



# A high-resolution transcriptomic atlas depicting nitrogen fixation and nodule development in soybean<sup>oo</sup>

Baocheng Sun<sup>1,2†</sup>, Yu Wang<sup>1,2†</sup>, Qun Yang<sup>1†</sup>, Han Gao<sup>1,2</sup>, Haiyu Niu<sup>1,2</sup>, Yansong Li<sup>1,2</sup>, Qun Ma<sup>1,2</sup>, Qing Huan<sup>1</sup>, Wenfeng Qian<sup>1,2\*</sup>  and Bo Ren<sup>1,2\*</sup> 

1. State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

<sup>†</sup>These authors contributed equally to this work.

\*Correspondences: Wenfeng Qian ([wfqian@genetics.ac.cn](mailto:wfqian@genetics.ac.cn)); Bo Ren ([bren@genetics.ac.cn](mailto:bren@genetics.ac.cn), Dr. Ren is fully responsible for the distributions of all materials associated with this article)



Baocheng Sun



Bo Ren

## ABSTRACT

Although root nodules are essential for biological nitrogen fixation in legumes, the cell types and molecular regulatory mechanisms contributing to nodule development and nitrogen fixation in determinate nodule legumes, such as soybean (*Glycine max*), remain incompletely understood. Here, we generated a single-nucleus resolution transcriptomic atlas of soybean roots and nodules at 14 days post inoculation (dpi) and annotated 17 major cell types, including six that are specific to nodules. We identified the specific cell types responsible for each step in the ureides synthesis pathway, which enables spatial compartmentalization of biochemical reactions during soybean nitrogen fixation. By utilizing RNA velocity analysis, we reconstructed the differentiation

dynamics of soybean nodules, which differs from those of indeterminate nodules in *Medicago truncatula*. Moreover, we identified several putative regulators of soybean nodulation and two of these genes, *GmbHLH93* and *GmSCL1*, were as-yet uncharacterized in soybean. Overexpression of each gene in soybean hairy root systems validated their respective roles in nodulation. Notably, enrichment for cytokinin-related genes in soybean nodules led to identification of the cytokinin receptor, *GmCRE1*, as a prominent component of the nodulation pathway. *GmCRE1* knockout in soybean resulted in a striking nodule phenotype with decreased nitrogen fixation zone and depletion of leghemoglobins, accompanied by downregulation of nodule-specific gene expression, as well as almost complete abrogation of biological nitrogen fixation. In summary, this study provides a comprehensive perspective of the cellular landscape during soybean nodulation, shedding light on the underlying metabolic and developmental mechanisms of soybean nodule formation.

Keywords: cytokinin, determinate nodule, nitrogen fixation, nodulation, RNA velocity, single nucleus RNA sequencing (snRNA-seq), soybean, ureide

Sun, B., Wang, Y., Yang, Q., Gao, H., Niu, H., Li, Y., Ma, Q., Huan, Q., Qian, W., and Ren, B. (2023). A high-resolution transcriptomic atlas depicting nitrogen fixation and nodule development in soybean. *J. Integr. Plant Biol.* **00**: 1–17.

## INTRODUCTION

Biological nitrogen fixation (BNF) is a crucial biological process that allows plants to efficiently obtain the nitrogen they need to grow. In this process, legume plants form a beneficial relationship with soil bacteria called rhizobia, which colonize in special nodules and fix atmospheric nitrogen into a form the plants can use. The nodule is composed of plant cells and symbionts that have undergone a series of complex genetic and molecular changes in response to signals from the rhizobia and plants. While nitrogen fixation provides an important source of nitrogen for legume plants, the gene expression regulation of nodule development is complex and not yet fully understood (Oldroyd, 2013; Roy et al., 2020).

Legume nodulation is a complex process regulated by various factors such as phytohormones, microRNAs, small peptides, reactive oxygen species (ROS), light, and nutrients (Ferguson et al., 2019; Roy et al., 2020; Wang et al., 2021; Xu et al., 2021; Ji et al., 2022; Luo et al., 2023; Yun et al., 2022). For example, phytohormones play vital roles in all nodulation processes, including epidermal infection, infection thread invasion, nodule organogenesis, and nitrogen fixation activation (Velandia et al., 2022). Cytokinin is known to be one of the crucial phytohormones involved in nodulation, as it has been shown to be necessary and sufficient to initiate nodule formation (Cooper and Long, 1994; Heckmann et al., 2011). Furthermore, many other hormones regulate legume nodulation in a cytokinin-dependent way. For example, cytokinin signaling appears to act upstream of auxin biosynthesis and transport to regulate nodule organogenesis (Plet et al., 2011; Ng et al., 2015; Gao et al., 2021). Genetic data positions cytokinin signaling upstream of NIN, NSP2 and DELLAs, the repressors of GA signal pathway, as the spontaneous nodule phenotype of *snf2* (*cre1*) mutant requires these transcription regulators (Tirichine et al., 2007; Jin et al., 2016). The positive role of cytokinin in nodulation was supported by a series of evidence from the cytokinin receptor mutants' characterization (Murray et al., 2007; Tirichine et al., 2007; Plet et al., 2011). *CRE1* knockdown causes dramatically reduced nodule number in *M. truncatula* and the *L. japonicus* cytokinin receptor *lhk1 lhk1a-1 lhk3* triple mutant has no nodule (Gonzalez-Rizzo et al., 2006; Held et al., 2014). Reduction of the endogenous active cytokinin level by over-expressing a *CYTOKININ OXIDASE/DEHYDROGENASE* (*CKX*) or knockdown *LOG1* gene leads to decreased nodule number in *L. japonicus* and *M. truncatula*, respectively (Lohar et al., 2006; Mortier et al., 2014). However, some seemingly contradictory results were also reported. For example, epidermis-specific cytokinin depletion led to increased nodule number in *M. truncatula*, and a mutation of *L. japonicus* *CKX3* gene causes evaluated endogenous cytokinin level and reduced nodule number (Jardinaud et al., 2016; Reid et al., 2016). In contrast, knockdown a cytokinin pathway positive regulator, *GmRR11*, causes increased nodule number in soybean (Chen et al., 2022). These observations underscore

the complexity of cytokinin's role and the delicate balance of signaling during nodulation, necessitating further investigation into these intricate relationships.

Legume roots form two types of nodules, determinate and indeterminate, based on their morphological characteristics (Hirsch, 1992). These two types differ in primordium origin, meristem maintenance, and mature nodule shape (Hadri et al., 1998; Xiao et al., 2014). Determinate nodules are spherical in shape, lack persistent meristem, and originate from the outer cortex (Doyle et al., 2000; Smith and Atkins, 2002). Most species of the tropical legume tribes such as soybean, cowpea (*Vigna unguiculata*), common bean (*Phaseolus vulgaris*), and mung bean (*Vigna radiata*), form determinate nodules. On the other hand, indeterminate nodules have a more persistent meristem, resulting in a cylindrical shape, and originate from the inner cortex. Indeterminate nodules are typically found in temperate species in Papilionoideae, such as *Medicago* (*Medicago truncatula*), clover (*Trifolium repens*), and pea (*Pisum sativum*).

Legume nodules can also be classified according to their exported forms of fixed nitrogen: amide-forming and ureide-forming. Amide-forming nodules assimilate fixed nitrogen as asparagine (Asn) and glutamine (Gln), which are then transported out of nodules, while ureide-forming nodules export allantoin (Aln) and allantoinic acid (Alc) out of nodules. Indeterminate nodules are all amide-forming, while determinate nodules are almost exclusively ureide-forming (Tegeeder, 2014; Valentine et al., 2017). An exception is the nodule of *Lotus japonicus*, which is determinate but uses Asn as the exported fixed nitrogen (Smith and Atkins, 2002). Despite the significant differences between these two types of nodules, the underlying metabolic and developmental mechanisms are still not clear.

Studying nodulation is challenging due to the intricate heterogeneity encompassed within nodules. Therefore, investigating legume nodules at the single-cell level is necessary for a thorough understanding of the process. Indeed, the advances in single-cell RNA sequencing (scRNA-seq) and single-nucleus RNA sequencing (snRNA-seq) have revolutionized our ability to investigate the transcriptomes of highly heterogeneous cell types (Denyer et al., 2019; Liu et al., 2021; Wang et al., 2021). Recent studies in *M. truncatula* have demonstrated the utility of these techniques in exploring the differentiation trajectories of symbiotic and un-symbiotic cells (Cervantes-Pérez et al., 2022; Ye et al., 2022), as well as the early response to rhizobial infection (Cervantes-Pérez et al., 2022). Additionally, a study performing transcriptome profiling for individual cell types has revealed genes responsible for carbon/nitrogen metabolism and transport in infected cell (IC) and uninfected cell (UC) in *L. japonicus* (Wang et al., 2022). Therefore, it is important to perform scRNA-seq/snRNA-seq on soybean nodules to unravel the heterogeneity of cells and understand the specific genes involved in metabolism and development.

Here, we used snRNA-seq to construct a comprehensive, high resolution transcriptomic atlas of soybean roots and nodules. This dataset was used to identify all major cell types present in the underground organs of soybean, including six nodule-

specific cell types. Comparison of these snRNA-seq data with publicly available scRNA-seq/snRNA-seq datasets from *M. truncatula* nodules revealed significant differences in the molecular regulatory mechanisms controlling nodule development and nitrogen metabolism between determinate and indeterminate nodules. In line with previous studies, pathways involved in ureide, rather than amide, biosynthesis and transportation were prominently enriched in determinate nodules. In addition, we examined the progenitor cells and differentiation trajectories of nodule-specific cell lineages, and experimentally demonstrated the role of cytokinin receptor, *GmCRE1*, in nodule formation and biological nitrogen fixation. These findings address long-standing questions about the cellular composition and development of determinate nodules and can serve as a resource to guide biotechnological improvements to nitrogen fixation in soybean and other determinate nodule legume crops.

## RESULTS

### Construction of a transcriptional atlas of soybean roots and nodules at single-nucleus resolution

To investigate the molecular mechanisms underlying nodule organogenesis in soybean, we inoculated roots of 5-d-old *Glycine max* cv. Williams 82 (W82) seedlings with or without rhizobium (hereafter referred to as N group, including both nodules and roots, and R group, including roots only, respectively). The complete underground organs were collected from each group at 14 dpi for snRNA-seq to identify nodule-specific patterns of gene expression spanning all stages of nodule development.

Nuclei were isolated using fluorescence-activated nucleus sorting stained with DAPI (4',6-diamidino-2-phenylindole, [Figure S1](#)). High-quality snRNA-seq libraries were prepared using the 10x Genomics Chromium single-cell microfluidics platform and subjected to high-throughput sequencing. We obtained 23,063 and 19,712 transcriptomes from the N and R groups, respectively, with a median number of 1,471 and 1,411 expressed genes per nucleus ([Table S1](#)). To validate the efficiency of mRNA capture in our snRNA-seq results, bulk RNA-seq was conducted in parallel for the N and R samples. The gene expression levels estimated from the pooled nuclei in the snRNA-seq experiment (i.e., pseudobulk) were highly correlated with corresponding expression levels in the bulk RNA-seq ( $\rho > 0.8$  in both groups, Spearman's correlation, [Figure S2](#)), supporting the reliability and accuracy of our snRNA-seq data.

