

Single-cell transcriptome atlases of soybean root and mature nodule reveal new regulatory programs that control the nodulation process

Sergio Alan Cervantes-Pérez^{1,9,11}, Prince Zogli^{1,10,11}, Sahand Amini^{2,3}, Sandra Thibivilliers^{2,3}, Sutton Tennant^{2,3}, Md Sabbir Hossain^{2,3}, Hengping Xu^{2,3}, Ian Meyer¹, Akash Nooka¹, Pengchong Ma⁴, Qiuming Yao⁴, Michael J. Naldrett⁵, Andrew Farmer⁶, Olivier Martin⁷, Samik Bhattacharya⁸, Jasper Kläver⁸ and Marc Libault^{2,3,*}

¹Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68503, USA

²Division of Plant Science and Technology, College of Agriculture, Food, and Natural Resources, University of Missouri-Columbia, Columbia, MO 65211, USA

³Interdisciplinary Plant Group of Missouri-Columbia, Columbia, MO 65211, USA

⁴School of Computing, University of Nebraska-Lincoln, Lincoln, NE 68503, USA

⁵Proteomics and Metabolomics Facility, Center for Biotechnology, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

⁶National Center for Genome Resources, Santa Fe, NM 87505, USA

⁷INRAE, Université Paris-Saclay, Institut des Sciences des Plantes de Paris Saclay, IPS2, Batiment 630 Plateau du Moulon, Rue Noetzlin, 91192 Gif sur Yvette Cedex, France

⁸Resolve Biosciences, Monheim am Rhein, Germany

⁹Present address: The School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

¹⁰Present address: GALY Co., 100 Hood Park Dr., Boston, MA 02126, USA

¹¹These authors contributed equally to this article.

*Correspondence: Marc Libault (libaultm@missouri.edu)

<https://doi.org/10.1016/j.xplc.2024.100984>

ABSTRACT

The soybean root system is complex. In addition to being composed of various cell types, the soybean root system includes the primary root, the lateral roots, and the nodule, an organ in which mutualistic symbiosis with N-fixing rhizobia occurs. A mature soybean root nodule is characterized by a central infection zone where atmospheric nitrogen is fixed and assimilated by the symbiont, resulting from the close cooperation between the plant cell and the bacteria. To date, the transcriptome of individual cells isolated from developing soybean nodules has been established, but the transcriptomic signatures of cells from the mature soybean nodule have not yet been characterized. Using single-nucleus RNA-seq and Molecular Cartography technologies, we precisely characterized the transcriptomic signature of soybean root and mature nodule cell types and revealed the co-existence of different sub-populations of *B. diazoefficiens*-infected cells in the mature soybean nodule, including those actively involved in nitrogen fixation and those engaged in senescence. Mining of the single-cell-resolution nodule transcriptome atlas and the associated gene co-expression network confirmed the role of known nodulation-related genes and identified new genes that control the nodulation process. For instance, we functionally characterized the role of GmFWL3, a plasma membrane microdomain-associated protein that controls rhizobial infection. Our study reveals the unique cellular complexity of the mature soybean nodule and helps redefine the concept of cell types when considering the infection zone of the soybean nodule.

Key words: soybean, root, nodule, single-cell RNA-seq, nanodomains

Cervantes-Pérez S.A., Zogli P., Amini S., Thibivilliers S., Tennant S., Hossain M.S., Xu H., Meyer I., Nooka A., Ma P., Yao Q., Naldrett M.J., Farmer A., Martin O., Bhattacharya S., Kläver J., and Libault M. (2024). Single-cell transcriptome atlases of soybean root and mature nodule reveal new regulatory programs that control the nodulation process. *Plant Comm.* 5, 100984.

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

INTRODUCTION

Legumes engage in a mutualistic symbiotic interaction with rhizobia, a group of nitrogen-fixing soil bacteria (e.g., *Bradyrhizobium diazoefficiens* for soybean) (Li et al., 2020). This biological process, termed nodulation, has economic and ecological impacts on agriculture, as it helps to mitigate the application of nitrogen fertilizers, reduces environmental pollution, and supports sustainable agricultural practices. For most legume species (e.g., *Medicago truncatula*, *Lotus japonicus*, *Glycine max*, *Phaseolus vulgaris*, *Pisum sativum*), the nodulation process is initiated by rhizobial infection of the root hair cells. Concomitant with this infection, the root inner cortical cells engage in *de novo* cell divisions, leading to the formation of nodule primordia. Ultimately, a new root organ, the nodule, emerges, in which differentiated bacteria called bacteroids fix and assimilate atmospheric nitrogen for the plant (Udvardi and Poole, 2013; De La Peña et al., 2018). On the basis of their organogenesis and cellular organization, mature legume nodules are generally classified into two types: indeterminate and determinate (Brewin, 1991; Ferguson et al., 2010). Mature indeterminate nodules (e.g., *M. truncatula* and *P. sativum*) can be divided into four biologically and microscopically distinct zones. In zone I, at the tip of the nodule, a permanent nodule meristem persists. In zone II, rhizobia infect the plant cells. Zone III is the nitrogen fixation zone, where bacteroids fix and assimilate atmospheric nitrogen for the plant. In zone IV, the nodule cells senesce. Unlike indeterminate nodules, mature determinate nodules (e.g., *G. max*, *L. japonicus*, and *P. vulgaris*) are not organized into visually distinct zones associated with different stages of interaction between plant cells and rhizobia. As a result, all the plant cells colonized by rhizobia are located in the center of the nodule, and, over its lifetime, the determinate nodule will senesce outwards from the center (Puppo et al., 2005). At the molecular level, senescing nodule cells are characterized by induction of the expression of a NAC/CYP regulatory module and decreases in leghemoglobin content and nitrogenase activity (Buono et al., 2019; Doll, 2023; Yu et al., 2023).

During the past 20 years, numerous -omics studies of different legume species have led to the identification of many genes that control the nodulation process, including those involved in the symbiosis between infected nodule cells and bacteroids (Libault et al., 2010a; Breakspear et al., 2014; Clarke et al., 2015; Velicković et al., 2018; Mergaert et al., 2020; Roy et al., 2020; Shimoda et al., 2020; Islam et al., 2022). The emergence of high-throughput sequencing technologies has led to the construction of several legume transcriptomic atlases that have notably revealed differences in the transcriptomic profiles of the root and the nodule and have enabled the identification of hundreds of nodule-specific genes (Benedito et al., 2008; Libault et al., 2010c; Severin et al., 2010; Verdier et al., 2013). For example, the transcriptome of each of the four zones of the *M. truncatula* nodule was established using laser microdissection (Roux et al., 2014). This technology has recently been superseded by emerging single-cell and single-nucleus transcriptomic technologies (sc- and sNucRNA-seq) (Denyer et al., 2019; Jean-Baptiste et al., 2019; Lee et al., 2019; Ryu et al., 2019; Wendrich et al., 2020; Liu et al., 2021; Xu et al., 2021; Shahan et al., 2022). For instance, single-cell RNA-

seq was recently used to capture the transcriptomic profile of indeterminate *M. truncatula* nodule cells and confirmed the zone-specific transcriptomic programs of infected nodule cells (Ye et al., 2022). A similar approach was recently applied to developing and maturing soybean nodules (i.e., 12-, 14-, and 21-days post rhizobial inoculation [dpi]; Liu et al., 2023; Sun et al., 2023). Here, to complement these studies, we report the use of single-nucleus RNA-seq technology on soybean roots and 28-dpi nodules, which are transitioning to their senescence phase as reported by Yu et al. (2023).

Our study provides a new perspective on the cellular and molecular complexity of the infection zone of the soybean nodule. Specifically, we report that different transcriptomic programs are specifically activated in three different sub-populations of rhizobial-infected soybean cells: those not actively fixing atmospheric dinitrogen, those fixing and assimilating atmospheric nitrogen, and those already engaged in senescence. Single-cell-resolution gene co-expression networks not only reveal interactions between known nodulation-related genes but also support the identification of new candidate genes that control the symbiosis between soybean cells and rhizobia. For instance, we provide functional evidence for the role of *GmFWL3*, a homolog of the plasma membrane microdomain-encoding gene *GmFWL1* (Thibivilliers et al., 2020a), in control of soybean cell infection by *B. diazoefficiens*. Our study reveals the unique cellular complexity of the mature soybean nodule, enabling a deeper understanding of the molecular processes that govern nodulation.

RESULTS

Single-cell-resolution transcriptome atlas of the soybean root

To establish the transcriptomic profile of each cell type in the soybean root, we applied sNucRNA-seq technology to three independent soybean root replicates (Supplemental Figure 1). After independent processing to eliminate doublets and background contamination (see Supplemental Figure 1A–1C), we found that the transcriptomes of the three root replicates were strongly correlated (Supplemental Figure 1D). Subsequently, the replicates were integrated, followed by dimensional reduction (see methods). The single-nucleus root transcriptome atlas is composed of 14 369 high-quality nuclei, captures an average of 1949 unique molecular identifiers (UMIs) and 1363 expressed genes per nucleus, and covers the expression of 75.8% of the predicted soybean protein-coding genes (42 390 out of 55 897; Supplemental Table 1). This root atlas nicely overlaps with but also better covers the previously reported bulked transcriptomes of the soybean root and root tip, which identified 39 709 and 36 354 expressed genes, respectively (Libault et al., 2010c) (i.e., 88.1% and 81.6% of the expressed genes identified using sNucRNA-seq were also identified in the root and root-tip bulk transcriptomes, respectively; Supplemental Figure 2A). After application of the Uniform Manifold Approximation and Projection (UMAP) dimensional reduction technique, the root nuclei were distributed in 16 distinct cell clusters according to their transcriptomic profiles (Figure 1A). Except for root hair cell cluster 3, which is characterized by slightly higher numbers of UMIs and

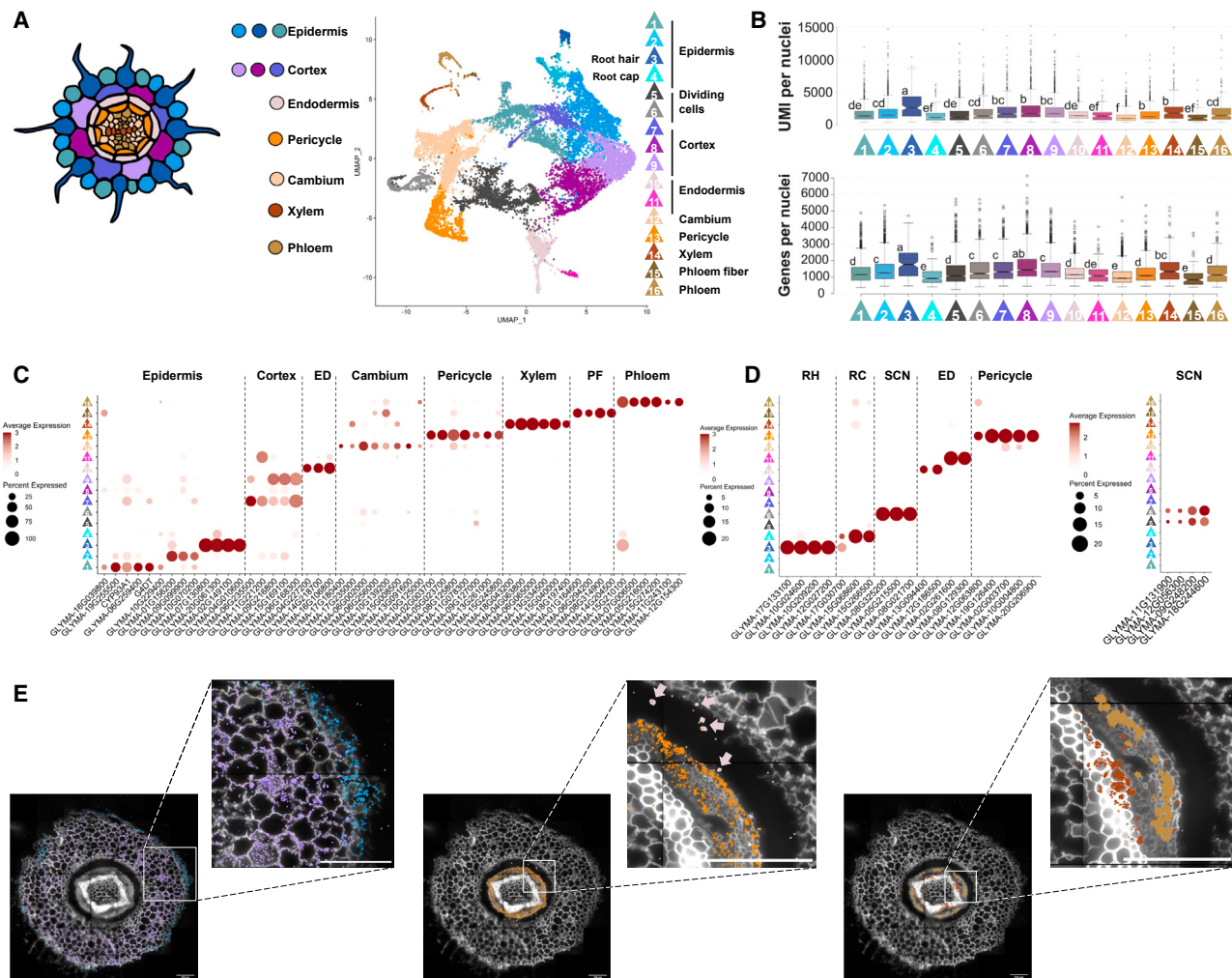


Figure 1. Establishment of a single-nucleus transcriptome atlas of the soybean root.

(A) Uniform Manifold Approximation and Projection (UMAP) plot of 14 369 soybean root nuclei based on their transcriptomic profiles. The nuclei were clustered into 16 different groups.

(B) Distribution of the number of unique molecular identifiers (UMIs) and expressed genes per root cluster (Tukey's test with $p < 0.05$ reported to highlight differences between clusters).

(C) Dotplot representation of the expression of 52 soybean root cell-type-specific marker genes validated using M.C. technology (Supplemental Figure 3).

(D) Dotplot representations of the expression of root cell-type-specific marker genes identified on the basis of previous functional genomics studies and their orthologous relationships with root cell-type-specific marker genes from *Arabidopsis thaliana* and *Medicago truncatula* (Supplemental Figure 4). For these two dotplot figures, the percentage of nuclei expressing the gene of interest (circle size) and the mean expression of the gene (circle color) are shown.

(E) Integrated analysis of the expression of several soybean root marker genes using M.C. technology on a soybean root cross-section. Left panel: detection of transcripts from epidermal (blue) and cortical (purple) marker genes. Central panel: detection of transcripts from endodermal (light pink color; see arrows in the magnified picture) and pericycle (orange) marker genes. Right panel: detection of transcripts from xylem (red) and phloem (brown) marker genes. ED, endodermis; PF, phloem fiber; RH, root hair cells; RC, root cap cells; SCN, stem cell niche (see methods for details).

expressed genes, the remaining 15 clusters have similar transcriptional activities (Figure 1B). From the saturation of our sequencing, we estimated that the transcriptome coverage of each root cluster varies from 57.5% (i.e., cluster 15) to 99.1% (i.e., clusters 2, 8, and 9). Not surprisingly, this coverage depends on the size of the population of nuclei per cluster. Except for clusters 11 (123 nuclei) and 15 (89 nuclei), the transcriptomic coverage of the root clusters is greater than 85% (Supplemental Table 2). By examining the expression of the 55 897 protein-coding genes across the 16 root clusters, we identified 14 088 ubiquitously expressed genes (i.e., genes

found expressed in all clusters). Among them, 2753 are constitutively expressed across the 16 clusters (i.e., less than a 4-fold change in activity between the clusters in which the gene shows the highest and lowest expression) (Supplemental Table 1). Applying very stringent criteria (i.e., fold-change >20 between the two clusters in which the gene is most highly expressed; minimum expression of 0.1 UMIs for the gene considered/10 000 sequenced UMIs, expressed in at least 20% of nuclei in the cluster where the gene was found to be specifically expressed), we also identified 424 root cell-type marker genes (Supplemental Table 1).



