



universität
wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

Transcriptional requirements and diversifications in early
embryos of *Arabidopsis thaliana*

verfasst von / submitted by

Ping Kao

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on the student
record sheet:

UA 794 685 490

Dissertationsgebiet lt. Studienblatt /
field of study as it appears on the student record sheet:

Betreut von / Supervisor:

Molekulare Biologie

Mitbetreut von / Co-Supervisor:

Michael D. Nodine, PhD



This thesis is dedicated to all scientists,
may we keep being passionate and thriving,
and let it be a reminder that getting scooped is not the end.

Acknowledgements

I would like to thank Michael Nodine for giving me the opportunity to work on this exciting project. I am grateful to my PhD committee members Frédéric Berger, Kikuë Tachibana and Martin Bayer for their input, support and advice. Special thanks to the great VBC facilities, especially to the BioOptics, Next Generation Sequencing, Advanced Microscopy and Plant Sciences for their amazing technical support. Thanks to Tomokazu Kawashima, Sean Montgomery, Michael Borg, Zdravko Lorkovic, Bingkun Lei, Ortrun Mittelsten Scheid, Ruben Gutzat, Mattia Dona, Marco Incarbone, Magnus Nordborg and all the great colleagues at the VBC for the discussions, suggestions and all the funs. I am also grateful to all the current and former Nodine group members for the great experience in GMI. Last but not least, I would like to thank Kotoha Tanaka, Kaori Sakuramori, Yuhan Kao, all my friends and my family for the emotional support throughout the six-year journey.

Transcriptional requirements and diversifications in early embryos of *Arabidopsis thaliana*

1 Abstract	1
2 Zusammenfassung	2
3 Synopsis	3
4 Introduction	3
4.1 Zygotic embryogenesis and maternal-to-zygotic transition	3
4.2 Activations and diversifications of gene expression in Arabidopsis embryogenesis.....	4
4.3 Developments and applications of techniques.....	5
5 Goals of the thesis	5
6 Transcriptional Activation of Arabidopsis Zygotes Is Required for Initial Cell Divisions	6
Abstract	7
Introduction	7
Results.....	8
Expansion microscopy improves whole-mount fluorescent immunostaining.	8
Visualization of transcriptional activities in eggs and zygotes.	9
Active transcription is physiologically required for zygote cell division.....	9
Discussion.....	12
Materials and Methods	13
References.....	15
Acknowledgements.....	18
Author contributions.....	18
Additional information.....	18
Supplementary Information	19
7 Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis	22
Abstract.....	23
Background	23
Abstract.....	23
Background	23
Results.....	24
Embryos are enriched for transposon-derived small RNAs	24
Small RNAs from euchromatic and heterochromatic transposons exhibit distinct developmental dynamics.....	26
Embryonic methylome dynamics.....	28
Small RNA-directed methylation of transposons during embryogenesis.....	30
Chromatin regulates small RNA transcription.....	31
Homeostasis of transposon-derived siRNAs.....	33
Discussion.....	35
Conclusions.....	36

Methods.....	37
Supplementary information	39
Acknowledgements.....	39
Peer review information	39
Review history	39
Authors' contributions	40
Funding	40
Availability of data and materials	40
References.....	40
Publisher's Note.....	43
Additional file	44
8 Gene Expression Variation Across Arabidopsis Embryos at Single-Nucleus Resolution	51
SUMMARY.....	52
INTRODUCTION	53
RESULTS	54
DISCUSSION	59
METHODS.....	60
FIGURES AND LEGENDS	63
SUPPLEMENTAL INFORMATION.....	68
REFERENCES.....	75
9 Discussion	79
10 References	80

1 Abstract

Embryogenesis is a fundamental developmental process common to eukaryotes conducting sexual reproduction. Due to technical difficulties, our knowledge of the regulatory mechanisms during embryogenesis in flowering plants is limited compared to animals. In animal embryogenesis, a crucial period known as maternal-to-zygotic transition consisting of elimination of maternal genetic products and zygotic genome activation is required for cell specification. While maternal-to-zygotic transition is observed in multiple species, the timing, scale and the underlying mechanisms seem to vary among species. Although similar events have been reported in flowering plants, it is risky to assume any mechanism is conserved between the two kingdoms. In the model eudicots *Arabidopsis thaliana*, the timing and the scale of zygotic genome activation have been under debate for the past decade and was only resolved by a recent transcriptome study. Although transcriptional changes are often coupled with developmental processes, there has yet been genome-wide evidence indicating the requirement of active transcription in early embryogenesis. On the other hand, revealing the diverse gene expression profiles is fundamental to understanding the embryo body axes formation and cell specification mechanisms. Several studies revealed transcriptome dynamics along developmental stages, but a comprehensive gene expression atlas for various embryonic cell types is still required.

In this thesis, I investigate the requirements and diversifications of gene expression during early embryogenesis in *Arabidopsis thaliana*. Visualization of active transcription markers with expansion microscopy provides cytological evidence for globally active transcription in zygotes. Live-cell imaging on developing seed cultured with transcriptional inhibitors provides physiological evidence indicating active transcription is required in zygotes before the asymmetric cell division. These observations are consistent with the recently reported global post-fertilization transcriptome changes and indicates active transcription is required at the very beginning of embryogenesis. In addition, embryonic methylomes from a developmental time series reveal the cytosine methylation dynamics on euchromatin and heterochromatin transposable elements. Visualization of zygotic nuclei shows post-fertilization heterochromatin decondensation and suggests the production of 24-nt siRNA contributing to cytosine methylation. On the other hand, the single nucleus RNA-seq libraries provide contamination-free embryonic cell-type specific transcriptomes. Examining the expression dynamics of epigenetic regulators across embryonic cell types reveals the lineage-specific expression patterns and suggests hypomethylated suspensor genomes. Transcription factor binding site analyses reveal enriched transcription factors which may involve in the function or specification of individual cell types. The cell type-specific gene expressions could serve as a cornerstone for embryogenesis studies, and techniques I introduced to overcome the technical barriers would expand the toolkits for plant biologists

2 Zusammenfassung

Die Embryogenese ist ein grundlegender Entwicklungsprozess, der allen Eukaryoten gemeinsam ist, die sexuelle Fortpflanzung betreiben. Aufgrund technischer Schwierigkeiten ist unser Wissen über die Regulationsmechanismen während der Embryogenese bei blühenden Pflanzen im Vergleich zu Tieren begrenzt. In der tierischen Embryogenese ist eine entscheidende Periode, die als maternal-to-zygotischer Übergang bekannt ist und aus der Eliminierung der mütterlichen genetischen Produkte und der Aktivierung des zygotischen Genoms besteht, für die Zellspezifikation erforderlich. Während der maternale-zu-zygotische Übergang bei mehreren Arten beobachtet wird, scheinen der Zeitpunkt, das Ausmaß und die zugrundeliegenden Mechanismen zwischen den Arten zu variieren. Obwohl ähnliche Vorgänge bei Blütenpflanzen berichtet wurden, ist es riskant anzunehmen, dass irgendein Mechanismus zwischen den beiden Königreichen konserviert ist. In der eudikotylen Modellpflanze *Arabidopsis thaliana* sind der Zeitpunkt und das Ausmaß der zygotischen Genomaktivierung seit einem Jahrzehnt umstritten und wurden erst durch eine aktuelle Transkriptomstudie geklärt. Obwohl transkriptionelle Veränderungen oft mit Entwicklungsprozessen gekoppelt sind, gab es bisher keine genomweiten Hinweise auf die Notwendigkeit einer aktiven Transkription in der frühen Embryogenese. Andererseits ist die Aufdeckung der verschiedenen Genexpressionsprofile von grundlegender Bedeutung für das Verständnis der Bildung der Körperachsen des Embryos und der Zellspezifikationsmechanismen. Mehrere Studien haben die Transkriptomdynamik entlang der Entwicklungsstadien aufgedeckt, aber ein umfassender Genexpressionsatlas für verschiedene embryonale Zelltypen ist noch erforderlich.

In dieser Arbeit untersuche ich die Anforderungen und Diversifikationen der Genexpression während der frühen Embryogenese in *Arabidopsis thaliana*. Die Visualisierung aktiver Transkriptionsmarker mit Expansionsmikroskopie liefert zytologische Beweise für global aktive Transkription in Zygoten. Live-Cell-Imaging an sich entwickelnden Samen, die mit Transkriptioninhibitoren kultiviert wurden, liefert physiologische Hinweise darauf, dass aktive Transkription in Zygoten vor der asymmetrischen Zellteilung erforderlich ist. Diese Beobachtungen stimmen mit den kürzlich berichteten globalen Transkriptomveränderungen nach der Befruchtung überein und weisen darauf hin, dass eine aktive Transkription bereits zu Beginn der Embryogenese erforderlich ist. Zusätzlich zeigen embryonale Methylome aus einer Entwicklungszeitreihe die Cytosin-Methylierungsdynamik auf Euchromatin und Heterochromatin transponierbaren Elementen. Die Visualisierung der zygotischen Kerne zeigt die Dekondensation des Heterochromatins nach der Befruchtung und deutet auf die Produktion von 24-nt siRNA hin, die zur Cytosin-Methylierung beiträgt. Andererseits liefern die Einzelkern-RNA-seq-Bibliotheken kontaminationsfreie embryonale zelltypspezifische Transkriptome. Die Untersuchung der Expressionsdynamik epigenetischer Regulatoren über embryonale Zelltypen hinweg offenbart die linienspezifischen Expressionsmuster und lässt auf hypomethylierte Suspensor-Genome schließen. Analysen von Transkriptionsfaktor-Bindungsstellen zeigen eine Anreicherung von Transkriptionsfaktoren, die möglicherweise an der Funktion oder Spezifikation einzelner Zelltypen beteiligt sind. Die zelltypspezifischen Genexpressionen könnten als Eckpfeiler für Studien zur Embryogenese dienen, und die von mir eingeführten Techniken zur Überwindung der technischen Barrieren würden den Werkzeugkasten für Pflanzenbiologen erweitern

3 Synopsis

This thesis presents the requirements and diversifications of gene expression in early embryos of *Arabidopsis thaliana*. It also demonstrates the techniques we developed for solving biological questions. The first study focuses on the requirement of active transcription in early embryogenesis (Kao and Nodine, 2019). It introduces the expansion microscopy (ExM) and chemical incubation techniques, and presents cytological and physiological evidence showing that active transcription is required for the asymmetric zygotic division. The second study presents the methylome dynamics of developing embryos and demonstrates the less compact chromatin in early embryos with ExM (Papareddy *et al.*, 2020). The third study focuses on resolving the transcriptional profiles of different cell types established at the globular stage. It presents a single nucleus RNA-seq technique to acquire clean embryonic transcriptomes at single cell resolution. It then identifies and validates the cell-type specific expression profiles and highlights a few obvious features.

4 Introduction

4.1 Zygotic embryogenesis and maternal-to-zygotic transition

Sexual reproduction could increase gene variety and provide adaptive diversification and is common among eukaryotes (Clark, Aagaard and Swanson, 2006; Bai, 2015). Maternal genome from eggs and paternal genome from sperms fuse to create the zygotic genome after fertilization, and the epigenetic modifications specific to parental gametes are reprogrammed in order to establish new cell types (Wilmsen and Scheres, 2004; Shi and Wu, 2009; Gerri *et al.*, 2020). The resulting zygotes go through a developmental process known as zygotic embryogenesis, during which a single cell zygote proliferates and differentiates under specific programmes to form a complex embryo. Flowering plants have evolved unique sporophyte-gametophyte life cycles and the beginning of sporophytic phase is marked by the fertilization between eggs and sperms (Walbot and Evans, 2003). The zygotes and future embryos are embedded in versatile maternal structures collectively called seeds, which is the hallmark of flowering plants (Linkies *et al.*, 2010). The seeds protect the interior embryos and germinate in suitable environments to start post-embryonic developments, including vegetative growth and reproductive growth and eventually the production of the next generation.

In animals, the cell divisions and polarity establishments in early embryogenesis are directed by the inherited maternal gene products as the zygotic genome remains transcriptionally silent (Baroux *et al.*, 2008; Tadros and Lipshitz, 2009; Walser and Lipshitz, 2011; Lee, Bonneau and Giraldez, 2014). Later in embryogenesis, a crucial event known as maternal-to-zygotic transition (MZT) replaces the maternal controls with zygotic controls. Two independent events consist of MZT: the elimination of inherited maternal gene products (maternal elimination, ME) and the activation of zygotic genome (zygotic genome activation, ZGA). After MZT, the pro-embryos undergo more detailed specifications and form the diverse tissues, organs and eventually the embryos under the instructions of zygotic programmes (St Johnston and Nüsslein-Volhard, 1992; Beddington and Robertson, 1999). Interestingly, while MZT universally exists in most if not all species conducting sexual reproduction, the timing, the scale, the order of ME and ZGA and the underlying mechanisms are diverse among species. Revealing the unique aspects of MZT is essential to understand embryogenesis in any species.

In flowering plants, molecular and histological evidence in several species indicate increased transcriptional activities and transcriptome reconstruction soon after fertilization in zygotes (Sauter *et al.*, 1998; Dresselhaus *et al.*, 1999; Okamoto *et al.*, 2005; Sprunck *et al.*, 2005; Ning *et al.*, 2006; Meyer and Scholten, 2007; Pięciński *et al.*, 2008; Zhao *et al.*, 2011; Niedojadło *et al.*, 2012; Abiko *et al.*, 2013; Anderson *et al.*, 2013, 2017; Domoki *et al.*, 2013; Ohnishi, Hoshino and Okamoto, 2014; Chen *et al.*, 2017; Ohnishi *et al.*, 2019). These observations support the idea that ZGA occurs in zygotes. While it is conventionally assumed that flowering plants also have MZT during embryogenesis, there is little evidence indicating a transient, if any, genome-wide transcriptional quiescence in zygotes. Moreover, the ability of conducting asexual reproduction (Schmidt, Schmid and Grossniklaus, 2015; Wibowo *et al.*, 2018) and somatic embryogenesis (Zimmerman, 1993; Wójcikowska and Gaj, 2016; Horstman, Bemer and Boutilier, 2017) from various tissues suggests the the plant genomes are highly plastic and the transcriptional profiles can be reshaped when the right stimuli are presented. The evidence leads to a hypothesis that a direct gametophyte-to-sporophyte transition instead of an off-and-on maternal-to-zygotic transition occurs during zygotic embryogenesis in flowering plants (Kao

and Nodine, 2019). Nonetheless, the term ZGA is used to represent the global increase of transcriptional activity during embryogenesis.

4.2 Activations and diversifications of gene expression in *Arabidopsis* embryogenesis

Arabidopsis thaliana (*Arabidopsis*) has been a model organism for the studying zygotic embryogenesis in flowering plants for decades. Embryogenesis presumably consists of an early morphogenesis phase where the body plans and basic cell types are established, and a later maturation phase where the embryos prepare themselves for dormancy within seeds (Park and Harada, 2008). In brief, the zygotes divide asymmetrically and produce a smaller apical cell which is the precursor of most embryonic cells, and a larger basal cell which will become the extraembryonic tissue known as suspensor and contribute to the root tips of embryos. The apical cell will go through a series of well-defined vertical and horizontal cell divisions and establish the body axes, while the basal cell will divide horizontally to produce a file of suspensor cells which connect and nurture the developing embryos (Long and Barton, 1998; Capron *et al.*, 2009; ten Hove, Lu and Weijers, 2015). The precursors of the fundamental cell types are established at the globular stage and continue to differentiate to form a variety of distinct and well-defined tissues and stages (Park and Harada, 2008). The stereotypic cell division patterns, the short life cycle and the abundant genetic resources make *Arabidopsis* an ideal model organisms to study embryogenesis in eudicots. However, because the plant embryos are protected by layers of maternal tissue known as integuments (or seed coats after fertilization) and tightly connected to the surrounding endosperms, it is technically challenging to observe, manipulate or isolate developing embryos.

Transcriptional changes are coupled with developmental events and cell type specifications (Huijser and Schmid, 2011; ten Hove, Lu and Weijers, 2015). Decades of genetic analyses identified over 500 embryonic defective (EMB) genes and their mutants resulted in defective embryonic phenotypes at various developmental stages (Meinke, 2019). The mutant phenotypes, especially the ones in early stages, indicate that active expressions of EMB genes are required early in embryogenesis (Meinke *et al.*, 2008; Meinke, 2019). More detailed studies reveal a variety of genes contributing to zygotic embryogenesis and lead to two conflicting models differ in the timing and scale of ZGA. The first model proposes that the embryogenesis in *Arabidopsis* is similar to animals in that the zygotic genome remains transcriptionally quiescent until the globular stage, and the inherited maternal genetic products are directing the early embryogenesis (Vielle-Calzada, Baskar and Grossniklaus, 2000; Pillot *et al.*, 2010; Autran *et al.*, 2011; Del Toro-De León, García-Aguilar and Gillmor, 2014; García-Aguilar and Gillmor, 2015; Armenta-Medina and Gillmor, 2019). This model suggests the body axes of embryos and the initial cell types are defined by the maternal genome and imprinted maternal genes. The second model proposes that the zygotic genome is active soon after fertilization, similar to the proposed models of other flowering plants (Sauter *et al.*, 1998; Dresselhaus *et al.*, 1999; Okamoto *et al.*, 2005; Sprunck *et al.*, 2005; Ning *et al.*, 2006; Meyer and Scholten, 2007; Pięciński *et al.*, 2008; Zhao *et al.*, 2011; Niedojadło *et al.*, 2012; Abiko *et al.*, 2013; Anderson *et al.*, 2013, 2017; Domoki *et al.*, 2013; Ohnishi, Hoshino and Okamoto, 2014; Chen *et al.*, 2017; Ohnishi *et al.*, 2019). This model suggests the body axes establishments and cell specifications are directed by the zygotic genome. The recessive segregation pattern of a number of mutants showing defects in early stages (Liu and Meinke, 1998; Ronceret *et al.*, 2005; Xu *et al.*, 2005; Lin *et al.*, 2007; Arnaud Ronceret *et al.*, 2008; A. Ronceret *et al.*, 2008; Yu *et al.*, 2012, 2016; Guo *et al.*, 2016) and equal paternal contributions to zygotic and early embryonic transcriptomes (Nodine and Bartel, 2012; Zhao *et al.*, 2019) favor the second model and indicate genome-wide transcriptional changes before the asymmetric zygotic division. While the transcriptome conflicts between the two models has been resolved in a recent publication (Schon and Nodine, 2017), there is limited information connecting the transcriptional changes to developmental requirements on a genome-wide scale.

Several studies reveal the transcriptome dynamics among different developmental stages (Belmonte *et al.*, 2013; Hofmann, Schon and Nodine, 2019). On the other hand, the transcriptome diversities among the embryonic cell types are less known. Studies utilizing the cell-specific fluorescent reporters and various target nuclei enrichment methods provide more details to the spatial transcriptional dynamics (Slane *et al.*, 2014; Palovaara *et al.*, 2017), but the limitation of available reporters and the maternal contaminations compromise the datasets. Other studies reveal the transcriptional dynamics between embryo proper and suspensors (Belmonte *et al.*, 2013; Zhou *et al.*, 2020), but finer resolution is required in order to reveal the transcriptional features among individual cell types.

4.3 Developments and applications of techniques

Expansion microscopy (ExM) is a technique to visualize cytological events (Chen, Tillberg and Boyden, 2015; Tillberg *et al.*, 2016). The ExM technique embeds samples in gel matrices and expands samples uniformly in three dimensions by osmotic pressure. This results in samples with higher microscopic resolution while retaining relative spatial positions and increased reagent accessibility. I adapted and modified ExM to Arabidopsis developing seeds and established a protocol to compromise the cell wall while retaining the protein content. The gel matrices result in better morphology and yield by providing physical support to cavities such as embryo sacs and reducing autofluorescence backgrounds. Applying ExM to plant tissues provides a more robust option for plant biologists to visualize proteins of interest or cellular events.

A seed culture and live-cell imaging system was introduced to increase the survival rate of embryos and allowed real time observation (Gooh *et al.*, 2015). I expanded the application of this system by applying various chemicals to culture media, which allowed me to study the embryonic responses to global transcriptional or translational inhibition. By mimicking the classical studies done in animal embryos (Braude *et al.*, 1979; Newport and Kirschner, 1982; Edgar, Wolf and Wood, 1994; Edgar and Datar, 1996; Zamir, Kam and Yarden, 1997), I was able to pinpoint when the global activation is required during zygotic embryogenesis.

The blooming next generation sequencing (NGS) technologies promote the development of single cell RNA-seq techniques (Hwang, Lee and Bang, 2018). Different systems have different requirements for the input samples and instruments. I modified published protocols (Picelli, Björklund, *et al.*, 2014; Picelli, Faridani, *et al.*, 2014; Habib *et al.*, 2016) for single nucleus RNA-seq (snRNA-seq) and optimized the nuclei isolation (Galbraith *et al.*, 1983) and fluorescent activated nuclei sorting (FANS) protocols to increase the purity of samples. I utilized this system to gain non-contaminated embryonic transcriptomes at single cell resolution and investigated the transcriptional diversities among embryonic cell types.

5 Goals of the thesis

There are two main goals of this thesis. The first goal is to investigate the requirements of active transcription in early zygotic embryogenesis. It will provide the missing evidence linking the observed ZGA in zygotes to developmental processes. To achieve the goal, I utilized the ExM immunofluorescent staining and live-cell imaging techniques to gain histological and physiological evidence for transcriptional activities in zygotes and early embryos. The second goal is to investigate the diverse gene expression programs in various embryonic cell types. It will provide a detailed expression atlas of globular embryos necessary for revealing the functions and specification mechanisms. To achieve the goal, I utilized the nuclei isolation, FANS enrichment and snRNA-seq techniques to gain embryonic transcriptomes at single cell resolution. The subsequent analyses will identify cell type-specific transcriptomes as well as prominent gene expression diversities.

6 Transcriptional Activation of Arabidopsis Zygotes Is Required for Initial Cell Divisions

Type	<i>published</i> Scientific reports, 9(1), p. 17159. (2019) doi: 10.1038/s41598-019-53704-2
Authors	Ping Kao & Michael D. Nodine
Contributions	Conceptualization, P.K. and M.D.N.; Methodology, P.K.; Formal Analysis, P.K.; Investigation, P.K.; Writing – Original Draft, P.K.; Writing – Review & Editing, M.D.N. and P.K.; Validation, M.N.; Visualization, P.K.; Supervision, M.N.

open

Transcriptional Activation of Arabidopsis Zygotes is Required for Initial Cell Divisions

Ping Kao & Michael D. Nodine *

Abstract

Commonly referred to as the maternal-to-zygotic transition, the shift of developmental control from maternal-to-zygotic genomes is a key event during animal and plant embryogenesis. Together with the degradation of parental gene products, the increased transcriptional activities of the zygotic genome remodels the early embryonic transcriptome during this transition. Although evidence from multiple flowering plants suggests that zygotes become transcriptionally active soon after fertilization, the timing and developmental requirements of zygotic genome activation in *Arabidopsis thaliana* (*Arabidopsis*) remained a matter of debate until recently. In this report, we optimized an expansion microscopy technique for robust immunostaining of Arabidopsis ovules and seeds. This enabled the detection of marks indicative of active transcription in zygotes before the first cell division. Moreover, we employed a live-imaging culture system together with transcriptional inhibitors to demonstrate that such active transcription is physiologically required in zygotes and early embryos. Our results indicate that zygotic genome activation occurs soon after fertilization and is required for the initial zygotic divisions in Arabidopsis.

Introduction

The transition of developmental control from parental-to-zygotic genomes is a pivotal event during animal and plant development. In animals, maternally inherited gene products regulate early embryogenesis while the zygotic genome remains transcriptionally quiescent until the maternal-to-zygotic transition (MZT). Two interdependent events constitute the MZT in animals: the degradation of inherited maternal gene products and zygotic genome activation (ZGA) when the zygotic genome breaks transcriptional quiescence to produce transcripts that instruct subsequent embryogenesis. While the MZT is universal in multiple species, the underlying mechanisms, scale and timing of ZGA are diverse (reviewed in ref.^{1–4}). Investigating how different species evolved various mechanisms to initiate ZGA is crucial to understanding embryogenesis.

Compared to animals, the knowledge of timing and requirements of ZGA in flowering plants is limited. Histological and molecular evidence in *Hyacinthus orientalis* (hyacinth)^{5,6}, *Nicotiana tabacum* (tobacco)^{7,8}, *Oryza sativa* (rice)^{9–13}, *Triticum aestivum* (wheat)^{14,15} and *Zea mays* (maize)^{16–20} altogether indicate that large-scale transcriptional activities increase in zygotes after fertilization and prior to the first division. These results suggest that, similar to animals, plant zygotic genomes may also transition from a transcriptionally quiescent to active state. However, plant and animal life cycles are fundamentally different, where plants alternate between haploid gametophytic and diploid sporophytic phases²¹. More specifically, a subset of sporophytic cells undergo meiosis to produce haploid spores, which divide mitotically to generate multicellular gametophytes containing eggs and sperms. Fertilization of the egg cell contained within each female gametophyte marks the onset of the sporophytic generation. Although it is unclear how similar the gametophytic-to-sporophytic transition in plants is to the MZT in animals, we have referred to the large-scale increase of transcriptional activities after fertilization as ZGA in the following text.

Although ZGA has been partially characterized in the model flowering plant *Arabidopsis thaliana* (*Arabidopsis*), the timing, parental contributions and requirements of ZGA was debatable. One model proposed that Arabidopsis zygotes are transcriptionally quiescent²² and early embryos mostly rely on maternal gene

products for growth and division^{23–27}. However, several mutants exhibiting defects in the initial asymmetric division of the zygote segregate in a recessive manner consistent with transcriptional activities of either parental allele being sufficient for the first zygotic division^{28–35}. Moreover, transcriptome analyses indicated equal parental genomic contributions to the embryonic transcriptome as early as the 1-cell/2-cell stage³⁶. Based on these results, it was

Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030, Vienna, Austria. *email: michael.nodine@gmi.oeaw.ac.at

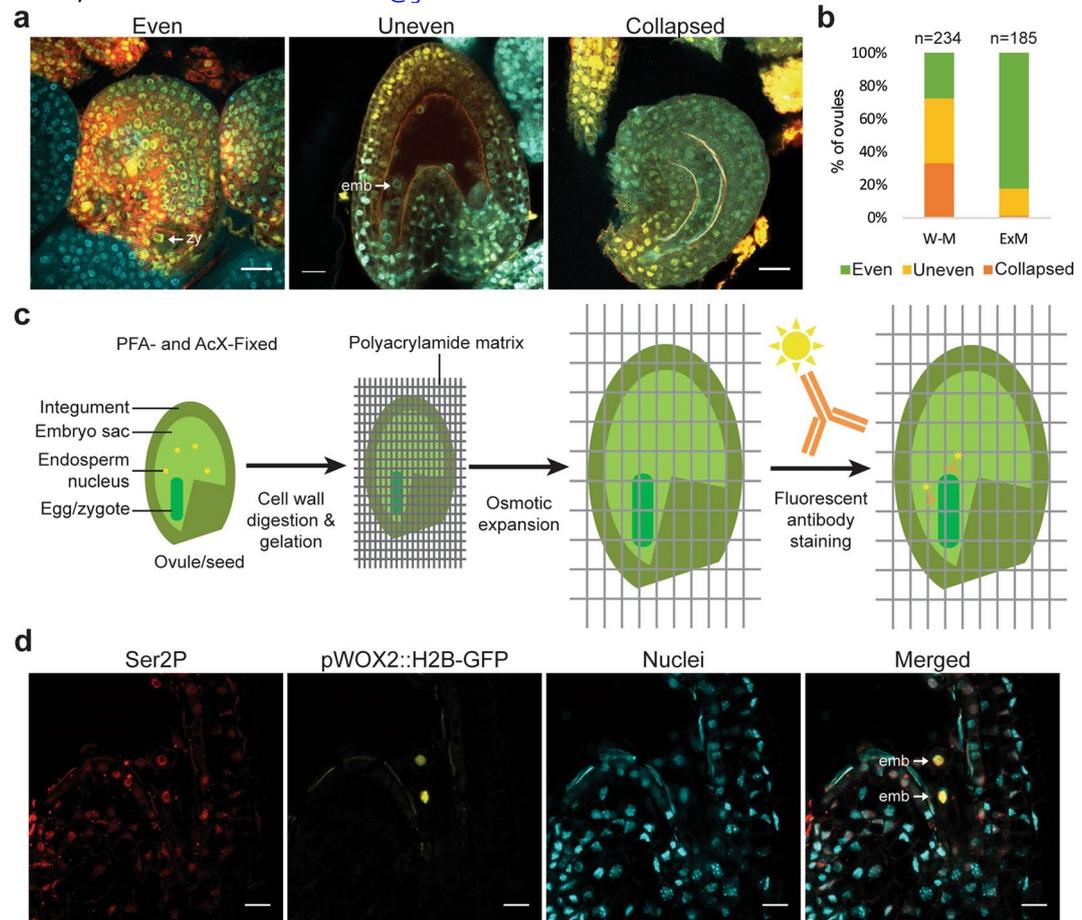


Figure 1. Expansion microscopy on Arabidopsis seeds. (a) Representative images of evenly stained samples (Even), unevenly stained samples (Uneven), or collapsed samples (Collapsed). Tubulin (red) and RNAPII Ser2P (yellow) were detected with immunofluorescence, and nuclei were stained with DAPI (cyan). (b) Quantification of the number of seeds with either even or uneven staining, or that were collapsed when using conventional whole-mount (W-M) or expansion microscopy (ExM) protocols. The total number of seeds examined with each method is indicated. (c) Schematic of expansion microscopy workflow. Ovules or seeds were fixed, incubated with cell wall digestion enzymes, embedded in a polyacrylamide gel matrix and osmotically expanded before fluorescent immunostaining (See Materials and Methods for details). (d) Protein retention after expansion in embryos. Immunofluorescent signal from DAPI-stained nuclei (cyan), RNAPII Ser2P (red) or HISTONE 2B-tagged GFP expressed with the embryo-specific WOX2 promoter (pWOX2::H2B-GFP; yellow) are shown. Scale bars represent 20 μm . zy, zygote; emb, embryo.

proposed that the zygotic genome is activated within the first few hours after fertilization with equal contributions of maternal and paternal alleles to the transcriptome³⁶. Although the maternal transcriptome dominance reported in a conflicting publication²⁵ can be readily explained by the amount of maternal RNA contamination in the samples³⁷, the precise timing and requirements of zygotic genome activation was unresolved until recently³⁸. Here, we provide independent evidence by expansion microscopy and live-cell imaging to demonstrate that transcriptional activities are markedly increased soon after fertilization in Arabidopsis and that zygotic transcription is essential for the initial embryonic cell divisions.

Results

Expansion microscopy improves whole-mount fluorescent immunostaining.

Phosphorylation of serine 2 on the carboxy-terminal domain of RNA polymerase II (RNAPII Ser2P) indicates elongating polymerase^{39,40}. Therefore, we used conventional whole-mount fluorescent immunostaining^{22,41} on fertilized ovules (seeds) to detect evidence for RNAPII transcriptional activities in zygotes and embryos. We also stained against tubulin with antibodies and chromatin with 4',6-diamidino-2-phenylindole (DAPI) to unambiguously identify egg and zygote nuclei because tubulin separates the zygote nucleus from surrounding endosperm nuclei. We obtained several samples with uniform and high signals, but found that the conventional protocol produced inconsistent results (Fig. 1a). Namely, 92/234 (39.3%) of samples exhibited uneven or no signal likely due to limited antibody accessibility (Fig. 1b). Embryos in particular had low signals because they were embedded within seeds. Moreover, 77/234 (32.9%) samples had collapsed embryo sacs, which contain the embryos (Fig. 1c), and thus were impossible to analyze (Fig. 1b). We therefore could not robustly detect RNAPII Ser2P with the conventional immunostaining protocol.

To improve the whole-mount fluorescent immunostaining method for Arabidopsis ovules and seeds, we adapted an expansion microscopy protocol (ExM)^{42,43}. The ExM technique physically expands specimen uniformly in three dimensions to increase reagent accessibility and microscopic resolution while retaining the relative spatial positions of the signals. Samples were fixed in 4% paraformaldehyde and 0.1 mg/mL Acryloyl-X, SE (6-((acryloyl)amino)hexanoic acid, succinimidyl ester) (AcX), incubated with cell wall-digesting enzymes and embedded in an expandable polyacrylamide gel matrix (Fig. 1c). Samples were then expanded by osmotic pressure to improve antibody penetration and increase specimen size before immunostaining. We examined protein retention in expanded samples in seeds expressing a previously described embryo-specific reporter line (pWOX2::H2B-GFP, pWOX2::tdTomato-RCI2b)⁴⁴. As expected, the nuclear-localized H2B-GFP signal was confined to embryonic nuclei indicating that our protocol successfully retained proteins in their proper subcellular localizations (Fig. 1d). We also stained against RNAPII Ser2P and found that the epitopes were retained and detectable with antibodies (Fig. 1d). To examine expansion ratios in three dimensions, we stained cell walls with SCR1 Renaissance 2200 and nuclei with DAPI, and examined samples before and after expansion. Although expansions were not uniform in all three dimensions, ranging from 1.3 to 2.0 times in length depending on the experiment, the RNAPII Ser2P signal was high and consistent among samples. With the ExM technique we found that only 2/185 (1.1%) samples had collapsed embryo sacs and 31/185 (16.7%) had uneven staining, while 152/185 (82.2%) samples had high and consistent signal (Fig. 1b). Therefore, by physically expanding the specimens with the modified ExM technique before staining, we were able to make interior tissues more accessible to reagents. This method produced more consistent staining and enabled the robust detection of subcellular marks within zygotes.

Visualization of transcriptional activities in eggs and zygotes.

To infer transcriptional activities, we then employed the ExM approach described above to visualize RNAPII Ser2P in eggs and zygotes. We also stained against tubulin with antibodies and chromatin with DAPI to identify egg and zygote nuclei, and observed staining patterns that were consistent with previous reports^{45,46} (Supplementary Videos S1 and S2). In unfertilized ovules, the RNAPII Ser2P signal was low in 31 out of 37 (83.8%) of egg nuclei examined, but high in surrounding sporophytic integument nuclei (Fig. 2a). In contrast, after fertilization the RNAPII Ser2P signal was high in 35 out of 38 (92.1%) of zygote nuclei examined, as well as in surrounding endosperm and integument nuclei (Fig. 2a).

Trimethylation marks on histone H3 lysine 36 (H3K36me3) are deposited during RNAPII transcription elongation⁴⁷. Therefore, we examined H3K36me3 levels in ovules and seeds using ExM to further inspect transcriptional activities in eggs and zygotes, respectively. Similar to RNAPII Ser2P, H3K36me3 signal was low in the majority of egg nuclei (30/36, 83.3%), but high in the majority of zygote nuclei (30/33, 90.9%), as well as in endosperm and surrounding sporophytic integument nuclei (Fig. 2b). The detection of H3K36me3 and RNAPII Ser2P in eggs and zygotes indicated that RNAPII transcription was relatively low in eggs and became highly active in zygotes after fertilization.

Active transcription is physiologically required for zygote cell division.

Classic transcriptional inhibition experiments in mouse⁴⁸, *Drosophila melanogaster*⁴⁹, *Caenorhabditis elegans*⁵⁰, *Xenopus laevis*⁵¹ and *Danio rerio*⁵² early embryos elegantly demonstrated that inherited gene products were sufficient for early embryogenesis. Therefore, to test whether active transcription is required for Arabidopsis zygote development, we cultured seeds in the presence of transcriptional inhibitors and examined zygotic cell division patterns with live-imaging microscopy⁴⁴. If inherited gene products were sufficient for early embryogenesis, then we expected no developmental delay or arrest when transcriptional inhibitors were included in the culture media. Seeds were cultured in the presence or absence of transcriptional inhibitors for one hour before acquiring the first image, and zygotes were labelled with nuclear-localized GFP and plasma membrane-localized tdTomato both of which were under the control of the embryo-specific WOX2 promoter (pWOX2::H2B-GFP, pWOX2::tdTomato-RCI2b)⁴⁴. The nuclear-localized GFP signal can be observed in elongated zygotes before they divide and served as a marker specific to zygotes and early embryos. Similar to what was originally reported for this culture system⁴⁴, we found that approximately 80% of embryos survived in our system under all conditions (Fig. 3i). That is, regardless of culture conditions approximately 20% of embryos lost fluorescent signal likely due to technical limitations of the system. We classified these embryos as “dead” and the remaining ones as “alive”, and only these surviving embryos were informative to test our hypothesis. We further categorized the alive embryos as “normal”, “delayed”, or “arrested” if the observed cell cycle duration was <12, ≥12 or ≥20 hours, respectively. Embryos with only one cell division observed were conservatively estimated to be delayed and categorized as “delayed (est.)”. As a positive control, we first compared cell division patterns of embryos cultured with or without the fla-

vopiridol (FLP) kinase inhibitor, which causes cell-cycle arrest⁵³. As expected, when cultured in the control Nitsch medium with 5% Trehalose (N5T) most (37/39, 94.9%) embryos divided normally similar to previous reports^{44,46,54} (Fig. 3a,j and Supplementary Video S3) with a 7.5-hour median cell cycle duration (Fig. 3k). In contrast, most embryos cultured with 100 μ M FLP had arrested cell division (30/32, 93.8%; Fig. 3b,j and Supplementary Videos S4, P -value = 1.1×10^{-14} , chi-squared test). Furthermore, the distribution of the cell cycle duration was significantly greater compared to N5T (Fig. 3k; P -value = 3.8×10^{-15} , two-sample Kolmogorov-Smirnov test). This control test indicated that our system was capable of capturing cell cycle arrest in early embryogenesis.

