Variety is the spice of life: engineering tools for visualizing transcriptional heterogeneity and its effect on cell differentiation in Arabidopsis

Cassandra Maranas

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Jennifer Nemhauser, Chair

James Carothers

Adam Steinbrenner

Neda Bagheri

Program Authorized to Offer Degree

Molecular Engineering and Sciences

University of Washington

Abstract

Variety is the spice of life: engineering tools for visualizing transcriptional heterogeneity and its effect on cell differentiation in Arabidopsis

Cassandra Maranas

Chair of the Supervisory Committee:
Jennifer Nemhauser
Department of Biology

Two genetically identical cells exposed to the same signals will have differences in gene expression. Despite this variability, multicellular development proceeds remarkably robustly in most cases. How and when variation on the cellular level manifests in multicellular coordination and organogenesis is not well understood. Cell-to-cell variation in gene expression can be highly detrimental and actively buffered out; however, in other contexts, it is crucial and actively amplified. For example, variation must be minimized to build organs with consistent size and shape, yet the initiation of organogenesis requires a subset of cells to take on a new fate, a process that often relies on small differences between cells. Aiming to uncover the manifestation of cell-to-cell variation in development, I built a series of serine integrase-based transcriptional recorders to visualize cell-to-cell variation, and then used these tools to analyze Arabidopsis stomatal development and root initiation. In addition, I helped engineer tissue-specific Integrase Erasers to study essential genes. Finally, I showed that the plant hormone auxin acts as a gene expression driver and coordinator in root initiation. Auxin triggers cell-to-cell expression variation by assigning some cells a root precursor fate, while at the same time ensures that these precursor cells are coordinated for robust root formation. My thesis has provided genetic tools for synthetic biologists to apply across applications, and has also demonstrated how application of these tools can elucidate the role, manifestation, and management of variation in development.

Table of Contents

Acknowledgements

I have been fortunate through my PhD career to have no shortage of supportive mentor figures. First, I would like to thank my advisor, Dr. Jennifer Nemhauser for her constant support and belief in me. I am grateful to have had a PI who is invested in my personal growth and wellbeing in addition to my professional development. During my PhD, I underwent multiple difficult personal and family circumstances and her support through these events was constant and always comforting. I would also like to thank Dr. Sarah Guiziou, who provided excellent mentorship through my first couple years, teaching me all about integrases and tremendously impacting my approach to engineering projects. I also thank Dr. Alexander Leydon for his mentorship, always answering my (and everyone else's) questions and also I thank him for playing a key role in establishing a positive and welcoming lab culture. I would like to thank my committee members: Dr. James Carothers, Dr. Adam Steinbrenner, and Dr. Neda Bagheri for their support and guidance. My thesis is better for the insight and questions raised by my committee members and by everyone else who I had science conversations with or who offered feedback or asked questions about my work.

I would like to thank my other labmates throughout the years for being great to work with and to talk to/commiserate with: Dr. Roman Baez, Dr. Eric Yang, Joey Zempke, Jonah Chu, Wesley George, Janet Solano Sanchez, Benjamin Downing, and Julie Ray. I really have enjoyed talking about science, life, and everything else with all of you. We also have been lucky in the lab to have many amazing undergraduates. I thank my undergraduate mentees, Sarah Scallon, Sydney VanGilder, and Linda Nguyen, for their time and effort contributing to my projects, as well as the great company they provided during the process. I also thank all the other undergraduate

researchers who worked in the Nemhauser lab during my time here for the energy and positivity they brought and the great conversations I got sucked into constantly and inevitably.

To my Nook friends, I am really grateful we managed to stick together through so many years after graduating and no longer living a few rooms away from each other. Having you all as a constant support through this time means more than I can say. To my sports degenerate friends (except Lucas), thank you for always loving sports discourse as much as I do and for letting me relieve stress by ranting about the Kansas City Chiefs. To Nick, a specific thank you for inspiring my thesis title during a heated discussion of NFL playoff seeding. To my Roosevelt roommates, thank you for ensuring I was never missing community during the COVID years. To my fellow grad student friends, thank you for understanding what it's like and commiserating. To Tara and Ankita, thank you for the constant silliness and making me be social and go on trips and whatnot.

To my sister, Tina, thank you for understanding me better than anyone else and always being there for me in the bad moments and the good ones. Thank you to my dad for always pushing me to be my best. Thank you to my stepmother for all the support and for your help during some hard times. Rest in peace to my mother, who died during my PhD studies and who herself was a scientist. I try my best to protect the parts of me that came from you and I know you would be proud of me. And the biggest thanks of all to our cat Scarlet who is the best cat who ever lived and whose cuddles and purrs could melt away anyone's worries.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Building resilience by cultivating difference: A role for noise in development

Cassandra Maranas and Jennifer L. Nemhauser

Abstract

Across all biological life, cells in the same environment, with exposure to the same signals and cues exhibit differences in gene expression patterns. This phenomenon is deemed noise and it has long been a question whether it serves a functional role. In plants, recent advances indicate that noise enables many cell fate decisions and thus triggers organogenesis. Additionally, evidence suggests that noise allows organisms to adapt to dynamic environmental conditions and stressors. Given these recent findings and the increasing pressures of climate change on agriculture, efforts to understand the sources and effects of noise are crucial for future projects in engineering resilient, adaptable crops. In this review, we discuss how plants manage noisy gene expression, in some cases buffering and in some cases amplifying natural transcriptional noise. We also discuss the downstream implications of cell to cell heterogeneity on developmental outcomes and robustness. We describe recent progress in this area and present the possibility of using gene expression variability as an engineering target.

Address

Department of Biology, University of Washington, Seattle, 98195-1800, United States

Corresponding author: Nemhauser, Jennifer L. (jn7@uw.edu)

Current Opinion in Plant Biology 2025, **88**:102809

This review comes from a themed issue on **Epigenetics and gene regulation_2025**

Edited by Dr. **Sara Farrona** and **Dr. Ralf Müller-Xing**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.pbi.2025.102809

Introduction

When trying to model the highest performing crop variety, one must choose which metric to optimize (e.g., maximum potential yield or most consistent yield) as well as what environmental challenges are likely to predominate (e.g., flooding, drought, macro- or micronutrient limitations). For roots, there has been enormous progress in identifying an ideal ideotype ("steep, deep, and cheap" [1]), and in validating its performance in multiple crops exposed to limited water or nutrient conditions [2]. For the most part, phenotypic plasticity in this context is considered maladaptive, where, for example, a burst of lateral root initiation is a drag on yield as it limits the growth of primary/main roots while introducing intraplant competition for the same mobile resources [1]. In well-resourced environments, this approach has proven highly effective, and annual average yields of crops like corn have continued to climb [3].

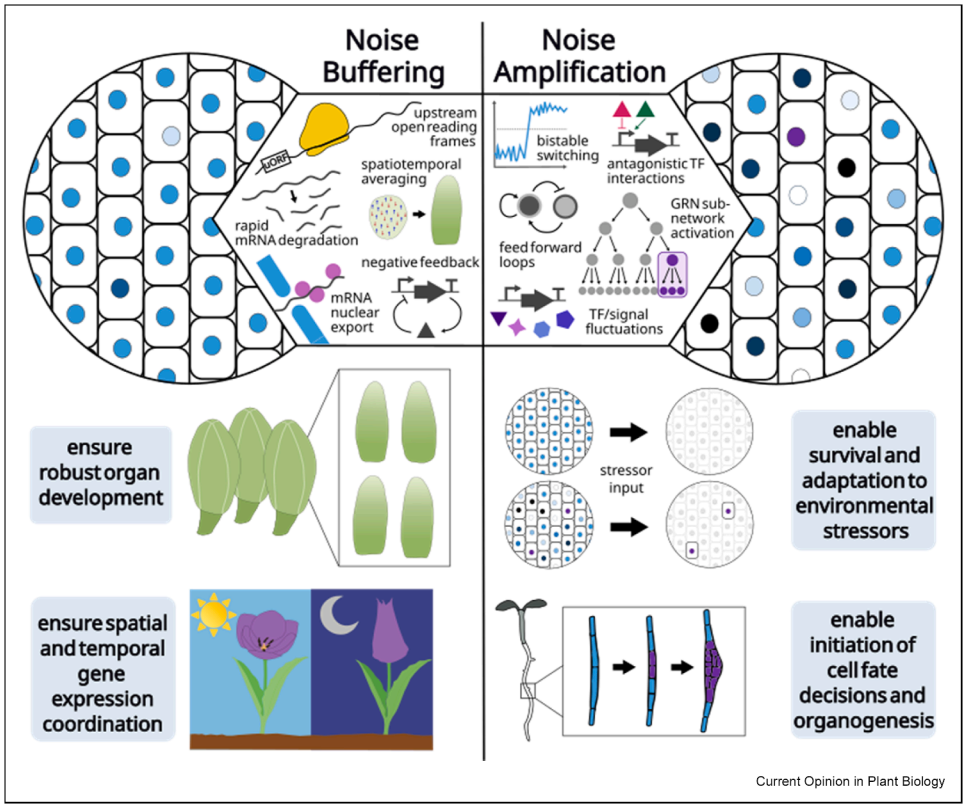
Two recent trends suggest that plant scientists may want to reconsider the quest for the single supreme performer, particularly for less-consistently resourced agricultural settings. First, both biotic and abiotic environmental conditions are becoming less predictable as climate change continues apace [4]. Second, multiple recent studies have highlighted the ways in which heterogeneity or noise across biological scales is critical for maintaining robust developmental programs (Box 1) [5,6]. The loss of noise, in other words, could lead to phenotypes that are highly vulnerable to environmental fluctuations. Taken together, these findings argue that we may need to shift our focus from maximizing yields to maximizing resilience—a shift that could mean embracing diverse ideotypes within the same field, making it possible to hedge our bets.

Main

Marching to the beat of your own transcriptional drum

Transcription levels of genetically identical cells growing in identical environments can differ, sometimes dramatically [7]. The heterogeneity found at individual loci can manifest in overall differences in cell state, which can in turn lead to striking differences in morphology and behavior [8]. This inherent variability is termed "noise", and there is a vigorous debate about whether it is useless or problematic, a residue of incomplete evolutionary optimization, or whether it could actually enhance fitness [9]. Arguments supporting the latter hypothesis have been particularly well-documented in studies of solid tumors, where cell identity is highly heterogeneous, and where this noise allows for resistance to therapeutics [10], as well as aggressive expansion and metastasis [11].

Figure 1



Mechanisms and implications of buffering and amplifying transcriptional noise. Depending on circumstances, noise can be either buffered to reduce cell-to-cell heterogeneity in gene expression, or amplified to increase heterogeneity. Buffering (left) can be achieved by modulating aspects of translational regulation (such as upstream open reading frames), mRNA degradation rates, and mRNA nuclear export rates. It can also be achieved through gene regulation, particularly using negative feedback, and on broader scales, buffering occurs through spatiotemporal averaging. Buffering is key to enable robust organ development and spatial/temporal coordination of gene expression patterns. Noise amplification (right) can also be achieved through genetic regulation, particularly feed forward loops and bistable switching, as well as through activation of genetic subnetworks. Fluctuations in transcription factor and chemical signal concentrations also amplify noise in gene expression, as do antagonistic transcription factor interactions. Amplifying noise is important to enable adaptation to stressors and environmental conditions, and to enable cell fate transitions and trigger organogenesis.

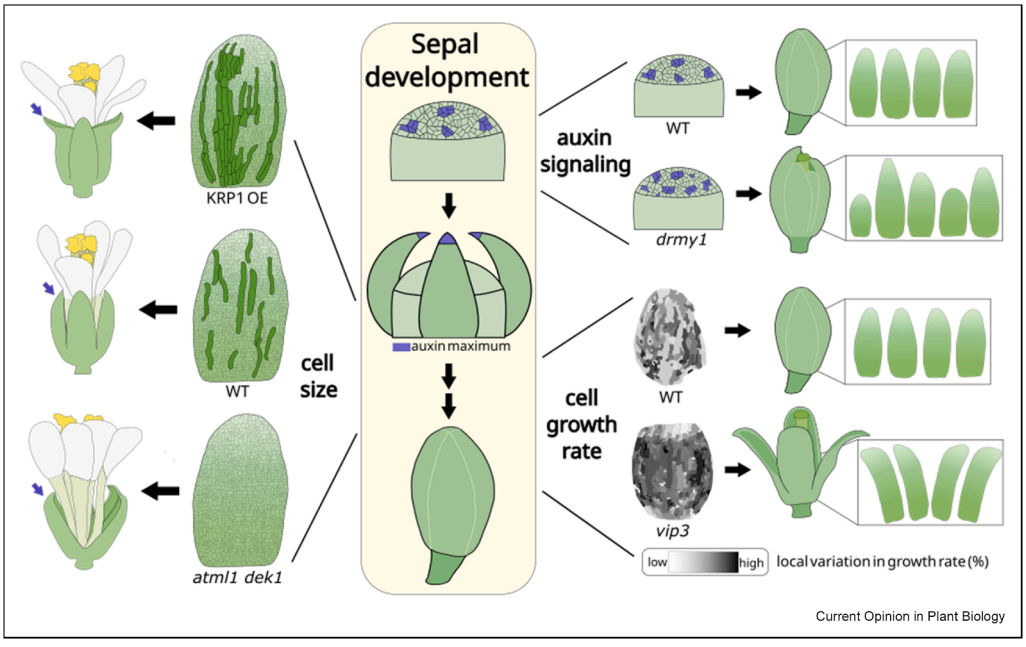
Transcriptional noise has a well-documented role in development, such as in mammalian blastocyst organization [12], mouse pancreatic development [13], and *Drosophila* retinal development [14]. It has been shown that, in some cases, noise is under positive selection, for example in the expression of plasma membrane transporters in yeast [15]. There are also several documented molecular mechanisms for noise amplification, including fluctuations in chemical signals or transcription factors, antagonistic transcription factor interactions, bistable switching between cell states, feed forward loops, and activation of gene regulatory subnetworks [16,17] (Figure 1). In plant development, the role of noise in development is a relatively new topic, but one of much interest. Stochasticity likely plays a critical role in the earliest stages of initiating cell fate determination in plants, and thus has been incorporated as a key parameter in many computational models of stem cell differentiation [18]. An example can be seen in stomatal development, where stochastic expression of the transcription factor SPEECHLESS (SPCH) directs leaf epidermal cells towards a stomatal cell fate. Initially low levels of stochastic *SPCH* expression are amplified by a feed forward loop, thereby generating a bistable switch for stomatal cell fate specification [19].

Recently, researchers have begun to connect epigenetic DNA modifications to plasticity in gene expression and phenotype in the context of development [20], adaptation [21], and even cancer progression [22]. In Arabidopsis, one study [23] has shown that DNA modifications, namely methylation, can affect gene expression noise, reporting the discovery of a new type of gene body methylation (GbM), termed Dynamic GbM. Dynamic GbM genes undergo frequent removal and addition of methyl groups by constitutive expression of

various methylation modifiers. This results in higher cell-to-cell variation in GbM at these genes (in contrast to stably methylated genes), and thus, higher variability and a larger dynamic range in gene expression. Genes associated with adaptive responses, such as those involved in stress response and temperature regulation, were enriched among the Dynamic GbM genes. Additionally, homologs of Dynamic GbM genes in Arabidopsis were also likely to be Dynamic in earlier diverged species, indicating a selective evolutionary pressure for this type of regulation. The authors posit that this behavior is retained because it allows for rapid switching between gene expression states.

While amplifying noise is sometimes advantageous, there are also multiple molecular mechanisms to buffer out stochasticity in gene expression. These mechanisms include modulation of mRNA stability and nuclear export rate, as well as the use of certain gene regulatory architecture, particularly negative feedback loops [17,24] (Figure 1). While chromatin structure is important in maintaining adaptive responses, it is also implicated in noise buffering [25]. On a broader scale, buffering of noise can also be achieved through spatiotemporal averaging, wherein local heterogeneous fluctuations in gene expression are averaged out over larger areas and longer timescales [24] as has been shown to be the case in sepal development in Arabidopsis [26]. Translational regulation can also buffer noise [27]. One recent study [28] connects upstream open reading frames (uORFs) to the maintenance of daily transcriptional oscillations. The presence of a uORF reduced both the level and variation of reporter output in transgenic Arabidopsis lines. A stochastic computational model based on mass action kinetics indicates that the uORF slows translation rate by approximately 6-fold, while the mRNA degradation rate was similar in both cases. Next, the authors focus on the impact of the four uORFs found in the gene *TIMING OF CAB EXPRESSION 1* (*TOC1*), a key regulator of the circadian clock with a pronounced daily oscillation in expression. Loss of *TOC1* uORFs increases average levels of gene expression and increases the variability in circadian period length. The circadian period variability manifests both as a loss of cell-to-cell synchrony in *TOC1* expression within each transgenic line, and as an increased variability in circadian period length between lines.

Figure 2



Sources and effects of noise in sepal development. Sepal development is a commonly used model for studying the effects of noise. (left) Sepals are patterned with giant cells (dark green) and small cells (light green). Overexpression of *KRP1* causes an increased proportion of giant cells and results in an outward curvature of the mature sepal, whereas the *atml1 dek1* knockout mutant has no giant cells and an inward curvature of the mature sepal. This is in contrast to WT sepals, which are relatively flat. (top right) In WT sepal development, four precise auxin maxima give way to four sepals with consistent size and shape. Increased heterogeneity in auxin patterning in the *drmy1* mutant results in flowers with variable sepal number and sepals with variable shape and size. (bottom right) The correct pattern of local variation in cell growth rates is important for proper sepal development. In the *vip3* mutant, there is increased local growth variation compared to the WT, resulting in sepals with an outward curvature, leaving floral organs exposed.

Moving up the (biological) scales

How transcriptional noise affects larger-scale phenotypes is a question of active study by multiple groups, with sepals becoming the model organ of choice in many studies. Sepals are well-suited for studying noise, as they are a relatively simple organ with highly consistent morphology that follows a stereotypical progression through developmental stages [29]. Cell-to-cell heterogeneity in gene expression is necessary for the proper patterning of giant and small cells on the sepal epidermis [30]. In mutants with more homogeneous patterning, where sepals are made up of either mostly giant cells (such as in the KIP RELATED PROTEIN 1 (KRP1) overexpression mutant), or solely small cells (such as in the double knockout mutant of ARABIDOPSIS THALIANA MERISTEM LAYER 1 (ATML1) and DEFECTIVE KERNAL 1 (DEK1)), sepals are curved throughout development (Figure 2). This is in contrast with wild-type sepals which flatten as they mature, showing that heterogeneity is necessary for proper sepal development [31]. In a landmark study in this field, local stochasticity was found to be essential to maintain robust sepal development [26]. Following a mutant screen looking for variation in sepal size, the researchers focused on one mutant, *variable organ size and shape 1* (*vos1*) that led to overall reduced levels of reactive oxygen species (ROS) and less cell-to-cell variability in growth rates. Live imaging combined with mechanical models revealed that wild-type sepal shape depends on spatiotemporal averaging of variability in cellular growth rates.

Interestingly, too much variability in cellular growth rates can also prevent robust sepal development. Following up on work in *Saccharomyces cerevisiae* [32] that found that the Polymerase-associated Factor 1 Complex (Paf1C) is a driver of both transcriptional elongation and noise, researchers investigated the effect of loss of activity of the Arabidopsis Paf1C subunit *VERNALIZATION INDEPENDENCE 3* (*VIP3*) [33]. *vip3* mutants show more variation in reporter signal between sister guard cells, as well as more variation in local growth rates during sepal development. This led to more variation in mature sepal size and shape, when compared with wild-type plants (Figure 2). The authors attribute this phenotype to "error propagation", wherein lower levels of heterogeneity in sepal initiation are propagated as cells grow and divide, resulting in highly variable mature sepals. Live imaging of sepals confirmed that there were significant differences in growth patterns between the WT and *vip3* mutant. For example, cells at the tip of *vip3* sepals reinitiated growth at the same time when growth had essentially stopped in wild-type sepal tip cells. The researchers proposed that in *vip3*, increased noise during sepal initiation is propagated throughout sepal development. This increase in noise prevents the formation of regional growth domains, which interferes with the establishment of mechanical cues for sepal growth arrest.

Venturing outside of sepal development, gene expression noise is buffered by microRNAs associated with ARGONAUTE 1 (AGO1), with mutant *ago1* Arabidopsis plants showing increased variability in hypocotyl length and asymmetry in rosette structure [34].

We can play it fast or we can play it together

Organogenesis is often triggered by a subset of cells changing cell identity. Such processes, including the initiation of a new root or a flower, proceed in a remarkably robust manner given the stochastic nature of their induction [35]. Another recent study of sepal development [36] drove home the trade-offs between speed, noise, and developmental robustness. Four precisely spaced sepal primordia arise from four precisely spaced auxin maxima (Figure 2). The *development related myb-like 1* (*drmy1)* mutant increases noise in auxin signaling via increasing expression of *CUP-SHAPED COTYLEDON1* (*CUC1*) and the resulting sepals are more variable in number, size, and shape (Figure 2). In wild-type plants, feedback between auxin and *CUC1* establishes precisely positioned boundaries between sepals. Overexpression of *CUC1* reduces sepal robustness, largely phenocopying the *drmy1* mutant. In the *drmy1 cuc1* double mutant, robustness in sepal number and positioning is restored, but sepal initiation timing and size remain more variable than in wild-type flowers. A 2D model of cell growth and division revealed that when cellular growth rate was increased, there was more variation in auxin maxima in both the wild-type and *drmy1* backgrounds. Conversely, reducing the growth rate restores robust auxin patterning in *drmy1*. This relationship was validated experimentally in Arabidopsis buds cultured on varying levels of sucrose, where reduced growth rates were sufficient to restore robustness in auxin patterning and sepal development in *drmy1* mutants. In this way, sepal development toes the line between developmental speed and robustness.

Another compelling example of the speed vs. robustness tradeoff was revealed during an exploration of cis-regulatory control of the orthologous *ANATHA (AN)* and *UNUSUAL FLORAL ORGANS (UFO)* genes in *Solanum lycopersicum* (tomato) and Arabidopsis, respectively [37]. Mutating cis-regulatory elements in the promoter of *AN* causes precocious expression and earlier initiation of flowering, likely reflecting the loss of binding by a repressing transcription factor. Analogous mutations to the promoter of *UFO*, result in flowers with increased variability in petal size, shape, and number. Differences in penetrance and expressivity in mutants in both systems suggest significant roles for developmental speed, genetic dosage, and environmental conditions in ultimately shaping morphological outcomes.

There also appears to be a tradeoff between the number of cells and robustness, at least in certain circumstances. Auxin patterning is variable early in sepal initiation due

to noise, but later develops into robustly positioned maxima [38]. The authors hypothesize that the increase in the number of cells as the meristem grows allows for spatial averaging of noisy expression patterns and therefore robust sepal development. To test this, they treated sepal buds with oryzalin, which inhibits cell division, thus producing meristems with far fewer cells. The oryzalin-treated meristems show higher variability in auxin patterning compared to untreated meristems, supporting the connection between cell number and patterning robustness.

Facing the (sometimes stressful) music

Supporting a context-dependent linkage of variation to an adaptive advantage, a study of trichome patterning found a competitive inhibitory relationship between transcriptional activators and mobile repressors [39]. The authors integrated computational modeling and experimental approaches to demonstrate that stochastic fluctuations in the production, movement, and activity of transcription factors allow for variable patterning of trichome cells that arises from an initially uniform population. Trichomes are known to be important in protecting against stressors such as drought and UV radiation, as well as insect herbivory, and there are high levels of natural variation in trichome patterning [40]. As such, trichome development is likely dynamically tuned to the environment, and the stochastic nature of initiation could be a driver of adaptive capacity.

Another example of stochasticity allowing for adaptive responses is seen in protoplast regeneration. One study of Arabidopsis mesophyll protoplast regeneration [41] performed single nucleus RNA sequencing of leaf cells before and after isolating protoplasts. They found significantly higher variability in gene expression in the protoplasts compared to the intact leaf cells. A majority of the highly variably expressed genes were related to stress response, but regeneration-associated transcription factors such as *WUSCHEL* (*WUS*) and *DORNRÖSCHEN* (*DRN*; also known as *ENHANCER OF SHOOT REGENERATION1* or *ESR1*) also showed highly variable expression in protoplasts and in the resultant regenerated microcalli. The authors then used transposase-accessible chromatin sequencing (ATAC-seq) to connect higher variability in gene expression in protoplasts to increased chromatin accessibility. Here, noisy gene expression is likely essential for producing the rare cells capable of regeneration.

Adaptive capacity also manifests on a plant-to-plant basis. For example, germination time is highly variable across Arabidopsis accessions, likely due to different sensitivities to abscisic acid (ABA) [42]. A stochastic model where ABA and gibberellins (GA) antagonistically regulate a group of transcription factors reveals a positive correlation between ABA and variation in germination time and the opposite effect for GA. Exogenous treatment of seeds with ABA or GA confirms this prediction. In another study, a time course transcriptomic analysis of individual Arabidopsis seedlings found many highly variable genes (HVGs) with divergent expression across multiple time points [43]. They found that many HVGs are involved in stress and/or light response, potentially connecting them to resilience. A similar transcriptomic analysis on individual *Zea mays* (corn) plants grown in the field also identified a large number of HVGs, although field-grown corn showed nearly 2.5X more variation in gene expression levels on average than lab-grown Arabidopsis [44]. Both studies also found that HVGs tend to be shorter in length and are targeted by more transcription factors on average.

Discussion: fighting noise with noise

A growing body of evidence from both experimental data and computational models indicates that noisy gene expression should not be written off as solely an obstacle to biological fitness. The potential connection between noise and adaptive capacity is a particularly attractive prospect for the engineering of resilient crops. Global environmental conditions are rapidly changing, with temperatures increasing and periods of drought lasting longer and longer [45]. However, this overall trend is itself noisy. Longer drought periods are interspersed

Box 1. Glossary of terms

Noise: The inherent cell-to-cell variability in gene expression, even between cells of the same type in the same environment. Noise can be defined as intrinsic or extrinsic, depending on the source [51]. Intrinsic sources of noise are due to the inherent stochasticity of processes involved in gene expression (transcription, translation, mRNA/protein degradation), while extrinsic sources are independent of these processes.

Plasticity: The capacity to enact varied genetic programming given a fixed genotype, often in response to environmental signals. In a stressful environment, plasticity can be realized through altered expression patterns within certain genetic programming or by activating/deactivating entirely separate genetic modules.

Robustness: The ability of a cell or organism to maintain consistent enactment of biological programming, resulting in consistent phenotypes, despite fluctuations or changes in external conditions. Robustness can also describe consistency in programming and phenotypes between separate individuals when exposed to different environmental cues.

Resilience: The ability to survive and thrive in many different environments, including those with exposure to stressors.

Adaptability: The ability of a cell/organism/species to adjust its programming to increase the odds of survival in adverse conditions. In extreme cases, this can be in the form of "bet hedging", where increased cell-to-cell or organism-to-organism variance in gene expression patterns allows a small minority to survive a stressful environment and then rebuild the population.

with increased instances of flooding and overall climate trends are incredibly variable and unpredictable [46].

Many efforts to engineer crops focus on enhancing traits that confer resilience to specific stressors, such as drought and heat. However, these crops will have a reduced capacity for varied responses and, as a result, will likely not be well suited to highly variable environmental conditions [47]. With the development of new technologies to visualize and measure gene expression noise [48,49], the field is now well equipped for an alternative approach, adopting gene expression variability as an engineering target. Perhaps plants with more noisy expression of genes needed to survive a given stress will better manage the trade-offs between defense and growth—a shift that may make for less precision in the ideotypes found in a single field but more consistency in annual yields. Or perhaps stochastic rotation through a suite of defense responses might be more effective than constitutive expression of all responses at once [50]. Similarly to some of the discussed developmental processes, it could be that the best way to fight a noisy climate, and to ensure our collective survival, is to turn up the biological noise (Box 1).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
Jennifer Nemhauser reports financial support was provided by National Institutes of Health. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Our work on auxin's impact on developmental noise is supported by the National Institutes of Health (R35-GM148135) and the Benjamin D. Hall Endowed Chair in Basic Life Sciences. We thank the other members of the Nemhauser laboratory, as well as members of the Steinbrenner, Imaizumi and Di Stilio groups, for productive discussions.

Data availability

No data was used for the research described in the article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Lynch JP: **Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems**. *Ann Bot* 2013, **112**:347–357.

2. Del Bianco M, Kepinski S: **Building a future with root architecture**. *J Exp Bot* 2018, **69**:5319–5323.

3. *United States department of agriculture: Corn Yield, United States*. 2025.

4. *Climate adaptation plan 2024-2027*. United States Department of Agriculture; 2024.

5. Sanchez A, Choubey S, Kondev J: **Regulation of noise in gene expression**. *Annu Rev Biophys* 2013, **42**:469–491.

6. Cortijo S, Locke JCW: **Does gene expression noise play a functional role in plants?** *Trends Plant Sci* 2020, **25**:1041–1051.

7. Elowitz MB, Levine AJ, Siggia ED, Swain PS: **Stochastic gene expression in a single cell**. *Science* 2002, **297**:1183–1186.

8. Nicholson DJ: **Is the cell really a machine?** *J Theor Biol* 2019, **477**:108–126.

9. Roeder AHK, Otegui MS, Dixit R, Anderson CT, Faulkner C, Zhang Y, Harrison MJ, Kirchhelle C, Goshima G, Coate JE, *et al.*: **Fifteen compelling open questions in plant cell biology**. *Plant Cell* 2022, **34**:72–102.

10. Janku F: **Tumor heterogeneity in the clinic: is it a real problem?** *Ther Adv Med Oncol* 2014, **6**:43–51.

11. Kapoor-Narula U, Lenka N: **Cancer stem cells and tumor heterogeneity: deciphering the role in tumor progression and metastasis**. *Cytokine* 2022, **157**, 155968.

12. Holmes WR, Reyes de Mochel NS, Wang Q, Du H, Peng T, Chiang M, Cinquin O, Cho K, Nie Q: **Gene expression noise enhances robust organization of the early mammalian blastocyst**. *PLoS Comput Biol* 2017, **13**, e1005320.

13. Larsen HL, Martín-Coll L, Nielsen AV, Wright CVE, Trusina A, Kim YH, Grapin-Botton A: **Stochastic priming and spatial cues orchestrate heterogeneous clonal contribution to mouse pancreas organogenesis**. *Nat Commun* 2017, **8**:605.

14. Wernet MF, Mazzoni EO, Çelik A, Duncan DM, Duncan I, Desplan C: **Stochastic spineless expression creates the retinal mosaic for colour vision**. *Nature* 2006, **440**:174–180.

15. Zhang Z, Qian W, Zhang J: **Positive selection for elevated gene expression noise in yeast**. *Mol Syst Biol* 2009, **5**:299.

16. Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, Babu MM: **Interplay between gene expression noise and regulatory network architecture**. *Trends Genet* 2012, **28**:221–232.

17. Bhalerao RP: **Getting it right: suppression and leveraging of noise in robust decision-making**. *Quant Plant Biol* 2024, **5**, e10.

18. Stumpf PS, Smith RCG, Lenz M, Schuppert A, Müller F-J, Babtie A, Chan TE, Stumpf MPH, Please CP, Howison SD, *et al.*: **Stem cell differentiation as a non-markov stochastic process**. *Cell Syst* 2017, **5**:268–282. e7.

19. Lau OS, Davies KA, Chang J, Adrian J, Rowe MH, Ballenger CE, Bergmann DC: **Direct roles of SPEECHLESS in the specification of stomatal self-renewing cells**. *Science* 2014, **345**:1605–1609.

20. Nemec S, Kilian KA: **Materials control of the epigenetics underlying cell plasticity**. *Nat Rev Mater* 2020, **6**:69–83.

21. Bogan SN, Yi SV: **Potential role of DNA methylation as a driver of plastic responses to the environment across cells, organisms, and populations**. *Genome Biol Evol* 2024, **16**, evae022.

22. Feinberg AP, Levchenko A: **Epigenetics as a mediator of plasticity in cancer**. *Science* 2023, **379**, eaaw3835.

23. ** Williams CJ, Dai D, Tran KA, Monroe JG, Williams BP: **Dynamic DNA methylation turnover in gene bodies is associated with enhanced gene expression plasticity in plants**. *Genome Biol* 2023, **24**:227.
This publication contributed the discovery of a new scheme of gene body methylation (GbM), deemed Dynamic GbM. Dynamic GbM genes were more variably expressed and were enriched with genes involved in stress response and temperature regulation. Their

Dynamic regulation was conserved evolutionarily, suggesting Dynamic GbM enables adaptive responses by increasing variability in gene expression.

24. Urban EA, Johnston Jr RJ: **Buffering and amplifying transcriptional noise during cell fate specification**. *Front Genet* 2018, **9**:591.

25. Fraser LCR, Dikdan RJ, Dey S, Singh A, Tyagi S: **Reduction in gene expression noise by targeted increase in accessibility at gene loci**. *Proc Natl Acad Sci U S A* 2021, **118**, e2018640118.

26. Hong L, Dumond M, Tsugawa S, Sapala A, Routier-Kierzkowska A-L, Zhou Y, Chen C, Kiss A, Zhu M, Hamant O, *et al.*: **Variable cell growth yields reproducible organ development through spatiotemporal averaging**. *Dev Cell* 2016, **38**: 15–32.

27. Munro V, Kelly V, Messner CB, Kustatscher G: **Cellular control of protein levels: a systems biology perspective**. *Proteomics* 2024, **24**, e2200220.

28. Wu H-W, Fajiculay E, Wu J-F, Yan C-CS, Hsu C-P, Wu S-H: **Noise reduction by upstream open reading frames**. *Nat Plants* 2022, **8**:474–480.

29. Roeder AHK: **Arabidopsis sepals: a model system for the emergent process of morphogenesis**. *Quant Plant Biol* 2021, **2**, e14.

30. Meyer HM, Teles J, Formosa-Jordan P, Refahi Y, San-Bento R, Ingram G, Jönsson H, Locke JCW, Roeder AHK: **Fluctuations of the transcription factor ATML1 generate the pattern of giant cells in the arabidopsis sepal**. *eLife* 2017, **6**.

31. Roeder AHK, Cunha A, Ohno CK, Meyerowitz EM: **Cell cycle regulates cell type in the arabidopsis sepal**. *Development* 2012, **139**:4416–4427.

32. Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, François J, Yvert G: **Cell-to-cell stochastic variation in gene expression is a complex genetic trait**. *PLoS Genet* 2008, **4**, e1000049.

33. ** Trinh D-C, Martin M, Bald L, Maizel A, Trehin C, Hamant O: **Increased gene expression variability hinders the formation of regional mechanical conflicts leading to reduced organ shape robustness**. *Proc Natl Acad Sci U S A* 2023, **120**, e2302441120.

This publication highlights a case where variation in sepal development is required to ensure robust sepal morphology. Reducing the natural cell-to-cell variation in growth rates during sepal development counterintuitively results in sepals with highly variable shape and size. They propose that spatiotemporal averaging of variable cell growth rates is key to robust sepal development.

34. Lemus T, Mason GA, Bubb KL, Alexandre CM, Queitsch C, Cuperus JT: **AGO1 and HSP90 buffer different genetic variants in Arabidopsis thaliana**. *Genetics* 2023, **223**, iyac163.

35. Hong L, Dumond M, Zhu M, Tsugawa S, Li C-B, Boudaoud A, Hamant O, Roeder AHK: **Heterogeneity and robustness in plant morphogenesis: from cells to organs**. *Annu Rev Plant Biol* 2018, **69**:469–495.

36. ** Kong S, Zhu M, Pan D, Lane B, Smith RS, Roeder AHK: **Tradeoff between speed and robustness in primordium initiation mediated by auxin-CUC1 interaction**. *Nat Commun* 2024, **15**:5911.

This publication connects heterogeneity in auxin signaling in the earliest stages of sepal initiation to downstream variability in sepal development and eventual morphology. Using a combined experimental and computational approach, they further connect this initial heterogeneity to changes in cell growth rates, drawing conclusions regarding the connection and tradeoff between developmental speed and robustness

37. * Lanctot A, Hendelman A, Udilovich P, Robitaille GM, Lippman ZB: **Antagonizing cis-regulatory elements of a conserved flowering gene mediate developmental robustness**. *Proc Natl Acad Sci U S A* 2025, **122**, e2421990122.

This publication uses CRISPR/Cas9-based genome editing of the promoters of a regulator of flower development in both tomato and Arabidopsis to connect the presence of specific cis regulatory elements to variation in expression. The role of developmental speed is explored, focusing on the impact on robustness in the morphology of floral organs.

38. * Kong S, Rusnak B, Zhu M, Roeder AHK: **Stochastic gene expression in auxin signaling in the floral meristem of Arabidopsis thaliana**. *Nat Commun* 2025, **16**:4682.

This publication explores how early stage variability in auxin patterning is resolved through the stages of sepal initiation to enable robust sepal development. The relative contributions of extrinsic and intrinsic noise on this observed variability is a topic of focus, as is the connection between number of cells in the meristem and auxin patterning variability. The authors present a model wherein increased cell number allows spatiotemporal averaging of initial stochastic fluctuations to ensure robust sepal development.

39. Digiuni S, Schellmann S, Geier F, Greese B, Pesch M, Wester K, Dartan B, Mach V, Srinivas BP, Timmer J, *et al.*: **A competitive complex formation mechanism underlies trichome patterning on arabidopsis leaves**. *Mol Syst Biol* 2008, **4**:217.

40. Bickford CP: **Ecophysiology of leaf trichomes**. *Funct Plant Biol* 2016, **43**:807–814.

41. Xu M, Du Q, Tian C, Wang Y, Jiao Y: **Stochastic gene expression drives mesophyll protoplast regeneration**. *Sci Adv* 2021, **7**, eabg8466.

42. Abley K, Formosa-Jordan P, Tavares H, Chan EYT, Afsharinafar M, Leyser O, Locke JCW: **An ABA-GA bistable switch can account for natural variation in the variability of arabidopsis seed germination time**. *eLife* 2021, **10**.

43. Cortijo S, Aydin Z, Ahnert S, Locke JC: **Widespread inter-individual gene expression variability in Arabidopsis thaliana**. *Mol Syst Biol* 2019, **15**, e8591.

44. Cruz DF, De Meyer S, Ampe J, Sprenger H, Herman D, Van Hautegem T, De Block J, Inzé D, Nelissen H, Maere S: **Using single-plant-omics in the field to link maize genes to functions and phenotypes**. *Mol Syst Biol* 2020, **16**, e9667.

45. Malhi GS, Kaur M, Kaushik P: **Impact of climate change on agriculture and its mitigation strategies: a review**. *Sustainability* 2021, **13**:1318.

46. Thornton PK, Ericksen PJ, Herrero M, Challinor AJ: **Climate variability and vulnerability to climate change: a review**. *Glob Change Biol* 2014, **20**:3313–3328.

47. Ross Ashby W: **Requisite variety and its implications for the control of complex systems**. *Cybernetica* 1958, **1**:83–99.

48. Tripathi RK, Wilkins O: **Single cell gene regulatory networks in plants: opportunities for enhancing climate change stress resilience**. *Plant Cell Environ* 2021, **44**:2006–2017.

49. Maranas CJ, George W, Scallon SK, VanGilder S, Nemhauser JL, Guiziou S: **A history-dependent integrase recorder of plant gene expression with single-cell resolution**. *Nat Commun* 2024, **15**:9362.

50. Park J, Dies M, Lin Y, Hormoz S, Smith-Unna SE, Quinodoz S, Hernández-Jiménez MJ, Garcia-Ojalvo J, Locke JCW, Elowitz MB: **Molecular time sharing through dynamic pulsing in single cells**. *Cell Syst* 2018, **6**:216–229. e15.

51. Araújo IS, Pietsch JM, Keizer EM, Greese B, Balkunde R, Fleck C, Hülskamp M: **Stochastic gene expression in Arabidopsis thaliana**. *Nat Commun* 2017, **8**:2132.

nature communications

Article

https://doi.org/10.1038/s41467-023-37607-5

An integrase toolbox to record gene-expression during plant development

Received: 25 October 2022

Accepted: 23 March 2023

Published online: 03 April 2023

Check for updates

Sarah Guiziou[1], Cassandra J. Maranas[1,2], Jonah C. Chu[1,2] & Jennifer L. Nemhauser[1]

There are many open questions about the mechanisms that coordinate the dynamic, multicellular behaviors required for organogenesis. Synthetic circuits that can record in vivo signaling networks have been critical in elucidating animal development. Here, we report on the transfer of this technology to plants using orthogonal serine integrases to mediate site-specific and irreversible DNA recombination visualized by switching between fluorescent reporters. When combined with promoters expressed during lateral root initiation, integrases amplify reporter signal and permanently mark all descendants. In addition, we present a suite of methods to tune the threshold for integrase switching, including: RNA/protein degradation tags, a nuclear localization signal, and a split-intein system. These tools improve the robustness of integrase-mediated switching with different promoters and the stability of switching behavior over multiple generations. Although each promoter requires tuning for optimal performance, this integrase toolbox can be used to build history-dependent circuits to decode the order of expression during organogenesis in many contexts.

Biologists have long been fascinated by the molecular pathways that support the development of complex multicellular organisms. Plants are particularly intriguing subjects to study, as the development programs that start in their embryos persist throughout their lifespan, strongly influenced by environmental cues. The growing environmental pressures resulting from climate change make this adaptability increasingly important[1]. A better understanding of the mechanisms that underlie plant developmental plasticity will help guide the engineering of traits that can face current and future challenges[2].

To fully understand the molecular trajectory underlying fate transitions that enable de novo organogenesis and regeneration in plants, we need methods that can sense and relay information in a manner that can be dynamically and quantitatively read out by an observer. Current methods enable precise quantification of DNA[3], RNA[4], and proteins[5] allowing the capture of a snapshot of the molecular state of studied organisms. Combining these approaches with single-cell methods has led to the discovery of new plant cell types and a more detailed view of cell-fate transitions[6–12].

A challenge of current single-cell methods is that they require the destruction of samples, and, therefore, do not allow for real-time readouts, reports from the same sample across multiple timepoints, or preserve spatial relationships. With recent advances in high-throughput and high-precision microscopy, fluorescent reporters and sensors have allowed imaging at cellular resolution of transcription level, protein and molecule concentration and localization in a continuous manner in their native context[13]. However, detection is limited to a reduced amount of information at a time due to the limited number of fluorescent tags and to short timescales due to photobleaching and stress to the organisms. Recently, the development of synthetic, DNA-based recording systems has overcome some of the technical challenges of 'omic and microscopy techniques, allowing the sensing and relaying of multiple signals simultaneously during animal development (reviewed in[2]).

Serine integrases, used by bacteriophages to mediate their own integration into the bacterial genome, were critical to the success of one of the most promising synthetic recorders[14]. In a synthetic system,

[1]Department of Biology, University of Washington, Seattle, WA 98195, USA. [2]These authors contributed equally: Cassandra J. Maranas, Jonah C. Chu. e-mail: jn7@uw.edu

serine integrases are used to invert or excise DNA in a site-specific and irreversible manner, referred to here on as an integrase switch. The integrase recognizes two DNA sites of around 40 bp known as attB and attP sites. If the sites are in the same orientation, the DNA region between them is excised and if the sites are in the opposite orientation, the region is inverted. Gene-regulatory parts, such as promoters or terminators, can be placed between integrase sites to mediate a specific gene expression pattern dependent on integrase expressions. Complex genetic circuits have been developed using serine integrases, implementing Boolean logic (in bacteria[15,16], mammalian cells[17], and plant protoplasts[18]), history-dependent logic (in bacteria[19,20]), and cell-lineage tracing (in animals[14]). Serine integrases can also be used to induce the expression of toxic genes at a specific time, to mediate site-specific DNA integration[21]. To date, serine integrases have not been used extensively in plant systems, although they have been shown to work in principle in *Arabidopsis*[22], *Nicotiana benthamiana*[21,23], barley[24], and wheat[25]. One study in *N. benthamiana* used a recombination directionality factor (RDF), which when combined with the integrase, allowed reversing of the integrase reaction[23].

Tyrosine integrases have been more extensively used in Arabidopsis and other plants. Cre recombinase, for example, has been used to perform cell-lineage tracing through DNA excision using a single pair of identical integrase sites[26,27]. Cre has also been used with integrase-site mutants to generate stochasticity in the output[28]. Recently logic circuits have been engineered using DNA excision mediated by a combination of FLP, Cre, and B3RT integrases[29]. Many tyrosine integrases, including FLP and Cre, act on two identical sites, and so, in principle, can catalyze both a forward and reverse recombination reaction[30]. Serine integrases, in contrast, use two distinct sites, leading to a directional and irreversible recombination event. Another limitation on the use of tyrosine integrases in plants has been low efficiency and silencing, which has recently been connected to CHH methylation of tyrosine recombinase sites[31].

Here, we develop a toolbox of well-characterized parts to build synthetic circuits in *Arabidopsis* using PhiC31 and Bxb1 serine integrases. We express integrases from promoters of well-characterized transcription factors essential for lateral root development as a test-case for building synthetic recorders. To optimize the specificity and robustness of our tools when using different promoters, we build and test a variety of methods to tune the threshold for integrase switching-tools that could be used to tune the activity of any protein of interest. Finally, we characterize two methods that allow for further fine-tuning of the timing and level of integrase activity: split-intein integrase and estradiol-inducible integrase. Collectively, these modular parts make it possible to record gene expression at specific times and spaces during plant development, as well as contribute to an accelerated design-test-build-learn cycle for other plant synthetic biology devices.

Results

Orthogonal and efficient DNA switches in *Arabidopsis*

Our first goal was to test the efficiency of three serine integrases (PhiC31, Bxb1, and Tp901) in *Arabidopsis* transgenic lines. To do this, we needed two constructs: the target and the integrase. These two constructs cannot be on the same plasmid, as even low levels of integrase produced in bacteria during cloning will enable target site inversion. For our target construct, we used a constitutive promoter (pUBQ10[32];) flanked by integrase sites positioned between two reporter genes: the mScarlet and mTurquoise2 (mTurq) fluorescent proteins (Fig. 1a). To more closely match the expression level of most developmentally-relevant genes, we opted to use the promoter of PROTEIN PHOSPHATASE 2 A SUBUNIT 3 (pPP2AA3), a gene which has been widely used as a qPCR control due to its constitutive nature and medium expression level[33]. If expressed and functional, the integrase should mediate the inversion of the promoter resulting in the switch of expression from mTurq to mScarlet. Targets containing either PhiC31 or Bxb1 integrase sites strongly expressed mTurq and not mScarlet in roots and leaves (Fig. 1b). When either the PhiC31 or Bxb1 integrase was constitutively expressed alongside the target with its cognate integrase sites, we observed exactly the opposite reporter expression (strong mScarlet and no mTurq) in all tissues. Moreover, we confirmed that PhiC31 and Bxb1 integrases are orthogonal to one other, as Bxb1 integrase does not mediate an integrase switch in the PhiC31 target line nor vice versa (Fig. 1c). The Tp901 integrase, known to be less efficient than Bxb1 and PhiC31[34], did not cause any switching in targets carrying its target sequences, even with strong promoters (p35S and pUBQ10), and codon optimization (Supplementary Fig. 1). We also tested a switch using YFP and Luciferase reporters that had been used previously in *N. benthamiana*[23], and confirmed that the pPP2AA3 promoter allows constitutive integrase switch with this target as well (Supplementary Fig. 2).

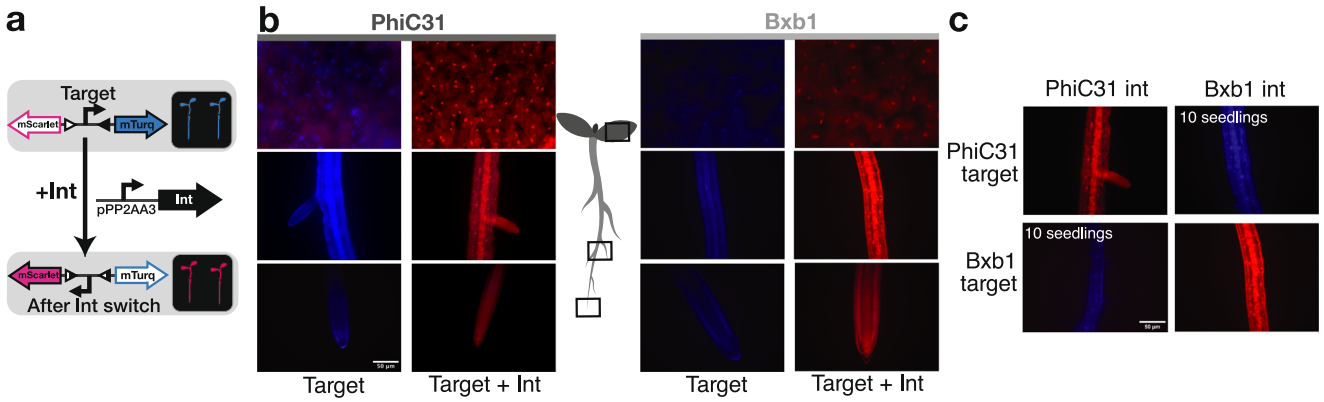


Fig. 1 | Integrase mediates orthogonal DNA-switch in *Arabidopsis*. a Design of the integrase target. The target is composed of two integrase sites (triangles) surrounding a constitutive promoter (pUBQ10) and two fluorescent reporters (mTurquoise2 and mScarlet). In absence of integrase, mTurquoise2 is expressed. In presence of integrase, the integrase mediates inversion of the DNA between the integrase sites, inverting the promoter, and leading to mScarlet expression. The expression of the integrase is mediated by the constitutive promoter pPP2AA3. **b** Constitutive integrase switch characterization. On the left side, *Arabidopsis* seedlings with PhiC31 target alone and PhiC31 target with pPP2AA3:PhiC31 construct. On the right side, Bxb1 target alone, and Bxb1 target with pPP2AA3:Bxb1 construct. Microscopy images are an overlay of mTurq (in blue) and mScarlet (in red) fluorescence, from top to bottom are representative images of the leaf, a lateral root, and the root tip ($n = 20$ seedlings screened). The scale bar in the left bottom image applies to all images. **c** Orthogonality test of the integrase switch. Each integrase target line was transformed with both integrases. Microscopy images are overlays of mTurq (in blue) and mScarlet (in red) fluorescence, and are representative images ($n = 10$ seedlings screened). The scale bar in the left bottom image equals 50 µM and applies to all images.

https://doi.org/10.1038/s41467-023-37607-5

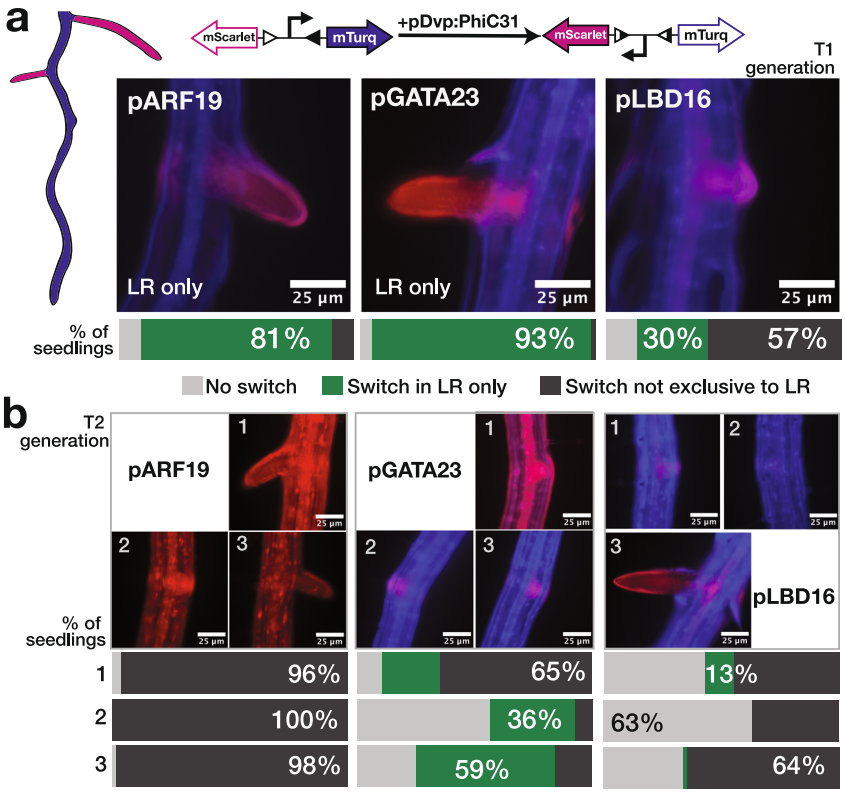


Fig. 2 | PhiC31 integrase switch under the control of developmental promoters. **a** Developmental (Dvp) promoters drove the expression of PhiC31 integrase with a target that switches from mTurquoise2 to mScarlet when the integrase is active. Target lines were transformed with the integrase constructs, and at least 20 T1 seedlings per integrase constructs were characterized. Representative images of emerged lateral roots are shown for each promoter-integrase construct, as well as a bar representing the percentage of seedlings in each phenotypic category: no switch (light gray), switch in LR only (green), or switch not exclusive to LR (dark gray). **b** Characterization of T2 seedlings. For each construct, we selected 3 T1 lines with an LR-only switch phenotype, and characterized 20 T2 seedlings per T1 line. For each T1 line, a representative T2 seedling is shown above bar graphs displaying the percentage of seedlings in each phenotypic category. The percentage of the phenotype represented by the T2 image is displayed numerically on the relevant portion of the graph. Source data are provided as a Source Data file.

To record developmental events, switching of the integrase targets must be consistently restricted to a narrow range of time and space. For the next round of design and testing, we focused on the PhiC31 integrase and lateral root development, a well-characterized example of de novo organogenesis[35]. Lateral root development is a good model for applying synthetic tools to study gene expression because it is a well-studied pathway with defined transcriptional control points[36]. The density and placement of lateral roots are also features of plant architecture that are linked to climate resilience and therefore a strong candidate for synthetic engineering[37]. Lateral roots initiate from a small population of founder cells at the xylem pole of the pericycle layer[38], and follow a fairly stereotyped pattern through morphogenesis[39].

As test drivers for integrase expression, we selected the promoters of several well-studied transcription factors expressed in the early stages of lateral root initiation: AUXIN RESPONSE FACTOR 7 (ARF7)[40], AUXIN RESPONSE FACTOR 19 (ARF19)[40], LATERAL ORGAN BOUNDARIES DOMAIN 16 (LBD16)[41], and GATA TRANSCRIPTION FACTOR 23 (GATA23)[42]. Because the integrase switch is heritable, we would expect that if the integrase is expressed in lateral root founder cells and works efficiently, all cells in the new root should be in the switched state as well. Simply put, all of the cells in the main root should express mTurq, while all of the cells of the lateral roots should express mScarlet. We characterized approximately 20 independent transformants (T1s) per integrase construct, and categorized each seedling by the following categories: (1) No-switch: expression of mTurq only; (2) LR-only: expression of mScarlet in lateral root only; or (3) Non-exclusive: any expression of mScarlet in the main root. For pARF19 and pGATA23, a majority of the T1s was switched in the lateral root only (81% and 92% of the seedlings, respectively) (Fig. 2a), showing that the integrase switch can record the transcription of a development-related gene. Additionally, this data prove that the integrase system can faithfully trace cell lineage, as all cells, even those in fully emerged lateral roots, continued to express mScarlet only.

Other promoters did not fulfill the specifications. For pLBD16, we observed only 30% LR-only seedlings, while 57% of T1s showed non-exclusive expression of mScarlet (Fig. 2a). Most of the seedlings in the non-exclusive category (70%) did not display switching in the entire seedling, but instead had mScarlet expression in a few cells in the vasculature in addition to the lateral root (Supplementary Fig. 3). This "weak" non-exclusive switching pattern (Supplementary Fig. 3) corresponds to the known expression pattern of *LBD16*[43–45]. For *ARF7*, 79% of seedlings were switched in all tissues of the root (Supplementary Fig. 4). Roughly half of the non-exclusive seedlings showed a full switch and half showed some expression of mTurq (in addition to mScarlet) in the entire root. This result is consistent with *ARF7* being expressed in other tissues and other times of development[46].

Our next question was whether the integrase system would remain robust over subsequent generations, or whether some low level of leakiness would lead to plants where every cell was in the switched state. For these experiments, we selected three T1 lines where PhiC31 was driven by pARF19, pGATA23 or pLBD16, and which were characterized as having LR-only switching. From each line, we characterized 20 progeny (T2 seedlings). In all cases, we observed a decrease in LR-only seedlings in the T2 generation (Fig. 2b). For pGATA23 T2s, we observed LR-only switches, but at a lower proportion than in T1s, and obtained seedlings displaying no-switch and non-exclusive switches. This pattern is not surprising, as in the T1

generation, plants are hemizygous for the integrase transgene insertion events, meaning that some T2s may end up with no integrase and others may have different numbers of insertions leading to a range of expression levels. For *ARF19*, the majority of T2 seedlings are fully switched (96%, 100%, 98%), also consistent with an increased dosage of integrase in many of the T2s. The lack of no-switch category T2s for this construct suggests that the integrase expression may be happening during gamete development and then transmitted to all of the cells in the T2 generation. For pLBD16 T2s, most of the seedlings are either no-switch or non-exclusive, with 66% or more of the non-exclusive seedlings having a weak switch in the main root similar to the T1 generation (Supplementary Fig. 3). To obtain a robust, cell-type specific switch, the expression level of the gene of interest in those cells should be significantly greater than that in other cell types. With *LBD16*, it seems that the expression level in LR cells is similar enough to that in the phloem pole pericycle (BAR Webservices[43],) to make LR-only switching rare.

A suite of tools to optimize switch sensitivity by tuning integrase activity

The integrase switch is a binary output, while gene expression is analog and conventionally defined relative to a standard "background" or "basal" level. For example, low-level gene expression is often "rounded down" to be defined as off when it falls below an arbitrary threshold and is considered specific to a developmental event when enriched above a similarly arbitrary threshold. To be able to record events marked by different promoters, each with their own relative levels of "off" and "on", we needed to be able to tune the sensitivity of the integrase switch (e.g., at what level of promoter expression the integrase switch is activated; Fig. 3a). While there is a rich literature of characterizing modular modifications for tuning protein activity in other systems, there are relatively few such parts available for plant synthetic biologists. We decided to characterize modifications that were predicted to work at the transcriptional, translational, and post-translational levels (Fig. 3b).

