Supplementary Information for

Regulation of Stomatal Development by Stomatal Lineage miRNAs

Jiali Zhu, Ji-Hwan Park, Seulbee Lee, Jae Ho Lee, Daehee Hwang, June M. Kwak and Yun Ju Kim

Correspondence: Yun Ju Kim (yjkim77@ibs.re.kr); June M. Kwak (jkwak@dgist.ac.kr)

This PDF file includes:

Supplementary text

Fig. S1 to S7

SI references

Other supplementary materials for this manuscript include the following:

Dataset S1 to S6

SI Materials and Methods

Plant materials and growth conditions

Arabidopsis thaliana ecotype Columbia-0 (Col-0) was used as WT in this study. Plants were grown on $^1/_2$ Murashige and Skoog (MS) agar media without sucrose or in soil at 22℃ under long-day (16h-light/8h-dark) conditions.

Plasmid construction and plant transformation

The molecular constructs were generated using the Gateway Cloning Technology (Invitrogen). Detailed information and the primers used for the constructs are listed in Dataset S6. For the knock down line of miR861 (STTM-miR861), plasmid pFGC5941-PacI-ath-STTM861-3p was a gift from Guiliang Tang (Addgene plasmid #84180; http://n2t.net/addgene:84180; RRID:Addgene_84180) (1). Floral dipping method was used to introduce plasmid constructs (2). In brief, *Arabidopsis* floral buds were dipped into the solution containing Agrobacterium tumefaciens (GV3101), 0.05% surfactant Silwet L-77, and 5% sucrose. The transformed plants were kept for 24 hours under high humidity and dark conditions.

Microscopy

Seedlings were stained with freshly prepared FM4-64 solution (Invirtrogen, F34653) in the dark for 30~40min and rinsed by water before visualization of the plasma membrane. Images were acquired on a Confocal laser scanning microscope (Olympus FV1200) with excitation at 500-540nm for GFP and 600-645nm for FM4-64.

RNA Immunoprecipitation and small RNA isolation

RNA immunoprecipitation experiments were conducted as previously described (Carbonell, 2012) with a minor modification. In brief, four grams of *Arabidopsis* leaves from 6 day-old seedlings were ground in liquid nitrogen and dissolved in 12-ml protein lysis buffer (50mM Tris-HCl, PH 7.4, 2.5mM $MgCl_2$, 100mM KCl, 0.1% Nonident P-40, 0.5mM PMSF, proteinase inhibitor cocktails (EDTA-free protease

inhibitor cocktail Tablet/50ml lysis buffer, Roche), 50 units/ml RNase inhibitor (Thermo Scientific)). Cell extracts were centrifuged two times at 12,000 rcf at 4℃ for 15 min to remove cell debris. The supernatant was incubated with binding control agarose beads (Chromotek, Cat. No. bab-20) without antibody for 30min at 4℃ to remove the non-specific binding, then incubated with anti-GFP agarose beads (Chromotek, Cat. No. gta-20) for 2hr at 4℃. The beads were washed 3 times by protein lysis buffer at 4℃. RNA was recovered by incubating the beads in same volume of proteinase K buffer (50mM Tris-HCl, PH 7.4, 25mM EDTA, 300 mM NaCl, 1.25% SDS, 1μg/ul proteinase K (Thermo Scientific, Cat. No. AM2546)) for 15min at 65℃ and purified using the Zymo micro RNA kit (Zymo Research, Cat. No. R1061). Small RNAs were quantified using the Agilent 2100 Bioanalyzer system (Aglient small RNA Kit, Cat. No. 5067-1548).

Western blot analysis

Proteins were extracted with 1X SDS buffer containing 50mM Tris-HCl, pH6.8, 10% glycerol, 2%SDS, 12.5mM EDTA, 1% *β*-mercaptoethanol, and 0.02% bromophenol blue. The protein was loaded on a 6% (w/v) SDS-PAGE gel and transferred to Hybond-C Extra membrane (GE Healthcare). For GFP detection, the membranes were probed with eGFP Tag Antibody (1:1000 v/v, Invitrogen, Cat. No. CAB4211) overnight at 4℃ and then secondary anti-rabbit antibody (1:3000 v/v, Sigma, Cat. No. A3687) for 1hr at RT. The signal was detected using the Supersignal West Pico PLUS Chemiluminescent Substrate (Thermo-Fisher Scientific, Cat. No. 34577) and developed using an Auto Film Processor (TongYang Medical).

