). This stratification was

achieved by utilizing reported marker genes (Figure 1b, Figure S5, Table S4) to authenticate the different cell types involved in leaf development. The largest cell cluster was displayed by mesophyll cells, divided by the enrichment of reported marker genes involved in photosynthesis like the *RUBISCO BISPOSPHATE CARBOXYLASE SMALL SUBUNIT* (*RBCS*) and *LIGHT-HARVESTING CHLOROPHYLL A/B-BINDING PROTEIN* (*LHCB*) that were dominantly expressed in mesophyll palisade and spongy tissue, respectively. The *GLUTATHIONE S-TRANSFERASE F SUBUNIT* (*GSTF*) was detected in the epidermal cluster and *GLUTATHIONE S-TRANSFERASE L1-LIKE* (*GSTL*) was enriched in primordium group. Phloem cell identity based on the transcript enrichment of highly expressed *SUCROSE TRANSPORT PROTEIN SUC2-LIKE 2* (*SUC2*). Bundle sheath cell surrounded the cell populations of xylem and phloem, which identified with the expression level of *KETOACYL-CoA SYNTHESIS 11* (*KCS11*). Parenchyma cells exist in higher plants composed of thin-walled cells that remain capable of cell division even when mature, and *LIPID TRANSFER PROTEIN 7* (*LTP7*) transcripts were characterized to abundantly gather in parenchyma. Stomata cell distributes on the epidermal surface to modulate the gas exchange; this population was identified by specific expression of *BASICPENTACYSTEINE 1* (*BPC1*) (Liu et al., 2020b; Wang et al., 2021; Zhang et al., 2021; Zhang et al., 2019).

All single cells were assigned into eight major cell clusters based on cluster-characteristically expressed DEGs corresponding to distinct molecular roles (Table S4) in palisade mesophyll, spongy mesophyll, epidermal, primordium, phloem, bundle sheath, parenchymal and stomata guard cells (Figure 1c). Each subgroup contained a number of single cells ranging from 28 to 3216 (Table S2). Furthermore, despite the attempts to elaborate sub-clusters within each major cell cluster, only the palisade mesophyll and spongy mesophyll cell clusters were divided into two and three sub-clusters (Figure 1d–e), respectively. Identification of different cell clusters not only contributed to gathering of basic information regarding cell type, but it also aided further characterization of marker genes and analysis of cell heterogeneity.

### Characterization of represented marker gene in each cell cluster

In order to characterize marker genes, relevant up-regulated differentially expressed genes (DEGs) counts, distributed at a range from 72 to 1513 (Table S2, Table S5), were assessed in each cluster. The largest percentage of elevated DEGs occurred in palisade mesophyll cells. As expected, the majority of up-regulated DEGs participated in mechanisms associated with photosynthesis as mesophyll cells play a pivotal role in light capture. Following the identification of DEGs, we calculated the output of forty novel marker genes (Table S6) in the distinct cell types. Top five genes with the highest expression level in each cluster were collected and the expression profiles were described in a heatmap (Figure 2a, Figure S6). These identified DEGs will be useful in making a distinction between cell types in any future peanut leaf scRNA-seq studies. Meanwhile, eight representative genes with the highest expression in each cluster were selected for display in the t-SNE map (Figure 2b). Using GO analysis, it was identified that cluster-specific DEGs were part of diverse biological processes, majorly metabolic and cellular processes (Figure 2c, Table S8). Based on the KEGG presentation, the enriched pathways mostly involved ribosomal activity, metabolism and photosynthesis (Figure 2d, Table S7). Overall, the visualization of



**Figure 1** Development and application of an scRNA-seq approach for the peanut leaf. (a) Brief process of isolation of peanut leaf protoplast cells and their loading onto the 10×Genomics platform. (b) Violin plots show the expression pattern and distribution of cluster-specific genes. (c) Classification of distinct peanut leaf cell clusters (t-SNE map). Each dot indicates individual cells that are coloured based on cell type. (d–e) t-SNE map embedding of sub-cluster assignments following a second round of clustering within palisade and spongy tissues.

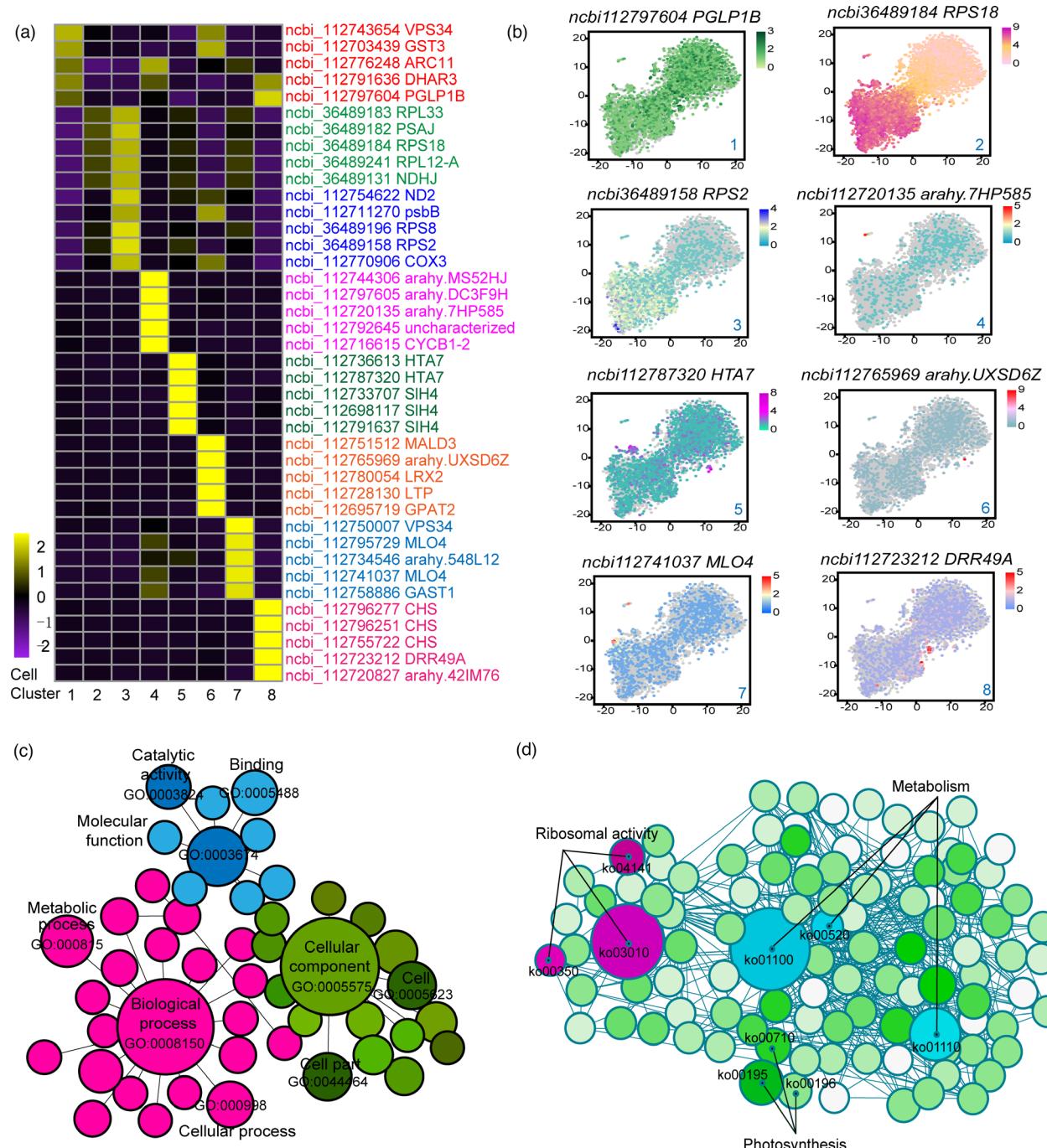
variation in expression in each cell provided us with a more intuitive method to interrogate gene expression in the context of different cell types.

#### Pseudo-time trajectory analysis of cells of the peanut leaf

In order to dissect the spatial and temporal distribution of the assessed single cells, a pseudo-time trajectory was used to visualize the cell distribution of each cluster along the main stem (Figure 3a, Table S9). Four small branches were identified near

the main stem and distinct sub-cluster cells displayed time heterogeneity with the main pathway of cell differentiation. Additionally, pseudo-time analysis revealed ten critical marker genes that could be used to divide all single cells into nine states of leaf cell development and differentiation (Figure S7). Moreover, 4486 DEGs were implicated in cell growth (Table S10), a finding that will enable further investigation into leaf cell development.