To functionally annotate the root clusters, we first used Molecular Cartography (M.C.) technology developed by Resolve Biosciences, a multiplexed and high-resolution RNA *in situ* hybridization technique, on a cross-section of the soybean root. Specifically, we analyzed the transcriptional activity of 52 soybean cluster-specific genes selected from the 16 root clusters in the morphological context of the soybean root. The transcriptional patterns of these genes led to annotation of the epidermal, cortical, endodermal, pericycle, cambial, xylem, and phloem clusters (Figure 1C and 1E and Supplemental Figure 3). We assume that the small population of nuclei in cluster 15 and its lower transcriptomic coverage compared with other root clusters (Supplemental Table 2) result from the enucleation of phloem cells during their maturation. To accurately annotate the root hair cell cluster, which is a difficult cell type to assess from our M.C. experiments because of its unique morphology and peripheral localization in the root, we also examined the activity of previously and newly functionally validated root-hair-specific genes (Figure 1D and Supplemental Figure 4). To annotate the soybean “root cap” and “stem cell niche” clusters, cell types not represented on the root M.C. cross-section, we analyzed the activity of soybean genes orthologous to *M. truncatula* and *Arabidopsis thaliana* marker genes (Figure 1D and Supplemental Figure 4; Supplemental Table 3). Using the same strategy, we also confirmed the identity of clusters 10 and 11 as “endodermal” clusters and refined the identity of cluster 13 as the “pericycle” cluster. We thus functionally annotated the 16 soybean root clusters and experimentally validated the expression of a large collection of new marker genes of soybean root cell types. By examining the distribution of the 424 single-cell-type marker genes across the 16 annotated soybean root clusters, we found that these genes were restricted to 9 clusters (i.e., root hair cluster 3, root cap 4, dividing cells 6, cortex 7, endodermis 10 and 11, xylem 14, phloem fiber 15, and phloem 16; Supplemental Table 1), likely reflecting the biological specialization of these cell types.

Single-cell-resolution transcriptome atlas of the soybean nodule

The soybean nodule and the physiology of its cells change over time, a consequence of its continuous development from initiation of the nodule primordium to senescence of the mature nodule. To complement recent single-cell transcriptomic studies performed on the developing stages of the soybean nodule (i.e., 12, 14, and 21 dpi) (Liu et al., 2023; Sun et al., 2023), we

established a single-cell transcriptome atlas of 28-dpi mature soybean nodules. The 28-dpi nodule atlas was generated from two independent replicates (Supplemental Figure 5). As we did for the “root” sample, we independently processed the “nodule” replicates to eliminate doublets and background contamination (see Supplemental Figure 5A and 5B), confirmed the correlation between the two replicates (Supplemental Figure 5C), and performed dimensional reduction upon integration (see methods). The nodule atlas is composed of 7830 nuclei with an average of 1058 UMIs and 647 expressed genes per cell and a total of 37 119 expressed genes (~66.4% of the soybean protein-coding genes) (Supplemental Table 1). Like the root datasets, the sNucRNA-seq datasets from the soybean nodules overlap well with a previously published soybean nodule bulk transcriptome (Libault et al., 2010c) (i.e., 87.3% of the expressed genes identified using sNucRNA-seq technology were also identified in the nodule bulk transcriptome; Supplemental Figure 2C). The UMAP of the nodule nuclei revealed 11 different cell clusters named A to K (Figure 2A). Clusters F and G are characterized by significantly higher numbers of UMIs and expressed genes per nucleus compared with the other clusters, suggesting the higher transcriptomic activity of cells in these two clusters (Figure 2B). The transcriptomic coverage of the nodule clusters varies from 34.4% (i.e., cluster D, which contains only 21 nuclei) to 98.9% (i.e., clusters F and G). Except for clusters D and C (79.1% coverage, 115 nuclei), the transcriptomic coverage of the remaining nodule clusters is greater than 85% (Supplemental Table 2). Using the same parameters described above, we identified 250 nodule-cluster marker genes and 950 ubiquitously expressed genes across the 11 nodule clusters, but only 16 constitutively expressed genes (Supplemental Table 1). Among the 16 constitutively expressed genes of the nodule, 8 genes encoding proteins with fundamental biological functions were also identified as constitutively expressed in the soybean root system (i.e., Glyma.17G073300 [SRPR protein that supports protein translation at the endoplasmic reticulum], Glyma.19G248000 [ankyrin repeat-containing protein], Glyma.09G156600 [small subunit ribosomal protein S3e], Glyma.07G091800, Glyma.18G222400 [spastin, a microtubule-severing protein], Glyma.12G056500 [AN1-type zinc finger protein], Glyma.13G212500 [proline-rich nuclear receptor coactivator], and Glyma.08G262900 [UV excision repair protein RAD23]). The low number of ubiquitously and constitutively expressed genes in the nodule results from the lower transcriptomic depth of cluster D, a consequence of the very small population of

(C) Dotplot representation of the expression of 34 cell-type-specific marker genes of the soybean nodule validated using M.C. technology (Supplemental Figures 6 and 7).

(D) Dotplot representation of the expression of six nodule cluster I-specific marker genes. The dot sizes in (C) and (D) represent the percentage of cells in which each gene is expressed. For these two dotplot figures, the percentage of nuclei expressing the gene of interest (circle size) and the mean expression (circle color) of the gene are shown.

(E–H) Integrated analysis of the expression of several soybean nodule marker genes using M.C. technology on a soybean nodule cross-section.

(E) Detection of transcripts from the inner/outer cortical cells (blue) and the sclereid layer (pink). (F) Detection of transcripts from the vascular endodermis (light pink) and vascular bundle (orange). (G) Detection of *B. diazoefficiens* transcripts from nine different genes in infected nodule cells (yellow). (H) Detection of plant transcripts in *B. diazoefficiens*-infected (red) and uninfected cells (green; green arrows) (see methods for details).

(I) Identification of the population of 968 rhizobia-infected (yellow circles) and 1769 uninfected cells (black stars) of the nodule through a principal-component analysis (PCA) plot of nodule cells analyzed by M.C. technology. To generate these plots, the transcript numbers of the 10 *B. diazoefficiens* genes were taken into consideration.

(J) Violin plots of the density of the number of 10 different bacterial (left) and Glyma.17G195900 (right) transcripts in the population of 968 infected (yellow) and 1769 uninfected (gray) cells of the soybean nodule. A two-tailed Student's *t*-test of 0 supports the significant difference in expression of 10 bacterial genes and Glyma.17G195900 between *B. diazoefficiens*-infected and uninfected cells.

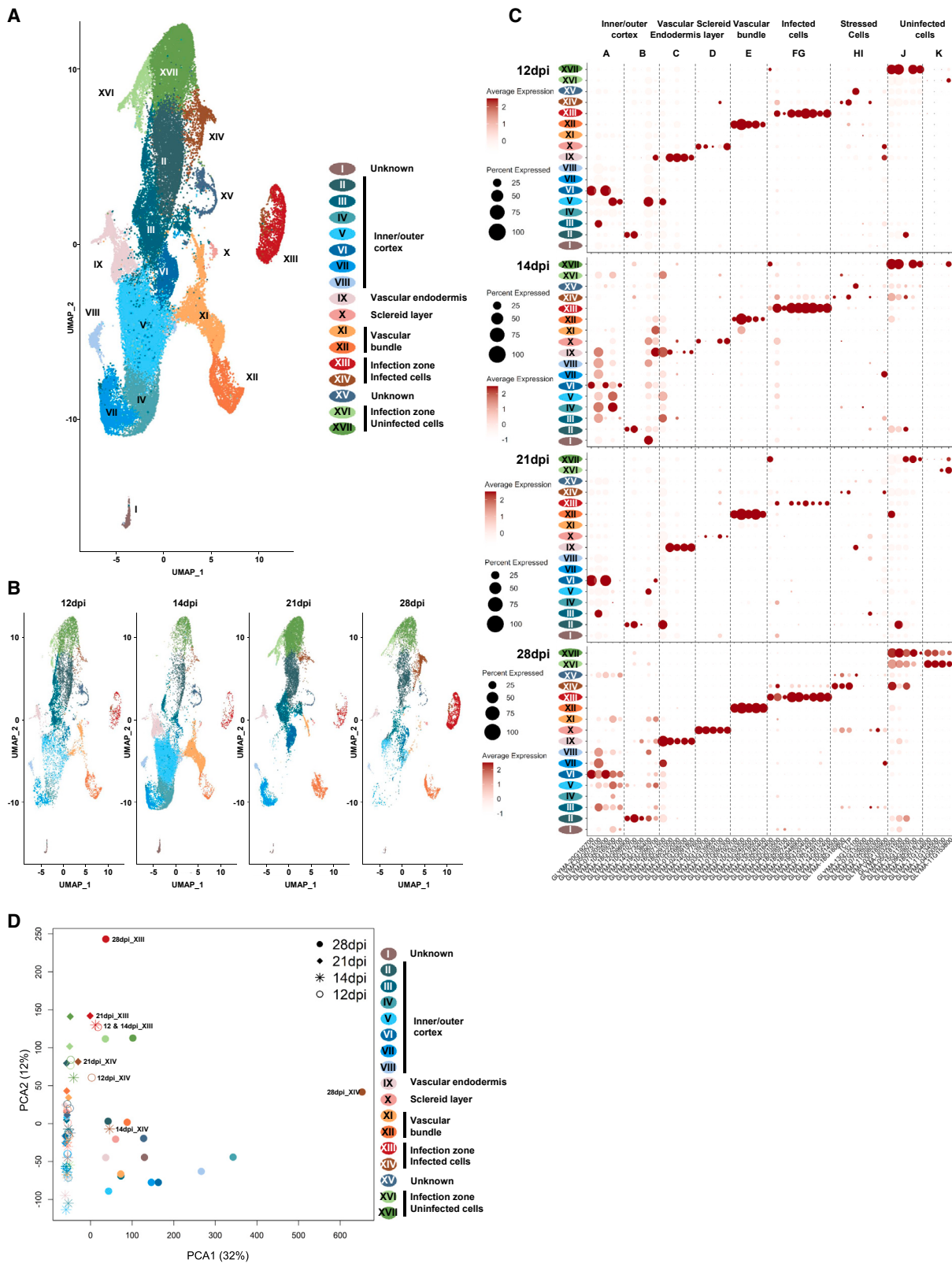


Figure 3. Integrated analysis of single-cell gene expression datasets during development of the soybean nodule.
(A) UMAP projection and integration of soybean nodule transcriptomes at single-cell resolution for nodules at 12 dpi (Liu et al., 2023), 14 dpi (Sun et al., 2023), 21 dpi (Liu et al., 2023), and 28 dpi (this study). The 17 clusters of this integrated UMAP (I to XVII) were functionally annotated on the basis of the

(legend continued on next page)

nuclei in this cluster (i.e., 21 nuclei). Therefore, we estimate that the 16 constitutively expressed nodule genes identified here are highly and stably expressed in all cell types of the nodule.

To annotate the 11 clusters, we applied M.C. technology to a cross-section of a mature nodule. After analysis of the spatial activity of 28 soybean genes (Figure 2C and Supplemental Figure 6), we annotated clusters A and B as the “inner/outer cortical cell” cluster (Figure 2E, blue), cluster C as the “vascular endodermis” cluster (Figure 2F, light pink), cluster D as the “sclereid layer” cluster (Figure 2E, pink), and cluster E as the “vascular bundle” cluster (Figure 2F, orange). As with the root phloem cells, we assume that the very limited number of nuclei in cluster D and its low transcriptomic coverage result from enucleation of the cells in the sclereid layer of the nodule. Using M.C. technology with probes designed against nine *B. diazoefficiens* genes, we identified both infected cells (Figure 2G and Supplemental Figure 7) and uninfected cells in the infection zone of the nodule (Figure 2G, white arrow). Our M.C. experiments also revealed the spatial activity of two soybean genes expressed in the uninfected cells (Glyma.06G235500 [tyrosine aminotransferase] and Glyma.06G002000 [encoding an MLO protein, an inhibitor of plant defense responses; (Colebatch et al., 2004)]) (Figure 2H, green arrows, and Supplemental Figure 8) and four soybean genes expressed in the rhizobia-infected nodule cells (Glyma.17G195900 [casein kinase], Glyma.01G164600 [L-ascorbate peroxidase], Glyma.15G210100 [trehalose-6-phosphate synthase, which controls the biosynthesis and accumulation of trehalose, a carbon source for the symbiont], and Glyma.05G216000) (Figure 2H, red, and Supplemental Figure 8). Owing to its high transcriptional activity, we found that Glyma.17G195900 is an excellent marker for rhizobia-infected nodule cells (Figure 2I and 2J). Therefore, we named this gene *RIM* (i.e., *Rhizobia-Infected Marker* gene). The preferential activity of Glyma.06G235500 and Glyma.06G002000 in clusters J and K, in conjunction with the preferential expression of *RIM*, Glyma.01G164600, and Glyma.15G210100 in clusters F and G and Glyma.05G216000 in cluster H (Figure 2C), led to the annotation of clusters J and K as “uninfected cells” and clusters F, G, and H as “infected cells” of the nodule. To further support these annotations, we analyzed the expression profiles of soybean genes that control the biosynthesis of purines and ureides (Supplemental Figure 9A). We observed that rhizobia-infected cells of clusters F and G, and to a lesser extent cluster H, strongly express genes that control purine biosynthesis, whereas a ureide permease is most highly expressed in uninfected cells of the infection zone in clusters J and K (i.e., Glyma.02g116300). Our results nicely support the known compartmentation of the ureide biosynthesis pathway between infected and uninfected nodule cells (Bergersen, 1965; Ohshima and Kumazawa, 1978; 1979; Kouchi et al., 1988, 1990; Tajima

et al., 2004). The functional annotation of clusters F, G, and H was further confirmed on the basis of expression of soybean genes orthologous to *M. truncatula* transcription factor (TF) genes with regulatory roles in the nodulation process (Supplemental Figure 9B; Supplemental Table 4).

To annotate cluster I, we analyzed the biological functions of genes specifically and preferentially expressed in this cluster. Interestingly, we found that expression of NAC039 (Glyma.06G157400) and NAC018 (Glyma.04G208300) genes, marker genes of nodule senescence (Yu et al., 2023), was induced in cluster I (Supplemental Figure 10A). We also identified several genes strongly and almost exclusively expressed in cluster I that are involved in oxidative stress response and nodule senescence. These included genes encoding a metallothionein (Fonseca-García et al., 2022) (Glyma.18G180800), two CAP (Cysteine-rich secretory proteins Antigen 5 and Pathogenesis-related 1) proteins, proteins associated with soybean nodulation-related traits (Zhu et al., 2019) (Glyma.15G062300 and Glyma.13G252600), a γ -thionin antimicrobial protein (Zasloff, 2002) (Glyma.16G100400), and two TCTPs (translationally controlled tumor proteins), which hamper programmed cell death (Kiirika et al., 2014) (Glyma.09G044200 and Glyma.15G148900) (Figure 2D and Supplemental Figure 10B). Taken together, our data suggest that cluster I is composed of nuclei isolated from senescing nodule cells. This assumption is further supported by the decreased activity of leghemoglobin genes (Figure 4D). Considering the nodule UMAP, we functionally annotated the 11 cell clusters (Figure 2A), including clusters F, G, H, and I that are associated with infected nodule cells actively fixing atmospheric dinitrogen and those engaged in senescence.

Single-cell-type transcriptomic changes during nodule development

Recent studies have reported the transcriptome profiles of developing soybean nodules (i.e., 12, 14, and 21 dpi) at the single-cell level (Liu et al., 2023; Sun et al., 2023). To examine changes in the cellular complexity and transcriptomic profiles of soybean cells during nodule development, we processed (Supplemental Figure 11) and integrated into a single UMAP the transcriptomes of single nuclei isolated from each of the four developmental time points (Figure 3A). The 9248, 22 332, 14 000, and 7830 nuclei isolated from nodules at 12, 14, 21, and 28 dpi, respectively, were clustered into 17 groups in an integrated nodule UMAP (Figure 3A and 3B).

Hypothesizing that the expression of nodule cell-type marker genes should be conserved at least to some extent during nodule development, we looked at the expression patterns of 51 marker genes isolated from each of the eleven 28-dpi nodule clusters

expression of 51 cell-type marker genes identified from the 28-dpi sNucRNA-seq datasets (i.e., $p \leq 0.01$; expression in $\geq 25\%$ of the nuclei in the cluster under consideration; Supplemental Figure 12).

(B) Split UMAPs and distribution of the number of nuclei per cluster in percentages for each developmental stage of the nodule (i.e., 12, 14, 21, and 28 dpi). (C) Dotplot representation of the expression of the 51 nodule cell-type marker genes (Supplemental Figure 12) at 4 developmental stages of the soybean nodule. The percentage of nuclei in the cluster expressing the gene of interest (circle diameter) and the mean of gene expression (circle color) are both displayed.

(D) Principal-component analysis of the transcriptomes of the 17 clusters of the integrated soybean nodule UMAP and for each nodule developmental stage (12-, 14-, 21-, and 28-dpi). The transcriptomes of the cells infected by rhizobia (clusters XIII and XIV) are specifically labeled.