We tested our tuning methods by expressing the integrase construct constitutively and observing the resulting level of switching in the roots of T1 seedlings. While in theory, the constitutively expressed integrase should cause every cell to behave in the same manner, stochasticity in transcription, translation, and integrase activity results in cell-to-cell variation in the precise timing when switching occurs. This variation makes it possible to use the level of switching at a given time point as a performance metric that serves as a proxy for integrase activity. To capture the range of variation observed, each seedling was assigned to one of five classes, capturing the relative level of mTurq to mScarlet observed (Fig. 3c, Supplementary Fig. 5). The classes ranged from no switching (mTurq only) to full switching (mScarlet only).

To mimic integrase expression under developmental promoters of different strengths and to capture the impact of transcriptional control modifications, we used three constitutive promoters of increasing strengths: pPP2AA3, pUBQ10, and p35S (Fig. 3d). We observed subtle differences in the switching behavior among the various promoters with p35S-driven integrase lines showing the highest percentage of seedlings in the full-switch class (Fig. 3d). As a further test of transcriptional control and in recognition of recent work documenting the striking impact of terminator sequences on gene expression[47–51], we switched the UBQ1 terminator for one of our promoters, pPP2AA3, to the 35 S terminator. We found that the constructs with t35S showed a decreased switch sensitivity compared to those with tUBQ1 ($p < 0.001$) (Fig. 3e). This result further highlights the importance of promoter-terminator interactions, which could involve loop formation or the preferential localization of transcription factors to different terminator regions[52,53].

For post-transcriptional modifications, we studied the impacts of an SV40 T antigen-derived nuclear localization signal (NLS)[54] predicted to increase integrase activity[55] and an RNA destabilization tag (DST) from SMALL AUXIN UP-REGULATED RNA (SAUR) genes[56] predicted to decrease activity (Fig. 3b). For pPP2AA3, the addition of the NLS appeared to increase the proportion of fully switched seedlings when compared to the construct with the integrase alone, although the difference was not statistically significant ($p = 0.15$) (Fig. 3d). For the stronger promoters pUBQ10 and p35S, the addition of the NLS did not significantly affect the switching threshold, which was likely already at a maximum level. The addition of the DST significantly decreased the switch sensitivity for all three promoters (pPP2AA3: $p < 0.01$, pUBQ10, p35S: $p < 0.001$) (Fig. 3d). In all cases, the addition of the DST increased the proportion of seedlings in the weaker switch categories and, in the case of pUBQ10 and p35S, reduced the proportion of fully switched seedlings.

As a final option for post-translational tuning, we tested two ubiquitin (Ub)-based protein destabilization tags (Fig. 3e). These tags work by exposing an N-terminal residue which triggers the degradation of the protein by ubiquitin ligases[57]. Previously characterized in *Saccharomyces cerevisiae* (yeast), the N-end rule states that the identity of the N terminal residue determines the half-life of the protein, thus different Ub degrons confer varying levels of instability[58]. We chose a Ub-Arginine (UbR) and a Ub-Glutamine (UbQ) degron to test as in yeast they had a strong or modest impact on protein turnover, respectively[59]. While the N-degron pathway has been characterized in plants[60], it has not been used in synthetic circuits *in planta* to tune protein levels. Consistent with the yeast results, we found that both degrons significantly increased the threshold for the integrase switch when compared to the integrase alone ($p < 0.001$ for both comparisons), with UbR acting more strongly than UbQ.

As transient expression in *N. benthamiana* is a favorite testbed for plant synthetic biology applications, we wanted to know whether this toolbox of tuning strategies would be useful in that context as well. In addition to the pPP2AA3, p35S, and pUBQ10 promoters, we also tested the collection of tuning options with pARF19, classified here as a weak constitutive promoter. In addition to its well-known role in initiating new lateral roots, ARF19 is also important for leaf expansion in *Arabidopsis*, so we hoped that it might be expressed in *N. benthamiana* leaves as well[61]. The NLS had a similar effect as in *Arabidopsis*, increasing the switch sensitivity for the integrase expressed under pPP2AA3 and pARF19 but not pUBQ10 or p35S (Supplementary Fig. 6). Unlike in *Arabidopsis*, the DST did not have a significant effect on switching in *N. benthamiana* (Supplementary Fig. 6). The 35 S terminator with the NLS significantly increased switching and without the NLS tag increased switching with approaching statistical significance ($p = 0.17$) as compared to the UBQ1 terminator in *N. benthamiana* (Supplementary Fig. 6). This is in contrast to our findings in *Arabidopsis*. In addition, the effect of the Ub degrons was quite different from what was observed in *Arabidopsis* (Supplementary Fig. 6). UbR, which drastically reduced switching in *Arabidopsis*, did not significantly affect the switching in *N. benthamiana*. Even more surprisingly, UbQ which conferred a more modest, but significant reduction in switch sensitivity in *Arabidopsis*, increased switching in *N. benthamiana*. These differences have practical implications for optimizing synthetic devices, but also point to potentially fundamental differences in N-end rule dynamics and control between the two plants.

Engineering robust integrase switches with developmental promoters

We next wanted to test the impact of the tuning modifications for developmental promoters, and focused on the impact of the NLS and DST in combination with pGATA23 and pARF19 constructs. As expected, the NLS increased the proportion of seedlings showing non-exclusive switching from less than 10% to above 86% (Fig. 4a). Conversely, the addition of the DST led to the absence of non-exclusive

https://doi.org/10.1038/s41467-023-37607-5

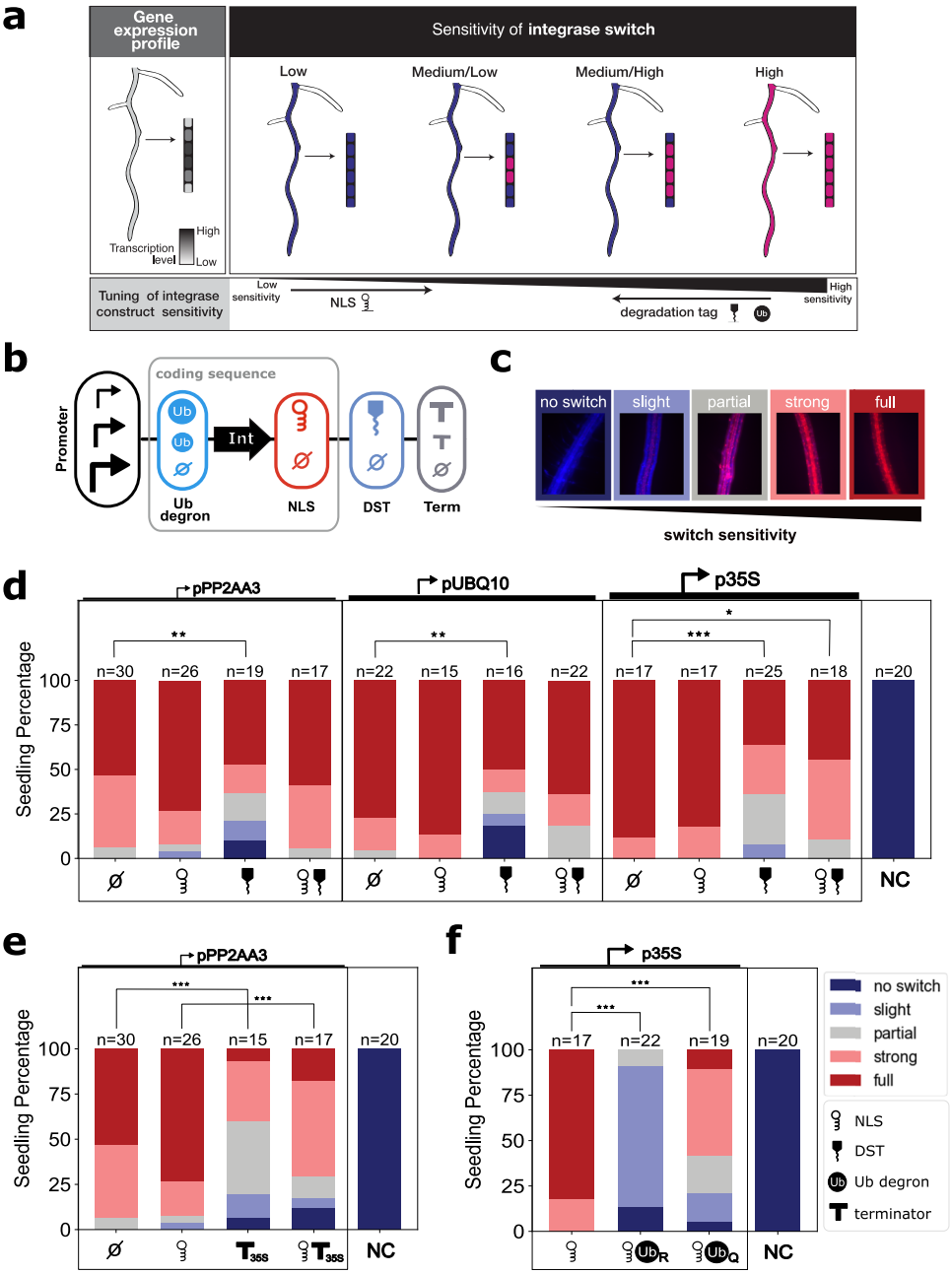


Fig. 3 | Methods for tuning integrase switch sensitivity. a For the gene expression profile of a given gene, a low-sensitivity integrase switch will result in little or no switching in any cells while a high-sensitivity switch will result in complete switching even in cells with relatively low expression of the gene. Different sensitivities of the integrase switch can lead to switches occurring at different levels of transcription. This sensitivity must be tuned to achieve the desired specificity for a given gene expression profile. **b** The integrase tuning constructs consist of a constitutive promoter controlling the integrase expression with tuning add-ons including a nuclear localization signal (NLS), RNA destabilization tag (DST), ubiquitin (Ub) degrons, and varied terminator. **c** The level of sensitivity is sorted into one of five categories, evaluated based on the level of mScarlet compared to mTurquoise fluorescence (Supplementary Fig. 5). **d**, **e** To evaluate statistical significance, each switching category was assigned a number from 1 through 5 (1 = no switch, 5 = full switch). Significance was determined using a two-sided analysis of variance (ANOVA) with a post-hoc Tukey's Honest Significance Difference test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). The negative control (NC) is the target line without any transformed integrase construct. **d** Tuning results using 3 constitutive promoters (pPP2AA3, pUBQ10, p35S) with NLS, DST, and both. From left to right, the p values are as follows: 0.0099, 0.0057, 0, 0.0235. **e** Tuning results from varying the terminator from tUBQ1 to t35S. p values from left to right are 0.0001 and 0.0017. **f** Tuning results from and Ub degrons with all p values equal to 0. Source data are provided as a Source Data file.

switching, and a slight increase in seedlings with no observed switch (from 5% no-switch in pGATA23 alone to 9% with DST; from 9% no-switch in pARF19 alone to 10% with DST) (Fig. 4a). Similarly, for pLBD16, the addition of the NLS leads to non-exclusive switching in 100% of seedlings, with 97% of the seedlings fully switched to mScarlet expression in the entire seedling (as opposed to the small number of non-LR cells showing switching in the transgenics expressing unmodified integrase from pLBD16) (Fig. 4a and Supplementary Fig. 3). When pLBD16 was used to drive expression of an integrase modified with the DST, no seedlings were categorized as LR-only (Fig. 4a), but there were a higher proportion of seedlings with a switch only in the LR and in a few cells in the main root (76% with DST, 47% without; Supplementary Fig. 3). This trend is consistent with the DST allowing recording of which cells express the highest level of *LBD16*.

https://doi.org/10.1038/s41467-023-37607-5

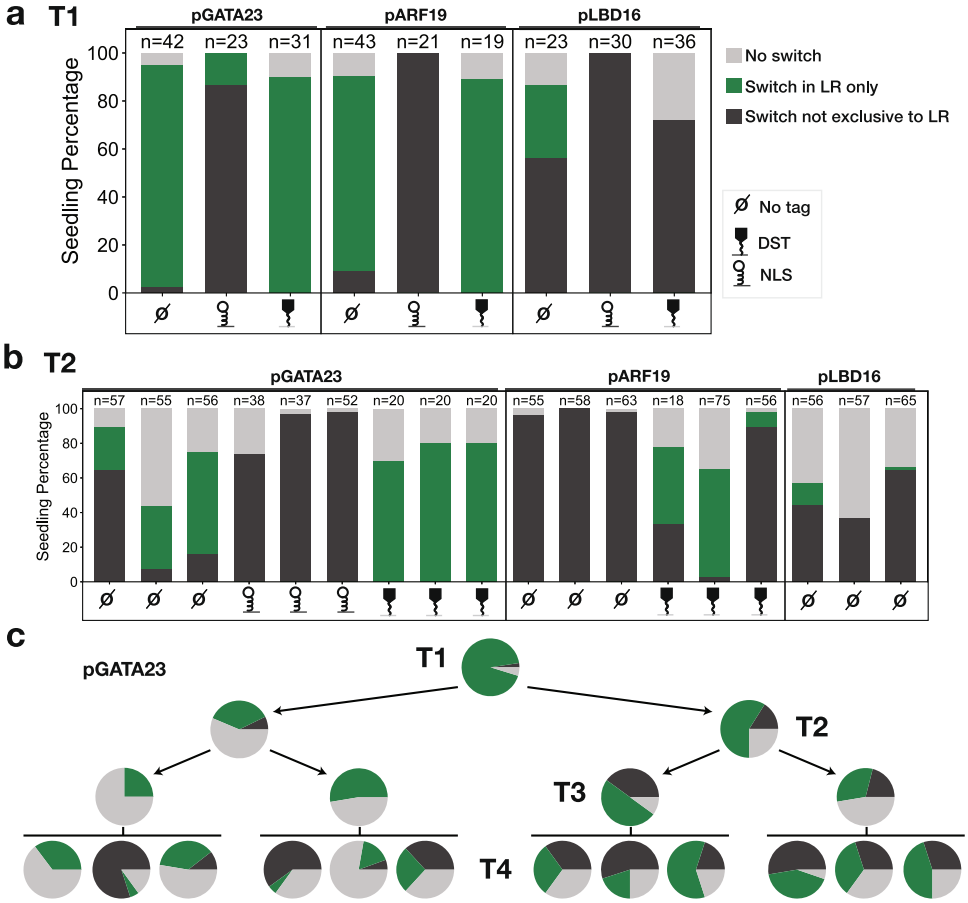


Fig. 4 | Developmental promoters with tuning tags are stable over multiple generations. a Phenotyping of T1 seedlings with constructs in PhiC31 target line, PhiC31 integrase is expressed from the indicated developmental promoters in combination with various tuning tags (legend on the right). The percentage of seedlings in each of the defined phenotypic categories is shown. **b** Phenotyping of T2 seedlings from a subset of the T1 lines represented in (**a**). T2 from three T1 lines per construct were characterized, all the T1 lines selected for T2 were switching only in the LR. **c** Percentage of pGATA23:PhiC31 (no tag) seedlings in each phenotypic category over four generations. The pie charts for T1 and T2 are derived from the same data as displayed in (**a**) and (**b**). From each generation, three seedlings categorized as LR-only were selected for propagation. Source data are provided as a Source Data file.

As for the constructs without tags, we tracked the stability of integrase switching over multiple generations. As seen previously, the ratio between switching categories changed somewhat between T1s and T2s. For pGATA23, the addition of the DST increased the stability of the phenotypic ratio of the T2 generation (Fig. 4b). We did not observe any T2 seedlings with non-exclusive switches while using the DST. The presumed increase in integrase efficiency with the NLS led to a complete absence of T2 seedlings with an LR-only switch, consistent with the T1 pattern. For pARF19, the integrase switch was no longer LR-specific in T2s (Fig. 2b). The addition of the DST increased the stability of the switch in the T2 generation, leading to an LR-only switch of up to 47% of seedlings in one T1 family (Fig. 4b). We still observed a high variability between families, with some showing mostly non-exclusive switching (Fig. 4b and Supplementary Fig. 7). For pLBD16, the T2s had a higher proportion of non-exclusive seedlings, although the addition of the DST narrowed the extent of mScarlet expression outside the LR (Fig. 4b and Supplementary Fig. 3). We also investigated the extent to which no-switch T2s represented individuals that had lost the integrase (as T2s were not selected on an antibiotic before characterization). After performing a post-characterization selection, we found that the proportion of no-switch seedlings was highly reduced (Supplementary Fig. 8), meaning that we are likely underreporting the stability of the lines in T2s.

In summary, tuning allowed us to obtain at least one T1 line for each promoter that accurately recorded the expression of the corresponding native gene. As mentioned previously, approximately one quarter of T2 seedlings from each T1 line lost the integrase construct via segregation and therefore was not switched. These unswitched seedlings can be easily removed from the further analysis without requiring the use of antibiotic selection, as we did not see switching in the absence of integrase in any case; however, this background could be an issue for some applications.

To analyze later generations, we followed a T1 carrying pGATA23:PhiC31 (no tag) over four generations, propagating two LR-only seedlings at each generation. In the fourth generation, we obtained in median 32.5% (ranging from 5 to 60%) of seedlings with an LR-only phenotype (Fig. 4c). While there was clear line-to-line variability and some loss of phenotypic robustness, we could find many lines where the integrase-based recorder was still working well even in the T4 generation. The addition of the DST appeared to further stabilize the recorder function, as 65 to 100% of T3 seedlings were LR-only and there were no seedlings with non-exclusive switching phenotypes (Supplementary Fig. 7c).

In addition to wanting performance stability across generations, we also wanted to make sure that an integrase-based recorder would faithfully record the spatiotemporal pattern of developmental gene expression, as there is an inherent lag between the induction of a promoter of interest and the time when the switched target reporter is detectable. To test whether this time gap was relevant to the timescales of lateral root development, we compared the expression pattern of our genes of interest using a traditional transcriptional reporter with the expression pattern of the integrase switch system driven by the same promoter. While our initial characterization shown in Fig. 2

https://doi.org/10.1038/s41467-023-37607-5

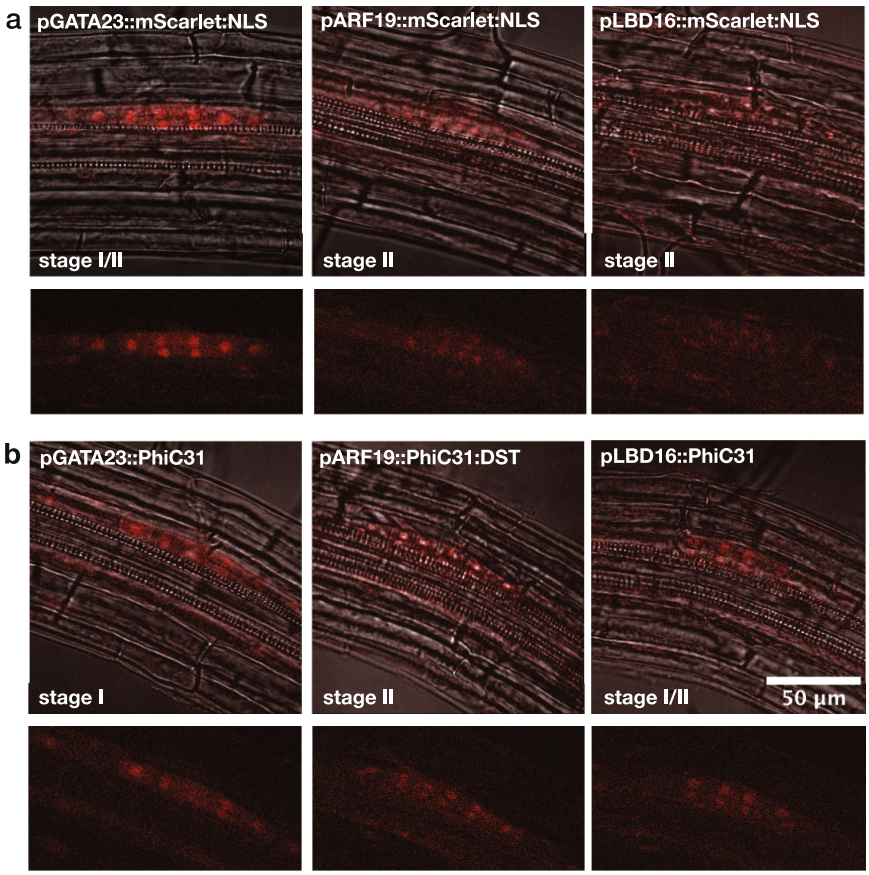


Fig. 5 | Confocal imaging of transcriptional reporters and integrase-based recorder in early-stage lateral roots. a, **b** Overlays of brightfield and red fluorescence channels from a single frame are shown above an image of the red fluorescence channel alone. For each image, the developmental stage of the lateral root primordium is indicated. **a** Transcriptional reporter lines, composed of the promoter of interest driving expression of mScarlet fused to an NLS. **b** PhiC31 integrase-based recorder lines (pDvp:PhiC31 in PhiC31 target line switching from mTurq to mScarlet) with any modifications indicated in each panel. 3 seedlings per line were imaged. Source data are provided as a Source Data file.

revealed the overall pattern of integrase-based recorder activity, for these comparisons we focused our attention at the earliest stage of lateral root development. Onset of expression for transcriptional reporters and integrase constructs appeared essentially identical (Fig. 5), indicating that the integrase system records the spatiotemporal pattern of gene expression with no significant delay. Beyond allowing for heritable gene expression in all daughter cells, an additional benefit of the integrase system was an amplification of the developmental promoter signals. This was most obvious with the weakest promoter, pLBD16. By the same logic, the integrase system could be of great use for any application requiring normalization of output levels from multiple promoters or across multiple input signals.

Increasing the potential applications of the integrase-based recorder

Another challenge with the integrase-based recorder is that many cellular events of potential interest may not have well-characterized promoters associated with them, or may rely on promoters that are activated in multiple cell types or conditions. For example, any promoter active in the embryo could trigger the switch of the integrase target, making any subsequent recording impossible. To overcome this limitation, we built additional tools that allow induction of the integrase at a user-determined time in development.

The first of these tools is the split intein integrase system which has already been characterized in vitro[62]. Inteins are sequences that trigger autocatalytic splicing, making it possible to reconstitute proteins from fragments expressed from two separate constructs[63], a technique that has been used previously in plants[64]. In the split intein integrase system, we applied here, the PhiC31 integrase is split into two extein domains: the N-terminal sequence fused to the intein N-term: Npu DnaEN and the PhiC31 C-terminal sequence fused to the intein C-term: Ssp DnaEC (Fig. 6a). Expression of the two components in a single cell triggers post-transcriptional trans-splicing, generating a fully functional PhiC31 integrase. We tested the split-intein integrase system in *Arabidopsis* with the PhiC31 integrase using strong constitutive promoters for the expression of the two components: pUBQ10 for the N-term, and p35S for the C-term, and found that it worked well (Fig. 6b). We compared the level of integrase switch from this construct to the full integrase under the control of each of the constituent promoters, and found that the split-intein system led to a decrease in switch efficiency. While 90% of split-intein T1 seedlings showed some level of switching, no full switch was observed. In contrast, when the full integrase was expressed under the control of either pUBQ10 or p35S, 75% or more of the seedlings were fully switched. This is consistent with reports that the trans-splicing approach delays the integrase switch in *E. coli*[62]. We additionally tested the split-intein integrase system with pARF19 driving both integrase components (Supplementary Fig. 9a). We observed LR specific switches in T1 seedlings (Supplementary Fig. 9b). Even with its lower switching efficiency, the split-intein integrase system allows the recording of developmental gene expression. Using pARF19:Int without tags led to most T2 seedlings having non-specific switching outside of the lateral root. By combining pARF19 with the split-intein system, most of the seedlings now showed the desired LR-specific switching phenotype

https://doi.org/10.1038/s41467-023-37607-5

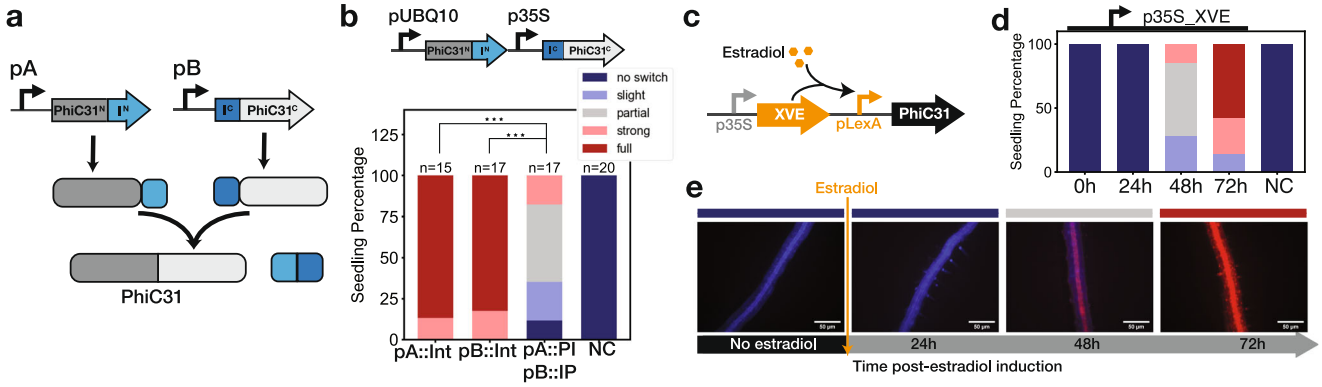


Fig. 6 | Split-intein and inducible promoters as additional tools to tune and induce the integrase switch. a Schematic of the split-intein integrase system. The system is composed of two constructs: (i) promoter A driving the N-terminal half of the integrase (PhiC31^{N}, dark gray) fused to the N-terminal portion of the intein protein (I^{N}: Npu DnaE^{N}, light blue) and (ii) promoter B driving the C-terminal portion of the intein (I^{C}: Ssp DnaE^{C}, dark blue) fused to C-terminal half of the integrase (PhiC31^{C}, light gray). When the two constructs are expressed, the inteins autocatalyze trans-splicing, covalently joining the two parts of PhiC31. **b** The split-intein system reduces the efficiency of the integrase switch. Following the nomenclature of (**a**), pUBQ10 = promoter A and p35 = promoter B. The integrase switch efficiency of the split-intein system is compared with the full integrase expressed with promoter A alone (pUBQ10: PhiC31) and with promoter B alone (p35S:PhiC31). NC corresponds to the target line without integrase. T1 seedling phenotypes are determined with fluorescent microscopy images and categorized from no switch to full switch. The data were tested for significance using an two-sided ANOVA and post-hoc Tukey's HSD test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$), with all p values equal to 0. **c** The estradiol inducible integrase construct is composed of p35S:XVE (transcriptional activator composed of a DNA-binding domain of LexA, the transcription activation domain of VP16, and the regulatory region of the human estrogen receptor[66];) and pLexA-minimal 35S driving expression of PhiC31. **d** Characterization of the estradiol inducible integrase construct shows induction as early as 48 h after treatment. 7 T1 Seedlings with the estradiol integrase construct in the PhiC31 target line were characterized just before estradiol treatment and every 24 h following. The bar graph represents the percentage of seedlings with a given level of switching (classes are color coded as in (**b**), $n = 14$ seedlings. **e** Representative images of a seedling at the specified time point relative to estradiol induction. Source data are provided as a Source Data file.

(Supplementary Fig. 9c). In this way, the split-intein system can be used as a tuning mechanism, as well as allowing the combination of different promoters.

The split-intein integrase system could be used to induce the integrase recording system at a specific stage of seedling development, thereby avoiding recording at earlier stages. This would be done by placing one component under the control of a developmental promoter and the other under the control of an inducible promoter. We could then activate the integrase system through the inducible promoter at the beginning of an experiment to record the expression of genes only after a specific time point, reducing issues with genes expressed in embryonic tissues. As a proof-of-principle for this design, we used the heterologous estradiol-inducible system[65,66] to drive integrase expression. This same system has been used to induce the expression of Cre recombinase in the context of cell-lineage tracing[26–28]. Before estradiol induction, we did not observe any switch, confirming that the estradiol system had an undetectable level of background activity (Fig. 6c). After induction, we analyzed seedlings every 24 h and observed the earliest signs of switching at 48 h with more than 50% of seedlings fully switched by 72 h. This timing fits well with reports that estradiol induction of a reporter peaks at 24 h[66], and would suggest that it takes approximately 24 h after promoter activation for the integrase to become active, mediate the switch and then allow expression and maturation of mScarlet. We also characterized our inducible integrase construct in T2s and confirmed that the baseline of expression without inducer is low enough to prevent any integrase switching in subsequent generations (Supplementary Fig. 10).

Discussion

Integrase-based recorders of gene expression have a number of advantages over current methods of tracking transcription in individual cells. Among the most prominent of these is that early events can be read-out much later in development, and, in the designs presented here, there is no need to disrupt the spatial relationship between cells. Even more, the use of serine integrases over tyrosine integrase brings new design possibilities to build history-dependent trackers and complex integrase-based devices. We have added a suite of characterized parts to help synthetic biologists build serine integrase devices in *Arabidopsis*. In addition to the PhiC31 and Bxb1 integrases and cognate targets, we built and tested tuning modifiers like RNA and protein destabilization tags and a split-intein control module. This entire suite of standardized tools can be directly implemented in any system where fine control of protein levels is needed to optimize performance. We also provided proof-of-principle that integrase-based recorders can be used to capture the history of gene expression at specific times and spaces during plant development. While we observed line to line variability in the robustness of our integrase switch, we successfully obtained at least one T1 integrase reporter line per promoter which faithfully records the spatiotemporal pattern of the gene expression. Importantly, we also found that switches functioned robustly over multiple generations. In all these experiments, we set a high standard for transparency around performance variability in plant synthetic devices, including showing quantitative characterization of many seedlings from different independent insertion events over several generations. Additionally, we observed differences in the effect of tuning parts between *Arabidopsis* stable lines and *N. benthamiana* transient assays, adding another note of caution in developing synthetic devices for use in multiple plants.

The integrase-based recording system characterized here can be readily adapted to the tracing activity of other promoters, including those expressed in other tissues and developmental processes. Because the integrase acts as a signal amplifier, the integrase system could be of interest in the following expression of any genes that are difficult or impossible to observe with traditional reporters. Additionally, the integrase system could be used to record the expression of genes in situations where live imaging is not available. For example, while many labs have at least some access to fluorescent microscopy, most do not have sophisticated live-imaging setups. There are also conditions, such as roots growing in natural soil conditions, where it would be highly advantageous to read out early expression events much later in development. Moreover, in situations where imaging is not compatible with other protocols (e.g., some fixation techniques),

it is also possible to detect the state of the integrase targets used here by sequencing.

Additional synthetic devices should now be accessible working from the toolbox described here. For example, one challenge in producing a developmental recorder is that many promoters of interest are expressed at multiple points in development. One solution would be to combine our inducible integrase and split-intein integrase system, where one part of the integrase is under the control of the externally inducible promoter, and the other is expressed from the developmental promoter of interest. Another use of the split-intein integrase would be to use it as an AND gate by placing the two split-intein components under the control of two promoters from genes of interest. This will allow the recording of when and in which cells two different genes are simultaneously expressed. By using both PhiC31 and Bxb1 integrases, a history-dependent tracker could be constructed with the capacity to record on a single cell level if, and in what order, two genes are expressed. A similar design has been shown previously to work in bacteria[19].

In addition to contributing to our understanding of existing organisms, integrase-based devices can also enable the engineering of novel forms or functions by driving the expression of genes other than reporters. For example, integrase switches could be used to induce the expression of a toxic gene under certain conditions. Cre recombinase, a tyrosine integrase, has already been used to generate homozygous fertilization-defective mutants in plants[67], and to activate a large-tumor antigen in mice[68,69]. A particularly exciting application to imagine is to replace reporter genes in integrase targets with transcription factors able to initiate entire response cascades. Root development could be re-coded by implementing history-dependent synthetic signaling circuits that used integrases to activate developmental regulators, potentially helping plants survive drought or flooding.

Methods

Construction of plasmids

Our cloning strategy was based on Golden Gate assembly using appropriate spacers (Supplementary Fig. 11)[70] and BsaI-HFv2 (NEB) as the restriction enzyme. Candidate promoter sequences (ARF7: AT5G20730, ARF19: AT1G19220, LBD16: AT2G42430, GATA23: AT5G26930) were amplified from Col-0 genomic DNA to add specific Golden Gate spacers. After gel purification, each level0 promoter sequence was cloned using a Zero Blunt PCR Cloning Kit (ThermoFisher Scientific). The PhiC31 integrase sequence was a gift from the Orzeaz lab. Bxb1 and Tp901 sequences were a gift from the Bonnet lab. Integrases were amplified using primers with golden gate compatible spacers to generate level 0 integrase parts (primer list available in Supplementary Data 1). Constitutive plant promoters and terminators were purchased from Addgene as part of the MoClo Toolbox for Plants[70]. Some level0 parts were ordered from Twist Bioscience: a mutated version of the pPP2AA3 promoter without BsaI sites, the DST, the Ub-tags, and the mTurq-tUBQ10 level0 construct for target construction. The mScarlet-tRBCs level0 construct was amplified from a transcriptional reporter[10]. Other level0 fragments were ordered from IDT as Gblocks: the codon-optimized Tp901 integrase sequence, the two split-intein PhC31 constructs, and the integrase target sequences without promoters. For the integrase target level0 sequences, the pUBQ10 promoter was added by Golden Gates using BbsI sites.

Construction of constitutive and lateral root-specific level 1 integrase constructs was performed via Golden Gate reaction in the modified pGreenII-Hygr vector containing compatible Golden Gate sites[71]. The construction of integrase targets was performed with the same methods in a modified pGreenII-Kan vector. Construction of level 2 integrase constructs, such as the split-intein system construct, was performed by amplifying completed level 1 integrase constructs using primers with golden gate compatible spacers, then performing Golden Gate reactions in the modified pGreenII-Hygr vector containing compatible Golden Gate sites. Construction of promoter reporters was performed by assembling through Golden gate reaction the mScarlet with NLS, tRBCs terminator, and promoter in the modified pGreenII-Hygr vector[10]. Details on constructs and primers can be found in Supplementary Data 1 and 2.

Enzymes for Golden Gate assembly were purchased from New England Biolabs (NEB, Ipswich, MA, USA). PCR was performed using 2X Q5 PCR master mix (NEB) and GoTaq master mix for colony PCR (Promega, Madison, WI, USA). Primers were purchased from IDT (Louvain, Belgium), and DNA fragments from Twist Bioscience or IDT. Plasmid extraction and DNA purification were performed using Monarch kits (NEB). Sequences were verified with Sanger sequencing by Azenta Life Sciences (Seattle, USA). Chemically-competent cultures of the *E. coli* strain DH5alphaZ1 (laciq, PN25-tetR, SpR, deoR, supE44, Delta(lacZYA-argFV169), Phi80 lacZDeltaM15, hsdR17(rK−, mK+), recA1, endA1, gyrA96, thi-1, relA1) were transformed with plasmid constructs containing kanamycin resistance. Transformed *E. coli* was grown in LB media (LB broth, Miller) with kanamycin (Millipore Sigma, 50 μg/mL).

Plant growth conditions

Arabidopsis seedlings were sown in 0.5 X Linsmaier and Skoog nutrient medium (LS) (Caisson Laboratories) and 0.8% w/v agar, stratified at 4 °C for 2 days, and grown in constant light at 22 °C. Phyto agar (PlantMedia/bioWORLD) was used when imaging seedlings and Bacto agar (ThermoFisher) was otherwise.

Construction and selection of transgenic *Arabidopsis* lines

Agrobacterium tumefaciens strain GV3101 was transformed by electroporation, and subsequently grown in LB media with rifampin (Millipore Sigma, 50 μg/mL), gentamicin (Millipore Sigma, 50 μg/mL), any antibiotics carried on the specific plasmid(s), most often kanamycin (Millipore Sigma, 50 μg/mL). The floral dip method[72] was used to generate integrase target lines in Col-0, and then used to introduce each integrase construct into these established target lines. For T1 selection: 120 mg of T1 seeds (~2000 seeds) were sterilized using 70% ethanol and 0.05% Triton-X-100 and then washed using 95% ethanol. Seeds were resuspended in 0.1% agarose and spread onto 0.5X LS Bacto selection plates, using 25 μg/mL of kanamycin for target lines and 25 μg/mL kanamycin and 25 μg/mL hygromycin for lines with both the integrase and the target. The plates were stratified at 4 °C for 48 h then light pulsed for 6 h and covered for 48 h[73]. They were then grown for 4–5 days. To select transformants, tall seedlings with long roots and a vibrant green color were picked from the selection plate with sterilized tweezers and transferred to a new 0.5X LS Phyto agar plate for characterization.

Characterization of integrase switch in *Arabidopsis* transgenic lines

T1 seedlings for each line were grown 4–5 days after transformant selection. Each selected seedling was imaged at 10X magnification using an epifluorescence microscope (Leica Biosystems, model: DMI 3000) using the RFP (exposure 500 ms, gain 1.6) and CFP (exposure 300 ms, gain 1.6) channels. Selected T1 seedlings were then transferred to soil, and at maturation T2 seeds were selected. For later generations, seedlings were sterilized similarly to T1s, stratified, plated on an LS agar plate, grown for 4–5 days, and characterized using the epifluorescence microscope as for T1.

For the target lines, the seedlings with the highest level of mTurq expression were selected and transferred to soil to generate T2 seeds. The brightest among these lines was maintained as the target line for each integrase, and used for all later transformations of integrase constructs.

For the constitutive integrase constructs in a target line, around 20-30 T1 seedlings were analyzed per construct. Each seedling was categorized into one of five classes as seen in Supplementary Fig. 5

based on the level of switching. Representative images for each category were taken using the RFP and CFP channels and merged for final images. For each construct, the percentage of seedlings in each category was plotted in a bar plot with the number of seedlings tested mentioned at the top of the bar. To evaluate statistical significance, each switching category was assigned a number from 1 through 5 (1 = no switch, 5 = full switch). Significance was determined using analysis of variance (ANOVA) with a post-hoc Tukey's Honest Significance Difference test.

For the YFP to Luciferase PhiC31 target, the target line and the target with pPP2AA3-PhiC31 construct were characterized. T2 seedlings from target lines with and without integrase were grown on LS plates, 7 days old seedlings were imaged with Azure c600 Gel imaging system for YFP fluorescence. Then, 100 μM of Luciferin were sprayed on seedling, after one hour kept in the dark, and seedlings were imaged using NightOwl LB 983 in vivo imager with an exposure time of 10 min.

For the developmental promoter integrase constructs in PhiC31 target line, at least 20 T1 seedlings were analyzed per construct. Each seedling was categorized into one of three classes based on specificity of switching (LR-only, non-exclusive to LR, no-switch). Representative images for each construct were taken using the RFP and CFP channels and merged for final images. A selected number of T1 seedlings with LR-only switch were transplanted to the soil to characterize the T2 generation. For each T1 line, 20 T2 seedlings were characterized in an identical way than for T1s, and similarly for T3 and T4 generations. For each construct, the percentage of seedlings in each of the three categories were plotted in a bar plot with the number of seedlings tested mentioned at the top of the bar.

Python data analysis script which includes statistical tests and plotting functions was run in version 3.9.1 and with the following package dependencies: pandas (version 1.5.3), scipy.stats (version 1.10.0), matplotlib.pyplot (version 3.6.3), matplotlib.colors (version 3.6.3), scikit_posthocs (version 0.21), and numpy (version 1.24.2). All images taken during seedling characterization were opened and processed using the ImageJ program (version 1.53c). Each.tif image file contained the images of a seedling's RFP and CFP channels.tif files were processed through an ImageJ macro to adjust the color lookup table, brightness, and contrast of each channel (RFP: Red, Min: 200, Max: 3000) (CFP: Blue, Min: 200, Max: 4000). After adjustment, the macro overlaid the two channels to create a composite image, rotated the image, added a scale bar, and flattened the image to produce our final processed images.

Testing the hygromycin resistance of seedlings post characterization

To select T2 hygromycin-resistant seedlings after characterization without selection, the roots of 7 days old seedlings were removed with a razor blade, and seedlings were then transferred onto 0.5X LS BactoAgar plates containing hygromycin. Seedlings were screened for root regrowth after seven days. In our extensive testing of control plants, not all hygromycin resistant seedlings are able to regrow roots after this stressful intervention, but all seedlings that grow roots are truly resistant.

Characterization of the tuning constructs in *Nicotiana benthamiana*

Integrase target integrated *Nicotiana benthamiana* seeds were acquired from the Orzaez lab[23]. This line has a stably integrated integrase target which switches from LUC firefly luciferase to YFP upon integrase expression. The plants were grown 25 days before injection. *Agrobacterium*-mediated transient transformation of *N. benthamiana* was performed using the *A. tumefaciens* strains GV3101[74]. For each injection in addition to the *A. tumefaciens* with the integrase constructs, we injected an RFP injection efficiency control consisting of constitutively expressed mCherry (donated by Jennifer Brophy) and a construct containing a P19 gene silencing suppressor protein for enhanced transient transformation[75]. Each *A. tumefaciens* strain was grown overnight in LB at 30 °C, pelleted and incubated in MMA media (10 mM MgCl2, 10 mM MES pH 5.6, 100 μM acetosyringone) for 3 h at room temperature with rotation. Strain density was normalized to an OD600 of 1.5 for each strain in the final mixture of strains before injection. For each integrase construct, the integrase strain, the RFP control, and P19 were injected together; we also injected as control the RFP control and P19 together, as well as the negative control P19 alone. Each *A. tumefaciens* solution was injected into 3–4 different leaves from separate plants. Four days later, hole punches were taken from each injected leaf at 3 locations, and the punches were placed in a 96 well plate. Plate reader measurements of YFP (excitation wavelength: 506 nm, emission wavelength: 541) and mCherry (excitation wavelength: 584 nm, emission wavelength: 610) fluorescence were taken using a Spark® Multimode Microplate Reader by Tecan. Twelve measurements were taken at different locations within the punch. Three tobacco injection replicates per construct were performed and, in each replicate, three leaves were injected. For each punch, the median of the ratio of YFP over RFP fluorescence was calculated and plotted. The box corresponds to the quartile and the median between the different punches for one construct. The data were tested for significance using an ANOVA and post-hoc Tukey's HSD test. The tobacco injection data was plotted and statistically analyzed using a Python data analysis script.

Confocal imaging of reporter and integrase lines

Arabidopsis transgenic reporter lines for *LBD16, ARF19*, and *GATA23* with mScarlet nuclear localized were generated as for integrase switch transgenic lines. After characterization of T1 seedlings, seedlings expressing mScarlet were fixed in 4% formaldehyde using vacuum infiltration followed by ClearSee solution[76]. Fixed and cleared seedlings were mounted on microscope slides using 50% glycerol and Parafilm edges to prevent the coverslips from pressing on the root.

For the integrase lines, for each promoter *LBD16, ARF19*, or *GATA23*, one construct showing a reliable LR-only integrase switch was selected. For each construct, two T1 lines representative of other characterized T1 lines were selected to perform the root bend essay. For each line, 20 T2 seeds of the corresponding T1 line were placed on plates following a specific pattern to avoid a seedling collision after the rotation of the plate. The seeds were stratified for 120 h, grown vertically for 96 h at 22 °C, rotated 90° while keeping the plate vertical, and grown for an additional 20 h. Seedlings were fixed and mounted as mentioned in the previous paragraph.

Imaging of the seedlings were performed using Nikon A1R HD25 laser scanning confocal microscope with 561 laser and 578-623 detector for RFP imaging. For the integrase lines, seedlings were imaged at the bend region, while for the reporter lines, seedlings were scanned to find early-developed lateral roots. Imaging was processed using FIJI. For each imaging, a Z-stack was recorded. First, a maximum average of the Z-stack in the RFP channel was generated. Additionally, we selected one Z-location focusing on the LR nucleus and generated both an image of the RFP channel and the RFP and brightfield merged. The main figure uses the merged RFP/brightfield images.

Estradiol induction time course

For estradiol induction in T1s, antibiotic selection was performed as described in the method section about *A. thaliana* transgenic lines. Four days after transplanting resistant seedlings onto 0.5X LS Phyto plates, the seedlings are imaged via microscopy in RFP and CFP channels with identical settings as described in the method section about integrase switch seedling characterization. Then the seedlings were transferred onto new 0.5X LS Phyto plates with 10 μM β-estradiol. Each seedling was imaged 24, 48, and 72 h after transplanting onto estradiol and categorized into the appropriate switching category for each timepoint. Data were processed for tuning seedlings.

For estradiol induction in T2s, seeds were plated onto 0.5X LS Phyto plates, stratified for 48 h, and left to grow for 6 days. Then they were transplanted onto 10 μM estradiol 0.5X LS Phyto plates and imaged and categorized for T1 seedlings.

Reporting summary

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

Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. Plasmids and plant materials are available upon request from JLN (jn7@uw.edu; please expect a response within 3 weeks). Source data are provided with this paper.

Code availability

Python and ImageJ scripts are available on GitHub [https://doi.org/10.5281/zenodo.7612666].

References

1. Gray, S. B. & Brady, S. M. Plant developmental responses to climate change. *Dev. Biol.* **419**, 64–77 (2016).
2. Guiziou, S., Chu, J. C. & Nemhauser, J. L. Decoding and recoding plant development. *Plant Physiol.* **187**, 515–526 (2021).
3. Dumschott, K., Schmidt, M. H.-W., Chawla, H. S., Snowdon, R. & Usadel, B. Oxford nanopore sequencing: new opportunities for plant genomics? *J. Exp. Bot.* **71**, 5313–5322 (2020).
4. Seyfferth, C. et al. Advances and opportunities in single-cell transcriptomics for plant research. *Annu. Rev. Plant Biol.* **72**, 847–866 (2021).
5. Mergner, J. & Kuster, B. Plant proteome dynamics. *Annu. Rev. Plant Biol.* **73**, 67–92 (2022).
6. Jean-Baptiste, K. et al. Dynamics of gene expression in single root cells of Arabidopsis thaliana. *Plant Cell* **31**, 993–1011 (2019).
7. Mironova, V. & Xu, J. A single-cell view of tissue regeneration in plants. *Curr. Opin. Plant Biol.* **52**, 149–154 (2019).
8. Nelms, B. & Walbot, V. Defining the developmental program leading to meiosis in maize. *Science* **364**, 52–56 (2019).
9. Shulse, C. N. et al. High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* **27**, 2241–2247.e4 (2019).
10. Gala, H. P. et al. A single-cell view of the transcriptome during lateral root initiation in Arabidopsis thaliana. *Plant Cell* **33**, 2197–2220 (2021).
11. Liu, Z. et al. Global dynamic molecular profiling of stomatal lineage cell development by single-cell rna sequencing. *Mol. Plant* **13**, 1178–1193 (2020).
12. Oliva, M. et al. An environmentally-responsive transcriptional state modulates cell identities during root development. Preprint at bioRxiv 2022.03.04.483008 https://doi.org/10.1101/2022.03.04.483008 (2022).
13. Sadoine, M. et al. Designs, applications, and limitations of genetically encoded fluorescent sensors to explore plant biology. *Plant Physiol.* **187**, 485–503 (2021).
14. Chow, K.-H. K. et al. Imaging cell lineage with a synthetic digital recording system. *Science* **372**, eabb3099 (2021).
15. Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P. & Endy, D. Amplifying genetic logic gates. *Science* **340**, 599–603 (2013).
16. Guiziou, S., Mayonove, P. & Bonnet, J. Hierarchical composition of reliable recombinase logic devices. *Nat. Commun.* **10**, 456 (2019).
17. Weinberg, B. H. et al. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* **35**, 453–462 (2017).
18. Gomide, M. S. et al. Genetic switches designed for eukaryotic cells and controlled by serine integrases. *Commun. Biol.* **3**, 255 (2020).
19. Zúñiga, A. et al. Rational programming of history-dependent logic in cellular populations. *Nat. Commun.* **11**, 1–14 (2020).
20. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science* **353**, aad8559 (2016).
21. Hou, L. et al. An open-source system for in planta gene stacking by Bxb1 and Cre recombinases. *Mol. Plant* **7**, 1756–1765 (2014).
22. Thomson, J. G., Chan, R., Thilmony, R., Yau, Y.-Y. & Ow, D. W. PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the Arabidopsis genome. *BMC Biotechnol.* **10**, 17 (2010).
23. Bernabé-Orts, J. M. et al. A memory switch for plant synthetic biology based on the phage ϕC31 integration system. *Nucleic Acids Res.* **48**, 3379–3394 (2020).
24. Kapusi, E., Kempe, K., Rubtsova, M., Kumlehn, J. & Gils, M. phiC31 integrase-mediated site-specific recombination in barley. *PLoS One* **7**, e45353 (2012).
25. Kempe, K. et al. Transgene excision from wheat chromosomes by phage phiC31 integrase. *Plant Mol. Biol.* **72**, 673–687 (2010).
26. Efroni, I. et al. Root regeneration triggers an embryo-like sequence guided by hormonal interactions. *Cell* **165**, 1721–1733 (2016).
27. Wachsman, G., Heidstra, R. & Scheres, B. Distinct cell-autonomous functions of RETINOBLASTOMA-RELATED in Arabidopsis stem cells revealed by the Brother of Brainbow clonal analysis system. *Plant Cell* **23**, 2581–2591 (2011).
28. Smetana, O. et al. High levels of auxin signalling define the stem-cell organizer of the vascular cambium. *Nature* **565**, 485–489 (2019).
29. Lloyd, J. P. B. et al. Synthetic memory circuits for stable cell reprogramming in plants. *Nat. Biotechnol*. https://doi.org/10.1038/s41587-022-01383-2 (2022).
30. Olorunniji, F. J., Rosser, S. J. & Stark, W. M. Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochem. J.* **473**, 673–684 (2016).
31. Liu, R., Long, Q., Zou, X., Wang, Y. & Pei, Y. DNA methylation occurring in Cre-expressing cells inhibits loxP recombination and silences loxP-sandwiched genes. *N. Phytol.* **231**, 210–224 (2021).
32. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* **139**, 5–17 (2005).
33. Lilly, S. T., Drummond, R. S. M., Pearson, M. N. & MacDiarmid, R. M. Identification and Validation of Reference Genes for Normalization of Transcripts from Virus-Infected Arabidopsis thaliana. *Mol. Plant. Microbe Interact.* **24**, 294–304 (2011).
34. Chao, C.-Y.-Y. G. Engineering of tools for De Novo Assembly of Human Cells. (Massachusetts Institute of Technology, 2020).
35. Banda, J. et al. Lateral root formation in arabidopsis: a well-ordered LRexit. *Trends Plant Sci.* **24**, 826–839 (2019).
36. Torres-Martínez, H. H., Napsucialy-Mendivil, S. & Dubrovsky, J. G. Cellular and molecular bases of lateral root initiation and morphogenesis. *Curr. Opin. Plant Biol.* **65**, 102115 (2022).
37. Brophy, J. A. N. Toward synthetic plant development. *Plant Physiol.* **188**, 738–748 (2022).
38. von Wangenheim, D. et al. Rules and self-organizing properties of post-embryonic plant organ cell division patterns. *Curr. Biol.* **26**, 439–449 (2016).
39. Malamy, J. E. & Benfey, P. N. Organization and cell differentiation in lateral roots of Arabidopsis thaliana. *Development* **124**, 33–44 (1997).
40. Okushima, Y. et al. Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* **17**, 444–463 (2005).
41. Okushima, Y., Fukaki, H., Onoda, M., Theologis, A. & Tasaka, M. ARF7 and ARF19 regulate lateral root formation via direct

activation of LBD/ASL genes in Arabidopsis. *Plant Cell* **19**, 118–130 (2007).
42. De Rybel, B. et al. A novel aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity. *Curr. Biol.* **20**, 1697–1706 (2010).
43. Winter, D. et al. An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**, e718 (2007).
44. Goh, T., Joi, S., Mimura, T. & Fukaki, H. The establishment of asymmetry in Arabidopsis lateral root founder cells is regulated by LBD16/ASL18 and related LBD/ASL proteins. *Development* **139**, 883–893 (2012).
45. Goh, T. et al. Lateral root initiation requires the sequential induction of transcription factors LBD16 and PUCHI in Arabidopsis thaliana. *N. Phytol.* **224**, 749–760 (2019).
46. Orosa-Puente, B. et al. Root branching toward water involves posttranslational modification of transcription factor ARF7. *Science* **362**, 1407–1410 (2018).
47. Andreou, A. I., Nirkko, J., Ochoa-Villarreal, M. & Nakayama, N. Mobius assembly for plant systems highlights promoter-terminator interaction in gene regulation. Preprint at bioRxiv 2021.03.31.437819 https://doi.org/10.1101/2021.03.31.437819 (2021).
48. Diamos, A. G. & Mason, H. S. Chimeric 3′ flanking regions strongly enhance gene expression in plants. *Plant Biotechnol. J.* **16**, 1971–1982 (2018).
49. Nagaya, S., Kawamura, K., Shinmyo, A. & Kato, K. The HSP terminator of Arabidopsis thaliana increases gene expression in plant cells. *Plant Cell Physiol.* **51**, 328–332 (2010).
50. Felippes, F. de et al. The key role of terminators on the expression and post-transcriptional gene silencing of transgenes. *Plant J.* **104**, 96–112 (2020).
51. Ingelbrecht, I. L., Herman, L. M., Dekeyser, R. A., Van Montagu, M. C. & Depicker, A. G. Different 3′ end regions strongly influence the level of gene expression in plant cells. *Plant Cell* **1**, 671–680 (1989).
52. Tan-Wong, S. M., French, J. D., Proudfoot, N. J. & Brown, M. A. Dynamic interactions between the promoter and terminator regions of the mammalian *BRCA1* gene. *Proc. Natl Acad. Sci. USA* **105**, 5160–5165 (2008).
53. Al-Husini, N., Medler, S. & Ansari, A. Crosstalk of promoter and terminator during RNA polymerase II transcription cycle. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194657 (2020).
54. Adam, S. A., Lobl, T. J., Mitchell, M. A. & Gerace, L. Identification of specific binding proteins for a nuclear location sequence. *Nature* **337**, 276–279 (1989).
55. Lange, A. et al. Classical nuclear localization signals: definition, function, and interaction with importin α. *J. Biol. Chem.* **282**, 5101–5105 (2007).
56. Newman, T. C., Ohme-Takagi, M., Taylor, C. B. & Green, P. J. DST sequences, highly conserved among plant SAUR genes, target reporter transcripts for rapid decay in tobacco. *Plant Cell* **5**, 701–714 (1993).
57. Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186 (1986).
58. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20**, 1298–1345 (2011).
59. Chassin, H. et al. A modular degron library for synthetic circuits in mammalian cells. *Nat. Commun.* **10**, 2013 (2019).
60. Gibbs, D. J., Bacardit, J., Bachmair, A. & Holdsworth, M. J. The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. *Trends Cell Biol.* **24**, 603–611 (2014).
61. Wilmoth, J. C. et al. NPH4/ARF7 and ARF19 promote leaf expansion and auxin-induced lateral root formation. *Plant J.* **43**, 118–130 (2005).
62. Olorunniji, F. J. et al. Control of ϕC31 integrase-mediated site-specific recombination by protein trans-splicing. *Nucleic Acids Res.* **47**, 11452–11460 (2019).
63. Li, Y. Split-inteins and their bioapplications. *Biotechnol. Lett.* **37**, 2121–2137 (2015).
64. Evans, T. C. Jr, Xu, M.-Q. & Pradhan, S. Protein splicing elements and plants: from transgene containment to protein purification. *Annu. Rev. Plant Biol.* **56**, 375–392 (2005).
65. Siligato, R. et al. MultiSite gateway-compatible cell type-specific gene-inducible system for plants. *Plant Physiol.* **170**, 627–641 (2016).
66. Zuo, J., Niu, Q. W. & Chua, N. H. Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *Plant J.* **24**, 265–273 (2000).
67. Nagahara, S., Takeuchi, H. & Higashiyama, T. Generation of a homozygous fertilization-defective gcs1 mutant by heat-inducible removal of a rescue gene. *Plant Reprod.* **28**, 33–46 (2015).
68. Lakso, M. et al. Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc. Natl Acad. Sci. U. S. A.* **89**, 6232–6236 (1992).
69. Pichel, J. G., Lakso, M. & Westphal, H. Timing of SV40 oncogene activation by site-specific recombination determines subsequent tumor progression during murine lens development. *Oncogene* **8**, 3333–3342 (1993).
70. Engler, C. et al. A golden gate modular cloning toolbox for plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
71. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765 (2011).
72. Clough, S. J. & Bent, A. F. Transformation of Arabidopsis thaliana. *Plant J.* **16**, 735–743 (1998).
73. Harrison, S. J. et al. A rapid and robust method of identifying transformed Arabidopsis thaliana seedlings following floral dip transformation. *Plant Methods* **2**, 19 (2006).
74. Yang, Y., Li, R. & Qi, M. In vivo analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *Plant J.* **22**, 543–551 (2000).
75. Lakatos, L., Szittya, G., Silhavy, D. & Burgyán, J. Molecular mechanism of RNA silencing Supplementary ression mediated by p19 protein of tombusviruses. *EMBO J.* **23**, 876–884 (2004).
76. Kurihara, D., Mizuta, Y., Sato, Y. & Higashiyama, T. ClearSee: a rapid optical clearing reagent for whole-plant fluorescence imaging. *Development* **142**, 4168–4179 (2015).

Acknowledgements

We thank Wesley George, Eric Yang, Dr. Alexander Leydon, as well as other members of the Nemhauser, Imaizumi, and Steinbrenner groups, for feedback and discussions. We thank Eric Yang for developing and gifting the pPP2AA3 promoter; Dr. Jennifer Brophy for sharing the RFP injection efficiency control; members of the Bonnet lab for sending us Bxb1 and Tp901 integrase plasmids; and members of the Orzáez lab for useful discussions, as well as the *N. benthamiana* PhiC31 target line and the YFP to Luc PhiC31 target plasmid. This work was supported by grants from the National Institutes of Health (grant no. GM107084), the National Science Foundation (grant no. IOS-1546873), and the Howard Hughes Medical Institute Faculty Scholars Program. In addition, support to S.G. was provided by EMBO (grant no. ALTF 409-2019).