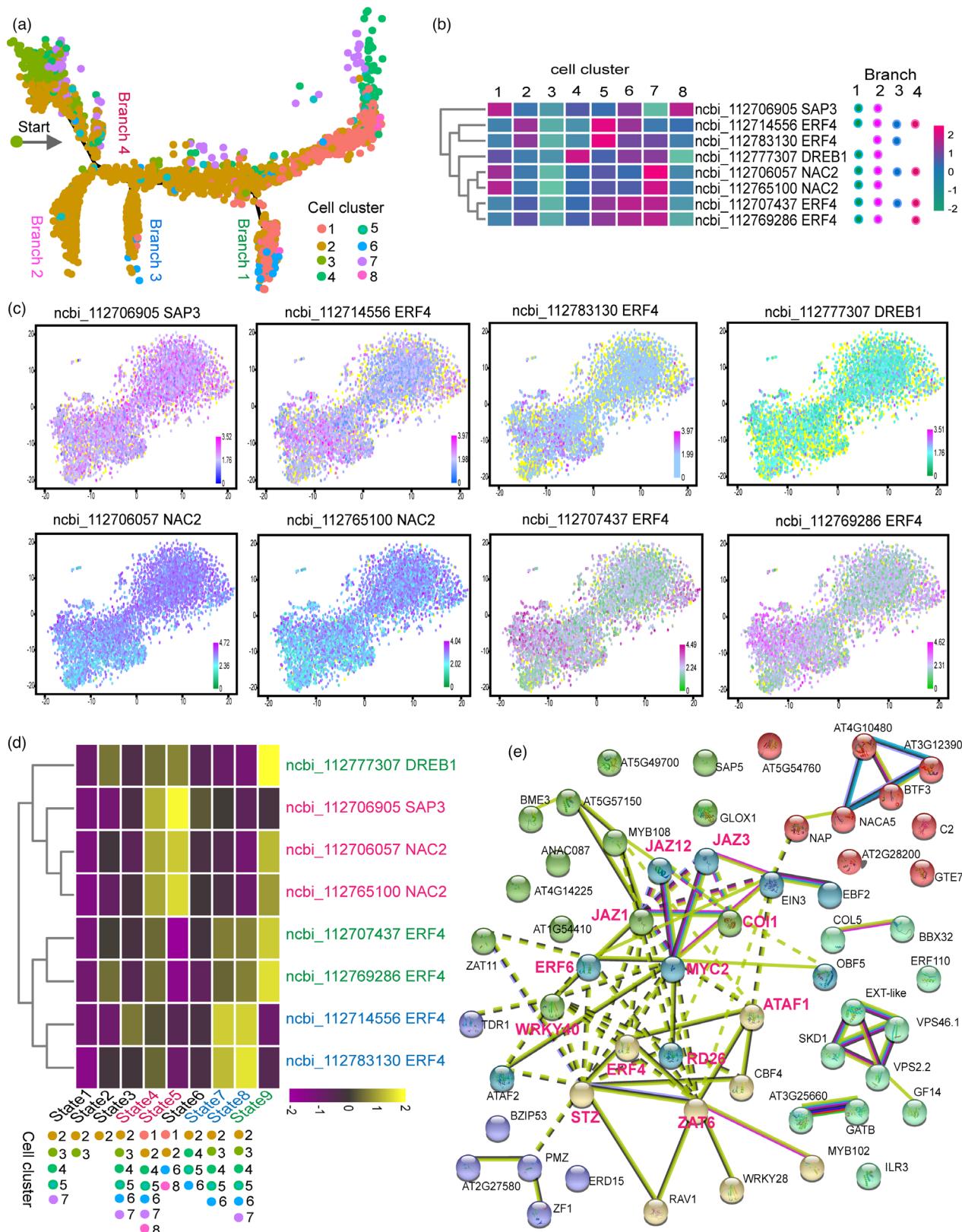
Furthermore, a total of 2069 DEGs were identified as determinants of cell fate in four cell differentiation branches, including



**Figure 2** Identification of new marker genes within cell-type clusters. (a) Heatmap shows the top five DEGs with the highest expression levels ( $\log^2$  Fold Change) in each sub-cluster (Table S6). (b) The expression patterns of eight new marker genes distributed in a t-SNE map. (c-d) GO and KEGG pathway enrichment analysis of all DEGs.

305, 1417, 238 and 109 distributed from branch 1 to 4 (Table S11). The majority of DEGs (chloroplast protein) identified were involved in the pathways of photosynthesis, carbohydrate metabolism or energy transfer. However, the conduction of cell differentiation was generally attributed to transcriptional regulation, thus, the identification of critical transcription factors (TFs) correlated with cell fate was important for subsequent studies reported herein. Accordingly, data regarding 117 TFs are collectively presented in each branch of the pseudo-time analysis,

including their expression level and differential expression mode in distinct cell cluster (Table S12). The expression distribution of eight core TFs in every cell demonstrated that the transcripts *NAC002* and *SAP3* are highly expressed in most cells, while *ERF4* is up-regulated in spongy tissue, *DREB1* is down-regulated in mesophyll cells and overall gene expression distribution suggested that *NAC002* may determine many aspects of the mechanisms underpinning leaf cell growth (Figure 3C). Besides, *SAP3* and *NAC002* were identified as associated with differentiation state 4



**Figure 3** Pseudo-time trajectory analysis of cell types in the peanut leaf. (a) Distribution of cells in each cluster along with pseudo-time trajectory. (b) The relative expression of eight representative TFs in the cell differentiation branch. (c) The distribution of expression of eight TFs in all cell subsets, yellow dots as background that represented the cells with no expression of the given transcript. (d) Heatmap displaying the average expression level of eight TFs in distinct cell differentiation states. (e) All identified TFs in an interaction network constructed by using their homologs in the *Arabidopsis* genome.

and 5, *ERF4* was related to state 7 and 8 while the other three TFs showed correlation with state 9. This indicated that TFs potentially modulate the relationship between cell-cluster placement and differentiation fate (Figure 3d). Moreover, 91 TFs were specifically presented in branch 2 (Figure S8). Branch 2 mainly contained cells of the spongy mesophyll that had undergone the eighth state of cell differentiation. However, the 91 identified TFs tended to be down-regulated in spongy tissue when compared to other clusters, implying that transcription suppression of TF expression may determine the differentiation state of the spongy cells of the peanut plant leaf. To discern the detailed pathways and identify factors involved in the outcome of cell fate, we performed homologous interactive network analysis of TFs in *Arabidopsis*. The results of the analysis indicated that the ethylene (ETH) and jasmonic acid (JA) pathways demonstrated pivotal characteristics and led to the elucidation that *ERF4/6*, *WRKY40*, *MYC2* and multiple *JAZ* proteins consisted of a core sophisticated protein–protein interaction network (Figure 3e). Based on these findings, we concluded that peanut leaf cell differentiation relies on the mechanisms of phytohormone modulation in dark conditions.

#### Peanut leaf mesophyll cell development process

The mesophyll layer is the primary location of photosynthesis. It has the largest cell number and consists of two distinct cell types making up palisade and spongy tissues (Evans, 2021). Currently, it is considered that the mesophyll layer originates as part of the primordium, which in turn, develops into the parenchyma before finally forming the mesophyll layer (Runions et al., 2017). To test this hypothesis, primordium, parenchymal and mesophyll cell clusters were analysed to describe the developmental trajectory of the mesophyll layer (Figure S9). By investigating cell differentiation orientation along the timeline, the pseudo-time trajectory data revealed that the primordium was firstly developed into parenchymal cells, which further developed into the palisade layer. However, we found little evidence to support the hypothesis that the primordium directly converted into spongy tissue (Figure 4a). A total of 3942 DEGs were identified along the main stem of the pseudo-time trajectory (Table S13, Figure S10) and 193 TFs existed in the differential expression profile (Table S14, Figure 4c), of which 73 TFs were identified as being involved in the circadian clock and/or in the plant hormone pathways (Figure 4d). Additionally, all cells could be classified into five differentiation states (Figure 4b, Table S15). On the other hand, 269 TFs were characterized to correlate with these distinct states (Table S16). The majority of TFs were highly expressed in state 1, and it was clear that JA and ETH TFs appeared to regulate state 4 and state 5, respectively (Figure 4e). Two branch points appeared in the main stem of the development trajectory, the palisade cell layer subset split into spongy tissue concentrated at branch point one while the primordium branched into parenchymal and spongy cells gathered at point two (Figure 4a–b, Figure S10). Thirty-two (32) TFs that participated in branch differentiation were filtered from the DEG profile (Table S17). Interestingly, *MYB102* was simultaneously present in both branches and, along with other JA TFs, dominated the process of palisade differentiation into spongy cells (branch 2), as shown by high levels of expression (Figure 4f, Table S18). Furthermore, an interaction network of 25 critical TFs associated with JA/ETH, ABA and nitrogen metabolism pathways were shown to determine the development of the mesophyll layer (Figure 4g). Overall, a hypothetical mode of differentiation from primordium and

parenchyma to mesophyll cell types was described in accordance with trajectory analysis. Parenchymal cells maintained the ability to transform into both spongy and palisade cells, and a concept of palisade cells being capable of developing into spongy tissue in peanut leaf was proposed (Figure 4h). These results could contribute to future understanding of the developmental mechanisms of the mesophyll layer in the leaf blade.