Total RNA extraction and RT-qPCR analysis

Total RNA was extracted from seedlings using the TRIzol Reagent (Invirtrogen) and Quick-RNA Plant Mini-prep Kit (Zymo Research). cDNA synthesis was conducted using the cDNA EcoDry Premix Kit (Takara), and qPCR was performed using the iQ SYBR Green Supermis (Bio-Rad) in a CFX384 Real-Time PCR System (Bio-Rad). The expression levels of the genes analyzed were normalized to *ACTIN2*. Three

biological replicates were conducted. The primer sequences are listed in Dataset S6.

Stomatal phenotype analysis

Seedlings grown on 1/2 Murashige and Skoog (MS) agar media without sucrose were used to analyze stomatal development phenotype. Stomatal pairs and numbers were counted in the same position on the abaxial side of cotyledons of 10-day-sold seedling in an area of 780 X 780 μm^2. To calculate the percentage of plants having stomatal pairs per unit area (780x780 μm^2), ten cotyledons from ten independent seedlings were used. For example, when ten independent cotyledons were analyzed, five cotyledons showed no stomatal pair, three cotyledons showed one stomatal pair, and two cotyledons showed more than two stomatal pairs. Then the percentage of seedlings displaying stomatal pairs per unit of area would be 50 % (no pair), 30 % (1 pair), and 20 % (≥2 pairs). Experiments were repeated three times with ten independent plants for each experiment.

Small RNA library sequencing and data analysis

Small RNAs were converted to cDNA libraries using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs, MA, USA). 5'- and 3'-adapters were ligated to the purified small RNA, followed by reverse transcription using Protoscript II reverse transcriptase and incorporation of index tags by PCR. The resulted libraries were subjected to size selection (145–160 bp) using Sage Pippin Prep (Sage science, MA, USA) on a 3% agarose gel. The cDNA libraries were the sequenced using the Illumina Hiseq 2500 system, generating 50 bp single-end reads.

After the library sequencing, the adapter sequences were trimmed, and the reads were filtered out with length < 15 bp. The remained reads were then aligned to the *Arabidopsis thaliana* genome (TAIR10) using Bowtie 1 (3) with the option allowing for all possible multiple alignments without any mismatches. Then, the number of reads that were aligned onto the annotated loci of 1) coding and non-coding genes excluding miRNA genes in TAIR10 and 2) mature miRNAs in miRBase 21 (4)

were calculated by using coverageBed of BEDTools (5). For each sample, reads per million mapped reads (RPM) values of the individual loci were computed, and quantile normalization (6) was applied to the RPM values.

Identification of DE miRNAs

The small RNAs expressed in at least one of the six stomatal development stages were identified as those with RPMs > 1 in at least two of the triplicates. A pseudo-value was added to the RPM values of the individual samples and converted to the $\log_2$-RPM values. To identify DE miRNAs from the expressed small RNAs, all pairwise comparisons were performed among the six stages. For each expressed small RNA in the individual comparisons, a paired *t*-test was conducted to obtain T-value, and the empirical null distribution for T-value was generated by applying random sampling experiments 1,000 times (7). The empirical false discovery rate (FDR) for the observed T-value of each small RNA was estimated by performing the two-tailed test on the empirical distribution. Among the expressed small RNAs, we selected the mature miRNAs with their FDRs < 0.1 and absolute $\log_2$-median ratio > 0.58 in at least one of the comparisons. Finally, miRNAs were removed whose maximum expression changes ($\log_2$-median ratio) across the six stages were less than one, and the remaining miRNAs were identified as DE miRNAs.

To investigate the dynamic expression patterns of stomatal lineage miRNAs, we clustered the expression patterns of DE miRNAs in the two paths (*SPCH–MUTE–FAMA* and *EPF2–EPF1*) by using correlation-based k-means clustering with k = 12 and k = 8 for *SPCH–MUTE–FAMA* and *EPF2–EPF1* paths, respectively. The twelve and eight clusters were further grouped into three (stomatal entry and differentiation) and two groups (stomatal entry, commitment and differentiation) by applying a hierarchical clustering method to centroid values of the clusters with complete linkage and Euclidean dissimilarity measure.