Author contributions

S.G. and J.L.N. designed the project. S.G. designed the constructs. C.M. performed and analyzed the tuning and estradiol inducible experiments, S.G. and J.C. performed and analyzed the developmental promoter experiments. S.G., C.M., and J.L.N. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

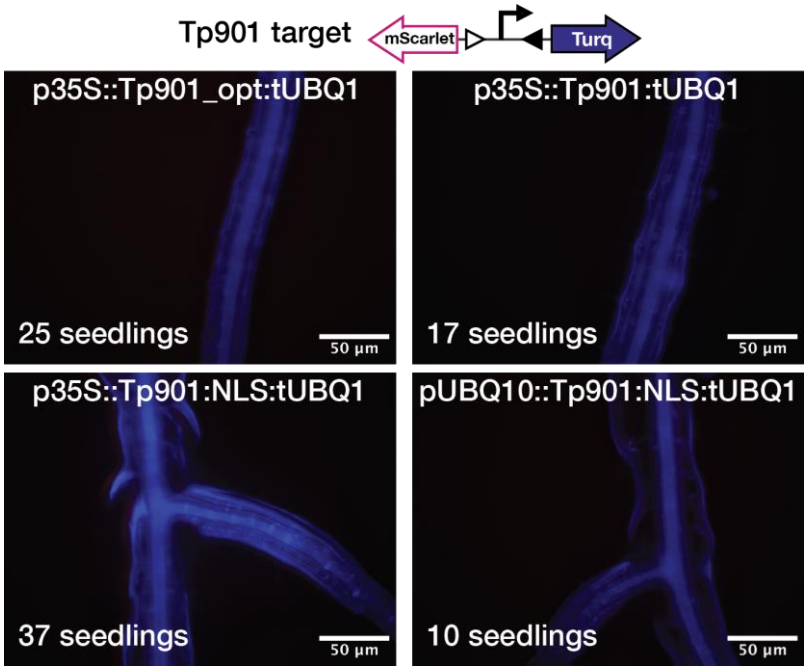
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-37607-5.

Correspondence and requests for materials should be addressed to Jennifer L. Nemhauser.

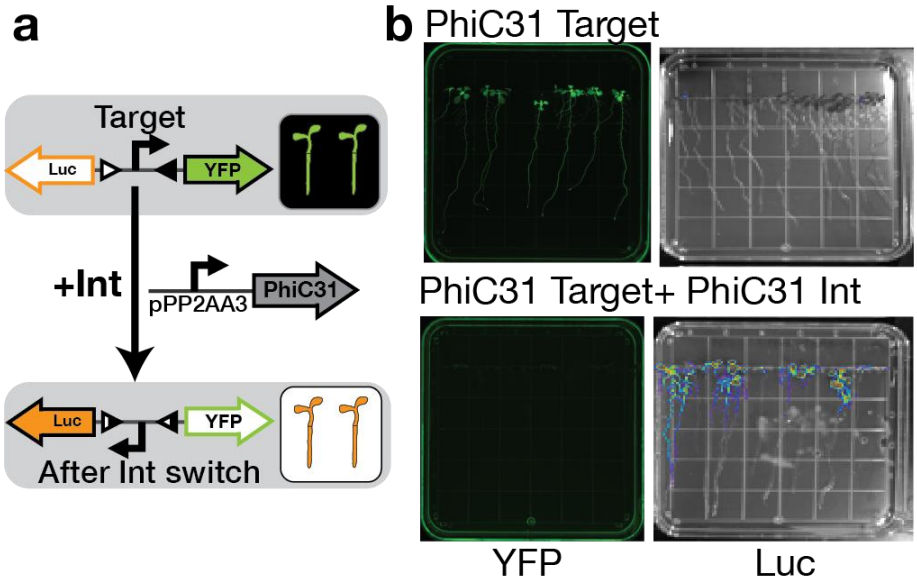
Peer review information *Nature Communications* thanks Cheng-Hsun Ho and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

An integrase toolbox to record gene-expression during plant development

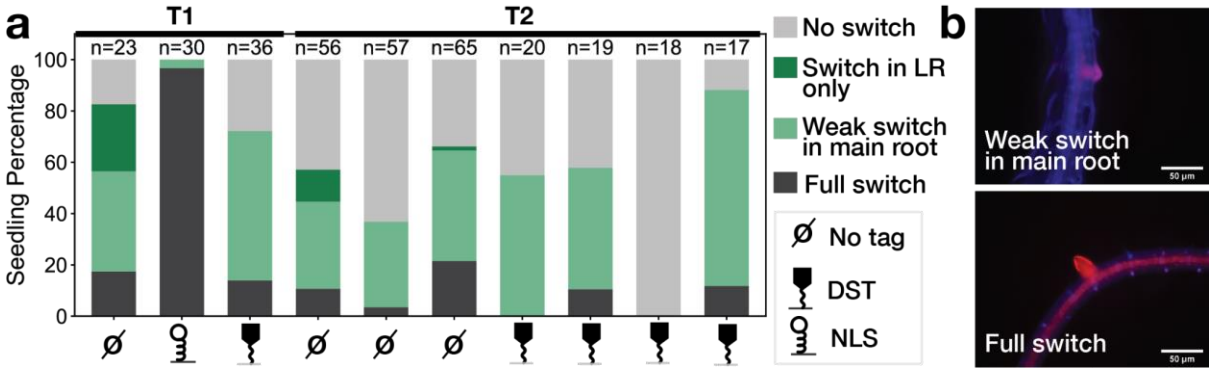
Guiziou *et al.*



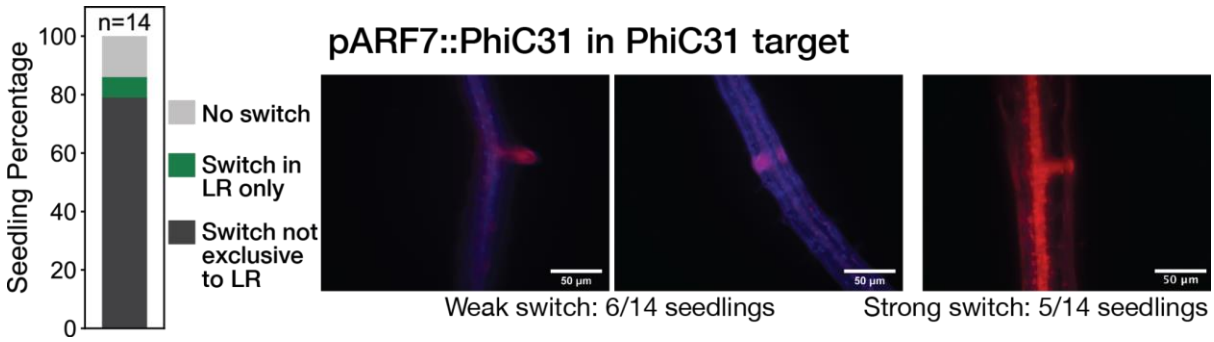
Supplementary Figure 1. Tp901 integrase does not work well in *Arabidopsis*. Tp901 target line was transformed with Tp901 integrase constructs. If the integrase was active the target should swtch from mTurq to mScarlet expression. No mScarlet was ever detected. T1 expression patterns were analyzed by microscopy. Each image corresponds to a representative image for each integrase construct, the integrase construct is mentioned on the top and the number of seedling characterization on the bottom left. The microscopy image is an overlay of the blue and red fluorescence channel. The number of seedlings indicated corresponds to the number of seedlings characterized.



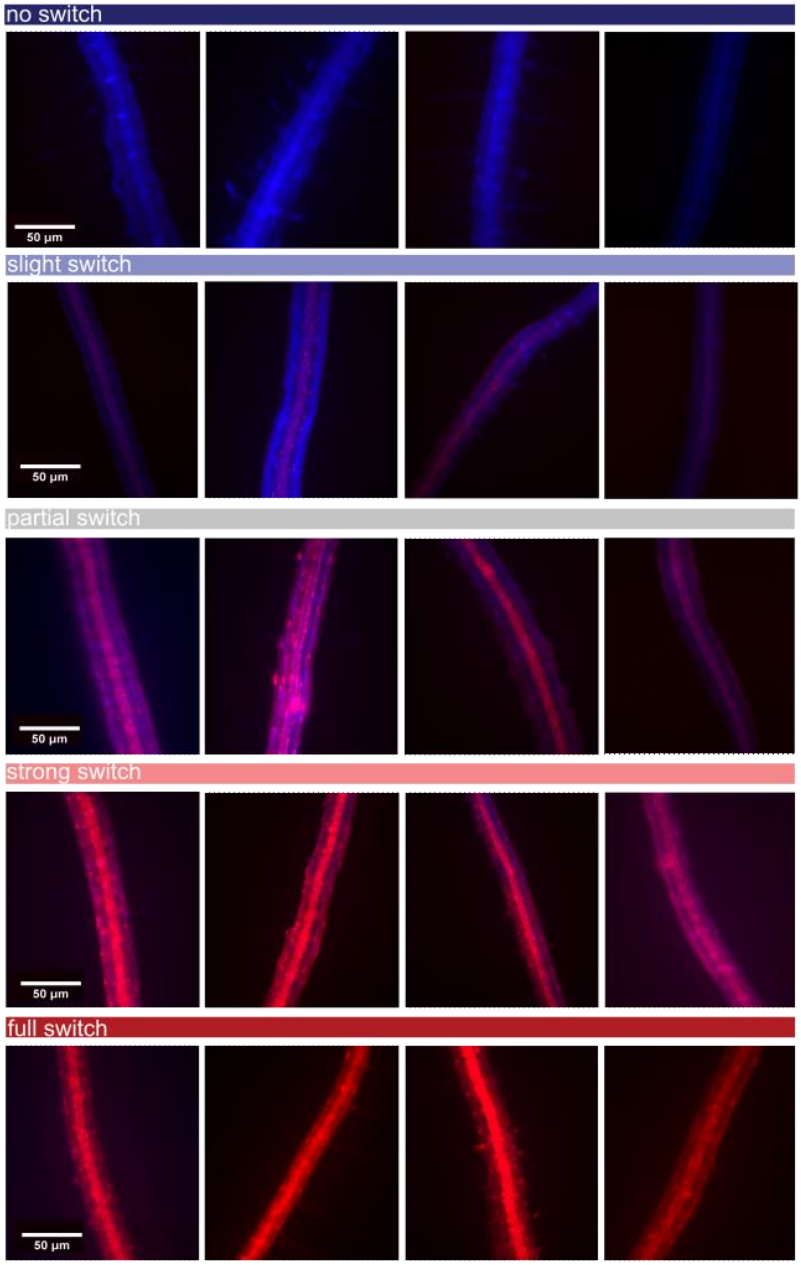
Supplementary Figure 2. Integrase target for macroscopic analysis. (a) The PhiC31 target [1] switches from YFP to Luciferase expression, allowing to image the full seedling under gel imager for YFP or night owl after luciferin treatment. We used pPP2AA3 to drive PhiC31 integrase expression constitutively. (b) Images of T2 seedling under an Azure c600 Gel imaging system (for YFP fluorescence) (left) and NightOWL LB 983 *in vivo* imaging system (for Luciferase) after luciferin treatment (right), on the top are seedlings with target and no integrase at the bottom seedlings with target and integrase.



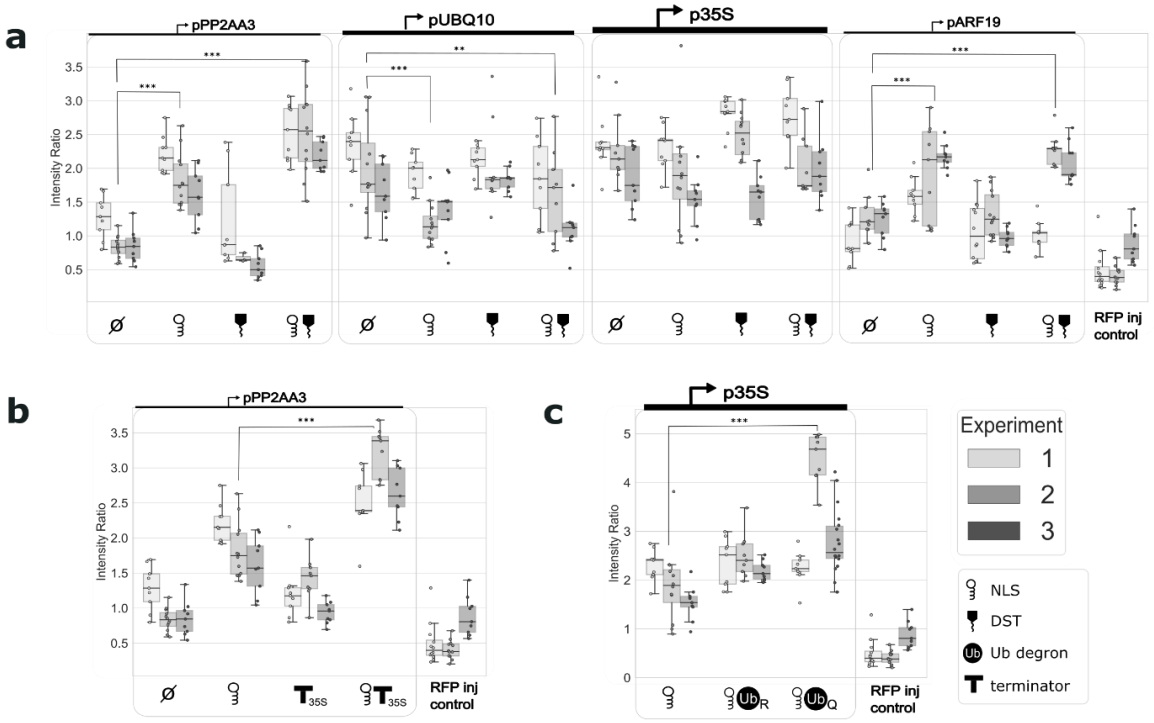
Supplementary Figure 3. Characterization of pLBD16 integrase switch constructs. (a) Phenotyping of T1 seedlings with constructs in PhiC31 target line, constructs are PhiC31 with pLBD16, and various tuning tags (label at the bottom of the graph): no tag, DST, or NLS (legend on the right). The graph corresponds to the percentage of seedlings in each of the defined phenotype categories, such as no switch corresponding to no mScarlet expression in the root, switch in LR only: mScarlet expression only in the lateral root, weak switch in the main root: mScarlet expression in few cells in the main root (corresponding to the image in b), full switch: mScarlet expression everywhere in the root (corresponding to the image in b). The number of seedlings characterized for each construct is mentioned at the top of the bar in the graph. (b) Representative image of the seedling in the weak switch in main root and full switch phenotypic categories. Microscopy images are an overlay of the blue and red channels. Source data are provided as a Source Data file.



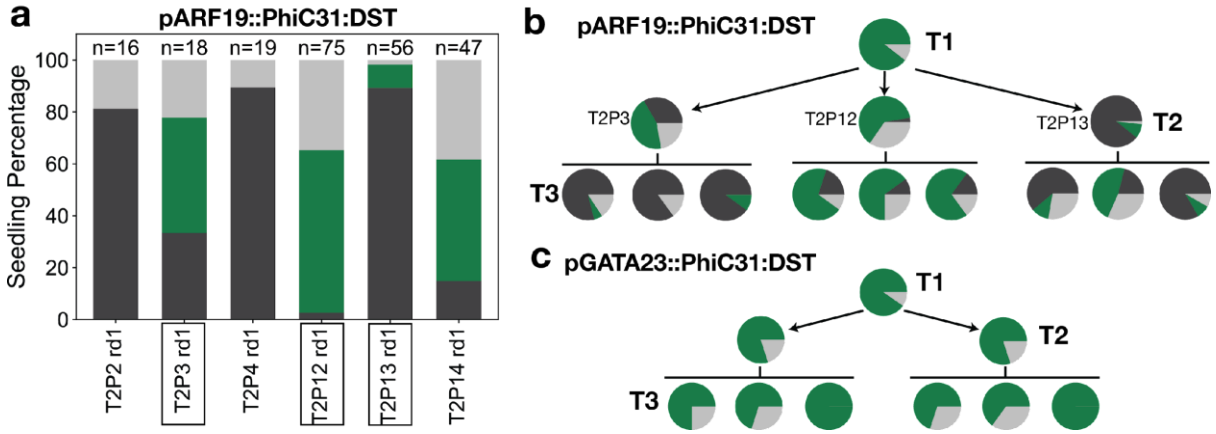
Supplementary Figure 4. Characterization of pARF7::PhiC31 switch. 14 T1 seedlings of pARF7::PhiC31 in PhiC31 target line were characterized. On the left, the bar graph corresponds to the percentage of seedlings in each of the phenotypic categories, no switch in light gray, switch only in LR in green, and switch not exclusive to LR in dark gray. On the right are representative images of seedlings with switch not exclusive to LR, either a weak switch (6/14 seedlings) or a strong switch everywhere (5/14 seedlings). Source data are provided as a Source Data file.



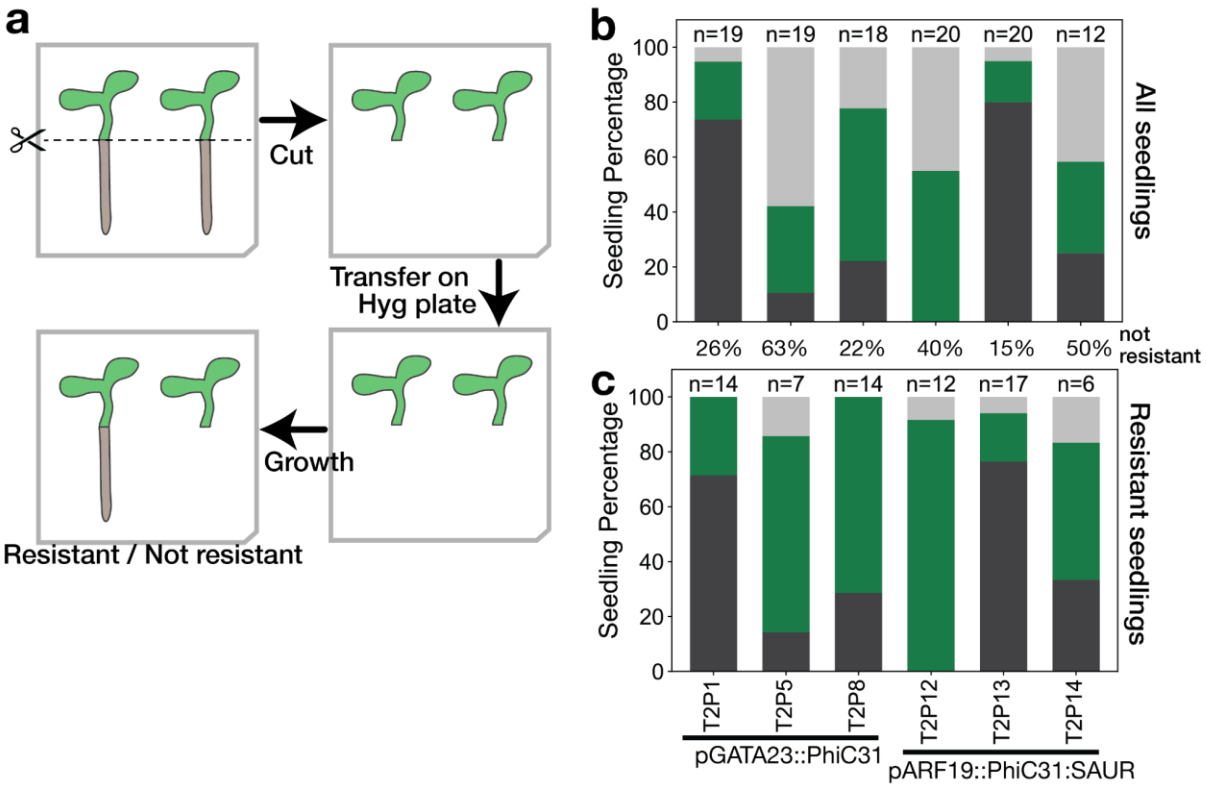
Supplementary Figure 5. Example images of switching categories for tuning data. For evaluating the tuning results we sorted each seedling into one of five categories (no switch, slight switch, partial switch, strong switch, full switch). Within each category variation is present. The 20 seedlings pictured above represent the spectrum of constitutive switching levels we observed in characterizing integrase tuning parts. The scale bar is shown in the left-most image in each row and the scale applies through the whole row.



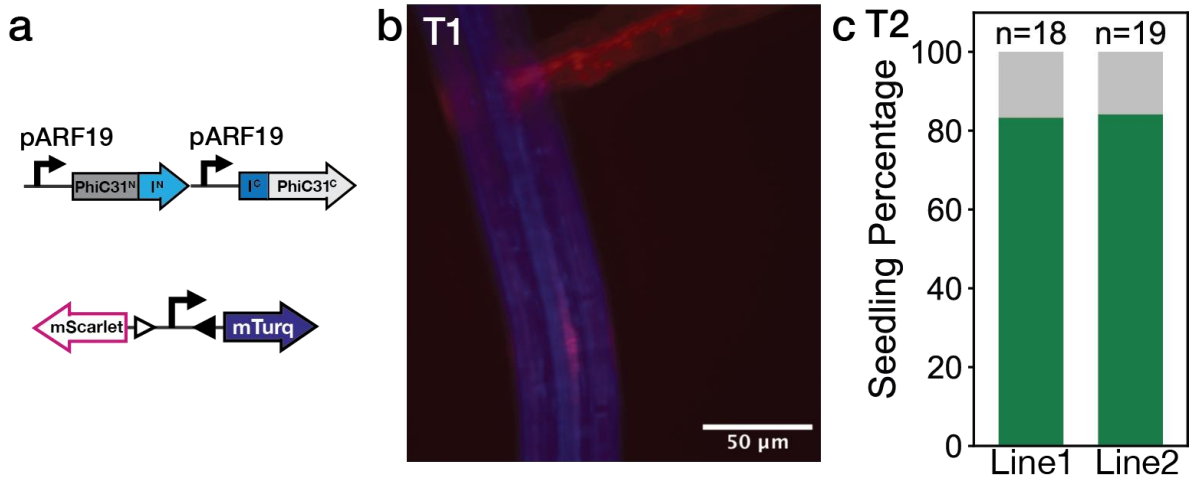
Supplementary Figure 6. The effect of tuning modifications differ between *N. benthamiana* and *Arabidopsis*: (**a-c**) Tuning was tested in *N. benthamiana* through tobacco injection. The integrase target switches from a Luc reporter to YFP and an RFP injection control was co-injected with each construct. The metric for level of switching is the ratio of YFP to RFP. Each tuning construct was injected into 3 leaves per experiment. Three punches were taken from each leaf and the resultant fluorescence was measured with a plate reader. Each point on the boxplot represents one leaf punch. Each box represents one of three replicate experiments performed for each construct. The box represents the middle two quartiles of the data and the whiskers represent the lowest and highest quartiles. The horizontal line within the box represents the median. The minimum and maximum values are shown by the termination of the whiskers, except in the case of some outliers which are shown as individual points. Tuning parts tested were (**a**) NLS and DST (*p* values left to right are 0, 0, 0.0009, 0.0097, 0, and 0). , (**b**) varied terminator (*p* value = 0), and (**c**) Ub degron (*p* value = 0). The data were tested for significance using a two-sidedn ANOVA and post-hoc Tukey's HSD test (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$). Source data are provided as a Source Data file.



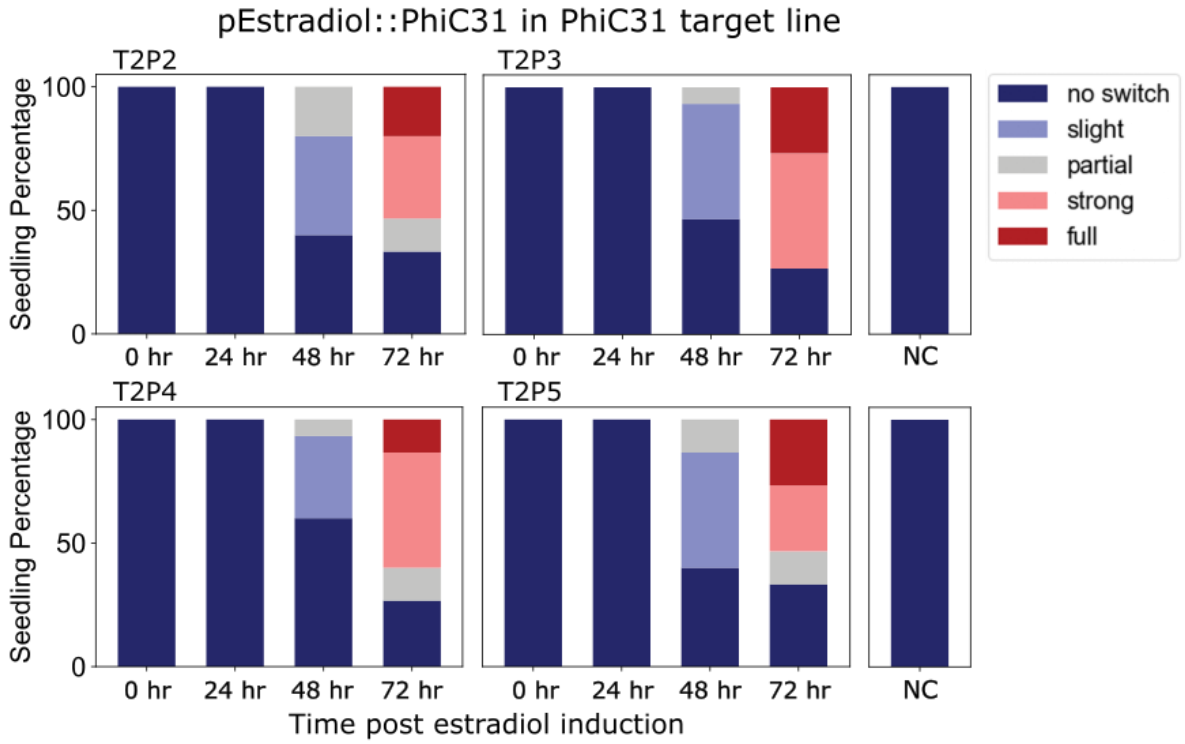
Supplementary Figure 7. Additional data for pARF19::PhiC31:DST and pGATA23::PhiC31:DST. (a) Phenotyping of T2 seedlings from T1 lines with pARF19::PhiC31:DST in PhiC31 target line. The graph corresponds to the percentage of seedlings in each of the defined phenotype categories, such as no switch corresponding to no mScarlet expression in the root, switch in LR only: mScarlet expression only in the lateral root, switch not exclusive to LR: mScarlet expression in the main root. The number of seedlings characterized for each construct is mentioned at the top of the bar in the graph. The lines used in Figure 4 have their name boxed. (b) and (c) Phenotype of T1, T2, T3 plants from pARF19::PhiC31:DST (b) and pGATA23::PhiC31:DST (c) constructs in PhiC31 target line. The pie charts are another representation of the previous phenotype bar graph representing the percentage of seedlings in each of the defined phenotype categories. From each generation, three seedlings with the LR-only switch phenotype were kept to generate the next generation. Source data are provided as a Source Data file.



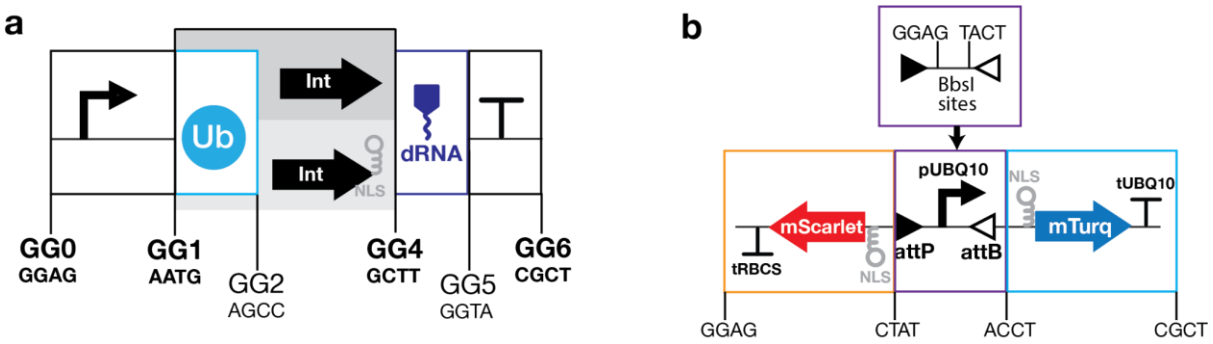
Supplementary Figure 8. Post-phenotyping selection of seedlings. (a) Process to determine post-characterization if a seedling is resistant to hygromycin. Post microscope characterization, seedlings were cut between the hypocotyl and the root. The aerial tissue (hypocotyl and cotyledons) were transferred to a hygromycin plate and grown for seven days. Resistant seedlings grew new roots. (b) and (c) Phenotype of T2 seedlings for pGATA23::Phic31 and pARF19::PhiC31:SAUR constructs in PhiC31 target line. Labels of the lines are at the bottom, T2PX corresponds to the name of the specific line. (b) is the phenotype characterization of one round of T2 seedlings for those lines (number at the top of the bars). (c) is the same data from b with only Hyg resistant seedlings included. The percentage of seedlings which are not resistant is represented between both graphs. The graph corresponds to the percentage of seedlings in each of the defined phenotype categories: no switch (light gray); LR only (green); not exclusive (dark gray). Source data are provided as a Source Data file.



Supplementary Figure 9. Split-intein system with ARF19 promoter. (a) Schematic of the ARF19 promoter driving the two split-intein parts and the PhiC31 integrase target composing this Arabidopsis line. **(b)** 10 T1 seedlings were imaged, here a representative image of a seedling with switch observed in a fully emerged lateral root and in an early stage of LR development. **(c)** Phenotyping of T2 seedlings from two T1 lines with pARF19 driving the two split-intein parts in PhiC31 target line. The graph corresponds to the percentage of seedlings in each of the defined phenotype categories, such as no switch corresponding to no mScarlet expression in the root (in gray), switch in LR only: mScarlet expression only in the lateral root (in green), switch not exclusive to LR: mScarlet expression in the main root (in black - none here). The number of seedlings characterized for each construct is mentioned at the top of the bar in the graph. Source data are provided as a Source Data file.



Supplementary Figure 10. Estradiol induction of PhiC31 integrase in T2 seedlings. T2 seeds were collected from T1 plants which were not treated with estradiol. Four T2 lines were tested: T2P2, T2P3, T2P4, and T2P5 with 15 seedlings screened for each line. Seedlings were imaged and classified based on switching level as per Supplementary Figure 5 at 0hr, 24hr, 48hr, and 72hr post-estradiol induction. Source data are provided as a Source Data file.



Supplementary Figure 11. Cloning strategy based on golden gate assembly. (a) Cloning strategy for the integrase construct with the BsaI spacer between each part specified. (b) Cloning strategy for the integrase construct. The central part with the integrase sites and promoter is constructed by golden gate assembly with BbsI enzyme to add the promoter to the synthetic fragment with the integrase sites only.

Supplementary reference

1. Bernabé-Orts, J. M. *et al.* A memory switch for plant synthetic biology based on the phage ϕC31 integration system. *Nucleic Acids Res.* **48**, 3379–3394 (2020).

nature communications

Article

https://doi.org/10.1038/s41467-024-53716-1

A history-dependent integrase recorder of plant gene expression with single-cell resolution

Received: 5 June 2024

Accepted: 18 October 2024

Published online: 30 October 2024

Check for updates

Cassandra J. Maranas[1], Wesley George[1], Sarah K. Scallon[1,3], Sydney VanGilder[1,3], Jennifer L. Nemhauser[1] & Sarah Guiziou[2]

During development, most cells experience a progressive restriction of fate that ultimately results in a fully differentiated mature state. Understanding more about the gene expression patterns that underlie developmental programs can inform engineering efforts for new or optimized forms. Here, we present a four-state integrase-based recorder of gene expression history and demonstrate its use in tracking gene expression events in *Arabidopsis thaliana* in two developmental contexts: lateral root initiation and stomatal differentiation. The recorder uses two serine integrases to mediate sequential DNA recombination events, resulting in step-wise, history-dependent switching between expression of fluorescent reporters. By using promoters that express at different times along each of the two differentiation pathways to drive integrase expression, we tie fluorescent status to an ordered progression of gene expression along the developmental trajectory. In one snapshot of a mature tissue, our recorder is able to reveal past gene expression with single cell resolution. In this way, we are able to capture heterogeneity in stomatal development, confirming the existence of two alternate paths of differentiation.

The sequential expression of genes during development progressively restricts and specifies a cell's fate. Many of the genes involved in cell differentiation processes have been identified across a host of model and non-model organisms, and, in many cases, these genes have been assembled into cell fate pathways. However, a full understanding of development will require combining the establishment of these genetic regulatory networks with information about the extent and effect of variation in how these networks are experienced by individual cells. The key limiting technology to achieve this aim is a method to monitor gene expression history with the single-cell resolution that maintains spatial information.

Existing single-cell RNA sequencing (scRNAseq) technologies provide a wealth of gene expression data in a variety of contexts[1]. However, scRNAseq provides only a snapshot of the cell's transcriptional state at one point in time, fails to capture the expression of lowly expressed genes, and requires the destruction of the sample, thus losing spatial resolution. Recent methods combining scRNAseq with additional techniques (such as metabolically labeling mRNAs[2,3], evaluating mRNA splice status[4], and pseudo-time analysis[5]) have been developed to allow users to infer some temporal information and have been applied to study cell specification in *Zea mays*[6] and *Arabidopsis thaliana*[7]. As a different approach to gain more information from scRNAseq, efforts to add spatial information include positional barcoding in tandem with next-generation sequencing in mammalian[8] and plant[9] systems; and in situ sequencing combined with in situ fluorescent hybridization in mammals[10] and plants[11]. Such approaches provide some spatial context, but no temporal information and are quite costly and complex. Alternatively, keeping spatial and temporal information, methods that utilize fluorescent expression as a marker of genetic events have been widely used for decades in a diverse array

[1]Department of Biology, University of Washington, Seattle, WA, USA. [2]Engineering Biology, Earlham Institute, Norwich, UK. [3]These authors contributed equally: Sarah K. Scallon, Sydney VanGilder. e-mail: sarah.guiziou@earlham.ac.uk

of organisms, including mouse models[12] and plants[13]. These approaches are limited in information due to the finite number of fluorescent tags that can be used at one time, the stress to the organism due to repeated irradiation, the long half-lives, the slow maturation time, and the photobleaching risk of fluorescent proteins. One previous study[14] implemented a fluorescence-based live imaging approach over multiple days for studying lateral root development in *Arabidopsis* but was able to track only six individuals due to the complexity of the imaging setup.

Recently, DNA-based recording systems have been developed and overcome some of the challenges of 'omic and microscopy-based technologies. Systems based on CRISPR-Cas9 mediated mutations and subsequent DNA sequencing enabled the recording of the Wnt signaling pathway in mammalian cells[15], although this approach lacked spatial resolution. DNA-based recording devices utilizing CRISPR have also been used for lineage tracing, for example, in reconstructing lineages in *Arabidopsis* tissues[16], zebrafish organs[17], and in the development of a lineage tracing mouse line[18]. Researchers have also utilized the DNA recombination abilities of integrases for similar DNA-based recording approaches. For example, the Cre-LoxP system has been employed to trace phloem and xylem vascular cell lineage in *Arabidopsis*[19]. Other work employs stochastic integrase recombination to generate barcodes which are then deconvoluted to construct cell lineages in mouse cell lines and *Drosophila melanogaster*[20]. Synthetic memory switches using tissue-specific integrase expression have recently been implemented in *Arabidopsis*[21,22]. More complex integrase circuits have been designed and implemented in bacterial cells[23–25], mammalian cells[26,27], and plant protoplasts[21,27].

Plants are particularly interesting organisms for studying the differentiation process as they continue to generate new organs throughout their lifetime. Decoding plant developmental programs can also inform the engineering of novel crops with enhanced properties to adapt to environmental stresses accelerated by climate change[28].

Here, we build on previous work using single serine integrase switches prototyped in *Nicotiana benthamiana*[29] and implemented in *Arabidopsis* for tracking the expression of genes underlying lateral root initiation[22]. Following the logic of a previously engineered recorder in bacterial cells[25], we engineer a four-state history-dependent integrase-based recorder of gene expression in *Arabidopsis*. We first validate the design of the recorder target using constitutive expression of two serine integrases: PhiC31 and Bxb1. We then test the efficacy of the circuit to record two well-known transcriptional trajectories, those occurring during the development of lateral roots and stomata. In both cases, we are able to capture the order of expression of two genes in individual cells. In addition, the recorder reveals heterogeneity in the gene expression history of mature stomata, highlighting the alternative differentiation pathways that they had experienced.

Results

An integrase-based history-dependent recorder functions in plants

Serine integrases mediate site-specific DNA inversions or excisions based on the presence and orientation of specific integrase sites (inversion when sites are in opposite orientation and excision when in the same orientation). Our integrase-based, history-dependent recorder of gene expression leverages these recombination abilities, with four possible fluorescent expression outputs, each indicative of a specific order of occurrence of signal inputs. The recorder is composed of two parts: a target (composed of integrase sites, fluorescent proteins, and gene regulatory elements) and an integrase construct (mediating integrase expression). We designed our target construct (Fig. 1a) for *Arabidopsis* with a strong constitutive p35S promoter[30] and three genes encoding fluorescent proteins: ER-localized mtagBFP2[31] (named BFP from here on), nuclear-localized mCherry[32] (named RFP from here on), and nuclear-localized NeonGreen[33] (named GFP from here on). The promoter and reporter genes are flanked by integrase sites for the orthogonal serine integrases: PhiC31 (Fig. 1a, gray triangles) and Bxb1 (Fig. 1a, black triangles). The target has four possible DNA states which result in different expression outputs: State 0 (BFP expression), State 1 (RFP expression), State 2 (GFP expression), and State α (no fluorescent expression). The switches between states are order-dependent and mediated by the PhiC31 and Bxb1 integrases.

The 'PhiC31 then Bxb1' target allows tracking of the PhiC31 then Bxb1 lineage, switching progressively from State 0 to State 2 when PhiC31 is expressed before Bxb1. In this target, State 0 represents the initial DNA state with no integrase expression. Expression of PhiC31 integrase mediates the switch to State 1 via inversion of the BFP-RFP cassette, resulting in a switch from BFP to RFP expression. Subsequent expression of Bxb1 integrase in these State 1 cells mediates the subsequent switch from State 1 to 2 via inversion of the p35S promoter and results in GFP expression. If Bxb1 is expressed 'out of order' prior to PhiC31 in a cell, it excises 75% of the target, including the p35S promoter, corresponding to a switch from state 0 to state α and resulting in a loss of fluorescence (Fig. 1a). The integrase switches are irreversible and heritable, so descendent cells inherit the target DNA state from the mother cell.

To characterize this target, we used a constitutive promoter to drive each integrase separately, the promoter for PROTEIN PHOSPHATASE 2 A SUBUNIT 3 (pPP2AA3, commonly used as a control for qPCR[34]). pPP2AA3 is not as strongly expressed as other commonly used plant constitutive promoters such as pUBQ10 or p35S, making it a better match for the expression level of most developmental genes. We first transformed the target in *Arabidopsis*, to obtain a stable target line. We confirmed strong expression of only BFP in both root and leaf tissue (Fig. 1b, left; Fig. 1c, left) in the target plant line selected for the rest of the experiments. We subsequently transformed this plant line with a PhiC31 constitutively expressed construct. We observed a switch in expression to RFP in both root and leaf tissue (Fig. 1b, middle), as expected. To test the State 1 to State 2 switch in isolation, we designed a switched target construct where the starting state of the target is the DNA State 1. We confirmed that this switched target shows strong RFP expression in the leaf and root (Supplementary Fig.1). When transformed with a pPP2AA3-expressed Bxb1 construct, this switched target exhibited strong constitutive GFP expression in both the leaf and root (Fig. 1b, right) confirming the switch from State 1 to State 2. Transformation of the State 0 target with pPP2AA3-expressed Bxb1 resulted in a strong reduction in BFP expression in the T1 generation (Fig. 1c), as expected. To confirm the switch from State 0 to State α, we genotyped this seedling and confirmed the presence of the target in a DNA state corresponding to the length of the State α target (Supplementary Fig. 2).

In addition to the 'PhiC31 then Bxb1' target, we also generated a 'Bxb1 then PhiC31' target (Supplementary Fig. 3) wherein the Bxb1 then PhiC31 lineage is tracked, switching from State 0 to 1 then 2 when Bxb1 then PhiC31 are expressed sequentially. We validated the initial switches for this target, showing Bxb1 expression causes a switch to RFP expression, therefore from State 0 to State 1 in the target, and PhiC31 expression to no fluorescent expression, therefore, to State α of the target (Supplementary Fig. 3). We also developed and made publicly available an alternate target for tracking the PhiC31 then Bxb1 lineage in which the reporters are expressed with pUBQ10 instead of p35S. We designed and implemented history-dependent integrase targets in *Arabidopsis* with identifiable fluorescent outputs for the different DNA states.

History-dependent recording of gene expression during lateral root initiation

For the next step in prototyping, we used well-characterized developmental promoters in the integrase construct to control the

https://doi.org/10.1038/s41467-024-53716-1

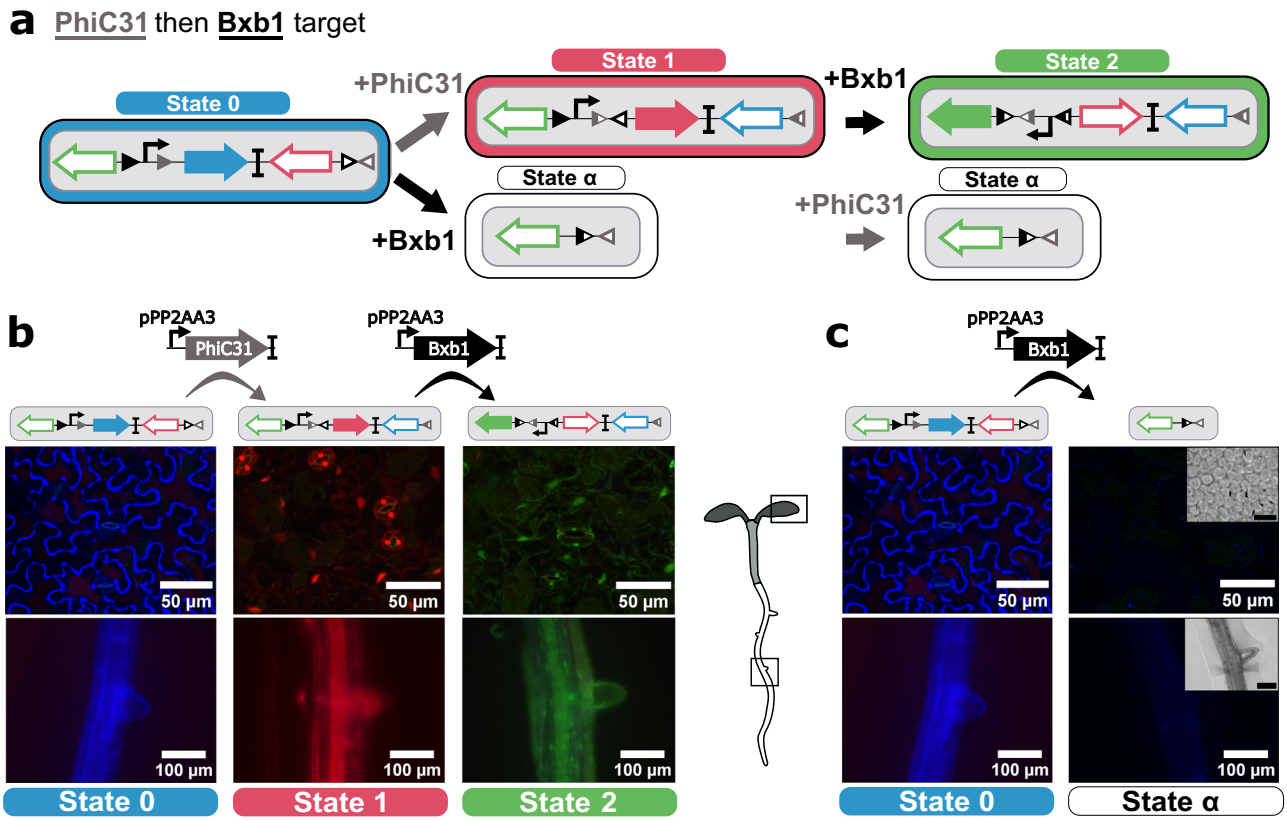


Fig. 1 | The history-dependent integrase circuit switches between fluorescent states based on the order of inputs. a Schematic of the integrase target circuit. The target consists of two sets of integrase sites for the PhiC31 and Bxb1 integrases; three genes for fluorescent proteins (mtagBFP2 (BFP), mCherry (RFP), NeonGreen (GFP)); and a p35S strong constitutive promoter for fluorescent expression. The initial state of the target (State 0) results in the expression of mtagBFP2. Upon expression of PhiC31, an inversion causes the target to switch to State 1 and results in the expression of RFP. Subsequent Bxb1 expression causes another inversion, flipping the promoter direction and mediating the switch to State 2, resulting in GFP expression. Alternatively, the addition of Bxb1 in a cell with the State 0 target results in a DNA excision, causing a switch to State α and loss of fluorescence. **b** The PhiC31-Bxb1 integrase order mediates a series of switches from State 0 to 1 to 2. (left) The State 0 target shows strong BFP expression in the leaf and the root. (middle) Constitutive PhiC31 expression mediates a switch of the target to State 1 and strong RFP fluorescence in leaf and root tissue. (right) Subsequent constitutive expression of Bxb1 mediates a switch from State 1 to 2 and results in GFP expression in the leaf and root. **c** Expression of Bxb1 prior to PhiC31 mediates an 'out of order' switch to State α and loss of fluorescence. (left) The initial target in State 0 shows strong BFP expression in the leaf and the root. (right) Constitutive expression of Bxb1 with the target in State 0 results in a switch to State α and a loss in fluorescence in the leaf and root. The root and leaf tissue of $n = 10$ seedlings from at least 2 T1 lines were characterized for each of the target States 0, 1, 2, and α. Black scale bars in the brightfield images correspond to 100 μm (root) and 50 μm (leaf). Source data are provided as a Source Data file.

expression of the integrases. To note, the integrase recorder captures gene expression in a digital manner, where each gene is either expressed (1) or not expressed (0). This is in contrast to biological systems where gene expression is analog. Gene expression levels are continuous with a cell-type-specific distribution of expression around a mean. As a result, expression levels that might be considered background or basal can be sufficient for integrase activity. In practice, in the integrase recorder, the digital state 0 corresponds to low gene expression, and the digital state 1 corresponds to expression above a specific threshold. This threshold is dependent on the level of integrase activity, and tuning parts can be used to adjust the switching threshold for the expression level of the gene of interest[22]. The switch threshold is affected by factors such as integrase efficiency, promoter strength, and other factors affecting expression, such as the insertion location of the integrase construct. The approach we took for deciding which tuning parts to use starts with generating the integrase construct without any tuning parts. Then, based on the switch result with this construct, we added a tuning part if needed, either a nuclear localization signal tag[22] (NLS, shown to increase integrase switching efficiency) if underswitched or an RNA destabilization tag from SMALL AUXIN UP-REGULATED RNA genes[35] (DST) if overswitched.

For an initial test of the recorder, we turned to the development of lateral roots, as many promoters for genes involved in lateral root cell fate specification have been already characterized using reporter constructs[36,37] and, for a few promoters, using integrase switches[22]. Beyond the depth of knowledge about lateral root developmental programming, it is also of high interest for engineering efforts aimed at increasing drought resilience[38]. Lateral roots develop from a small subset of xylem pole pericycle (XPP) cells, designated as 'founder cells'. These founder cells then undergo a series of proliferative, asymmetrical cell divisions, establishing the new lateral root[39]. Our lateral root recorder was designed to track the expression of two genes: *ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN 6* (*AHP6*) and *GATA TRANSCRIPTION FACTOR 23* (*GATA23*). AHP6 is a negative regulator of cytokinin signaling, and it plays an important role in guiding protoxylem formation[40] and orienting cell divisions during lateral root initiation. It is expressed in the xylem poles and XPP cells[41]. *GATA23* is strongly expressed in lateral root founder cells, setting into motion the series of cell divisions that form the lateral root[42]. For the integrase construct component of the lateral root recorder, we placed the first integrase (PhiC31) under the control of pAHP6 and the second integrase (Bxb1) under the control of pGATA23.

Table 1 | Summary overview of integrase lines and characterization categories

	As expected	Underswitched	Overswitched
pAHP6 single switch	RFP in xylem and lateral roots BFP otherwise	RFP in fewer cells than xylem and LR	RFP in more cells than xylem and LR
pGATA23 single switch	RFP in LR BFP otherwise	RFP in fewer cells than LR	RFP in more cells than LR
Lateral root history-dependent tracker	GFP in LR, RFP in xylem, BFP otherwise	GFP in fewer cells than LR and/or RFP in fewer cells than xylem	GFP in more cells than LR and/or RFP in more cells than xylem and LR
pSPCH single switch	RFP in guard cells and surrounding epidermal cells, BFP otherwise	RFP in fewer cells than guard cells and surrounding epidermal cells	RFP in more cells than guard cells and surrounding epidermal cells
pMUTE single switch	RFP in guard cellsBFP otherwise	RFP in fewer cells than guard cells	RFP in more cells than guard cells
Stomatal history-dependent tracker	GFP in guard cells RFP surrounding epidermal cells, BFP otherwise	GFP in fewer cells than guard cells and/or RFP in fewer cells than surrounding epidermal cells	GFP in more cells than guard cells and/or RFP in more cells than surrounding epidermal cells

First, we characterized the single switches for pAHP6 and pGATA23. To characterize the performance of the single switches, we categorized seedlings as (1) 'as expected', meaning the only RFP-expressing cells were those that expressed the recorded gene (or descended from such a cell); (2) over-switched, meaning additional cells were expressing RFP; and (3) under switched meaning that fewer cells were expressing RFP (see Table 1 for a streamlined overview of these categories for each single integrase switch and full recorder and Supplementary Table 1 for more details on each construct). To note, in our system overswitching corresponds to switching happening at basal expression levels, meaning the integrase switch threshold is too low. Underswitching corresponds to an integrase switch threshold that is too high resulting in fewer cells than expected undergoing the switch. For our single switches and recorder lines, we characterized all T2 seedlings without selection for integrase constructs or homozygosity. Therefore, we observed unswitched T2 seedlings with all cells expressing BFP. We attributed this to the loss of the integrase construct between generations, as the proportion of these unswitched seedlings is consistent with that for prior integrase switches[22] and is consistent with Mendelian heritability (around 25%)[43], so we omitted these seedlings from our characterization.

For pAHP6, we transformed the 'PhiC31 then Bxb1' target line with a single integrase (PhiC31) under the control of the pAHP6 promoter, which we called the pAHP6 single integrase switch, as it only uses the switch from State 0 to State 1 of the target (Fig. 2a). In any cell where *AHP6* has been expressed, that is, cells in the xylem poles and XPP, we expected to see a switch from BFP to RFP expression, indicating a switch of the target from State 0 to State 1. Because the switch is heritable, we would predict that in a plant with this single pAHP6 switch, all xylem pole cells, XPP cells, and their descendants (e.g., lateral roots) would express RFP, representing the 'as expected' switch pattern (Table 1). Because no switch was observed initially in the xylem poles or XPP in any T1 seedlings, we added an NLS onto PhiC31 to increase its DNA recombination activity, resulting in four out of seven T1 seedlings with the 'as expected' switch pattern (Supplementary Table 2). Using this integrase construct with the NLS, we were able to consistently generate 'as expected' T2 seedlings (Fig. 2a). In the best performing pAHP6 single integrase switch T1 line (with NLS), 78% of T2 seedlings showed the expected switch pattern and the worst performing line had 18% with the expected pattern (Fig. 2a and Supplementary Fig. 4a). Part of this seedling to seedling variability is due to the fact that we characterized a mix of hetero and homozygous seedlings for the integrase construct. There were also seedlings consistently in the underswitched category (the line with the highest prevalence had 59% underswitched), in this case meaning the switch occurred in the lateral root but not in the xylem poles or XPP cells. This indicates stronger expression of *AHP6* in lateral root precursor cells, consistent with the fact that *AHP6* expression is induced by the plant hormone auxin[41], which is found at high levels in founder cells during early lateral root initiation[44].

GATA23 has been previously used to drive a single PhiC31 integrase switch as a recorder of lateral root initiation[22]. To characterize a Bxb1-mediated single switch, we transformed the 'Bxb1 then PhiC31' target (Supplementary Fig. 3) with a single Bxb1 construct driven by pGATA23 such that Bxb1 mediates the switch from State 0 to 1 (Fig. 2b). As only one switch of the target is utilized, we refer to this as the pGATA23 single integrase switch. Consistent with our knowledge of *GATA23* expression and previously characterized *GATA23* integrase switch[22], we expected to observe RFP expression only in cells within the lateral root, constituting the 'as expected' switch category (Fig. 2b and Table 1). The pGATA23::Bxb1 single integrase switch resulted in T1 seedlings with RFP expression in cells outside the lateral root (overswitched) (Supplementary Table 2). To achieve switch specificity consistently in the T1 and T2 generations, we added a DST to the Bxb1 construct[22]. This pGATA23::Bxb1-DST single integrase switch resulted in five out of eight T1 seedlings with the 'as expected' switch pattern (Supplementary Table 2). From these 'as expected' T1s, we obtained three T1 lines out of five with T2 seedlings behaving 'as expected' (ranging from 15% to 71% of seedlings) (Fig. 2b and Supplementary Fig. 4). The majority of the T2 seedlings in most of the lines were overswitched. The T1 line with the lowest proportion of overswitching was 29%, and the highest line was 100% overswitched. The full T2 characterization of the single switches for the lateral root genes can be found in Supplementary Fig. 4a.

Next, we generated the full lateral root recorder, building the dual integrase construct with pAHP6-driven PhiC31 and pGATA23-driven Bxb1 and transforming it into the 'PhiC31 then Bxb1' target line. We added an NLS to the pAHP6::PhiC31 construct and no tag for the pGATA23::Bxb1 construct. According to the known expression patterns of *AHP6* and *GATA23*, we expected to observe xylem pole and XPP cells expressing RFP, lateral root cells expressing GFP, and the rest of the cells in the root should remain in their initial, BFP-expressing state (Fig. 2c). We were able to consistently generate seedlings with the 'as expected' pattern (Fig. 2d). In the T1 generation (Supplementary Table 3), four out of nine seedlings were over switched with the rest being some combination of 'as expected' and under switched. In the context of the full lateral root recorder, over-switched refers to seedlings with PhiC31 and/or Bxb1 over-switching, such as with GFP-expressing cells outside the lateral root and/or RFP-expressing cells outside the xylem poles and XPP (Supplementary Fig. 5).

We characterized T1 lines from the 'as expected' and under-switched T1s because subsequent generations are likely to have stronger integrase expression due to higher copy number in homozygous individuals. In the T2 generation, an average of 23% of seedlings screened per line showed the expected output. Those that deviated from the expected phenotype consisted primarily of overswitched seedlings (Supplementary Fig. 6a). In contrast to the pGATA23 single switch where the pGATA23::Bxb1 construct led to overswitching in every T1 seedling (Supplementary Table 2), in the full recorder we were able to achieve switching specific to *GATA23* expression without

https://doi.org/10.1038/s41467-024-53716-1

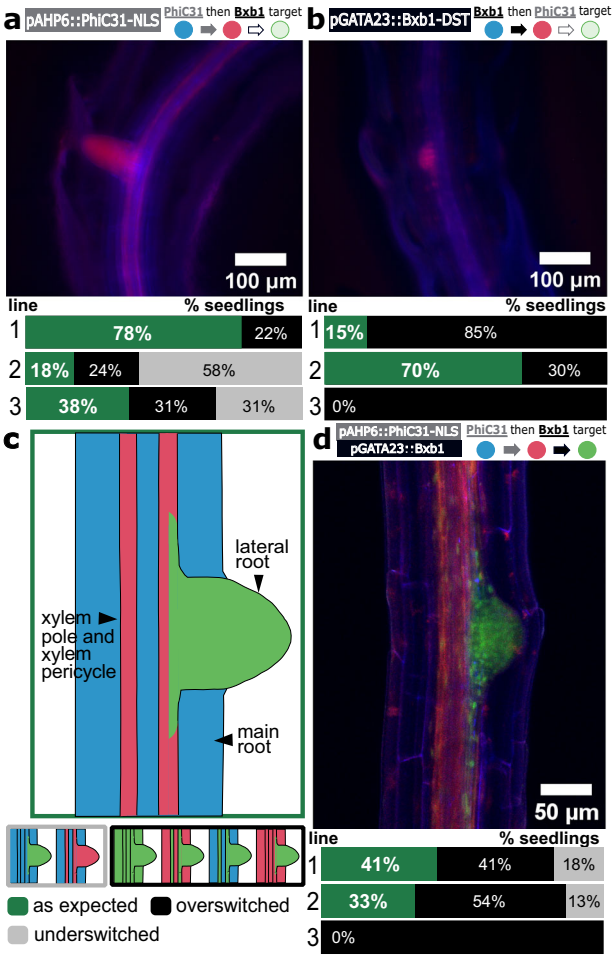


Fig. 2 | Application of the history-dependent recorder in lateral roots. a A pAHP6 single integrase switch. Plants carrying the 'PhiC31 then Bxb1' target were transformed with pAHP6::PhiC31-NLS. A representative image of a T2 seedling is shown with the switch characterization for three representative lines (line 1: T2P5 ($n = 17$), 2: T2P2 ($n = 16$), 3: T2P7 ($n = 9$)), see Supplementary Fig. 4a), where n represents the number of seedlings characterized in each line, omitting seedlings with no switching. The bar graphs represent the proportion of seedlings in each T1 line, showing the 'as expected', overswitched, and underswitched switch pattern (as defined in Table 1). **b** A pGATA23 single integrase switch. Plants carrying the 'Bxb1 then PhiC31' target (Supplementary Fig. 3) were transformed with pGATA23::Bxb1-DST. A representative image of a T2 seedling is shown with the switch characterization for three representative lines (line 1: T2P5 ($n = 17$), 2: T2P3 ($n = 20$), 3:T2P1 ($n = 16$), see Supplementary Fig. 4a). The bar graphs represent the proportion of seedlings in each T1 line showing the 'as expected', over switched, and under switched switch pattern (as defined in Table 1). **c** Predicted recorder output. Xylem pole and xylem pole pericycle cells should express RFP (State 1, red), the cells in the lateral root should express GFP (State 2, green), and any other cell should remain in the initial BFP-expressing (State 0, blue). Below are schematic representations of the underswitched (left) and overswitched (right) patterns (see Supplementary Fig. 5 for more detail). **d** Lateral root recorder output. T2 seedlings from 5 lines, which came from T1 plants with an 'as expected' or under-switched recorder output, were screened. A representative image for one of these T2 seedlings is shown, as well as a characterization of the recorder output for three representative T1 lines (line 1: T2P1 ($n = 17$), 2: T2P3 ($n = 15$), 3: T2P9 ($n = 13$), see Supplementary Fig. 6a). The bar graphs represent the proportion of seedlings in each T1 line showing the 'as expected', over switched, and under switched switch pattern (as defined in Table 1). Source data are provided as a Source Data file.

addition of a DST to Bxb1. In our dual integrase construct, the Bxb1 transcriptional unit is immediately downstream of the PhiC31 unit, and this construct architecture has been shown to result in reduced expression of the downstream gene[45], possibly due to transcription-induced DNA supercoiling[46]. We observed negligible intra-plant variation in the fluorescent output of the *AHP6* single integrase switch, the *GATA23* single integrase switch, and the full lateral root recorder. Lateral roots of similar developmental stages on the same seedling showed the same fluorescent output, indicating low levels of intrinsic noise in the integrase switches. Schematics of all the observed switch patterns can be seen in Supplementary Fig. 6. Overall, we implemented a recorder tracking the expression patterns of AHP6 and GATA23 during lateral root initiation and consistently generated T2 seedlings with the expected switch pattern.