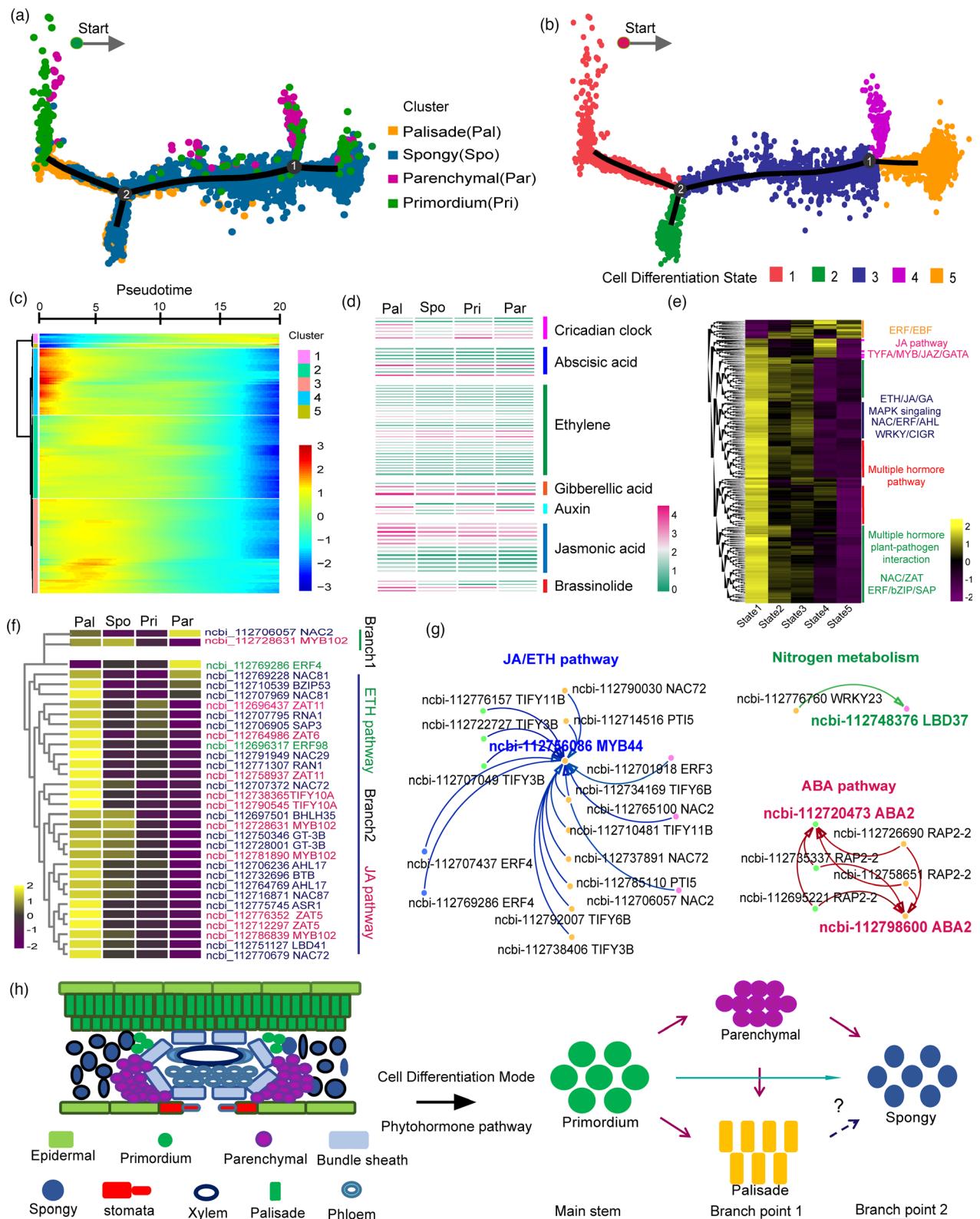
#### Deciphering the trajectory of development from the primordium to epidermal cell

Epidermal tissue surrounds the mesophyll layer of the leaf to protect it from immediate physical trauma caused by environment. The outer epidermis is covered with a cuticle and wax layer formed as a protective seal. The epidermal layer is, therefore, important for leaf survival (von Wangenheim et al., 2017). Previous hypotheses suggested that the epidermis originated from the primordium. To test whether this was in fact the case, we re-constructed the trajectory relationship between the primordium and epidermis (Figure S9). By observing individual cell distribution along with the trajectory of the main stem, we discovered that the epidermal layer originated earlier than the primordium and is, therefore, not directly developed from the primordium in its entirety. However, a small portion of epidermal tissue did appear to have evolved from the primordium (Figure 5a), and we found 11 TFs involved in this process after filtering of the DEG profile (Table S19, Figure S11). The expression level of these 11 TFs gradually decreased as primordium differentiated into epidermal tissue (Figure 5b), and were specifically highly expressed in the primordium cluster (Figure 5c).

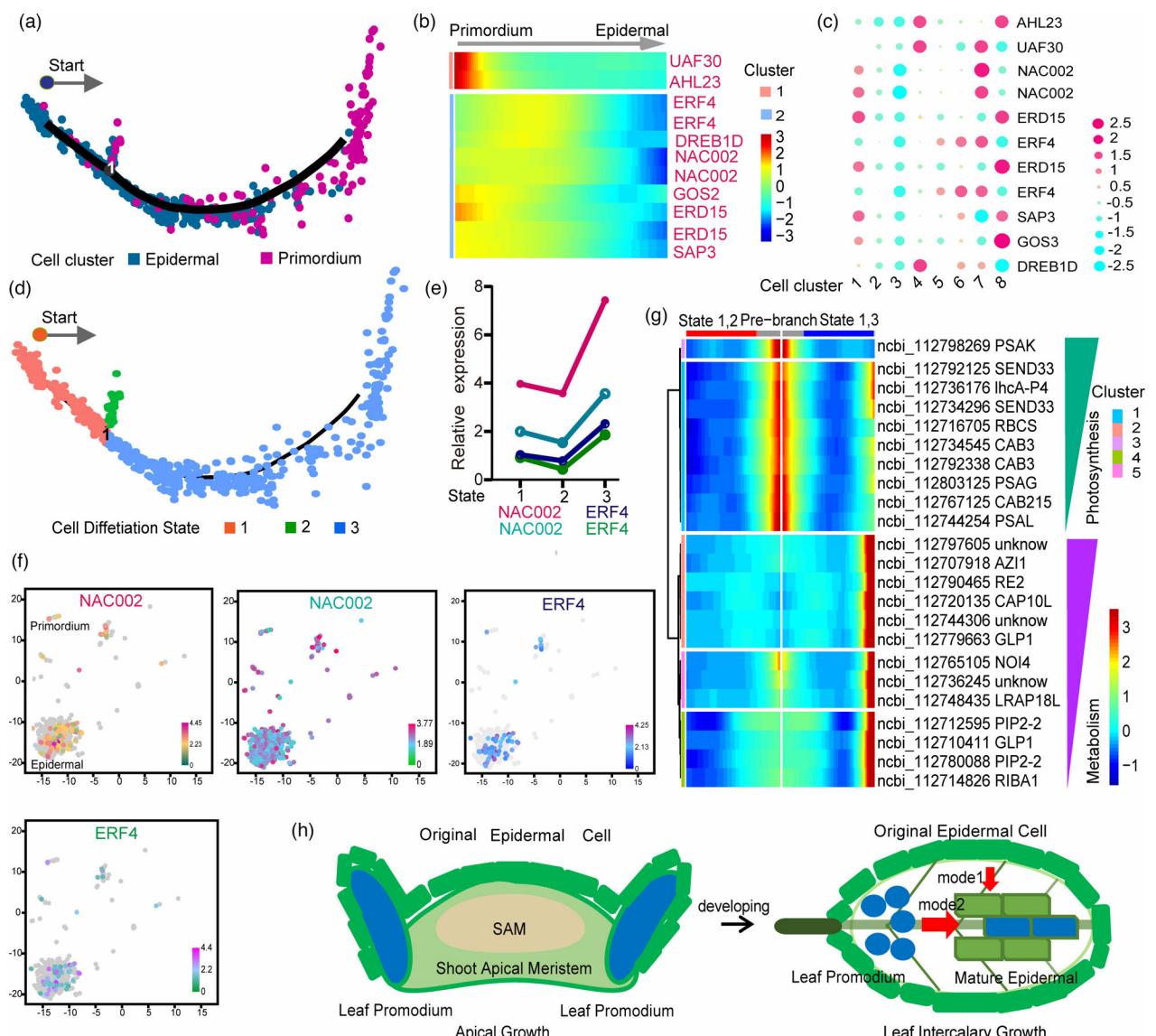
Meanwhile, all cells studied were divided into three distinct cell differentiation states (Figure 5d, Table S20), with *NAC002* and *ERF4* being differentially expressed depending on the state (Figure 5e). Cell expression distribution suggested that *NAC002* was expressed more broadly than *ERF4*, implying that *NAC002* may have a stronger influence on epidermal differentiation when compared to *ERF4* (Figure 5f). Of the 23 specific DEGs that represented marker genes for determination of cell fate and the different corresponding differentiation states, the majority were considered to contribute to the function of photosynthesis and energy metabolism as shown by pathway enrichment analysis (Figure 5g). In summary, we proposed putative modes for epidermal development. One mode was the original epidermal cells driven from the shoot apical meristem (SAM), which continually proliferated and turned into a part of the mature epidermal cells during leaf growth. A second mode was proposed whereby the original leaf primordium evolved into epidermal cell directly during leaf size expansion (Figure 5h), which consequently expanded the current understanding of epidermal origination and development in peanut.

#### Validation of the scRNA-seq result at the cell population level

To validate whether the true expression level of identified DEGs reflected the scRNA-seq data obtained from the divergent cell clusters, we developed an approach to detect their expression. Briefly, leaf strips were degraded using a cellulase enzyme solution for 30-minutes. Incomplete digestion resulted in cells maintaining their original morphology and cell walls including those of elliptical spongy tissue, baculiform plastid tissue and irregular epidermis (Figure 6a). Next, glass capillaries were used to capture live single cells and mesophyll (spongy, palisade) and epidermal cells were separated using microscopy to visually



**Figure 4** Developmental trajectory of the mesophyll layer from the primordium. (a-b) Cell cluster and differentiation state distributions followed the pseudo-time trajectory of mesophyll development. (c) Clustering and expression kinetics of representative TFs along a pseudo-time progression of differentiation from the primordium to the mesophyll layer. (d) Heatmap visualization of the expression level of representative TFs involved in circadian clock and plant hormone pathways. (e) Average expression heatmap of identified differentially expressed TFs associated with five cell differentiation states. (f) Representative TFs involved in branch 1-2 differentiation. (g) A constructed critical TF interaction network involved in the process of mesophyll differentiation. (h) Putative mode of progression of primordium differentiation into mesophyll in the peanut leaf, based on trajectory analysis.

