PCP
PLANT & CELL PHYSIOLOGY

Regular Paper

Single-Cell RNA Sequencing of Arabidopsis Leaf Tissues Identifies Multiple Specialized Cell Types: Idioblast Myrosin Cells and Potential Glucosinolate-Producing Cells

Taro Maeda[1], Shigeo S. Sugano[2], Makoto Shirakawa[3], Mayu Sagara[3], Toshiro Ito[3], Satoshi Kondo[4,5] and Atsushi J. Nagano[1,6,*]

[1]Institute for Advanced Biosciences, Keio University, Kakuganji 246-2, Mizukami, Tsuruoka, Yamagata, 997-0052 Japan
[2]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Higashi 1-1-1, Tsukuba, Ibaraki, 305-8566 Japan
[3]Division of Biological Science, Graduate School of Science and Technology, Nara Institute of Science and Technology (NAIST), Takayama 8916-5, Ikoma, Nara, 630-0192 Japan
[4]Agriculture and Biotechnology Business Division, Toyota Motor Corporation, Toyota 1, Toyota, Aichi, 471-8571 Japan
[5]Genesis Research Institute, Inc., Noritake-Shinmachi 4-1-35, Nishi-ku, Nagoya, Aichi, 451-0051 Japan
[6]Faculty of Agriculture, Ryukoku University, Yokotani 1-5, Seta Oe-cho, Otsu, Shiga, 520-2194 Japan

*Corresponding author: E-mail, anagano@agr.ryukoku.ac.jp
(Received 27 July 2022; Accepted 25 November 2022)

The glucosinolate–myrosinase defense system (GMDS), characteristic of Brassicales, is involved in plant defense. Previous single-cell transcriptomic analyses have reported the expression profiles of multiple GMDS-related cell types (i.e. myrosinase-rich myrosin idioblasts and multiple types of potential glucosinolate synthetic cells as well as a candidate S-cell for glucosinolate accumulation). However, differences in plant stages and cell-type annotation methods have hindered comparisons among studies. Here, we used the single-cell transcriptome profiles of extended Arabidopsis leaves and verified the distribution of previously used markers to refine the expression profiles of GMDS-associated cell types. Moreover, we performed beta-glucuronidase promoter assays to confirm the histological expression patterns of newly obtained markers for GMDS-associated candidates. As a result, we found a set of new specific reporters for myrosin cells and potential glucosinolate-producing cells.

Keywords: Arabidopsis ● Glucosinolate–myrosinase complexes ● Myrosin cell ● S-cell ● scRNA-Seq ● WRKY23

Introduction

Glucosinolates (GLs) are sulfur-rich secondary metabolites characteristic of Brassicales. Myrosinases hydrolyze GLs into unstable metabolites, which then contribute to the plant defense against herbivores and pathogens (Sønderby et al. 2010, Shirakawa and Hara-Nishimura 2018, Shirakawa et al. 2022). GLs, biosynthesized from tryptophan and methionine, have a high diversity, with over 130 different documented structures. Furthermore, indole, aliphatic and benzoyl glucosinolates are known to have distinct biological functions (Sønderby et al. 2010, Nintemann et al. 2018).

Although the histological distribution differs from species to species, GLs and myrosinase accumulate in different cells in the plant (Andréasson et al. 2001, Koroleva et al. 2010, Shirakawa and Hara-Nishimura 2018, Shirakawa et al. 2022). In Arabidopsis leaves, GL-rich 'S-cells' and myrosinase-rich 'myrosin idioblasts (MI)' are adjacent to each other in the leaf veins (Koroleva et al. 2010, Berrío et al. 2021). Upon damage, infection or herbivore attack, the two components of GLs and myrosinases are exposed to each other, stimulating rapid GL hydrolysis by myrosinase (Kliebenstein et al. 2005). This chemical defense system is called the glucosinolate–myrosinase defense system (GMDS) (Mocniak et al. 2020).

GLs and myrosinase are among the most well-studied secondary metabolites of plants, but the physiological and developmental details of the GMDS-involved cells are unexplored at the single-cell level (Koroleva et al. 2000, Sønderby et al. 2010). Previous studies revealed a near-complete inventory of GL biosynthetic genes. However, the transport of GLs from production to storage sites remains obscure (Koroleva et al. 2000, Sønderby et al. 2010). The development process of MI is partly common to that of stomatal guard cells (GCs) (Li and Sack 2014, Shirakawa et al. 2014a, 2016a, 2016b, Shirakawa and Hara-Nishimura 2018), and only part of the MI-specific developmental process has been revealed.

Single-cell level analysis is crucial in distinguishing gene expression among the diverse cell types in the leaf veins. Previously, four single-cell RNA-Seq (scRNA-Seq) studies successfully analyzed Arabidopsis leaves (Berrío et al. 2021, Kim et al. 2021, Lopez-Anido et al. 2021, Procko et al. 2022). Two of the four studies, Kim et al. (2021) and Berrío et al. (2021), have discussed the GMDS-related cells. However, the previous results are difficult to compare with each other. One reason is the identical naming for the cell clusters characterized by different genetic

Plant Cell Physiol. 64(2): 234–247 (2023) doi:https://doi.org/10.1093/pcp/pcac167, Advance Access publication on 28 November 2022, available online at https://academic.oup.com/pcp

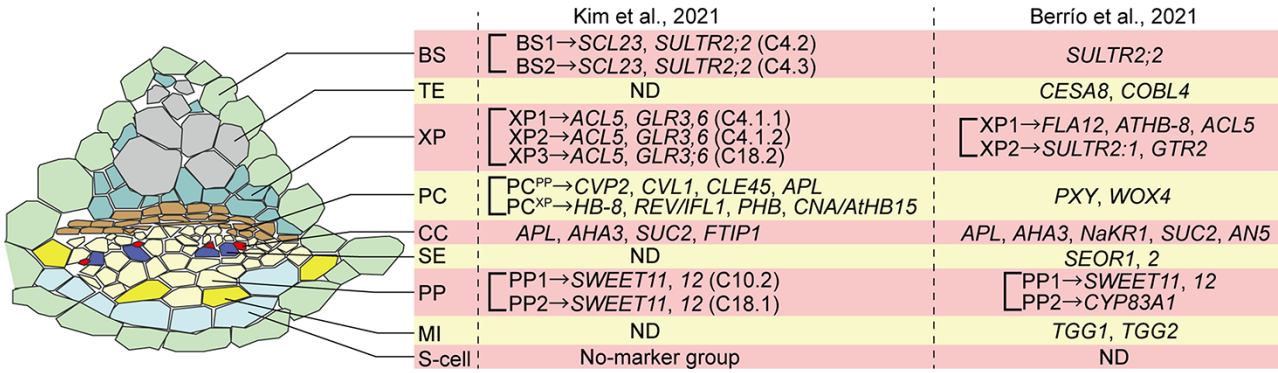

Fig. 1 Schematic hypothetical vasculature section and previously used marker genes for scRNA-Seq. The left side is a schematic hypothetical vasculature section (modified from Berrío et al. 2021). The right side shows the previously used marker genes. For the tissue divided into multiple clusters in the previous scRNA-Seq, the marker genes are described with the previously used cluster-ID (e.g. C4.2).

markers. Moreover, these studies analyzed leaves at different developmental stages via different protoplast construction methods.

The Arabidopsis vascular system contains multiple components: sieve element (SE), companion cell (CC), xylem parenchyma (XP), bundle sheath (BS), procambium cell (PC), tracheary element (TE), phloem parenchyma (PP), MI and S-cells (**Fig. 1**) (Oparka and Turgeon 1999, Leegood 2008, Berrío et al., 2021, Kim et al. 2021). The two papers commonly list PP, PC and BS cells as GL-producing cells, and Kim et al. (2021) additionally noted XP as GL-producing cells. However, the difference between their annotation markers confuses the discussion of these cells.

Two studies by Kim et al. (2021) and Berrío et al. (2021) used leaves at different developmental stages, and their experimental procedures differed as well. Kim et al. (2021) enriched the vascular cell via a modified epidermal peeling method. Berrío et al. (2021) used the whole developing leaves for their protoplast construction. Although only Berrío et al. (2021) detected the clusters corresponding MI cells, the possible causes of the difference have not been discussed.

Herein, we performed scRNA-Seq of matured Arabidopsis leaves to reveal the transcriptomic profiles and markers of GMDS-related cells. We segregated the obtained scRNA-Seq data according to the overall expression profile and correspondence with conventional cell types. This method allowed us to separate the inner cells of the leaf based on gene expression profiles. To determine the correspondence to previous studies, we verified the marker distribution from previous research and their histological expression by beta-glucuronidase (GUS) analysis.

Result

Identification of myrosin, epidermal and phloem cell clusters in Arabidopsis leaves

Our scRNA-Seq identified 17,373 cells in protoplasts isolated from long-day-grown Arabidopsis rosette leaves (3- to 4-weeks old; **Fig. 2A**). We used standard computational pipelines (Cell Ranger provided by 10× Genomics) to align the raw sequencing data to the *Arabidopsis thaliana* genome (TAIR10) and derived a gene expression matrix of 21,376 genes. The scRNA-Seq yielded 12,845 reads per cell. After quality control of the dataset (see Materials and Methods), 19,110 genes across 3,264 cells were retained. The number of median expressing genes per cell (=1,328) was lower than that found in previous papers [Kim et al. 2021 (3,342), Berrío et al. 2021 (2,017)].

To identify distinct cell-type populations and compare cell-type identities, we analyzed all QC-treated cells via unsupervised clustering analysis and partitioned the cells into seven transcriptionally distinct clusters (**Fig. 2B**). The results of the unsupervised clustering attempted to annotate these seven clusters with unique molecular identifier (UMI) count distribution and previously used marker genes for leaf cell annotation (**Fig. 2C–E** and **Supplementary data S1**). In addition, as a distinct analysis from the clustering, we visualized local similarities and global structures in the cell populations with a uniform manifold approximation and projection (UMAP) algorithm. We then overlapped the clustering results with the visualized local similarities. As a result, our clustering corresponds with the UMAP-based similarity of transcriptomic profiles (**Fig. 2B**).

We identified the origin of two cell types based on single-cell transcriptome profiles: epidermal cells (cluster-4) and MI (cluster-6; **Fig. 2B, E**). These clusters were separated from others on the UMAP plot, suggesting their unique transcriptome signatures. Our UMAP plot showed that cluster-3 also had distinct expression profiles from other clusters (**Fig. 2B**), although this cluster contained multiple cell markers, including CC markers.

