CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL

INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

Activación del genoma cigótico en *Arabidopsis Thaliana*

Tesis que presenta:

José Jaime Alaniz Fabián

Para obtener el grado de:

Maestría en ciencias

En la especialidad de:

Biotecnología de plantas

Director de la tesis: Dr. C. Stewart Gillmor III

Irapuato, Guanajuato Febrero de 2019

CENTER FOR RESEARCH AND ADVANCED STUDIES OF THE NATIONAL POLYTECHNIC INSTITUTE

IRAPUATO UNIT

Zygotic genome activation in *Arabidopsis Thaliana*

Thesis presented by:

José Jaime Alaniz Fabián

To obtain the degree of:

Master of science

Specializing in:

Plant biotechnology

Advisor: Dr. C. Stewart Gillmor III

Irapuato, Guanajuato February, 2019

Este trabajo fue realizado en el Laboratorio de Desarrollo y Morfogénesis de Plantas bajo la dirección del Dr. Stewart Gillmor en el departamento del Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO) del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN)

Comité tutorial

Dra. Katarzyna Oktaba Sosin

Dr. Cei Abreu Goodger

Agradecimientos

Al consejo nacional de ciencia y tecnología (CONACyT), por el financiamiento durante la realización del proyecto.

Al centro de investigación y estudios avanzados del IPN (CINVESTAV), por la estructura y administración del posgrado.

Al laboratorio nacional de genómica para la biodiversidad (Langebio, Unidad de Genómica Avanzada), por facilitar los equipos, recursos, servicios y estructura que permitieron la realización del proyecto.

Al Dr. Stewart Gillmor, por la dirección del proyecto y la libertad con que me ha otorgado constantes oportunidades de crecimiento.

A la Dra. Daphné Autran, por su contribución intelectual a mi formación.

Al Dr. Gerardo Del Toro, por su asesoría teórica y técnica.

A los miembros de mi comité tutorial: Dra. Katarzyna Oktaba Sosin y Dr. Cei Abreu Goodger, por su asesoría, críticas y discusiones, que han aportado robustez al proyecto como sin duda al resto del posgrado.

1e- Resumen

En las angiospermas, el gametofito femenino contiene 2 gametos: la célula central, que al ser fecundada da origen al endospermo (un tejido que rodea y nutre al embrión durante su desarrollo), y la célula huevo, que da origen al embrión. La embriogénesis es sostenida por la alocación de recursos maternos al nuevo individuo y, adicionalmente, durante el desarrollo temprano, por productos que la madre ha depositado en la célula huevo antes de la fecundación, por lo que se dice que la embriogénesis temprana ocurre bajo control materno. La emancipación del embrión se conoce como transición materno-cigótica y requiere la concusión de 2 procesos: la degradación del restante de los productos depositados en la célula huevo, y el arranque de la producción de los productos propios (activación del genoma cigótico). En *Arabidopsis thaliana,* se cree que el embrión es transcripcionalmente activo desde la fecundación y los productos cigóticos y maternos coexisten en el embrión hasta el agotamiento o degradación de estos últimos en la etapa globular (aproximadamente 3 días después de la fecundación). La distinción de los productos cigóticos de aquellos heredados por la célula huevo es crucial para el estudio de la activación del genoma cigótico y las contribuciones parentales a ésta, no obstante, esta distinción ha resultado sumamente compleja. Sumado a esto, distintos estudios transcriptómicos en etapas pre globulares han generado resultados contradictorios en cuanto a la abundancia relativa de transcritos maternos y paternos; las diferencias en estos estudios han sido atribuidas a contaminación de las muestras por los tejidos que rodean al embrión y a efectos de hibridación generados por los distintos ecotipos de los parentales usados. En este trabajo, utilizando genes reporteros y parentales isogénicos, se ha corroborado el sesgo parental observado en estudios transcriptómicos, excluyendo la contaminación e hibridación como origen del mismo. Se ha explorado también la expresión de los genes reporteros antes y después de la fecundación (en célula huevo y embrión), con resultados aún no conclusivos. Por último, se ha demostrado la utilidad de una versión inestable del gen reportero YFP (proteína amarilla fluorescente por sus siglas en inglés) como herramienta para distinguir entre productos maternalmente depositados y cigóticos.

1- Abstract

In angiosperms, the female gametophyte contains 2 gametes: the central cell, which gives rise to the endosperm (an embryo nourishing tissue) when fertilized, and the egg cell, which gives rise to the embryo. Embryogenesis is sustained by allocation of maternal resources to the seed, and, additionally, during early development, by maternal products deposited in the egg cell prior to fertilization, for this reason it is said that early embryogenesis is under maternal control. The emancipation of the embryo is known as the maternal to zygotic transition and consist of 2 individual processes: the degradation of the maternally-deposited products, and the onset of zygotic genome activity (zygotic genome activation). In *Arabidopsis Thaliana*, it is thought that the embryo is transcriptionally active immediately after fertilization, and maternal and zygotic products coexist in the embryo until the exhaustion or degradation of carried-over products at the globular stage (approximately 3 days after fertilization). Distinction of zygotic and egg-deposited products is crucial for the study of zygotic genome activation and parental contributions to it, nevertheless, this distinction has proven challenging. Further, different transcriptomic studies on preglobular embryos have generated

conflicting results regarding the relative abundance of maternal and paternal transcripts; differences in these results have been attributed to contamination of the embryo samples by surrounding tissues and to hybridization effects generated by the distinct ecotypes used in these studies. In this work, using reporter genes and isogenic parentals, I have corroborated the parental bias observed in transcriptomic studies, while excluding contamination and hybridization as the cause of such bias. Additionally, the expression of reporter genes before and after fertilization (in the egg cell and embryo) has been explored, with results so far inconclusive. Lastly, the functionality of an unstable version of the YFP (yellow fluorescent protein) reporter gene to distinguish between maternally-inherited and zygotic products has been shown.

2- Introduction

Sexual reproduction in plants and animals occurs when gametes fuse to produce the zygote. The egg and the sperm are highly differentiated cell types whose epigenetic states and cytoplasmic contents must be reconciled in the zygote, so that a new individual can begin the cell division and differentiation required for early embryo development.

2.1- The maternal-to-zygotic transition (MZT) and zygotic genome activation (ZGA) in animals

During the first cell cycles after fertilization in animals, embryo development is supported by maternal products deposited in the egg. This maternal support continues until the maternal to zygotic transition (MZT), which is defined by two events: zygotic genome activation (ZGA), when the embryo starts transcribing genes that are required for early development; and degradation of maternally-deposited mRNAs (Telford *et al.*, 2008; Tadros and Lipshitz, 2009; Jukam et al., 2017). Notably, the mammalian sperm has also been shown to provide the zygote with critical development regulators such as mRNAs, miRNAs (microRNAs), NAT siRNAs (natural antisense short interfering RNA) and centrosomes (reviewed by Kumar *et al.*, 2013.)

An extreme example of maternal support of early development is the sea urchin *Arbacia punctulata,* whose eggs (which are laid externally) are able to develop without fertilization. Surprisingly, even after removal of egg nuclei (using centrifugal force to break apart the nucleated fraction of the egg), *A. punctulata* embryos are able to develop for 2 days, forming up to 500 cells (Harvey, 1936). However, later experiments measuring incorporation of tritium-labeled uridine into RNA showed that zygotic transcription was detected before the end of the first cell cycle (Poccia *et al.*, 1985). Thus, *A. punctulata* is an example of an organism where zygotic transcription begins long before it is actually required to support development.