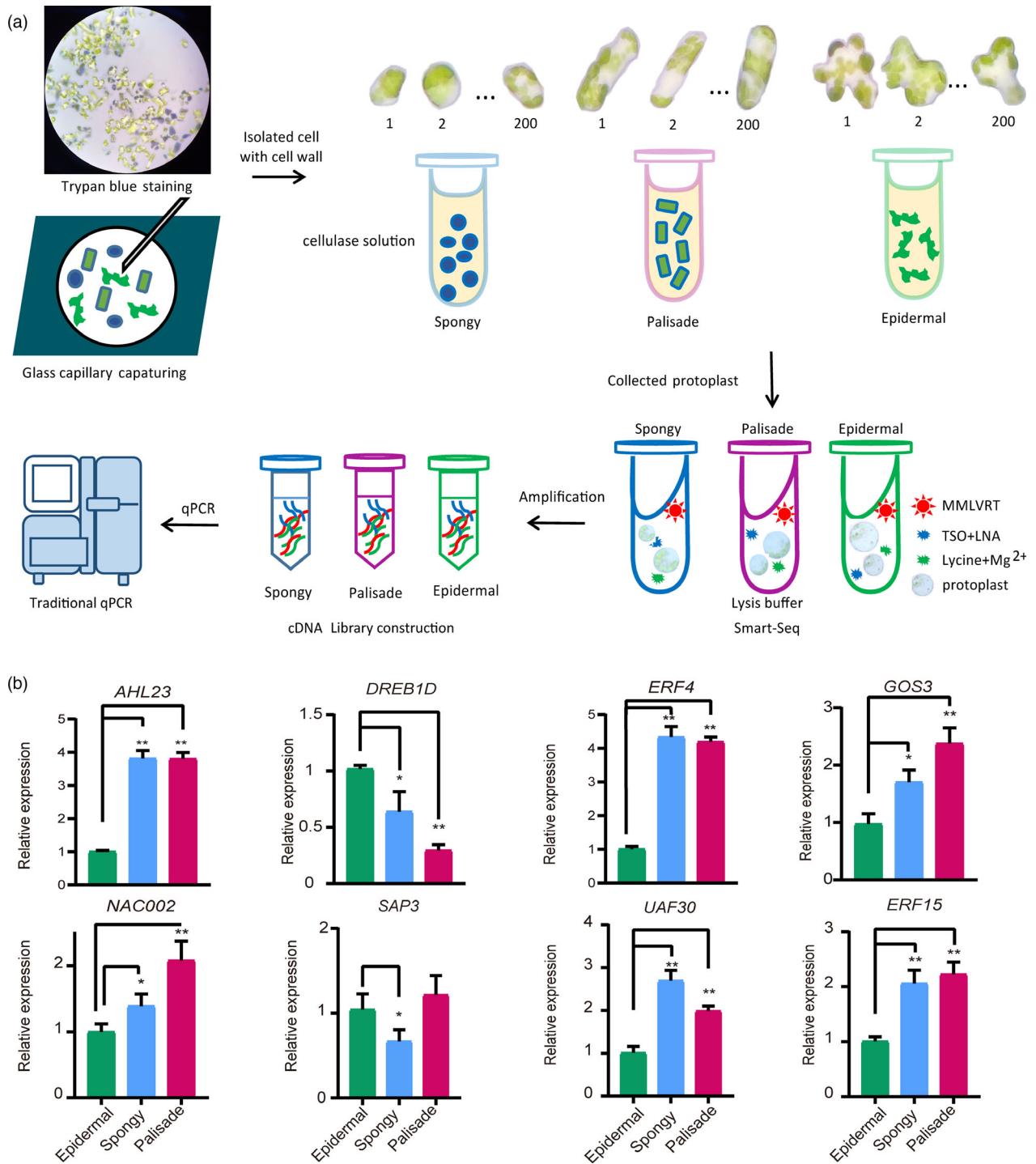


**Figure 5** Pseudo-time trajectory analysis of epidermal cells origination. (a) Cell cluster distribution along the pseudo-time trajectory. (b) Clustering and expression kinetics of TFs along the pseudo-time progression from the primordium to epidermal cells. (c) Relative expression of representative TFs in each cell cluster. (d) Distribution of cell differentiation states along the pseudo-time trajectory. (e) Average expression level of NAC002 and ERF4 in three epidermal differentiation states. (f) Expression distribution of NAC002 and ERF4 in all epidermal and primordium cells, grey dots represent the cells with no expression of the specified transcript. (g) Clustering and expression kinetics of representative marker genes for the pseudo-time process of cell differentiation fate in epidermal cells. (h) Putative models of the development and differentiation pattern of epidermal cells in the peanut leaf.

distinguish their morphology after staining with trypan blue. Isolated single cells were placed into new tubes within cell-type groups (200 cells per tube) before they were further digested by cellulase and pectinase solution in order to completely degrade the cell wall. Finally, different cell populations without their cell walls had their RNA extracted in order to construct the cDNA library by following the SMART-seq protocol. This cDNA library served as a template for conventional quantitative PCR examination (Figure 6a).

We utilized epidermal cells as a reference sample to detect the true expression abundance of aforementioned TFs that affected cell differentiation, from primordium to epidermis, at a cell-group

level. In addition to *DREB1D* and *SAP3*, expression levels of other TFs were in accordance with scRNA-seq detected values in divergent cell clusters. In particular, *AHL23* and *ERF4/15* expression presented similar patterns, with a high level in mesophyll when compared to epidermal cells. *NAC002* and *GOS3* TFs were particularly upregulated in palisade tissue, and *UAF30* expression was selectively elevated in spongy tissue (Figure 6b). Overall, our method integrated microseparation, cell wall digestion, SMART-seq library construction and traditional real-time PCR to examine gene expression in homogenous cell groups. This protocol provided an easy, fast, high throughput and low-cost method for further validation of the scRNA-seq result in the peanut leaf.



**Figure 6** A novel validation method for the scRNA-seq result in distinct cell populations. (a) Leaf cells were captured by glass capillary tubing and their morphology was observed under microscopy before tandem cellulose solution was utilized to digest the cell wall and generate protoplasts for the construction of a SMART-seq cDNA library and qPCR examination. (b) The scRNA-seq validation method detected the relative expression levels of multiple TFs highlighted in epidermal development in distinct cell types. In all histograms, the mean  $\pm$  SD of three biological repeats are plotted, asterisks indicate significant differences when compared to epidermal cells (T-test, \* $P$  < 0.05, \*\* $P$  < 0.01).

#### Peanut *AHL23* promoted leaf growth in ectopically expressed *Arabidopsis*

In the above sections, we have reported multiple transcription factors important for mesophyll and epidermal development.

With an objective of establishing their molecular functionality in leaf growth, we checked the expression levels of all TFs in our previously published transcriptome profile. As a result, we identified that the *AHL23* (*Arahy.BT3IUC*) transcript abundance (Figure S12) was differentially expressed in distinct peanut variety

(Liu et al., 2019). Subsequently, we focussed on describing the mechanism of *AHL23* function. The coding sequence of peanut *AHL23* (*AT-HOOK MOTIF NUCLEAR LOCALIZED PROTEIN 23*) was 882 bp. The protein structure contained typical AT-hook and DUF296 domains (Figure 7a, Figure S12). Phylogenetic analysis indicated that *AHL23* was similar to *MtAHL23* and *SIAHL23* in group III (Figure 7b) and transiently expressed *AHL23* protein was mainly localized at the nucleus of *Arabidopsis* protoplast cells (Figure 7c). The *AHL23* transcript was highly enriched in peanut pod and root tissue, but the expression was lower in other areas such as in the leaf or stem, which suggested that *AHL23* expression might be activated by dark conditions (Figure S12D). Furthermore, *AHL23* was ectopically expressed in *Arabidopsis* to generate *AHL23-OX* (overexpression) transgenic lines (Figure S12E). At the seedling stage in overexpression lines, it was identified that *AHL23* activated hypocotyl elongation (Figure 7d-e) and in germinated seeds the cotyledon size was slightly larger as compared to wild type (WT) under normal growth conditions (Figure 7f-g). At the vegetative growth stage, *AHL23-OX* lines grew at a significantly enhanced rate, both in terms of leaf expansion and overall plant size, in comparison with WT (Figure 7h-i). Furthermore, microscopic observation of the leaf blade revealed that *AHL23* overexpression increased epidermal cell number but did not result in an increase in cell size compared to WT (Figure S12). This suggested that *AHL23* was able to positively modulate epidermal proliferation. Interestingly, *AHL23* overexpression promoted leaf size expansion (Figure 7j), resulted in larger plant height (Figure 7m) and induced early flowering at reproductive growth stage (Figure 7k). Peanut *AHL23* overexpression showed more vegetative and reproductive growth in *Arabidopsis*, and the flowering time occurred at least 10 days sooner than WT plants (Figure 7l). At this point, the mechanisms by which *AHL23* regulated leaf growth was still unclear. Therefore, we collected the leaf tissue of the transgenic plants in order to detect the 88 phytohormone species by LC-MS. *AHL23* overexpression increased the contents of ethylene precursor ACC (Figure 7n), JA and OPDA (Figure 7o), cytokinin derivatives IPR and iP9G (Figure 7p), auxin synthesis precursor TRP, auxin derivatives IAA-Leu and IAA-Val (Figure 7q) and strigolactones precursor 5DS (Figure 7r) but down-regulated the levels of GA19 (Figure 7s) and salicylic acid (Figure 7t). It did not affect the abscisic acid (ABA) pathway (Table S22). Peanut *AHL23* involved ETH/JA pathways presented obvious difference with their orthologues in *Arabidopsis* in that the regulated plant growth usually depended upon GA/IAA pathway. In conclusion, we hypothesize that *AHL23* positively promoted leaf cell growth, and in turn, a high growth index promoted the flowering via the phytohormone pathway (Figure 7u). Overall, this result expanded our knowledge of the molecular function of the *AHL* transcription factor in the peanut plant.