Identification of potential target DEGs for DE miRNAs

mRNA expression profiles across the four stages (ML1, SPCH, MUTE, FAMA) were

obtained from the raw RNA-sequencing data generated by Adrian et al. (8). The adapter sequence and the bases with PHRED scores < 20 from the reads were trimmed, and the reads with length of the remaining sequences < 30 bp were removed. Then, the reads were aligned to the TAIR10 genome using Bowtie 2 (9) with the default option. For each annotated gene in TAIR10, the number of reads was calculated using HTSeq-count (10), and the trimmed mean of M-values (TMM) normalization to the read counts was performed using edgeR (11), leading to the acquisition of counts per million mapped reads (CPM) values. Next, the genes with CPM > 1 in at least one sample were determined as expressed genes (12). After adding one pseudo-value to the CPM values, they were converted into $\log_2$-CPM values, and the quantile normalization was conducted. To identify DEGs, the similar statistical test used for the identification of DE miRNAs was applied to the all pairwise comparisons between the stages. For each gene in the individual comparisons, FDR for Student's T-statistics was computed by applying the two-tailed test with the empirical distribution of T-values. The DEGs were identified as the genes with their FDRs and absolute $\log_2$-median ratios > 10^{th} and 90^{th} percentiles of the empirical distributions for T-value and $\log_2$-median-ratio, respectively. The cutoffs for FDR and $\log_2$-median-ratio were determined through manual validation after the trials of several cutoff values (0.05, 0.1, 0.2, 0.25 and 0.3).

To identify putative target DEGs for DE miRNAs, two criteria were considered; 1) base-pairing relationships and 2) anti-correlation of expression profiles during stomatal lineage between miRNAs and mRNAs. Predicted and/or collected 33,743 base-pairing relationships between 426 miRNAs and 16,074 target mRNAs were predicted and/or collected from 1) the Small RNA Target Prediction tool (13), 2) imiRTP (14), 3) TargetFinder (15), and 4) psRNATarget (16). To identify significantly anti-correlated expression profiles between DE miRNAs and DEGs, Pearson's correlation coefficient was measured and empirical null distribution for the correlation coefficient was generated by random permutation. Then, the significantly anti-correlated relationships were determined as those with the correlation coefficients < -0.5, which is 25^{th} percentile of the empirical distribution.

Functional enrichment analysis of Gene Ontology biological processes (GOBPs)

To understand cellular processes enriched by the three groups of potential target DEGs in the *SPCH-MUTE-FAMA* path, the enrichment analysis by was performed using DAVID software (17). GOBPs significantly represented by the genes were identified as those with the number of involved genes $\geq$ 3 and $P < 0.1$.

Construction of miRNA-target mRNA regulatory network

To select key DE miRNAs that significantly regulate the DEGs during stomatal lineage progression, the hypergeometric test was applied with the inferred relationships between DE miRNAs and their putative target DEGs. Key DE miRNAs were determined with the resulted $P < 0.01$. To construct a regulatory network of key DE miRNAs and their target DEGs, Cytoscape software (18) was used with the inferred regulatory relationships as well as experimentally validated protein-protein interactions that were collected from CCSB interactome database(19), HitPredict (20), iNID (21), the IntAct molecular interaction database (22), TAIR (23), and STRING (24).

RNA ligase-mediated rapid amplification of 5' cDNA ends

Total RNA was extracted from 10-day-old seedlings using the TRIzol reagent (Invirtrogen) and Quick-RNA plant Mini prep Kit (Zymo Research, Cat. No. R2024) followd by DNAse I digestion. Poly(A) RNA was purified from total RNA (25μg) using the MagJET mRNA Enrichment Kit (Thermo Scientific, Cat. No. K2811). Then the poly(A) RNA was ligated with GeneRacer RNA oligo using T4 RNA ligase for 1 hr at 37℃. The ligated RNA was cleaned and concentrated to 6μl using the RNA Clean & Concentrator Kit (Zymo research, Cat. No. R1016). The mRNA was reverse transcribed using the Prime Script 1st strand cDNA Synthesis Kit (Takara, Cat. No. 6110A) to generate RACE-ready first strand cDNA with known 5' ends. To obtain 3'ends, the first strand cDNA was used for PCR amplification using a forward gene-specific primer (GSP) and the GeneRacer 5' primer. The RACE PCR products were

purified using the HiGene Gel & Pcr purification System Kit (Biofact, Cat. No. GP104-100). The PCR products were cloned into PGEM-T Easy Vector Systems, and 10 individual colonies were selected for sequencing to detect the miRNA cleavage site.