Cluster-4 was annotated as epidermal cells based on expression of *FDH*, *KCS1* and *LTPG1* (**Fig. 2E** and **Supplementary data S1**). *FDH* and *KCS1* encode 3-ketoacyl-CoA synthases related to the cuticular wax and suberin biosynthesis (Wu et al. 2011). *LTPG1* is a marker of epidermis cells in Berrío et al. (2021). The expression of these genes was restricted to cluster-4, suggesting that no other clusters contain epidermal cells.

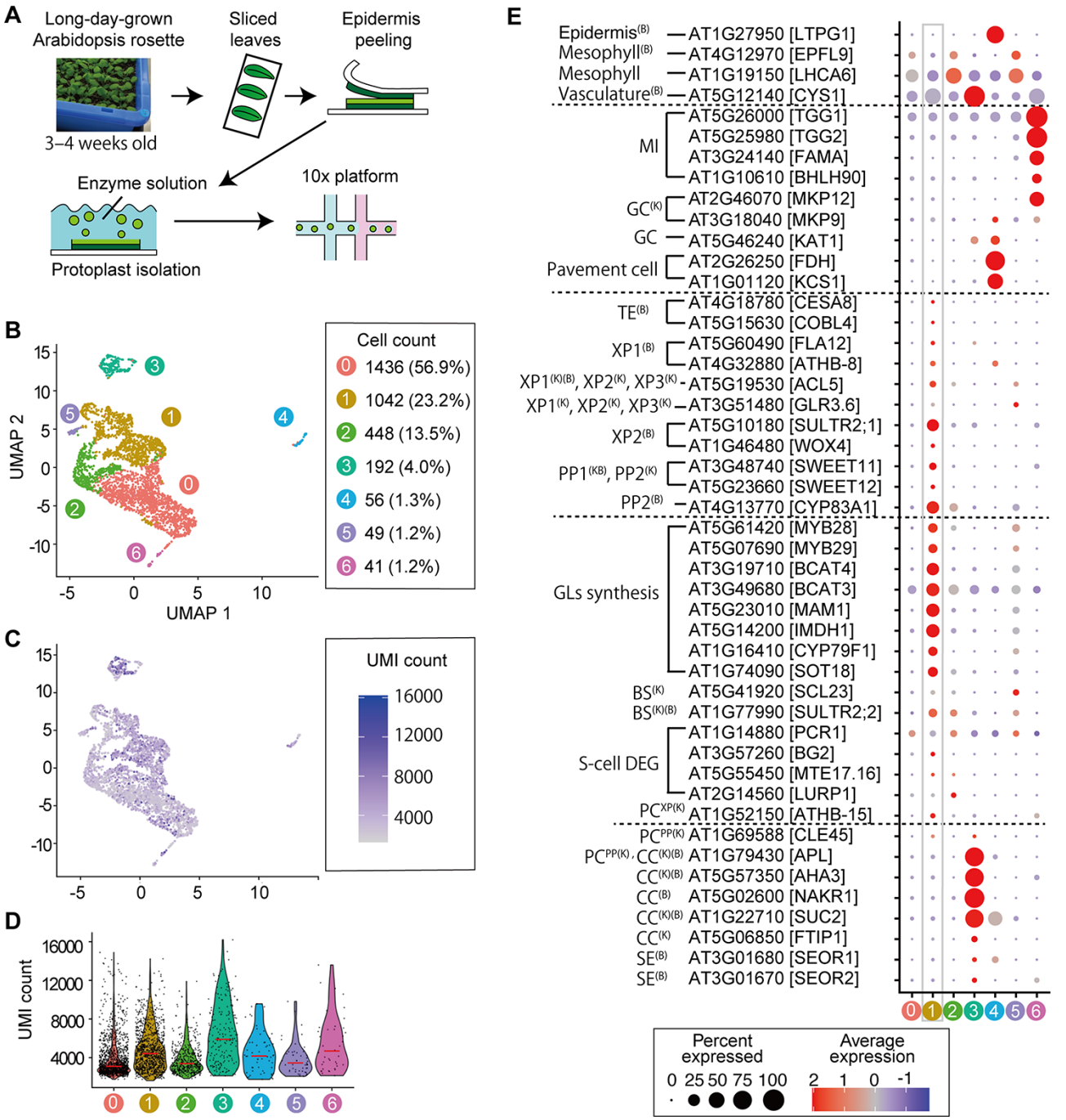

Fig. 2 Clustering and annotation of *Arabidopsis thaliana* single-cell transcription profiles. (A) Overview of the protoplast isolation procedure. (B) Two-dimensional embedding of the transcription profile of cells that passed the quality check. UMAP visualized transcription profile similarity among cells. Cells are colored by cluster identity via graph-based clustering. The number of cells belonging to each cluster was noted in the box. (C) The UMAP-visualized transcription with the cell coloration by the number of UMI counts. The coloration of the UMI count for each cell was noted in the box. (D) Violin plot of the UMI count distribution in each cluster. (E) Expression of the adopted cell-type markers. Dot area is the proportion of cluster cells expressing a given gene. The color indicates expression across cells in that cluster. Annotation information is described on the left side of each marker. The superscript of the annotation information indicates the references (K = Kim et al. 2021, B = Berrío et al. 2021).

Although cluster-4 consisted of the third-lowest number of cells (1.3%; **Fig. 2B**), this small amount would result from epidermis peeling prior to protoplast isolation.

We identified cluster-6 as MI with expression patterns of *TGG1*, *TGG2*, *FAMA*, *bHLH090* and *KAT1* (**Fig. 2E** and **Supplementary data S1**). *TGG*s encode myrosinases. *FAMA* and

bHLH090 are transcription factors that promote MI differentiation. Although *TGGs* and *FAMA* are commonly expressed in the MI and GC (Ohashi-Ito and Bergmann 2006, Li and Sack 2014, Shirakawa et al. 2014a), *bHLH090* shows a myrosin-cell-specific expression. *KAT1*, which encodes a potassium channel protein, is known as a GC-specific gene (Nakamura et al. 1995). We found no expression of *KAT1* in cluster-6. The expression pattern of these five genes supported cluster-6 containing MIs but no GCs. Cluster-6 corresponded to the minor cell type (1.2% of the total cells, **Fig. 2B**). This ratio is consistent with the microscopic observation of leaf MIs.

We then compared our MI cluster (cluster-6) with the previously indicated GC cluster by Kim et al. (2021). Using a similar epidermis peeling method, Kim et al. (2021) detected a cluster corresponding to GC but no MI cluster. To clarify the discussion in this study, we have specified the reference information for the cell type. For example, $GC^{(K)}$ means GC cell types by Kim et al. (2021), hereafter. Kim et al. (2021) used three markers (*FAMA*, *MKP12* and *MKP9*) for their $GC^{(K)}$ annotation. However, as mentioned above, *FAMA* is commonly expressed in GC and MI. Regarding *MKP12* and *MKP9* expression, no previous study investigated MI. Our MI cluster, cluster-6, showed *MKP12* expression but no *MKP9* expression. The lack of *MKP9* expression suggests that our MI cluster is not completely identical to $GC^{(K)}$, which expresses both *MKP12* and *MKP9*. However, other markers suggest that $GC^{(K)}$ was similar to our MI cluster, although Kim et al. (2021) provided no information concerning *KAT1* expression.

The expression of multiple cell markers indicated that cluster-3 is a mixture of CCs, $PC^{PP(K)}$ and $SE^{(B)}$. Most cluster-3 cells (70–84%) expressed markers for CCs (*APL*, *AHA3*, *NAKR1* and *SUC2*; **Fig. 2E** and **Supplementary data S1**). For *FTIP1*, which was adopted as a CC marker by Kim et al. (2021), the percentage of expressing cells was lower (4%) than that of other CC markers (**Fig. 2E** and **Supplementary data S1**). A portion of the cluster-3 cells showed expression of the *CLE45* marker for $PC^{PP(K)}$ and the *SEOR1* and *SEOR2* markers for $SE^{(B)}$ (**Fig. 2E**). This result suggested that majority of cluster-3 is CC, although this cluster also contains $PC^{PP(K)}$ and $SE^{(B)}$.

Cell clusters with unclear identification

EPFL9 and *LCHA6* expression suggested that cluster-0, cluster-1, cluster-2 and cluster-5 include mesophyll cells (**Fig. 2E** and **Supplementary data S1**). These four clusters included cells that express the mesophyll markers *EPFL9* (percent expressed cells were 6%, 2%, 9% and 12%, respectively) and *LCHA6* (30%, 22%, 50% and 43%, respectively) (**Fig. 2E** and **Supplementary data S1**).

Cluster-1, however, showed expression of multiple cell markers in addition to mesophyll markers, indicating that this cluster is a mixture of multiple cell types. Several cluster-1 cells showed marker gene expression for vasculature (*CYS1*), $TE^{(B)}$ (*CESA8* and *COBL4*), XPs (*FLA12*, *ATHB-8*, *ACL5*, *SULTR2;1* and *WOX4*), PPs (*SWEET11*, *SWEET12* and *CYP83A1*), BSs (*SULTR2;2*) and $PC^{XP(K)}$ (*ATHB-15*). Moreover, we detected the expression of previously reported 'potential S-cell'-specific genes (*BG2* and *MTE17.16*) (Kim et al. 2021) and multiple genes for GL synthesis (*MYB28* and *MYB29* as positive regulators for GLs biosynthesis and *BCAT4*, *BCAT3*, *MAM1*, *IMDH1*, *CYP79F1*, *CYP83A1* and *SOT18* as genes directly involved in GL biosynthesis; **Fig. 2E**). Accordingly, we considered that cluster-1 contains multiple types of GMDS-related cells and reclustered it as in the following sections.

Cell cluster annotation for GL synthesis

To determine the cell clusters related to GMDS, we extracted cluster-1 cells and reanalyzed them based on the newly selected top 3,000 highly variable genes within cluster-1. A new unsupervised clustering partitioned the cells into seven transcriptionally distinct subclusters (**Fig. 3A**).