Translation and degradation of maternally-deposited mRNAs must be regulated to supply the required products at the right moment through a number of cell cycles. In *Drosophila,* this regulation is carried out by the RNA-binding proteins ME31B (Wang *et al.*, 2017) and SMAUG (Tadros *et al.*, 2007). Depending on their interacting proteins, ME31B and SMAUG can either protect transcripts from degradation, block their translation, or promote their clearance (Tadros *et al.*, 2007; Wang *et al.*, 2017). Other factors influencing the translation of maternal transcripts in animals are “codon optimality” (the preference for certain codons over others during translation) (Radhakrishnan *et al.*, 2016), and polyadenylation, which has been found

to be a widespread mechanism for regulation of mRNA translation and degradation before zygotic genome activation (ZGA) (Aanes *et al.*, 2011; Harvey *et al.*, 2013) (reviewed by Avilés-Pagán and Orr-Weaver, 2018). In zebrafish, a conserved microRNA, MiR-430, is transcribed in the embryo and promotes degradation of maternally-deposited transcripts at the MZT (Giraldez *et al.*, 2006), while also inhibiting the translation of its targets (Bazzini et al., 2012). In *Drosophila*, miR-309 plays a similar role (Bushati et al., 2008).

During the first cell divisions the zygotic genome is kept transcriptionally quiescent, most likely by repressive histone variants and modifications. However, the transcriptional machinery is competent, as shown by the quick transcription of microinjected plasmids before ZGA in *Xenopus*, mouse and zebrafish. These microinjected plasmids are subsequently silenced, presumably by the same mechanisms that silences transcription of endogenous genes until ZGA, when the embryo is released from its transcriptional repression (reviewed by Lee *et al., 2014*.)

In vertebrates, activation of the zygotic genome has been shown to be regulated by the nuclear-to-cytoplasmic ratio, which can act through titration of components such as histones by the increasing amount of DNA. This ratio exponentially increases as nuclei undergo division without increase in the embryo size. Evidence for this mechanism in various animals has been provided by experiments that increase or decrease cell-volume and ploidy, with concomitant effects on the timing of ZGA (reviewed by Jukam et al., 2017). It has also been proposed that ZGA is regulated in a time-dependent manner. In this model, a continuous process, such as the translation of maternal mRNA, the polyadenylation of mRNAs, or the import of transcription factors to the nucleus eventually reaches a critical point, triggering ZGA (reviewed by Jukam et al., 2017). It must be noted that the nuclear-to-cytoplasmic ratio and time-dependent models are not mutually exclusive. Consistent with time-dependent regulation, the mRNA of the Zelda zinc-finger transcription factor, which binds a motif present in many early-transcribed genes in *Drosophila*, is detected at high levels in the female germline and young embryo before the ZGA (Liang *et al.*, 2008).

Interestingly, the earliest transcribed genes in *Drosophila* and zebrafish are short and intron poor, most likely as a consequence of the constraints imposed by the brief cell cycles (8 and 15 min, respectively) before the mid-blastula transition (Heyn *et al.*, 2014). This constraint is explained by the fact that progression of the cell cycle through mitosis leads to the abortion of nascent transcripts (Shermoen and O'Farrell, 1991).

The timing of the two events defining the MZT vary greatly between species, and do not always happen simultaneously, nor in the same order. By inhibiting transcription in animal embryos, the MZT has been defined as the development point were the embryo arrests, which ranges from one cell cycle for the mouse embryo to 15 cycles in *Xenopus* (reviewed by Baroux *et al.,* 2008). Nevertheless, the beginning of transcription in these species has been detected earlier than the arrest in its absence; i.e. immediately after fertilization in mouse and as early as 8-cell stage in *Xenopus* (reviewed by Jukam et al., 2017), and for the special case of the mitochondrial transcripts in zebrafish (and probably in other species), immediately after fertilization (Heyn *et al.*, 2014).

2.2- The MZT and ZGA in plants

In flowering plants, the female gametophyte consists of seven cells: two gametes (the homodiploid central cell and the haploid egg cell), and five accessory cells (two synergids and three antipodal cells). The male gametophyte is composed of three haploid cells: two sperm cells and one accessory cell (the vegetative cell). One sperm fuses with the egg to form the zygote, and the second sperm fuses with the central cell to form the endosperm (an embryo-nourishing tissue). In contrast to animals, plant zygotes are transcriptionally active, as shown by recessive mutations affecting the first cell division in the progeny of heterozygous plants (Mayer et al., 1993; Lukowitz *et al.*, 2004; Ueda et al., 2011), and the large number of transcripts that increase during early embryo development, many of which are also not detected in the egg cell (Autran *et al.*, 2011; Nodine and Bartel, 2012; Zhou *et al.*, 2014; Anderson *et al.*, 2017; Chen *et al.*, 2017).

Similar to animals, zygotic transcription in plants may not be required as soon as it is detected. This was shown by inhibition of RNA Pol-II through RNAi, where 85% of embryos made it to the octant stage, and 60% past the globular stage (3 days after pollination), indicating that egg cell products are sufficient to support the first divisions. In contrast, the endosperm stayed at the 1 nucleus stage 40% of the time, indicating higher transcriptional requirements for this tissue (Pillot *et al.*, 2010).

Knowledge of translational regulation of egg-deposited transcripts in plants is limited to a single study of maize-*Tripsacum* hybrids. Microarray analysis found no detectable differences in transcript populations between mature ovules, and ovules with apomictic pro-embryos, suggesting that very little transcription occurs during the first few days of development of apomictic embryos. However, the ratio of polyadenylated RNA to total RNA did increase during this period, suggesting that immature transcripts in the egg cell are gradually polyadenylated to promote translation and support development (Grimanelli *et al.*, 2005). Other studies have suggested the existence of this phenomena but gave no insights on its regulation. For example, an experiment using suppression subtractive hybridization and mirror orientation selection (a method that allows analysis of differences in very similar samples by subtracting the common transcripts by hybridizing them to an excess of the reference cDNA) allowed detection of differentially expressed transcripts, suggesting that some egg-deposited transcripts are actively degraded (Ning *et al.*, 2006). The existence of active degradation is also supported by significantly down-regulated genes in rice zygotes vs egg cells (Anderson *et al.*, 2017).

The extent of the paternal contributions to the plant zygote has not been extensively characterized, but transcripts present in the sperm or pollen, but not the egg cell, have been detected in tobacco zygotes at 96 HAP (hours after pollination) (Ning *et al.*, 2006), similarly, the GUS reporter protein was detected in pollen tubes and zygotes of arabidopsis (Xiang *et al.*, 2011), suggesting that some degree of sperm carry over may exist. Interestingly, some of the most highly detected transcripts in mature sperm cells were not detectable in the zygote at 60 or at 96 HAP (equivalent to 10 and 46 hours after fertilization). In contrast, some transcripts persisted despite not being so abundant in the sperm cell, suggesting some level of selection for carried-over transcripts (Xin *et al.*, 2011). These potentially carried-over

transcripts are thought to have a role in zygotic genome activation and/or early embryogenesis (reviewed by Luo *et al*., 2014). Indeed, some cytoplasmic content of the sperm cell seems to be required for successful karyogamy, since the injection of transcriptionally competent sperm nuclei into an egg won't produce karyogamy, unlike the electrically induced fusion of the sperm and egg cells (Matthews-Rochon *et al.*, 1994). One clear paternal contribution is mRNA of the *SHORT SUSPENSOR* gene, which is present in the sperm cell, but only translated in the zygote, triggering zygote elongation (Bayer *et al.*, 2009).

2.2.1- The effect of hybridization on maternal regulation of embryogenesis in plants

Surprisingly, one parent-of-origin study in early embryogenesis of Arabidopsis found that both parental genomes contribute equally to the global transcriptomic profile (Nodine and Bartel, 2012). This observation was apparently in conflict with a previous report that found primarily maternal transcripts in early embryos (Autran et al., 2011). This disagreement was later attributed to the different accessions used to produce the hybrid embryos required for parent-of-origin transcript profiling (Col x Cvi for Nodine and Bartel, and Col x Ler for Autran et al.). Using a functional test for paternal rescue of maternally-inherited mutations, Del Toro-De León et al. (2014) showed that the ecotypes of the parents could have a large effect on the activity of paternal alleles. In agreement with this, the ecotype used by Nodine and Bartel (Cvi), is placed as an outgroup when clustering with 150 accessions using either sequence polymorphisms or CG methylation polymorphisms (Schmitz et al., 2013).