Similar to N5T control media, most embryos within seeds cultured with 0.5% dimethyl sulfoxide (DMSO) divided normally (56/60, 93.3%; Fig. 3c,j and Supplementary Video S5; P -value = 0.99, chi-squared test) and had a 7.8-hour median cell cycle duration (P -value = 0.99, two-sample Kolmogorov-Smirnov test). Triptolide (TPL) induces protease-dependent degradation of RNAPII components^{55,56}. We cultured seeds with TPL to test whether active transcription was required for zygotic divisions and early embryogenesis. When cultured with 500 μ M TPL and 0.5% DMSO, 51/86 (59.3%) embryos were arrested, 32/86 (37.2%) had delayed cell

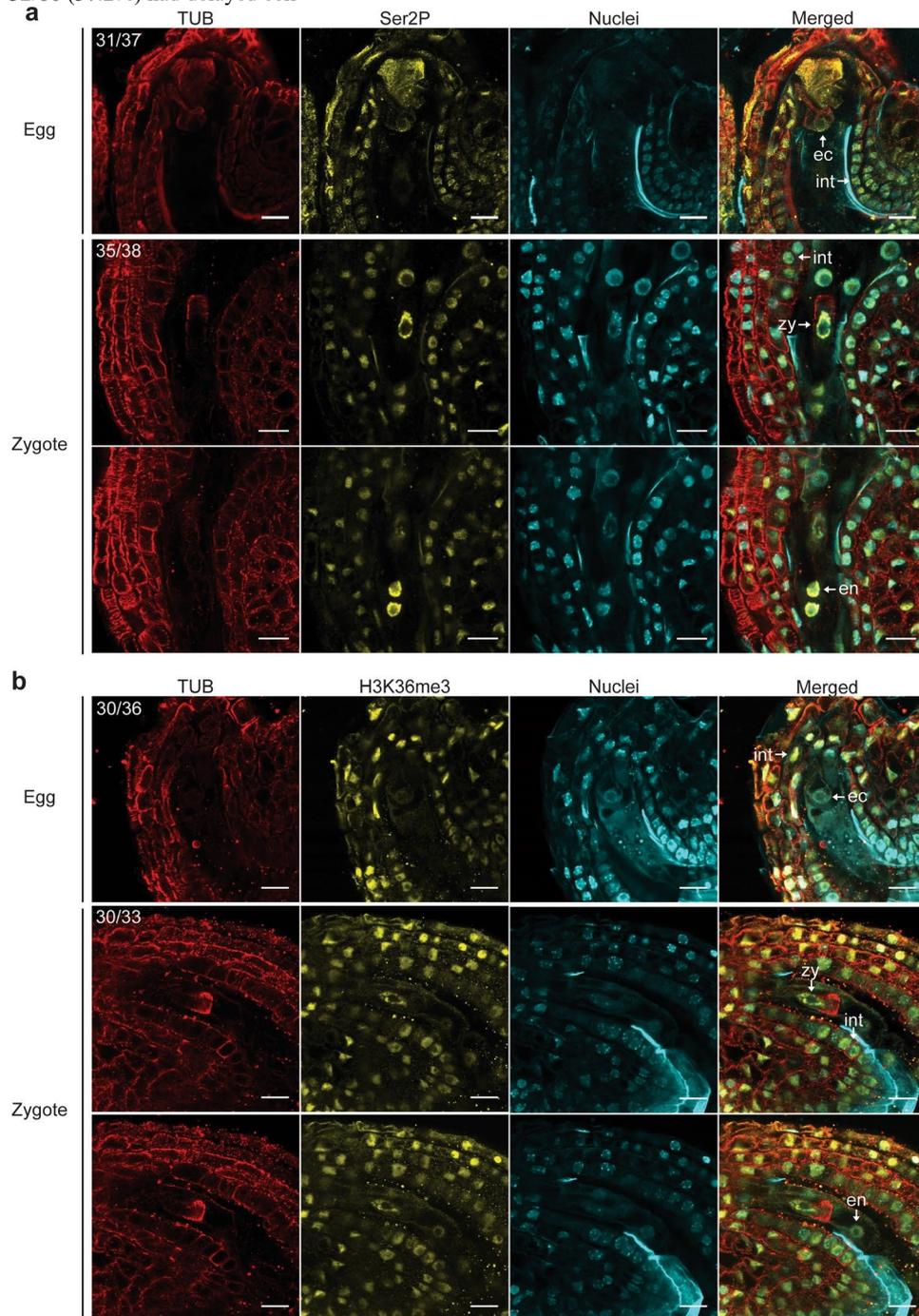


Figure 2. Visualization of transcriptional activities in eggs and zygotes. (a) Representative ExM images of tubulin (red), RNAPII Ser2P (yellow) and DAPI-stained nuclei (cyan) in ovules containing eggs (*top*) or seeds containing zygotes with focal planes on zygotes (*middle*) or endosperm (*bottom*). The number of ovules or seeds with similar staining patterns out of the total number examined are indicated. (b) Representative ExM images of tubulin (red), H3K36me3 (yellow) and nuclei (cyan) in ovules/seeds containing eggs (*top*) or zygotes with focal planes on zygotes (*middle*) or endosperm (*bottom*). The number of ovules or seeds with similar staining patterns out of the total number examined are indicated. Scale bars represent 20 μm . ec, egg cell; zy, zygote; en, endosperm; int, integument.

divisions and only 3/86 (3.5%) developed normally (Fig. 3d,j and Supplementary Video S6). The proportion of embryos that were arrested or delayed was significantly greater than observed for the DMSO control condition (P -value = 8.8×10^{-27} , chi-squared test), and the distribution of cell cycle duration was significantly greater compared to DMSO (Fig. 3k, P -value = 2.2×10^{-16} , two-sample Kolmogorov-Smirnov test). Moreover, TPL had a dose-dependent effect on embryo development whereupon when seeds were cultured with 250 μM TPL, 20/63

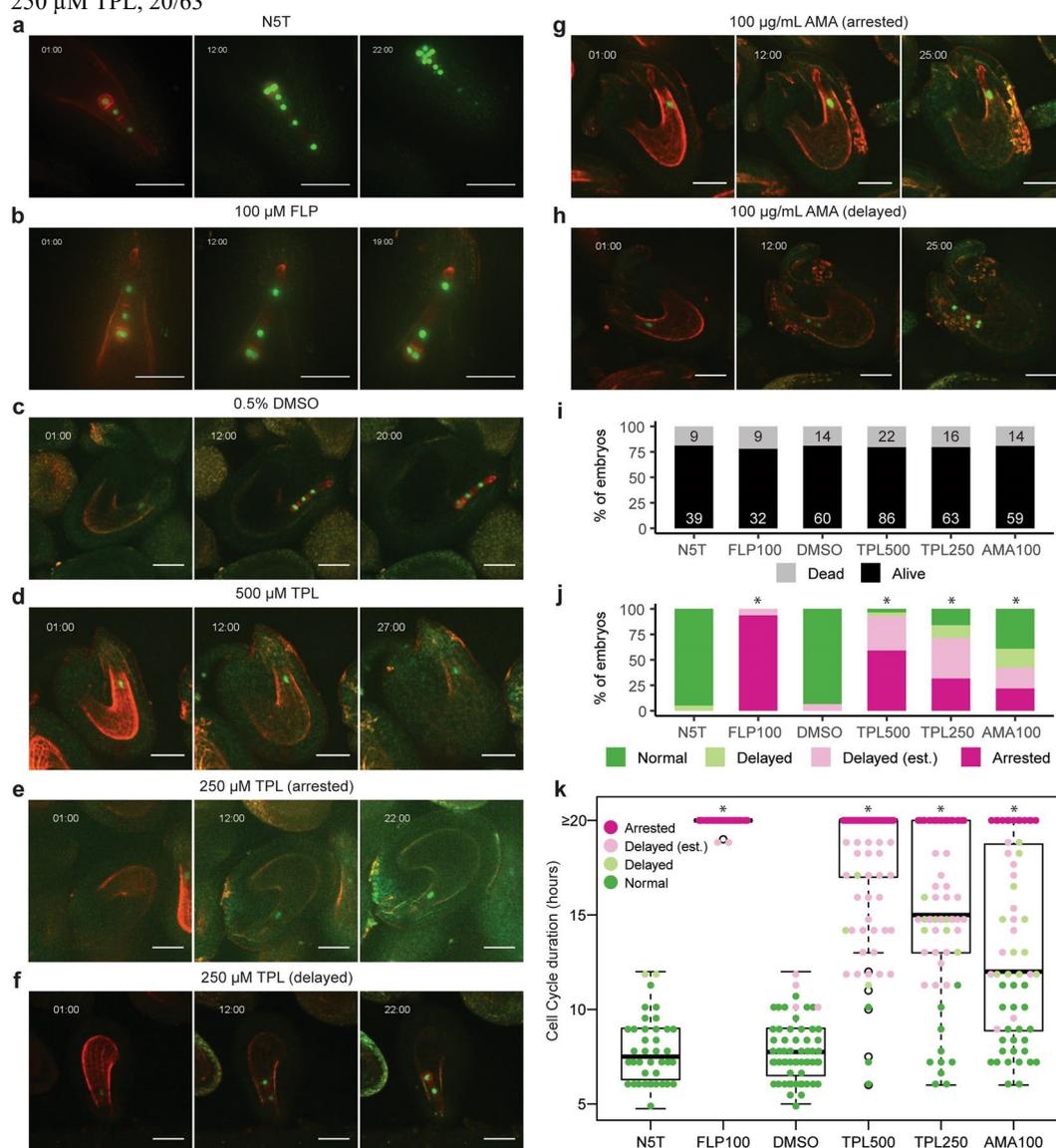


Figure 3. Transcription is required for initial embryonic cell divisions. (a–h) Time-lapse observations of embryos expressing pWOX2::H2B-GFP, pWOX2::tdTomato-RCI2b cultured in (a) N5T, (b) N5T with 100 μM FLP, (c) N5T with 0.5% DMSO, (d) N5T with 0.5% DMSO and 500 μM TPL, (e,f) N5T with 0.5% DMSO and 250 μM TPL or (g,h) N5T with 0.5% DMSO and 100 $\mu\text{g}/\text{mL}$ AMA. The time after incubation is indicated as hours:minutes. (i) Bar chart illustrating the percentage of embryos that survived (Alive) or died (Dead) under various conditions. The number of dead and alive embryos are indicated in black and white, respectively. (j) Bar chart showing the percentage of embryos that divided normally or abnormally when cultured under various conditions. The total number of surviving embryos for each condition are indi-

cated in Supplementary Table S1. Asterisks indicate P values < 0.001 based on chi-squared tests that the observed differences in frequencies between test and control groups were related to different culture conditions. **(k)** Graph illustrating cell cycle durations for each culture condition based on the observed duration between cell divisions or the lower estimations on the duration (est.). Because of limitations on recording time, cell cycle durations for arrested embryos were set to an upper limit of 20 hours. The number of embryos examined for each condition are equal to the number of alive embryos labelled in (j), and the detail number of embryos for each phenotype are indicated in Supplementary Table S1. Each dot represents a biological replicate, thick horizontal bars indicate medians, and the top and bottom edges of the box indicate the 75th and 25th percentile, respectively. Asterisks indicate significantly different populations determined by two-sample Kolmogorov-Smirnov tests, P -values < 0.001 . The mean and median cell cycle durations are indicated in Supplementary Table S1. Scale bars represent 50 μm . N5T, Nitsch medium with 5% Trehalose; FLP, flavopiridol; DMSO, dimethyl sulfoxide; TPL, triptolide; AMA, α -amanitin.

(31.7%) embryos were arrested (Fig. 3e,j and Supplementary Videos S7), 33/63 (52.4%) had delayed divisions (Fig. 3f,j and Supplementary Videos S8) and 10/63 (15.9%) developed normally (P -value = 3.4×10^{-17} compared to DMSO control, chi-squared test). The distribution of cell cycle duration was significantly greater compared to DMSO but significantly less compared to 500 μM TPL (Fig. 3k, P -value = 2.2×10^{-16} and 1.5×10^{-5} , respectively, two-sample Kolmogorov-Smirnov test).

To further test whether de novo transcription was required for zygote divisions and early embryogenesis, we also cultured seeds with α -amanitin (AMA), which binds to RNAPII and prevents nucleotide incorporation and transcript translocation^{57,58}. In the presence of 100 $\mu\text{g}/\text{mL}$ (108.8 μM) AMA, 23/59 (39.0%) embryos divided normally, while 13/59 (22.0%) were arrested (Fig. 3g,j and Supplementary Video S9) and 23/59 (39.0%) were delayed (Fig. 3h,j and Supplementary Video S10) in their development (P -value = 1.2×10^{-9} compared to DMSO control, chi-squared test). The distribution of cell cycle duration was significantly greater compared to DMSO (Fig. 3k, P -value = 1.6×10^{-9} , two-sample Kolmogorov-Smirnov test) but similar to 250 μM TPL (P -value = 4.0×10^{-3} , two-sample Kolmogorov-Smirnov test). Full details of the total number of samples examined and the cell cycle durations are listed in Supplementary Table S1. Although samples were cultured and imaged during different time spans due to technical reasons, the varying end time points do not affect our conclusions. As demonstrated by our results, the 20-hour time span is sufficient to capture at least two rounds of cell division for embryos cultured under control conditions. Because zygotes cultured with transcriptional inhibitors exhibited no or delayed zygotic division when monitored over longer periods of time compared to zygotes cultured under control conditions, this indicated that the delayed or arrested division was not due to insufficient observation time. The arrested and delayed cell cycles in the presence of the TPL and AMA transcriptional inhibitors indicated that de novo transcription was essential for the onset of zygote division and early embryogenesis.

Discussion

Optimization of the ExM fluorescent immunostaining technique for seeds enabled the robust detection of RNAPII Ser2P and H3K36me3 in zygotes. Because RNAPII Ser2P and H3K36me3 are hallmarks of active transcription, this indicated that zygotes are transcriptionally active soon after fertilization and before the first division. Moreover, zygotes had arrested and delayed cell divisions when cultured in the presence of transcriptional inhibitors (i.e. TPL and AMA), which is consistent with ZGA being required for the first division. Our observations complement a recent transcriptome study demonstrating that ZGA occurs in zygotes and is required for elongation and division³⁸.

While the conventional whole-mount immunostaining protocol can result in even-stained samples, we found that the method produced variable and inconsistent results. To improve the robustness of histological detection in Arabidopsis ovules and seeds, we adapted ExM for plant tissues in this study. Constraints imposed by undigested cytoskeleton and cell wall components limited the ability to fully expand samples. We omitted the proteinase K treatment described in the original ExM protocol because we found that even a mild proteinase K treatment (i.e. 5 U/mL for 30 minutes) resulted in signal loss without improving expansion ratios or consistency. Nevertheless, we were able to produce expanded samples with the ExM method, which makes interior tissues accessible to antibodies and resulted in more consistent immunostaining. Sample expansion also helped reduce autofluorescence, which is common in plant tissues. Our ExM protocol therefore provides a more robust and efficient option for conducting whole-mount fluorescent immunostaining of eggs and zygotes contained in ovules and seeds, respectively.

The detection of low levels of RNAPII Ser2P and H3K36me3 in mature eggs indicated that there is low transcriptional activity before fertilization. After fertilization, signal corresponding to RNAPII Ser2P and H3K36me3 was high demonstrating that transcriptional activities clearly increased in the zygote before the first division. It was previously reported that endosperm, but not zygote nuclei were transcriptionally active²². However, our observations indicate that both endosperm and zygote nuclei are transcriptionally active. The missing RNAPII Ser2P signal in zygotes in a previous report²² may be due to the antibodies used (H5, ab24758; Abcam, discontinued) or the inconsistency of the conventional whole-mount protocol. Our new histological evidence is consistent with a recent transcriptome study³⁸ and indicates that ZGA occurs soon after fertilization in Arabidopsis.

As revealed by live-cell imaging, transcriptional inhibition resulted in delayed or arrested development and provided evidence that ZGA is required for initial zygotic division and early embryogenesis. The differ-

ent responses of zygotes to TPL and AMA were likely due to the incubation timing because TPL and AMA are known to be fast-response and slow-response inhibitors, respectively⁵⁹. The 250 μ M TPL treatment, which is at least 100-fold greater than the recommended concentration for tissue culture⁵⁹, showed more delayed than arrested zygote division. The requirement of high concentrations of transcriptional inhibitors to have a physiological effect may be due to the high transcriptional activities of zygotes, as well as the limited accessibility of transcriptional inhibitors to zygotes deeply-embedded within seeds. Consistent with our results using transcriptional inhibitors, previous RNAi-mediated knock-down of RNAPII resulted in delayed embryogenesis²². Although we cannot completely exclude that the transcriptional inhibition of ovule tissues supporting the developing embryo (e.g. integument and endosperm) are the primary cause of the embryo arrest we observed, it is clear that inherited parental transcripts are not sufficient and de novo transcripts are required for early embryogenesis regardless of their origin. Moreover, *in vitro* fertilized or isolated zygotes developed normally through early embryogenesis in the absence of surrounding maternal tissue in rice^{60,61}, maize⁶² and tobacco^{8,63}, and Arabidopsis embryos can develop to at least the globular stage in the absence of the endosperm^{44,64}. These reports are consistent with the delayed and arrested zygotic division observed upon culturing with transcriptional inhibitors being primarily caused by the inability to transcribe genes in zygotes.

Because it is difficult to differentiate de novo transcribed from maternally inherited transcripts, the activation of the zygotic genome can be inferred from the transcriptional activities of paternal alleles. Several studies reported that early Arabidopsis embryos rely on maternal factors with little or no paternal activity^{23,24,65-69} and suggest a quiescent state in zygotes²². In contrast, other studies reported paternal allele expression in early embryos⁷⁰⁻⁷⁵. Additionally, mutants defective in the asymmetric division of the zygote were reported to be recessive, suggesting that both parental alleles are transcriptionally active after fertilization^{28-35,76,77}. Moreover, a recent genome-wide study reported significant upregulation of 4,436 genes in zygotes compared to eggs³⁸. Altogether these results indicate that genes involved in early embryogenesis are transcriptionally active in zygotes, and thus do not support gradual ZGA or reliance on maternal gene products during early Arabidopsis embryogenesis as previously proposed^{26,27}.

Genetic, microscopic and genomic studies in additional flowering plants, including maize^{16-20,78}, wheat^{14,15}, tobacco^{7,8,79} and rice⁹⁻¹³, indicate that ZGA occurs soon after fertilization in zygotes of multiple flowering plant species. In plants, zygotes mark the transition from the haploid gametophytic to diploid sporophytic phase of the life cycle. Because there is no clear evidence of prolonged transcriptional quiescence after fertilization and parentally-inherited gene products are not sufficient for early embryogenesis, the transcriptome remodeling observed in plant zygotes during the gametophytic-to-sporophytic transition is fundamentally different than the maternal-to-zygotic transition in early animal embryos as previously proposed⁸⁰. In further contrast to animals^{2,3}, the apparent similarities in the timing of transcriptome remodeling across plant species is intriguing and may indicate similar underlying mechanisms shared among flowering plants. For example, maternally and paternally inherited gene products converge to rapidly activate *WUS* *HOMEBOX 8* (*WOX8*) gene expression in Arabidopsis zygotes^{75,76,80,81}, and additional mechanisms integrating biparentally inherited information may exist in Arabidopsis and other plants. With published egg and zygote transcriptomes in multiple plants, as well as the increasing number of genome-wide approaches applicable to low amounts of input material, the field is poised to identify the key regulatory genes involved in transcriptome remodeling and initiation of embryogenesis.

Materials and Methods

plant materials and growth conditions. *Arabidopsis thaliana* accession Columbia (Col-0) and pWOX2::H2B-GFP, pWOX2::tdTomato-RCI2b transgenic Col-0 plants were grown at 20–22 °C temperature and 16 h light/8 h dark cycles under incandescent lights (130–150 μ mol/m²/s) in a climate-controlled growth chamber.

Sample collection and staging. For egg-containing ovule collection, we carefully emasculated flower buds from stage 11–12⁸² under a dissection microscope and then kept in growth chamber for 24 hours. The emasculated pistils were then examined under a dissection microscope and only the pistils with no pollen on stigmatic papillae were collected for experiments. For zygote-containing seed collection, we selected self-pollinated siliques from stage 14–15⁸² and dissected the siliques as described in the following sections. The detailed staging were determined by morphology after image acquisition according to previous reports^{44,46,75,83}.

cell wall digestion enzymes. We tested several cell wall digestion enzymes from multiple manufacturers and found the performance of the enzymes varied between manufacturers, as well as between batches from the same manufacturers. We tested driselase (Sigma), cellulase (Sigma), cellulase R10 (Duchefa, Yakult), cellulase RS (Duchefa, Yakult), pectolyase (Duchefa; discontinued), pectinase (Sigma), macerozyme R10 (Duchefa) and hemicellulase (Sigma) for conventional whole-mount protocol. We chose cellulase RS (Duchefa), hemicellulase (Sigma) and pectinase (Sigma) or macerozyme R10 (Duchefa) for expansion microscopy based on their performance and availability.

Antibodies. We used primary antibodies detecting RNAPII Ser2P (ab5095, Abcam), H3K36me3 (ab9050, Abcam), tubulin (ab89984, Abcam), and secondary antibodies anti-rabbit-Alexa488 (ab150077, Abcam) and anti-chicken-Alexa555 (ab150170, Abcam) because these antibodies are commercially available, commonly

used to detect corresponding epitopes^{84–89} and produced minimal non-specific signals as shown in Supplementary Video S11.

Conventional whole-mount fluorescent immunostaining. Conventional whole-mount fluorescent immunostaining was performed according to a published protocol⁴¹. Seeds of self-fertilized Col-0 siliques at stages 14–15⁸² were isolated under a dissection scope. Isolated seeds were collected in 4% PFA, 0.1% Triton X-100 and 1× PBS solution. Seeds were then briefly vacuum infiltrated and incubated at room temperature for one hour. Fixed seeds were washed three times with 0.1% Triton X-100 and 1× PBS before incubation with enzyme mix (1% driselase, 0.5% cellulase, and 1% pectolyase in water). Alternatively, seeds were washed once more with protoplast salt solution (20 mM MES, pH 5.0, 0.4 M mannitol, 20 mM KCl and 10 mM CaCl₂) before incubation with protoplast enzyme solution (3% cellulase, 1% hemicellulase, 1% macerozyme, 20 mM MES, pH 5.0, 0.4 M mannitol, 20 mM KCl and 10 mM CaCl₂, 0.1% BSA and 1% β-mercaptoethanol) before use. Enzyme solution was prepared as previously described⁹⁰. Seeds were incubated in either enzyme solution at 37 °C for 2 hours with gentle agitation. Digested seeds were washed twice with 0.2% Triton X-100, 1× PBS and embedded in 3% polyacrylamide matrix on adhesive slides as described⁴¹. Slides were incubated with 1% Triton X-100, 1× PBS at 4 °C for two hours with gentle agitation. Permeabilized samples were incubated with 1× PBS, 2% BSA, 0.1% Triton X-100 at room temperature for one hour. Samples were then incubated with 1× PBS, 2% BSA, 0.1% Triton X-100 and 1:500 dilution of primary antibodies against RNAPII Ser2P (ab5095, Abcam) and tubulin (ab89984, Abcam) at 4 °C overnight with gentle agitation. On the next day, samples were washed with 0.2% Triton X-100, 1× PBS at 4 °C for one hour at least five times. For secondary antibody incubation, samples were incubated in 1:500 dilution of anti-rabbit-Alexa488 (ab150077, Abcam) and anti-chicken-Alexa555 (ab150170, Abcam) in 1× PBS, 2% BSA, 0.1% Triton X-100 at 4 °C overnight with gentle agitation. Samples were then washed with 10 µg/ mL 4',6-diamidino-2-phenylindole (DAPI), 0.2% Triton X-100, 1× PBS at 4 °C in the dark for one hour twice and washed three times without DAPI. Samples were mounted in Vectashield Mounting Medium (H-1200, Vector) and imaged by ZEISS LSM700/780 with 25× oil objective at maximal resolution (>1024 × 1024) as 8-bit images at 13 z-stacks with 1 µm intervals by ZEN software. DAPI signals were excited by 405 nm laser and passed through SP490 filters. Alexa488 signals were excited by 488 nm laser and passed through BP490-635 filters. Alexa555 signals were excited by 555 nm laser and passed through 560–1000 nm filters. Pinhole sizes were kept as 1 airy unit for each color, and color channels were scanned separately to avoid false signal.

expansion microscopy. We modified published ExM protocols for plant tissues^{42,43}. For each condition, the immunostainings were conducted at least three times and are considered technical replicates. For each technical replicate ≥10 ovules or seeds from 1-2 flowers on ≥4 individual plants were recorded as biological replicates. Pistils containing unfertilized ovules or siliques containing fertilized seeds were carefully sliced open longitudinally under a dissection scope and transferred to 1× PBS, 0.1% Triton X-100, 4% PFA and 0.1 mg/ mL Acryloyl-X, SE (6-((acryloyl)amino)hexanoic acid, succinimidyl ester; Thermo Fisher). After brief vacuum infiltration, samples were incubated at 4 °C overnight as recommended by a previous report⁴³. Samples were washed with water twice and then washed once more with protoplast salt solution (20 mM MES, pH 5.0, 0.4 M mannitol, 20 mM KCl and 10 mM CaCl₂) before incubation with protoplast enzyme solution (3% cellulase, 1% hemicellulase, 1% macerozyme, 20 mM MES, pH 5.0, 0.4 M mannitol, 20 mM KCl and 10 mM CaCl₂, 0.1% BSA and 1% β-mercaptoethanol) before use⁹⁰. Samples were incubated at 37 °C for 2–3 hours (depending on developmental stage) with gentle agitation. Cell wall-digested samples were washed twice with 0.2% Triton X-100, 1× PBS and then permeabilized with 1% Triton X-100, 1× PBS at 4 °C for two hours. Pistils or siliques were then carefully dissected on depression slides under a dissection scope. Ovules/seeds attached to septums (i.e. ovule/ seed strings) were detached from pistils/siliques and transferred to 0.2% Triton X-100, 1× PBS. Isolated ovule/ seed strings were drained briefly and incubated in monomer solution (1× PBS, 2 M NaCl, 8.625% (w/w) sodium acrylate, 2.5% (w/w) acrylamide, 0.15% (w/w) N,N'-methylenebisacrylamide) at 4 °C overnight. Samples were then polymerized with 0.2% ammonium persulfate (APS) and tetramethylethylenediamine (TEMED) on glass slides with an adequate spacer (200–300 µm) at room temperature for one hour. Proteinase K treatment was omitted because it resulted in massive loss of epitopes. Gel slices containing ovule/seed strings were removed with razor blades and every two slices were transferred to 1 mL of water in 2 mL microtubes for expansion for one hour. The expansion was repeated at least twice. Expanded samples were then immunostained as described above for conventional immunostaining. For image acquisition, stained samples were mounted in 0.2% Triton X-100, 1× PBS with Frame-SealTM (BIO-RAD, SLF0601) and imaged by ZEISS LSM700/780 with 25× oil objective at maximal resolution (>1024 × 1024) as 8-bit images at 13 z-stacks with 1 µm intervals by ZEN software. DAPI signals were excited by 405 nm laser and passed through SP490 filters. Alexa488 signals were excited by 488 nm laser and passed through BP490-635 filters. Alexa555 signals were excited by 555 nm laser and passed through 560–1000 nm filters. Pinhole sizes were kept as 1 airy unit for each color, and color channels were scanned separately to avoid false signal.

Live-cell imaging. The ovule culture was performed as previously reported with slight modifications^{44,46}. For each condition we performed ≥ 3 technical replicates where we collected ≥100 self-pollinated seeds at stage 14–15⁸² from four individual plants. Seeds were carefully isolated from siliques under a dissection scope and transferred to N5T medium (5% trehalose dihydrate (Sigma), 1× Nitsch basal salt mixture (Duchefa), 1×

Gamborg's vitamin solution (Sigma), 0.05% MES-KOH, pH 5.8). A brief vacuum infiltration was applied to seeds followed by a 30-minute incubation in N5T medium or medium supplemented with α -amanitin (Sigma), flavopiridol (Sigma), triptolide (Cayman Chemical), or dimethyl sulfoxide (Fisher Scientific) at room temperature with gentle agitation to submerge seeds completely. The seeds were then transferred to micro-Insert 4 Well in μ -Dish (ibidi) or μ -Slide 8 Well (ibidi) with corresponding medium for live-cell imaging. Time series images were recorded at 10 z-stacks with 1 μ m intervals by Yokogawa CSU X1 spinning disc and Axio Observer (ZEISS) with 40 \times oil objective, Hamamatsu EMCCD 9100-13 camera and ZEN software, or Visioscope Spinning Disc Confocal (Visitron Systems GmbH) with 20 \times air objective, Andor Ixon Ultra 888 EMCCD camera and Visiview software. GFP was excited by 488 nm laser and emitting signals were passed through 525/550 nm filters. tdTomato was excited by 561 nm laser and resulting signals were passed through 605/670 nm filters. The first images were taken one hour after incubation and the time-lapsed images were taken every 30 to 60 minutes for at least 20 hours. **image processing.** The eggs and zygotes were distinguished according to their morphology as previously described^{44,46,75,83}. More specifically, eggs had an average length of 20–25 μ m and nuclei polarized to the chalazal end, while the zygotes elongated to approximately 70 μ m before dividing and the nuclei were apically localized approximately one-third from the chalazal end during elongation. We considered each egg or zygote as a biological replicate. All confocal microscope images were slightly adjusted for brightness by ZEN software to nearly saturate the signals in integument nuclei for qualitative comparison before exporting and scale bars were added by Fiji according to recorded pixel sizes. Spinning disc microscope images were processed by Fiji. For each timepoint the z-stacks were merged by maximum intensity projection, and then the contrast and brightness were adjusted for each channel. This procedure was necessary because the raw signals were kept to minimal levels to avoid phototoxicity. Color channels were then merged and images were cropped before adding scale bars and time stamps. For live-cell imaging experiments, the zygote/embryo phenotypes were determined according to the following criteria. "Normal" embryos had no morphological defects and <12 hours between cell divisions consistent with previous observations⁴⁴. "Dead" zygotes/embryos exhibited cell shrinkage or loss of fluorescent signals indicative of cell death. Otherwise, we considered the zygotes/embryos as "arrested" or "delayed" if there was either no cell division for >20 hours or the duration of the cell-cycle was 12–20 hours, respectively. When only one cell division was observed within the recorded videos, zygotes/embryos were conservatively estimated to be delayed "delayed (est.)" instead of "arrested" and the recorded cell-cycle durations were underestimated. To minimize potential bias, all movies were examined and classified by a person who did not perform the experiment and had no knowledge about their identities.

Statistics. The numbers of each phenotype for each condition (Supplementary Table S1) were used to perform chi-squared tests to determine if there was a significant difference between the frequencies observed in test groups (FLP100, TPL500, TPL250 or AMA100) compared to control groups (N5T or DMSO). The recorded cell cycle durations (Fig. 3k) were used to perform two-sample one-tailed Kolmogorov-Smirnov tests to determine if the distribution of cell cycle duration in a test condition was greater than or similar to a reference condition. R version 3.6.1 was used to perform all statistical tests with default settings and to generate charts presented in Fig. 3.

Received: 8 July 2019; Accepted: 4 November 2019;

Published online: 20 November 2019

References

1. Baroux, C., Autran, D., Gillmor, C. S., Grimanelli, D. & Grossniklaus, U. The maternal to zygotic transition in animals and plants. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 89–100 (2008).
2. Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annu. Rev. Cell Dev. Biol.* **30**, 581–613 (2014).
3. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042 (2009).
4. Walsler, C. B. & Lipshitz, H. D. Transcript clearance during the maternal-to-zygotic transition. *Curr. Opin. Genet. Dev.* **21**, 431–443 (2011).
5. Niedojadło, K., Pięciński, S., Smoliński, D. J. & Bednarska-Kozakiewicz, E. Transcriptional activity of *Hyacinthus orientalis* L. female gametophyte cells before and after fertilization. *Planta* **236**, 153–169 (2012).
6. Pięciński, S., Smoliński, D. J., Zienkiewicz, K. & Bednarska, E. Changes in poly(A) RNA and TMG snRNA distribution in the embryo sac of *Hyacinthus orientalis* L. before and after fertilization. *Sex. Plant Reprod.* **21**, 247–257 (2008).
7. Ning, J. *et al.* Differential gene expression in egg cells and zygotes suggests that the transcriptome is restructured before the first zygotic division in tobacco. *FEBS Lett.* **580**, 1747–1752 (2006).
8. Zhao, J. *et al.* Dynamic changes of transcript profiles after fertilization are associated with *de novo* transcription and maternal elimination in tobacco zygote, and mark the onset of the maternal-to-zygotic transition. *Plant J.* **65**, 131–145 (2011).
9. Abiko, M., Maeda, H., Tamura, K., Hara-Nishimura, I. & Okamoto, T. Gene expression profiles in rice gametes and zygotes: identification of gamete-enriched genes and up- or down-regulated genes in zygotes after fertilization. *J. Exp. Bot.* **64**, 1927–1940 (2013).
10. Anderson, S. N. *et al.* Transcriptomes of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *Plant J.* **76**, 729–741 (2013).
11. Anderson, S. N. *et al.* The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with Asymmetric Activation of Parental Genomes. *Dev. Cell* **43**, 349–358.e4 (2017).
12. Ohnishi, Y., Hoshino, R. & Okamoto, T. Dynamics of Male and Female Chromatin during Karyogamy in Rice Zygotes. *Plant Physiol.* **165**, 1533–1543 (2014).
13. Ohnishi, Y., Kokubu, I., Kinoshita, T. & Okamoto, T. Sperm Entry into the Egg Cell Induces the Progression of Karyogamy in Rice Zygotes. *Plant Cell Physiol.*, <https://doi.org/10.1093/pcp/pcz077> (2019).

14. Domoki, M. *et al.* Identification of genes preferentially expressed in wheat egg cells and zygotes. *Plant Cell Rep.* **32**, 339–348 (2013).
15. Sprunck, S., Baumann, U., Edwards, K., Langridge, P. & Dresselhaus, T. The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). *Plant J.* **41**, 660–672 (2005).
16. Chen, J. *et al.* Zygotic Genome Activation Occurs Shortly after Fertilization in Maize. *Plant Cell* **29**, 2106–2125 (2017).
17. Dresselhaus, T. *et al.* Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. *MGG - Molecular and General Genetics* **261**, 416 (1999).
18. Meyer, S. & Scholten, S. Equivalent Parental Contribution to Early Plant Zygotic Development. *Curr. Biol.* **17**, 1686–1691 (2007).
19. Okamoto, T., Scholten, S., Lörz, H. & Kranz, E. Identification of Genes that are Up- or Down-regulated in the Apical or Basal Cell of Maize Two-celled Embryos and Monitoring their Expression During Zygote Development by a Cell Manipulation- and PCR-based Approach. *Plant Cell Physiol.* **46**, 332–338 (2005).
20. Sauter, M., von Wiegen, P., Lörz, H. & Kranz, E. Cell cycle regulatory genes from maize are differentially controlled during fertilization and first embryonic cell division. *Sex. Plant Reprod.* **11**, 41–48 (1998).
21. Walbot, V. & Evans, M. M. S. Unique features of the plant life cycle and their consequences. *Nat. Rev. Genet.* **4**, 369–379 (2003).
22. Pillot, M. *et al.* Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis. *Plant Cell* **22**, 307–320 (2010).
23. Del Toro-De León, G., García-Aguilar, M. & Gillmor, C. S. Non-equivalent contributions of maternal and paternal genomes to early plant embryogenesis. *Nature* **514**, 624–627 (2014).
24. Vielle-Calzada, J.-P., Baskar, R. & Grossniklaus, U. Delayed activation of the paternal genome during seed development. *Nature* **404**, 91–94 (2000).
25. Autran, D. *et al.* Maternal Epigenetic Pathways Control Parental Contributions to Arabidopsis Early Embryogenesis. *Cell* **145**, 707–719 (2011).
26. García-Aguilar, M. & Gillmor, C. S. Zygotic genome activation and imprinting: parent-of-origin gene regulation in plant embryogenesis. *Curr. Opin. Plant Biol.* **27**, 29–35 (2015).
27. Armenta-Medina, A. & Gillmor, C. S. Genetic, molecular and parent-of-origin regulation of early embryogenesis in flowering plants. *Curr. Top. Dev. Biol.* **131**, 497–543 (2019).
28. Yu, T.-Y. *et al.* The Arabidopsis Receptor Kinase ZAR1 Is Required for Zygote Asymmetric Division and Its Daughter Cell Fate. *PLoS Genet.* **12**, e1005933 (2016).
29. Guo, L. *et al.* The anaphase-promoting complex initiates zygote division in Arabidopsis through degradation of cyclin B1. *Plant J.* **86**, 161–174 (2016).
30. Xu, J. *et al.* Embryonic Factor 1 encodes an AMP deaminase and is essential for the zygote to embryo transition in Arabidopsis: FAC1 in zygote to embryo transition. *Plant J.* **42**, 743–758 (2005).
31. Ronceret, A. *et al.* The first zygotic division in Arabidopsis requires de novo transcription of thymidylate kinase. *Plant J.* **53**, 776–789 (2008).
32. Ronceret, A., Gadea-Vacas, J., Guillemot, J. & Devic, M. The alpha-N-acetyl-glucosaminidase gene is transcriptionally activated in male and female gametes prior to fertilization and is essential for seed development in Arabidopsis. *J. Exp. Bot.* **59**, 3649–3659 (2008).
33. Ronceret, A. *et al.* Genetic analysis of two Arabidopsis DNA polymerase epsilon subunits during early embryogenesis. *Plant J.* **44**, 223–236 (2005).
34. Lin, Z. *et al.* AtCDC5 regulates the G2 to M transition of the cell cycle and is critical for the function of Arabidopsis shoot apical meristem. *Cell Res.* **17**, 815–828 (2007).
35. Liu, C.-M. & Meinke, D. W. The titan mutants of Arabidopsis are disrupted in mitosis and cell cycle control during seed development. *Plant J.* **16**, 21–31 (1998).
36. Nodine, M. D. & Bartel, D. P. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**, 94–97 (2012).
37. Schon, M. A. & Nodine, M. D. Widespread Contamination of Arabidopsis Embryo and Endosperm Transcriptome Data Sets. *Plant Cell* **29**, 608–617 (2017).
38. Zhao, P. *et al.* Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions. *Dev. Cell* **49**, 882–893.e5 (2019).
39. Hajheidari, M., Koncz, C. & Eick, D. Emerging roles for RNA polymerase II CTD in Arabidopsis. *Trends Plant Sci.* **18**, 633–643 (2013).
40. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
41. García-Aguilar, M. & Autran, D. Localization of Chromatin Marks in Arabidopsis Early Embryos. In 419–441 (Humana Press, New York, NY, 2018).
42. Chen, F., Tillberg, P. W. & Boyden, E. S. Optical imaging. *Expansion microscopy. Science* **347**, 543–548 (2015).
43. Tillberg, P. W. *et al.* Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. *Nat. Biotechnol.* **34**, 987–992 (2016).
44. Gooh, K. *et al.* Live-cell imaging and optical manipulation of Arabidopsis early embryogenesis. *Dev. Cell* **34**, 242–251 (2015).
45. Webb, M. C. & Gunning, B. E. The microtubular cytoskeleton during development of the zygote, proembryo and free-nuclear endosperm in Arabidopsis thaliana (L.) Heynh. *Planta* **184**, 187–195 (1991).
46. Kimata, Y. *et al.* Cytoskeleton dynamics control the first asymmetric cell division in Arabidopsis zygote. *Proc. Natl. Acad. Sci. USA* **113**, 14157–14162 (2016).
47. Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126 (2012).
48. Braude, P., Pelham, H., Flach, G. & Lobatto, R. Post-transcriptional control in the early mouse embryo. *Nature* **282**, 102–105 (1979).
49. Edgar, B. A. & Datar, S. A. Zygotic degradation of two maternal Cdc25 mRNAs terminates Drosophila's early cell cycle program. *Genes Dev.* **10**, 1966–1977 (1996).
50. Edgar, L. G., Wolf, N. & Wood, W. B. Early transcription in *Caenorhabditis elegans* embryos. *Development* **120**, 443–451 (1994).
51. Newport, J. & Kirschner, M. A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675–686 (1982).

52. Zamir, E., Kam, Z. & Yarden, A. Transcription-dependent induction of G1 phase during the zebra fish midblastula transition. *Mol. Cell Biol.* **17**, 529–536 (1997).
53. Dai, Y. & Grant, S. Cyclin-dependent kinase inhibitors. *Curr. Opin. Pharmacol.* **3**, 362–370 (2003).
54. Kimata, Y. *et al.* Polar vacuolar distribution is essential for accurate asymmetric division of *Arabidopsis* zygotes. *Proc. Natl. Acad. Sci. USA* **116**, 2338–2343 (2019).
55. Titov, D. V. *et al.* XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* **7**, 182–188 (2011).
56. Vispé, S. *et al.* Triptolide is an inhibitor of RNA polymerase I and II-dependent transcription leading predominantly to down-regulation of short-lived mRNA. *Mol. Cancer Ther.* **8**, 2780–2790 (2009).
57. Kaplan, C. D., Larsson, K.-M. & Kornberg, R. D. The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol. Cell* **30**, 547–556 (2008).
58. Brueckner, F. & Cramer, P. Structural basis of transcription inhibition by α -amanitin and implications for RNA polymerase II translocation. *Nat. Struct. Mol. Biol.* **15**, 811–818 (2008).
59. Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2**, 103–108 (2011).
60. Sato, A., Toyooka, K. & Okamoto, T. Asymmetric cell division of rice zygotes located in embryo sac and produced by *in vitro* fertilization. *Sex. Plant Reprod.* **23**, 211–217 (2010).
61. Uchiumi, T., Uemura, I. & Okamoto, T. Establishment of an *in vitro* fertilization system in rice (*Oryza sativa* L.). *Planta* **226**, 581–589 (2007).
62. Kranz, E. & Lorz, H. *In Vitro* Fertilization with Isolated, Single Gametes Results in Zygotic Embryogenesis and Fertile Maize Plants. *Plant Cell* **5**, 739–746 (1993).
63. He, Y.-C., He, Y.-Q., Qu, L.-H., Sun, M.-X. & Yang, H.-Y. Tobacco zygotic embryogenesis *in vitro*: the original cell wall of the zygote is essential for maintenance of cell polarity, the apical-basal axis and typical suspensor formation. *Plant J.* **49**, 515–527 (2007).
64. Ngo, Q. A. *et al.* *Arabidopsis* GLAUCE promotes fertilization-independent endosperm development and expression of paternally inherited alleles. *Development* **134**, 4107–4117 (2007).
65. Golden, T. A. *et al.* Short integuments1/suspensor1/carpel factory, a Dicer homolog, is a maternal effect gene required for embryo development in *Arabidopsis*. *Plant Physiol.* **130**, 808–822 (2002).
66. Grossniklaus, U., Vielle-Calzada, J. P., Hoepfner, M. A. & Gagliano, W. B. Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* **280**, 446–450 (1998).
67. Luo, M., Bilodeau, P., Dennis, E. S., Peacock, W. J. & Chaudhury, A. Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing *Arabidopsis* seeds. *Proc. Natl. Acad. Sci. USA* **97**, 10637–10642 (2000).
68. Guitton, A.-E. & Berger, F. Loss of Function of Multicopy Suppressor of IRA 1 Produces Nonviable Parthenogenetic Embryos in *Arabidopsis*. *Curr. Biol.* **15**, 750–754 (2005).
69. Pagnussat, G. C. *et al.* Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development* **132**, 603–614 (2005).
70. Aw, S. J., Hamamura, Y., Chen, Z., Schnittger, A. & Berger, F. Sperm entry is sufficient to trigger division of the central cell but the paternal genome is required for endosperm development in *Arabidopsis*. *Development* **137**, 2683–2690 (2010).
71. Xiang, D. *et al.* Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. *Plant Physiol.* **156**, 346–356 (2011).
72. Köhler, C., Page, D. R., Gagliardini, V. & Grossniklaus, U. The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat. Genet.* **37**, 28–30 (2005).
73. Weijers, D., Geldner, N., Offringa, R. & Jürgens, G. Early paternal gene activity in *Arabidopsis*. *Nature* **414**, 709–710 (2001).
74. Baroux, C., Blanvillain, R. & Gallois, P. Paternally inherited transgenes are down-regulated but retain low activity during early embryogenesis in *Arabidopsis*. *FEBS Lett.* **509**, 11–16 (2001).
75. Ueda, M., Zhang, Z. & Laux, T. Transcriptional Activation of *Arabidopsis* Axis Patterning Genes WOX8/9 Links Zygote Polarity to Embryo Development. *Dev. Cell* **20**, 264–270 (2011).
76. Lukowitz, W., Roeder, A., Parmenter, D. & Somerville, C. A MAPKK Kinase Gene Regulates Extra-Embryonic Cell Fate in *Arabidopsis*. *Cell* **116**, 109–119 (2004).
77. Tzafir, I. *et al.* Identification of genes required for embryo development in *Arabidopsis*. *Plant Physiol.* **135**, 1206–1220 (2004).
78. Scholten, S., Lörz, H. & Kranz, E. Paternal mRNA and protein synthesis coincides with male chromatin decondensation in maize zygotes. *Plant J.* **32**, 221–231 (2002).
79. Xin, H.-P. *et al.* Expressed sequence-tag analysis of tobacco sperm cells reveals a unique transcriptional profile and selective persistence of paternal transcripts after fertilization. *Sex. Plant Reprod.* **24**, 37–46 (2011).
80. Ueda, M. *et al.* Transcriptional integration of paternal and maternal factors in the *Arabidopsis* zygote. *Genes Dev.* **31**, 617–627 (2017).
81. Bayer, M. *et al.* Paternal Control of Embryonic Patterning in *Arabidopsis thaliana*. *Science* **323**, 1485–1488 (2009).
82. Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early flower development in *Arabidopsis*. *Plant Cell* **2**, 755–767 (1990).
83. Mansfield, S. G. & Briarty, L. G. Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo. *Can. J. Bot.* **69**, 461–476 (1991).
84. Antosz, W. *et al.* The Composition of the *Arabidopsis* RNA Polymerase II Transcript Elongation Complex Reveals the Interplay between Elongation and mRNA Processing Factors. *Plant Cell* **29**, 854–870 (2017).
85. Chen, C. *et al.* RNA polymerase II-independent recruitment of SPT6L at transcription start sites in *Arabidopsis*. *Nucleic Acids Res.* **47**, 6714–6725 (2019).
86. Liu, P. *et al.* The Histone H3K4 Demethylase JM16 Represses Leaf Senescence in *Arabidopsis*. *Plant Cell* **31**, 430–443 (2019).
87. Zheng, S. *et al.* The *Arabidopsis* H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor. *Nat. Commun.* **10**, 1303 (2019).
88. Wei, L. *et al.* Super-multiplex vibrational imaging. *Nature* **544**, 465–470 (2017).
89. Capalbo, L. *et al.* Coordinated regulation of the ESCRT-III component CHMP4C by the chromosomal passenger complex and centralspindlin during cytokinesis. *Open Biol.* **6** (2016).
90. Yoo, S.-D., Cho, Y.-H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).