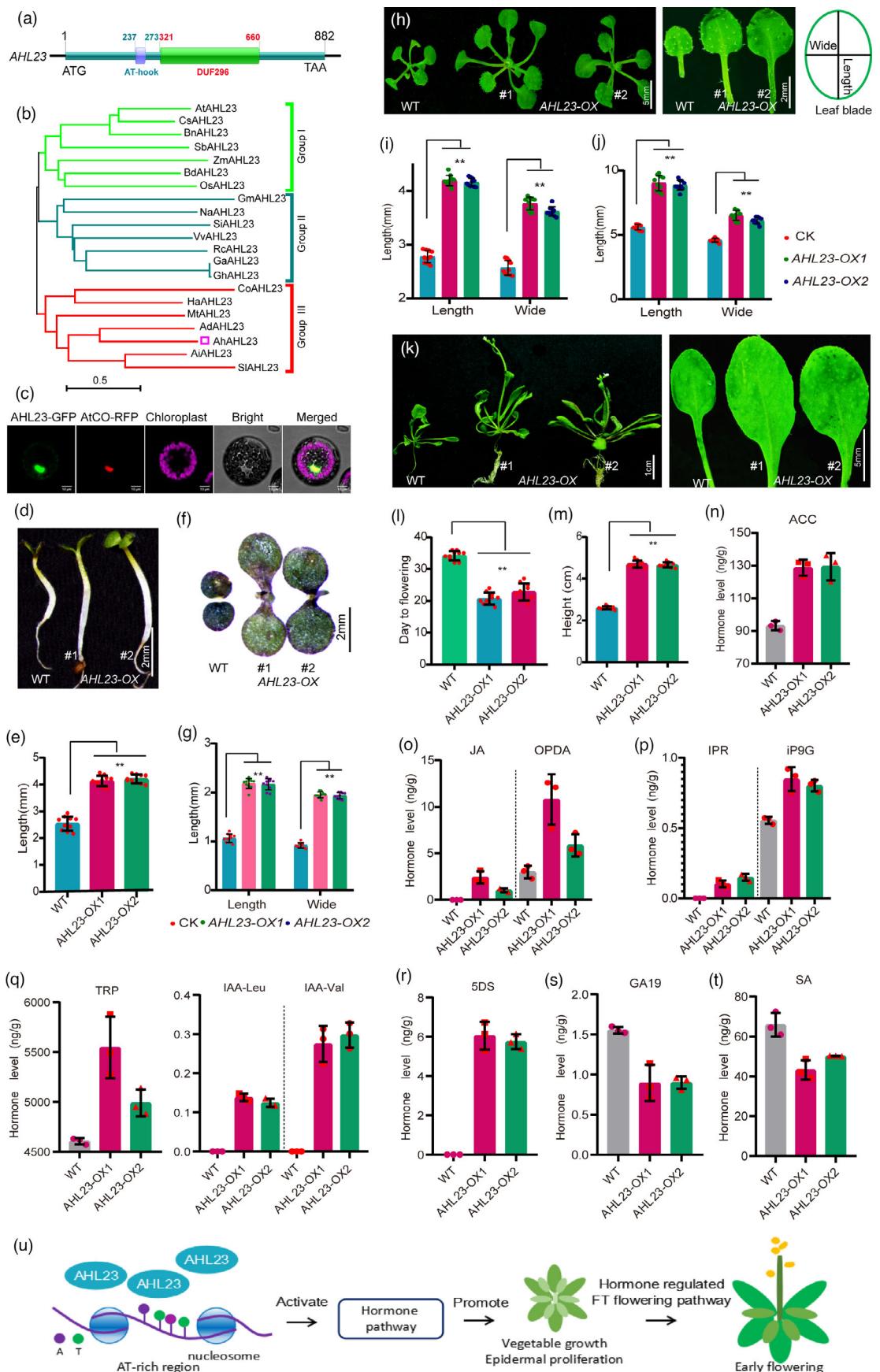
**Figure 7** Integrated analysis of the molecular function of peanut *AHL23*. (a) Schematic diagram of the genome structure of *AHL23*. (b) Phylogenetic tree of *AHL23* homologs in a diversity of plant species. (c) Subcellular localization of *AHL23* protein at the nucleus of an *Arabidopsis* protoplast cell, scale bar is 10 µm. (d) Hypocotyl phenotype of *AHL23-OX* lines and WT at seedling stage. (e) measurement of hypocotyl length in *AHL23* transgenic and WT plants ( $n = 10$ ). (f) Cotyledon phenotype of *AHL23-OX* lines and WT at seedling stage. (g) measurement of cotyledon length and width in *AHL23* transgenic and WT plants ( $n = 10$ ). (h) Leaf morphology in *AHL23-OX* and WT plants at vegetative (i) and reproductive (j) stages ( $n = 10$ ). (k) The phenotype of the entire plant and mature leaf in *AHL23-OX* and WT plants at reproductive stage. (l) Flowering time statistics in the different plant lines ( $n = 10$ ). (m) Plant height investigation in WT and *AHL23-OX* lines ( $n = 10$ ), (n-t) Plant hormone detection results in *AHL23-OX* and WT plants after application of LC-MS, histograms indicate the mean ± SD of three biological repeats. (u) Putative model of peanut *AHL23* regulated mechanism in plant growth. All asterisks indicate significant differences when compared with the WT plants (T-test, \* $P < 0.05$ , \*\* $P < 0.01$ ).

## Discussion

Large-scale high-throughput detection of scRNA-seq in hypostatic plant tissue depends on the dissociation of that tissue into single cells, in addition to the separation of cells according to their spatial and morphological context, to provide crucial insight into location-specific cell development. In past, enzyme-induced protoplast generation methods were used to dissociate plant tissues (McFaline-Figueroa et al., 2020; Zhang et al., 2019). In this study, leaf blade protoplasts were successfully isolated from peanut seedlings, which allowed the characterization of major cell types of the leaf. The robust cell isolation method aided a detailed depiction of the transcriptomic landscape of the leaf blade and enabled the tracking of single leaf cell developmental trajectories using scRNA-seq approach. In the future, it would be beneficial to demonstrate a graph-based approach for the mapping of single-cell transcriptome landscapes across the timeline of leaf development. Here, we have shown that scRNA-seq can be performed on non-model plants and this technology could profoundly benefit the functional study of leaf blade cells in the allotetraploid peanut and other plant species.

### Cell differentiation of mesophyll and epidermal in peanut leaf

Apart from describing the gene expression atlas in various cell types, scRNA-seq is a powerful tool to depict the pseudo-time trajectory of leaf cell development. In our study, scRNA-seq was utilized to produce a differentiation trajectory that provides a new method for understanding the molecular dynamics of leaf mesophyll and epidermal ontogeny. Initially, mesophyll authentication was manifested with specific marker genes, *RBCS* and *LHCB*. Their expression levels were highly accumulated in spongy and palisade tissues as the leaves contained a large amount of chlorophyll to capture light (Zhang et al., 2021). Identification of epidermis mainly depends upon *GSTF* that correlates with cuticle biosynthesis and influences numerous redox-dependent processes including hormonal and stress responses (Horváth et al., 2019). *GSTF* transcript was found abundant not only in epidermis, excluding the influence of mesophyll cluster, but was also highly distributed in epidermis-derived stomata cell. Secondly, aggregate new marker genes in mesophyll and epidermal clusters remained to influence the chloroplastic biological activity, like the ribosomal protein L/S subunit (*RPL/S*), photosystem subunit protein (*PSP*) and phosphoglycolate phosphatase 1B (*PGLP1B*). As the plant material was etiolated seedling, the photosynthesis reaction should be inhibited by dark environment. However, the expression profiles of these new marker genes of photosynthesis were enriched in the leaf cell. This could be attributed to chlorophyll proliferation required for the replenishment of vast chloroplastic-like gene translated protein. This process is mainly



occurred in mesophyll and epidermal tissues along with leaf expansion as both of them contain large quantities of chlorophyll. Besides, the abovementioned marker genes exhibited stable expression level in distinct developmental states. This implies that the roles of identified signature genes may be similar to housekeeping genes that undertake the responsibility of sustaining cell differentiation with higher expression abundance.

Pseudo-time analysis in our study identified that the primordium is directly developed into parenchymal and palisade tissue but the majority of spongy tissue cells are differentiated from palisade cells with a small contribution from the parenchyma. Moreover, we speculated that the leaf epidermis probably originates from shoot apical meristems (SAMs) and, subsequently, the primordium differentiates into epidermal tissue that composes part of the complete mature epidermis. Transcriptional regulation affects the dynamics of cell differentiation. Several critical *ERF* and *NAC* transcription factors (TFs) were simultaneously involved in the development of primordium into mesophyll tissue. Further the expression of these TFs is highly accumulated in palisade tissue, and is sparsely distributed in other clusters during the mesophyll differentiation (Figure 4). The likely reason for this phenomenon is the highly expressed palisade cell occupying an important position in the pseudo-time trajectory of primordium covets to mesophyll tissue (Figure 4a). The responsive-position analysis of palisade cluster in the trajectory of cell differentiation state, which suggesting that the position coordinated three cell differentiation states from 1 to 3 (Figure 4b) along with the developmental trajectory of primordium differentiates into mesophyll. The RNA velocities of bot of these TFs were obviously declined from primordium converted into epidermis, demonstrating that hormone-responsive TFs failed to generate more effect on the highly differentiated mature epidermal cell. Furthermore, according to conventional morphological observation, the young leaf undergoes five stages of growth, including leaf primordium development, followed by formation of an apical bulge and eventual marginal and intercalary growth, all of which finally result in generation of the mature leaf (Alvarez et al., 2016). However, our cell trajectory proposal is that the cells of the inner leaf have a diverse differentiation strategy, this hypothesis fills the gaps unexplained by morphological observations thus far in the field.