We combined the snRNA-seq data from the N and R groups and used the “sctransform” algorithm ([Hafemeister and Satija, 2019](#)) to normalize the transcriptomes among individual nuclei. Individual transcriptomes were then visualized in two-dimensional space by uniform manifold approximation and projection (UMAP) analysis, and a total of 21 cell clusters were identified by the Louvain method (clusters 0–20, [Figure 1A](#)). We observed considerable overlap between the N and R groups for most clusters, with the exception of clusters 18, 19, and 20, which almost exclusively contained N

A single-nucleus resolution atlas of soybean nodules

group nuclei ([Figure 1B](#)). This finding suggested that clusters 18–20 potentially contained transcriptomes of either nodule-specific cells or root cells with transcriptomes that were fundamentally altered during nodulation.

To identify the tissue origin of cells in clusters 18–20, we examined the expression patterns of well-known genes to participate in soybean nodulation, including *NIN1a*, *NFR1*, *NSP1*, *RPG*, *NF-YA*, *ENOD40*, *TML*, and *DNF2* ([Roy et al., 2020](#)). We found that most of these genes were specifically expressed in one or more of these three clusters, further supporting that these three clusters comprised nodule-specific cell types ([Figures 1C, S3](#)). These observations indicated that soybean nodule-specific cells were successfully captured by our snRNA-seq analysis.

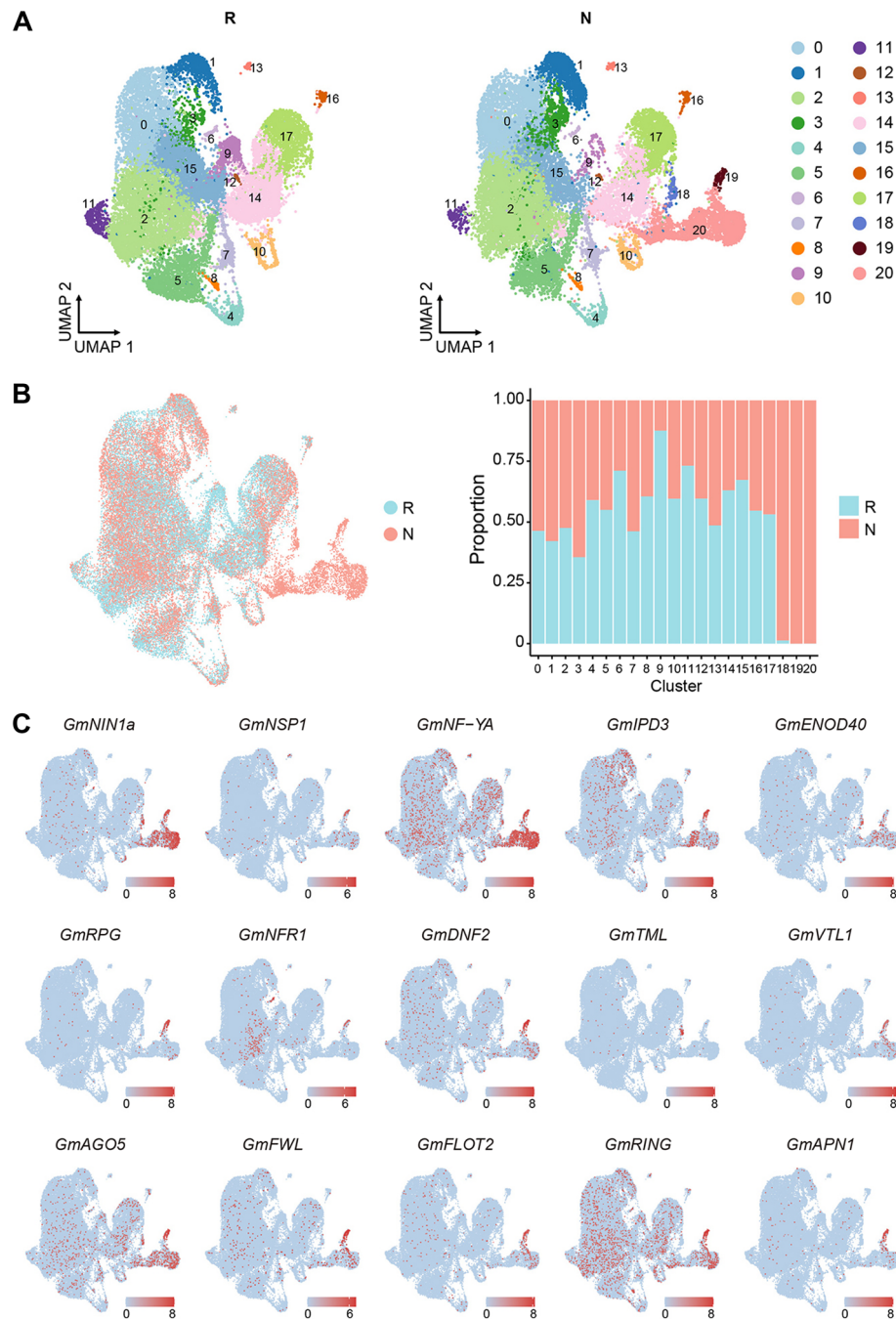
### Six cell types specific to nodules in soybean

We next annotated the various cell types in each cell cluster. Given the limited availability of cell-type specific marker genes reported for soybean roots and nodules, we utilized a strategy based on the cluster-specific expression of orthologs of *Arabidopsis* marker genes ([Denyer et al., 2019](#); [Zhang et al., 2019](#)) to infer soybean cell types. This analysis successfully identified four cell types, including trichoblast (cluster 6), meristem (cluster 10), phloem (clusters 12 and 13), and xylem (cluster 16, [Figures 2C, S4](#); [Table S2](#)).

To annotate cell types in the remaining cell clusters, we computationally identified 20 cluster-enriched genes in our snRNA-seq data ([Table S3](#)) and experimentally determined their expression pattern *in planta* using a  $\beta$ -glucuronidase (GUS) reporter system in transgenic hairy roots. For example, we detected the cluster 0-specific gene *GmWRKY72* ([Figure 2A](#)), which was exclusively expressed in the epidermis of both roots and nodules ([Figure 2B](#)), indicating that cluster 0 contained epidermal cells ([Figure 2C, D](#)). Similarly, clusters 2 and 3 were designated as cortical cells based on the high expression of *GmPLA2A* and *GmABCG37*, while cluster 7 was annotated as vascular cells based on enrichment with *GmFLA* transcripts, and clusters 14 and 15 were defined as procambial cells due to the high expression of *GmNDL2* and *GmTPC1*, respectively ([Figure 2](#)).

In addition, some nodule-specific cell types were identified in clusters 18–20. We designated clusters 18 and 19 as the nodule phloem pole pericycle and infected cells, respectively, based on enrichment with *GmTPS1* and *Gm17G050100* transcripts ([Figure 2](#)). As the irregular distribution of cluster 20 in the UMAP space, we hypothesized that this cluster might contain multiple cell types. To investigate this cell cluster, we further divided cluster 20 into seven subclusters ([Figure S5A](#)) and identified four subcluster-enriched genes. Detection with the GUS reporter system showed that these genes, *GmTAT7*, *GmCYP83B1*, *GmTOPII*, and *Gm05G212300*, were expressed in the outer cortex, inner cortex, pericycle related procambium, and uninfected cells/vascular cells, respectively, leading to the designation of cluster 20 as nodule ground tissue composed of at least the four aforementioned anatomical cell types ([Figures 2, S5A](#)).

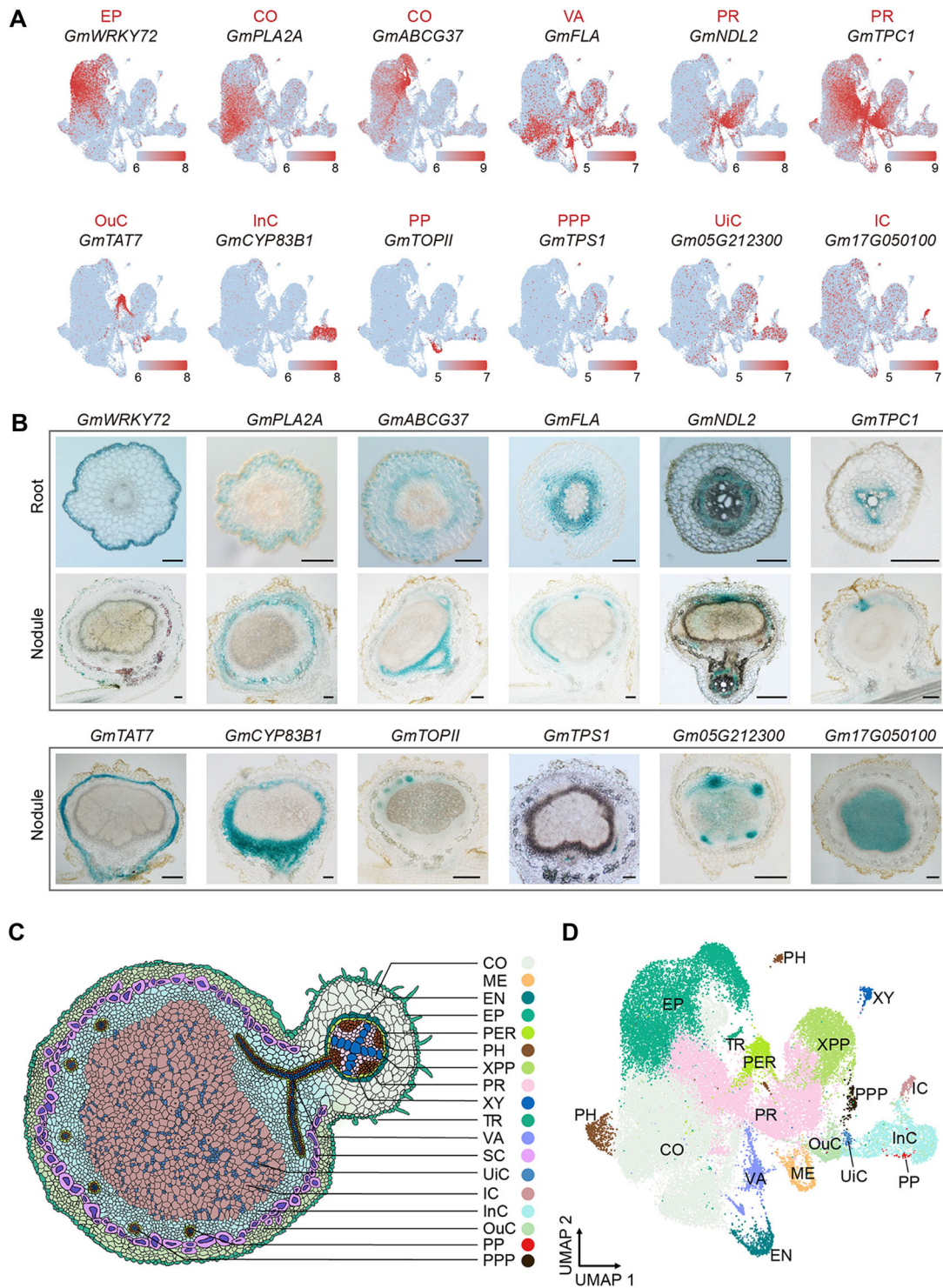




**Figure 1. Identification of 21 cell clusters in soybean roots and nodules using snRNA-seq**  
**(A)** Uniform manifold approximation and projection (UMAP) visualization of 21 cell clusters for 19,712 and 23,063 nuclei of soybean roots without (R group) or with (N group) rhizobium inoculation, respectively. Each dot in the UMAP visualization represents an individual cell. **(B)** Identification of common and specific clusters between R and N groups. Left: Visualization of cells from R (cyan) and N (coral) groups in the same UMAP space. Right: Proportion of R group and N group cells in all 21 clusters. **(C)** UMAP visualization for the expression of nodulation-related genes in soybean. Dots represent individual cells, which are colored according to the normalized gene expression levels (i.e.,  $\log(\text{CPM} + 1)$ ). Here, CPM refers to the number of transcripts of a particular gene in a cell, normalized to a million total transcripts.