Acknowledgements

This work was supported by the Österreichische Akademie der Wissenschaften (Austrian Academy of Sciences). Special thanks to Daisuke Kurihara and Tetsuya Higashiyama for generously sharing pWOX2::H2B-GFP, pWOX2::tdTomato-RCI2b lines and their instructions on live-cell imaging. We are also grateful to Frederic Berger for insightful discussions and Michael Schon for his input on figure generation. Additionally, we acknowledge the IMP-IMBA-GMI BioOptics Core Facility and the Plant Sciences Facility at Vienna BioCenter Core Facilities GmbH (VBCF) for support. Ping Kao also personally thanks Kotoha Tanaka and Kaori Sakuramori for their support.

Author contributions

Conceptualization, P.K. and M.D.N.; Methodology, P.K.; Formal Analysis, P.K.; Investigation, P.K.; Writing – Original Draft, P.K.; Writing – Review & Editing, M.D.N. and P.K.; Validation, M.N.; Visualization, P.K.; Supervision, M.N. **competing interests**

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-53704-2>.

Correspondence and requests for materials should be addressed to M.D.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access
License,



This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Supplementary Information

Transcriptional Activation of Arabidopsis Zygotes Is Required for Initial Cell Divisions

Ping Kao¹ and Michael Nodine^{1,*}

¹Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3,
1030 Vienna, Austria

*Corresponding Author: michael.nodine@gmi.oeaw.ac.at

LEGENDS FOR SUPPORTING INFORMATION

Supplementary Figure S1. Additional panel of images for Figure 2.

Supplementary Table S1. Sample number and cell cycle duration for live-cell imaging.

Supplementary Video S1. Unadjusted z-stack series of fluorescent immunostaining on expanded zygote samples (Example 1). The scale bar represents 20 μm . Staining on nuclei, RNAPII Ser2P and tubulin are presented in cyan, yellow and red, respectively.

Supplementary Video S2. Unadjusted z-stack series of fluorescent immunostaining on expanded zygote samples (Example 2). The scale bar represents 20 μm . Staining on nuclei, RNAPII Ser2P and tubulin are presented in cyan, yellow and red, respectively.

Supplementary Video S3. Live-cell imaging of embryos cultured in N5T medium. The scale bar represents 50 μm .

Supplementary Video S4. Live-cell imaging of embryos cultured with 100 μM FLP in N5T medium. The scale bar represents 50 μm .

Supplementary Video S5. Live-cell imaging of embryos cultured with 0.5% DMSO in N5T medium. The scale bar represents 50 μm .

Supplementary Video S6. Live-cell imaging of embryos cultured with 500 μM TPL, 0.5% DMSO in N5T medium. The scale bar represents 50 μm .

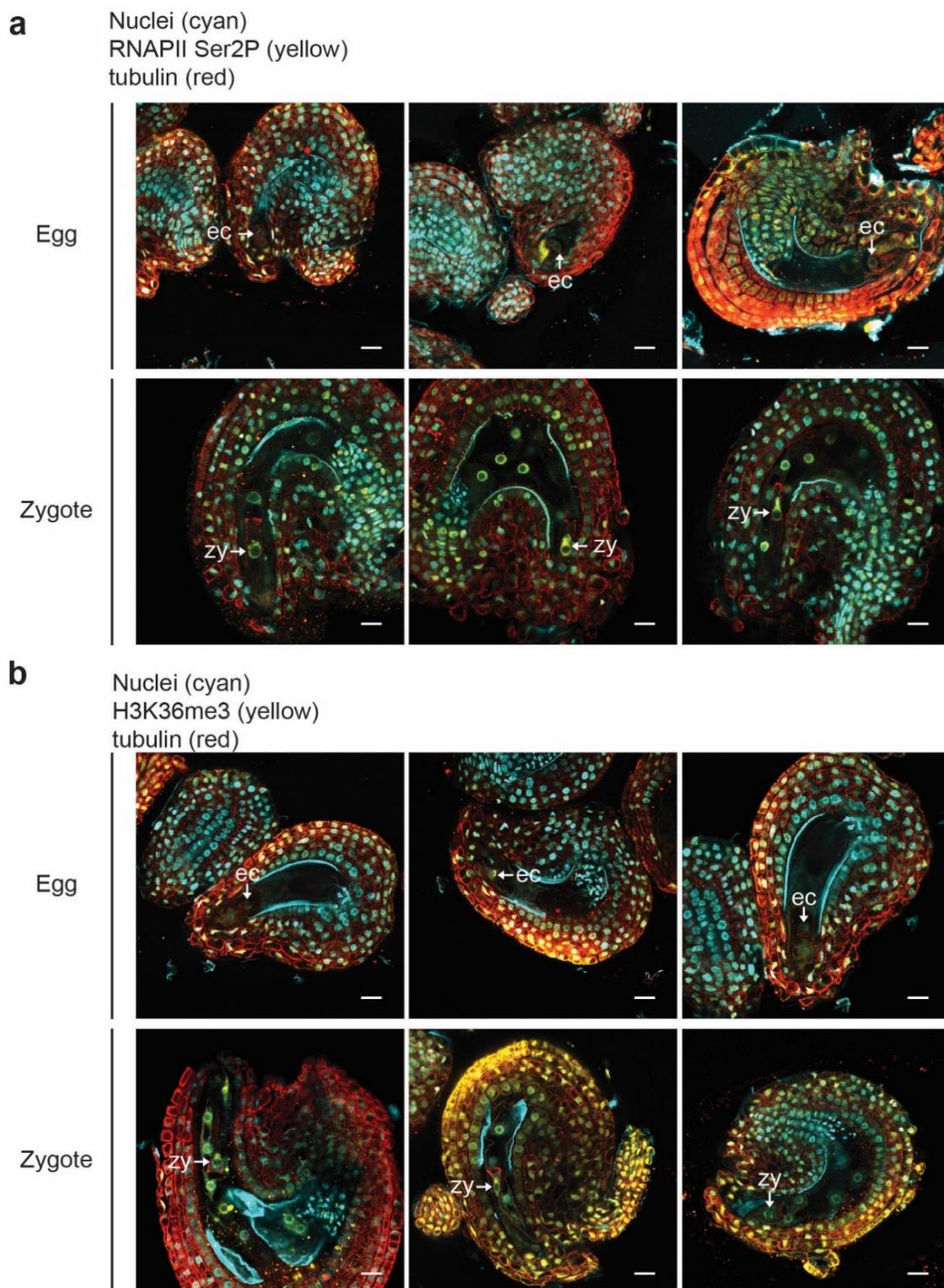
Supplementary Video S7. Live-cell imaging of embryos cultured with 250 μM AMA, 0.5% DMSO in N5T medium with arrested zygotes. The scale bar represents 50 μm .

Supplementary Video S8. Live-cell imaging of embryos cultured with 250 μM AMA, 0.5% DMSO in N5T medium with arrested zygotes. The scale bar represents 50 μm .

Supplementary Video S9. Live-cell imaging of embryos cultured with 100 $\mu\text{g/mL}$ AMA, 0.5% DMSO in N5T medium with arrested zygotes. The scale bar represents 50 μm .

Supplementary Video S10. Live-cell imaging of embryos cultured with 100 $\mu\text{g/mL}$ AMA, 0.5% DMSO in N5T medium with delayed division. The scale bar represents 50 μm .

Supplementary Video S11. Unadjusted z-stack series of fluorescent immunostaining without primary antibodies. The scale bar represents 20 μm . Staining on nuclei, anti-rabbit-Alexa488 and anti-chicken-Alexa555 are presented in cyan, yellow and red, respectively.



Supplementary Figure S1. Visualization of transcription activities in eggs and zygotes.

(a) Unadjusted expansion microscopy images showing one focal plane of staining against tubulin (red), RNAPII Ser2P (yellow) and DAPI-stained nuclei (cyan) in ovules containing eggs (top) or seeds containing zygotes (bottom). **(b)** Unadjusted expansion microscopy images showing one focal plane of staining against tubulin (red), H3K36me3 (yellow) and DAPI-stained nuclei (cyan) in ovules containing eggs (top) or seed containing zygotes (bottom). Scale bars represent 20 μm . ec, egg cell; zy, zygote.

7 Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis

Type	<i>published</i> Genome Biology volume 21, Article number: 251 (2020) doi: 10.1186/s13059-020-02163-4
Authors	Ranjith K. Papareddy, Katalin Páldi, Subramanian Paulraj, Ping Kao, Stefan Lutzmayer & Michael D. Nodine
Contributions	R.K.P. and M.D.N. conceived the project; R.K.P. developed the methodology, implemented software used, and performed formal analysis; R.K.P., S.P., K.P., P.K., S.L., and M.D.N. conducted the experiments; R.K.P. and M.D.N. wrote and edited the article; M.D.N. supervised the project and acquired funding. The authors read and approved the final manuscript.

RESEARCH

Open Access

Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis

Ranjith K. Papareddy, Katalin Páldi[†], Subramanian Paulraj[†], Ping Kao, Stefan Lutzmayer and Michael D. Nodine[†]

* Correspondence: michael.nodine@gmi.oeaw.ac.at [†]Katalin Páldi and Subramanian Paulraj contributed equally to this work. Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr.Bohr-Gasse 3, 1030 Vienna, Austria

Abstract

Background: Eukaryotic genomes are partitioned into euchromatic and heterochromatic domains to regulate gene expression and other fundamental cellular processes. However, chromatin is dynamic during growth and development and must be properly re-established after its decondensation. Small interfering RNAs (siRNAs) promote heterochromatin formation, but little is known about how chromatin regulates siRNA expression.

Results: We demonstrate that thousands of transposable elements (TEs) produce exceptionally high levels of siRNAs in Arabidopsis thaliana embryos. TEs generate siRNAs throughout embryogenesis according to two distinct patterns depending on whether they are located in euchromatic or heterochromatic regions of the genome. siRNA precursors are transcribed in embryos, and siRNAs are required to direct the re-establishment of DNA methylation on TEs from which they are derived in the new generation. Decondensed chromatin also permits the production of 24-nt siRNAs from heterochromatic TEs during post-embryogenesis, and siRNA production from bipartite-classified TEs is controlled by their chromatin states.

Conclusions: Decondensation of heterochromatin in response to developmental, and perhaps environmental, cues promotes the transcription and function of siRNAs in plants. Our results indicate that chromatin-mediated siRNA transcription provides a cell-autonomous homeostatic control mechanism to help reconstitute pre-existing chromatin states during growth and development including those that ensure silencing of TEs in the future germ line.

Keywords: Small RNAs, DNA methylation, Chromatin, Epigenetics, Linker histone H1, Plant embryogenesis, RNAi, Transposable elements

Background

Eukaryotic genomes are partitioned into euchromatic and heterochromatic domains [1, 2]. Euchromatic regions are enriched for genes and provide a transcriptionally permissive state. Heterochromatic regions are densely packed, or condensed, regions of the genome that are typically transcriptionally quiescent and characterized by highly



© The Author(s). 2020 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



repetitive DNA and transposable elements [3, 4]. Heterochromatin formation is promoted by various pathways including those affecting covalent modifications of histones, which package DNA into nucleosomes, as well as cytosine methylation [5–7]. Chromatin states are re-established after fertilization of egg and sperm in diverse animals [8–12], and histone reprogramming has also been observed in plants shortly after fertilization [13, 14]. However, little is known about how heterochromatin-promoting pathways respond to labile chromatin states shortly after fertilization to help reestablish euchromatic and heterochromatic states.

Small RNA-based pathways promote heterochromatin formation and associated transcriptional silencing in animals, fungi, and plants [15]. Plants employ 24-nucleotide (nt) small interfering RNAs (siRNAs) to help promote silencing of repetitive elements including TEs, and thus prevent DNA mutations caused by TE mobilization. During canonical RNA-directed DNA methylation (RdDM), RNA Polymerase IV (Pol IV) is recruited to target loci and generates transcripts that are co-transcriptionally converted into double-stranded RNAs by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) [16–21]. The resulting double-stranded RNAs are then processed into 23-nt/24-nt RNA duplexes by the DICER-LIKE 3 (DCL3) endoribonuclease [22]. The 24-nt strand of the duplex binds to ARGONAUTE 4 (AGO4) and guides AGO4 and associated proteins to siRNA-complementary sites contained within noncoding RNAs produced by RNA Polymerase V [23, 24]. DOMAINS REARRANGED METHYLTRANSFERASES 1/2 (DRM1/2) is then recruited to target loci, which results in the de novo methylation of cytosines in the CG, CHG, and CHH contexts (where H ≠ G) [25, 26]. Because CG and CHG methylation are maintained independently of siRNAs by DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3), respectively, CHH methylation is a good indicator of RdDM activities [20]. Nevertheless, CHH methylation can also be maintained independently of siRNAs by CMT2 in post-embryonic tissues, and this occurs mostly on the bodies of long TEs that are densely packed in nucleosomes and inaccessible to DRM2 [27].

Although sRNAs promote heterochromatin formation in diverse eukaryotic species, little is known about how chromatin states regulate small RNA production in animals and plants. Plant gametes and associated companion cells that support the gametes are contained within multicellular haploid gametophytes [28]. Current models propose that the large-scale chromatin decondensation observed in terminally differentiated companion cells facilitates TE transcription, and the resulting transcripts serve as substrates for RDR1/6 and DCLs to generate 20–22-nt siRNAs that move into gametes and stabilize TE silencing [29–33]. The endosperm and embryo are products of double-fertilization, and it has also been suggested that hypomethylation of endosperm promotes the production of siRNAs, which then move into the embryo to mediate TE methylation [31, 34]. RdDM is required for the progressive methylation of a few target loci during *Arabidopsis thaliana* (*Arabidopsis*) embryogenesis [35], and late-staged embryos are CHH hyper-methylated compared to other tissues [36–38]. However, the dynamics of embryonic RdDM have not been reported genome-wide in plants due to the difficulty in generating genome-wide profiles of siRNAs and methylomes from early embryos, which are small and deeply embedded within maternal seed tissues. More generally, it is virtually unknown how chromatin states may promote the de novo production of sRNAs to help re-establish TE silencing cell-autonomously including in the future germ-line.

Results

Embryos are enriched for transposon-derived small RNAs

We recently developed a low-input small RNA sequencing (sRNA-seq) method to profile small RNA dynamics during eight stages of *Arabidopsis* embryogenesis, as well as floral buds and leaves [39] (Fig. 1a and Additional file 1: Table S1). We previously

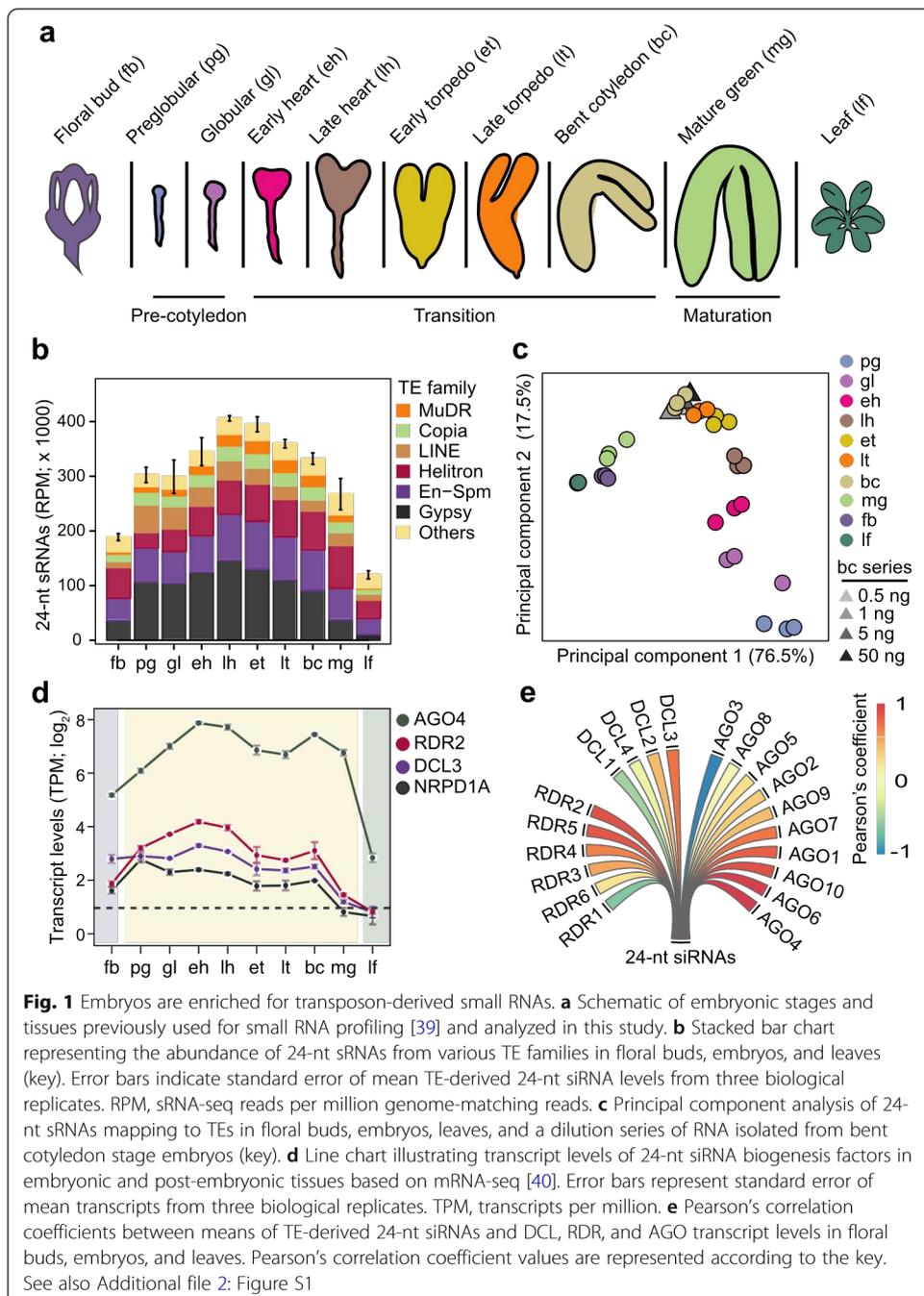


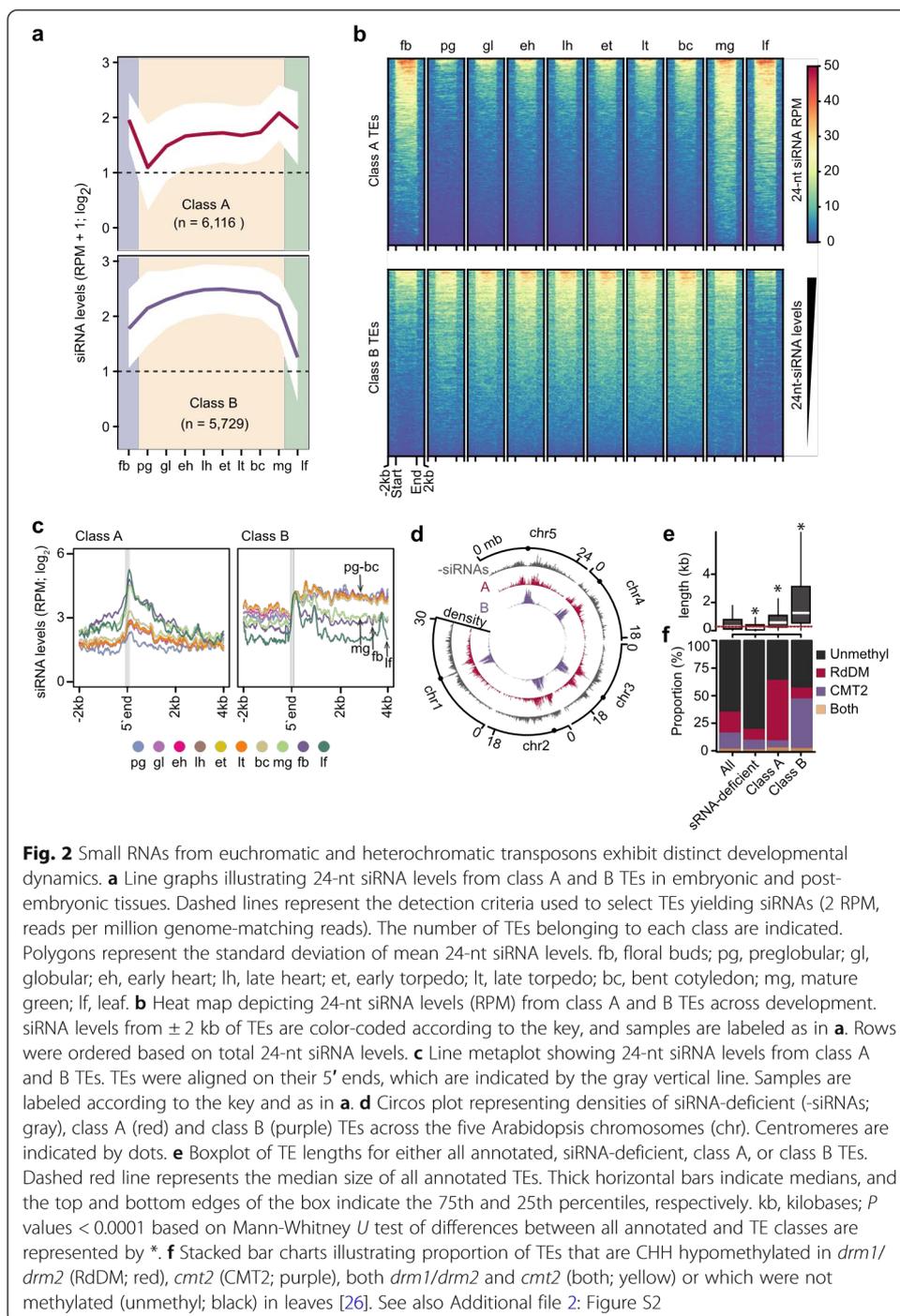
Fig. 1 Embryos are enriched for transposon-derived small RNAs. **a** Schematic of embryonic stages and tissues previously used for small RNA profiling [39] and analyzed in this study. **b** Stacked bar chart representing the abundance of 24-nt sRNAs from various TE families in floral buds, embryos, and leaves (key). Error bars indicate standard error of mean TE-derived 24-nt siRNA levels from three biological replicates. RPM, sRNA-seq reads per million genome-matching reads. **c** Principal component analysis of 24-nt sRNAs mapping to TEs in floral buds, embryos, leaves, and a dilution series of RNA isolated from bent cotyledon stage embryos (key). **d** Line chart illustrating transcript levels of 24-nt siRNA biogenesis factors in embryonic and post-embryonic tissues based on mRNA-seq [40]. Error bars represent standard error of mean transcripts from three biological replicates. TPM, transcripts per million. **e** Pearson's correlation coefficients between means of TE-derived 24-nt siRNAs and DCL, RDR, and AGO transcript levels in floral buds, embryos, and leaves. Pearson's correlation coefficient values are represented according to the key. See also Additional file 2: Figure S1

focused on the ~ 21-nt microRNA class of small RNAs involved in post-transcriptional regulation [39], but noticed that the vast majority of TE-derived small RNAs were 24nt long and highly enriched in embryos compared to floral bud or leaf tissues (Fig. 1b and Additional file 2: Figure S1A). Small RNAs were detected from 28,087 TEs and the highest amounts from several families including Gypsy, MuDR, and En-Spm were detected during mid-embryogenesis. The levels of TE-derived 24-nt sRNAs were highly correlated among biological replicates from floral bud, embryonic, and leaf tissues indicating stage- and tissue-specific sRNA populations (Fig. 1c and Additional file 2: Figure S1B). Moreover, principal component analysis revealed that 76.5% and 17.5% of the variation in TE-derived 24-nt siRNAs was accounted for by principal components 1 and 2, respectively. Principal component 1 distinctly separated post-embryonic and mature green embryo stages from pre-maturation embryonic stages, whereas principal component 2 stratified the pre-maturation embryonic samples according to developmental stage. Libraries prepared from 50, 5, 1, or 0.5 ng of RNA isolated from bent cotyledon embryos clustered together with the biological replicates generated from ≥ 500 ng of bent cotyledon RNA (Fig. 1c and Additional file 2: Figure S1B). This indicated that the vast majority of variation observed in 24-nt embryonic and post-embryonic siRNA populations was biological rather than technical.

Small interfering RNAs involved in RdDM are typically 24-nt long and begin with a 5' adenine [41]. Accordingly, adenosines were the dominant first base of 24-nt sRNAs in embryos (Additional file 2: Figure S1C), and the levels of embryonic 24-nt sRNAs were most highly correlated with the levels of transcripts encoding key canonical RdDM components such as RDR2, DCL3, and AGO4, which were also enriched in developing embryos (Fig. 1d, e) [20, 40]. Therefore, canonical 24-nt siRNAs are highly enriched in embryos, exhibit distinct developmental dynamics, and upon maturation become similar to post-embryonic siRNA populations.

Small RNAs from euchromatic and heterochromatic transposons exhibit distinct developmental dynamics

To examine the temporal dynamics of embryonic siRNAs mapping to TEs in more detail, we used mclust [42] to define eight unique clusters of the 31,189 TAIR10-annotated TEs based on their 24-nt siRNA levels (Additional file 2: Figure S2A, B). At least 2 sRNA-seq reads per million genome-matching reads were detected for 11,845 TEs, and these were grouped into either class A (6116 TEs; 19.6% of total) or class B (5729 TEs, 18.4% of total) based on the dynamics of their corresponding siRNAs between floral buds, developing embryos, and leaves (Fig. 2a, b, Additional file 3: Table S3, and Additional file 2: Figure S2B). The remaining 19,344 TEs (62.0% of total) were considered siRNA-depleted (Fig. 2a and Additional file 2: Figure S2B, C) and served as negative controls. Small interfering RNAs from class A TEs had low levels in preglobular embryos that were increased at the globular stage, and were stable until sharply increasing at the mature green stage and then remained at high levels in leaves and floral buds (Fig. 2a–c). In contrast, class B TEs produced large amounts of siRNAs already in preglobular embryos, then continued to gradually increase through mid-embryogenesis and were strongly reduced in mature embryos, leaves, and floral buds (Fig. 2a–c).



Class A and B TEs have distinct features. Class A TEs are short and dispersed along pericentromeric and euchromatic regions of chromosomes (Fig. 2d, e and Additional file 2: Figure S2D). In contrast, class B TEs are generally longer, concentrated in heterochromatic centromeres, and especially in embryos, siRNAs are generated from throughout whole TEs (Fig. 2c–e and Additional file 2: Figure S2D). Transposons targeted by DRM2 and CMT2 are distinct to euchromatic and heterochromatic domains of the genome, respectively [27]. The sizes and genomic locations of class A and B TEs are characteristic of TEs respectively methylated by either the RdDM or CMT2 pathways in post-embryonic tissues [26, 27, 43]. Indeed, class A TEs were enriched for TEs with reduced methylation in RdDM-defective *drm1/2* mutant leaves including HAT, SINE, SADHU, and other short TE families, whereas class B TEs were enriched for TEs with reduced methylation in *cmt2* mutant leaves including MuDR, En-Spm, and long terminal repeat families such as Gypsy and Copia (Fig. 2f and Additional file 2: Figure S2E, F). In addition, class B TEs have heterochromatic fea-

tures such as high levels of GC content, nucleosome occupancy, HISTONE 3 LYSINE 9 dimethylation (H3K9me2), and linker histone 1 (H1) compared to class A TEs (Additional file 2: Figure S2G). Altogether, we identified two distinct classes of TEs based on the levels of siRNAs they produce during development: euchromatic TEs (i.e., class A) that progressively generate siRNAs during embryogenesis and are methylated by the siRNA-dependent RdDM pathway in post-embryonic tissues, and heterochromatic TEs (i.e., class B) that produce very large amounts of siRNAs during embryogenesis prior to maturation and are methylated independently of siRNAs in post-embryonic tissues.

Embryonic methylome dynamics

During RdDM, 24-nt siRNAs are loaded onto ARGONAUTE proteins and serve as sequence-specific guides for the recruitment of methyltransferases to target loci. siRNA-directed methylation of TEs contributes to their transcriptional silencing and immobilization and limits their mutagenic potential [44–47]. To investigate the functions of embryonic 24-nt siRNAs, we adapted a whole-genome bisulfite sequencing approach called methylC-seq [48] to profile methylomes at single-base resolution from the low amounts of DNA available from early Arabidopsis embryos (see the “Methods” section) (Additional file 1: Table S1). Comparisons of methylomes generated with 0.1, 0.5, 1, or ~4 ng of genomic DNA isolated from bent cotyledon embryos had nearly identical cytosine methylation levels indicating that there was low variability of this method when using different amounts of input DNA (Additional file 2: Figure S3A). Therefore, we used this robust low-input methylC-seq method to profile methylomes from 8-cell/16-cell (preglobular; 3 days after pollination [DAP]), early heart (4 DAP) and bent-cotyledon (8 DAP) embryos, as well as leaves and floral buds. We compared these datasets with publicly available methylomes generated from sperm [32], or latestaged embryos from early torpedo [49], mid-torpedo to early maturation [31], or mature green [37] stages. Because methylation of cytosines in the CHH context (mCHH, where H ≠ G) is a hallmark of siRNA-directed DNA methylation [20], we focused on CHH methylation. More specifically, the Arabidopsis genome was divided into 50-kb bins, and the mean-weighted CHH methylation rates from reproductive, embryonic, and vegetative tissues were calculated for each bin (Additional file 2: Figure S3B). Consistent with previous studies, CHH methylation gradually increased during embryogenesis until peaking at the maturation stage and then decreased in leaves (Additional file 2: Figure S3B) [35, 37, 38]. As expected, CHH methylation was most prominent at pericentromeric and centromeric regions densely populated with euchromatic and heterochromatic TEs that were enriched for siRNAs during embryogenesis (Additional file 2: Figure S3B). Relative to sperm, euchromatic TEs had low CHH methylation levels in early embryos that increased during embryogenesis and were hypermethylated relative to leaves in 8 DAP bent-cotyledon-staged embryos (Fig. 3a). In contrast, CHH methylation was barely detectable from heterochromatic TEs in sperm, but then increased in early embryos and became hypermethylated relative to leaves by 6 DAP in early torpedo-staged embryos (Fig. 3b). Therefore, CHH methylation is established on both euchromatic and heterochromatic TEs during early embryogenesis, and TEs become hypermethylated relative to post-embryonic tissues at late stages of embryogenesis.

To examine embryonic methylation dynamics in more detail, we identified significant differentially methylated regions (DMRs) by pairwise comparisons between six embryonic stages (preglobular, early heart, early torpedo, bent cotyledon, late torpedo-toearly mature green and mature green) (see the “Methods” section). We found 21,361 embryonic CHH DMRs with a median size of approximately 100-bp (Additional file 4:

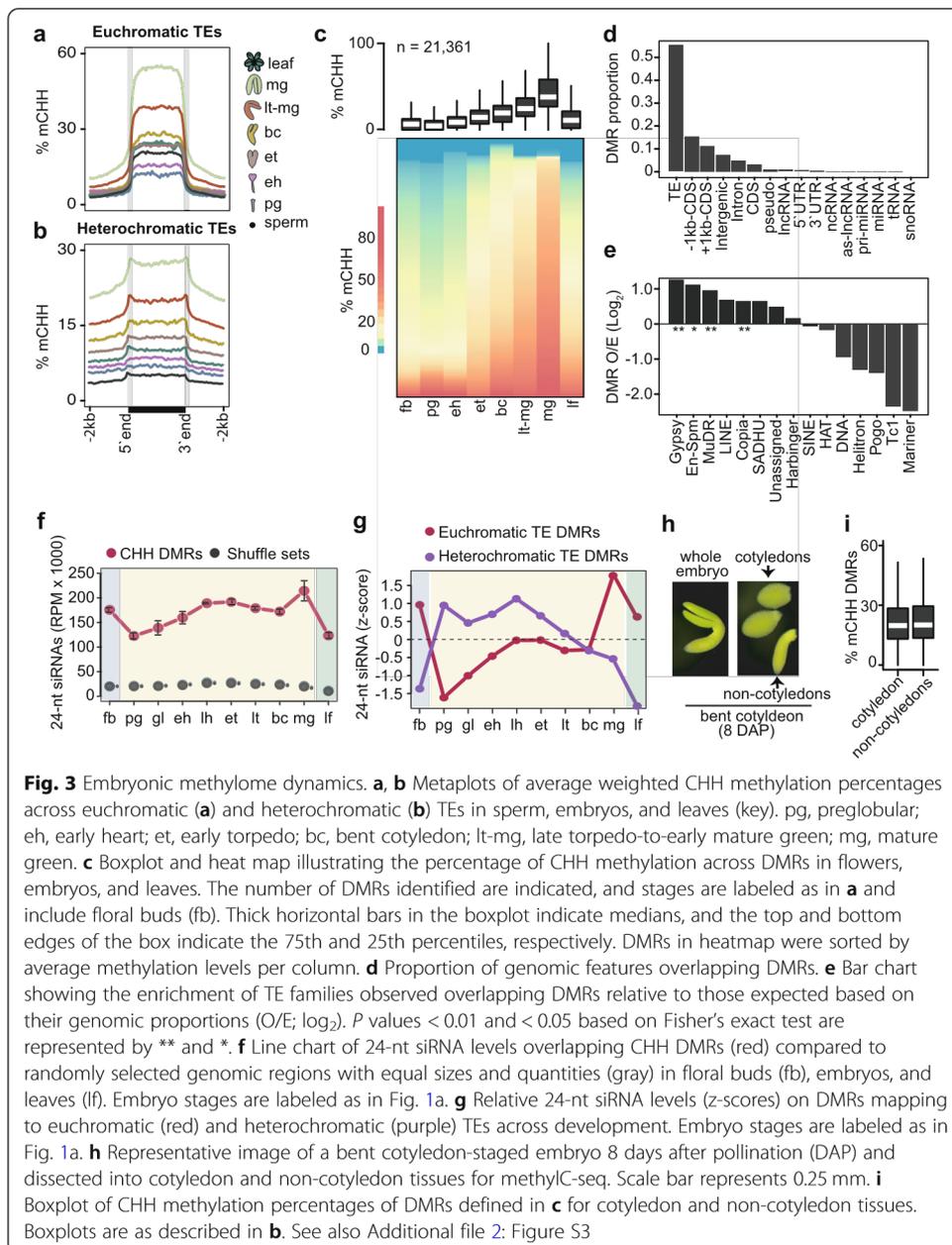


Table S4, Additional file 2: Figure S3d). Consistent with the genome-wide CHH methylation dynamics described above, embryonic CHH DMRs were hypomethylated in preglobular embryos, progressively methylated until embryo maturation, and then were sharply reduced in leaves and floral buds (Fig. 3c and Additional file 2: Figure S4A). Moreover, these embryonic DMRs were observed on various TE families especially Gypsy and Copia LTR retrotransposons, as well as MuDR and En-Spm family DNA TEs (Fig. 3d, e). The enrichment of mCHH DMRs across various TE families was generally consistent with the corresponding levels of embryonic 24-nt siRNAs (Fig. 3e and Additional file 2: Figure S2F). To examine the relationships between siRNAs and methylation further, we quantified 24-nt siRNA levels on mCHH DMRs across embryogenesis. Compared to randomized controls, CHH DMRs were highly enriched for 24-nt siRNAs (Fig. 3f). siRNAs overlapping euchromatic TE DMRs were lowest in early embryos, peaked at maturation and were also abundant in floral buds and leaves. In contrast, siRNAs overlapping heterochromatic TE DMRs were highly abundant at early-to-middle stages of embryogenesis and then strongly reduced in late-embryonic stages, leaves, and floral buds (Fig. 3g). Therefore, large-scale changes of DNA methylation occur on both euchromatic and heterochromatic TEs during embryogenesis and are associated with 24-nt siRNAs.

Notably, both euchromatic and heterochromatic TEs were hypermethylated in mature embryos relative to earlier stages and post-embryonic tissues (Fig. 3a–c and Additional file 2: Figure S3B). Because the proportion of embryonic tissue composed of cotyledons also

increases during embryo development, we tested whether the progressively increasing levels of embryonic CHH methylation could be merely due to increased proportions of cotyledon tissues in embryos as they develop. Namely, we dissected cotyledon and non-cotyledon tissues from bent-cotyledon-staged embryos, and profiled their methylomes (Fig. 3h). Both cotyledon and non-cotyledon tissues were similarly hypermethylated on DMRs indicating that hypermethylation occurs throughout late-staged embryos and is not confined to the terminally differentiated cotyledons (Fig. 3i). Interestingly, the CHH hypermethylation observed in mature embryos resembled the hypermethylation reported in root columella cells and pollen vegetative nuclei [30, 50]. Similar to these specific cell-types, mature embryos have also exited the cellcycle, which is further supported by increased and decreased levels of transcripts encoding negative and positive regulators of the cell-cycle, respectively (Additional file 2: Figure S3E) [51]. Therefore, CHH hypermethylation of mature embryos, as well as root columella cells and vegetative nuclei, is associated with cell cycle dormancy or exit.

Small RNA-directed methylation of transposons during embryogenesis

To test whether 24-nt siRNAs derived from euchromatic and heterochromatic TEs were necessary for progressive TE methylation during embryogenesis, we performed methylC-seq on 24-nt siRNA-deficient early heart (4 DAP) and bent-cotyledon embryos (8 DAP), as well as leaves and floral buds (Additional file 1: Table S1). NRPD1A encodes the largest subunit of RNA polymerase IV (Pol IV) [16], and accordingly, 24-nt siRNAs overlapping euchromatic and heterochromatic TEs were nearly eliminated in *nRPD1a*-3 mutants (Fig. 4a, c). Nearly all euchromatic TEs were completely hypomethylated in *nRPD1a* embryonic and post-embryonic tissues (Fig. 4b and Additional file 2:

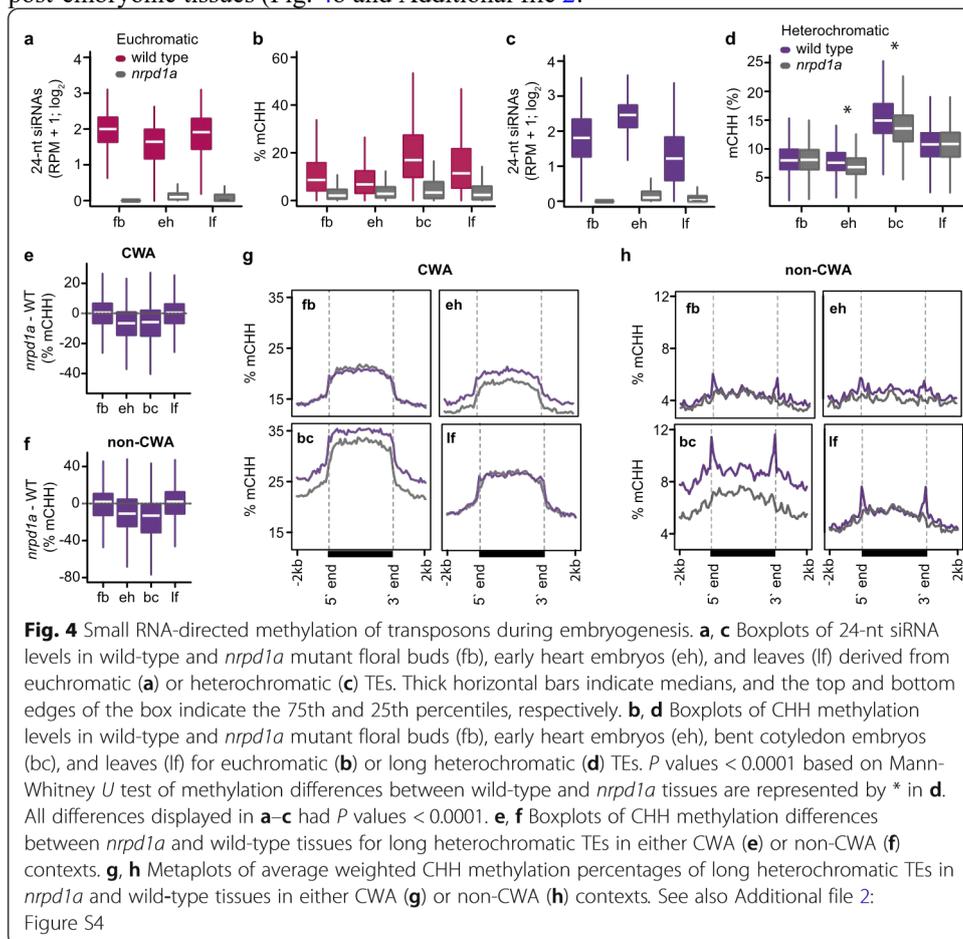


Figure S4). Because the 5729 heterochromatic TEs classified based on their embryonic siRNA dynamics also included short TEs methylated by the RdDM pathway in postembryonic tissues (Additional file 2: Figure S2D, E), we partitioned heterochromatic TEs into either short (≤ 723 bp), medium (724–2114 bp), or long (> 2114) and examined their meth-

ylation levels in *nripd1a* tissues (Fig. 4d and Additional file 2: Figure S5A-C). Consistent with previous observations from post-embryonic tissues [26, 27, 43], CHH methylation of short and medium TEs was significantly reduced compared to wild type in all *nripd1a* tissues tested including embryos (Additional file 2: Figure S5B, C). In contrast, long heterochromatic TEs were globally unaffected in *nripd1a* post-embryonic tissues, but significantly hypomethylated in *nripd1a* mutant embryos relative to wild type (Fig. 4d). Methylation of long heterochromatic TEs is thus partially dependent on siRNAs in embryonic, but not post-embryonic tissues.