Undoubtedly, transcriptional regulation determines cell fate during differentiation, as shown previously when a series of TFs were identified to correspond to cell fate transitions in cell-type ontogeny (Gaillochet et al., 2017). In consideration of the fact that our peanut material was etiolated leaf and that the constructed protein network and associated TFs were mainly focussed on the ethylene and jasmonic acid pathways, it is implied that leaf cell fate is controlled by phytohormone accumulation and the induction of skotomorphogenesis. Accordingly, future research to unravel the function of these TFs associated with leaf cell differentiation should focus on plant development during dark conditions.

### AHL23 positively modulates the leaf growth

*AHL* (*AT-HOOK MOTIF NUCLEAR LOCALIZED PROTEIN*) transcription factor contains the typical AT-hook and DUF296 domains. The AT-hook facilitates binding specificity to AT-rich DNA sequences at nuclear matrix attachment regions (MARs), whereas DUF296 enables protein-protein interactions. The AHL protein has diverse roles in the plant developmental process, including

hypocotyl elongation, gibberellin biosynthesis, pollen nexine formation and flower development (Liu et al., 2021a). AT-rich sequences are generally found in flanking regions such as the transcriptional units or regions close to *cis*-regulatory elements. AHL binds to AT-rich regions of DNA to regulate the expression of downstream target genes during epigenetic modification (Xi et al., 2020). Molecular role of several AHL TFs, such as *AHL15* inhibiting axillary meristem maturation, have been identified in *Arabidopsis* (Karami et al., 2020). Moreover, *TEK* (*AHL16*) acts in the maintenance of genomic integrity by silencing transposable elements and repeat-containing genes through epigenetic machinery (Jia et al., 2015; Xu et al., 2013). Although the functions of some members of the *AHL* protein family have been explored in *Arabidopsis*, systematic elucidations of their function have not been attempted in peanut. In this study, peanut *AHL23* in leaf scRNA-seq profiling showed its involvement in the process of differentiation of primordial cells into epidermal tissue. Furthermore, when ectopically expressed in *Arabidopsis*, *AHL23* overexpression lines exhibited significant propensity for enhanced growth, especially leaf cell expansion and development in addition to early flowering. Taken together, we conclude that *AHL23* is a positive regulator of plant development, the functional explanation of which may contribute to future agronomic trait improvement in peanut breeding. Up-regulation of *AHL23* resulted in increased leaf size that would enable the capturing of more sunlight to increase photosynthetic efficiency, which in turn could promote early flowering and reduce overall growth time. We propose to undertake full functional characterization of *AHL23* in future studies.

### Validation method for scRNA-seq results in plants

Despite the fact that scRNA-seq supports the identification of gene expression profile heterogeneity in divergent cell clusters, the current methods for subsequently validating the true expression levels of DEGs in different cell clusters are not up to mark especially with the pace of advances in scRNA-seq analysis in plants. Herein, we developed an easy, fast, cheap and high throughput approach to test the relative expression of DEGs at an isogenous cell population level. This method integrated multiple biotechnologies and user-friendly methods that could be utilized in the future to carry out scRNA-seq verification in plant leaves. The disadvantage of this method is that our process for distinguishing cell type is overly dependent on the morphology of the original cell. Therefore, we were only able to separate spongy, palisade and epidermal cell types. Due to lack of phenotype information, it is difficult to isolate primordium, parenchymal and other cell-type clusters from substantial numbers of cells. However, we consider that the visualization of spatial transcriptomics addressed the deficiencies of the scRNA-seq and validated obtained results successfully. The integrated analysis of scRNA-seq and spatial transcriptomics will have a prominent effect on future plant cell atlas studies.

### Future application of the scRNA-seq in plants

Single-cell RNA-seq has allowed development of cell expression atlases, integrated with computational algorithms, have opened up new avenues for investigating the cell-type composition of tissues and interpreting various development process, for example, stem-cell proliferation and differentiation (Rodriguez-Villalon and Brady, 2019). Despite the broad use of scRNA-seq for studying post embryonic organogenesis of root and shoot apical

meristem (Serrano-Ron *et al.*, 2021), isolating living individual cells from distinct plant tissues remains difficult. That notwithstanding, the technology is gradually improving to obtain single nucleus (snRNA-seq) to describe the cell atlas. Very recently, snRNA-seq has been used to generate a transcriptional atlas of developing seeds in *Arabidopsis* (Picard *et al.*, 2021). We believe that scRNA-seq and snRNA-seq will be widely applied in plant development study in near future. Furthermore, we anticipate that use of these approaches (scRNA-seq and snRNA-seq) not only for basic research on cell cluster-generated gene heterogeneity, but also in germplasm resources for a given crop to assay and utilize genetic variation in crop improvement programs. When performed on transcript landscape associated with multiple critical agronomic traits, scRNA-sequencing can help develop the Pan-scRNA-seq profile in different germplasms (Lei *et al.*, 2021), especially for vital gene regulators for use in crop breeding (Hirsch *et al.*, 2014).

In addition to revealing the gene expression atlas, scRNA-seq will be used in several areas of plant research, especially in researching response to biotic and abiotic stress. Environment stress triggered transcriptome profile changes are broadly detected in plant tissues, but specific factors like drought, ultraviolet radiation or chemical mutagenesis that induce transcriptome variations have not been explored at single-cell level. By coupling single-cell isolation with full-length transcriptome analysis, variations in transcript expression levels and structure in single somatic cell can be detected for better understanding of plant adaptation to environmental stress (Long *et al.*, 2021). Moreover, leaf is the most important aerial part of a plant, which is exposed to phytopathogen invasion. Further integrating scRNA-seq with bulk RNA-seq in plant leaves will be helpful for identification of major cell clusters involved in immune resistance pathways. Along the lines of conventional metagenome (Nayfach & Pollard, 2016), individual cell separation technology can help develop a novel interdisciplinary field, the single-cell metagenome, study the metagenome in single cell for future identification of advantageous endosymbiotic microorganisms in plants. Though scRNA-seq application remains restricted by the means for individual cell isolation in plants, we foresee rapid progress that will certainly shed light on hitherto unexplored mechanisms of plant development.

## Methods

### Peanut leaf protoplast isolation

Young seedlings of the peanut variety 'Hanghua2hao' were grown under dark conditions for one week at room temperature (25 °C), and then the leaf blades of these seedlings were collected. The leaf was cut into 1–2 mm strips and added to tubes containing 30 ml enzyme solution consisting of 3% cellulase R-10, 1.5% macerozyme R-10, 0.3% pectinase Y-23, 0.25% BSA, 5 mM MES and 8% (w/v) mannitol (without Ca<sup>2+</sup> and Mg<sup>2+</sup>) with shaking at 40 rpm for 2 hours at 25 °C. Cells were then filtered with a 40 µm cell strainer. Cell activity was detected by trypan blue staining and cell concentration was measured using a hemocytometer and a light microscope. Lastly, protoplasts were resuspended in 8% (w/v) mannitol solution in preparation for loading onto the chromium controller of the 10xGenomics platform. Approximately, 2×10<sup>4</sup> isolated single cells and enzyme gel-beads were packed into a single oil droplet for single-cell RNA-seq library construction.

### scRNA-seq Library construction

The scRNA-seq libraries were constructed with the Chromium Single-Cell 3' GEM (Gel Beads-in-Emulsion) Library & Gel Bead Kit v3, as per the user manual. The Single-Cell 3' protocol produces Illumina-ready sequencing libraries. A Single-Cell 3' Library comprises of standard Illumina paired-end constructs.

### Data processes

Cell Ranger software was used to perform splicing-aware alignment of reads of the peanut genome (GCF\_003086295 in Peanutbase.org) and for the assignment of cell barcodes. When combining data from multiple libraries, we recommend equalizing the read depth between libraries before merging, this reduces any batch effect introduced during sequencing. For cell clustering, Seurat was employed as a clustering tool for scRNA-seq data, was used to perform data quality control and exploration of our scRNA-seq data. To overcome the extensive technical noise in any single gene analysed, Seurat clustered the cells based on their PCA scores. Use of the Seurat software provided us with a graph-based clustering approach for analysis of the scRNA-seq data. Seurat utilizes t-SNE (t-distributed Stochastic Neighbor Embedding) as a powerful tool to visualize and explore scRNA-seq datasets. The t-SNE tool enabled the assignment of cells with similar local neighbourhoods in high-dimensional and low-dimensional spaces (Trapnell *et al.*, 2014; Xu and Su, 2015; Kobak and Berens, 2019). To identify the differentially expressed genes (DEGs), we used a likelihood-ratio test to identify differential expression when comparing a single cluster to all other cells. GO enrichment analysis provided all GO terms that were significantly enriched in differentially expressed genes compared to the background genome. Furthermore, we filtered for differentially expressed genes that corresponded to biological functions. All enriched genes were mapped to GO terms in the Gene Ontology database (geneontology.org), gene numbers were calculated for every term and significantly enriched GO terms in differentially expressed genes were defined by a hypergeometric test. KEGG pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in differentially expressed genes. For marker gene analysis, we further selected the top five genes according to the results of the analysis of DEGs. Next, the distribution of expression of each marker gene was demonstrated by using a heatmap and bubble diagram. TBtools was utilized to draw the heatmap of gene expression (Chen *et al.*, 2020).

