PNAS
www.pnas.org

Supplementary Information for

Cell-type specific transcriptome and histone modification dynamics during cellular reprogramming in the Arabidopsis stomatal lineage

Laura R. Lee[1], Diego L. Wengier [1,2,3] Dominique C. Bergmann [1, 2*]

[1]Department of Biology, Stanford University, Stanford, CA, USA 94305
[2]HHMI, Stanford University, Stanford, CA, USA 94305
[3]Present address, INGEBI – CONICET Vuelta de Obligado 2490, 1428 – CABA, Argentina
* Correspondence: dbergmann@stanford.edu

This PDF file includes:

Supplementary text
Figures S1 to S9
Table S1
Legends for Datasets S1 to S2
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S2

www.pnas.org/cgi/doi/10.1073/pnas.1911400116

Supplementary Information Text

Extended Materials and Methods

Plant Growth Conditions

All lines used in this study were in the Col-0 ecotype. Seedlings were grown on 0.5 Murashige and Skoog (MS) medium at 22°C under 16 hr-light/8 hr-dark cycles. Plants grown for FACS were grown horizontally while plants grown for confocal imaging were grown vertically.

Plasmid Construction

Primers for generating an artificial microRNA against CLF and SWN were designed using WMD3 (http://wmd3.weigelworld.org/) (http://wmd3.weigelworld.org/cgi-bin/webapp.cgi). and amplified using the pRS300 vector as a template. Primers A and B were modified from the original protocol to allow the resulting PCR product to be compatible with pENTR/D-TOPO (Life Technologies, Carlsbad, CA).

YFP No Stop (YFP NS) was amplified with Phusion Polymerase and YFP F/YFP NS R primers using R4pGWB440 as template. YFP NS was cloned into pENTR D-TOPO (pE-YFP NS). A binary vector expressing YFP-YFP under FAMAp (2.5 kb, subcloned from our pENTR FAMAp clone) was built by combining R4pGWB440 with pE-YFP NS and a vector carrying FAMAp flanked by attL4 and attR1 in a multiple Gateway recombination reaction.

The sequence for WIND3 (AT1G36060) was amplified from gDNA with Phusion Polymerase and CACC adaptors were added to make the amplicon compatible with pENTR/D-TOPO (Life Technologies, Carlsbad, CA). WIND3 was then subcloned under the SPCHp into R4pGWB440 in a multiple Gateway recombination reaction.

Sequences for all primers used in this work can be found in SI Appendix Table S1.

Microscopy

Confocal microscopy images were taken with a Leica SP5 microscope and processed in ImageJ. Cell outlines were visualized by 0.1 mg/ml propidium iodide in water (Molecular Probes, P3566).

Seedlings were collected at 12 dpg. Samples were cleared in 7:1 ethanol:acetic acid, treated for 15 min with 1 N potassium hydroxide, rinsed in water, and mounted in Hoyer's medium. Whole seedlings were mounted for quantification of the lobing phenotype. Differential contrast interference (DIC) images were obtained from cotyledons at 20× and 40x on a Leica DM2500 microscope.

RNA extraction from seedlings and RT-qPCR

Aerial tissue of seedlings was collected at 12 dpg. For studies that correlated *WIND3* expression level with reprogramming enhancement, one cotyledon was saved for DIC imaging and the remaining aerial tissue was used for RNA extraction with the RNeasy plant mini kit (QIAGEN, Valencia, CA) with on-column DNAse digestion. 700 ng of total RNA was used for cDNA synthesis using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA). The qPCR reactions were performed on a CFX96 Real-Time PCR detection system (Bio-Rad, Hercules, CA) with the Ssoadvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA). Three technical replicates were performed per seedling and three seedlings were used per line. Expression values were normalized to the reference gene PP2A using the ΔC_T method and relative expression of the target was calculated from the ratio of $FAMA^{LGK}$+SPCHp:WIND3 to $FAMA^{LGK}$. The standard error for ΔC_T was propagated via the formula Error(target+reference) =

$(Error(target)^2+Error(reference)^2)^{1/2}$ where the target is the target gene and the reference is PP2A. Error is expressed as $2^{-\Delta\Delta CT}$±Error(target+reference).

Protoplasting and Fluorescence Activated Cell Sorting

Protoplasts were generated as described previously (1) with minor changes. Briefly, the aerial tissue of approximately 300 12 dpg plants were harvested and then incubated in 25 mL protoplasting buffer – 1.25% cellulase (Yakult), 0.3% macerozyme (Yakult), 0.4M mannitol, 20mM MES, 20mM KCl, 10 mM $CaCl_2$, 0.1% BSA, adjusted to pH 5.7 with 1M TrisCl pH 7.5 – with shaking for 3 hours to ensure maximum digestion. Cells were strained through a 40 µm cell strainer (BD Falcon), pelleted by spinning at 500g for 5 minutes, and gently resuspended in 2 mls of protoplasting buffer. Protoplasts were then immediately used for FACS.

Col-0 protoplasts were used as a negative control for YFP signal and plants expressing ML1p:YFP-RC12A (an abundant and bright marker) as a positive control for YFP expression to enable us to design a gating strategy. YFP signal from FAMAp:YFP-YFP was then used to isolate WT and LGK GC protoplasts. Protoplast were sorted on a FACSAria II (BD Biosciences) fitted with a 100 µm nozzle. Cell sorting/flow cytometry analysis for this project was done on instruments in the Stanford Shared FACS Facility. FACSAria II settings were established as described previously (1) with minor adjustments to voltage settings that varied with routine system calibrations.

For RNA-seq experiments, 20,000 GC protoplasts were captured per replicate per genotype. Sorted protoplasts were collected into 350µl of RNA extraction buffer (Buffer RLT+β-mercapto ethanol from Qiagen). Protoplasts were collected on multiple days and were stored in RNA extraction buffer at -80ºC until all RNA could be processed in tandem.

For ChIP-seq experiments, 100,000 GC protoplasts were captured per replicated per genotype. Sorted protoplasts were collected into 350µl of PBS with 2x Roche Protease inhibitor cocktail (#4693159001) held at 4ºC. Samples were brought to 1ml with PBS supplemented with protease inhibitor and crosslinked by adding 37% formaldehyde to a final concentration of 1% and incubating on ice for 7 minutes. The reaction was quenched by adding glycine to a final concentration of 0.125M and incubating on ice for a further 5 minutes. Samples were then flash frozen in liquid nitrogen and stored at -80ºC until all samples could be processed in tandem.

One concern about our method of capturing cells to analyze for reprogramming is that the protoplasting process itself could induce a general reprogramming. To test the extent of this potential issue, we compared our data to list of genes differentially expressed due to protoplasting (2). Only thirteen genes associated with H3K27me3 that is differentially enriched during LGK reprogramming are also differentially expressed due to protoplasting (2) (Dataset S1, p-value = 0.028). In addition, the most differentially depleted loci in WT GCs relative to whole aerial tissue are specifically implicated in the development of the *Arabidopsis* epidermis (Table 1). Therefore, protoplasting does not seem to mask the cell type-specific identity signature of WT GCs and does not preclude the comparison of protoplasts with whole tissue.

RNA-seq Library Preparation

RNA was extracted from protoplasts using the RNEasy Micro extraction kit (Qiagen) according to manufacturer's instructions. Of the resulting 12 microliters of extracted RNA, 3 were used for quantitation with a Bioanalyzer 2 Plant RNA Pico Assay. Once we confirmed all RNA samples had RNA Integrity Numbers (RIN) of 9 or better we used 5 microliters of total RNA per sample for cDNA synthesis with the Ovation RNA-seq System V2 from Nugen according to manufacturer's instructions. The resulting cDNA concentration was quantified using the Qubit 2.0. Aliquots of

each cDNA sample were diluted to a 100µl volume with a 10 ng/µl concentration in TE pH 8.0 for sonication with a diagenode bioruptor. The following settings were used to generate a 200 bp average fragment size: 4ºC, power setting low, cycle conditions (seconds on/off) 30/90 for 60 minutes. Fragment size distribution was confirmed with the BioAnalyzer High Sensitivity DNA Assay (Agilent) prior to library generation with the Ovation Ultra Low System V2 (NuGEN), which was performed according to manufacturer's instructions. The resulting libraries were sequenced in 50bp single end reads on the HiSeq2500 (Illumina).