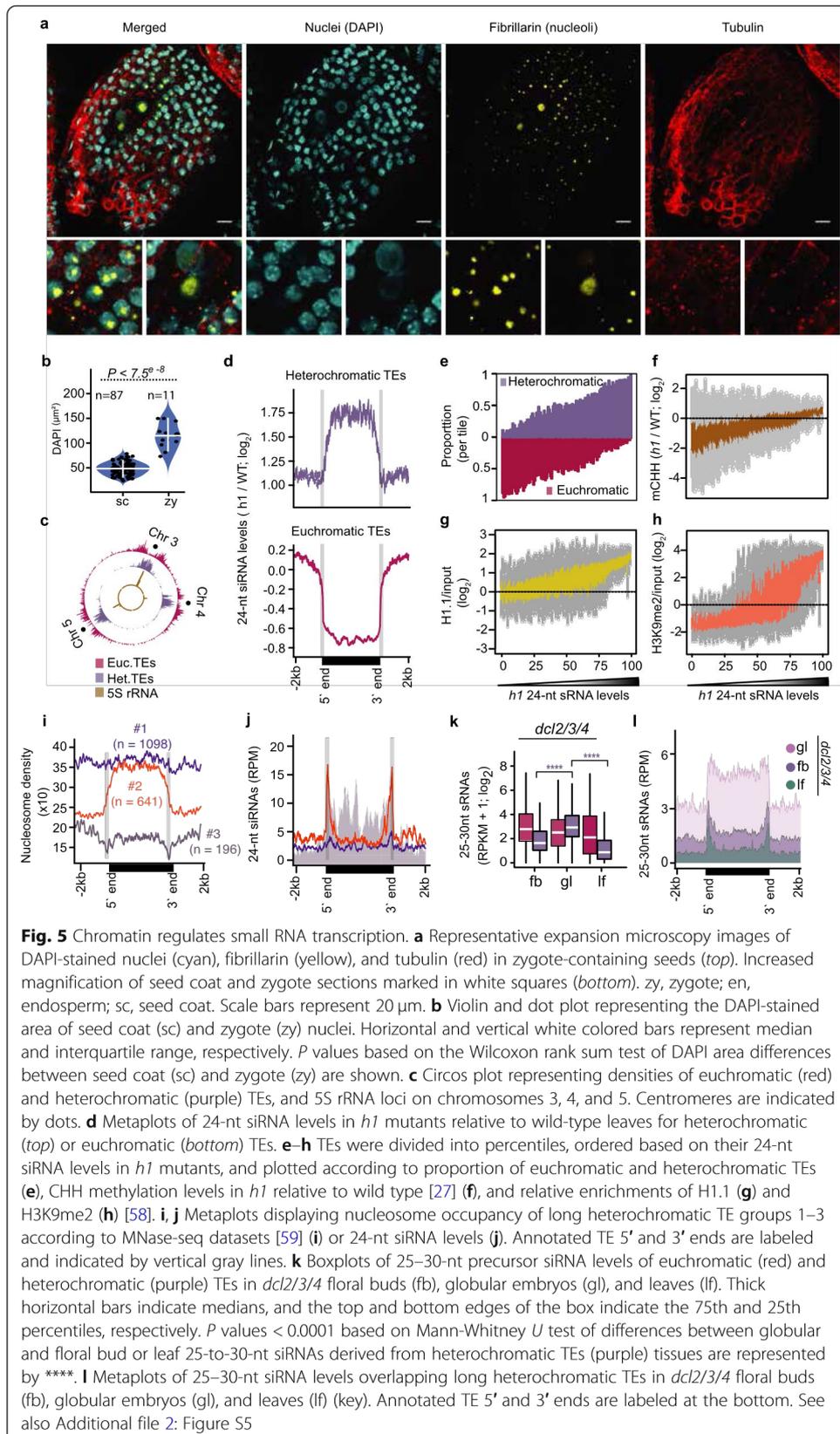
Long heterochromatic TEs are methylated by CMT2 in post-embryonic tissues, and their highly condensed chromatin states were proposed to inhibit siRNA-directed DRM2-mediated methylation [26, 27, 43, 52]. CHH methylation can be classified as CWA (W = A or T) or non-CWA. DRM2 methylates CWA and non-CWA sites, and CMT2 preferentially methylates CWA nucleotides [53–55]. Only the edges of long heterochromatic TEs were hypomethylated in non-CWA contexts of post-embryonic *nripd1a* mutant tissues relative to wild type. However, both edges and bodies of long heterochromatic TEs were hypomethylated in all CHH contexts in early heart and especially bent-cotyledon *nripd1a* embryos compared to wild type (Fig. 4g, h). Therefore, 24-nt siRNAs originating from both bodies and edges of long heterochromatic TEs rapidly increase in early embryos and are partially required for TE methylation.

Chromatin regulates small RNA transcription

Heterochromatin prevents access to de novo methyltransferases [27, 56], and thus, it may also impede Pol IV access and resulting siRNA transcription. Consistent with relatively low transcript levels of heterochromatin-promoting factors in early embryos [57], reduction of DAPI-stained chromocenters and enlarged nuclei in zygotes compared to somatic tissues indicated that zygotic chromatin is decondensed (Fig. 5a, b and Additional file 2: Figure S6A). A marked increase in zygotic nucleoli size, as well as the co-localization of hundreds of 5S rRNAs and heterochromatic TEs in centromeric regions (Copenhaver, 1999; Simon et al., 2018), further suggested that heterochromatic TEs could be decondensed in early embryos (Fig. 5b, c). Therefore, post-fertilization heterochromatin decondensation, potentially associated with rRNA production, may permit Pol IV accessibility to heterochromatic TEs and corresponding transcription of 24-nt siRNA precursors soon after fertilization. We then used a more experimentally tractable post-embryonic tissue to test how chromatin may generally regulate siRNA production.

Because linker histone 1 (H1) inhibits RNA polymerases from binding to chromatin [60, 61], and its depletion results in the loss of both chromocenters and chromatin compaction [59, 62, 63], we tested whether decreased H1 levels during postembryogenesis were sufficient to increase siRNA biogenesis throughout long heterochromatic TEs similar to what we observed in embryos. That is, we performed sRNAseq on leaves carrying null mutations in the two expressed H1.1 and H1.2 isoforms (i.e., *h1.1-1/h1.2-1* or *h1* mutants) and found that 24-nt siRNAs from heterochromatic TEs were significantly increased by more than 3.7-fold and were predominantly derived from TE bodies (Fig. 5d: Additional file 2: Figure S6B). Strikingly, hierarchical clustering of TE-derived 24-nt siRNAs demonstrated that *h1* leaf siRNA populations were more similar to siRNA populations from wild-type mature embryos instead of leaves indicating that H1 depletion was sufficient to induce an embryo-like siRNA population in a post-embryonic tissue (Additional file 2: Figure S6C).

To investigate the effects of H1 depletion on TE-derived siRNAs in more detail, we ranked euchromatic and heterochromatic TEs based on their 24-nt siRNA levels in *h1* mutant leaves and found that increased 24-nt siRNA levels from long heterochromatic TEs in *h1* mutants were positively correlated with TE length, CHH methylation, and H1 occupancy (Fig. 5e–g and Additional file 2: Figure S6D). Because TEs with increased 24-nt siRNAs in *h1* mutants were also enriched for HISTONE 3 LYSINE 9 dimethylation (H3K9me2) (Fig. 5h), we tested whether loss of H3K9me2 can also affect heterochromatic siRNA production. We examined the levels of siRNAs from long heterochromatic TEs in leaves triple mutant for *SU(VAR)3-9* *HOMOLOG 4/5/6* histone methyltransferases (*svh4/5/6*), which are deficient in H3K9me2 levels (Additional file 1: Table S1) [43]. Long heterochromatic TEs produced only 1.2-fold more 24-nt siRNAs in *svh4/5/6* leaves compared to wild type, and the global 24-nt siRNA populations were similar to wild type (Additional file 2: Figure S6B, C). These results suggest that



reducing heterochromatin, rather than H3K9me2 marks associated with heterochromatin, is sufficient for siRNA production from long heterochromatic TEs. Accordingly, H1 promotes nucleosome occupancy on heterochromatic regions [58, 59], and we found that heterochromatic TEs were enriched for H1 (Fig. 5g and Additional file 2: Figure S6D) and had reduced nucleosome occupancy in *h1* mutants (Additional file 2: Figure S6F). Our results indicate that depletion of H1 is sufficient to decrease nucleosome occupancy of heterochromatic TEs, as well as increase corresponding siRNA and CHH methylation levels.

Because 24-nt siRNAs were enriched on the bodies of long heterochromatic TEs in *h1* mutants with reduced chromatin compaction (Fig. 5d and Additional file 2: Figure S6B), we next examined the relationships between nucleosome occupancy and 24-nt siRNA levels of long heterochromatic TEs in wild-type leaves. We employed an iterative k-means clustering approach to generate three groups of long heterochromatic TEs based on their nucleosome occupancy using publicly available micrococcal nuclease sequencing data [59] (Fig. 5i). Group 1 comprised 1098 TEs (56.7% of total) that had high densities of nucleosomes and were devoid of 24-nt siRNAs throughout their lengths (Fig. 5i, j). Group 2 contained 641 TEs (33.1% of total) and had low and high nucleosome occupancy over the edges and bodies, respectively, and were enriched for 24-nt siRNAs only on the edges (Fig. 5i, j). Group 3 consisted of only 196 TEs (10.2% of total) and had very low nucleosome levels, but abundant 24-nt siRNAs, on both their edges and bodies similar to euchromatic TEs (Fig. 5i, j). Altogether, these results suggest that increased nucleosome occupancy restricts RNA Pol IV activity, and thus, chromatin states alone appear to explain 24-nt siRNA production from TEs.

RNA Pol IV transcribes ~25-to-40-nt RNAs that are co-transcriptionally converted to double-stranded RNAs by RNA-dependent RNA Polymerases and rapidly processed into 23-nt/24-nt duplexes by DICER-LIKE (DCL) endoribonucleases [17, 18, 22]. These transient Pol IV-dependent 24-nt siRNA precursors can be robustly detected in *dcl2/3/4* mutants [17, 18, 64], and thus, 24-nt siRNA precursor levels indicate Pol IV transcriptional activities. We performed sRNA-seq on *dcl2/3/4* flowers, globular embryos, and leaves and compared levels of 24-nt siRNA precursors from euchromatic and heterochromatic TEs (Additional file 1: Table S1). Compared to leaves and flowers, we respectively detected 11-fold and 3.8-fold significantly more 24-nt precursors from heterochromatic TEs in *dcl2/3/4* early embryos (Fig. 5k). Importantly, the 24-nt siRNA precursors mostly originated from the bodies of long heterochromatic TEs in embryos, but were strongly reduced in floral buds and leaves (Fig. 5l). Together with the observations that 24-nt siRNAs were also enriched on the bodies of long heterochromatic TEs in early embryos and *h1* mutant leaves (Figs. 2c, and 5d), our results are consistent with a model whereby decompaction of heterochromatin in early embryos and *h1* mutant leaves permits Pol IV access and transcriptional activities to produce 24-nt siRNAs.

Homeostasis of transposon-derived siRNAs

In contrast to heterochromatic TEs, we found 2.7-fold significantly less 24-nt siRNAs from euchromatic TEs in *h1* leaves compared to wild type, which was also associated with their CHH hypomethylation (Fig. 5d, f and Additional file 2: Figure S6B). Unlike heterochromatic TEs, euchromatic TEs were lowly enriched for H1 in wild-type leaves and nucleosome occupancy was further reduced in *h1* mutants (Fig. 5g and Additional file 2: Figure S6D, F). Therefore, siRNA depletion from euchromatic TEs is likely an indirect consequence of sequestering Pol IV to accessible heterochromatic TEs. Moreover, we observed the greatest enrichment of siRNAs derived from heterochromatic compared to euchromatic TEs at the preglobular stage of embryogenesis, which is the earliest post-fertilization sRNA-seq dataset available (Fig. 6a). This was reduced during mid-embryogenesis and then further decreased to almost post-embryonic levels during maturation (Fig. 6a). Together with our siRNA precursor analysis (Fig. 5k, l), this indicates that Pol IV is more efficiently recruited to heterochromatic TEs compared to euchromatic TEs during the initial stages of embryogenesis. The enrichment of heterochromatic TE-derived 24-nt siRNAs in preglobular embryos surpassed what we observed in *h1* mutant leaves (Fig. 6a), suggesting that reduced nucleosome occupancy alone does not fully account for the extreme enrichment of heterochromatic siRNAs in preglobular embryos.

Based on available sRNA-seq datasets, euchromatic, but not heterochromatic, TE-derived 24-nt siRNAs were substantially reduced in mutants deficient in CG and CHH methylation (Fig. 6b and Additional file 2: Figure S7) [21, 43, 69]. Therefore, low CHH

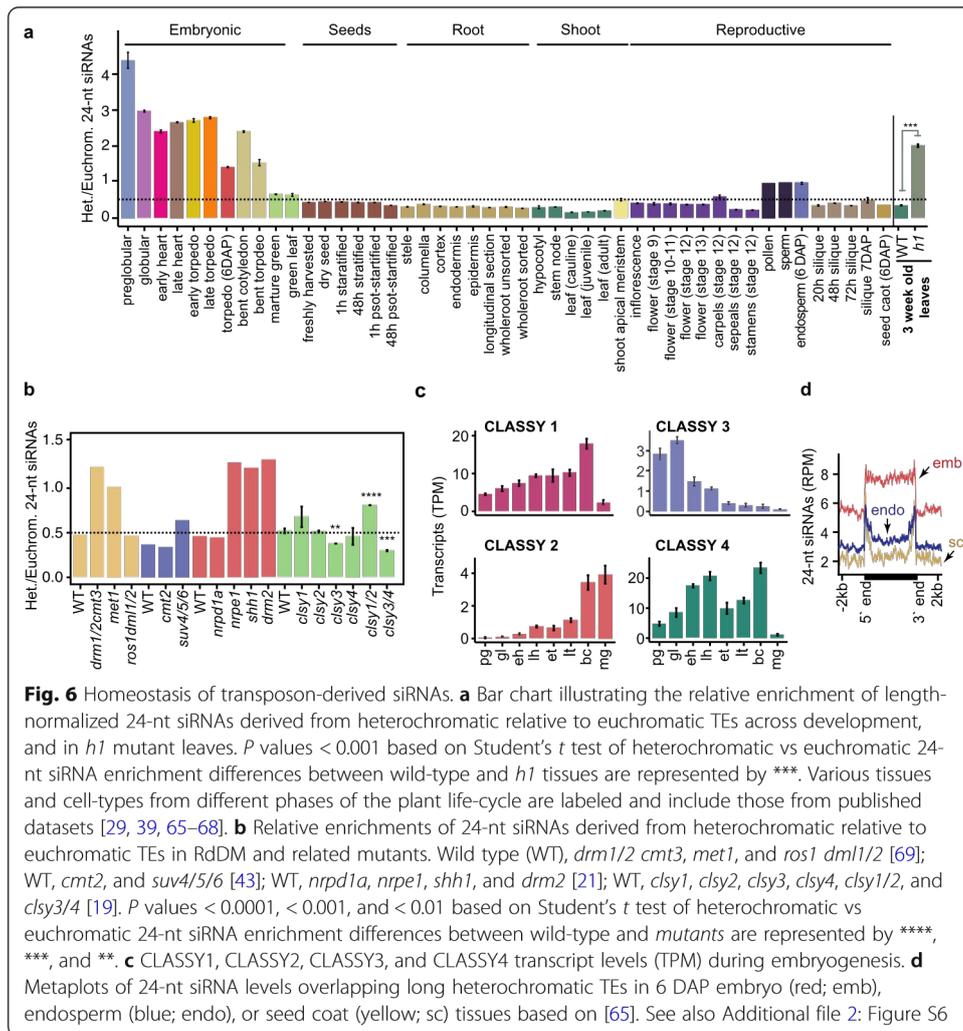


Fig. 6 Homeostasis of transposon-derived siRNAs. **a** Bar chart illustrating the relative enrichment of length-normalized 24-nt siRNAs derived from heterochromatic relative to euchromatic TEs across development, and in *h1* mutant leaves. *P* values < 0.001 based on Student's *t* test of heterochromatic vs euchromatic 24-nt siRNA enrichment differences between wild-type and *h1* tissues are represented by ***. Various tissues and cell-types from different phases of the plant life-cycle are labeled and include those from published datasets [29, 39, 65–68]. **b** Relative enrichments of 24-nt siRNAs derived from heterochromatic relative to euchromatic TEs in RdDM and related mutants. Wild type (WT), *drm1/2 cmt3*, *met1*, and *ros1 dml1/2* [69]; WT, *cmt2*, and *suv4/5/6* [43]; WT, *nrdp1a*, *npe1*, *shh1*, and *drm2* [21]; WT, *cly1*, *cly2*, *cly3*, *cly4*, *cly1/2*, and *cly3/4* [19]. *P* values < 0.0001, < 0.001, and < 0.01 based on Student's *t* test of heterochromatic vs euchromatic 24-nt siRNA enrichment differences between wild-type and mutants are represented by ****, ***, and **. **c** CLASSY1, CLASSY2, CLASSY3, and CLASSY4 transcript levels (TPM) during embryogenesis. **d** Metaplots of 24-nt siRNA levels overlapping long heterochromatic TEs in 6 DAP embryo (red; emb), endosperm (blue; endo), or seed coat (yellow; sc) tissues based on [65]. See also Additional file 2: Figure S6

methylation in preglobular embryos (Fig. 3a, b) may reduce methylation-dependent feedback loops that facilitate production of siRNAs from euchromatic TEs in preglobular stages. CLASSY (CLSY) chromatin remodeling factors also promote 24-nt siRNA production: CLSY 1/2 and CLSY 3/4 help recruit Pol IV to euchromatic and heterochromatic regions, respectively (Fig. 6b) [19, 21, 43, 69]. Dynamic chromatin states and corresponding establishment of methylation/CLSY-dependent transcription of siRNA precursors likely contribute to the unique siRNA populations observed in embryos.

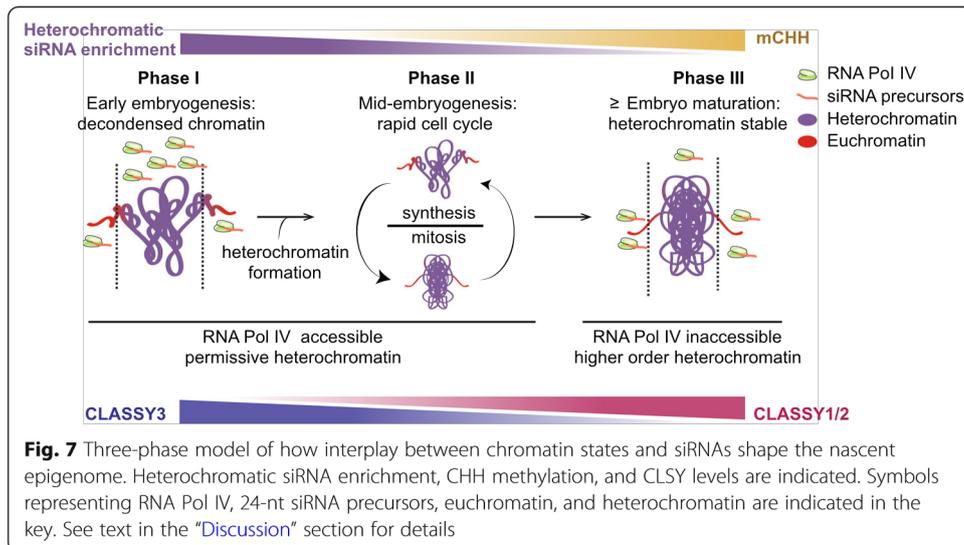
Remarkably, the relative amounts of euchromatic and heterochromatic TE-derived siRNAs remained stable throughout post-embryonic development (Fig. 6a). Root columella and pollen vegetative cells are depleted for H1 [50, 70], but were not depleted for euchromatic siRNAs (Fig. 6a). This may be due to these terminally differentiated celltypes being derived from a single mitotic division, and thus, they may retain the ability to recruit Pol IV to euchromatic TEs. For example, these cell types are CHH hypermethylated [30, 50], and thus, CHH methylation-dependent positive feedback loops on euchromatic TEs may counteract the loss of H1. Importantly, TE-derived siRNA populations in embryos were distinct from endosperm and seed-coat tissues (Fig. 6d). Consistent with chromatin states primarily regulating siRNA production, the endosperm has reduced cell division compared to embryos by 6 DAP [71]. Accordingly, endosperm siRNAs overlapped mostly edges, but not bodies, of heterochromatic TEs typical of other non-embryonic populations (Fig. 6d). Similar to pollen, but in contrast to other non-embryonic tissues, endosperm had similar levels of siRNAs from heterochromatic and euchromatic TEs (Fig. 6a). Because nearly equal siRNA levels from euchromatic and heterochromatic siRNAs were also observed in *met1* mutants (Fig. 6b and Additional file 2: Figure S7), this balance may be due to loss of CG methylation-dependent euchromatic siRNA production in CG hypomethylated pollen vegetative nuclei and endosperm [32]. Altogether our data indicate that the homeostasis of 24-nt siRNA production from euchromatic and heterochromatic TEs are affected by dynamic chromatin states including, but not restricted to, those associated with early embryogenesis.

Discussion

Although siRNAs direct faithful re-establishment of methylation genome-wide across generations [72], the dynamics of embryonic siRNAs and how they contribute to the nascent epigenome have not been reported. In this study, we demonstrated that thousands of TEs produce exceptionally high levels of 24-nt siRNAs in embryos (Fig. 1) and can be classified into two distinct groups based on their developmental dynamics (Fig. 2). siRNAs from euchromatic TEs gradually increase to post-embryonic levels during embryogenesis and are constitutively required to direct TE methylation in embryonic and post-embryonic tissues (Fig. 4). In contrast, heterochromatic TEs produce a burst of siRNAs soon after fertilization, and specifically during embryogenesis, to help establish TE methylation *de novo*, which is then maintained independent of siRNAs during post-embryogenesis (Fig. 4) [26, 27, 43]. Interestingly, the levels of siRNAs from these euchromatic and heterochromatic bipartite-classified TEs are regulated according to their chromatin states (Fig. 5). Decondensed chromatin permits transcription of 24nt siRNAs, and this contributes to cell autonomous homeostatic control mechanisms that normalize chromatin states.

We propose a three-phase model for how chromatin states, and resulting siRNA dynamics, help shape the nascent epigenome (Fig. 7). After fertilization, zygotic chromatin is decondensed and this appears to be associated with transcriptional activation of rRNAs, including hundreds of 5S rRNA loci that co-localize with heterochromatic TEs near centromeric regions (Fig. 5a, c). Arabidopsis zygotes require *de novo* synthesis of gene products directly after fertilization [57, 73], and genes involved in rRNA biogenesis produce high levels of transcripts in preglobular embryos relative to later stages [40]. In contrast to Arabidopsis, maternally donated proteins drive early embryogenesis in *Xenopus* and H1 dynamics mediate transcriptional activation of rRNA loci in oocytes and their silencing in somatic tissues [74, 75]. In pollen vegetative cells, decondensation of rRNA loci can also permit their transcription and concomitant cell growth [76] and may be required for the rapid cell divisions in early endosperm, which also have enlarged nucleoli (Fig. 5a). Therefore, reduced heterochromatin in a variety of cell types, including those producing large amounts of protein such as early embryos, endosperm, and pollen vegetative cells, may be permissive for Pol IV-mediated transcription of siRNA precursors from TEs that are typically in a deep heterochromatic state during other developmental phases (Fig. 7; phase 1). Consistent with decondensation of heterochromatin facilitating *de novo* production of siRNAs, Pol IV-dependent 24-nt siRNAs were sharply increased throughout heterochromatic TEs in h1 mutant leaves with reduced heterochromatin (Fig. 5d and Additional file 2: Figure S6). Production of siRNAs from euchromatic TEs is delayed relative to those from heterochromatic TEs (Fig. 6a). Because euchromatic, but not heterochromatic, TEs require CHH methylation and CLSY1/2 chromatin remodelers to produce full siRNA levels (Fig. 6b) [19, 21, 43, 69], both of which are gradually increased during embryogenesis (Figs. 3a–c and 6c), the developmental time-lag in euchromatic compared to heterochromatic siRNA production may also be partially due to the delay establishing methylation/CLSY dependent positive feedback loops during early embryogenesis.

Embryos divide rapidly through the bent cotyledon stage [77] and dynamic chromatin condensation and decondensation associated with such increased cell division [78, 79] likely allows access of Pol IV to both heterochromatic and euchromatic TEs (Fig. 7;



Phase 2). Production of embryonic siRNAs from heterochromatic relative to euchromatic TEs is steady between the globular and bent cotyledon stages. Heterochromatic TE-derived siRNAs are rapidly reduced upon maturation when heterochromatin becomes highly condensed [80] and euchromatic domains containing genes encoding seed storage and oil body biogenesis proteins are transcriptionally activated [40] (Fig. 6a). As a consequence, Pol IV access to heterochromatic TEs is likely greatly reduced, and this results in more Pol IV being readily available to produce siRNAs from euchromatic TEs (Fig. 7; Phase 3). Consistently, we observed a burst of siRNAs from euchromatic TEs, and their associated hypermethylation, at the mature stage (Fig. 2a, b). Heterochromatic TEs also become hypermethylated at the mature stage, which appears to be largely independent of siRNAs, but rather dependent on CMT2 as is typical of subsequent post-embryonic development (Fig. 4). Therefore, CHH hypermethylation throughout mature embryos may largely be a consequence of the rapid shift of chromatin states upon maturation.

Similar to h1 mutants [58, 59], heat stress also causes reduced nucleosome occupancy and decondensed heterochromatin [81, 82]. Moreover, we classified the heat-activated ONSEN/AT5TE15240 as a class B/heterochromatic TE based on its siRNA dynamics, and ONSEN transcription and transposition are greatly enhanced in RdDM-defective mutants [83]. Interestingly, 24-nt siRNAs were increased throughout the body of ONSEN TEs after heat stress, and this was further enhanced in rapidly dividing undifferentiated calli [84], which has increased chromatin accessibility in rice [84, 85]. Based on these results, as well as our previous observation that heat-stress-related genes are significantly enriched in preglobular embryos [40], we suggest that the chromatin dynamics caused by heat stress and subsequent recovery are analogous to what we observed in early embryos. That is, heat stress and fertilization may decrease nucleosome occupancy across heterochromatic TEs, which enables Pol IV-mediated siRNA production and subsequent reconstitution of proper heterochromatin to help limit TE mobilization in the genome. Strong upregulation of TE-specific endo-siRNAs and piwi-interacting RNAs observed in H1-depleted *Drosophila* [86] suggests that H1-dependent regulation of chromatin states may also facilitate heterochromatic small RNA transcription in animals.

Conclusions

Reprogramming of heterochromatin during early embryogenesis occurs in diverse metazoa including flies, mammals, worms, and zebrafish [8–12], and we observed that CHH methylation is also reprogrammed during early plant embryogenesis (Fig. 3). For example, CHH methylation was essentially lost on the bodies of heterochromatic TEs in sperm and subsequently fully re-established by both siRNA-dependent and siRNA-independent pathways (Fig. 4). Reduced heterochromatin in early animal embryos has been associated with increased developmental potential [87, 88], and similar relationships have also been observed in plants including reprogramming associated with plant regeneration and heat-stress induced somatic embryogenesis [89–95]. However, decreased heterochromatin would also increase the risk of TE mobilization and resulting mutations, and this could be especially dangerous in plant zygotes because they are the precursors of all cell types including the gametes. Our results indicate that embryos produce 24-nt siRNAs according to their chro-

matin states including those that are permissible for Pol IV transcription soon after fertilization. These de novo produced 24-nt siRNAs direct re-methylation of both euchromatic and heterochromatic TEs in the new generation. Therefore, decondensed chromatin permits transcription of early embryonic siRNAs to help promote cell-autonomous TE silencing. More generally, reduced heterochromatin due to sharp increases in rRNA production requirements during growth (e.g., early embryos, endosperm, and pollen vegetative cells), and perhaps in response to external cues such as heat stress, enables the synthesis and functions of siRNAs that can help reconstitute proper chromatin states.

Methods

Plant material and growth conditions

All genotypes were in the Columbia-0 (Col-0) *Arabidopsis thaliana* background including *dcl2/3/4* mutants composed of *dcl2-1*, *dcl3-1* and *dcl4-2t* [96], *h1.1-1/h1.2-2* [27], *nrpd1a-3* [97], and *suv4/5/6* [98]. Plants were grown in a climate-controlled growth chamber at 20 to 22 °C under a 16-h light/8-h dark cycle with incandescent lights at 130 to 150 $\mu\text{mol}/\text{m}^2/\text{s}$.

Embryo isolation and nucleic acid extraction

Embryos were dissected from siliques either 3 days after pollination (DAP) (preglobular), 4 DAP (early heart/transition), or 8 DAP (bent cotyledon). Siliques were opened with forceps and seeds were collected in 2-ml Eppendorf tubes containing nucleasefree water and kept on ice. Seeds were then crushed with pestles, and embryos were selected under an inverted microscope using a microcapillary tube. Isolated embryos, as well as cotyledon and non-cotyledon portions of bent-cotyledon embryos, were thoroughly and serially washed 4 \times with nuclease-free water and stored at -80 °C. RNA was isolated as previously described [39, 99]. Genomic DNA was extracted from ≥ 50 embryos per stage, floral buds, and leaves using Quick-DNA™ Micro prep Kit (Zymo D3020) according to the manufacturer's recommendations.

Small RNA profiling

siRNA-seq libraries were generated as previously described [39]. Briefly, total RNA from each sample was size selected for 18 to 30-nucleotide RNAs using denaturing polyacrylamide-urea gels. Size-selected RNA was used to ligate adapters and synthesize cDNA with the NEBNext Multiplex Small RNA Library Prep Set for Illumina kit (cat. no E7300; New England Biolabs) according to the manufacturer's recommendations. Various numbers of PCR cycles were used to amplify cDNAs: 18, 20, 22, and 24 PCR cycles for globular and 14, 16, 18, and 20 PCR cycles for early heart and 3-week-old leaf samples. Final PCR amplicons were initially run on a 90% (v/v) formamide/8% (w/v) acrylamide gel for 30 min at 5 W, followed by 30 W for ≥ 2 h, and stained with SYBR Gold (1:10,000; Thermo Fisher Scientific). PCR amplicons between 137 and 149 bp corresponding to 18- to 30-nucleotide siRNAs with adapters, respectively, were inspected under a UV transilluminator, and amplicons with non-saturated signals generated from PCR cycles were gel-purified. Gel-purified small RNA libraries were resuspended in 15 μL of Elution Buffer (Qiagen). Finally, small RNA libraries were quality checked for the expected size range with Agilent High sensitivity NGS fragment Kit (DNF-4741000) and were sequenced on a HiSeq 2500 instrument (Illumina) in 50-base singleend mode.

Small RNA sequencing analysis

Small RNA-seq library datasets generated in this study or downloaded from NCBI's Sequence Read Archive (SRA) were subjected to the same small RNA analysis pipeline. First, raw fastq files were adapter trimmed with Cutadapt [100] and sequences between 18 and 30 bases in length, and that contained an adapter were retained. The trimmed sequences were then aligned to the *Arabidopsis thaliana* TAIR10 genome [101] with STAR [102] requiring zero mismatches and allowing up to 100 multiple end-to-end alignments. Multi-mapping reads from aligned SAM files were re-assigned with a "richget-richer" algorithm using the custom python script "readmapIO.py" as described previously [103]. Resulting output bed-Files were then sorted, condensed, and normalized for total genome-matching reads. The

BEDtools [104] map function was then used to quantify the sum of the normalized reads per million (RPM) mapping to TAIR10 annotated Transposable elements (TEs).

For the model-based clustering of transposon-derived 24-nt siRNAs, mean RPM of 24-nt siRNAs from biological triplicates of floral bud, embryonic, and leaf samples mapping to TEs were calculated and used as input for R library Mclust [42] to identify the optimal Gaussian mixture model (GMM). By employing Mclust function `mclust-BIC(.,G=seq(2,20),by=2)` in sequential increments of two until twenty components, we identified the VEV (Variable volume, Equal shape, Variable orientation) ellipsoidal distribution model to be optimal with the minimum number of components (i.e., eight) containing maximum Bayesian Information Criterion (BIC). Finally to yield eight transposon clusters with VEV ellipsoidal distribution model, the `Mclust(.,G=8, modelNames="VEV")` function was applied on mean TE-derived 24-nt siRNAs.

Principal component analysis of 24-nt siRNAs was performed with the R `prcomp` function using default parameters. Hierarchical clustering of transposon-derived 24-nt siRNAs was performed by calculating Euclidean distances between samples and the distance matrix was subjected to the R function `hclust(*,"complete")`. Heatmaps and metaplots of TE-derived siRNAs were generated with deepTools [105]. Briefly, a matrix containing normalized 24-nt siRNA scores per genome regions for tissue types or genotypes were generated (`computeMatrix scale-regions -bs 5 -m 4000 -b 2000 -a 2000 --averageTypeBins mean`). The obtained matrix was used to generate heatmaps (`deepTools plotHeatmap`) or metaplots (`deepTools plotProfile`). For Fig. 5, regions without siRNA signals were removed and the remaining genomic regions were used to calculate matrix containing nucleosome signal and 24-nt siRNA levels. This matrix then served as input to employ Iterative K-means clustering with deepTools function `plotProfile --kmeans`.

DNA methylation profiling

MethylC-seq libraries were generated using post-bisulfite adapter tagging (PBAT) to avoid the bisulfite-induced loss of intact sequencing templates as described [48] with the following modifications. Briefly, genomic DNA was subjected to bisulfite treatment for 200 min with EZ DNA Methylation-DirectTM Kit (Zymo D5020). Bisulfite-treated DNA was then preamplified for two cycles with primers (5'-CCCTACACGACGCTCT TCCGATCT-NNNNNN-3') containing random hexamers and purified using the Zymo DNA Clean and Concentrator kit. Adaptor primers (5'-CAGACGTGTGCTCTCCG ATCTNNNNNN-3') were added to preamplified products and then amplified for 12 PCR cycles with indexing primers for Illumina sequencing. Methylome libraries were purified using Beckman Coulter AMPureXP DNA beads. Libraries quality checked for fragment length between 200 and 600 bp were used for sequenced in single-read mode on an Illumina HiSeq2500 or Nextseq instrument.

DNA methylation analysis

Sequenced reads were quality filtered and trimmed using Trim Galore with default settings. In addition, the first six bases of each read were removed to exclude random hexamers from the pre-amplification step of library construction and to also reduce 5' methylation-bias (m-bias). Reads were aligned against the C-to-T converted TAIR10 genome using Bismark in non-directional mode to original top strand (OT), original bottom strand (OB), complementary to OT (CTOT) and OB (CTOB) (`bismark --non_directional -q --score-min L,0,-0.4`) [106]. Aligned BAM files containing clonal duplicates were removed with function `deduplicate_bismark -s --bam`, and uniquely mapped reads were then used as input for the Methylpy software [107]. Weighted methylation rates at each covered cytosine was extracted using command `methylpy call-methylation-state --pairedend FALSE`. Bisulfite conversion rates were calculated using the unmethylated chloroplast genome or spiked-in unmethylated Lambda phage DNA controls (European Nucleotide Archive Accession Number J02459, Promega catalog number D1521). FASTQ files obtained from publicly available methylomes generated from sperm [32], early torpedo [49], mid-torpedo to early maturation [31], mature green [37] embryos, and H1 mutant tissues [27] were also processed in the similar manner; except only 5' end nucleotides of the reads with m-bias were removed and aligned in directional mode to OT and OB strands.

Differentially methylated regions (DMRs) were defined using Methylpy as described [36]. Briefly, biological replicates were pooled and differentially methylated sites (DMSs)

were identified by the root mean square tests with false discovery rates ≤ 0.01 . Cytosine sites with ≥ 4 overlapping reads were retained for all samples except for preglobular in which DMSs with ≥ 3 overlapping reads were retained. Differentially methylated sites within 100-bp were collapsed into DMRs. CHH-DMRs were further filtered by discarding regions with < 4 DMSs and methylation differences $< 20\%$. Using these parameters, DMRs were identified in all 10 pairwise combinations across embryonic samples (preglobular, early heart, early torpedo, bent cotyledon, mature green) and merged using the BEDtools merge function [104]. DMRs were used to calculate the weighted CHH methylation rate on all analyzed tissue types. CHH methylation metaplots for class A, B, and siRNA-deficient TEs were plotted using the R library Seqplots [108]: Body, upstream, and downstream regions of TEs were split into equal-sized bins, and the average weighted mCHH level for each bin was calculated and plotted.

Expansion microscopy and DAPI quantification

The expansion microscopy technique [109] optimized for Arabidopsis seeds was conducted as previously described [73]. Anti-Fibrillar antibody (ab4566, Abcam) and anti-alpha Tubulin antibody (ab89984, Abcam) were used in 1:500 dilution as primary antibodies. Goat Anti-Mouse IgG H&L (Alexa Fluor® 488) (ab150113, Abcam) and goat Anti-Chicken IgY H&L (Alexa Fluor® 555) (ab150170, Abcam) were used in 1:500 dilution as secondary antibodies. For each sample, a stack of nine images with 1- μm intervals were recorded by ZEISS LSM700 with 25 \times oil objective and ZEN software at 1024 \times 1024 resolution in 8-bit. DAPI signals were excited by 405-nm laser and passed through SP490 filters. Alexa488 signals were excited by 488-nm laser and passed through BP490-635 filters. Alexa555 signals were excited by 555-nm laser and passed through 560–1000-nm filters. Pinhole sizes were kept as 1 airy unit for each color, and color channels were scanned separately. FIJI software was used for image processing and nuclear size quantification. Each stack of images was first Z-projected on maximum intensity and then the nuclear areas were determined based on DAPI signals. The zygotic nuclei were distinguished from the endosperm nuclei according to position and tubulin staining patterns.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-02163-4>.

Additional file 1: Table S1. Datasets and general mapping statistics.

Additional file 2: Figure S1. Characteristics of embryonic 24-nt siRNAs and their similarities across samples. Figure S2. siRNA dynamics and characteristics. Figure S3. Benchmarking low-input methylC-seq, methylomes and cell-cycle transcripts. Figure S4. Genome browser screenshots of 24-nt siRNAs and DNA methylation levels. Figure S5. Size-based partitioning of heterochromatic TEs and small RNA-directed methylation. Figure S6. TE-derived siRNA accumulation and association with chromatin. Figure S7. TE-derived siRNAs in methylation mutants.

Additional file 3: Table S3. Euchromatic and heterochromatic TE classifications.

Additional file 4: Table S4. Differentially methylated regions.

Additional file 5. Review history.

Acknowledgements

We thank the Vienna Biocenter Core Facilities GmbH (VBCF) Next Generation Sequencing and Plant Sciences Facilities for next-generation sequencing and plant growth chamber access, respectively, and the Institute of Molecular Pathology-Institute of Molecular Biology-Gregor Mendel Institute Molecular Biology Services for instrument access and support. We also thank Alexander Vogt for help in optimizing low-input methylC-seq library preparation; Anna Smolka for technical assistance; Patrick Hüther and Claude Becker for advice on methylation analysis; Zdravko Lorkovi, Michael Borg, and Frédéric Berger for sharing reagents; and Michael Schon, Balaji Enugutti, and other members of the Nodine lab for valuable input.

Peer review information

Kevin Pang was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 5.

Authors' contributions

R.K.P. and M.D.N. conceived the project; R.K.P. developed the methodology, implemented software used, and performed formal analysis; R.K.P., S.P., K.P., P.K., S.L., and M.D.N. conducted the experiments; R.K.P. and M.D.N. wrote and edited the article; M.D.N. supervised the project and acquired funding. The authors read and approved the final manuscript.

Funding

This work was supported by the European Research Council under the European Union's Horizon 2020 Research and Innovation Program (grant 637888 to M.D.N.).

Availability of data and materials

All sequencing data generated in this study are available at the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO, <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE152971 [110]. Publicly available next-generation sequencing data were downloaded from NCBI, GEO, and are listed along with general mapping statistics in Additional file 1: Table S1. The software code used for the sRNA, methylome, and transcriptome analysis including a nextflow pipeline is available at <https://github.com/mnodine/Papareddy.2020> [111]. BigWig files of processed datasets generated either as part of this study or publicly available can be downloaded at https://github.com/mnodine/Papareddy.2020/tree/master/processed_BigWigs [111].