History-dependent tracking of gene expression during stomatal development

We next applied our history-dependent recorder to a different well-characterized differentiation program: stomatal development. Stomata are each composed of two guard cells which together form a small pore in the leaf surface. These pores are required for efficient gas exchange and are a target of interest for engineering increased carbon capture with higher water use efficiency[47].

Each stoma descends from a proliferating meristemoid cell on the leaf epidermis, which serves as the precursor of the stomatal cell lineage. Stomata develop following a stereotypical progression wherein the meristemoid undergoes some number of asymmetrical cell divisions before dividing symmetrically exactly once into the two guard cells that comprise the stoma. Many of the genes involved in this progression from meristemoid to guard cell have been extensively studied[48]. We focused on recording the expression of two genes in the context of stomatal development: *SPEECHLESS* (*SPCH*), which enables the asymmetrical cell divisions of the meristemoid, and *MUTE*, which is expressed later and is essential to trigger the single symmetric division into two guard cells[49]. We, therefore, used pSPCH to drive the expression of the first integrase (PhiC31) and pMUTE to drive the expression of the second integrase (Bxb1). The final recorder line contains this integrase construct and the 'PhiC31 then Bxb1' target.

We first separately tested the individual integrase switches with pSPCH driving PhiC31 and pMUTE driving Bxb1. For the pSPCH single switch, we transformed the pSPCH-driven PhiC31 construct into the 'PhiC31 then Bxb1' target line, expecting to observe a switch from BFP to RFP expression (State 0 to 1) in the guard cells plus any surrounding epidermal cells that resulted from asymmetric meristemoid divisions (the 'as expected' switch category) (Fig. 3a and Table 1). In this pSPCH single integrase switch, over-switched means that additional leaf epidermal cells were expressing RFP and under-switched means that fewer cells were expressing RFP (Table 1). In the T1 generation, four out of the ten seedlings showed 'as expected' switching (Supplementary Table 4). In the T2 generation, we obtained up to 50% of seedlings per line switching 'as expected' with a high variability between lines. Most of the T2 seedlings showed an overswitched phenotype, with some or all leaf epidermal cells not in contact with stomata expressing RFP. The most over-switched T1 line had 100% over-switched seedlings, while the lowest rate of overswitching observed was 14%. One line had 64% of the seedlings under-switched, with only the guard cells of the stomata expressing RFP (Fig. 3a and Supplementary Fig. 4b).

To generate the pMUTE single switch, we built a pMUTE-driven Bxb1 construct and transformed it into the 'Bxb1 then PhiC31' target (Supplementary Fig. 3) line. We characterized the lines similarly to the pSPCH single switch lines in the T1 and T2 generations, with the 'as expected' phenotype being RFP expression in the guard cells of the stomata only (Table 1). In the T1 generation, five out of ten seedlings showed the 'as expected' switch pattern (Supplementary Table 4). In the T2 generation, we consistently obtained a majority of seedlings expressing RFP only in the guard cells (73% and 53%) (Fig. 3b). Every seedling that did not fit this pattern showed an over-switched phenotype, with RFP expression in additional leaf epidermal cells (Fig. 3b and Supplementary Fig. 4b). For both the single integrase switches, we observed negligible variation in the switch pattern across locations on

https://doi.org/10.1038/s41467-024-53716-1

the leaf, with leaf epidermal cells throughout showing a consistent switch output, whether it was under switched, over switched, or 'as expected'.

We then generated the full stomatal recorder by building the dual integrase construct with pSPCH-driven PhiC31 and pMUTE-driven Bxb1 and transforming it into the 'PhiC31 then Bxb1' target line. We expected to see expression of GFP in the guard cells of the stomata, expression of RFP in any surrounding epidermal cells that resulted from asymmetric meristemoid divisions, and expression of BFP in the remaining leaf epidermal cells that are not involved in stomatal differentiation. Of the nine T1 seedlings generated, six showed GFP expression specific to stomata, and the lines from these six seedlings were characterized (Supplementary Table 3). In five out of the six lines, we obtained seedlings with the expected phenotype (Fig. 3d and Supplementary Fig. 6b), with a proportion of up to 50%, and a low of 14% 'as expected' seedlings per line, with three out of six lines having an 'as expected' proportion greater than 33%. The majority of the not 'as expected' seedlings were over-switched, meaning they were expressing GFP in cells other than the guard cells and/or expressing RFP in excess epidermal cells. The over-switched category accounted for the majority of seedlings in most of the stomatal recorder T1 lines, including 100% of seedlings in one line, with a low of 35% and a median of 68% of seedlings. Most of the overswitched seedlings (median of 67%) showed both PhiC31 and Bxb1 overswitching (Supplementary Fig. 7). A small proportion of seedlings (15% or less) showed PhiC31 underswitching, meaning the guard cells were expressing GFP and the rest of the epidermal cells were expressing BFP, but no cells were actively expressing RFP. For one line, we obtained only seedlings with non-specific switching (both PhiC31 and Bxb1 over switching), meaning some or all non-stomata epidermal cells were expressing GFP (switched to State 2) and/or some or all epidermal cells not in contact with at least one stoma were expressing RFP (switched to DNA State 1). A schematic summary of all the observed recorder outputs is in Supplementary Fig. 7. As highlighted previously, some of the variability between seedlings is due to differences in the zygosity of the integrase construct. For the stomatal recorder, we performed characterization of homozygous and non-homozygous T3 lines and found the homozygous line to be significantly more overswitched, likely due to higher integrase expression (Supplementary Fig. 8). Despite the observed variability, we consistently generated seedlings in the 'as expected' category for the stomata recorder and the *SPCH* and *MUTE* single integrase switches.

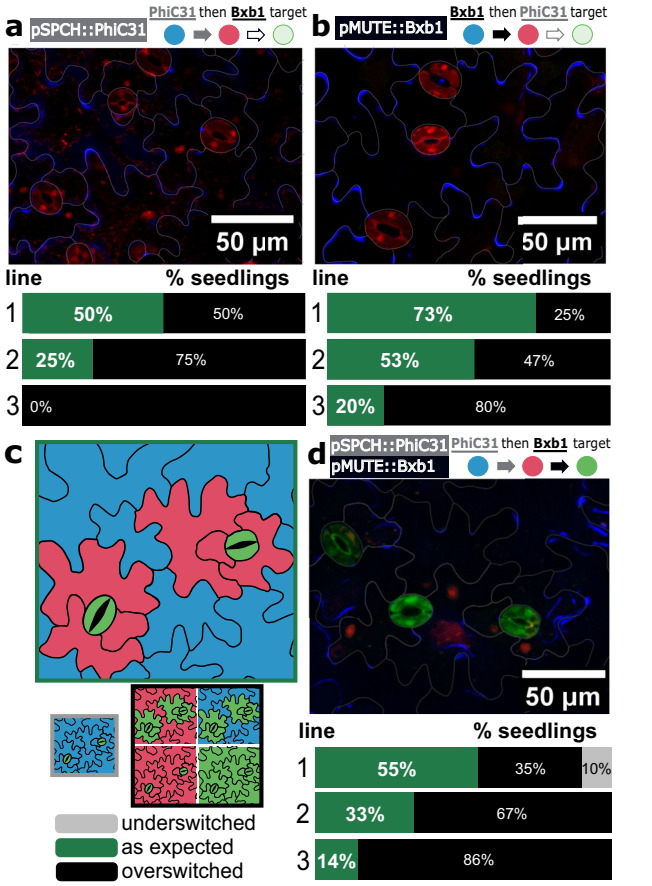


Fig. 3 | Application of the history-dependent recorder during stomata development. a A pSPCH single integrase switch. Plants carrying the 'PhiC31 then Bxb1' target were transformed with pSPCH::PhiC31. A representative image of a T2 seedling is shown, as well as the switch characterization for three representative lines (line 1: T2P3 ($n = 12$), 2: T2P8 ($n = 14$), 3: T2P1 ($n = 14$), see Supplementary Fig. 4b) where n represents the number of seedlings characterized in each line omitting seedlings with no switching. **b** A pMUTE single integrase switch. Plants carrying the target were transformed with pMUTE::Bxb1. A representative image of a T2 seedling is shown with the switch characterization for three representative lines (line 1: T2P2 ($n = 15$), 2: T2P7 ($n = 15$), 3: T2P1 ($n = 15$), see Supplementary Fig. 4b). **c** Prediction of stomatal recorder output. Guard cells are predicted to be in State 2 and expressing GFP (State 2, green), surrounding epidermal cells that are the result of meristemoid division should be expressing RFP (State 1, red), and any other cell should sustain the initial BFP expression (State 0, blue) Below are schematic representations of the underswitched (left) and overswitched (right) patterns (see Supplementary Fig. 7 for more detail). **d** Experimental stomatal development recorder output. The 'PhiC31 then Bxb1' target is transformed with the dual stomatal integrase construct. One representative image of a T2 seedling matching the expected output is shown. The recorder performance is shown for three representative lines (line 1: T2P1 ($n = 20$), 2: T2P4 ($n = 18$), 3: T2P2 ($n = 20$), see Supplementary Fig. 6b), with percentages indicating the proportion of seedlings matching the expected recorder output and n representing the number of seedlings characterized in each T1 line omitting seedlings with no switching. Source data are provided as a Source Data file.

The history-dependent recorder can identify cells that have undergone an alternate developmental path

For the recorder seedlings with an 'as expected' phenotype, the majority of the stomata expressed GFP as expected; however, there was a subset of stomata that did not express any fluorescent protein (Fig. 4a). In the design of the recorder, a cell which has no fluorescence indicates that the target has been switched to State α due to Bxb1 recombination before PhiC31 recombination. This result suggests the possibility that there are two stomatal populations that differ in how likely they are to switch to State α versus to proceed through the series of switches to State 2. We did not observe such variation in the single *SPCH* and *MUTE* integrase switches, so we hypothesized that this discrepancy was due to differences in the relative timing of *SPCH* and *MUTE* expression. A closer relative timing of expression would increase the chances of Bxb1 mediating the switch to State α, while a longer timing would give PhiC31 more of a head start, increasing the likelihood of the cell switching to State 1 and then State 2. Indeed, temporal logic gates based on similar integrase designs have been employed previously in *E. coli* to reveal the timing between events[50].

It has been shown that the timing between *SPCH* and *MUTE* expression differs between the development of two types of stomata: anisocytic (A) and nonanisocytic (NA)[51]. During stomatal differentiation, *SPCH* initiates, and *MUTE* terminates the asymmetrical divisions of the meristemoid[49]. Closer timing of expression between *SPCH* and *MUTE* permits fewer asymmetrical divisions, resulting in an NA stoma, whose development can include zero, one, or two asymmetric meristemoid divisions. In contrast, a longer delay between *SPCH* and *MUTE* expression allows the full three asymmetrical cell divisions needed for development of an A stoma (Fig. 4b). As a result, A stomata, the majority of stomata in most cases, are easily identified because the series of three asymmetrical divisions results in a stoma

https://doi.org/10.1038/s41467-024-53716-1

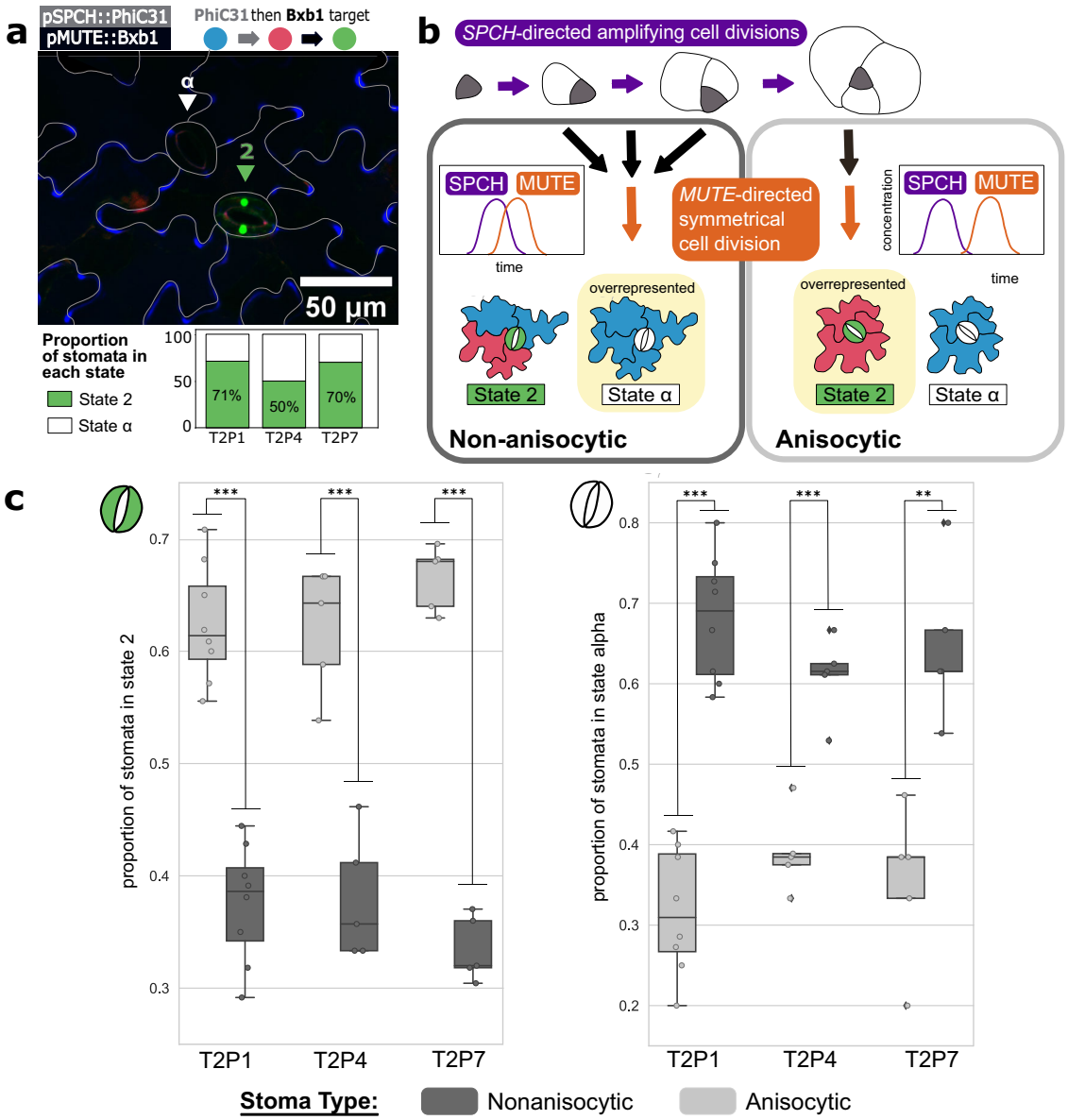


Fig. 4 | The history-dependent recorder reveals two distinct stomatal populations. a State 2 and State α stomata. A confocal image shows two stomata: one in State 2 (green triangle) and one in State α (white triangle). Below the image, the total proportion of stomata in each state is shown in a barplot, with each bar representing one of the three best-performing stomata recorder lines (labeled on the *x*-axis). $n = 257$, $n = 172$, and $n = 174$ stomata were characterized for T2P1, T2P4, and T2P7. **b** Alternate paths of stomatal development. The stereotypical understanding of stomatal development follows an anisocytic (A) scheme, with the meristemoid dividing asymmetrically three times before dividing into two guard cells. Non-anisocytic (NA) stomata divide asymmetrically zero, one, or two times before the terminal division. Recorder output is predicted to differ between stomata types due to varied timing of integrase onset. An A stomata is surrounded by exactly three cells of unequal size, while an NA stomata can have two, four, five, or three equally sized neighbor cells. **c** Characterization of recorder state in A vs. NA stomata. The boxes represent the proportion of each stoma type (anisocytic in light gray and non-anisocytic in dark gray) within each of the three stomata recorder lines (*x*-axis) that are in State 2 (left) and State α (right). Each box represents the middle quartiles of the respective dataset, and the whiskers represent the highest and lowest quartiles. The center of each box is the median. Each point represents the proportion of stomata in each state for one T2 seedling. 10 seedlings were counted per line. For each seedling, a minimum of 15 of each stomata type was counted (at least 30 total stomata per seedling). To evaluate the statistical significance between the result for A and NA stomata populations, a two-sided Student's *t* test was performed (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). From left to right in the plots, *p*-values are $p = 0$, $p = 0.0001$, $p = 0$, $p = 0$, $p = 0.0001$, $p = 0.001$. Source data are provided as a Source Data file.

surrounded by exactly three cells of unequal size, whereas an NA stomata can have two, four, or five neighbor cells or three neighbor cells of roughly equal size[52] (Fig. 4b).

We hypothesized these two developmental lineages, with their differences in relative gene expression timing, led to two stomata populations with different proportions of cells in each recorder state. We predicted that the State 2 stomata should be overrepresented in the A population and State α stomata should be overrepresented in the NA population. To test this hypothesis, we counted 25 stomata per T2 seedling in 10 individuals. This was repeated for the three best-performing stomata recorder lines, only characterizing seedlings matching the 'as expected' phenotype (Supplementary Fig. 7). For each counted stoma, we categorized it as A or NA and State 2 or State α (corresponding to GFP expression or no expression, respectively). We then calculated for both stomatal populations, the percentage of stomata in State 2 versus in State α. For A stomata, an average of 65% across all three lines were in State 2 and 35% in State α. In contrast, for NA stomata, an average of 33% across the three lines were in State 2,

with 67% in State α (Fig. 4c). Therefore, as predicted, State 2 stomata were enriched in the A population compared to NA where a majority of stomata were in State α. In theory, the State α stomata could be a result of stomatal differentiation, which skips *SPCH* expression entirely. If this were the case, the pSPCH single integrase switch result (Fig. 3a) would include unswitched stomata at a comparable rate, however we observed that all stomata in the 'as expected' pSPCH single integrase switch lines expressed RFP and therefore include *SPCH* expression in their lineage. This result supports our hypothesis that the disparity in recorder output was split probabilistically based on the extent of asymmetrical cell divisions prior to *MUTE* expression. Overall, we showed that the state of the recorder in a given stoma correlated strongly with its developmental history and stomatal type. The successful encoding of differences in the relative timing of *SPCH* and *MUTE* expression during alternative developmental trajectories demonstrates the potential utility of this strategy in connecting patterns of gene expression to differentiation in many other contexts.

Discussion

Developmental biology has made enormous strides in identifying gene regulatory networks that specify cell fate at a population scale (i.e., for the 'average' cell with a given identity). There is increasing interest in understanding the history of individual cells, including quantifying cell-to-cell variation in gene expression dynamics and context-specific, cell non-autonomous influences on cell fate. We have built a tool that can help with these efforts. Our integrase-based history-dependent recorders worked in two distinct developmental contexts: the initiation of a new root from the pericycle cells of the primary root and the differentiation of stomata in the leaf epidermis. In both cases, we could readily visualize cells in State 0, State 1, and State 2, and we consistently generated seedlings whose fluorescent expression matched our expectations based on known gene expression patterns. In tracking stomatal development, we also detected guard cells with no fluorescence, indicating a switch to State α. This finding supports prior evidence for the existence of two populations of stomata. Our results supported the hypothesis that the difference between stomata in State 2 and State α reflect differences in the relative time between the expression of the two genes of interest. Collectively, our work demonstrates the utility of the recorder for revealing variation in the timing of gene expression events even in seemingly uniform cell differentiation processes.

For a targeted study of the gene expression patterns underlying cell development paths of interest, our recorder provides an accessible and powerful methodology. Our recorder was sensitive enough to capture a disparity in gene expression timing in a very well-studied context, so it should prove even more beneficial for those working in less-studied cell development contexts. This sensitivity, combined with the strong constitutive, sustained expression of the reporters, allows the recorder to act as an amplifier of the gene expression signal, which should prove useful for studying low-expressing genes not suited for traditional transcriptional reporters and enables evaluation of developmental events over timescales far exceeding the degradation time of the fluorescent proteins. While we applied our recorder design to well-characterized differentiation processes, it can be adapted to processes wherein the patterns of gene expression are not well understood. This would necessitate additional optimization of the integrase switch specificity to reduce plant-to-plant and cell-to-cell variability combined with gene expression validation using a transcriptional reporter or spatial transcriptomics data. With this extra optimization and validation step, our recorder offers unique advantages for less well-studied genes, including sustained fluorescent expression to enable recording even when the exact timing of expression is unknown; high sensitivity to enable recording even at low levels of expression; the ability to discern variation from cell to cell; and the tracing of the cell lineage which had expressed the gene of interest. Our recorder enables a snapshot readout of each cell's gene expression history with only standard molecular biology techniques and access to a fluorescent microscope to take images at just a single or small number of time points.

While much information could be gleaned from the recorders analyzed here, future engineering efforts, especially those focusing on less well-characterized pathways, may require a systematic approach to reduce variability. This issue is a persistent one for engineering in plants. Line-to-line variation is likely due to the random insertion location of transgenes leading to different levels of gene expression. This issue could be addressed using targeted DNA transformation techniques. Variability in performance between seedlings carrying the same transgene insertions could be addressed with iterative rounds of optimization with our previously characterized tuning parts[22]. In the future, information from a thorough characterization of integrase efficiency combined with knowledge of gene expression strength could inform the choice of tuning parts, expediting the process of generating robust and specific integrase switches. In this study, variability is exacerbated by varied integrase expression levels among seedlings within T1 lines due to differences in the zygosity of the integrase construct. For future applications, characterization and optimization of T3 homozygous lines will be required.

A four-state recorder like the designs used here can capture the history of expression for two genes. Using our current target design, researchers could generate multiple recorders within the same cell specification pathway to capture the expression patterns for more gene combinations. Moreover, the current recorder infrastructure could be adapted following existing integrase target designs to generate a target with a higher number of inputs[25]. The design capabilities of integrase circuits[23] combined with the large number of characterized orthogonal integrases[27,53] should enable the construction of circuits with higher numbers of outputs. The possible scope is limited in principle only by the availability of spectrally distinct reporters, and this limit could be overcome with multiplexing techniques[54] or switching to DNA barcodes[20].

In theory, our recorder designs should be transferable to any plant that can be transformed, for applications such as those facilitating evo-devo studies of key developmental processes such as stomata. Our recorder can be applied in plants that undergo different paths of stomatal development, such as in amphibious plants which, largely skip the *SPCH*-driven, asymmetrical cell divisions[55], and grasses which, have a completely different arrangement of the subsidiary cells surrounding the stomata compared to *Arabidopsis*[56]. In addition, integrase switches can be adapted to allow the study of essential genes in non-essential organs[57]. Porting this system into crop plants could enable optimization and quantification of traits associated with climate resilience, facilitating marker-assisted breeding efforts. Our history-dependent recorder should also function across different non-plant organisms and should be applicable to challenges such as studying differentiation paths involving lowly-expressed genes or evaluating cell-to-cell variability in gene expression trajectory. In applications beyond tracking gene expression, integrase circuits like the ones described here could be modified for engineering applications. If the reporters were replaced with developmental genes and integrase expression was tied to chemical signals, specific stress responses, or specific tissues—interactive synthetic genetic programming of development should be entirely feasible for applications such as programming climate-resilient root architectures. The relative ease and accessibility of our method should enable rapid prototyping and optimization for use in elucidating and reprogramming cell states.

Methods

Construction of plasmids

Our cloning strategy was based on Golden Gate assembly using an appropriate spacer and BsaI-HFv2 (NEB) and BbsI-HF (NEB) as the

restriction enzymes (Supplementary Fig. 9). Details on all primers and constructs can be found in Supplementary Datas 1 and 2 respectively. Candidate promoter sequences (SPCH: AT5G53210, MUTE: AT3G06120, AHP6: AT1G80100, GATA23: AT5G26930) were amplified from Col-0 genomic DNA (primer list available in Supplementary Data 1) to add specific Golden Gate spacers. After PCR purification, each level 0 promoter sequence was cloned using a Zero Blunt PCR Cloning Kit (ThermoFisher Scientific). The PhiC31 integrase sequence was a gift from the Orzeaz lab. The Bxb1 sequence was a gift from the Bonnet lab. Integrases were amplified using primers with golden gate compatible spacers to generate level 0 integrase parts. Constitutive plant promoters and terminators were purchased from Addgene as part of the MoClo Toolbox for Plants[58]. A mutated version of the pPP2AA3 promoter without BsaI sites was ordered from Twist Bioscience. Other level 0 fragments were ordered from IDT as Gblocks: the fluorescent proteins NeonGreen-NLS and mtagBFP2-ER and combinations of integrase sites and restriction sites for the construction of various integrase targets. The mCherry sequence was amplified from a constitutively expressed mCherry construct (donated by Jennifer Brophy). For the integrase target level 0 sequences, the p35S or pUBQ10 promoter was added by Golden Gate using BbsI sites (Supplementary Fig. 9).

The construction of level 1 single integrase constructs (constitutive, stomata, and lateral root specific) was performed via Golden Gate reaction in the modified pGreenII-Hygr vector containing compatible Golden Gate sites[58]. The construction of integrase targets was performed with the same methods in a modified pGreenII-Kan vector. Construction of level 2 integrase constructs expressing both PhiC31 and Bxb1 for the full recorders was performed by amplifying completed level 1 integrase constructs using primers with Golden Gate compatible spacers, then performing Golden Gate reactions in the modified pGreenII-Hygr vector containing compatible Golden Gate sites. The addition of an NLS or DST to level 1 and level 2 integrase constructs was performed using PCR-mediated site-directed mutagenesis (NEB Q5 Site-Directed Mutagenesis, Cat #E0554S). Primers for mutagenesis were designed with NEBasechanger (primer list available in Supplementary Data 2). Enzymes for Golden Gate assembly were purchased from New England Biolabs (NEB, Ipswich, MA, USA). PCR was performed using a 2X Q5 PCR master mix (NEB) and a GoTaq master mix for colony PCR and genotyping (Promega, Madison, WI, USA). Primers were purchased from IDT (Coralville, IA, USA), and DNA fragments from Twist Bioscience (San Francisco, CA, USA) or IDT. Plasmid extraction and DNA purification were performed using Monarch kits (NEB). Sequences were verified with Sanger sequencing by Azenta Life Sciences or Primordium Labs for whole plasmid sequencing. Chemically-competent cultures of the *E. coli* strain DH5αZ1 (laciq, PN25-tetR, SpR, deoR, supE44, Delta(lacZYA-argFV169), Phi80 lacZ-DeltaM15, hsdR17(rK −, mK +), recA1, endA1, gyrA96, thi-1, relA1) were transformed with plasmid constructs containing kanamycin resistance. Transformed *E. coli* was grown in LB media (LB broth, Miller) with kanamycin (Millipore Sigma, 50 μg/mL).

Plant growth conditions

Arabidopsis seedlings were sown in 0.5 x Linsmaier and Skoog nutrient medium (LS) (Caisson Laboratories) and 0.8% w/v Phyto agar (PlantMedia/bioWORLD), stratified at 4 °C for 2 days, and grown in constant light at 22 °C.

Construction and selection of transgenic *Arabidopsis* lines

Agrobacterium tumefaciens strain GV3101 was transformed by electroporation, and subsequently grown in LB media with rifampin (Millipore Sigma, 50 μg/mL), gentamicin (Millipore Sigma, 50 μg/mL), and kanamycin (Millipore Sigma, 50 μg/mL). The floral dip method[59] was used to generate integrase target lines in Col-0, and then transformation of the integrase constructs into the target lines was performed via an adjusted dual dipping approach wherein plants are dipped on two occasions approximately a week apart (with the first dipping occurring when the stems are around 3 inches tall) and the agrobacterium were incubated with MMA (10 mM MgCl2, 10 mM MES pH 5.6, 100 μM acetosyringone) for an hour prior to dipping, as acetosyringone is known to improve agrobacterium-mediated transformation efficiency in *Arabidopsis* explants[60]. For T1 selection: 150 mg of T1 seeds (~2500 seeds) were sterilized using 70% ethanol and 0.05% Triton-X-100 and then washed using 95% ethanol. Seeds were resuspended in 0.1% agarose and spread onto 0.5X LS Phyto selection plates, using 25 μg/mL of kanamycin for target lines or 25 μg/mL kanamycin and 25 μg/mL hygromycin for lines with both the integrase construct and the target. The plates were stratified at 4 °C for 48 h and then grown for 7, 8 days. To select transformants, tall seedlings with long roots and vibrant green color were picked from the selection plate with sterilized tweezers and transferred to a new 0.5X LS Phyto agar plate for characterization.

Plant genotyping

For the genotyping of plants, leaves from approximately 20 day old plants were used. For plant DNA extraction, leaf sections approximately 1 cm^2 in area were frozen on dry ice and ground for 1 min into a fine powder. 10 μm 0.5 M NaOH was added to each sample before boiling them for 1 min and then adding 100 μL neutralization solution (1 part 1 M Tris-OH/HCl, pH 8.0 and 4 parts 0.1 M TE, pH 8.0). Genotyping PCRs were set up using the GoTaq master mix (Promega, Madison, WI, USA), making sure to gently stir the DNA extraction with the pipette tip before pipetting into the reaction.

Hygromycin selection to determine integrase zygosity in T3 lines

T3 seeds were plated on 0.5 X LS and 0.8% w/v Phyto agar with 50 μg/mL hygromycin. Seeds were then stratified at 4 °C for 2 days, and then grown in constant light at 22 °C for 10 days before hygromycin resistance was evaluated. Seedlings with substantial root growth were characterized as resistant, while those with extremely stunted root growth were considered non-resistant.

Imaging of reporter and integrase lines

Initial screening of root and leaf tissue was performed using a Leica Biosystems microscope (model: DMI 3000) with a 10x objective for imaging roots and a 40x objective for imaging leaf tissue. Further characterization and images in the manuscript were obtained via confocal imaging.

Confocal imaging of the seedling root and leaf tissue was performed using a Nikon A1R HD25 laser scanning confocal microscope with a Plan Apochromat Lambda 20x objective. Three channels were used: 561 laser and 578-623 detector for RFP imaging; 488 laser and 503–545 detector for GFP imaging; 405 laser and 419–476 detector for BFP imaging. For each image, a Z-stack was recorded. All images were processed using ImageJ FIJI.

Confocal imaging of lateral root recorder was done after characterization of T1 seedlings and subsequent generation of T2 seeds. Ten- day old T2 seedlings were mounted on slides with water and with Parafilm edges to prevent the coverslip from pressing on the root. The main root of each seedling was scanned for pre-emergence lateral roots to image. The gain was set at 75, 45, and 40 respectively for the BFP, RFP, and GFP channels. The laser power was set to 3, 10, and 4, respectively.

Confocal imaging of stomata single switches and recorder was performed after the characterization of T1 seedlings and subsequent generation of T2 seeds. The first true leaves from 12-16-day-old T2 seedlings were cut off and mounted on slides with 50% glycerol for imaging of the abaxial side of the leaf. The edges of the coverslip were painted with clear nail polish to prevent movement of the sample

during imaging. The leaves were scanned to locate the flattest areas to take images. The gain was set at 75, 50, and 35, respectively, for the BFP, RFP, and GFP channels. The laser power was set to 2, 12, and 3, respectively.

Characterization of the history-dependent recorder in Arabidopsis transgenic lines

T1 seedlings for each line were grown 4–5 days after transformant selection. Each selected seedling was imaged at 10x magnification using an epifluorescence microscope (Leica Biosystems, model: DMI 3000) using the GFP (exposure 400 ms, gain 2), RFP (exposure 500 ms, gain 2), and CFP (exposure 300 ms, gain 2) channels. Selected T1 seedlings were then transferred to soil, and at maturation, T2 seeds were collected. For later generations, seedlings were sterilized similarly to T1s, stratified, plated on an LS agar plate, and grown for either 10 days (for characterizing roots) or 12–16 days (for characterizing leaves). Target characterization was done using the epifluorescence microscope as for T1. For the target lines, the seedlings with the highest level of mtagBFP2 expression in the root (or, in the case of the pre-switched target, the highest level of mCherry expression) were selected and transferred to soil to generate T2 seeds. The line with the most consistently bright seedlings was maintained as the target line for each integrase target and used for all later transformations of integrase constructs.

For the constitutive integrase constructs in a target line, around 5-10 T1 seedlings were analyzed per construct, and the ones that displayed any level of switching were transplanted to the soil for characterization in the T2 generation, where around 20 seedlings were characterized per line. Representative root images were taken using the RFP, GFP, and CFP channels and merged for final images. Representative leaf images were generated using a Nikon A1R HD25 laser scanning confocal microscope as described above.

Characterization in the context of lateral root development

For the pAHP6 and pGATA23 single integrase switches, at least 15 T1 seedlings were analyzed per construct. Each seedling was categorized into one of three classes based on the specificity of switching: 'as expected' (RFP expression in xylem pole, XPP, and lateral root cells); over switched (RFP expression in additional cells); under switched (RFP expression in fewer cells). The seedlings in the no switch category represented around 25% of seedlings and were omitted from analysis as this proportion is consistent with a loss of the integrase construct through gene segregation[21]. Representative images were taken in the epifluorescence microscope using the GFP, RFP, and CFP channels and merged for final images. For the *AHP6* switch, T1 seedlings, which showed switching in the xylem poles, xylem pericycle, and lateral root cells, were transplanted to the soil along with T1 seedlings which showed switching only in the lateral root. For the pGATA23 single integrase switch, only T1 seedlings that showed switching specific to the lateral root were transplanted for future T2 characterization. For each selected T1 line, 15-20 T2 seedlings were characterized, categorizing each as follows: 'as expected' (RFP expression in lateral root cells); overswitched (RFP expression in additional cells); underswitched (RFP expression in fewer cells) (Table 1).

For the full lateral root recorder to track both *AHP6* and *GATA23* expression, nine T1 seedlings were analyzed and those whose switch pattern was either 'as expected' or some form of underswitched (Supplementary Fig. 6), were transplanted to soil for future T2 characterization. From each T1 line, 15–20 seedlings were characterized and categorized based on switch pattern (Supplementary Fig. 5). For Fig. 2, these categories were consolidated into under-switched, 'as expected', and over-switched categories (Table 1) as for the single switches, with any seedlings with Bxb1 and/or PhiC31 over switching fitting into the over switched category and any seedlings with Bxb1 and/or PhiC31 under switching fitting into the under switched category (Supplementary Fig. 5). Representative images for the full lateral root recorder were taken with the Nikon A1R HD25 laser scanning confocal microscope.

Characterization in the context of stomata development

For the single pSPCH and pMUTE T1 single integrase switches, leaves from at least 10 T1 seedlings were analyzed per construct. Due to the chlorophyll autofluorescence causing difficulties in imaging mCherry in the epifluorescence microscope, this characterization was done using the confocal microscope. For the pSPCH single integrase switch, T1 seedlings with switching in only stomata and some epidermal cells that border stomata were transplanted for T2 characterization. T1 seedlings with switching limited to only stomata were also transplanted. For the pMUTE single integrase switch, seedlings with switching only in the stomata were transplanted for T2 characterization. For the full stomata recorder for tracking *SPCH* and *MUTE* expression, nine T1 seedlings (Supplementary Table 3) were characterized, and any with GFP expression specific to stomata were transplanted for T2 characterization.

For the pSPCH and pMUTE single integrase switch T1 lines, leaves from 15-20 seedlings were characterized as described for T1. The pSPCH single integrase switch seedlings were sorted into one of three categories: 'as expected' (RFP expression in guard cells and surrounding epidermal cells); overswitched (RFP expression in additional cells); underswitched (RFP expression in fewer cells). For the pMUTE single integrase switch, the categories are as follows: 'as expected' (RFP expression in only guard cells); overswitched (RFP expression in additional cells); underswitched (RFP expression in fewer cells). For the full recorder the categories are as seen in Supplementary Fig. 7. For Fig. 3 these categories were consolidated into the same underswitched, 'as expected', and overswitched categories as for the single switches, with seedlings with either or both PhiC31 and Bxb1 overswitching fitting into the overswitched category and seedlings with either PhiC31 or Bxb1 underswitching fitting into the underswitched category. Representative images for the single switches and the full recorder were taken with the Nikon A1R HD25 laser scanning confocal microscope.

For quantifying the relationship between recorder output and stomata type, leaves from seedlings were stained with 2 mg/mL Calcofluor White for 30 minutes (Millipore Sigma Cat #18909) to better visualize the cell boundaries. To screen random stomata all across the leaf, we started from the top left of each leaf and shifted the viewing frame of the microscope across the leaf incrementally and then down at least a full frame before moving back across the leaf. In each frame, the centermost stomata were chosen to be categorized. This categorization involved determining the type of stomata based on the pattern of surrounding cells and then noting the recorder output (GFP expressing or no fluorescence). In the case that the type of stomata was not able to be determined, the frame was skipped. This process continued for each seedling until at least 15 A and 15 NA stomata were counted.

Analysis

For each single switch construct, the percentage of seedlings in each of the three categories was plotted in a bar plot with the number of seedlings tested mentioned at the top of the bar. For the full recorders, the percentage of seedlings in each of the expanded seven categories (Supplementary Figs. 5, 7) was plotted but consolidated into the same three categories for display in Figs. 2, 3.

For quantifying the relationship between stomata type and recorder output, the percentage of State 2 and State α stomata of each type was plotted. For comparisons between A and NA stomata, a student's *t* test was performed to evaluate statistical significance.

Python data analysis script which includes statistical tests and plotting functions was run in version 3.9.1 and with the following package dependencies: pandas (version 1.5.3), scipy.stats (version 1.10.0), matplotlib.pyplot (version 3.6.3), matplotlib.colors (version

3.6.3), scikit_posthocs (version 0.21), seaborn (v0.12.0), and numpy (version 1.24.2).

All images taken during seedling characterization were opened and processed using the ImageJ FIJI program (version 1.53c). Each.tif image file contained the images of a seedling's GFP, RFP, and CFP channels. tif files were processed with adjustments to the color lookup table, brightness, and contrast of each channel (GFP: Green, Min 200, Max 2500) (RFP: Red, Min: 300, Max: 3000) (CFP: Blue, Min: 100, Max: 3000).

Reporting summary

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

Data availability

Fully annotated sequences of each of the DNA constructs used in this study are available at Benchling [https://benchling.com/cjmaranas/f_/P8Coz4Vw-maranas-et-al-2024-a-history-dependent-integrase-recorder-of-plant-gene-expression-with-single-cell-resolution-/]. Constructs and seed lines used in this study are available by request from J.L.N. (jn7@uw.edu; please expect a response within 3 weeks). DNA constructs can be purchased from Addgene. Arabidopsis seeds are available through the Arabidopsis Biological Resource Center (ABRC). Addgene deposit numbers and ABRC stock numbers are listed in Supplementary Data 2. Source data are provided with this paper and through Figshare [https://doi.org/10.6084/m9.figshare.26824825]. Source data are provided in this paper.

Code availability

The plotting scripts used in this study have been deposited in Zendo [https://doi.org/10.5281/zenodo.13864568].

References

1. Jovic, D. et al. Single-cell RNA sequencing technologies and applications: A brief overview. *Clin. Transl. Med.* **12**, e694 (2022).
2. Bhat, P. et al. SLAMseq resolves the kinetics of maternal and zygotic gene expression during early zebrafish embryogenesis. *Cell Rep.* **42**, 112070 (2023).
3. Cao, J., Zhou, W., Steemers, F., Trapnell, C. & Shendure, J. Sci-fate characterizes the dynamics of gene expression in single cells. *Nat. Biotechnol.* **38**, 980–988 (2020).
4. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
5. Cannoodt, R., Saelens, W. & Saeys, Y. Computational methods for trajectory inference from single-cell transcriptomics. *Eur. J. Immunol.* **46**, 2496–2506 (2016).
6. Satterlee, J. W., Strable, J. & Scanlon, M. J. Plant stem-cell organization and differentiation at single-cell resolution. *Proc. Natl. Acad. Sci. USA* **117**, 33689–33699 (2020).
7. Shahan, R. et al. A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Dev. Cell* **57**, 543–560 (2022).
8. Rodriques, S. G. et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
9. Liu, C. et al. A spatiotemporal atlas of organogenesis in the development of orchid flowers. *Nucleic Acids Res.* **50**, 9724–9737 (2022).
10. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, https://doi.org/10.1126/science.aat5691 (2018).
11. Nobori, T., Oliva, M., Lister, R. & Ecker, J. R. Multiplexed single-cell 3D spatial gene expression analysis in plant tissue using PHYTOMap. *Nat. Plants* **9**, 1026–1033 (2023).
12. Doh, S. J. et al. Fluorescent reporter transgenic mice for in vivo live imaging of angiogenesis and lymphangiogenesis. *Angiogenesis* **21**, 677–698 (2018).
13. Sadoine, M. et al. Designs, applications, and limitations of genetically encoded fluorescent sensors to explore plant biology. *Plant Physiol.* **187**, 485–503 (2021).
14. von Wangenheim, D. et al. Rules and self-organizing properties of post-embryonic plant organ cell division patterns. *Curr. Biol.* **26**, 439–449 (2016).
15. Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, https://doi.org/10.1126/science.aap8992 (2018).
16. Donà, M. et al. A versatile CRISPR-based system for lineage tracing in living plants. *Plant J.* **115**, 1169–1184 (2023).
17. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
18. Bowling, S. et al. An engineered CRISPR-Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells. *Cell* **181**, 1410–1422 (2020).
19. Shi, D., Lebovka, I., López-Salmerón, V., Sanchez, P. & Greb, T. Bifacial cambium stem cells generate xylem and phloem during radial plant growth. *Development* **146**, https://doi.org/10.1242/dev.171355 (2019).
20. Chow, K. H. K. et al. Imaging cell lineage with a synthetic digital recording system. *Science* **372**, https://doi.org/10.1126/science.abb3099 (2021).
21. Lloyd, J. P. B. et al. Synthetic memory circuits for stable cell reprogramming in plants. *Nat. Biotechnol.* **40**, 1862–1872 (2022).
22. Guiziou, S., Maranas, C. J., Chu, J. C. & Nemhauser, J. L. An integrase toolbox to record gene-expression during plant development. *Nat. Commun.* **14**, 1844 (2023).
23. Guiziou, S., Mayonove, P. & Bonnet, J. Hierarchical composition of reliable recombinase logic devices. *Nat. Commun.* **10**, 456 (2019).
24. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science* **353**, aad8559 (2016).
25. Zúñiga, A. et al. Rational programming of history-dependent logic in cellular populations. *Nat. Commun.* **11**, 4758 (2020).
26. Weinberg, B. H. et al. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* **35**, 453–462 (2017).
27. Gomide, M. S. et al. Genetic switches designed for eukaryotic cells and controlled by serine integrases. *Commun. Biol.* **3**, 255 (2020).
28. Gray, S. B. & Brady, S. M. Plant developmental responses to climate change. *Dev. Biol.* **419**, 64–77 (2016).
29. Bernabé-Orts, J. M. et al. A memory switch for plant synthetic biology based on the phage ϕC31 integration system. *Nucleic Acids Res.* **48**, 3379–3394 (2020).
30. Holtorf, S., Apel, K. & Bohlmann, H. Comparison of different constitutive and inducible promoters for the overexpression of transgenes in Arabidopsis thaliana. *Plant Mol. Biol.* **29**, 637–646 (1995).
31. Subach, O. M., Cranfill, P. J., Davidson, M. W. & Verkhusha, V. V. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PLoS ONE* **6**, e28674 (2011).
32. Hu, G.-Y. et al. Optimizing the protein fluorescence reporting system for somatic embryogenesis regeneration screening and visual labeling of functional genes in cotton. *Front. Plant Sci.* **12**, 825212 (2021).
33. Shaner, N. C. et al. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **10**, 407–409 (2013).
34. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. Genome-wide identification and testing of superior reference

genes for transcript normalization in Arabidopsis. *Plant Physiol.* **139**, 5–17 (2005).
35. Newman, T. C., Ohme-Takagi, M., Taylor, C. B. & Green, P. J. DST sequences, highly conserved among plant SAUR genes, target reporter transcripts for rapid decay in tobacco. *Plant Cell* **5**, 701–714 (1993).
36. Péret, B. et al. Arabidopsis lateral root development: an emerging story. *Trends Plant Sci.* **14**, 399–408 (2009).
37. Gala, H. P. et al. A single-cell view of the transcriptome during lateral root initiation in Arabidopsis thaliana. *Plant Cell* **33**, 2197–2220 (2021).
38. Brophy, J. A. N. Toward synthetic plant development. *Plant Physiol.* **188**, 738–748 (2022).
39. Banda, J. et al. Lateral root formation in arabidopsis: A well-ordered LRexit. *Trends Plant Sci.* **24**, 826–839 (2019).
40. Mähönen, A. P. et al. Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science* **311**, 94–98 (2006).
41. Bishopp, A. et al. A mutually inhibitory interaction between auxin and cytokinin specifies vascular pattern in roots. *Curr. Biol.* **21**, 917–926 (2011).
42. De Rybel, B. et al. A novel aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity. *Curr. Biol.* **20**, 1697–1706 (2010).
43. Passricha, N., Saifi, S., Khatodia, S. & Tuteja, N. Assessing zygosity in progeny of transgenic plants: current methods and perspectives. *J. Biol. Methods* **3**, e46 (2016).
44. Dubrovsky, J. G. et al. Auxin acts as a local morphogenetic trigger to specify lateral root founder cells. *Proc. Natl. Acad. Sci. USA* **105**, 8790–8794 (2008).
45. Kallam, K. et al. Tunable control of insect pheromone biosynthesis in Nicotiana benthamiana. *Plant Biotechnol. J.* **21**, 1440–1453 (2023).
46. Johnstone, C. P. & Galloway, K. E. Supercoiling-mediated feedback rapidly couples and tunes transcription. *Cell Rep.* **41**, 111492 (2022).
47. Nguyen, T. B.-A., Lefoulon, C., Nguyen, T.-H., Blatt, M. R. & Carroll, W. Engineering stomata for enhanced carbon capture and water-use efficiency. *Trends Plant Sci.* **28**, 1290–1309 (2023).
48. Adrian, J. et al. Transcriptome dynamics of the stomatal lineage: birth, amplification, and termination of a self-renewing population. *Dev. Cell* **33**, 107–118 (2015).
49. Peterson, K. M., Rychel, A. L. & Torii, K. U. Out of the mouths of plants: The molecular basis of the evolution and diversity of stomatal development. *Plant Cell* **22**, 296–306 (2012).
50. Hsiao, V., Hori, Y., Rothemund, P. W. & Murray, R. M. A population-based temporal logic gate for timing and recording chemical events. *Mol. Syst. Biol.* **12**, 869 (2016).
51. Doll, Y., Koga, H. & Tsukaya, H. Experimental validation of the mechanism of stomatal development diversification. *J. Exp. Bot.* **74**, 5667–5681 (2023).
52. Pillitteri, L. J. & Dong, J. Stomatal development in Arabidopsis. *Arabidopsis Book* **11**, e0162 (2013).
53. Yang, L. et al. Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014).
54. Linghu, C. et al. Spatial multiplexing of fluorescent reporters for imaging signaling network dynamics. *Cell* **183**, 1682–1698 (2020).
55. Doll, Y., Koga, H. & Tsukaya, H. The diversity of stomatal development regulation in Callitriche is related to the intrageneric diversity in lifestyles. *Proc. Natl. Acad. Sci. USA* **118**, e2026351118 (2021).
56. Nunes, T. D. G., Zhang, D. & Raissig, M. T. Form, development and function of grass stomata. *Plant J.* **101**, 780–799 (2020).
57. Leydon, A. R. et al. A conserved function of corepressors is to nucleate assembly of the transcriptional preinitiation complex. Preprint at https://doi.org/10.1101/2024.04.01.587599 (2024).
58. Engler, C. et al. A golden gate modular cloning toolbox for plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
59. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* **16**, 735–743 (1998).
60. Sheikholeslam, S. N. & Weeks, D. P. Acetosyringone promotes high efficiency transformation of Arabidopsis thaliana explants by Agrobacterium tumefaciens. *Plant Mol. Biol.* **8**, 291–298 (1987).

Acknowledgements

We thank Janet Solano Sanchez, Ben Downing, and Dr. Alexander Leydon, as well as other members of the Nemhauser, Imaizumi, Di Stilio, Steinbrenner, and Patron groups, for feedback and discussion. We thank Eric Yang for developing and gifting the pPP2AA3 promoter, the Orzáez lab for the PhiC31 sequence, the Brophy lab for the mCherry sequence, and members of the Bonnet lab for sending us the Bxb1 integrase plasmid. We thank Jonah C. Chu for the construction of early versions of the targets and integrase constructs. We thank Keiko Torii for the discussions and advice about stomatal development. This work was supported by grants from the National Institutes of Health (grant no. GM107084, J.L.N.), the National Science Foundation (grant no. IOS-1546873, J.L.N.), and the Howard Hughes Medical Institute Faculty Scholars Program. In addition, support to S.G. was provided by the UK Research and Innovation (UKRI) Biotechnology and Biological Sciences Research Council (BBSRC) via the Earlham Institute Core Capability Grant (grant no. BB/CCG2220/1, S.G.).

Author contributions

C.M., S.G., and J.L.N. designed the project. C.M., W.G., and S.G. designed the constructs. C.M., W.G., and S.S. generated the constructs. C.M., S.S., and S.V. performed and analyzed the constitutive integrase switching experiments, C.M. performed and analyzed the lateral root and stomata differentiation experiments, C.M., S.G., and J.L.N. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

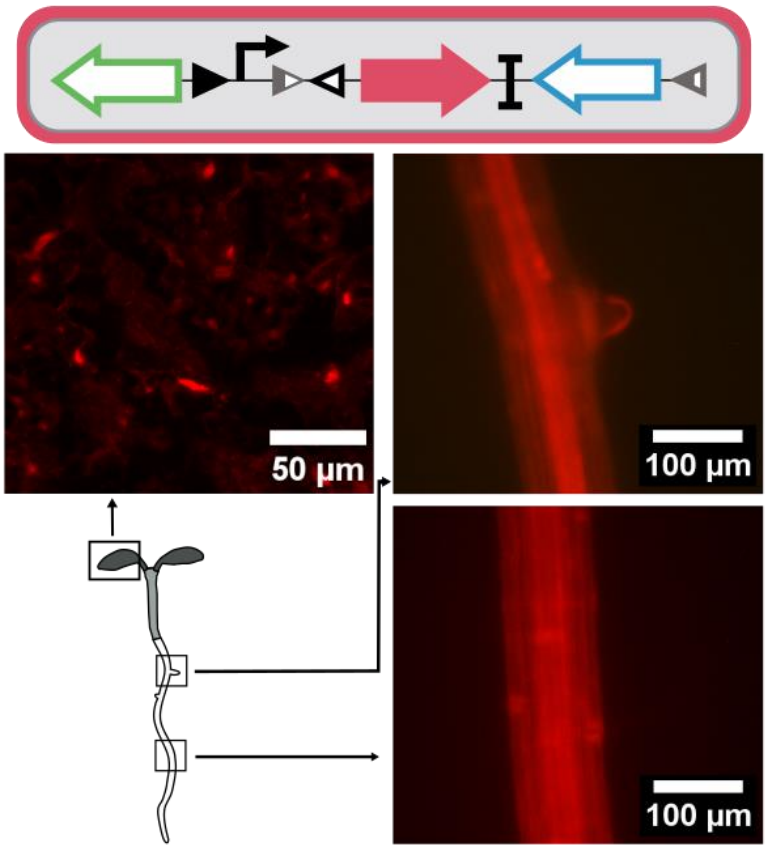
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53716-1.

Correspondence and requests for materials should be addressed to Sarah Guiziou.

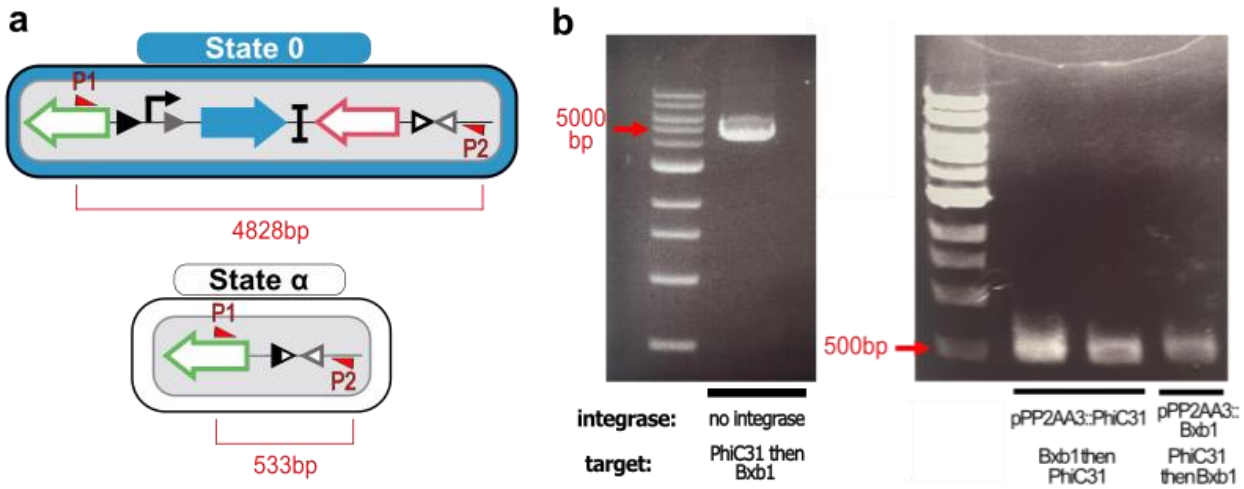
Peer review information *Nature Communications* thanks Diego Orzáez, who co-reviewed with Marta Vazquez-Vilar, Klaus Palme, and the other anonymous reviewer(s) for their contribution to the peer review of this work. [A peer review file is available].

A history-dependent integrase recorder of plant gene expression with single cell resolution

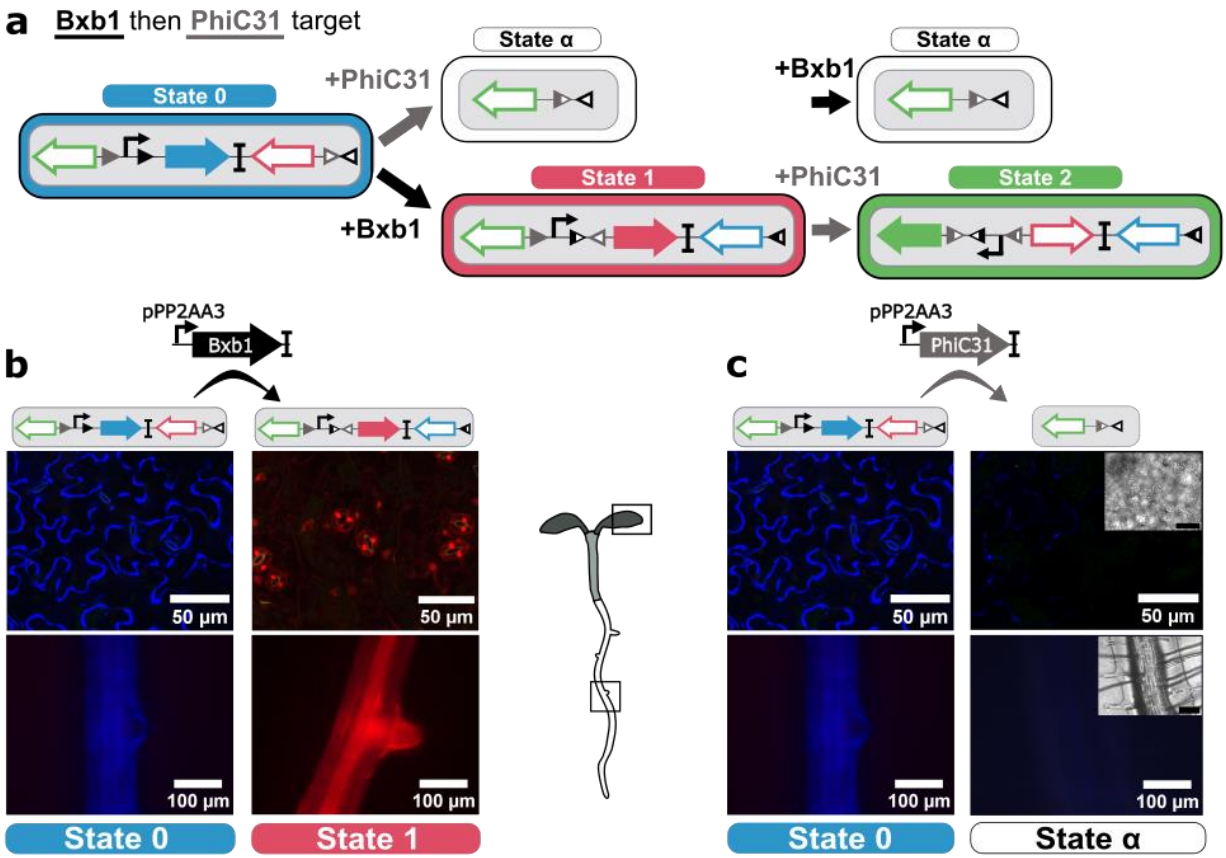
Maranas *et al.*



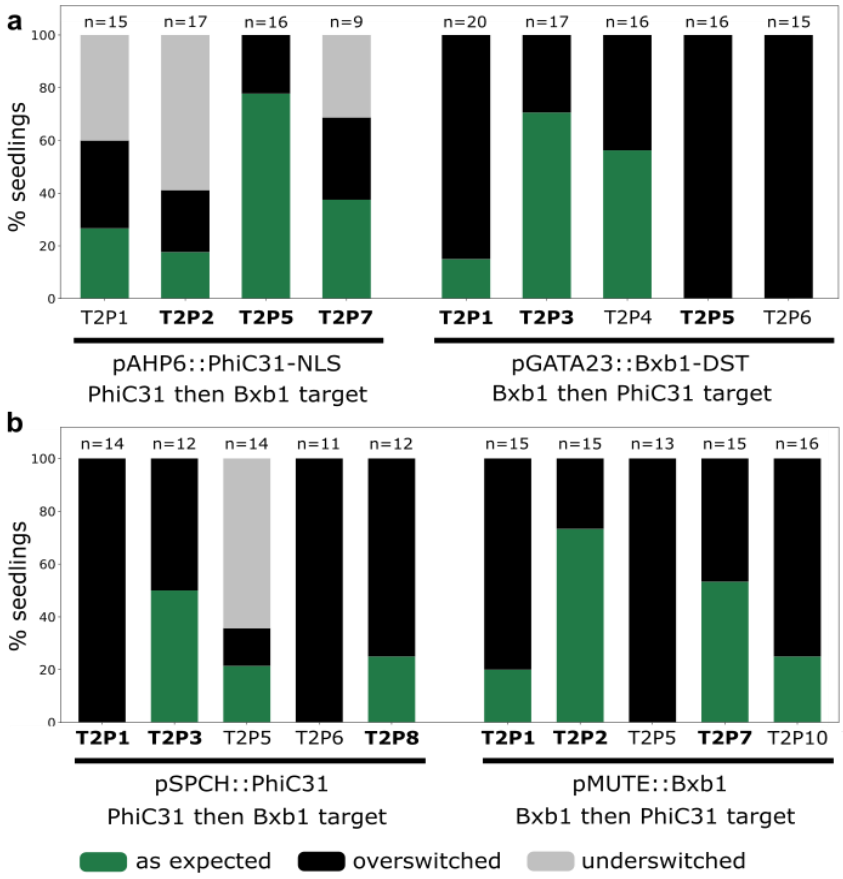
Supplementary Figure 1. Preswitched 'PhiC31 then Bxb1' integrase target circuit shows strong fluorescent expression in the leaf and root. The preswitched target was designed to test the State 1 to State 2 switch in the 'PhiC31 then Bxb1' target such that the target construct is initially in State 1, expressing RFP as per the shown target construct schematic (top). Representative leaf and root images of a T2 seedling transformed with the preswitched target (bottom).