Further experimental examination of eight additional cluster-enriched genes (Figure S5B-D) finally enabled the complete annotation of all 21 clusters into 17 distinct cell types, including epidermis, trichoblast, cortex, meristem, vascular cells, endodermis, procambium, pericycle,

pericycle related procambium, phloem, xylem, phloem pole pericycle, xylem pole pericycle, inner cortex, outer cortex, uninfected cells, and infected cells (Figure 2). Notably, we observed high variation in the number of cells obtained for each cell type, ranging from 181 to 11,750



**Figure 2. Cell-type annotation of 21 clusters identified in soybean roots and nodules**

**(A)** Uniform manifold approximation and projection (UMAP) visualization of cluster-enriched gene expression in soybean roots and nodules. Cells are colored according to the normalized gene expression levels (i.e.,  $\log(\text{CPM} + 1)$ ). **(B)** Determination of the expression patterns of the identified cluster-enriched genes using *promoter:GUS* staining. Scale bar: 100  $\mu\text{m}$ . **(C)** Schematic illustration of the anatomy of soybean roots and nodules, with different cell types shown in various colors. **(D)** UMAP visualization of cell type in soybean roots and nodules. Individual cells are represented dots and are colored according to the schematic in **(C)**. Sclerenchyma cells, characterized by their high lignin content and located between the inner and outer cortex in soybean nodules, were not captured in our snRNA-seq analysis, likely due to challenges associated with protoplasting these heavily lignified cells. CO, Cortex; ME, Meristem; EN, Endodermis; EP, Epidermis; PER, Pericycle; PP, Pericycle related Procambium; PH, Phloem; PPP, Phloem Pole Pericycle; PR, Procambium; XY, Xylem; TR, Trichoblast; VA, Vascular; XPP, Xylem Pole Pericycle; SC, Sclerenchyma; IC, Infected Cell; InC, Inner Cortex; OuC, Outer Cortex; UiC, Un-infected Cell.

sequenced cells. Having constructed a cell atlas for soybean roots and nodules, it has instigated further inquiries regarding their concerted efforts in carrying out nitrogen fixation, as well as the origin of nodule-specific cells within the soybean roots.

### Spatial compartmentalization of ureides synthesis in soybean nodules

The discovery of the six nodule-specific cell types in soybean (i.e., determinate nodule) has raised questions about their relationship with nodule cell types in *M. truncatula* (i.e., indeterminate nodule). Previous studies in *M. truncatula* have identified 13 cell types in 14-dpi nodules by scRNA-seq (Ye et al., 2022), and 10 cell types in the 2-dpi root tips by snRNA-seq (Cervantes-Pérez et al., 2022). To compare the scRNA-seq/snRNA-seq data between soybean and *M. truncatula* nodules, we integrated the two datasets in *M. truncatula* with our soybean N group based on orthologous gene expression levels (see Materials and Methods). After applying dimensionality reduction using UMAP, we observed that the soybean and *M. truncatula* cell clusters generally overlapped rather than spatially separated on the UMAP graph, suggesting that experimental batch effects were largely eliminated (Figure S6A, B).

To investigate differences in gene expression between soybean and *M. truncatula*, we used Spearman's correlation coefficient among corresponding genes between the two species to calculate the pairwise expression similarity. This analysis showed that although the same or corresponding cell types shared similar transcriptomes between soybean and *M. truncatula* roots (Figure S6C), this transcriptomic similarity was greatly reduced for nodule cell types, except for that of infected cells (Figure S6D). In order to specifically identify differences between determinate and indeterminate nodules, we restricted data integration to the six nodule-specific cell types in soybean and cell types identified in the 14-dpi nodules of *M. truncatula*, removing one mix cell cluster and two unknown cell clusters (Figures 3A, S6E). This comparison showed that even the infected cells in soybean nodules were only moderately correlated with the infection cells, pre-infection cells, and nitrogen fixation cells identified in *M. truncatula* (all three Spearman's correlation coefficients  $\rho < 0.2$ ; Figure 3B). These observations further implied the substantial differences in nodule structures, nitrogen fixation products, and bacteroid characteristics between soybean and *M. truncatula* may contribute to the observed lower transcriptomic similarity in nodule cells compared to root cells.

To explore differences in nodule development and metabolism between determinate and indeterminate nodules, we conducted functional enrichment analysis on highly expressed genes in soybean and *M. truncatula* nodules. In contrast to *M. truncatula* nodules, soybean nodules were highly enriched with upregulated genes associated with the "Purine metabolism" (Figure 3C). Among the 37 annotated genes involved in purine metabolism and transportation in

soybean (Table S4), 32 were exclusively expressed in soybean nodule-specific cells (six genes in cluster 20, 21 in cluster 19, and seven in cluster 18, Figure 3D). For example, *GmNSH2a* (encoding a nucleoside hydrolase) was exclusively expressed in cluster 19, *GmALN1a* (encoding an allantoinase) was specifically expressed in cluster 18, *GmPUR3* (encoding a glycinamide ribonucleotide transferase) was highly expressed in clusters 19 and 20, while *GmUOX* (encoding a uricase) was expressed in clusters 18 and 20.

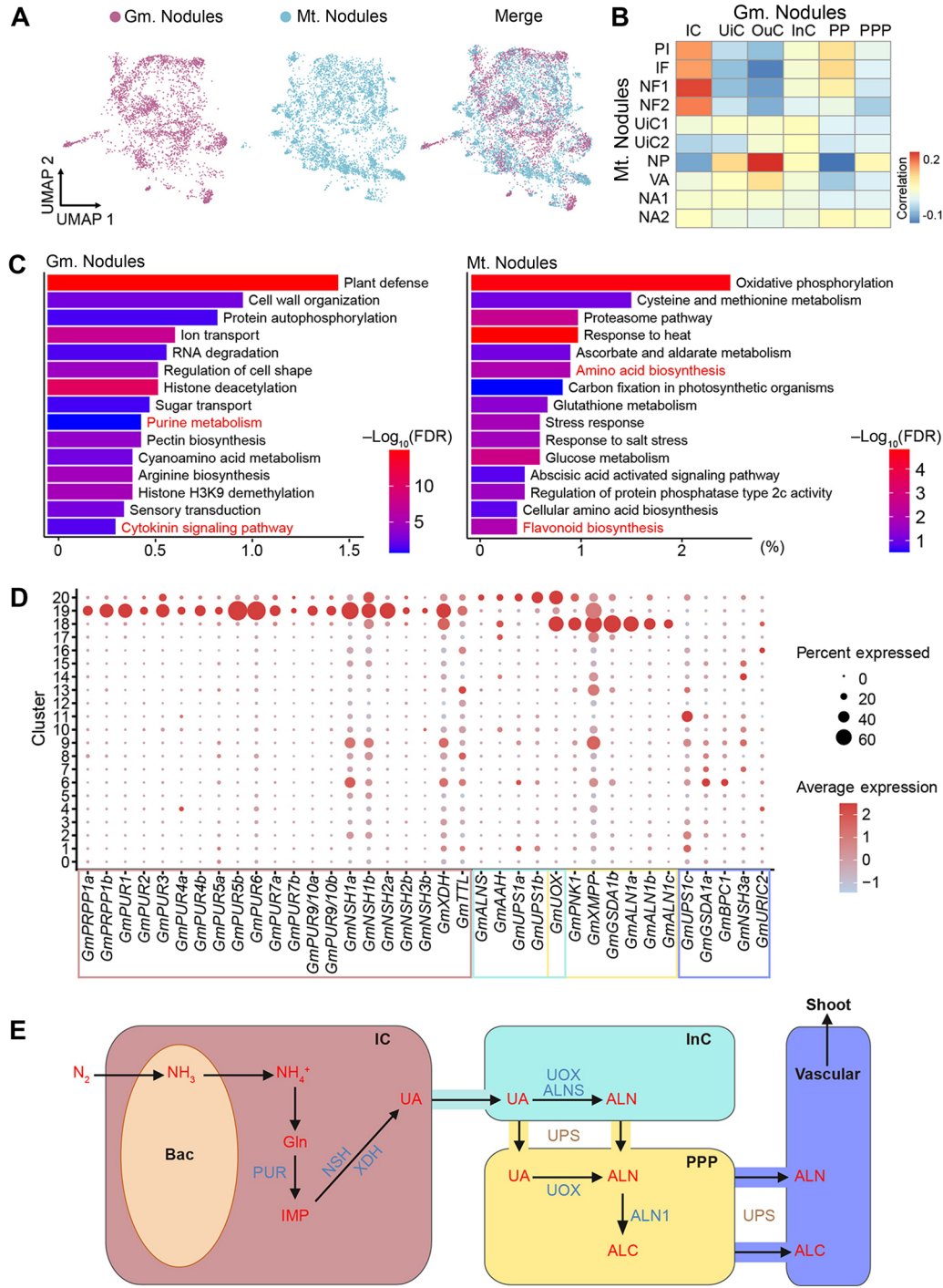
This striking overrepresentation of purine pathway-related genes in soybean nodules was particularly noteworthy because purine serves as the precursor for ureides, the primary form of assimilated and transported nitrogen in soybean BNF (Smith and Atkins, 2002). Our data showed that genes involved in the initial steps of *de novo* purine biosynthesis, from phosphoribosylpyrophosphate to uric acid (UA), were specifically expressed in infected cells (cluster 19, Figure 3D). Genes encoding UA transporters, such as *ureide permease 1* (*GmUPS1a* and *1b*), were upregulated in inner cortex of soybean nodules (cluster 20, Figure 3D). In addition, the gene encoding uricase (*GmUOX*), which catalyzes UA into Aln, was specifically expressed both in inner cortex (cluster 20) and phloem pole pericycle (cluster 18), and genes encoding allantoinase (*GmALN1a*, *GmALN1b*, and *GmALN1c*), which catalyze Aln into Alc, were exclusively expressed in phloem pole pericycle (cluster 18, Figure 3D; Table S4).

By integrating cell type-specific gene expression patterns with our understanding of biochemical reactions, we postulated a model of nitrogen assimilation in soybean nodules that incorporates spatial compartmentalization of ureides biosynthesis (Figure 3E). In this model, UA is synthesized in infected cells, then transported to the inner cortex of soybean nodules. In the inner cortex, UA is catalyzed into Aln by uricase, after which it is then transported into the nodule pericycle, along with any residual UA, where it is further transformed into Alc. Finally, Aln and Alc in the nodule pericycle are loaded into vascular bundles for transport to the shoot. Notably, this metabolic pathway is fundamentally distinct from that in *M. truncatula* nodules, which showed enrichment for genes involved in amino acid and flavonoid biosynthesis (Figure 3C), consistent with the previous report (Ye et al., 2022). Collectively, these results indicated that the ureide-dependent metabolic processes of nitrogen fixation and transport in determinate nodules of soybean fundamentally differ from the corresponding amide-dependent processes in indeterminate nodules of *M. truncatula*.