The enrichment in GL synthetic genes enabled us to annotate subcluster-α1, subcluster-α4 and subcluster-α5 as potential GL-producing cells (**Fig. 3B** and **Supplementary data S2**). The expression of a positive regulator for GL biosynthesis, *MYB28*, was determined on the three subclusters (**Fig. 3B**). Subcluster-α4 indicated another positive regulator for GL biosynthesis (*MYB29*) and seven genes for GL biosynthesis (*BCAT4*, *BCAT3*, *MAM1*, *IMDH1*, *CYP79F1*, *CYP83A1* and *SOT18*; **Fig. 3B**). Expression of the seven GL synthetic genes was indicated in subcluster-α1 and subcluster-α5, although the expression levels were weaker than that in subcluster-α4.

For subcluster-α1, the expression of previously used markers to annotate the XPs ($XP2^{(B)}$, $XP1^{(B)}$, $XP1^{(K)}$, $XP2^{(K)}$ and $XP3^{(K)}$) showed that most α1 cells are derived from the XPs (**Fig. 3B**). Major cells of α1 indicated the expression of *SULTR2;1* (68%), which was used to identify $XP2^{(B)}$ (**Fig. 3B** and **Supplementary data S2**). Although the *SULTR2;1* expression was present in α4 (54%), the expression level was weaker than that in α1 (**Fig. 3B**). Another $XP2^{(B)}$ marker, *WOX4*, was detected in a portion of Subclusters α1 and α4. Other markers for XP-related clusters (*FLA12*, *ATHB-8*, *ACL5* and *GLR3.6*) were detected in a portion of cells in α1 (1–16%, **Fig. 3B**). Although markers for $PC^{XP(K)}$, $PC^{PP(K)}$ and $TE^{(B)}$ (*ATHB-15*, *CLE45*, *CESA8* and *COBL4*) were detected in the α1 subcluster (1–6%) (**Fig. 3B**), we concluded that most of the α1 cells belong to XPs.

The subcluster-α4 seems to correspond with $PP2^{(B)}$, which is defined by *CYP83A1* expression (Berrío et al. 2021). Kim et al. (2021) and Berrío et al. (2021) used identical names for the several cell clusters defined by the different genetic markers (**Fig. 1**). Berrío et al., (2021) characterized a subtype of PP, 'PP2', as the cell indicating no signal of *SWEET11* and *12* but *CYP83A1* genes. However, Kim et al. (2021) used the same term, 'PP2', when naming the cell cluster expressing *SWEET11* and *12*. We found that most cells of subcluster-α4 (83%) indicated *CYP83A1* expression, although several α4 cells (1–54%) showed marker gene expressions for other cell types (i.e. *EPFL9*, *LHCA6*, *GLR3.6*, *SULTR2;1*, *WOX4*, *SCL23* and *SULTR2;2*) (**Fig. 3B**, **Supplementary data S2**).

For subcluster-α5, the specific expression of *SWEET11* and *10* indicated that this subcluster contains $PP1^{(K)(B)}$ and $PP2^{(K)}$

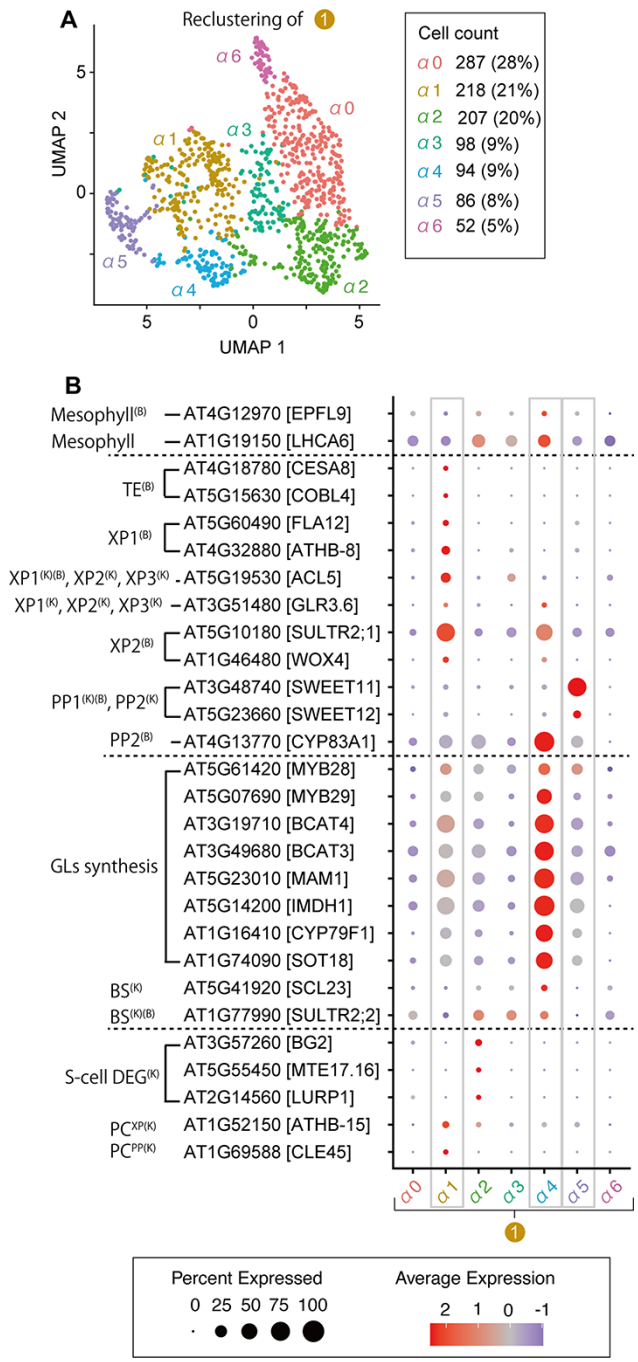

Fig. 3 Clustering and annotation of *Arabidopsis thaliana* single-cell transcription profiles. (A) UMAP-based two-dimensional embedding of the transcription profile of cluster-1 cells in **Fig. 2B**. Cells are colored by their reclustered identity. (B) Expression of the adopted cell-type marker genes on reanalyzed subclusters from cluster-1. Dot size and color have the same meaning as in **Fig. 2E**.

cells (8–70%, **Fig. 3B** and **Supplementary data S2**). Unlike the results in the study by Kim et al., our clustering did not split *SWEET11*- and *10*-expressing cells into two clusters (i.e. PP1$^{(K)}$ and PP2$^{(K)}$). Clustering parameters may cause this difference.

Myrosin-cluster-specific genes and *in situ* visualization

After the cell clustering and annotation, we identified a set of highly expressed genes in cluster-6 (annotated as MI) to find new marker genes for MI. We compared cluster-6 with all other clusters and detected 238 upregulated differentially expressed genes (DEGs) (adjusted *P*-value < 0.01, **Fig. 4A**, **Supplementary data S3** and **S4** and **Supplementary Fig. S1**). We performed the same analysis on the other clusters and removed overlapping DEGs to detect cluster-6-specific DEGs. As a result, 164 genes remained as MI-specific DEGs (**Fig. 4A** and **Supplementary data S5**). The 164 genes contained three previously known markers for MI (*TGG1*, *TGG2* and *FAMA*) (Barth and Jander 2006, Ueda et al. 2006, Li and Sack 2014, Shirakawa et al. 2014a). Previous scRNA-Seq analysis of developing leaves identified 78 genes as MI markers (Berrío et al. 2021); our 164 markers contained 20 of these 78 genes (**Fig. 4A** and **Supplementary data S5**). Hereafter, we refer to these 20 genes as conserved MI markers.

To confirm marker specificity, we visualized the expression profiles of several of the marker genes, *RAB18*, *UMAMIT10* and *WRKY23*, through a GUS reporter assay (**Fig. 4B-D**). We selected *RAB18* based on the high average log2FC value (**Fig. 4A**). Among the 20 conserved MI markers, we then selected *UMAMIT10* because of the high log2FC value and the clear physical function of UMAMIT family (Zhao et al. 2021). We also focused on *WRKY23* because this gene is directly connected with *TGG1* in ATTED-II (**Fig. 4E**).

In *ProRAB18:GUS*, GUS signals were detected at both MIs and GCs, similar to *TGG1* and *FAMA* (**Fig. 4B**) (Barth and Jander 2006, Li and Sack 2014, Shirakawa et al. 2014a). In *ProUMAMIT10:GUS* and *ProWRKY23:GUS*, GUS signals were detected in the idioblasts along leaf veins (i.e. MI) in the inner tissues similar to *TGG2* and *bHLH090* (**Fig. 4C, D**) (Barth and Jander 2006, Shirakawa et al. 2014a).

To examine whether FAMA is required for the expression of *WRKY23*, we introduced *ProWRKY23:GUS* into mutants of *FAMA* (*fama-1*), a master regulator of MI differentiation (Li and Sack 2014, Shirakawa et al. 2014a). *FAMA* loss dramatically reduced GUS signals along leaf veins (**Fig. 4F**). Our quantitative PCR analysis showed that *WRKY23* expression levels in *fama-1* seedlings were reduced to 23% of those in the wild type (WT) (**Fig. 4G**). These results indicated that *WRKY23* is a newly found *FAMA*-regulated gene and suggested its contribution to myrosin cell development. *WRKY23* is required for polar auxin transport (PAT), which is necessary for MI development (Grunewald et al. 2012, Shirakawa et al. 2014b, 2016b). Therefore, *WRKY23* may be a link between PAT and MI development.

Our scRNA-Seq provided insights into the MI unknown functions. From the cluster-6-specific DEGs, we found six genes related to water transport and response: *PIP2;3*, *RAB18*, *FAMA*, *ANN1*, *HAT22* and *P5CS1* (**Supplementary data S4**). Myrosin cells may play an unknown role in maintaining water responses in plants as they are distributed close to veins through which water circulates in plants (Shirakawa et al. 2014a, 2016a).