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests
The authors declare that they have no conflicts of interests.

Received: 10 April 2020 Accepted: 4 September 2020

Published online: 17 September 2020

References

1. Heitz E. Das heterochromatin der moose. *Jahrb Wiss Bot.* 1928;69:762–818.
2. Grewal SIS, Jia S. Heterochromatin revisited. *Nat Rev Genet.* 2007;8:35–46.
3. Elgin SCR, Grewal SIS. Heterochromatin: silence is golden. *Curr Biol.* 2003;13:R895–8.
4. Huisinga KL, Brower-Toland B, Elgin SCR. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma.* 2006;115:110–22.
5. Chan SW-L, Henderson IR, Jacobsen SE. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet.* 2005;6:351–60.
6. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res.* 2011;21:381–95.
7. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet.* 2019;20:207–20.
8. Laue K, Rajshakar S, Courtney AJ, Lewis ZA, Goll MG. The maternal to zygotic transition regulates genomewide heterochromatin establishment in the zebrafish embryo. *Nat Commun.* 2019. <https://doi.org/10.1038/s41467-019-09582-3>.
9. Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat Cell Biol.* 2018;20:620–31.
10. Rudolph T, Yonezawa M, Lein S, Heidrich K, Kubicek S, Schäfer C, et al. Heterochromatin formation in *Drosophila* is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol Cell.* 2007;26:103–15.
11. Mutlu B, Chen H-M, Moresco JJ, Orelo BD, Yang B, Gaspar JM, et al. Regulated nuclear accumulation of a histone methyltransferase times the onset of heterochromatin formation in *C. elegans* embryos. *Sci Adv.* 2018;eaat6224. <https://doi.org/10.1126/sciadv.aat6224>.
12. Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS One.* 2010;5:e10531.
13. Ingouff M, Hamamura Y, Gourgues M, Higashiyama T, Berger F. Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. *Curr Biol.* 2007;17:1032–7.
14. Ingouff M, Rademacher S, Holec S, Soljić L, Xin N, Readshaw A, et al. Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in *Arabidopsis*. *Curr Biol.* 2010;20:2137–43.
15. Martienssen R, Moazed D. RNAi and heterochromatin assembly. *Cold Spring Harb Perspect Biol.* 2015;7:a019323.
16. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. RNA polymerase IV directs silencing of endogenous DNA. *Science.* 2005;308:118–20.
17. Zhai J, Bischof S, Wang H, Feng S, Lee T-F, Teng C, et al. A one precursor one siRNA model for Pol IV-dependent siRNA biogenesis. *Cell.* 2015;163:445–55.
18. Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, et al. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. *Elife.* 2015;4:e09591.
19. Zhou M, Palanca AMS, Law JA. Locus-specific control of the de novo DNA methylation pathway in *Arabidopsis* by the CLASSY family. *Nat Genet.* 2018;50:865–73.
20. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* 2010;11:204–20.

21. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature*. 2013;498:385–9.
22. Singh J, Mishra V, Wang F, Huang H-Y, Pikaard CS. Reaction mechanisms of Pol IV, RDR2, and DCL3 drive RNA channelling in the siRNA-directed DNA methylation pathway. *Mol Cell*. 2019;75:576–89.e5.
23. Zilberman D, Cao X, Jacobsen SE. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*. 2003;299:716–9.
24. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*. 2008;135:635–48.
25. Cao X, Jacobsen SE. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Natl Acad Sci U S A*. 2002;99(Suppl 4):16491–8.
26. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell*. 2013;152:352–64.
27. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*. 2013;153:193–205.
28. Drews GN, Koltunow AMG. The female gametophyte. *Arabidopsis Book*. 2011;9:e0155.
29. Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*. 2009;136:461–72.
30. Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, et al. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*. 2012;151:194–205.
31. Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, et al. Genome-wide demethylation of Arabidopsis endosperm. *Science*. 2009;324:1451–4.
32. Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*. 2012;337:1360–4.
33. Feng X, Zilberman D, Dickinson H. A conversation across generations: soma-germ cell crosstalk in plants. *Dev Cell*. 2013; 24:215–25.
34. Mosher RA, Melnyk CW. siRNAs and DNA methylation: seedy epigenetics. *Trends Plant Sci*. 2010;15:204–10.
35. Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. *Curr Biol*. 2012;22:1825–30.
36. Kawakatsu T, Nery JR, Castanon R, Ecker JR. Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biol*. 2017;18:171.
37. Bouyer D, Kramdi A, Kassam M, Heese M, Schnittger A, Roudier F, et al. DNA methylation dynamics during early plant life. *Genome Biol*. 2017;18:179.
38. Lin J-Y, Le BH, Chen M, Henry KF, Hur J, Hsieh T-F, et al. Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. *Proc Natl Acad Sci U S A*. 2017;114:E9730–9.
39. Plotnikova A, Kellner MJ, Schon MA, Mosiolek M, Nodine MD. MicroRNA dynamics and functions during Arabidopsis embryogenesis. *Plant Cell*. 2019;31:2929–46.
40. Hofmann F, Schon MA, Nodine MD. The embryonic transcriptome of Arabidopsis thaliana. *Plant Reprod*. 2019;32:77–91.
41. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, et al. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*. 2008;133:116–27.
42. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J*. 2016;8:289–317.
43. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol*. 2014;21:64–72.
44. Cao X, Jacobsen SE. Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol*. 2002;12:1138–44.
45. Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*. 2006;443:1008–12.
46. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet*. 2009;41:630–4.
47. Bies-Etheve N, Pontier D, Lahmy S, Picart C, Vega D, Cooke R, et al. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep*. 2009;10:649–54.
48. Clark SJ, Smallwood SA, Lee HJ, Krueger F, Reik W, Kelsey G. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat Protoc*. 2017;12:534–47.
49. Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. Correction: Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife*. 2015;4. <https://doi.org/10.7554/eLife.08658>.
50. Kawakatsu T, Stuart T, Valdes M, Breakfield N, Schmitz RJ, Nery JR, et al. Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nat Plants*. 2016;2:16058.
51. Vandepoele K, Raes J, De Veylder L, Rouzé P, Rombauts S, Inzé D. Genome-wide analysis of core cell cycle genes in Arabidopsis. *Plant Cell*. 2002;14:903–16.
52. Lyons DB, Zilberman D. DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes. *eLife*. 2017. <https://doi.org/10.7554/eLife.30674>.
53. Gouil Q, Baulcombe DC. DNA methylation signatures of the plant chromomethyltransferases. *PLoS Genet*. 2016;12:e1006526.
54. Li X, Harris CJ, Zhong Z, Chen W, Liu R, Jia B, et al. Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proc Natl Acad Sci U S A*. 2018;115:E8793–802.
55. Wendte JM, Zhang Y, Ji L, Shi X, Hazarika RR, Shahryary Y, et al. Epimutations are associated with CHROMO-METHYLASE 3-induced de novo DNA methylation. *Elife*. 2019;8. <https://doi.org/10.7554/eLife.47891>.

56. Schoft VK, Chumak N, Mosiolek M, Slusarz L, Komnenovic V, Brownfield L, et al. Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. *EMBO Rep.* 2009;10:1015–21.
57. Zhao P, Zhou X, Shen K, Liu Z, Cheng T, Liu D, et al. Two-step maternal-to-zygotic transition with two-phase parental genome contributions. *Dev Cell.* 2019;49:882–93.e5.
58. Choi J, Lyons DB, Kim MY, Moore JD, Zilberman D. DNA methylation and histone H1 jointly repress transposable elements and aberrant intragenic transcripts. *Mol Cell.* 2019. <https://doi.org/10.1016/j.molcel.2019.10.011>.
59. Rutowicz K, Lirski M, Mermaz B, Teano G, Schubert J, Mestiri I, et al. Linker histones are fine-scale chromatin architects modulating developmental decisions in Arabidopsis. *Genome Biol.* 2019;20:157.
60. Russanova VR, Driscoll CT, Howard BH. Adenovirus type 2 preferentially stimulates polymerase III transcription of Alu elements by relieving repression: a potential role for chromatin. *Mol Cell Biol.* 1995;15:4282–90.
61. Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M, Kraus WL. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. *Science.* 2008;319:819–21.
62. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 1997;389:251–60.
63. Fyodorov DV, Zhou B-R, Skoultschi AI, Bai Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol.* 2018;19:192–206.
64. Yang D-L, Zhang G, Tang K, Li J, Yang L, Huang H, et al. Dicer-independent RNA-directed DNA methylation in Arabidopsis. *Cell Res.* 2016;26:1264.
65. Erdmann RM, Satyaki PRV, Klosinska M, Gehring M. A small RNA pathway mediates allelic dosage in endosperm. *Cell Rep.* 2017;21:3364–72.
66. Narsai R, Gouil Q, Secco D, Srivastava A, Karpievitch YV, Liew LC, et al. Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. *Genome Biol.* 2017;18:172.
67. Breakfield NW, Corcoran DL, Petricka JJ, Shen J, Sae-Seaw J, Rubio-Somoza I, et al. High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. *Genome Res.* 2012;22:163–76.
68. Xu L, Hu Y, Cao Y, Li J, Ma L, Li Y, et al. An expression atlas of miRNAs in Arabidopsis thaliana. *Sci China Life Sci.* 2018;61: 178–89.
69. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell.* 2008;133:523–36.
70. He S, Vickers M, Zhang J, Feng X. Natural depletion of histone H1 in sex cells causes DNA demethylation, heterochromatin decondensation and transposon activation. *eLife.* 2019. <https://doi.org/10.7554/elife.42530>.
71. Boissnard-Lorig C, Colon-Carmona A, Bauch M, Hodge S, Doerner P, Bancharel E, et al. Dynamic analyses of the expression of the HISTONE::YFP fusion protein in arabidopsis show that syncytial endosperm is divided in mitotic domains. *Plant Cell.* 2001;13:495–509.
72. Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, Claudio C, et al. A role for RNAi in the selective correction of DNA methylation defects. *Science.* 2009;1600–4. <https://doi.org/10.1126/science.1165313>.
73. Kao P, Nodine MD. Transcriptional activation of Arabidopsis zygotes is required for initial cell divisions. *Sci Rep.* 2019;9: 17159.
74. Bouvet P, Dimitrov S, Wolffe AP. Specific regulation of Xenopus chromosomal 5S rRNA gene transcription in vivo by histone H1. *Genes Dev.* 1994;8:1147–59.
75. Wolffe AP. Dominant and specific repression of Xenopus oocyte 5S RNA genes and satellite I DNA by histone H1. *EMBO J.* 1989;8:527–37.
76. Mérai Z, Chumak N, García-Aguilar M, Hsieh T-F, Nishimura T, Schoft VK, et al. The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes. *Proc Natl Acad Sci U S A.* 2014;111:16166–71.
77. Jenik PD, Jurkuta REJ, Barton MK. Interactions between the cell cycle and embryonic patterning in Arabidopsis uncovered by a mutation in DNA polymerase epsilon. *Plant Cell.* 2005;17:3362–77.
78. Sonnevile R, Craig G, Labib K, Gartner A, Blow JJ. Both chromosome decondensation and condensation are dependent on DNA replication in *C. elegans* embryos. *Cell Rep.* 2015;12:405–17.
79. Müller S, Almouzni G. Chromatin dynamics during the cell cycle at centromeres. *Nat Rev Genet.* 2017;18:192–208.
80. van Zanten M, Koini MA, Geyer R, Liu Y, Brambilla V, Bartels D, et al. Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. *Proc Natl Acad Sci U S A.* 2011;108: 20219–24.
81. Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten SO. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell.* 2010;22:3118–29.
82. Probst AV, Mittelsten SO. Stress-induced structural changes in plant chromatin. *Curr Opin Plant Biol.* 2015;27:8–16.
83. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature.* 2011;472:115–9.
84. Masuta Y, Nozawa K, Takagi H, Yaegashi H, Tanaka K, Ito T, et al. Inducible transposition of a heat-activated retrotransposon in tissue culture. *Plant Cell Physiol.* 2017;58:375–84.
85. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, et al. High-resolution mapping of open chromatin in the rice genome. *Genome Res.* 2012;22:151–62.
86. Lu X, Wontakal SN, Kavi H, Kim BJ, Guzzardo PM, Emelyanov AV, et al. Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9. *Sci Am Assoc Adv Sci.* 2013;340:78–81.
87. Bošković A, Eid A, Pontabry J, Ishiuchi T, Spiegelhalter C, Raghu Ram EVS, et al. Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev.* 2014;28:1042–7.
88. Becker JS, Nicetto D, Zaret KS. H3K9me3-dependent heterochromatin: barrier to cell fate changes. *Trends Genet.* 2016;32:29–41.

89. She W, Grimanelli D, Rutowicz K, Whitehead MWJ, Puzio M, Kotliński M, et al. Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants. *Development*. 2013;4008–19. <https://doi.org/10.1242/dev.095034>.
90. Rosa S, Ntoukakis V, Ohmido N, Pendle A, Abranches R, Shaw P. Cell differentiation and development in Arabidopsis are associated with changes in histone dynamics at the single-cell level. *Plant Cell*. 2014;26:4821–33.
91. Ikeuchi M, Favero DS, Sakamoto Y, Iwase A, Coleman D, Rymen B, et al. Molecular mechanisms of plant regeneration. *Ann Rev Plant Biol*. 2019:377–406. <https://doi.org/10.1146/annurev-arplant-050718-100434>.
92. Li H, Soriano M, Cordewener J, Muiño JM, Riksen T, Fukuoka H, et al. The histone deacetylase inhibitor trichostatin A promotes totipotency in the male gametophyte. *Plant Cell*. 2014;26:195–209.
93. Ikeda-Iwai M, Umehara M, Satoh S, Kamada H. Stress-induced somatic embryogenesis in vegetative tissues of *Arabidopsis thaliana*. *Plant J*. 2003;34:107–14.
94. Kamada H, Tachikawa Y, Saitou T, Harada H. Heat stress induction of carrot somatic embryogenesis. *Plant tissue culture letters*. Japanese Soc Plant Cell Mol Biol. 1994;11:229–32.
95. Li W, Liu H, Cheng ZJ, Su YH, Han HN, Zhang Y, et al. DNA methylation and histone modifications regulate de novo shoot regeneration in Arabidopsis by modulating WUSCHEL expression and auxin signaling. *PLoS Genet*. 2011;7:e1002243.
96. Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, et al. Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet*. 2006;38:721–5.
97. Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, Pikaard CS. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*. 2005;120:613–22.
98. Yelagandula R, Stroud H, Holec S, Zhou K, Feng S, Zhong X, et al. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. *Cell*. 2014;158:98–109.
99. Lutzmayer S, Enugutti B, Nodine MD. Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. *Sci Rep*. 2017;7:5913.
100. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011;17:10–2.
101. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res*. 2012;40:D1202–10.
102. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
103. Schon MA, Kellner MJ, Plotnikova A, Hofmann F, Nodine MD. NanoPARE: parallel analysis of RNA 5' ends from low-input RNA. *Genome Research*. 2018;28:1931–42.
104. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
105. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44:W160–5.
106. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27:1571–2.
107. Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*. 2015;523:212–6.
108. Stempor P, Ahringer J. SeqPlots - interactive software for exploratory data analyses, pattern discovery and visualization in genomics. *Wellcome Open Res*. 2016;1:14.
109. Chen F, Tillberg PW, Boyden ES. Expansion microscopy. *Science*. 2015:543–8. <https://doi.org/10.1126/science.1260088>.
110. Papareddy R, Nodine M. Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis. GSE152971. Gene Expression Omnibus. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152971>. Accessed 25 June 2020.
111. Papareddy R, Nodine M. Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis. Github. <https://github.com/mnodine/Papareddy.2020>. Accessed 26 June 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Additional file

Table of contents

Figure S1. Characteristics of embryonic 24-nt siRNAs and their similarities across samples

Figure S2. siRNA dynamics and characteristics

Figure S3. Benchmarking low-input methylC-seq, methylomes and cell-cycle transcripts

Figure S4. Genome browser screenshots of 24-nt siRNAs and DNA methylation levels

Figure S5. Size-based partitioning of heterochromatic TEs and small RNA-directed methylation

Figure S6. TE-derived siRNA accumulation and association with chromatin

Figure S7. TE-derived siRNAs in methylation mutants

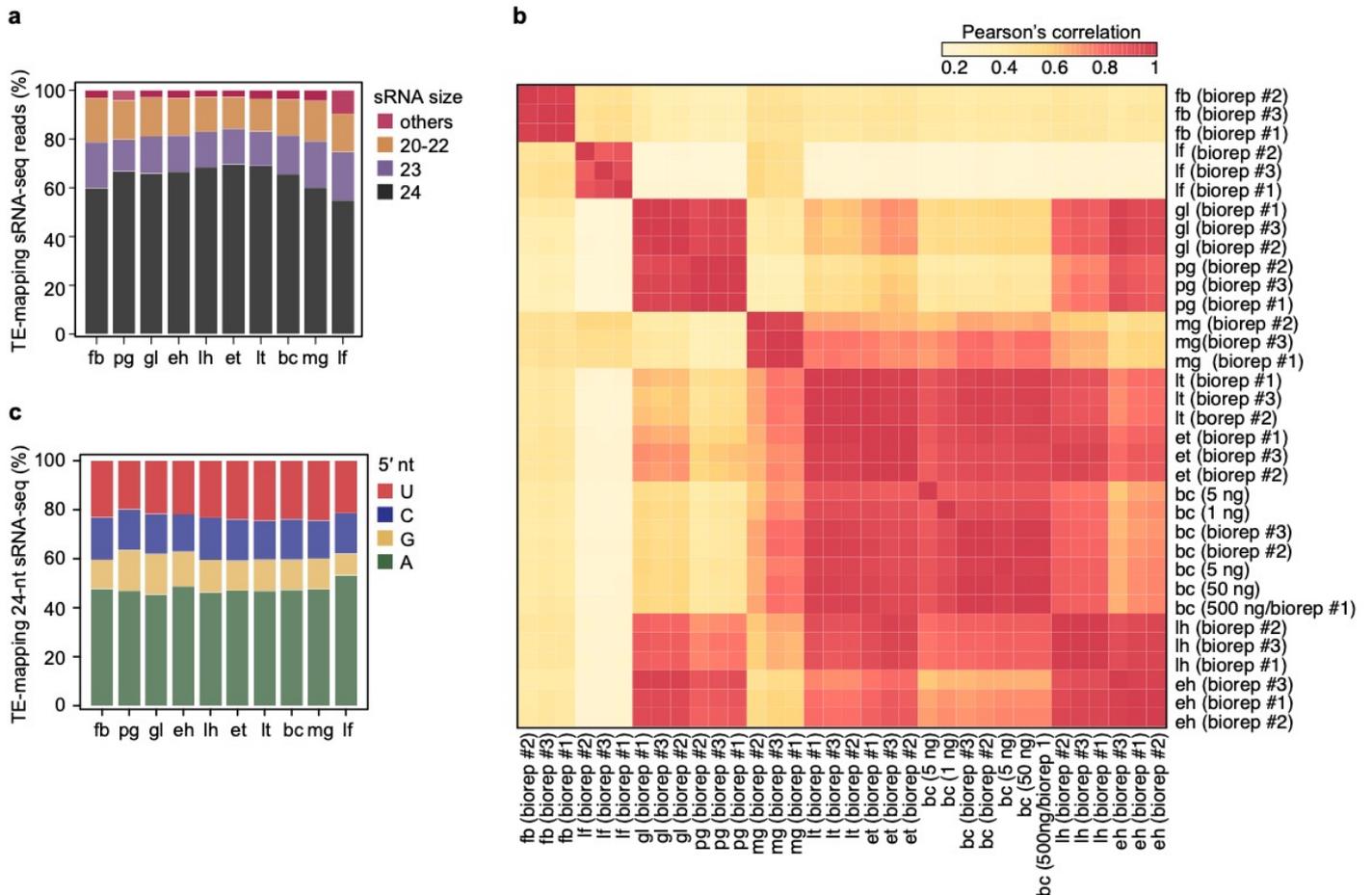


Figure S1. Characteristics of embryonic 24-nt siRNAs and their similarities across samples, related to Fig. 1. **a** Stacked bar chart depicting the percentage (%) of sRNA-seq reads mapping to TEs according to their sizes. Others include 18, 19 and 25-30 bases. **b** Heatmap of Pearson's correlation coefficients of 24-nt siRNAs among different biological replicates and tissue types. **c** Stacked bar chart showing the percentage (%) of 24-nt siRNAs with various 5'-most nucleotides in floral bud, embryonic and leaf tissues. Nucleotides are color-coded according to the key. Sample labels are as shown in Fig. 1a.

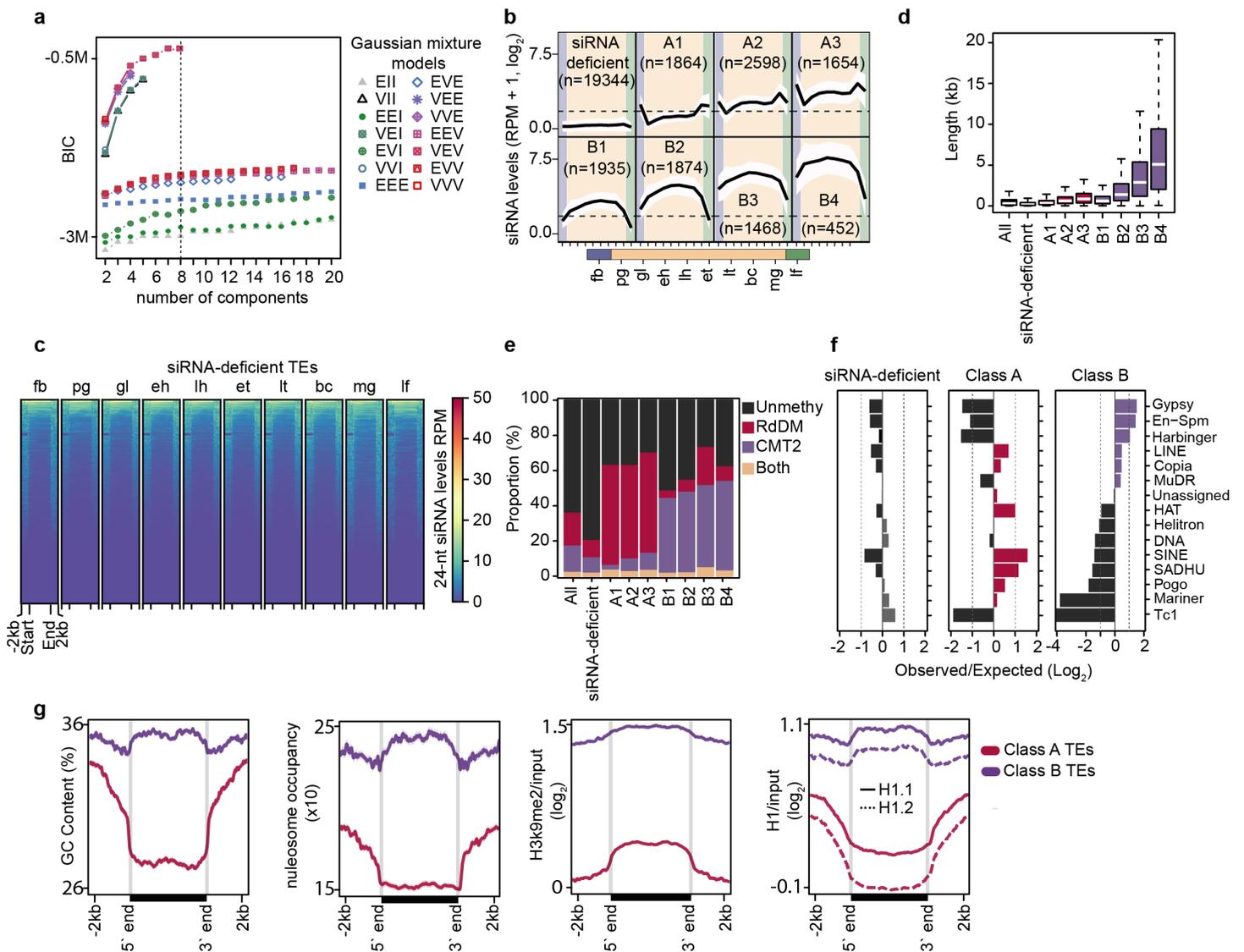


Figure S2. siRNA dynamics and characteristics, related to Fig. 2. **a** Bayesian information criterion (BIC) of various Gaussian mixture models of TE-derived 24-nt siRNAs as determined by *mClust*. EII, Equal volume, Equal shape, VII, Variable volume, Equal shape; EEI, Equal shape, Equal volume, coordinate axes orientation; VEI, Variable volume, Equal shape, coordinate axes orientation; EVI, Equal volume, Variable shape, coordinate axes orientation; VVI, Variable volume, Variable shape, coordinate axes orientation; EEE, Equal volume, Equal shape, Equal orientation; EVE, Equal volume, Variable shape, Equal orientation; VEE, Variable volume, Equal shape, Equal orientation; VVE, Variable volume, Variable shape, Equal orientation; EEV, Equal volume, Equal shape, Variable orientation; VEV, Variable volume, Equal shape, Variable orientation; EVV, Equal volume, Variable shape, Variable orientation; VVV, Variable volume, Variable shape, Variable orientation. **b** Line graphs illustrating 24-nt siRNA levels from siRNA-deficient TEs, as well as originally defined three class A and four class B TEs in embryonic and post-embryonic tissues. Dashed lines represent the detection criteria used to select TEs yielding siRNAs (2 RPM, reads per million genome-matching reads). The number of TEs belonging to each class are indicated. Polygons represent the standard deviation of mean 24-nt siRNA levels. fb, floral buds; pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green; lf, leaf. **c** Heat map showing 24-nt siRNA RPM from siRNA-deficient TEs across development. siRNA levels from ± 2 -kb of TEs are color-coded according to the key, and samples are labelled as in Fig. 1a. Rows were ordered based on total 24-nt siRNA levels. **d** Boxplot of TE lengths for *mClust*-defined groups

(*Figure S2, continued*) including all annotated, siRNA-deficient, class A and class B TEs. Thick horizontal bars indicate medians, and the top and bottom edges of the box indicate the 75th and 25th percentiles, respectively. kb, kilobases. **e** Stacked bar charts illustrating proportion of TE subclasses that are CHH hypomethylated in *drm1/drm2* (RdDM; red), *cmt2* (CMT2; purple), both *drm1/drm2* and *cmt2* (both; yellow) or which were not methylated (unmethylated; black) in leaves. **f** Enrichment of TE families observed in siRNA-deficient, class A or class B groups relative to their respective genomic backgrounds (observed/expected; log₂). **g** Metaplots of GC content, nucleosome occupancy (i.e. MNase-seq data) [62], and H3K9me2 and H1 levels [63] over class A and B TEs.

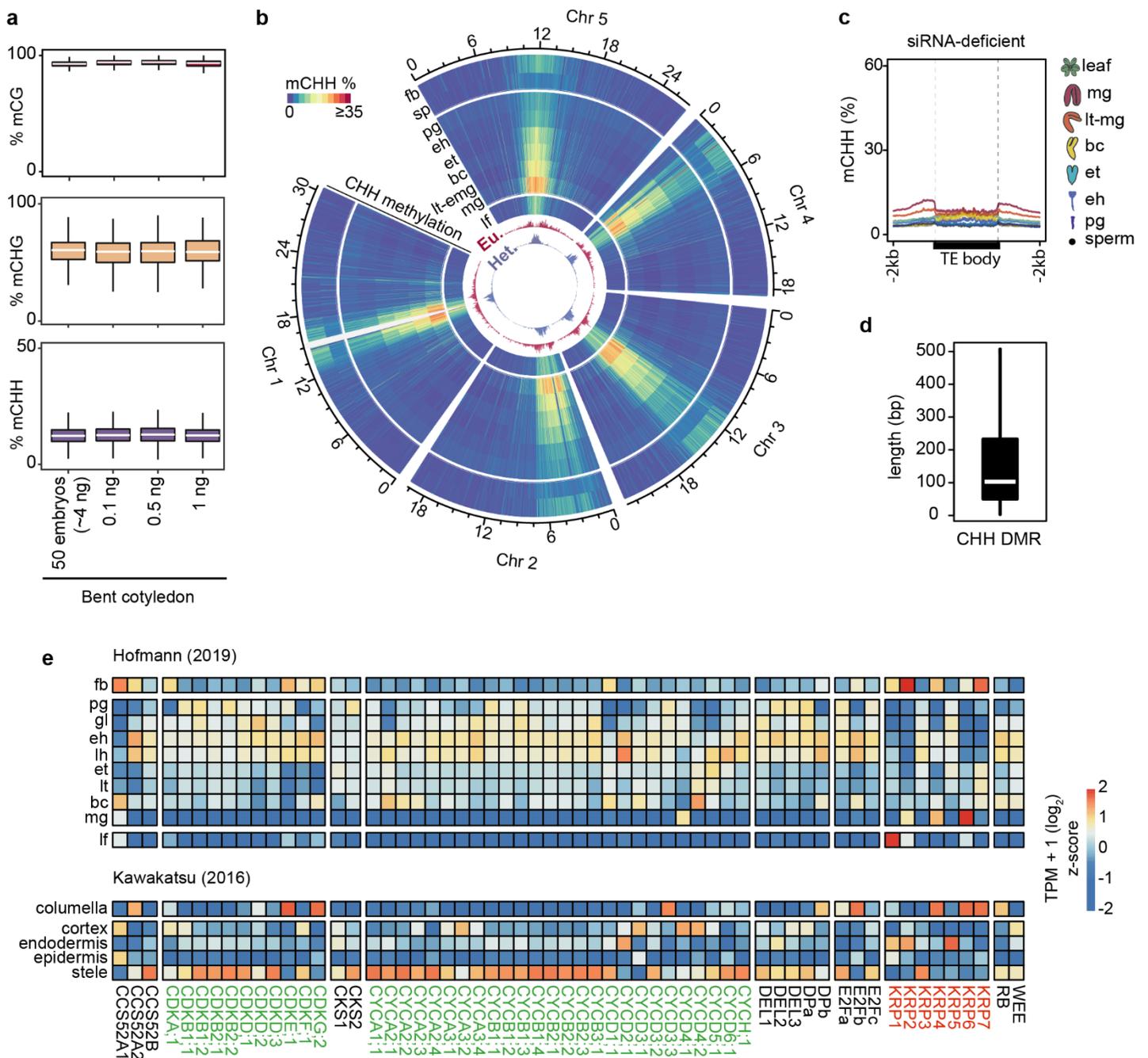


Figure S3. Benchmarking low-input methylC-seq, methylomes and cell-cycle transcripts, related to Fig. 3. **a** Box plots of average weighted CHH methylation percentages of TEs with at least 5 informative cytosines with $\geq 3\times$ coverage in either CG (*top*), CHG (*middle*) or CHH (*bottom*) contexts from various amounts of input genomic DNA. Labels on x-axis indicate whether these were from 50 bent cotyledon embryos (~ 4 ng), or a dilution series of genomic DNA from bent cotyledon embryos. Thick horizontal bars indicate medians, and the top and bottom edges of the box indicate the 75th and 25th percentiles, respectively. **b** Circos heat map representing average weighted CHH methylation of euchromatic (Eu.; red) or heterochromatic (Het.; purple) TEs in reproductive, embryonic and vegetative tissues across the five Arabidopsis chromosomes (Chr). Unmapped region in chromosome 1 (Chr 1) is shaded in white. fb, floral bud; sp, sperm; pg, preglobular; eh, early heart; lh, late heart; (Figure S3, continued) et, early torpedo; bc, bent cotyledon; lt-mg, late torpedo-to-early mature green; mg, mature green. **c** Metaplot of average weighted CHH methylation percentages of siRNA-deficient TEs sperm, embryos and leaves. Abbreviations are as in (b). **d** Box plot of CHH developmental DMR lengths, and as described in (a). **e** Heat maps of relative transcript levels from cell-cycle related genes across floral buds, embryos and leaves (*top*) or individual root cell types (*bottom*). Cell-cycle promoting and inhibitor gene labels are color-coded in green and red, respectively. Relative transcript levels are shown as z-scored TPMs according to the key.

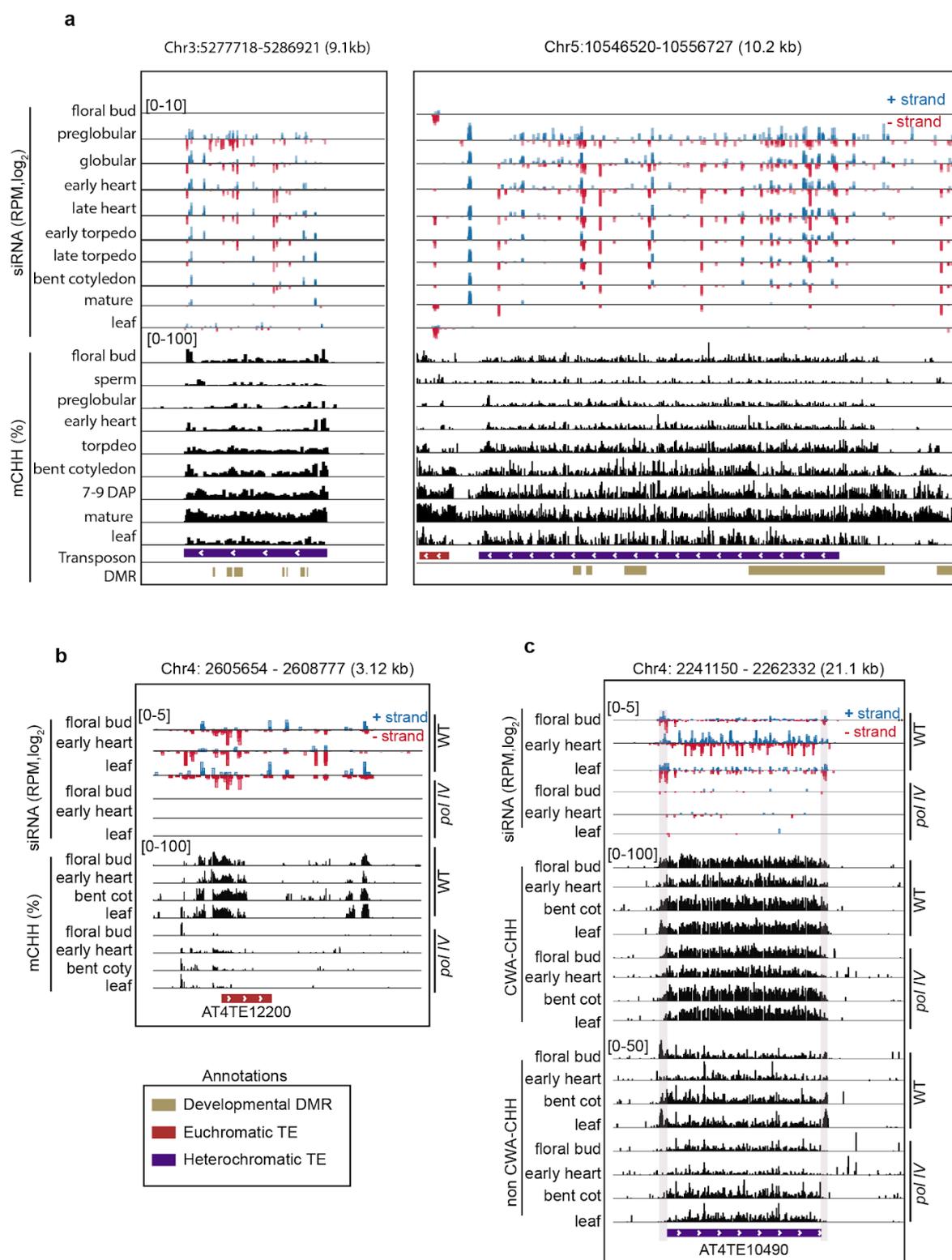


Figure S4. Genome browser screenshots of 24-nt siRNAs and DNA methylation levels, related to Figures 3 and 4. **a** Representative Integrative Genome Browser (IGV) screenshots of 24-nt siRNAs and weighted CHH methylation on developmental Differentially Methylated Regions (DMRs). **b** Representative IGV screenshot of 24-nt siRNAs and weighted CHH methylation in wild type and Pol IV mutants in floral bud, embryos and leaf (*Figure S4, continued*) tissues on euchromatic TEs (red) **c** Twenty-four nucleotide siRNAs and weighted CWA and nonCWA-CHH methylation in wild type and Pol IV mutants in floral bud, embryos and leaf tissues on heterochromatic TEs (navy). Edges of the heterochromatic TE are highlighted with background shading.

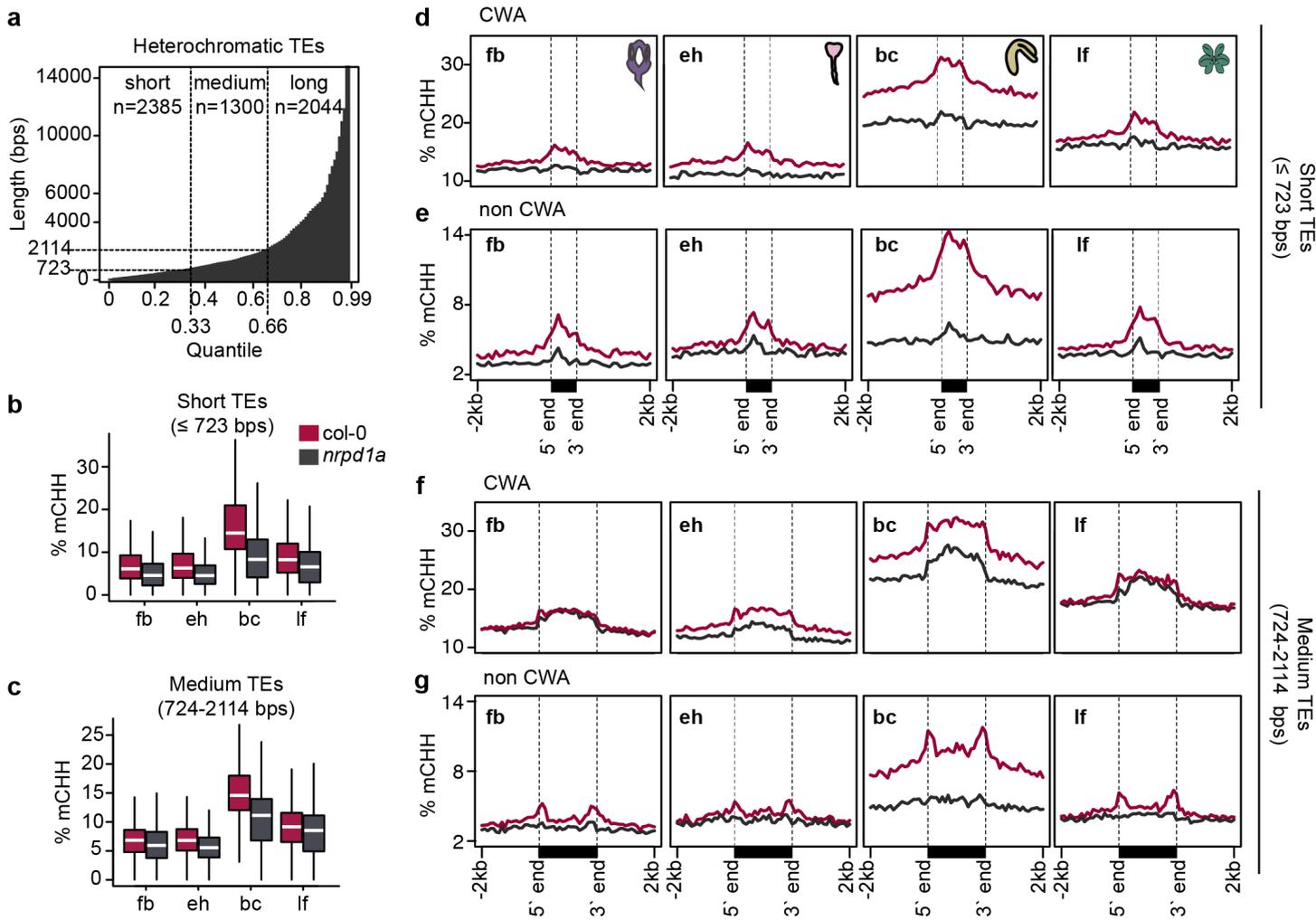


Figure S5. Size-based partitioning of heterochromatic TEs and small RNA-directed methylation, related to Fig. 4.

a Histogram of heterochromatic TE lengths across quantiles, which were used to separate TEs into short, medium or long classes. Vertical and horizontal dashed lines represent quantiles and sizes used for classifications. **b** and **c** Boxplots of CHH methylation levels in wild type and *nrpd1a* mutant floral buds (fb), early heart embryos (eh), bent cotyledon embryos (bc) and leaves (lf) for short (**b**) and medium (**c**) heterochromatic TEs. **d-f** Meta-plots of average weighted CHH methylation percentages of short (**d** and **e**) and medium (**f** and **g**) (*Figure S5, continued*) heterochromatic TEs in either CWA (**d** and **f**) or non-CWA (**e** and **g**) context.