### Reconstructing the dynamics of nodule-specific cell differentiation

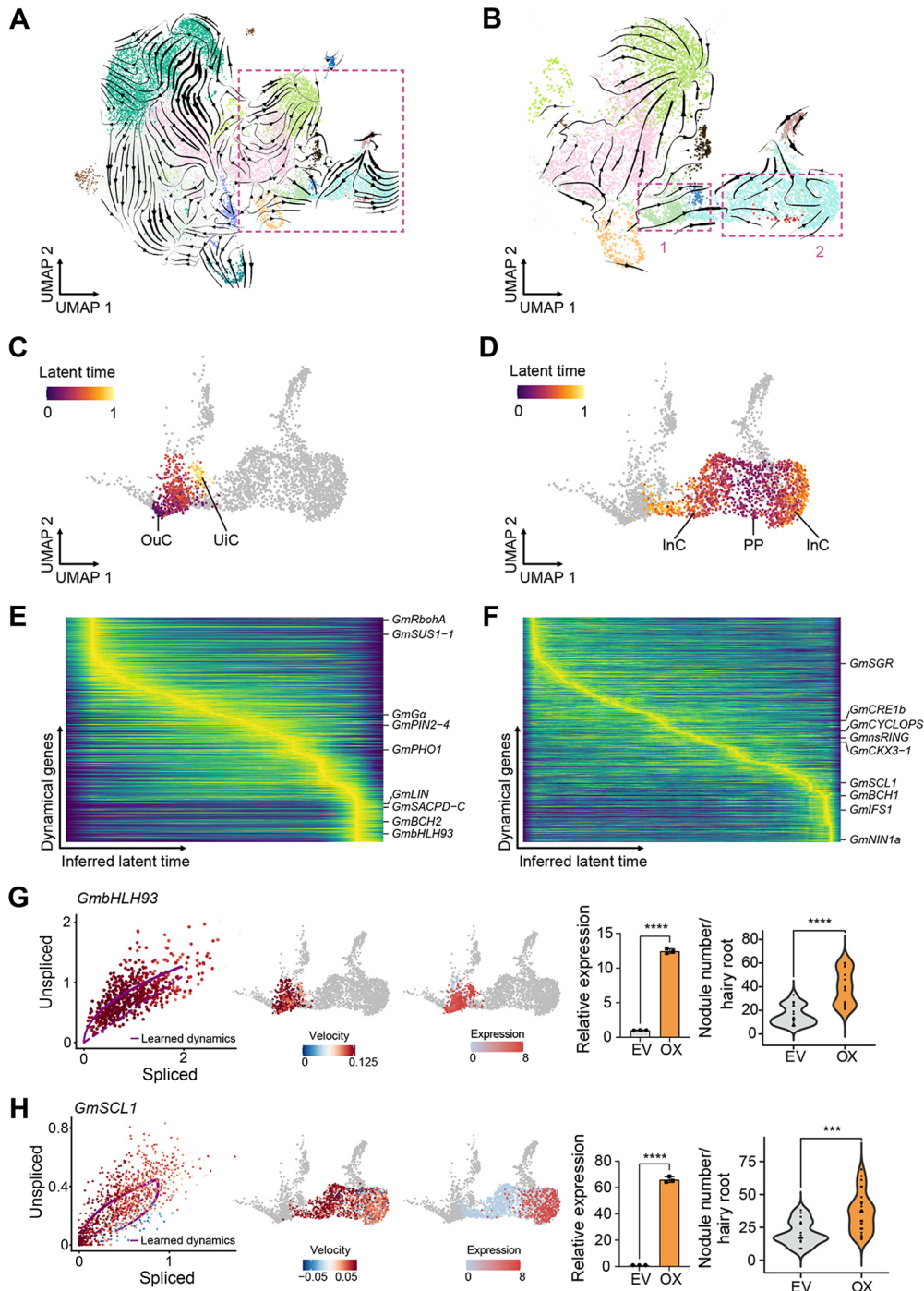
To explore the developmental origin(s) of soybean nodule tissues, a long-standing question in BNF, we used RNA velocity analysis to reconstruct the dynamic differentiation trajectory of the cell types in N group. This analysis showed root developmental streams in which young epidermal cells progressed to trichoblasts; cortical cells, which contain intercalary meristem,





**Figure 3. Comparison of single cell/nucleus transcriptomes between soybean and *M. truncatula* nodules**

(A) Uniform manifold approximation and projection (UMAP) visualization of soybean and *M. truncatula* nodule cells. Each dot represents an individual cell and is colored according to its species of origin. (B) Heat map shows the pairwise Spearman's correlation coefficients between cell types in soybean and *M. truncatula* nodules. Each row represents a cell type from *M. truncatula*, and each column represents a cell type from soybean. Abbreviations from soybean are based on those used in Figure 2D. Abbreviations from *M. truncatula* correspond to cell types identified in Ye et al. (2022): NF1, 2, Nitrogen Fixation 1, 2; PI, Pre-infection; IF, Infection; NP, Nodule Parenchyma; VA, Vascular tissue; UIC, 1, 2, Un-infected Cell, 1, 2; NA2, Nodule Apex 2. (C) Functional enrichment analysis of high expressed genes in soybean and *M. truncatula* nodules. (D) Expression pattern of genes involved in the purine metabolic pathway in soybean roots and nodules. Each dot is colored based on the scaled average gene expression level across all cells within a cluster, and the radius represents the percentage of cells expressing the gene within a cell cluster. (E) Schematic illustration of the nitrogen assimilation process in soybean nodules. Bac, Bacteroid; Gln, Glutamine; IMP, Inosine monophosphate; UA, Uric acid; ALN, Allantoin; ALC, Allantoic acid; PUR, Purine biosynthesis protein; NSH, Nucleoside hydrolase; XDH, Xanthine dehydrogenase; UOX, Uricase; ALNS, Allantoin synthase; ALN1, Allantoinase 1; UPS, Ureide permease.



**Figure 4. Reconstruction of developmental trajectory and identification of driver genes in soybean nodulation using RNA velocity**

(A) Visualization of velocities for cells in the N group as streams in the uniform manifold approximation and projection (UMAP) space. The red dashed rectangle highlights nodule-specific cells (clusters 18–20) and their adjacent cell types in the UMAP plot (clusters 9, 10, 14, and 17). (B) Visualization of velocities for nodule-specific cells and their adjacent cell types as streams in the UMAP space. The red dashed rectangles 1 and 2 highlight streams from nodule outer cortical cells to uninfected cells and from pericycle related procambial cells to the inner cortex, respectively. (C) UMAP visualization of nodule-specific cells. Outer cortical cells and uninfected cells are colored according to their scVelo's latent time, and other cells are shown in gray. (D) Similar to (C), inner cortical cells and pericycle related procambium cells are colored according to their scVelo's latent time. (E) Heat map showing the scaled expression levels of the 1,200 genes with highest likelihood of dynamic expression along the scVelo's latent time from outer cortex to uninfected cells. (F) Similar to (E), heat map showing the scaled gene expression levels along the scVelo's latent time from pericycle related procambium cell to the inner cortex. (G) From left to right: Unspliced–spliced phase portraits (cells colored according to individual gene's velocity), velocity and expression of *GmbHLH93* among cells, expression level of *GmbHLH93* in transgenic hairy roots, and nodule number per hairy root when *GmbHLH93* is overexpressed (OX) from the 35S promoter. The expression levels were measured using qRT-PCR and were normalized to the expression level of *GmABC1*. Data are presented as means  $\pm$  standard deviation (SD) from three independent experiments. Asterisks indicate statistical significance, compared to the empty vector (EV) control. Two-tailed Student's *t*-test, \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ . (H) Similar to (G), for *GmSCL1*.



differentiated into endodermis; and procambial cells formed xylem pole pericycle (Figure 4A). These observations were consistent with established knowledge of root development (Petricka et al., 2012; Denyer et al., 2019; Liu et al., 2021; Shahan et al., 2022). By contrast, some differentiation streams led to formation of nodule ground tissues (cluster 20, Figure 4A). Whereas the origins of all nodule cells had not been determined, RNA velocity analysis of nodule-specific cells (clusters 18–20) and their adjacent cell types in the UMAP plots (clusters 9, 10, 14, and 17) showed that uninfected cells in soybean nodules were likely derived from the nodule outer cortical cells while the inner cortex in soybean nodules was plausibly derived from pericycle related procambial cells (Figure 4B).

To gain a better understanding of gene expression dynamics along cell differentiation streams, we determined the latent time of individual cells using scVelo, a likelihood-based model that solves cell differentiation dynamics (Bergen et al., 2020). This latent time analysis showed that outer cortex and pericycle related procambium cells differentiated earlier in soybean nodules within their respective streams (Figure 4C, D). Since genes with pronounced transcriptional activation are likely to serve as drivers of cell differentiation, we identified putative driver genes of nodule-specific cell lineages along the estimated latent time. Notably, several well-known nodulation-related genes emerged, including *GmNIN1a*, *GmCYCLOPS*, *GmBCH1*, *GmIFS1*, *GmPIN* (Figure 4E, F).

To investigate other putative driver genes of nodule formation identified from scVelo modeling (Figure 4E, F), we examined two of them, *beta HLH protein 93* (*GmbHLH93*) and *SCREAM-like protein* (*GmSCL1*), which are expressed in outer cortex and inner cortex, respectively, and also exhibited dynamic expression patterns but their potential functions in nodulation have not been reported yet. To verify their roles in nodulation, we generated transgenic hairy roots overexpressing each gene individually. Phenotypic analyses showed that overexpression of either *GmbHLH93* or *GmSCL1* resulted in significantly greater nodule numbers than that in vector control hairy roots (Figure 4G, H), supporting their potential roles as regulators of soybean nodule organogenesis.

### Cytokinin receptors coding gene, *GmCRE1*, is a key regulator of soybean nodulation and nitrogen fixation

In light of overrepresentation with highly expressed genes related to “Cytokinin signaling” in soybean nodules (Figure 3C), we next investigated the potential roles of genes related to cytokinin signaling in soybean nodule organogenesis. Indeed, more than 26% of genes in the entire cytokinin pathway were exclusively expressed in the three nodule-specific cell clusters (clusters 18–20), which showed a higher proportion than that of all other phytohormone-related genes (Figure 5A). Furthermore, the results of functional enrichment analysis implied that the identified nodule-specific cytokinin pathway genes encompassed the complete cytokinin pathway, comprising genes related to cytokinin biosynthesis, metabolism, and signal transduction (Figure S7A; Table S5). We also noted that

some negative regulators of the cytokinin pathway, such as A-type *GmRRs* (suppressors of cytokinin signaling) and *CKXs* (encoding cytokinin oxidase/dehydrogenase, which degrades cytokinin), were also specifically expressed in the three nodule-specific cell clusters (Figure S7B). These observations together indicated that a sophisticated regulatory network modulated cytokinin signaling in soybean nodules.