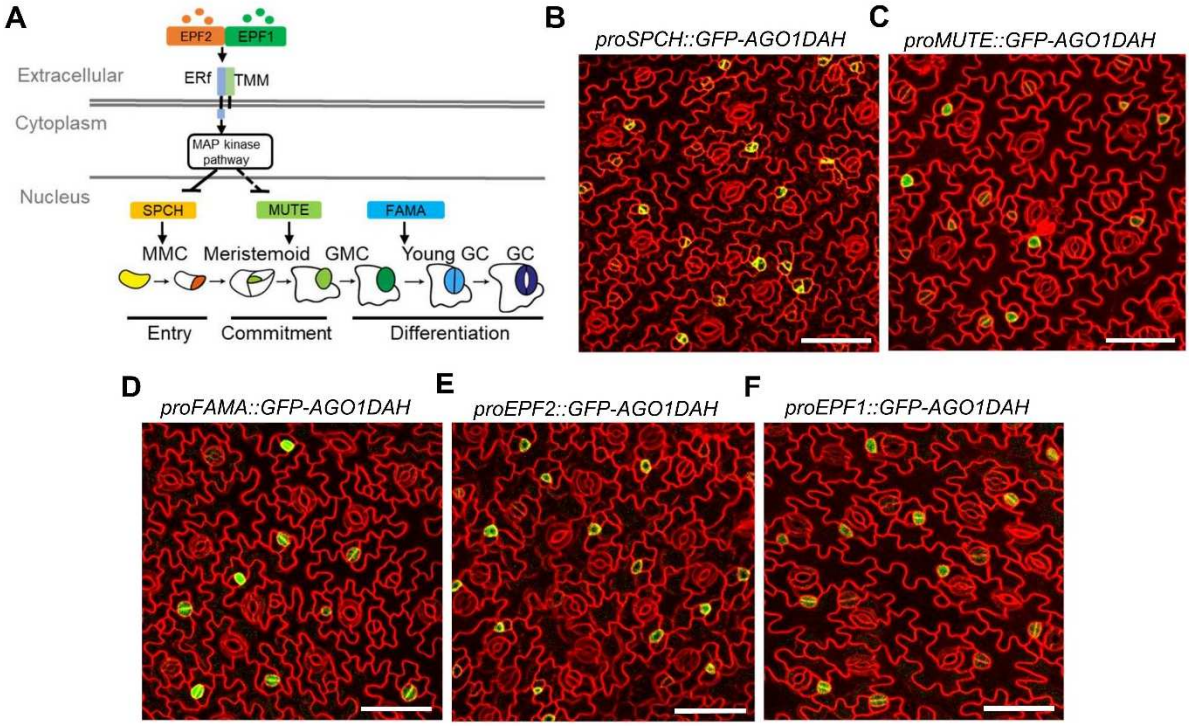

Fig. S1. Confocal images of reporter lines used for isolation of developmental stage-specific AGO-associated miRNAs in the stomatal lineage. (A) Schematic presentation of the genetic regulatory network in stomatal development modified from Adrian et al. (2015) (8). Binding of EPF1 and EPF2 peptide ligands to the TMM/ERf receptor kinases leads to the activation of the MAPK cascade modulating the key transcriptional factors, SPCH, MUTE, and FAMA, which in turn control stomatal development (entry, commitment, differentiation). MMC, meristemoid mother cell (yellow); Meristemoid (orange); GMC, guard mother cell (green); GC, guard cell (blue, dark blue). The promoters of the developmental stage-specific marker genes (color-coded) were used for the generation of transgenic plants. (B-F) Expression patterns of GFP-AGO1DAH under the stomatal lineage-specific promoters in the transgenic plants. Scale bar: 50 μm.