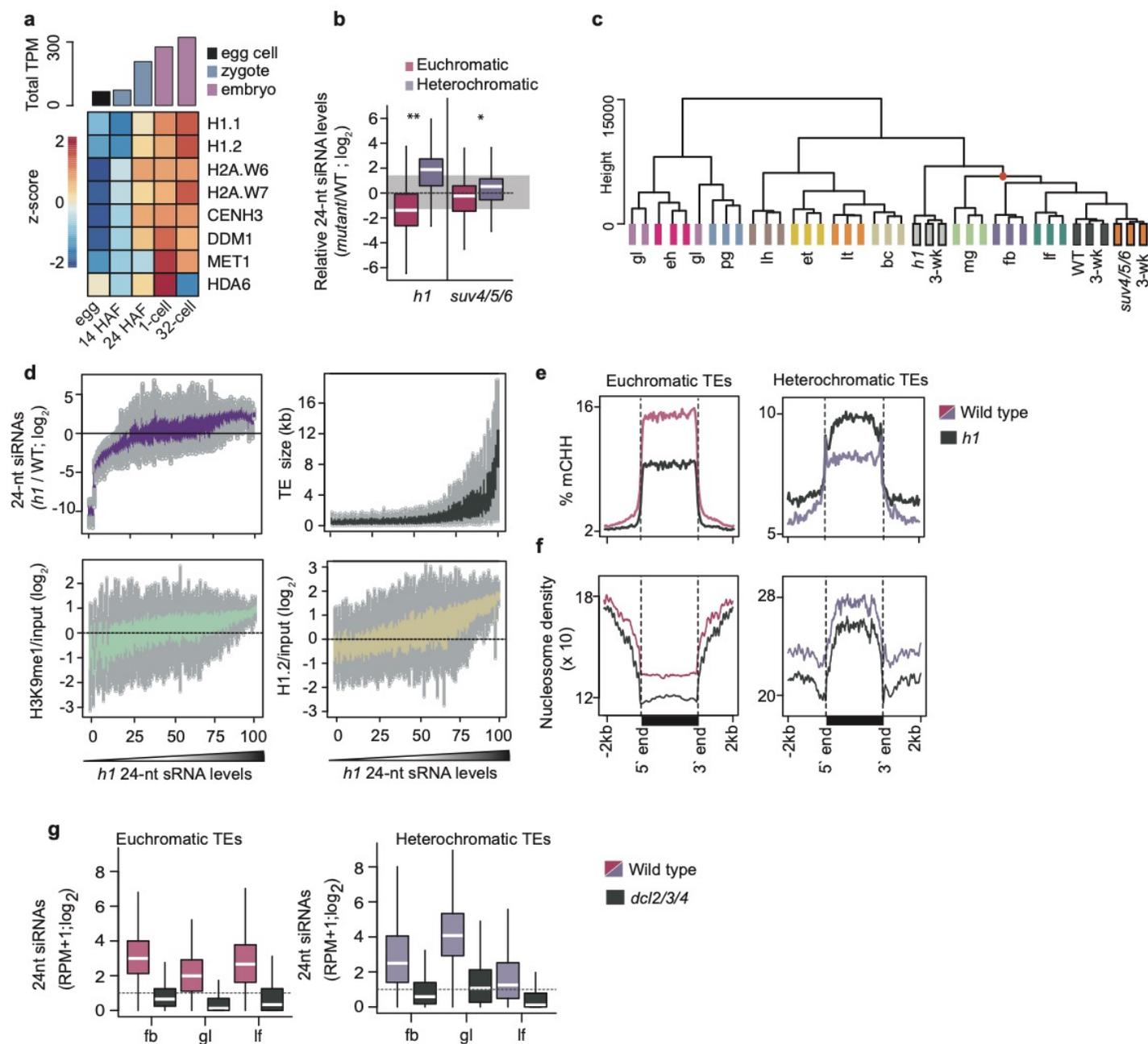


Figure S6. TE-derived siRNA accumulation and association with chromatin, related to Fig 5. **a** Bar chart depicting total (*top*) and heat map of individual (*bottom*) transcript levels from genes involved in heterochromatin formation in eggs, zygotes 14 or 24 hours after fertilization (HAF), and embryos at 1-cell and ~32-cell stages. **b** Boxplots of relative 24-nt siRNA levels overlapping euchromatic (red) or heterochromatic (purple) TEs in *h1* or *suv4/5/6* mutants relative to wild-type leaves. P values < 0.001 and < 0.0001 based on Mann-Whitney U test of 24-nt siRNA differences between wild-type and mutant tissues are represented by * and **, respectively. **c** Dendrogram based on hierarchical clustering of Euclidean distances of TE-derived 24-nt siRNAs from embryonic and post-embryonic tissues. pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green; lf, leaf; fb, floral bud; 3-wk; 3 week old leaves; WT, wild type. Branchpoint of embryonic and non-embryonic tissues is indicated by a red dot. **d** TEs were divided into percentiles, ordered based on their 24-nt siRNA levels in *h1* mutants as in Fig. 5e-h), and plotted according to 24-nt siRNA levels in (*Figure S6, continued*) *h1* relative to wild type, TE length, and enrichment of H1.2 or H3K9me2. **e** and **f** Metaplots of average weighted CHH methylation percentages (**e**) and nucleosome occupancy (**f**) of euchromatic (*left*) and heterochromatic (*right*) TEs. **g** Boxplots of 24-nt siRNA levels of euchromatic (*top*) and heterochromatic (*bottom*) TEs in wild type and *dcl2/3/4* floral buds (fb), globular embryos (gl) and leaves (lf).

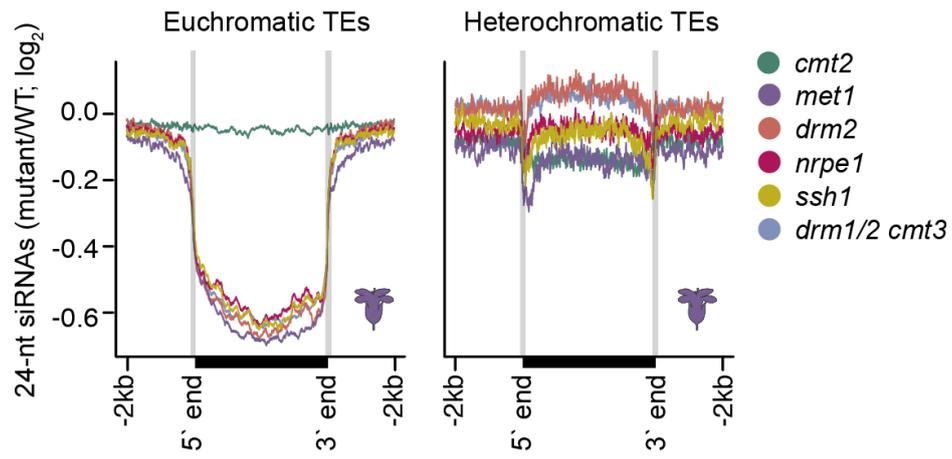


Figure S7. TE-derived siRNAs in methylation mutants, related to Fig. 6. Metaplots of 24-nt siRNA levels overlapping euchromatic (*left*) and heterochromatic (*right*) TEs in various RdDM and related mutants. Annotated TE 5' and 3' ends are labelled at the bottom.

8 Gene Expression Variation Across Arabidopsis Embryos at Single-Nucleus Resolution

Type	<i>in preparation</i>
Authors	Ping Kao, Michael A. Schon, Magdalena Mosiolek & Michael D. Nodine
Contributions	P.K., M.A.S. and M.D.N. conceived the project; P.K. and M.A.S. developed the methodology; P.K. and M.M. performed the experiments, prepared figures, and wrote and edited the article; M.D.N. acquired funding and supervised the project.

Title

Gene Expression Variation Across Arabidopsis Embryos at Single-Nucleus Resolution

Authors

Ping Kao, Michael A. Schon, Magdalena Mosiolek & Michael D. Nodine

Contributions:

P.K., M.A.S. and M.D.N. conceived the project; P.K. and M.A.S. developed the methodology; P.K. and M.M. performed the experiments, prepared figures, and wrote and edited the article; M.D.N. acquired funding and supervised the project.

SUMMARY

Characterizing gene expression programs in individual cell types is crucial to understand the pattern formation in Arabidopsis embryos. While the transcriptomes from various developmental stages yield valuable insights for the regulatory events during different embryonic phases, the knowledge on the gene expression diversities among embryonic cell types is limited. In this article, we utilized the single nucleus RNA-seq to generate non-contaminated embryonic transcriptomes at single cell resolution and presented the representative expression profiles for embryonic cell types at the globular stage. The expression patterns of epigenetic regulators indicated that zygotic small RNA dependent and independent pathways establish distinct chromatin environments in the initial embryonic cell lineages. Transcription factor binding site analyses revealed potential activators or inhibitors for the genes preferentially expressed in different cell types. We presented a comprehensive set of cell type-specific transcriptomes which would serve as a robust reference toward characterizing gene regulations for pattern formation and cell type specifications.

INTRODUCTION

Metazoans and land plants establish their body plans during embryogenesis (Gerri et al., 2020; Park and Harada, 2008). Gene-regulatory mechanisms have evolved and been elaborated upon independently in these two major eukaryotic lineages to generate the immense morphological diversity observed in multicellular organisms (Bai, 2015; Clark et al., 2006). For example, in animals it has been long recognized that maternal gene products control initial pattern formation before the transition of control from the maternal to the zygotic genome (Lee et al., 2014; Tadros and Lipshitz, 2009; Walser and Lipshitz, 2011). By contrast, transcriptional activation of the zygotic genome soon after fertilization is necessary for zygote elongation and initial divisions in *Nicotiana tabacum* (tobacco) (Ning et al., 2006; Zhao et al., 2011) and the model flowering plant *Arabidopsis thaliana* (Arabidopsis) (Kao and Nodine, 2019; Zhao et al., 2019). In addition, the vast majority of genes regulating Arabidopsis embryo morphogenesis are zygotically expressed (Nodine and Bartel, 2012; Zhao et al., 2019) and required (Meinke, 2019). Therefore, genes are expressed from the zygotic genome during initial stages of embryo development and the diversification of gene expression programs across plant embryonic cell-types contributes to the formation of the basic body plan. Characterizing how gene expression programs are established in individual cell types of early embryos is critical to understand the molecular basis of pattern formation in plant embryos, and more broadly the general and unique principles of embryonic patterning in multicellular organisms.

Forward genetic screens successfully identified many genes required for proper plant embryogenesis (Lukowitz et al., 2004; Mayer et al., 1998; Meinke, 2019), but relatively few mutations in genes encoding cell-specific transcriptional regulators were recovered. This is at least partially due to the high degree of genetic redundancy among plant transcription factors that typically belong to multigene families (Riechmann, 2002). Alternatively, RNA populations can be characterized to infer gene-regulatory processes underlying cellular differentiation events. Transcriptomes generated from early embryos at various stages of development have accordingly yielded insights into the biological processes operating during different embryonic phases (Hofmann et al., 2019; Zhao et al., 2019). However, these transcriptomes were generated from whole embryos. Additional studies have revealed genes that are preferentially expressed in broad (Belmonte et al., 2013; Casson et al., 2005; Slane et al., 2014; Zhou et al., 2020) or more specific regions (Palovaara et al., 2017) of plant embryos, but either lack cellular resolution or were contaminated with RNAs derived from the maternal seed coat that encompasses the developing embryo (Schon and Nodine, 2017).

Single-cell mRNA-seq (scRNA-seq) has been instrumental towards understanding developmental events at cellular resolution over the past decade (Chen et al., 2019; Hwang et al., 2018). Several studies have applied these approaches to plant tissues (Brennecke et al., 2013; Efroni and Birnbaum, 2016; Ryu et al., 2019; Satterlee et al., 2020; Shulse et al., 2019; Song et al., 2020; Tang and Tang, 2019; Zhang et al., 2019), but scRNA-seq has yet to be reported for plant embryos. This is primarily due to the presence of rigid cell walls that hold plant cells together. Although cell walls can be removed by enzymatic treatment of tissues that are easy to access, such protoplasting techniques remain impractical for early embryos because they are deeply embedded within maternal seed tissues. Single-nucleus mRNA-seq (snRNA-seq) (Habib et al., 2016) offers an alternative method to inspect transcriptomes at single-cell resolution in plants and has been recently applied to endosperm tissues within seeds (Picard et al., 2020). Here, we present a workflow to obtain contamination-free high-quality transcriptomes from individual early embryonic nuclei followed by their assignment to the respective cell types. Remarkably, these initial cell types already preferentially express sets of genes that are characteristic of their post-embryonic derivatives, have different evolutionary trajectories and appear to be regulated by distinct epigenetic and transcriptional mechanisms.

RESULTS

Acquisition of contamination-free transcriptomes from individual embryonic nuclei

To acquire single-cell transcriptomes of early embryos, we utilized fluorescence-activated nuclei sorting (FANS) coupled with single-nucleus RNA-seq (snRNA-seq) (Fig. 1A,B). More specifically, we used a transgenic line expressing nuclear-localized green fluorescent protein under the control of the embryo-specific *WUSCHEL-RELATED HOMEBOX 2* promoter (pWOX2::NLS-GFP) to fluorescently label nuclei in embryos but not the surrounding endosperm or maternal tissues (Fig. 1A) (Gooh et al., 2015). Briefly, we fixed siliques or seeds containing globular stage embryos with a low concentration of dithiobis(succinimidyl propionate) (DSP) before nuclei isolation to preserve RNA. Nuclei were also stained with 4',6-diamidino-2-phenylindole (DAPI) and intact nuclei were selected based on DAPI profiles (Fig. S1A). Embryonic nuclei were then isolated based on their strong GFP signal and sorted individually into 96-well plates. Fixed nuclei were decrosslinked with dithiothreitol (DTT) to enable the generation of cDNA and subsequent next-generation sequencing (NGS) compatible libraries from single-nuclei using the Smart-seq2 protocol (Picelli et al., 2014a; Picelli et al., 2014b). The resulting NGS libraries were sequenced on an Illumina HiSeq 2500 followed by the alignment of NGS reads to the Araport11 transcriptome (Cheng et al., 2017) and transcript quantification by kallisto (Bray et al., 2016) (Fig. 1B). After quality controls (see Methods), 534 out of 944 (72%) nuclei were retained for further analyses. A total of 24,591 genes were detected from all nuclei with an average of 440,289 aligned reads and 2,576 detected genes per snRNA-seq library (Fig. S1B, Supplementary Data1). Therefore, our approach allowed us to acquire high-quality RNA-seq libraries from hundreds of individual embryonic nuclei.

Contamination of early embryonic mRNA-seq datasets with RNAs from surrounding maternal seed tissues has been a major limitation to seed transcriptomics (Schon and Nodine, 2017). To evaluate the level of maternal contamination in individual snRNA-seq libraries, we applied the tissue enrichment test (Schon and Nodine, 2017). Although we used stringent selection criteria during FANS, 20-50% of the snRNA-seq libraries per plate were still significantly enriched for either seed coat or endosperm transcripts while remaining libraries were enriched for embryonic transcripts or had ambiguous identities (Fig. S2A). To systematically identify contaminated snRNA-seq libraries, we conducted unsupervised clustering on all libraries and labeled them according to their tissue enrichment scores (Fig. S2B). Because clusters 12 and 13 were enriched for libraries with seed coat labels, we excluded them from subsequent analyses (Fig. S2B). To further evaluate how well we could remove non-embryonic nuclei, we combined the expression levels of snRNA-seq libraries from each plate and performed tissue enrichment tests (Fig. 1C). Retained and excluded nuclei were enriched for embryonic and seed coat cell types, respectively. Moreover, transcriptomes from the retained nuclei were more similar to published embryonic transcriptomes (Hofmann et al., 2019; Nodine and Bartel, 2012) than those from discarded nuclei (Fig. S2C). Altogether, our stringent criteria allowed us to successfully remove non-embryonic snRNA-seq libraries and obtain 486 high-quality snRNA-seq libraries from embryonic cells.

Identification of embryonic cell types

The precursors of the most fundamental plant tissues are already established along apical-basal and radial axes of the embryo by the globular stage (Park and Harada, 2008; ten Hove et al., 2015) (Fig. 2A). After the asymmetric division of the zygote, the smaller apical cell will generate the apical-cell lineage (ACL) while the larger basal cell will produce the basal-cell lineage (BCL). The suspensor (sus) of the BCL connects the embryo proper with maternal tissues and provides support, but undergoes programmed cell death and does not contribute to post-embryonic tissues. The apical-most cells of the BCL include the quiescent center initials (qc) and columella initials (col) which will become incorporated into the embryo proper at the globular stage and are the precursors of the corresponding cell types in the root meristem. Unlike the BCL, the ACL undergoes divisions along the radial axis of the developing embryo to form concentric tissue layers. The outermost protoderm (upper, upd; lower, lpd), middle ground tissue initials (grd) and innermost vascular initials (vas) produce the epidermal, ground and vascular tissues of the root, respectively. The shoot meristem initials (smi) will produce all aerial tissues after germination while the upd and cells within the upper inner periphery (uip) of embryos contribute to the cotyledons that support the seedling. Because these tissue precursors are present in globular embryos, we developed a method to assign each snRNA-seq cluster to the most likely cell type. Namely, we identified 174 reference genes expressed in embryos from the literature and recorded their expression patterns as either strongly, weakly or not expressed in the nine cell types found in globular embryos listed above (Supplementary TableS1). We reasoned that the overrepresentation of cell-specific reference transcripts should indicate how likely each nucleus was from one of these cell types, and thus calculated a cell-type score for each snRNA-seq dataset using hypergeometric tests (see Methods). After recording cell-type scores for each of the 486 embryonic nuclei, we performed uniform manifold approximation and projection (UMAP) clustering to resolve their identities. In contrast to unsupervised clustering, which was able to distinguish embryo proper and suspensor nuclei, but not individual cell types (Fig. S2C), the reference-guided clustering identified 12 clusters that were each enriched for

a specific cell type (Fig. 2A, B). We also identified one cluster (cluster 4) that was enriched for multiple cell types and substantially less genes detected per nuclei compared to the other clusters. We discarded the snRNA-seq libraries belonging to this cluster from subsequent analyses because of their poor quality, which may be due to being generated from multiple or damaged nuclei. Cell-specific reference transcripts tended to co-localize to the same cluster (Figs. 2C, S3B). For example, WUSCHEL-RELATED HOMEBOX 5 (WOX5) and JACK-DAW (JKD) transcripts are highly enriched in the qc (Haecker et al., 2004; Shimotohno et al., 2018; Welch et al., 2007) and co-localize to cluster 12. Therefore, by highlighting the differences among cell types based on a reference gene set we were able to infer which cell type each snRNA-seq cluster represented.

Although snRNA-seq libraries can generate insights into the transcriptome variations among individual cells, the information provided can be complex and difficult to interpret. To provide a concise and uniform parameter to examine gene expression patterns, we calculated “enrichment scores” in each of the twelve clusters for the 13,893 transcripts detected in $\geq 10\%$ of nuclei in ≥ 1 cluster. Enrichment scores are a combination of the deviations of a transcript’s mean levels and the percentage of nuclei it was detected in for each cluster relative to the other 11 clusters (see Methods), and thus concisely summarize the relative abundance of each transcript in each cluster. The 250 genes with the highest enrichment scores (top-ranked 250) from each cluster were considered to be preferentially enriched for that cluster (Supplementary Data 2). For example, 74 of the 118 (62.7%) reference transcripts detected in $\geq 10\%$ nuclei in ≥ 1 cluster had enrichment scores within the top-ranked 250 genes in at least one cluster, and four were the top-ranked gene (*PINI* in cluster 1, *KANI* in cluster 2, *WOX5* in cluster 12 and *WOX8* in cluster 13).

We compared the transcriptomes of each cell cluster with published transcriptomes from the embryo proper and suspensor regions of globular embryos (Belmonte et al., 2013) as an initial test of our marker-based cell type assignments. In agreement with the marker-based assignments, clusters 8, 10 and 13 were exclusively enriched for suspensor transcripts based on tissue enrichment tests (Fig. 2D). Also consistent with the cell type assignments, clusters 1, 2, 5, 6, 7 and 11 were enriched for only embryo proper transcripts. Cluster 9 had mixed cell type assignments and accordingly was enriched for both embryo proper and suspensor transcripts. Most of the nuclei in clusters 3 and 12 were respectively labelled as col and qc cell types, which are situated between the suspensor and embryo proper. Whereas cluster 3 was only enriched for suspensor transcripts, cluster 12 was enriched for both suspensor and embryo proper transcripts. We confirmed these results with another published transcriptome dataset generated from embryo proper and suspensors of globular embryos (Zhou et al., 2020) (Fig. 2D).

We then used RNA in situ hybridization (ISH) to further evaluate the marker-based assignments of snRNA-seq clusters to individual cell types. We selected 33 genes without reported expression patterns from the 250 most highly enriched genes of each cluster for ISH (Fig. S4). The expression patterns (Fig. 2E) and enrichment scores (Fig. S3C) indicated that the candidates represented a specific cell or group of cells. We could detect ISH signal in at least 50% of embryos for 26 of the 33 probes tested (78.8%) and compared the ISH and snRNA-seq expression patterns for these in more detail (Fig. 2E and Fig. S4). AT3G13690, AT4G29020 and SERINE CARBOXYPEPTIDASE-LIKE 25 (SCPL25) transcripts were expressed at high levels in clusters 2, 5 and 6, and detected by ISH almost exclusively in the protoderm. AT5G01870 was also highly expressed in clusters 2, 5 and 6, as well as cluster 12, and was detected in the protoderm and columella initial; whereas AT1G044880 was enriched in clusters 6 and 9, and detected in the upper protoderm. AT1G80133, AT2G42660, AT3G54780 and GH3.2 were highly expressed in clusters 3 and 12, and ISH signals were detected in the columella, quiescent center, as well as throughout the suspensors for GH3.2. LITTLE ZIPPER 4 (ZPR4) was highly enriched in cluster 7 based on snRNA-seq and detected in the shoot meristem initials by ISH. RESPONSIVE TO DEHYDRATION 19 (RD19) transcripts were also detected by ISH in shoot meristem initials, but were moderately expressed in all clusters. AT4G38370 transcripts were also expressed throughout the clusters, albeit strongly enriched in cluster 8, but ISH was stronger in the embryo proper. These two apparent discrepancies may also be due to cell-specific stabilization of these transcripts after nuclear export. ALCATRAZ-INTERACTING PROTEIN 1 (ACI1), AT3G15680 and AT3G15720 transcripts were expressed most highly in cluster 11 and detected in vascular initials with ISH. ROOT MERISTEM GROWTH FACTOR 8 (RGF8) was highly expressed in clusters 11 and 12, and RGF8 transcripts were detected in vascular and columella initials. AT5G61412, BETA GLUCOSIDASE 17 (BGLU17), COBRA-LIKE PROTEIN 6 PRECURSOR (COBL6), CYSTEINE ENDOPEPTIDASE 1 (CEP1), EARLY NODULIN-LIKE PROTEIN 2 (ENODL2), MAJOR LATEX PROTEIN 28 (MLP28) and SPERMIDINE DISINAPOYL ACYLTRANSFERASE (SDT) transcripts were within the top 250-ranked genes of clusters 8, 10 or 13 based on enrichment scores, and all were detected in suspensors by ISH. AT3G13230, MITOTIC ARREST-DEFICIENT 2 (MAD2) and TARGETING PROTEIN FOR XKLP2 (TPX2) were highly expressed in clusters 1, 5 and/or 9, and ‘salt-and-pepper’ ISH patterns were detected. Such patterns are indicative of cell-cycle regulated genes, and accordingly we observed that clusters 1, 5 and 9 were enriched for mitotic (M) phase regulated transcripts (Menges et al., 2003) (Fig. S3E). Transcripts enriched in clusters 1, 5 and 9 also tended to be localized to the subprotoderm, protoderm or both layers, respectively. Therefore, we predict that cluster 1 is dividing subprotoderm (dividing inner; div.i), cluster 5 is dividing protoderm (dividing outer; div.o) and cluster 9 is dividing cells in general (div) (Fig. 2G). Altogether, our in silico and in situ validations indicate that we have identified major cell types present in globular embryos: the suspensor (sus1, cluster

10; *sus2*, cluster 8; *sus3*, cluster 13); columella initials (*col*; cluster 3), quiescent center initials (*qc*; cluster 12); vascular initials (*vas*; cluster 11); shoot meristem initials (*smi*; cluster 7); and the lower and upper protoderm (*lpd*, cluster 2; *upd*, cluster 6) (Fig. 2G).

General characteristics of transcriptomes from embryonic cell-types

To gain insights into which biological processes are enriched in each embryonic cell-type, we then conducted gene ontology (GO) term enrichment analyses on the top-250 ranked genes of each cluster (Fig. 3A and Supplementary TableS2). Significantly enriched GO-terms were identified for the top-ranked 250 genes in the *div*, *vas*, *div.i*, *smi*, *lpd*, *div.o* and *upd* clusters, but not *sus1/2/3*, *col* or *qc* clusters. The inability to detect enriched terms in the BCL (i.e. *sus1/2/3*, *col* and *qc*) is likely due to the limited annotation of genes specifically expressed in these cell types. Consistent with the *div*, *div.i* and *div.o* clusters representing actively dividing cells, GO terms related to progression through mitotic phases (*div* and *div.o*) and microtubules (*div.i* and *div.o*) were enriched. GO-terms related to body axis specification were also enriched in the top-250 ranked genes of the *div.i* cluster, as well as the *vas* cluster. The *lpd* and *upd* protodermal clusters were both enriched for specification of axis polarity and cutin biosynthesis terms within their top-250 ranked genes. Moreover, the top-250 ranked genes of the *lpd* and *upd* clusters could be distinguished from each other by their over-representation of epidermal and cotyledon development GO terms, respectively. The top-250 ranked genes of the *smi* cluster was enriched for DNA replication processes including pre-replicative complex assembly, which is consistent with the *smi* cluster being depleted for mitosis-phase markers (Fig. S3D). Overall, the enriched GO terms were consistent with the assigned cluster identities (Fig. 2F) and indicate that we have classified embryonic cell types with distinct functions.

Next, we tested whether genes essential for embryogenesis are preferentially enriched within the top-250 ranked genes of each cluster. EMBYRO-DEFECTIVE (EMB) genes are a set of genes required for normal embryo development in Arabidopsis (Meinke, 2019). EMB genes were enriched in the top-250 genes of the ACL clusters including the *smi*, *div.o* and *upd* clusters which were significantly enriched. In contrast, EMB genes were depleted from top-250 genes of the BCL clusters, which includes the significantly depleted *col* and *sus2* (Fig. 3B). Further supporting that genes preferentially expressed in the ACL are more likely to be required for proper development than those in the BCL, we found that the top-250 ranked genes within the ACL, and especially the *div*, *vas* and *lpd* clusters, were more highly conserved across Brassicacea species and land plants in general compared to BCL clusters (Haudry et al., 2013; Tian et al., 2020) (Figs. 3C and S5A, B). Also consistent with the EMB analyses, the top-250 ranked genes within the BCL clusters were more poorly conserved especially genes enriched in the *sus1*, *sus2*, *col* and *qc* clusters. Altogether, these results suggested that genes preferentially expressed in ACL clusters, and especially the *smi*, *vas* and *div* clusters are under stronger purifying selection than those in BCL clusters especially the *col* cluster, which are mutating at a faster rate. This is consistent with the more variable morphologies of suspensors relative to embryo proper in plants (Chen et al., 2020). Because the DEMETER-LIKE1/2 (DML1/2) transcripts encoding mutagenic DNA glycosylases are among the most poorly conserved genes in our dataset and expressed in the BCL, it is also possible that the higher mutation rates in genes preferentially enriched in the BCL are partially related to the distinct epigenetic processes that occur in this lineage.

Transcripts encoding epigenetic regulators vary across embryonic cell types

Soon after fertilization of egg and sperm, epigenetic states are reprogrammed in the new generation (Gehring, 2019). This includes replacement of histones, as well as re-establishment of DNA methylation landscapes genome-wide by small RNA dependent and independent pathways (Bouyer et al., 2017; Ingouff et al., 2010; Jullien et al., 2012; Nagasaki et al., 2007; Papareddy et al., 2020). Because such differential chromatin states can strongly influence gene expression, we examined the transcript levels of genes previously implicated in chromatin regulation. More specifically, we found that 50/191 genes involved in general chromatin features, histone modifications (i.e. acetylation, methylation and ubiquitination), polycomb repressive complexes, DNA methylation or demethylation or small RNA production or activities had enrichment scores ≥ 2.5 in ≥ 1 embryonic cell cluster (Erdmann and Picard, 2020; Pikaard and Mittelsten Scheid, 2014) (Fig. 4A). General chromatin factors and components of the polycomb repressive complex tended to vary between the embryo proper and suspensor. The HAC1 histone acetyltransferase was enriched in the suspensor clusters while the HDA3/4 histone deacetylases were enriched in the embryo proper. Moreover, the JM16/27/29 and JM22 histone demethylases were respectively enriched in the suspensor and embryo proper. Interestingly, the terminally differentiated suspensor was enriched for transcripts encoding proteins required for the production of 24-nt small interfering RNAs (siRNAs) such as CLSY1, NRPD1A and DCL3. By contrast, genes encoding ARGONAUTE (AGO; AGO1/5/8/9/10) proteins, which bind to small RNAs and mediate their gene repressive activities, were enriched in the precursors to the shoot meristem initials. The enrichment of AGOs in shoot meristem initials is supported by previous reports (Gutzat et al., 2020; Jullien et al., 2012; Tucker et al., 2008) and could indicate that the precursors to aerial tissues of the plant, including the gametes, are enriched for small RNA-mediated surveillance pathways that prevent transposon mobilization and other genome

de-stabilizing events. Altogether, these results indicate that zygotic small RNA dependent and independent pathways establish distinct chromatin environments in the initial embryonic cell lineages.

The most striking cell-specific enrichments were in pathways affecting cytosine methylation, which is typically associated with transcriptional silencing of transposons and repression of gene promoters (Law and Jacobsen, 2010). CMT3 and MET1 encode DNA methyltransferases that maintain cytosine methylation in the CHG (H ≠ G) and CG contexts, respectively, and both were enriched in the embryo proper. In stark contrast, transcripts encoding the ROS1, DML2 and DML3 DNA glycosylases required for the removal of methylated cytosines were highly enriched in the basal cell lineage (BCL) including the suspensor, columella and quiescent center initials. Recently, 275 genes were found to be hypermethylated and downregulated in *ros1 dml2 dml3* triple mutant seedlings and are considered direct targets of ROS1/DML2/DML3 (i.e. RDD targets) (Lin et al., 2020). We detected 50/275 RDD targets in ≥10% of nuclei in ≥1 embryonic cell cluster with enrichment scores ≥2 (Fig. S6). Sixteen of these RDD targets were highly enriched in the BCL and tended to have annotations related to extracellular signaling activities (e.g. signal peptides, extracellular domains, transmembrane domains). Notably, DML3 is also a putative embryonic RDD target, and thus may establish a positive feedback loop that enhances DNA demethylation in the BCL. ROS1, DML2 and DML3 transcripts were increased specifically in the BCL between the 1-cell and 32-cell stages (Zhou et al., 2020) (Fig. 4C). Consistently, most embryonic RDD target candidates were also increased in the BCL during these early embryonic stages (Fig. 4D). Based on these results, we suggest that DNA demethylases become activated in the BCL by the globular stage whereupon they catalyze the removal of methyl groups from a specific set of gene promoters, and correspondingly derepress their expression to potentially regulate intercellular signaling processes.

Differential enrichment of transcription factor binding motifs

To gain insights into the transcriptional processes that help define these embryonic cell-specific transcriptomes, we tested whether any consensus DNA motifs from the CIS-BP database of transcription factor binding experiments were overrepresented in the promoters of the top-250 ranked genes of each cluster. A total of 18 TF motif families were overrepresented in at least one cluster (Fig. 5A). WRKY2 is a major transcriptional activator in the basal cell lineage (BCL) and was shown to directly activate WOX8 and WOX9 during polarization of the zygote (REF: Ueda 2011 Dev Cell). Consistent with this report, the most overrepresented motif in BCL clusters is the W-box bound by WRKY TFs, and this correlates well with the expression enrichment of WRKY2 (Pearson's $r = 0.86$; Supplemental TableS3). In addition to WRKY2, the expression pattern of two other WRKY TFs (WRKY28 and WRKY19) strongly correlated with enrichment of the WRKY motif (Pearson's $r = 0.94$ and 0.96 , respectively). The WOX family binding motif was similarly concentrated in basal clusters, matching the observed expression pattern of WOX8 and to a lesser extent WOX9. The B3 domain transcription factor FUSCA3 is highly specific to basal nuclei in the globular atlas, and the RY motif bound by FUS3 is similarly enriched only in basal clusters. Maintenance of QC identity in the embryo requires JACKDAW (JKD), a member of the INDETERMINATE DOMAIN (IDD) subfamily of C2H2 zinc finger TFs (Welch et al., 2007). The IDD motif is enriched exclusively in the QC cluster, where JKD is the second highest ranked gene behind only WOX5. Class IV HD-ZIPs include the L1 layer marker genes ATML1 and PDF2, and their binding site is overrepresented in the three protoderm clusters.

To elucidate potential cooperative or competitive relationships between TF binding sites in the globular embryo, the 18 motif families were tested for co-occurrence with each of the other motifs within the set of promoters from top-250 ranked genes (Fig. 5D). The major suspensor/embryo proper spatial partition was again observed, with a tendency of basal suspensor motifs to co-occur in the same promoters and apical embryo proper motifs to co-occur in a separate set. An exception to this pattern is the HD-ZIP IV motif, which is enriched in clusters of the outer cell layer, but it is largely absent from promoters containing other apically-enriched motifs. Instead, it tends to occur together in promoters containing more basal motifs, such as FUS3 and WOX binding sites.

One of the strongest pairs of co-occurring motifs was the ETHYLENE RESPONSE FACTOR (ERF) family's GCC-box and the TELOMERE BINDING PROTEIN (TBP) telobox. In promoters that contain both motifs, the GCC-box tended to occur at or just downstream of the transcription start site, while teloboxes were concentrated immediately upstream (Fig. 5E). For such promoters containing both motifs, nucleotides within a GCC-box or telobox are significantly more conserved in a 63-species alignment of land plants relative to promoter regions that do not include either motif (Haudry et al., 2013; Tian et al., 2020) (Fig. 5F). As a specific example, the two expressed histone deacetylases, HDA3 and HDA4, are both enriched in all embryo proper tissues and contain one or more conserved copies of the telobox and GCC-box elements in their promoters (Fig. 5G). Putative GCC-box binding proteins include the transcriptional activator *DORNRÖSCHEN* (*DRN*) and four *CYTOKININ RESPONSE FACTORS* (*CRF1/2/3/10*) which are expressed in embryo proper tissues enriched for the GCC-box. In contrast, TBP family member transcripts generally were negatively correlated with telobox-containing gene transcripts (Supplemental TableS3) with the most highly expressed member, *TRP2*, mostly restricted to the basal cell lineage. Teloboxes have been reported as functional motifs in Polycomb Responsive

Elements (Xiao et al., 2017), and members of the TBP family were recently shown to recruit the Polycomb Repressive Complex PRC2 to teloboxes (Zhao et al., 2018). Therefore the co-occurrence of teloboxes and GCC-boxes at genes active in the embryo proper is consistent with antagonism between activation by *DRN/CRFs* and repression by PRC2.

DISCUSSION

We presented a contamination-free cell type specific expression atlas with FANS and snRNA-seq. Our sample size was relatively small in comparison with other scRNA-seq studies utilizing droplet-based methods (REF roots). The conventional scRNA-seq analysis pipelines were not readily applicable to distinguish detailed cell types because we observed the variations among nuclei were dominated by housekeeping genes and cell cycle-related genes. While it was possible to resolve the inter-cell type variations by adding another hundreds or thousands of nuclei, we successfully identified the embryonic cell types by highlighting the variations of known marker genes collectively. The lack of representations for ground tissue initials and upper inner peripheries could be either due to the fewer specific markers reported for these two cell types or the two cell types were not fully established or distinct at the early globular stage. Our dataset provided a more precise and detailed expression atlas than the previous apical-basal resolution and enabled the distinction of the BCL quiescent center and columella in embryo proper which was previously impossible. The identified clusters could be sub-clustered and analyzed individually for further investigations such as searching for rare cell types. Moreover, we identified three clusters representing dividing cells, suggesting the embryonic cell type specifications are dynamic and need to be maintained. Further analyses on the dividing cells and the corresponding cell types could reveal the cell specification and maintenance mechanisms.

The ACL and BCL clusters had distinct epigenetic landscapes according to the expressions of the epigenetic regulators. Most variations were observed between ACL and BCL instead of individual clusters, suggesting the epigenetic regulations in early embryos were established after the asymmetric zygotic division and inherited in the corresponding lineages. Interestingly, the QC cluster showed intermediate epigenetic landscapes, suggesting a genome-wide shift from the inherited BCL states to acquired states similar to embryo proper. The depletion of DNA methylases (MET1 and CMT3) and the enrichment of DNA demethylases (ROS1, DML2 and DML3) suggested CG and CHG content in BCL were hypomethylated. Several key factors involved in the 24-nt siRNA production were also enriched in BCL, suggesting suspensors were actively producing 24-nt siRNA for RNA-directed DNA methylation (RdDM). On the other hand, the genes involved in siRNA activities were enriched in ACL, especially in the shoot meristem initials which will be the precursors of all above ground tissues as well as future gametes. Because of the plasmodesmatal connections between the AC and BC lineages, as well as the terminally differentiated fate of BCL, it is tempting to speculate that one of the functions of the suspensor is to generate large amounts of 24-nt siRNAs that move into the embryo proper to stabilize the epigenome, similar to what has been proposed in the pollen vegetative nuclei and central cell/endosperm. The Active DNA demethylation in BCL also appeared to derepress a subset of the RDD target genes which involved in extra-cellular signaling functions.

It is interesting that the BCL enriched genes were less conserved than the ACL enriched genes. The higher mutation rates of the BCL enriched genes might not be directly related to the hypomethylated genome because the suspensor genome will not contribute to the next generation. Instead, it could suggest the BCL enriched genes were more redundant or their functions had higher mutation tolerance. The hypomethylated genome could accumulate more mutations due to error-prone DNA repair mechanisms, and our observations suggested the ACL and BCL established different epigenetic landscapes after the zygotic division. In normal cases where the embryos were from the less mutated ACL, only the non-deleterious mutations from the zygotic stage could be passed to the next generation and contribute to the population genetic variations. Nonetheless, it is known that the top tier suspensor possesses the potential to develop into a new embryo proper when the old one is impaired or removed. In this case the more mutated suspensor genome could introduce more genetic variations to the next generation, and such events may facilitate adaptation or even speciation in nature.

METHODS

Plant materials and growth conditions

The wild type *Arabidopsis thaliana* accession Columbia (Col-0) and the transgenic Col-0 plants pWOX2::H2B-GFP, pWOX2::tdTomato-RCI2b were grown at 20°C-22°C temperature and 16h light/8h dark cycles under incandescent lights (130-150 $\mu\text{mol}/\text{m}^2/\text{s}$) in a climate-controlled growth chamber.

Nuclei isolation and fluorescence activated nuclei sorting (FANS)

We isolated developing seeds containing globular or pre-globular embryos from WOX2p::NLSGFP lines and Col-0 for sorting. All buffers used in the nuclei isolation and sorting contained 0.4 U/mL RNase inhibitor murine (NEB, M0314L). For each replicate, developing seeds from 30 (pre-globular) or 20 (globular) siliques were carefully isolated with tungsten needles under a stereomicroscope. The isolated seeds were immediately transferred to 600 μL cooled fixative buffer consisted of 1x Galbraith's buffer (20 mM MOPS, pH 7.0, 30 mM sodium citrate, 1% Triton X-100, 45 mM MgCl_2) and 500 μM dithiobis(succinimidyl propionate) (DSP, ThermoFisher, 22586). The seeds were isolated between 2-6 p.m. and incubate at 4°C overnight. Cross-linked samples were incubated with 800 μL quenching buffer consisted of 1 M Tris-HCl, pH 7.0, 30 mM sodium citrate, 1% Triton X-100 and 45 mM MgCl_2 at room temperature for fifteen minutes with gentle shaking. The quenched samples were washed with 600 μL HG-GB (1x Galbraith's buffer and 1 M hexylene glycerol, Sigma 112100) twice. The seeds were then gently homogenized with micro-pestles in 1.5 mL microtubes with 200 μL HG-GB. Micro-pestles were rinsed with 400 μL HG-GB and the homogenized samples were gently pipetted ten times, and then incubated at 4°C for fifteen minutes to maximize nuclei release. The homogenized samples were then filtered with 30 μm filters and collected in 2 mL microtubes. Another 600 μL HG-GB was added to the homogenizing 1.5 mL microtube, filtered through the same 30 μm filter and collected in the same 2 mL microtube to maximize nuclei recovery. The filtered samples were then centrifuge at 4°C at 1000 g for ten minutes. The supernatant was removed carefully without disturbing the gray-ish pellets. A fresh aliquot of 1 mL HG-GB and 1 μL of 10 mg/mL DAPI were added into microtubes and the pellet were gently re-suspended. Samples were then washed for five times, each time consisted of a ten-minute centrifuge at 4°C at 1000 g and replacement of supernatant with fresh aliquot of 1 mL 1x Galbraith's buffer. After washes, the nuclei were then re-suspended in 800 μL 1x Galbraith's buffer for sorting.

The isolated nuclei were sorted with BD FACSAria™ III Cell Sorter (BD Biosciences). DAPI signals were activated by 375 nm laser and collected with 450/40 nm filter. GFP signals were activated by 488 nm laser and collected with 530/30 nm filter. To maximize purity, only the droplets containing a DAPI signal falling into the two peak regions (representing 2C and 4C nuclei) were considered for GFP gating. For GFP gating, a region with low auto-fluorescence and high GFP signal was drawn and the region had near zero events for Col-0 samples while it had at least 200 events for WOX2p::NLSGFP samples. Each nucleus passing both DAPI and GFP gating was collected with single cell settings into 4 μL of lysis buffer on 96-well plates. The lysis buffer was provided by Vienna Biocenter Core Facilities Next Generation Sequencing team (VBCF NGS) with the addition of 25 mM dithiothreitol (DTT).