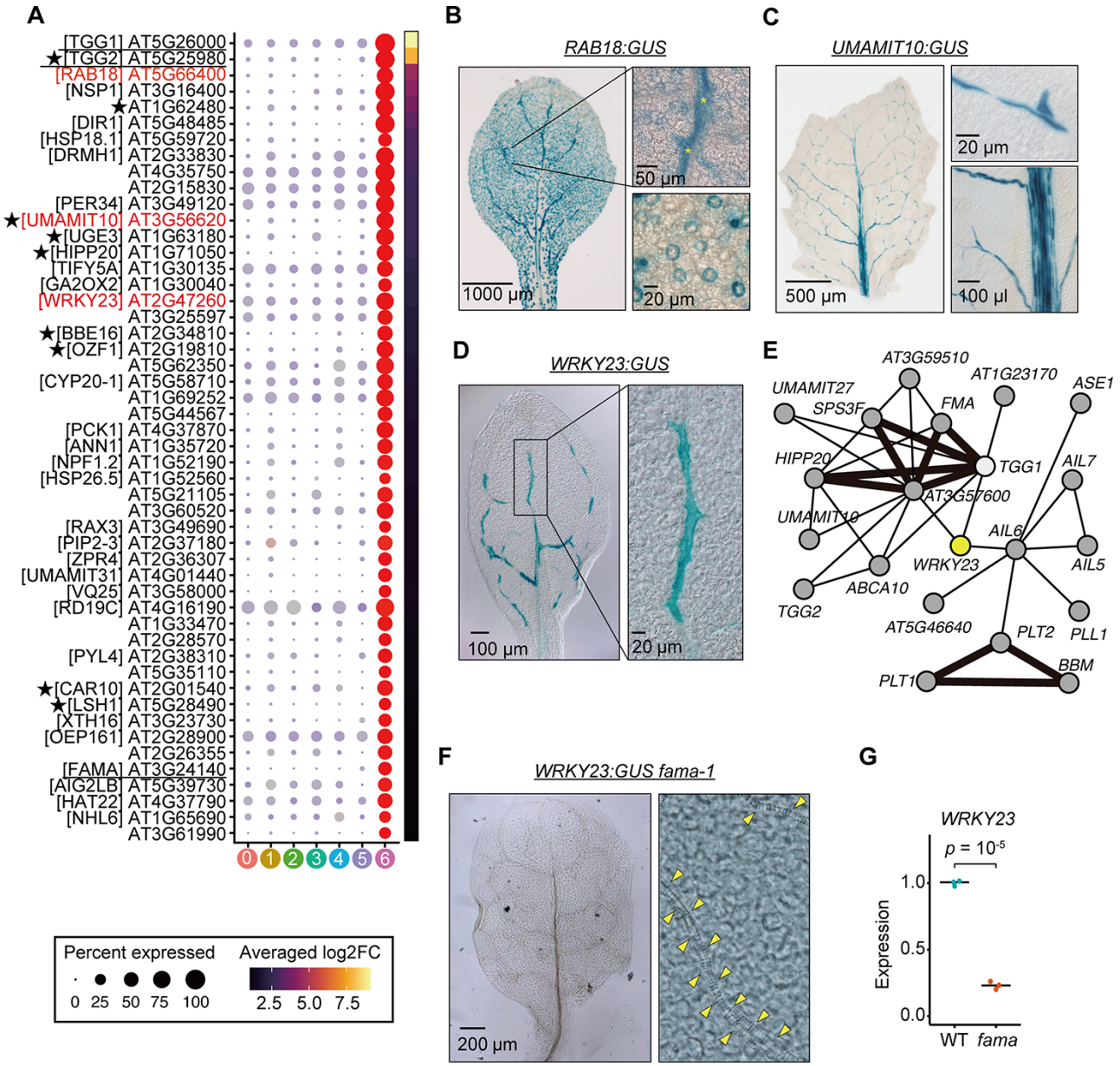

Fig. 4 Significantly upregulated genes in the myrosin cell cluster. (A) Dot plot profiling of the cluster-6-specific upregulated DEGs. The top 50 genes of the detected DEGs are plotted in order of the log fold changed (log2FC) value. Dot size and color mean the same as those in **Fig. 2E**. The heat map on the right side indicates the average log2FC in cluster-6 cells. The star indicates that Berrío et al. (2021) selected the genes as MI markers. The genes used for the GUS staining was colored red. (B–D and F) Histochemical assay by GUS activity and tissue-specific localization. Transgenic plants harboring each construct were grown (B–D) under the WT background or (F) *FAMA* knock-out strain. Yellow triangles denote the leaf veins. (E) Co-expressed gene network of *TGG1* via ATTED-II. Edges were drawn based on the rank of correlation (bold lines = 1–5, normal lines = <5). (G) The expression level of *WRKY23* in Arabidopsis leaves. *fama; FAMA* knock-out strain.

Enriched gene ontology (GO) analysis of DEGs detected the enrichment of two GO terms, GO:1901700:response to the oxygen-containing compound and GO:0005783:endoplasmic reticulum (adjusted *P*-value < 0.01). We found 32 previously obtained upregulated genes in MI-overproduction mutant, *syp22* (Shirakawa et al. 2014a) from our cluster-6-specific DEGs (**Supplementary data S4**).

Glucosinolate synthesis–related cell types

Our clustering identified cluster-1 where leaf cells expressed GL biosynthesis–related (GBR) genes. Reanalysis of cluster-1 narrowed down major GL-producing (MGP) subclusters to α1, α4 and α5. To clarify whether MGP subclusters can complete GL synthesis, we examined the expression of previously known GBR genes (Sønderby et al. 2010, Nour-Eldin et al. 2012).

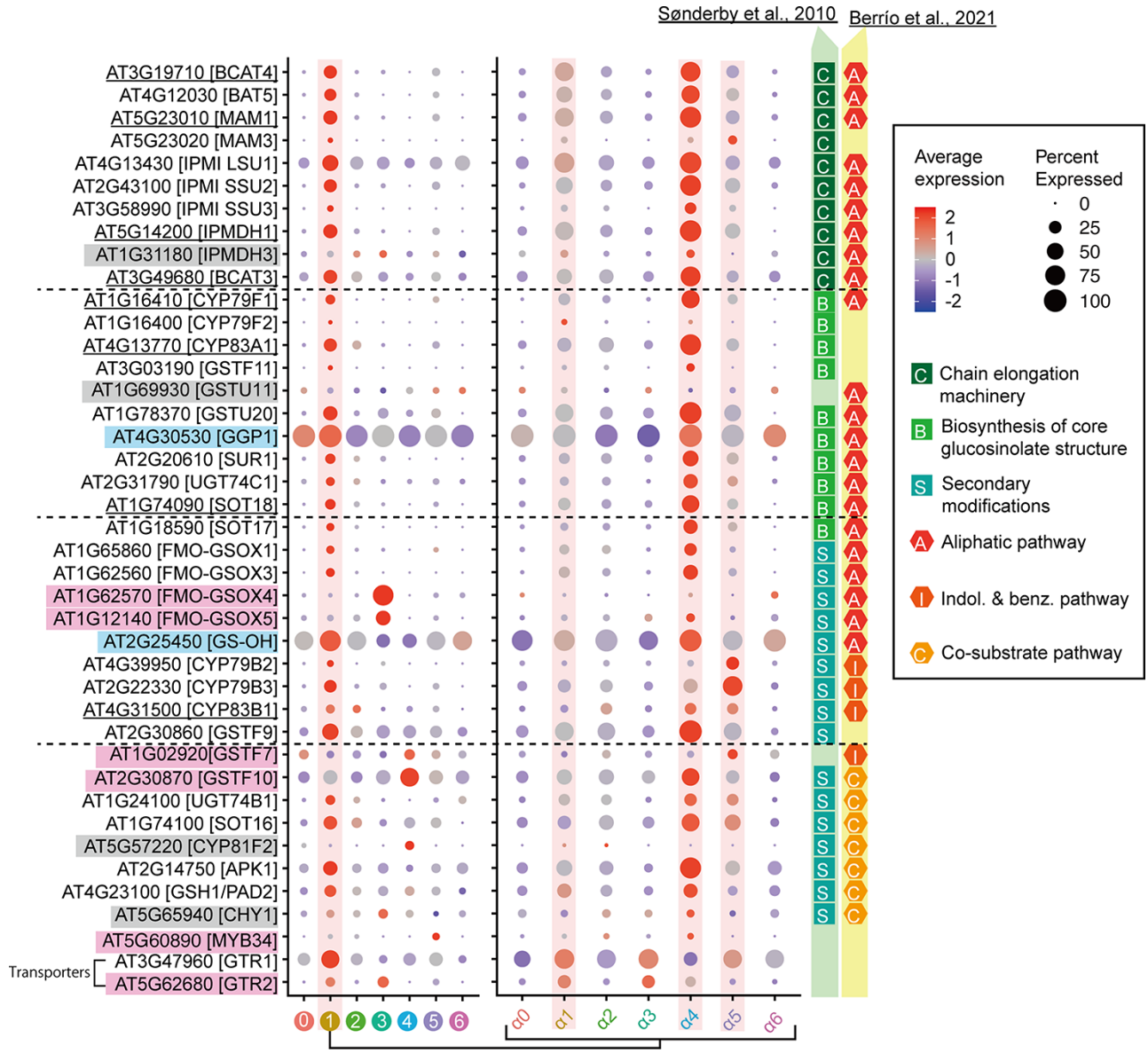

Fig. 5 Expression of the previously known GBR genes in each cell cluster. Dot plot profiling of the genes for GL biosynthesis pathways. The corresponding step in the GL synthesis pathway by Sønderby et al. (2010) is shown on the far left. Dot size and color mean the same as those in **Fig. 2E**. The results of the first step of graph-based clustering are shown on the left, and the results of the second step of clustering are shown in the box on the right. Magenta boxes show genes strongly expressed in a cluster other than cluster-1 in the first step clustering. Cyan boxes represent ubiquitously expressed genes among the clusters. Gray boxes show that genes lack a clear expression pattern. The black line indicates the genes used for the cluster annotation.

We selected 45 genes as GBR based on the previous studies (Sønderby et al. 2010, Berrío et al. 2021) (**Supplementary data S6–8**). Our scRNA-Seq data contained 38 GBR genes. Two genes were not contained in the Col-0 genome (*AOP2* and *APK2*), and five genes did not show expression beyond the threshold (*FMO-GSOX2*, *AOP3*, *CYP79A2*, *AAO4* and *BZO1*) (**Supplementary data S8**). The 38 GBR genes included seven of our annotation marker genes (**Fig. 5**).

In our first clustering, we detected 28 GBR genes in cluster-1. This result suggests that many steps of GL biosynthesis are covered by cluster-1-belonging cells (**Fig. 5**). The other five GBR genes (*FMO-GSOX4*, *FMO-GSOX5*, *GSTF7*, *GSTF10* and *MYB34*) are located in the non-cluster-1 group. *FMO-GSOX4* and *FMO-GSOX5* were commonly expressed in cluster-3 (CC). *GSTF7* and *GSTF10* expression was increased in cluster-4 (epidermis). *MYB38* was detected in cluster-5. This result suggests that several non-cluster-1 cells (CC and epidermis) can synthesize a number of derivative GL products (e.g. *S*-oxygenated aliphatic glucosinolates). *GGP1* and *GS-OH* were ubiquitously detected but especially enriched in cluster-1. The expression of the four

remaining GBR genes (*IPMDH3*, *GSTU11*, *CYP81F2* and *CHY1*) was weak in cluster-1 and lacked a clear expression pattern (**Fig. 5**).