Chromatin Immunoprecipitation and ChIP-seq Library Preparation

All ChIP samples were generated from pools of 100,000 GC protoplasts. Two pools of protoplasts were used as replicates for each biological condition and antibody , i.e. two pools of 100,000 WT GC protoplasts were used as biological replicates to produce the WT GC H3K27me3 profile. Each ChIP sample was produced with a corresponding input control, i.e. one per biological replicate. ChIP was performed as previously described(3) beginning with the nuclear pelleting and lysis steps and ending with the addition of protease buffer. Minor changes were made, mainly to sample volumes, as follows. Previously crosslinked protoplast samples were removed from -80ºC and thawed on ice. Nuclei were pelleted by centrifugation for 1 hour at 16,000g at 4ºC. The supernatant was removed, and samples were resuspended in 30 µl of nuclei lysis buffer with gentle pipetting. Samples were incubated on ice for 10 minutes and then flash frozen in liquid nitrogen. After thawing on ice again, samples were brought to 100µl for sonication as previously described. Prior to immunoprecipitation, 10µl of sonicated chromatin were removed per sample and held at -20ºC for input controls. ChIP and input samples were then processed in tandem beginning with the reverse-cross linking step through DNA purification. Each biological replicate was produced with an input control.

During the immunoprecipitation phase of the protocol, antibodies against H3K27me3 (Millipore 07-449, lot # 2652971) and H3K4me3 (Millipore 07-473, lot # 2664283) were used for ChIP at 5µg/ml and 2.2µl/ml respectively. 50µl of Dynabeads Protein A (ThermoFisher 10001D) per sample were used to isolate antibody-bound chromatin. Wash steps – high salt, LiCl, and TE – were carried out at 1mL volumes. Immune complexes were eluted into 100µl of ChIP elution buffer. Input aliquots were thawed on ice and then brought to 100µl with ChIP elution buffer. Volumes of NaCl and protease buffer were scaled to reflect the smaller ChIP elution volume (4µl and 10.4µl respectively). Following protease digestion, immunoprecipitated and input DNA were purified using SPRI beads (Beckman Coulter, AMPURE XP, A63880) following manufacturer's instructions. Samples were eluted into 10µl of nuclease free water and then stored at -20ºC.

The eluted DNA was then used for library preparation with the Ovation Ultra Low System V2 (NuGEN) following manufacturer's instructions, beginning with the End Repair phase of the protocol. The resulting libraries were sequenced in 50bp single end reads on the HiSeq2500 (Illumina).

RNA-seq Data Analysis

RNA-seq was performed using two biological replicates, i.e. independently sorted pools of 20,000 GCs, per genotype. Reads were mapped to TAIR10.31 with Bowtie2 using default settings. The number of mapped reads were counted in genes using HTSeq-count ignoring strandedness with otherwise default settings. The resulting raw count matrices were then analyzed in DESeq2 with design=~genotype. Gene expression was considered differential if they had an adjusted p value threshold of 0.05 and a log2 fold change threshold of 1. The count matrices were also converted from raw counts to transcript per million (TPM) and these normalized values were used for all visualizations with the exception of PCAs, which were generated using DESeq2. The ICI was calculated as described previously (4) using the following samples from the stomatal lineage transcriptional map(5): FGF, MMG, ML1Y. The following parameters were set to calculate the

matrix of spec scores: l=10, u=10, with no median filter. Normalized RNA-seq reads are available in Dataset S1.

ChIP-seq Data Analysis

Reads were mapped to TAIR10.31 with Bowtie2 using default settings. Peaks were called using as in the ENCODE Uniform Processing Pipeline for replicated histone ChIP-seq experiments. Briefly, peaks were called with MACS2 and rejected unless they appeared in both replicates or at least two pseudoreplicates generated from the pooled samples. Biological replicates and input controls were incorporated as is described in the ENCODE Uniform Processing Pipeline for replicated histone ChIP-seq experiments. Briefly, peaks are called from ChIP samples for areas that are enriched over input using MACS2 with the following settings: -g 135000000 -p 1e-2 --nomodel --shift 0 --keep-dup all -B --SPMR. In addition, the --extsize for each sample was set and empirically determined by strand cross correlation analysis of mapped reads in the ChIP sample. Peaks are retained if they appear in both biological replicates (samples generated from individual protoplast pools) or two pseudo replicates. Pseudoreplicates are peak sets called on half of the pooled reads, chosen at random without replacement. Peaks were annotated to genes, including landmarks such as the TSS or TTS, using default settings with HOMER (6).

For differential binding analysis, the number of reads in peaks for each sample were normalized to the number of reads in the same region for each samples' respective input control using DiffBind (7). This analysis was performed using the high-confidence peak set derived as described above. Read enrichment analysis within the high-confidence peak set was conducted using mapped reads from the two biological replicates described above. Differential histone peaks were identified using DiffBind with default settings to compare WT and LGK GCs with a false discovery rate threshold of 0.05. The same analysis was used to compare aerial tissue H3K27me3 with WT GC H3K27me3 with the exception that full library size was ignored during normalization to render the reads within peak sets more comparable. Gene ontology singular enrichment analysis was performed with Agrigo (http://bioinfo.cau.edu.cn/agriGO/analysis.php) using default settings. Full GO term lists are in Dataset S2.

Other Statistical Analysis and Data Visualization

Statistical significance of overlapping gene sets was conducted with the GeneOverlap package implemented in R 3.3. Read density enrichment heat maps, average profiles, and Spearman correlation coefficient calculations of sequencing data sets were conducted with deeptools. Signal coverage tracks were generated using MACS2 as implemented in the ENCODE uniform processing pipeline. All other data visualizations were generated with ggplot2 in R 3.3.

Data Availability

The raw data, bed files containing ChIP peaks, bigWig files showing ChIP fold change over input, and TPM-normalized RNA abundance measurements are deposited in GEO under the accession GSE118138.