A large number of hybridization effects have been described in plants and other species, including differences between parents and hybrid progeny in transcriptomic profiles, DNA methylation, non-coding RNA abundance, dosage sensitivity, and phenotypic differences, among others (reviewed by Bar-Zvi et al., 2017). Nevertheless, the extent of hybridization effects on parental contribution to early embryogenesis has not been assessed; and the equal contribution reported by Nodine and Bartel is still surprising, since it would imply, besides equal transcription from both alleles, either little to no carry-over from the egg cell or a quick degradation of carried-over transcripts after fertilization. Despite the different compaction levels of parental genomes, equal transcription could be expected if both gamete's chromatin were remodeled quickly after fertilization, as it is the case for histone H3 variants (Ingouff et al., 2010). Quick degradation of maternally-deposited transcripts could also be possible if, as in zebrafish (Giraldez et al., 2006), the onset of zygotic transcription triggered maternal clearing, and transcription began earlier in Col x Cvi hybrid embryos.

2.3- Imprinting

Besides the storage of products in the gametophyte, maternal (or paternal) control can also happen through mono-allelic or biased allelic expression in a parent-of-origin-dependent manner, a process known as imprinting. Imprinting can be driven by differences between parental alleles in DNA methylation, histone modifications, and non-coding RNAs. It has been mainly found in the placenta in animals and the endosperm in plants- both tissues serve as an interface between the mother and the embryo (reviewed by Rodrigues and Zilberman, 2015). This phenomenon evolved independently in plants and animals, suggesting its emergence in response to similar selective pressures and in a close relatedness to reproduction and placental habit (reviewed by Pires and Grossniklaus, 2014).

2.3.1- Non-equivalence of parental genomes

It was recently shown that the parental genomes are not functionally equivalent in rice zygotes. Using electrofusion of gametes to generate zygotes with different content of paternal and maternal genome copies, Okamoto's group showed the non equivalence of parental genomes, since polyploid embryos survived and developed normally more frequently if their parental nuclear DNA contents were balanced 1:1, or the excess was maternal (even with increased ploidy), than if they had a paternal excess (Toda, et al., 2018).

In mammals, nuclear transplantation experiments have also shown that parental genomes are not equivalent. In the most drastic example in mammals, maternal and paternal pro-nuclei are maintained in different compartments after gamete fusion until the four-cell embryo, allowing them to go through separate re-programming processes. Not surprisingly, these epigenetic differences cause uniparental mammalian embryos to be nonviable (unless particular differentially methylated regions are deleted), despite having a competent cytoplasm (reviewed by Haaf et al., 2004).

The driving force producing and maintaining uniparental (or biased) expression is under debate. Considering that uniparental expression exposes the individual to deleterious recessive mutations in the expressed alleles (sometimes referred to as functional haploidy), it is unlikely to have evolved independently in animals and plants without functional haploidy conferring more advantages than disadvantages (discussed by Pires and Grossniklaus, 2014). Several phenomena have been proposed to be driving biased expression and non-equivalence of parental alleles:

I. The parental conflict hypothesis proposes that, in species in which the mother allocates resources to the progeny of different males, any particular male might benefit from drawing more resources to its own offspring, while the mother would benefit from equally distributing the nutrients. This conflict would push for the evolution of mechanisms enhancing the expression of growth-promoting genes when they are inherited paternally and diminishing it when inherited maternally (discussed by Pires and Grossniklaus, 2014.)

II. The co-adaptation hypothesis argues that, for some traits to which both contribute, the maternal and embryo genomes are pressured to cooperate, for example to balance the birth weight in humans. For the genes involved, the selective pressure to co-adapt can generate maternally biased expression (discussed by Rodrigues & Zilberman, 2015.)

III. The cytonuclear interaction theory suggests that the predominant inheritance of maternal organelles may lead to imprinting of genes that interact with them because of the pressure to co-adapt (discussed by Spencer & Clark, 2014.)

IV. The ovarian time-bomb theory proposes that imprinting arose to prevent deleterious consequences of parthenogenetic development. Imprinting would make both genomes necessary and thus prevent development of unfertilized eggs (discussed by Spencer & Clark, 2014.)

It has also been proposed that the biased expression of many genes did not arise as a consequence of any selective pressure to do so, but rather as a side effect of the activity of the demethylase DEMETER (DME) and the histone-methylation-catalyzing POLYCOMB REPRESSIVE COMPLEX 2 (PRC2), which have other functions unrelated to imprinting (discussed by Rodrigues & Zilberman, 2015). Or even as a consequence of epigenetic modifications silencing adjacent genes in addition to transposable elements, where modification of expression of these adjacent genes generated an advantage for the endosperm (discussed by M. Gehring et al., 2011). It must be mentioned that these last two ideas can't by themselves explain the conservation of imprinting, but only its emergence.

2.4- Evidence for maternal bias in zygotic transcription

In plants, the overwhelming detection of maternal over paternal mRNAs in transcriptomic studies of early embryos in Arabidopsis and rice (Autran *et al.*, 2011; Anderson *et al.*, 2017) suggested that maternal control exists in early plant embryo development in the form of a maternal bias in *de novo* transcription. Since a considerable number of these maternal zygotic mRNAs were not detected in the respective egg cells, these results are in agreement with the maternally biased expression of reporter constructs in Arabidopsis for genes not detected in microarray analysis of egg cells (Wuest *et al.*, 2010; Raissig *et al.*, 2013).

This maternal transcript bias in early embryos has also been ascribed to maternal (seed coat) contamination in embryo preparations (Schon and Nodine, 2017). Though it is possible that maternal tissue contamination could account for some of the maternal bias seen in embryo transcriptomes, maternal contamination as the main explanation for maternal transcript bias in embryos is strongly argued against by reporter genes that also show a maternal bias in embryo expression (Vielle-Calzada et al., 2000; Autran *et al.*, 2011; Raissig *et al.*, 2013), semi-qRT-PCR detecting selected maternally-biased genes in embryo but not in surrounding tissues (Anderson *et al.*, 2017), RT-PCR not detecting seed coat or endosperm transcripts in RNA samples extracted from embryos (Raissig *et al.*, 2013), and functional evaluation of mutants for genes required for normal embryogenesis, that demonstrated a delay in paternal allele functionality (Del Toro-De León et al., 2014).

The mechanism for the regulation of the maternal transcript-bias has been linked to the RNA-directed DNA methylation (RdDM) pathway, through studies with KRYPTONITE (KYP). KYP encodes a Histone 3 Lysine 9 di-methyltransferase that has been implicated in histone methylation that acts to reinforce the RdDM pathway (Matzke & Mosher, 2014). In RNA-seq from 2-4 cell embryos maternally mutant for *kryptonite (kyp)* in the Ler background, and paternally wild type for the Col ecotype, the paternal contribution was greatly increased compared to Ler x Col hybrid embryos where both genomes were wild type (Autran *et al.*, 2011). The importance of the RdDM pathway in silencing the reporters for maternally biased genes in the embryo, was confirmed by assessing the effect of mutations in genes involved the RdDM pathway (*DRM2* [*DOMAINS REARRANGED METHYL TRANSFERASE 2*]*, CMT3* [*CHROMOMETHYLASE 3*], *NRPD1A* [*NUCLEAR RNA POLYMERASE D 1A*]*, NRPD1B* [*NUCLEAR RNA POLYMERASE D 1B*]*, DCL4* [*DICER LIKE 4*], *RDR2* [*RNA-DIRECTED RNA POLYMERASE 2*] and *AGO4* [*ARGONAUTE 4*]), the effect of mutations in genes involved in CG methylation maintenance was also tested (*MET1* [*METHYLTRANSFERASE 1*] and *DDM1*

[*DECREASE IN DNA METHYLATION 1*]) . Mutants in genes involved in the RdDM pathway but not those involved in CG methylation maintenance increased paternal allele expression (Autran *et al.*, 2011). Similarly, the Polycomb Repressive Complex 2 (PRC2) has been linked to paternal allele silencing, and *MET1* was implicated in the expression of a paternally expressed gene (Raissig *et al.*, 2013). Unfortunately, the interpretation of the genome-wide experiment using the *kyp* mutant (Autran *et al.*, 2011) must be treated with caution, since the plants used were subsequently shown to have a mixed background of Ler and Col, possibly biasing the analysis (Schon and Nodine, 2017). Since Ler was used as a mother and Col as a father, an apparent increase of paternal contribution could arise from the Col portions of the mother.

