

Single-nucleus transcriptomes reveal spatiotemporal symbiotic perception and early response in *Medicago*

Received: 8 September 2022

Zhijian Liu^{1,6}, Jun Yang^{1,6}, Yanping Long^{1,6}, Chi Zhang^{1,2,3,6}, Dapeng Wang², Xiaowei Zhang², Wentao Dong^{1,6}, Li Zhao⁴, Chengwu Liu^{1,6}, Jixian Zhai¹✉ & Ertao Wang^{1,2,5}✉

Accepted: 25 August 2023

Published online: 25 September 2023

 Check for updates

Establishing legume–rhizobial symbiosis requires precise coordination of complex responses in a time- and cell type-specific manner. Encountering *Rhizobium*, rapid changes of gene expression levels in host plants occur in the first few hours, which prepare the plants to turn off defence and form a symbiotic relationship with the microbes. Here, we applied single-nucleus RNA sequencing to characterize the roots of *Medicago truncatula* at 30 min, 6 h and 24 h after nod factor treatment. We found drastic global gene expression reprogramming at 30 min in the epidermis and cortex and most of these changes were restored at 6 h. Moreover, plant defence response genes are activated at 30 min and subsequently suppressed at 6 h in non-meristem cells. Only in the cortical cells but not in other cell types, we found the flavonoid synthase genes required to recruit rhizobia are highly expressed 30 min after inoculation with nod factors. A gene module enriched for symbiotic nitrogen fixation genes showed that *MtFER* (*MtFERONIA*) and *LYK3* (*LysM domain receptor-like kinase 3*) share similar responses to symbiotic signals. We further found that MtFER can be phosphorylated by LYK3 and it participates in rhizobial symbiosis. Our results expand our understanding of dynamic spatiotemporal symbiotic responses at the single-cell level.

Legumes can symbiotically fix nitrogen with rhizobia^{1–5}, enabling substantially reduced use of fertilizer required for agricultural production⁶. The establishment of symbiotic nitrogen fixation (SNF) relies on mutual recognition of legumes and rhizobia⁷. Under nitrogen-limited conditions, legume roots release a class of flavonoids into the rhizosphere,

inducing rhizobia to secrete lipo-chitooligosaccharides called nod factors (NFs)⁸. Perception of NFs by the plasma membrane-localized receptor complex consisting of nod factor perception (NFP), LysM domain receptor-like kinase 3 (LYK3) in *Medicago truncatula*^{9,10}, nod factor receptor 1 (LjNFR1) and LjNFR5 in *Lotus japonicus*^{11–13} leads to a

¹Institute of Plant and Food Science, Department of Biology, School of Life Sciences, Southern University of Science and Technology (SUSTech), Shenzhen, China. ²New Cornerstone Science Laboratory, National Key Laboratory of Plant Molecular Genetics, Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China. ³State Key Laboratory for Managing Biotic and Chemical Treats to the Quality and Safety of Agro-Products, Key Laboratory of Biotechnology in Plant Protection of MOA of China and Zhejiang Province, Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou, China. ⁴School of Life Sciences, Division of Life Sciences and Medicine, MOE Key Laboratory for Membraneless Organelles and Cellular Dynamics, University of Science and Technology of China, Hefei, China. ⁵School of Life Science and Technology, ShanghaiTech University, Shanghai, China. ⁶These authors contributed equally: Zhijian Liu, Jun Yang, Yanping Long, Chi Zhang. ✉e-mail: zhaijx@sustech.edu.cn; etwang@cemps.ac.cn

series of symbiotic responses, including root hair curling, infection thread (IT) formation and cortical cell division¹⁴. NF-induced remodelling of the microtubule cytoskeleton is associated with root hair curling and the formation of ITs¹⁵.

In the past several years, several high-throughput sequencing-based studies on whole root or specific cell types, such as epidermal cells and root hairs, have revealed that cell cycle- and plant hormone-related genes are reactivated after inoculation with NFs^{16–21}. Specifically, one study²¹ found that genes involved in immune response rapidly increase at 1 h after *Rhizobium* inoculation and then decrease at 24 h after inoculation on the basis of their gene microarray data of whole root. Others found activation of cell-cycle genes and cytokinin-related pathways induced by nodulation factors¹⁹. Another study discovered changes in cell-cycle genes and auxin-related pathways during *Rhizobium* infection¹⁶. However, the response of legumes to NFs can be cell type-specific and the bulked measurements may have difficulties in detecting changes in expression level that occur only in certain cell types.

In recent years, rapid developments in single-cell sequencing technology have transformed our perspective of transcriptional regulation. However, the plant cell wall must be enzymatically digested to obtain free single cells, a process called protoplasting^{22,23}. Unfortunately, the expression levels of many genes, especially those involved in defence and stress responses, change during this treatment^{24,25}, therefore making protoplasting undesirable for studying genes that participate in early symbiotic interactions. Single-nucleus RNA sequencing (snRNA-seq) has offered a route to bypass the prerequisite of protoplasting for single-cell sequencing and shown great promise^{26–29}, including a full-length snRNA profiling method previously developed by us²⁹. Recently, others³⁰ profiled the 48 h rhizobial-inoculated root apex of *M. truncatula* using snRNA-seq and identified many new genes and pathways which are responsive to *Rhizobium* inoculation at each cell type. However, the response induced by NFs alone, especially at the extremely early stage from 0.5 h post treatment, has not been completely uncovered. In this study, we applied time-course snRNA-seq analysis of symbiotic perception in *M. truncatula* to examine early cell type-specific responses to nodulation.

Results

snRNA-seq reveals the cell heterogeneity in *Medicago* root

To analyse the transcriptome reprogramming of early response, we inoculated NFs on 7-day-old *M. truncatula* seedling. We then collected the susceptible zone (to NFs) of roots at 0.5 hours posttreatment (hpt), 6 hpt and 24 hpt, and roots inoculated in buffered nodulation medium (BNM) as mock control, to prepare snRNA-seq libraries (Fig. 1a). We obtained a total of 25,276 high-quality single-nucleus transcriptomes in the four libraries, which covered 30,469 genes, with median genes per nucleus at 1,018 and median unique molecular identifiers (UMIs) per nucleus at 1,390 (Supplementary Data 1). After integrating the data by scVI, ten cell clusters were identified (Fig. 1b and Supplementary Fig. 1). However, we found that cluster 0 and cluster 3 were indistinguishable in the uniform manifold approximation and projection (UMAP) plots; even the most specifically expressed genes in cluster 0 tended to be upregulated in cluster 3 (Supplementary Fig. 1), which led us to merge these two cell clusters, thus obtaining nine cell clusters (Fig. 1c). For cluster annotation, we used 15 cell type-specific markers^{31–41} previously described in *M. truncatula* (Supplementary Table 1) and we were able to assign most of clusters to a specific cell type (Fig. 1d and Extended Data Fig. 1). For example, *MtPLT1*, *MtPLT3* and *MtPLT4* have previously been reported to exhibit high activity in meristem³⁸ and our data show that these three genes are primarily expressed in cluster 6, hence it is annotated as meristem. Similar procedures were applied for cluster 1 as the epidermal cells using *MtPT1* (ref. 35), *MtERN2* (ref. 36) and *MtPUB1* (ref. 37); and cluster 3 as cortex using *MtIFS1* and *MtIFS3* genes which are specifically expressed in the root cortex³¹. Additionally, we noticed that

cluster 3 also contains *MtPT5* which is shown to be expressed in both the cortex and epidermal cells³². We also annotated cluster 5 as endodermal cells using *MtCASPL-1* and *MtCASPL-2* (ref. 34) and cluster 6 as meristem using *MtLAX2* (ref. 41). In particular, we found that *MtLAX2* expression is also high in clusters 2 and 8. Because *MtPHO1.1* (ref. 40) and *MtSUNN*³⁹, which are primarily expressed in vascular bundles, exhibit high specificity in clusters 4 and 8, we annotate clusters 2, 4 and 8 as stele. The annotation results were further confirmed by using annotation transferred from an *Arabidopsis* root single-cell dataset (Extended Data Fig. 2a,b). We then identified each cell cluster-specific gene, including 41 known SNF genes (Fig. 1e, Extended Data Fig. 3, Supplementary Fig. 2 and Supplementary Data 2). By gene ontology (GO) analysis, we identified the pathways of upregulated genes for each cluster (Supplementary Fig. 3). The enriched calcium-related pathways for cluster 1 (epidermis) and cell cycle-related pathways for cluster 6 (meristem), further confirmed the accuracy of the cellular annotation (Supplementary Fig. 3). Further analysis revealed that on the basis of single-cell data from *Arabidopsis* roots, as well as GO term enrichment results, the three stele clusters can be distinguished (Extended Data Fig. 2a and Supplementary Fig. 3). For cluster 2, the enriched GO terms are involved in auxin signalling and transport. The label transfer results imply that the cluster is procambium. Additionally, we found that two homologous genes of the procambium-specific expressed genes identified in *Arabidopsis*, *AtSMXL5* (ref. 42) are mainly expressed in cluster 2 (Supplementary Fig. 4). Regarding cluster 4, although the label transfer results indicate its similarity to the pericycle in *Arabidopsis*, the enrichment of multiple GO terms related to metal ion transport suggests its xylem identity. For cluster 8, the label transfer results indicate that almost all nuclei within this cluster are more similar to phloem cells in the *Arabidopsis* root data. This is further supported by the enrichment of specifically expressed genes associated with the GO term 'phloem development'. Therefore, these findings suggest that cluster 2 is probably the procambium, cluster 4 is a mixture of pericycle and xylem and cluster 8 is probably the phloem. In summary, we successfully revealed the cell heterogeneity and classified the major cell types in *Medicago* roots.

Analysis of the ubiquitously expressed gene response pattern

To investigate the response patterns of genes in early signalling pathways, we grouped the 1,868 genes detected in all cell clusters at four timepoints into six gene categories according to their expression at different timepoints in each cluster (Fig. 2a and Supplementary Data 3). A total of 838 genes that could not be accurately assigned to these six categories were discarded (Methods). Five gene categories, from categories 2 to 6, showed similar expression patterns in all cell clusters (Fig. 2b). Categories 2 and 5 were activated from 0.5 hpt (Fig. 2b). GO enrichment analysis showed that these two categories are involved in all layers of RNA transcription and processing, as well as transcription-associated histone modifications and chromatin remodelling (Supplementary Fig. 5), suggesting that transcription is activated after brief inoculation with NFs. Both categories 3 and 6 transiently increased within 0.5 hpt and decreased afterward, as opposed to category 4 (Fig. 2b). In agreement with their response patterns, category 3 was enriched in intracellular signal transduction after NF inoculation and the top GO terms of category 6 were mainly involved in stress response and phospholipid metabolism. Category 4 was related to carbohydrate phosphorylation, indicating the subsequent use of nutrients during the early symbiotic stage. Although all clusters showed similar trends in these five gene categories, the change of gene expression in epidermal cells was always more dramatic and discrepant with that in other cell types (Fig. 2b). In addition, category 1 showed highly heterogeneous expression patterns, in which the epidermal cell cluster possessed a unique pattern compared to other clusters (Fig. 2b), with a sharp downregulation at 0.5 hpt and gradual recovery to the uninoculated level at 24 hpt. Several endocytosis-related genes, including orthologues of *CLATHRIN HEAVY CHAIN 1 (CHC1)* from *L. japonicus*⁴³,

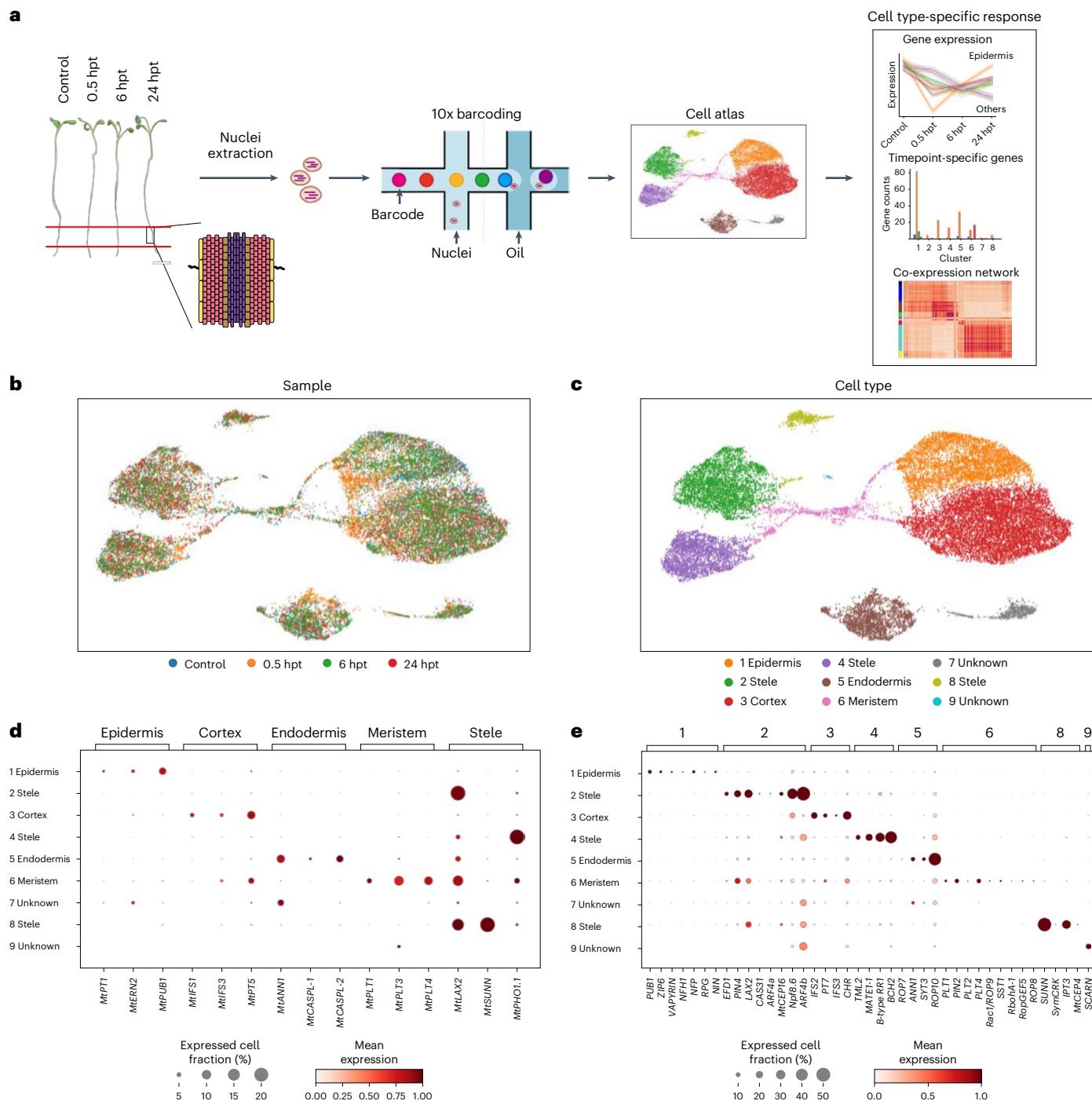


Fig. 1 | Single-nucleus transcriptomes reveal *Medicago* root heterogeneity during the inoculation time course. **a**, Schematic diagram of the single-nucleus transcriptomics library preparation and analysis. **b**, Integration of four single-nucleus datasets. **c**, UMAP visualization of nine identified cell clusters. Clusters 0 and 3 are merged because the upregulated expressed genes of cluster 0 were simultaneously upregulated in cluster 3 and the boundary between

the two clusters was ambiguous on the UMAP plot. More details are shown in Supplementary Fig. 1. **d**, Dot plot of cell type-specific marker genes. The size of the dot represents the percentage of cells expressing the gene in each cluster. The color intensity represents the level of gene expression. Gene expression is the average expression per cluster after min-max normalization. **e**, Dot plot of partial known SNF genes.

GRAVITROPISM DEFECTIVE2 (*GRV2*) from *Arabidopsis*⁴⁴ and two adaptor protein complex subunit genes *AP-1 COMPLEX SUBUNIT GAMMA-LIKE2* (*AP1G2*)⁴⁵ and *AP-4 COMPLEX SUBUNITE EPSILON* (*AP4E*)⁴⁶, were included in category 1. GO analysis showed that endocytosis is the most enriched pathway in category 1, indicating that the response of endocytosis in the epidermis is different from that of other cell types (Supplementary Fig. 5). Cell endocytosis has been widely recognized to be closely related

to symbiosis in the uptake of rhizobia cells⁴³. The presence of NFs within ITs also implies that endocytosis is required for the NF-induced signalling pathway⁴⁷. *CHC1* is involved in clathrin-mediated endocytosis, which affects the expression of multiple early nodulation genes in *L. japonicus* and influences rhizobial infection⁴³. Homologues of *AP1G2*, *AP4E* and *GRV2* are required for endocytosis but their roles in symbiosis remain unknown. In conclusion, we found that the epidermis

displays a more dramatic response to induction by NFs relative to other cell types and the 0.5 hpt timepoint is distinct from other timepoints.

snRNA-seq reveals cell type-specific response induced by NFs

Next, we investigated the response of each cell type to NFs. By integrating the four snRNA datasets from different timepoints, consistent with the above results, we found that the transcriptome at 0.5 hpt shows the lowest similarity with other timepoints, indicating that the root undergoes extensive transcriptome reprogramming rapidly after NFs stimulation and gradual recovery within 6 hpt (Fig. 2c and Supplementary Fig. 6). Previous whole root study²¹ has also revealed the transient reprogramming at 1 h after rhizobia inoculation and our results indicate that the transient reprogramming is most evident in epidermal and cortical cells, highlighting the essential role of these two cell types in NFs perception (Supplementary Fig. 6). We also found that the numbers of timepoint-specific upregulated genes at 0.5 hpt were highest in all clusters, confirming the uniqueness of 0.5 hpt in early response (Fig. 2d and Supplementary Data 4). We have validated these findings by adding a new biological replicate and obtained highly consistent results by analysing the two-replicates datasets using a fully combination-based method (Methods) (Extended Data Fig. 4 and Supplementary Data 5). Notably, GO analysis of these induced genes showed that plant defence-related pathways were induced in almost all differentiated cell types (except cluster 8, a subtype of stele) at 0.5 hpt, especially in the epidermis and cortex. These remained active in the epidermis at 6 hpt (Fig. 2e and Supplementary Figs. 7–9), indicating that plant defences were transiently activated in roots after the application of NFs. In addition, the pathways for synthesis of flavonoid, polyketide and chalcone were enriched at 0.5 hpt in cortical cells and only a few cells in other cell types (Fig. 2e and Supplementary Fig. 10), suggesting that the local accumulation of flavonoid is also activated during early symbiotic response. Furthermore, at 0.5 hpt, xylitol transferase, which is essential for the rearrangement of the cell wall, had higher expression activity in the endodermis (Supplementary Fig. 11). In addition, the activation pattern of the cell wall biogenesis and organization pathway is similar to that of xylitol transferase (Fig. 2e and Supplementary Figs. 7 and 8), indicating that the cells are ready for remodelling of the cell wall, although ITs are formed only after several hours of root hair curling during rhizobial infection.