In our second clustering step, the 28 cluster-1-enriched GBR genes were expressed in our MGP subclusters (subcluster-α1, subcluster-α4 or subcluster-α5). These results support that subcluster-α1, subcluster-α5 or subcluster-α6 are the major cell types involved in GL biosynthesis in cluster-1. The expression of 10 non-cluster-1-enriched GBR genes (*IPMDH3*, *GSTU11*, *GGP1*, *FMO-GSOX4*, *FMO-GSOX5*, *GS-OH*, *GSTF7*, *GSTF10*, *CYP81F2* and *CHY1*) was not restricted to the three MGP subclusters, supporting that these genes are differently regulated than the 28 cluster-1-specific GBR genes.

The three MGP clusters differed in GBR gene expression. The average expression of genes for GL chain elongation machinery was higher in subcluster-α4 than that in subcluster-α1 or subcluster-α5 (**Fig. 5**). The genes for GL secondary modifications via the indole pathway, *CYP79B2*, *CYP79B3* and *GSTF7*, tended to indicate high expression in subcluster-α5. These results suggest that the three types play slightly different physiological roles in GL metabolism.

Furthermore, we focused on two GL transporter genes, *GTR1* and *GTR2* (**Fig. 5**). We found *GTR1* and *GTR2* from cluster-1 and cluster-3, respectively. In our reclustering of cluster-1, subcluster-α1 and subcluster-α3 indicated high expression of *GTR1* and *GTR2*.

Upregulated genes in MGP subclusters and *in situ* visualization

We identified a set of upregulated DEGs in each cluster to find markers for the three MGP subclusters (α1, α4 and α5). We then confirmed their specificity via the GUS promoter assay. Our scRNA-Seq data give 536 DEGs from three MGP subclusters, and the GUS promoter assay found four suitable markers for the detection of MGP subclusters.

The 'FindAllMarker' function of the Seurat package found a total of 1,289 DEGs from the seven subclusters (adjusted *P*-value < 0.01, **Fig. 6A** and **Supplementary data S9**). DEG intersections identified 536 MGP-subcluster-specific upregulated DEGs (**Fig. 6A**). The three MGPs shared two DEGs. Subcluster-α4 and subcluster-α5 shared six DEGs, subcluster-α1 and subcluster-α5 shared 11 DEGs and subcluster-α1 and subcluster-α4 shared 28 DEGs (**Fig. 6**). Subcluster-α1 contained the most abundant subcluster-specific DEG, 157 (**Fig. 6**). Subcluster-α4 and subcluster-α5 contained 97 and 50 specific DEGs, respectively (**Fig. 6** and **Supplementary data S10** and **S11**). Among the 536 DEGs, 26 were GBR genes (**Fig. 6A** and **Supplementary data S10** and **S11**).

To confirm marker specificity, we visualized the expression of six MGP-specific DEGs, namely, *DEG26*, *AT1G21440*, *GAPC2*, *AT3G57050*, *SULTR2;1* and *AT5G44720*, through a GUS reporter assay (**Fig. 7**). As a result, GUS signals were detected specifically in the leaf veins in *ProSULTR2;1:GUS*, *ProAT3G57050:GUS*, *ProAT5G44720:GUS* and *ProAT1G21440:GUS* (**Fig. 7A**). In *ProDEG26:GUS* and *ProGAPC2:GUS*, ubiquitous signals were found in the leaf tissues including the veins (**Fig. 7A**). Consistent with our GUS analyses, the latter two genes showed weak expression in non-MGP subclusters in our scRNA-Seq (**Fig. 7B**). Our GUS staining might detect those weak expressions as a signal. Under our experimental conditions, those broadly expressed DEGs seemed unsuitable as markers for MGP subclusters, despite the upregulated normalized expression levels in the MGP subclusters (**Fig. 7C**). Using the four exclusively expressed genes in MGP subclusters, we obtained suitable markers for the clusters and confirmed that the corresponding cells were located in the leaf vein.

GO-based differences among the three subclusters

Our GO enrichment analysis suggested physiological differences among MGP subclusters. Focusing on the differences among the three MGP subclusters, we summarized DEG intersections among the three subclusters without considering the non-MGP subcluster DEG calling (**Supplementary Fig. S1**). We found 435 DEGs in total, and these DEGs showed that the three MGP subclusters indicated the enrichment of 165 GOs in total.

As a result of the DEG calling, we found 180 (α1-DEGs), 150 (α4-DEGs) and 22 (α5-DEGs) subcluster-specific DEGs, respectively. The three MGPs shared two DEGs (α1α5α6-DEGs). Subcluster-α1 and subcluster-α4 shared 29 DEGs (α1α4-DEGs), subcluster-α1 and subcluster-α6 shared 14 DEGs (α1α5-DEGs), and subcluster-α4 and subcluster-α5 shared 38 DEGs (α4α5-DEGs) (**Supplementary Fig. S1**).

Our GO enrichment analysis showed the enrichment of 43 GOs from α1-DEGs, 87 GOs from α4-DEGs, two GOs from α5-DEGs and 33 GOs from α1α4-DEGs. From other DEGs, we found no GO enrichment. Among α4-DEGs and α1α4-DEGs, enriched GOs contained multiple GL-related GO terms (e.g. 'glycosyl compound metabolic process' from α4-DEG and α1α4-DEGs) (**Supplementary Fig. S1**). These results support the contribution of the MGP subclusters to GL synthesis.

Enriched GO terms differed slightly among the subcluster-specific DEGs (α1-, α4- and α5-DEGs). In α4-DEGs, the enrichment was detected for photosynthesis-related GOs (e.g. 'photosystem'). For α1-DEGs, the enrichment was detected in various terms related to cell wall and ribosomes (e.g. 'cell wall' and 'small ribosomal subunit'). Although α4- and α1α4-DEGs shared 20 GO terms (**Supplementary Fig. S1**), these data suggest that subcluster-α4 shows higher photosynthetic activity and that subcluster-α1 indicates protein synthesis. In α5-DEGs, we found the GO enrichment for 'response to water deprivation' and 'response to inorganic substance'.

Discussion

We here provided a single-cell transcriptome analysis from 3- to 4-week-old Arabidopsis leaves. The small representation of cells expressing known epidermal makers indicated that our epidermal peeling before protoplast isolation successfully enriched for cells inside leaves. Of the detected 3,264 cells, epidermal cells accounted for 1.3% (56 cells) of the total (**Fig. 2B**).

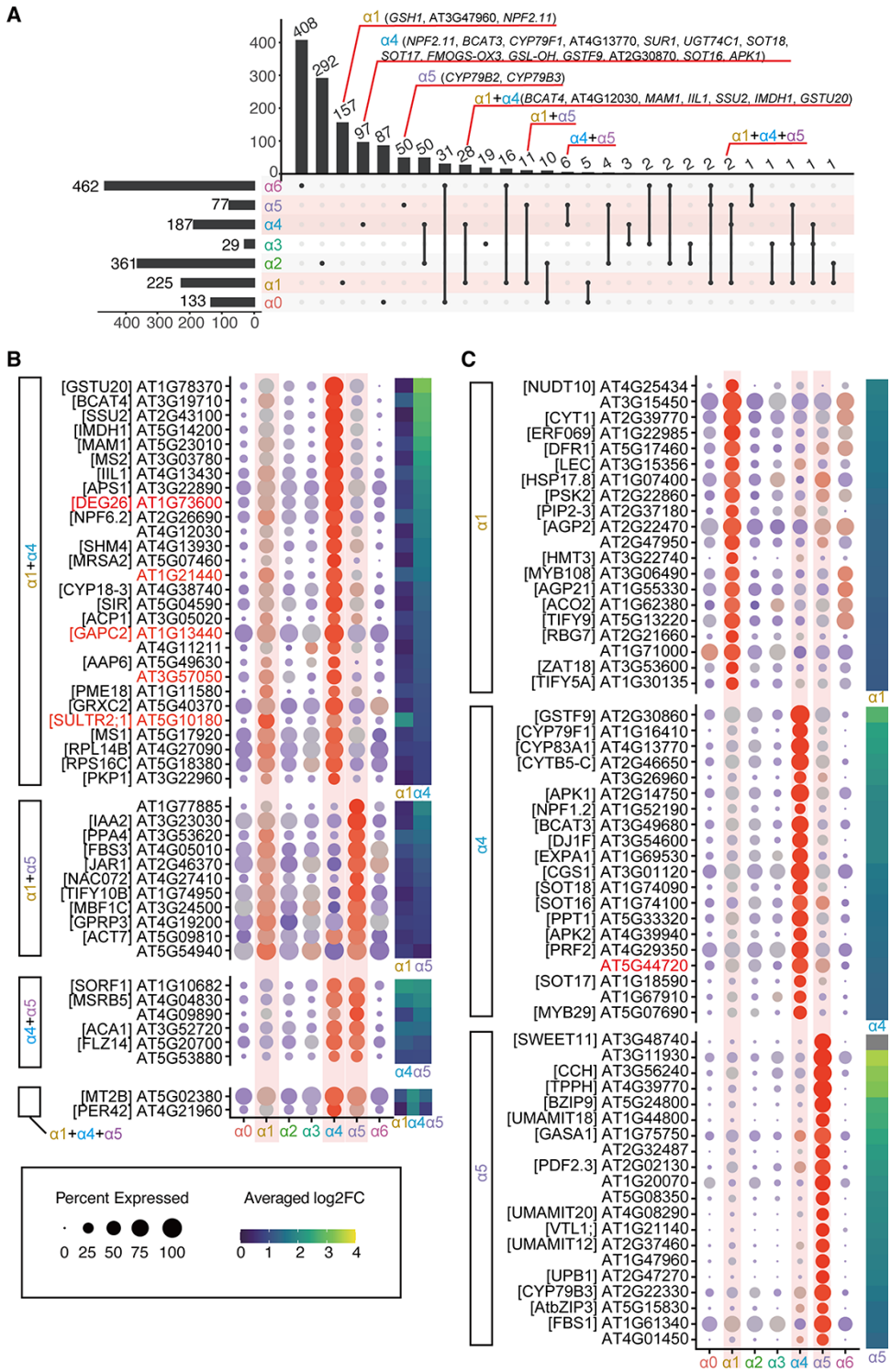

Fig. 6 Significantly upregulated genes in the MGP subclusters. (A) An UpSet plot of DEGs in each cell cluster. Adjusted *P*-value < 0.01, log2FC > 0.25. Colorations and cluster IDs correspond with those in **Figs. 2** and **3**. The vertical bar chart indicates the number of DEG conserved among the subclusters. Intersect connectors indicate the subcluster composition in a given number of genes (vertical bar chart). Connections corresponding to no gene were omitted. The gene names in the figure indicate the GL-related genes detected as DEGs in each connection. (B and C) Expression of obtained DEGs across the subclusters. Dot size and colors mean the same as those in **Fig. 2E**. The heat map on the right side indicates the averaged log2-fold change in each subcluster. (B) Expression of upregulated DEGs commonly detected in $\alpha1 + \alpha4$ and $\alpha1 + \alpha4 + \alpha5$ subclusters. (C) Significantly expressed genes in $\alpha1$, $\alpha4$ and $\alpha5$ subclusters, respectively. The top 20 DEGs were selected based on the averaged log2FC value.