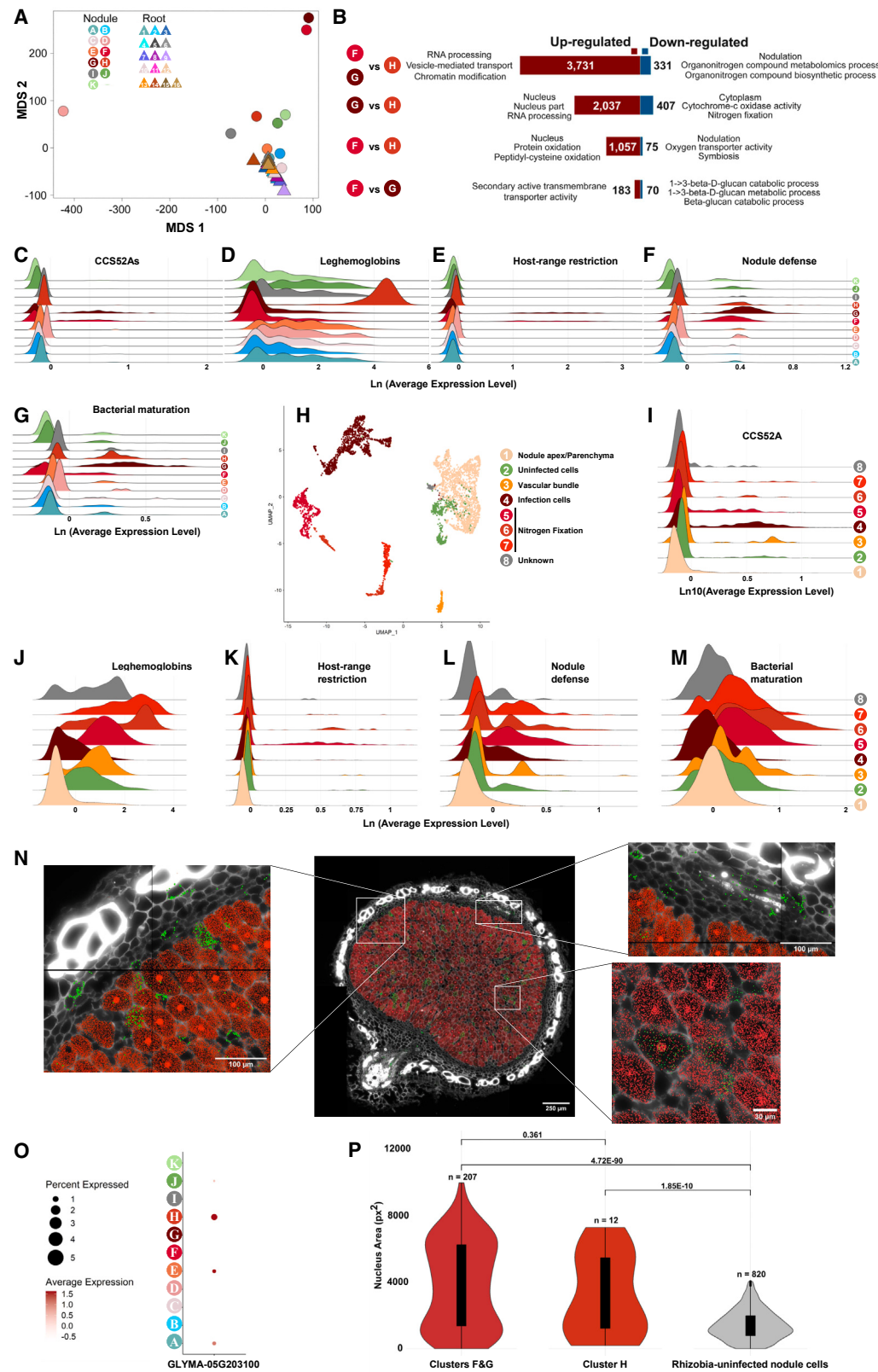


Figure 4. Identification of distinct populations of *B. diazoefficiens*-infected cells based on their transcriptional profiles.

(A) Comparative cell multi-dimensional scaling (MDS) plot of the 16 root and 11 nodule clusters.

(B) Identification of differentially expressed genes (DEGs) between the *B. diazoefficiens*-infected cell clusters of the soybean nodule (i.e., clusters F, G, and H). For each population of DEGs, we have highlighted the top enriched gene ontology categories.

(legend continued on next page)

(Figure 2A and Supplemental Figure 12) to functionally annotate the 17 clusters of the integrated nodule UMAP (Figure 3C). This approach revealed co-expression of nodule-cell-type marker genes for many clusters (i.e., for the vascular endodermis [cluster IX], the sclereid layer [cluster X], the vascular bundle [cluster XII], the infected cells [cluster XIII, which is enriched in non-nitrogen-fixing cells, and cluster XIV which is enriched in nitrogen-fixing and senescing cells] [Supplemental Figure 13], and the uninfected cells in the infection zone of the nodule [i.e., clusters XVI and XVII, clusters J and K in the 28-dpi UMAP]) (Figure 3C). On the other hand, clusters II to VIII share expression of marker genes of the inner and outer cortical nodule cells (Figure 3C), suggesting more plasticity in the transcriptomic profile of the nodule cortical cells in and between the developmental stages of the nodule.

To estimate the levels of conservation of the transcriptomes for each cell type during nodule development, we performed a principal-component analysis. Interestingly, this analysis revealed limited changes in the transcriptomes of cells infected by rhizobia (i.e., cluster XIII and XIV) during nodule development at 12-, 14-, and 21-dpi (Figure 3D). Whereas PCA1 variance was driven by transcriptomic changes during nodule development, notably from 28-dpi nodules, PCA2 variance reflected transcriptomic differences between different cell types of the nodule. Interestingly, when exclusively considering the PCA1 dimension, we noted the absence of significant transcriptomic variance between the clusters representing the 12-, 14-, and 21-dpi nodules compared with the 28-dpi nodule clusters (Figure 3D). This result suggests that only minor transcriptomic changes occurred in the nodule cell clusters from 12 to 21 dpi. However, at 28 dpi, we observed significant transcriptomic changes for each cell type of the soybean nodule, especially in the nitrogen-fixing and rhizobia-infected cells of cluster XIV, likely reflecting the physiological changes that occur at this developmental stage. Taken together, our analyses reveal similar transcriptomic profiles of 12-, 14-, and 21-dpi nodule cell types but drastic and global transcriptomic changes in 28-dpi nodules, supporting the focus of our study on 28-dpi soybean nodules to better capture the transcriptomic changes that occur in different populations of rhizobia-infected cells. This conclusion is further strengthened by the fact that 71% and 59% of the 1985 and 1359 nuclei of rhizobia-infected cell clusters XIII and XIV were identified from the 28-dpi nodule, respectively, and that over 51% of the cells expressing at least one of the five soybean leghemoglobin genes, and, *a fortiori*,

actively fixing atmospheric nitrogen, were isolated from the 28-dpi nodule. Accordingly, we performed a more comprehensive analysis of the transcriptome of infected nodule cells isolated from 28-dpi nodules.

The infection zone of the soybean nodule is composed of different populations of rhizobia-infected cells

Our single-cell transcriptomic analysis revealed differences and similarities in the transcriptional profiles of different cell types in the root and nodule organs (Figures 1A and 2A). Analysis of a multidimensional scaling (MDS) plot revealed that most of the 28-dpi nodule cell clusters differ transcriptionally from root cell clusters (Figure 4A), confirming the unique functions of the different cell types that comprise the nodule organ. The infected cell clusters F and G, as well as cells of the sclereid layer (cluster D), have the most unique transcriptomic profiles. Uninfected nodule cells (clusters J and K), another sub-population of infected cells (cluster H), and senescing nodule cells (i.e., cluster I) are also characterized by specific transcriptomic profiles. This MDS analysis also demonstrated that the infected cells of clusters F and G differ transcriptionally from the infected cells of cluster H (Figure 4A).

To further characterize these differences between infected nodule cells, we performed a comparative transcriptomic analysis of clusters F, G, and H (Figure 4B). We identified only 253 differentially expressed genes (DEGs) between clusters F and G (Figure 4B; Supplemental Table 5), confirming their transcriptional similarity. Among these DEGs, we identified several upregulated genes in cluster G that control the catabolism and metabolism of cyclic β -glucans, bacterial carbohydrates that function in suppression of host defense, regulation of osmotic potential, and interaction with host membranes (Bhagwat et al., 1999; Poole and Ledermann, 2022) (Supplemental Table 6). When comparing the activity of soybean genes between clusters F and G vs. H, we identified 3731 upregulated genes in clusters F and G (Supplemental Table 5) associated with “RNA processing,” “vesicle-mediated transport,” and “chromatin modification” functions (Supplemental Table 6), including genes orthologous to *MtCCS52A*, which controls the endoreduplication rate of infected cells during rhizobial differentiation (Cebolla et al., 1999; Vinardell et al., 2003) (Figure 4C and Supplemental Figure 14 for details; Supplemental Table 7). In cluster H, we identified 331 upregulated genes (Supplemental Table 5) associated with the “nodulation” process and “biosynthesis of

(C–G) Ridge plot distributions of the expression of soybean *CCS52A* (**C**) and leghemoglobin genes (**D**), as well as genes that control host-range restriction (**E**), nodule defense (**F**), and bacterial maturation (**G**) (x axis) as defined by Roy et al. (2020) (Supplemental Table 4). The number of cells expressing the gene(s) in each cluster is represented on the y axis.

(H) UMAP plot of 4368 *Medicago* nodule nuclei based on their transcriptomic profiles. The raw single-cell RNA-seq datasets were obtained from Ye et al. (2022) and reprocessed before generating the UMAP (see methods). These nuclei were clustered into 8 different groups.

(I–M) Ridge plot distributions of the expression of *Medicago CCS52A* (**I**) and leghemoglobin genes (**J**), as well as genes that control host-range restriction (**K**), nodule defense (**L**), and bacterial maturation (**M**) (x axis) as defined by Roy et al. (2020) (Supplemental Table 4). The number of cells expressing the gene(s) in each cluster is represented on the y axis.

(N) M.C. images of the expression of Glyma.05G203100 (green) and Glyma.17G195900 (red) genes used as markers of rhizobia-infected nodule cells. Glyma.05G203100 transcripts were specifically detected in a subset of infected nodule cells (right panel), in the cells of the vascular bundle (top-left panel), and in a sub-population of cortical cells (bottom-left panel).

(O) Dotplot representations of the expression of Glyma.05G203100. The percentage of nuclei in the cluster expressing the gene of interest (circle diameter) and the mean of gene expression (circle color) are both displayed.

(P) Distribution of the nuclear area of rhizobia-infected cells of clusters F and G and cluster H and uninfected nodule cells. A two-tailed Student's *t*-test was used to estimate the significance of differences in nuclear area among these clusters.

organonitrogen compounds” (Supplemental Table 6). This result suggests that cluster H cells are involved in an active nitrogen fixation process. This conclusion is further confirmed by the transcriptional activity of four soybean leghemoglobin genes in cluster H (Figure 4D and Supplemental Figure 14 for details; Supplemental Table 7) and the differential activity of soybean orthologs of legume genes that control various aspects of the legume–rhizobia symbiosis. For instance, clusters F and G are characterized by the preferential expression of genes that control host-range restriction and the plant defense system as defined previously (Roy et al., 2020) (Figure 4E and 4F and Supplemental Figure 14 for details; Supplemental Table S7). Interestingly, genes that control bacterial maturation were most highly expressed in cluster G and, to a lesser extent, in cluster H (Figure 4G and Supplemental Figure 14 for details; Supplemental Table S7).

We hypothesized that the rhizobia-infected cells of clusters F and G and cluster H in the soybean nodule are physiologically and biologically similar to those of the infection zone and nitrogen fixation zone in the *M. truncatula* nodule (i.e., zones II and III, respectively). To verify this hypothesis, we analyzed the expression patterns of known nodulation-related genes in *M. truncatula* (Supplemental Table S8). Accordingly, we first reanalyzed published single-cell RNA-seq datasets from the *M. truncatula* nodule (see methods for details; Figure 4H and Supplemental Figure 15A). We annotated cluster 4 and clusters 6 and 7 as the infection zone (i.e., zone II) and the nitrogen fixation zone (zone III) of the *M. truncatula* nodule on the basis of the expression of selected marker genes (Ye et al., 2022) (Figure 4H and Supplemental Figure 15A; Supplemental Table 8). Cluster 5 seems to be composed of cells in a transitional stage between cluster 4 and clusters 6 and 7, as indicated by the expression of nodulation marker genes (Supplemental Figure 15A). We examined the transcriptional activity of *M. truncatula* *CCS52A*, leghemoglobin genes, and genes characterized as regulators of host range, defense, and bacterial maturation responses (Roy et al., 2020) and found that *MtCCS52A* was most broadly and highly expressed in cluster 4 and was expressed at a lower level in cluster 5 (Figure 4I and Supplemental Figure 15B for details; Supplemental Table S7). On the other hand, leghemoglobin genes were most highly expressed in nitrogen-fixing clusters 6 and 7 (Figure 4J and Supplemental Figure 15B for details; Supplemental Table S7). We also observed the preferential activity of host-range restriction and nodule defense genes in cluster 5 and found that genes that control bacterial maturation were preferentially expressed in cluster 6 and, to a lesser extent, in clusters 5 and 7 (Figure 4K–4M and Supplemental Figure 15B for details; Supplemental Table 7). The similar single-cell-type expression profiles of nodulation-related genes in nodule clusters of soybean (Figure 4C–4G) and *M. truncatula* (Figure 4I–4M) suggests that the sub-populations of clusters F and G and cluster H in soybean have similar biological functions to those described previously for the infection zone II (clusters 4 and 5) and the nitrogen fixation zone III of the *M. truncatula* nodule (clusters 6 and 7), respectively. Our single-cell transcriptomic analysis thus redefines current knowledge of cellular diversity in the population of infected cells in the mature soybean nodule.

Our results suggest that at least two distinct populations of cells are present in the infected zone of the mature soybean nodule: the infected but non-nitrogen-fixing cells of clusters F and G and the infected nitrogen-fixing cells of cluster H. However, our single-cell transcriptomic datasets do not enable us to determine whether these populations co-exist in the same nodule. To test this possibility, we looked for molecular markers of sub-populations of rhizobia-infected cells. Upon careful review of the spatial activity of soybean genes using M.C. technology, we determined that Glyma.05G203100, which encodes a DNA-repair DEK protein, is expressed in a subset of rhizobia-infected cells (Figure 4N). According to the 28-dpi UMAP, Glyma.05G203100 is most highly expressed in the nitrogen-fixing cluster H and, to a lesser extent, in the cortical and vascular bundle cells of the nodule (i.e., clusters A and D, respectively; Figure 4O). The sNucRNA-seq transcriptional profile of Glyma.05G203100 was perfectly confirmed by M.C. (Figure 4N). Confirmation of the expression pattern of Glyma.05G203100, a gene that shows relatively low expression in soybean nodule cells, by two independent technologies further supports the biological relevance of the soybean root and nodule sNucRNA-seq transcriptomes. Our results also support the existence of different sub-populations of rhizobia-infected soybean cells that differ in the differential expression of thousands of genes (Figure 4B), including Glyma.05G203100 (Figure 4N and 4O), leading to active endoreduplication and transcriptional activity in the cells of clusters F and G and active fixation of atmospheric nitrogen in the cells of cluster H. To gain further insight into their biology and confirm that the cells of clusters F and G are endoreduplicated, as suggested by *GmCCS52a* expression (Figure 4C), we analyzed the nuclear area of the non-nitrogen-fixing cells of clusters F and G, the infected nitrogen-fixing cells of cluster H, and the uninfected nodule cells by taking advantage of the DAPI staining of the nuclei used in M.C. Our results revealed that the nuclei of cells from clusters F and G and cluster H were significantly larger than those of uninfected nodule cells, reflecting their endoreduplication (Figure 4P). We assume that the cells of clusters F and G must enter cycles of endoreduplication prior to becoming active nitrogen-fixing cluster-H cells. Taken together, our sNucRNA-seq and M.C. results support previously published spatial metabolomics observations that revealed the biochemical heterogeneity of the nodule infection zone (Veličković et al., 2018; Agtuca et al., 2020).