Phytohormone signalling is required for the establishment of SNF⁴⁸ and previous studies found that the concentration of various phytohormones changes during the first 24 h after treatment with either purified NFs or rhizobia^{16,17,19,49}. Analysis of phytohormone-related genes revealed that the expression of several genes involved in the biogenesis, sensing and metabolism of ethylene, auxin, cytokinin and gibberellin are highly expressed at 0.5 hpt (Supplementary Data 6). We found that *MtACS2*, *MtACS3* and *MtACO*, which are essential for ethylene synthesis⁵⁰, are highly expressed at 0.5 hpt but in different cell types. *MtACS3* could be identified in almost all cell types, whereas *MtACS2* and

MtACO were only identified in the epidermis and cortex, respectively. In addition, *ETHYLENE RECEPTOR-LIKE 3* (*MtETR3*) was identified in the endodermis. For auxin, *YUCCA8*, which is involved in synthesis of indole-3-acetic acid (IAA), was identified in the stele (cluster 4). We also identified *AUXIN RESPONSE FACTOR 16a* (*MtARF16a*) in the epidermis and cortex, *AUXIN RESPONSIVE GH3 FAMILY* (*MtGH3.1*) and *PIN-FORMED AUXIN EFFLUX CARRIER COMPONENT 10* (*MtPIN10*) in the endodermis, *MtPIN3* in the cortex, *auxin importer LIKE AUX1-5* (*MtLAX5*) in the stele (cluster 2) and *AUXIN CONJUGATE HYDROLASE* (*MtIAR31*) in the cortex and stele (cluster 4). For cytokinin, *ISOPENTENYL TRANSFERASE 2* (*MtIPT2*) and *MtIPT3*, which are involved in cytokinin synthesis, were found in the stele (cluster 4) and cortex, respectively. *LONELY GUY 3* (*LOG3*), which encodes another cytokinin synthesis enzyme, was upregulated in the epidermis and endodermis. *LOG-like 4* also showed its peak expression in the epidermis. For gibberellin, we found that *GIBBERELLIN C20-GA2-OXIDASE 7* (*MtGA2ox7*) is upregulated in almost all cell types like *MtACS3* but *MtGA2ox4*, *MtGA2ox5* and *MtGA2ox6* were only identified in epidermis, cortex and stele (cluster 4), respectively.

We also identified genes that were specifically downregulated by NF at 0.5 hpt (Supplementary Fig. 12a and Supplementary Data 4). However, unlike the upregulated genes, few GO terms were found to be enriched. We found only one enriched term each for clusters 4 and 5 (Supplementary Fig. 12b). Specifically, we observed that genes for ‘methyl indole-3-acetate esterase activity’ were enriched in cluster 4, suggesting that methyl-IAA hydrolysis is repressed at this timepoint. However, we also identified *YUCCA8*, which is required for IAA synthesis, as a timepoint-specific upregulated gene in cluster 4 (Supplementary Data 4). This result suggests a complex regulation network for IAA synthesis in the early response.

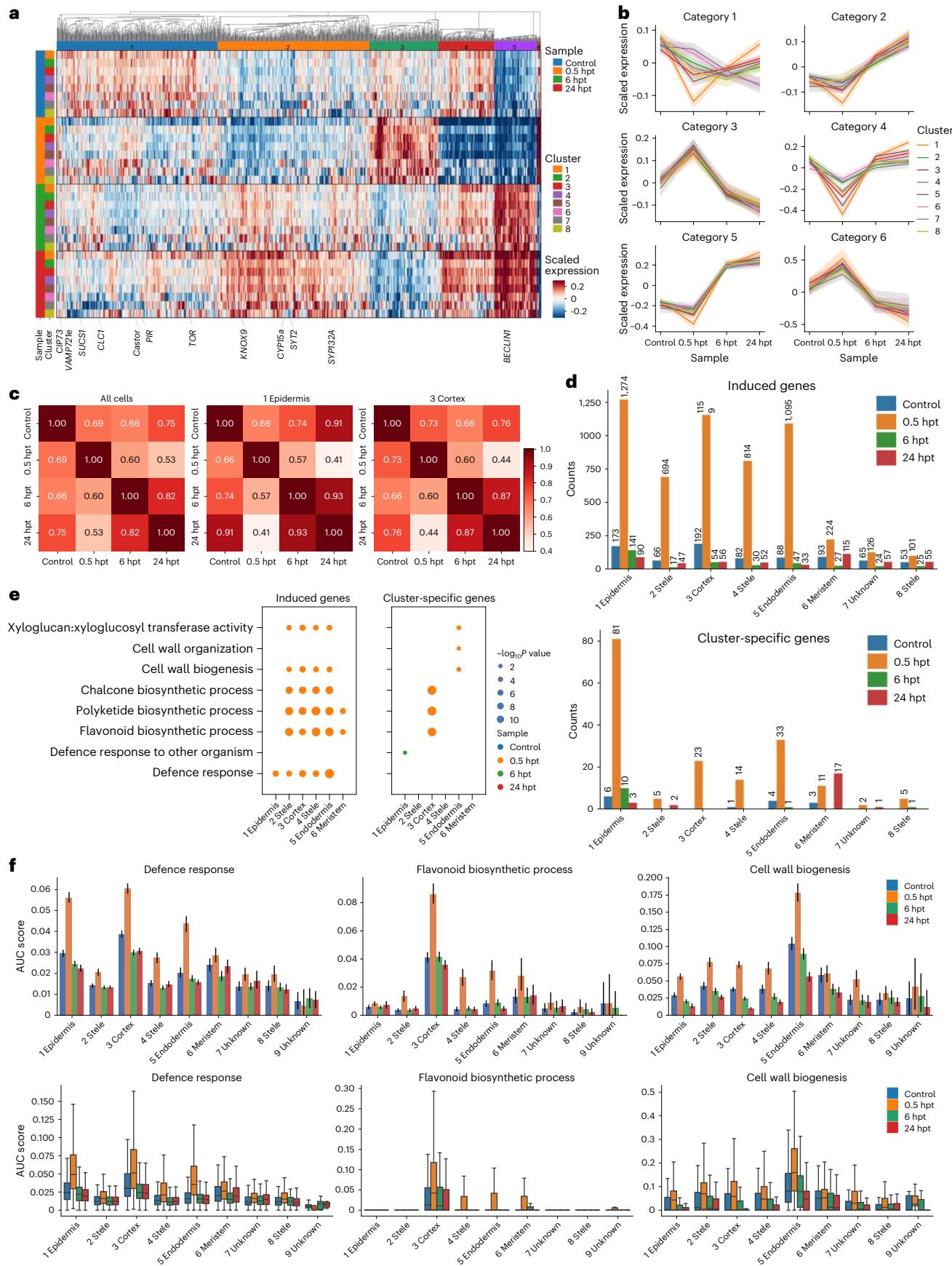
Taken together, our data suggested that the pathways involved in defence, endocytosis, flavonoid accumulation and cell wall rearrangement respond rapidly to NFs signals within 0.5 hpt. Epidermal, cortical and endodermal cells showed the most dramatic reprogramming (Supplementary Fig. 6b), suggesting their central role in the early response.

A subtype of epidermal cells is essential for SNF

By reclustering, the epidermis can be further divided into two sub-cell types (subcluster 1-0 and 1-1) (Fig. 3a). Both subclusters were identified in all four samples, indicating that the difference contributed to the divergence of the two subclusters before the application of the NFs (Fig. 3b). We then compared our results with data generated from specific cell types: ref. 19 used laser microdissection to obtain epidermal cells at 4 and 24 hpt, while ref. 16 obtained root hair cells by brushing liquid nitrogen-treated roots with a paint brush at 24 hpt and ref. 18 isolated root hair cells at 4 and 20 hpt by vortexing the root segments in liquid nitrogen. We then identified genes upregulated by NF treatments at each timepoint relative to control sample for each subcluster and compared them with these earlier epidermis- or root hair-specific studies^{16,18,19} (Extended Data Fig. 5 and Supplementary Data 4). We observed

Fig. 2 | Single-nucleus transcriptomes reveal cell type-specific responses induced by NFs. **a, b**, Six major gene patterns varied over inoculation time. Only genes that are expressed in all cell clusters are used for analysis. Genes were scaled according to their average expression in each cell cluster. The heatmap (**a**) shows the expression pattern of each gene and the line plot (**b**) shows the average of all genes in each gene category. The shading represents 90% confidence interval of average values. Refer to Methods for details. **c**, Heatmap represents the alignment score between each sample pair. The alignment score is the percentage of cells with a mutual nearest neighbour for each dataset pair. The high scores indicate a high similarity of cell expression between the two datasets. **d**, The number of timepoint-specific genes identified in each cell cluster for the NF-treatment time course. Upper: genes which are specifically expressed at each treatment timepoint (or control sample). Lower: timepoint-specific genes with cluster-specific expression patterns. These genes are not only expressed at a particular timepoint after treatment but are also cluster specific, that is

spatiotemporal-specific genes. The numbers above the bars are the numbers of the corresponding genes. **e**, GO enrichment analysis of genes listed in **d**. The complete GO analysis results are shown in Supplementary Figs. 7 and 8. The *P* values were calculated using the one-sided Fisher's exact test (hypergeometric test) wrapped in clusterProfiler. Subsequently, the default parameters were used to filter the GO entries, specifically those with *P* < 0.05 and BH-adjusted *P* < 0.2. The unadjusted *P* values are plotted in the panel. **f**, The expression pattern of different pathway genes induced in 0.5 hpt. Expression levels of gene sets are measured by AUC score. In total 25,276 biologically independent nuclei are plotted in this figure. Upper: the mean value of the AUC score among nuclei and the error bars represent 99% confidence interval of average values. Lower: the boxplot visualization represents the AUC score distribution among nuclei. Boxplot centre, median; bounds of box, lower quartile (Q1) and upper quartile (Q3); minima, Q1 – 1.5 (Q1–Q3); maxima, Q1 + 1.5 (Q1–Q3).



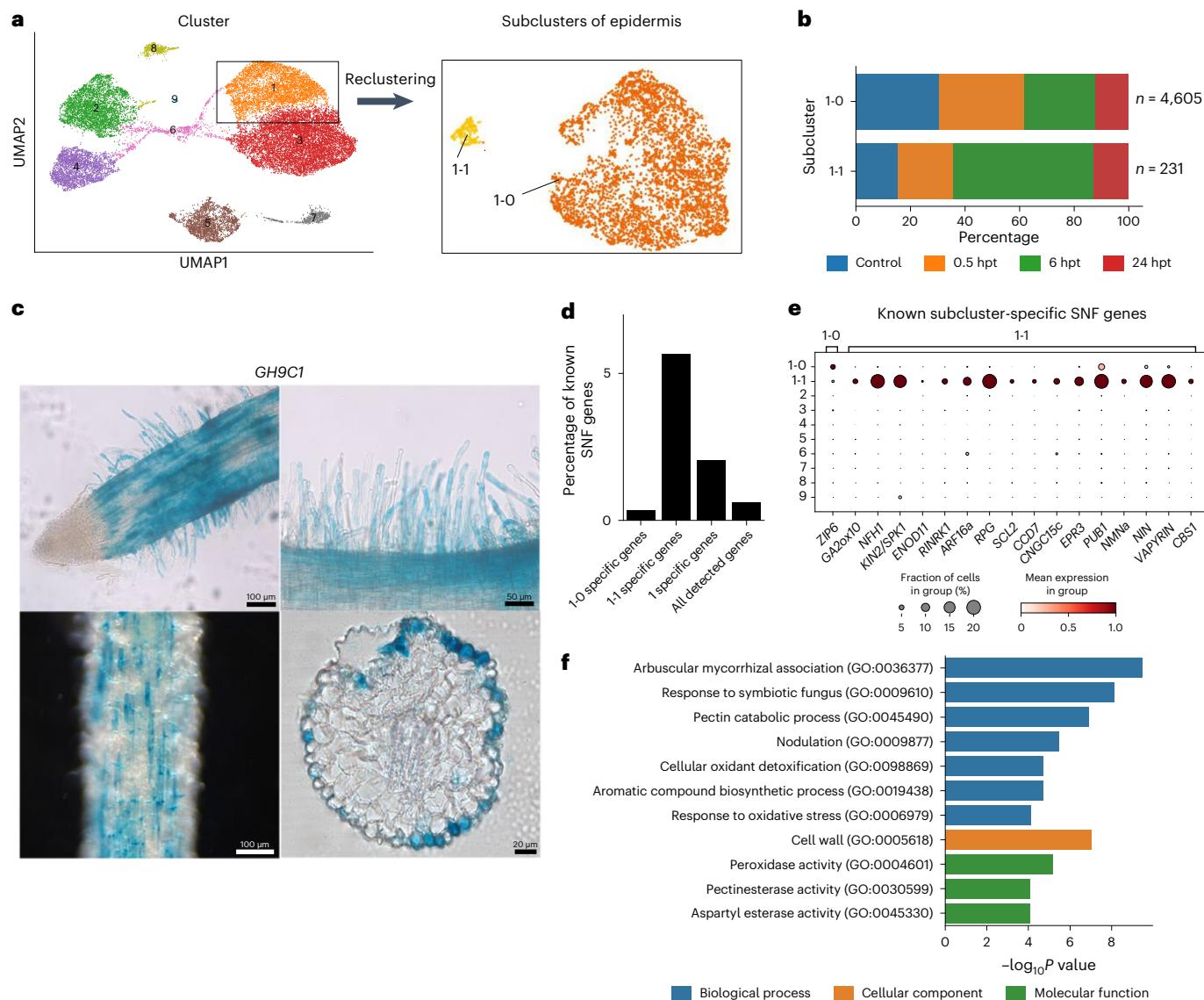


Fig. 3 | A subtype of epidermal cells is essential for SNF. **a**, UMAP visualization of epidermis reclustering result. **b**, Bar chart representing the percentage of cells from different timepoints in each sub-cell type. *n*, cell number. **c**, Validation of cluster 1-1 identity by GUS-reporter lines. Experiments were independently repeated three times with similar results. **d**, Percentage of known SNF genes in different cell type-specific gene sets. **e**, Dot plot representing the expression

pattern of specific SNF genes for each epidermis subcluster. **f**, GO enrichment analysis of cluster-specific genes in cluster 1-1. The *P* values were calculated using the one-sided Fisher's exact test (hypergeometric test) wrapped in clusterProfiler. Subsequently, the default parameters were used to filter the GO entries, specifically those with a *P* < 0.05 and BH-adjusted *P* < 0.2. The unadjusted *P* values are plotted in the panel.

the highest proportion of overlap (50.6%) between cluster 1-1 at 6 hpt and the RNA-seq data of ref. 18 at 4 hpt, while there was limited overlap with the data from ref. 16. The limited overlap may be due to our use of only the NF-susceptible zone of the roots for snRNA-seq, together with differences in the NFs inoculation time and method used to isolate epidermal cells. We next found that the two root hair-specific genes *MtNFH1* (ref. 51) and *RPG*⁵², which are reported as essential genes for rhizobial infection, were highly enriched in subcluster 1-1 (Supplementary Data 2). In addition, the GUS signal of 1-1 specifically expressed gene *GH9C1* in our list primarily labelled the root hair and epidermal cells in the mature zone (Fig. 3c), which is consistent with the expression pattern of the root hair marker gene *AtGH9C1* in *Arabidopsis* root hair and epidermis⁵³. Thus, we assigned the identity of subcluster 1-1 to the root hair. For subcluster 1-0, the label transfer results from *Arabidopsis* single-cell datasets by scANVI suggested that it was a mixture of epidermal non-hair cells (atrichoblasts) and immature root hair cells (trichoblasts in the meristem and elongation zone) (Extended Data Fig. 2c).

In agreement with the important role of root hair at the beginning of rhizobia infection, we found that the proportion of known SNF genes in subcluster 1-1 was significantly higher than that in 1-0 (Fig. 3d,e and Supplementary Data 2). In addition, the GO terms associated with the arbuscular mycorrhizal symbiosis (AMS) were also enriched (Fig. 3f). Previously, ref. 54 discovered that ~60% of differentially expressed genes induced by Myc-LCOs were also induced by NFs. Our results corroborate this finding, suggesting that NFs also activate AMS signalling pathways through the common symbiotic signalling pathway because of structural similarity between NFs and Myc-LCOs⁵⁵.

Co-expression of *MtFER* and *LYK3*

In addition to epidermal cells, cortical cells also underwent extensive transient reprogramming at 0.5 hpt (Fig. 2c). To further explore their functions, we extracted epidermal and cortical cells at 0.5 hpt to construct a co-expression network (Fig. 4a and Supplementary Data 7). Nine co-expression modules were identified (Supplementary Fig. 13a).