Despite the evidence for maternal bias in *de novo* transcription in the early embryo, it is plausible that the available datasets for the egg cell in Arabidopsis (Wuest *et al.*, 2010) and rice (Anderson *et al.*, 2017) are not detecting a set of transcripts with very short poly(A) tails, since these data were generated using poly(A) enriching procedures. In zebrafish, poly(A) tail length was shown to heavily influence the results of RNA-seq experiments: using qRT-PCR with both oligo d(T) and random primers, only about 30% of egg-deposited transcripts were estimated to be polyadenylated prior to fertilization (Aanes *et al.*, 2011). Thus, an analysis of maternal bias which accounts for the variable length in poly(A) tails in egg cell and zygote transcriptomes (Wuest *et al.*, 2010; Anderson *et al.*, 2017) is still required. In addition, most of the reporter genes previously used to monitor maternal expression in the zygote are actually expressed in the egg cell (Vielle-Calzada, Baskar and Grossniklaus, 2000; Autran *et al.*, 2011; Xiang *et al.*, 2011), or their egg cell expression has never been tested (Raissig *et al.*, 2013).

In animals there is also a general bias for transcripts being of maternal origin (40% to 75% of protein coding genes across various species and 60% to 70% of mRNA in zebrafish [reviewed by Lee et al. 2014)]. This has been interpreted as a consequence of egg cell carry over in rapidly-dividing animal embryos, since animals are mostly transcriptionally quiescent at early stages (discussed by Harvey et al., 2013, and Avilés-Pagán & Orr-Weaver, 2018). The detection of paternal transcripts has been often used as a proxy for the onset of zygotic transcription, which has been the phenomenon of interest (reviewed by Lee et al., 2014). In one-cell mice embryos, microinjection of reporters and incorporation of bromo UTP have shown that the male pronucleus is more transcriptionally active than the female pronucleus (Aoki et al., 1997). However, the biological relevance of this is not clear, since at this stage transcription seems to be promiscuous, occurring at low levels for more than 90% of genes, as well as for intergenic regions and transposons. Surprisingly, transcription in the male pronucleus is independent of enhancers (Wiekowski et al., 1997). This is likely a consequence of the uneven localization of DNA methylation, histone variants and histone modifications in male pronuclei. It is not clear if this epigenetic disorganization is a consequence of the divergent remodeling processes required by the maternal and paternal pronuclei, or if there is a developmental role for this disorganization (reviewed by Funaya & Aoki, 2017.)

3- Justification

The strategies taken so far to distinguish gametophytic and zygotically-transcribed maternal products in plants have had limited success, due to the fact that most zygotic transcripts are shared between the egg and zygote, preventing the source of the maternal bias from being determined. In rice, egg cell transcriptomes have been compared with zygote transcriptomes hours after fertilization (Anderson *et al.*, 2017); in Arabidopsis for embryos 2 and 3 days after fertilization (Raissig *et al.*, 2013); and in tobacco by suppression subtractive hybridization (Ning *et al.*, 2006). The rice comparisons showed a massive early maternal transcript bias, but only profiled poly(A)-tailed-transcripts. The Arabidopsis comparisons were done with embryo transcriptomes which may have some maternal contamination, and the tobacco studies were not designed to distinguish the parent of origin of the transcripts. Additionally, the rice and Arabidopsis studies used hybrid transcriptomes to distinguish between maternal and paternal reads, and hybridization can (at least in some cases) have a strong effect on parent-of-origin dependent expression (Del Toro-De León, García-Aguilar and Gillmor, 2014).

Reporter gene analysis allows examination of gene expression in an isogenic background, and is not sensitive to contamination of maternal tissues, which have been argued to be the source of the detected maternal bias (Nodine and Bartel, 2012; Schon and Nodine, 2017). So far, reporter genes have been implemented to distinguish the parental bias in *de novo* zygotic expression for only a handful of genes (Raissig *et al.*, 2013). Thus, examining more reporters for parent of origin expression in the zygote will contribute valuable data for a better understanding of early zygotic genome activation. As part of a collaboration with Raju Datla's lab, our group has generated and analyzed a set of parent-of-origin transcriptomes for several time points along early Arabidopsis development using the Tsu and Col accessions (D. Xiang, G. Del Toro-de León, C. Abreu, S. Gillmor, R. Datla; unpublished). These transcriptomes show a strong maternal bias in early embryogenesis, as well as other types of parent-of-origin bias. Experimental validation of parent-of-origin expression for specific genes will be an important test of these transcriptome results.

4- Hypothesis

Maternal and paternal alleles of many genes show non-equivalent expression in the zygote and early embryo.

5- Objectives

1. Validate transcriptomic bias in parent of origin dependent expression in the early embryo.
2. Obtain a reporter gene system that will allow differentiation between transcripts carried over from gametes, and expressed *de novo* in the zygote.
3. Explore the association between epigenetic modifications and maternal bias in the early embryo.

5.1 Strategies

1. Use genetic and transcriptome data to identify genes with a strong parent-of-origin expression bias in the zygote and early embryo.
2. Build reporters for selected genes, which can be visualized in the early embryo.
3. Determine if there is a parent-of-origin-dependent effect on the expression of reporters and compare the results with RNA-seq data.
4. Build and test a time-resolved reporter using the promoter of EC1 (egg cell 1) and the destruction box of cyclin B1 fused to YFP, and compare it to the stable YFP one.
5. Analyze publicly available data to search for insights into the mechanism that generates parentally biased gene expression in the early embryo.

6- Materials and methods

6.1- Gene selection for reporter analysis

6.1.1- Genes selected based on transcriptomic data

The following criteria were considered when selecting the genes to be assayed:

1) The number of parent-of-origin-assigned reads, to be certain that the bias detected is not due to insufficient sampling of the transcript population.

2) The difference in maternal proportion of reads, detected between both directions of the cross, to avoid selecting genes whose bias is due to ecotype dominance, i.e. the ecotype being the source of the bias instead of the parent of origin.

3) Egg cell presence or absence of the transcript, to consider if the bias detected is due to carry over or *de novo* transcription.

4) The size of the promoter region, since a gene with a very small promoter is likely to have the information for its expression coded elsewhere (for example in enhancers), and would be thus difficult to reproduce (S. Gillmor and G. del Toro; personal communication).

5) The abundance of the corresponding transcript in the embryo, since its transcriptional activity might be rather low (Pillot et al., 2010), making the expression difficult to detect.

6) Whether the gene is predicted to be expressed in the surrounding maternal tissues, for ease of visualization of zygotic expression.

According to the proportion of reads mapping to each of the parental alleles in the Col x Tsu hybrid, genes were chosen for being either maternally biased, paternally biased, or equally detected in both directions of the cross. After applying all the filters, using a squared chi test,

all parentally-biased genes had a $p<=0.05$ when testing the reads mapping uniquely to the maternal or paternal alleles, therefore, a filter for *p* value was not applied.

6.1.2- Maternally biased genes

Maternally biased genes were filtered based on the following criteria, using the data for zygote and octant stages:

1) At least 20 parent-of-origin-assigned reads in both directions of the cross.