Gene regulatory network analysis reveals intra-cluster heterogeneity within rhizobia-infected nodule cells

To further characterize the molecular heterogeneity between cell populations of the nodule infection zone and reveal the principal regulatory pathways that govern the biology of these cells, we inferred gene regulatory networks (GRNs) for clusters F and G and cluster H. Leveraging sNucRNA-seq data from 28-dpi nodules, we used the 2000 most differentially expressed genes between clusters F and G and cluster H, known SNF genes, and predicted soybean TF genes to infer GRNs (Figure 5). We then identified the TFs with the largest number of targets (21 for the cluster H network, as hubs 20 and 21 had the same number of targets; Supplemental Table S9) as the top 20 regulators for each network and visualized them alongside known SNF-related genes (Figure 5A and 5B and Supplemental Figure 16) (Roy et al., 2020). Comparison of these two lists of regulators

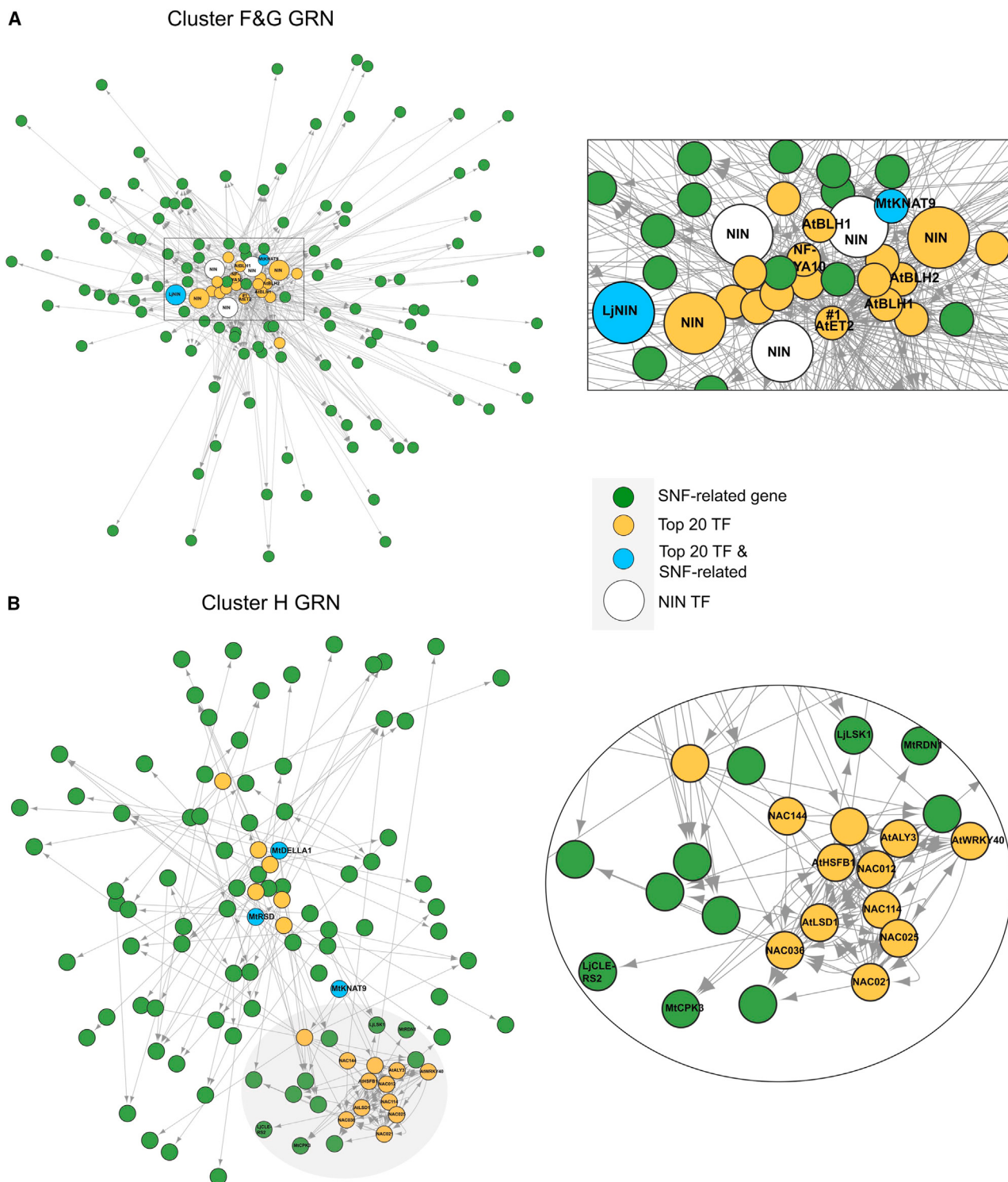


Figure 5. Gene regulatory networks of 28-dpi rhizobia-infected cells.

(A) Simplified visualization of the inferred gene regulatory network for nodule cell clusters F and G. This network shows only interactions that involve the top 20 regulators (orange nodes, or blue if they are also SNF related) of this network and/or known SNF-related genes (green nodes, or blue if they are also among the top 20 regulators). In this network, to underscore the enrichment of NIN TFs, we also visualized three NINs (large white nodes) that are neither among the top 20 hubs nor the SNF-related genes but are among the top 100 regulators in the network. Genes referenced in the main text have been labeled by gene names and the remaining nodes by their soybean gene IDs.

(legend continued on next page)

revealed only three shared TFs (Supplemental Table S9, green cells), further supporting the functional divergence between cells of clusters F and G and cells of cluster H in infected nodules (Figure 4B). Notably, a larger number of target genes were under the control of the top 20 TF regulators in clusters F and G (median of 234.5 targets) than in cluster H (median of 147 targets), further demonstrating the higher transcriptional activity occurring in the infected cells of clusters F and G (Figure 2B), likely through an increased involvement of TFs or transcriptional-level control. This inference drawn from contrasting target numbers is also supported by the most significantly enriched gene ontology term for clusters F and G compared with cluster H, namely “RNA processing” (Figure 4B).

In the cluster F and G network, the top 20 regulators form a prominent central cluster (Figure 5A, orange and blue nodes), in contrast to the cluster H network, in which the 21 regulators are split into two major and distinct groups (Figure 5B, orange and blue nodes). This observation implies functional homogeneity among cells of clusters F and G but within-cluster heterogeneity among cells of cluster H. Notably, the F and G network exhibits an overrepresentation of NIN TFs, with 6 NIN genes identified among the top 100 regulators (Figure 5A, large nodes), three of which—including an *LjNIN* ortholog (Figure 5A, large blue node)—are among the top 20 regulators (Figure 5A, large orange and blue nodes). Conversely, no NIN TFs are found among the top 100 regulators in the cluster H network (Figure 5B). Prior investigations have revealed the role of the *LjNIN* protein as a primary regulator of *LjNF-YA1* (Soyano et al., 2013; Laffont et al., 2020). Besides the three NINs among the top 20 regulators in clusters F and G, our network analysis also identified *NF-YA10* (Glyma.10G082800) in this network (Figure 5A). Consistent with its role during the nodulation process, *NF-YA10* shows the highest expression in nodules, followed by soybean root hairs (Yu et al., 2020). Taken together, the presence of *GmNF-YA10* and three *NIN* TFs among the top regulators of the F and G network suggests potential regulatory interactions among these TFs in cluster F and G cells. Although *NF-YA10* is also among the top regulators of the cluster H network (Figure 5B), lack of a central role for NIN TFs in network H likely implies a different regulatory pathway for *NF-YA10* in F and G vs. H.

Another unique feature in the F and G network is the significant disparity between the number of target genes of the first (Glyma.07G128700; 1409 targets) and second regulators (Glyma.10G081700, 437 targets). This gap suggests a predominant and pivotal regulatory role for Glyma.07G128700 in cluster F and G cells. This gene encodes an Effector of Transcription 2 (*ET2*) protein that has been reported to regulate the expression of several KNAT TFs to control cell differentiation through regulation of the cell cycle (Ivanov et al., 2008). *ET2* is also associated with DNA methylation and repair processes (Tedeschi et al., 2019), aligning well with another enriched gene ontology term, “chromatin modification,” in clusters F and G (Figure 4B). Interestingly, we identified the KNOX family

member Glyma.17G104800 among the top 20 regulators in the cluster F and G network. Glyma.17G104800 is orthologous to *MtKNAT9*, a known SNF-related TF in *Medicago* (Di Giacomo et al., 2017). Although Glyma.17G104800 also features among the top regulators in the cluster H network, it has a significantly lower number of targets (Supplemental Table 9), underscoring its more central role in clusters F and G than in cluster H. Supporting the central role of Glyma.17G104800 in clusters F and G, we identified three BEL1-like homeodomain 1/2 TF genes, which are homologous to Arabidopsis *BLH1* or *BLH2*. In Arabidopsis, *BLH1* works with *AtKNAT3* to control organ development (Kim et al., 2013), and *BLH2* affects the expression of multiple KNOX TFs (Kumar et al., 2007; Jeon and Byrne, 2020). Together, our data suggest that the identified NIN/NF-YA and BLH/KNAT regulons play a central role in the control of rhizobial infection processes in the F and G cell clusters of the soybean nodule. We hypothesize that, among these TF genes, Glyma.07G128700 plays a central role in regulating the NIN/NF-YA/KNAT/BLH network, either directly or indirectly, and potentially through chromatin modification and DNA methylation.

The topography of the cluster H network differs from that of the F and G network, in that the top regulators are split into two groups (Figure 5B). Upon examining the functional categorization of targets in the major “NIN” regulatory group of the F and G network (Figure 5A) and the two distinct regulatory groups of the H network (Figure 5B), we observed that the “autoregulation of nodule number” category (i.e., orthologs of *LjCLE-RS2*, *LjLSK1*, *MtCPK3*, and *MtRDN1*, according to Roy et al., 2020) was overrepresented in one group of the H network (highlighted in gray in Figure 5B). In this same highlighted group, 12 genes were associated with plant stress response and the cell death program. These included *GmNAC012/NAC021/NAC036*, which were annotated as Arabidopsis *NTL9* (NAC Transcription factor-Like 9), a calmodulin-regulated transcriptional repressor linked with regulation of leaf senescence and defense response (Yoon et al., 2008; Block et al., 2014); Glyma.15G192000, a homolog of *AtLSD*, which controls the plant cell death pathway and immune response (Mühlenbock et al., 2008); and *GmNAC114*, a homolog of *AtSND2*, which regulates cell wall organization (Hussey et al., 2011) and is intricately linked with cell growth and death in various organs (Dauphin et al., 2022). Besides the four NAC TFs mentioned above (i.e., *GmNAC012/021/036/114*), *GmNAC144* and *GmNAC025* are homologous to *AtSOG1*, a TF that controls endocycling and the crosstalk between immune response and DNA damage (Adachi et al., 2011; Yoshiyama et al., 2020). This overrepresentation of NAC genes in one group of the H network (highlighted in gray in Figure 5B) is supported by previous studies showing the roles of several legume NAC genes in nodule senescence. For instance, *MtNAC969*, which is phylogenetically close to Arabidopsis *ANAC092* (Wang et al., 2023b), implicated in leaf senescence (Guo et al., 2021), is a regulator of nodule senescence in *Medicago* (de Zelicourt et al., 2012). Recently, *LjNAC094* was also reported to regulate nodule senescence in

(B) Simplified visualization of the inferred gene regulatory network for nodule cell cluster H. This network shows only interactions that involve the top 21 regulators (orange nodes, or blue if they are also SNF related) of this network and/or known SNF-related genes (green nodes, or blue if they are also among the top 21 regulators). In this network, a cluster of 12 regulators at the bottom of the network, including the top eight hubs of the list (see Supplemental Table 9), is highlighted in gray. Genes referenced in the main text and blue nodes are labeled by gene names and the remaining nodes by their soybean gene IDs.

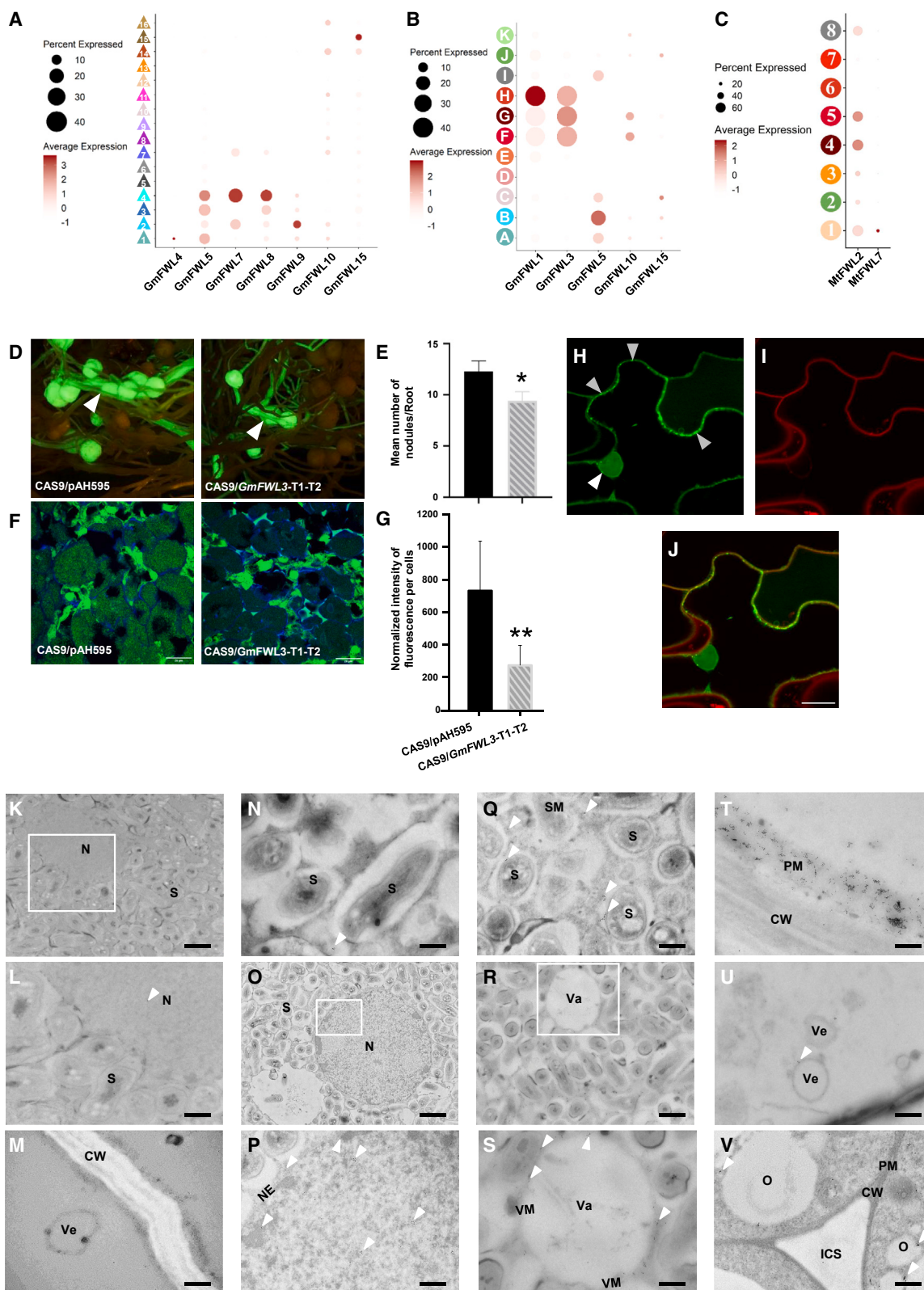


Figure 6. Functional characterization of *GmFWL3*, a new microdomain-associated protein-coding gene.

(A–C) Dotplot representations of the expression of *GmFWL* genes expressed in at least 1 of the 16 root (A) and 11 nodule clusters (B), and the expression of *MtFWL2* and 7, orthologs of *GmFWL3* and 1, respectively, in the *M. truncatula* nodule clusters (Figure 4H) (C). The percentage of nuclei in the cluster expressing the gene of interest (circle diameter) and the mean of gene expression (circle color) are both displayed.

(legend continued on next page)

L. japonicus (Wang et al., 2023b). Therefore, these six NAC TFs are potential candidates for the regulation of nodule senescence in soybean. Finally, a few other TFs in this highlighted group have Arabidopsis orthologs that are associated with defense/immunity response (AtHSFB1 and AtWRKY40) or the cell cycle (AtALY3). These findings, together with the close transcriptional relationship between clusters H and I (senescing cells; Figure 2A) support the hypothesis that cluster H cells, at least to some extent, are involved in autoregulation of nodule number via defense–response signaling, as well as regulatory pathways associated with the cell wall, the cell cycle/endocycling, and cell death.

Overall, this GRN analysis underscores the diversity within and between clusters of infected nodule cells, shedding light on novel candidates potentially implicated in various signaling/regulatory processes in soybean nodules, particularly in cells of the infection zone.

GmFWL3, like GmFWL1, controls infection of soybean nodule cells by *B. diazoefficiens*

GmFWL1, a member of the soybean plasma membrane microdomain-associated FWL family, has been identified as a microdomain-associated protein that controls chromatin accessibility and the infection of nodule cells by *B. diazoefficiens* (Libault et al., 2010b; Qiao et al., 2017a). Our single-cell RNA-seq approach enabled us to revisit the expression pattern of *GmFWL1* in mature soybean nodules, revealing its preferential expression in cells of rhizobia-infected clusters F, G, and H (Figure 6A and 6B). Given the over-representation of GO terms related to “RNA processing” and “chromatin condensation” in the pool of cluster F and G DEGs (Figure 4C), our results suggest that the microdomain fraction of the plant plasma membrane controls the infection of nodule cells by rhizobia, potentially through changes in chromatin condensation (Libault et al., 2010b).