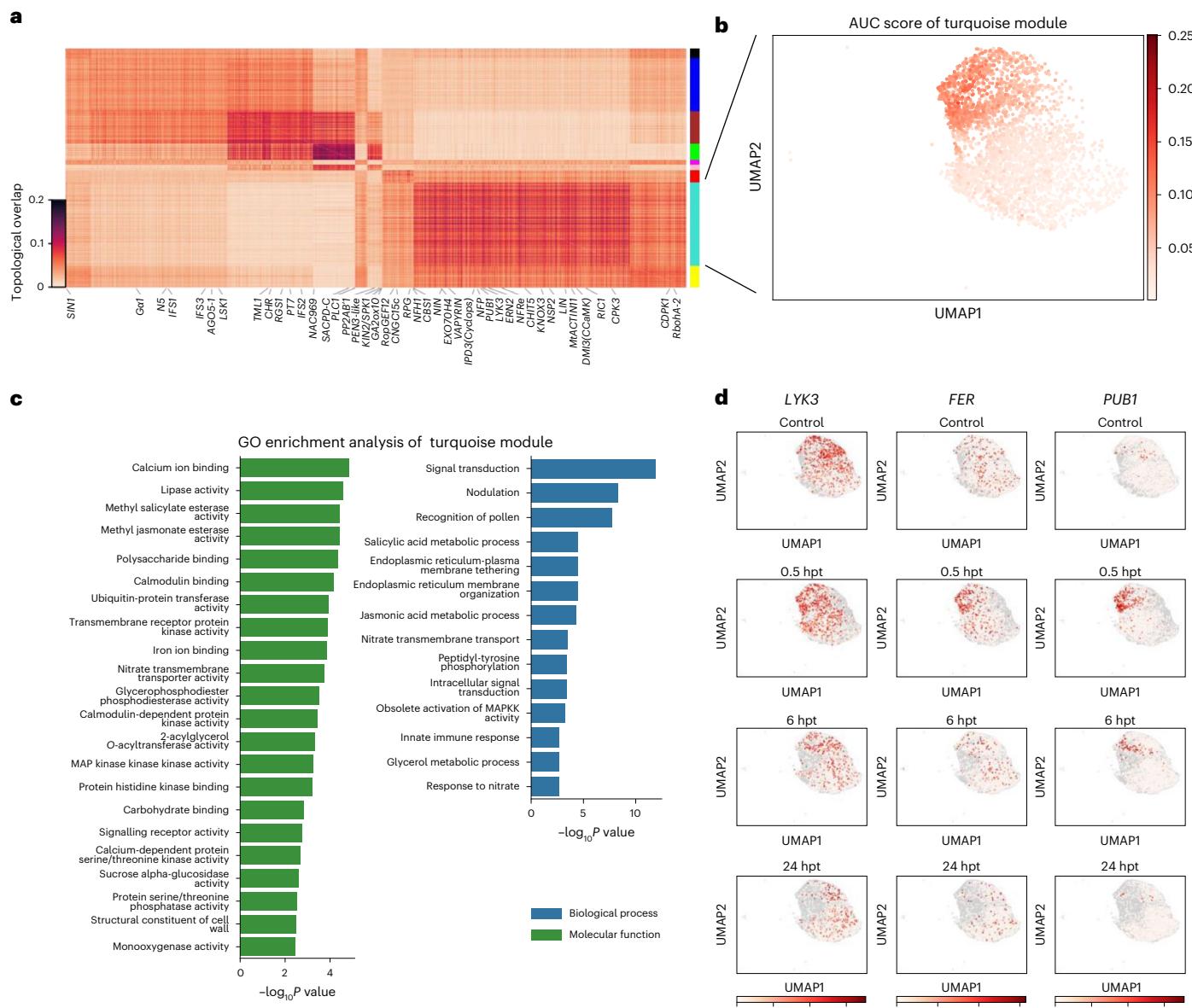


Fig. 4 | Single-cell data reveal a nodulation-related co-expression module shared by FER, PUB1 and LYK3. **a**, Co-expression modules identified by WGCNA in epidermal and cortical cells at 0.5 hpt. The colour scale reflects the topological overlap, ranging from white to deep red. A higher value on the scale indicates a greater level of similarity between genes. **b**, UMAP visualization of activation degree of the turquoise module. Expression levels of gene sets are measured by AUC score. **c**, GO enrichment analysis of genes within the turquoise module.

The *P* values were calculated using the one-sided Fisher's exact test (hypergeometric test) wrapped in clusterProfiler. Subsequently, the default parameters were used to filter the GO entries, specifically those with $P < 0.05$ and BH-adjusted $P < 0.2$. The unadjusted *P* values are plotted in the panel. **d**, UMAP visualization of the expression patterns of the LYK3, PUB1 and FER. These three genes belong to the same co-expression module.

The blue module, which is associated with flavonoid synthesis, was highly activated in the left part of cortical cells in the UMAP embedding space (Supplementary Fig. 13). Module yellow, which is characterized by calcium-binding pathways, was co-expressed in the left part of cortex and epidermis (Supplementary Fig. 13b), which is consistent with the specific activation of Ca^{2+} signalling after the perception of bacterial signalling factors. These results indicate that different subtypes of the same cell type also differ in their functions in NF-induced early response. Besides, we found the module turquoise is upregulated in epidermis at 0.5 hpt and enriched in 'signal transduction' and 'nodulation' pathway, containing many known SNF genes including *NODULATION SIGNALLING PATHWAY 2 (NSP2)*³⁶, *LUMPY INFECTIONS (LIN)*⁵⁶, *CHITNASE5 (CHIT5)*⁵⁷, *DOES NOT MAKE INFECTION 3 (DMI3)*⁵⁸,

*INTERACTING PROTEIN OF DMI3 (IPD3)*⁵⁹, *NFP*¹², *VAPYRIN*⁶⁰, *ETHYLENE RESPONSE FACTOR REQUIRED FOR NODULATION 2 (ERN2)*³⁶, *PUB1* (ref. 37) and *LYK3* (ref. 13) (Fig. 4a,c). Interestingly, we found that *MtFER* (Medtr7g073660), an apparent orthologue of *Arabidopsis FERONIA* (Extended Data Fig. 6), was also found in the turquoise module with *LYK3*. *LYK3* can phosphorylate *PUB1* and regulate nodulation in early response³⁷. Besides, we found all three genes peaked in terms of expression levels at 0.5 hpt and declined gradually subsequently, suggesting potential interactions of *MtFER* with *PUB1* and *LYK3* (Fig. 4d).

MtFER shows similar expression pattern as *LYK3*

FER regulates plant pollen tube reception⁶¹, root hair growth^{62,63} and immunity response⁶⁴. As *LYK3* and *MtFER* co-express in the

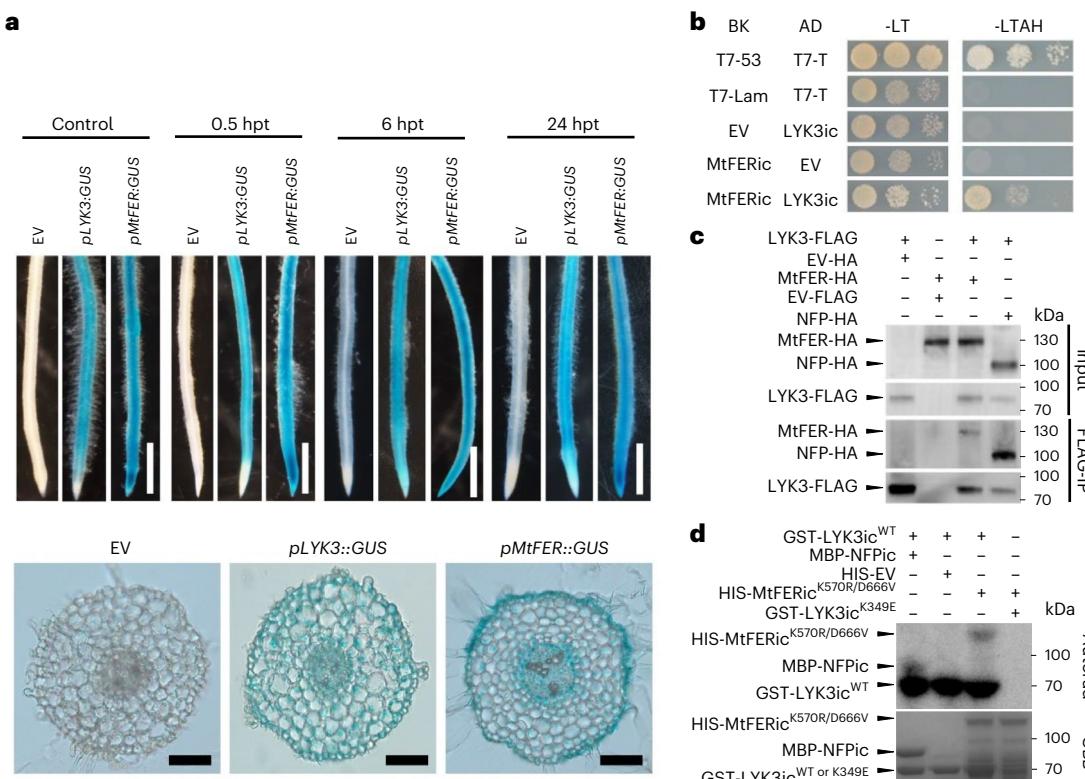


Fig. 5 | MtFER is phosphorylated by LYK3. **a**, *pLYK3::GUS* and *pMtFER::GUS* show similar expression patterns with or without NF treatment. Scale bars, 500 µm for roots (top) and 100 µm for root sections (bottom). Experiments were independently repeated three times with similar results. **b**, A Y2H assay was used to test for interactions between MtFER and LYK3. Yeast cells carrying different combinations of constructs are listed on the left. Serial dilutions (ten times) of yeast cells expressing the indicated proteins from the pGBT7 (BK) and pGAD7 (AD) vectors were plated on SD2/-Leu-Trp (-LT) medium or SD4/-Leu-Trp-Ade-His (-LTAH) medium with 30 mM 3-amino-1,2,4-triazole (3AT). T7-T/T7-53 and T7-T/T7-Lam served as positive and negative controls, respectively. **c**, LYK3 co-immunoprecipitates with MtFER in *Nicotiana benthamiana*. FLAG-tagged LYK3 and

HA-tagged MtFER/NFP were co-expressed in *N. benthamiana* leaves. Proteins were immunoprecipitated (IP) with FLAG-M2 beads and separately analysed by western blotting with anti-FLAG and anti-HA. NFP served as the positive control. Experiments were independently repeated twice with similar results. **d**, MtFER is phosphorylated by LYK3 in vitro. All proteins were expressed in *E. coli*. The phosphorylation assay was performed using purified proteins and [γ -³²P]-ATP and phosphorylation was detected by autoradiography. NFP served as the positive control and the kinase-dead form LYK3ic^{K349E} served as the negative control for the in vitro phosphorylation assays. Experiments were independently repeated twice with similar results.

same module, we then investigated their expression pattern during symbiotic response and nodule development by analysing their *promoter::GUS*-transformed hairy roots. Our results showed that *pMtFER::GUS* has a similar expression pattern in roots/nodules as *pLYK3::GUS* and is mainly expressed in epidermal cells, cortical cells and stele in the NFs/rhizobial response zone and developing nodules inoculated with or without NFs or *Sm1021* (Fig. 5a and Extended Data Fig. 7). We also noticed an exception, *pMtFER::GUS* is highly expressed in the root tip while *pLYK3::GUS* is weakly expressed in this region (Fig. 5a and Extended Data Fig. 7).

Interestingly, our data showed that the MtFER protein can interact with LYK3 in the yeast two-hybrid (Y2H) assay (Fig. 5b). We further confirmed the interaction by a co-immunoprecipitation (Co-IP) assay when LYK3-FLAG and MtFER-HA were co-expressed in *Nicotiana benthamiana* leaves (Fig. 5c). Finally, we found that LYK3 can phosphorylate MtFER in an in vitro kinase assay (Fig. 5d). Our results indicate that the requirement for MtFER in the rhizobial symbiosis might be related to its interaction with LYK3.

MtFER is required for root-nodule symbiosis

To further investigate whether MtFER is required for root-nodule symbiosis, we analysed a *Tnt1* insertion mutant of MtFER (NF4385) from the *M. truncatula* mutant database (<https://medicago-mutant.dasnr.okstate.edu/mutant/orderseed.php>). Homozygote *Mtfer* mutant plants

were not isolated (Supplementary Fig. 14), indicating that MtFER shares a similar role to that of *Arabidopsis* in fertilization and plant growth. However, MtFER overexpression had no effect on root and nodule development (Supplementary Fig. 15).

Therefore, we generated transformed hairy roots expressing three different RNAi hairpins, allowing the RNA-mediated interference of MtFER. Root growth of RNAi plants was suppressed and their root length was only ~60% of that of the empty vector (EV) control (Fig. 6a–c). Markedly reduced root hair elongation and density were also observed in the RNAi plants (Extended Data Fig. 8). Interestingly, the infection events were significantly repressed in RNAi plants with fewer infection foci/elongating ITs in root hairs and fewer ramified ITs in nodule primordia (Fig. 6d). Total nodule number, the number of functional pink nodules and nodule density were also dramatically reduced in MtFERi plants (Fig. 6e,f and Extended Data Fig. 9a–d). The symbiotic phenotype of transgenic hairy roots was further assessed by inoculation with *hemA:LacZ*-marked *Sinorhizobium meliloti* strain *Sm1021*. X-Gal-staining of *Sm1021-LacZ* (blue) revealed that nodules of EV hairy roots harbour more rhizobia than those of MtFERi hairy roots at 21 days post inoculation (dpi) (Extended Data Fig. 9e–h). Nearly the whole nodule is stained blue in EV hairy roots (Extended Data Fig. 9e), while rhizobia in MtFERi hairy roots were observed mainly in the distal zone of nodules where they seem to be restricted to the infection zone (Extended Data Fig. 9f–h). Our data indicated that knockdown of MtFER

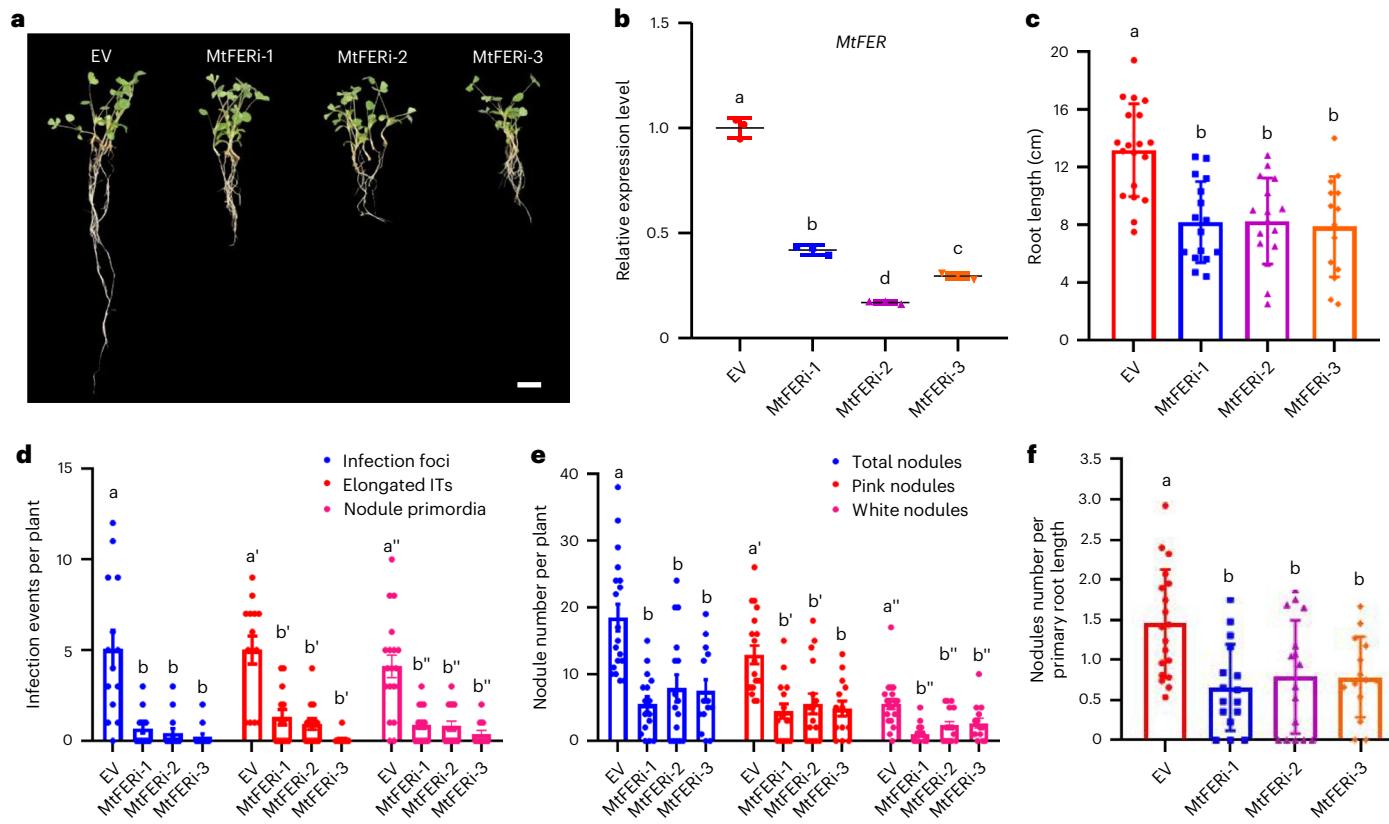


Fig. 6 | Suppression of *MtFER* expression inhibits root and nodule development.

a, Representative photograph of root phenotype at 21 dpi in wild-type hairy roots transformed with an EV and MtFERi (MtFERi-1/-2/-3) ($n \geq 14$). Scale bar, 1 cm. **b**, Relative expression levels of *MtFER* in transgenic hairy roots ($n = 3$). Data are mean \pm s.d. Expression levels were normalized against the reference gene *EF-1*. **c**, Root length of transgenic hairy roots (EV, $n = 19$; MtFERi-1/-2, $n = 16$; MtFERi-3, $n = 14$). Data are mean \pm s.d. **d**, Quantification of different infection events in transgenic hairy roots 7 dpi with *Sm1021-LacZ* (EV, $n = 14$; MtFERi-1/-2/-3, $n = 13$). Data are mean \pm s.d. **e**, Quantification of nodules

in transgenic hairy roots 21 dpi with *Sm1021* (EV, $n = 19$; MtFERi-1/-2, $n = 16$; MtFERi-3, $n = 14$). Data are mean \pm s.d. **f**, Density of nodules in transgenic hairy roots (EV, $n = 19$; MtFERi-1/-2, $n = 16$; MtFERi-3, $n = 14$). Data are mean \pm s.d. Experiments in **a-f** were independently repeated three times with similar results. Statistically significant differences between EV and MtFERi-1/-2/-3 in experiments **b-f** were determined by one-way ANOVA followed by Duncan's multiple range tests ($P < 0.05$); different letters indicate significant difference. The exact P values of Duncan's multiple range tests can be found in Supplementary Data 8.

expression impairs both rhizobial epidermal infection and rhizobial proliferation within nodules.