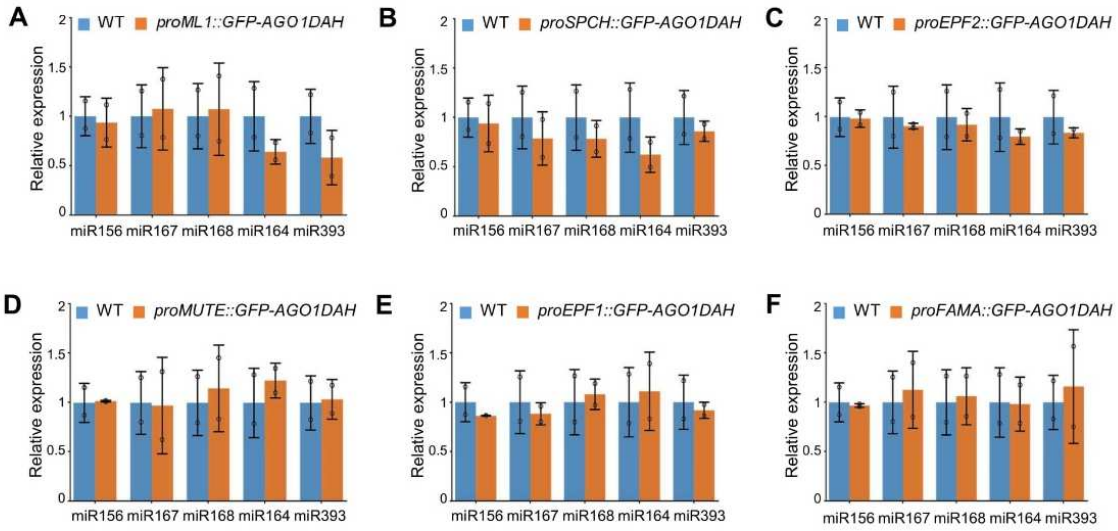

Fig. S2. Endogenous miRNA expression levels in GFP-AGO1DAH reporter lines used for stomatal lineage miRNA profiling. Quantitative RT-PCR analysis of miR156, miR167, miR168, miR164, and miR393 in each stage-specific GFP-AGO1DAH transgenic line. Six-day-old seedlings were used for the analysis. The expression values were normalized to U6. Error bars represent mean ± s.d. calculated from two independent biological repeats.

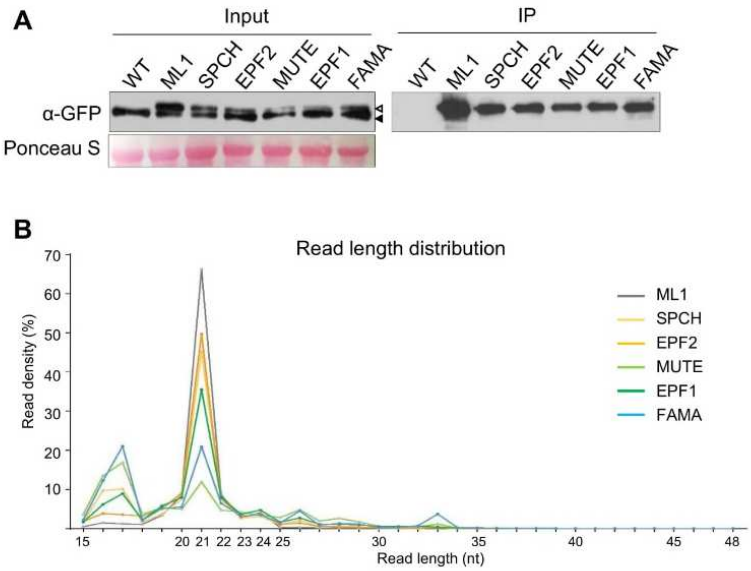

Fig. S3. Isolation of developmental stage-specific AGO1 and associated miRNAs. (A) Immunoprecipitation of GFP-AGO1DAH from each stage of stomatal lineage cells. Immunoprecipitated GFP-AGO1DAH proteins were detected using anti-GFP antibody (top). Ponceau-S staining shows the amount of proteins loaded in each lane (bottom). The blank arrowhead indicates the GFP-AGO1DAH protein that is absent in WT control, and the black arrowhead indicates non-specific proteins detected. (B) Size distribution of the developmental stage-specific AGO1-associated miRNAs in the stomatal lineage cells.