GmFWL3 is the only other member of the FWL family that shares a pattern of expression with *GmFWL1* (i.e., *GmFWL3* is not ex-

pressed in the root [Figure 6A] but is preferentially expressed in infected nodule cells) (Figure 6B; clusters F, G, and H). Although *MtFWL7*, the *M. truncatula* ortholog of *GmFWL1*, was not expressed in infected nodule cells, *MtFWL2*, the *M. truncatula* ortholog of *GmFWL3*, was most highly expressed in *Medicago* nodule clusters 4 and 5, which were annotated as *S. meliloti*-infected and nitrogen-fixing cells (Figure 6C). On the basis of these findings, we hypothesize that *GmFWL3* encodes another microdomain-associated protein that regulates rhizobial infection of soybean cells. This hypothesis is supported by previous studies in which GmFWL1 and GmFWL3 proteins were localized to the symbiosome membrane (Clarke et al., 2015; Qiao et al., 2017a). To verify this hypothesis, we deleted the conserved PLAC8 domain of *GmFWL3* by expressing two guide RNAs (i.e., GmFWL3T1 and T2) using CRISPR-Cas-mediated genome-editing technology (Supplemental Figure 17). Nodule number was significantly lower on GFP-positive CAS9/GmFWL3T1-T2 transgenic roots than on GFP-positive Cas9/pAH595 control transgenic roots (Figure 6D and 6E; Student's *t*-test, $p = 0.009$). At the cellular level, staining of bacteroids with green-fluorescent SYTO13 dye revealed significantly lower intensity of the fluorescent signal in CAS9/GmFWL3T1-T2 transgenic roots than in control roots (Figure 6F and 6G; Student's *t*-test, $p = 0.009$). We concluded that microbial infection of nodule cells was impaired upon mutagenesis of *GmFWL3*.

To support the role of GmFWL3 as a microdomain-associated protein, we analyzed its cellular and subcellular localization by expressing N- and C-terminal translational fusions between GFP and GmFWL3 in tobacco leaf cells. Both N- and C-terminal GFP-GmFWL3 chimeric proteins were localized in puncta at the plasma membrane (Figure 6H–6J, gray arrows, and Supplemental Figure 18A–18I). Plasmolysis assays (Supplemental Figure 18J–18R) and protoplast isolation experiments (Supplemental Figure 18S–18U) provided further evidence of the punctate localization of GmFWL3-GFP at the plasma membrane. Given its punctate localization similar to that of GmFWL1 (Libault et al., 2010b), GmFWL3 is likely a plasma membrane microdomain-associated protein. However, in contrast to GmFWL1, which was exclusively localized at the

(D) Stereoscope images of representative transgenic soybean roots upon mutagenesis of *GmFWL3* using CRISPR-Cas9 technology (CAS9/GmFWL3-T1-T2 transgene) and of control roots (CAS9/pAH595 transgene). The GFP signal was used as a reporter to identify the transgenic roots (white arrow). (E) Average number of nodules on 150 pUB-CAS9-pAH595 and 201 pUB-CAS9/GmFWL3-T1-T2 GFP-positive transgenic roots (Student's *t*-test: $p = 0.009$). (F) Representative SYTO13 staining of transgenic soybean roots transformed with CAS9/pAH595 control (pUB-CAS9-pAH595) and CAS9/GmFWL3-T1-T2 transgenes. Staining reveals the density of rhizobial bacteria in infected cells of the soybean nodules. (G) Quantification of the infection rate of CAS9/pAH595 control and CAS9/GmFWL3-T1-T2 transgenic nodule cells upon SYTO13 staining of bacterial and nuclear DNA (ANOVA single factor test $\alpha = 0.05$, $*p < 1e-100$, $n \approx 350$). (H–J) Subcellular localization of GmFWL3 in tobacco leaves using confocal microscopy and in soybean nodules using transmission electron microscopy after immunogold labeling. Cross-sections of tobacco leaf cells transiently expressing p35S::GmFWL3-GFP (H, gray and white arrows highlight the punctate localization of GmFWL3 on the plasma and nuclear membranes) and counterstained with the membrane dye SynaptoRed (FM-64, I) reveal the punctate plasma membrane localization of GFP-GmFWL3 fusion proteins. Co-localization of the GFP-GmFWL3 signal with the membrane dye FM64 confirms its membrane localization (J). Scale bar, 20 μ m. (K–V) Subcellular localization of GmFWL3 in soybean nodules observed by transmission electron microscopy after immunogold labeling. Representative images of gold particle distribution in the plasma membrane (PM), symbiosome (S), vacuole (Va), vesicles (Ve), membrane-bound organelles (O), nuclear envelope (NE), and nucleus (N) after immunogold labeling against a c-myc epitope tag alone (control, K–N) and myc-tagged GmFWL3 chimeric protein (O–V). Compared with c-myc alone, which had no or a few randomly distributed gold particles (K–N), the myc-GmFWL3 protein was strongly detected in the nucleus (O and P), nuclear envelope (P), symbiosome membranes (Q), vacuolar membrane (R and S), plasma membrane (T), and vesicular and other membrane-bound organelles (U and V). White arrowheads point to the gold particles. CW, cell wall; PM, plasma membrane; S, symbiosome; SM, symbiosome membrane; Va, vacuole; VM, vacuolar membrane; Ve, vesicle; N, nucleus; NE, nuclear envelope; ICS, intercellular space. Scale bars correspond to 400 nm (A–K), 1 μ m (L, zoom in from K, see white box; and S, zoom in from R, see white box), and 2 μ m (P, zoom in from O, see white box).

plasma membrane, GmFWL3-GFP was also associated with the nucleus and the nuclear membrane (Figure 6H, white arrows). This nuclear membrane localization was confirmed by colocalization of GmFWL3-GFP with the nuclear envelope marker CFP-AtSUN1 (Graumann et al., 2010) in tobacco epidermal leaf cells (Supplemental Figure 19A–19H) and by epifluorescent confocal microscopy in soybean root epidermal cells and root hairs (Supplemental Figure 19I–19Q). To further assess the subcellular localization of GmFWL3 in infected cells of the soybean nodule, cell types in which *GmFWL3* is most highly expressed, we used high-resolution transmission electron microscopy (TEM) combined with immunogold labeling to examine transgenic soybean nodules expressing N-terminal and C-terminal myc-tagged GmFWL3 proteins under the control of the pFMV promoter. We observed localization of GmFWL3 at the plasma, symbiosome, vacuolar, and vesicular membranes and the nuclei of infected nodule cells (white arrows, Figure 6K–6V).

GmFWL3 and GmFWL1 belong to the same large membrane microdomain protein complexes

We hypothesized that GmFWL3 is a microdomain-associated protein. To verify this hypothesis, we performed co-immunoprecipitation assays on 30-dpi GFP-positive transgenic soybean nodules expressing N- and C-terminal myc-tagged GmFWL3 proteins to identify GmFWL3 interaction partners (Supplemental Figure 20A). Across three independent biological replicates, a total of 321 proteins co-immunoprecipitated with N-terminal or C-terminal myc-tagged GmFWL3 proteins in at least two replicates but not with the myc-tag alone (Supplemental Table 10). Among these GmFWL3 interactors were five FWL/PLAC8 proteins including GmFWL1, six vacuolar ATPases, nine SPFH-domain microdomain-associated proteins (i.e., prohibitin/flotollin/remorin), 12 aquaporins, and 21 proteins with functions related to vesicle trafficking (Figure 7A; Supplemental Table 10). We examined the transcriptional activity of the 321 genes encoding these proteins (Figures 7B) and found that 10 were preferentially expressed in clusters F, G, and H (fold-change ≥ 4 between the expression level in the most highly expressed F, G, and H cluster vs. the most highly expressed cluster in the remaining nodule clusters and the root clusters; Figure 7C). These genes encode three FWL/PLAC8 proteins (including GmFWL1 and GmFWL3), three prohibitin/flotollin/remorin proteins, which are well-characterized microdomain-associated proteins, one CASP-like protein, one sulfate transporter, one vesicle-associated membrane protein, and one receptor-like kinase 1. Among these 10 genes, 9 were co-expressed in clusters F and G, and, to a lesser extent, in cluster H (Figure 7C, red characters, and Supplemental Figure 20B). The co-expression of these genes in infected nodule cells and the interactions of their proteins with GmFWL3 suggest the formation of a quaternary protein structure localized in the microdomain fraction of biological membranes to control the symbiosis between soybean nodule cells and bacteroids. GmFWL1 and GmFWL3 proteins play a central and likely redundant role in formation of this protein complex, as revealed by integration of the lists of proteins co-immunoprecipitated with GmFWL3 and GmFWL1 (Qiao et al., 2017a). When comparing the proteins co-immunoprecipitated by the two FWL proteins, we found that 63 proteins that interacted with GmFWL3 also interacted with GmFWL1 (Graumann et al., 2010) (Supplemental Table 10).

These included seven SPFH-domain microdomain-associated proteins, four vacuolar ATPases, 15 proton-ATPase/GTPases, four aquaporins, four receptor kinases, four integral membrane proteins, three proteins with transport activity, and three proteins associated with vesicle trafficking. Interestingly, among these shared binding partners was GmFLOT2/4 (Glyma.06G065600), the soybean ortholog of *M. truncatula* MtFLOT2 and MtFLOT4. This result shows that GmFWL1 and GmFWL3 proteins cooperate to support a network of microdomain-associated proteins to control infection of soybean nodule cells by *B. diazoefficiens*.

DISCUSSION

The nodule is a root organ specialized for the fixation and assimilation of atmospheric nitrogen. This biological process is the product of the symbiotic interaction between nitrogen-fixing rhizobia and the plant. Microscopy observations have revealed differences in cellular organization between determinate and indeterminate nodules. Indeterminate nodules (e.g., *M. truncatula*) are organized into different zones that reflect differences in the developmental stages of the plant cells and their relationship with the symbiont. These zones include the active meristematic zone at the tip of the indeterminate nodule (zone I), the zone of infection by the bacteria (zone II), the nitrogen fixation zone (zone III), and the plant cell senescence zone (zone IV). Recent molecular studies have revealed differences in transcriptomic and epigenomic profiles among the different zones of the indeterminate *M. truncatula* nodule (Mergaert et al., 2020; Pecrix et al., 2022). Such zonation does not exist in determinate nodules (e.g., soybean). Except for the central location of the senescing zone that emerges in 4-week-old nodules (Yu et al., 2023) and the classification into infected and uninfected cells, there have been no reports of biologically different populations of infected cells, leading to the assumption that *B. diazoefficiens*-infected soybean cells are biologically homogenous (Dupont et al., 2012). This assumption has recently been challenged through the spatial profiling of over 100 metabolites using matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry (MS) imaging (Stopka et al., 2017). In our study, single-cell-resolution transcriptomic analyses (Figure 4) and M.C. experiments (Figure 2) revealed cellular heterogeneity among rhizobia-infected cells in the infection zone of the soybean nodule. In addition to distinguishing uninfected (clusters J and K) and senescing cells (cluster I) on the basis of their transcriptomic profiles, we also observed different transcriptomic signatures between co-existing sub-populations of cells comprising the rhizobia-infected clusters F and G and cluster H (i.e., endoreduplication and strong transcriptomic activity in clusters F and G; strong induction of leghemoglobin gene expression in cluster H). This result suggests that only cells of cluster H are actively fixing atmospheric dinitrogen. Previous studies on 12-, 14-, and 21-dpi developing nodules did not reveal this cellular heterogeneity among *B. diazoefficiens*-infected cells. Specifically, our comparative analysis of gene expression in clusters at different nodule developmental stages revealed that the transcriptomic profiles of 12-, 14-, and 21-dpi nodule cell types were similar (Figure 3D). However, drastic and global transcriptomic changes occurred in 28-dpi nodules, especially in *B. diazoefficiens*-infected cells (Figure 3D). Therefore, we conclude that unique transcriptomic programs are activated in

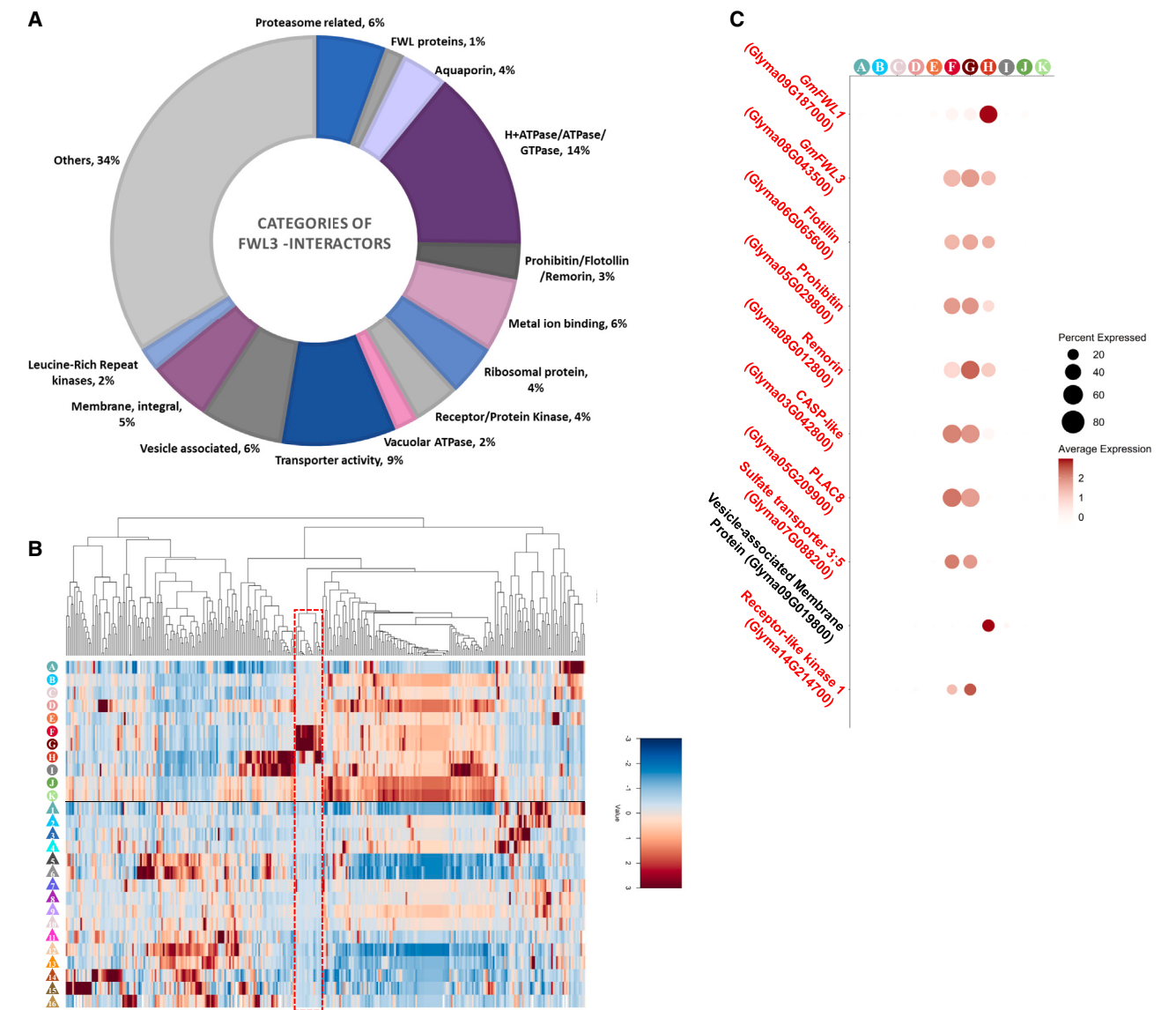


Figure 7. Co-expression of genes encoding proteins co-immunoprecipitated with GmFWL3 at the single-cell level.

(A) Donut chart of the distribution of putative GmFWL3 interaction partners according to their biological functions.

(B) Heatmap representation of the expression of genes encoding the 321 proteins proposed to interact with GmFWL3. Gene expression is displayed for each of the 16 soybean root clusters (1–16; Figure 1A) and 11 28-dpi soybean nodules (A–K; Figure 2A). The set of genes highlighted in the red dashed square are preferentially expressed in clusters F, G, and H.

(C) Dotplot representation of the expression of 10 soybean genes that interact with GmFWL3, including GmFWL3 itself. Nine are preferentially expressed in clusters F, G, and H (Supplemental Figure 20). The percentage of nuclei expressing the gene of interest (circle size) and the mean expression (circle color) of the genes are shown.

different sub-populations of *B. diazoefficiens*-infected cells later in nodule development.