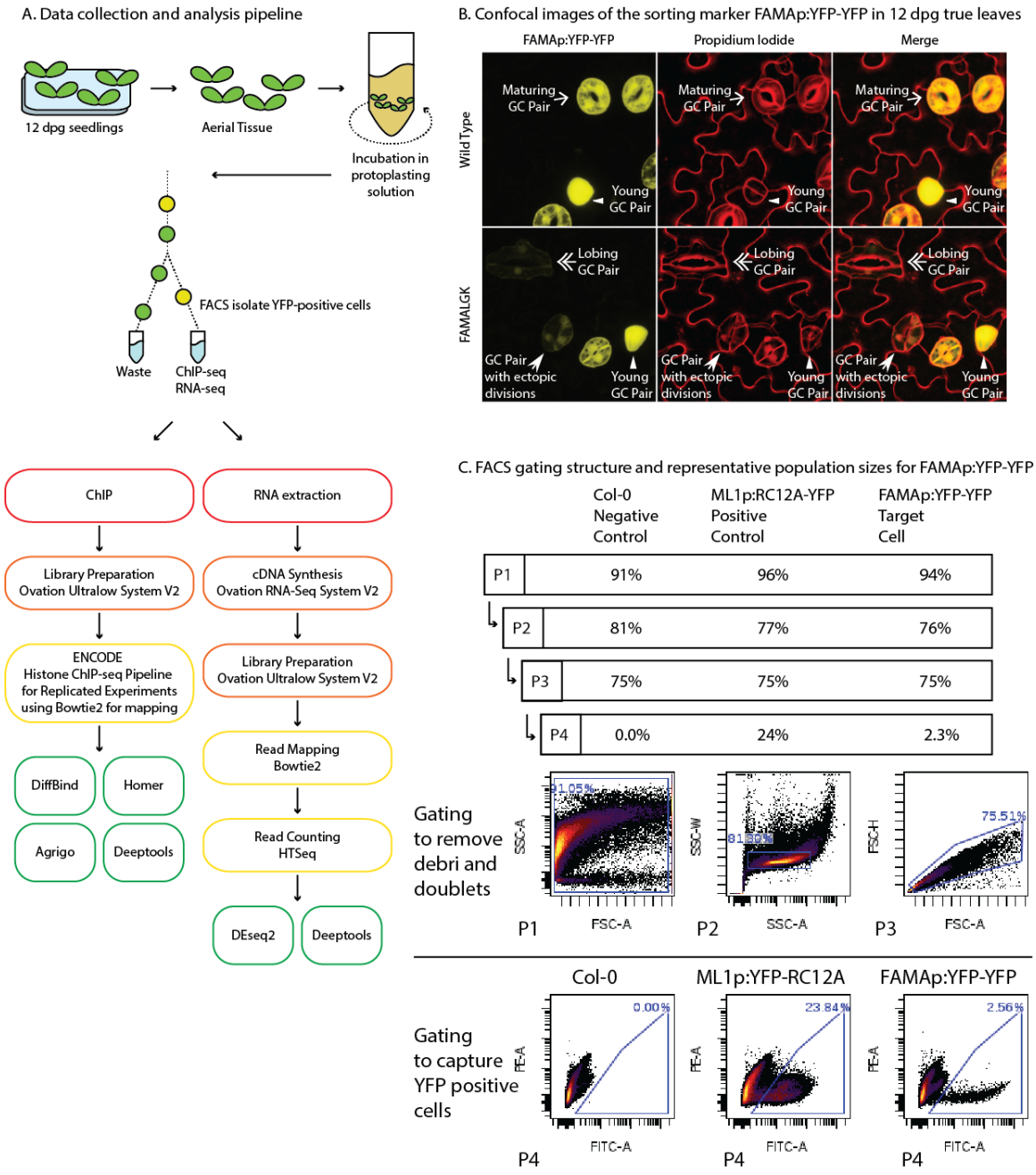


Figure S1: Cell collection and data analysis pipeline

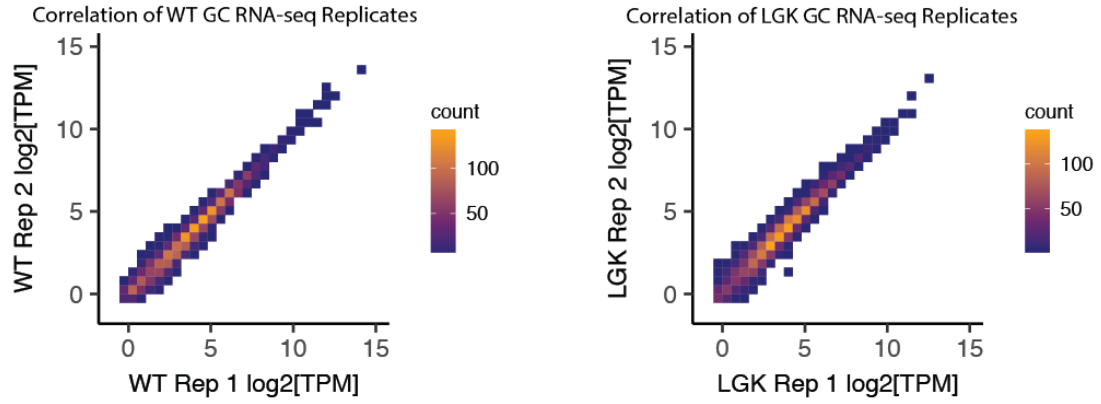
A: Overall scheme of cell selection, processing and downstream analysis pipeline: nucleic acid collection (red), library synthesis (orange), initial (yellow) and secondary (green) data processing. B: Confocal microscopy images showing the expression of FAMAp:YFP-YFP in Col-0 and FAMA^{LGK} backgrounds at 12 dpg in true leaves. White arrowheads indicate peak expression in young GCs, single and double arrows point to GCs that show evidence of reprogramming and lower YFP. Cell walls are visualized with propidium iodide C: Gating structure used to capture YFP-positive GCs and representative population of cells within selective gates represented as the percentage of cells from the parental gate (i.e. 81% of cells captured by the P2 gate were captured in the parental P1 gate). 100,000 total cells were analyzed to generate each plot. Col-0

was used as a negative control for YFP. Plants expressing ML1:YFP-RC12A, an abundant bright YFP marker, were used as a positive control to establish initial gating strategy.

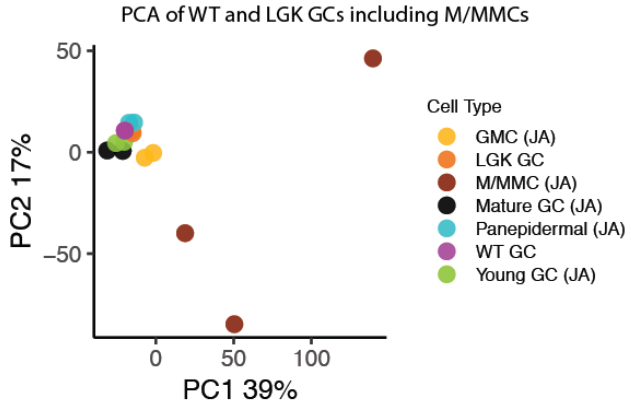
A. Mapping statistics for RNA-seq experiments

	WT		LGK	
	rep1	rep2	rep1	rep2
Total Reads	$2.5x10^7$	$3.0x10^7$	$2.7x10^7$	$2.5x10^7$
Mapped Reads	$2.3x10^7$	$2.6x10^7$	$2.4x10^7$	$2.2x10^7$

B. Correlations of gene expression values between replicates



C. PCA of WT and LGK GCs with other stomatal lineage cells



D. Classification of WT and LGK GCs among lineage and non-lineage cell types

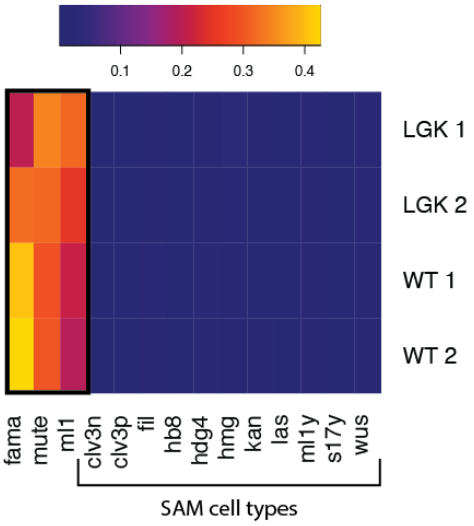


Figure S2: RNA-seq quality control and ICI based cell-type identification

A: Table summarizing the total and mapped fraction of RNA-seq reads. B: Correlation between RNA-seq replicates (R^2 value for WT replicates is 0.97, R^2 for LGK replicates is 0.95). C: PCA plots similar to Fig 1B but including (SPCHp:SPCH-YFP, SSY) samples, showing how variability

among SSY replicates masked PCs of the remaining samples. D: A heatmap of ICI-defined similarities between cells sorted for this work (y axis) and previously defined cell types in the shoot from (8, 9) (x axis). WT GCs are most similar to previously sorted young GCs (FAMAp:GFP-FAMA, FGF) and are decreasingly similar to GMCs (MUTEp:MUTE-GFP, MG) and pan-epidermal samples (ML1p:YFP-RC12A, ML1Y). LGK GCs share mixed similarity to other stomatal lineage cell types, but do not resemble shoot meristem. Because previous work (Adrian et al., 2015) showed that stomatal lineage cells more closely these samples than root samples, we did not display the root comparisons here.