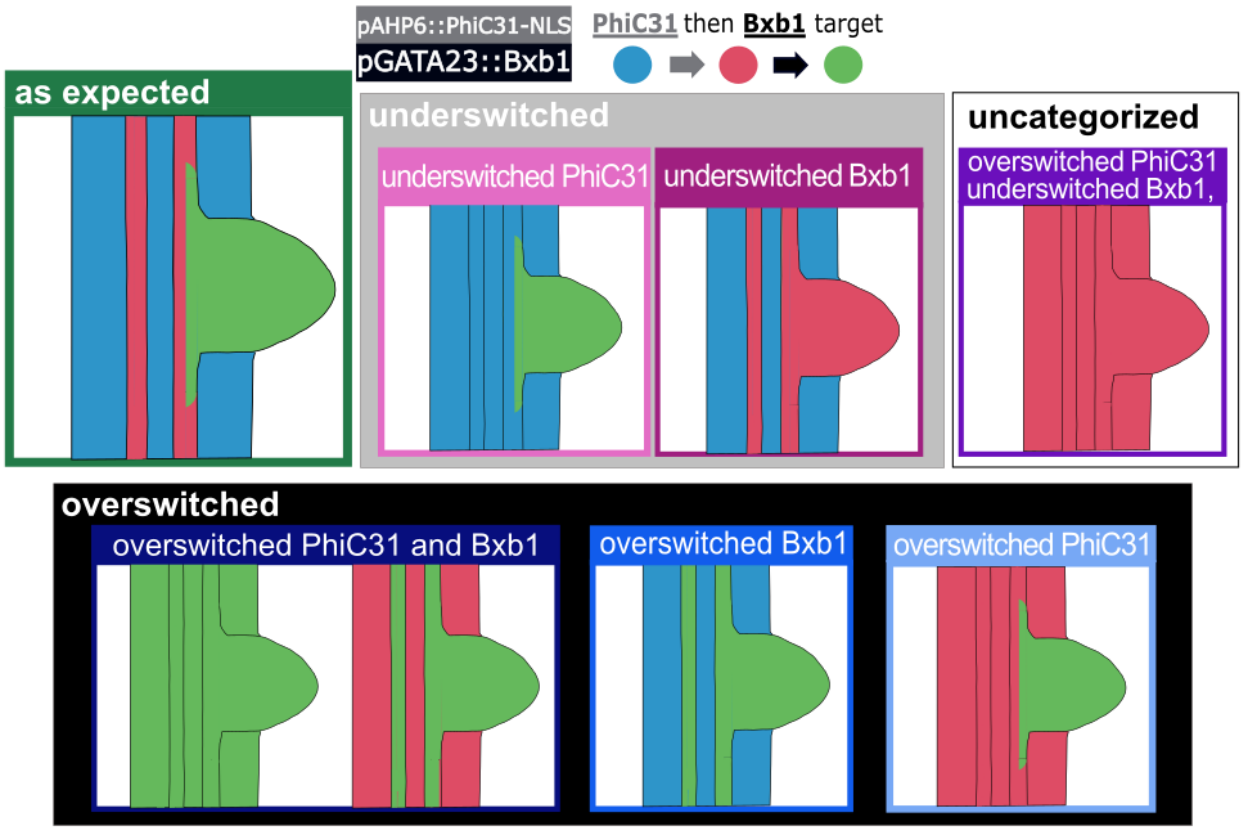
Supplementary Figure 2. State α genotyping. a Schematics of the primer binding on the full length target (shown as State 0, but will be the same size for State 1 or 2) and of the State α target. The PCR of the full length target would result in a band length of 4828 bp whereas PCR of the excised target would result in a band size of 533 bp. **b** Genotyping results. (left) Control PCR for the 'PhiC31 then Bxb1' full length target in State 0. (right) Genotyping results for pPP2AA3::PhiC31 in the 'Bxb1 then PhiC31' target (Supplementary Figure 3) (left two lanes, corresponding to two different T1 seedlings) and pPP2AA3::Bxb1 in the 'PhiC31 then Bxb1' target (right lane, one T1 seedling). Both sets of genotyping were performed twice with replicable results. Source data are provided as a Source Data file.



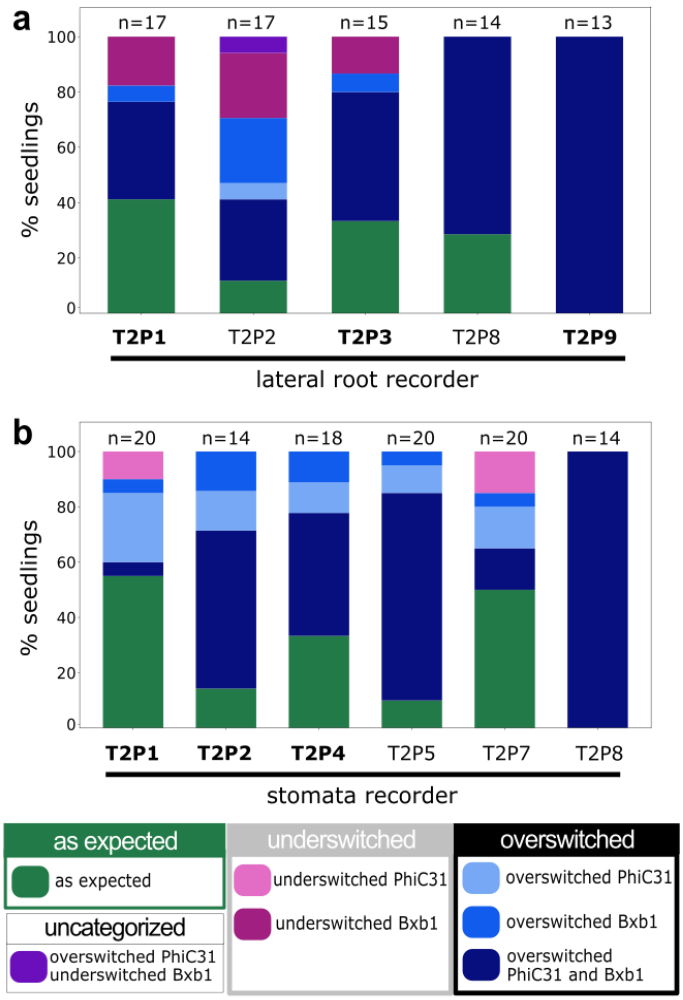
Supplementary Figure 3. The reverse order history dependent integrase circuit switches between fluorescent states. a Schematic of integrase target circuit. The design of the target is the same as in Fig. 1, but Bxb1 mediates the switch from State 0 to 1 and PhiC31 mediates the switch from State 1 to 2. Addition of the PhiC31 integrase to the State 0 target configuration results in a DNA excision and a loss of fluorescence (State α). **b** The Bxb1 integrase mediates switch from State 0 to State 1. (left) The initial target in State 0 shows strong BFP mtagBFP2 expression in the leaf and the root. (middle) Addition of constitutive Bxb1 expression to the target mediates a switch to State 1 and strong RFPmCherry fluorescence in leaf and root tissue. **c** Addition of PhiC31 prior to Bxb1 mediates an 'out of order' switch to State α and loss of fluorescence. (left) The initial target in State 0 shows strong BFPmtagBFP2 expression in the leaf and the root. (right) Addition of constitutively expressed PhiC31 to the target in State 0 results in a switch to State α and a loss in fluorescence in the leaf and root. Root and leaf tissue from 10 seedlings each from at least 2 T1 lines for each of State 0, 1, 2, and α were characterized. n=10 seedlings were screened per T1 line with at least 2 T1 lines per each of State 0, 1, and α. Black scale bars in the brightfield images correspond to 100 µm (root) and 50 µm (leaf).



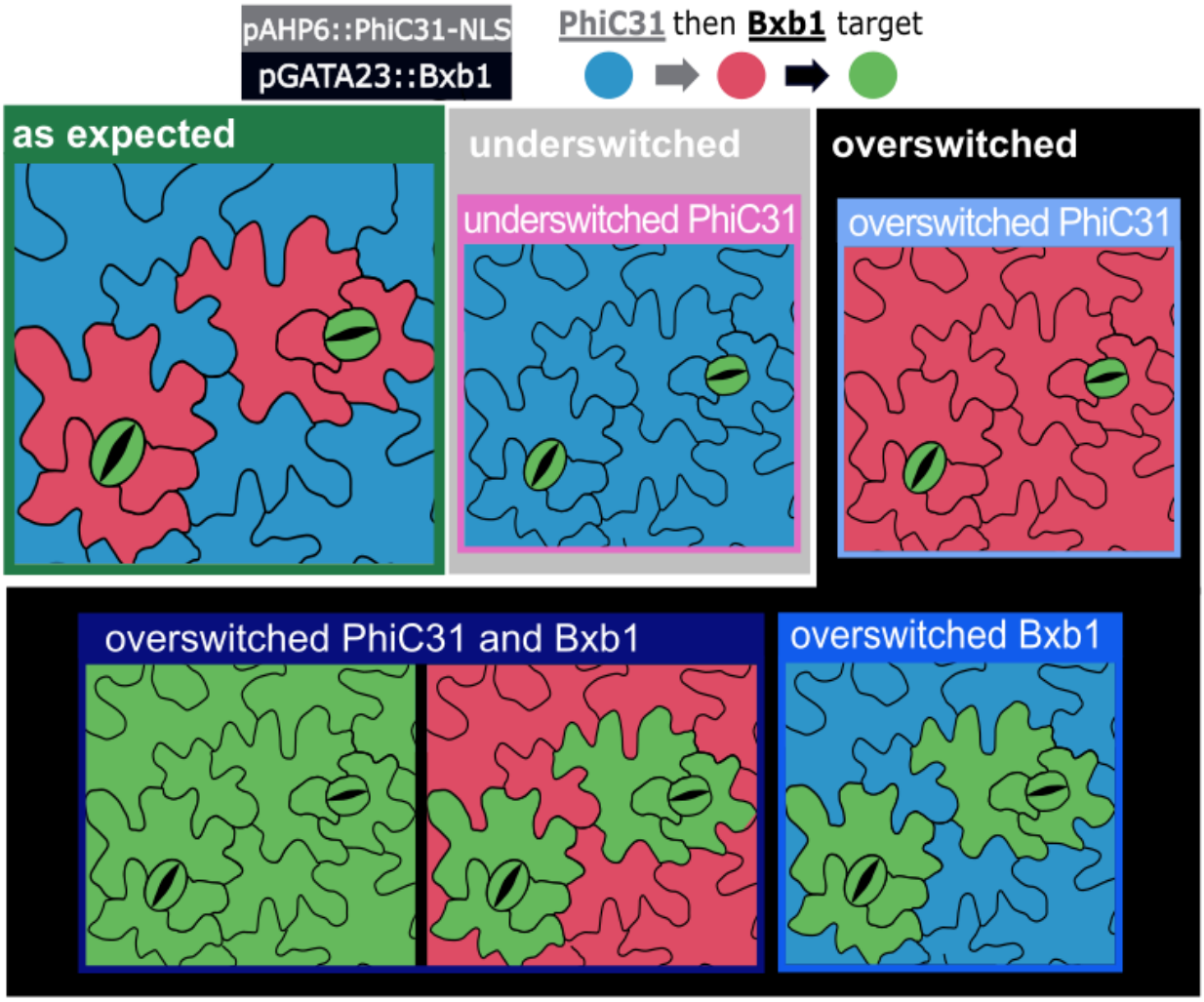
Supplementary Figure 4. Single integrase switch characterization for lateral root and stomata related genes. The bar graph corresponds to the percentage of seedling per T1 line in each category: 'as expected' (in green), overswitched (in black), and underswitched (in light grey). **a** Single integrase switch performance for lateral root genes. Seedlings were screened from T1 lines which came from T1 plants with an 'as expected' or underswitched switch pattern. The 'PhiC31 then Bxb1' target was transformed with the pAHP6::PhiC31-NLS integrase construct (left). The reverse order 'Bxb1 then PhiC31' target (Supplementary Figure 3) was transformed with the pGATA23::Bxb1-DST integrase construct (right). The number of screened seedlings for each T1 line are indicated above the bars. The bolded T1 lines correspond to the lines shown in Figures 2a and 2b. **b** Single integrase switch performance for stomata genes. Seedlings were screened from T1 lines which came from T1 plants with an 'as expected' or underswitched switch pattern. The 'PhiC31 then Bxb1' target was transformed with the pSPCH::PhiC31 integrase construct (left). The 'Bxb1 then PhiC31' target (Supplementary Figure 3) was transformed with the pGATA23::Bxb1 integrase construct (right). The number of screened seedlings for each T1 line are indicated above the bars. The bolded T1 lines correspond to the lines shown in Figures 3a and 3b. Source data are provided as a Source Data file.



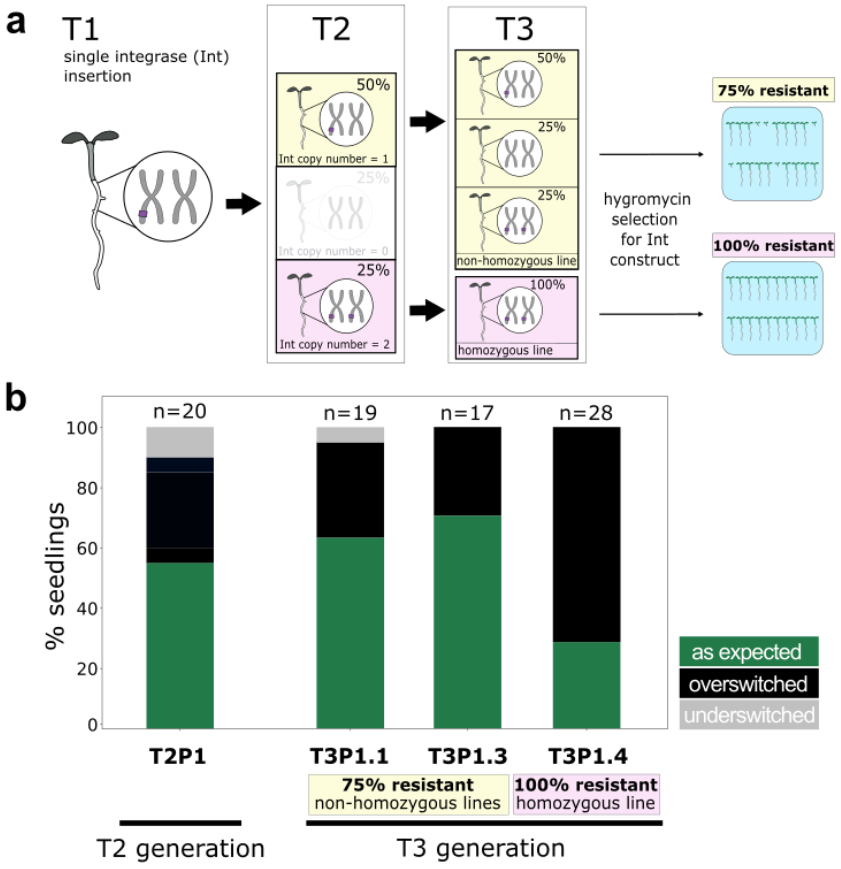
Supplementary Figure 5. Schematics for observed lateral root recorder output. The diagrams represent the lateral root recorder outputs observed for the 'PhiC31 then Bxb1' target transformed with the pAHP6::PhiC31-NLS-pGATA23::Bxb1 integrase construct. The box colors correspond to the key in Supplementary Figure 5 and the background boxes correspond to the categories in Figure 2.



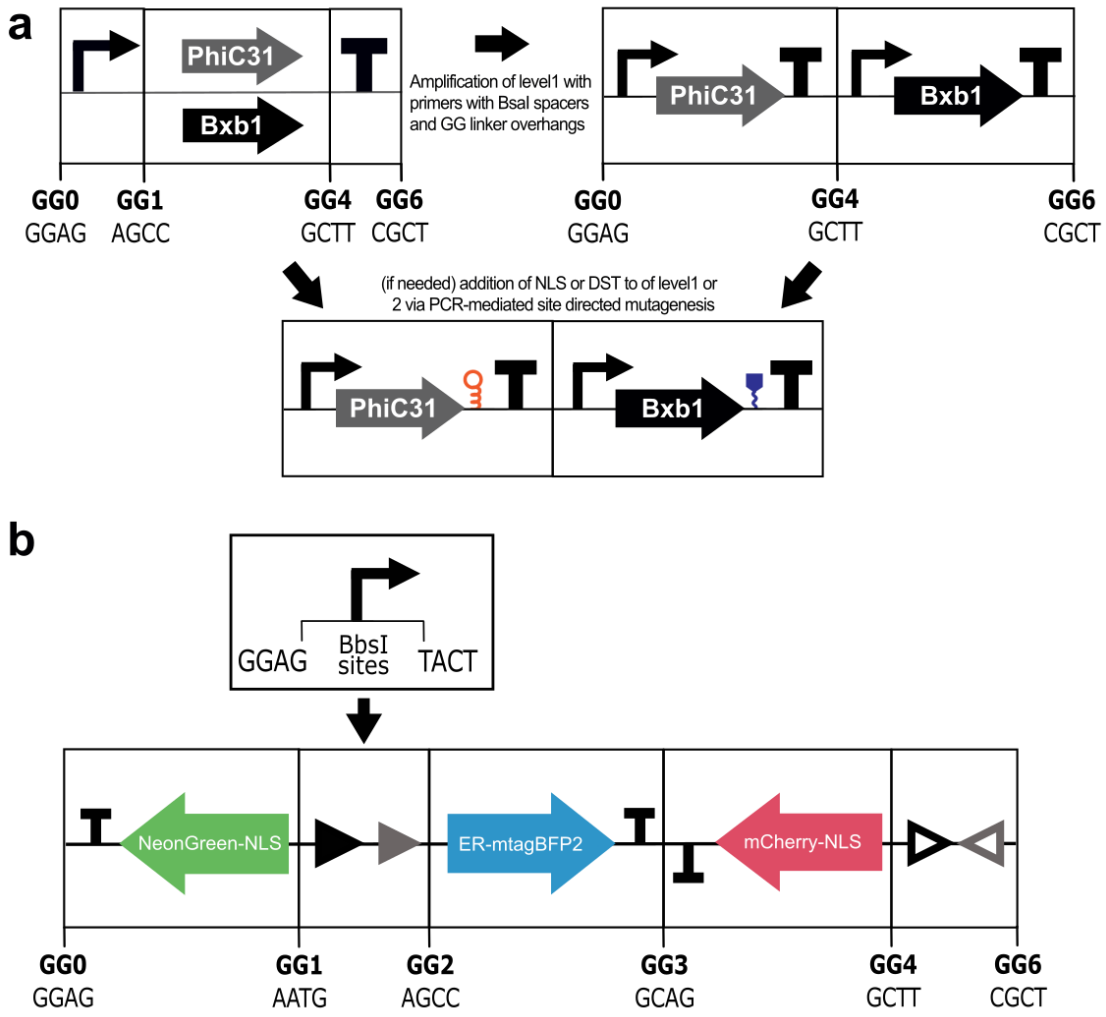
Supplementary Figure 6. Detailed characterization of recorder performance. The bar graph corresponds to the percentage of seedling per T1 line in each category. **a** Lateral root recorder performance. Seedlings were screened from 5 T2 seedlings from T1 lines which came from T1 plants with an 'as expected' or underswitched recorder output. The 'PhiC31 then Bxb1' target was transformed with the pAHP6::PhiC31-NLS-pGATA23::Bxb1 dual integrase construct. The number of seedlings screened for each T1 line is indicated above the bars. Bolded T1 lines correspond to the lines shown in Fig. 2d. The detailed switch categories correspond to the representations in Supplementary Figure 6. **b** Stomata recorder performance. Seedlings were screened from 6 T1 lines which came from T1 plants with an 'as expected' or underswitched recorder output. The 'PhiC31 then Bxb1' target was transformed with the pSPCH::PhiC31-pMUTE::Bxb1 dual integrase construct. The number of seedlings screened for each T1 line is indicated above the bars. Bolded T1 lines correspond to the lines shown in Fig 3d. The detailed switch categories correspond to the representations in Supplementary Figure 7. Source Data are provided as a Source Data file.



Supplementary Figure 7. Schematics for observed stomata recorder output. The diagrams represent the lateral root recorder outputs observed for the 'PhiC31 then Bxb1' target transformed with the pSPCH::PhiC31-pMUTE::Bxb1 integrase construct. The box colors correspond to the key in Supplementary Figure 5 and the background boxes correspond to the categories in Figure 3.



Supplementary Figure 8. Characterization of stomatal recorder in homozygous and non-homozygous T3 lines. a Schematic of integrase construct heredity from the T1 to T3 generation. The integrase construct exists as a single copy in the T1 plant (inserted via floral dip). Plants in the resulting T2 line have a mix of integrase copy numbers (one: 50%, two: 25%, zero: 25%). Plants within T3 lines descending from a homozygous parent will all be homozygous for the integrase, while plants within T3 lines descending from a heterozygous parent will show the same distribution of integrase copy number as the parental T2 line. Zygosity can be assessed in T3 via hygromycin selection wherein homozygous lines will be 100% resistant and non-homozygous lines will be 75% resistant. **b** Results of T3 characterization and hygromycin selection. Seedlings were screened from T3 lines which came from the best performing stomata recorder T2 line (T2P1, plotted for reference on the left). Only the T3 lines which showed switching were included. The number of seedlings screened for each line is indicated above the bars. Of the three T3 lines, two of them (T3P1.1 and T3P1.3) are non-homozygous (75% hygromycin resistant) and one (T3P1.4) is homozygous (100% hygromycin resistant). The 'no switch' seedlings are represented as corresponding to the seedlings which have lost the integrase construct. Source data is provided as a Source Data file.



Supplementary Figure 9. Cloning strategy. a Cloning strategy for the integrase level 1 and level 2 constructs. The level 1 is built using golden gate cloning with BsaI restriction sites. The level 1 integrase constructs are then amplified using primers with overhangs including a BsaI restriction site and the appropriate GG sticky end. Then the level 2 is assembled using golden gate assembly with BsaI. If desired, tuning parts (NLS, DST) are added to the level 1 or level 2 via PCR-mediated site directed mutagenesis. **b** Cloning strategy for the target construct. The part with the integrase sites and fluorescent proteins is constructed by golden gate assembly with the BsaI enzyme. The promoter is added to the synthetic fragment via golden gate cloning using BbsI restriction sites to form the completed target.

Supplementary Table 1. Detailed overview of integrase lines and characterization categories.

	Integrase construct name	PhiC31 construct	Bxb1 construct	Target construct	As expected	Underswitched	Overswitched
pAHP6 single switch	L1P137	pAHP6::PhiC31-NLS:tUBQ1	n/a	PhiC31 then Bxb1 target (Fig. 1)	RFP in xylem and lateral roots BFP otherwise	RFP in less cells than xylem and LR	RFP in more cells than xylem and LR
pGATA23 single switch	L1B42	n/a	pGATA23::Bxb1-DST:tUBQ10	Bxb1 then PhiC31 target (Supplementary Figure 3)	RFP in LR BFP otherwise	RFP in less cells than LR	RFP in more cells than LR
Lateral root history-dependent tracker	L1PB10	pAHP6::PhiC31-NLS:tUBQ1	pGATA23::Bxb1-DST:tUBQ10	PhiC31 then Bxb1 target (Fig. 1)	GFP in LR RFP in xylem BFP otherwise	GFP in less cells than LR and/or RFP in less cells than xylem	GFP in more cells than LR and/or RFP in more cells than xylem and LR
pSPCH single switch	L1P139	pSPCH::PhiC31:tUBQ1	n/a	PhiC31 then Bxb1 target (Fig. 1)	RFP in guard cells and surrounding epidermal cells BFP otherwise	RFP in less cells than guard cells and surrounding epidermal cells	RFP in more cells than guard cells and surrounding epidermal cells
pMUTE single switch	L1B39	n/a	pMUTE::Bxb1:tUBQ10	Bxb1 then PhiC31 target (Supplementary Figure 3)	RFP in guard cells BFP otherwise	RFP in less cells than guard cells	RFP in more cells than guard cells
Stomatal history-dependent tracker	L1PB1	pSPCH::PhiC31:tUBQ1	pMUTE::Bxb1:tUBQ10	PhiC31 then Bxb1 target (Fig. 1)	GFP in guard cells RFP in surrounding epidermal cells BFP otherwise	GFP in less cells than guard cells and/or RFP in less cells than surrounding epidermal cells	GFP in more cells than guard cells and/or RFP in more cells than surrounding epidermal cells

Supplementary Table 2. Characterization of single integrase switches for lateral root genes in the T1 generation.

Tracked gene	Integrase construct	T1 seedling	Switch categorization
AHP6	pAHP6::PhiC31-NLS	T1P1	Overswitched
AHP6	pAHP6::PhiC31-NLS	T1P2	As expected
AHP6	pAHP6::PhiC31-NLS	T1P3	As expected
AHP6	pAHP6::PhiC31-NLS	T1P4	Overswitched
AHP6	pAHP6::PhiC31-NLS	T1P5	Underswitched
AHP6	pAHP6::PhiC31-NLS	T1P6	Overswitched
AHP6	pAHP6::PhiC31-NLS	T1P7	As expected
AHP6	pAHP6::PhiC31	T1P1	Underswitched
AHP6	pAHP6::PhiC31	T1P2	Underswitched
AHP6	pAHP6::PhiC31	T1P3	Underswitched
AHP6	pAHP6::PhiC31	T1P4	Underswitched
GATA23	pGATA23::Bxb1	T1P1	Overswitched
GATA23	pGATA23::Bxb1	T1P2	Overswitched
GATA23	pGATA23::Bxb1	T1P3	Overswitched
GATA23	pGATA23::Bxb1	T1P4	Overswitched
GATA23	pGATA23::Bxb1	T1P5	Overswitched
GATA23	pGATA23::Bxb1	T1P6	Overswitched
GATA23	pGATA23::Bxb1	T1P7	Overswitched
GATA23	pGATA23::Bxb1	T1P8	Overswitched
GATA23	pGATA23::Bxb1	T1P9	Overswitched
GATA23	pGATA23::Bxb1	T1P10	Overswitched
GATA23	pGATA23::Bxb1-DST	T1P1	As expected
GATA23	pGATA23::Bxb1-DST	T1P2	Overswitched
GATA23	pGATA23::Bxb1-DST	T1P3	As expected
GATA23	pGATA23::Bxb1-DST	T1P4	As expected
GATA23	pGATA23::Bxb1-DST	T1P5	As expected
GATA23	pGATA23::Bxb1-DST	T1P6	As expected
GATA23	pGATA23::Bxb1-DST	T1P7	Overswitched
GATA23	pGATA23::Bxb1-DST	T1P8	Overswitched

Supplementary Table 3. Characterization of the lateral root and stomatal recorder in the T1 generation.

Cell lineage	Tracked genes	PhiC31 construct	Bxb1 construct	T1 seedling	Switch categorization
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P1	Underswitched PhiC31
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P2	Underswitched PhiC31
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P3	As expected
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P4	Overswitched PhiC31 and Bxb1
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P5	Overswitched PhiC31 and Bxb1
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P6	Overswitched PhiC31 and Bxb1
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P7	Overswitched PhiC31 and Bxb1
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P8	Underswitched PhiC31
lateral root	*AHP6, GATA23*	pAHP6::PhiC31-NLS	pGATA23::Bxb1	T1P9	As expected
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P1	As expected
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P2	As expected
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P3	Overswitched Bxb1
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P4	As expected
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P5	As expected
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P6	Overswitched Bxb1
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P7	As expected
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P8	As expected
stomata	*SPCH, MUTE*	pSPCH::PhiC31	pMUTE::Bxb1	T1P9	Overswitched Bxb1

Supplementary Table 4. Characterization of single integrase switches for stomatal genes in the T1 generation.

Tracked gene	Integrase construct	T1 seedling	Switch categorization
SPCH	pSPCH::PhiC31	T1P1	As expected
SPCH	pSPCH::PhiC31	T1P2	Overswitched
SPCH	pSPCH::PhiC31	T1P3	As expected
SPCH	pSPCH::PhiC31	T1P4	Overswitched
SPCH	pSPCH::PhiC31	T1P5	Underswitched
SPCH	pSPCH::PhiC31	T1P6	As expected
SPCH	pSPCH::PhiC31	T1P7	Overswitched
SPCH	pSPCH::PhiC31	T1P8	As expected
SPCH	pSPCH::PhiC31	T1P9	Overswitched
SPCH	pSPCH::PhiC31	T1P10	Overswitched
MUTE	pMUTE::Bxb1	T1P1	As expected
MUTE	pMUTE::Bxb1	T1P2	As expected
MUTE	pMUTE::Bxb1	T1P3	Overswitched
MUTE	pMUTE::Bxb1	T1P4	Overswitched
MUTE	pMUTE::Bxb1	T1P5	As expected
MUTE	pMUTE::Bxb1	T1P6	Overswitched
MUTE	pMUTE::Bxb1	T1P7	As expected
MUTE	pMUTE::Bxb1	T1P8	Overswitched
MUTE	pMUTE::Bxb1	T1P9	Overswitched
MUTE	pMUTE::Bxb1	T1P10	As expected

ACS SyntheticBiology

pubs.acs.org/synthbio Research Article

A Hot-Swappable Genetic Switch: Building an Inducible and Trackable Functional Assay for the Essential Gene MEDIATOR 21

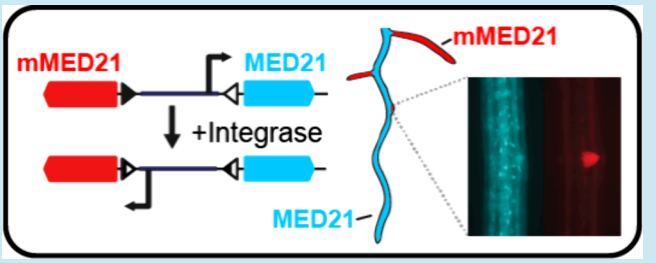
Published as part of ACS Synthetic Biology special issue "Plant Synthetic Biology 2.0".

Isabella J. Watson, Cassandra Maranas, Jennifer L. Nemhauser, and Alexander R Leydon*

Cite This: *ACS Synth. Biol.* 2025, 14, 2170–2180 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Essential genes, estimated to be approximately 20% of the *Arabidopsis* genome, are broadly expressed and required for reproductive success. They are difficult to study, as interfering with their function leads to premature death. Transcription is one of the essential functions of life, and the multiprotein Mediator complex coordinates the regulation of gene expression at nearly every eukaryotic promoter. In this study, we focused on a core mediator component called MEDIATOR21 (MED21), which is required for activation of transcription. Our previous work has also shown a role for MED21 in repression of gene expression through its interaction with a corepressor protein. Here, we sought to differentiate the role MED21 plays in activation versus repression using the model plant *Arabidopsis*. As mutations in MED21 lead to embryo-lethal phenotypes, we constructed a set of synthetic switches using PhiC31 serine integrases to create an "on-to-off" inducible loss-of-function MED21 in a nonessential tissue. Our technology, which we call Integrase Erasers, enabled *med21* mutant plants to survive into adulthood by ablating protein expression selectively in lateral root primordia, allowing quantification and characterization of *med21* mutant phenotypes in a postembryonic context. In addition, we engineered chemical induction of the Integrase Eraser to ablate MED21 expression in whole seedlings at a user-specified time point. Finally, we extended this technology to build a hot-swappable Integrase Isoform Switch where expression of the integrase toggled cells from expressing wild-type MED21 to expressing MED21 sequence variants. Our analysis of the entire set of new integrase-based tools demonstrates that this is a highly efficient and robust approach to the study of essential genes.



KEYWORDS: *synthetic biology, integrases, MED21, inducible knockouts, variant switching*

INTRODUCTION

Essential genes perform functions required for survival. They are also likely to be broadly expressed, show up as hubs in protein networks, and be components of stable complexes with other biomolecules.[1−3] Examples like general transcription factors (GTFs) and the Mediator complex play significant roles in regulating transcription at most genes via interactions with gene-specific DNA-binding transcription factors.[4−7] The difficulty in assigning function to mediator subunits is well-documented in plants.[4,5,8,9] Previously, we reported that the activity of the plant corepressor TOPLESS (TPL) depends on interaction with the mediator complex through direct contact with the middle domain subunit MEDIATOR21 (MED21).[10,11] One consequence of the embryo-lethal phenotype of *med21* loss of function mutants[12] has been difficulty differentiating its mechanism of action during gene activation versus gene repression.

There are several approaches that have been employed to study essential genes in yeast, such as temperature-sensitive alleles,[13] chemically induced nuclear depletion (anchor away),[14] and auxin-induced degradation (AID).[15] Adaptations of these approaches have been employed in plants,[16−18] yet the majority of essential genes have been characterized by the embryo-lethal phenotypes found in T-DNA insertion mutants.[19] Recent advances in the application and optimization of serine integrase activity in *Arabidopsis thaliana* provide an alternative means to efficiently, precisely, and irreversibly alter relatively large genomic regions.

Serine integrases, hereafter termed integrases, induce recombination events at asymmetric DNA sequences that are specific to each integrase. Two integrases, Bxb1 and PhiC31, were recently optimized to build a modular toolkit for *Arabidopsis*.[20] In that work, the integrase target site contained a ubiquitous promoter flanked by integrase sites, such that expression of the cognate integrase switched expression from

Received: February 4, 2025
Revised: April 23, 2025
Accepted: April 24, 2025
Published: May 9, 2025

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

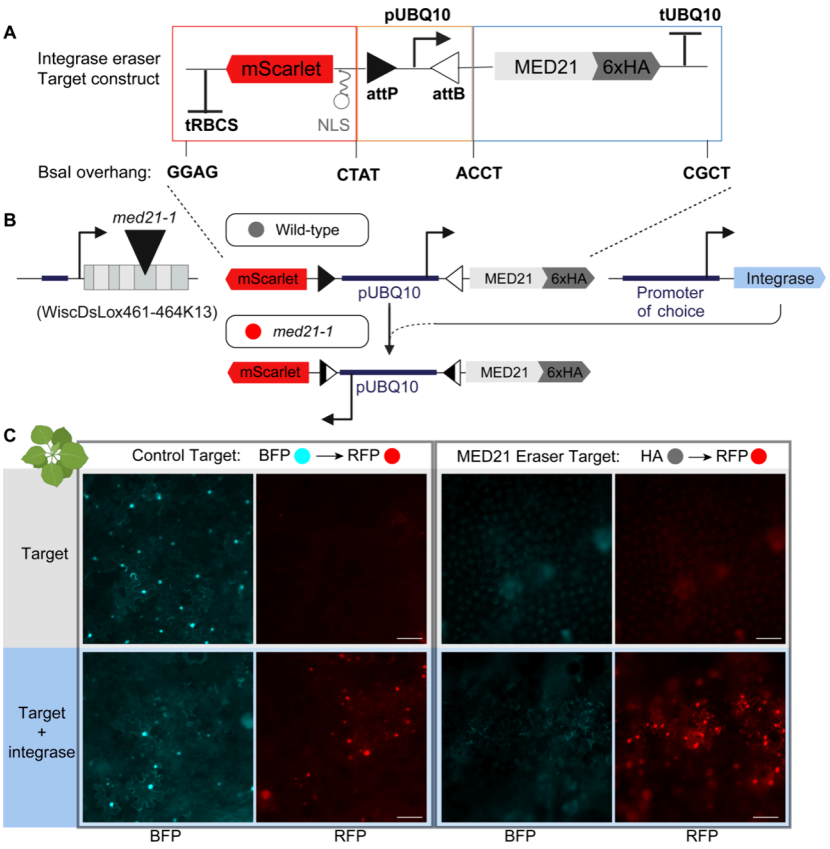


Figure 1. Engineering and rapid prototyping of the integrase eraser approach. (A) Cloning strategy based on (Guiziou et al.[20]) for the target integrase target with BsaI adapters highlighted for use in golden gate cloning strategies, see Plasmid Maps. (B) Design of the integrase target. The target is composed of two PhiC31 integrase sites (triangles) surrounding a constitutive promoter (pUBQ10), the rescue coding sequence for MED21:6xHA and the fluorescent reporter (mScarlet). In the absence of integrase, MED21:6xHA is expressed. In the presence of integrase, the integrase mediates inversion of the DNA between the integrase sites, inverting the promoter and leading to mScarlet expression. The expression of the integrase is mediated by the selected promoter selected. (C) Rapid prototyping of the MED21 Eraser target in transient transfections of *Nicotiana benthamiana* at 2 days after injection. On the left side is a control target that switches from mTurquoise to mScarlet alone (top) and with a *p35S:PhiC31* construct (bottom). On the right side shows the MED21 target alone (top) and MED21:HA target with a *p35S:PhiC31* construct (bottom). The BFP channel for the MED21:HA is shown to demonstrate the level of background expected for the BFP channel in the negative control. Microscopy images were taken on a 20x objective to allow a wide view of switching efficiency, and the 50-μm scale bar in the RFP channels applies to all paired BFP images.

one fluorescent protein to another. Further work using more complicated target design and cell-type-specific developmental promoters allowed order-of-expression cell lineage tracing.[22]

Here, we extended the work on initial integrase prototypes[20] to study the essential gene *MEDIATOR21* (*MED21*). We constructed an "on-to-off" inducible loss-of-function integrase target, which we term an Integrase Eraser. This target switches expression from MED21 to a fluorescent reporter, thereby permanently labeling cells with altered genomes. We tested target behavior using two different PhiC31 expression schemes in planta: (1) an integrase driver that induces loss of function specifically in lateral root primordia and (2) an inducible integrase driver that enables synchronized switching in all cells at a given time point. By expression of these constructs in a *med21* mutant plant in combination with the MED21 target, both approaches successfully circumvent the embryonic lethal phenotype. In addition, we engineered an "isoform switch" integrase target, where MED21 is expressed as a P2A-linked self-cleaving fusion with a fluorescent protein, differentially marking cells expressing wild-type or mutant isoform. This new collection of tools, and the ready adaptation of its components to most contexts of potential interest, unlocks the study of essential genes in postembryonic plants while specifically advancing our understanding of the Mediator complex in coordinating transcriptional regulation.

RESULTS AND DISCUSSION

To better understand the function of *MED21* in *Arabidopsis*, we built a synthetic circuit that would allow for selective loss of *MED21* function in lateral roots, a nonessential organ. This Integrase Eraser technology is composed of two parts: an "on-to-off" inducible loss-of-function integrase target and a driver line that expresses the PhiC31 integrase from a cell-type-specific promoter, pGATA23. To build our target, we adapted a golden gate assembly approach from Guiziou et al.,[20] that yielded a final construct with a constitutive promoter (pUBQ10) flanked by integrase sites, positioned between a wild-type copy of *MED21* fused to an HA epitope tag and a reporter gene (mScarlet) (Figure 1A, Plasmid Map pNL3779). We also cloned a version of the target lacking the HA tag to control for effects of tag placement on the MED21 protein (Plasmid Map pNL3857). Both constructs were transformed into a heterozygous T-DNA loss-of-function *med21−1* mutant. The target constitutively complements the loss of MED21 until

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

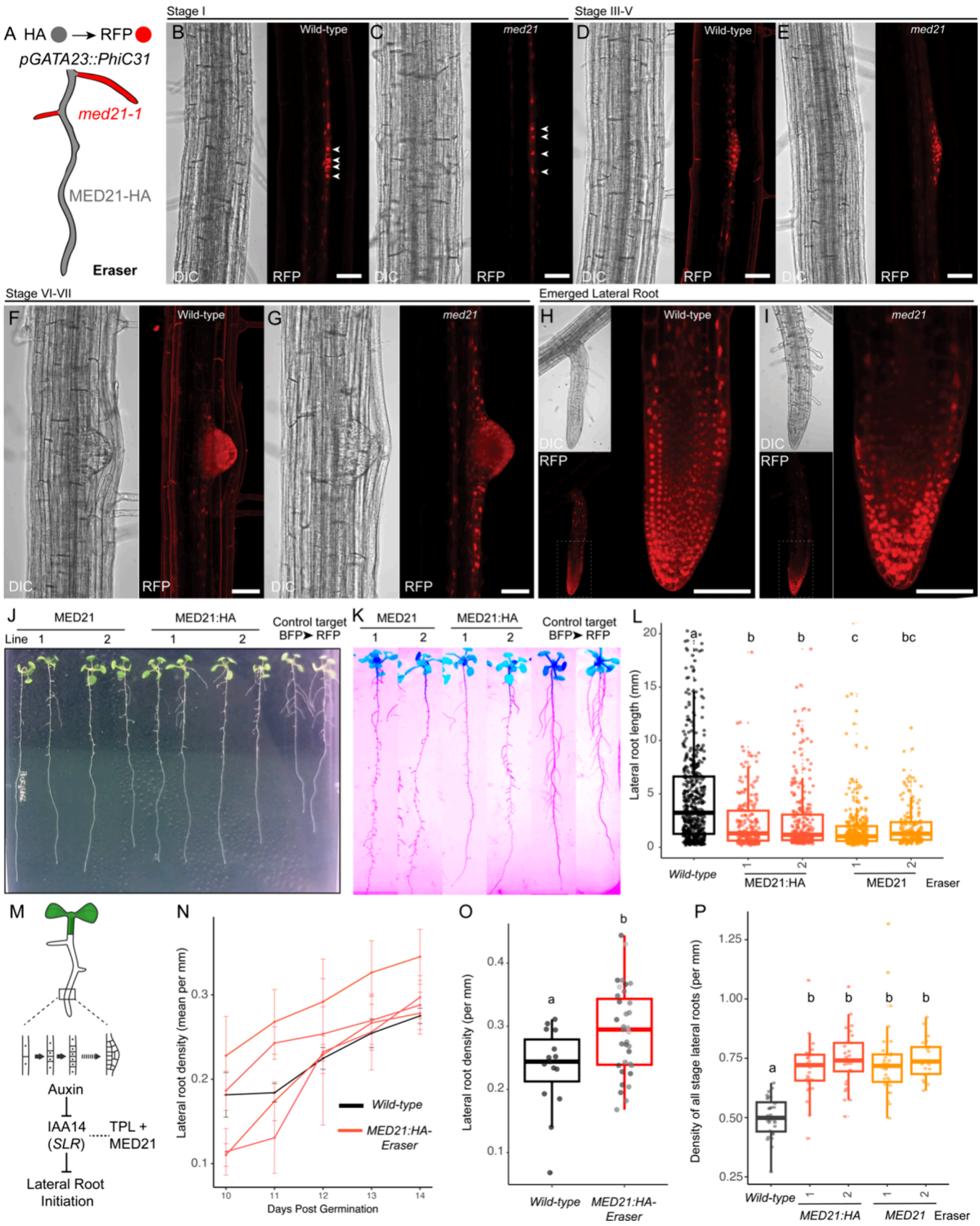


Figure 2. Cell-type-specific integrase eraser implemented in lateral root primordium results in *MED21* loss of function and increases in lateral root initiation. (A) Schematic of the predicted behavior for the MED21 eraser approach. Lateral root primordium cells (red) will lose expression of the rescue construct, rendering them an effective knockout for MED21. (B–J) Confocal microscopy analysis of the wild-type control switch (from Figure 1C) and MED21 eraser lateral root initiation. All scale bars are 50 μm. (J) Still image from Supplemental Movie 1 of the root growth phenotype of MED21 eraser lines compared to the wild type at day 11. (K) Whole seedling epifluorescent image of MED21 eraser lines at day 14.

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

Figure 2. continued

Images have been color inverted to allow visualization of the short root phenotype. Natural autofluorescence was captured for whole seedlings on plates, allowing all root tissue to be observed. (L) Lateral root lengths were quantified at day 14 from individual selected T2 lines for each eraser type. The experiment was performed in triplicate for each genotype, with 6 seedlings per plate (n = 18 per genotype, at least 180 lateral roots per genotype). (M) Auxin-induced degradation of IAA14 is strictly required for the initiation of lateral root development. IAA14 recruits TPL, which in turn inhibits the mediator complex through interaction with MED21. Cartoon is modified from ref 10. (N) Lateral root density over time for MED21:HA eraser lines. Data are presented as means ± standard error. The experiment was performed in triplicate for each genotype, with 6 seedlings per plate (n = 18 per genotype). (O) Lateral root density at day 14 of growth for the MED21:HA eraser lines. Each gray color represents an individual lateral root density from 4 to 6 independent lines and is representative of the day 14 data from N. (P) Lateral roots of all stages (emerged and nonemerged) were quantified in homozygous mutant lines at day 14. For (L,O,P), letters indicate significant difference (ANOVA and Tukey HSD multiple comparison test; $p < 0.001$).

the integrase is active, so it is transformed into mutant plants before the introduction of the driver line (Figure 1B). All cells where the integrase mediates an inversion in the target (a "switch") are marked by the expression of mScarlet (Figure 1B, lower arrows). Therefore, the presence of mScarlet indicates the loss of wild-type *MED21* activity (Figure 1B).

We used rapid prototyping in *Nicotiana benthamiana* to test the functionality of the MED21 Integrase Eraser target. As a control, we used a previously characterized target that switches from mTurquoise to mScarlet in the presence of the PhiC31 integrase expressed from the strong viral 35S promoter (Figure 1C). Without the addition of PhiC31 integrase, no mScarlet was detected for either the control or MED21 Eraser targets, indicating no stochastic switching occurred. In contrast, in the presence of PhiC31, we observed the presence of cells that expressed nuclear mScarlet. These results confirmed that the MED21 Integrase Eraser was functional and specific.

Next, we generated stable transgenic *Arabidopsis* plants expressing the *MED21* Integrase Eraser constructs. We first focused on erasing *MED21* activity during lateral root initiation, a well-characterized example of de novo organogenesis (Banda et al., 2019).[21] Lateral roots are an attractive candidate for this Integrase Eraser technology, as they are not required for plants to survive in controlled environments like growth rooms or greenhouses. Thus, the impact of losing essential gene function can be studied while maintaining largely wild-type levels of plant health and reproductive fitness. Lateral roots develop from a small subset of the xylem pole pericycle (XPP) termed founder cells. *GATA TRANSCRIPTION FACTOR 23 (GATA23)* is strongly and largely specifically expressed in these founder cells, setting into motion the series of asymmetric cell divisions that form the new root.[23] Previous use of a *pGATA23* integrase driver produced many lines with switching only in cells in the lateral root, that is, cells from a lineage involving *GATA23* expression.[20] Our use of this same promoter in the *MED21* Integrase Eraser should "erase" *MED21* in all of the cells that compose the lateral roots (Figure 2A, red lateral roots). It has been recently demonstrated that *Arabidopsis* promoters behave in a predominantly unidirectional manner,[24] and we ensured that pUBQ10 behaved unidirectionally in the context of the integrase target. Indeed, in all MED21 eraser lines, we observed no detectable expression of the upstream RFP gene by microscopy (0/43 MED21:HA target, 0/49 MED21 target) in plants that do not carry an integrase expression construct. We confirmed this by RT-qPCR and observed no detectable expression of the RFP at the molecular level in the absence of the integrase (Figure S1A,B). These results demonstrate that pUBQ10 acts as a strong unidirectional promoter, consistent with our other implementations in *Arabidopsis*.[20,22]

After introducing the *pGATA23::PhiC31* driver into *med21−1* plants complemented by the MED21 eraser target (Figure 1B), we identified lines with mScarlet expression predominantly restricted to the lateral roots (Figure 2B; 5 independent MED21:6xHA eraser lines and 9 independent untagged MED21 eraser lines), and compared these to the control switch (Figure 1C) through confocal microscopy. All selected lines demonstrated essentially wild-type pre-emergence lateral root development (Figure 2B−G; stages I−VII). However, the loss of MED21 did lead to mild disruptions of architecture in postemergence roots. As early as 10 days post germination (dpg), we observed that MED21 eraser lateral roots had larger root cap zones and abnormally shaped root hairs (Figure 2H−I), as well as a reduced length compared with our wild-type controls (Figure 2J). The latter phenotype was even more striking at 14 dpg (Figure 2K,L). Dynamic tracking of lateral root development over 16 days (Supplemental Video 1) highlights the impact of loss of MED21 activity. We observed switching primarily from the early stages of the lateral root initiation; however, we observed some nuclear RFP signal in the plant pericycle outside the developing lateral root, especially in older roots, consistent with the pGATA23 promoter's native expression behavior. Therefore, in future implementations, the pGATA32 promoter could be tuned to reduce integrase activity to improve specificity through the inclusion of protein degrons to the integrase, or RNA destabilization tags (DSTs) to its transcript.[20] Importantly, we also observed no evidence of bidirectional transcription of the MED21:HA gene in dissected lateral root cells that have fully switched to the reverse orientation (Figure S1A,C), again supporting the conclusion that pUBQ10 is unidirectional. If a target needed to have an excisable portion, for example, to eliminate low levels of transcription that may occur in unforeseen circumstances, the reorientation of the target sites to face in the same direction would allow construction of a guaranteed "zero background" integrase target site (Figure S1D).

Previous work has shown that the corepressor TPL interacts with the Mediator complex directly through contacts with MED21[10] (Figure 2N). The interaction between the Mediator complex and TPL is critical in maintaining repression of transcription in yeast and plants.[10,11] One of the best-studied TPL-regulated pathways is auxin response.[25] In low auxin environments, the TPX family (TPL/TPRs) inhibits the activity of auxin response factor (ARF) transcriptional activators through the mutual binding of Aux/IAA proteins (AUXIN/INDOLE-3-ACETIC ACID). When auxin levels rise, Aux/IAA proteins are degraded, disrupting the connection between ARF and TPX proteins, thereby activating expression of ARF target genes.[26] Following this logic, loss of MED21

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

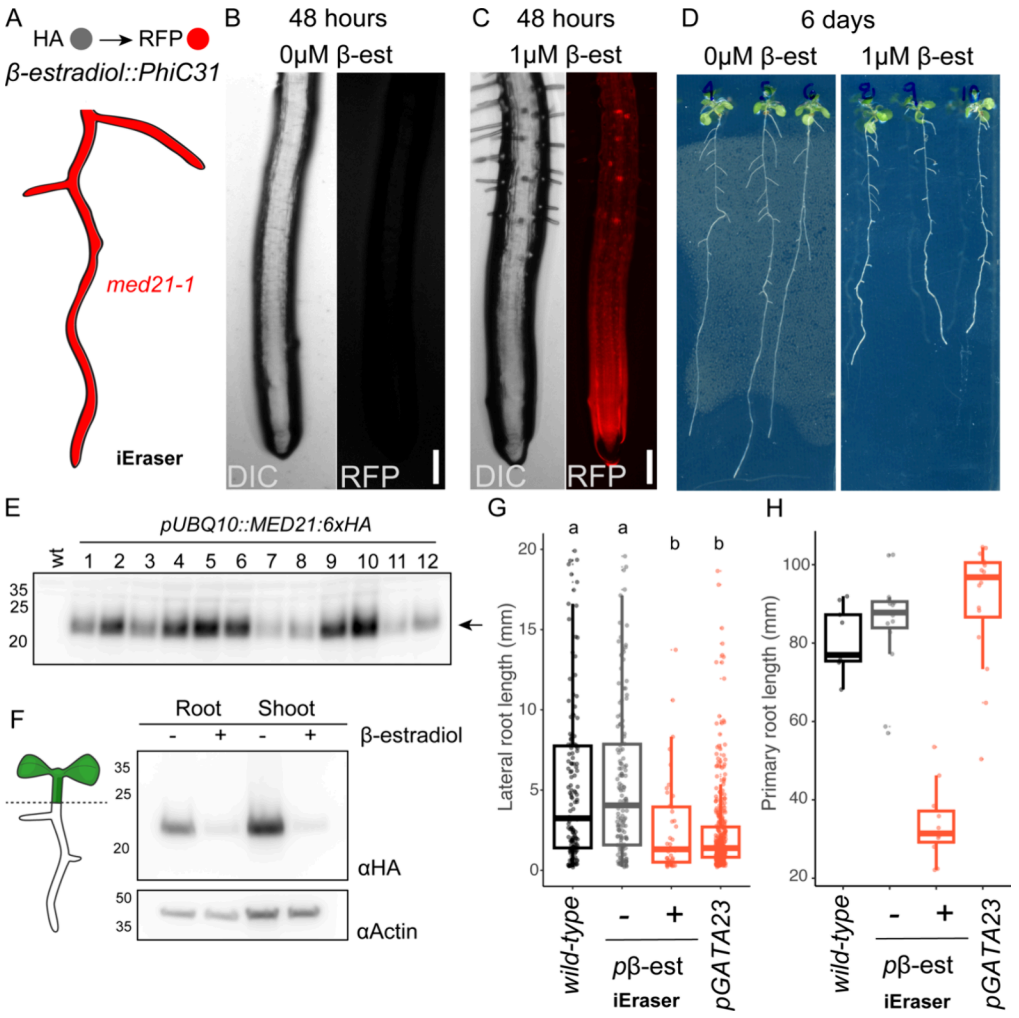


Figure 3. A chemically inducible MED21 eraser. (A) Schematic of the predicted behavior for the MED21 iEraser. The estradiol inducible integrase construct is composed of p35S:XVE (transcriptional activator composed of a DNA-binding domain of LexA, the transcription activation domain of VP16, and the regulatory region of the human estrogen receptor) and pLexA-minimal 35S driving expression of PhiC31. (B,C) Characterization of the iEraser by fluorescence microscopy shows induction as early as 48 h after treatment. The scale bar represents 50 μm (C). (D) Growth phenotypes were visible in estradiol-induced switched plants grown on 1 μM β-estradiol induction plates for 6 days. Seedlings were first grown for 6 days on LS media lacking β-estradiol before being transplanted to induction media (12 days total growth). (E) Protein expression analysis by Western blot for 12 independent iEraser lines. (F) Protein expression analysis by Western blot for roots and shoots isolated from plants grown with or without β-estradiol. Seedlings were grown on LS media for 6 days, and then transplanted to either control or induction media for another 6 days of growth. The experiment was performed in triplicate for each genotype. (G,H) Lengths of lateral (G) and primary (H) roots were quantified at 6 days from the indicated Integrase Eraser type. Seedlings were grown as in (F). Letters indicate significant difference (ANOVA and Tukey HSD multiple comparison test; $p < 0.001$).

should reduce TPL activity and phenocopy the addition of auxin. In lateral roots, the loss of MED21 should therefore promote initiation, an auxin-regulated process (Figure 2M). To test our model, we performed a time course growth analysis of MED21:HA eraser lines to determine whether lateral roots were initiated more frequently than controls (lateral root density, Figure 2N). We observed that over time, the MED21:HA eraser lines did indeed have a higher lateral root density than wild-type controls (Figure 2N,O). When we included pre-emergence lateral roots, this trend was even more clear (Figure 2P).