2) A maximum difference of 0.3 in the maternal fraction of reads detected between the two directions of the cross.

3) No transcripts detected in the egg cell (called absent in 3 out of 3 biological replicates) as determined by Wuest et al. (2010).

4) A maximum of 1 positive presence call in the non embryonic seed regions in the preglobular and globular stages (each of the 5 tissues has at least 2 replicates for each stage, each one of which has a presence/absence call), using the Belmonte dataset of gene expression in different regions of the seed (Belmonte *et al.*, 2013).

5) At least 70% of the reads mapping to the maternal allele in both zygote and octant stages.

6) No gene located less than 1 Kb upstream.

The four most highly detected genes in the octant and zygote stages, within this group, were selected (Table 1).

6.1.3- Equally expressed genes

Equally expressed genes were filtered based on the same criteria as maternally-biased ones, but with between 40% to 60% of the reads mapping to the maternal allele in both stages.

The 3 most highly detected ones were selected (Table 1).

6.1.4- Paternally biased genes

Due to the scarcity of paternally-biased genes, some of the filters applied to the other groups of genes had to be dropped or relaxed. Data were considered only for the zygote stage:

1) At least 20 parent-of-origin-assigned reads in both directions of the cross.

2) A maximum difference of 0.3 in the parental proportion of paternal reads detected between the two directions of the cross.

3) At least 70% of the reads mapping to the paternal allele.

4) No gene located in the proximal 750 bp upstream region of the annotated transcriptional start site.

The 2 most highly expressed genes were selected (Table 1).

6.1.5- Genes based on functional genetic data

A subset of genes was selected based on delayed paternal allele complementation in a functional test in isogenic Columbia (Del Toro-De León et al., 2014). 4 genes not detected in the egg cell data (Wuest *et al.*, 2010) were selected with the purpose of analyzing *de novo* expression bias and 3 detected were selected as positive controls and to possibly gain insights into the extent of carry-over. Parent-of-origin reads in the hybrid transcriptome were not required for this class of genes.

6.1.6- Genes based on special criteria

6.1.6.1- Maternally biased

L41, a gene coding for a ribosomal protein, was selected because its upstream region contains several motifs that had been previously associated with maternally-biased expression (Del Toro-De León, 2017 [Doctoral dissertation]). There is no probe in the ATH1 microarray (used for egg cell) for this gene, and it shows a reciprocal maternal bias in our hybrid transcriptome. The L41 gene has a very small (250bp) upstream region, which would have excluded this gene for consideration based solely on the aforementioned criteria.

EMB1674, a gene that became equally expressed in one direction of the cross at the octant stage, but passed other filters, was selected because it has been shown to be required for normal embryogenesis (Tzafrir *et al.*, 2004).

6.1.6.2- Paternally biased

FUS3, a transcriptional activator that regulates many genes during seed development (Wang and Perry, 2013), passed the filters applied to paternally biased genes but was only the 10th most abundant with 5.1 TPMs (average between both directions of the cross). It was selected due to its previously reported paternal bias (Raissig *et al.*, 2013).

6.2- Reporter construction

6.2.1- Vector backbone selection and construction

Since the detection of gene expression in the embryo has proven difficult, most likely due to low transcriptional activity of individual genes (Pillot et al., 2010; Del Toro-De León, 2017 [Doctoral dissertation]), this was taken into account when choosing gene reporters. The GUS (uidA Beta Glucuronidase) protein has been successfully used to visualize expression in the early embryo (Raissig et al., 2013; Del Toro-De León, 2017 [Doctoral dissertation]). The advantage of using the GUS gene as a reporter is that GUS encodes an enzyme, and the enzymatic reaction can continue for up to 5 days, making it very sensitive. The disadvantage of GUS is that the process for its visualization is more time consuming than that for a fluorescent protein such as YFP, where the reporter can be directly visualized under the microscope, without the need for an additional staining procedure. Additionally, the signal of

the GUS reporter can spread to the surrounding tissues if incubated for enough time. Thus, GUS and YFP each have advantages and disadvantages when used as markers.

For GUS reporter fusions, I used the pBGWFS7 plasmid (Karimi et al., 2002) used by Raissig (Raissig et al., 2013), which contains a translational fusion of GFP and GUS. This reporter has the advantage that GFP fluorescence can be visualized as well as the colorimetric GUS product.

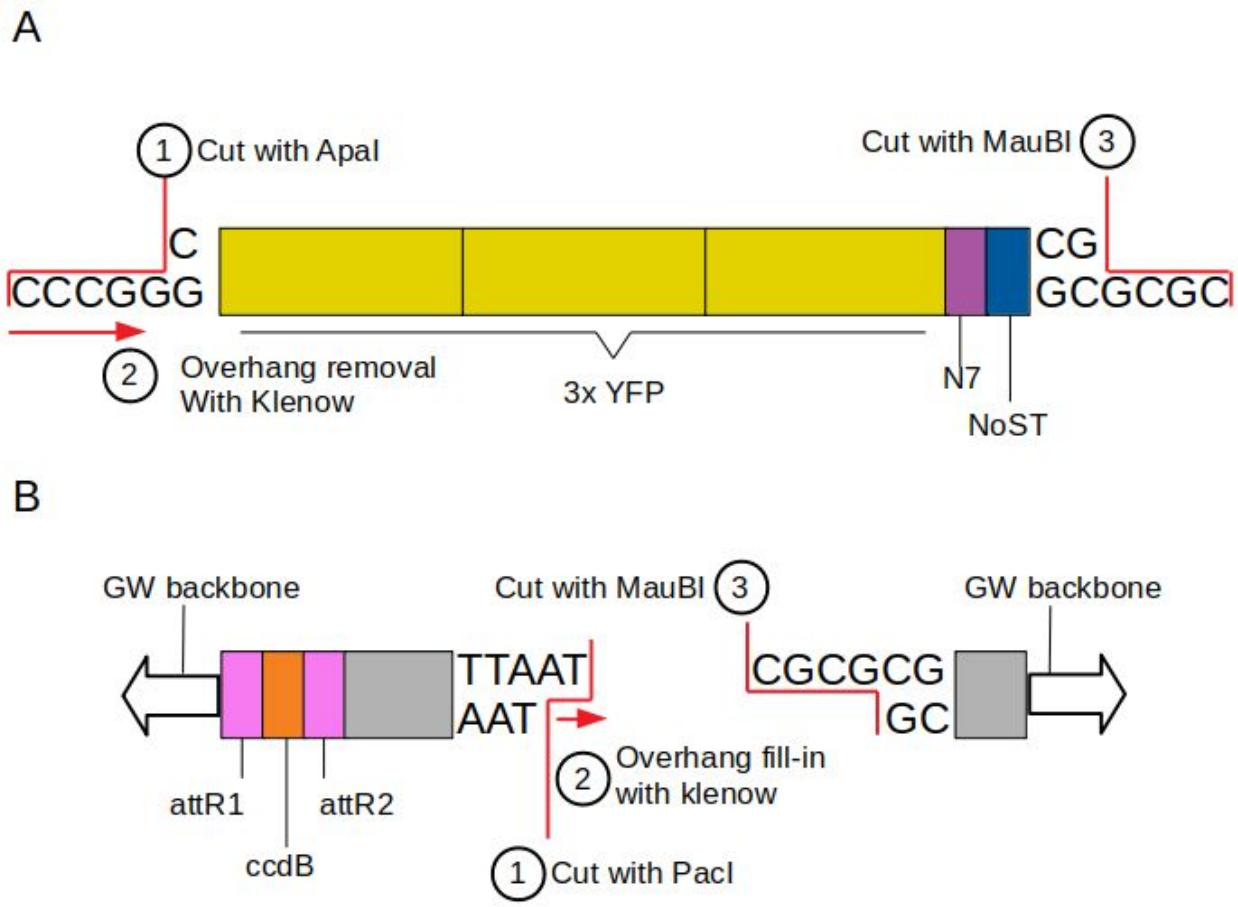


Figure 1: Diagram of expression vector construction. (A) Process followed for extracting the reporter from the plasmid donated by W. Lukowitz (U. Georgia, USA). (B) Process followed to remove the stop codon formed between recombination sites and reporter insertion site, while also generating compatible ends for reporter ligation.