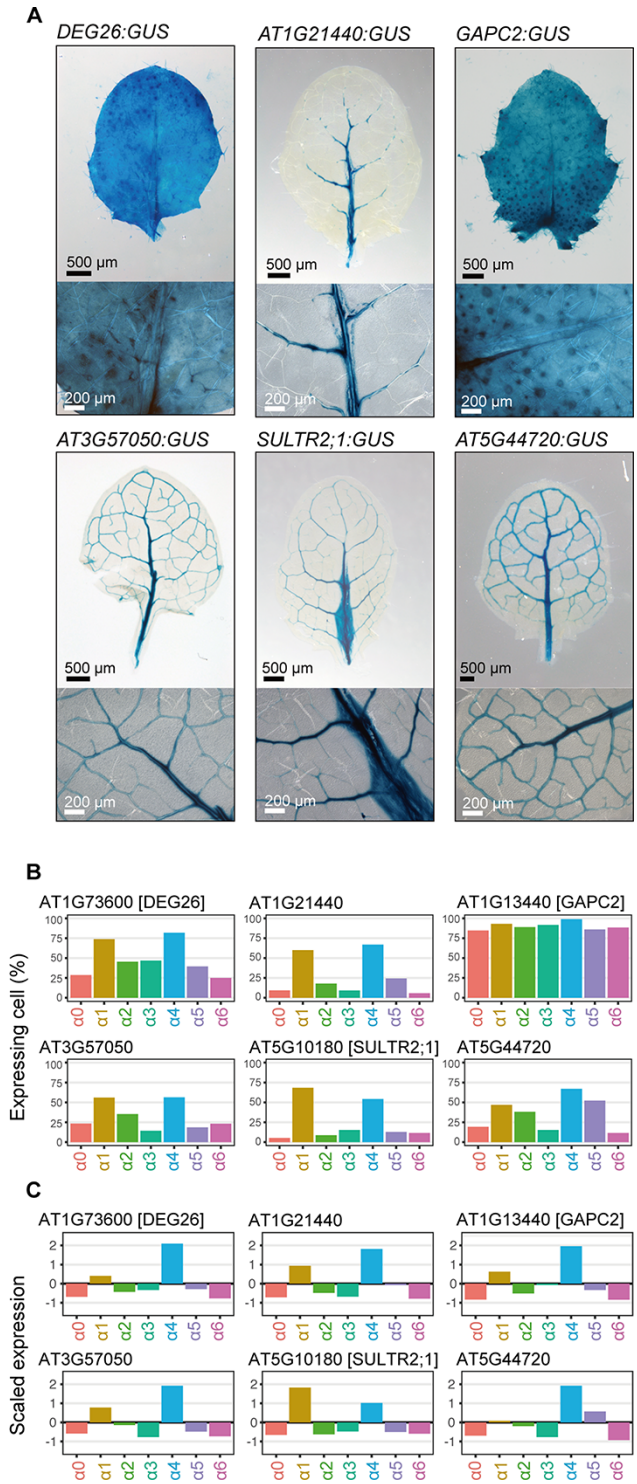

Fig. 7 Histochemical assay and scRNA-Seq expression profiles of the representative DEGs in MGP subclusters. (A) Histochemical assay by GUS activity and tissue-specific localization. Transgenic plants harboring each construct were grown under the WT background. (B and C) Expression profiles of analyzed genes in cluster-1-subclusters. (B) Percentage of the expressing cell in each subcluster. (C) Averaged normalized expression degree in each subcluster.

We successfully divided the cluster corresponding to MI via graph-based clustering. Expression of *KAT1*, a marker of GCs, indicated that our MI cluster contains no GC. Although the modified 'tape sandwich' by Kim et al. (2021) detected no MI but GC$^{(K)}$, we could not revisit their annotation because they did not mention *KAT* expression. Our scRNA-Seq, however, showed that standard peeling achieves both epidermal cell removal and comprehensiveness of inner cell types.

We detected 164 specific upregulated DEGs for MI, including 20 of the 78 markers found by Berrío et al. (2021). This difference in detected MI markers may be because of the early developmental stages of leaves and the absence of the epidermis peeling step in Berrío et al. (2021), although the feature studies for comprehensive histological analysis using different developmental stages will be required to clarify the background of these differences. Our GUS staining of the representative three MI markers supported the high expression of our MI markers in MI-marked leaf veins (**Fig. 2B–D**). The specificity, however, differed among markers. *RAB18*, excluded from the markers in the study by Berrío et al., identified not only MI but also GC. *UMAMIT10*, one of the 20 shared markers, had a specific expression in the leaf vein (=MI). The 20 conserved makers would be more conservative MI markers because they are commonly detected between different developmental stages, whereas the 78 markers from the study by Berrío et al. exclude the DEGs determined in epidermal cells.

One of our newly detected markers, *WRKY23*, showed a highly specific expression in MI via GUS staining (**Fig. 4D**), although this gene was undetected by Berrío et al. (2021). Previously, we identified *WRKY23* in the gene set that was upregulated in *FAMA* overexpression cell lines and an MI-overproduction mutant, *syp22* (Shirakawa et al. 2014a). In contrast to the expression in the *FAMA* overexpression cell lines, the *fama* mutant showed a significant reduction in *WRKY23* expression levels (**Fig. 4F, G**), suggesting that *FAMA*, the master MI regulator, is strongly required for the expression of *WRKY23*. Although *FAMA* also regulates GC development (Ohashi-Ito and Bergmann 2006), the expression of *WRKY23* was undetectable in GC, suggesting that *WRKY23* is a potential MI-specific regulator; unknown MI-lineage specific factor(s) may permit the expression of *WRKY23* in MI, or unknown stomatal lineage specific factor(s) may inhibit the expression of *WRKY23* in GC. Collectively, *WRKY23* is a promising target to reveal MI differentiation. These results suggest that combining single-cell, reporter and mutant analyses can create a more comprehensive understanding of the differentiation of rare cell types.

We found no clear localization of previously reported DEGs in the potential 'S-cell' clusters identified by Kim et al. (2021). We selected the top five S-cell DEGs based on the false discovery rate they used, and our data contained the expression of four DEGs (*PCR1*, *BG2*, *MTE17.16* and *LURP1*). In our dataset, *PCR1* was detected mostly in cluster-0, cluster-2 and cluster-5. *MTE17.16* was found in cluster-1 and cluster-2. *BG2* and *LURP1*

were detected in cluster-1 and cluster-2, respectively (**Fig. 2E**). Therefore, we could not discuss which clusters correspond to the S-cells identified by Kim et al. (2021). To reveal the correspondence between our scRNA-Seq data and S-cells in the study by Kim et al., future experimental results (e.g. targeted gene expression analysis) are needed.

Only cluster-1 showed the expression of GBR genes for the early stages of GL synthesis (e.g. *BCAT4*). Although other clusters (e.g. cluster-3 and cluster-4) expressed several genes for derivative GL synthesis (e.g. *GSTF10*), our results suggest that these non-cluster-1 cell types had no direct contribution to *de novo* GL synthesis. Transport of cluster-1-derived intermediates may be involved in derivative GL synthesis in non-cluster-1 cells (**Fig. 5**).

Despite the difficulty in comparing between studies because of differences in definition and cell-type range, our results for the GBR gene expression are mainly consistent with previous studies. The expression pattern of GBR genes indicated that MGP cells in the cluster-1 could be distinguished into three types, α1, α4 and α5. The expression patterns of marker genes indicate that these subclusters correspond to $TE^{(B)} + XPs$ ($XP1^{(B)}$, $XP2^{(B)}$, $XP1^{(K)}$, $XP2^{(K)}$ and $XP3^{(K)}$), $PP2^{(B)}$, and other PPs ($PP1^{(B)}$, $PP1^{(K)}$ and $PP2^{(K)}$), respectively. α1 and α4 contained minor cells corresponding to PCs ($PC^{XP(K)}$ and $PC^{PP(K)}$) and $BS^{(K)(B)}$, respectively. Kim et al. (2021) detected GBR genes in their cluster-C10 ($PP1^{(K)} + PC^{(K)}$), cluster-C18 ($PP2^{(K)} + XP3^{(K)}$) and cluster-C4 ($BS^{(K)} + XP^{(K)}$). Berrío et al. (2021) found GBR gene expression in $PP1^{(B)}$, $PP2^{(B)}$ and $BS^{(B)}$. Our result indicated that MGP cells correspond to BS, PP, XP and some PCs. These transcript-informed predictions concerning MGP cells will give a promising target for future microscale studies of GL distribution via X-ray or mass spectrometry imaging. UMAP plotting (**Fig. 3A**) indicated that cluster division was not as clear among the subclusters as among cluster-3, cluster-4 and cluster-5. This result suggests that the transcriptional profile continuously changes among these cells, as indicated in previous scRNA-Seq studies (Berrío et al. 2021, Kim et al. 2021, Lopez-Anido et al. 2021).