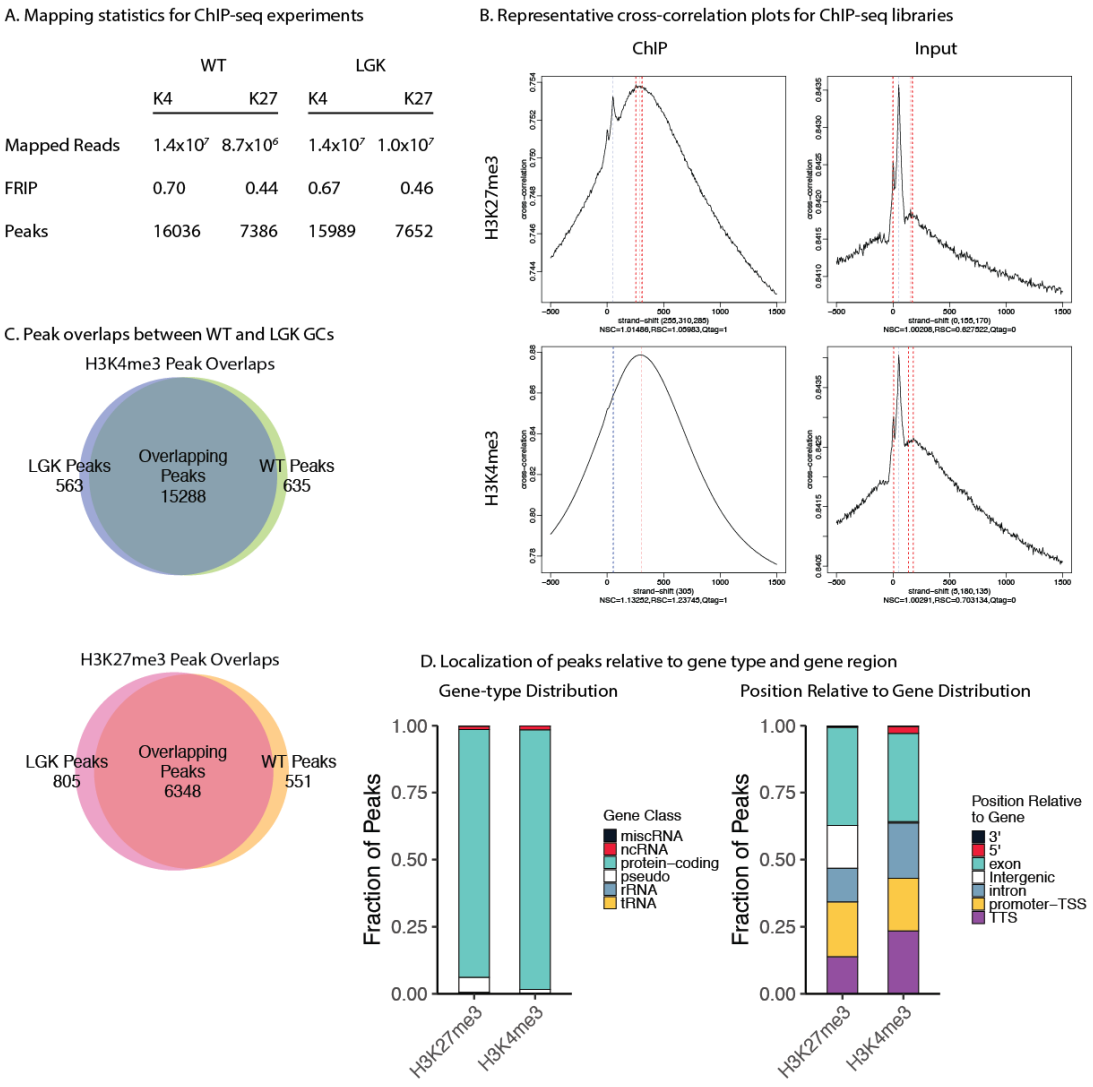


Figure S3: Supporting information on ChIP-seq quality and histone mark distribution

A: Table summarizing quality and mapping statistics, including total mapped reads, fraction of reads in peaks (FRIP), and total numbers of peaks. B: Cross correlation plots of representative datasets comparing ChIP to input control, demonstrating enrichment of specific fragments in the ChIP datasets. Red lines mark the center of the cross-correlation peak. Blue lines indicate the phantom peak that correlates to read length. C: Overlapping genomic intervals for H3K4me3 or H3K27me3 among genotypes. These overlaps were identified in DiffBind, which defines peaks as overlapping if they share at least one base pair. D: Distribution of H3K27me3 and H3K4me3 peaks in WT GCs relative to gene type and gene region. Both H3K27me3 and H3K4me3 are overwhelmingly associated with protein-coding genes. The major distinction between H3K4me3 and H3K27me3 peak position relative to genes is that H3K27me3 peaks are found in intergenic space (white), while H3K4me3 peaks are not. Both gene-type and positions relative to genes were defined by HOMER.

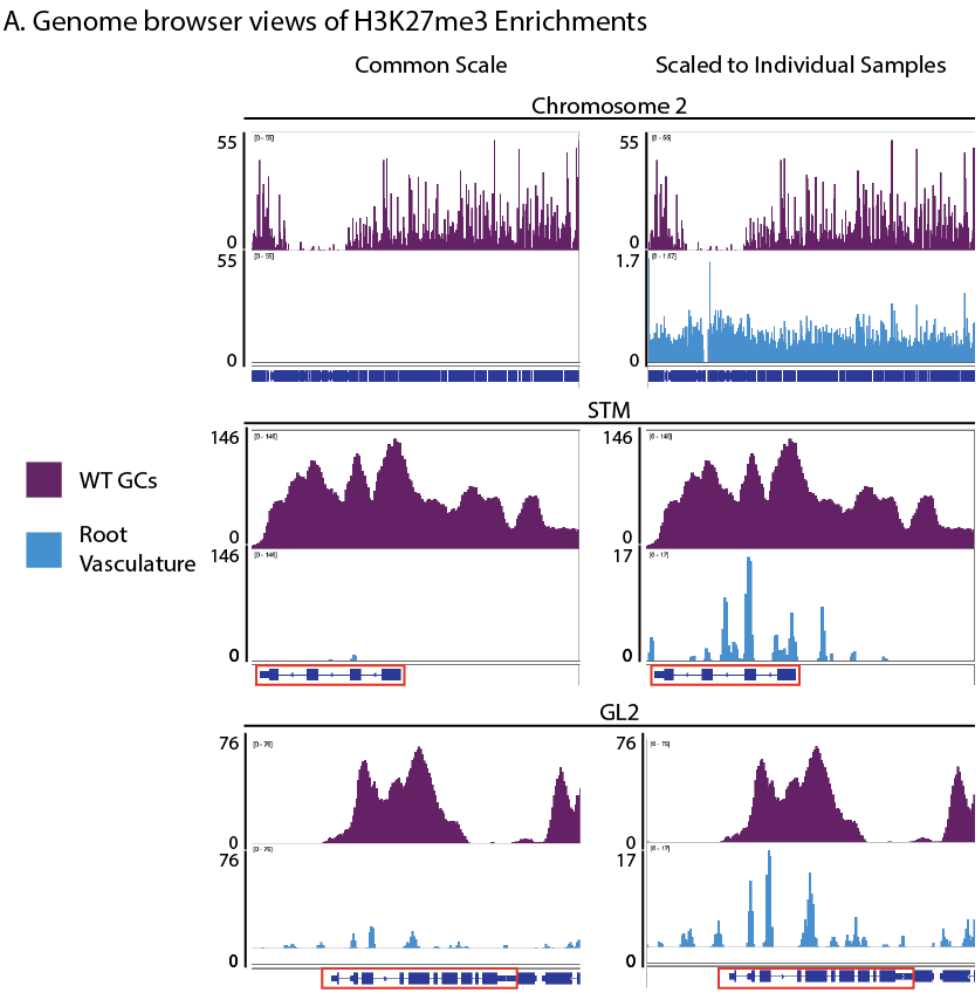
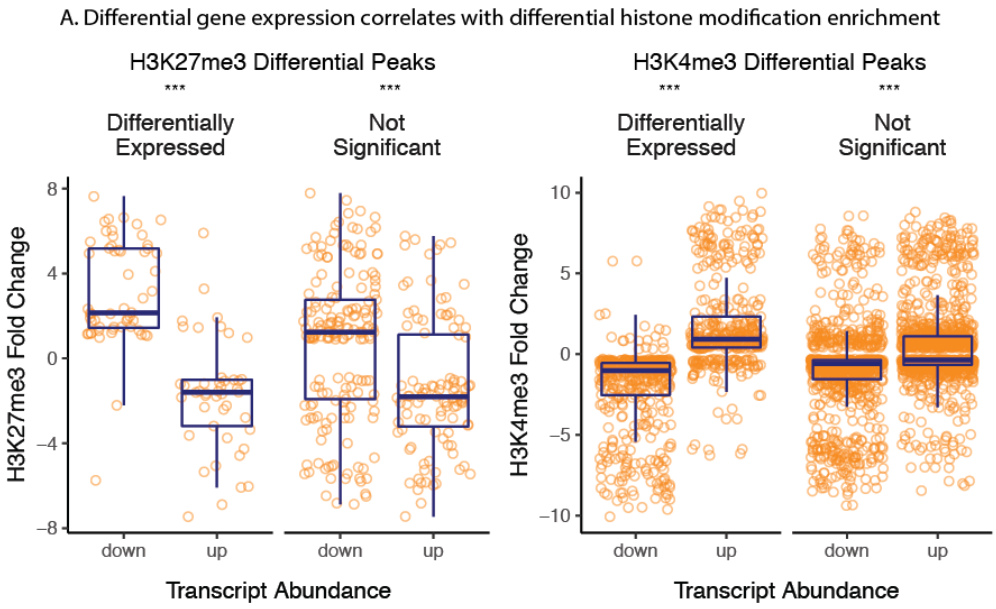