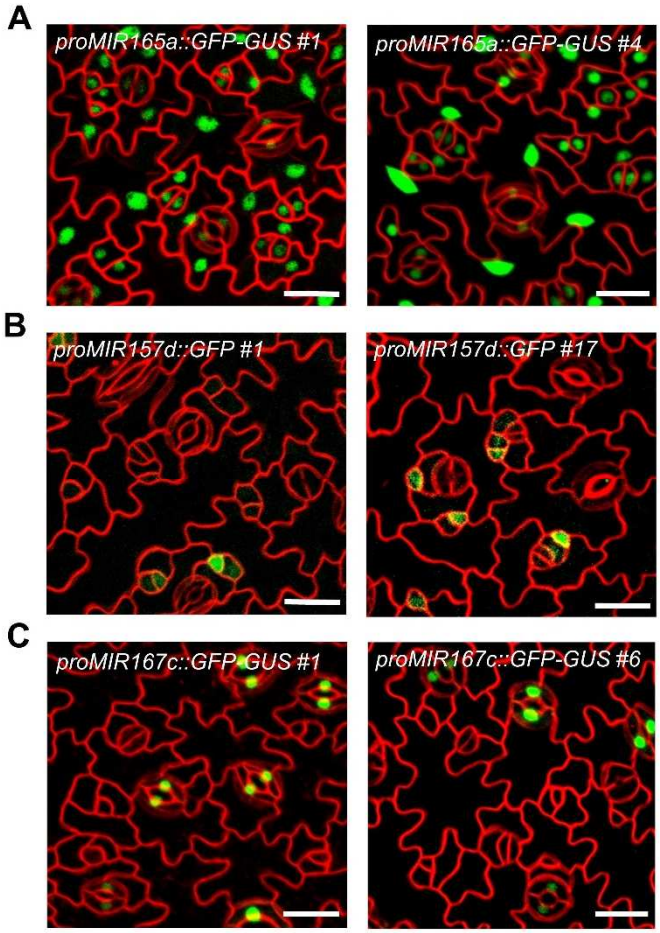

Fig. S4. Confocal images of the epidermis of *proMIR165a::GFP-GUS*, *proMIR157d::GFP* and proMIR167c::GFP-GUS transgenic plants. At least three independent plant lines were analyzed. Cell outlines are visualized by FM4-64. Scale bar: 20 μm.

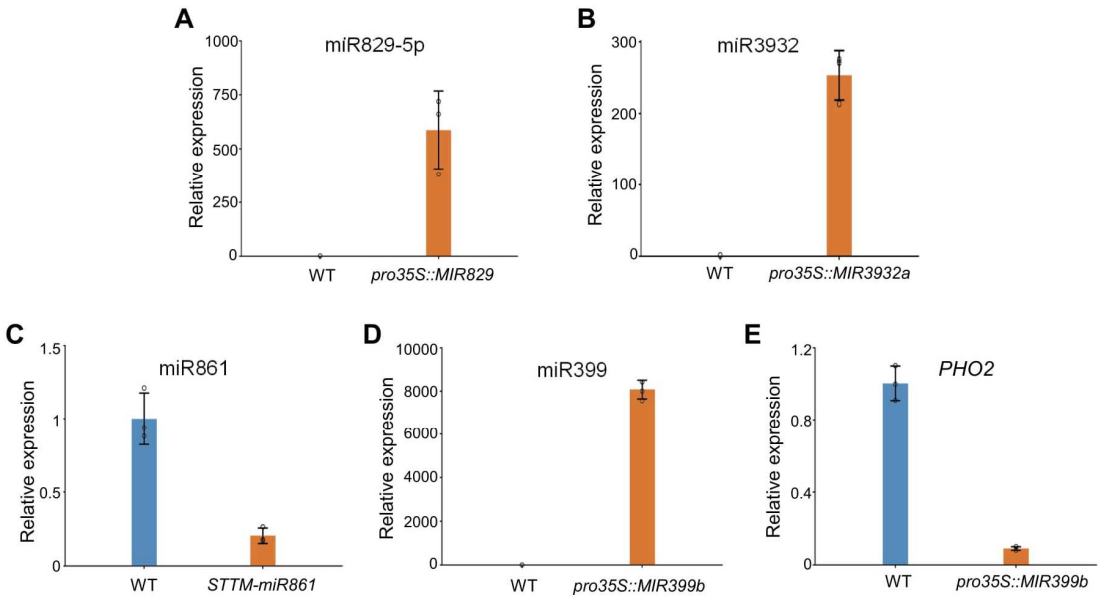

Fig. S5. Accumulation levels of stomatal lineage miRNAs in the overexpression or knock-down plants. Quantitative RT-PCR analysis of miR829 (A), miR3932 (B), miR861 (C), miR399 (D), and *PHO2* (E) in the miRNA overexpression and knock-down plants. Error bars represent mean ± s.d. calculated from three independent biological repeats.