We then investigated whether the rhizobial infection and nodule development defects observed in MtFERi hairy roots are associated with changes in symbiotic or defence gene expression. Expression levels of marker genes for defence (*WRKY* and *Chitinase*) and symbiosis (*NIN*, *ENOD11*, *Vapyrin* and *FLOT4*) increased after NF treatment in EV hairy roots, especially at 0.5 and 6 hpt (Extended Data Fig. 10). This was mirrored by the single-nucleus transcriptome data (Supplementary Fig. 16). Notably, expression of both sets of these marker genes was significantly repressed in MtFERi hairy roots after NF treatment (Extended Data Fig. 10). These data imply that loss of *MtFER* might directly impede early symbiotic responses. Our results further indicate that LYK3-MtFER module-mediated NFs signalling might play an important role in rhizobial symbiosis.

Discussion

The establishment of rhizobial symbiosis requires precise coordination of different cells in a time- and region-specific manner. The cell-type specificity of many SNF genes demonstrate the desirability of finely dissecting roots at the single-cell level. Although protoplast-based plant single-cell sequencing strategies have been widely used in the study of plant developmental biology, protoplast preparation is often difficult for complex tissues and the expression levels of thousands

of genes are altered during the process. Here, we used our previously established nuclei-based method to divide root cells into nine clusters, thus revealing cell type-specific changes in the early response to nodulation. Our data show a widespread response induced by NFs in almost all cell types, including vesicular transport in the epidermis and flavonoid synthesis in the cortex. In addition, we found that there is a transient but dramatic reprogramming at 0.5 hpt, which includes the activation and re-inhibition of plant immunity. We also set up a web server (http://zhailab.bio.sustech.edu.cn/sc_medicago) to make our research accessible to the community. These data resources will contribute to a more comprehensive understanding of the nodulation early response network.

Notably, for cell type annotation, several marker genes were not specific in our data; such results have also been reported in previous single-cell research⁶⁵. There are two possible explanations for this observation. Although single-nuclei sequencing has been well established in animals and plants^{26–29,66–68}, in specific tissues, the expression levels of a few genes differ substantially between the nucleus and whole cell⁶⁹. The second possible reason is unsuitable regulatory sequences such as promoters for the transgenic line or ‘probe’ for RNA in situ hybridization. Recently, ref. 65 used the GUS promoter-aided method to determine that cell cluster-specific expressed genes may not be expressed in the corresponding cell types. They suggested that this may be due to the failure to use the actual gene regulatory sequence

when constructing the vector. Nevertheless, this phenomenon poses challenges for the accurate annotation of cell types. To address this problem and to cope with the rarity of nodule marker genes, we also used datasets generated from *Arabidopsis* to validate our annotation (Extended Data Fig. 2). Previous reports indicate that gene expression in different cell types is not conserved between monocot and dicot⁷⁰. The data from *Arabidopsis* are sufficient to annotate the cell types of *Medicago*, indicating that, at least in dicots, gene expression in the roots is conserved enough to allow cross-species annotation. For other non-model dicots, published single-cell datasets will reduce the difficulty of annotation.

Compared with previous bulk-based transcriptome studies^{16,18,19,71}, most of the upregulated genes identified in our data are new and the proportion of overlap between these genes and each of the bulk datasets varies significantly, with a maximum of 50.6% (Extended Data Fig. 5c). However, for downregulated genes, there is little overlap among these datasets from different studies, suggesting that the identification of downregulated genes is more sensitive to differences in culture conditions, sampling and transcriptome profiling technologies (Extended Data Fig. 5d). We also conducted a comparison between our data with the snRNA-seq data of 48 h *S. meliloti*- and mock-inoculated roots obtained by ref. 30. The UMAP analysis reveals that the two snRNA-seq datasets could be generally well mixed and that cells of the same type tended to cluster together, indicating high similarity between these two datasets (Supplementary Fig. 17a). The further comparison revealed that, unlike the maximum overlap of approximately half of the response genes in bulk data, only a limited overlap between the two datasets was detected (Supplementary Fig. 17b). This discrepancy in overlap between the two different snRNA-seq datasets may be due to different inoculation times or because the response induced by NFs alone is not completely identical to that induced by rhizobia. Previously, ref. 16 profiled transcriptome change of *M. truncatula* root hair during the early stage of rhizobia infection. We would like to note that a limited degree of overlap was similarly observed by comparing the datasets of ref. 30 (1, 3 and 5 dpi) and ref. 16 (2 dpi) (Supplementary Fig. 18). The differences in processing times (0.5, 6 and 24 hpt versus 48 hpi) and treatment condition (NFs versus rhizobia) between these two different scRNA-seq datasets may lead to a further reduction in the degree of overlap.

In addition, single-cell sequencing allows us to distinguish signals from different tissues as much as possible, as enzymes required for flavonoid synthesis are induced in the cortex and cell wall biogenesis in the endodermis. To confirm whether these small numbers of non-cortical cells were doublets of other cell types and cortical cells, we assigned cells to either an activated or repressed state of the flavonoid synthesis pathway on the basis of the expression of genes required for flavonoid synthesis. We found little difference in the number of UMIs detected in these two states in non-cortical cells, suggesting that this was not caused by doublets (Supplementary Fig. 10c,d). Interestingly, we determined that some of these genes, including two orthologues of *F-BOX STRESS INDUCED 1* (*FBS1*), shared with pathways activated by pathogenic bacteria, were also resiled, indicating that the suppression of plant defence pathways may be due to the recognition of nitrogen-fixing bacteria by the plant. Polysaccharides secreted by rhizobia play a central role in the negative regulation of plant defence⁷². Inoculation with NFs alone led to transient activation of plant defence, suggesting that other mechanisms play a role in the negative regulation of plant immunity.

In the step of identifying the regulatory network at 0.5 hpt, we found a co-expression module enriched with known SNF genes and upregulated in the epidermis. This module (Supplementary Data 7) contained ten *PUB* genes (*PUB1*, *PUB2*, *PUB3*, *PUB4*, *PUB13*, *PUB30*, *PUB35*, *PUB40*, *PUB41*, *PUB49*) and two *LYK* genes (*LYK3* and *LYK4*).

On the basis of co-expression gene module analysis, we found *MtFER* and *LYK3* share similar responses to NF treatment (Fig. 4d), which

was further confirmed by *promoter::GUS* analysis (Fig. 5a and Extended Data Fig. 7). MtFER participates in rhizobial symbiosis by interacting and being phosphorylated by LYK3 (Fig. 5b-d and Fig. 6), indicating spatiotemporal symbiotic perception during rhizobial infection. In our present study, we show that the function of MtFER in promoting root growth is conserved in *M. truncatula*^{62,63}, the length of roots and root hairs as well as root hair density is severely inhibited in the *MtFER* knock-down roots (Fig. 6a,c and Extended Data Fig. 8), the reduction of infection events and nodule development may due to the lack of interaction points for rhizobia and rhizobial signal transduction. *Arabidopsis* fer mutants are more susceptible to infection by *Pseudomonas syringae*⁶⁴ and *Pto DC3000 COR*' (ref. 73) but ref. 74 show increased resistance to powdery mildew infection. Here, we found that defence marker genes (*WRKY* and *Chitinase*) were repressed in *MtFER* when hairy roots were treated with NFs (Extended Data Fig. 10a,b), indicating a role for *MtFER* in plant immunity. However, we also found that the expression of several symbiotic genes (*NIN*, *ENOD11*, *Vapyrin* and *FLOT4*) is repressed after NF treatment at 0.5 and 6 hpt in *MtFER* roots compared to EV roots (Extended Data Fig. 10c-f), demonstrating that *MtFER* affects the early symbiotic response. This implies that *MtFER* is required for the appropriate regulation of defence- and symbiosis-related gene expression at the early stages of the symbiosis to ensure rhizobia infection and nodule organogenesis. So, the observed nodulation phenotypes (Fig. 6d-f) in *MtFER* hairy roots may result from either a reduced symbiotic response or altered root development, including root hair formation or a combination of these two factors.

FER can interact with RAC/ROP and functions as central regulators for root hair growth⁶². Recently, RAC/ROPs are reported to be involved in root hair development, symbiotic signalling and root hair deformation during root-nodule symbiosis^{75,76}. Here, we showed that *MtFER* plays important roles in the root hair development (Extended Data Fig. 8), infection events (Fig. 6d) and nodule development (Fig. 6e-f and Extended Data Fig. 9). FER-RAC/ROP signalling pathways may play conserved roles in root hair development and symbiotic signalling during rhizobial infection. In conclusion, taking advantage of the single-nucleus extraction method, we successfully established the first single-cell level spatial transcriptomic atlas of *M. truncatula* roots in response to NFs, suggesting spatiotemporal symbiotic perception and early response during legume–rhizobial symbiosis.

Methods

Plant growth and NF treatment

M. truncatula seeds (A17 and *pENOD11::GUS*) were firstly scarified with concentrated sulfuric acid for 3–5 min, then surface-sterilized with 10% (v/v) bleach for 3 min, rinsed in sterile water five times and finally plated on 1% water agar medium for germination. Seeds were incubated at 4 °C for 3 d for cold stratification and kept at 24 °C overnight for germination in the dark (plates upside down). The germinated seedlings were transferred onto BNM (Supplementary Table 2) for growth in an environmentally controlled chamber with a 16 h light/8 h dark period at 22 °C (photon flux density of 250 $\mu\text{mol m}^{-2} \text{s}^{-1}$). NFs purified from *S. meliloti* strain 1021 as described by ref. 77 were dissolved in BNM to a final concentration of 5×10^{-9} M. Roots of 7-day-old seedlings were immersed in either BNM for 24 h (mock controls) or BNM with 5×10^{-9} M NFs for 24 h. For the 6 or 0.5 h NF treatments, the roots were first immersed in BNM for 18 or 23.5 h, at which point the medium was replaced with BNM containing 5×10^{-9} M NFs for another 6 or 0.5 h, respectively. For identification of differentially expressed genes (DEGs) the mock control plants were taken as control plants for all three NF treatments. Genes that were upregulated at one timepoint relative to all other timepoints were defined as timepoint-specific genes and the intersection of these timepoint-specific genes with cluster-specific genes was used to identify spatiotemporal-specific genes. The susceptible zone to NFs of roots was collected for nuclear isolation.

Plasmid construction and hairy roots transformation

The 3 kb promoter of *MtFER* and the 2.6 kb promoter of *LYK3* were amplified by PCR (primer information in Supplementary Table 3) from A17 genomic DNA. The candidate sequences that could be targeted by MtFERi were selected according to *pssRNAi* (<https://www.zhaolab.org/pssRNAi>) and were amplified by PCR (primer information in Supplementary Table 3) from A17 genomic DNA. The amplicons were cloned into pENTR/SD/D-Topo and then transferred to pBGWFS7 (for promoter::GUS analysis) and pK7GWIWG2D(II)-R (for RNAi analysis) by LR reactions (Invitrogen), respectively. The promoter::GUS and RNAi constructs were transferred into *Medicago* seedlings by *Agrobacterium rhizogenes* strain *Arqua-1* mediated hairy root transformation protocol⁷⁸. Ten days later, after removal of untransformed roots based on DsRED fluorescence, the composite transgenic plants were planted in pure vermiculite and inoculated with 5 ml of *S. meliloti* strain *Sm1021* or *Sm1021-LacZ* (*Sm1021* carrying a *ProHemA::LacZ* reporter) (optical density OD₆₀₀ = 0.03).

Histological analysis and microscopy

For GUS-staining analysis, transgenic roots were put into the staining buffer (containing 10 mM EDTA disodium salt, 100 mM NaH₂PO₄, 0.5 mM K₄Fe(CN)₆, 0.5 mM K₃Fe(CN)₆, 0.1% Triton-X100, 0.5 mg ml⁻¹ of X-Gluc, adjusting pH to 7.0) and vacuumed for 10 min, then incubated in the fresh GUS-staining buffer at 37 °C for 0.5 to 6 h. The GUS-stained roots were photographed by Nikon SMZ1270 microscope and then embedded in 3% (m/v) low-gelling temperature agarose (A9414, Sigma) and sectioned at 70 µm using a Leica VT1200S. For LacZ-staining analysis, transgenic roots were stained with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) as described by ref. 79. GUS/LacZ-stained roots and sections were taken by Zeiss Axio Scope A1 microscope.

Nuclei isolation and 10× snRNA-seq library construction

The nuclei of the roots were isolated as described in a previous publication⁸⁰. Briefly, the roots were placed in ice-cold 1× nuclei isolation buffer (NIB; Sigma-Aldrich, CELLYTPN1) supplemented with 1 mM dithiothreitol (DTT), 1× protease inhibitor (Roche) and 0.4 U µl⁻¹ murine RNase inhibitor (Vazyme, R301-03) in a 60 mm Petri dish on a frozen block and chopped thoroughly using a razor blade. The lysate was then applied to a pre-wet 40 µm strainer and centrifuged at 500g for 5 min at 4 °C. The nuclei pellet was resuspended in 500 µl of NIB stained with 4',6-diamidino-2-phenylindole (DAPI). For cell sorting, the nuclei were loaded into a flow cytometer with a 100 µm nozzle. A total of 100,000 nuclei for each sample were then sorted into a 15 ml tube with 1 ml of collection buffer (1× PBS with 1% BSA). The sorted nuclei were centrifuged at 500g for 5 min at 4 °C and the pellet was resuspended in 50 µl of 1× PBS with 1% BSA. The nuclei were visualized under a fluorescence microscope using the DAPI channel to check the quality and yield. A total of 40,000 nuclei were loaded into the 10x Genomics chip. Library construction for Illumina sequencing was performed with 10x Chromium Single Cell 3' Solution v.3.1 kit, as described previously²⁹.

Single-nuclei data analysis

We used Cell Ranger (v.6.0.0) to preprocess the Illumina reads and aligned them to the Medtr_A17_4 genome. The rest of the parameters are default, except for the ‘include-introns’ argument which is enabled. Subsequent analysis was performed using the Scanpy suite⁸¹. For quality control, we first used ScDblFinder⁸² to eliminate potential doublets, after which we maintained cells with gene counts of 300–3,000 and UMI counts of 500–5,000. We then reprocessed the raw data using the MtrunA17r5.0 genome and the resulting gene expression patterns, so both V4 and V5 versions of annotation can be explored on our website (https://zhaolab.bio.sustech.edu.cn/sc_medicago).

The scVI⁸³ was used to integrate the resulting matrices as official tutorial. We then used the Leiden algorithm (resolution 0.4) to perform

clustering on nearest neighbour graphs (n_neighbors = 15) constructed on the scVI embedding space and the UMAP algorithm (min_dist = 0.5) to visualize the distribution of the data in scVI space. After that we used existing experimentally validated marker genes to unveil the identity of each cell cluster and used the scANVI⁸⁴ algorithm to validate the annotation results using public single-cell data of *Arabidopsis* roots as the official tutorial. After clustering algorithm, we used Cellex⁸⁵ to identify specifically expressed genes in each cell cluster. In this step, only genes with a specificity score >0.8 were retained. To recluster the epidermis data, we first filtered the gene expression matrix to include only nuclei from cluster 1. We then performed principal component analysis on this subset of nuclei. After data integration, we used the Leiden algorithm to perform clustering on the neighbour graph (n_neighbors = 15) constructed on the low-dimension embedding space.

To identify cell type-specific responses, we used Scanorama⁸⁶ to calculate the transcriptome similarity of samples at different sampling timepoints. Then the Wald test implemented in Diffxpy to identify genes which specifically expressed in particular inoculation timepoints or differentially expressed compared with control sample. To maintain the quality of our analysis, we filtered out genes that were identified in fewer than three nuclei in each comparison. Only genes with adjusted P < 0.05 and fold change >2 will be retained. To identify the cell type-specific response of genes expressed in all clusters, only genes that could be detected as expressed in no less than 10% of the cells in at least one timepoint in all clusters were used for gene clustering. Genes were first scaled according to the average expression of each cluster and then clustered into six different categories using k-means implemented in scikit-learn⁸⁷. If the distance of a gene from the centre of the category to which it belongs is >0.8, the gene is discarded. Then, we performed hierarchical clustering separately using cosine similarity.

We used WGCNA⁸⁸ to perform the gene co-expression analysis. Briefly, we first extracted the top 6,000 highly variable genes by triku⁸⁸ and constructed metacells⁸⁹ based on them, thus overcoming the difficulties posed by the sparsity of single-cell data for co-expression analysis. Then, we use the function ‘blockwiseConsensusModules’ (corType = ‘pearson’, mergeCutHeight = 0.4, networkType = ‘signed’) to extract the co-expressed gene modules. The smallest power with >0.8 ‘SFT.R.sq’ value is selected for WGCNA. The obtained topological overlap matrix was used to calculate the gene pair connectivity.

In all of the above steps, we used the pycenic⁹⁰ package to calculate the area under the curve (AUC) score, clusterProfiler⁹⁰ to perform GO analysis and the rpy2 package to implement the invocation of R packages. The source code used to reproduce this project can be accessed at https://github.com/ZhaiLab-SUSTech/sc_medicago.