With respect to PP, our result supported the existence of $PP2^{(B)}$ (subcluster-α4) and its contribution to GL synthesis. Kim et al. (2021) did not identify the corresponding group with $PP2^{(B)}$ and instead used *Sweet11* and *Sweet12* expressions as the marker genes for PPs ($PP1^{(K)}$ and $PP2^{(K)}$). Berrío et al. (2021) detected a single cluster expressing *Sweet11* and *Sweet12* as $PP1^{(B)}$ and defined the $PP2^{(B)}$, which exhibits no *Sweet11* or *Sweet12* expression, based on *CYP83A1* expression. In our result, the *Sweet11* or *Sweet12* expression was limited to a single cluster (α5). We thus considered that our α5 will correspond with $PP1^{(B)}$ and will be a mixture of $PP1^{(K)}$ and $PP2^{(K)}$. Although Kim et al. (2021) reported physiological differences between $PP1^{(K)}$ and $PP2^{(k)}$, we could not analyze them. Our subcluster-α4, corresponding to $PP2^{(B)}$, indicated high expression of the GBR genes. We cannot explain why this group was undetected by Kim et al. *ProAT5G44720:GUS* suggested that α4 is located in the leaf vein as other MGS cells (**Fig. 7A**). The distinct characteristics of α4 (e.g. cellular-scale localization in leaf vain) and its role in the GMDS require further research.

We demonstrated differences in *GTR* expression among cell clusters. *GTR1* and *GTR2* are known to transport apoplastic GLs into the cell. In our scRNA-Seq, there was high expression of *GTR2* in cluster-1 and cluster-3. Despite the weak *GTR2* expression from cluster-1, cluster-1 subclustering indicated that *GTR2* expression is concentrated in subcluster-α1 and subcluster-α3. The enrichment in cluster-3 (CC) matches with *GTR2* function, the efflux of leaf-producing GLs into the seed. Nour-Eldin et al. (2012) reported that GTR2 concentrates on 'phloem-associated cells'. GTR2-enriched cells might be CC and part of XP. CC cells may intake apoplastic GLs produced by other vascular cells and transport them to the phloem. With respect to the *GTR1*, we observed the gene expression signal from cluster-1 and several of its subclusters (e.g. α1, α3 and α5). Nour-Eldin et al. (2012) showed that *GTR1* and *GTR2* have a redundant function in the long-distance transportation of GLs. However, they indicated that *GTR1* is expressed in both vasculature-associated and vasculature-adjacent mesophilic cells in 5- to 6-week-old leaves. Moreover, based on the weak effect of a *GTR1* knockdown on long-distance transport, they considered that *GTR2* plays a major role in GL phloem loading and that *GTR1* may primarily be involved in GL distribution within the leaf. Our result showed that *GTR1* is not expressed in CC but in a portion of MGS cells (α1 and α5) and unknown leaf tissues (α3), supporting Nour–Eldin's hypothesis. Their *GTR1*-expressing vasculature-associated cells may be XP (α1) + PPs (α5), and vasculature-adjacent mesophilic cells may correspond with α3. Subcluster-α1 and subcluster-α5 seem to produce GLs on their own and incorporate those produced by other cells via apoplasts.

Three MGP subclusters indicated different DEGs among each other and suggested functional differences among them. The three MGPs only shared two DEGs. The DEGs, which were found only in a particular MGP, contained 50–157 DEGs (**Supplementary data S10**). The expression of the already known genes for indole and benzoyl glucosinolate biosynthesis (*CYP79B2*, *CYP79B3*, *CYP83B1* and *GSTF7*) suggested that α5 is a major source of the indole and benzoyl glucosinolates. This result is consistent with the study by Berrío et al. (2021), which detected gene expression for indole and benzoyl glucosinolates from α5-corresponding $PP1^{(B)}$. Our α5-specific 97 DEGs may give a feature target to reveal the benzoyl glucosinolate biosynthesis. **Fig. 5** suggests that α4 is a major source of aliphatic glucosinolates. Our α4-specific DEG contained several previously reported GMDS-related genes (*CYTB5-C* and *AT5G44720*). *CYTB5-C* is a gene that influences the GL profile of *A. thaliana*, and *AT5G44720* has been reported as a coexpression gene of *CB5-C* (Vik et al. 2016). We did not find any GMDS-related genes in the α1-specific 157 DEGs. However, the shared DEGs between α1 and α4 contained multiple candidates of GMDS-related genes (*NPF6.2*, *APS1*, *AT1G21440*, *AT1G73600* and *MS2*). *NPF6.2* encodes NRT/PTR transporters, such as *GTR1* and *GTR2*. Although no transporter has been reported to release GLs from

the cell to the apoplast, the above-mentioned *NPF6.2* shared a DEG between α1 and α4 and is a promising GL exporter candidate. The *APS1*-encoded protein produces adenosine 5′-phosphosulfate, a source for GL biosynthesis. *AT1G21440* and *AT1G73600* are potential GL-related genes based on coexpression analysis (Bekaert et al. 2012, Harun et al. 2021). MS2 synthesizes methionine (Met), a source for the biosynthesis of Met-derived GLs (Schuster et al. 2006). With respect to other DEGs, we found no reported relationships with GMDS. These results may provide a basis for understanding the physiological differences and similarities among these MGPs. The mutant phenotype of newly identified markers and microscale distribution of GLs are important questions that remain to be explored for the GMDS study.

Materials and Methods

Plant material and growth conditions

Arabidopsis thaliana Columbia (Col-0) ecotype was used for all lines except for *ProWRKY23:GUS* (C24). The *fama-1* mutant (SALK_100073) (Shirakawa et al. 2014a) was obtained from the ABRC at Ohio State University. *ProWRKY23:GUS* (Grunewald et al. 2012) was provided by W. Grunewald and T. Beekman (VIB). *ProWRKY23:GUS fama-1* was generated by crossing. *Arabidopsis* seeds were grown on 0.5% gellan gum with Murashige Skoog. Plates were cultivated under constant light conditions. For plant preparation for protoplast isolation, Col-0 seeds were sown in the soil (3:1 mix of SuperMixA and vermiculite) and cultured for 3–4 weeks under 16/8 h day/night cycle at 22°C.

Protoplast isolation

Protoplast isolation from long-day-grown Arabidopsis rosette leaves was performed as previously described (Yoshida et al. 2013). Strips of leaves between 0.5 and 1.0 mm in width, whose epidermis was peeled using a tape, were incubated in an enzyme solution containing 1% (w/v) cellulase 'onozuka' R10 (Yakult Pharmaceutical Inc., Japan), 0.25% (w/v) macerozyme 'onozuka' R10 (Yakult), 0.4 M mannitol, 20 mM MES, 20 mM KCl, 10 mM $CaCl_2$ and 5 mM 2-mercaptoethanol for 1 h with shaking at 50 rpm, 22°C. The solution was filtered through a 75 μm nylon mesh (BD Biosciences Inc., USA) and collected by centrifugation at 100 × *g* for 10 min. The pellet was washed with a W5 buffer containing 150 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl and 2 mM MES, followed by centrifugation at 100 × *g* for 5 min. Washing was repeated twice. After removing the supernatant, the protoplast cells were resuspended in PBS and used for 10× library preparation (**Fig. 2A**).

Generation of single-cell transcriptomes

A commercially available droplet-based system from 10× Genomics Inc. (Zheng et al. 2017) was used to isolate protoplasts according to the manufacturer's instructions (Chromium Single-Cell 3′ Reagent Kits v3 CG000184 Rev C). The protoplast suspension was loaded into Chromium Chip B with v3 chemistry and barcoded with a Chromium Controller (10× Genomics). RNA from barcoded cells was subsequently reverse-transcribed, and sequencing libraries were constructed with reagents from a Chromium Single-Cell kit (10× Genomics) according to the manufacturer's instructions. Sequencing was performed with BGISEQ-500 according to the manufacturer's instructions (BGI).

Raw reads were demultiplexed and mapped to the TAIR10 reference genome by the 10× Genomics Cell Ranger pipeline (v6.0) using default parameters. Downstream single-cell analyses were performed using Seurat unless specifically mentioned (v4.0.1) (Stuart et al. 2019). Briefly, for each gene and each cell barcode (filtered by Cell Ranger), UMIs were counted to construct digital expression matrices. A gene expressed in >3 cells was considered as expressed, and each cell was required to have 200 expressed genes. Because cell multiplets may exhibit an aberrantly high gene count, we removed cells with >3,000 expressed genes. Cells containing a high rate of plastid-DNA-derived genes (>20%) were removed because droplets containing collapsed cells or free plastid may exhibit a high rate of plastid-DNA-derived genes. Even after these treatments, our dataset contained many cells containing a small number of UMIs and genes. These low-quality data may confuse the cell clustering and annotation, and we therefore removed the cells containing low UMI count (<1,000 UMI) or low gene number (<1,000 genes) from the analysis.

Cluster visualization and annotation

A set of the top 2,000 highly variable genes from each dataset expressed in a sample were used by Seurat. After selecting highly variable 2,000 genes, we used principal component analysis and selected the top 13 principal components for a k-nearest neighbor graph construction and divided the cells into transcriptionally distinct clusters via a shared nearest neighbor modularity optimization-based clustering algorithm. Clusters were identified using Seurat FindClusters function (resolution = 0.1, **Fig. 2B**). We used the UMAP algorithm to reveal local similarities and global structures. Several traditional marker genes were used to facilitate cell-type annotation (**Fig. 2C**). For the subset analysis of GBR cells, we used a set of the top 3,000 highly variable genes from 'cluster-1' and the top 19 aligned correlated components (FindClusters resolution = 0.5).

Plasmid construction and transgenic plants

The Gateway Cloning System (Thermo Fisher Scientific) was used for plasmid construction. For transcriptional GUS fusion constructs, the 2-kb promoter of *RAB18* (AT5G66400) was cloned into pENTR D-TOPO. The promoter sequence was introduced into the binary vector pBGWFS7 (BASTA selection for plants) using the LR reaction to generate pBGWFS7 ProRAB18. The Hybrid-gateway system was used to generate all other constructs. Promoters were cloned into 5′-TOPO and used for LR reactions with R4pGWB601 and pENTR D-TOPO GUS-GFP. *Agrobacterium tumefaciens* (strain GV3101) was transformed with binary vectors; plants were transformed with agrobacteria using the floral dip method (Clough and Bent 1998). T1 seeds were selected using a medium containing 10 mg L^{-1} BASTA (Sigma–Aldrich).

GUS staining

Samples were fixed in 90% acetone for 15 min on ice and then stained with GUS staining solution. The staining method was described previously (Shirakawa et al. 2014a). Representative images were acquired with an AX-70 light microscope (OLYMPUS) and an AXIO Zoom V16 (ZEISS) microscope.