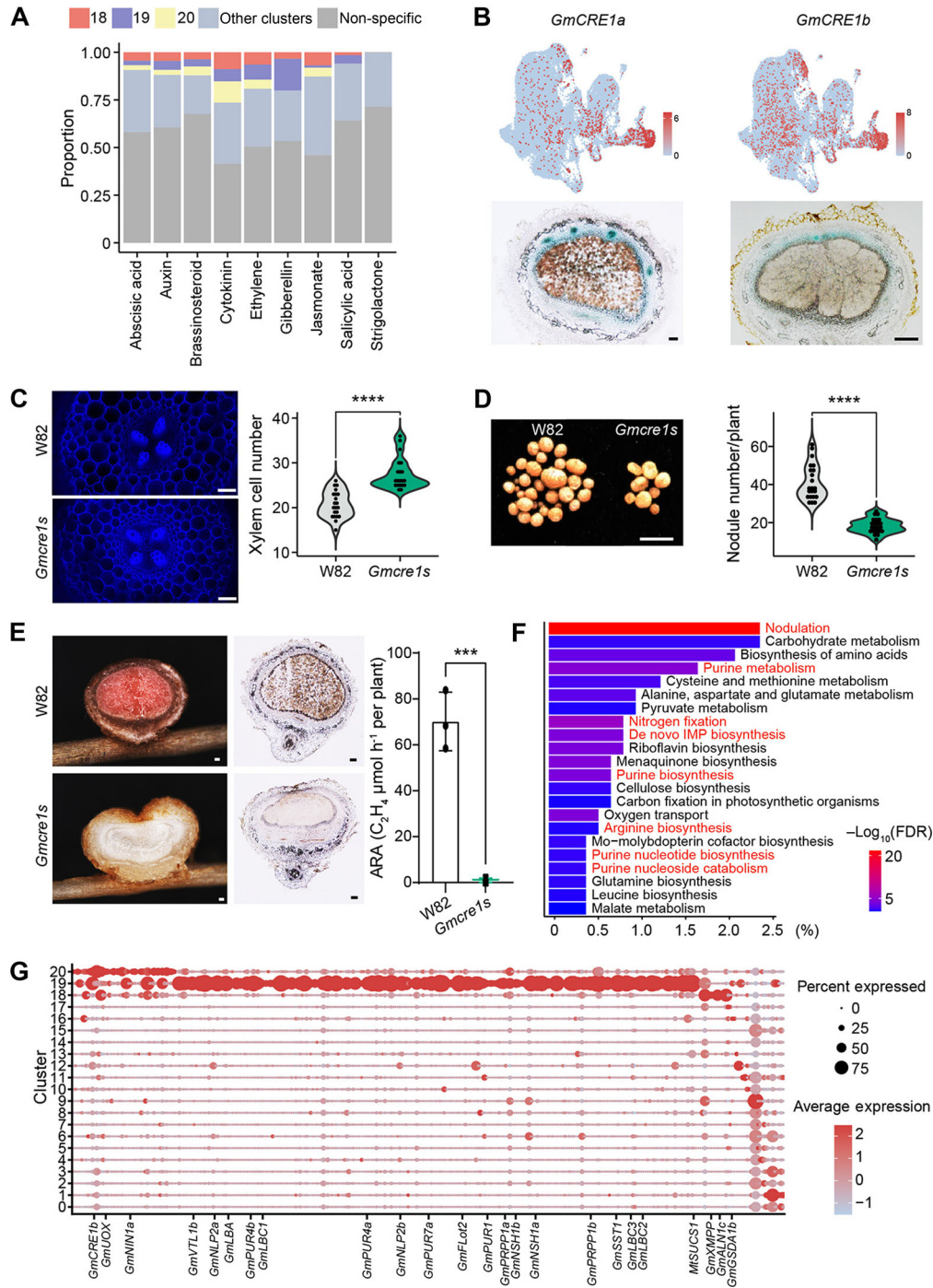
Notably, genes encoding cytokinin receptors, *GmCRE1a/b/c/d*, were highly expressed in nodule inner cortex and vascular cells (cluster 20, Figures 5B, S7C), and in particular, *GmCRE1b* (*Gm08G049000*) exhibited dynamic expression during soybean nodulation (Figures 4F, S7D), suggesting their function in soybean nodulation. We thereby focused on exploring the role of *GmCRE1* in soybean nodule formation through experiments *in planta*. To this end, we generated a *Gmcre1a/b/c/d* quadruple knockout line in soybean (*Gmcre1s* hereafter) through CRISPR-Cas9 gene editing (Figure S8A).

Compared to the wild-type (WT) W82 soybean plants, *Gmcre1s* plants had a greater number of xylem cells (Figure 5C), shorter primary roots and less lateral roots (Figure S8B), suggesting that *GmCRE1* participated in regulating soybean root development. Moreover, nodule number was significantly lower in *Gmcre1s* plants compared to WT (Figure 5D), and *Gmcre1s* nodules had a sunken distal region, rather than spherical as in WT nodules (Figure 5E). In addition, *Gmcre1s* nodules had substantially smaller nitrogen fixation zones than WT, with pale white interior. Indeed, nitrogenase activity indicated by acetylene reduction assay (ARA) confirmed that the capacity for nitrogen fixation was almost completely abolished in *Gmcre1s* nodules (Figure 5E). Consistent with these findings, we further observed that leghemoglobin genes (*LGBs*) expression, which is essential for nitrogen fixation, was significantly decreased in *Gmcre1s* nodules relative to WT (Figure S8C).

To further investigate the molecular mechanisms by which *GmCRE1* participates in soybean nodule organogenesis, we performed bulk RNA-seq on WT and *Gmcre1s* roots and nodules. We found that down-regulated genes in *Gmcre1s* plants were significantly overrepresented in pathways related to “nodulation”, “nitrogen fixation”, and “purine-related metabolism” (Figure 5F), which highly overlapped with the pathways activated in soybean nodule-specific cell types (Figure 3C, left panel). These down-regulated genes in *Gmcre1s* plants were specific to nodule-specific cells types, especially the infected cells (cluster 19, Figure 5G). Collectively, the observation that knockout of *GmCRE1* could almost completely block BNF in soybean demonstrated the critical role of *GmCRE1*-dependent cytokinin signaling in the formation of soybean nodules.

## DISCUSSION

The process of nitrogen fixation in legume nodules requires precise coordination of transcriptional programs across cells. Here, we applied snRNA-seq to construct a comprehensive transcriptomic atlas of soybean roots and nodules, identifying six



**Figure 5. Experimental investigation of the cytokinin signal pathway in soybean nodulation and nitrogen fixation**

(A) The proportion of genes exclusively expressed in soybean roots and nodules for each phytohormone pathway. (B) Uniform manifold approximation and projection (UMAP) visualization and histological analysis of *GmCRE1a* and *GmCRE1b* expression. The GUS staining results are presented for 14-dpi nodules of transgenic hairy roots. Scale bar: 100  $\mu$ m. (C) Phenotypic changes in *Gmcre1s* roots. The autofluorescence of xylem was observed from the UV channel. Scale bar: 100  $\mu$ m. The number of xylem cells was quantified based on root sections. Three independent experiments ( $n > 20$ ) were performed and similar results were obtained. (D) Phenotypic changes in *Gmcre1s* nodules. Nodule number was quantified from three independent experiments ( $n > 20$ ). Scale bar: 0.5 cm. Asterisks in (C) and (D) indicate statistically significant difference compared to the wild-type W82. Two-tailed Student's *t*-test, \*\*\*\*,  $P < 0.0001$ . (E) Characterization of the nitrogen fixation zone in soybean nodules. Left panel: blade slides observed using a stereoscopic microscope; middle panel: low-gelling temperature agarose slides observed using a microscope. Scale bar: 100  $\mu$ m. Right panel: Acetylene reduction assay (ARA) of nodules from W82 and *Gmcre1s* plants grown 14 dpi without external nitrogen sources. Data are presented as a mean  $\pm$  SD from three independent experiments. (F) Functional enrichment analysis of down-regulated genes in *Gmcre1s* plants. (G) Expression enrichment of down-regulated genes in *Gmcre1s* plants across 21 cell clusters in soybean roots and nodules. The figure is similar to Figure 3D except that it shows the expression of down-regulated genes in *Gmcre1s* plants.

nodule-specific cell types as well as their likely cellular progenitors. Moreover, we characterized the precise cell types specifically responsible for each step in the ureides synthesis pathway. Furthermore, our analyses highlight the crucial role of cytokinin in soybean nodulation and identify putative driver genes in this process. In particular, we experimentally validate the role of cytokinin receptor, GmCRE1, *in planta* as an essential component of nodule development and nitrogen fixation. Overall, our study addresses several as-of-yet unknown questions in the development of determinate nodules, and can serve as a resource for molecular breeding or biotechnology strategies to enhance legume crop productivity and sustainability.

In this study, we used snRNA-seq instead of scRNA-seq for two main reasons. Firstly, the process of protoplasting in plant scRNA-seq is susceptible to the introduction of erroneous gene expression and the loss of specific cell types (Birbaum et al., 2003; Han et al., 2017; Denyer et al., 2019). Secondly, snRNA-seq is favorable over scRNA-seq as it provides a higher fraction of unspliced pre-mRNA, owing to the subcellular localization from which RNA was extracted. This feature is crucial for determining the transient cellular dynamics of gene expression with RNA velocity analysis (La Manno et al., 2018; Bergen et al., 2020; Weiler et al., 2023). Indeed, we observed a higher fraction of unspliced pre-mRNA in our snRNA-seq data (~33% across all relevant cell types, Figure S9A, B) compared to less than 5% commonly seen in plant scRNA-seq experiments (Jean-Baptiste et al., 2019; Zhang et al., 2021). Despite the potential difference between nucleus and whole-cell RNA, snRNA-seq allows for greater accuracy in tracking the trajectory of gene expression changes over time, and is particularly valuable for characterizing cell fate decisions during development. Indeed, the experimental validation employed by this study confirmed the computational analysis results.

In contrast to a recent scRNA-seq study of indeterminate nodules in *M. truncatula* that uncovered two groups of apical meristem cell types differentiating into symbiotic or non-symbiotic cells (Ye et al., 2022), our study reveals the presence of nodule-specific phloem pole pericycle and pericycle related procambial progenitors within determinate nodules in soybean. Such progenitors exhibit transient meristematic activity, which is a distinctive feature of intercalary meristem, unlike the indeterminate nodules in *M. truncatula* that originate from a persistent apical meristem.

This study shows that the majority of genes in the ureides metabolism pathway are expressed at high levels in soybean nodules, aligning well with previous reports establishing the prevalence of ureides metabolism in soybean nodules (Tegeeder, 2014; Valentine et al., 2017; Fan et al., 2022). Our results enhance our understanding of the cell type-specific expression of these genes within soybean nodules (Figure 3D), illustrating the compartmentalization of biochemical reactions across different cell types within nodules (Figure 3E), to physically separate pathway reactions and thereby prevent deleterious reversible reactions caused by accumulation of intermediates or products. Furthermore, these findings based on high-resolution transcriptomics and GUS staining results, show that UA is likely

catalyzed into Aln and Alc in inner cortex and pericycle cells of the nodule, respectively (Figure 3E), contradicting the opinion that Aln and Alc are synthesized in uninfected cells (Smith and Atkins, 2002; Carter and Tegeeder, 2016).

As the complex ureides synthesis pathway requires approximately 20 enzymes, in contrast to only four enzymes required for Asn synthesis (Witte and Herde, 2020), it is intriguing that many agriculturally important legume crops, such as soybean, cowpea, common bean, and mung bean, all employ ureides for nitrogen assimilation and long-distance transport. One possible advantage of ureides is their low carbon:nitrogen (C:N) ratio of 1.0, meaning that one ureides molecule comprises four carbon and four nitrogen atoms. In contrast, Gln and Asn have C:N ratios of 2.5 and 2.0, respectively, necessitating more carbon for nitrogen fixation via amides than ureides. This hypothesis of C:N economy in BNF is supported by seminal work showing that cowpea nodules, which depend on ureides, consumed significantly less carbon and lost less CO<sub>2</sub> than white lupine (*Lupinus albus*) nodules, which rely on amides (Layzell et al., 1979). As BNF is an energy-intensive process, the coordination of carbon and nitrogen in legume nodules likely represents a crucial target of evolutionary selection (Sprent and McKey, 1994; Lavin et al., 2005). Indeed, a previous study showed that over-expression of common bean *UPS1*, which encodes a ureide transporter, led to improve BNF, shoot nutrition, and seed yield in soybean (Carter and Tegeeder, 2016). The importance of ureides metabolism and transportation in soybean BNF highlights a potential strategy to enhance symbiotic nitrogen fixation and ultimately increasing legume crop productivity, via optimizing the ureide pathway.