Verification by additional set of biological replicates

For the analysis of biological replicates, we first used the same pipeline to preprocess the Illumina reads and obtained the gene matrix. We then applied the scVI algorithm for data integration and the Leiden algorithm for clustering the integrated dataset. The parameters used were identical to those used in the analysis of replicate 1 only. And the clusters obtained from the combined datasets were renamed on the basis of their similarity to the replicate 1 only clustering (Extended Data Fig. 4b). To identify induced genes specific to each timepoint, we first grouped nuclei from the same replicates, timepoints and clusters together to form the pseudobulk datasets and then used the likelihood-ratio test wrapped in edgeR⁹¹ to identify the genes upregulated at each timepoint when compared with all other timepoints. Only genes with adjusted P < 0.05 and a fold change >2 were retained.

Phylogeny reconstruction of FER and FER-like genes

Genomes of *Aeschynomene evenia*, *Arabidopsis thaliana*, *Bauhinia variegata*, *L. japonicus*, *Lupinus albus*, *M. truncatula*, *Oryza sativa*, *Phaseolus lunatus*, *Populus trichocarpa* and *Vitis vinifera* were selected, representing species ranging from monocots to basal core eudicots to

legumes. Longest transcripts for each gene were then extracted and used in the following analyses. The genome versions of each species are listed in Supplementary Table 4. To correctly identify orthologues of AtFER (AT3G51550.1) from paralogues of AtFER (AtFER-like genes) in *M. truncatula*, we first identified AtFER-like genes by aligning protein sequences of AtFER against all *A. thaliana* proteins with MMseqs2 (ref. 92; v.13.45111). Top ten hits were retained as putative AtFER-like genes. Then, we identified orthologues of both AtFER and AtFER-like genes in other species with reciprocal best hits (RBH) of MMseqs2 search using getRBH.pl (<https://github.com/Computational-conSequences/SequenceTools/blob/master/getRBH.pl>). Protein sequences of the resulting orthologues of AtFER and AtFER-like genes, together with AtFER and AtFER-like genes in *A. thaliana*, were aligned using mafft-linsi⁹³ (v.7.487), which were then converted to codon alignment of nucleotide sequences using pal2nal⁹⁴ (v.14) and used to infer phylogenetic relationship with maximum-likelihood approach using RAxML⁹⁵ (v.8.2.12) based on codon alignment with bootstrap set to 100. Mid-point rooting were performed with FigTree v.1.1.4 (<https://github.com/rambaut/figtree/>) and long branches were cut with TreeShrink⁹⁶ (v.1.3.9) with quantile set to 0.1. Speciation nodes and duplication nodes were identified with duplication-loss-coalescence model of dlcpar⁹⁷ (v.2.0.1) using species topology extracted from Tree of Life 2.0 (ref. 98) with parameter ‘search’. Genes connected via duplication nodes to AtFER were considered as AtFER paralogues and genes connected to AtFER absent of duplication nodes were considered as AtFER orthologues. The final reconciled tree was illustrated with iTOL v.5 (ref. 99).

Yeast two-hybrid and Co-IP assays

CDS of *MtFER* and *LYK3* fragments were ligated into the *pGBKT7GW* and *pGADT7GW* vectors, respectively, and cotransformed into the yeast strain AH109 according to the LiAc transformation method¹⁰⁰. After growing 2 days in a 28 °C incubator, 7 µl of yeast suspension was dropped onto minimal synthetic dropout agar medium lacking leucine/tryptophan (-LT) or leucine/tryptophan/adenine/histidine (-LTAH) containing 30 mM 3-amino-1,2,4-triazole (Clontech). Protein–protein interactions were determined by the growth of yeast on the selected medium. Yeast growth was monitored for up to 7 days. For the Co-IP assays, *pUBGFP-3×FLAG* and *pUBGFP-3×HA* were used to express FLAG-tagged LYK3 and HA-tagged MtFER/NFP proteins, the respective constructs, were transformed into *Agrobacterium tumefaciens* GV3101 via electroporation. *Agrobacterium*-mediated transformation of *N. benthamiana* leaves and Co-IP assays were performed according to ref. 101, except that the desired OD₆₀₀ of all *Agrobacterium* transformants we used in the infiltration medium is OD₆₀₀ = 0.2. Co-immunoprecipitated proteins were detected by western blot analysis using 1:2,000 dilution of anti-FLAG (F1804-1MG, Sigma) or anti-HA (H6908-2ML, Sigma).

In vitro phosphorylation assay

MtFERic lacking autophosphorylation ability by two-site mutations at K570R and D666V (*MtFERic*^{K570R/D666V}) was expressed from pCold TF in the *Escherichia coli* strain Rosetta (TransGen Biotech). LYK3ic^{WT} and kinase dead form LYK3ic^{K349E} were expressed from pGEX-4T-1 in the *E. coli* strain Rosetta (TransGen Biotech). NFP was expressed from pMAL-c2X in the *E. coli* strain Rosetta (TransGen Biotech). Purification of GST-LYK3ic^{WT}, GST-LYK3ic^{K349E}, MBP-NFPic and HIS-MtFERic^{K570R/D666V} was performed according to the protocols of E8200 (New England Biolabs) and BioSprint96 (Qiagen), respectively. The phosphorylation of *MtFERic*^{K570R/D666V} was performed at 30 °C for 30 min in the presence of LYK3ic^{WT} or LYK3ic^{K349E} in a reaction volume of 20 µl containing 50 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 1 mM DTT, 100 µM ATP, 2.5 mCi [γ^{32} P] ATP and 1 µg of purified proteins. The reaction was terminated with 1× loading buffer and subjected to SDS-polyacrylamide gel electrophoresis. Radiolabelled bands were visualized with a Bio-imaging analyser (BAS-2500) (Fujifilm Life Science; <https://www.fujifilm.com>).

RNA extraction and qRT-PCR analysis

Total RNA extraction and first-strand complementary DNA synthesis were performed according to the manufacturer’s protocol of TRI-zol reagent (15596018, Invitrogen) and PrimeScript RT reagent kit (RR047A, TaKaRa), respectively. Quantitative PCR with reverse transcription (qRT-PCR) was performed using the CFX Connect Real-Time PCR Detection System (Bio-Rad) and 2× Hieff qPCR SYBR Green Master Mix (11143ESS0, Yeasen). Gene expression level was normalized against the reference genes *EF-1* or *Histone 2A* for symbiosis and defence marker genes, respectively. Primers are listed in Supplementary Table 3.

Statistical analysis

Statistical analyses for the root and nodule phenotypes and genes expression of *MtFERic* plants were performed using GraphPad Prism 8.0.1 and SPSS statistics 26.0. Dot plots were used to show individual datapoints whenever possible. Statistically significant differences between the control and experimental groups were determined by two-tailed Student’s *t*-tests and one-way analysis of variance (ANOVA) followed by Duncan’s multiple range test (*P* < 0.05).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing data generated in this study were deposited in China National Center for Bioinformation with accession PRJCA011245. Source data are provided with this paper.

Code availability

The source code to reproduce this project can be accessed at https://github.com/ZhaiLab-SUSTech/sc_medicago.

References

1. Nutman, P. S. Centenary lecture. *Phil. Trans. R. Soc. B* **317**, 69–106 (1987).
2. Kelly, S., Radutoiu, S. & Stougaard, J. Legume LysM receptors mediate symbiotic and pathogenic signalling. *Curr. Opin. Plant Biol.* **39**, 152–158 (2017).
3. Roy, S. et al. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *Plant Cell* **32**, 15–41 (2020).
4. Zipfel, C. & Oldroyd, G. E. Plant signalling in symbiosis and immunity. *Nature* **543**, 328–336 (2017).
5. Antolín-Llovera, M., Ried, M. K., Binder, A. & Parniske, M. Receptor kinase signaling pathways in plant–microbe interactions. *Annu. Rev. Phytopathol.* **50**, 451–473 (2012).
6. Fields, S. Global nitrogen: cycling out of control. *Environ. Health Perspect.* **112**, A556–A563 (2004).
7. Oldroyd, G. E. Speak, friend and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**, 252–263 (2013).
8. Long, S. R. *Rhizobium* symbiosis: nod factors in perspective. *Plant Cell* **8**, 1885–1898 (1996).
9. Arrighi, J.-F. et al. The *Medicago truncatula* lysine motif-receptor-like kinase gene family includes NFP and new nodule-expressed genes. *Plant Physiol.* **142**, 265–279 (2006).
10. Smit, P. et al. *Medicago LYK3*, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiol.* **145**, 183–191 (2007).
11. Limpens, E. et al. LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* **302**, 630–633 (2003).
12. Madsen, E. B. et al. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**, 637–640 (2003).

13. Radutoiu, S. et al. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**, 585–592 (2003).
14. Yang, J. et al. Mechanisms underlying legume–rhizobium symbioses. *J. Integr. Plant Biol.* **64**, 244–267 (2022).
15. Timmers, A. C., Auriac, M.-C. & Truchet, G. Refined analysis of early symbiotic steps of the *Rhizobium–Medicago* interaction in relationship with microtubular cytoskeleton rearrangements. *Development* **126**, 3617–3628 (1999).
16. Breakspear, A. et al. The root hair ‘infectome’ of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for auxin signaling in rhizobial infection. *Plant Cell* **26**, 4680–4701 (2014).
17. Larainzar, E. et al. Deep sequencing of the *Medicago truncatula* root transcriptome reveals a massive and early interaction between nodulation factor and ethylene signals. *Plant Physiol.* **169**, 233–265 (2015).
18. Damiani, I. et al. Nod factor effects on root hair-specific transcriptome of *Medicago truncatula*: focus on plasma membrane transport systems and reactive oxygen species networks. *Front. Plant Sci.* **7**, 794 (2016).
19. Jardinaud, M.-F. et al. A laser dissection-RNAseq analysis highlights the activation of cytokinin pathways by Nod factors in the *Medicago truncatula* root epidermis. *Plant Physiol.* **171**, 2256–2276 (2016).
20. Pecrix, Y. et al. Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nat. Plants* **4**, 1017–1025 (2018).
21. Lohar, D. P. et al. Transcript analysis of early nodulation events in *Medicago truncatula*. *Plant Physiol.* **140**, 221–234 (2005).
22. Rich-Griffin, C. et al. Single-cell transcriptomics: a high-resolution avenue for plant functional genomics. *Trends Plant Sci.* **25**, 186–197 (2020).
23. Libault, M., Pingault, L., Zogli, P. & Schiefelbein, J. Plant systems biology at the single-cell level. *Trends Plant Sci.* **22**, 949–960 (2017).
24. Denyer, T. et al. Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* **48**, 840–852 (2019).
25. Birnbaum, K. et al. A gene expression map of the *Arabidopsis* root. *Science* **302**, 1956–1960 (2003).
26. Farmer, A., Thibivilliers, S., Ryu, K. H., Schiefelbein, J. & Libault, M. Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in *Arabidopsis* roots at the single-cell level. *Mol. Plant* **14**, 372–383 (2021).
27. Sunaga-Franze, D. Y. et al. Single-nucleus RNA sequencing of plant tissues using a nanowell-based system. *Plant J.* **108**, 859–869 (2021).
28. Neumann, M. et al. A 3D gene expression atlas of the floral meristem based on spatial reconstruction of single nucleus RNA sequencing data. *Nat. Commun.* **13**, 2838 (2022).
29. Long, Y. et al. FlsnRNA-seq: protoplasting-free full-length single-nucleus RNA profiling in plants. *Genome Biol.* **22**, 66 (2021).
30. Cervantes-Pérez, S. A. et al. Cell-specific pathways recruited for symbiotic nodulation in the *Medicago truncatula* legume. *Mol. Plant* **15**, 1868–1888 (2022).
31. Biała, W., Banasiak, J., Jarzyniak, K., Pawela, A. & Jasiński, M. *Medicago truncatula* ABCG10 is a transporter of 4-coumarate and liquiritigenin in the medicarpin biosynthetic pathway. *J. Exp. Bot.* **68**, 3231–3241 (2017).
32. Liu, J. et al. Closely related members of the *Medicago truncatula* PHT1 phosphate transporter gene family encode phosphate transporters with distinct biochemical activities. *J. Biol. Chem.* **283**, 24673–24681 (2008).
33. De Carvalho-Niebel, F., Timmers, A. C., Chabaud, M., Defaux-Petas, A. & Barker, D. G. The Nod factor-elicited annexin MtAnn1 is preferentially localised at the nuclear periphery in symbiotically activated root tissues of *Medicago truncatula*. *Plant J.* **32**, 343–352 (2002).
34. Xiao, T. T. et al. Fate map of *Medicago truncatula* root nodules. *Development* **141**, 3517–3528 (2014).
35. Chiou, T. J., Liu, H. & Harrison, M. J. The spatial expression patterns of a phosphate transporter (MtPT1) from *Medicago truncatula* indicate a role in phosphate transport at the root/soil interface. *Plant J.* **25**, 281–293 (2001).
36. Cerri, M. R. et al. *Medicago truncatula* ERN transcription factors: regulatory interplay with NSP1/NSP2 GRAS factors and expression dynamics throughout rhizobial infection. *Plant Physiol.* **160**, 2155–2172 (2012).
37. Mbengue, M. et al. The *Medicago truncatula* E3 ubiquitin ligase PUB1 interacts with the LYK3 symbiotic receptor and negatively regulates infection and nodulation. *Plant Cell* **22**, 3474–3488 (2010).
38. Franssen, H. J. et al. Root developmental programs shape the *Medicago truncatula* nodule meristem. *Development* **142**, 2941–2950 (2015).
39. Laffont, C. et al. MtNRLK1, a CLAVATA1-like leucine-rich repeat receptor-like kinase upregulated during nodulation in *Medicago truncatula*. *Sci. Rep.* **8**, 2046 (2018).
40. Nguyen, N. N. et al. PHO1 family members transport phosphate from infected nodule cells to bacteroids in *Medicago truncatula*. *Plant Physiol.* **185**, 196–209 (2021).
41. Roy, S. et al. MtLAX2, a functional homologue of the *Arabidopsis* auxin influx transporter AUX1, is required for nodule organogenesis. *Plant Physiol.* **174**, 326–338 (2017).
42. Wendrich, J. R. et al. Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* **370**, eaay4970 (2020).
43. Wang, C. et al. *Lotus japonicus* clathrin heavy Chain1 is associated with Rho-Like GTPase ROP6 and involved in nodule formation. *Plant Physiol.* **167**, 1497–1510 (2015).
44. Silady, R. A. et al. The GRV2/RME-8 protein of *Arabidopsis* functions in the late endocytic pathway and is required for vacuolar membrane flow. *Plant J.* **53**, 29–41 (2008).
45. Robinson, M. S. Adaptable adaptors for coated vesicles. *Trends Cell Biol.* **14**, 167–174 (2004).
46. Barois, N. & Bakke, O. The adaptor protein AP-4 as a component of the clathrin coat machinery: a morphological study. *Biochem. J.* **385**, 503–510 (2005).
47. Timmers, A., Auriac, M.-C., de Billy, F. & Truchet, G. Nod factor internalization and microtubular cytoskeleton changes occur concomitantly during nodule differentiation in alfalfa. *Development* **125**, 339–349 (1998).
48. Oldroyd, G. E. & Downie, J. A. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu. Rev. Plant Biol.* **59**, 519–546 (2008).
49. Lin, J., Frank, M. & Reid, D. No home without hormones: how plant hormones control legume nodule organogenesis. *Plant Commun.* **1**, 100104 (2020).
50. Wang, K. L.-C., Li, H. & Ecker, J. R. Ethylene biosynthesis and signaling networks. *Plant Cell* **14**, S131–S151 (2002).
51. Cai, J. et al. Role of the Nod factor hydrolase MtNFH1 in regulating Nod factor levels during rhizobial infection and in mature nodules of *Medicago truncatula*. *Plant Cell* **30**, 397–414 (2018).
52. Arrighi, J.-F. et al. The RPG gene of *Medicago truncatula* controls Rhizobium-directed polar growth during infection. *Proc. Natl Acad. Sci. USA* **105**, 9817–9822 (2008).
53. del Campillo, E., Gaddam, S., Mettle-Amuah, D. & Heneks, J. A tale of two tissues: AtGH9C1 is an endo- β -1, 4-glucanase involved in root hair and endosperm development in *Arabidopsis*. *PLoS ONE* **7**, e49363 (2012).
54. Camps, C. et al. Combined genetic and transcriptomic analysis reveals three major signalling pathways activated by Myc-LCOs in *Medicago truncatula*. *New Phytol.* **208**, 224–240 (2015).