### Pseudo-time trajectory analysis

Single-cell trajectories were constructed using a matrix of cells and their gene expression profiles by Monocle (Version3.0). Once the cells were ordered, we were able to visualize the trajectory in the reduced dimensional space. The trajectory was visualized in a tree-like structure, including tips and branches. Differential Expression Analysis in Monocle was used to identify genes that were differentially expressed between groups of cells and to assess the statistical significance of those findings. We identified key genes related to the development and differentiation process, and grouped genes with similar trends in expression, with the reasoning that such groups might share common biological functions and regulators (Qiu *et al.*, 2017; Trapnell *et al.*, 2014).

### Validation of scRNA-seq by qPCR

Peanut seedlings were grown under dark conditions for one week at room temperature (25 °C) and the leaf blade of each plant was cut into small strips and degraded by an enzyme solution (1.5% cellulase R-10, 0.75% macerozyme R-10, 0.3% pectinase Y-23, 0.25% BSA, 5 mM MES, 8% mannitol) for 30 mins. Next, the cells were filtered using a 40 µm cell strainer before being stained by trypan blue to select the dead cells by microscopic evaluation. A glass capillary was next used to capture the living single cells of spongy, plastid and epidermal tissue. Cells of each type were distinguished by observation of their morphology. Each isolated cell population (~200 cells) was added to pre-prepared 2ml centrifuge tubes containing a cellulase-pectinase enzyme solution (1.5% cellulase R-10, 0.75% macerozyme R-10, 0.25% BSA, 5 mM MES and 8% mannitol) in order to digest the cell wall completely after 1 hour. Subsequently, we collected the different cell groups by using a centrifuge and spinning at 300g for 5 min. RNA was extracted from each cell group to construct libraries by following the SMART-seq (switching mechanism at 5' end of the RNA transcript sequencing) protocol (SMART-Seq HT Kit, Takara) (Picelli et al., 2014). Lastly, a cDNA library served as a template for detecting the gene expression level by applying conventional quantitative PCR with the ABI step one plus system (Liu et al., 2017). The epidermal cell population was used as the reference sample and RBCS were used as the internal reference control (Table S21).

### Experimental processes of AHL23 functional study

*AHL23* was cloned from the peanut leaf cDNA library. NCBI Protein BLAST search was used to identify the homologous *AHL23* gene in a diversity of plant species and MEGA version 7 was used to construct the phylogenetic tree. For subcellular localization analysis, an *AHL23* CDS sequence without the terminator bases was inserted into the pNA580 vector in order to construct the *AHL23*-GFP plasmid (Figure S13), which is driven by a 3SS promoter. *AHL23*-GFP was transiently expressed in *Arabidopsis* protoplast cells as per a previously published method (Liu et al., 2019). To obtain transgenic plants, the *AHL23* coding sequence was cloned into the pBWA(V)HS vector to construct a 35S::*AHL23* plasmid (Figure S13) that was transfected into Agrobacterium EHA105 to facilitate the *Arabidopsis* inflorescence. The generated T0 seed was selected by hygromycin (Hyg) until we obtained stable *AHL23* overexpression lines at T2 generation. PCR was used to identify expression levels of *AHL23* in *AHL23*-OX transgenic lines by using the transgenic plants' DNA as a template (Figure S12). We utilized the *AHL23*-OX lines to record the leaf phenotype and growth state. The Columbian *Arabidopsis* species served as a wild-type plant. Plant hormone examination was done by using LC-MS with three biological repeats (Figure S14).

### Acknowledgements

This study was supported by Key-Area Research and Development Program of the Guangdong Province (2020B020219003), National Natural Science Foundation of China (32001442), Guangdong Basic and Applied Basic Research Foundation (2020A1515010021), China Agriculture Research System of MOF and MARA (CARS-13), Technology Special Fund of Guangdong Province Agriculture and Rural Affairs Department (2019KJ136-02), Agricultural Competitive Industry Discipline

Team Building Project of Guangdong Academy of Agricultural Sciences (202104TD), Special Fund for Scientific Innovation Strategy-Construction of High Level Academy of Agriculture Science (R2016PY-JX001, R2017YJ-YB2001, R2018PY-JX001, R2018QD-049, R2018QD-050, R2019PY-QF006). R.K.V. is thankful to Science & Engineering Research Board (SERB) of Department of Science & Technology (DST), Government of India for providing the J C Bose National Fellowship (SB/S9/Z-13/2019).

### Conflict of interest

The authors declare that there is no conflict of interest.

### Author contributions

Hao Liu and Yanbin Hong developed the protoplast isolation approach, Dongxiu Hu and Puxuan Du performed experiment on scRNA-seq validation and *AHL23* function analysis, Qing Lu and Haifen Li performed the scRNA-seq experiment, Shaoxiong Li and Haiyan Liu analysed the scRNA-seq data. Liping Wang processed the experimental figure. Xiaoping Chen facilitated funding. Xuanqiang Liang supervised the study. Rajeev K Varshney contributed to conceptualization of the study, intellectual guidance, data analysis and interpretation, and Yanbin Hong and Xiaoping Chen conceived the study.

### References

- Alvarez, J.P., Furumizu, C., Efroni, I., Eshed, Y. and Bowman, J.L. (2016) Active suppression of a leaf meristem orchestrates determinate leaf growth. *Elife*, **5**, e15023.
- Bertioli, D.J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. et al. (2019) The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.*, **51**, 877–884.
- Chen, C.J., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y.H. and Xia, R. (2020) TBTools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant*, **13**, 1194–1202.
- Chen, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., Li, H. et al. (2019) Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. *Mol. Plant*, **12**, 920–934.
- Chen, X.P., Yang, Q.L., Li, H.F., Li, H.Y., Hong, Y.B., Pan, L.J., Chen, N. et al. (2016) Transcriptome-wide sequencing provides insights into geocarpy in peanut (*Arachis hypogaea* L.). *Plant Biotech. J.*, **14**, 1215–1224.
- Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K. and Timmermans, M. (2019) Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev Cell*, **48**, 840–852.
- Evans, J.R. (2021) Mesophyll conductance: walls, membranes and spatial complexity. *New Phytol.*, **229**, 1864–1876.
- Gaillochet, C., Stiehl, T., Wenzl, C., Ripoll, J.J., Bailey-Steinitz, L.J., Li, L., Pfeiffer, A. et al. (2017) Control of plant cell fate transitions by transcriptional and hormonal signals. *Elife*, **6**, e30135.
- Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Peñagaricano, F. et al. (2014) Insights into the maize pan-genome and pan-transcriptome. *Plant Cell*, **26**, 121–135.
- Horváth, E., Bela, K., Holinka, B., Riyazuddin Riyazuddin, R., Gallé, A., Hajnal, A., Hurton, A. et al. (2019) The *Arabidopsis* glutathione transferases, AtGSTF8 and AtGSTU19 are involved in the maintenance of root redox homeostasis affecting meristem size and salt stress sensitivity. *Plant Sci.*, **283**, 366–374.
- Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorritt, M.W., Saunders, L., Bubb, K.L., Trapnell, C. et al. (2019) Dynamics of gene expression in single root cells of *arabidopsis thaliana*. *Plant Cell*, **31**, 993–1011.
- Jia, Q.S., Zhu, J., Xu, X.F., Lou, Y., Zhang, Z.L., Zhang, Z.P. and Yang, Z.N. (2015) *Arabidopsis* AT-hook protein TEK positively regulates the expression of arabinogalactan proteins for Naxine formation. *Mol. Plant*, **8**, 251–260.