Single nucleus RNA-seq (snRNA-seq)

The sorted nuclei were submitted to VBCF NGS for library preparation following the published SmartSeq2 single-cell protocol (REF) with an addition of a 30-minute incubation at 37°C before reverse transcription. The tagged libraries were sequenced by VBCF NGS on Illumina HiSeq 2500 with 50 bp single ends mode.

Sequence alignment and quantification

The Fastq files and the Araport11 annotations were provided to kallisto for pseudo-alignment with the supplemented codes.

Quality control and census count conversion

The TPM table, cell data and gene data (Supplementary Data 3) were imported to Monocle3. Libraries with less than 100,000 aligned reads or less than 1000 detected genes were considered as of low quality and excluded from subsequent analyses. The TPM values were then converted to census counts with the census conversion algorithm (Qiu et al., 2017). The census counts were used as gene expression levels for the subsequent analyses.

Maternal contamination removal and tissue enrichment tests

The single or combined expressions were used to perform tissue enrichment tests with all default settings as described (Schon and Nodine, 2017). The census count expression and the metadata of 1116 snRNA-seq libraries from 12 plates were constructed as a cell data set (CDS) for Monocle3 with R version 3.6.3. The first quality control was done as Monocle3 guideline instructed. Nuclei with more than 1,000,000 pseudo-aligned reads and more than 1000 detected genes per nucleus were considered in subsequent analyses. Genes passing the Monocle3 function `detect_genes(CDS, min_expr = 0.1)` and expressed in at least 3 nuclei were considered in subsequent analyses. This quality control step resulted in a CDS with 24591 genes and 534 nuclei. An unsupervised dimension reduction and clustering performed on this CDS resulted in twenty clusters. Two of the clusters (12 and 13) were dominated by nuclei most resembling the seed coat reference according to tissue enrichment tests, and the nuclei from these two clusters were excluded from subsequent analyses. A CDS containing 486 globular embryonic nuclei and expressing 23,959 genes was then used for subsequent cell type-score calculation and clustering.

Calculation of cell type-scores for globular nuclei and clustering

A set of 174 marker genes were collected from literature (supplementary table S1). Each gene had at least one primary literature indicating the expression patterns in embryos by either in situ hybridization or transgenic reporters (transcriptional-/translational-FP or GUS). The expressions were assigned to nine cell types: upper protoderm (upd), lower protoderm (lpd), shoot meristem initials (smi), upper inner periphery (uip), vascular initials (vas), ground tissue initials (grd), quiescent center initials (QC), columella initials (col) and suspensors (sus). The expressions were recorded as strongly expressed (s), weakly expressed (w), not expressed (n) or not informed (NA). The 174×9 matrix was intersected with expressed genes in our globular snRNA-seq libraries, which had at least one census count in at least 7 nuclei. The resulting 135 expressed marker genes served as the reference for cell type-score calculations, with 56, 52, 38, 43, 62, 43, 56, 51 and 29 positive markers and 79, 83, 97, 92, 73, 92, 79, 84 and 98 negative markers for upd, lpd, smi, uip, vas, grd, QC, col and sus, respectively. We utilized the two-tailed hypergeometric tests and assumed that a nucleus expressing the positive markers of a cell type were more likely from that cell type, while expressing the negative from a cell type were less likely from that cell type. The cell type-scores were the $-\log(\text{p-value})$ calculated by two-tailed hypergeometric tests with a homemade script. The 486×9 cell type-scores were incorporated into the globular CDS, and the dimension reduction and clustering were built on these nine cell type-scores. The cluster identities were predicted based on the cell type labels within each cluster for further validations.

In silico validation of cluster identities

The mean expressions of all nuclei within each cluster were used as the cluster expressions. The cluster expressions were then used to perform tissue enrichment tests as previously described (REF Schon&Nodine) and to calculate the Spearman correlation with the three replicates of embryo proper (32E) and suspensor (32S) samples from Zhou et. al..

The three genes (PEAR1, DOF6 and GATA20) were not included in our reference marker list for tissue score calculation and were reported by Smit et. al. showing vascular-specific expression patterns in embryos. The schematic expression patterns of the three genes were adopted from Smit et. al. and the expressions of the three genes were plotted as dotplots with Monocle3 function `plot_genes_by_group()`.

In situ validation of cluster identities

Thirty four DEGs without previously reported embryonic expression patterns were selected according to their expression patterns and probe availability. The in situ primers were designed, probes were synthesized and in situ hybridizations were conducted as previously described (REF).

Ranked Gene Enrichment

For each cluster of nuclei, a ranked gene enrichment strategy was written in the R programming language as *enrichment_ranking.R* (Supplemental Code), and is defined as follows: let G be the set of “expressed” genes, defined as all genes with ≥ 1 RNA-seq read count in $\geq 10\%$ of nuclei in ≥ 1 cluster, excluding non-nuclear and non-pollII genes. For each gene i in each nucleus j , $CPM_{ij} = 10^6 \times \text{counts}_{ij} / \sum_{g \in G} \text{counts}_{gj}$. Let C be a set of nuclei in a cluster and $|C|$ the number of nuclei in cluster C . Mean CPM of gene i in cluster C is defined as $\mu_{iC} = \sum_{j \in C} CPM_{ij} / |C|$. Proportion detected p is defined for each gene i in each cluster C as the number of

nuclei in which gene i was detected: $p_{iC} = \left(\sum_{j \in C} \begin{cases} 1 & \text{if } \text{counts}_{ij} \geq 1 \\ 0 & \text{if } \text{counts}_{ij} < 1 \end{cases} \right) / |C|$. Using one cluster C as an in-group and all other clusters as outgroup O , a mean CPM log₂ fold change of each gene i is calculated as $F_{iC} = \log_2 \left(\frac{1 + \mu_{iC}}{1 + \sum_{o \in O} \mu_{io} / |O|} \right)$, and a mean proportion difference $D_{iC} = p_{iC} - \sum_{o \in O} p_{io}$. Both sets F_C and D_C were centered and mean-scaled so that $\widehat{F}_C = \frac{F_C - \bar{F}_C}{\sigma(F_C)}$, and $\widehat{D}_C = \frac{D_C - \bar{D}_C}{\sigma(D_C)}$, where \bar{x} is the mean and $\sigma(x)$ the standard deviation. Enrichment magnitude E_{iC} of gene i in cluster C is the combined deviation from the mean of F_C and D_C :

$$E_{iC} = \sqrt{\widehat{F}_{iC}^2 + \widehat{D}_{iC}^2} \times \begin{cases} 1 & \text{if } \widehat{F}_{iC} + \widehat{D}_{iC} > 0 \\ -1 & \text{if } \widehat{F}_{iC} + \widehat{D}_{iC} \leq 0 \end{cases}$$

In each cluster genes are ranked from highest to lowest enrichment magnitude; the first 250 genes are considered “top-ranked genes” for that cluster, and the last 250 genes are “bottom-ranked genes”.

Gene ontology analyses

The IDs of the top250 enriched genes for each cluster were submitted to TAIR GO Term enrichment (https://www.arabidopsis.org/tools/go_term_enrichment.jsp) powered by PANTHER with Fisher’s Exact test and calculated false discovery rate. The full enriched terms were presented in Supplementary Table S2. The first five terms with lowest p-values were highlighted in Fig. 3.

Transcription factor binding site analyses

Promoters for all genes were defined as the region 500bp upstream to 100bp downstream of the most common 5’ end in nanoPARE datasets of globular stage embryos (Plotnikova 2019 Plant Cell). For genes with no nanoPARE signal, the most upstream 5’ end annotated in TAIR10 v.46 was used. Transcription factor binding motifs for *Arabidopsis thaliana* were downloaded from CIS-BP (<http://cisbp.cabr.utoronto.ca>). All directly determined motifs were tested for statistical overrepresentation using Analysis of Motif Enrichment (AME, <http://meme-suite.org/doc/ame.html>) in each cluster by comparing the 250 top ranked gene promoters against a background set of the bottom 250 ranked gene promoters with default parameters. Motifs that were significantly enriched in at least one cluster were collapsed into motif families.

FIGURES AND LEGENDS

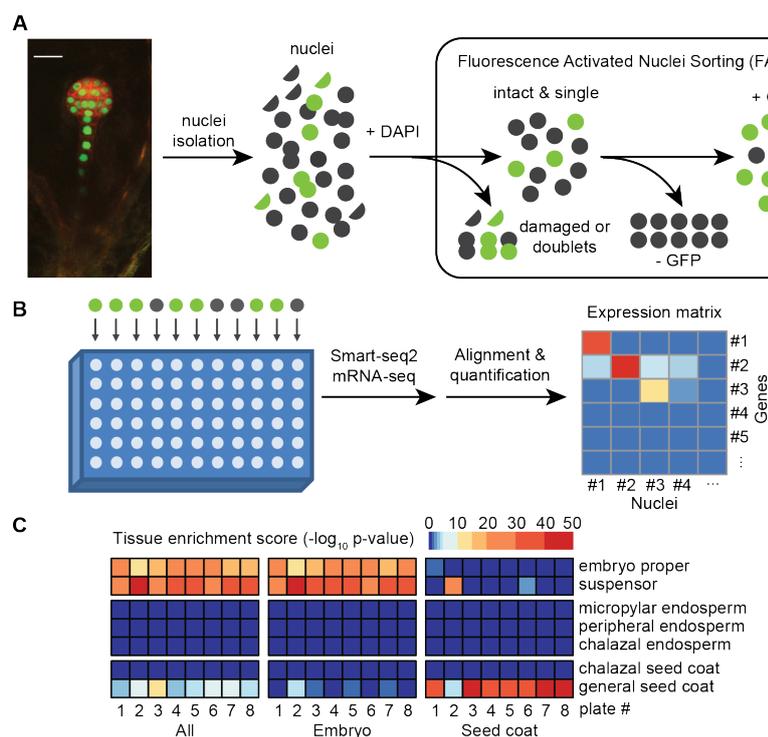


Figure 1. Acquisition of contamination-free transcriptomes from individual embryonic nuclei.

(A) Collection of singular embryonic nuclei. Fixed nuclei isolated from the pWOX2::NLS-GFP transgenic plants were analyzed by FACS machine to exclude debris or aggregates before enriching for GFP⁺ nuclei. The scale bar represents 20 μ m.

(B) Sequencing and quantification of single-nucleus libraries. Single-nucleus libraries were prepared from each sorted nucleus with a modified SmartSeq2 protocol (see Methods) and sequenced by Illumina platforms. The resulting reads were aligned to Araport11 and quantified by kallisto algorithm.

(C) Maternal contamination removal and assessment. Nuclei from each plate were assigned as embryonic or contaminated according to the unsupervised clustering results (Supplementary Fig S1C). The mean expressions of all nuclei under each category were considered as representative expressions and were tested tissue enrichment. The tissue enrichment test results indicated the filtered embryonic snRNA-seq libraries had little maternal contamination.

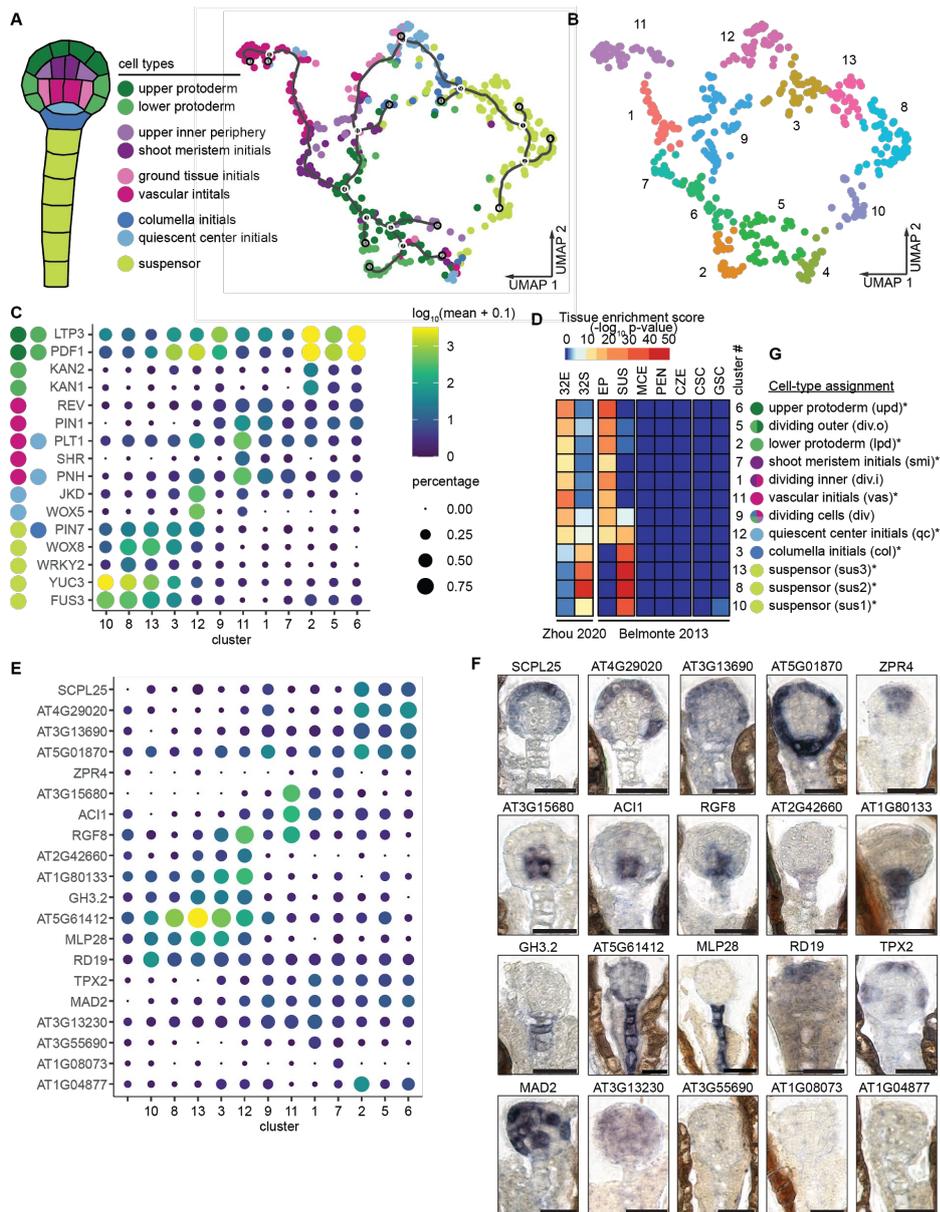


Figure 2. Identification of embryonic cell types.

(A) Resolving nine defined cell types by supervised clustering. The marker genes expressed in at least one of the nine cell types were used for calculating cell-type scores with hypergeometric tests. Each dot represented a nucleus and nuclei were labeled as the cell type with the highest cell-type score.

(B) The thirteen clusters corresponding to Fig 2A.

(C) The expression pattern of known cell type-specific markers (left) and in situ hybridization (ISH) candidates (right). The sizes of dots represented the percentage of expressing nuclei in each cluster, and the colors represented the log-transformed mean expression levels of each cluster.

(D) A selection of images of the in situ hybridization (ISH)-validated genes presented in Fig 3C (right). The corresponding clusters which the validated genes were enriched in were labeled on each image. The scale bars represent 20 μ m.

(E) A Summary of cluster identities. The number of experimentally tested ISH and the number of ISH patterns matching the assigned cluster identities were listed in the first two columns. The in silico combined cluster expressions were used to conduct tissue enrichment tests against Belmonte et.al. microarray data as previously described. The resulting tissue enrichment test heatmaps were shown in the middle columns. The same in silico combined cluster expressions were used to calculate Spearman correlations with the 32E (globular embryo proper) and 32S (globular suspensor) dataset from Zhou et.al. The resulting Spearman correlation heatmaps were shown in the rightmost columns. The cluster identities were assigned based on in situ and in silico validations and were color-coded with the cell type colors presented in Fig 2A. The asterisks indicated the more confident clusters representing globular embryo cell types..

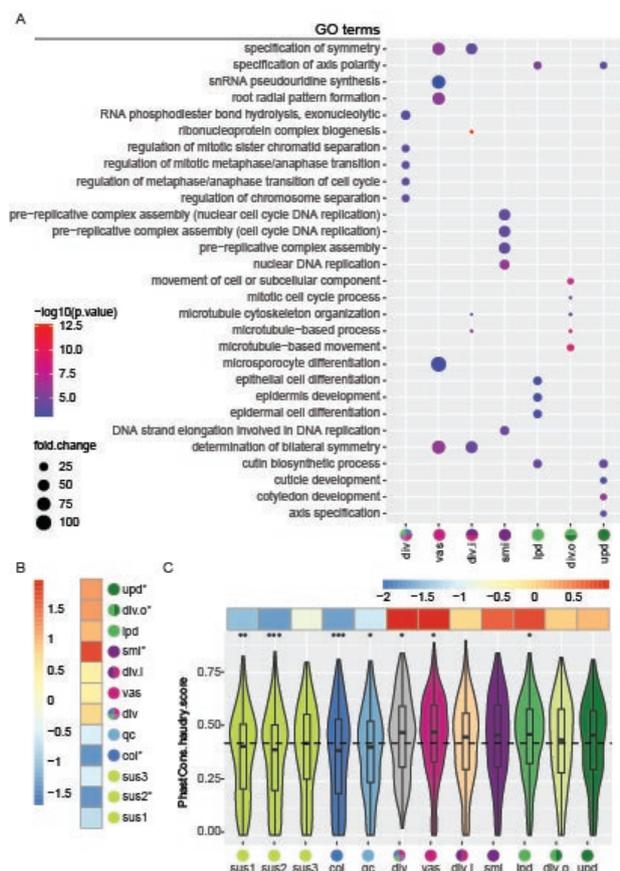


Figure 3. General characteristics of transcriptomes from embryonic cell-types.

(A) The top five enriched gene ontology (GO) terms identified by PANTHER for each cluster according to the top250 enriched genes for each cluster. The suspensor clusters (8,10,13) and hypophysis clusters (3,12) had no enriched GO terms and thus excluded in this graph. The sizes of dots represented the fold changes and the colors of dots represented the $-\log_{10}$ transformed p-value.

(B) The overrepresentation of embryo-defective (EMB) genes for the top250 enriched genes for each cluster. Asterisks indicate significant ($p \leq 0.05$) enrichment or depletion of EMB genes than expected. upd: upper protoderm, div.o: dividing protoderm, lpd: lower protoderm, smi: shoot meristem initials, div.i: dividing inner cells, vas: vascular initials, div: dividing cells, qc: quiescent center initials, col: columella initials, sus: suspensor.

(C) The PhastCons conservation scores (Haudry et al., 2013) of top250 enriched genes for each cluster. The dashed line indicates the mean conservation scores of all expressed genes and the deviations from it were highlighted for each cluster in the upper panel. The asterisks indicated p-values ≤ 0.05 (*), ≤ 0.01 (**) or ≤ 0.001 (***) by two-sided KS tests with the alternative hypothesis that the cluster conservation scores were not equal to the score of all expressed genes in embryos. The PhastCons and PhyloP scores (Tian et al., 2020) had similar trends and were showed in Supplementary Fig. S5.

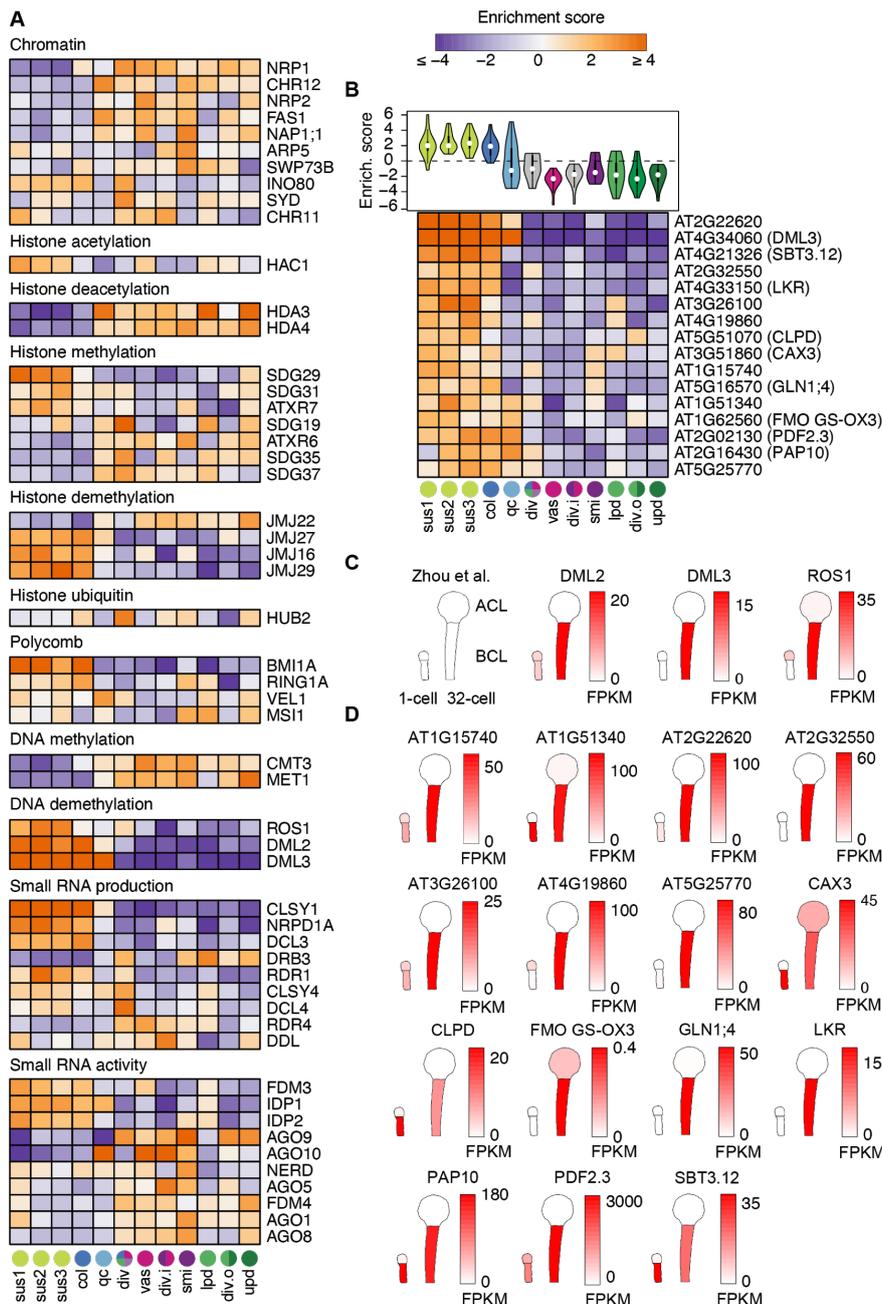


Figure 4. Transcripts encoding epigenetic regulators vary across embryonic cell types.

(A) Heatmap illustrating enrichment scores in 12 clusters corresponding to different embryonic cell types. Transcripts with enrichment scores ≥ 2.5 in ≥ 1 cell cluster are shown and enrichment scores are colored according to key. Gene names are indicated and cluster identities are marked and color-coded at the bottom. sus1/2/3, suspensor1/2/3; col, columella initials; qc, quiescent center; div, dividing cells; smi, shoot meristem initials; div.i, dividing inner cells; lpd, lower protoderm; div.o, dividing outer cells; upd, upper protoderm

(B) Violin plot (*top*) and heatmap (*bottom*) of enrichment scores for 16/50 ROS1/DML2/DML3 (RDD) targets detected and enriched in the basal cell lineage.

(C and D) Schematic representations of RDD transcripts (C) and their putative embryonic targets (D) based on published mRNA-seq from apical and basal cell lineages in 1-cell and 32-cell stage embryos (REF). Transcript levels (FPKM; fragments per kilobase of transcript per million mapped reads) are colored according to the keys.

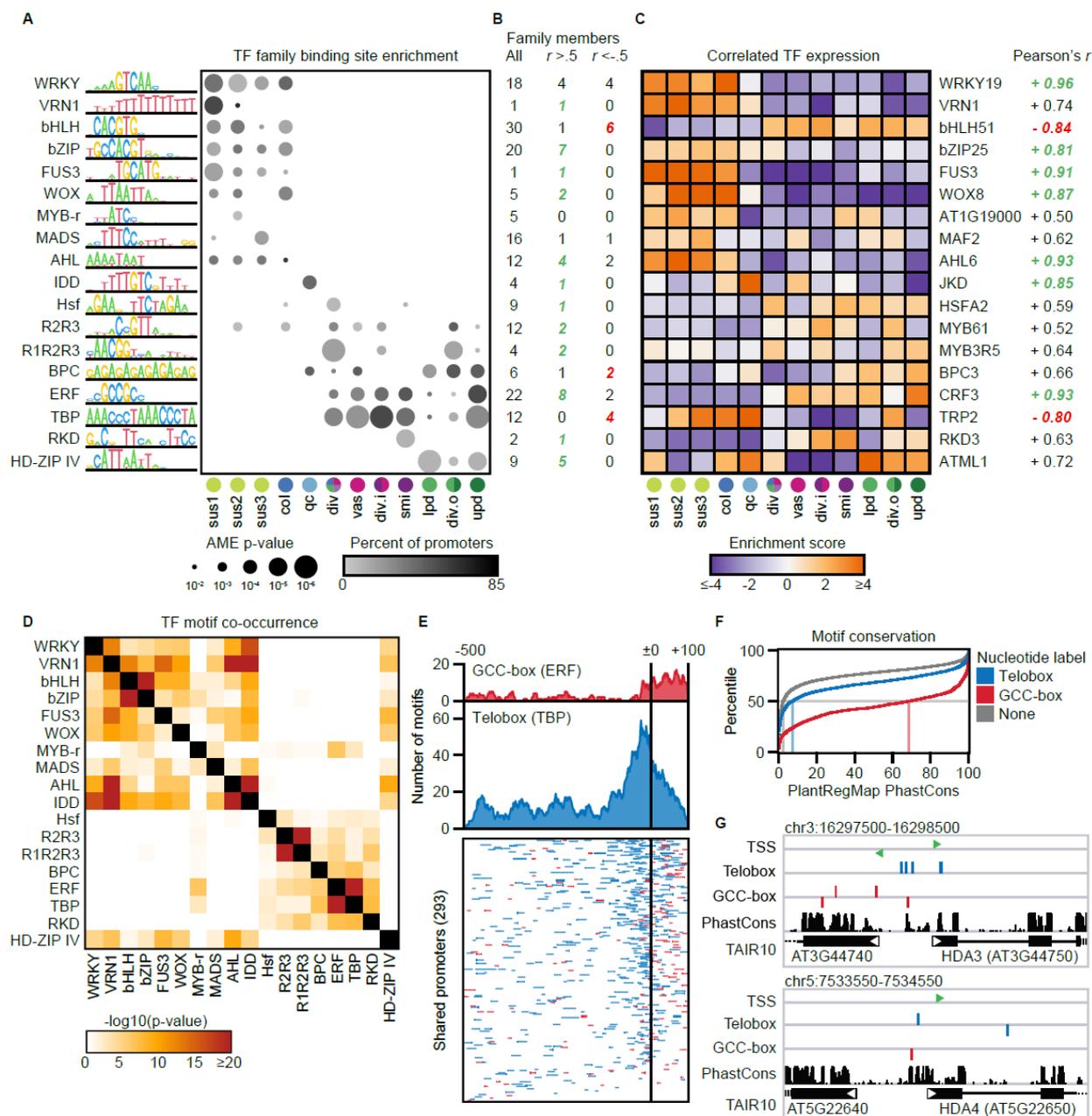


Figure 5. Cluster-enriched transcription factor binding motifs.

(A) Dot plot of transcription factor families with DNA binding motifs significantly enriched in at least one cluster. Dot size shows the most significant enrichment ($-\log_{10}$ p-value, AME) of a motif in the family; dot color depicts the percentage of the 250 top-ranked genes whose promoter contains the specified motif.

(B) Number of transcription factors in each family which (left) are detected in the globular atlas, (center) have a Pearson's correlation between expression enrichment and motif enrichment across clusters greater than 0.5, or (right) less than -0.5.

(C) Expression enrichment of the transcription factor within each family whose expression enrichment correlates most strongly to motif enrichment.

(D) Pairwise analysis of overrepresentation of a second TF motif co-occurring within the set of top-250 promoters containing a specific motif (p-value, hypergeometric distribution).

(E) (top) Metaplots of the distribution of GCC-box motifs (red) and telobox motifs (blue) relative to the transcription start sites (± 0) of 293 top-250 ranked promoters that possess both motifs. (bottom) Location of individual motifs in the set of 293 promoters, colored as above.

(F) Cumulative frequency of evolutionary conservation score for the promoters in (E), subdivided by whether the nucleotide was contained in a telobox (blue), a GCC-box (red), or neither (gray). Median conservation score of each group is demarcated by a vertical line.

(G) IGV browser snapshots of the ± 500 bp region around the TSS of HDA3 (top) and HDA4 (bottom), with putative telobox and GCC-box motifs marked in blue and red, respectively.

SUPPLEMENTAL INFORMATION

Supplementary Figures and legends

Supplementary Figure S1. Embryonic nuclei enrichment by FACS and the quality of single nuclear libraries.

Supplementary Figure S2. Assessing and mitigating maternal contamination.

Supplementary Figure S3. Resolving cell types by clustering.

Supplementary Figure S4. Quantification of RNA in situ Patterns.

Supplementary Figure S5. Conservation score distribution among clusters.

Supplementary Figure S6. RDD target candidate expression across embryonic cell types.

Supplementary Tables

Supplementary Table S1. Curated marker genes used for cell-type score calculation.

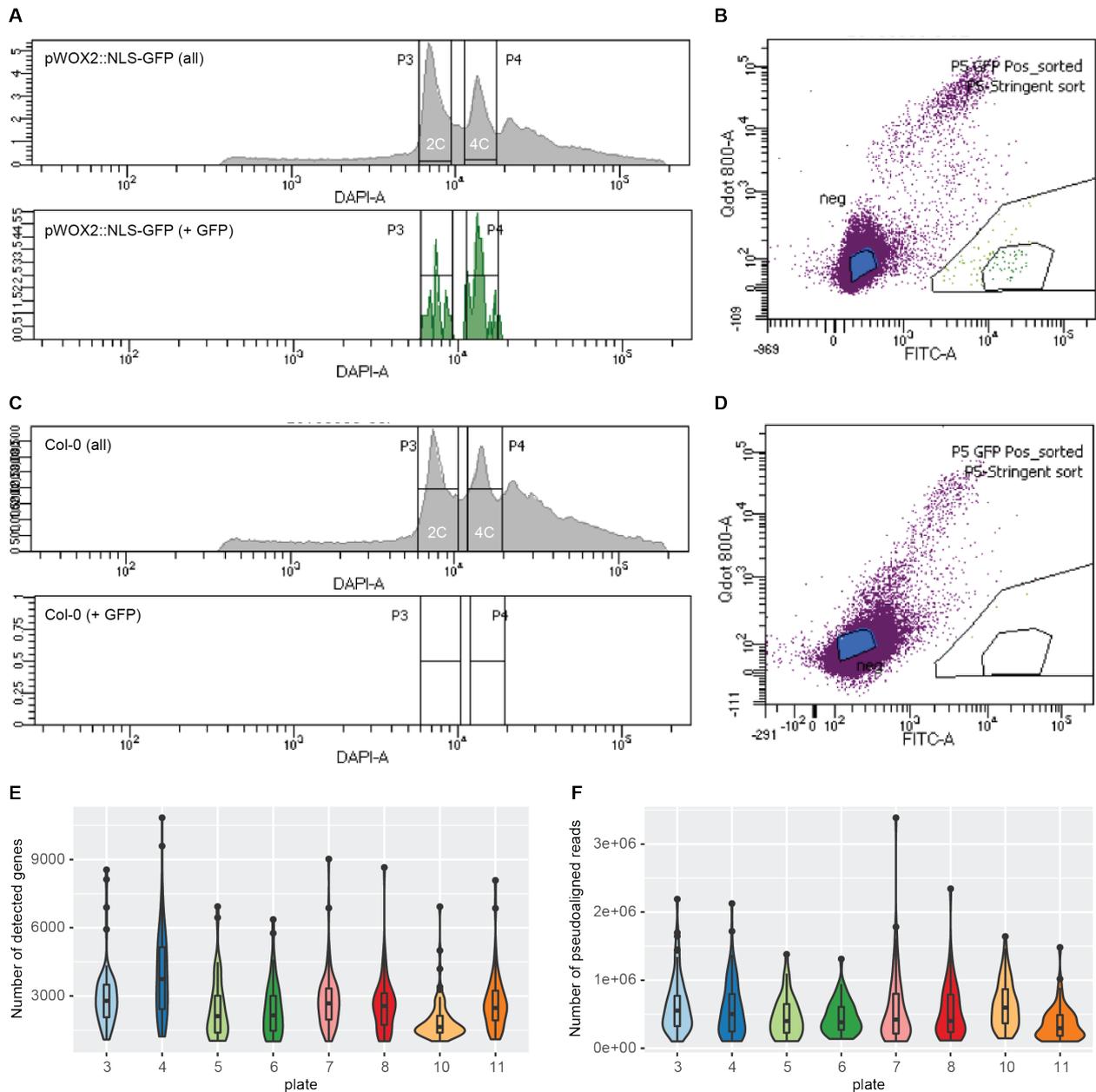
Supplementary Table S2. Gene ontology analyses results for the 12 clusters.

Supplementary Table S3. Transcription factor motif correlations.

Supplementary Data

Supplementary Data 1. General information of nuclei and genes

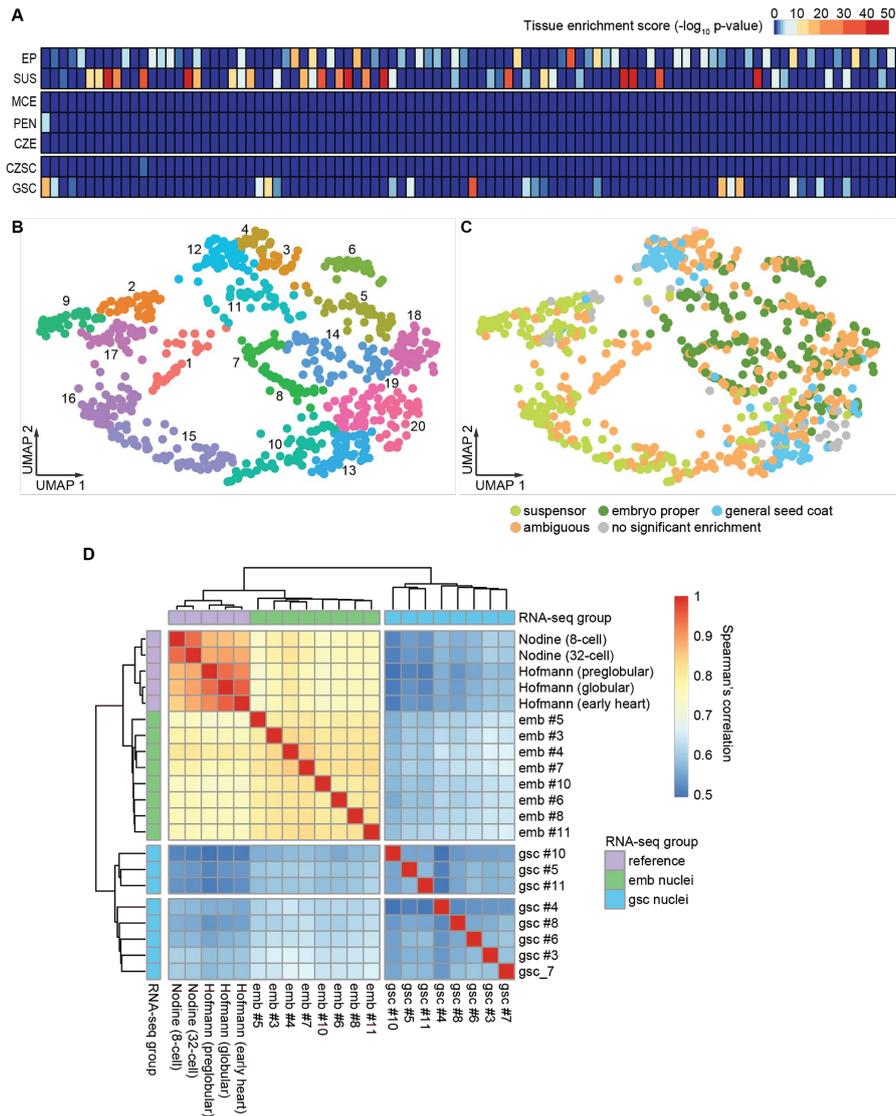
Supplementary Data 2. Gene expression data and ranks



Supplementary Figure S1. Embryonic nuclei enrichment by FACS and the quality of single nuclear libraries.

(A) The DAPI profile of all nuclei (left-up), GFP+ nuclei (left-down) and the selection of GFP+ nuclei (right) for the pWOX2::NLS-GFP transgenic line and wildtype Col-0. The two peaks in the DAPI profile represented the 2C and 4C nuclei, and only droplets within these two peaks were funneled to GFP selection. In GFP selection, each nuclei was plotted based on the collected emission of GFP wavelength on the x-axis and infrared on the y-axis. Only the nuclei with high GFP signal and low infrared signal (the bottom right inner circle) were collected for sequencing. A less stringent selection (corresponding to the outer circle) resulted in >70% of libraries were considered as maternal contamination in our preliminary tests.

(B) The number of detected genes (up) and the number of aligned reads (bottom) per nucleus for each plate. The quantiles for all 534 nuclei were listed on the right.

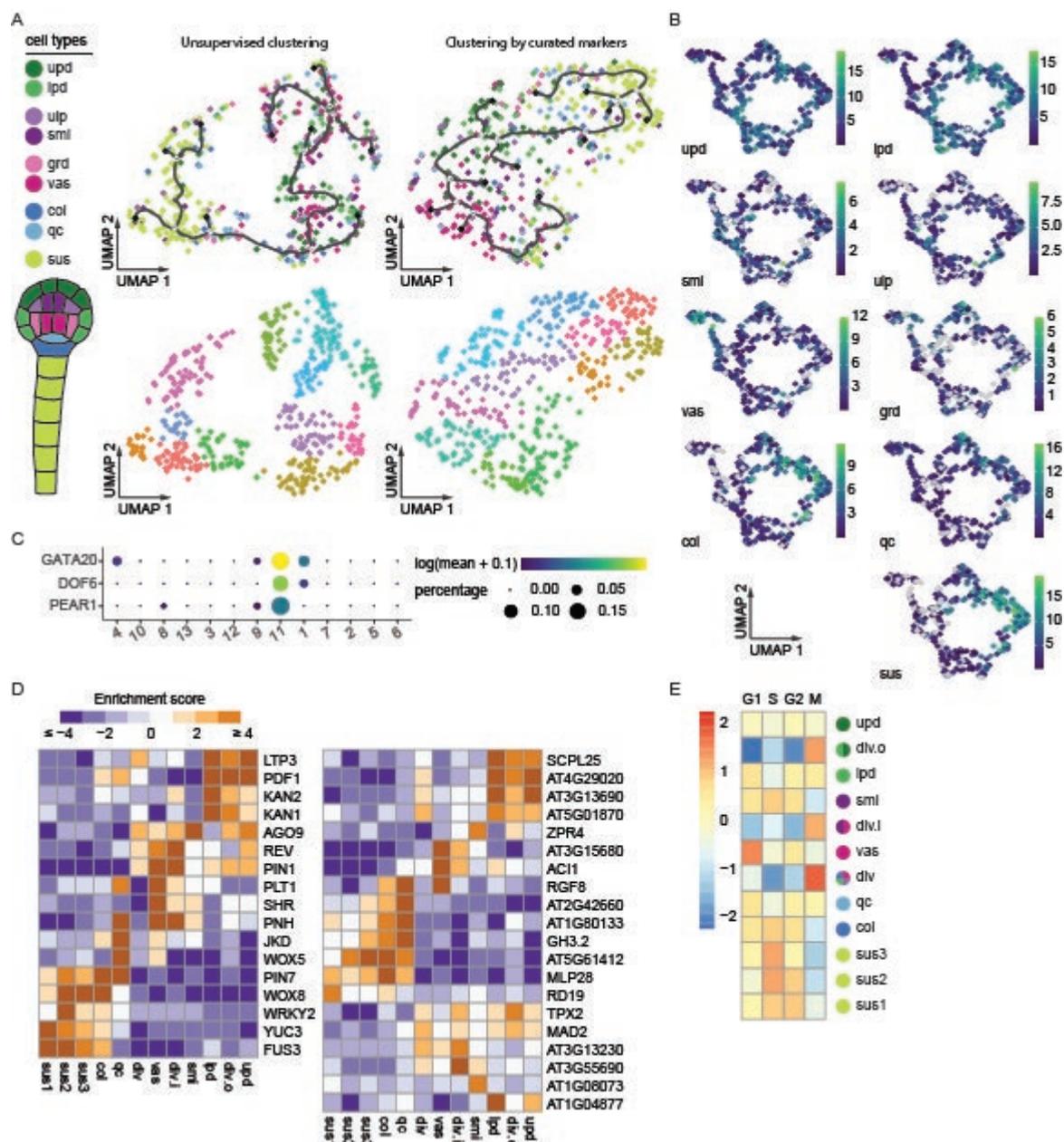


Supplementary Figure S2. Assessing and mitigating maternal contamination.

(A) A representative result of tissue enrichment tests on single nuclear RNA-seq libraries. Each row represented one of the seed tissue types, and each column represented a nucleus. Most nuclei were enriched for one tissue type, indicating that the major source of maternal contamination was falsely sorted maternal nuclei instead of the ambient RNA.