To have a fluorescent reporter construct that would be brighter than a single GFP (as in the pBGWFS7 construct above), I used a Gateway-based backbone to make a reporter with three fused Venus (YFP) proteins and a nuclear localization signal (N7). This construct was built from 3xYFP-N7 in the pCAMBIA backbone (Bayer et al. 2009) and an 'empty' Gateway backbone (Figure 1). The resulting 3xYFP-N7 reporter, along with the pBGWFS7 GUS-GFP reporter, are shown in Figure 3.

6.2.2 Promoter region selection

Introns have been reported to greatly influence gene expression, especially when located near the beginning of the gene (Ueda, Zhang and Laux, 2011; Gallegos and Rose, 2017). Gallegos and Rose recently published a study on the effect of the intron distance from the transcriptional start site (Gallegos and Rose, 2017). They found that the expression of *UBQ10* was greatly reduced when removing its first intron or placing it more than 1 kb away from the transcriptional start site. In keeping with the results of this study, and of other studies where introns in the 5' end of the coding sequence were shown to have a large effect on gene expression, I included gene regulatory sequences as follows. For all genes, I cloned 2kb upstream of the ATG, this region is reportedly sufficient to reproduce the expression pattern of most genes (Xiao *et al.*, 2010; Liu, 2013). When the upstream gene was located less than 2kb upstream of the gene of interest, in which case I cloned the sequence up to the upstream gene. When the gene had no introns within 1 kb downstream of the translational start codon, only the ATG codon and 1 extra base (or n*3+1 bases to improve primer binding) to correct the reading frame with the reporter proteins were added. When the gene had an intron within 1 kb downstream of the ATG, I cloned this 1 kb downstream sequence. Primers were designed using the NCBI primer-design tool (Jian *et al.*, 2012). For a flow chart of the expression-driving-sequences selection see Figure S6.

6.2.3 Reporter construction and transformation

Promoter regions were amplified from genomic DNA extracted from *Arabidopsis thaliana,* Col-0 ecotype, using the high fidelity Phusion Hot-Start II polymerase. Amplicons were A-tailed by adding Taq DNA polymerase, and ligated to TOPO entry vector. Ligations were then transformed into *E. coli* and plated in 100 µg/mL spectinomycin LB plates. Direction of the promoters inserted in the TOPO vector was checked by either PCR or restriction enzyme digestion from plasmid DNA extracted from spectinomycin-resistant colonies for all of the plasmids, and was confirmed along with the integrity of the sequence by sanger sequencing for 7 of the plasmids. Plasmids with the desired inserts were recombined to the expression vectors constructed using the Gateway LR Clonase II enzyme mix, and transformed into *E coli* to select non-recombinant plasmids out by the action of the ccdb gene. In-direction, recombinant plasmids were then introduced to *Agrobacterium tumefaciens* strain GV3101 by electroporation, transformed CFUs were selected in 100 µg/mL rifampicin 50 µg/mL kanamycin YEB plates. Finally, plasmidic DNA was extracted from individual *Agrobacterium* colonies, transformed back to *E. coli*, and checked again for presence of desired insert by restriction enzyme digestion or by PCR.

6.3- Plant transformation

Plants were pruned when they started bolting to promote the growth of secondary inflorescences. Plant transformation was done as in Logemann et al. 2006, with small modifications:

1) Agrobacterium plates were grown at room temperature for 2 days in the dark.

2) Bacteria were lifted from the plate using a razor and suspended in 5 mL of YEB.

3) Suspension was then mixed with 20 mL of a 0.03% Silwet L-77 (V/V), and 5% sucrose (W/V) solution.

4) The solution was then dispensed on inflorescences using a pipette.

5) Plants were kept in covered in wet trays overnight.

6) T1 plants were selected by spraying the seedlings with diluted Finale (Bayer) (to be at 100 mg/L in ammonium glufosinate) at 1 day after germination and repeated at 7 days after germination.

6.4 Selection of transformation events

T1 embryos were screened for expression, and representative reporter lines were selected based on the intensity and expression patterns in the embryo, and, when possible, in the seed tissues. For most of the reporters the intensity of expression was rather low, complicating reporter expression characterization due to autofluorescence in maternal tissues. For those reporters, embryos were excised from seed tissues before observation.

6.5- Plant growth and crosses

For transformation, *Arabidopsis thaliana* ecotype Col-0 plants were grown in the greenhouse in Sunshine mix #4, with alternated NPK fertilizer watering.

For the crosses, transformed plants were grown as WT plants and then emasculated removing only the anthers from closed flowers, plants were then transferred to a growth chamber with long days (16h light), after 2 days plants were hand pollinated and kept in the growth chamber for the duration of the experiment.

For the crosses,T2 plants were grown and homozygous plants were identified. Reciprocal crosses were done with T2 and/or T3 plants, and penetrance was determined by squeezing the embryos out of their seeds and examining them with a compound microscope under mercury lamp illumination, with appropriate filters. Reporter gene expression was determined at 2, 3 and 5 days after pollination (DAP), roughly corresponding to the 2-8 cell, globular, and heart stages.

6.6- Microscopy

For the examination of penetrance, a DM6000B Leica microscope with a mercury fluorescence lamp was used. For the quantitative analysis of the fluorescence, the following settings were used for the pictures. The same settings were used for the maternally and paternally-inherited reporter crosses.

6.7 GUS stains

For the *L41* reporter, seeds were incubated in a permissive GUS staining solution (1 mM EDTA [pH 8], 1 mM $K_4Fe[CN]_6$, 1 mM $K_3Fe[CN]_6$, 1% [V/V] Triton-X-100, 1 mg/mL X-Gluc and 50-mM-in-phosphate phosphate buffer [pH 7]). GUS stainings were incubated at 37°C for 3

days for the 2 DAP stage and overnight for 3 and 5 DAP. Stains were cleared in Hoyer's solution overnight (37.5 mg/mL gum arabic, 0.5 g/mL chloral hydrate, 5 % [V/V] glycerol).

6.8 Bioinformatics

Bioinformatic analysis were done using R (R core team; 2013), using the packages: data.table (Dowle and Srinivasan, 2018), ggplot2 (Wickham, 2016), stringi (Gagolewski M. et al., 2018), and stringr (Wickham, 2018). Code is available upon request.

For association analyses between maternal bias and epigenetic marks, only transcripts with no more than 0.2 difference in maternal bias between directions and at least 10 informative reads were used. Additionally, for the association of egg cell and maternal bias analysis, only genes called present in the egg cell, and with at least 3 TPMs in the zygote were used. Other particular data treatments are specified in the respective figures.

7- Results

7.1- Strategy 1: Use genetic and transcriptome data to identify genes with a strong parent-of-origin expression bias in the zygote and early embryo.

Given the ecotype effect on parental contributions to the transcriptomes (Del Toro-De León et al., 2014), and several limitations of a previous study (Autran et al., 2011), the Tsu x Col hybrid, which is functionally similar to isogenic Columbia (Del Toro-De León et al., 2014), was used to make parent-of-origin-specific transcriptomes of early embryos (Xiang et al., manuscript in preparation). The Tsu genome has polymorphisms in more than 17 thousand genes respective to the Col genome, allowing the recognition of the origin of these polymorphic transcripts. RNA-seq profiling of the hybrid in the zygote, octant, globular, bent and mature stages was performed in both directions of the cross and reads were assigned based on the polymorphisms (Xiang et al., manuscript in preparation). This dataset showed a heavy transcript bias towards the maternal parent, regardless of the direction of the cross. This data was used to select the genes to test with reporters as described below.