By identifying different sub-populations of rhizobia-infected soybean cells, our study enables us to reconsider the definition of a plant cell type based not only on morphological or physiological differences but also on molecular attributes (i.e., gene expression) and the nature of environmental interactions (in this case, with rhizobia). From a functional perspective, our data suggest that a subset of rhizobia-infected cells are actively engaged in fixing atmospheric nitrogen (cluster H cells), whereas others are engaged in events of endoreduplication (cluster F and

G cells). Our M.C. data suggest that these two populations of cells co-exist in the same nodule. For instance, the differential expression of Glyma.05G203100 between cluster H and clusters F and G (Figure 4O) was nicely confirmed by M.C. with the detection of Glyma.05G203100 transcripts in a small population of the infected nodule cells (Figure 4N), suggesting the co-existence of two populations of rhizobia-infected cells in the same nodule. These two populations share similar endoreduplication rates (Figure 4P), supporting the fact that 4C endoreduplication of the soybean nodule cells is a prerequisite for rhizobial infection, as suggested previously (Fan et al., 2022). However, the physiology of these two populations

differs. We found that cells of the F and G clusters are characterized by active endoreduplication (Figure 4B and 4C) and higher transcriptional activity (Figure 2B) but do not actively express leghemoglobin genes, suggesting that they are not actively fixing atmospheric nitrogen (Figure 4D). The putative predominant role of the ET2 TF (Glyma.07G128700) in cells of clusters F and G, as reflected by the large number of predicted targets in our network (Figure 5A), supports the presence of intense and broad transcriptional activity in cells of the F and G clusters. By contrast, and despite their endoreduplicated nature, cells of the H cluster are characterized by lower transcriptional activity and high nitrogen-fixing activity (Figure 4B and 4D). We hypothesize that the major function of F and G cells is to create the conditions for successful nitrogen fixation by generating a large pool of transcripts. Then, when entering active nitrogen fixation (cluster H), the infected cells will capitalize on this large pool of transcripts to maximize protein translation and metabolism, a requirement for active nitrogen fixation. The level of stress induced by the symbiosis (e.g., superoxide radicals and reactive oxygen species [Dalton et al., 1991; Puppo et al., 1991; Davies and Puppo, 1992; Moreau et al., 1996; Puppo et al., 2005]), as reflected in the identification of multiple NAC TFs in the gene network of cluster H (Figure 5B), ultimately leads to senescence of the infected cells (cluster I). We assume that these senescing cells belong to the central senescing zone of the soybean nodule that emerges as early as 4 weeks after bacterial inoculation (Yu et al., 2023). Thus, based on our knowledge, our study establishes for the first time the transcriptomes of different co-existing populations of rhizobia-infected soybean cells, including those engaged in senescence. Such cellular complexity of a determinate nodule is similar to that reported for the indeterminate nodule, but the latter is divided into various zones of infected cells (i.e., the rhizobia-infected zone II, the nitrogen fixation zone III, and the senescing zone IV).

Although soybean and *Medicago* nodules differ in morphology, we hypothesize that the soybean nodule is, from a cellular composition point of view, similar to the *M. truncatula* nodule. It is composed of infected but non-nitrogen-fixing cells (clusters F and G of the soybean nodule; zone II of the *Medicago* nodule), nitrogen-fixing cells (cluster H of the soybean nodule; zone III of the *Medicago* nodule), and senescing cells (cluster I of the soybean nodule; zone IV of the *Medicago* nodule). Our data (Figure 3) and those of Yu et al. (2023) support the coexistence of these cells. We hypothesize that the purpose of the co-existence of these different populations is to ensure a steady supply of nitrogen to the plant by balancing the energy cost associated with nitrogen fixation with the impact of nitrogen fixation on plant cell viability (i.e., nitrogen fixation triggers the formation of various oxidizing species).

The single-nucleus transcriptome atlases of the root and nodule also help us to refine the identification of genes that control biological processes. For instance, in addition to their critical roles in the infection of plant cells by different types of symbiotic and pathogenic microbes, including bacteria, fungi, and viruses, microdomain-associated proteins are also central to control of the nodulation process (e.g., flotillins, remorins, and FWL

protein-coding genes have previously been reported to control legume nodulation) (Haney and Long, 2010; Thibivilliers et al., 2020a; Yu, 2020). Here, we identified a new microdomain-associated protein-coding gene, *GmFWL3*, that is preferentially expressed in infected nodule cells (i.e., clusters F, G, and H). Although the expression profiles of *GmFWL1* and *GmFWL3* overlap (Figure 6B), the subcellular localization of the *GmFWL3* protein in infected nodule cells is broader than that reported previously for *GmFWL1* (Libault et al., 2010b). Specifically, in addition to its localization at the plasma membrane of infected nodule cells, *GmFWL3* was also found in the symbiosome membrane, as reported previously (Clarke et al., 2015). This observation, together with the significant decrease in bacterial infection of nodule cells upon *GmFWL3* knockout, suggests that the microdomain fraction of the symbiosome membrane plays a critical role in the communication between plant cells and bacteroids. In addition to this first role, we assume that *GmFWL3* might also have another symbiotic function that requires its high expression in infected cells of the soybean nodule. Specifically, considering previous reports that microdomain-associated proteins regulate the transport of different biochemical compounds such as sugars and auxin (Teale et al., 2006; Krecek et al., 2009), we hypothesize that *GmFWL3* contributes to nutrient transport between bacteroids and plant cells throughout the lifetime of the symbiosis. The identification of other microdomain-associated proteins that interact with *GmFWL3*, including *GmFWL1*, and their co-expression in the same cell type of the nodule, support their interaction and the formation of a protein network in the microdomain fraction of biological membranes of the rhizobia-infected cells. Our findings highlight the ability of single-cell genomics to dissect complex biological processes, such as refining the biological concept of “cell type” by including high-resolution molecular attributes in this definition.

METHODS

Bacterial culture

Escherichia coli, *Agrobacterium rhizogenes* (K599), and *Agrobacterium tumefaciens* (GV3101) strains were grown in LB medium supplemented with appropriate antibiotics at 37°C for *E. coli* and 30°C for *Agrobacterium* strains. *B. diazoefficiens* USDA110 was grown in HM medium supplemented with 50 µg/mL chloramphenicol. *B. diazoefficiens* cultures were grown for 3 days and pelleted at 4000 *g* for 10 min, then washed and diluted to an OD_{600nm} of 0.1 in nutritive NPNS solution for inoculation (Broughton and Dilworth, 1971).

Plant material

Soybean (*Glycine max* Williams 82) seeds were sterilized as described previously (Pingault et al., 2018). The seeds were then placed on agar B&D medium in the absence of nitrogen and germinated in a growth chamber. The roots of 6-day-old plants (i.e., the meristematic, elongation, and maturation zones of the root, which include emerging and fully elongated root hair cells) were collected and processed to generate the sNucRNA-seq libraries. To isolate mature nodules, 3-day-old seedlings were germinated on B&D agar medium without nitrogen and inoculated with a suspension of *B. diazoefficiens* USDA 110 (OD_{600nm} = 0.1). After 72 h of incubation in the dark, the seedlings were transferred to a vermiculite:perlite mix (3:1) and grown in the growth chamber (16 h light/8 h dark) at 20°C–26°C. Mature nodules were collected 28 days after bacterial inoculation.

Plant Communications

Nucleus isolation, sNucRNA-seq library preparation, and sequencing

For nucleus isolation, roots and nodules were chopped and passed through 30- and 40- μ m cell strainers as described previously (Thibivilliers et al., 2020b). The filtered nuclei were purified by cell sorting using a FACS Aria II 603 cell sorter (BD Biosciences). The three sNucRNA-seq root libraries and two nodule libraries were constructed following the protocol of the Chromium Single Cell 3' Library & Gel Bead Kit v3.1 (10X Genomics). Sequencing of single-indexed, paired-end libraries was performed on the Illumina NovaSeq 6000 platform according to the 10X Genomics recommendations (see Supplemental Table 11 for detailed information).

Cloning and molecular constructs

Conventional cloning procedures

All constructs were generated using either classical restriction-enzyme ligation or Gateway cloning strategies (www.lifetechnologies.com). For Gateway cloning, *GmFWL3* cDNA and *FWL3-GFP* were cloned into the pDONR-Zeo vector by the BP clonase reaction. Constructs were sequenced to confirm the integrity of the cloned genes. The *GmFWL3*-pDONR plasmid was used with the pMDC43 and pMDC83 destination vectors in an LR reaction to generate p35S::GFP-FWL3 and p35S::FWL3-GFP translational fusion constructs. The p35S::CFP-AtSUN1 construct was described previously (Graumann et al., 2010).

To generate myc-tagged chimeric proteins, AttR1-CmR-ccdb-AttR2-10xMyc and 10xMyc-AttR1-CmR-ccdb-AttR2 cassettes were amplified from pGWB20 and pGWB21 (Nakagawa et al., 2007), respectively, and cloned in place of the HA tag in the CGT3304 plasmid using BamHI and Eco53kI restriction sites. After DNA sequence confirmation, the resulting promoter FMV::AttR1-CmR-ccdb-AttR2-10xmyc-tnos and pFMV:10xmyc-AttR1-CmR-ccdb-AttR2-tnos cassettes were excised by SbfI restriction enzyme digestion and ligated into AKK1467B at the SbfI restriction site, thereby creating AKK1467B-10myc-GW and AKK1467B-GW-10myc Gateway-compatible destination plasmids (see Supplemental Figure 21). The *GmFWL3*-pDONR-Zeo plasmid was used in an LR reaction to clone *GmFWL3* cDNA into the modified AKK1467B-10myc-GW and AKK1467B-GW-10myc Gateway destination plasmids, generating pFMV:10myc-*GmFWL3* and pFMV::*GmFWL3*-10myc constructs. The pFMV:10myc-FWL3, pFMV::FWL3-10myc, p35S::FWL3-GFP, and p35S::GFP-FWL3 constructs and their respective controls, pFMV:10Myc and p35S::GFP, were transformed into *A. rhizogenes* (strain K599) for hairy root transformation.

CRISPR-Cas9 design and screening for mutation by band shift and sequencing

The guide RNAs used to knock out selected genes via CRISPR-Cas9 technology were designed using the guide RNA designer website (Doench et al., 2014). Two *GmFWL3* target sequences (referred to hereafter as *GmFWL3*-T1 and T2) were independently cloned into the Esp3I and BsaI sites of the pAH595 guide RNA entry vector under the control of the AtU6 and At7SL promoters, respectively, using the Golden Gate method (Curtin et al., 2018) to create the pAH595-*GmFWL3*-T1-T2 donor plasmid. The pAH595-*GmFWL3*-T1-T2 entry vector was then used with the pNJB184-CAS9 entry vector in a two-fragment multi-site Gateway LR clonase reaction. The cassettes were recombined into the pUB-GW-GFP binary vector (Maekawa et al., 2008), which carries a GFP selectable marker for screening transgenic FWL3-CRISPR-Cas9 hairy roots (Xie and Yang, 2013). The empty pAH595 donor plasmid combined with pNJB184-CAS9 in an LR reaction was used as the control. The resulting pLjUB::Cas9-pU6/At7SL::*GmFWL3*-T1-T2 binary plasmid (hereafter CAS9/*GmFWL3*-T1-T2) was transformed into *A. rhizogenes* (K599) for hairy root transformation and nodule phenotyping.

To characterize the nature of the mutations induced by CRISPR-Cas9 in the *FWL3* gene, the genomic DNA of transgenic soybean roots expressing GFP was extracted. The regions spanning the target sites were amplified

Single-cell RNA-seq atlas of soybean root organs

by PCR using the *GFWL3*-5'utr forward (5'-CCAAGTCCAATAAC-TATGCTTGAG-3') and reverse primers (5'-TCAACGGCTCATGCC-3') and then sequenced for analysis.

Plant transformation and confocal laser scanning microscopy

Tobacco leaf infiltration and protoplast isolation

Nicotiana benthamiana leaves were co-infiltrated with *A. tumefaciens* (GV3101) expressing the virus RNA-silencing suppressor protein HC-Pro to enhance expression of the transgene, together with the following constructs: p35S::GFP-*GmFWL3*, p35S::*GmFWL3*-GFP, and p35S::GFP. Three days after infiltration, the tobacco leaf cells were imaged using a Nikon A1 confocal microscope. To produce transgenic tobacco leaf protoplasts, infiltrated epidermal leaf cells were incubated in MKM medium (9% mannitol, 0.037% KCl, 0.2 M MOPS [pH 6.0] supplemented with 0.05% driselase, 0.02% macerozyme R10, and 0.1% onozuka R10 [all Sigma, <http://www.sigmaaldrich.com/>]) for 3 h in the dark at room temperature. Tobacco leaf epidermal protoplasts expressing the GFP-FWL3 and FWL3-GFP translational fusions, as well as the GFP control, were imaged using a Nikon A1 confocal microscope. Plasma membrane staining was performed by infiltrating 5 μ M of FM4-64 (SynaptoRed C2, Biotium no. 70020) prior to microscopy.

Soybean hairy root transformation

Eleven-day-old soybean Williams 82 plants were used for hairy root transformation (Pingault et al., 2018). Three days after transformation, shoot explants on rockwool cubes were watered with nutritive NPNS solution and allowed to grow for an additional 10 days. The plants were then transferred to a 3:1 autoclaved mixture of vermiculite and perlite and grown for an additional 14 days to allow plants to develop roots. Plants were then inoculated with USDA110. Thirty-day-old transgenic nodules, characterized by the expression of the GFP reporter, were collected under a Nikon SMZ25 epifluorescence stereoscope.

Co-immunoprecipitation assay

Co-immunoprecipitation experiments were performed as described previously (Qiao et al., 2017a) using transgenic nodules isolated from GFP-positive transgenic roots (see above). Total protein extracts for co-immunoprecipitation assays were obtained by grinding transgenic nodules in protein extraction buffer (50 mM Tris-MES [pH 7.5], 300 mM sucrose, 150 mM NaCl, 10 mM potassium acetate, 5 mM EDTA, Sigma plant protease inhibitor cocktail, and 1% Triton X-100). After a 30-min incubation in protein extraction buffer, the proteins were filtered through a 40- μ m filter (Fisherbrand no. 22-363-547) and centrifuged (15 000 g, 10 min, 4°C). Prior to the co-immunoprecipitation assay, western blot assays were performed to detect the tagged proteins using anti-Myc-HRP antibodies (Fisher no. R951-25) diluted 1:1000 in TBS-0.05% Tween20-2% skimmed milk. Co-immunoprecipitated proteins were isolated by applying total protein extracts to 50 μ L of anti-myc Tag MicroBeads (Miltényi Biotech, no. 130-091-284) on ice for 30 min and separated through a μ Column (Miltényi Biotech no. 130-042-701) in the magnetic field of the μ MACS Separator system (Miltényi Biotech, nos. 130-042-602 and 130-042-303) according to the manufacturer's protocol. Samples were eluted with SDS-PAGE sample buffer for further analysis as described below. Three replicates each for both N- and C-terminal c-myc fusions to *GmFWL3* were compared with three replicates of c-myc alone as controls.

MS and identification of *GmFWL3*-binding protein partners

Co-immunoprecipitated proteins were denatured at 95°C for 5 min and loaded onto a 10% Bolt Bis-Tris Plus gel. After brief electrophoresis to concentrate the proteins at the top of the gel, the proteins were fixed and stained with colloidal Coomassie blue (Sigma). The areas of gel that contained the co-immunoprecipitated proteins were excised and subjected to reduction and alkylation with DTT and iodoacetamide, respectively, then washed with ammonium bicarbonate/acetonitrile to remove stain and SDS.

Trypsin digestion was carried out overnight at 37°C. Peptides were extracted from the gel pieces, dried down, and resuspended in 0.1% trifluoroacetic acid. Samples were desalted using an Oasis HLB μ Elution solid-phase extraction plate (Waters, Milford, MA). Eluates were dried down and resuspended in 2.5% acetonitrile and 0.1% formic acid. Peptides were then run by nanoLCMS/MS using a 2-h gradient on a 0.075 $\times$ 250 mm CSH C18 column (Waters) feeding into a Q-Exactive HF mass spectrometer.

All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.6.1). Mascot was set up to search the *Glycine max* protein database (Wm82.a4.v1; 92 226 records) for tryptic peptides. Mascot was searched with a fragment ion mass tolerance of 0.060 Da and a parent ion tolerance of 10.0 PPM. Deamidated asparagine and glutamine, oxidized methionine, and carbamidomethylated cysteine were specified as variable modifications in Mascot. Scaffold (version 4.8.9, Proteome Software, Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted with a probability of 80% or greater and <1% FDR by the Peptide Prophet algorithm (Keller et al., 2002) with Scaffold delta-mass correction. Protein identifications were accepted with a probability greater than 99.0%, <1% FDR, and at least two peptides per protein.