Our study elucidates the predominant role of the cytokinin receptor, GmCRE1, in late developmental stages or mature soybean nodules, which is in contrast to a recent snRNA-seq study on *M. truncatula*. The latter study showed that many early nodulation genes, which participate in cytokinin signaling, exhibit cell type-specific expression patterns, highlighting the crucial function of cytokinin in initiating nodulation (Cervantes-Pérez et al., 2022). In contrast to *Mtcre1* mutant nodules, which maintain nitrogen-fixing capacity similar to that of WT nodules (Boivin et al., 2016), our results show that *Gmcre1s* mutant nodules in soybean have almost completely lost their ability to fix nitrogen (Figure 5E). Moreover, the soybean *Gmcre1s* mutant exhibited an increase in the number of xylem cells in primary roots and decreased in the number of lateral roots compared to the WT, whereas two allelic mutants of *Mtcre1* showed a reduction in xylem poles and increase in lateral roots (Laffont et al., 2015). Taken together, these observations suggest that the role of cytokinin in both roots and nodules varies between legumes with determinate and indeterminate nodules.

A recent report established a cell atlas of soybean nodules and roots using single-nucleus and spatial transcriptomics (Liu et al., 2023). Their study provided valuable insights into the specialization of uninfected cells into functionally distinct subgroups and the identification of a transitional subtype of infected cells during nodule development. Our study expands



findings from this work by applying RNA velocity analysis to reconstruct the dynamic differentiation trajectory of cell populations and predict driver genes in soybean nodulation. In line with previous studies, we identified several putative driver genes, whose functions were known during nodulation, such as *Rboh*, *PIN*, *PHO1*, *NIN1*, and *IFS1* (Schäuser et al., 1999; Huo et al., 2006; Subramanian et al., 2006; Marino et al., 2011; Arthikala et al., 2017; Gao et al., 2021; Nguyen et al., 2021). Furthermore, the expression dynamics of these genes during nodulation were elucidated using latent time analysis, revealing their temporal roles in distinct developmental stages (Figure 4). For instance, *GmRbohA* is highly expressed in the early stage of nodule-specific cell differentiation (Figure 4E), which is in agreement with the previous work showing that reactive oxygen species produced by RBOH participates in nodule inception (Arthikala et al., 2017). Likewise, dramatic changes in the expression of *GmNIN1a* in the late stage of nodule formation (Figure 4F) corroborate the established function of NIN protein in the nitrogen fixation stage (Feng et al., 2021).

Identification of driver genes that control nodulation is crucial as it provides the potential for biotechnological approaches to enhance nitrogen fixation efficiency in legume crops. Our study identifies several genes that could be exploited to improve nitrogen fixation. In particular, the cytokinin receptor, GmCRE1, which plays a critical role in nodule development and nitrogen fixation, was confirmed through phenotypic characterization of the *Gmcre1s* soybean mutant (Figure 5). We also investigated the functions of two previously uncharacterized putative driver genes, *GmbHLH93* and *GmSCL1*. Overexpression of these genes in transgenic hairy roots resulted in a higher number of nodules than that in vector control plants, supporting a likely role in promoting nodulation.

In summary, our study provides a valuable resource for understanding determinate nodules at single-cell resolution, revealing the metabolic mechanisms that promote BNF efficacy in soybeans as well as the developmental mechanisms that drive cell differentiation in nodulation. Overall, this snRNA-seq analysis with experimental assays provide potential strategies for agronomic improvement of BNF efficiency in legumes and expands our understanding of determinate nodule development as it differs from the indeterminate nodule model legume, *M. truncatula*.

## MATERIALS AND METHODS

### Plant materials and growth conditions

*Glycine max* cv. Williams 82 was used as the wild type for all snRNA-seq and hairy root transformation experiments. The soybean seeds were sowed into sterile vermiculite for germination. Plants were grown in a phytotron set to a 16-h-light/8-h-dark period at 25°C, 65% relative humidity. The 5 d post germination seedlings were inoculated with a suspension of *Sinorhizobium fredii* CCBAU45436 strain (OD<sub>600</sub> = 0.05), and nodules were collected at 14 dpi. As a negative control, uninoculated seedlings were grown under the same conditions.

### Nuclei isolation, library construction, and snRNA-sequencing

We isolated soybean nuclei following a published protocol (Thibivilliers et al., 2020) with minor modification. Briefly, soybean seedlings with their aerial parts removed were collected in a precooled 90 mm Petri dish containing NIB (Sigma). Roots were quickly chopped with a razor blade and then incubated in the cold room for 15 min with gentle horizontal shaking. The isolated nuclei were filtered through a 40- $\mu$ m strainer (Falcon) into a 50 mL centrifuge tube (Corning). Subsequently, 1 mL of the isolated nuclei was transferred to a 1.5 mL centrifuge tube and stained with DAPI for 10 min. The stained nuclei were sorted using a BD FACASaria II cell sorter, and the purified nuclei were counted using trypan blue in a count board.

Cell capture was performed using the commercially available “Chromium Single Cell B Chip Kit”, following the recommended protocol provided by 10x Genomics. Library construction was performed using “Single Cell 3' Library” and “Gel Bead Kit V3.1”, and the resulting libraries were sequenced using the “Illumina Novaseq 6000” platform, targeting a sequencing depth of at least 100,000 reads/cell, based on the PE150 protocol.

### Raw data processing and generation of single-nucleus gene expression matrices

The fastq files of snRNA-seq were generated from Illumina binary base call (BCL) files using the “mkfastq” function of Cell Ranger (version 7.0.0, 10x Genomics) (Zheng et al., 2017). These fastq files were subsequently aligned to the soybean reference genome (version v2.1, annotated by the U.S. DOE Joint Genome Institute) using STAR (version 2.7.1a) (Dobin et al., 2013) embedded in the “count” function of Cell Ranger. The reads that shared the same cell barcode, gene identity, and unique molecular identifier (UMI, with one mismatch allowed) were collapsed into a UMI count. Finally, a gene expression matrix was generated for each snRNA-seq sample, including only cells identified as genuine by the cell-calling algorithm (Lun et al., 2019). Each row of the matrix represents a gene and each column represents a genuine cell.

Given that the N group sample has the potential to contain symbiotic rhizobia, it is possible that the plant cell nuclei were inadvertently contaminated during the preparation of the snRNA-seq library. To ascertain whether transcripts from rhizobia were indeed present, we conducted an alignment of the R2 (i.e., read 2) fastq file of the N group sample to the rhizobium genome (*S. fredii* CCBAU25509). Only a small fraction of the reads (18,388 reads, ~0.004%) could be uniquely mapped to the rhizobium genome, indicating that rhizobial contamination is unlikely to pose a significant issue. Since only the sequencing reads aligned to the soybean genome were used in our study, any sequencing reads originating from the rhizobia were effectively eliminated from further downstream analysis.

### Normalization, dimensionality reduction, cell clustering, and cell-type annotation for snRNA-seq data

All analyses of single-nucleus gene expression matrices were performed using the R package “Seurat” (version 3.2.3) (Stuart et al., 2019). The gene expression matrices for the N group and R group were merged and then normalized using function “SCTransform” (Hafemeister and Satija, 2019). The normalized data were subsequently subjected to linear dimensionality reduction using function “RunPCA”, and the top 30 principal components were chosen for non-linear dimensionality reduction and visualization using function “RunUMAP” (McInnes et al., 2018) (with parameters “min.dist = 0.5” and “n.neighbors = 30”). Based on the Louvain algorithm (Blondel et al., 2008), cell clustering was performed using functions “FindNeighbors” and “FindClusters” (with parameter “resolution = 0.3”), according to the top 30 principal components from the linear dimensionality reduction. The cell types of each cluster were annotated based on genes previously reported to be expressed in specific cell types and the genes identified by GUS staining (Table S3). The cluster-enriched genes were identified using function “FindAllMarkers” (with parameters “min.pct = 0.1” and “test.use = wilcox”), with the criteria of a fold difference > 2 and an adjusted *P*-value < 0.01 between the two groups of cells.

### Integration of *M. truncatula* and soybean scRNA/snRNA-seq datasets

To compare gene expression across different species, the orthologous pairs of the *Medicago* (MtrunA17r5.0) and soybean were predicted on their protein sequences by software “ortho-finder” (Emms and Kelly, 2019) (version 2.2.7, with the parameter “-S diamond”). Taking into account gene duplication events and the resulting paralogs, homologous genes were combined into groups. The UMI counts for all homologous genes within each group were summed to obtain a total UMI count for a species. An expression matrix for each species was then generated, with each row representing a gene group, and each column representing a genuine cell.

The gene expression matrices for samples from the two species were normalized using function “SCTransform”. The top 3,000 genes that showed highest variability across samples were selected using functions “SelectIntegrationFeatures” and “PrepSCTIntegration”. The “reference-based” Seurat integration workflow (version 3.2.3) was then used to integrate the expression datasets, with the N group sample as the reference. Specifically, functions “FindIntegrationAnchors” and “IntegrateData” were used to identify common variation across datasets while removing batch effects. This computational method involves a combination of canonical correlation analysis (CCA) and mutual nearest neighbors (MNN).

### Functional enrichment analysis

The functional enrichment analysis of gene sets was performed on the website DAVID (<https://david.ncifcrf.gov/tools>).

A single-nucleus resolution atlas of soybean nodules

(jsp) (Huang et al., 2009a, 2009b). The terms in “GO-TERM\_BP\_DIRECT”, “KEGG\_PATHWAY”, “UP\_KW\_BIOLOGICAL\_PROCESS” categories were used.

### RNA velocity analysis

A file containing spliced and unspliced UMI counts for each gene (i.e., the loom file) was generated using function “run10x” in software “velocyto” (version 0.17.17) (La Manno et al., 2018). The resulting loom file was then imported into the Python environment and all downstream analyses were conducted with Python package “scVelo” (version 0.2.3). Data preprocessing, including gene selection, normalization, moment computation among nearest neighbors in PCA space, was performed using functions “scv.pp.filter\_and\_normalize” (with min\_shared\_counts = 30, n\_top\_genes = 3,000) and “scv.pp.moments” (with n\_pcs = 30, n\_neighbors = 20). RNA velocities were then estimated using either the default model with function “scv.tl.velocity” (mode = “stochastically”) or the dynamical model with functions “scv.tl.recover\_dynamics” and “scv.tl.velocity” (mode = “dynamical”). The dynamical model simultaneously generated a model-based estimation of the extent to which a gene is dynamically expressed, which was then used to select putative driver genes. The latent time of individual cells along a developmental trajectory was estimated using function “scv.tl.latent\_time”.

### Bulk RNA-seq

Total RNA was extracted from fresh plant tissues or nuclei using TRIzol. RNA-seq libraries were constructed and sequenced by the Illumina NovaSeq. 6000 in paired-end mode with 150-nt reads. Sequencing reads of low quality (>50% bases with a Phred score < 20) were filtered. After sequencing adapters were removed by software “fastp” (Chen et al., 2018), the clean reads were aligned to the soybean reference genome (version v2.1) and counted for each gene using STAR (version 2.6.0a, with parameter “--quantMode TranscriptomeSAM GeneCounts”). Differentially expressed genes between W82 and *Gmcre1s* plants were identified from three biological repeats using R package “edgeR” (version 3.28.1) (Robinson et al., 2010), with criteria of >2-fold difference and FDR < 0.05.