To enable loss of MED21 activity throughout the plant at a specific time point, we engineered a chemically inducible *MED21* Eraser (iEraser) (Figure 3A). To do this, we replaced *pGATA23* with the well-characterized β-estradiol system, as it is small, fast-acting in nearly every plant cell after treatment, and is fully orthogonal to known plant hormones.[20,27] We used our *MED21:HA* target for these experiments, as it functioned similarly to untagged *MED21* (Figure 2) and makes it possible to assay protein levels. Complemented *med21−1*, MED21:HA Eraser lines were transformed with *pβ-estradiol::PhiC31* and characterized for any leakiness in the expression of the integrase to eliminate these lines from further analysis (Figure S2A−C). Progeny from selected transformants (T2 plants) were then transplanted onto estradiol plates for analysis. By 48 h, we observed strong and specific induction of MED21:HA switching only in seedlings exposed to β-estradiol (Figure 3B,C). We defined a seedling as having switched when nuclear RFP was detected in the root tip region by microscopy and considered a line specific when all roots show at least some cell switching (see Methods). Previous integrase induction by β-estradiol was detectable around 48 h, with more than 50% of seedlings fully switched by 72 h.[20] In our experiments, we detected 100% switching by 48 h. The efficiency of switching

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

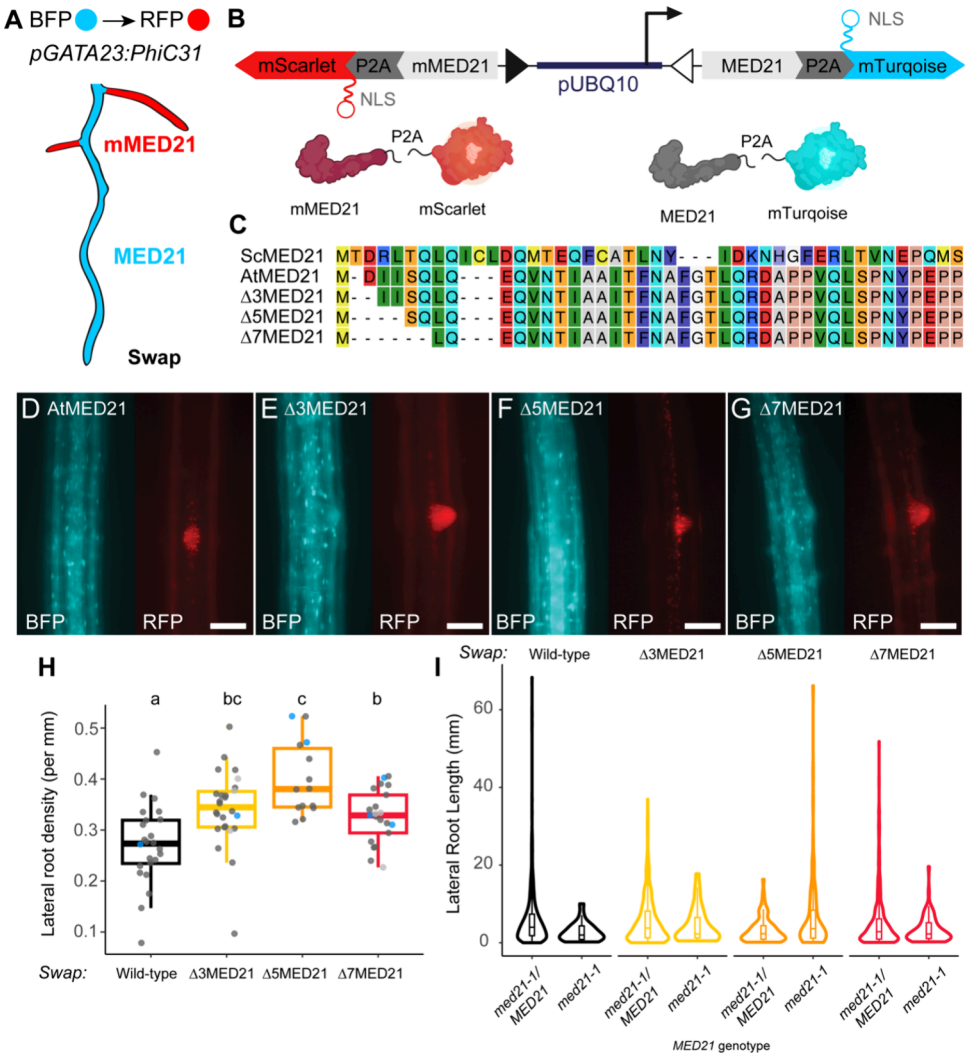


Figure 4. Engineering and rapid prototyping of the "hot-swap" isoform switch approach. (A) Schematic of the predicted behavior for the MED21 Swap. Lateral root primordia will switch the expression of the rescue construct from wild-type (blue) to mutant isoform (red). (B) Design of the integrase target. The target is composed of two integrase sites (triangles) surrounding a constitutive promoter (pUBQ10), and each MED21 isoform is fused to the P2A self-cleaving peptide sequence and a fluorescent reporter (mTurquoise or mScarlet). In the basal state, MED21 and mTurquoise are produced. Once the PhiC31 integrase is expressed, it mediates inversion of the DNA between the integrase sites, inverting the promoter and leading to production of mMED21 and mScarlet. (C) Alignment of the AtMED21 N-terminal mutants. (D−G) Epifluorescence microscopy analysis of wild-type and mMED21 swap lateral root initiation in *Arabidopsis* primary transformants. Microscopy images were taken on a 20x objective, and the scale bar represents 50 μm. (H) Lateral root density (LRD per mm) and (I) lengths of lateral roots (Lateral root length in mm) from primary transformant lines were quantified at 14 days post germination from the indicated Integrase Swap type. At least 15 independent primary transformant lines were tested for each swap type. Lines were transplanted to soil and PCR genotyped for the *med21* genotype. In (H), the LRD data points from med21 homozygous plants are colored in blue, while heterozygotes are colored dark gray, and ungenotyped samples are colored light gray. Letters indicate significant difference (ANOVA and Tukey HSD multiple comparison test; $p < 0.001$). In (I), data are presented as violin plots with nested boxplots to demonstrate that medians and interquartile ranges are comparable across conditions.

in both tagged and untagged MED21 iEraser lines was similar, with little nonspecific switching found (Figure S2B). The rate of uninduced lines demonstrating leaky switching is consistent with our own observations that the β-estradiol inducible system (XVE, a transcriptional activator composed of a DNA-binding domain of LexA, the transcription activation domain of VP16, and the regulatory region of the human estrogen receptor)[27] has a low but detectable transcriptional leakiness and highlights the importance of thorough screening of multiple independent lines. MED21:HA iEraser seedlings demonstrated increased root hair development (Figure 3B,C) and growth reductions following estradiol treatments compared with controls lacking β-estradiol (Figure 3D).

We characterized the expression of *MED21* in 12 independent iEraser lines, using the HA epitope tag. All lines had detectable protein expression (Figure 3E), so we focused further characterization of the MED21 iEraser on the progeny of one of these lines. Loss of MED21:HA could be detected in root and shoot tissues isolated from iEraser seedlings exposed to β-estradiol for 6 days, although a low level of protein was still visible on Western blots (Figure 3F). It is possible that the UBQ10 promoter drives high MED21:HA protein levels and therefore longer MED21:HA turnover times, and this could be tuned down by the addition of RNA destabilization tags to achieve lower abundances.[20] This raises the possibility that β-estradiol may not be reaching a subset of cells, as has been seen

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

for certain cell types like meristematic tissues.[28,29] To test this hypothesis, we took advantage of the irreversible nature of the integrase-mediated switch. We reasoned that if all cells were switched in a parent plant, then the cells of the subsequent gametophyte and germ cells would also be switched, leading to a fully switched seedling in the next generation. Instead, we observed no switching in the progeny of seedlings exposed to β-estradiol (Figure S1D), indicating that at least a subset of meristematic cells must escape the induction of PhiC31.

Significant growth reduction in both primary and lateral roots could be measured in MED21 iEraser seedlings treated with β-estradiol (Figure 3D). This phenotype is consistent with the predicted impact of the loss of this essential gene. As expected, we found that *pβ-estradiol::PhiC31* phenocopied the reductions in lateral root length (Figure 3G) observed in *pGATA23::PhiC31*. Consistent with the difference in expected expression patterns of each promoter, *pGATA23* Eraser lines had a similar primary root length to wild type, while primary root length in iEraser plants was sharply reduced (Figure 3H).

To extend the Integrase Eraser platform, we engineered a new technology that we termed Swap, which uses integrase-based recombination to switch between sequence variants (isoforms) of the same gene. We built a prototype of Swap that enabled a switch from the expression of wild-type MED21 to the expression of a mutant form of MED21 (mMED21) selectively in lateral roots (Figure 4A, Plasmid Maps pNL4174-4177). We employed a golden gate assembly approach to build our isoform target constructs, reminiscent of what we used to build our original constructs in Figure 1. The Swap target has a constitutive promoter (pUBQ10) flanked by integrase sites, positioned between the MED21 variants: wild-type MED21 marked with mTurquoise and a mutant form of MED21 marked with mScarlet. To avoid altering protein function through direct fusions, we instead introduced a P2A sequence between the MED21 and reporter sequences (Figure 4B). The P2A peptide induces ribosomal skipping,[30] resulting in untagged versions of both MED21 variants and reporters.

We selected mutant alleles of MED21 that are known to disrupt its binding with TPL (Figure 4C). Previous work in our lab has shown that deletions in the first seven amino acids of MED21 compromise TPL-binding activity and interfere with TPL-based repression.[10] Ectopic expression of any of these mutant *med21* isoforms in xylem pole pericycle cells competes with wild-type MED21, resulting in increased lateral root density.[10] Here, we created four Swaps, in which each wild-type MED21 is switched to a different *med21* mutant (Δ3MED21, Δ5MED21, Δ7MED21), in addition to the control, which has a wild-type allele on either side of the switch (Figure 4C). We performed rapid prototyping in *Nicotiana benthamiana* and found that switching of all Swap targets depended on the addition of the integrase (Figure S3A). Once exposed to integrase, Swap targets successfully switched, as indicated by expression of mScarlet.

We introduced the MED21 isoform switch targets into *med21−1/MED21, pGATA23::PhiC31* plants and performed fluorescence microscopy to identify primary transformant lines (T1) with specific switching to mScarlet in only the lateral roots (Figures 4E−G and S3B−E). We hypothesized that plants with mMED21 isoform switches should lead to increased lateral root density, as was observed previously (Leydon et al.,[10]). Additionally, the mMED21 isoforms should retain their activation activity in the Mediator complex, leading to more normal lateral root lengths compared to the short lateral root phenotype observed in the MED21 Eraser experiments (Figures 2 and 3). Therefore, we quantified lateral root density in primary transformant mMED21 isoform switch lines (lateral root density, Figure 4H). We observed that at mMED21 N-terminal mutation lines demonstrated higher lateral root density (Figure 4F) and that these lateral roots were similar in length to the wild type (Figure 4I).

CONCLUSIONS

Essential genes are estimated to comprise approximately 20% of the genome,[31] and further techniques to study, modify, and engineer these genes are needed to better understand and model their functions. Integrases are a promising tool, especially when compared to artificial microRNAs[32] or CRISPR/Cas9-based tissue-specific knock out (TSKO),[33] which suffer from low efficiency. In addition, integrase-driven recombination events are highly precise and less likely to produce the heterogeneity and diversity of repair-based mutations that result from Cas-based cleavage. While the tyrosine integrases such as Cre recombinase have been used more extensively in *Arabidopsis* and other plants to perform lineage tracing,[34,35] implement logic circuits[36] and study an essential gene HAP2/GCS1,[37] there are several advantages that make serine integrases attractive. One clear advantage is that serine integrases bind to two distinct DNA-binding sites, making their modification of DNA irreversible. This is contrary to tyrosine integrases, such as FLP and Cre, which act on two identical sites and are able to catalyze both a forward and reverse reaction.[38] Another limitation of tyrosine integrase application in plants has been the observation of low efficiency activity due to CHH context methylation of the tyrosine recombinase sites.[39]

The current major drawback to the Integrase Eraser approach is the requirement for a heterozygote mutation in the gene of interest; however, large numbers of mutants from gametophyte[40] and embryo-lethal[19] studies are ripe for reanalysis with such techniques. Another engineering constraint is that the high efficiency of integrase activity can lead to "overswitching" where recombination occurs in cells expressing only a background level of promoter expression. This potential challenge can be addressed by screening multiple lines, as is standard protocol for characterizing transgenic lines, and by using the integrase tuning toolkit described previously[20] to attach protein degrons to the integrase, or RNA destabilization tags (DSTs) to its transcript. As a new suite of publicly available tools that directly address the obstacles of pleiotropy and essentiality, Integrase Erasers should enable researchers to gain experimental access to whole new swaths of the genome that have stood just out of reach.

METHODS

Cloning. Our cloning strategy was based on Golden Gate assembly using appropriate spacers and BsaI-HFv2 (NEB) as the restriction enzyme. The *Arabidopsis* MED21:HA sequence was amplified by PCR with appropriate Golden Gate restriction sites and the construction of integrase targets was performed by Golden Gate reaction in the modified pGreenII-Hygr vector containing compatible Golden Gate sites defined in Guiziou et al.[20] Plasmids used for construction of the novel targets are available at Addgene: 5 prime side mScarlet cassette - L0-T4 tRBSC_mScarlet (Addgene #195887), Integrase site flanked pUBQ10 - L0-T1 Target PhiC31−(Addgene

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

#195890). Control target DNA - T1P/pUBQ10(mTurq→mScarlet) - (Addgene #195945). Each target was constructed into the Plant MoClo backbone vector pICH86966 (Addgene #48075), see Plasmid Maps below. pGATA23 driven PhiC31—L1_GATA23_1—(Addgene #195913) and estradiol inducible promoter: p35S_XVE-PhiC31-tUBQ1−(Addgene # 195932) were made previously. Enzymes for Golden Gate assembly were purchased from New England Biolabs (NEB, Ipswich, MA, USA). PCR was performed using 2X Q5 PCR master mix (NEB) and GoTaq master mix for colony PCR (Promega, Madison, WI, USA). Primers were purchased from IDT (Louvain, Belgium).Sequences were verified with Sanger sequencing by Azenta Life Sciences (Seattle, USA). Chemically competent cultures of the *E. coli* strain DH5alphaZ1 (laciq, PN25-tetR, SpR, deoR, supE44, Delta(lacZYA-argFV169), Phi80 lacZDeltaM15, hsdR17(rK−, mK+), recA1, endA1, gyrA96, thi-1, relA1) were transformed with plasmid constructs containing kanamycin resistance. Transformed *E. coli* was grown in LB media (LB broth, Miller) with kanamycin (Millipore Sigma, 50 μg/mL).

Western Blot. Protein was collected into 1.5 mL tubes with one steel bead per tube (MN Beads Type D−3 mm steel beads, Machery-Nagel), snap frozen in liquid nitrogen, and homogenized on max settings on a tissue homogenizer (MM400, Retsch). Homogenized tissue was resuspended in 2x sample buffer, boiled for 10 min, and spun down before being run on handmade 10% acrylamide SDS-PAGE gels and Western blotted with anti-HA-HRP antibodies from Roche/Millipore Sigma (RRID:AB_390917, REF-12013819001, Clone 3F10), or anti-Actin antibody (Abcam, ab197345). In all experiments, an equivalent mass of protein samples was used for extraction for the compared sample types, i.e., root and shoots.

Plant Growth. *Arabidopsis* seedlings were sown in 0.5× Linsmaier and Skoog nutrient medium (LS) (Caisson Laboratories) and 0.8% w/v agar, stratified at 4 °C for 2 days, and grown in constant light at 22 °C. Phyto agar (PlantMedia/bioWORLD) was used when imaging seedlings, and Bacto agar (ThermoFisher) was otherwise. For *Arabidopsis thaliana* experiments T2 plant lines harboring T-DNAs for MED21 (*med21−1*, WiscDsLox461-464 K13) were grown on media supplemented with 25 μg/mL Glufosinate-ammonium (Oakwood Chemical, SC).

Construction and Selection of Transgenic *Arabidopsis* Lines. *Agrobacterium tumefaciens* strain GV3101 was transformed by electroporation and subsequently grown in LB media with rifampin (Millipore Sigma, 50 μg/mL), gentamicin (Millipore Sigma, 50 μg/mL), any antibiotics carried on the specific plasmid(s), most often kanamycin (Millipore Sigma, 50 μg/mL). The floral dip method[41] was used to generate integrase target lines in Col-0, and then used to introduce each integrase construct into these established target lines. For T1 selection: 120 mg of T1 seeds (∼2000 seeds) were sterilized using 70% ethanol and 0.05% Triton-X-100 and then washed using 95% ethanol. Seeds were resuspended in 0.1% agarose and spread onto 0.5X LS selection plates using 25 μg/mL kanamycin for target lines and 25 μg/mL kanamycin and 25 μg/mL hygromycin for lines with both the integrase and the target. The plates were stratified at 4 °C for 48 h, then light pulsed for 6 h, and covered for 48 h. They were then grown for 4−5 days. To select transformants, tall seedlings with long roots and a vibrant green color were picked from the selection plate with sterilized tweezers and transferred to a new 0.5X LS Phyto agar plate for characterization.

Imaging of Reporter and Integrase Lines. T1 seedlings for each line were grown 4−5 days after transformant selection. Each selected seedling was imaged at 20× magnification using an epifluorescence microscope (Leica Biosystems, model: DMI 3000) using the RFP (exposure 700 ms, gain 2) and CFP (exposure 700 ms, gain 2) channels. Selected T1 seedlings were then transferred to soil, and at maturation, T2 seeds were selected. For later generations, seedlings were sterilized similarly to T1s, stratified, plated on an LS agar plate, grown for 4−5 days, and characterized using the epifluorescence microscope as for T1. Confocal Imaging of the Eraser seedlings were performed using Nikon A1R HD25 laser scanning confocal microscope with 561 laser and 578−623 detector for RFP imaging. For the reporter lines, seedlings were scanned to find early developed lateral roots. Imaging was processed using FIJI.[42] For each imaging, a Z-stack was recorded, and a maximum average of the Z-stack in the RFP channel was generated.

Root Growth Phenotyping. 6−8 T2 seedlings from five selected *A. thaliana* lines were grown on antibiotic LS Phyto plates for 14 days. Each plate was scanned on days 10, 11, 12, 13, and 14 of growth by using a flatbed scanner (Epson America, Long Beach, CA). Using FIJI, every lateral and primary root was measured in units of millimeters, and every individual lateral root was counted. Then for each of the selected seedlings, we divided the number of lateral roots on the primary root by the root length. The data was graphed through the use of R studio software. Fluorescence microscopy was used to identify and quantify the number of emerging lateral roots for every seedling.

For estradiol induction in T2s, antibiotic selection was performed as described in the section about *A. thaliana* transgenic lines. Four days after transplanting Kanamycin, Hygromycin resistant seedlings onto 0.5X LS Phyto plates, the seedlings were imaged via microscopy in RFP channels with identical settings as described in the method section about integrase switch seedling characterization. Then the seedlings were transferred onto new 0.5X LS Phyto plates with 10 μM β-estradiol. Each seedling was imaged 24, 48, and 72 h after transfer by both epifluorescence (for switching phenotype) and on a flatbed scanner as described above (for root growth phenotyping). A seedling was considered to exhibit a switch when any nuclear RFP was detected in the root tip region (manual scanning at least 10 mm up each individual root tip). No distinction was made as to the proportion of cells switched, and unswitched seedlings were required to show no evidence of nuclear RFP.

RNA Analysis. Seedlings for RNA analysis were grown for 14 days before collection of entire roots or microdissected lateral roots that were homogenized by bead beating. The RNA was extracted from these tissues using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany), including an on-column DNA digestion procedure RNase-Free DNase Set (Cat. No. 79254, Qiagen, Hilden, Germany). Complementary DNA (cDNA) was obtained with iScript cDNA synthesis kit (Bio-Rad, Hercules CA). We analyzed expression levels of the gene of interest with qPCR. The qPCR reaction was carried out on 2 μL of 1/5-diluted cDNA using 2X SYBR Super Mix (Bio-Rad, Hercules CA) with locus-specific primers. We used the reference gene to normalize expression in transgenic *Arabidopsis* for SERINE/THREONINE PROTEIN PHOS-

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

PHOTASE 2A (PP2A). The qPCR primers were as follows: PP2A (F GCGGTTGTGGAGAACATGATACG, R GAACCAAACACAATTCGTTGCTG), mScarlet (F ACCTTGAACCGCATGAACTC, R AGAGGAAAGTCGGAAGTGGTAG), MED21:HA (F ACAAGCTGCTGACAATTGCC, R AGCAGCTCTAGTGGCGTAATC). All primers were optimized for >90% linear efficiency.

Protein Alignments. The MED21 protein sequence was aligned to homologues using CLC Sequence Viewer 7 (QIAGEN, Aarhus, Denmark), a tree was constructed using a neighbor-joining method, and bootstrap analysis performed with 10,000 replicates.

Quantification and Statistical Analysis. All quantification and statistical analyses were performed in R (4.4.2), and the corresponding code has been deposited into GitHub: https://github.com/achillobator/Hot-Swappable_Genetic_Switch

Plasmid Maps. pNL3779_L2_mScarlet-pUBQ10-MED21-HA — https://benchling.com/s/seq-Db4peeJ0ujTdVFozrQrZ?m=slm-V2AY2UyqGBWm9GCGQFgp

pNL3857_L2_mScarlet-pUBQ10-MED21—https://benchling.com/s/seq-2733YDOGQuioeAFeeMtF?m=slm-vZcALggBA2HcJQ27XG0m

pNL4174_L2_ mScarlet—P2A-MED21-pUBQ10-MED21-P2A-mTurq—https://benchling.com/s/seq-GUNlRpAHMR84NYfQsDux?m=slm-PGEFPSnUIDR0XS4CEpD3

pNL4175_L2_ mScarlet—P2A-del3MED21-pUBQ10-MED21-P2A-mTurq—https://benchling.com/s/seq-ZBgrtLvyusbezOjT5XJc?m=slm-Em28XAtBr4AWN9bXCajj

pNL4176_L2_ mScarlet—P2A-del5MED21-pUBQ10-MED21-P2A-mTurq—https://benchling.com/s/seq-RJrmwYXp8ZWQ1pxBfLv2?m=slm-AQekxodEH0LI71MDA1Z5

pNL4177_L2_ mScarlet—P2A-del7MED21-pUBQ10-MED21-P2A-mTurq—https://benchling.com/s/seq-lB7kvjDcc7z8euYfNX1N?m=slm-usEImXgg6j4Gl00Sag89

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.5c00085.

Confirmation that the pUBQ10 promoter behaves unidirectionally in *Arabidopsis* (Supplemental Figure 1); characterization of independent MED21 iEraser lines, quantification of microscopy analysis (Supplemental Figure 2); characterization of swap targets in plants, in microscopy of rapid prototyping in *Nicotiana benthamiana*, and in *Arabidopsis thaliana* primary transformants (Supplemental Figure 3) (PDF)

S5. Supplemental DNA Sequences. Plasmid maps of all PhiC31 target constructs in FASTA and Genbank format (ZIP)

S4. Supplemental Movie 1. Time-lapse imaging of MED21 Eraser root growth phenotype (AVI)

■ AUTHOR INFORMATION

Corresponding Author

Alexander R Leydon − *Department of Biology, University of Washington, Seattle, Washington 98195-1800, United States;* orcid.org/0000-0003-3034-1482; Email: aleydon@uw.edu

Authors

Isabella J. Watson − *Department of Biology, University of Washington, Seattle, Washington 98195-1800, United States;* orcid.org/0009-0003-1848-4787

Cassandra Maranas − *Department of Biology, University of Washington, Seattle, Washington 98195-1800, United States*

Jennifer L. Nemhauser − *Department of Biology, University of Washington, Seattle, Washington 98195-1800, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.5c00085

Author Contributions

Conceptualization: A.R.L., J.L.N., Methodology: I.J.W., A.R.L., Software: I.J.W., A.R.L., Validation: I.J.W., A.R.L., Formal analysis: I.J.W., A.R.L., Investigation: I.J.W., C.J.M., A.R.L., Resources: J.L.N., Data Curation: I.J.W., A.R.L., Writing—Original Draft: I.J.W., Writing—Review and Editing, I.J.W., C.J.M., A.R.L., J.L.N., Visualization: I.J.W., C.J.M., A.R.L., Supervision: A.R.L., J.L.N., Project administration: A.R.L., J.L.N., Funding acquisition: A.R.L., J.L.N.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank current and former members of the Nemhauser group, including Dr. Sarah Guiziou, Janet Solano Sanchez, Benjamin Downing for constructive discussions. We thank Dr. Takato Imaizumi, Dr. Adam Steinbrenner, and Dr. Veronica Di Stillio for insightful suggestions. This work was supported by the NIH (R01-GM107084 and R35-GM148135-01 to JLN) and a Faculty Scholar Award from the Howard Hughes Medical Institute (to J.N.L.), A.R.L. was supported as a Simons Foundation Fellow of the Life Sciences Research Foundation.

■ REFERENCES

(1) Cohen, O.; Oberhardt, M.; Yizhak, K.; Ruppin, E. Essential Genes Embody Increased Mutational Robustness to Compensate for the Lack of Backup Genetic Redundancy. *PloS One* **2016**, *11* (12), No. e0168444.

(2) He, X.; Zhang, J. Why Do Hubs Tend to Be Essential in Protein Networks? *PLoS Genet.* **2006**, *2* (6), No. e88.

(3) Rancati, G.; Moffat, J.; Typas, A.; Pavelka, N. Emerging and Evolving Concepts in Gene Essentiality. *Nat. Rev. Genet.* **2018**, *19* (1), 34−49.

(4) Chen, J.; Yang, S.; Fan, B.; Zhu, C.; Chen, Z. The Mediator Complex: A Central Coordinator of Plant Adaptive Responses to Environmental Stresses. *Int. J. Mol. Sci.* **2022**, *23* (11), 6170.

(5) Dolan, W. L.; Chapple, C. Conservation and Divergence of Mediator Structure and Function: Insights from Plants. *Plant Cell Physiol.* **2017**, *58* (1), 4−21.

(6) Maji, S.; Dahiya, P.; Waseem, M.; Dwivedi, N.; Bhat, D. S.; Dar, T. H.; Thakur, J. K. Interaction Map of Arabidopsis Mediator Complex Expounding Its Topology. *Nucleic Acids Res.* **2019**, *47* (8), 3904−3920.

(7) Malik, N.; Agarwal, P.; Tyagi, A. Emerging Functions of Multi-Protein Complex Mediator with Special Emphasis on Plants. *Crit. Rev. Biochem. Mol. Biol.* **2017**, *52* (5), 475−502.

(8) Jaskolowski, A.; Iñigo, S.; Arellano, S. M.; Arias, L. A.; Fiol, D. F.; Sede, A. R.; Oldra, M. B.; Lorenzi, H.; Muschietti, J. P.; Pagnussat, G. C.; Cerdán, P. D. The MED30 Subunit of Mediator Complex Is Essential for Early Plant Development and Promotes Flowering in Arabidopsis Thaliana. *Dev. Camb. Engl.* **2019**, *146* (10), dev175224.

(9) Zhang, L.; Guo, C. The Important Function of Mediator Complex in Controlling the Developmental Transitions in Plants. *Int. J. Mol. Sci.* **2020**, *21* (8), 2733.
(10) Leydon, A. R.; Wang, W.; Gala, H. P.; Gilmour, S.; Juarez-solis, S.; Zahler, M. L.; Zemke, J. E.; Zheng, N.; Nemhauser, J. L. Repression by the Arabidopsis TOPLESS Corepressor Requires Association with the Core Mediator Complex. *eLife* **2021**, *10*, No. e66739.
(11) Leydon, A. R.; Downing, B.; Sanchez, J. S.; Loll-krippleber, R.; Belliveau, N. M.; Rodriguez-mias, R. A.; Bauer, A.; Watson, I. J.; Bae, L.; Villén, J.; Brown, G. W.; Nemhauser, J. L. A Conserved Function of Corepressors Is to Nucleate Assembly of the Transcriptional Preinitiation Complex. bioRxiv April 1, 2024; p 2024.04.01.587599. .
(12) Dhawan, R.; Luo, H.; Foerster, A. M.; Abuqamar, S.; Du, H.-N.; Briggs, S. D.; Mittelsten scheid, O.; Mengiste, T. HISTONE MONOUBIQUITINATION1 Interacts with a Subunit of the Mediator Complex and Regulates Defense against Necrotrophic Fungal Pathogens in Arabidopsis. *Plant Cell* **2009**, *21* (3), 1000−1019.
(13) Li, Z.; Vizeacoumar, F. J.; Bahr, S.; Li, J.; Warringer, J.; Vizeacoumar, F. S.; Min, R.; Vandersluis, B.; Bellay, J.; Devit, M.; Fleming, J. A.; Stephens, A.; Haase, J.; Lin, Z.-Y.; Baryshnikova, A.; Lu, H.; Yan, Z.; Jin, K.; Barker, S.; Datti, A.; Giaever, G.; Nislow, C.; Bulawa, C.; Myers, C. L.; Costanzo, M.; Gingras, A.-C.; Zhang, Z.; Blomberg, A.; Bloom, K.; Andrews, B.; Boone, C. Systematic Exploration of Essential Yeast Gene Function with Temperature-Sensitive Mutants. *Nat. Biotechnol.* **2011**, *29* (4), 361−367.
(14) Haruki, H.; Nishikawa, J.; Laemmli, U. K. The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. *Mol. Cell* **2008**, *31* (6), 925−932.
(15) Gosink, M. M.; Vierstra, R. D. Redirecting the Specificity of Ubiquitination by Modifying Ubiquitin-Conjugating Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (20), 9117−9121.
(16) Hoeberichts, F. A.; Vaeck, E.; Kiddle, G.; Coppens, E.; Van de cotte, B.; Adamantidis, A.; Ormenese, S.; Foyer, C. H.; Zabeau, M.; Inzé, D.; Périlleux, C.; Van breusegem, F.; Vuylsteke, M. A Temperature-Sensitive Mutation in the *Arabidopsis Thaliana* Phosphomannomutase Gene Disrupts Protein Glycosylation and Triggers Cell Death*. *J. Biol. Chem.* **2008**, *283* (9), 5708−5718.
(17) Vidali, L.; Augustine, R. C.; Fay, S. N.; Franco, P.; Pattavina, K. A.; Bezanilla, M. Rapid Screening for Temperature-Sensitive Alleles in Plants. *Plant Physiol.* **2009**, *151* (2), 506−514.
(18) Winkler, J.; Mylle, E.; De meyer, A.; Pavie, B.; Merchie, J.; Grones, P.; Van damme, D. L. Visualizing Protein-Protein Interactions in Plants by Rapamycin-Dependent Delocalization. *Plant Cell* **2021**, *33* (4), 1101−1117.
(19) Meinke, D.; Muralla, R.; Sweeney, C.; Dickerman, A. Identifying Essential Genes in Arabidopsis Thaliana. *Trends Plant Sci.* **2008**, *13* (9), 483−491.
(20) Guiziou, S.; Maranas, C. J.; Chu, J. C.; Nemhauser, J. L. An Integrase Toolbox to Record Gene-Expression during Plant Development. *Nat. Commun.* **2023**, *14* (1), 1844.
(21) Banda, J.; Bellande, K.; von Wangenheim, D.; Goh, T.; Guyomarc'h, S.; Laplaze, L.; Bennett, M. J. Lateral Root Formation in Arabidopsis: A Well-Ordered LRexit. *Trends Plant Sci.* **2019**, *24* (9), 826−839.
(22) Maranas, C. J.; George, W.; Scallon, S. K.; Vangilder, S.; Nemhauser, J. L.; Guiziou, S. A History-Dependent Integrase Recorder of Plant Gene Expression with Single Cell Resolution. bioRxiv June 4, 2024; p 2024.06.04.597320. .
(23) De rybel, B.; Vassileva, V.; Parizot, B.; Demeulenaere, M.; Grunewald, W.; Audenaert, D.; Van campenhout, J.; Overvoorde, P.; Jansen, L.; Vanneste, S.; Möller, B.; Wilson, M.; Holman, T.; Van isterdael, G.; Brunoud, G.; Vuylsteke, M.; Vernoux, T.; De veylder, L.; Inzé, D.; Weijers, D.; Bennett, M. J.; Beeckman, T. A Novel Aux/IAA28 Signaling Cascade Activates GATA23-Dependent Specification of Lateral Root Founder Cell Identity. *Curr. Biol. CB* **2010**, *20* (19), 1697−1706.
(24) Silver, B. D.; Willett, C. G.; Maher, K. A.; Wang, D.; Deal, R. B. Differences in Transcription Initiation Directionality Underlie Distinctions between Plants and Animals in Chromatin Modification Patterns at Genes and Cis-Regulatory Elements. *G3 Bethesda Md* **2024**, *14* (3), jkae016.
(25) Yu, Z.; Zhang, F.; Friml, J.; Ding, Z. Auxin Signaling: Research Advances over the Past 30 Years. *J. Integr. Plant Biol.* **2022**, *64* (2), 371−392.
(26) Calderon-villalobos, L. I.; Tan, X.; Zheng, N.; Estelle, M. Auxin Perception—Structural Insights. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (7), No. a005546.
(27) Zuo, J.; Niu, Q. W.; Chua, N. H. Technical Advance: An Estrogen Receptor-Based Transactivator XVE Mediates Highly Inducible Gene Expression in Transgenic Plants. *Plant J. Cell Mol. Biol.* **2000**, *24* (2), 265−273.
(28) Hu, Y.; Omary, M.; Hu, Y.; Doron, O.; Hoermayer, L.; Chen, Q.; Megides, O.; Chekli, O.; Ding, Z.; Friml, J.; Zhao, Y.; Tsarfaty, I.; Shani, E. Cell Kinetics of Auxin Transport and Activity in Arabidopsis Root Growth and Skewing. *Nat. Commun.* **2021**, *12*, 1657.
(29) Borghi, L.; Gutzat, R.; Fütterer, J.; Laizet, Y.; Hennig, L.; Gruissem, W. Arabidopsis RETINOBLASTOMA-RELATED Is Required for Stem Cell Maintenance, Cell Differentiation, and Lateral Organ Production. *Plant Cell* **2010**, *22* (6), 1792−1811.
(30) De lima, J. G. S.; Lanza, D. C. F. 2A and 2A-like Sequences: Distribution in Different Virus Species and Applications in Biotechnology. *Viruses* **2021**, *13* (11), 2160.
(31) Lloyd, J. P.; Seddon, A. E.; Moghe, G. D.; Simenc, M. C.; Shiu, S.-H. Characteristics of Plant Essential Genes Allow for Within- and between-Species Prediction of Lethal Mutant Phenotypes. *Plant Cell* **2015**, *27* (8), 2133−2147.
(32) Tiwari, M.; Sharma, D.; Trivedi, P. K. Artificial microRNA Mediated Gene Silencing in Plants: Progress and Perspectives. *Plant Mol. Biol.* **2014**, *86* (1−2), 1−18.
(33) Decaestecker, W.; Buono, R. A.; Pfeiffer, M. L.; Vangheluwe, N.; Jourquin, J.; Karimi, M.; Van isterdael, G.; Beeckman, T.; Nowack, M. K.; Jacobs, T. B. CRISPR-TSKO: A Technique for Efficient Mutagenesis in Specific Cell Types, Tissues, or Organs in Arabidopsis[OPEN]. *Plant Cell* **2019**, *31* (12), 2868−2887.
(34) Efroni, I.; Mello, A.; Nawy, T.; Ip, P.-L.; Rahni, R.; Delrose, N.; Powers, A.; Satija, R.; Birnbaum, K. D. Root Regeneration Triggers an Embryo-like Sequence Guided by Hormonal Interactions. *Cell* **2016**, *165* (7), 1721−1733.
(35) Wachsman, G.; Heidstra, R.; Scheres, B. Distinct Cell-Autonomous Functions of RETINOBLASTOMA-RELATED in Arabidopsis Stem Cells Revealed by the Brother of Brainbow Clonal Analysis System. *Plant Cell* **2011**, *23* (7), 2581−2591.
(36) Lloyd, J. P. B.; Ly, F.; Gong, P.; Pflueger, J.; Swain, T.; Pflueger, C.; Fourie, E.; Khan, M. A.; Kidd, B. N.; Lister, R. Synthetic Memory Circuits for Stable Cell Reprogramming in Plants. *Nat. Biotechnol.* **2022**, *40* (12), 1862−1872.
(37) Nagahara, S.; Takeuchi, H.; Higashiyama, T. Generation of a Homozygous Fertilization-Defective Gcs1Mutant by Heat-Inducible Removal of a Rescue Gene. *Plant Reprod.* **2015**, *28* (1), 33−46.
(38) Olorunniji, F. J.; Rosser, S. J.; Stark, W. M. Site-Specific Recombinases: Molecular Machines for the Genetic Revolution. *Biochem. J.* **2016**, *473* (6), 673−684.
(39) Liu, R.; Long, Q.; Zou, X.; Wang, Y.; Pei, Y. DNA Methylation Occurring in Cre-Expressing Cells Inhibits loxP Recombination and Silences loxP-Sandwiched Genes. *New Phytol.* **2021**, *231* (1), 210−224.
(40) Johnson, M. A.; Von besser, K.; Zhou, Q.; Smith, E.; Aux, G.; Patton, D.; Levin, J. Z.; Preuss, D. Arabidopsis Hapless Mutations Define Essential Gametophytic Functions. *Genetics* **2004**, *168* (2), 971−982.
(41) Clough, S. J.; Bent, A. F. Floral Dip: A Simplified Method for Agrobacterium-Mediated Transformation of *Arabidopsis Thaliana*. *Plant J.* **1998**, *16* (6), 735−743.
(42) Schindelin, J.; Arganda-carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.;

Schmid, B.; Tinevez, J.-Y.; White, D. J.; Hartenstein, V.; Eliceiri, K.; Tomancak, P.; Cardona, A. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* **2012**, *9* (7), 676−682.

https://doi.org/10.1021/acssynbio.5c00085
ACS Synth. Biol. 2025, 14, 2170−2180

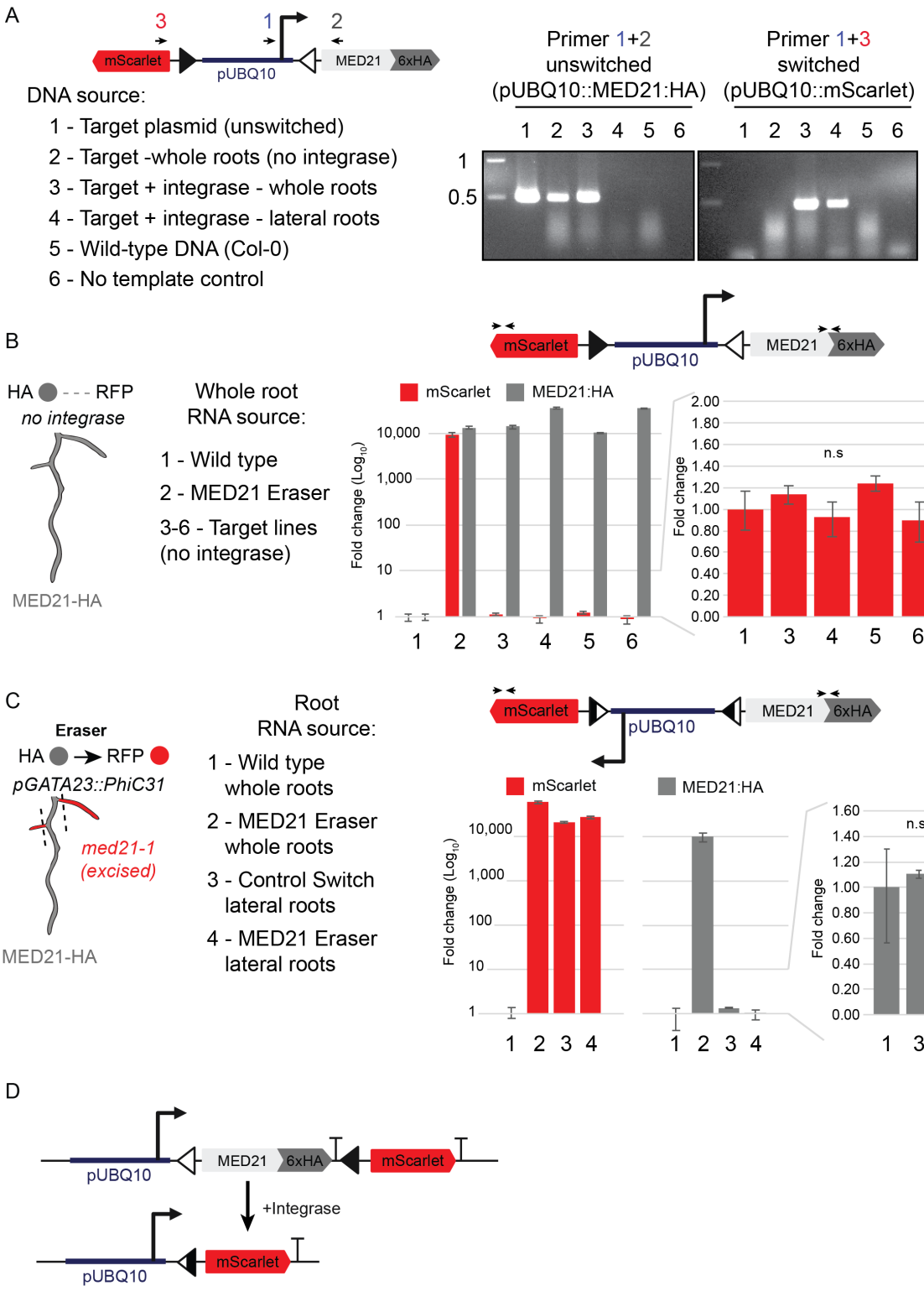
Supplemental Information for:

A Hot-Swappable Genetic Switch: Building an inducible and trackable functional assay for the essential gene MEDIATOR 21

Isabella J. Watson[1], Cassandra Maranas[1], Jennifer L Nemhauser[1], Alexander R Leydon[1*]

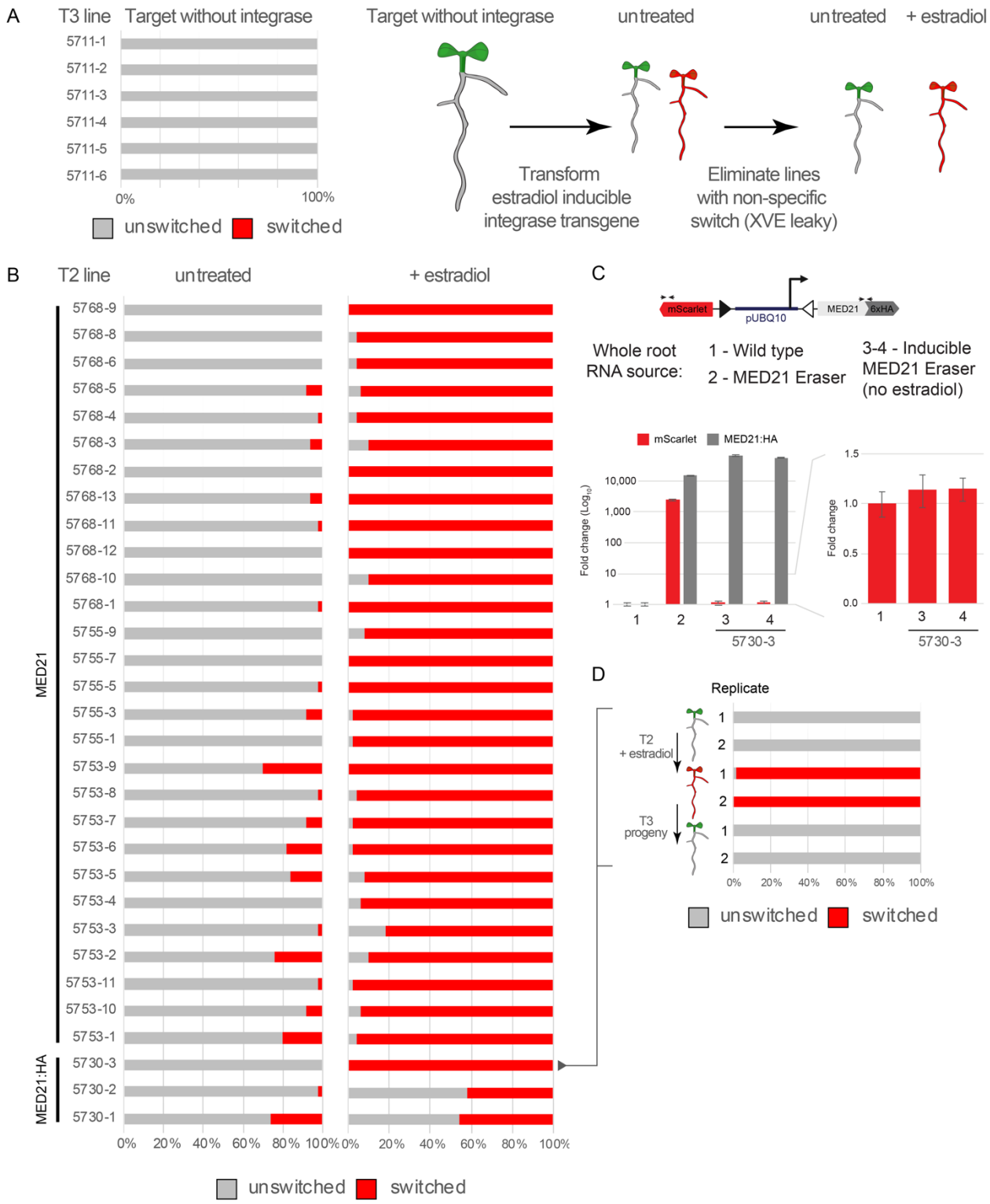
[1]Department of Biology, University of Washington, Seattle, WA 98195-1800 USA

[*]E-mail: aleydon@uw.edu



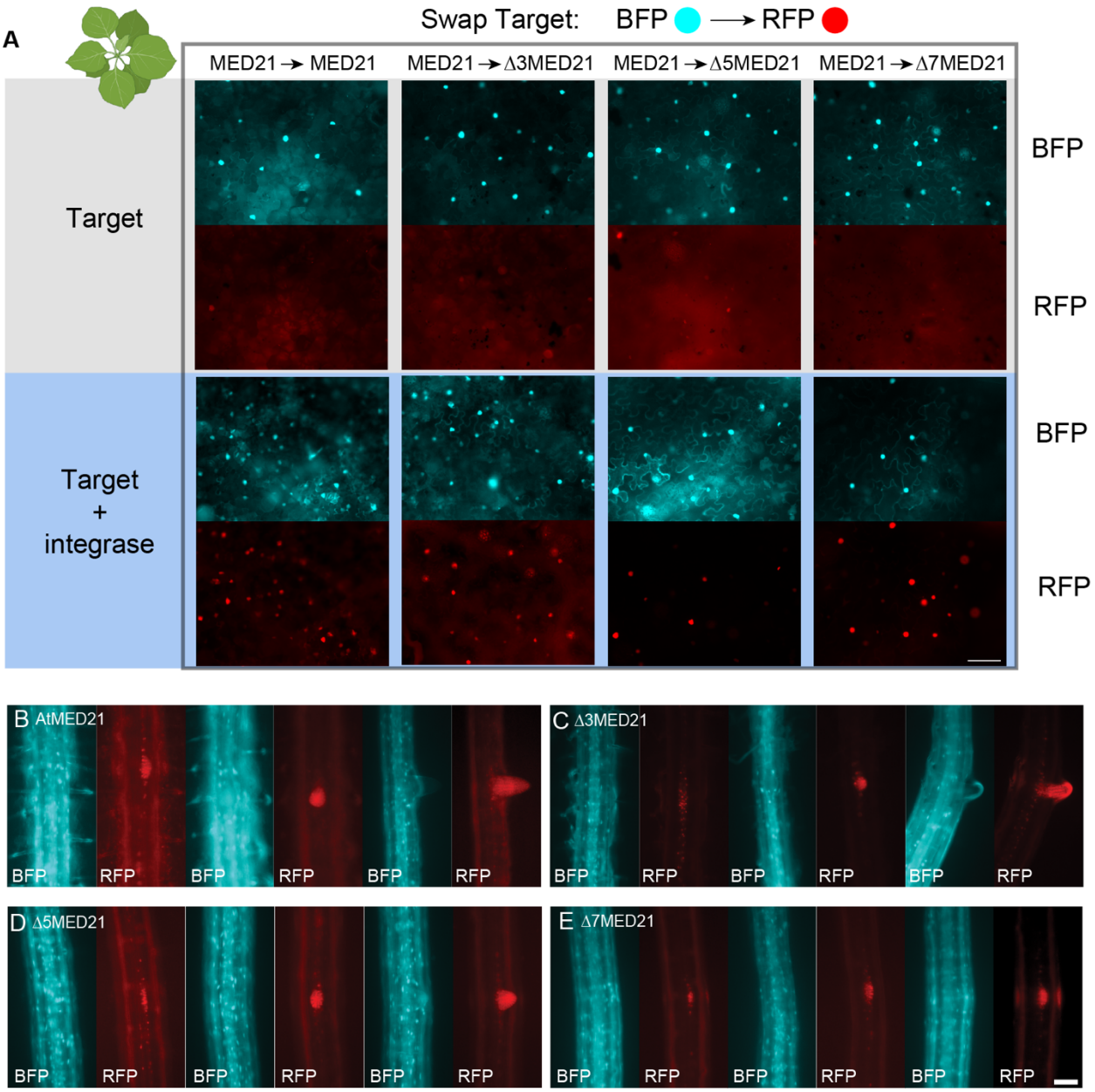
S1. Confirmation that the pUBQ10 promoter behaves unidirectionally in Arabidopsis. A. Schematics of the primer binding on the unswitched target. The PCR of the unswitched target would result in a band for primers 1+2 (left DNA gel) whereas PCR of the switched target would result in a band for primers 1+3 (right DNA gel). **B.** RT-qPCR experiments performed on whole roots containing only the MED21:HA Eraser

target. MED21:HA Eraser target lines with no integrase were compared to either wild type (no target) and MED21:HA eraser lines that contain both the eraser and the *pGATA23:PhiC31*integrase construct, where both states should exist in whole roots. Whole roots were excised, and RNA was extracted, cDNA was generated, and RT-qPCR was performed. Four biological replicates of 8 14-day old seedlings each were tested, and strong MED21:HA RNA expression was detected. mScarlet expression was not statistically different than background detected from the wild type (Col-o) seedling controls (n.s. not significantly different - ANOVA). **C.** RT-qPCR experiments performed on dissected lateral roots containing only the MED21:HA Eraser target. For lateral root dissections, 4 biological replicates of 10 seedlings each were grown for 14 days and then manually dissected under a dissecting microscope and pooled for RNA extraction. MED21:HA expression was not statistically different than background detected from the wild type (Col-o) seedling controls (n.s. not significantly different - ANOVA). For B & C. The first graph has the expression (y-axis) as fold change in $\log_{10}$ to allow visualization of the highly expressed gene that is downstream of the pUBQ10 promoter, while the second graph zooms in using only log fold change. In both experiments the experimental primers are normalized to the *AtPP2AA3* subunit as the reference gene. **D.** Schematic of an excisable version of the MED21 eraser construct. In this orientation the rescue construct can be completely excised from the genome upon action of the integrase, yielding a “zero background” design.



S2. Characterization of independent MED21 iEraser lines. A. To highlight the variability in the leakiness of the XVE control system, we first screened stable transformants carrying the MED21 Eraser target for red fluorescence and observed absolutely no background switching (>100 seedlings for each line). We then

transformed the XVE:PhiC31 controlled expression cassette into plants and screened these T2 lines for any background non-specificity in the response to estradiol as shown in the flowchart to the right. **B.** Characterization of T2 seedlings by fluorescence microscopy, data is quantified by stacked bar plots. For the untagged MED21 eraser construct, we selected 28 T1 lines that were Kanamycin and Hygromycin resistant, and characterized 50 T2 seedlings per T1 line. Three MED21:HA eraser lines were included as controls (bottom three lines). Seedlings were seeded on Kanamycin and Hygromycin plates and grown for 5 days before being sprayed with 1μM estradiol. Plants were screened for RFP by fluorescence microscopy before and after application (3 days) and scored as either unswitched (gray) or switched (red). Data is presented as stacked bar graphs to highlight the proportion of the population undergoing a switch. **C**. RT-qPCR experiments performed on whole roots containing the MED21:HA Eraser target and the XVE:PhiC31 controlled expression cassette in the absence of estradiol. Two biological replicates were tested pooling whole root tissue from 8 seedlings for each replicate from 14 day old seedlings. The first graph has the expression (y-axis) as fold change in $\log_{10}$ to allow visualization of the highly expressed gene that is downstream of the pUBQ10 promoter, while the second graph zooms in using only log fold change. In both experiments the experimental primers are normalized to the *AtPP2AA3* subunit as the reference gene. **D**. Characterization of transmission of MED21:HA switch phenotypes to progeny. Switched seedlings for the highly specific line 5730-3 were transplanted (two replicates) to soil and allowed to produce seeds. These seeds were plated on media with no estradiol and screened for RFP by fluorescence microscopy. 50 seedlings were evaluated for each replicate of the experiment.



S3. Characterization of swap targets in plants. A. Rapid prototyping of the MED21 swap target in transient transfections of *Nicotiana benthamiana* at 2 days after injection. Top panel (gray backing) is the target alone. Bottom panel (blue backing) shows the impact of co-expressing the target with p35S:PhiC31. Microscopy images were taken on a 20x objective to allow a wide view of switching efficiency, the 50µm scale bar in the bottom right RFP channel applies to all images. **B-E.** Epifluorescence microscopy analysis of wild-type and MED21 swap lateral root initiation in *Arabidopsis* primary transformants. Microscopy images were taken on a 20x objective, and the 50µm scale bar in the bottom right RFP channel applies to all images.

S4. Supplemental Movie S1. Time-lapse imaging of MED21 Eraser root growth phenotype. Seedlings carrying the *pGATA23::PhiC31* driver were initially selected for Target and Driver constructs and T3 plants were confirmed to be homozygous for the *med21-1* mutation. Seedlings were grown for 5 days prior to hand transplantation on large square plates for time course analysis. All efforts were made to age and size match plants before transplantation. Images were taken every hour for 16 days, and the movie was cropped at the point at which the primary root neared the bottom of the plate. The genotypes are as follows: MED21 Eraser: 5750-5, 5754-9, MED21-HA Eraser: 5727-4, 5727-10, Control mTurquoise → mScarlet Target: 444.

Chapter 5: Auxin coordinates cell states during root development

Cassandra Maranas[1], Sydney VanGilder[1], Linda Nguyen[1], Jennifer Nemhauser*[1]

1 Biology department, UW

* corresponding author: jn7@uw.edu

ABSTRACT

Two genetically identical cells exposed to the same signals will have differences in gene expression. Despite this variability, multicellular development proceeds remarkably robustly in most cases. Cell-to-cell variation in gene expression can be highly detrimental and actively buffered out; however, in other contexts, it is crucial and actively amplified. For example, variation must be minimized to build organs with consistent size and shape, yet the initiation of organogenesis requires a subset of cells to take on a new fate, a process that often relies on small differences between cells. In plants, much of development is controlled by the hormone auxin, which has been hypothesized to coordinate cell responses by inducing degradation of transcriptional repressors. To quantify the level of cell-to-cell variation and directly test its connection to auxin signaling, we assessed variation in expression of a lateral root founder cell marker *GATA23* when auxin levels or responsiveness was modulated. We found that auxin acted as both an amplifier and a constrainer of transcriptional variation during the initiation of a new root. We then extended this work to analysis of root regeneration, where auxin was also found to play a critical role in coordinating cells during fate transitions.

INTRODUCTION

Stochastic cell-to-cell variation in gene expression is well established across organisms and cell type[1]. For many genes, cell-to-cell variation in expression is deleterious to functioning, and cellular mechanisms to buffer it are employed on the gene, transcript, and protein levels[2–6]. But there is evidence that expression variability is advantageous and even selected for in some cases[7] and many different regulatory strategies have been implicated in promoting it[8]. The importance of cell-to-cell variation has been established for many cell differentiation processes[9–11] because they often rely on existing cell-to-cell differences to trigger cell fate transitions in a subset of cells[12]. In fact, existing cell-to-cell variation can be amplified during organogenesis to form defined boundaries between the identities of differentiating and non-differentiating cells and ensure coordinated development[2]. Studies using single cell RNA sequencing have identified many genes associated with cell differentiation as being variably expressed[13,14].

In studies of cell differentiation trajectories, one idea that has been explored is that of a "transition cell state" in which a cell's gene expression profile bears similarities to both the undifferentiated and differentiated cell states and is more heterogeneous than either state[15,16]. Because of limitations in current experimental techniques for assessing transient cell states, an alternative approach is stochastic modeling of gene expression networks driving differentiation. One such study[17] developed a model of cell differentiation and transition state dynamics based on Waddington's epigenetic landscape[18] wherein cell states were represented as energy potentials, together forming a probabilistic developmental landscape, and cell differentiation was a cell's trajectory through this landscape. Changing the concentration of a signaling molecule distorted the energy landscape of the transition state and thus also distorted cell differentiation trajectories through it. Other work has also linked extracellular signaling to heterogeneous

transition state dynamics and cell differentiation trajectories[19,20] but the mechanisms describing how cells variably process these signals to make cell fate decisions are not yet clear.

Despite the established variability in cell states and differentiation trajectories, development is mostly predictable and robust, resulting in consistently formed tissues and organs. *Arabidopsis* sepal development has been a particularly useful model for studying the relationship between cell-to-cell heterogeneity and developmental robustness. First, in a landmark study, it was shown that when more cells are present (such as in a mature sepal), cell-to-cell variation in growth rates is spatiotemporally averaged, buffering these differences and enabling consistent sepal growth[4]. As such, initial cell-to-cell variability is resolved over time and space. Second, rapid and coordinated amplification of cell-to-cell variation is key to sepal initiation, but increasing the speed and sensitivity of the response comes at the expense of cell-to-cell coordination and robustness[21]. Third, if there is sufficient variation in the earliest stages of sepal development (patterning of sepal primordia sites by the plant hormone auxin), the lack of cell-to-cell coordination is propagated through developmental stages, resulting in inconsistently formed sepals[21,22]. Taken together, this work shows that proper development is predicated on the management of cell-to-cell variation, first amplifying it to initiate differentiation and later attenuating it to ensure robust progression through developmental stages.

Like sepals, lateral root (LR) development is governed by auxin signaling. Lateral roots are also an excellent model for understanding variation, as their initiation is regulated by one of the best understood plant morphogenetic processes and the roots themselves are completely dispensable in lab-grown plants[23]. LRs are initiated when a subset of xylem pole pericycle (XPP) cells receive a sufficiently high auxin signal. They then undergo reprogramming to take on an LR founder cell identity and begin to divide (Fig. 1A)[24]. The number of LR founder cells has been

historically described as 2-3[25]; however, more recent work has revealed that it can vary from as few as 2 to as many as 11[26]. In the absence of auxin, the expression of auxin-responsive genes is repressed by the TPL/TPR corepressors which inhibit the activity of AUXIN RESPONSE FACTOR (ARF) activators through association with coreceptors/adaptors from the AUXIN/INDOLE-3-ACETIC ACID (Aux/IAA) family (Fig. 1Bi). Auxin signaling facilitates rapid Aux/IAA degradation through formation of a complex between auxin, the Aux/IAA and auxin receptors like TRANSPORT INHIBITOR RESPONSE 1 (TIR1) (Fig. 1Bi).