Figure S4: Improved technical performance of ChIP-seq in sorted Arabidopsis GC protoplasts relative to previous single cell type profiles

A: Genome browser views of H3K27me3 data generated for GC (this study, purple) or vascular cells (previous study (10), blue). H3K27me3 is nearly undetectable genome-wide and at high confidence targets in previous experiments. Target genes are denoted with a red box. B: Pairwise overlaps of peaks for (1) aerial tissue and WT GCs, (2) root protoplasts and WT GCs, and (3) root protoplasts and aerial tissue. As shown here, the previously published root protoplast dataset

recapitulates only 11% of peaks from a bulk tissue dataset. Our GC protoplast dataset recapitulates 80% of the peaks called from a bulk tissue dataset. The data used for comparison are from the sole published ChIP-seq histone profile generated from a single Arabidopsis vegetative cell type (10) and have significantly lower resolution than ours. Profiles of H3K27me3 from root hair and non-hair cells was done by ChIP-chip with a NimbleGen array (11), which is not recommended for comparison to ChIP-seq derived data. Therefore we consider our dataset to represent a technical advance for Arabidopsis research.



B. H3K27me3 versus differential expression summary

		Differential Expression	
Gene Expression	H3K27me3 Levels	Significant	Not Significant
Down	Increase	54	116
Down	Decrease	2	61
Up	Increase	9	28
Up	Decrease	36	76
Total		101	281
Chi square p-value		3.26E-15	4.04E-10

C. H3K4me3 versus differential expression summary

		Differential Expression	
Gene Expression	H3K4me3 Levels	Significant	Not Significant
Down	Increase	19	229
Down	Decrease	463	788
Up	Increase	334	584
Up	Decrease	91	630
Total		907	2231
Chi square p-value		3.83E-117	6.78E-36

Figure S5: Differential gene expression is predictive of changing H3K27me3 and H3K4me3 levels.

A: Differentially expressed genes have more divergent histone modification enrichments as compared to genes that are not differentially expressed. However, even among genes where expression is changing but does not pass a significance test, gene expression changes are predictive of histone modification enrichment changes. Asterisks represent distributions that are significantly different by Welch's t-test ($p<E-7$). B: Counts of genes that have differential gene expression and differential H3K27me3. This association between Gene expression change direction and H3K27me3 change direction is significant by Chi square test. This is true for genes that were statistically significantly expressed or not. C: Counts of genes that have differential gene expression and differential H3K4me3. This association between Gene expression change direction and H3K4me3 change direction is significant by Chi square test. This is true for genes that were statistically significantly expressed or not.

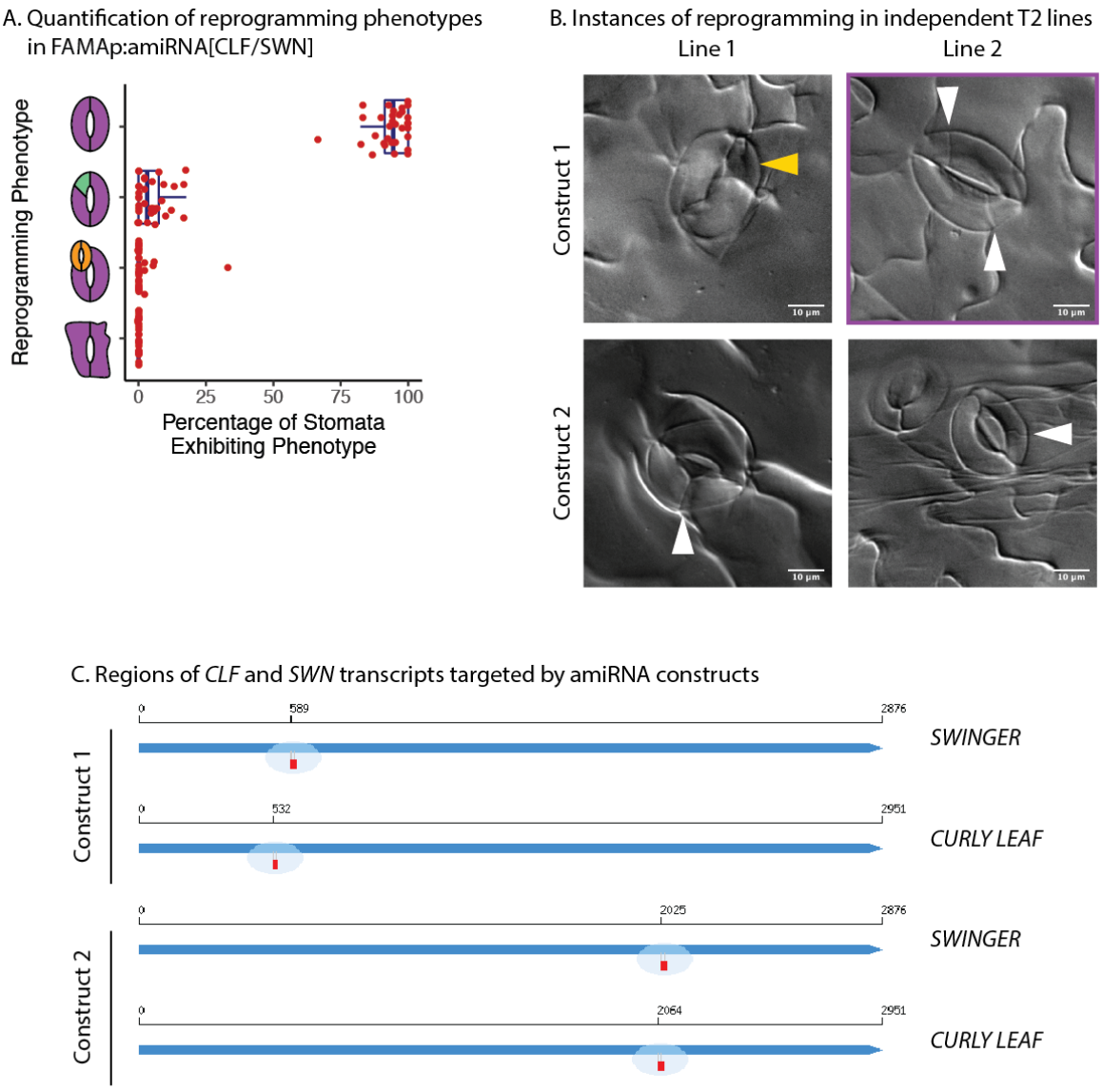


Figure S6: Quantification of reprogramming caused by FAMAp:amiRNA(CLF/SWN)

A: Quantification of reprogramming in FAMAp:ami[CLF/SWN] at 9 dpg using fixed, cleared tissue. Phenotypic categories depicted on the y-axis are the same as used in (12). Red dots represent each individual image scored (6 T2 seedlings, 33 images, 1065 total stomatal complexes counted) for quantification. Position of red dots on the x axis is determined by the number of complexes in an image with the indicated phenotype. Here, reprogrammed complexes account for 6% of total complexes, while the same quantification in FAMA^{LGK} showed 65% of complexes were reprogrammed. B: Images showing GC reprogramming caused by two different artificial microRNA constructs targeting different regions of *CLF* and *SWN* transcripts with two independent T2 lines per construct. DIC images were taken of cleared cotyledon tissue from 12 dpg seedlings. Construct 1 line 2 (boxed in purple) was used for quantification in (A). C: Visualizations from WMD3 showing where each artificial microRNA construct targets the *CLF* and *SWN* transcripts.