- Karami, O., Rahimi, A., Khan, M., Bemer, M., Hazarika, R.R., Mak, P., Compier, M. et al. (2020) A suppressor of axillary meristem maturation promotes longevity in flowering plants. *Nat. Plants*, **6**, 368–376.
- Kim, J.Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M. et al. (2021) Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell*, **33**, 511–530.
- Kobak, D. and Berens, P. (2019) The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **10**, 5416.
- Kumar, R., Pandey, M.K., Roychoudhry, S., Nayyar, H., Kepinski, S. and Varshney, R.K. (2019) Peg biology: deciphering the molecular regulations involved during peanut peg development. *Front. Plant. Sci.* **10**, 1289.
- Lei, L., Goltzman, E., Goodstein, D., Wu, G.A., Rokhsar, D.S. and Vogel, J.P. (2021) Plant pan-genomics comes of age. *Annu. Rev. Plant. Biol.* **72**, 411–435.
- Liu, H., Dong, S., Gu, F., Liu, W., Yang, G., Huang, M., Xiao, W. et al. (2017) NBS-LRR protein Pik-H4 interacts with OsBHD1 to balance rice blast resistance and growth by coordinating ethylene-brassinosteroid pathway. *Front. Plant Sci.* **8**, 127.
- Liu, H., Dong, S., Li, M., Gu, F., Yang, G., Guo, T., Chen, Z. et al. (2021a) The class III peroxidase gene OsPrx30, transcriptionally modulated by the AT-hook Protein OsATH1, mediates rice bacterial blight-induced ROS accumulation. *J. Integr. Plant. Biol.* **63**, 393–408.
- Liu, H., Gu, J., Lu, Q., Li, H., Gu, J., Ren, L., Deng, L. et al. (2019) Transcriptomic analysis reveals the high-oleic acid feedback regulating the homologous gene expression of stearoyl-ACP desaturase 2 (SAD2) in peanuts. *Int. J. Mol. Sci.* **20**, 3091.
- Liu, H., Hong, Y., Lu, Q., Li, H., Gu, J., Ren, L., Deng, L. et al. (2020a) Integrated analysis of comparative lipidomics and proteomics reveals the dynamic changes of lipid molecular species in high-oleic acid peanut seed. *J. Agric. Food. Chem.* **68**, 426–438.
- Liu, Q., Liang, Z., Feng, D., Jiang, S., Wang, Y., Du, Z., Li, R. et al. (2021b) Transcriptional landscape of rice roots at the single cell resolution. *Mol. Plant*, **14**, 384–394.
- Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J. et al. (2020b) Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol. Plant*, **13**, 1178–1193.
- Long, Y.P., Liu, Z.J., Jia, J.B., Mo, W.P., Feng, L., Lu, D.D., Liu, B. et al. (2021) FlsnRNA-seq: protoplasting-free full-length single-nucleus RNA profiling in plants. *Genome Bio.* **22**, 66.
- McFaline-Figueroa, J.L., Trapnell, C. and Cuperus, J.T. (2020) The promise of single-cell genomics in plants. *Curr. Opin. Plant Biol.* **54**, 114–121.
- Nayfach, S. and Pollard, K.S. (2016) Toward Accurate and quantitative comparative metagenomics. *Cell*, **166**, 1103–1116.
- Nelms, B. and Walbot, V. (2019) Defining the developmental program leading to meiosis in maize. *Science*, **364**, 52–56.
- Picard, C.L., Povilus, R.A., Williams, B.P. and Gehring, M. (2021) Transcriptional and imprinting complexity in *Arabidopsis* seeds at single-nucleus resolution. *Nat. Plants*, **7**, 730–738.
- Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S. and Sandberg, R. (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181.
- 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. Meth.* **14**, 979–982.
- Rodriguez-Villalon, A. and Brady, S.M. (2019) Single cell RNA sequencing and its promise in reconstructing plant vascular cell lineages. *Curr. Opin. Plant Biol.* **48**, 47–56.
- Runions, A., Tsiantis, M. and Prusinkiewicz, P. (2017) A common developmental program can produce diverse leaf shapes. *New Phytol.* **216**, 401–418.
- Ryu, K.H., Huang, L., Kang, H.M. and Schiefelbein, J. (2019) Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* **179**, 1444–1456.
- Satterlee, J.W., Strable, J. and Scanlon, M.J. (2020) Plant stem-cell organization and differentiation at single-cell resolution. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33689–33699.
- Serrano-Ron, L., Perez-Garcia, P., Sanchez-Corronero, A., Gude, I., Cabrera, J., Ip, P.L., Birnbaum, K.D. et al. (2021) Reconstruction of lateral root formation through single-cell RNA-seq reveals order of tissue initiation. *Mol. Plant*. <https://doi.org/10.1016/j.molp.2021.05.028>. [Epub ahead of print].
- Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M. et al. (2019) High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* **27**, 2241–2247.
- Song, Y.H., Shim, J.S., Kinmonth-Schultz, H.A. and Imaizumi, T. (2015) Photoperiodic flowering: time measurement mechanisms in leaves. *Annu. Rev. Plant Biol.* **66**, 441–464.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J. et al. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386.
- Von Wangenheim, D., Wells, D.M. and Bennett, M.J. (2017) Adding a piece to the leaf epidermal cell shape puzzle. *Dev. Cell*, **43**, 255–256.
- Wang, Y., Huan, Q., Li, K. and Qian, W.F. (2021) Single-cell transcriptome atlas of the leaf and root of rice seedlings. *J. Genet. Genomics*. <https://doi.org/10.1016/j.jgg.2021.06.001>. [Epub ahead of print].
- Xi, Y.P., Park, S.R., Kim, D.H., Kim, E.D. and Sung, S. (2020) Transcriptome and epigenome analyses of vernalization in *Arabidopsis thaliana*. *The Plant J.* **103**, 1490–1502.
- Xu, C. and Su, Z. (2015) Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics*, **31**, 1974–1980.
- Xu, X., Crow, M., Rice, B.R., Li, F., Harris, B., Liu, L., Demesa-Arevalo, E. et al. (2021) Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell*, **56**, 557–568.
- Xu, Y.F., Wang, Y.Z., Stroud, H., Gu, X.F., Sun, B., Gan, E.S., Ng, K.H. et al. (2013) A matrix protein silences transposons and repeats through interaction with retinoblastoma-associated proteins. *Curr. Biol.* **23**, 345–350.
- 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.
- Zhang, T.Q., Xu, Z.G., Shang, G.D. and Wang, J.W. (2019) A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant*, **12**, 648–660.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1** Brief chart of Cell Ranger software report.

**Figure S2** Data quality control determined the medium number of gene and UMI. A, violin plots. B, dot plots. C, Correlation coefficient between the UMI and gene, before filter. D, Correlation coefficient of the UMI and gene, after filter.

**Figure S3** The number of cells in each peanut leaf cell cluster.

**Figure S4** Circular visualization displaying the global transcriptional analysis of all identified genes (relative expression level) in peanut leaf cell by scRNA-seq (Table S3). Form outer circular to inner circular represented the cell cluster 1 to 8, respectively.

**Figure S5** Dot and violin plots of 44 reported marker genes for cell cluster.

**Figure S6** Novel 40 marker gene distribution in t-SNE map, which selected from each cell cluster (Top five genes in each cluster).

**Figure S7** Pseudotime trajectory analysis identified the total trajectory map of all cell. A, development trajectory of all leaf cells. B, the distribution of each separated cell clusters in trajectory map. C, Clustering and expression kinetics of identified DEGs along with the pseudotime trajectory. D-E, All single-cells can be divided into nine state profiles of cell differentiation based on the trajectory analysis. F, Pseudo-time trajectory analysis identified the expression pattern of ten critical marker genes in nine state profiles along with peanut leaf development and differentiation. G, Clustering and expression kinetics of identified DEGs in each cell differentiation branch.

**Figure S8** Heatmap displayed the average expression level of represented TFs in branch 2 of all cell trajectory.

**Figure S9** PPseudotime trajectory of mesophyll and epidermal cell development. A, primordium cell developed into mesophyll cell. B, primodium developed into epidermal cell.

**Figure S10** Pseudotime trajectory analysis of primordium and parenchymal developed into mesophyll cell. A, Clustering and expression kinetics of identified DEGs (Table S13) along with the main stem of pseudotime trajectory. B, expression distribution of top ten DEGs with highest level. C, Clustering and expression kinetics of identified DEGs in cell differentiation branch 1. D, expression distribution of top ten DEGs at the point 1 of cell differentiation state. E, Clustering and expression kinetics of identified DEGs in cell differentiation branch 2. F, expression distribution of top ten DEGs at the point 2 of cell differentiation state.

**Figure S11** Pseudotime trajectory analysis of primordium developed into epidermal cell. A, Clustering and expression kinetics of identified DEGs (Table S19) along with the main stem of pseudotime trajectory. B, expression distribution of top ten DEGs with highest level. C, expression distribution of top ten DEGs correlated to cell differentiation state. D, expression distribution of top ten DEGs at the point 1 of the cell differentiation state.

**Figure S12** Molecular function analysis of peanut AHL23. A, previous transcriptome (RNA-seq) analysis identified the AHL23 transcript abundance during the seed development of normal oleic acid variety (L70) and high oleic acid variety (H176). B, real-time PCR validated the truth expression level of AHL23 in L70 and H176 seed. C, SMART database (smart.embl-heidelberg.de) predicted the conserved domain in AHL23 protein. D, eGFP displayed the relative expression of AHL23 in peanut plant, the value of AHL23 transcript abundance downloaded from the Peanutbase.org. E, PCR cloned the AHL23 coding sequence in AHL23-OX lines. F, Transverse sections of leaf blades that collected from wild-type and AHL23-OX seedlings grown at vegetative stage, scale bar 50 $\mu$ m. G, Scanning electron microscope (SEM) observed the size of epidermal cell in WT and AHL23-OX transgenic lines, scale bar 50 $\mu$ m.

**Figure S13** Vector construction for AHL23 function study.

**Figure S14** Brief process of phytohormone examination.

**Table S1** Total individual cell number.

**Table S2** Statistical computation of cell clusters.

**Table S3** All identified genes.

**Table S4** Reported marker gene for cell clusters.

**Table S5** All identified DEGs in different cell cluster.

**Table S6** Top 40 novel marker genes in different cell cluster.

**Table S7** List of identified DEGs involved all KEGG pathway.

**Table S8** List of identified DEGs involved all GO pathway.

**Table S9** Trajectory analysis identified the DEGs along with the timeline.

**Table S10** Trajectory identified nine states of all cell differentiation.

**Table S11** Identification of DEGs in different cell differentiation branches (four branches).

**Table S12** Identification of important transcription factors in different cell differentiation branches.

**Table S13** Trajectory analysis identified the DEGs in the pseudotime of primordium developed into mesophyll.

**Table S14** Trajectory analysis identified the important TFs in the pseudo-time of primordium developed into mesophyll.

**Table S15** Relative expression of phytohormone TFs in primordium, parenchymal, and mesophyll cell cluster.

**Table S16** Expression level of important TFs in distinct cell differentiation state of primordium developed into mesophyll.

**Table S17** Trajectory analysis identified the DEGs in cell differentiation branch of primordium developed into mesophyll.

**Table S18** Expression level of TFs in two cell differentiation branch of primordium developed into mesophyll.

**Table S19** Trajectory analysis identified the DEGs in the pseudotime map of primordium developed into epidermal cell.

**Table S20** Expression level of DEGs in distinct cell differentiation state of primordium developed into epidermal cell.

**Table S21** Primers were used in this study.

**Table S22** Hormones examination in WT and AHL23-OX lines.