The rate of auxin-induced degradation controls the speed of transcriptional responses[27], as well as the rate of LR morphogenesis[28]. We hypothesized that, beyond acting as a timer for events within a single cell, auxin might also act to coordinate the response of multiple cells undergoing parallel fate changes during development. To test this hypothesis, we used a variety of chemical and genetic tools to perturb auxin signaling and measure the impact on variation in initiation of a new root in the context of the primary root as well as during regeneration. Our results support a role for auxin as a coordinator of multi-cell behaviors during development.

RESULTS AND DISCUSSION

As a metric for variation in LR initiation and an early step in auxin signaling, we measured the number of LR founder cells. As *GATA TRANSCRIPTION FACTOR 23* (*GATA23*) expression is one of the earliest markers of LR founder cell identity[29], we counted the number of *GATA23*-expressing cells (GECs) in early stage LRs as a proxy for the number of founder cells (Fig. 1C). In our standard growth conditions, the overall distribution of GECs was roughly symmetrical, with a peak around 5, and a range of 2-8, consistent with previous studies[26,30,31]. The coefficient of variation (CV), a normalized expression of data spread, makes it possible to compare the extent of variation between different data sets. The CV of GECs in WT plants was 28% (Fig. 1D). We observed some seedling-to-seedling variation in the distribution of GEC number, but the median number of GECs was relatively consistent between seedlings (Fig. 1E). Thus, the observed variability in GEC number is inherent to the LR initiation process, as it is recapitulated in each seedling.

To assess the impact of dampening the auxin signal, we grew seedlings in the presence of low levels of auxinole, a compound that competes with auxin for binding with TIR1 (Fig1Bii)[32]. To find the ideal concentration for these assays, we quantified emerged lateral root density in a range of doses (Fig. S1). From this result, we selected 0.5 μM auxinole, as it was able to mildly reduce but not block lateral root development. The distribution of number of GECs in the auxinole-treated LRs was considerably wider than in the control (Fig. 1D), with the maximum increasing to 12 (1.5x the control). Auxinole treatment also increased the variability by approximately 20%, resulting in a CV of 34%. As in the control conditions, the median number of GECs for LRs within each seedling is relatively consistent (Fig. 1E). Auxinole shifted the distribution of CV across seedlings upwards, while maintaining a similar range with control

conditions (Fig. 1F). From these results, we can conclude that the correct level of auxin response is required to maintain robustness in the absolute number and variation in number of LR founder cells.

We were next interested in examning the impact of auxinole treatment in later stages of LR development. To capture LRs across all stages, we measured the width of each LR on each *GATA23* reporter seedling screened. Plotting these LR widths by stage (Fig. 1G) we found that early stage LRs were the most variable in width, with variation decreasing through each of the later stages. Treatment with auxinole resulted in increased average width and variation therein for early and mid-stage LRs, but in later stages the distribution grew closer to that of the control. This trend indicates that as LR development progresses, initial cell-to-cell variation is resolved, allowing robust root formation, even when the coordinating auxin signal is dampened by auxinole treatment. This trend could be a function of the increased number of cells in later stage LRs, allowing for spatiotemporal averaging of each cell's growth and cell cycle fluctuations, similar to what is observed in sepal development[4].

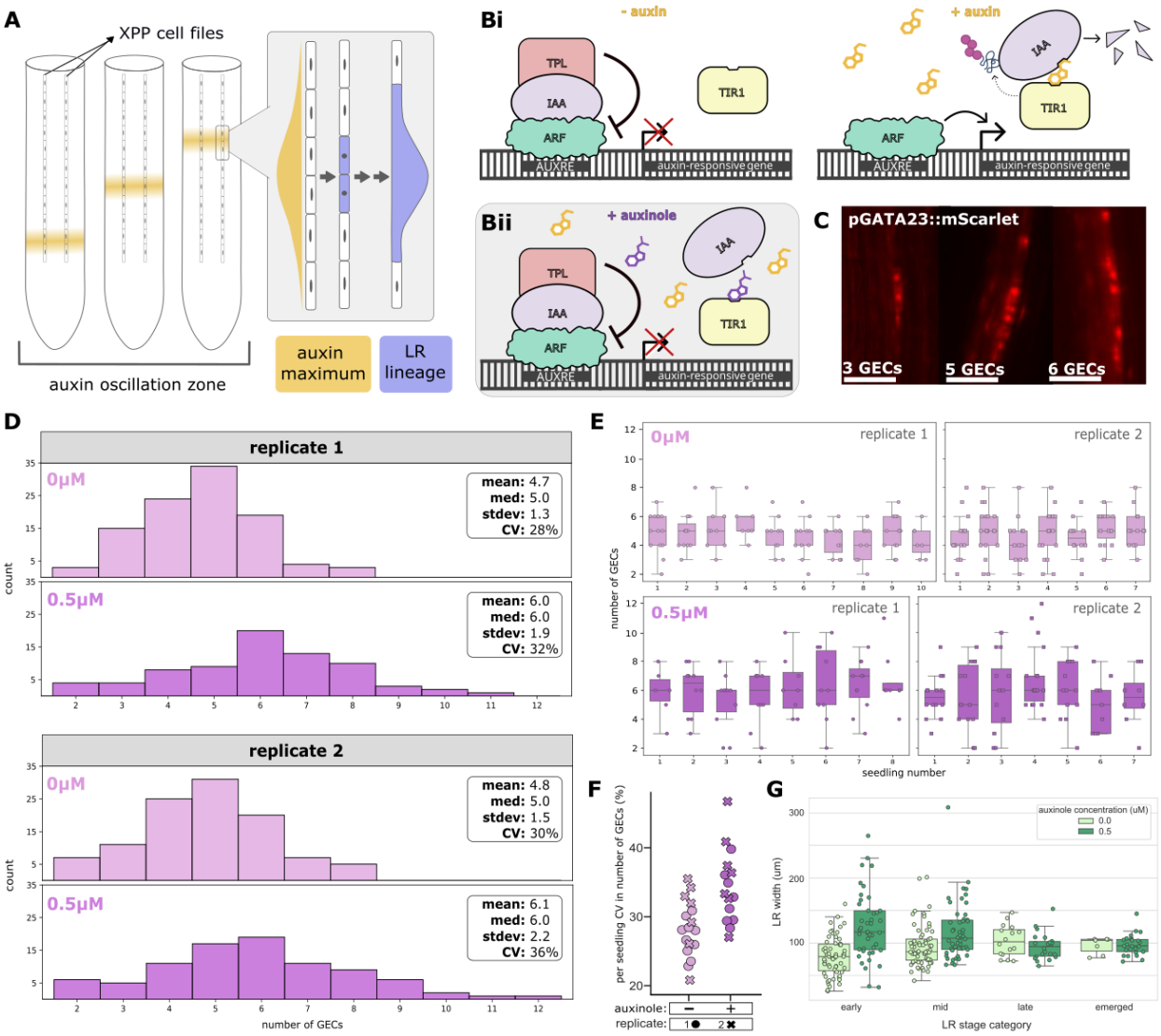


Figure 1: Dampening the auxin signal strength with auxinole treatment increases variability in the early stages of lateral root initiation. **A.** Specification of LR founder cells. (left) Auxin concentration oscillates along the length of the root defining LR branch sites. (right) A subset of XPP cells which are exposed to high auxin signal undergo a cell fate transition, becoming LR founder cells and dividing to form the LR. **Bi.** Auxin signaling schematics. (left) Without auxin present, TPL represses gene expression through contact with Aux/IAA. (right) When auxin is present, it mediates ubiquitination and degradation of Aux/IAA through contact with TIR1, relieving repression of auxin-responsive genes. **Bii.** With both auxin and auxinole present, auxinole competes for binding with TIR1, reducing IAA degradation rate and slowing relief of repression of auxin-responsive genes. **C.** Examples of counting GECs using the pGATA23::mScarlet reporter line. The GEC data is used to produce panels D-F. **D.** Overall distribution in GEC number in the control (light pink) and 0.5μM auxinole treatment (magenta) for two cell counting replicates. Summary stats are included for each distribution. **E.** Per seedling distribution in GEC number for control (top) and auxinole (bottom) treatments across two replicates. **F.** Per seedling CV in GEC number for the control and 0.5μM auxinole treatment. **G.** Overall distribution of LR widths by developmental stage. Stage categories are as follows – early: stages 1-2; mid: stages 3-5; late: stages 6-7, according to Banda et. al 2019[23]. For early stage LRs, width was measured based on the extent of GATA23 reporter expression. For the later LR categories, width was measured based on LR protrusion from the main root.

One model to explain our results is that a weakened auxin response is ineffective in repressing LR fate in cells adjacent to the founder cells, a role that has been previously documented[33]. This is consistent with the observation that auxinole treatment significantly increased the high end of the GEC distribution (Fig. 1D). Following this logic, unrepressed neighboring cells might have lower expression of LR-initiation genes like *GATA23* than cells at the core of the initiating LR. We tested this prediction with a previously characterized integrase-based durable recorder of *GATA23* expression, a tool we call an integrase switch[34]. The integrase switch has two components: (1) the target which can switch between expression of one gene to another by inverting the direction of the promoter, and (2) a driver that directs the expression of the PhiC31 serine integrase (Fig, 2A). In the integrase switch used here, PhiC31 accumulates in cells when *GATA23* is expressed, and, past a certain accumulation threshold, the target is switched so that cells express mScarlet instead of mTurq (Fig. 2B). Crucially, unlike a transcriptional reporter, the integrase switch is a binary response: only those cells that surpass a threshold level of *GATA23* expression will switch (*GATA23*-switched cells, GSCs) (Fig. 2A), and the change in the target and associated reporter expression is permanent and durable in these cells. Because of the binary nature and imposed expression threshold of the switch, we predicted that the number of GSCs would not be affected by auxinole to the same degree as GECs, because the level of *GATA23* expression in ineffectively repressed adjacent cells would likely be below the threshold at which the integrase switch occurs.

Repeating the cell counting process in our *GATA23* switch line, evaluating the number of GSCs, we found a comparable distribution to that found in the transcriptional reporter. The distributions of GECs (Fig. 1C) and GSCs (Fig. 2B) had the same median of 5 cells and the same minimum (2) and maximum (8). Seedling specific plotting of the number of GSCs (Fig. 2C) also reveals

similar patterns to that of GECs, with every seedling across both experiments having a median number of cells between 4 and 6 (Fig. 2C). However, unlike for the GECs, application of auxinole did not result in a significant increase in median number of GSCs. We computed the average GEC or GSC number for all counted LRs on each screened seedling in each condition and, with auxinole addition, we found a significant upwards shift in average GEC number and no significant difference in GSCs (Fig. 2D).

The variability in number of GSCs was modestly increased with auxinole treatment, but not to the same level as was found in GECs (an overall 6% increase in CV compared to 21% for GECs) (Fig. 2E). On a per seedling basis, each auxinole-treated seedling had a median number of GSCs between 4 and 6 (Fig. 2C), just like the control. As we predicted, the effect of auxinole on GSC numbers was weaker compared to GECs. Additionally, as would be expected if auxinole treatment impairs repression of the auxin response in neighboring cells, auxinole treatment led to a higher frequency of "double LRs", where two LRs were formed in very close proximity to one another (Fig. 2F). Loss of function of the peptide RALFL34 is known to interfere with the repression of LR neighboring cells, with the *ralfl34* mutant showing increased divisions of LR-neighboring XPP cells (indicating founder cell identity) and an increased incidence of double LRs[35]. Both phenotypes are similar to our results with auxinole treatment (Fig. 2D, Fig. 2F), further supporting a model where auxin signaling levels in central GECs are optimized for lateral inhibition of auxin responsiveness in neighboring cells.

Another difference we noted with the integrase recorder was a higher relative frequency of low GSC numbers. This could be attributed to the same dampened neighbor cell repression if this phenomenon evenly affected LRs regardless of GSC number, but this is not what we observed. Another possibility is that some proportion of these low GSC LRs are LR terminations. Because

many means of LR arrest occur in early stages, such as through inhibition of LR founder cell division by cytokinin[36], we posited that the switch line should have a higher proportion of early stage LRs when compared with the transcriptional reporter, which is indeed what we observed. While the overall LR density was similar between the switch and reporter lines (Fig. 2G), Stage 1 density was higher in the switch line (Fig. 2H). Additionally, auxinole treatment increased the relative proportion of Stage 1 LRs in the recorder seedlings (Fig. S2A), but not in the switch seedlings (Fig. S2B). This suggests that auxinole treatment increases the likelihood of an LR to terminate early. In LR development, cytokinin signaling works antagonistically to auxin signaling and it is the balance between the two that guides root development. Mutually antagonistic hormone signaling dynamics are common in plants and are known to affect variability in plant traits, such as in *Arabidopsis* seed germination, where the relative signaling strength of the hormones abscisic acid and gibberellic acid sets the level of variability in germination timing[37]. Similarly, our results suggest that shifting the balance between auxin and cytokinin signaling (in this case by reducing auxin signal strength) also affects variability in cell responses.

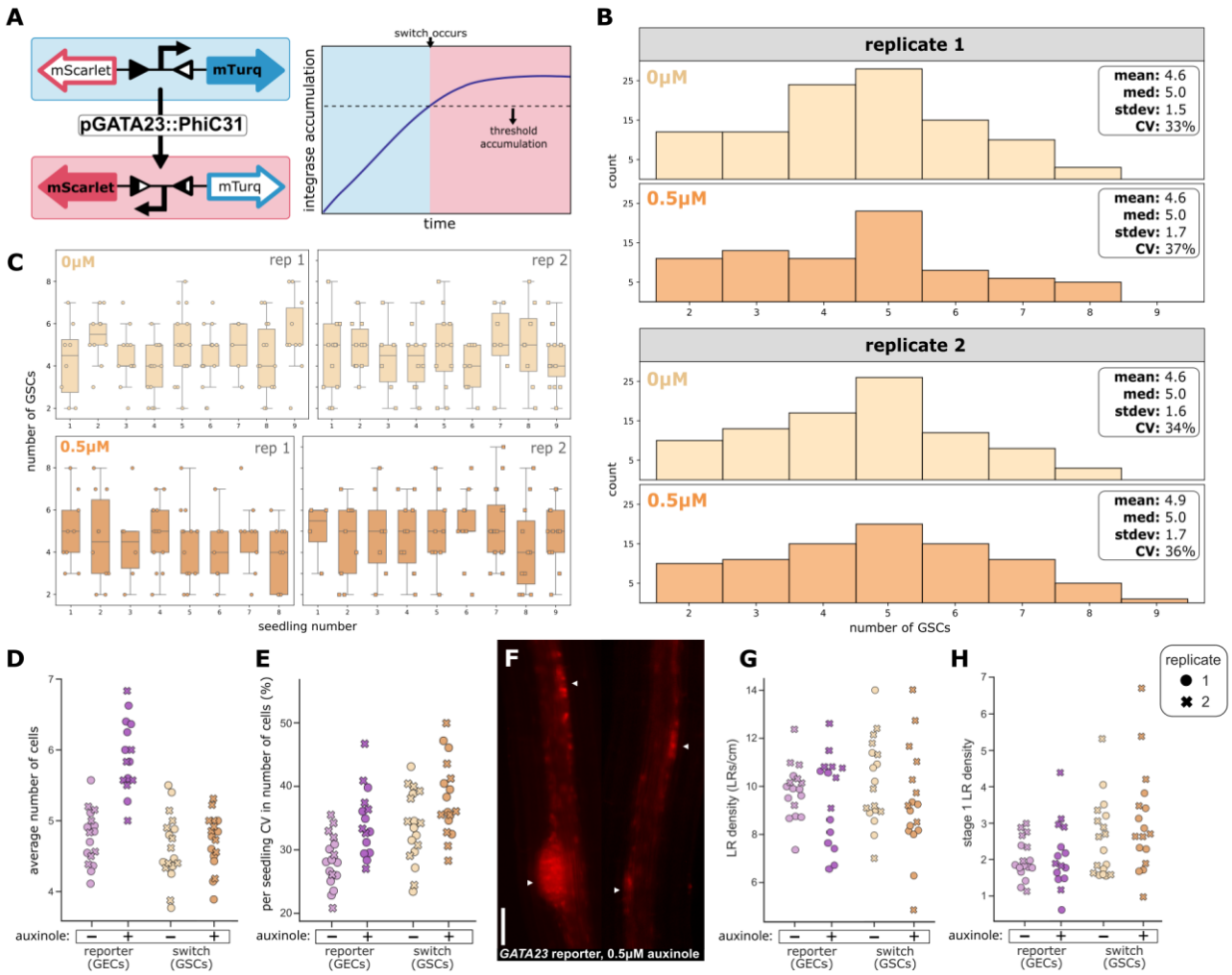


Figure 2: *GATA23* integrase switch shows that dampening the auxin signal during LR initiation results in incomplete repression of the auxin response in neighboring cells and an increased rate of early LR arrests. **A.** *GATA23* integrase recorder schematic. (left) The recorder starting state is expression of mTurquoise. With sufficient accumulation, the PhiC31 integrase mediates DNA recombination, inverting the promoter direction and switching expression to mScarlet. The switch is permanent, thus mScarlet expression is sustained indefinitely. **B.** Overall distribution in GSC number (as measured by counting the number of mScarlet-expressing cells in the *GATA23* switch line) in the control (light orange) and 0.5μM auxinole treatment (dark orange) for two cell counting replicates. Summary stats are included for each distribution. **C.** Per seedling distribution in GSC number for control (top) and auxinole (bottom) treatments across two replicates. **D.** Per seedling average number of cells for the *GATA23* reporter (GECs, pink) and recorder (GSCs, orange) in the control and auxinole treatment. **E.** Per seedling CV in number of cells for the *GATA23* reporter (GECs, pink) and recorder (GSCs, orange) in the control and auxinole treatment. **F.** Example images of "double LRs" seen more frequently in auxinole-treated seedlings. Scale bar is 100μm. **G, H.** Total (**G**) and stage 1 only (**H**) LR densities for the *GATA23* reporter and recorder, with 0 and 0.5μM auxinole added.

As an independent test of the impact of lowered auxin sensitivity on variation, we analyzed a previously characterized line expressing an IAA14 mutant with a lower binding affinity for auxin (*IAAslow*)[30]. While the overall LR density in *IAAslow* seedlings was not significantly changed compared to WT, the variation was greatly increased (Fig. 3A). The CV in number of GECs in *IAAslow* was 42%, compared to 10% in WT. With auxinole treatment, the WT CV in LR density was 15%, which increased to 47% in *IAAslow*. As reported previously, the distribution of LR stages was quite different in *IAAslow* with the majority of LRs in Stage 1 (Fig. 3B, Fig. S2C) and a dramatic reduction in emerged LRs (Fig. 3C, Fig. S2C). We predicted we would find similar increases in median and variability in number of GECs in *IAAslow* as we did with auxinole treatment (Fig. 1C). Instead, we found that the median number and range of GECs in *IAAslow* was the same as in WT (Fig. 3D,E), albeit with a slight increase in the average from 4.8 to 5.3 GECs. The range of GEC values increased in *IAAslow*, with a maximum number of GECs of 11 compared to 8 in WT.

The *IAAslow* mutant and auxinole should act independently to reduce auxin sensitivity, so in combination might reveal the upper limits of variation. When *IAAslow* seedlings were exposed to auxinole (*IAAslow*+auxinole), the median for the number of GECs in *IAAslow*+auxinole was 6, an increase from the no auxinole control but relatively unchanged compared to WT+auxinole (Fig. 3D). Each individual seedling had a median number of GECs between 4 and 7 (Fig. 3E). The mean showed the same pattern, where the *IAAslow*+auxinole mean of 5.8 GECs was comparable to the WT+auxinole mean of 6.0 GECs. The distribution in per seedling average GEC numbers in *IAAslow*+auxinole was not significantly changed compared to *IAAslow* (Fig. 3F). This contrasts with the significant upward shift seen in WT+auxinole compared to WT alone (Fig. 3F). Looking at the variation in the number of GECs during LR initiation, we found

that *IAAslow*+auxinole was the most variable, with a CV of 47%. This compares to a CV of 40% for the no auxinole *IAAslow* control and 34% for WT+auxinole. Plotting CV in GEC number on a per seedling basis for each line and in each auxinole treatment (Fig. 3G), we found that auxinole addition consistently increased the average CV, shifting the distribution upwards in both WT and *IAAslow*. So, while the combinatorial effect on variability in GEC number by *IAAslow* and auxinole was synergistic, the effect on GEC number itself was not. This pattern suggests that there is a functional limit to the number of LR initial cells (a number already close to saturation in both mutant and auxinole treatment), but perhaps not in the potential for variation.

In addition to altered number of GECs, we noticed the majority of the nuclei in GECs in the *IAAslow* background were unusually shaped, often resembling the lens-like shape of nuclei in non-LR XPP cells (Fig. 3H). These XPP-like cells made up 58% of the screened LRs and another 10% were comprised of a mix of XPP-like and more typical LR-like cells (Fig. 3I). Interestingly, the appearance of the cells in the LR had little bearing on the distribution of number of GECs. It could be that these XPP-like GECs represent a transition state within the LR initiation process. Evidence has shown that changes to a cell's signaling environment can alter the stability and durability of the transition state[17]. Recently, the capacity of precursor cells to undergo mixed cell fate transitions and the importance of auxin in controlling these transitions has been shown in stomatal development[38]. In *IAAslow*, auxin signal strength is much reduced, and it would make sense that a transition state could arise from a reduction in auxin signal strength in LR precursor XPP cells and/or a reduction in repression of the auxin signal strength in XPP cells neighboring LR precursors. In either case, a cell's trajectory could be diverted towards a relatively stable transition state. If this were the case, we would expect that further

dampening of the auxin signal with auxinole would increase the likelihood of cells entering this transition state. Consistent with this prediction, auxinole treatment increased the number of both XPP-like GECs to 64% (Fig. 3I).

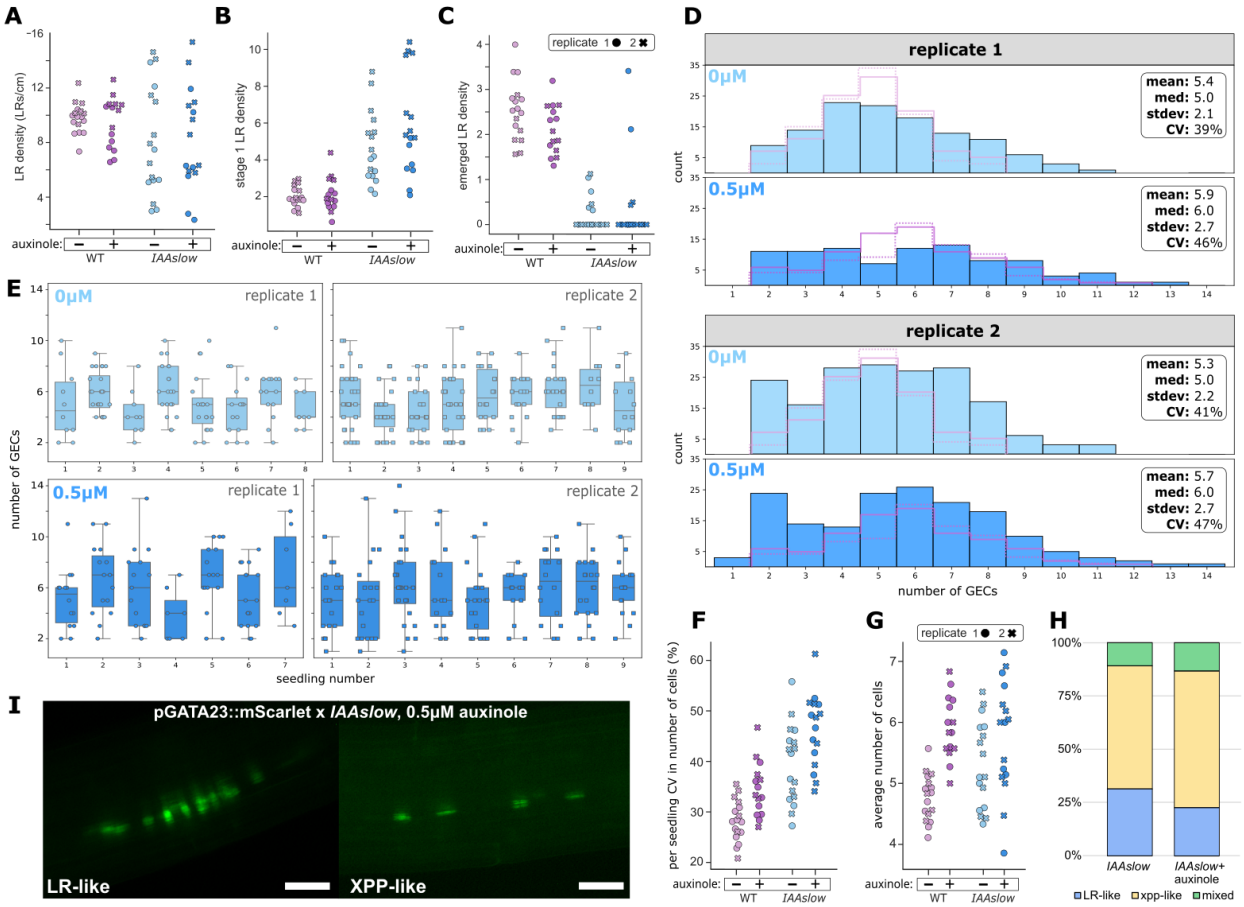


Figure 3: Further dampening of auxin signaling with auxinole and *IAAslow* shows differing effects on GEC number and variation therein, and potentially muddied transition state dynamics. **A-C.** Total (**A**), stage 1 only (**B**), and emerged only (**C**) LR densities in WT and *IAAslow* backgrounds, with 0 and 0.5μM auxinole added, over two replicates. **D.** Overall distribution in GEC number in *IAAslow* in the control (light blue) and 0.5μM auxinole treatment (dark blue) for two cell counting replicates. GECs for the *IAAslow* background were determined by counting GECs in a pGATA23::mScarlet x *IAAslow* cross line. Summary stats are included for each distribution. For reference, the 0 and 0.5μM auxinole distributions in WT (Fig. 1D) are represented as a dotted (replicate 1) and solid (replicate 2) outlines over the *IAAslow* 0 and 0.5μM auxinole distributions, respectively. **E.** Per seedling distribution in GEC number in *IAAslow* for control (top) and 0.5μM auxinole (bottom) treatments across two replicates. **F, G.** Per seedling CV in (**F**) and average (**G**) GEC number in WT and *IAAslow* and with 0 and 0.5μM auxinole treatment. **H.** Proportion of cell appearances in *IAAslow* GECs. **I.** Example images of LR-like (left) and XPP-like (right) GECs. Scale bar is 100μm.

If auxin is indeed a controller of variation during root initiation, we reasoned that similar trends should be found in regenerated roots (RRs) (Fig. 4A), which involve expression of many of the same developmental genes[39]. In root regeneration from leaf explants, the wound site is flooded with high levels of auxin which are required for RR initiation[40]. Root regeneration is well suited to analyzing phenotype variation because, compared to LR development, root regeneration is not robust. Leaf explants of the same age, and even from the same seedling, show natural variation in the number of RRs and the timing of their development[41]. In pre-emerged stages, RRs are less consistent in size and shape, while LRs have a more consistent morphology [23]. This initial variation gets buffered, however, as mature RRs are relatively consistent in morphology[42].

We began by testing the effect of different concentrations of auxinole on various metrics of root regeneration, such as timing of RR emergence and number of RRs. After removing and culturing the leaf explants, we screened the number of RRs from each explant every day for 13 days. The auxinole treatment delayed the onset of RRs, with the control explants starting to show emerged RRs as early as day 4, while the seedlings treated with 5μM auxinole began to show emerged RRs on day 6 (Fig. 4B). Regardless of the treatment, once the first RRs emerged, there was a steady increase in the average number of RRs over the 13 days of screening, the slope of which decreased with higher auxinole concentration. To complement the auxinole treatments, we also analyzed the *IAAslow* mutant and a *tir1* double mutant which is hypersensitive to auxin[43]. In the *IAAslow* background, the earliest emerged RRs were delayed similarly to the auxinole treated seedlings (Fig. 4C). In contrast, the *tir1* mutant was no delayed and showed a large jump in the average number of RRs per explant between day 5 and day 7 before leveling off to reach a more modest incline from day 8 to day 13 (Fig. 4C). Total number of roots was also correlated to auxin status, with a lower average number of RRs on day 13 for both auxinole treatments and

IAAslow mutants when compared to control, and a higher average number of roots in *tir1* mutants (Fig. 4D). The trend did not reflect viability differences between the conditions, as the trend was similar when analyzing all surviving explants (green) or only explants that regenerated at least 1 RR (yellow).

We next plotted the distributions of number of RRs on day 13 for each explant in auxinole treatment (Fig. 4E). In the 0μM treatment, the distribution was the most spread out, with explants regenerating anywhere from 0 to 4 roots at comparable frequencies. With 1 or 2μM, surviving explants either regenerated 0 roots or they regenerated 2-3 roots. With 5μM auxinole, over half the explants did not survive to 13 days, and the ones that did regenerated either 0 or 1 roots. Repeating this for the *IAAslow* and *tir1* mutants (Fig. 4F), we found the *IAAslow* result to be similar to that of the 5μM auxinole treatment, where explants either regenerated 0 or 1 roots, but a higher proportion of *IAAslow* explants survived the 13 days compared to the 5μM auxinole. In the *tir1* mutant, somewhat counterintuitively, we found the distribution to be similar to that seen for the 1 and 2μM auxinole treatments. A higher proportion of *tir1* explants survived to 13 days compared to the 1 and 2μM auxinole treatments, but among the surviving explants, around half regenerated zero roots in all three cases, while the other half regenerated multiple roots (2-3 RRs for 1 and 2μM auxinole, 2-4 for *tir1*).

For every case tested, there were a number of explants that, by day 13, had not regenerated any roots but remained alive, hereafter deemed non-regenerating (NR) explants. Because the NR explants were noticeably lighter in color than the regenerating explants and had a hard, brittle texture (Fig. 4G) matching descriptions of plant callus[44], we posited that these explants may be on a callus forming trajectory. Callus formation is another means of regeneration where a high auxin signal induces the formation of a pluripotent mass of cells from which regeneration of

organs can occur if cultured on appropriate media after callus induction[42]. In lab settings, callus is typically induced by culturing plant tissue in a high auxin media, but callus formation can also be triggered by high auxin after wounding[45]. To more thoroughly assess the regeneration outcomes for the NR explants, we cultured all the explants for an additional week (to day 20). We posited that NR explants committed to a callus trajectory would remain alive without regenerating additional roots by day 20, while NR explants not undergoing callus formation would die off in that time. Because callus formation is associated with very high auxin signaling, we predicted that the *tir1* explants, which should have the highest auxin signal, would be most likely to form callus. Indeed, we found that *tir1* explants had the highest proportion of NR explants on day 20 (Fig. 4H). Additionally, the proportion did not change between day 13 and day 20, meaning every NR explant on day 13 ended up in the callus forming trajectory. The 0μM WT, representing the second highest auxin signaling condition tested, also did not show a reduction in the proportion of NR explants from day 13 to day 20. In contrast, for every other condition (Fig. 4H), there was a reduction in the proportion of NR explants.

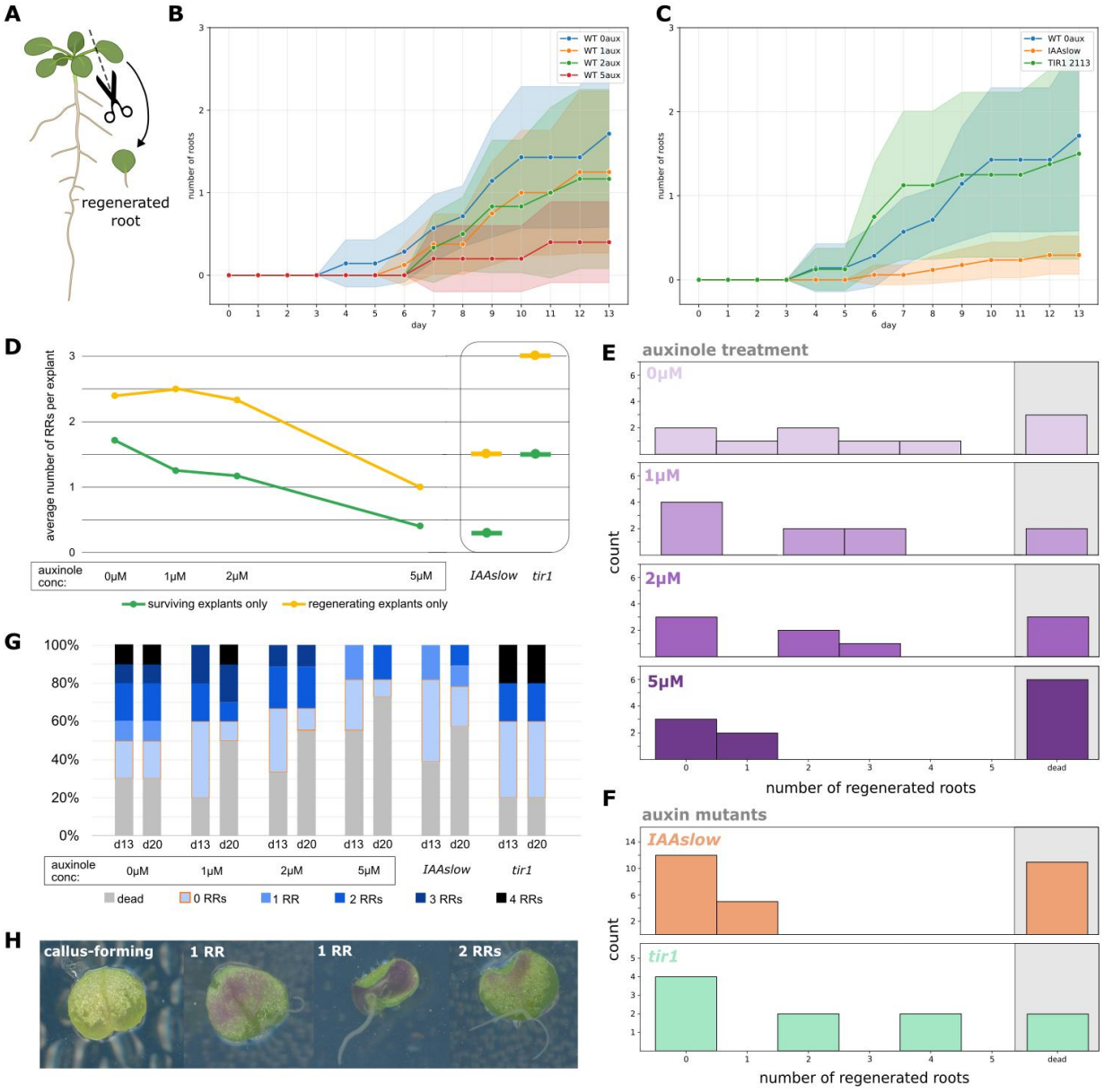


Figure 4: Auxin signaling conditions impact frequency and robustness of root regeneration outcomes. **A.** Schematic of root regeneration from *Arabidopsis* leaf explants. The leaf is removed from the seedling and cultured, where it may regenerate a new root/s near the wounding site. **B, C.** Regeneration timelines for (**B**) auxinole-treated and (**C**) auxin mutant explants. Day 0 is the day of explant excision. Shaded areas represent standard error. **D.** Average number of RRs on day 13 for each auxinole treatment (left, line plot) and auxin mutant line (right). Green includes all explants surviving to day 13 and yellow only included explants which regenerated at least one root by day 13. **E, F.** Overall distribution in number of RRs for the auxinole treatments (**E**) and the auxin mutants (**F**) taken on day 13. **G.** Changes in regeneration outcomes from day 13 (d13) to day 20 (d20) for each auxinole treatment and auxin mutant. **H.** Example images of a callus-forming explant (left), 1 RR explant (middle), and 2 RRs explant (right).

CONCLUSION

Auxin serves to coordinate cell responses both by rapidly turning on LR initiation genes in founder cells, and by repressing the expression of these genes in neighboring XPP cells. Proper balance between these processes is needed to establish defined boundaries between differentiating and non-differentiating cells, both physically and in terms of cell identities. In the WT auxin conditions, robust establishment of these boundaries enables rapid and coordinated progression from an undifferentiated cell state, through the transition state, eventually reaching a differentiated cell state (Fig. 5, top). Dampening the auxin signal reduces both effects, resulting in a muddying of the cell identities and, therefore, more cells expressing the founder cell marker *GATA23* and increased variability in expression patterns. This dampening, in the case of the *IAAslow* mutant, was associated with increased frequency of *GATA23*-expressing cells which appeared more like XPP cells than LR cells. We posited that this effect was due to stabilized transition state dynamics imbued by the perturbation to auxin signaling and that these cells represent a transition state in LR development. Overall, in reduced auxin signaling conditions, cells were less likely to differentiate and more likely to take on a transition state identity, potentially due to stabilization of the undifferentiated state and the transition state and destabilization of the differentiated cell state (Fig. 5, middle). Further reduction in the auxin signal all but eliminates establishment of boundaries between differentiating and non-differentiating cells, making differentiation much less likely and much more variable if it does occur (Fig. 5, bottom).

Stochastic models of cell differentiation have supported the relationship between signaling conditions and transition state stability and shown that a cell's differentiation trajectory depends on these dynamics[15–17]. Cell fate decisions in plant cells are quite plastic, with cells along the

same lineage undergoing mixed cell fate transitions in certain circumstances[39]. Perhaps a consequence of this plasticity, plant cells also have remarkable regenerative capacity, with differentiated cells under the right conditions becoming totipotent[46]. Auxin promotes cell plasticity, promoting mixed cell fate decisions and regenerative capacity. Work in *Arabidopsis* sepals has shown that cell-to-cell variability in auxin response is needed to initiate development, yet too much initial variability manifests as loss of phenotype robustness, as the variability is too high to be attenuated through typical strategies.

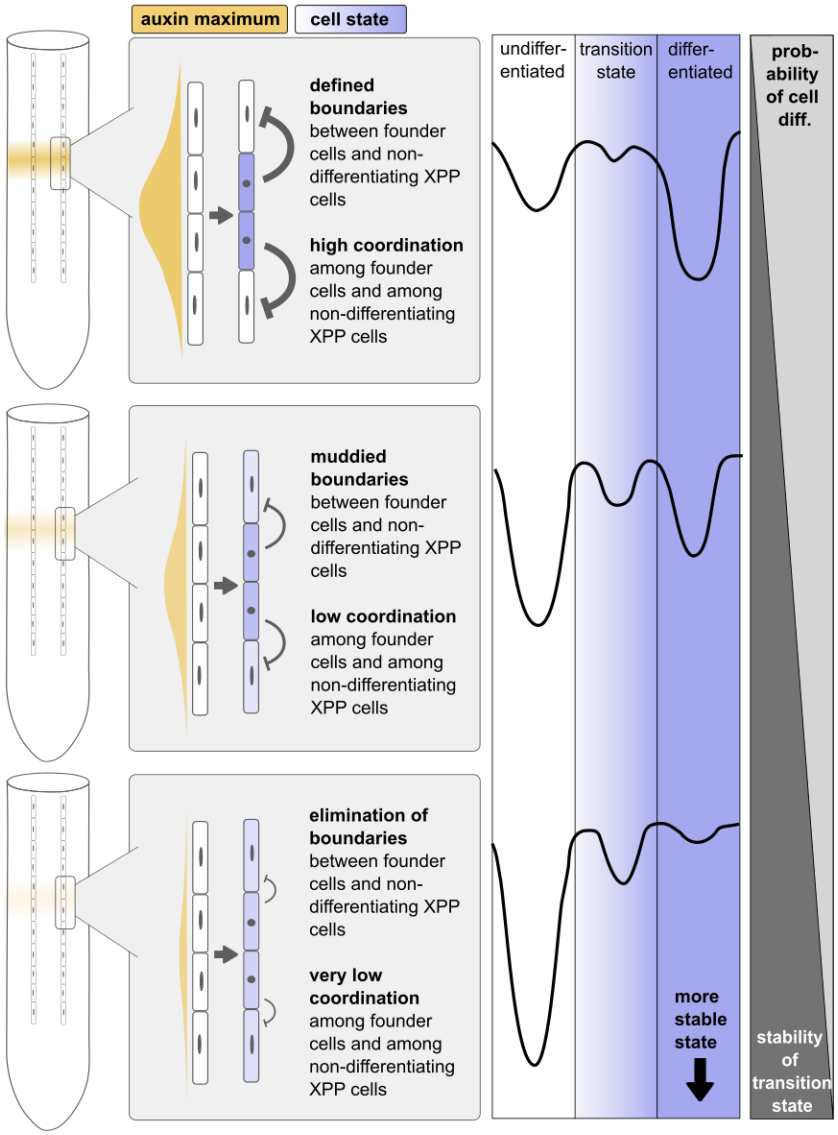


Figure 5: Model of cell differentiation dynamics in different auxin signaling conditions during LR initiation. On the right, graphical representation of the establishment of boundaries between differentiating (purple) and non-differentiating (white) cells in different auxin (yellow) signaling conditions. On the left, representation of energy dynamics underlying undifferentiated, transition, and differentiated states during LR initiation, with deeper valleys representing more stable cell states. Cells are more coordinated when moving through unstable transition states. In a typical auxin signaling context (top) the physical boundaries between XPP cells and LR founder cells is well established through a balance between auxin signal response in founder cells and the repression of this response in neighboring XPP cells. This results in rapid and coordinated progression of founder cells through the transition state to become differentiated. Reduction of the auxin signal (middle) reduces both the strength of the response in founder cells and the repression strength in neighboring XPP cells, resulting in muddied boundaries between differentiating and non-differentiating cells. This reduces the likelihood of differentiation which could be due to stabilization of the undifferentiated and transition cell states and destabilization of the differentiated state. Extreme reduction in the auxin signal (bottom) exacerbates these effects further, resulting in elimination of boundaries and making differentiation even less likely.

In plants, there is an established connection between hormone signaling dynamics, cell-to-cell variation, cell differentiation, and developmental outcomes, yet engineering efforts typically focus on individual signaling components to achieve desired phenotypes. Additionally, crops are often intentionally engineered with traits that are optimized for a given growth condition[47], resulting in low variation in these traits which may not be ideal for growth in increasingly unpredictable environmental conditions due to climate change[48]. Approaches targeting cell-to-cell variation could serve as an alternative approach for engineering developmental traits. Variation could be modulated through changes to the growth environment that alter hormone signaling dynamics and thus may also affect cell-to-cell variation. Alternatively, genetic modifications to components of hormone signaling pathways could be screened for their effect on cell-to-cell variation, with the goal of identifying lines with increased variation which could confer developmental flexibility or resilience to environmental stressors. This engineering philosophy aligns with Ashby's Law of Requisite Variety[49], a foundational principle of cybernetics which states "Only variety destroys variety". That is, as the complexity of the environment grows, so too must the complexity of the biological response, and engineering cell-to-cell variation could help to achieve this.

METHODS

Plant growth conditions

Arabidopsis seeds were sown in 0.5 x Linsmaier and Skoog nutrient medium (LS) (Caisson Laboratories) and 0.8% w/v Phyto agar (PlantMedia/bioWORLD), stratified at 4 °C for 2 days, and grown in constant light at 22 °C.

Cell counting experiments

Seeds from our cell counting Arabidopsis lines (GATA23 reporter, GATA23 integrase switch34, and GATA23 reporter x IAAslow30) were sown on 0.5X LS Phyto agar plates as described above, one plate with no auxinole added and one plate with 0.5µM auxinole (MedChemExpress, solid+solvent). After 8 days of growth in constant light at 22 °C, the plates were scanned using a flatbed 767 scanner (Epson America, Long Beach, CA) to assess root length and emerged LR density for each seedling. Seedlings were mounted on slides using diH2O and imaged with a Leica Biosystems DMI 3000 fluorescent microscope (using the RFP channel for the reporter, CFP and RFP channels for the switch, and the GFP channel for the IAAslow reporter cross line). To enable LR width measurements, for the first GATA23 reporter replicate, an image of every LR on each seedling was taken. For all other replicates, all LRs on each seedling were staged as per Banda et al. 2019, but images were only taken of early stage LRs (stage I or II). All LR images were taken at 20X magnification, with the exception of some LRs with enough GECs/GSCs that reduction to 10X was needed to fit them all in view. For a given cell counting line, imaging of both the mock and the auxinole-treated seedlings was done on the same day. Two independent replicates were performed on separate days for each line.

Regeneration assay

Arabidopsis seeds were sown as described above and grown for 12 days in constant light. The first true leaves from 12 day old seedlings were removed as explants with a doubled edged razor blade (Personna). Explants were then cultured on 0.5X LS plates with 0.8% w/v Phyto agar with concentrations of 0, 1, 2, and 5μM auxinole. The plates were scanned daily for 13 days of explant culturing using a flatbed 767 scanner and the number of regenerated roots of each explant was also recorded daily. Explants were cultured for an additional week (to day 20) at which point survival, number of regenerated roots, and any callus formation was evaluated and recorded. Callus formation was indicated by a surviving explant without any regenerated roots, with these explants appearing lighter in color and with a brittle texture in comparison to explants with regenerated roots.

LR/cell counting imaging analysis

All microscope image analysis was performed using the Fiji ImageJ program (version 1.53c). For measuring LR widths the straight line tool was used to span the width of the LR and the measure function was used to measure the distance. For early stage LRs without protrusion, width was assessed by the extent of bright GATA23 reporter expression. For later stage LRs, width was measured from the points of protrusion of the LR from the main root. To assess the number of GECs/GSCs, cells in early stage LRs were counted as GECs if they were sufficiently brighter in reporter expression than any background level expression in the area, and counted as GSCs if they had any mScarlet expression. For stage I LRs, GEC/GSC number was simply the number of expressing cells in the LR. For stage II LRs, having undergone the first periclinal division of LR development, GEC/GSC number was assessed by backtracking to the expected number in stage I, using the well understood patterns of cell divisions in early LR development.

For the IAAslow cross line, the shape of the nuclei in each imaged LR was also recorded along with GEC number. LRs containing cells with round nuclei were labeled “LR-like”, while LRs containing cells with oblong nuclei were labeled “XPP-like”. LRs containing cells with both appearances were labeled “mixed”. To assess LR density, seedling plate scans were opened in Fiji ImageJ and root lengths were measured using the segmented line tool.

All plots were generated using Python scripts with plotting functions and was run in version 3.9.1 and with the following package dependencies: pandas (version 1.5.3), scipy.stats (version 1.10.0), matplotlib.pyplot (version 3.6.3), matplotlib.colors (version 3.6.3), and numpy (version 1.24.2).

REFERENCES

1. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
2. Urban, E. A. & Johnston, R. J., Jr. Buffering and amplifying transcriptional noise during cell fate specification. *Front. Genet.* **9**, 591 (2018).
3. Fraser, L. C. R., Dikdan, R. J., Dey, S., Singh, A. & Tyagi, S. Reduction in gene expression noise by targeted increase in accessibility at gene loci. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2018640118 (2021).
4. Hong, L. *et al.* Variable Cell Growth Yields Reproducible OrganDevelopment through Spatiotemporal Averaging. *Dev. Cell* **38**, 15–32 (2016).
5. Munro, V., Kelly, V., Messner, C. B. & Kustatscher, G. Cellular control of protein levels: A systems biology perspective. *Proteomics* **24**, e2200220 (2024).
6. Wu, H.-W. *et al.* Noise reduction by upstream open reading frames. *Nat. Plants* **8**, 474–480 (2022).
7. Zhang, Z., Qian, W. & Zhang, J. Positive selection for elevated gene expression noise in yeast. *Mol. Syst. Biol.* **5**, 299 (2009).
8. Chalancon, G. *et al.* Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* **28**, 221–232 (2012).
9. Wernet, M. F. *et al.* Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174–180 (2006).
10. Holmes, W. R. *et al.* Gene Expression Noise Enhances Robust Organization of the Early Mammalian Blastocyst. *PLoS Comput. Biol.* **13**, e1005320 (2017).
11. Larsen, H. L. *et al.* Stochastic priming and spatial cues orchestrate heterogeneous clonal contribution to mouse pancreas organogenesis. *Nat. Commun.* **8**, 605 (2017).
12. Snijder, B. & Pelkmans, L. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* **12**, 119–125 (2011).
13. Osorio, D. *et al.* Single-cell expression variability implies cell function. *Cells* **9**, 14 (2019).
14. Gatlin, V., Gupta, S., Romero, S., Chapkin, R. S. & Cai, J. J. Exploring cell-to-cell variability and functional insights through differentially variable gene analysis. *NPJ Syst. Biol. Appl.* **11**, 29 (2025).
15. MacLean, A. L., Hong, T. & Nie, Q. Exploring intermediate cell states through the lens of single cells. *Curr. Opin. Syst. Biol.* **9**, 32–41 (2018).

16. Moris, N., Pina, C. & Arias, A. M. Transition states and cell fate decisions in epigenetic landscapes. *Nat. Rev. Genet.* **17**, 693–703 (2016).

17. Brackston, R. D., Lakatos, E. & Stumpf, M. P. H. Transition state characteristics during cell differentiation. *PLoS Comput. Biol.* **14**, e1006405 (2018).

18. Waddington, C. H. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. (Allen & Unwin, London, 1957).

19. Huang, S., Guo, Y.-P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713 (2007).

20. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547 (2008).

21. Kong, S. *et al.* Tradeoff between speed and robustness in primordium initiation mediated by auxin-CUC1 interaction. *Nat. Commun.* **15**, 5911 (2024).

22. Trinh, D.-C. *et al.* Increased gene expression variability hinders the formation of regional mechanical conflicts leading to reduced organ shape robustness. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2302441120 (2023).

23. Banda, J. *et al.* Lateral Root Formation in Arabidopsis: A Well-Ordered LRexit. *Trends Plant Sci.* **24**, 826–839 (2019).

24. Torres-Martínez, H. H., Hernández-Herrera, P., Corkidi, G. & Dubrovsky, J. G. From one cell to many: Morphogenetic field of lateral root founder cells in Arabidopsis thaliana is built by gradual recruitment. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 20943–20949 (2020).

25. Van Norman, J. M., Xuan, W., Beeckman, T. & Benfey, P. N. To branch or not to branch: the role of pre-patterning in lateral root formation. *Development* **140**, 4301–4310 (2013).

26. von Wangenheim, D. *et al.* Rules and Self-Organizing Properties of Post-embryonic Plant Organ Cell Division Patterns. *Curr. Biol.* **26**, 439–449 (2016).

27. Pierre-Jerome, E., Jang, S. S., Havens, K. A., Nemhauser, J. L. & Klavins, E. Recapitulation of the forward nuclear auxin response pathway in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9407–9412 (2014).

28. Guseman, J. M. *et al.* Auxin-induced degradation dynamics set the pace for lateral root development. *Development* **142**, 905–909 (2015).

29. Yadav, S. R., Bishopp, A. & Helariutta, Y. Plant development: early events in lateral root initiation. *Curr. Biol.* **20**, R843–5 (2010).

30. Vanstraelen, M. & Benková, E. Hormonal interactions in the regulation of plant development. *Annu. Rev. Cell Dev. Biol.* **28**, 463–487 (2012).

31. Dubrovsky, J. G. *et al.* Auxin acts as a local morphogenetic trigger to specify lateral root founder cells. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8790–8794 (2008).

32. Hayashi, K.-I. *et al.* Rational Design of an Auxin Antagonist of the SCFTIR1 Auxin Receptor Complex. *ACS Chem. Biol.* **7**, 590–598 (2012).

33. Cavallari, N., Artner, C. & Benkova, E. Auxin-regulated lateral root organogenesis. *Cold Spring Harb. Perspect. Biol.* **13**, a039941 (2021).

34. Guiziou, S., Maranas, C. J., Chu, J. C. & Nemhauser, J. L. An integrase toolbox to record gene-expression during plant development. *Nat. Commun.* **14**, 1844 (2023).

35. Murphy, E. et al. RALFL34 regulates formative cell divisions in Arabidopsis pericycle during lateral root initiation. J. Exp. Bot. 67, 4863–4875 (2016).

36. Li, X., Mo, X., Shou, H. & Wu, P. Cytokinin-mediated cell cycling arrest of pericycle founder cells in lateral root initiation of Arabidopsis. *Plant Cell Physiol.* **47**, 1112–1123 (2006).

37. Abley, K. *et al.* An ABA-GA bistable switch can account for natural variation in the variability of Arabidopsis seed germination time. *Elife* **10**, (2021).

38. Tang, L. P. *et al.* Time-resolved reprogramming of single somatic cells into totipotent states during plant regeneration. *Cell* (2025) doi:10.1016/j.cell.2025.08.031.

39. Serrano-Ron, L. *et al.* Reconstruction of lateral root formation through single-cell RNA sequencing reveals order of tissue initiation. *Mol. Plant* **14**, 1362–1378 (2021).

40. Kareem, A. *et al.* Water availability positions auxin response maxima to determine plant regeneration fates. *Nat. Plants* **11**, 1367–1379 (2025).

41. Sugimoto, K., Jiao, Y. & Meyerowitz, E. M. Arabidopsis regeneration from multiple tissues occurs via a root development pathway. *Dev. Cell* **18**, 463–471 (2010).

42. Xu, L. De novo root regeneration from leaf explants: wounding, auxin, and cell fate transition. *Curr. Opin. Plant Biol.* **41**, 39–45 (2018).

43. Yu, H. *et al.* Mutations in the TIR1 auxin receptor that increase affinity for auxin/indole-3-acetic acid proteins result in auxin hypersensitivity. *Plant Physiol.* **162**, 295–303 (2013).

44. Ikeuchi, M., Sugimoto, K. & Iwase, A. Plant callus: mechanisms of induction and repression. *Plant Cell* **25**, 3159–3173 (2013).

45. Ikeuchi, M. *et al.* Wounding triggers callus formation via dynamic hormonal and transcriptional changes. *Plant Physiol.* **175**, 1158–1174 (2017).

46. Xu, M., Du, Q., Tian, C., Wang, Y. & Jiao, Y. Stochastic gene expression drives mesophyll protoplast regeneration. *Sci. Adv.* **7**, eabg8466 (2021).

47. Kong, X., Zhang, M., De Smet, I. & Ding, Z. Designer crops: optimal root system architecture for nutrient acquisition. Trends Biotechnol. 32, 597–598 (2014).

48. Ikeuchi, M. et al. Wounding triggers callus formation via dynamic hormonal and transcriptional changes. Plant Physiol. 175, 1158–1174 (2017).

49. Ross Ashby, W. Requisite variety and its implications for the control of complex systems. Cybernetica 1, 83–99 (1958).

Chapter 5 Supplementary Figures

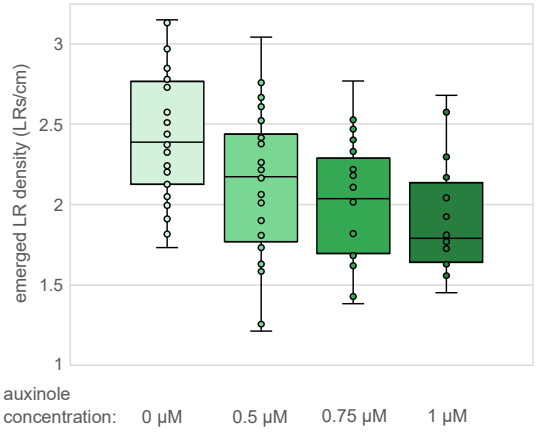


Figure S1: Dose response for emerged LR density in different auxinole concentrations. Seedlings were screened for the number of emerged LRs per cm of main root in 0, 0.5, 0.75, and 1 µM auxinole concentrations. Each point represents the LR density of one seedling in the given auxinole condition.

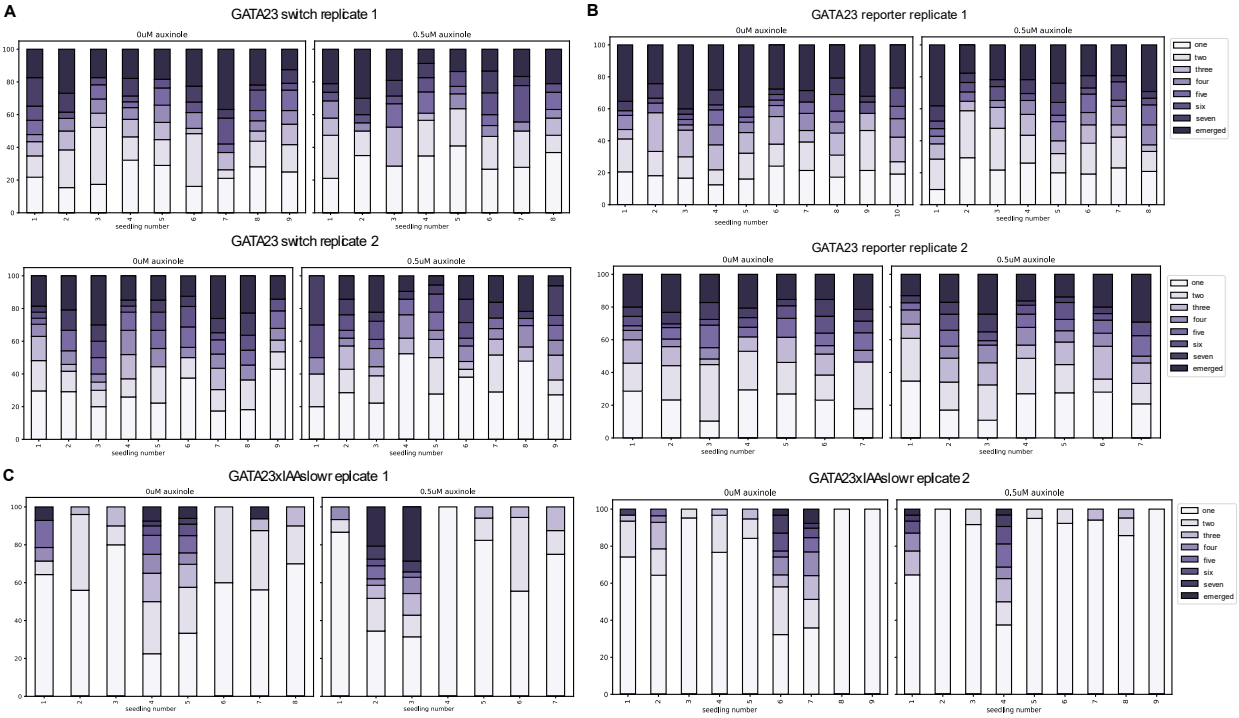


Figure S2: LR staging distributions for *GATA23* switch, *GATA23* reporter, and *GATA23xIAAslow*, auxinole treatment and control. LR staging distributions for each screened *GATA23* switch (**A**), *GATA23* reporter (**B**), and *GATA23xIAAslow* (**C**) seedling across 2 replicates and treated with 0 and 0.5 µM auxinole. Later LR stages are indicated with darker shades of purple as indicated in the legend.