### Plasmid construction

To generate the *promoter:GUS* constructs, the promoter regions (~3,000 bp) of soybean cluster-enriched genes were amplified and subcloned into the entry vector pGWCm. The promoter constructs were transferred from the entry vectors to pBGWFS7, which contains the open reading frame of GUS, via LR reactions (Invitrogen). For overexpression constructs, the coding sequences of *GmbHLH93* and *GmSCL1* were cloned into vector pUB-GFP-Flag by a Seamless Assembly cloning kit (CloneSmarter). The detailed information regarding promoter lengths and primers can be found in Table S6.

### Hairy roots transformation and the generation of stable transgenic soybean

The overexpression constructs were transformed into *Agrobacterium rhizogenes* strain K599, which was further used to infect soybean seedlings following a previously published transformation protocol (Kereszt et al., 2007), with minor modifications. Briefly, hypocotyls of 5-d-old soybean seedlings were injected with K599 and kept at high humidity (>90%) with plastic lids, until hairy roots at the injection sites had grown to 5–10 cm long. The primary roots were then cut and the hairy roots were immersed in water for 7 d before being transferred into pots with sterilized vermiculite and inoculated with 50 mL of *Sinorhizobium fredii* (CCBAU45436) suspension ( $OD_{600} = 0.05$ ) per pot.

The *Gmcre1s* mutant (*GmCRE1a/b/c/d* quadruple knockout line) was created using CRISPR/Cas9 gene editing. The CRISPR/Cas9 vector contained gRNAs (AACCTGGTCAATGGCA-GAAGGG and GGTTGAAACAAGGATGGCAAGGG, with the underlined sequences representing the target regions shared by *GmCRE1a/b/c/d*), which were transformed into W82 using the *A. tumefaciens* EHA105 strain. The putative editing region of the resulting transgenic plants were amplified using gene-specific oligos, and the PCR products were sequenced to identify transgenic lines with gene editing. The stable transgenic line used for analysis was homozygous for all four genes (*GmCRE1a/b/c/d*) without Cas9 elements. The primer sequences are listed in Table S6.

### DNA extraction

DNA was extracted from fresh soybean samples using the CTAB (cetyltrimethylammonium bromide) method. Briefly, 600  $\mu$ L of CTAB buffer was added to each sample, followed by grinding and incubation at 65°C for 20 min. After cooling, 200  $\mu$ L of chloroform was added and mixed thoroughly to extract genomic DNA. The aqueous phase was collected by centrifugation at 12,000 r/min for 15 min and transferred to 1.5 mL tubes. The DNA was further purified by precipitation using ethanol.

### Quantitative reverse transcription PCR

Total RNA was extracted from soybean hairy roots with TRIzol. First-strand cDNA was synthesized from 1  $\mu$ g of RNA using the HiScript III RT SuperMix for qPCR Kit with gDNA Wiper (Vazyme) according to the manufacturer's instructions. The reaction mixture was prepared following the protocol of SYBR® Green *Pro Taq* HS kit (Accurate Biology). The qRT-PCR experiments were carried out using the qTOWER<sup>3</sup> real-time PCR detection system (Analytik Jena). The reference gene *GmABCT* was used for normalization. The primer sequences are listed in Table S6.

### Histological analysis and microscopy

For GUS staining analysis, transgenic hairy roots were submerged in a 100 mM PBS solution (pH = 7.0) containing 10 mM EDTA disodium salt, 100 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.5 mM K<sub>4</sub>Fe(CN)<sub>6</sub>, 0.5 mM K<sub>3</sub>Fe(CN)<sub>6</sub>, 0.1% Triton-X100, and 1 mg/mL X-gluc (1758–0600, Inalco) at 37°C for at least 2 h. The stained roots and nodules were embedded in 3% (m/v) low-gelling temperature agarose (111860, Biowest), sectioned at 50  $\mu$ m using a Leica

VT1100S, and observed and photographed with a ZEISS Axio Scope A1 pol microscope.

### Acetylene reduction assay

Nodules (14 dpi) were harvested and placed in 20-mL vials, followed by injection with 2 mL of acetylene. The vials were then incubated for 1 h at 25°C. The amount of ethylene produced was measured using gas chromatography (GC9310-VI) through an acetylene reduction assay (ARA).

### Data availability statement

All sequencing data and process data have been deposited to the National Genomics Data Center (<https://ngdc.cncb.ac.cn/>), Beijing Institute of Genomics, Chinese Academy of Sciences, under the BioProject number PRJCA015369. Codes to analyze the data and generate figures are available at GitHub ([https://github.com/Yuwang-art/snRNA\\_seq\\_in\\_soybean](https://github.com/Yuwang-art/snRNA_seq_in_soybean)).

## ACKNOWLEDGEMENTS

We thank C. Tian (China Agricultural University) for providing the *Sinorhizobium fredii* strain CCBAU45436. This work was supported by the CAS Project for Young Scientists in Basic Research (YSBR-011) and National Key Research and Development Program of China (2021YFF1000103).

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

B. S., Q. Y., and B. R. conceived the research, B. S., H. G., H. N., Y. L., Q. Y., and Q. M. performed the experiments, Y. W., B. S., W. Q., and B. R. analyzed the data, W. Q. and B. R. wrote the manuscript with input from B. S., Y. W. and Q. H. All authors read and approve this manuscript.

**Edited by:** Zhizhong Gong, China Agricultural University, China

**Received** Apr. 14, 2023; **Accepted** Apr. 17, 2023; **Published** Apr. 19, 2023

**OO:** OnlineOpen

## REFERENCES

- Arthikala, M.K., Montiel, J., Sánchez-López, R., Nava, N., Cárdenas, L., and Quinto, C. (2017). Respiratory burst oxidase homolog gene A is crucial for *Rhizobium* infection and nodule maturation and function in common bean. *Front. Plant Sci.* **8**: 2003.
- Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**: 1408–1414.



- Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the *Arabidopsis* root. *Science* **302**: 1956–1960.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**: P10008.
- Boivin, S., Kazmierczak, T., Brault, M., Wen, J., Gamas, P., Mysore, K. S., and Frugier, F. (2016). Different cytokinin histidine kinase receptors regulate nodule initiation as well as later nodule developmental stages in *Medicago truncatula*. *Plant Cell Environ.* **39**: 2198–2209.
- Carter, A.M., and Tegeder, M. (2016). Increasing nitrogen fixation and seed development in soybean requires complex adjustments of nodule nitrogen metabolism and partitioning processes. *Curr. Biol.* **26**: 2044–2051.
- Cervantes-Pérez, S.A., Thibivilliers, S., Laffont, C., Farmer, A.D., Frugier, F., and Libault, M. (2022). Cell-specific pathways recruited for symbiotic nodulation in the *Medicago truncatula* legume. *Mol. Plant* **15**: 1868–1888.
- Chen, J., Wang, Z., Wang, L., Hu, Y., Yan, Q., Lu, J., Ren, Z., Hong, Y., Ji, H., Wang, H., et al. (2022). The B-type response regulator GmRR11d mediates systemic inhibition of symbiotic nodulation. *Nat. Commun.* **13**: 7661.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **17**: i884–i890.
- Cooper, J.B., and Long, S.R. (1994). Morphogenetic rescue of *Rhizobium meliloti* nodulation mutants by *trans*-zeatin secretion. *Plant Cell* **6**: 215–225.
- Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P. (2019). Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* **48**: 840–852.e845.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
- Doyle, J.J., Chappill, J.A., Bailey, C.D., and Kajita, T. (2000). Towards a comprehensive phylogeny of legumes: Evidence from RBCL sequences and non-monoclecular data. In *Advances in Legume Systematics*, P.S. Herendeen, A. Bruneau, eds. (Kew: Royal Botanic Gardens). pp. 1–20.
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**: 238.
- Fan, W., Xia, C., Wang, S., Liu, J., Deng, L., Sun, S., and Wang, X. (2022). Rhizobial infection of 4C cells triggers their endoreduplication during symbiotic nodule development in soybean. *New Phytol.* **234**: 1018–1030.
- Feng, J., Lee, T., Schiessl, K., and Oldroyd, G.E.D. (2021). Processing of NODULE INCEPTION controls the transition to nitrogen fixation in root nodules. *Science* **374**: 629–632.
- Ferguson, B.J., Mens, C., Hastwell, A.H., Zhang, M.B., Su, H.A., Jones, C.H., Chu, X.T., and Gresshoff, P.M. (2019). Legume nodulation: The host controls the party. *Plant Cell Environ.* **42**: 41–51.
- Gao, Z., Chen, Z., Cui, Y., Ke, M., Xu, H., Xu, Q., Chen, J., Li, Y., Huang, L., Zhao, H., et al. (2021). GmPIN-dependent polar auxin transport is involved in soybean nodule development. *Plant Cell* **33**: 2981–3003.
- Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. (2006). The *medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. *Plant Cell* **18**: 2680–2693.
- Hadri, A.-E., Spaink, H.P., Bisseling, T., and Brewin, N.J. (1998). Diversity of root nodulation and rhizobial infection processes. In *The Rhizobiaceae: Molecular Biology of Model Plant-Associated Bacteria*, H.P. Spaink, A. Kondorosi, P.J.J. Hooykaas, eds. Dordrecht: Springer Netherlands. pp. 347–360.
- Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**: 296.
- Han, Y., Chu, X., Yu, H., Ma, Y., Wang, X., Qian, W., and Jiao, Y. (2017). Single-cell transcriptome analysis reveals widespread monoallelic gene expression in individual rice mesophyll cells. *Sci. Bull. (Beijing)* **62**: 1304–1314.
- Heckmann, A.B., Sandal, N., Bek, A.S., Madsen, L.H., Jurkiewicz, A., Nielsen, M.W., Tirichine, L., and Stougaard, J. (2011). Cytokinin induction of root nodule primordia in *Lotus japonicus* is regulated by a mechanism operating in the root cortex. *Mol. Plant Microbe. Interact.* **24**: 1385–1395.
- Held, M., Hou, H., Miri, M., Huynh, C., Ross, L., Hossain, M.S., Sato, S., Tabata, S., Perry, J., Wang, T.L., and Szczyglowski, K. (2014). *Lotus japonicus* cytokinin receptors work partially redundantly to mediate nodule formation. *Plant Cell* **26**: 678–694.
- Hirsch, A.M. (1992). Developmental biology of legume nodulation. *New Phytol.* **122**: 211–237.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**: 1–13.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**: 44–57.
- Huo, X., Schnabel, E., Hughes, K., and Frugoli, J. (2006). RNAi phenotypes and the localization of a Protein::GUS fusion imply a role for *Medicago truncatula* PIN genes in nodulation. *J. Plant Growth Regul.* **25**: 156–165.
- Jardinaud, M.-F., Boivin, S., Rodde, N., Catrice, O., Kisiala, A., Lepage, A., Moreau, S., Roux, B., Cottret, L., Sallet, E., et al. (2016). 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.
- Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorrity, M.W., Saunders, L., Bubba, K.L., Trapnell, C., Fields, S., Queitsch, C., and Cuperus, J.T. (2019). Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* **31**: 993–1011.
- Ji, H., Xiao, R., Lyu, X., Chen, J., Zhang, X., Wang, Z., Deng, Z., Wang, Y., Wang, H., Li, R., et al. (2022). Differential light-dependent regulation of soybean nodulation by papilionoid-specific HY5 homologs. *Curr. Biol.* **32**: 783–795.
- Jin, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., Zhang, X., Dai, H., Yang, J., and Wang, E. (2016). DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. *Nat. Commun.* **7**: 12433.
- Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C.D., Nontachaiyapoom, S., Kinkema, M., and Gresshoff, P.M. (2007). *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nat. Protoc.* **2**: 948–952.
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastrioti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. *Nature* **560**: 494–498.
- Laffont, C., Rey, T., André, O., Novero, M., Kazmierczak, T., Debelle, F., Bonfante, P., Jacquet, C., and Frugier, F. (2015). The CRE1 cytokinin pathway is differentially recruited depending on *Medicago truncatula* root environments and negatively regulates resistance to a pathogen. *PLoS ONE* **10**: e0116819.
- Lavin, M., Herendeen, P., and Wojciechowski, M. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst. Biol.* **54**: 575–594.
- Layzell, D.B., Rainbird, R.M., Atkins, C.A., and Pate, J.S. (1979). Economy of photosynthate use in nitrogen-fixing legume nodules: Observations on two contrasting symbioses. *Plant Physiol.* **64**: 888–891.