55. Maillet, F. et al. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **469**, 58–63 (2011).
56. Kuppusamy, K. T. et al. LIN, a *Medicago truncatula* gene required for nodule differentiation and persistence of rhizobial infections. *Plant Physiol.* **136**, 3682–3691 (2004).
57. Malolepszy, A. et al. A plant chitinase controls cortical infection thread progression and nitrogen-fixing symbiosis. *eLife* **7**, e38874 (2018).
58. Rival, P. et al. Epidermal and cortical roles of NFP and DMI3 in coordinating early steps of nodulation in *Medicago truncatula*. *Development* **139**, 3383–3391 (2012).
59. Yano, K. et al. CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proc. Natl Acad. Sci. USA* **105**, 20540–20545 (2008).
60. Murray, J. D. et al. Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*. *Plant J.* **65**, 244–252 (2011).
61. Gao, Q. et al. A receptor-channel trio conducts Ca^{2+} signalling for pollen tube reception. *Nature* **607**, 534–539 (2022).
62. Duan, Q., Kita, D., Li, C., Cheung, A. Y. & Wu, H.-M. FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development. *Proc. Natl Acad. Sci. USA* **107**, 17821–17826 (2010).
63. Zhu, S. et al. The RALF1-FERONIA complex phosphorylates eIF4E1 to promote protein synthesis and polar root hair growth. *Mol. Plant* **13**, 698–716 (2020).
64. Guo, H. et al. FERONIA receptor kinase contributes to plant immunity by suppressing jasmonic acid signaling in *Arabidopsis thaliana*. *Curr. Biol.* **28**, 3316–3324 (2018).
65. Procko, C. et al. Leaf cell-specific and single-cell transcriptional profiling reveals a role for the palisade layer in UV light protection. *Plant Cell* **34**, 3261–3279 (2022).
66. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
67. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
68. Grindberg, R. V. et al. RNA-sequencing from single nuclei. *Proc. Natl Acad. Sci. USA* **110**, 19802–19807 (2013).
69. Thrupp, N. et al. Single-nucleus RNA-seq is not suitable for detection of microglial activation genes in humans. *Cell Rep.* **32**, 108189 (2020).
70. Zhang, T.-Q., Chen, Y., Liu, Y., Lin, W.-H. & Wang, J.-W. Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat. Commun.* **12**, 2053 (2021).
71. Czaja, L. F. et al. Transcriptional responses toward diffusible signals from symbiotic microbes reveal MtNFP-and MtDMI3-dependent reprogramming of host gene expression by arbuscular mycorrhizal fungal lipochitooligosaccharides. *Plant Physiol.* **159**, 1671–1685 (2012).
72. Mithöfer, A. Suppression of plant defence in rhizobia-legume symbiosis. *Trends Plant Sci.* **7**, 440–444 (2002).
73. Stegmann, M. et al. The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science* **355**, 287–289 (2017).
74. Kessler, S. A. et al. Conserved molecular components for pollen tube reception and fungal invasion. *Science* **330**, 968–971 (2010).
75. Wang, M. et al. Phosphorylation of MtRopGEF2 by LYK3 mediates MtROP activity to regulate rhizobial infection in *Medicago truncatula*. *J. Integr. Plant Biol.* **63**, 1787–1800 (2021).
76. Gao, J.-P. et al. Nod factor receptor complex phosphorylates GmGEF2 to stimulate ROP signaling during nodulation. *Curr. Biol.* **31**, 3538–3550 (2021).
77. Kohring, B., Baier, R., Niehaus, K., Puhler, A. & Flaschel, E. Production of nodulation factors by *Rhizobium meliloti*: fermentation, purification and characterization of glycolipids. *Glycoconj. J.* **14**, 963–971 (1997).
78. Boisson-Dernier, A. et al. *Agrobacterium rhizogenes*-transformed roots of *Medicago truncatula* for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Mol. Plant Microbe Interact.* **14**, 695–700 (2001).
79. Boivin, C., Camut, S., Malpica, C. A., Truchet, G. & Rosenberg, C. *Rhizobium meliloti* genes encoding catabolism of trigonelline are induced under symbiotic conditions. *Plant Cell* **2**, 1157–1170 (1990).
80. Thibivilliers, S., Anderson, D. & Libault, M. Isolation of plant root nuclei for single cell RNA sequencing. *Curr. Protoc. Plant Biol.* **5**, e20120 (2020).
81. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
82. Germain, P.-L., Lun, A., Meixide, C. G., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research* **10**, 979 (2021).
83. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
84. Xu, C. et al. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).
85. Timshel, P. N., Thompson, J. J. & Pers, T. H. Genetic mapping of etiologic brain cell types for obesity. *eLife* **9**, e55851 (2020).
86. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
87. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
88. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
89. Baran, Y. et al. MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol.* **20**, 206 (2019).
90. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
91. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
92. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
93. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
94. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
95. Stamatakis, A. Using RAxML to infer phylogenies. *Curr. Protoc. Bioinformatics* **51**, 6.14.1–6.14.14 (2015).
96. Mai, U. & Mirarab, S. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* **19**, 272 (2018).
97. Du, H. et al. Multiple optimal reconciliations under the duplication-loss-coalescence model. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **18**, 2144–2156 (2019).
98. Baker, W. J. et al. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst. Biol.* **71**, 301–319 (2022).
99. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

100. Daniel Gietz, R. & Woods, R. A. in *Methods in Enzymology* Vol. 350 (eds Guthrie, C. & Fink, G. R.) 87–96 (Academic Press, 2002).
101. Dong, W. et al. An SHR-SCR module specifies legume cortical cell fate to enable nodulation. *Nature* **589**, 586–590 (2021).

Acknowledgements

We thank J. Murray and H. Liu for helpful suggestions on the manuscript. This work is supported by the National Key R&D Program of China (grant nos. 2019YFA0903903 and 2019YFA0904703); National Natural Science Foundation of China (32070270, 32050081, 32088102 and 31825003); the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2016ZT06S172); the Shenzhen Sci-Tech Fund (KYTDPT20181011104005); the Key Laboratory of Molecular Design for Plant Cell Factory of Guangdong Higher Education Institutes (2019KSYS006); CAS Project for Young Scientists in Basic Research (YSBR-011); and the Stable Support Plan Program of Shenzhen Natural Science Fund (20200925153345004).

Author contributions

E.W., J.Z., Z.L. and J.Y. directed the research. Z.L., J.Y., Y.L. and C.Z. performed most of the experiments and analysis. D.W., X.Z., W.D., L.Z. and C.L. contributed to the analytical, molecular cloning and transformation work. E.W. and J.Z. oversaw the entire study. Z.L., J.Y., Y.L., E.W. and J.Z. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41477-023-01524-8>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41477-023-01524-8>.

Correspondence and requests for materials should be addressed to Jixian Zhai or Ertao Wang.

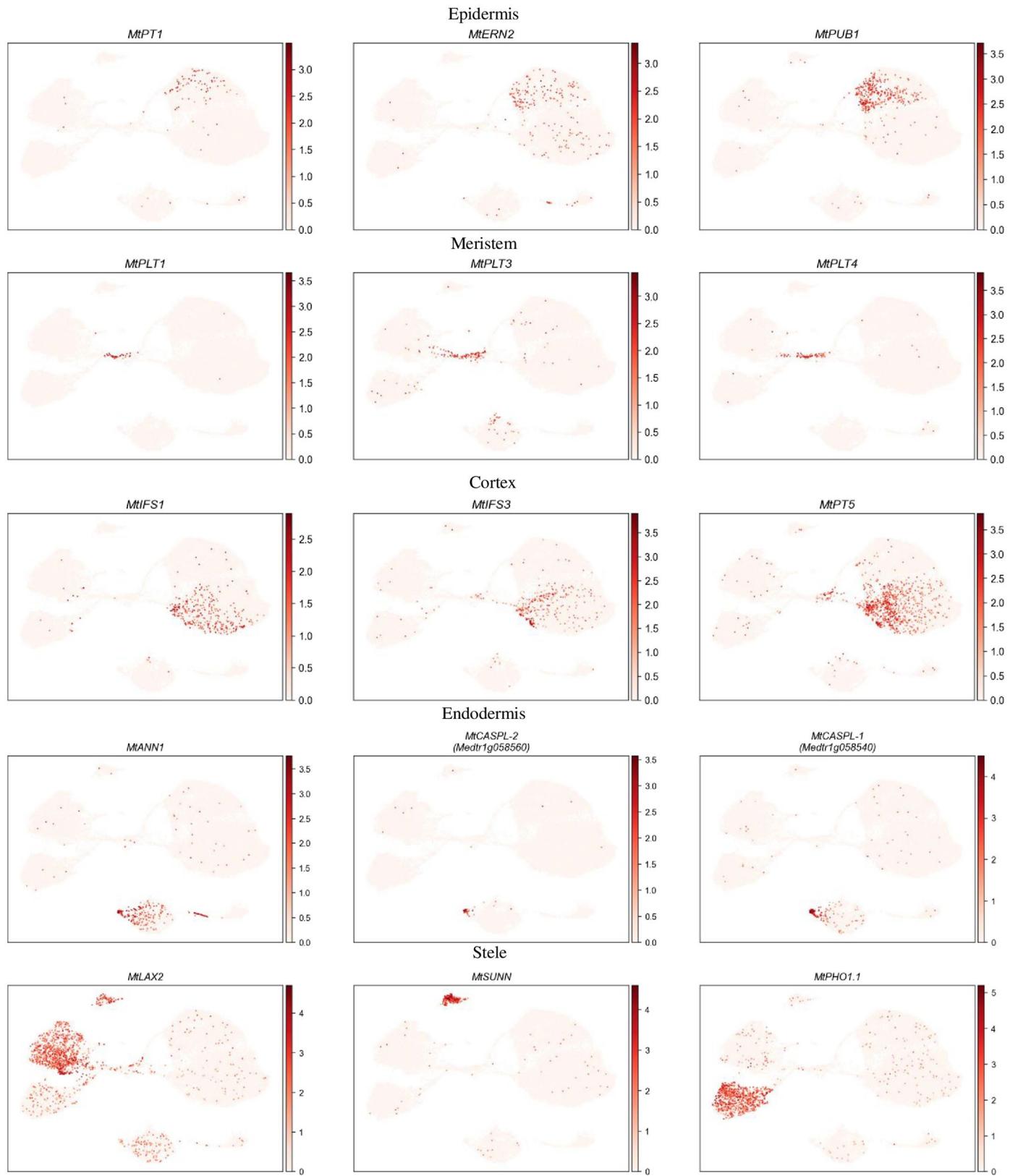
Peer review information *Nature Plants* thanks Oswaldo Valdes-Lopez, Maria Eugenia Zanetti and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

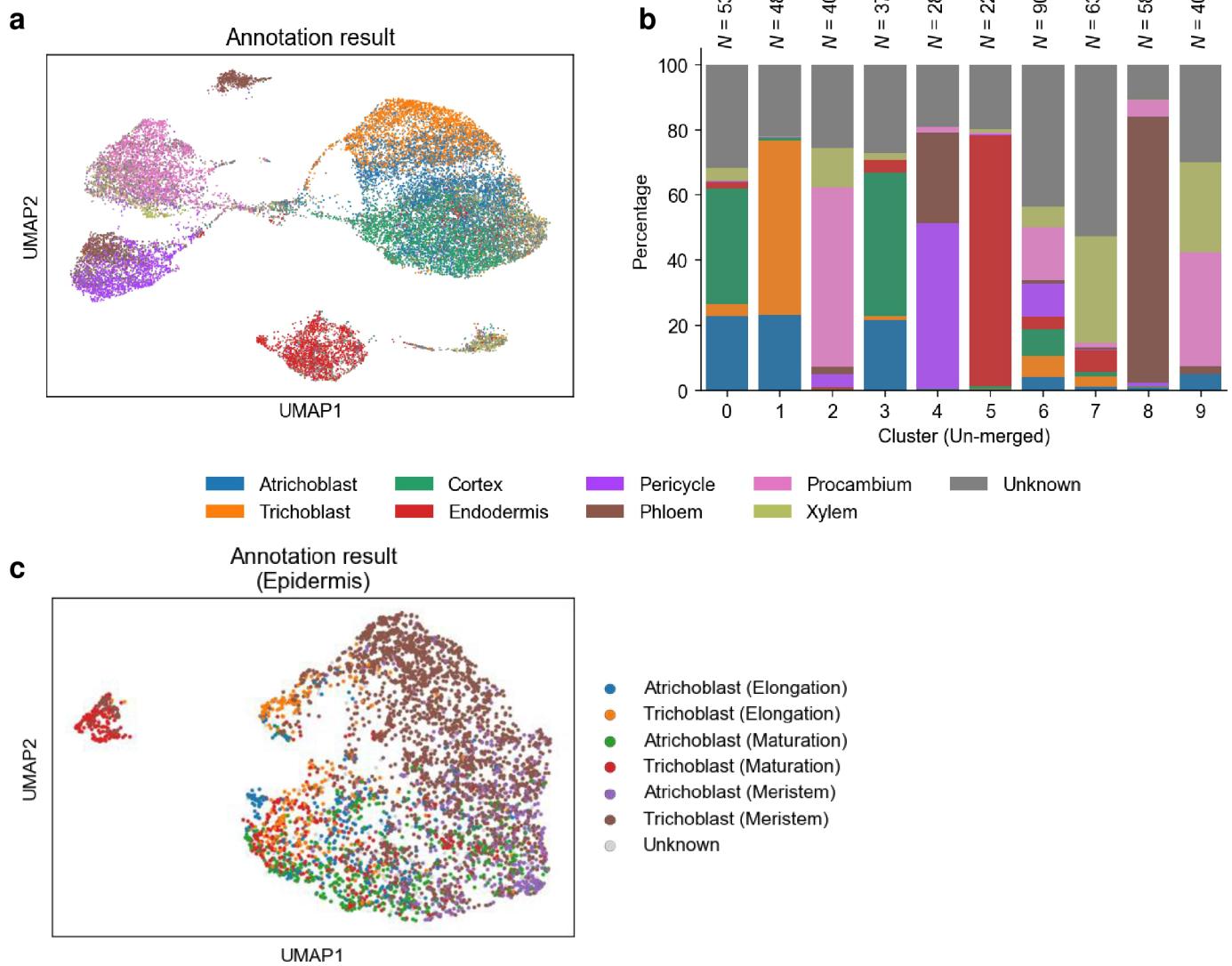
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

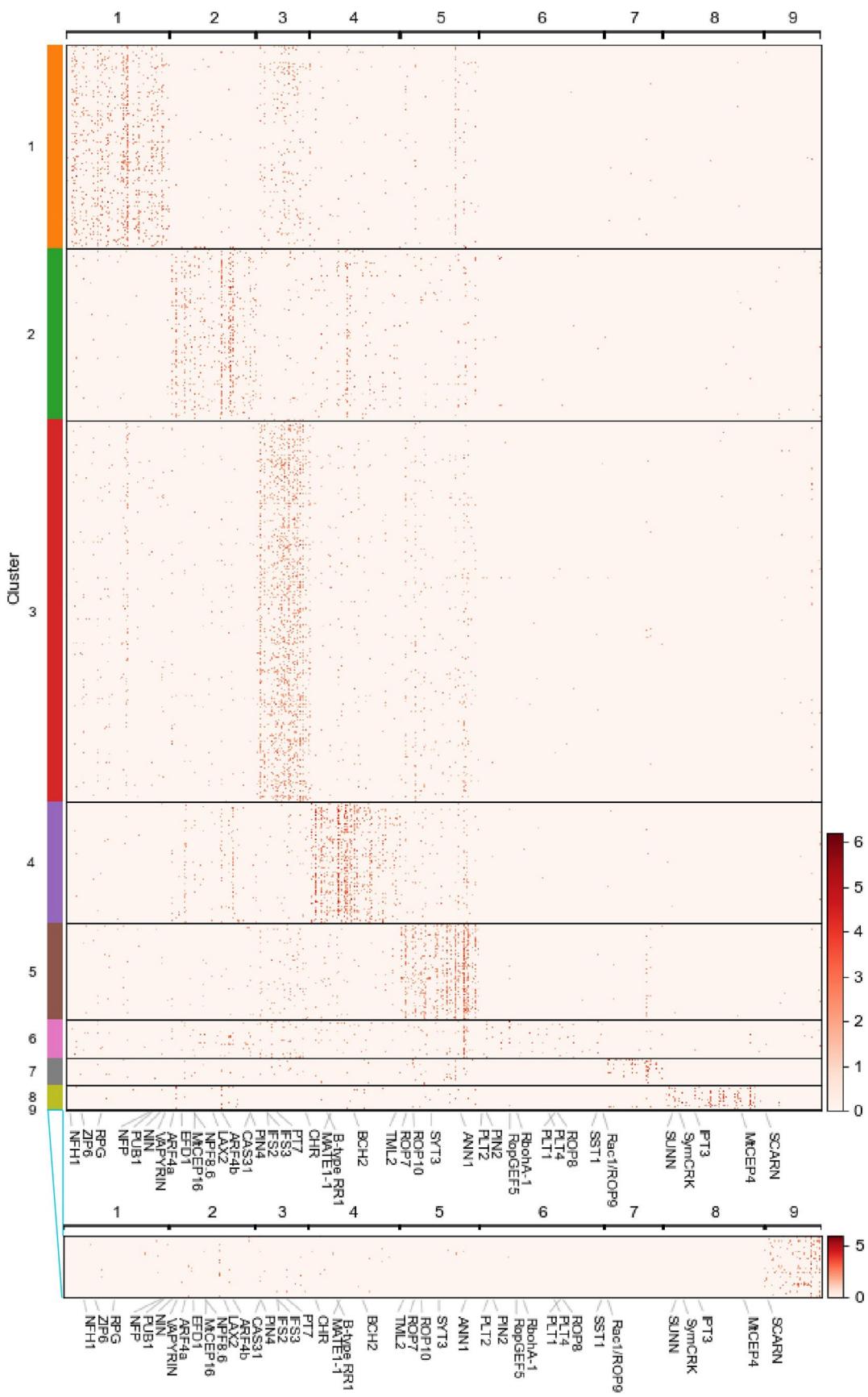
© The Author(s), under exclusive licence to Springer Nature Limited 2023



Extended Data Fig. 1 | UMAP visualizations of known cell type-specific marker genes. For each cell type, three genes are used to annotate its corresponding clusters. The gene locus and references for each gene can be found in Supplementary Table 1. This figure is related with Fig. 2d.