Reflection

The field of synthetic biology has been hugely impactful to our ability to engineer organisms for desired traits and functionality. Synthetic biology tools including CRISPR and recombinase technology have enabled incredibly fine control over cellular processes, revolutionizing fields like agricultural and metabolic engineering. However, a less discussed but equally promising application of synthetic biology is in developing tools that improve our understanding of biological processes. The improved control over cell functions enabled by synthetic biology offers an avenue to address specific biological questions in ways that were not possible previously. One example is in lineage tracing[1], wherein researchers can track the division patterns and parentage of cells during developmental processes. Both CRISPR[2] and recombinase[3] technology has been applied here. In both approaches, DNA barcodes serve as markers of cell lineage, and subsequent single cell sequencing can reveal every cell's history, elucidating dynamics of developmental processes such as neural development[2,3] and embryonic development[4] and allowing monophyletic lineage reconstruction[5]. My thesis contributes to this area, generating integrase-based recorders of gene expression dynamics[6,7] and the integrase eraser for study of essential gene function[8]. Many engineering breakthroughs have come because of insights gained through fundamental biological research, such as CRISPR, which was discovered through the study of bacterial immune systems[9]. An approach where synthetic biology is increasingly integrated into fundamental biology research should expedite both the development of new technologies and the discovery of new biological phenomena, some of which would inspire new technologies and engineering applications. As such, the union of fundamental biological research and synthetic biology/engineering would drive both fields forward.

Integrase technology has a lot of promise both in engineering and in driving forward fundamental biological research. It currently has been applied in the recording of gene expression dynamics[6,7], lineage tracing[3], engineering developmental processes[10], biosensors[11], more. Integrase recorders to track gene expression events can be quite powerful, but one limitation is that each recorder must be optimized individually for each gene or combination of genes recorded. A concerted focus on quantifying the efficiency of different integrases, the effect of tuning parts on that efficiency, and the strength of expression of the recorded gene/s, should allow a more informed design process and less need for extensive optimization. In plants, control over the integrase recorder construct insertion location should help greatly reduce variability in output[12]. One general limitation in integrase-based applications is in the limited number of orthogonally functioning integrases. While screening the dozens of integrases native to different viruses has yielded a good number with orthogonal function[13], they often have differences in efficiency which are not well characterized and can introduce unintended variation. Orthogonality depends on each integrase only recognizing its own specific set of binding sites, so engineering efforts to design similarly functioning integrases which each specifically recognize their own distinct binding sequences would go a long way towards maximizing the potential of integrase technology.

One application where integrase technology has proven beneficial is in assessing cell-to-cell variation in gene expression dynamics[7]. More and more recent evidence points to the conclusion that many biological processes depend on cell-to-cell variation [14]. However, assessing it is difficult because cell-to-cell variation can manifest in many ways, such as in expression of different genes, different timing in expression of the same genes, different strength of gene expression, and all combinations therein. Currently, experimental approaches to assess cell-to-

cell variation with spatial and temporal resolution are quite limited and most of the time involve single cell, spatially resolved RNA sequencing with some trajectory analysis to gain temporal insights[15]. This is not a technique that is accessible to most labs, as it is expensive and requires specialized equipment. Additionally, in some contexts, cell-to-cell variation is short-lived and only exists among a small number of cells, making detection all but impossible using current RNA sequencing methods. As such, there is a strong need for accessible experimental methods for revealing and exploring cell-to-cell variation. This could come by expanding the integrase recorder abilities or through new methods altogether.

Because of the difficulty in experimentally assessing cell-to-cell variation, an attractive approach is to combine experimental data with stochastic computational modeling of cell states. Such models have been successfully employed in the study of cell differentiation [16] and have the potential to elucidate how cells take on new identities in response to signals. Stochastic models could be used to generate hypotheses regarding cell-to-cell variation and its effect on cell decision making that can then be tested experimentally. The power of computational modeling in biology has grown exponentially in recent years with major advances in computational power, machine learning, and accessibility of high quality experimental data to inform or train models. The union of experimental and computational approaches to address biological questions has grown in popularity in recent years and has had quite a lot of success[17,18]. Such approaches should be very well suited to addressing questions about cell decision making and cell-to-cell variation. This also should lay the groundwork for studies relating cell-to-cell variation to variability in phenotypes, a connection that is not yet well understood.

In plants, cell-to-cell variation can be related to hormone signaling dynamics which guide many plant processes. Further study into the nature of this connection is warranted, as plant hormone

signaling networks are very complex and interconnected. In fact, in many cases, plant processes are not controlled just by one hormone, but the antagonistic interaction between two hormone signaling networks (abscisic acid/gibberellin, auxin/cytokinin as a couple examples)[19]. The balance between these antagonistic hormone signaling networks is what defines the cell state and the cell trajectory. Previous work[20,21], along with this work in this thesis have provided potential mechanisms for this hormone-mediated regulation of variation in specific developmental contexts, but a unifying mechanism tying hormone signaling dynamics to variation is still lacking. This could be tremendously impactful, as both hormone signaling and cell-to-cell variation relate to cell plasticity or stemness, which could have major implications in engineering regenerative capacity for agricultural or medical applications.

Traditionally, cell-to-cell variation was thought of as an obstacle to robust functioning, and some evidence supports this, with the regulation of certain genes including mechanisms to buffer expression variation. However, the regulation of other genes actually promotes and amplifies variation and, as these are commonly genes involved in development and response to stressors, variable expression is hypothesized to drive initiation of developmental processes and adaptive capacity to environmental conditions. As a result, it stands to reason that targeting variation in engineering efforts could enhance resilience and adaptability. With climate change causing increasing unpredictability in weather patterns[22], crops with traits optimized for a specific condition (i.e. drought or high temperatures) may not perform well when those conditions change. In this case, what may instead be needed are crops with enhanced resilience, which could be accomplished through modulating cell-to-cell variation.

References

1. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. Nat. Rev. Genet. 21, 410–427 (2020).

2. Zafar, H., Lin, C. & Bar-Joseph, Z. Single-cell lineage tracing by integrating CRISPR-Cas9 mutations with transcriptomic data. Nat. Commun. 11, 3055 (2020).

3. Chow, K.-H. K. et al. Imaging cell lineage with a synthetic digital recording system. Science 372, (2021).

4. Bowling, S. et al. An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. Cell 181, 1410–1422.e27 (2020).

5. Choi, J. et al. A time-resolved, multi-symbol molecular recorder via sequential genome editing. Nature 608, 98–107 (2022).

6. Guiziou, S., Maranas, C. J., Chu, J. C. & Nemhauser, J. L. An integrase toolbox to record gene-expression during plant development. Nat. Commun. 14, 1844 (2023).

7. Maranas, C. J. et al. A history-dependent integrase recorder of plant gene expression with single-cell resolution. Nat. Commun. 15, 9362 (2024).

8. Watson, I. J., Maranas, C., Nemhauser, J. L. & Leydon, A. R. A hot-swappable genetic switch: Building an inducible and trackable functional assay for the essential gene MEDIATOR 21. ACS Synth. Biol. 14, 2170–2180 (2025).

9. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012).

10. Williams, R. L. & Murray, R. M. Integrase-mediated differentiation circuits improve evolutionary stability of burdensome and toxic functions in E. coli. Nat. Commun. 13, 6822 (2022).

11. Essington, E. A. et al. An autonomous microbial sensor enables long-term detection of TNT explosive in natural soil. Nat. Commun. 15, 10471 (2024).

12. Lowry, G. V. et al. Towards realizing nano-enabled precision delivery in plants. Nat. Nanotechnol. 19, 1255–1269 (2024).

13. Gomide, M. S. et al. Genetic switches designed for eukaryotic cells and controlled by serine integrases. Commun Biol 3, 255 (2020).

14. Osorio, D. et al. Single-cell expression variability implies cell function. Cells 9, 14 (2019).

15. Cannoodt, R., Saelens, W. & Saeys, Y. Computational methods for trajectory inference from single-cell transcriptomics. Eur. J. Immunol. 46, 2496–2506 (2016).

16. Brackston, R. D., Lakatos, E. & Stumpf, M. P. H. Transition state characteristics during cell differentiation. PLoS Comput. Biol. 14, e1006405 (2018).

17. von Wangenheim, D. et al. Rules and Self-Organizing Properties of Post-embryonic Plant Organ Cell Division Patterns. Curr. Biol. 26, 439–449 (2016).

18. Abley, K. et al. An ABA-GA bistable switch can account for natural variation in the variability of Arabidopsis seed germination time. Elife 10, (2021).

19. Wang, Y. H. & Irving, H. R. Developing a model of plant hormone interactions. Plant Signal. Behav. 6, 494–500 (2011).

20. Kong, S., Rusnak, B., Zhu, M. & Roeder, A. H. K. Stochastic gene expression in auxin signaling in the floral meristem of Arabidopsis thaliana. Nat. Commun. 16, 4682 (2025).

21. Kong, S. et al. Tradeoff between speed and robustness in primordium initiation mediated by auxin-CUC1 interaction. Nat. Commun. 15, 5911 (2024).

22. Bolan, S. et al. Impacts of climate change on the fate of contaminants through extreme weather events. Sci. Total Environ. 909, 168388 (2024).

G3, 2023, 13(11), jkad206
https://doi.org/10.1093/g3journal/jkad206
Advance Access Publication Date: 11 September 2023
Plant Genetics and Genomics

A comparative analysis of stably expressed genes across diverse angiosperms exposes flexibility in underlying promoter architecture

Eric J.Y. Yang, Cassandra J. Maranas, Jennifer L. Nemhauser*

Department of Biology, University of Washington, Seattle, WA 98105-1800, USA

*Corresponding author: Jennifer L. Nemhauser, Department of Biology, University of Washington, 3747 West Stevens Way NE, Seattle, WA 98105-1800, USA.
Email: jn7@uw.edu

Promoters regulate both the amplitude and pattern of gene expression—key factors needed for optimization of many synthetic biology applications. Previous work in *Arabidopsis* found that promoters that contain a TATA-box element tend to be expressed only under specific conditions or in particular tissues, while promoters that lack any known promoter elements, thus designated as Coreless, tend to be expressed more uniformly. To test whether this trend represents a conserved promoter design rule, we identified stably expressed genes across multiple angiosperm species using publicly available RNA-seq data. Comparisons between core promoter architectures and gene expression stability revealed differences in core promoter usage in monocots and eudicots. Furthermore, when tracing the evolution of a given promoter across species, we found that core promoter type was not a strong predictor of expression pattern. Our analysis suggests that core promoter types are correlative rather than causative in promoter expression patterns and highlights the challenges in finding or building constitutive promoters that will work across diverse plant species.

Keywords: core promoter elements; constitutive promoters; TATA-box; Y patch; coreless; Plant Genetics and Genomics

Introduction

Precise control over gene expression is essential for development and survival. One of the first regulatory steps in expression regulation is transcription initiation, which is controlled by DNA regions designated as promoters. Current understanding of eukaryotic promoters is still remarkably limited, and we have difficulty even identifying a precise promoter region given an arbitrary sequence (Donczew and Hahn 2017). A core promoter region is functionally defined as the minimal region required for transcription initiation, associated with binding of RNA polymerase II (RNAPII) and general transcription factors (GTFs). Proximal and distal cis-regulatory elements contribute to the modulation of the core promoter's activity and give it its characteristic expression profile. A sequence containing the proximal cis-regulatory elements as well as the core promoters is often referred to as the "promoter" region (Biłas *et al.* 2016; Haberle and Stark 2018; Andersson and Sandelin 2020; Schmitz *et al.* 2022). In practice, cloning and analysis projects often pick an arbitrary length (e.g. up to 2,000 base pairs or until the next coding sequence) upstream of the transcription start site (TSS) to define as the promoter region (Andersson and Sandelin 2020; Schmitz *et al.* 2022).

Many core promoter elements have been identified within the core promoter region, which are important in directing RNAPII and determining the TSS. The TATA-box motif is the most well-understood of the core promoter elements, yet TATA-box-containing promoters only account for about 20% of eukaryotic promoters and about 30% of *Arabidopsis* promoters (Molina and Grotewold 2005; Donczew and Hahn 2017). In plants, additional core promoter types were proposed by Yamamoto et al. (2007, 2009) based on their identification of overrepresented motifs around a fixed distance from the TSS. Y patch, or pyrimidine patch, motifs are C and T rich motifs whose presence had been recently shown experimentally to associate with stronger expression (Jores *et al.* 2021). CA and GA are additional core promoter elements, represented in approximately 20 and 1% of genic promoters, respectively (Yamamoto *et al.* 2009). Unlike the TATA-box, which has a known GTF-binding protein associated with it, the molecular mechanisms of the Y patch, CA, and GA elements remain largely unknown. Core promoters that do not contain any of the identified core promoter types have been termed Coreless (Yamamoto *et al.* 2009, 2011). In *Arabidopsis*, Coreless promoters tend to be expressed more weakly but more broadly than those that contain TATA-boxes (Yamamoto *et al.* 2011; Das and Bansal 2019).

Constitutive promoters, defined here as promoters that are expressing in all tissues at all times, are versatile tools in synthetic biology due to their desirable expression pattern (Yang and Nemhauser 2022; Zhou *et al.* 2023). They are often used to drive expression of components used in synthetic circuits or metabolic engineering (Wu *et al.* 2014; South *et al.* 2019; Patron 2020; Brophy *et al.* 2022). Core promoter regions of constitutive promoters (such as the Cauliflower Mosaic Virus 35S promoter) have often been used as the starting point to build synthetic promoters by

Received on 25 August 2023; accepted on 01 September 2023

introducing natural cis-elements or synthetic TF-binding sites upstream of these core promoter regions to artificially tune expression strength or confer new expression patterns (Brückner et al. 2015; Ali and Kim 2019; Belcher et al. 2020; Cai et al. 2020; Brophy et al. 2022; Moreno-Giménez et al. 2022). However, a lack of understanding of the design constraints around promoters had made engineering synthetic promoters challenging. Current approaches often require trial and error or high throughput screening to identify functional synthetic promoters (Brückner et al. 2015; Belcher et al. 2020; Cai et al. 2020; Brophy et al. 2022; Moreno-Giménez et al. 2022). A better understanding of the contributions and limitations of core promoters in controlling expression patterns can therefore be essential in engineering better synthetic promoters.

Here, by leveraging publicly available RNA-seq atlases of 15 angiosperms, we were able to map gene expression pattern onto core promoter type in multiple genomic contexts. While TATA-box-containing promoters are overrepresented in conditionally expressed genes in all of the species we examined, the pattern for Coreless promoters was less clear. In most eudicots, Coreless promoters were overrepresented in uniformly expressed genes, but the opposite trend was observed in monocots. Additionally, by identifying orthologous gene groups within these species, we were able to track changes in core promoter type and expression pattern for groups of evolutionarily related promoters. We found that stably expressed genes are also more likely to have orthologs in other species compared to unstably expressed genes, and the orthologs tend to retain similar expression patterns. Last, we show that changes in core promoter types do not explain changes in expression pattern. This evolution-guided approach reveals design rules surrounding core promoter architecture and expression patterns.

Methods

Phylogenetic tree

A phylogenetic tree was constructed referencing NCBI's Taxonomy Browser and Li et al. 2021.

RNA-seq dataset processing

(Relevant files: 0_Slurm_Pipeline)

RNA-seq atlases were located in the NCBI Sequence Read Archive database. The references for the datasets can be found in Supplementary Table 1. The individual datasets were retrieved using sratoolkit-3.0.1 prefetch followed by fasterq-dump functions. Fastqc-0.11.9 were used to generate a QC report for each dataset. Trimmomatic-0.39 were used for adaptor and low-quality ends trimming using the following settings: "SLIDINGWINDOW:4:20 MINLEN:36." ILLUMINACLIP files TruSEq3-PE-2.fa were supplied for paired end data and TruSEq3-SE.fa were supplied for single end data. Reference transcriptomes were downloaded from Ensembl Plants (http://plants.ensembl.org/index.html) for *Arabidopsis thaliana, Camelina sativa, Cucumis melo, Glycine max, Phaseolus vulgaris, Pisum sativum, Vigna unguiculata, Sorghum bicolor, Zea mays, Solanum lycopersicum, Actinidia chinensis, Triticum aestivum* and Phytozome (https://phytozome-next.jgi.doe.gov) for *Arachis hypogaea, Cicer arietinum, and S. tuberosum* (Goodstein et al. 2012; Cunningham et al. 2021). An index file was generated, and the reads aligned and counted using Kallisto-0.44.0 with "-o counts -b 500." For single end data, fragment length and standard deviation were required, but the information is difficult to locate, and so a default value of "-l 200 -s 20" was used across the board.

Another Fastqc was performed on the trimmed files, and a final MultiQC-1.13 was run on the entire folder encompassing all the log files that Fastqc, Trimmomatic, and Kallisto generated. The MultiQC report was inspected to ensure the trimming step improved read quality and there were no major warnings.

Normalizing count, calculating CV, and percent ranking

(Relevant files: 1_Metadata_from_RUNselector.Rmd, 2_MOR_Normalization.Rmd)

Using an R script, the raw counts for each species were normalized using the DESeq2 package using a metadata file curated from the original study for the RNA-seq datasets. The coefficient of variation (CV) across all samples for a given atlas was used as a metric for stability for each gene, and the percentile ranking for each gene was calculated. The geometric mean for each gene was also calculated across all samples.

Extracting intergenic region and 5′UTR

(Relevant files: 3_ExtractPromUTR(ALL_Transcripts).ipynb, 8_Extract PromUTR(Orthologs).ipynb)

Gff3 annotation files and reference genomes were downloaded from Ensembl or Phytozome depending on where the reference transcriptomes were retrieved from. Forty percent of transcripts were selected from the total transcriptome, and their intergenic region and 5′UTR were extracted from the Gff3 annotation. Intergenic region and 5′UTRs of identified orthologs were extracted in a similar manner.

Labeling core promoter types

(Relevant files: 4_Label_Promoters.Rmd, 9_Motif_Scan.Rmd, 10_Octamer_Scan.ipynb)

Motif scan: Intergenic regions and 5′UTR sequences are trimmed to those regions to be scanned for each core promoter types: TATA box (−100 to TSS), Y patch (−100 to +100), and Inr (−10 to +10). Intergenic regions shorter than 100 bp were excluded from analysis. Each region was scanned for their respective motifs using motif files as well as methods outlined in Jores et al. (2021). A motif is considered to be present when the relative motif scores are above 0.85.

Octamer scan: Intergenic regions and 5′UTR sequences are trimmed based on the positions relative to the TSS outlined in Yamamoto et al. (2009) (TATA, −45 to −18; Y Patch, −50 to +50; CA, −35 to −1; GA, −35 to +75). Each region was scanned for the presence of octamer motifs from the TATA, Y patch, GA, and CA lists outlined in Yamamoto et al. (2009). If the specified region contained at least one motif for a given promoter type, it was labeled as positive.

Ortholog analysis

(Relevant files: 5_At_gene_ranking.Rmd, 6_Identifying_orthologs.Rmd, 7_Processing_orthologs.Rmd)

The *Arabidopsis* transcriptome was filtered to only include primary transcripts, and mitochondria as well as chloroplast transcripts were removed. Top 5% stable genes by CV, bottom 5% stable genes by CV, and a random set of 1,343 genes (5%) were randomly selected.

Using biomaRt in R, the Ensembl and Phytozome databases were queried for orthologs for the selected set of *Arabdiopsis* genes for each species (Durinck et al. 2009). Orthologs from *A. hypogaea, C. arietinum*, and *S. tuberosum* were retrieved from Phytozome, and the rest of the species from Ensembl. For an analysis in Fig. 3b, significance tests were done by ANOVA followed by Tukey's HSD (honestly significant difference). For each target gene that matched to an *Arabidopsis* transcript, only the highest expressing

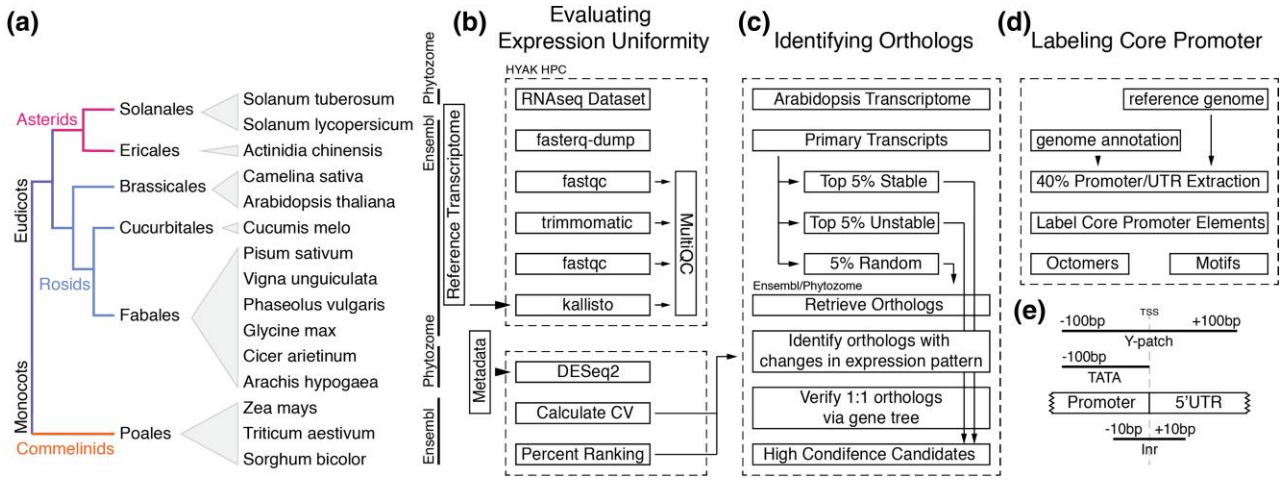


Fig. 1. An outline of the bioinformatics pipelines. a) The 15 angiosperms included in this study and their phylogenetic relationship. b–d) The 3 major data processing steps performed in the study. Detailed parameters are included in the *Methods* section. Reference genomes, transcriptomes, and gene orthologs were retrieved via either Ensembl (Cunningham et al. 2021) or Phytozome (Goodstein et al. 2012) databases depending on the species. e) Regions searched for each core promoter motif.

transcript was kept. If an *Arabidopsis* transcript retrieved more than one orthologs from a target species, these pairs of orthologs were removed from analysis. We only kept orthologous gene groups that had a "change" in expression pattern, defined as crossing the 50th percentile CV, in 2 target species, and the remaining candidates were manually mapped onto the phylogenetic tree to identify gene groups that had changes in expression patterns that are consistent with the tree. This means having changes in expression patterns that are mostly found in the same clade. Gene trees were built for these candidates using blast-align-tree (https://github.com/steinbrennerlab/blast-align-tree), and the candidate lists were further trimmed based on the gene trees to ensure a 1:1 relationship between all members in the gene group.

Results

We began this project by identifying species with RNA-seq Atlases, which we defined as datasets containing at least 10 different tissue samples and with samples that represented at least 2 distinct developmental stages. Although RNA-seq measures RNA levels and thus can be affected by posttranscriptional regulation, it is the best available proxy for transcriptional activity (Wang *et al.* 2009). Details regarding the dataset and their references can be found in Supplementary Table 1 (Libault *et al.* 2011; Kaeppler *et al.* 2012; Stelpflug *et al.* 2012; Potato Genome Sequencing Consortium *et al.*, 2013; Loraine *et al.* 2013a, 2013b; Moscow State University 2014; Kagale *et al.* 2014; Peanut Genome Consortium 2015; Sudheesh *et al.* 2015; Klepikova *et al.* 2016b; Vlasova *et al.* 2016; Kudapa *et al.* 2017; University of Tsukuba 2017; Yao *et al.* 2017; McCormick *et al.* 2017a, 2017b, 2017c; Penin *et al.* 2018; Ramírez-González *et al.* 2018; Brian *et al.* 2021). Figure 1a shows a phylogenetic tree of the 15 species that fit our criteria, which spans a range of angiosperms including multiple monocots and eudicots. The datasets were processed through a custom pipeline (Fig. 1b–d). In brief, Kallisto was used for RNA-seq quantification, and MultiQC was used to summarize all the outputs up till DESeq2 (Supplementary Data 1) (Bray *et al.* 2016; Ewels *et al.* 2016). For each species, normalized counts from each tissue were then converted to expression uniformity information using the CV as a metric. In this analysis, lower CV corresponds to more uniform expression, meaning comparable expression in all tissues. Higher CV, on the other hand, means less uniform and more tissue-specific expression. To facilitate comparison between species, we used percentile rank of CV as the primary metric, which represents the percentage of CVs that are less than or equal to a given value.

To determine whether the characteristic differences in expression patterns between different core promoter types seen in *Arabidopsis* hold across all the species in our dataset, we extracted the −100 bp to +100 bp region around the TSS as the "core promoter region" for 40% of all promoters in each species (Fig. 1d). TATA-box, Y patch, and Inr motifs were screened according to the methods detailed in Jores *et al.* (2021). The regions scanned for each motif are more relaxed than their known regions in *Arabidopsis*, as we applied the scan to multiple species and wanted to avoid falsely labeling promoters as Coreless. Illustration of the regions scanned for each core promoter type is illustrated in Fig. 1e.

The subset of promoters for each species was labeled as either TATA or Y patch. If a promoter did not contain either element, we labeled them as "Coreless." It is important to note that the definition of Coreless promoters introduced by Yamamoto et al. (2009) is somewhat more strict than the definition used here, as they also screened for the relatively rare CA and GA core promoter elements. We then plotted the distribution of CV for each species, broken down by core promoter types (Fig. 2). Similar results for Y patch, Inr, and a random set of promoters that serve as a control are in Supplementary Fig. 1.

Using microarray data, Yamamoto et al. (2011) had found that Coreless promoters are underrepresented in genes that respond to stimulus (i.e. more constitutively expressed). However, we did not see the same trend until we removed the lowest expressing transcripts from the analysis (transcripts with an average of less than 1 read). These extremely low read counts are likely to be unreliable, and an analysis of the weak-expressing genes that we removed revealed that they bias toward higher CV when compared to the rest of the genes in the dataset (Supplementary Fig. 2). This same minimum read number requirement was then applied to the rest of the species.

Overall, the expected trend of TATA-box-containing promoters being overrepresented in conditionally expressed genes is observed across all the species analyzed (Fig. 2a). In contrast, the

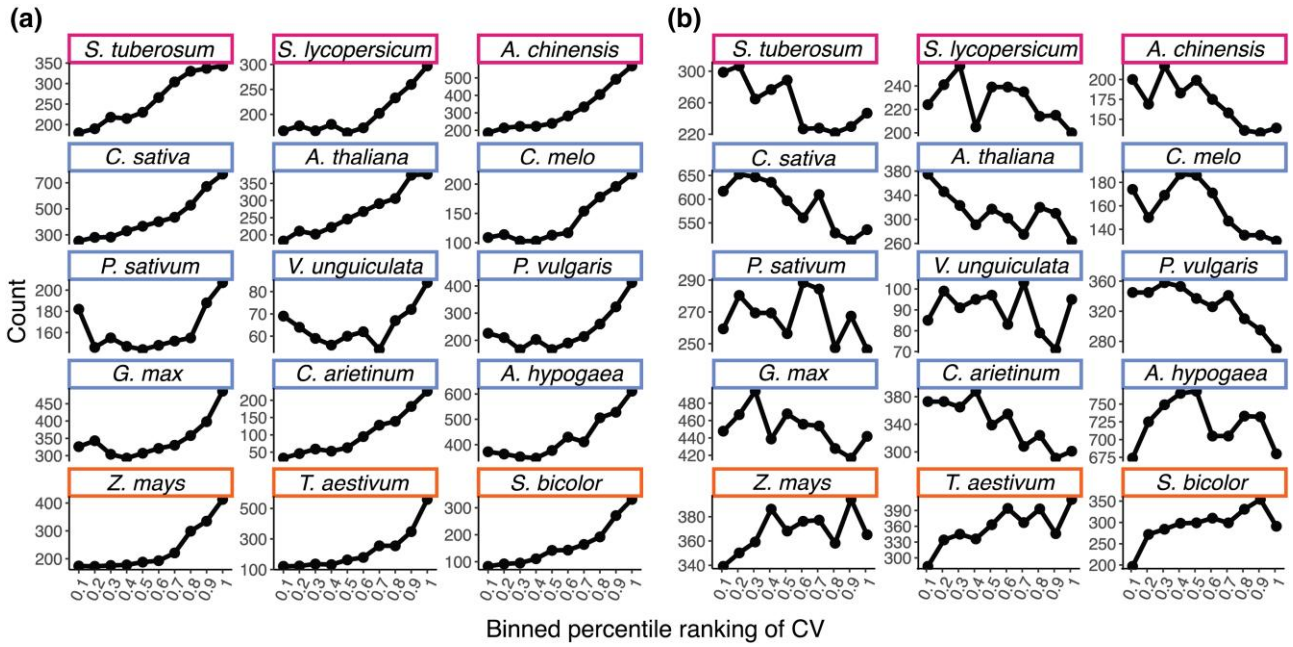


Fig. 2. Distribution of relative specificity or uniformity of TATA-box-containing and Coreless promoters. Higher CV rankings indicate more specificity, while lower CV rankings indicate more uniformity. A random subsampling of 40% of promoters from each species are shown here. a) TATA-box containing promoters, and b) promoters termed Coreless as they lacked both TATA-box and Y-path motifs. Colors correspond to phylogeny shown in Fig. 1a.

trend of Coreless promoters being associated with more uniformly expressed genes was weaker and only observed in a subset of the angiosperms. The monocots (*Z. mays*, *T. aestivum*, and *S. bicolor*) all exhibited a strong trend of Coreless promoters associating with conditionally expressed genes (e.g. those with higher CV values), along with an enrichment of Y patch-containing promoters being associated with uniform expression (Fig. 2b and Supplementary Fig. 1). This inverted pattern could be explained in 2 ways given that a promoter not labeled as containing a TATA-box or Y patch is labeled as Coreless. Under this classification scheme, an apparent enrichment by one category of promoters could reflect a surplus of that type of promoter in a particular CV ranking bin or a depletion of the other 2 promoter categories in that same bin. The latter explanation seems more likely for the Y patch promoters in monocots, but further experimental tests are required to fully resolve this question. The surprising pattern of Coreless genes "flipping" their behavior in monocots might also reflect an as-yet-undefined promoter element that is lumped into the Coreless category here. For example, there may be slight differences in TATA motif, as has been described for maize (Mejía-Guerra *et al.* 2015). Accounting for this known source of variation, we did not see any significant decrease in the Coreless trend toward conditionally expressed genes (Supplementary Fig. 1).

Our analysis was able to identify correlations between promoter type and expression pattern across many genes and led us to wonder whether the presence or absence of a specific core promoter type was sufficient to determine expression pattern. To test this hypothesis, we decided to focus on orthologous genes found across the species examined in this study (Fig. 1c). This approach allows us to test if changes in core promoter architecture during evolution led to changes in the uniformity of expression. We started by finding orthologs of *Arabidopsis* genes, as *Arabidopsis* has the most well-annotated genome and has 47,684 transcripts with a nonzero transcript count in at least one of the sampled tissues. Of this total, we retained only the primary transcripts of each nonmitochondrial and nonchloroplast genes, resulting in a final total of 26,842 genes. The top 5% most uniformly expressed and top 5% most conditionally expressed genes were selected based on CV, along with a randomly selected control set of equal size ($n = 1{,}343$ genes in each category). The sets of genes were used to query the Ensembl or Phytozome database for orthologs in the rest of the 14 species in our dataset (Goodstein *et al.* 2012; Cunningham *et al.* 2021). The orthologs were searched for in the database where their reference transcriptome was downloaded to ensure matching of the target transcript name with the transcript counts. Orthologs of *A. hypogaea*, *C. arietinum*, *and S. tuberosum* were found using Phytozome, and the remaining species were found in Ensembl.

Orthologous genes tended to retain their expression pattern across species (Fig. 3a). While orthologs corresponding to the random set of *Arabidopsis* genes were spread quite uniformly across distribution of CV rankings, the orthologs of the top 5% uniformly expressed set of *Arabidopsis* genes were skewed heavily toward the more uniform, lower percentage CV rankings. The orthologs of the 5% most conditionally expressed set of *Arabidopsis* genes showed a more subtle skew toward higher CV ranking. This trend was more visible in some species than others, partially due to the overall lower gene counts. One notable trend was that the most conditionally expressed gene set retrieved significantly fewer orthologs compared to the random or most uniformly expressed gene sets (Fig. 3b). This is possibly because uniformly expressed genes are associated with more fundamental cellular functions and therefore more likely to be conserved across species (Klepikova *et al.* 2016a). Following a similar logic, conditionally expressed genes tend to be more tissue-specific and therefore are more easily lost during species divergence.

Even when looking at genes that fell at the tail ends of the expression uniformity distribution from *Arabidopsis*, we could find orthologs positioned across the full range of CV rankings

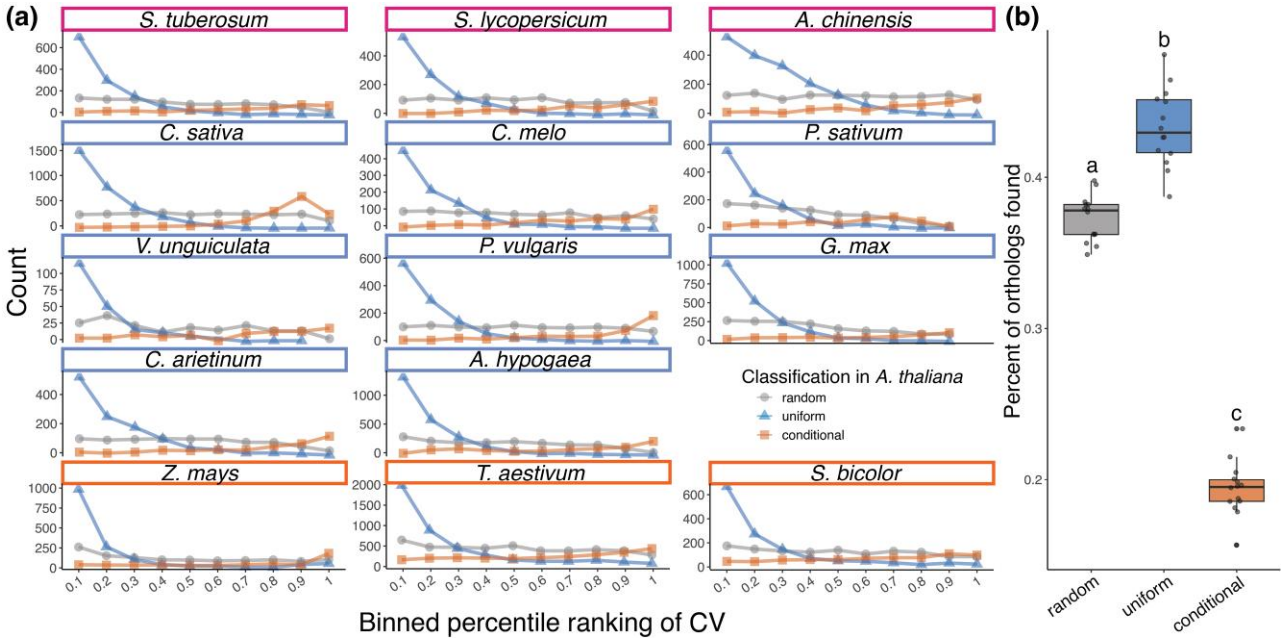


Fig. 3. Genes that show uniform expression in *A. thaliana* tend to behave similarly in other species. a) Distribution of CVs for orthologs of uniformly expressed (triangle), conditionally expressed (square), or random (circle) *A. thaliana* genes. The color of boxes around species names corresponds to Fig. 1a. b) Percent of orthologs found for each set of *A. thaliana* genes for each species. Each dot corresponds to a single species. Statistical tests were performed by 1-way ANOVA followed by Tukey HSD. All 3 groups are significantly different from one another.

(Fig. 3a). In other words, expression uniformity of a given gene can vary dramatically across species. These instances give us a unique opportunity to examine if the change in expression uniformity is predictive of changes in core promoter architecture. To investigate this further, we curated a set of evolutionarily related genes that showed this type of switching behavior. We limited our analysis to a set of 1:1 orthologs, where each species contributes a maximum of one gene for each orthologous gene set (Zahn-Zabal *et al.* 2020). This is to maximize confidence in evolutionary relatedness and minimize complications from gene duplications. Starting with the set of all the orthologs retrieved through Ensembl and Phytozome, we first filtered the target orthologs to count only the highest expressing transcript for each gene, thereby limiting each gene to a single representative transcript. We filtered the list of orthologs to include *Arabidopsis* transcripts that had only a single ortholog found in the transcriptome of each other species. We considered any target transcripts that crossed the 50th percentile in CV as "changing expression pattern," and we limited the *Arabidopsis* transcripts to those where transcripts changed expression pattern in at least 2 different species. These changes were mapped onto the phylogenetic tree to identify clusters where changes could be associated with a specific phylogenetic node.

For the most promising *Arabidopsis transcripts*, de novo gene trees were built by performing BLAST searches against the rest of the species. When more than one ortholog was found in any of the species, that species was removed from the set (Fig. 1c). With our stringent selection criteria, 7 high-confidence orthologous gene groups were found with 3 *Arabidopsis* transcripts (AT3G17020.1, AT3G18215.1, and AT4G40045.1) that are from the top 5% uniformly expressed genes list and 4 *Arabidopsis* transcripts (AT1G04700.1, AT5G17400.1, AT5G18910.1, and AT5G20410.1) from the top 5% conditionally expressed genes list. A summary of the filters and numbers of target orthologs as well as *Arabidopsis* query transcripts left after each step can be found in Supplementary Table 2.

The promoters for these 7 sets of orthologs were extracted, and TATA, Y patch, and Inr motifs were screened as described above (for clarity, this analysis will be referred to as motif scan) (Fig. 1d). In parallel, these promoters were also screened for TATA, Y patch, Inr, CA, and GA octamers as defined in Yamamoto *et al.* (2009) (octamer scan), and an illustration of the regions scanned for each octamers can be found in Supplementary Fig. 3. Comparing the 2 methods, the motif scan resulted in more identified core promoters due to its more relaxed parameters. Only 2 promoters were labeled as Y patch by the octamer scan but not the motif scan. A core promoter element was considered present if either method returned a positive result. The identification of the genes and a complete list of core promoter elements identified can be found in Supplementary Table 3. Within each orthologous gene group, changes in the presence of TATA or Y patch elements did not appear to correlate with changes in expression patterns (Fig. 4). In each group, there are examples of promoters having the same core promoter type but different expression patterns, as well as cases of promoters having the same expression pattern but different core promoter types. Since there were only 7 TATA-box-containing promoters (~15.5% of the promoters), we were not able to observe instances where 2 related TATA-box-containing promoters have different expression patterns, but there are multiple instances where changes in presence of TATA motif did not change expression pattern. This result suggests that the presence or absence of a TATA or Y patch is not sufficient to change expression pattern.

Discussion

Understanding the rules that govern the performance of natural promoters could inspire the construction of synthetic promoters that are able to retain their behavior over multiple generations in transgenic plants. Here, we mined RNA-seq atlases from 15

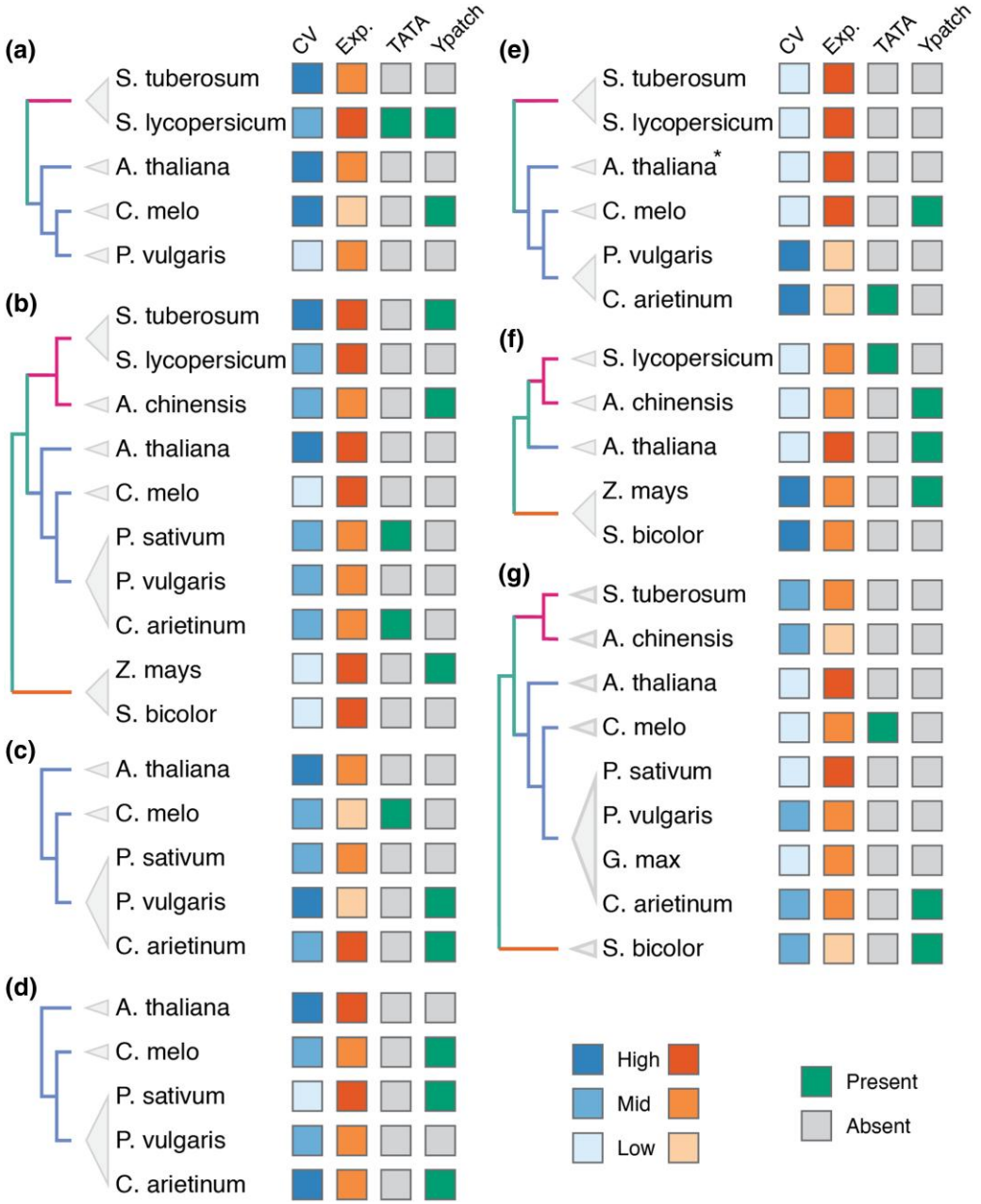


Fig. 4. Individual gene trees where expression uniformity changes can be observed. a–d) The gene is conditionally expressed in *A. thaliana* but uniformly expressed in another species. e–g) The gene is uniformly expressed in *A. thaliana* but conditionally expressed in another species. The *Arabidopsis* genes in each group are as follows: a) AT1G04700, b) AT5G17400, c) AT5G18910, d) AT5G20410, e) AT3G17020, f) AT3G18215, GAT4G40045. CV and expression strength (Exp.) are grouped by percentile ranking of 0.66–1.00 (high), 0.33–0.66 (mid), or 0.00–0.33 (low) and color coded accordingly. Presence or absence (gray) of TATA and Y patch motifs is indicated. *A. *thaliana* has no identifiable core promoter as the intergenic region is only 8 bp.

different angiosperms to extract patterns connected to the relative specificity or uniformity of gene expression across developmental stages and tissue types. We found that the previously observed trend that TATA-box-containing promoters are overrepresented in conditionally expressed genes is highly conserved. In contrast, the relative uniformity vs specificity of expression from Coreless promoters is not as well conserved. Coreless promoters from eudicots analyzed in this study were, in general, more highly associated with uniform expression patterns. Coreless promoters from monocot species, however, exhibited the opposite trend. In addition, we found that promoters tend to maintain their expression pattern across species, with the caveat that uniformly expressed genes are more likely to have identifiable orthologs when compared to conditionally expressed genes. Last, by tracking expression pattern and promoter type within the evolutionary trajectory of individual genes, we could test the hypothesis that promoter architecture is responsible for the level and pattern of gene expression. We found that none of the core promoter types screened for in this work is consistently associated with changes in expression pattern or strength. This suggests that while there may be a correlation between promoter architecture and transcription parameters, the underlying molecular mechanism that determines whether a gene is conditionally or specifically expressed remains unknown.

From a synthetic biology perspective, there are 2 major implications from the analysis described here. First, the hope of finding strong, constitutive natural promoters that work across diverse species may be even more challenging than we originally thought. For example, it is unlikely that there are natural promoter architectures that will work equally well as constitutive promoters in monocot and eudicot crops. Second, and more hopefully, our analysis suggests that the approach currently being taken by multiple labs for engineering synthetic promoters (Belcher *et al.* 2020; Brophy *et al.* 2022; Cai *et al.* 2020; Moreno-Giménez *et al.* 2022) is likely to find solutions that work well across species. The overall scheme of many of these groups is to take a core promoter region containing a TATA-box and then add natural cis-elements or synthetic transcription factor target sequences upstream of the core promoter to modulate expression strength or pattern. We found that the same core promoter could support widely varied expression patterns across evolution.

While the general trend that TATA-box-containing promoters are found in genes that are expressed in specific times and/or locations was highly conserved, close study of single gene phylogenies revealed that core promoters are not the determinant for expression pattern. The overall lack of pattern for TATA and Y patch motifs on the phylogenetic tree also suggests that the gain and loss of these promoter elements, at least in the genes studied here, are sporadic events that do not experience strong positive selection for maintenance. Our analysis leaves us with the question of why there is a discrepancy between the observed general preferences for core promoter types regarding expression uniformity but simultaneously a lack of contribution of core promoters to expression pattern, and whether there are mechanistic differences between Coreless and TATA promoters when they can achieve similar expression patterns. It is likely that constitutive expression can be achieved in at least 2 ways: by combining multiple tissue-specific elements that work together to achieve constitutive expression or by including "universal" elements that are broadly recognized across tissues. The analysis of the Cauliflower Mosaic Virus 35S promoter (p35S) showed that progressive deletion reduced promoter activity without affecting expression pattern and identified a short enhancer element that conferred constitutive expression (Odell *et al.* 1985; Fang *et al.* 1989; Hayashi *et al.* 1992). A more recent analysis of the p35S and other constitutive promoters, however, revealed multiple tissue-specific transcription factor binding sites (Cai *et al.* 2020). Performing a similar functional analysis on multiple Coreless promoters will be needed to determine whether the 2 classes of promoters achieve uniform expression using similar mechanisms. In addition, a more granular deletion analysis targeting individual cis-elements for both classes of promoters in multiple species, along with close examination of expression pattern, will be needed to fully map out promoter logic sufficiently to guide future engineering efforts.

Data availability

All scripts and datasets necessary to perform the analysis in the article, as well as supplemental data, are available at https://doi.org/10.5061/dryad.9w0vt4bmk.

Acknowledgments

We thank Dr Alexander Leydon and Janet Solano Sanchez for careful reading of the manuscript and Dr Adam Steinbrenner for advice on identifying orthologs. We also thank other members of the Di Stilio, Imaizumi, Steinbrenner, and Nemhauser lab for their feedback on this project.

Funding

This work was supported by the National Science Foundation (IOS-1546873), the National Institute of Health (R01-GM107084), and the Howard Hughes Medical Institute Faculty Scholar Award.

Conflicts of interest

The author(s) declare no conflict of interest.

Author contributions

Experimental design and analysis by EJYY, CJM, and JLN. Research performed by EJYY and CJM. Manuscript written by EJYY, CJM, and JLN.

Literature cited

Ali S, Kim W-C. 2019. A fruitful decade using synthetic promoters in the improvement of transgenic plants. Front Plant Sci. 10:1433. doi:10.3389/fpls.2019.01433.

Andersson R, Sandelin A. 2020. Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet. 21(2): 71–87. doi:10.1038/s41576-019-0173-8.

Belcher MS, Vuu KM, Zhou A, Mansoori N, Agosto Ramos A, Thompson MG, Scheller HV, Loqué D, Shih PM. 2020. Design of orthogonal regulatory systems for modulating gene expression in plants. Nat Chem Biol. 16(8):857–865. doi:10.1038/s41589-020-0547-4.

Biłas R, Szafran K, Hnatuszko-Konka K, Kononowicz AK. 2016. Cis-regulatory elements used to control gene expression in plants. Plant Cell, Tissue Organ Cult. 127(2):269–287. doi:10.1007/s11240-016-1057-7.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 34(5):525–527. doi:10.1038/nbt.3519.

Brian L, Warren B, McAtee P, Rodrigues J, Nieuwenhuizen N, Pasha A, David KM, Richardson A, Provart NJ, Allan AC, *et al.* 2021. Actinidia chinensis expression RNA sequencing. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA691387.

Brophy JAN, Magallon KJ, Duan L, Zhong V, Ramachandran P, Kniazev K, Dinneny JR. 2022. Synthetic genetic circuits as a means of reprogramming plant roots. Science. 377(6607): 747–751. doi:10.1126/science.abo4326.

Brückner K, Schäfer P, Weber E, Grützner R, Marillonnet S, Tissier A. 2015. A library of synthetic transcription activator-like effector-activated promoters for coordinated orthogonal gene expression in plants. Plant J. 82(4):707–716. doi:10.1111/tpj.12843.

Cai Y-M, Kallam K, Tidd H, Gendarini G, Salzman A, Patron NJ. 2020. Rational design of minimal synthetic promoters for plants. Nucleic Acids Res. 48(21):11845–11856. doi:10.1093/nar/gkaa682.

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, *et al.* 2021. Ensembl 2022. Nucleic Acids Res. 50(D1):D988–D995. doi:10.1093/nar/gkab1049.

Das S, Bansal M. 2019. Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. PLoS One. 14(3):e0212678. doi:10.1371/journal.pone.0212678.

Donczew R, Hahn S. 2017. Mechanistic differences in transcription initiation at TATA-less and TATA-containing promoters. Mol Cell Biol. 38(1):e00448-17. doi:10.1128/MCB.00448-17.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 4(8):1184–1189. doi:10.1038/nprot.2009.97.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32(19):3047–3048. doi:10.1093/bioinformatics/btw354.

Fang RX, Nagy F, Sivasubramaniam S, Chua NH. 1989. Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. Plant Cell. 1(1):141–150. doi:10.1105/tpc.1.1.141.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, *et al.* 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40(D1):D1178–D1186. doi:10.1093/nar/gkr944.

Haberle V, Stark A. 2018. Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Biol. 19(10):621–637. doi:10.1038/s41580-018-0028-8.

Hayashi H, Czaja I, Lubenow H, Schell J, Walden R. 1992. Activation of a plant gene by T-DNA tagging: auxin-independent growth in vitro. Science. 258(5086):1350–1353. doi:10.1126/science.1455228.

Jores T, Tonnies J, Wrightsman T, Buckler ES, Cuperus JT, Fields S, Queitsch C. 2021. Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. Nat Plants. 7(6):842–855. doi:10.1038/s41477-021-00932-y.

Kaeppler SM, de Leon N, Sekhon RS, Hansey C, Buell CR, Vaillancourt B, Springer N. 2012. B73 RNA-Seq Atlas. Bethesda (MD): NCBI Sequence Read Archive (SRA). SRP010680.

Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C, *et al.* 2014. Camelina sativa strain:DH55 Genome sequencing and assembly. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA231618.

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. 2016a. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. Plant J. 88(6):1058–1070. doi:10.1111/tpj.13312.

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. 2016b. High resolution transcriptomic development map of Arabidopsis thaliana based on RNA-seq. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA314076.

Kudapa H, Garg V, Chitikineni A, Varshney RK. 2017. Cicer arietinum gene expression atlas. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA413872.

Li H-T, Luo Y, Gan L, Ma P-F, Gao L-M, Yang J-B, Cai J, Gitzendanner MA, Fritsch PW, Zhang T, *et al.* 2021. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biol. 19(1):232. doi:10.1186/s12915-021-01166-2.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G. 2011. Glycine max transcriptome or gene expression. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA79597.

Loraine A, McCormick S, Estrada A, Patel K, Qin P. 2013a. High-throughput sequencing of Arabidopsis thaliana pollen cDNA uncovers novel transcription and alternative splicing. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA194429.

Loraine AE, McCormick S, Estrada A, Patel K, Qin P. 2013b. RNA-Seq of Arabidopsis pollen uncovers novel transcription and alternative splicing1[C][W][OA]. Plant Physiol. 162(2):1092–1109. doi:10.1104/pp.112.211441.

McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, *et al.* 2017a. Sorghum bicolor BTx623 gene atlas plate 8. Bethesda (MD): NCBI Sequence Read Archive (SRA). SRA558272.

McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, *et al.* 2017b. Sorghum bicolor BTx623 gene atlas plate 10. Bethesda (MD): NCBI Sequence Read Archive (SRA). SRA558539.

McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, *et al.* 2017c. Sorghum bicolor BTx623 gene atlas plate9. Bethesda (MD): NCBI Sequence Read Archive (SRA). SRA558514.

Mejía-Guerra MK, Li W, Galeano NF, Vidal M, Gray J, Doseff AI, Grotewold E. 2015. Core promoter plasticity between maize tissues and genotypes contrasts with predominance of sharp transcription initiation sites. Plant Cell. 27(12):3309–3320. doi:10.1105/tpc.15.00630.

Molina C, Grotewold E. 2005. Genome wide analysis of Arabidopsis core promoters. BMC Genomics. 6(1):25. doi:10.1186/1471-2164-6-25.

Moreno-Giménez E, Selma S, Calvache C, Orzáez D. 2022. GB_Synp: a modular dCas9-regulated synthetic promoter collection for fine-tuned recombinant gene expression in plants (p. 2022.04.28.489949). bioRxiv. doi:10.1101/2022.04.28.489949.

Moscow State University. 2014. RNA-seq analysis of an apical meristem time series reveals a critical point in Arabidopsis thaliana flower initiation. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA268115.

Odell JT, Nagy F, Chua N-H. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature. 313(6005):810–812. doi:10.1038/313810a0.

Patron NJ. 2020. Beyond natural: synthetic expansions of botanical form and function. New Phytol. 227(2):295–310. doi:10.1111/nph.16562.

Peanut Genome Consortium. 2015. Arachis hypogaea cultivar: Tifrunner transcriptome or gene expression. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA291488.

Penin AA, Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD. 2018. Transcriptome map of Solanum lycopersicum. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA507622.

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, *et al.* 2018. BCS cv-1 development. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJEB25639.

Schmitz RJ, Grotewold E, Stam M. 2022. Cis-regulatory sequences in plants: their importance, discovery, and future challenges. Plant Cell. 34(2):718–741. doi:10.1093/plcell/koab281.

South PF, Cavanagh AP, Liu HW, Ort DR. 2019. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. Science. 363(6422):eaat9077. doi:10.1126/science.aat9077.

Stelpflug SC, Sekhon RS, Vaillancourt B, Hirsch CN, Buell CR, de Leon N, Kaeppler SM. 2012. Zea mays transcriptome or gene expression. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA171684.

Sudheesh S, Sawbridge TI, Cogan NO, Kennedy P, Forster JW, Kaur S. 2015. Pisum sativum subsp. Sativum strain:Parafield transcriptome or gene expression. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA277076.

University of Tsukuba. 2017. RNA-seq study in 30 different tissues of muskmelon. NCBI Sequence Read Archive (SRA), Bethesda, MD. PRJDB6414.

Vlasova A, Capella-Gutiérrez S, Rendón-Anaya M, Hernández-Oñate M, Minoche AE, Erb I, Câmara F, Prieto-Barja P, Corvelo A, Sanseverino W, *et al.* 2016. Phaseolus genome sequencing and assembly. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA221782.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 10(1):57–63. doi:10.1038/nrg2484.

Wu Y, Wang Y, Li J, Li W, Zhang L, Li Y, Li X, Li J, Zhu L, Wu G. 2014. Development of a general method for detection and

quantification of the P35S promoter based on assessment of existing methods. Sci Rep. 4(1):7358. doi:10.1038/srep07358.

Potato Genome Sequencing Consortium, Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, *et al.* 2013. Transcriptome analysis of the potato (genotype RH89-039-16). Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJEB2430.

Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, Seki M, Shinozaki K, Abe T. 2007. Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genom. 8(1):67. doi:10.1186/1471-2164-8-67.

Yamamoto YY, Yoshioka Y, Hyakumachi M, Obokata J. 2011. Characteristics of core promoter types with respect to gene structure and expression in *Arabidopsis thaliana*. DNA Res. 18(5): 333–342. doi:10.1093/dnares/dsr020.

Yamamoto YY, Yoshitsugu T, Sakurai T, Seki M, Shinozaki K, Obokata J. 2009. Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. Plant J. 60(2):350–362. doi:10.1111/j.1365-313X.2009.03958.x.

Yang EJY, Nemhauser JL. 2022. Expanding the synthetic biology toolbox with a library of constitutive and repressible promoters (p. 2022.10.10.511673). bioRxiv. https://doi.org/10.1101/2022.10.10.511673.

Yao S, Jiang C, Huang Z, Torres-Jerez I, Chang J, Zhang H, Udvardi M, Liu R, Verdier J. 2017. Black-eyed pea gene expression atlas, Apr 11, 2017. Bethesda (MD): NCBI Sequence Read Archive (SRA). PRJNA389300.

Zahn-Zabal M, Dessimoz C, Glover NM. 2020. Identifying orthologs with OMA: a primer. F1000Res. 9:27. doi:10.12688/f1000research.21508.1.

Zhou A, Kirkpatrick LD, Ornelas IJ, Washington LJ, Hummel NFC, Gee CW, Tang SN, Barnum CR, Scheller HV, Shih PM. 2023. A suite of constitutive promoters for tuning gene expression in plants. ACS Synth Biol. 12(5):1533–1545. doi:10.1021/acssynbio.3c00075.

Editor: E. Akhunov