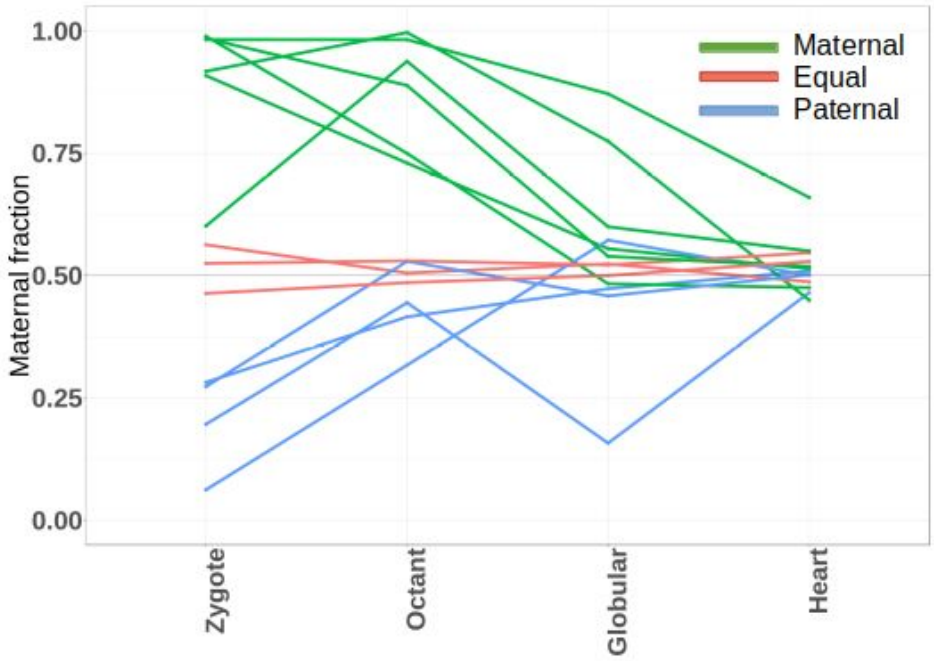


Figure 2: Maternal fraction of the reads for the selected genes along development. Green, orange and blue lines represent the maternally biased, equally detected and paternally biased genes; respectively. Genes with no informative reads are not represented.

AGI	Symbol	Selection criteria	Maternal fraction of reads			Transcript abundance (TPM)			Presence call	
			Zygote	Octant	Globular	Zygote	Octant	Globular	Egg	Sperm
AT5G15210	*HB30*	RNA-seq	0.98	0.89	0.54	173.5	176.5	175.0	A	A
AT5G42380	*CML37*	RNA-seq	0.92	1.00	0.77	316.5	51.4	28.5	A	A
AT5G47230	*ERF5*	RNA-seq	0.79	0.79	0.58	23.0	16.8	16.7	A	A
AT1G58210	*EMB1674*	RNA-seq, S	0.91	0.73	0.56	18.9	12.4	16.0	A	A
AT2G28400	N/A	RNA-seq	0.88	0.84	0.75	78.1	693.5	88.2	A	A
AT3G44260	*CAF1A*	RNA-seq	0.98	0.98	0.87	401.5	712.0	170.0	A	A
AT1G01600	*CYP86A4*	RNA-seq	0.99	0.75	0.48	51.9	58.9	80.2	A	P
AT5G13680	*ELO2*	Functional	0.67	0.70	0.52	21.6	15.6	28.7	P	A
AT5G19310	*MINU2*	Functional	ND	ND	ND	11.9	10.2	19.8	A	P
AT2G38670	*PECT*	Functional	ND	ND	ND	74.9	84.6	109.8	P	A

AT3G20070	*TTN9*	Functional	ND	ND	ND	3.0	8.7	8.5	A	A
AT1G49400	*EMB1129*	Functional	0.80	0.66	0.49	10.2	37.0	43.9	P	A
AT3G56020	*L41*	RNA-seq, S	0.69	0.68	0.51	44.5	87.9	70.3	ND	ND
AT5G22370	*EMB1705*	Functional	ND	ND	ND	55.2	24.6	42.4	A	A
AT5G07180	*ERL2*	RNA-seq	0.56	0.51	0.52	56.3	146.5	321.0	A	A
AT2G24762	*GDU4*	RNA-seq	0.53	0.53	0.52	336.0	334.5	154.8	A	A
AT3G20840	*PLT1*	RNA-seq	0.46	0.49	0.50	43.9	58.7	76.1	A	A
AT3G26790	*FUS3*	RNA-seq, S	0.27	0.53	0.46	5.1	26.7	36.8	A	A
AT1G67830	*ATFXG1*	RNA-seq	0.06	0.32	0.57	42.9	14.5	10.6	A	A
AT3G54940	N/A	RNA-seq	0.28	0.42	0.47	143.4	390.5	212.0	A	A
AT1G65090	N/A	RNA-seq	0.20	0.44	0.16	23.2	2.1	5.5	A	A

Table 1: Selected genes, selection criteria, presence in gametes, and parental bias and transcript abundance in the early embryo. Presence/absence calls are based on a Fisher combination of the *p* values for the three biological replicates in the respective tissues ($p <= 0.05$) (Wuest et al, 2010 for egg cell; Borges et al., 2008 for sperm cell); *p* values were calculated using the MAS5 method implemented in the affy package for R. ND: no reads/probe available. S: special criteria. Cells in green (maternal), yellow (equal), and blue (paternal) refer to parental transcript bias.