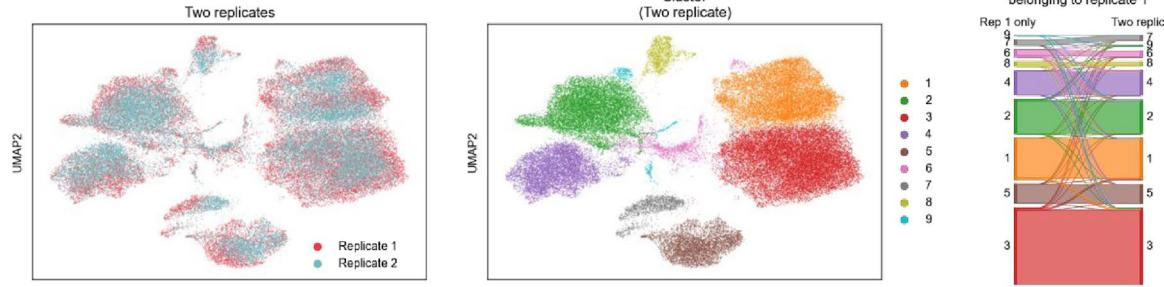
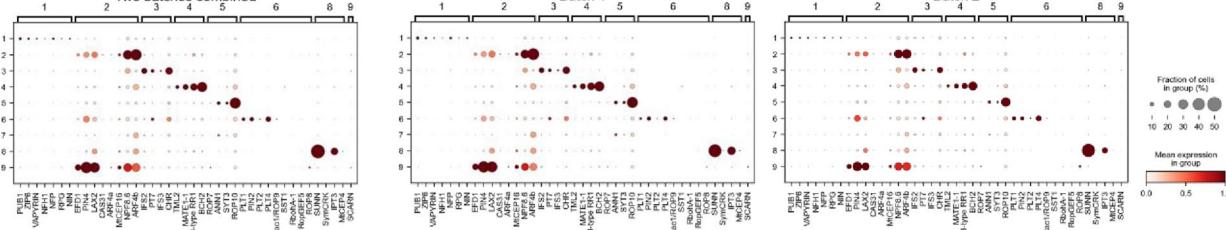
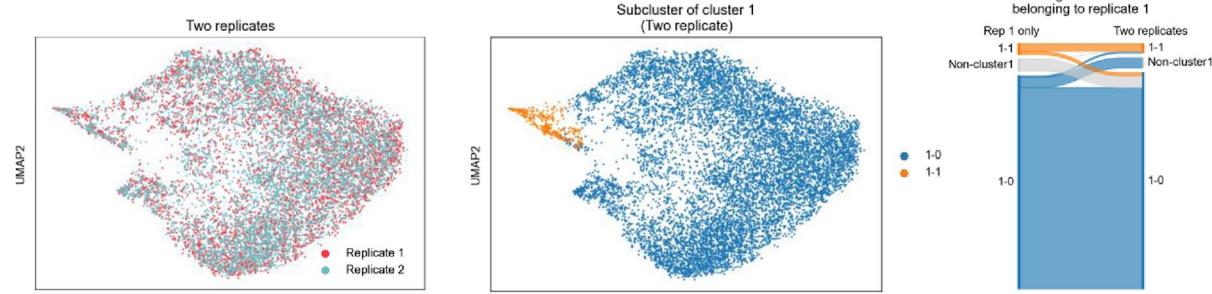
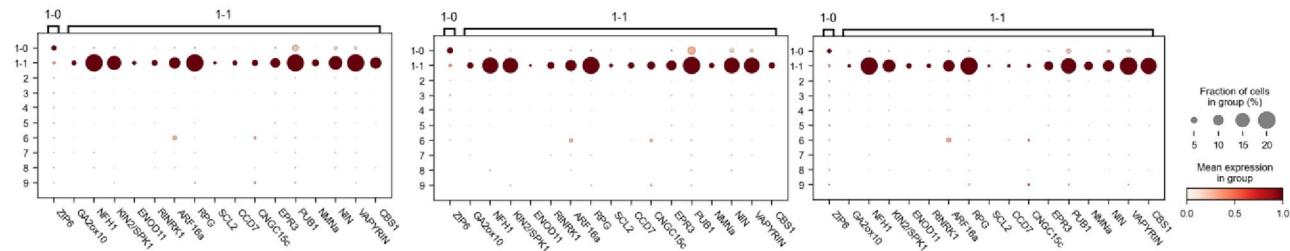
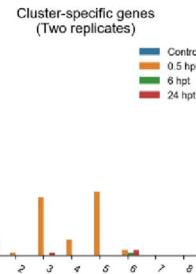
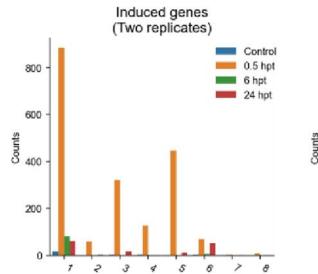
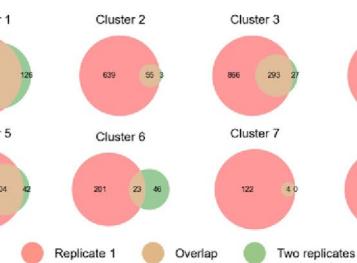
Extended Data Fig. 2 | Label transfer results from *Arabidopsis* single-cell datasets by scANVI. **a.** UMAP visualization of transferred annotation. **b.** The bar plots represent the predicted percentage of cells of different cell types in each cell cluster. N indicates the cell number. **c.** The predicted results for epidermal cells.



Extended Data Fig. 3 | Heatmap representing the expression pattern of cluster-specific genes for each cluster. The bottom panel shows a zoomed-in view of the gene expression pattern in cluster 9. The full genes list can be obtained in Supplementary Data. 2. This figure is related with Fig. 2e.

a

	Replicate 1			Replicate 2				
	Control	0.5 hpt	6 hpt	24 hpt	Control	0.5 hpt	6 hpt	24 hpt
Nuclei counts (After QC)	6292	6088	6867	6035	6378	7673	5501	10855
Number of genes detected	26663	26867	26220	25782	25434	26391	25882	27182
median genes/nucleus	1083	1087	953	965	694	780	792	827
mean genes/nucleus	1188	1201	1065	1072	815	888	892	910
median UMIs/nucleus	1483	1520	1308	1298	836	988	1012	1067
mean UMIs/nucleus	1707	1779	1520	1518	1040	1174	1201	1226

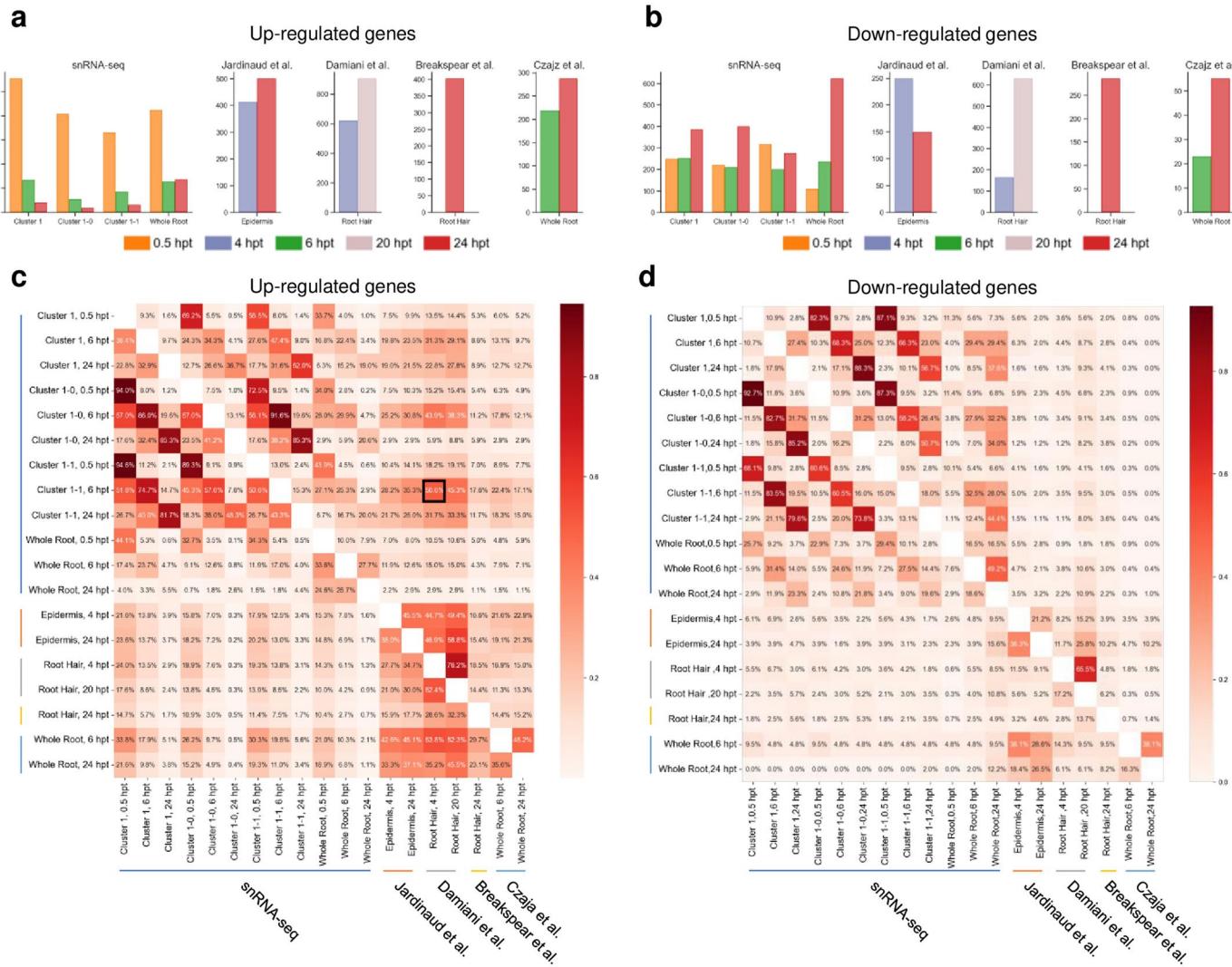
b**c****d****e****f****g**

Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Data validation using a second biological replicate.

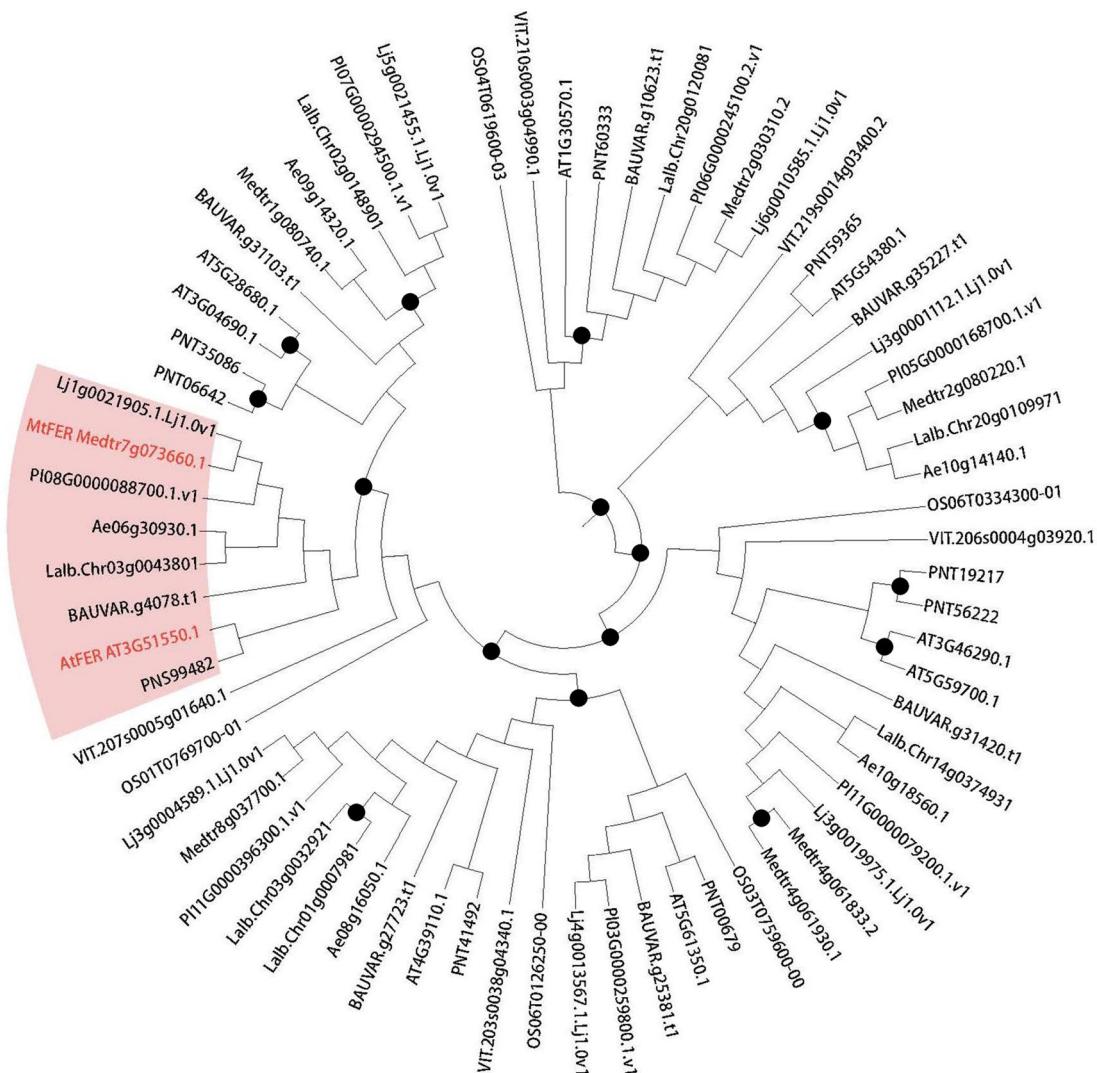
a. Quality control information for the snRNA-seq dataset. **b.** UMAP visualization of data integration and clustering results. The right panel shows the change in nuclei belonging to the replicate 1 between the clustering results of replicate 1 only analysis and the two-replicates analysis. The sequencing data from the second replicate were preprocessed using the same pipeline as the first replicate. We then applied the scVI algorithm for data integration and the Leiden algorithm for clustering the integrated dataset. The parameters used were identical to those used in the analysis of replicate 1 only. Then the clusters obtained from the combined datasets were renamed based on their similarity to the replicate 1 only clustering. **c.** Dot plot of partial SNF genes with cluster-specific expression patterns. The genes are identical to those shown in Fig. 1e. **d.** Left: UMAP visualization of the epidermis subcluster. Right: The change in

nuclei belonging to replicates between the replicate 1 only clustering and the combined clustering of two replicates. **e.** Dot plot of partial SNF genes with cluster-specific expression patterns. The genes are identical to those shown in Fig. 3e. **f.** The number timepoint-specific genes identified from the NF-treatment time course with biological replicates. Left, all genes. Right, timepoint-specific genes with cluster-specific expression patterns (that is spatiotemporal-specific genes). To identify genes specific to each timepoint, we first grouped nuclei from the same replicates, timepoints and clusters together to form the pseudobulk datasets and then used the likelihood-ratio test wrapped in edgeR to perform the differential expression analysis. Only genes with adjusted p-values less than 0.05 and a fold change greater than 2 were retained. The full gene list is provided in Supplementary Data. 5. **g.** The overlap with the timepoint-specific genes identified in replicate 1.



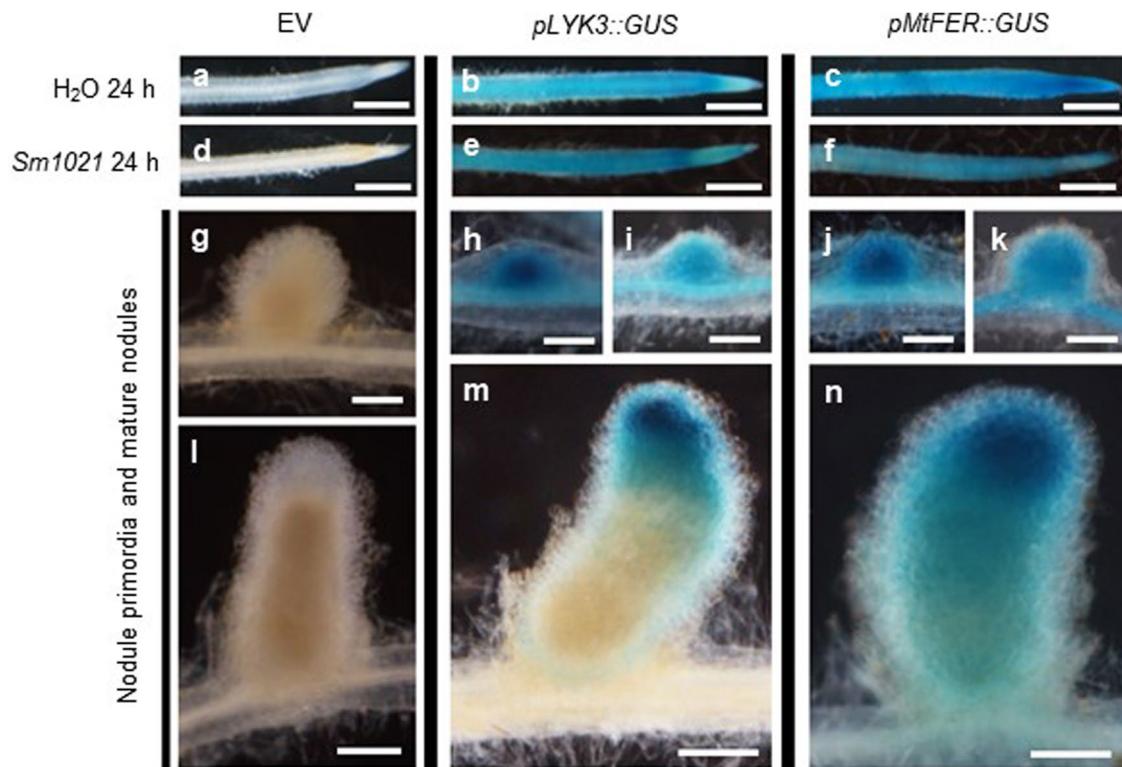
Extended Data Fig. 5 | Comparison of genes differentially expressed in response to NF treatment for snRNA-seq data and public single-cell-type transcriptome datasets. We identified DEGs in snRNA-seq data by comparison with the control sample. Whole roots DEGs were obtained by directly comparing gene expression in all nuclei at a given inoculation timepoint vs the control, rather than in a particular cluster. We used the two-sided Wald test implemented in `diffxpy` to identify DEGs and the full list of DEGs is provided in supplementary dataset 4. The p-values for publicly available data are obtained from their original publications. To make the results comparable, we used the following thresholds:

fold change > 2 and adjusted p-value < 0.05. **a, b.** The counts of upregulated genes (a) and downregulated genes (b) in different studies. **c, d.** Pairwise comparison of upregulated genes (c) and downregulated genes (d) identified by different studies. The colour represents the percentage of DEGs identified in the data corresponding to the row that were also identified in the data corresponding to the column. For example, the black box represents the 50.6% of upregulated expression genes identified in the snRNA-seq that were also upregulated in the Damiani et al.'s data.