A. WIND3 Expression levels in stomatal lineage cell types

	ATH1 Microarray		
	SPCH	MUTE	**WIND3**
Pan-Epidermal	2.3766	1.5436	**2.4008**
SPCH-stage	9.1470	1.6883	**2.5185**
MUTE-stage	4.0043	2.4100	**2.4009**
FAMA-stage	2.2968	1.5045	**2.4579**
	RNA-seq		
	SPCH	MUTE	**WIND3**
Pan-Epidermal	6.1397	4.5431	**4.5990**
SPCH-stage	5.4616	4.8011	**5.1362**
MUTE-stage	6.3381	8.7550	**4.1046**
FAMA-stage	7.4696	4.1874	**4.4685**

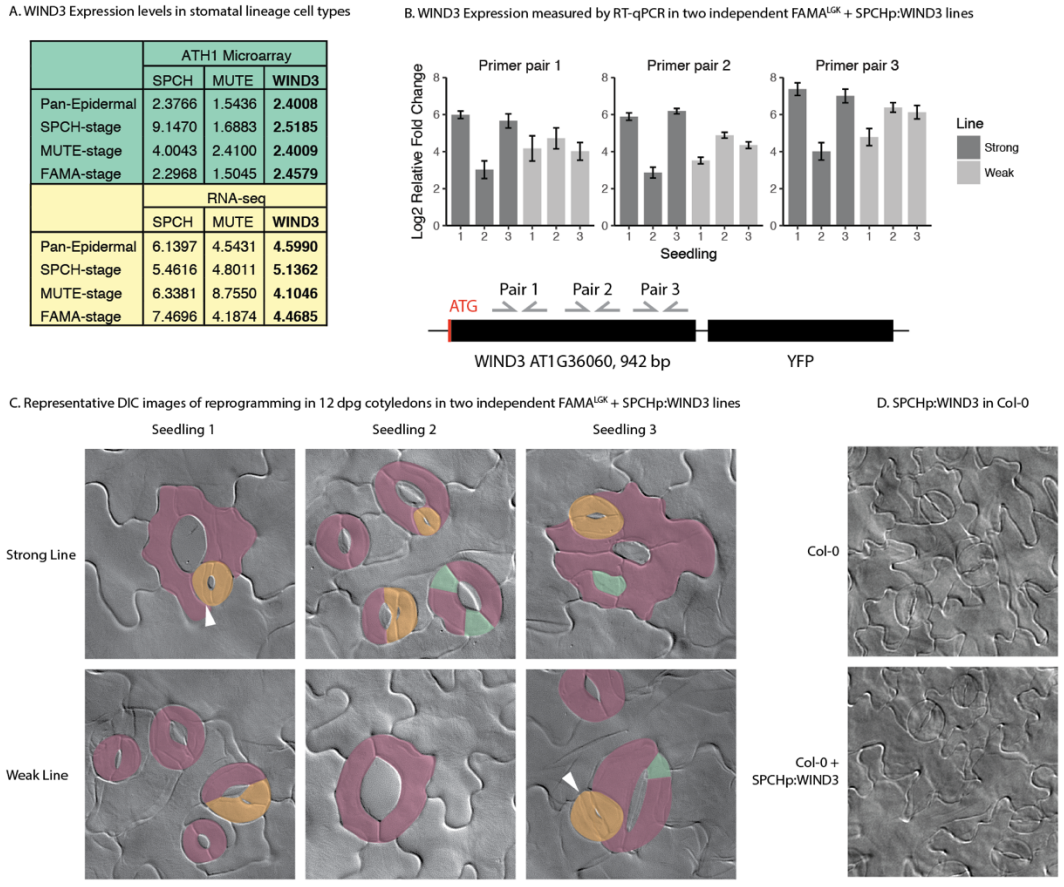
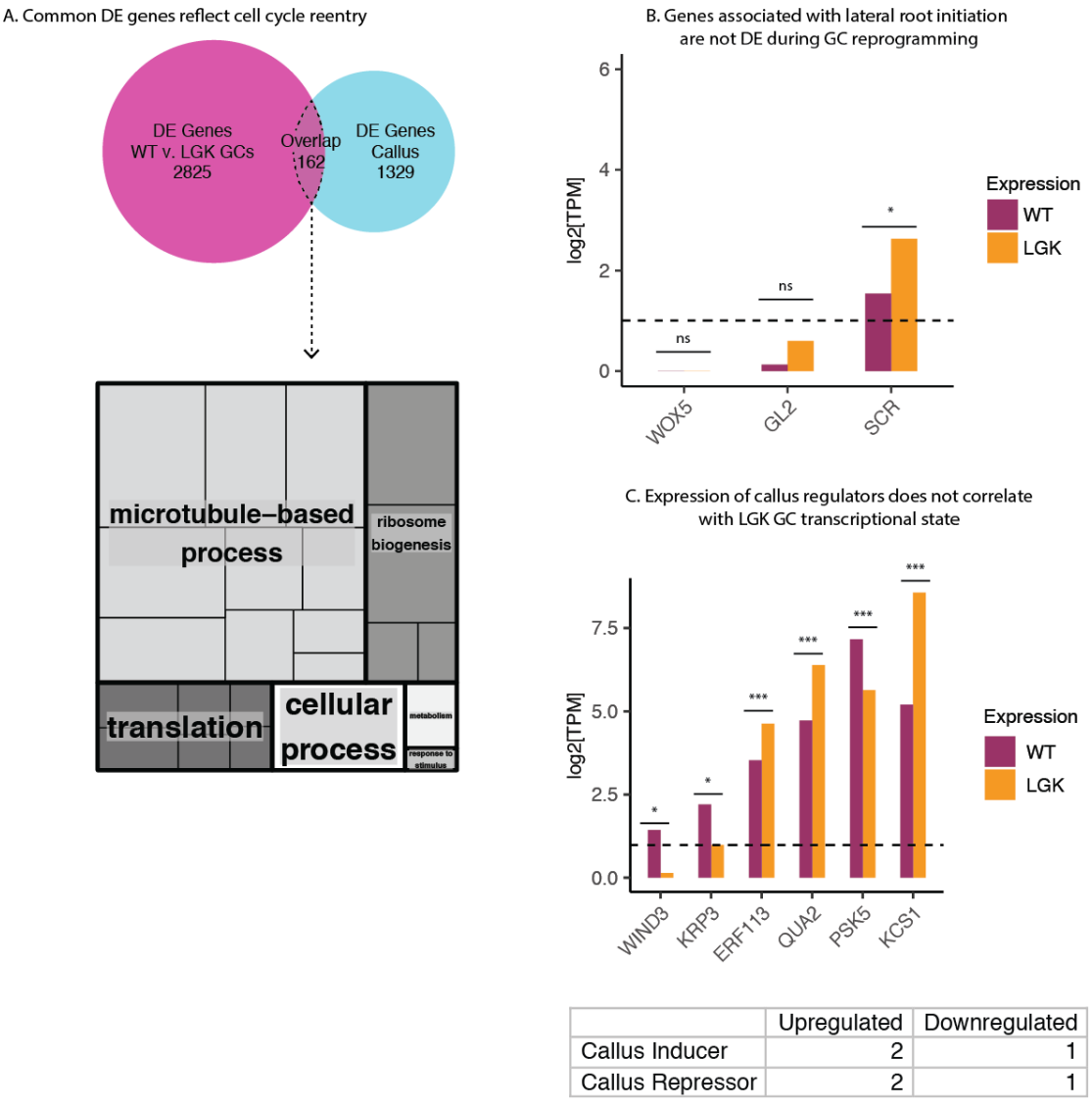