Reverse transcription PCR and quantitative RT-PCR

Samples were immediately frozen in liquid nitrogen. The RNeasy Plant Mini Kit (QIAGEN) was used to extract total RNA. The RNase-Free DNase Set (QIAGEN) was used to eliminate genomic DNA contamination in RNA samples. Reverse transcription PCR was performed using PrimeScript™ RT Master Mix (TaKaRa). Quantitative RT-PCR was performed as described previously (Shirakawa et al. 2021). Arabidopsis *ACTIN2* (At3g18780) was used as internal reference. Each experiment was repeated ≥3 times. The relative expression level of each gene was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). The following primer sets were used: *WRKY23*, forward: 5′-AGTCTCGGTAATGGTTGCTTTGG-3′ and reverse: 5′-TGTTGCTGCTGTTGGTGATGG-3′; *ACTIN2*, forward: 5′-GGCGATGAAGCT CAATCCAAACG-3′ and reverse: 5′-GGTCACGACCAGCAAGATCAAGACG-3′.

Supplementary Data

Supplementary data are available at *PCP* online.

Data Availability

The data underlying this article will be shared upon reasonable request to the corresponding author. All of the raw sequence data obtained in this research have been deposited in the DDBJ Sequence Read Archive under DRA014584. The data and the program codes have been made available on FigShare, at DOI: 10.6084/m9.figshare.20375205.

Funding

Japan Society for the Promotion of Science KAKENHI Grant-in-Aid for Scientific Research on Innovative Areas (22H04723); Japan Science and Technology Agency 'Precursory Research for Embryonic Science and Technology' (JPMJPR22D3); the Takeda Science Foundation and the Kato Memorial Bioscience Foundation (M. Shirakawa).

Acknowledgements

We thank Wim Grunewald and Tom Beekman for sharing materials (*WRKY23:GUS*). We also thank Kyoko Sunuma and Mikiko Higashiura for technical assistance. Computational resources were provided by the Data Integration and Analysis Facility, National Institute for Basic Biology.

Author Contributions

A.J.N. and S.K. designed the experiments. S.S.S. and T.M. performed the protoplast and single-cell RNA-Seq experiments. M.S. performed the GUS staining, mutant assay and quantitative RT-PCR. T.M. analyzed the data. T.M., A.J.N., S.S.S. and M.S. wrote the manuscript.

Disclosures

The authors have no conflicts of interest to declare.

References

Andréasson, E., Bolt Jørgensen, L., Höglund, A.S., Rask, L. and Meijer, J. (2001) Different myrosinase and idioblast distribution in Arabidopsis and *Brassica napus*. *Plant Physiol.* 127: 1750–1763.

Barth, C. and Jander, G. (2006) Arabidopsis myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *Plant J.* 46: 549–562.

Bekaert, M., Edger, P.P., Hudson, C.M., Pires, J.C. and Conant, G.C. (2012) Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis. *New Phytol.* 196: 596–605.

Berrío, R.T., Verstaen, K., Vandamme, N., Pevernagie, J., Achon, I., Van Duyse, J., et al. (2021) Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells. *Plant Physiol.* 188: 898–918.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16: 735–743.

Grunewald, W., De Smet, I., Lewis, D.R., Löfke, C., Jansen, L., Goeminne, G., et al. (2012) Transcription factor WRKY23 assists auxin distribution patterns during Arabidopsis root development through local control on flavonol biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 109: 1554–1559.

Harun, S., Afiqah-Aleng, N., Karim, M.B., Altaf Ul Amin, M., Kanaya, S. and Mohamed-Hussein, Z.-A. (2021) Potential *Arabidopsis thaliana* glucosinolate genes identified from the co-expression modules using graph clustering approach. *Peer J.* 9: e11876.

Kim, J.-Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., et al. (2021) Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* 33: 511–530.

Kliebenstein, D.J., Kroymann, J. and Mitchell-Olds, T. (2005) The glucosinolate-myrosinase system in an ecological and evolutionary context. *Curr. Opin. Plant Biol.* 8: 264–271.

Koroleva, O.A., Davies, A., Deeken, R., Thorpe, M.R., Tomos, A.D. and Hedrich, R. (2000) Identification of a new glucosinolate-rich cell type in Arabidopsis flower stalk. *Plant Physiol.* 124: 599–608.

Koroleva, O.A., Gibson, T.M., Cramer, R. and Stain, C. (2010) Glucosinolate-accumulating S-cells in Arabidopsis leaves and flower stalks undergo programmed cell death at early stages of differentiation. *Plant J.* 64: 456–469.

Leegood, R.C. (2008) Roles of the bundle sheath cells in leaves of C3 plants. *J. Exp. Bot.* 59: 1663–1673.

Li, M. and Sack, F.D. (2014) Myrosin idioblast cell fate and development are regulated by the Arabidopsis transcription factor FAMA, the auxin pathway, and vesicular trafficking. *Plant Cell* 26: 4053–4066.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25: 402–408.

Lopez-Anido, C.B., Vatén, A., Smoot, N.K., Sharma, N., Guo, V., Gong, Y., et al. (2021) Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. *Dev. Cell* 56: 1043–1055.e4.

Mocniak, L.E., Elkin, K. and Bollinger, J.M., Jr (2020) Lifetimes of the aglycone substrates of specifier proteins, the autonomous iron enzymes that dictate the products of the glucosinolate-myrosinase defense system in Brassica plants. *Biochemistry* 59: 2432–2441.

Nakamura, R.L., McKendree, W.L., Jr, Hirsch, R.E., Sedbrook, J.C., Gaber, R.F. and Sussman, M.R. (1995) Expression of an Arabidopsis potassium channel gene in guard cells. *Plant Physiol.* 109: 371–374.

Nintemann, S.J., Hunziker, P., Andersen, T.G., Schulz, A., Burow, M. and Halkier, B.A. (2018) Localization of the glucosinolate biosynthetic enzymes reveals distinct spatial patterns for the biosynthesis of indole and aliphatic glucosinolates. *Physiol. Plant* 163: 138–154.

Nour-Eldin, H.H., Andersen, T.G., Burow, M., Madsen, S.R., Jørgensen, M.E., Olsen, C.E., et al. (2012) NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature* 488: 531–534.

Ohashi-Ito, K. and Bergmann, D.C. (2006) Arabidopsis FAMA controls the final proliferation/differentiation switch during stomatal development. *Plant Cell* 18: 2493–2505.

Oparka, K.J. and Turgeon, R. (1999) Sieve elements and companion cells-traffic control centers of the phloem. *Plant Cell* 11: 739–750.

Procko, C., Lee, T., Borsuk, A., Bargmann, B.O.R., Dabi, T., Nery, J.R., et al. (2022) Leaf cell-specific and single-cell transcriptional profiling reveals a role for the palisade layer in UV light protection. *Plant Cell* 34: 3261–3279.

Schuster, J., Knill, T., Reichelt, M., Gershenzon, J. and Binder, S. (2006) Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in Arabidopsis. *Plant Cell* 18: 2664–2679.

Shirakawa, M. and Hara-Nishimura, I. (2018) Specialized vacuoles of myrosin cells: chemical defense strategy in Brassicales plants. *Plant Cell Physiol.* 59: 1309–1316.

Shirakawa, M., Morisaki, Y., Gan, E.S., Sato, A. and Ito, T. (2021) Identification of a devernalization inducer by chemical screening approaches in *Arabidopsis thaliana*. *Front. Plant Sci.* 12: 634068.

Shirakawa, M., Tanida, M. and Ito, T. (2022) The cell differentiation of idioblast myrosin cells: similarities with vascular and guard cells. *Front. Plant Sci.* 12: 829541.

Shirakawa, M., Ueda, H., Nagano, A.J., Shimada, T., Kohchi, T. and Hara-Nishimura, I. (2014a) FAMA is an essential component for the differentiation of two distinct cell types, myrosin cells and guard cells, in Arabidopsis. *Plant Cell* 26: 4039–4052.

Shirakawa, M., Ueda, H., Shimada, T. and Hara-Nishimura, I. (2016a) Myrosin cells are differentiated directly from ground meristem cells and are developmentally independent of the vasculature in Arabidopsis leaves. *Plant Signal. Behav.* 11: e1150403.

Shirakawa, M., Ueda, H., Shimada, T. and Hara-Nishimura, I. (2016b) FAMA: a molecular link between stomata and myrosin cells. *Trends Plant Sci.* 21: 861–871.

Shirakawa, M., Ueda, H., Shimada, T., Kohchi, T. and Hara-Nishimura, I. (2014b) Myrosin cell development is regulated by endocytosis machinery and PIN1 polarity in leaf primordia of *Arabidopsis thaliana*. *Plant Cell* 26: 4448–4461.

Sønderby, I.E. Geu-Flores, F. and Halkier, B.A. (2010) Biosynthesis of glucosinolates: gene discovery and beyond. *Trends Plant Sci.* 15: 283–290.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M. 3rd, et al. (2019) Comprehensive integration of single-cell data. *Cell* 177: 1888–1902.e21.

Ueda, H., Nishiyama, C., Shimada, T., Koumoto, Y., Hayashi, Y., Kondo, M., et al. (2006) AtVAM3 is required for normal specification of idioblasts, myrosin cells. *Plant Cell Physiol.* 47: 164–175.

Vik, D., Crocoll, C., Andersen, T.G., Burow, M. and Halkier, B.A. (2016) CB5C affects the glucosinolate profile in *Arabidopsis thaliana*. *Plant Signal. Behav.* 11: e1160189.

Wu, R., Li, S., He, S., Wassmann, F., Yu, C., Qin, G., et al. (2011) CFL1, a WW domain protein, regulates cuticle development by modulating the function of HDG1, a class IV homeodomain transcription factor, in rice and Arabidopsis. *Plant Cell* 23: 3392–3411.

Yoshida, K., Sakamoto, S., Kawai, T., Kobayashi, Y., Sato, K., Ichinose, Y., et al. (2013) Engineering the *Oryza sativa* cell wall with rice NAC transcription factors regulating secondary wall formation. *Front. Plant Sci.* 4: 383.

Zhao, C., Pratelli, R., Yu, S., Shelley, B., Collakova, E. and Pilot, G. (2021) Detailed characterization of the UMAMIT proteins provides insight into their evolution, amino acid transport properties, and role in the plant. *J. Exp. Bot.* 72: 6400–6417.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8: 14049.