Immunogold labeling TEM

Thirty-dpi transgenic nodules expressing the pFMV:myc-GmFWL3, pFMV::GmFWL3-myc, and pFMV:myc transgenes were isolated on the basis of GFP observation under a Nikon SMZ25 fluorescence stereoscope. The nodules were immediately fixed in 4% glutaraldehyde, 1% paraformaldehyde, and 0.2 M sodium cacodylate buffer (pH 7.2) for 1 h at room temperature. Fixed nodules were cut in half and stored at 4°C overnight in the buffer. For immunogold labeling, samples were embedded in LR-White resin and then sectioned to generate 100-nm cross-sections. The sections were collected onto nickel grids, blocked in 1 $\times$ PBS–Tween–BSA (0.05% [v/v] Tween 20 and 3% BSA) for 30 min, and rinsed in PBS–0.05% Tween (PBST). The grids were then labeled with anti-myc antibodies (R950-25, ThermoFisher) diluted 1:50 in PBS–0.05% Tween–1% BSA for 1 h. After three washes in PBST, the labeled samples were incubated for 1 h with the secondary antibody conjugated with 10 nm colloidal gold (A-31561, ThermoFisher) and diluted 1:100 in PBS–0.05% Tween–1% BSA. The sections were rinsed 3 times in PBS–0.05% Tween 20, then once in deionized water. Electron microscopy images were collected with a Hitachi H7500 TEM operated at 80 kV, focusing on rhizobia-infected cells of the soybean nodule. When necessary, sections were stained with uranyl acetate and lead citrate. For each transgene, three independent biological replicates were processed and observed under the transmission electron microscope.

sNucRNA-seq data pre-processing, integration, and clustering

Each sNucRNA-seq library was processed individually using 10X Genomics Cell Ranger software v6.1.1.0 for demultiplexing and alignment to the soybean reference genome from the Ensembl Plants database (*Glycine_max_v2.1.52*; http://ftp.ensemblgenomes.org/pub/plants/release-52/fasta/glycine_max/). Background contamination was subtracted using SoupX after read alignment (Young and Behjati, 2020), and doublets were filtered out using the DoubletDetection prediction method (Gayoso and Shor, 2022). Finally, we applied a minimum threshold of 500 UMIs to remove nuclei with lower transcript content. Upon normalization, integration anchors were defined for the combined set of three sNucRNA-seq root datasets and two sNucRNA-seq nodule datasets using Seurat V4 (Hao et al., 2021). The dimensional reduction was performed using the UMAP method with the first 40 principal components, selecting the top 2000 variable genes for clustering using the FindClusters method in Seurat V4. We used the Seurat object for the soybean nodule UMAP to generate expression distribution plots

of the gene set using the RidgePlot function in Seurat V4. The DotPlot function in Seurat V4 was used to generate the DotPlot expression figures.

Cell type annotation using M.C.

To annotate the soybean root and nodule cell clusters, we identified the most highly expressed genes with specific expression for each cluster using the FindAllMarkers function in Seurat V4. Probes were designed by Resolve Biosciences (Monheim am Rhein, Germany) and hybridized against 10- μ m cross-sections of fixed and paraffin-embedded roots and nodules. Upon hybridization, the cross-sections were stained with DAPI and calcofluor white, and microscopy observations were performed to reveal the positions of nuclei and cells, respectively. The microscopy images with transcript locations were analyzed using the Molecular Cartography plugin for ImageJ analysis software provided by Resolve Biosciences.

For the root, the epidermal marker genes were Glyma.19G255500, CYP93A1, Glyma.06G259400, G4DT, Glyma.09G099900, Glyma.10G070200, Glyma.01G156200, Glyma.07G130800, Glyma.20G061300, Glyma.02G149100, Glyma.04G010600, and Glyma.17G133100; the cortical marker genes were Glyma.06G235500, Glyma.11G221200, Glyma.09G216800, and Glyma.15G169100; the endodermal marker genes were Glyma.14G218700 and Glyma.16G106800; the pericycle marker genes were Glyma.02G003700, Glyma.05G023700, Glyma.11G078300, Glyma.08G125800, and Glyma.09G127000; the xylem marker genes were Glyma.15G245800, Glyma.04G063800, Glyma.06G065000, Glyma.13G334500, Glyma.15G040000, Glyma.18G197400, and Glyma.15G179500; and the phloem marker genes were Glyma.07G006500, Glyma.05G216000, Glyma.15G274200, Glyma.11G243100, and Glyma.12G154300 (see Supplemental Figures 2 and 3 for details).

For the nodule, the inner/outer cortical marker genes were Glyma.16G039800, Glyma.19G255500, CYP93A1, Glyma.06G259400, G4DT, Glyma.01G156200, Glyma.20G061300, and Glyma.03G079500; the sclereid layer marker genes were Glyma.04G063800, Glyma.06G065000, Glyma.13G334500, and Glyma.15G040000; the vascular endodermis marker genes were Glyma.04G218700, Glyma.14G227200, Glyma.16G106800, and Glyma.20G151700; the vascular bundle marker genes were Glyma.06G256000, Glyma.10G139200, Glyma.02G003700, Glyma.07G231500, Glyma.18G062100, Glyma.08G125800, Glyma.11G078300, Glyma.09G127000, Glyma.13G334500, and Glyma.15G040000; the infected cell marker genes were Glyma.17G195900 (*RIM*), Glyma.01G164600, Glyma.15G2100100, and Glyma.05G216000; and the uninfected cell marker genes were Glyma.06G235500 and Glyma.06G002000 (see Supplemental Figure 6 for details). We also used *B. diazoefficiens* probes against BAC45727, BAC46169, BAC47034, BAC48395, BAC51072, BAC51722, BAC52602, BAC52793, and BAC52805 to annotate the infected nodule cells (see Supplemental Figure 7 for details).

Cell type annotation using orthologous gene markers

To support the annotations of the soybean root cell clusters, we identified soybean orthologs of functionally validated cell-type-specific gene markers from roots of *A. thaliana* (Farmer et al., 2021) and *M. truncatula* (Cervantes-Pérez et al., 2022) (Supplemental Table 3). Gene orthology was based on identification of syntenic regions between the genome of *G. max* and those of *A. thaliana* and *M. truncatula* using CoGe (<https://genomevolution.org/coge/>).

UMAP visualization

For visualization, all sNucRNA-seq libraries for the root, the nodule, and the integration of both were combined using the Cell Ranger aggr function from 10X Genomics, and Loupe software from 10X Genomics was used to visualize the integrations.

Plant Communications

Differential gene expression and gene ontology analyses

To identify DEGs between clusters, we used DEsingle software (Miao et al., 2018), a zero-inflated negative binomial distribution method (Wang et al., 2019), with the raw read counts and thresholds of $p < 0.05$ and fold-change greater than 1.5. Gene ontology enrichment analyses were performed on the DEGs using the PlantRegMap GO Enrichment tool with a threshold of $p \leq 0.01$ <http://plantregmap.gao-lab.org/go.php>.

Comparison of soybean sNucRNA-seq and bulk RNA-seq

To evaluate the depth and sensitivity of the soybean root and nodule single-nucleus transcriptome atlases, we compared our pseudo-bulk sNucRNA-seq datasets with previously published root and nodule bulk RNA-seq datasets (Libault et al., 2010c). Using the legume information system database, we extracted bulk expression datasets (2022/11/14; https://data.legumeinfo.org/Glycine/max/expression/Wm82.gnm2.ann1.expr.Wm82.Libault_Farmer_2010/; identifiers SRR037385, SRR037386, and SRR037387 for the nodule, root tip, and entire root, respectively) and then compared the numbers of expressed genes between the bulk and pseudo-bulk RNA-seq libraries.

MDS analysis of the soybean root and nodule transcriptomes and 12-, 14-, 21-, and 28-dpi soybean nodules

To evaluate the level of similarity across the single-nucleus transcriptomes of soybean roots (16 cell clusters) and nodules (11 cell clusters) and developmental stages of the soybean nodule (i.e., 12-, 14-, 21-, and 28-dpi nodules), we first defined the integration anchors for the combined set of sNucRNA-seq datasets of the root and nodule together using Seurat V4 (Hao et al., 2021) and then performed an MDS analysis using the *ggfortify* library in R (version 4.2.2). Specifically, classical multidimensional scaling was performed to calculate a distance matrix between the different objects.

Reanalysis of single-cell RNA-seq datasets from *M. truncatula* nodules

Upon mining single-cell RNA-seq datasets from *M. truncatula* nodules (i.e., SAMC899255 and SAMC899256 from the National Genomics Data Center), we individually processed both datasets using the 10X Genomics Cell Ranger v6.1.1.0 pipeline to map the sequencing reads against the *M. truncatula* reference genome (<https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/>). After removing the “MIX” cluster 0 as described previously (Ye et al., 2022), we used the same analytical methods described above. For visualization purposes, we generated expression Ridge plots for selected *M. truncatula* and soybean genes using the RidgePlot function in Seurat V4.

GRN inference

To predict TF–target interaction pairs from gene expression data and infer the GRN, we used our 28-dpi nodule sNucRNA-seq data and the GENIE3 tool (Huynh-Thu et al., 2010). In this analysis, we defined cell subsets as cluster F and G cells and cluster H cells; gene subsets encompassed 2000 highly variable genes (specific to each cell subset), together with SNF-related genes (Roy et al., 2020) and all soybean TFs (Feng et al., 2022; Wang et al., 2023a). Duplicate genes and those with zero expression across all cells within a subset were excluded, resulting in 3133 genes for F and G and 3276 genes for H. For downstream analyses on the inferred networks, within each network, we used the top 50 000 interactions based on the importance measure.

DATA AND CODE AVAILABILITY

The sNucRNA-seq data generated in this study have been deposited at the National Center for Biotechnology Information (NCBI) under BioProject number PRJNA938968 and GEO number GSE226149. The webpage that provides access to the data from this manuscript is <http://soybeancellatlas.org>.

Single-cell RNA-seq atlas of soybean root organs

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This research was funded by the Nebraska Soybean Board, NSF awards 2127485 and 1854326, and the Nebraska Research Initiative.

AUTHOR CONTRIBUTIONS

SAC-P, PZ, and ML designed the experiments. SThibivilliers generated the single-cell RNA-seq libraries. PZ performed the molecular cloning, co-immunoprecipitation, and confocal observations and supervised the biochemical and TEM experiments. MSH, HX, IM, and AN performed molecular cloning and plant transformation to test the activity of cell-type-specific promoters. MJN carried out the mass spectrometry identification of the GmFWL3 co-immunoprecipitated proteins JK, SB, and S. Tennant performed M.C. experiments. QY and PM created the web-based tools for data visualization. SAC-P, PZ, SThibivilliers, STennant, SA, AF, OM, and ML analyzed the results. SAC-P, PZ, SA, and ML wrote the manuscript.

ACKNOWLEDGMENTS

We thank Dr. Shao-shan Carol Huang (New York University) for her support in developing and interpreting the gene networks, Dr. David Evans (Oxford Brookes University, UK), and Dr. Tsuyoshi Nakagawa (Shimane University, Matsue, Japan) for sharing the p35S:CFF-AtSUN1 and pGWB20/21 plasmids, respectively. We also thank Ms. Rosa Angelica Bedolla-Gaxiola for her artistic work (Figure 1A and 2A). We thank Dirk Anderson, Jules Russ, and Dr. You Zhou of the Flow Cytometry and Microscopy Core Research Facilities at the University of Nebraska-Lincoln Center for Biotechnology for help with sorting nuclei, confocal microscopy, and transmission electron microscopy. M.L. is a consultant for INARI, a plant biotechnology company.

Received: April 23, 2024

Revised: May 21, 2024

Accepted: June 3, 2024

Published: June 6, 2024

SUPPORTING CITATIONS

The following reference appears in the supplemental information: Qiao et al. (2017b)

REFERENCES

- Adachi, S., Minamisawa, K., Okushima, Y., Inagaki, S., Yoshiyama, K., Kondou, Y., Kaminuma, E., Kawashima, M., Toyoda, T., Matsui, M., et al. (2011). Programmed induction of endoreduplication by DNA double-strand breaks in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **108**:10004–10009.
- Agtuca, B.J., Stopka, S.A., Evans, S., Samarah, L., Liu, Y., Xu, D., Stacey, M.G., Koppelaar, D.W., Paša-Tolić, L., Anderton, C.R., et al. (2020). Metabolomic profiling of wild-type and mutant soybean root nodules using laser-ablation electrospray ionization mass spectrometry reveals altered metabolism. *Plant J.* **103**:1937–1958.
- Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S., Kakar, K., Wandrey, M., Verdier, J., Zuber, H., Ott, T., et al. (2008). A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* **55**:504–513.
- Bergersen, F. (1965). Ammonia? An Early Stable Product of Nitrogen Fixation by Soybean Root Nodules. *F.J. Bergersen. Aust. J. Biol. Sci.* **18**:1–9.
- Bhagwat, A.A., Mithö, A., Pfeffer, P.E., Kraus, C., Spickers, N., Hotchkiss, A., Rgen Ebel, J., and Keister, D.L. (1999). Further Studies of the Role of Cyclic-Glucans in Symbiosis. An *ndvC* Mutant of *Bradyrhizobium japonicum* Synthesizes Cyclodecakis-(133)-Glucosyl 1.

- Block, A., Toruño, T.Y., Elowsky, C.G., Zhang, C., Steinbrenner, J., Beynon, J., and Alfano, J.R. (2014). The *Pseudomonas syringae* type III effector HopD1 suppresses effector-triggered immunity, localizes to the endoplasmic reticulum, and targets the Arabidopsis transcription factor NTL9. *New Phytol.* **201**:1358–1370.
- Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., Morieri, G., Mysore, K.S., Wen, J., Oldroyd, G.E.D., et al. (2014). 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.
- Brewin, N.J. (1991). Development of the Legume Root Nodule.
- Broughton, W.J., and Dilworth, M.J. (1971). Control of leghaemoglobin synthesis in snake beans. *Biochem. J.* **125**:1075–1080.
- Buono, R.A., Hudecek, R., and Nowack, M.K. (2019). Plant proteases during developmental programmed cell death. *J. Exp. Bot.* **70**:2097–2112.
- Cebolla, A., María Vinardell, J., Kiss, E., Oláh, B., Roudier, F., Kondorosi, A., and Kondorosi, E. (1999). The mitotic inhibitor ccs52 is required for endoreduplication and ploidy-dependent cell enlargement in plants. *EMBO J.* **18**:4476–4484.
- 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**.
- Clarke, V.C., Loughlin, P.C., Gavrin, A., Chen, C., Brear, E.M., Day, D.A., and Smith, P.M.C. (2015). Proteomic analysis of the soybean symbiosome identifies new symbiotic proteins. *Mol. Cell. Proteomics* **14**:1301–1322.
- Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J., and Udvardi, M.K. (2004). Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant J.* **39**:487–512.
- Curtin, S.J., Xiong, Y., Michno, J.M., Campbell, B.W., Stec, A.O., Čermák, T., Starker, C., Voytas, D.F., Eamens, A.L., and Stupar, R.M. (2018). CRISPR/Cas9 and TALENs generate heritable mutations for genes involved in small RNA processing of *Glycine max* and *Medicago truncatula*. *Plant Biotechnol. J.* **16**:1125–1137.
- Dalton, D.A., Post, C.J., and Langeberg, L. (1991). Effects of ambient oxygen and of fixed nitrogen on concentrations of glutathione, ascorbate, and associated enzymes in soybean root nodules. *Plant Physiol.* **96**:812–818.
- Dauphin, B.G., Ranocha, P., Dunand, C., and Burlat, V. (2022). Cell-wall microdomain remodeling controls crucial developmental processes. *Trends Plant Sci.* **27**:1033–1048.
- Davies, M.J., and Puppo, A. (1992). Direct detection of a globin-derived radical in leghaemoglobin treated with peroxides. *Biochem. J.* **281**:197–201.
- De La Peña, T.C., Fedorova, E., Pueyo, J.J., and Mercedes Lucas, M. (2018). The symbiosome: Legume and rhizobia co-evolution toward a nitrogen-fixing organelle? *Front. Plant Sci.* **8**.
- 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.e5.
- de Zélicourt, A., Diet, A., Marion, J., Laffont, C., Ariel, F., Moison, M., Zahaf, O., Crespi, M., Gruber, V., and Frugier, F. (2012). Dual involvement of a *Medicago truncatula* NAC transcription factor in root abiotic stress response and symbiotic nodule senescence. *Plant J.* **70**:220–230.
- Di Giacomo, E., Laffont, C., Sciarra, F., Iannelli, M.A., Frugier, F., and Frugis, G. (2017). KNAT3/4/5-like class 2 KNOX transcription factors are involved in *Medicago truncatula* symbiotic nodule organ development. *New Phytol.* **213**:822–837.
- Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**:1262–1267.
- Doll, N.M. (2023). Eternal nodules to fix the nitrogen issue: Promotion of soybean nodule senescence by a NAC/CYP module. *Plant Cell* **35**:2709–2710.
- Dupont, L., Alloing, G., Pierre, O., Msehli, S.E., Hopkins, J., Hérouart, D., and Frendo, P. (2012). The legume root nodule: from symbiotic nitrogen fixation to senescence. In *Senescence*, T. Nagata, ed. (IntechOpen). p. Ch. 8.
- 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.
- Farmer, A., Thibivilliers, S., Ryu, K.H., Schiefelbein, J., and Libault, M. (2021). 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.
- Feng, X., Fang, F., Long, H., Zeng, R., and Yao, Y. (2022). Single-cell RNA-seq data analysis using graph autoencoders and graph attention networks. *Front. Genet.* **13**:1003711.
- Ferguson, B.J., Indrasumunar, A., Hayashi, S., Lin, M.H., Lin, Y.H., Reid, D.E., and Gresshoff, P.M. (2010). Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant Biol.* **52**:61–76.
- Fonseca-García, C., López-García, C.M., Pacheco, R., Armada, E., Nava, N., Pérez-Aguilar, R., Solís-Miranda, J., and Quinto, C. (2022). Metallothionein1A Regulates Rhizobial Infection and Nodulation in *Phaseolus vulgaris*. *Int. J. Mol. Sci.* **23**.
- Gayoso, A., and Shor, J. (2022). JonathanShor/DoubletDetection: Doubletdetection v4.2 Advance Access, published March 12, 2022. <https://doi.org/10.5281/ZENODO.6349517>.
- Graumann, K., Runions, J., and Evans, D.E. (2010). Characterization of SUN-domain proteins at the higher plant nuclear envelope. *Plant J.* **61**:134–144.
- Guo, Y., Ren, G., Zhang, K., Li, Z., Miao, Y., and Guo, H. (2021). Leaf senescence: progression, regulation, and application. *Mol. Hortic.* **1**:5.
- Haney, C.H., and Long, S.R. (2010). Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proc. Natl. Acad. Sci. USA* **107**:478–483.
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* **184**:3573–3587.e29.
- Hussey, S.G., Mizrahi, E., Spokevicius, A.V., Bossinger, G., Berger, D.K., and Myburg, A.A. (2011). SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in Arabidopsis fibres and increases fibre cell area in Eucalyptus. *BMC Plant Biol.* **11**:173.
- Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**:e12776.
- Islam, N., Krishnan, H.B., and Natarajan, S. (2022). Quantitative proteomic analyses reveal the dynamics of protein and amino acid accumulation during soybean seed development. *Proteomics* **22**:e2100143.
- Ivanov, R., Tiedemann, J., Czihal, A., Schallau, A., Diep, L.H., Mock, H.-P., Claus, B., Tewes, A., and Bäumlein, H. (2008). EFFECTOR OF TRANSCRIPTION2 is involved in xylem differentiation and