Figure S7: *WIND3* is not differentially expressed during normal stomatal lineage progression, but experimentally elevated levels positively correlate with reprogramming enhancement

A: Expression levels of *FAMA*, *SPCH*, and *WIND3* in stomatal lineage cell types from on RNA-seq and microarray platforms (5). B: WIND3 expression measured by RT-qPCR in three individual seedlings from two independent lines of FAMA^{LGK}+SPCHp:WIND3 at 12 dpg. Three technical replicates per seedling were used. Fold change values were calculated as 2^-ΔΔCq, where the parental line was the reference sample, PP2A was the reference gene, and Cq values are the mean of three technical replicates. Standard deviation of the ΔCq was calculated from the standard deviation among technical replicates for the target and reference genes. The error bars represent the range of 2^-ΔΔCq values that are obtained from ΔΔCq ± standard deviation. The position of each primer pair along the gene is illustrated below. C: DIC images taken from the same seedlings used in 7A. SIS are false-colored in orange. Meristemoids in stomata are false-colored in green. Original stomatal complexes are false-colored in purple. White arrows indicated divisions within SIS. D: Representative images of the epidermis of Col-0 and SPCHp:WIND3 in the Col-0 background at 7 dpg. Eleven seedlings from each of two independent transgenic lines were examined and no phenotypic consequence of the SPCHp:WIND3 transgene in this background was observed. Over 1000 stomatal complexes were observed.



	Upregulated	Downregulated
Callus Inducer	2	1
Callus Repressor	2	1

Figure S8: GC reprogramming is not transcriptionally similar to callus induction from shoot tissue

A: Genes that are commonly differentially regulated between GC reprogramming and callus induction do not co-occur more frequently than expected by chance. Genes in the overlapping set are enriched for GO terms that reflect cell cycle activation. B: LGK GCs do not exhibit the lateral root gene expression profile previously postulated to be an intermediate step in reprogramming. WOX5, GL2, and SCR were used as markers of lateral root identity in callus. Among these genes, only SCR is weakly upregulated in LGK GCs p=0.012. C: Above, only 6 of 107 known callus regulators are DE during GC reprogramming. Of these, only *WIND3* and *KRP3* are not differentially regulated during stomatal development. Below, a contingency table summarizes the expression behavior of these 6 regulators, including whether they induce or repress callus. Callus inducers are just as likely as repressors to be up or downregulated in LGK GCs relative to WT. A

triple asterisk denotes an adjusted p-value of less than 0.05, while a single asterisk denotes an adjusted p-value less than 0.1.

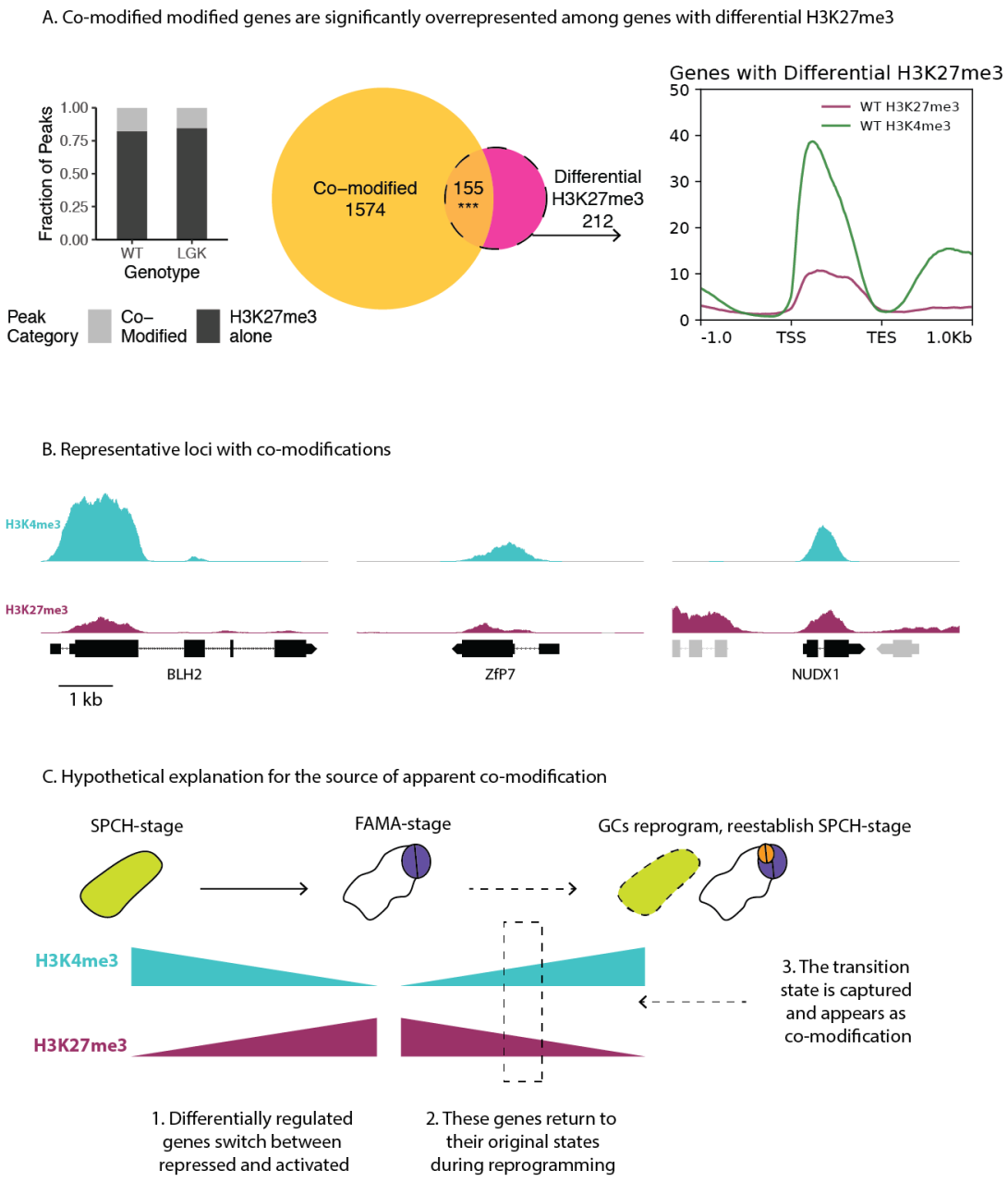


Figure S9: Genes marked by both H3K4me3 and H3K27me3 are overrepresented among genes with differential H3K27me3 between WT and LGK GCs

A: Co-modified genes are equally represented in WT and LGK GCs, but are statistically overrepresented among genes with differential H3K27me3 and these two marks can be shown to co-localize on average at differential H3K27me3 genes. B: Representative loci showing this overlap on a per locus level. The Y axis is uniformly scaled in all images. C: Model for why apparent co-modification is captured in our datasets. We hypothesize genes with differential H3K27me3 between WT and LGK GCs are essentially PRC2-target genes that are differentially regulated between the early and late stomatal lineage.