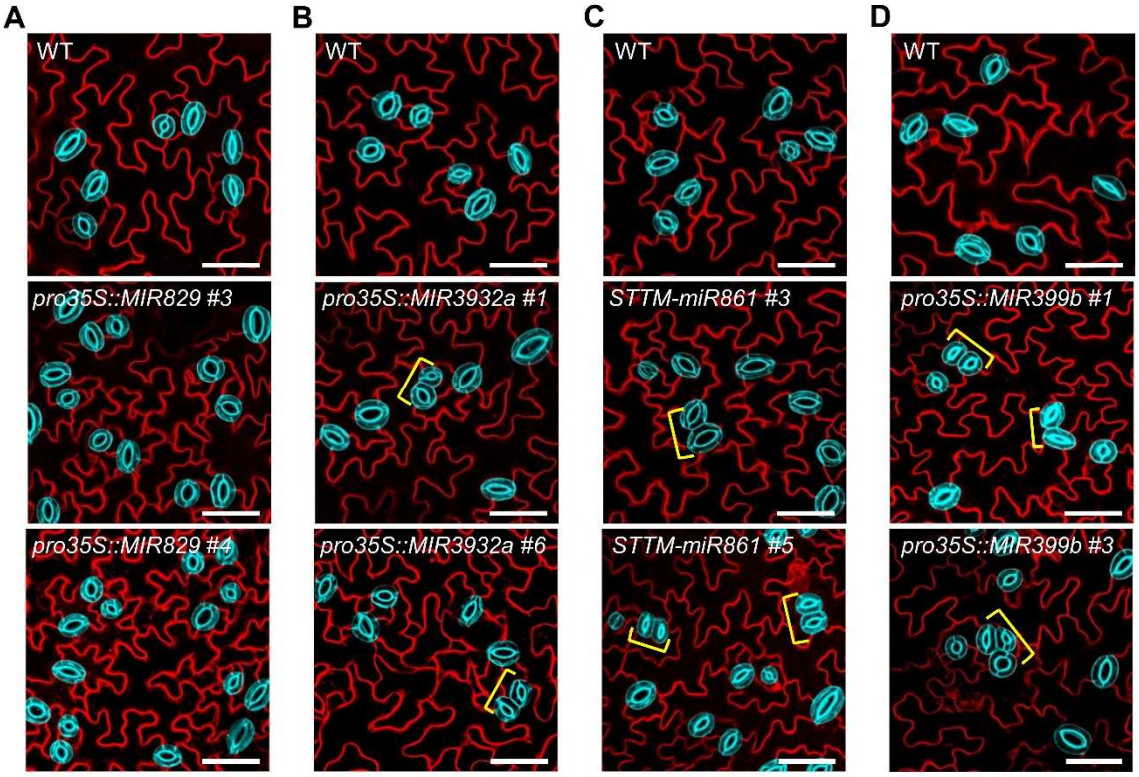

Fig. S6. Stomatal phenotypes of transgenic plants harboring *pro35S::MIR829*, *pro35S::MIR3932a*, *STTM-miR861,* or *pro35S::MIR399b* transgenes. At least three independent transgenic lines were analyzed. Mature guard cells are highlighted in blue for elucidation, and the brackets indicate stomatal clusters. Cell outlines are visualized by FM4-64. Scale bar: 50 μm.