- Liu, Z., Kong, X., Long, Y., Liu, S., Zhang, H., Jia, J., Cui, W., Zhang, Z., Song, X., Qiu, L., et al. (2023). Integrated single-nucleus and spatial transcriptomics captures transitional states in soybean nodule maturation. *Nat. Plants* **9**: 515–524.
- Liu, Q., Liang, Z., Feng, D., Jiang, S., Wang, Y., Du, Z., Li, R., Hu, G., Zhang, P., Ma, Y., et al. (2021). Transcriptional landscape of rice roots at the single-cell resolution. *Mol. Plant* **14**: 384–394.
- Lohar, D.P., Sharopova, N., Endre, G., Penuela, S., Samac, D., Town, C., Silverstein, K.A., and VandenBosch, K.A. (2006). Transcript analysis of early nodulation events in *Medicago truncatula*. *Plant Physiol.* **140**: 221–234.
- Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., Marioni, J.C., and participants in the 1st Human Cell Atlas, J. (2019). EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**: 63.
- Luo, Y., Liu, W., Sun, J., Zhang, Z.-R., and Yang, W.-C. (2023). Quantitative proteomics reveals key pathways in the symbiotic interface and the likely extracellular property of soybean symbiosome. *J. Genet. Genomics* **50**: 7–19.
- Marino, D., Andrio, E., Danchin, E.G., Oger, E., Gucciardo, S., Lambert, A., Puppo, A., and Pauly, N. (2011). A *Medicago truncatula* NADPH oxidase is involved in symbiotic nodule functioning. *New Phytol.* **189**: 580–592.
- McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for dimension reduction. [arXiv:1802.03426](https://arxiv.org/abs/1802.03426).
- Mortier, V., Wasson, A., Jaworek, P., De Keyser, A., Decroos, M., Holsters, M., Tarkowski, P., Mathesius, U., and Goormachtig, S. (2014). Role of *LONELY GUY* genes in indeterminate nodulation on *Medicago truncatula*. *New Phytol.* **202**: 582–593.
- Murray, J.D., Karas, B.J., Sato, S., Tabata, S., Amyot, L., and Szczygłowski, K. (2007). A cytokinin perception mutant colonized by *Rhizobium* in the absence of nodule organogenesis. *Science* **315**: 101–104.
- Ng, J.L., Hassan, S., Truong, T.T., Hocart, C.H., Laffont, C., Frugier, F., and Mathesius, U. (2015). Flavonoids and auxin transport inhibitors rescue symbiotic nodulation in the *Medicago truncatula* cytokinin perception mutant *cre1*. *Plant Cell* **27**: 2210–2226.
- Nguyen, N.N.T., Clua, J., Vetal, P.V., Vuarambon, D.J., De Bellis, D., Pervent, M., Lepetit, M., Udvardi, M., Valentine, A.J., and Poirier, Y. (2021). PHO1 family members transport phosphate from infected nodule cells to bacteroids in *Medicago truncatula*. *Plant Physiol.* **185**: 196–209.
- Oldroyd, G.E.D. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**: 252–263.
- Petricka, J.J., Winter, C.M., and Benfey, P.N. (2012). Control of *Arabidopsis* root development. *Annu. Rev. Plant Biol.* **63**: 563–590.
- Plet, J., Wasson, A., Ariel, F., Le Signor, C., Baker, D., Mathesius, U., Crespi, M., and Frugier, F. (2011). MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in *Medicago truncatula*. *Plant J.* **65**: 622–633.
- Reid, D.E., Heckmann, A.B., Novak, O., Kelly, S., and Stougaard, J. (2016). CYTOKININ OXIDASE/DEHYDROGENASE3 maintains cytokinin homeostasis during root and nodule development in *Lotus japonicus*. *Plant Physiol.* **170**: 1060–1074.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Roy, S., Liu, W., Nandety, R.S., Crook, A., Mysore, K.S., Pislariu, C.I., Frugoli, J., Dickstein, R., and Udvardi, M.K. (2020). Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *Plant Cell* **32**: 15–41.
- Schauser, L., Roussis, A., Stiller, J., and Stougaard, J. (1999). A plant regulator controlling development of symbiotic root nodules. *Nature* **402**: 191–195.
- Shahan, R., Hsu, C.W., Nolan, T.M., Cole, B.J., Taylor, I.W., Greenstreet, L., Zhang, S., Afanassiev, A., Vlot, A.H.C., Schiebinger, G., et al. (2022). A single-cell *Arabidopsis* root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Dev. Cell* **57**: 543–560.e549.
- Smith, P.M., and Atkins, C.A. (2002). Purine biosynthesis. Big in cell division, even bigger in nitrogen assimilation. *Plant Physiol.* **128**: 793–802.
- Sprent, J.I., and McKey, D. (1994). The nitrogen factor. In *Advances in Legume Systematics*, J.I. Sprent, D. McKey, eds. (Kew: Royal Botanic Gardens). pp. 241.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* **177**: 1888–1902.e21.
- Subramanian, S., Stacey, G., and Yu, O. (2006). Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *Plant J.* **48**: 261–273.
- Tegeder, M. (2014). Transporters involved in source to sink partitioning of amino acids and ureides: Opportunities for crop improvement. *J. Exp. Bot.* **65**: 1865–1878.
- Thibivilliers, S., Anderson, D., and Libault, M. (2020). Isolation of plant root nuclei for single cell RNA sequencing. *Curr. Protoc. Plant Biol.* **5**: e20120.
- Tirichine, L., Sandal, N., Madsen, L.H., Radutoiu, S., Albrektsen, A.S., Sato, S., Asamizu, E., Tabata, S., and Stougaard, J. (2007). A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* **315**: 104–107.
- Valentine, A.J., Kleinert, A., and Benedito, V.A. (2017). Adaptive strategies for nitrogen metabolism in phosphate deficient legume nodules. *Plant Sci.* **256**: 46–52.
- Velandia, K., Reid, J.B., and Foo, E. (2022). Right time, right place: The dynamic role of hormones in rhizobial infection and nodulation of legumes. *Plant Commun.* **3**: 100327.
- Wang, L., Zhou, Y., Li, R., Liang, J., Tian, T., Ji, J., Chen, R., Zhou, Y., Fan, Q., Ning, G., et al. (2022). Single cell-type transcriptome profiling reveals genes that promote nitrogen fixation in the infected and uninfected cells of legume nodules. *Plant Biotechnol. J.* **20**: 616–618.
- Wang, T., Guo, J., Peng, Y., Lyu, X., Liu, B., Sun, S., and Wang, X. (2021). Light-induced mobile factors from shoots regulate rhizobium-triggered soybean root nodulation. *Science* **374**: 65–71.
- Wang, Y., Huan, Q., Li, K., and Qian, W. (2021). Single-cell transcriptome atlas of the leaf and root of rice seedlings. *J. Genet. Genomics* **48**: 881–898.
- Weiler, P., Van den Berge, K., Street, K., and Tiberi, S. (2023). A guide to trajectory inference and RNA velocity. In *Single Cell Transcriptomics: Methods and Protocols*, R.A. Calogero, V. Benes, eds. (New York, NY: Springer US). pp. 269–292.
- Witte, C.P., and Herde, M. (2020). Nucleotide metabolism in plants. *Plant Physiol.* **182**: 63–78.
- Xiao, T.T., Schilderink, S., Moling, S., Deinum, E.E., Kondorosi, E., Franssen, H., Kulikova, O., Niebel, A., and Bisseling, T. (2014). Fate map of *Medicago truncatula* root nodules. *Development* **141**: 3517–3528.
- Xu, H., Li, Y., Zhang, K., Li, M., Fu, S., Tian, Y., Qin, T., Li, X., Zhong, Y., and Liao, H. (2021). miR169c-NFYA-C-ENOD40 modulates nitrogen inhibitory effects in soybean nodulation. *New Phytol.* **229**: 3377–3392.
- Ye, Q., Zhu, F., Sun, F., Wang, T.-C., Wu, J., Liu, P., Shen, C., Dong, J., and Wang, T. (2022). Differentiation trajectories and biofunctions of symbiotic and un-symbiotic fate cells in root nodules of *Medicago truncatula*. *Mol. Plant* **15**: 1852–1867.
- Yun, J., Sun, Z., Jiang, Q., Wang, Y., Wang, C., Luo, Y., Zhang, F., and Li, X. (2022). The miR156b-GmSPL9d module modulates nodulation by

targeting multiple core nodulation genes in soybean. *New Phytol.* **233**: 1881–1899.

**Zhang, T., Chen, Y., and Wang, J.** (2021). A single-cell analysis of the *Arabidopsis* vegetative shoot apex. *Dev. Cell* **56**: 1056–1074.

**Zhang, T., Xu, Z., Shang, G., and Wang, J.** (2019). A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant* **12**: 648–660.

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

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: <http://onlinelibrary.wiley.com/doi/10.1111/jipb.13495/supinfo>

**Figure S1.** DAPI and trypan blue staining of nuclei isolated from soybean roots and nodules

## A single-nucleus resolution atlas of soybean nodules

**Figure S2.** The correlation of the expression level estimated from snRNA-seq (i.e., pseudobulk) and bulk RNA-seq among genes

**Figure S3.** Expression of reported nodulation-related genes (summarized by Roy et al., 2020) in the snRNA-seq soybean data

**Figure S4.** Expression of soybean homologs of previously reported marker genes in *Arabidopsis* roots in the snRNA-seq soybean data

**Figure S5.** Expression of cluster-enriched genes in soybean roots and nodules

**Figure S6.** Comparison of single cell/nucleus transcriptomes between soybean and *M. truncatula*

**Figure S7.** Expression of genes in the cytokinin pathway among cells

**Figure S8.** Characterization of soybean *Gmcre1s* mutants

**Figure S9.** The proportion of spliced and unspliced transcripts in soybean roots and nodules

**Table S1.** Statistics of snRNA-seq in soybean roots and nodules

**Table S2.** List of marker genes reported in *Arabidopsis*

**Table S3.** List of marker genes used in this study

**Table S4.** The soybean genes in the purine pathway

**Table S5.** Cytokinin related genes showing enriched expression in soybean snRNA-seq

**Table S6.** Primers used in this study



Scan using WeChat with your smartphone to view JIPB online



Scan with iPhone or iPad to view JIPB online