- includes a functional DNA single strand cutting domain. *Dev. Biol.* **313**:93–106.
- Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorrity, M.W., Saunders, L., Bubb, 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.
- Jeon, H.-W., and Byrne, M.E. (2021). SAW homeodomain transcription factors regulate initiation of leaf margin serrations. *J. Exp. Bot.* **72**:1738–1747.
- Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical Statistical Model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**:5383–5392.
- Kiirika, L.M., Schmitz, U., and Colditz, F. (2014). The alternative *Medicago truncatula* defense proteome of ROS—defective transgenic roots during early microbial infection. *Front. Plant Sci.* **5**:341.
- Kim, D., Cho, Y.-h., Ryu, H., Kim, Y., Kim, T.-H., and Hwang, I. (2013). BLH1 and KNAT3 modulate ABA responses during germination and early seedling development in *Arabidopsis*. *Plant J.* **75**:755–766.
- Kouchi, H., Tsukamoto, M., and Tajima, S. (1990). Differential expression of nodule-specific (nodulin) genes in the infected, uninfected cortical cells of soybean (*Glycine max*) root nodules. *J. Plant Physiol.* **135**:608–617.
- Kouchi, H., Fukai, K., Katagiri, H., Minamisawa, K., and Tajima, S. (1988). Isolation and enzymological characterization of infected and uninfected cell protoplasts from root nodules of *Glycine max*. *Physiol. Plantarum* **73**:327–334.
- Křeček, P., Skúpa, P., Libus, J., Naramoto, S., Tejos, R., Friml, J., and Zajímalová, E. (2009). The PIN-FORMED (PIN) protein family of auxin transporters. *Genome Biol.* **10**:249.
- Kumar, R., Kushalappa, K., Godt, D., Pidkowich, M.S., Pastorelli, S., Hepworth, S.R., and Haughn, G.W. (2007). The *Arabidopsis* BEL1-LIKE HOMEODOMAIN proteins SAW1 and SAW2 act redundantly to regulate KNOX expression spatially in leaf margins. *Plant Cell* **19**:2719–2735.
- Laffont, C., Ivanovici, A., Gautrat, P., Brault, M., Djordjevic, M.A., and Frugier, F. (2020). The NIN transcription factor coordinates CEP and CLE signaling peptides that regulate nodulation antagonistically. *Nat. Commun.* **11**:3167.
- Lee, L.R., Wengier, D.L., and Bergmann, D.C. (2019). Cell-type-specific transcriptome and histone modification dynamics during cellular reprogramming in the *Arabidopsis* stomatal lineage. *Proc. Natl. Acad. Sci. USA* **116**:21914–21924.
- Li, R., Chen, H., Yang, Z., Yuan, S., and Zhou, X. (2020). Research status of soybean symbiosis nitrogen fixation. *Oil Crop Science* **5**:6–10.
- Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R.J., Bilgin, D.D., Radwan, O., Neece, D.J., Clough, S.J., May, G.D., et al. (2010). Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiol.* **152**:541–552.
- Libault, M., Zhang, X.C., Govindarajulu, M., Qiu, J., Ong, Y.T., Brechenmacher, L., Berg, R.H., Hurley-Sommer, A., Taylor, C.G., and Stacey, G. (2010). A member of the highly conserved *FWL* (tomato *FW2.2-like*) gene family is essential for soybean nodule organogenesis. *Plant J.* **62**:852–864.
- Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G., and Stacey, G. (2010c). An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* **63**:86–99.
- 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.
- 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.
- Maekawa, T., Kusakabe, M., Shimoda, Y., Sato, S., Tabata, S., Murooka, Y., and Hayashi, M. (2008). Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in *Lotus japonicus*. *MPMI (Mol. Plant-Microbe Interact.)* **21**:375–382.
- Mergaert, P., Kereszt, A., and Kondorosí, E. (2020). Gene expression in nitrogen-fixing symbiotic nodule cells in *Medicago truncatula* and other nodulating plants. *Plant Cell* **32**:42–68.
- Miao, Z., Deng, K., Wang, X., and Zhang, X. (2018). DEsingle for detecting three types of differential expression in single-cell RNA-seq data. *Bioinformatics* **34**:3223–3224.
- Moreau, S., Davies, M.J., Mathieu, C., Hérouart, D., and Puppo, A. (1996). Leghemoglobin-derived radicals. Evidence for multiple protein-derived radicals and the initiation of peribacteroid membrane damage. *J. Biol. Chem.* **271**:32557–32562.
- Mühlenbock, P., Szechynska-Hebda, M., Plaszczyca, M., Baudo, M., Mateo, A., Mullineaux, P.M., Parker, J.E., Karpinska, B., and Karpinski, S. (2008). Chloroplast signaling and LESION SIMULATING DISEASE1 regulate crosstalk between light acclimation and immunity in *Arabidopsis*. *Plant Cell* **20**:2339–2356.
- Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., et al. (2007). Improved gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci. Biotechnol. Biochem.* **71**:2095–2100.
- Ohyama, T., and Kumazawa, K. (1978). Incorporation of ¹⁵N into various nitrogenous compounds in intact soybean nodules after exposure to ¹⁵N₂ gas. *Soil Sci. Plant Nutr.* **24**:525–533.
- Ohyama, T., and Kumazawa, K. (1979). Assimilation and transport of nitrogenous compounds originated from ¹⁵N₂ fixation and ¹⁵NO₂ absorption. *Soil Sci. Plant Nutr.* **25**:9–19.
- Pecrix, Y., Sallet, E., Moreau, S., Bouchez, O., Carrere, S., Gouzy, J., Jardinaud, M.F., and Gamas, P. (2022). DNA demethylation and hypermethylation are both required for late nodule development in *Medicago*. *Nat. Plants* **8**:741–749.
- Pingault, L., Zogli, P., Brooks, J., and Libault, M. (2018). Enhancing phenotyping and molecular analysis of plant root system using ultrasound aeroponic technology. *Curr. Protoc. Plant Biol.* **3**:e20078.
- Poole, P.S., and Ledermann, R. (2022). Maintaining osmotic balance in legume nodules. *J. Exp. Bot.* **73**:8–10.
- Puppo, A., Herrada, G., and Rigaud, J. (1991). Lipid peroxidation in peribacteroid membranes from French-bean nodules. *Plant Physiol.* **96**:826–830.
- Puppo, A., Groten, K., Bastian, F., Carzaniga, R., Soussi, M., Lucas, M.M., De Felipe, M.R., Harrison, J., Vanacker, H., and Foyer, C.H. (2005). Legume nodule senescence: roles for redox and hormone signalling in the orchestration of the natural aging process. *New Phytol.* **165**:683–701.
- Qiao, Z., Brechenmacher, L., Smith, B., Strout, G.W., Mangin, W., Taylor, C., Russell, S.D., Stacey, G., and Libault, M. (2017). The *GmFWL1* (*FW2.2-like*) nodulation gene encodes a plasma membrane microdomain-associated protein. *Plant Cell Environ.* **40**:1442–1455.
- Qiao, Z., Pingault, L., Zogli, P., Langevin, M., Rech, N., Farmer, A., and Libault, M. (2017). A comparative genomic and transcriptomic analysis at the level of isolated root hair cells reveals new conserved root hair regulatory elements. *Plant Mol. Biol.* **94**:641–655.
- Roux, B., Rodde, N., Jardinaud, M.F., Timmers, T., Sauviac, L., Cottret, L., Carrère, S., Sallet, E., Courcelle, E., Moreau, S., et al. (2014). An integrated analysis of plant and bacterial gene expression

- in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**:817–837.
- 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.
- Ryu, K.H., Huang, L., Kang, H.M., and Schiefelbein, J. (2019). Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* **179**:1444–1456.
- Severin, A.J., Woody, J.L., Bolon, Y.T., Joseph, B., Diers, B.W., Farmer, A.D., Muehlbauer, G.J., Nelson, R.T., Grant, D., Specht, J.E., et al. (2010). RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. *BMC Plant Biol.* **10**:160.
- 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.e9.
- Shimoda, Y., Nishigaya, Y., Yamaya-Ito, H., Inagaki, N., Umehara, Y., Hirakawa, H., Sato, S., Yamazaki, T., and Hayashi, M. (2020). The rhizobial autotransporter determines the symbiotic nitrogen fixation activity of *Lotus japonicus* in a host-specific manner. *Proc. Natl. Acad. Sci. USA* **117**:1806–1815.
- Soyano, T., Kouchi, H., Hirota, A., and Hayashi, M. (2013). NODULE INCEPTION directly targets NF-Y subunit genes to regulate essential processes of root nodule development in *Lotus japonicus*. *PLoS Genet.* **9**:e1003352.
- Stopka, S.A., Agtuca, B.J., Koppelaar, D.W., Paša-Tolić, L., Stacey, G., Vertes, A., and Anderton, C.R. (2017). Laser-ablation electrospray ionization mass spectrometry with ion mobility separation reveals metabolites in the symbiotic interactions of soybean roots and rhizobia. *Plant J.* **91**:340–354.
- 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.* **65**:1536–1552.
- Tajima, S., Nomura, M., and Kouchi, H. (2004). Ureide biosynthesis in legume nodules. *Front. Biosci.* **9**:1374–1381.
- Teale, W.D., Paponov, I.A., and Palme, K. (2006). Auxin in action: signalling, transport and the control of plant growth and development. *Nat. Rev. Mol. Cell Biol.* **7**:847–859.
- Tedeschi, F., Rizzo, P., Huong, B.T.M., Czihal, A., Rutten, T., Altschmied, L., Scharfenberg, S., Grosse, I., Becker, C., Weigel, D., et al. (2019). EFFECTOR OF TRANSCRIPTION factors are novel plant-specific regulators associated with genomic DNA methylation in Arabidopsis. *New Phytol.* **221**:261–278.
- Thibivilliers, S., Farmer, A., and Libault, M. (2020a). Biological and cellular functions of the microdomain-associated FWL/CNR protein family in plants. *Plants* **9**:377.
- Thibivilliers, S., Anderson, D., and Libault, M. (2020b). Isolation of plant root nuclei for single cell RNA sequencing. *Curr. Protoc. Plant Biol.* **5**:e20120.
- Udvardi, M., and Poole, P.S. (2013). Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**:781–805.
- Veličković, D., Agtuca, B.J., Stopka, S.A., Vertes, A., Koppelaar, D.W., Paša-Tolić, L., Stacey, G., and Anderton, C.R. (2018). Observed metabolic asymmetry within soybean root nodules reflects unexpected complexity in rhizobacteria-legume metabolite exchange. *ISME J.* **12**:2335–2338.
- Verdier, J., Torres-Jerez, I., Wang, M., Andriankaja, A., Allen, S.N., He, J., Tang, Y., Murray, J.D., and Udvardi, M.K. (2013). Establishment of the *Lotus japonicus* Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. *Plant J.* **74**:351–362.
- Vinardell, J.M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., Tarayre, S., Roudier, F., Mergaert, P., Kondorosi, A., et al. (2003). Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* Nodules. *Plant Cell* **15**:2093–2105.
- Wang, T., Li, B., Nelson, C.E., and Nabavi, S. (2019). Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. *BMC Bioinf.* **20**:40.
- Wang, J., Chen, Y., and Zou, Q. (2023). Inferring gene regulatory network from single-cell transcriptomes with graph autoencoder model. *PLoS Genet.* **19**:e1010942.
- Wang, L., Tian, T., Liang, J., Li, R., Xin, X., Qi, Y., Zhou, Y., Fan, Q., Ning, G., Becana, M., et al. (2023). A transcription factor of the NAC family regulates nitrate-induced legume nodule senescence. *New Phytol.* **238**:2113–2129.
- Wendrich, J.R., Yang, B.J., Vandamme, N., Verstaen, K., Smet, W., Velde, C. V. de, Minne, M., Wybouw, B., Mor, E., Arents, H.E., et al. (2020). Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science*, 370.
- Xie, K., and Yang, Y. (2013). RNA-Guided genome editing in plants using a CRISPR-Cas system. *Mol. Plant* **6**:1975–1983.
- Xu, X., Crow, M., Rice, B.R., Li, F., Harris, B., Liu, L., Demesa-Arevalo, E., Lu, Z., Wang, L., Fox, N., et al. (2021). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell* **56**:557–568.e6.
- 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, published December 2022. <https://doi.org/10.1016/j.molp.2022.10.019>.
- Yoon, H.-K., Kim, S.-G., Kim, S.-Y., and Park, C.-M. (2008). Regulation of Leaf Senescence by NTL9-mediated Osmotic Stress Signaling in Arabidopsis. *Mol. Cell.* **25**:438–445.
- Yoshiyama, K.O., Aoshima, N., Takahashi, N., Sakamoto, T., Hiruma, K., Saijo, Y., Hidema, J., Umeda, M., and Kimura, S. (2020). SUPPRESSOR OF GAMMA RESPONSE 1 acts as a regulator coordinating crosstalk between DNA damage response and immune response in *Arabidopsis thaliana*. *Plant Mol. Biol.* **103**:321–340.
- Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* **9**:giaa151.
- Yu, Y. (2020). Remorins: Essential regulators in plant-microbe interaction and cell death induction. *Plant Physiol.* **183**:435–436.
- Yu, Y., Bai, Y., Wang, P., Wang, Y., Wan, H., Liu, C., and Ni, Z. (2020). Soybean nuclear factor YA10 positively regulates drought resistance in transgenic *Arabidopsis thaliana*. *Environ. Exp. Bot.* **180**:104249.
- Yu, H., Xiao, A., Wu, J., Li, H., Duan, Y., Chen, Q., Zhu, H., and Cao, Y. (2023). GmNAC039 and GmNAC018 activate the expression of cysteine protease genes to promote soybean nodule senescence. *Plant Cell* **35**:2929–2951.
- Zasloff, M. (2002). Antimicrobial Peptides of Multicellular Organisms.
- Zhu, J., Wang, J., Li, Q., Wang, J., Liu, Y., Li, J., Chen, L., Shi, Y., Li, S., Zhang, Y., et al. (2019). QTL analysis of nodule traits and the identification of loci interacting with the type III secretion system in soybean. *Mol. Genet. Genom.* **294**:1049–1058.