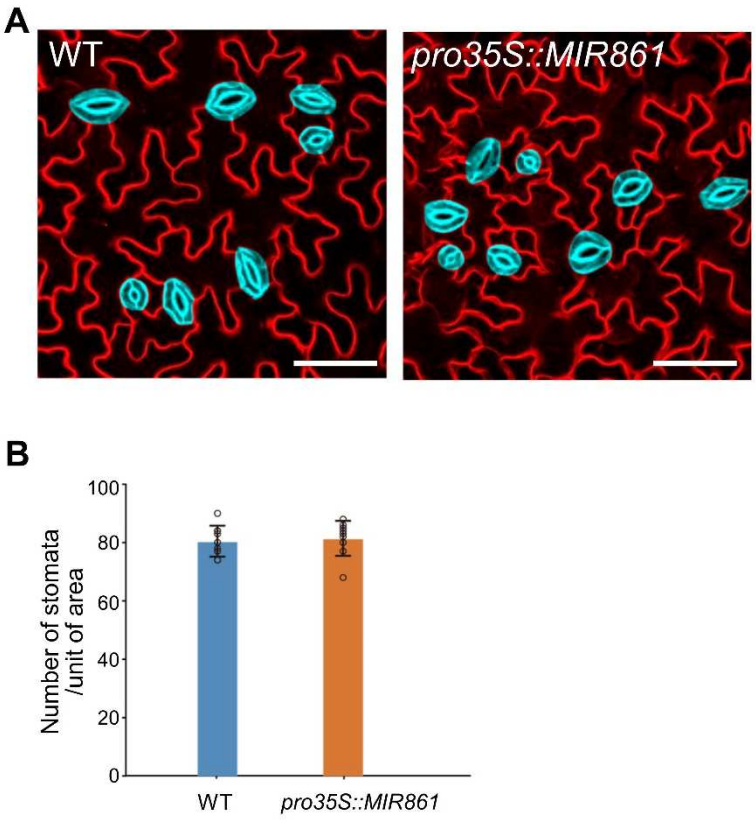

Fig. S7. Stomatal development phenotype of *pro35S::MIR861* transgenic plants. (A) Confocal image of WT and *pro35S::MIR861* transgenic plants. Mature guard cells are highlighted in blue for elucidation. Cell outlines are visualized by FM4-64. Scale bar, 50 μm. (B) Guard cell number per unit area (780 X 780 μm^2) in cotyledons of 8-day-old seedlings. Error bars represent mean ± s.d. calculated from nine plants.

SI Dataset S1 (Dataset_S1.xlsx)

List of DE miRNAs in SPCH-MUTE-FAMA and EPF2-EPF1 modules during stomatal lineage

SI Dataset S2 (Dataset_S2.xlsx)

List of DE miRNA–potential target DEG interaction pairs

SI Dataset S3 (Dataset_S3.xlsx)

GOBPs represented by potential target DEGs of DE miRNAs

SI Dataset S4 (Dataset_S4.xslx)

List of putative target genes of DE miRNAs presented in GO analysis

SI Dataset S5 (Dataset_S5.xlsx)

List of 10 hub-miRNAs in the stomatal entry group

SI Dataset S6 (Dataset_S6.xlsx)

List of primers used in this study

References

1. G. Tang *et al.* Construction of short tandem target mimic (STTM) to block the functions of plant and animal microRNAs. *Methods* **58**, 118-125 (2012).

2. M. Martinez-Trujillo, V. Limones-Briones, J. L. Cabrera-Ponce, L. Herrera-Estrella, Improving transformation efficiency ofArabidopsis thaliana by modifying the floral dip method. *Plant Mol. Biol. Rep.* **22**, 63-70 (2004).

3. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009)

4. A. Kozomara, S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68-D73 (2013).

5. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

6. B. M. Bolstad, R. A. Irizarry, M. Åstrand, T. P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).

7. H. J. Kim *et al.,* Time-evolving genetic networks reveal a NAC troika that negatively regulates leaf senescence in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4930-4939 (2018).

8. J. Adrian *et al.,* Transcriptome dynamics of the stomatal lineage: birth, amplification, and termination of a self-renewing population. *Dev Cell.* **33**,107-118 (2015).

9. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359 (2012).

10. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2014).

11. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2009).

12. K. Boo *et al.,* Pontin functions as an essential coactivator for Oct4-dependent lincRNA expression in mouse embryonic stem cells. *Nat. Commun.* **6**, 6810 (2015).

13. M. W. Jones-Rhoades, D. P. Bartel, Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* **14**, 787-799 (2004).

14. J. Ding, D. Li, U. Ohler, J. Guan, S. Zhou, Genome-wide search for miRNA-target interactions in Arabidopsis thaliana with an integrated approach. *BMC Genomics* **13**, S3 (2012)

15. N. Fahlgren *et al.,* High-throughput sequencing of arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS One* **2**, e219 (2007).

16. X. Dai, Z. Zhuang, P. X. Zhao, psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **46**, W49-W54 (2018).

17. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44-57 (2009).

18. P. Shannon *et al.,* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498-2504 (2003).

19. Arabidopsis Interactome Mapping Consortium, Evidence for network evolution in an arabidopsis interactome map. *Science* **333**, 601-607 (2011).

20. Y. López, K. Nakai, A. Patil, HitPredict version 4: comprehensive reliability scoring of physical protein–protein interactions from more than 100 species. *Database* **2015** (2015).

21. D. Choi*, et al.,* iNID: An analytical framework for identifying network models for

interplays among developmental signaling in arabidopsis. *Mol. Plant* **7**, 792-813 (2014).

22. H. Hermjakob *et al.,* IntAct: an open source molecular interaction database. *Nucleic Acids Res.* **32**, D452-D455 (2004).

23. T. Z. Berardini *et al.,* The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* **53**, 474-485 (2015).

24. D. Szklarczyk *et al.,* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362-D368 (2016).