Box 1: Using the GC H3K27me3 and H3K27me4 datasets for genomic explorations

Bivalent modifications – where H3K4me3 and H3K27me3 are found on the same histone tail – are prevalent in mammalian embryonic stem cells (ESCs) (13) and is associated with a "poised" state enabling a rapid change in gene expression, as well as possibly enabling pluripotency. To identify bivalently modified histones or nucleosomes where the combination of modifications on both H3 tails includes H3K4me3 and H3K27me3 with high confidence, it is necessary to perform re-ChIP experiments, where mono-nucleosomes are immunoprecipitated with first one antibody and then another. Bivalency is understudied in *Arabidopsis*; few bivalently modified nucleosomes have been identified by re-ChIP qPCR in chromatin isolated from complex tissues (14). This is a difficult technique and some studies have attempted to circumvent the necessity of re-ChIP, for instance by performing experiments identifying overlapping H3K27me3 and H3K4me3 domains by ChIP-seq in tissue that contains a limited number of cell types (15). Our single-cell-type data improves upon this resolution and can be used to provide context for earlier experiments identifying overlapping H3K27me3 and H3K4me3 domains. We present this analysis here as an example of how our dataset can be used to address questions about chromatin dynamics in *Arabidopsis* beyond the context of cell fate commitment and stomatal development.

We examined WT and LGK GCs for co-modified regions and found roughly 20% of H3K27me3 peaks overlap with H3K4me3 peaks in each sample (Supplemental Figure 9A-B). This indicates that both differentiated WT GCs and the reprogramming LGK GCs had fewer co-modified loci than even differentiated mammalian cells like mouse embryonic fibroblasts or macrophages, where approximately 50% of H3K27me3 peaks overlap with H3K4me3 according to a recent analysis (16). We did, however, find that co-modified genes are statistically overrepresented among genes where H3K27me3 is differential between WT and LGK GCs (Supplemental Figure 9A). We recognize that heterogeneity of individual loci, even within a single-cell-type population, may cause loci to appear co-modified when they are not. However, if loci were actually marked by H3K4me3 in some cells and H3K27me3 in others we would expect to over-estimate the degree of co-modification, and our studies of both terminal and reprogramming GCs, revealed relatively few potentially co-modified domains in either situation. Thus overall, in contrast to many animal studies, our data indicate co-modification does not correlate with re-established developmental potential in *Arabidopsis* cells. Among genes whose H3K27me3 enrichment changes between WT and LGK GCs, co-modified genes are statistically overrepresented. These loci, therefore, correspond to places where H3K27me3 was recently established or removed in WT GCs (Supplemental Figure 9C). In this context, rather than a mechanism to prime for such a change, as has been proposed for mammalian ESCs, co-modification may be the result of changing transcriptional state (13).

Table S1. Primers used in this study

Purpose	Gene name	AGI code or genome location	Primer name	Sequence (5' to 3')	Amplicon size (bp)	From study (if applicable)
Constructs	WIND3	AT1G36060	WIND3 Cloning F	CACCATGGCGGATCTCTTCGGTG		This study
			WIND3 Cloning R	CGATAAAATTGAAGCCCAATCTA TCTCATA		
	amiCLF/ SWN		A	CACCCCCAAACACACGCTCG		
			B	GCCGCTCTAGAACTAGTGGATCC		
			I	GATCTATCGCTGCATATTAGCGC TCTCTCTTTTGTATTCC		
			II	GAGCGCTAATATGCAGCGATAGA TCAAAGAGAATCAATGA		
			III	GAGCACTAATATGCACCGATAGT TCACAGGTCGTGATATG		
			IV	GAACTATCGGTGCATATTAGTGC TCTACATATATATTCCT		
	YFP-YFP		YFP-F	CACCATGGTGAGCAAGGGCGAG		
			YFP-R	AGCCTTGTACAGCTCGTCCAT		
qPCR	PP2A	AT1G13320	PP2A qPCR F	CAAGTGAACCAGGTTATTGGGA	101	Borghi et al. Plant Cell, 2010.
			PP2A qPCR R	ATAGCCAGACGTACTCTCCAG		
	WIND3	AT1G36060	WIND3 qPCR F1	TCCACAACCTCTTACCCGGA	100	This study
			WIND3 qPCR R1	GGAACTGAGATTGGTAGTTTTCG GG		
			WIND3 qPCR F2	TCGCGGAGATCCGTTTACCC	98	
			WIND3 qPCR R2	AAACGCGGCGCGATCATAAG		
			WIND3 qPCR F3	GGACGTCGAGGAAACGAGCT	103	
			WIND3 qPCR R3	GGTGAACCCGACCCATCACT		

Dataset S1 (separate file). This dataset is an excel spreadsheet that includes the following information for all annotated genes in the *Arabidopsis thaliana* genome: Gene ID, Gene NAME, Average WT GC Expression [TPM], Average LGK GC Expression [TPM], log2[fold change] (LGK/WT), H3K27me3 Peak WT, H3K4me3 Peak WT, H3K27me3 Peak LGK, H3K4me3 Peak LGK, H3K27me3 Peak in Aerial Tissue, H3K27me3 Fold Change WT GC v LGK GC, H3K4me3 Fold Change WT GC v LGK GC, H3K27me3 Fold Change WT GC v Aerial Tissue, Protoplasting Effect, WIND3 Motif.

Dataset S2 (separate file). This dataset is an excel spreadsheet with three tabs. Each tab contains enriched Gene Ontology terms. The first tab (WT K27) includes terms that are enriched among genes that have H3K27me3 peaks in WT GCs. The second tab (WT v LGK K27) includes terms enriched in the set of genes carrying differential levels of H3K27me3 in WT and LGK GCs. The third tab (WT v AT K27) includes terms enriched in the set of genes carrying differential levels of H3K27me3 in WT GCs and Aerial Tissue.

References

1. Bargmann BOR, Birnbaum KD (2010) Fluorescence Activated Cell Sorting of Plant Protoplasts. *J Vis Exp* (1):2–5.
2. Birnbaum K, et al. (2003) A Gene Expression Map of the Arabidopsis Root. *Science (80-)* 302(5652):1956–1960.
3. Lau OS, Bergmann DC (2015) MOBE-ChIP: a large-scale chromatin immunoprecipitation assay for cell type-specific studies. *Plant J* 84(2):443–450.
4. Efroni I, Ip P-L, Nawy T, Mello A, Birnbaum KD (2015) Quantification of cell identity from single-cell gene expression profiles. *Genome Biol* 16(1):9.
5. Adrian J, et al. (2015) Transcriptome Dynamics of the Stomatal Lineage: Birth, Amplification, and Termination of a Self- Renewing Population. *Dev Cell* 33:107–118.
6. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4):576–89.
7. Stark R, Brown G (2011) *DiffBind: differential binding analysis of ChIP-Seq peak data* Available at: http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf [Accessed July 1, 2019].
8. Yadav RK, Tavakkoli M, Xie M, Girke T, Reddy G V. (2014) A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. *Development* 141(13):2735–2744.
9. Yadav RK, Girke T, Pasala S, Xie M, Reddy G V. (2009) Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. *Proc Natl Acad Sci* 106(12):4941–4946.
10. de Lucas M, et al. (2016) Transcriptional Regulation of Arabidopsis Polycomb Repressive Complex 2 Coordinates Cell-Type Proliferation and Differentiation. *Plant Cell* 28(10):2616–2631.
11. Deal RB, Henikoff S (2010) A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. *Dev Cell* 18(6):1030–1040.
12. Matos JL, et al. (2014) Irreversible fate commitment in the Arabidopsis stomatal lineage requires a FAMA and RETINOBLASTOMA-RELATED module. *Elife* 3:1–15.
13. Bernstein BE, et al. (2006) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* 125(2):315–326.
14. Luo C, et al. (2013) Integrative analysis of chromatin states in Arabidopsis identified potential regulatory mechanisms for natural antisense transcript production. *Plant J* 73(1):77–90.
15. You Y, et al. (2017) Temporal dynamics of gene expression and histone marks at the Arabidopsis shoot meristem during flowering. 8:15120.
16. Liu Y, et al. (2016) Epigenetic profiles signify cell fate plasticity in unipotent spermatogonial stem and progenitor cells. *Nat Commun* 7:11275.