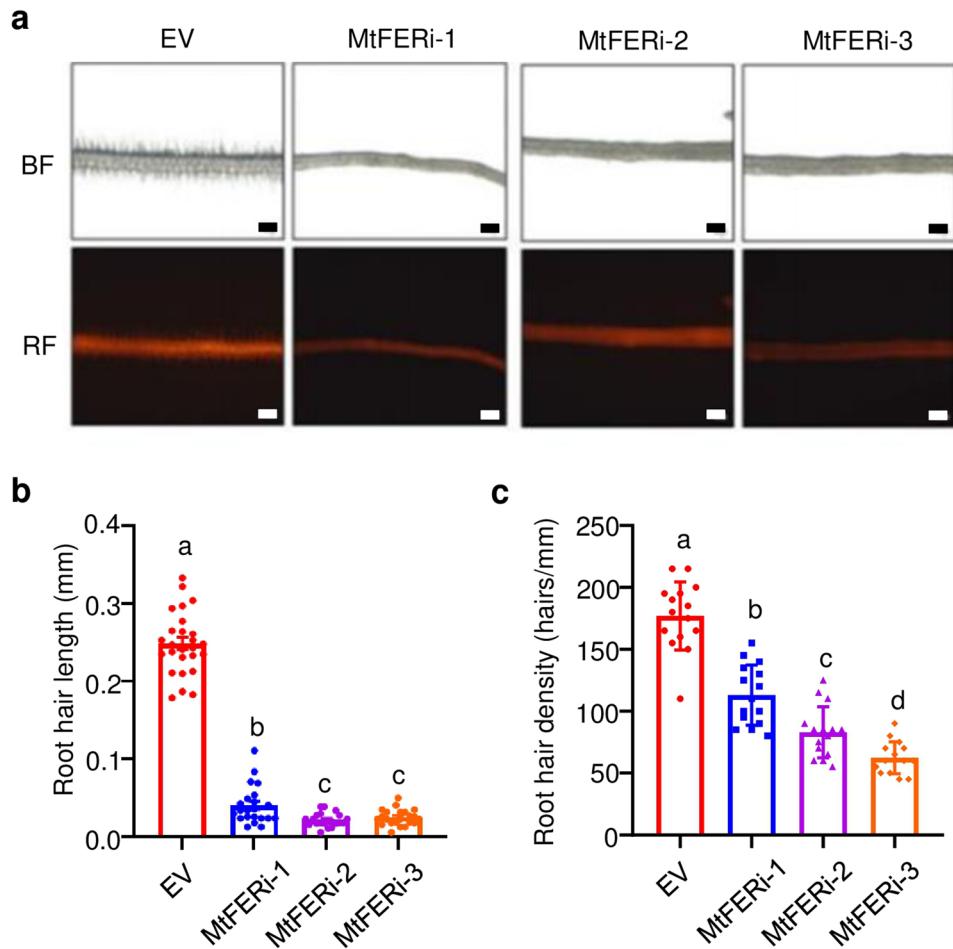
Extended Data Fig. 6 | Phylogeny of FER and FER-like genes. Genomes of *Aeschynomene evenia*, *Arabidopsis thaliana*, *Bauhinia variegata*, *Lotus japonicus*, *Lupinus albus*, *Medicago truncatula*, *Oryza sativa*, *Phaseolus lunatus*, *Populus trichocarpa* and *Vitis vinifera* were selected, representing species ranging from monocots, basal core eudicots, to legumes. Protein sequences of the orthologs of AtFER (AT3G51550.1) and AtFER-like genes were aligned using mafft-linsi, which were then converted to codon alignments of nucleotide sequences using pal2nal and used to infer phylogenetic relationship with maximum-likelihood approach using RAxML with bootstrap set to 100. Midpoint rooting were performed with

FigTree and long branches were cut with **TreeShrink** with quantile set to 0.1. Speciation nodes and duplication nodes were identified with Duplication-Loss-Coalescence Model of **dlcpar** using species topology extracted from Tree of Life 2.0 with parameter “search”. Genes connected via duplication nodes to AtFER were considered as AtFER paralogs and genes connected to AtFER absent of duplication nodes were considered as AtFER orthologs. The final reconciled tree was illustrated with iTOL. Duplication nodes were marked with black dots. The FER clade is highlighted in light red and AtFER and MtFER (*Medtr7g_073660.1*) are highlighted in red and bold.



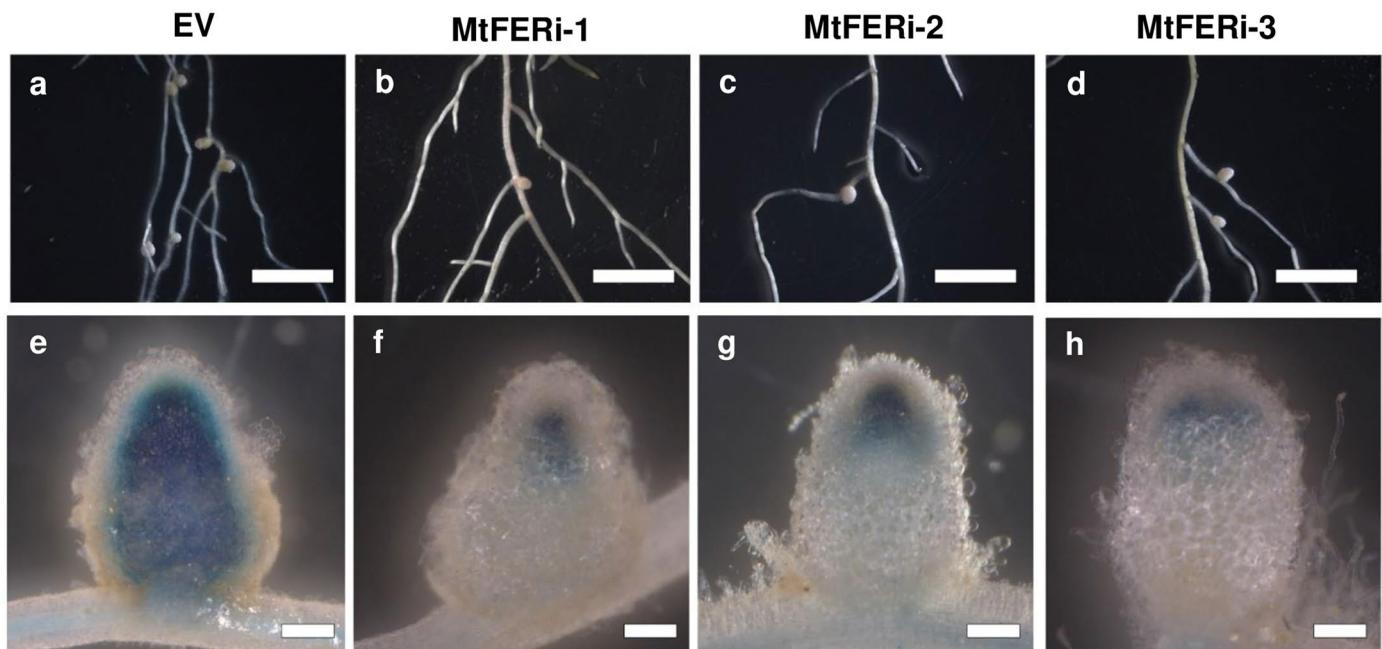
Extended Data Fig. 7 | *pLYK3::GUS* and *pMtFER::GUS* show similar expression patterns in nodules inoculated with *Sm1021*. *pLYK3::GUS* (b and e) and *pMtFER::GUS* (c and f) show similar expression patterns of roots 24 h after inoculation with *Sm1021* when compared with EV (a and d). *pLYK3::GUS* (h, i and

m) and *pMtFER::GUS* (j, k and n) also show similar expression patterns in nodule primordia and mature nodules when compared with EV (g and l). Scale bars, 2 mm (a–f) and 200 µm (g–n). Experiments in a–n were independently repeated three times with similar results.



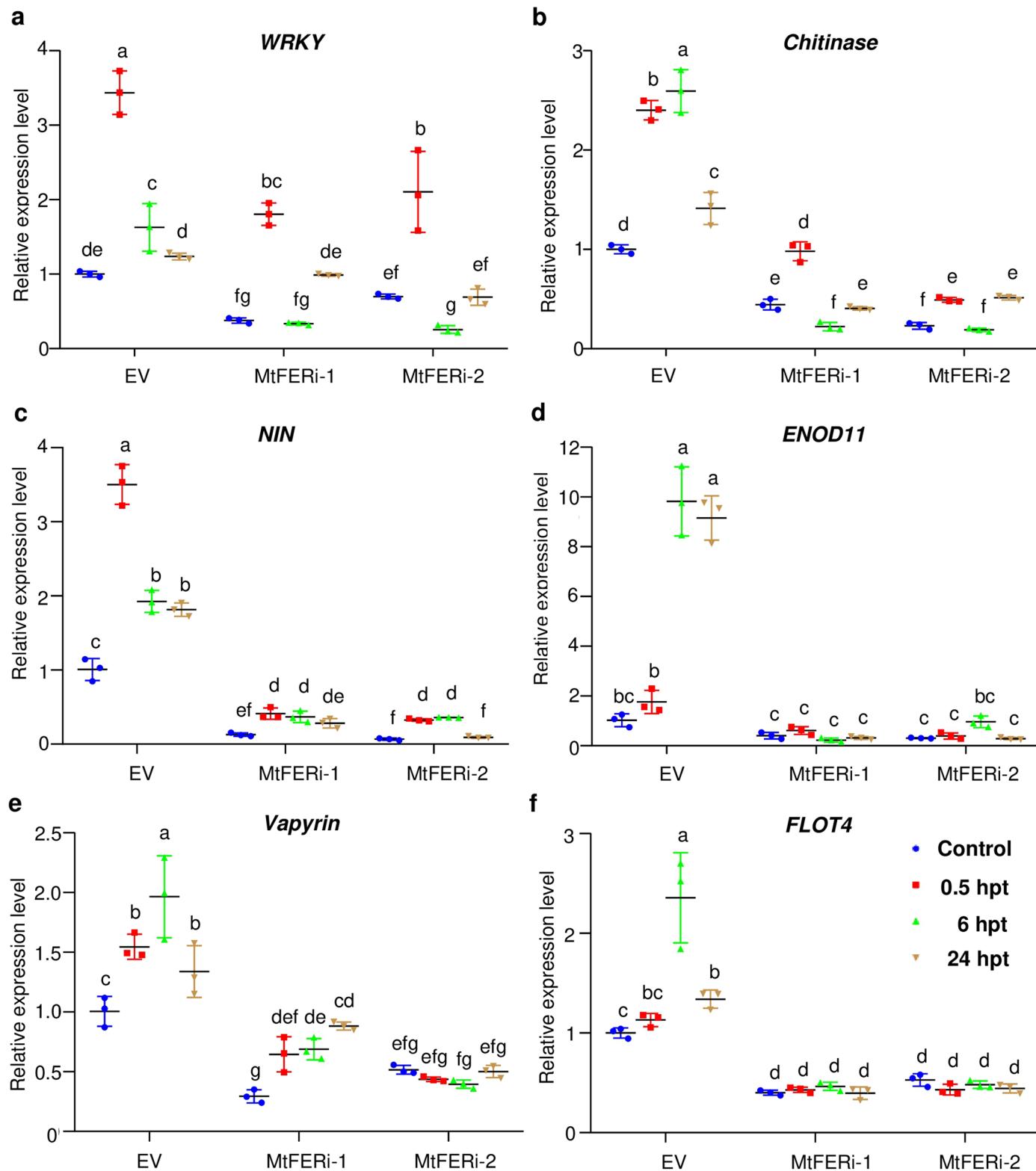
Extended Data Fig. 8 | Suppression of *MtFER* expression inhibits root hair growth. **a.** Representative photographs of root hair phenotype at 21 dpi in EV and MtFERi-1/-2/-3. BF, bright field. Scale bars, 100 μ m. **b.** Quantification of root hair length in EV and MtFERi-1/-2/-3 transgenic hairy roots (EV, n = 26; MtFERi-1/-3, n = 22; MtFERi-2, n = 21). Data are mean \pm SD. **c.** Root hair density of EV and MtFERi-1/-2/-3 transgenic hairy roots (EV, n = 15; MtFERi-1/-2/-3, n = 15). Data are

mean \pm SD. Experiments in b and c were independently repeated three times with similar results. Statistically significant differences between EV and MtFERi-1/-2/-3 groups in experiments b and c were determined by one-way ANOVA followed by Duncan's multiple range tests ($p < 0.05$), different letters indicate significant difference. The exact p values of Duncan's multiple range tests can be found in Supplementary Data 8.



Extended Data Fig. 9 | Representative roots and nodules from EV and MtFERi hairy roots. Representative photographs of roots and nodules from EV (a, e) and MtFERi hairy roots (b–d and f–h) at 21 dpi with *Sm1021* expressing the *LacZ* gene.

Rhizobia in the nodules (e–h) show blue colour when stained by X-Gal. Scale bars, 1 cm (a–d) and 200 µm (e–h). Experiments in e–h were independently repeated three times with similar results.



Extended Data Fig. 10 | Expression pattern of defence and symbiosis marker genes in EV and MtFERi hairy roots after NF treatment. Relative expression levels of *WRKY* (a), *Chitinase* (b), *NIN* (c), *ENOD11* (d), *Vapyrin* (e) and *FLOT4* (f) in EV and MtFERi-1/2 ($n \geq 15$) after NFs treatment. Expression levels of defence and symbiosis marker genes were normalized against the reference gene *Histone 2A* and *EF-1*, respectively. Data are mean \pm SD. Experiments were

repeated three times with similar results. Different letters indicate significant difference [Statistically significant difference between control and experimental groups were determined by one-way ANOVA (Duncan's multiple range tests; $p < 0.05$)]. The exact p values of Duncan's multiple range tests can be found in Supplementary Data 8.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection
Data analysis	snRNA-seq data preprocessing: CellRanger (6.0.0), STAR (2.7.2a) snRNA-seq analysis: SCANPY(1.8.0), Seurat(4.0.0) , scDblFinder(1.10.0), scvi-tools(0.16.0, including scVI, scANVI), cellex(1.2.2), AUCell(1.18), clusterProfiler(4.4), WGCNA(1.69), triku(2.1.3), pyscenic(0.11.2), pandas(1.4.3), scipy(1.8.1), numpy(1.21.0), diffxpy(0.7.4), batchglm(0.7.4), edgeR (3.40.0) Phylogeny Reconstruction: MMseqs2 (13.45111), mafft-linsi (7.487), pal2nal (14), RAxML (8.2.12), FigTree(1.1.4), TreeShrink (1.3.9), dlcpar (2.0.1), iTOL v5, Tree of Life (2.0) Visualization: Matplotlib(3.5), plotnine(0.8), seaborn(0.12), patchworklib(0.4.0) Others: Python(3.8), R(4.1.0), rpy2(3.5.1), SPSS 26, GraphPad Prism 8.0.1

The source code to reproduce this project can be accessed at https://github.com/ZhaiLab-SUSTech/sc_medicago

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw sequencing data generated in this study were deposited in China National Center for Bioinformation with accession PRJCA011245. The gene matrix generated in the work of Cervantes-Pérez et al. can be accessed in the Gene Expression Omnibus with the accession number GSE210881.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.

Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).

Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)

Please provide details about how you controlled for confounding variables in your analyses.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

A total of 600 roots were collected from the single-nucleus RNA-seq library construction, including 150 from the control group, 150 from the 0.5 hour post-treatment group, 150 from the 6 hour post-treatment group, and 150 from the 24 hour post-treatment group. Root length of transgenic hairy roots (EV, n=19; MtFERi-1/-2, n=16; MtFERi-3, n=14); Quantification of infection events in transgenic hairy roots (EV, n=14; MtFERi-1/-2/-3, n=13); Quantification of nodules in transgenic hairy roots (EV, n=19; MtFERi-1/-2, n=16; MtFERi-3, n=14) and Density of nodules in transgenic hairy roots (EV, n=19; MtFERi-1/-2, n=16; MtFERi-3, n=14).

Data exclusions

Low quality nuclei were filtered, refer to the supplementary material for details.

Replication

For the preparation of the single-nucleus RNA-seq library, two biological replicates were included for each timepoint, resulting in a total of eight libraries. all attempts at replication were successful.

Randomization

Not applicable, as samples were processed identically through standard and in some cases automated procedures (10X genomics, DNA/RNA isolation) that should not have bias outcomes.

Blinding

Not applicable in the analysis step, as the analysis is in genome-wide level and has no obvious bias for specific genes or cells. Seeds were assigned to different treatment groups in a completely randomized manner thus investigators were blinded to the group allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input type="checkbox"/>	<input checked="" type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

We use anti-FLAG (F1804-1MG, Sigma), anti-HA (H6908-2ML, Sigma) antibodies and anti-FLAG M2 agarose beads (A2220-25ML, Sigma) for Co-IP assay.

Validation

anti-FLAG antibody manufacturer's website <https://www.sigmaaldrich.cn/CN/zh/product/sigma/f1804>
 anti-HA antibody manufacturer's website <https://www.sigmaaldrich.cn/CN/zh/product/sigma/h6908>
 anti-FLAG M2 agarose beads antibody manufacturer's website <https://www.sigmaaldrich.cn/CN/zh/product/sigma/a2220>

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Public health
<input checked="" type="checkbox"/>	<input type="checkbox"/> National security
<input checked="" type="checkbox"/>	<input type="checkbox"/> Crops and/or livestock
<input checked="" type="checkbox"/>	<input type="checkbox"/> Ecosystems
<input checked="" type="checkbox"/>	<input type="checkbox"/> Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Demonstrate how to render a vaccine ineffective
<input checked="" type="checkbox"/>	<input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent
<input checked="" type="checkbox"/>	<input type="checkbox"/> Increase transmissibility of a pathogen
<input checked="" type="checkbox"/>	<input type="checkbox"/> Alter the host range of a pathogen
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enable evasion of diagnostic/detection modalities
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enable the weaponization of a biological agent or toxin
<input checked="" type="checkbox"/>	<input type="checkbox"/> Any other potentially harmful combination of experiments and agents