(B) Unsupervised clustering and identifying contaminated nuclei. All nuclei having $\geq 100,000$ aligned reads and ≥ 1000 expressed genes were clustered and the resulting UMAP plots were color-coded by clusters (left) or the significantly enriched tissue type according to tissue enrichment test (right). If a nucleus had significant enrichment for more than one or less than one tissue type, it was labeled as ambiguous or no significant enrichment, respectively.

(C) The Spearman correlation among the embryonic and contaminated samples corresponding to Fig 1C and published embryonic datasets (REF Hofmann, Nodine).



Supplementary Figure S3. Resolving cell types by clustering.

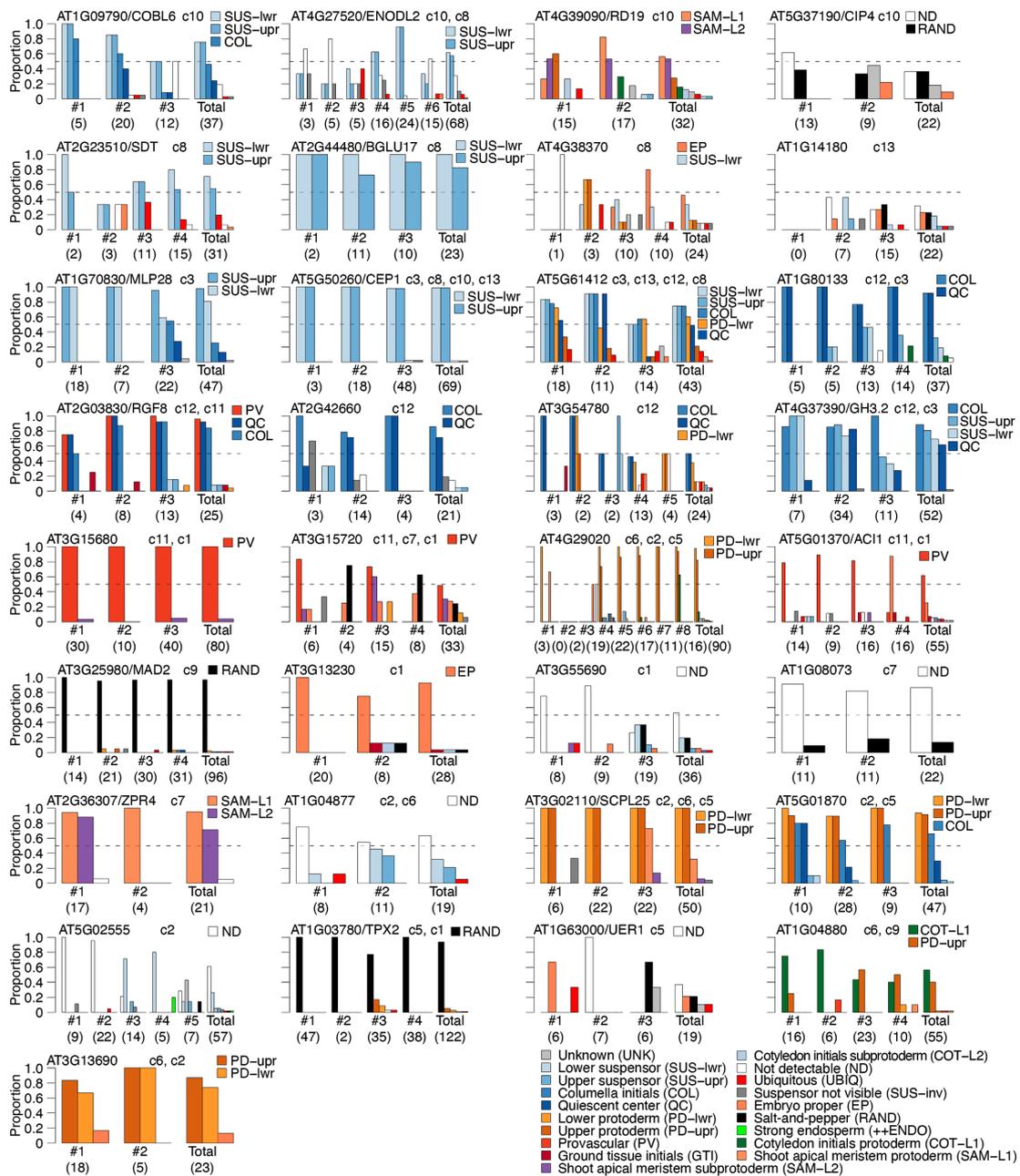
(A) The UMAP plots of unsupervised clustering (left) and clustering directly by the curated markers used for cell-type score calculation (right). The embryonic cell types were identical to Figure 2A. Each dot represented a nucleus and was labeled as the most resembling cell type (upper) or colored by cluster (lower).

(B) Accumulated expressions of markers selectively expressed in the corresponding cell types. Each dot represented a nucleus and was colored according to the sum of expressions of markers for the indicated cell types. Cluster identities are marked at the bottom left. sus1/2/3, suspensor1/2/3; col, columella initials; qc, quiescent center; div, dividing cells; smi, shoot meristem initials; div.i, dividing inner cells; lpd, lower protoderm; div.o, dividing outer cells; upd, upper protoderm.

(C) Expressions of three novel vascular-specific genes reported in Smit et. al.. The sizes of dots represented the percentage of expressing nuclei in each cluster, and the colors represented the log-transformed mean expression levels of each cluster.

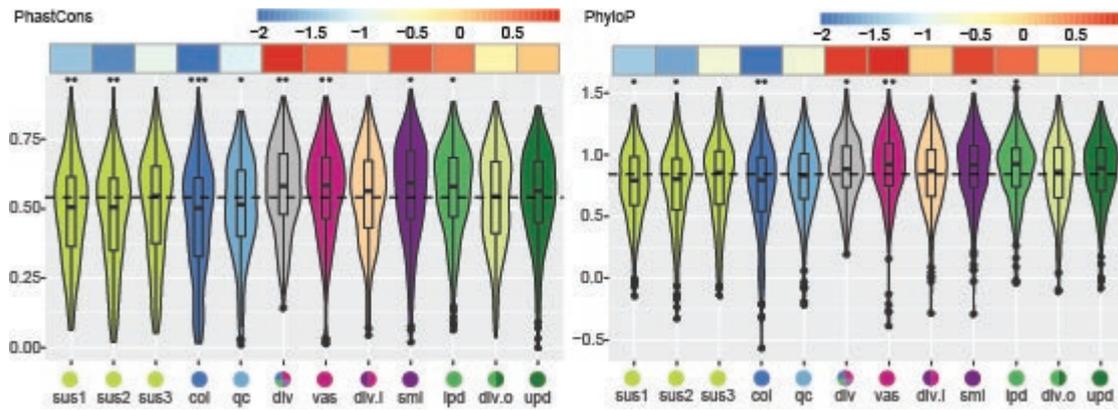
(D) The enrichment scores of the genes presented in Figure 2C. The expression ratios and expression levels were faithfully represented by the enrichment scores. Cluster identities are marked at the bottom. sus1/2/3, suspensor1/2/3; col, columella initials; qc, quiescent center; div, dividing cells; smi, shoot meristem initials; div.i, dividing inner cells; lpd, lower protoderm; div.o, dividing outer cells; upd, upper protoderm.

(E) The enrichment of cell cycle genes among the top250 expressed genes of each cluster. Three clusters had strong enrichment for M-phase genes and depletion of genes for other phases.



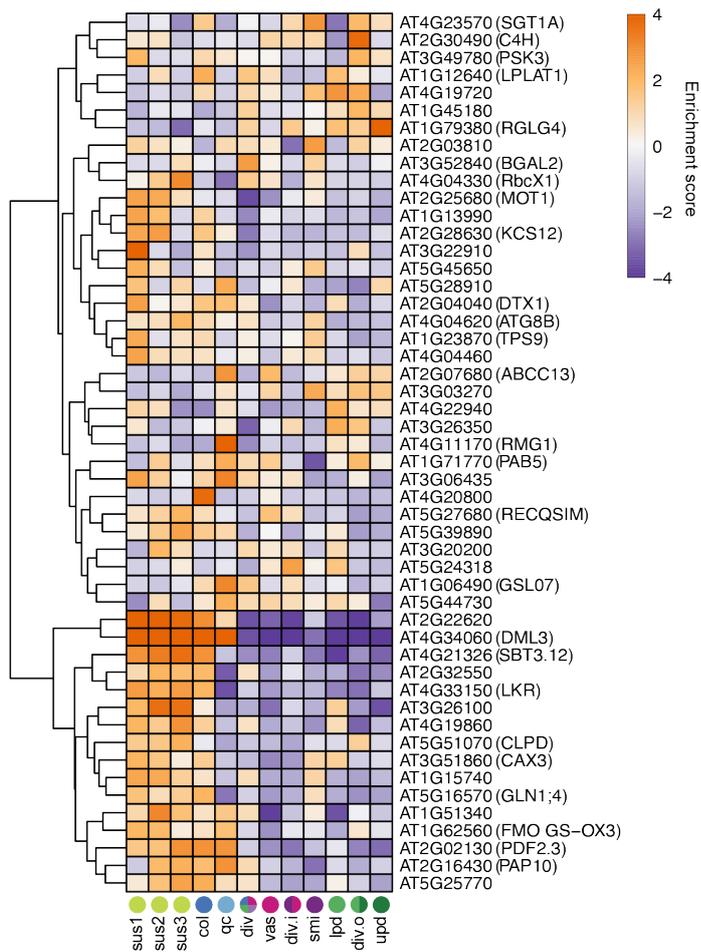
Supplementary Figure S4. Quantification of RNA in situ Patterns.

Bar plots illustrating the proportion of globular-staged embryos with RNA in situ signals in various cell types. Each panel summarizes patterns observed for individual transcripts. Corresponding unique gene identifiers (i.e. Arabidopsis Genome Initiative identifiers; AGIs) and, if annotated, common names are shown in the upper-left corner of each panel. Clusters for which transcripts are within the top 250 gene-ranked list are indicated at the top of each graph (e.g. c1, c2, etc.) in ascending order. Proportions of signals are shown for individual slides (i.e. technical replicates), as well as totals for each transcript, and the number of biological replicates are noted in parentheses. Legends in each panel are shown for patterns that occurred in >33% of embryos. The full legend is indicated in the bottom-right corner of the figure.



Supplementary Figure S5. Conservation score distribution among clusters.

The mean scores were highlighted in the upper panel. The conservation scores (Tian et al., 2020) showed similar trends. The asterisks indicated p-values ≤ 0.05 (*), ≤ 0.01 (**) or ≤ 0.001 (***) by one-sided KS tests with the alternative hypothesis that the cluster conservation scores were greater than the score of all expressed genes in embryos.



Supplementary Figure S6. RDD target candidate expression across embryonic cell types.

Heatmap of enrichment scores for 50 ROS1/DML2/DML3 (RDD) targets detected in $\geq 10\%$ nuclei in ≥ 1 cluster and with enrichment scores ≥ 2 in ≥ 1 cluster. Enrichment scores are colored according to key. Gene names are indicated and cluster identities are marked and color-coded at the bottom.

REFERENCES

- Bai, S.-N.** (2015). The concept of the sexual reproduction cycle and its evolutionary significance. *Front. Plant Sci.* **6**, 11.
- Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., Hashimoto, M., Fei, J., Harada, C. M., Munoz, M. D., et al.** (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E435–44.
- Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F. and Colot, V.** (2017). DNA methylation dynamics during early plant life. *Genome Biol.* **18**, 179.
- Bray, N. L., Pimentel, H., Melsted, P. and Pachter, L.** (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527.
- Brennecke, P., Anders, S., Kim, J. K., Kołodziejczyk, A. A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S. A., Marioni, J. C., et al.** (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095.
- Casson, S., Spencer, M., Walker, K. and Lindsey, K.** (2005). Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. *Plant J.* **42**, 111–123.
- Chen, G., Ning, B. and Shi, T.** (2019). Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front. Genet.* **10**, 317.
- Chen, M., Lin, J.-Y., Wu, X., Apuya, N. R., Henry, K. F., Le, B. H., Bui, A. Q., Pelletier, J. M., Cokus, S., Pellegrini, M., et al.** (2020). Comparative Analysis of Embryo Proper and Suspensor Transcriptomes in Plant Embryos With Different Morphologies. *Cold Spring Harbor Laboratory* 2020.11.30.404376.
- Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S. and Town, C. D.** (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **89**, 789–804.
- Clark, N. L., Aagaard, J. E. and Swanson, W. J.** (2006). Evolution of reproductive proteins from animals and plants. *Reproduction* **131**, 11–22.
- Efroni, I. and Birnbaum, K. D.** (2016). The potential of single-cell profiling in plants. *Genome Biol.* **17**, 65.
- Erdmann, R. M. and Picard, C. L.** (2020). RNA-directed DNA Methylation. *PLoS Genet.* **16**, e1009034.
- Gehring, M.** (2019). Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming? *New Phytol.* **224**, 91–96.
- Gerri, C., Menchero, S., Mahadevaiah, S. K., Turner, J. M. A. and Niakan, K. K.** (2020). Human Embryogenesis: A Comparative Perspective. *Annu. Rev. Cell Dev. Biol.* **36**, 411–440.
- Gooh, K., Ueda, M., Aruga, K., Park, J., Arata, H., Higashiyama, T. and Kurihara, D.** (2015). Live-cell imaging and optical manipulation of Arabidopsis early embryogenesis. *Dev. Cell* **34**, 242–251.
- Gutzat, R., Rembart, K., Nussbaumer, T., Hofmann, F., Pisupati, R., Bradamante, G., Daubel, N., Gaidora, A., Lettner, N., Donà, M., et al.** (2020). Arabidopsis shoot stem cells display dy-

dynamic transcription and DNA methylation patterns. *EMBO J.* **39**, e103667.

- Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J. J., Hession, C., Zhang, F. and Regev, A.** (2016). Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925–928.
- Haecker, A., Gross-Hardt, R., Geiges, B., Sarkar, A., Breuninger, H., Herrmann, M. and Laux, T.** (2004). Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Development* **131**, 657–668.
- Haudry, A., Platts, A. E., Vello, E., Hoen, D. R., Leclercq, M., Williamson, R. J., Forczek, E., Joly-Lopez, Z., Steffen, J. G., Hazzouri, K. M., et al.** (2013). An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat. Genet.* **45**, 891–898.
- Hofmann, F., Schon, M. A. and Nodine, M. D.** (2019). The embryonic transcriptome of *Arabidopsis thaliana*. *Plant Reprod.* **32**, 77–91.
- Hwang, B., Lee, J. H. and Bang, D.** (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96.
- Ingouff, M., Rademacher, S., Holec, S., Soljić, L., Xin, N., Readshaw, A., Foo, S. H., Lahouze, B., Sprunck, S. and Berger, F.** (2010). Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in *Arabidopsis*. *Curr. Biol.* **20**, 2137–2143.
- Jullien, P. E., Susaki, D., Yelagandula, R., Higashiyama, T. and Berger, F.** (2012). DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*. *Curr. Biol.* **22**, 1825–1830.
- Kao, P. and Nodine, M. D.** (2019). Transcriptional Activation of *Arabidopsis* Zygotes Is Required for Initial Cell Divisions. *Sci. Rep.* **9**, 17159.
- Law, J. A. and Jacobsen, S. E.** (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220.
- Lee, M. T., Bonneau, A. R. and Giraldez, A. J.** (2014). Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annu. Rev. Cell Dev. Biol.* **30**, 581–613.
- Lin, W., Sun, L., Huang, R.-Z., Liang, W., Liu, X., He, H., Fukuda, H., He, X.-Q. and Qian, W.** (2020). Active DNA demethylation regulates tracheary element differentiation in *Arabidopsis*. *Sci Adv* **6**, eaaz2963.
- Lukowitz, W., Roeder, A., Parmenter, D. and Somerville, C.** (2004). A MAPKK Kinase Gene Regulates Extra-Embryonic Cell Fate in *Arabidopsis*. *Cell* **116**, 109–119.
- Mayer, K. F., Schoof, H., Haecker, A., Lenhard, M., Jürgens, G. and Laux, T.** (1998). Role of WUSCHEL in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell* **95**, 805–815.
- Meinke, D. W.** (2019). Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in *Arabidopsis*. *New Phytol.*
- Menges, M., Hennig, L., Gruissem, W. and Murray, J. A. H.** (2003). Genome-wide gene expression in an *Arabidopsis* cell suspension. *Plant Mol. Biol.* **53**, 423–442.
- Nagasaki, H., Itoh, J.-I., Hayashi, K., Hibara, K.-I., Satoh-Nagasawa, N., Nosaka, M., Mukouhata, M., Ashikari, M., Kitano, H., Matsuoka, M., et al.** (2007). The small interfering RNA production pathway is required for shoot meristem initiation in rice. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14867–14871.

- Ning, J., Peng, X.-B., Qu, L.-H., Xin, H.-P., Yan, T.-T. and Sun, M.** (2006). Differential gene expression in egg cells and zygotes suggests that the transcriptome is restructured before the first zygotic division in tobacco. *FEBS Lett.* **580**, 1747–1752.
- Nodine, M. D. and Bartel, D. P.** (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**, 94–97.
- Palovaara, J., Saiga, S., Wendrich, J. R., van 't Wout Hofland, N., van Schayck, J. P., Hater, F., Mutte, S., Sjollem, J., Boekschoten, M., Hooiveld, G. J., et al.** (2017). Transcriptome dynamics revealed by a gene expression atlas of the early Arabidopsis embryo. *Nature Plants* **3**, 894–904.
- Papareddy, R. K., Páldi, K., Paulraj, S., Kao, P., Lutzmayer, S. and Nodine, M. D.** (2020). Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in Arabidopsis. *Genome Biol.* **21**, 251.
- Park, S. and Harada, J. J.** (2008). Arabidopsis Embryogenesis. In *Plant Embryogenesis* (ed. Suárez, M. F.) and Bozhkov, P. V.), pp. 3–16. Totowa, NJ: Humana Press.
- Picard, C. L., Povilus, R. A., Williams, B. P. and Gehring, M.** (2020). Single nucleus analysis of Arabidopsis seeds reveals new cell types and imprinting dynamics. 2020.08.25.267476.
- Picelli, S., Faridani, O. R., Björklund, A. K., Winberg, G., Sagasser, S. and Sandberg, R.** (2014a). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181.
- Picelli, S., Björklund, A. K., Reinius, B., Sagasser, S., Winberg, G. and Sandberg, R.** (2014b). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040.
- Pikaard, C. S. and Mittelsten Scheid, O.** (2014). Epigenetic regulation in plants. *Cold Spring Harb. Perspect. Biol.* **6**, a019315.
- Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A. and Trapnell, C.** (2017). Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315.
- Riechmann, J. L.** (2002). Transcriptional regulation: a genomic overview. *Arabidopsis Book* **1**, e0085.
- Ryu, K. H., Huang, L., Kang, H. M. and Schiefelbein, J.** (2019). Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiol.* **179**, 1444–1456.
- Satterlee, J. W., Strable, J. and Scanlon, M. J.** (2020). Plant stem-cell organization and differentiation at single-cell resolution. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 33689–33699.
- Schon, M. A. and Nodine, M. D.** (2017). Widespread Contamination of Arabidopsis Embryo and Endosperm Transcriptome Data Sets. *Plant Cell* **29**, 608–617.
- Shimotohno, A., Heidstra, R., Blilou, I. and Scheres, B.** (2018). Root stem cell niche organizer specification by molecular convergence of PLETHORA and SCARECROW transcription factor modules. *Genes Dev.* **32**, 1085–1100.
- Shulse, C. N., Cole, B. J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G. M., Zhu, Y., O'Malley, R. C., Brady, S. M., et al.** (2019). High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. *Cell Rep.* **27**, 2241–2247.e4.
- Slane, D., Kong, J., Berendzen, K. W., Kilian, J., Henschen, A., Kolb, M., Schmid, M., Harter, K., Mayer, U., De Smet, I., et al.** (2014). Cell type-specific transcriptome analysis in the early Arabidopsis thaliana embryo. *Development* **141**, 4831–4840.

- Song, Q., Ando, A., Jiang, N., Ikeda, Y. and Chen, Z. J.** (2020). Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. *Genome Biol.* **21**, 178.
- Tadros, W. and Lipshitz, H. D.** (2009). The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042.
- Tang, W. and Tang, A. Y.** (2019). Biological significance of RNA-seq and single-cell genomic research in woody plants. *Res. J. For.* **30**, 1555–1568.
- ten Hove, C. A., Lu, K.-J. and Weijers, D.** (2015). Building a plant: cell fate specification in the early Arabidopsis embryo. *Development* **142**, 420–430.
- Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J. and Gao, G.** (2020). PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* **48**, D1104–D1113.
- Tucker, M. R., Hinze, A., Tucker, E. J., Takada, S., Jürgens, G. and Laux, T.** (2008). Vascular signalling mediated by ZWILLE potentiates WUSCHEL function during shoot meristem stem cell development in the Arabidopsis embryo. *Development* **135**, 2839–2843.
- Walser, C. B. and Lipshitz, H. D.** (2011). Transcript clearance during the maternal-to-zygotic transition. *Curr. Opin. Genet. Dev.* **21**, 431–443.
- Welch, D., Hassan, H., Blilou, I., Immink, R., Heidstra, R. and Scheres, B.** (2007). Arabidopsis JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. *Genes Dev.* **21**, 2196–2204.
- Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J. D., Pai, A., Song, C., Zhuang, M., Klasfeld, S., He, C., et al.** (2017). Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat. Genet.* **49**, 1546–1552.
- Zhang, T.-Q., Xu, Z.-G., Shang, G.-D. and Wang, J.-W.** (2019). A Single-Cell RNA Sequencing Profiles the Developmental Landscape of Arabidopsis Root. *Mol. Plant* **12**, 648–660.
- Zhao, J., Xin, H., Qu, L., Ning, J., Peng, X., Yan, T., Ma, L., Li, S. and Sun, M.-X.** (2011). Dynamic changes of transcript profiles after fertilization are associated with *de novo* transcription and maternal elimination in tobacco zygote, and mark the onset of the maternal-to-zygotic transition. *Plant J.* **65**, 131–145.
- Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al.** (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* **50**, 278–284.
- Zhao, P., Zhou, X., Shen, K., Liu, Z., Cheng, T., Liu, D., Cheng, Y., Peng, X. and Sun, M.-X.** (2019). Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions. *Dev. Cell* **49**, 882–893.e5.
- Zhou, X., Liu, Z., Shen, K., Zhao, P. and Sun, M.-X.** (2020). Cell lineage-specific transcriptome analysis for interpreting cell fate specification of preembryos. *Nat. Commun.* **11**, 1366.

9 Discussion

To reveal whether transcriptional activity is required and if so, when it is required during Arabidopsis embryogenesis, I adapted several techniques to resolve the previous controversy. While genetics and parent-of-origin expression analyses on individual genes showed examples either supporting maternal dominance or zygotic dominance in early embryogenesis, there was no comprehensive large scale analyses until recently (Zhao *et al.*, 2019). Nonetheless, functional analyses are required to connect the transcriptome changes to physiological events. With modified ExM technique, I showed that the active transcription marks phosphorylation of serine 2 on the carboxy-terminal domain of RNA polymerase II (RNAPII Ser2P) and trimethylation on histone H3 lysine 36 (H3K36me3) are abundant in zygotes soon after fertilization and before the asymmetric zygotic division. The observations provided unambiguous evidence for large scale transcriptional activities which is consistent with recently published transcriptomes (Zhao *et al.*, 2019). This is also the first successful report of ExM application on plant materials, and the introduction of ExM to plants provides a superior option to visualize cellular events or proteins of interest in versatile tissues. I also applied the ExM to visualize the sizes of nucleoli and nuclei to evaluate the compactness of chromatin in zygotes. The reduced chromocenter and increased nucleoli in zygotes suggested less compacted chromatin. Together with the methylome data, the observations suggested post-fertilization heterochromatin decondensation and potential promotion of the production of 24-nt siRNA which contributes to the silencing of transposable elements (TEs). On the other hand, the modified seed culturing and live-cell imaging system for chemical treatment showed that the zygotes ceased to develop when culturing with RNAPII inhibitor triptolide. Similar to the classical experiments done in other species (Braude *et al.*, 1979; Newport and Kirschner, 1982; Edgar, Wolf and Wood, 1994; Edgar and Datar, 1996; Zamir, Kam and Yarden, 1997), our chemical treatments indicated that inherited genetic products are not sufficient to support early embryogenesis and active transcription is required for the asymmetric zygotic division. The cytological and physiological evidence were consistent with the observed post-fertilization transcriptome change (Zhao *et al.*, 2019) and provided unambiguous evidence to settle previously conflicting reports. While the two parental genomes contributed equally to the zygotic transcriptome globally, a subset of genes were preferentially expressed in zygotes or specific cell lineages (Zhao *et al.*, 2020). This suggested imprinting or other epigenetic regulations may target these genes, but how and why the expression of these genes had parental bias remained to be answered. While it is known that a subset of methylome resets during embryogenesis (Papareddy *et al.*, 2020) and the replacement of histone 3 after fertilization could contribute to epigenetic resetting (Ingouff *et al.*, 2010), further investigations are required to reveal the epigenetic dynamics and regulation mechanisms in pre- and post-fertilization cells.

To reveal the gene expression diversities in different embryonic cell types, a modified single nucleus RNA-seq system with low instrumental and technical barriers was introduced and acquired contamination-free embryonic transcriptomes at single cell resolution. Embryonic cell types at the globular stage were identified and validated, and representative expression profiles for the cell types were constructed for reference. Further analyses revealed the variations of epigenetic regulators across embryonic cell types. The expression patterns of siRNA-related genes suggested that the extraembryonic suspensors could be a production source of 24-nt siRNA, which might be transported to embryo proper and act as a surveillance system to silence TE and maintain genome integrity. The preferential expressions of DNA methylase and demethylases in different cell lineages suggested the cytosine methylation associated with gene promoter and TE silencing contributed to the epigenetic differences across cell types. Equally detailed cell type-specific methylomes are needed to investigate the epigenetic variations across embryonic cell types. Moreover, we identified transcription factor binding motifs enriched in different cell types and observed co-occurrence of antagonistic motifs which might serve as a dual gating system for embryonic genes. We demonstrated our single nucleus transcriptomes were effective in revealing the diversity among embryonic cell types, and more focus analyses could reveal more regulatory mechanisms in the future. I anticipate this study to be a cornerstone and facilitate plant embryogenesis studies.

10 References

- Abiko, M. *et al.* (2013) ‘Gene expression profiles in rice gametes and zygotes: identification of gamete-enriched genes and up- or down-regulated genes in zygotes after fertilization’, *Journal of experimental botany*, 64(7), pp. 1927–1940.
- Anderson, S. N. *et al.* (2013) ‘Transcriptomes of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization’, *The Plant journal: for cell and molecular biology*, 76(5), pp. 729–741.
- Anderson, S. N. *et al.* (2017) ‘The Zygotic Transition Is Initiated in Unicellular Plant Zygotes with Asymmetric Activation of Parental Genomes’, *Developmental cell*, 43(3), pp. 349–358.e4.
- Armenta-Medina, A. and Gillmor, C. S. (2019) ‘Genetic, molecular and parent-of-origin regulation of early embryogenesis in flowering plants’, *Current topics in developmental biology*, 131, pp. 497–543.
- Autran, D. *et al.* (2011) ‘Maternal Epigenetic Pathways Control Parental Contributions to Arabidopsis Early Embryogenesis’, *Cell*, 145(5), pp. 707–719.
- Bai, S.-N. (2015) ‘The concept of the sexual reproduction cycle and its evolutionary significance’, *Frontiers in plant science*, 6, p. 11.
- Baroux, C. *et al.* (2008) ‘The maternal to zygotic transition in animals and plants’, *Cold Spring Harbor symposia on quantitative biology*, 73, pp. 89–100.
- Beddington, R. S. and Robertson, E. J. (1999) ‘Axis development and early asymmetry in mammals’, *Cell*, 96(2), pp. 195–209.
- Belmonte, M. F. *et al.* (2013) ‘Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(5), pp. E435–44.
- Braude, P. *et al.* (1979) ‘Post-transcriptional control in the early mouse embryo’, *Nature*, 282(5734), pp. 102–105.
- Capron, A. *et al.* (2009) ‘Embryogenesis: pattern formation from a single cell’, *The Arabidopsis book / American Society of Plant Biologists*, 7, p. e0126.
- Chen, F., Tillberg, P. W. and Boyden, E. S. (2015) ‘Optical imaging. Expansion microscopy’, *Science*, 347(6221), pp. 543–548.
- Chen, J. *et al.* (2017) ‘Zygotic Genome Activation Occurs Shortly after Fertilization in Maize’, *The Plant cell*, 29(9), pp. 2106–2125.
- Clark, N. L., Aagaard, J. E. and Swanson, W. J. (2006) ‘Evolution of reproductive proteins from animals and plants’, *Reproduction*, 131(1), pp. 11–22.
- Del Toro-De León, G., García-Aguilar, M. and Gillmor, C. S. (2014) ‘Non-equivalent contributions of maternal and paternal genomes to early plant embryogenesis’, *Nature*, 514(7524), pp. 624–627.
- Domoki, M. *et al.* (2013) ‘Identification of genes preferentially expressed in wheat egg cells and zygotes’, *Plant cell reports*, 32(3), pp. 339–348.
- Dresselhaus, T. *et al.* (1999) ‘Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles’, *MGG - Molecular and General Genetics*, 261(2), p. 416.
- Edgar, B. A. and Datar, S. A. (1996) ‘Zygotic degradation of two maternal Cdc25 mRNAs terminates

- Drosophila's early cell cycle program', *Genes & development*, 10(15), pp. 1966–1977.
- Edgar, L. G., Wolf, N. and Wood, W. B. (1994) 'Early transcription in *Caenorhabditis elegans* embryos', *Development*, 120(2), pp. 443–451.
- Galbraith, D. W. *et al.* (1983) 'Rapid flow cytometric analysis of the cell cycle in intact plant tissues', *Science*, 220(4601), pp. 1049–1051.
- García-Aguilar, M. and Gillmor, C. S. (2015) 'Zygotic genome activation and imprinting: parent-of-origin gene regulation in plant embryogenesis', *Current opinion in plant biology*, 27, pp. 29–35.
- Gerri, C. *et al.* (2020) 'Human Embryogenesis: A Comparative Perspective', *Annual review of cell and developmental biology*, 36, pp. 411–440.
- Gooh, K. *et al.* (2015) 'Live-cell imaging and optical manipulation of *Arabidopsis* early embryogenesis', *Developmental cell*, 34(2), pp. 242–251.
- Guo, L. *et al.* (2016) 'The anaphase-promoting complex initiates zygote division in *Arabidopsis* through degradation of cyclin B1', *The Plant journal: for cell and molecular biology*, 86(2), pp. 161–174.
- Habib, N. *et al.* (2016) 'Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons', *Science*, 353(6302), pp. 925–928.
- Hofmann, F., Schon, M. A. and Nodine, M. D. (2019) 'The embryonic transcriptome of *Arabidopsis thaliana*', *Plant reproduction*, 32(1), pp. 77–91.
- Horstman, A., Bemer, M. and Boutilier, K. (2017) 'A transcriptional view on somatic embryogenesis', *Regeneration (Oxford, England)*, 4(4), pp. 201–216.
- ten Hove, C. A., Lu, K.-J. and Weijers, D. (2015) 'Building a plant: cell fate specification in the early *Arabidopsis* embryo', *Development*, 142(3), pp. 420–430.
- Huijser, P. and Schmid, M. (2011) 'The control of developmental phase transitions in plants', *Development*, 138(19), pp. 4117–4129.
- Hwang, B., Lee, J. H. and Bang, D. (2018) 'Single-cell RNA sequencing technologies and bioinformatics pipelines', *Experimental & molecular medicine*, 50(8), p. 96.
- Ingouff, M. *et al.* (2010) 'Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in *Arabidopsis*', *Current biology: CB*, 20(23), pp. 2137–2143.
- Kao, P. and Nodine, M. D. (2019) 'Transcriptional Activation of *Arabidopsis* Zygotes Is Required for Initial Cell Divisions', *Scientific reports*, 9(1), p. 17159.
- Lee, M. T., Bonneau, A. R. and Giraldez, A. J. (2014) 'Zygotic Genome Activation During the Maternal-to-Zygotic Transition', *Annual review of cell and developmental biology*, 30(1), pp. 581–613.
- Linkies, A. *et al.* (2010) 'The evolution of seeds', *The New phytologist*, 186(4), pp. 817–831.
- Lin, Z. *et al.* (2007) 'AtCDC5 regulates the G2 to M transition of the cell cycle and is critical for the function of *Arabidopsis* shoot apical meristem', *Cell research*, 17(9), pp. 815–828.
- Liu, C.-M. and Meinke, D. W. (1998) 'The titan mutants of *Arabidopsis* are disrupted in mitosis and cell cycle control during seed development', *The Plant journal: for cell and molecular biology*, 16(1), pp. 21–31.
- Long, J. A. and Barton, M. K. (1998) 'The development of apical embryonic pattern in *Arabidopsis*',

Development, 125(16), pp. 3027–3035.

Meinke, D. *et al.* (2008) ‘Identifying essential genes in *Arabidopsis thaliana*’, *Trends in plant science*, 13(9), pp. 483–491.

Meinke, D. W. (2019) ‘Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in *Arabidopsis*’, *The New phytologist*. doi: 10.1111/nph.16071.

Meyer, S. and Scholten, S. (2007) ‘Equivalent Parental Contribution to Early Plant Zygotic Development’, *Current biology: CB*, 17(19), pp. 1686–1691.

Newport, J. and Kirschner, M. (1982) ‘A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage’, *Cell*, 30(3), pp. 675–686.

Niedojadło, K. *et al.* (2012) ‘Transcriptional activity of *Hyacinthus orientalis* L. female gametophyte cells before and after fertilization’, *Planta*, 236(1), pp. 153–169.

Ning, J. *et al.* (2006) ‘Differential gene expression in egg cells and zygotes suggests that the transcriptome is restructured before the first zygotic division in tobacco’, *FEBS letters*, 580(7), pp. 1747–1752.

Nodine, M. D. and Bartel, D. P. (2012) ‘Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos’, *Nature*, 482(7383), pp. 94–97.

Ohnishi, Y. *et al.* (2019) ‘Sperm Entry into the Egg Cell Induces the Progression of Karyogamy in Rice Zygotes’, *Plant & cell physiology*. doi: 10.1093/pcp/pcz077.

Ohnishi, Y., Hoshino, R. and Okamoto, T. (2014) ‘Dynamics of Male and Female Chromatin during Karyogamy in Rice Zygotes’, *Plant physiology*, 165(4), pp. 1533–1543.

Okamoto, T. *et al.* (2005) ‘Identification of Genes that are Up- or Down-regulated in the Apical or Basal Cell of Maize Two-celled Embryos and Monitoring their Expression During Zygote Development by a Cell Manipulation- and PCR-based Approach’, *Plant & cell physiology*, 46(2), pp. 332–338.

Palovaara, J. *et al.* (2017) ‘Transcriptome dynamics revealed by a gene expression atlas of the early *Arabidopsis* embryo’, *Nature Plants*, 3(11), pp. 894–904.

Papareddy, R. K. *et al.* (2020) ‘Chromatin regulates expression of small RNAs to help maintain transposon methylome homeostasis in *Arabidopsis*’, *Genome biology*, 21(1), p. 251.

Park, S. and Harada, J. J. (2008) ‘*Arabidopsis* Embryogenesis’, in Suárez, M. F. and Bozhkov, P. V. (eds) *Plant Embryogenesis*. Totowa, NJ: Humana Press, pp. 3–16.

Picelli, S., Faridani, O. R., *et al.* (2014) ‘Full-length RNA-seq from single cells using Smart-seq2’, *Nature protocols*, 9(1), pp. 171–181.

Picelli, S., Björklund, A. K., *et al.* (2014) ‘Tn5 transposase and tagmentation procedures for massively scaled sequencing projects’, *Genome research*, 24(12), pp. 2033–2040.

Pięciński, S. *et al.* (2008) ‘Changes in poly(A) RNA and TMG snRNA distribution in the embryo sac of *Hyacinthus orientalis* L. before and after fertilization’, *Sexual plant reproduction*, 21(4), pp. 247–257.

Pillot, M. *et al.* (2010) ‘Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in *Arabidopsis*’, *The Plant cell*, 22(2), pp. 307–320.

Ronceret, A. *et al.* (2005) ‘Genetic analysis of two *Arabidopsis* DNA polymerase epsilon subunits during early embryogenesis’, *The Plant journal: for cell and molecular biology*, 44(2), pp. 223–236.

- Ronceret, A. *et al.* (2008) 'The alpha-N-acetyl-glucosaminidase gene is transcriptionally activated in male and female gametes prior to fertilization and is essential for seed development in Arabidopsis', *Journal of experimental botany*, 59(13), pp. 3649–3659.
- Ronceret, A. *et al.* (2008) 'The first zygotic division in Arabidopsis requires de novo transcription of thymidylate kinase', *The Plant journal: for cell and molecular biology*, 53(5), pp. 776–789.
- Sauter, M. *et al.* (1998) 'Cell cycle regulatory genes from maize are differentially controlled during fertilization and first embryonic cell division', *Sexual plant reproduction*, 11(1), pp. 41–48.
- Schmidt, A., Schmid, M. W. and Grossniklaus, U. (2015) 'Plant germline formation: common concepts and developmental flexibility in sexual and asexual reproduction', *Development*, 142(2), pp. 229–241.
- Schon, M. A. and Nodine, M. D. (2017) 'Widespread Contamination of Arabidopsis Embryo and Endosperm Transcriptome Data Sets', *The Plant cell*, 29(4), pp. 608–617.
- Shi, L. and Wu, J. (2009) 'Epigenetic regulation in mammalian preimplantation embryo development', *Reproductive biology and endocrinology: RB&E*, 7, p. 59.
- Slane, D. *et al.* (2014) 'Cell type-specific transcriptome analysis in the early Arabidopsis thaliana embryo', *Development*, 141(24), pp. 4831–4840.
- Sprunck, S. *et al.* (2005) 'The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.)', *The Plant journal: for cell and molecular biology*, 41(5), pp. 660–672.
- St Johnston, D. and Nüsslein-Volhard, C. (1992) 'The origin of pattern and polarity in the Drosophila embryo', *Cell*, 68(2), pp. 201–219.
- Tadros, W. and Lipshitz, H. D. (2009) 'The maternal-to-zygotic transition: a play in two acts', *Development*, 136(18), pp. 3033–3042.
- Tillberg, P. W. *et al.* (2016) 'Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies', *Nature biotechnology*, 34(9), pp. 987–992.
- Vielle-Calzada, J.-P., Baskar, R. and Grossniklaus, U. (2000) 'Delayed activation of the paternal genome during seed development', *Nature*, 404(6773), pp. 91–94.
- Walbot, V. and Evans, M. M. S. (2003) 'Unique features of the plant life cycle and their consequences', *Nature reviews. Genetics*, 4(5), pp. 369–379.
- Walser, C. B. and Lipshitz, H. D. (2011) 'Transcript clearance during the maternal-to-zygotic transition', *Current opinion in genetics & development*, 21(4), pp. 431–443.
- Wibowo, A. *et al.* (2018) 'Partial maintenance of organ-specific epigenetic marks during plant asexual reproduction leads to heritable phenotypic variation', *Proceedings of the National Academy of Sciences of the United States of America*, 115(39), pp. E9145–E9152.
- Willemsen, V. and Scheres, B. (2004) 'Mechanisms of pattern formation in plant embryogenesis', *Annual review of genetics*, 38, pp. 587–614.
- Wójcikowska, B. and Gaj, M. D. (2016) 'Somatic Embryogenesis in Arabidopsis', in Loyola-Vargas, V. M. and Ochoa-Alejo, N. (eds) *Somatic Embryogenesis: Fundamental Aspects and Applications*. Cham: Springer International Publishing, pp. 185–199.
- Xu, J. *et al.* (2005) 'EMBRYONIC FACTOR 1 encodes an AMP deaminase and is essential for the zygote to embryo transition in Arabidopsis', *The Plant journal: for cell and molecular biology*, 42(5),

pp. 743–756.

Yu, D. *et al.* (2012) ‘EMBRYONIC FACTOR 19 encodes a pentatricopeptide repeat protein that is essential for the initiation of zygotic embryogenesis in Arabidopsis’, *Journal of integrative plant biology*, 54(1), pp. 55–64.

Yu, T.-Y. *et al.* (2016) ‘The Arabidopsis Receptor Kinase ZAR1 Is Required for Zygote Asymmetric Division and Its Daughter Cell Fate’, *PLoS genetics*. Edited by G. P. Copenhaver, 12(3), p. e1005933.

Zamir, E., Kam, Z. and Yarden, A. (1997) ‘Transcription-dependent induction of G1 phase during the zebra fish midblastula transition’, *Molecular and cellular biology*, 17(2), pp. 529–536.

Zhao, J. *et al.* (2011) ‘Dynamic changes of transcript profiles after fertilization are associated with *de novo* transcription and maternal elimination in tobacco zygote, and mark the onset of the maternal-to-zygotic transition’, *The Plant journal: for cell and molecular biology*, 65(1), pp. 131–145.

Zhao, P. *et al.* (2019) ‘Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions’, *Developmental cell*, 49(6), pp. 882–893.e5.

Zhao, P. *et al.* (2020) ‘Equal parental contribution to the transcriptome is not equal control of embryogenesis’, *Nature plants*. doi: 10.1038/s41477-020-00793-x.

Zhou, X. *et al.* (2020) ‘Cell lineage-specific transcriptome analysis for interpreting cell fate specification of proembryos’, *Nature communications*, 11(1), p. 1366.

Zimmerman, J. L. (1993) ‘Somatic Embryogenesis: A Model for Early Development in Higher Plants’, *The Plant cell*, 5(10), pp. 1411–1423.