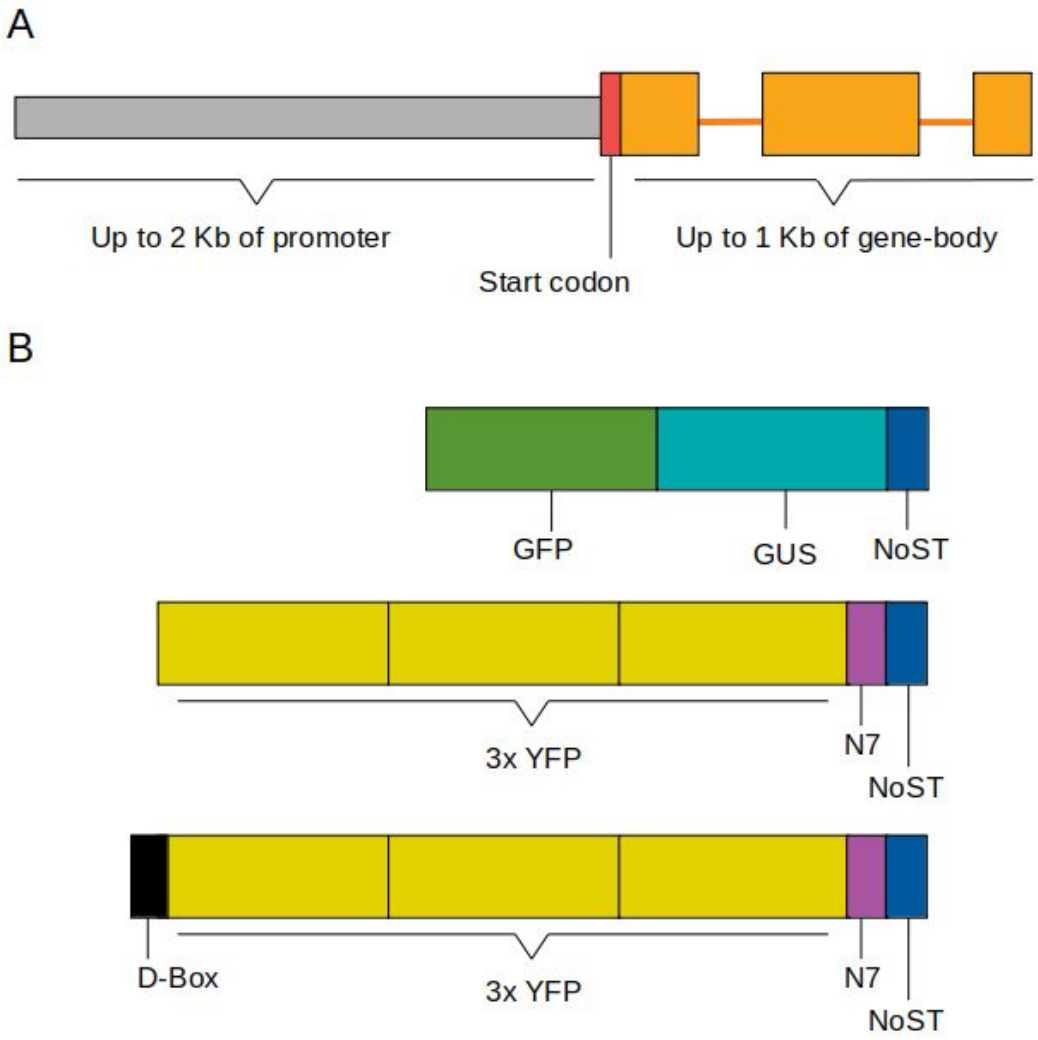


Figure 3: Reporter constructs used. (A) Scheme of the promoter regions used to drive the expression of reporters. (B) The three reporters used in this study. NoST: Nopaline synthase terminator.

According to the proportion of reads mapping to each of the parental alleles in the Col x Tsu hybrid, genes were chosen for being either maternally biased, paternally biased, or equally detected in both directions of the cross. Filters to select against ecotype-dominant genes (genes expressed more from one ecotype, regardless of the direction of the cross) and carried over transcripts (according to presence in the egg-cell [Wuest et al., 2010]) were applied (amongst others, see materials and methods).

A subset of maternally-biased genes was selected based on delayed paternal allele complementation in a functional test in isogenic Columbia (Del Toro-De León et al., 2014).

In total, 21 genes bonafine parentally-biased or equally-contributed (see materials and methods) were selected for reporter analysis (Table 1, Figure 2).

7.2- Strategy 2: Build reporters for selected genes, which can be visualized in the early embryo.

In animals, imprinted genes are commonly present in clusters, allowing *cis*-regulation of the cluster by imprinted control regions (ICRs). These regions are usually a few Kb long and contain parental-specific methylation and histone modifications that drive imprinted expression (reviewed by M. Bartolomei, 2009). The deletion of such regions removes the imprinting, and therefore a reporter construct would have to include them in order to reproduce the parent-of-origin-dependent expression. In plants, the modifications controlling imprinting have been mostly described in the endosperm, where maternal DNA hypomethylation in a site-specific manner is caused by action of the demethylase DEMETER (DME) in the central cell (Choi et al., 2002; reviewed by Rodrigues and Zilberman, 2015). In contrast to animals, imprinted genes in plants do not appear to be clustered (Gehring et al., 2011), and seem to be controlled by methylation patterns at individual genes (Reviewed by Rodrigues and Zilberman, 2015). In the embryo, the small amount of data available suggests that imprinting may be regulated by several mechanisms (Autran *et al.*, 2011; Raissig *et al.*, 2013; Del Toro-De León, 2017 [Doctoral dissertation]) (reviewed by García-Aguilar and Gillmor, 2015). Parent-of-origin-dependent expression seems to be associated with *cis* elements, since gene reporter constructs for specific genes reproduce the parent-of-origin expression seen with other methods (Autran *et al.*, 2011; Raissig *et al.*, 2013). Therefore, I considered that including near *cis* regulatory regions for driving the expression of reporter proteins should reproduce the pattern seen in transcriptomic assays (see materials and methods).

7.2.1- Vector backbone selection and construction

Considering the low transcriptional activity of the early Arabidopsis embryo (Pillot et al., 2010) and the difficulty in visualizing expression of reporter constructs in the early embryo (Del Toro-De León, 2017), 2 sensitive reporters were chosen for the study (Figure 3), additionally, to study carry over, a time-resolved reporter was also chosen, this reporter includes a D-box motif from cyclin B1 (Figure 3), which causes the reporter protein to be degraded during anaphase (Colón-Carmona et al., 1999).

7.2.2- Promoter region selection

Proximal promoter regions including up to 2 Kb of upstream and 1 kb of downstream sequences were selected for cloning (Figure 3, Table 2, Figure S6, for detailed procedure see materials and methods).

7.2.3- Reporter construction and transformation

The putative regulatory regions for all selected genes were successfully cloned into the expression vectors pBGWFS7 (Karimi, Inzé and Depicker, 2002) and pSG120, with the reporter proteins GFP-GUS and 3xYFP, respectively (Figure 3; for a more detailed description see Materials and Methods).Transformed plants resistant to BASTA (ammonium glufosinate) were obtained for 15 out of 20 promoters, with at least one of the two reporters (Table 3).

Gene	Upstream sequence cloned (bp)	Gene-body taken (bp)	Total amplicon size (bp)
HB30	2513	16	2529
CML37	1962	19	1981
ERF5	2102	17	2119
*EMB1674**	970	1508	2478
AT2G28400	1354	7	1364
CAF1a	1466	4	1470
CYP86A4	1796	581	2377
ELO2	314	460	774
MINU2	771	879	1640
PECT	590	567	1157
TTN9	2148	910	3058
EMB1129	2177	357	2534
L41	205	201	406
EMB1705	821	432	1253
ERL2	2123	1116	3239
GDU4	2153	34	2184
PLT1	1909	1107	3016
FUS3	2229	507	2736
ATFXG1	1181	807	1988
AT3G54940	1371	958	2329
AT1G65090	1227	700	1927

Table 2: Summary of selected proximal promoter regions to drive reporter expression. *:The reverse primer for *EMB1674* was accidentally designed for amplifying the whole gene, including the stop codon, thus impeding the correct translation of the reporter protein, and will no longer be considered.

7.2.4- Embryo expression of reporter constructs

Multiple transformation events were obtained and screened for a common expression pattern for most reporter lines (Table 3). Single transformation events were selected for the parent-of-origin expression experiment.

Promoter	T1 YFP plants	T1 GFP-GUS plants
HB30	9	0
CML37	0	0
ERF5	16	0
AT2G28400	3	0
CAF1a	4	9
CYP86A4	0	16
ELO2	13	0
MINU2	0	2
PECT	0	0
TTN9	11	0
EMB1129	1	0
L41	12	4
EMB1705	4	0
ERL2	13	0
GDU4	3	0
PLT1	1	3
FUS3	8	6
ATFXG1	0	0
AT3G54940	13	0
AT1G65090	0	0

Table 3: Number of analyzed transformation events for each promoter in each plasmid. Young embryos and seed regions for each plant were examined for expression of fluorescent proteins or GUS.

Reporter	Predicted localization of cloned CDS	N7 NLS in vector?	Observed localization
AT3G54940	Golgi apparatus, nucleus.	Yes	Endoplasmic reticulum
PLT1	Golgi apparatus, nucleus.	No	Cytoplasmic, nucleus
TTN9	Cell wall	Yes	Nucleus
CAF1a	Golgi apparatus, nucleus.	No	Cytoplasmic
L41	Cell membrane	No	Endoplasmic reticulum
PECT	Cell membrane, golgi apparatus.	No	Not observed
MINU2[*]	Nucleus	No	Nucleus
FUS3	Golgi apparatus	No	Nucleus, endoplasmic reticulum
EMB1129	Cell Membrane,golgi apparatus, nucleus	Yes	Endoplasmic reticulum
HB30	None	Yes	Nucleus
CYP86A4	Golgi apparatus, nucleus.	No	Nucleus, cytoplasm
AT1G65090	Golgi apparatus	Yes	Not observed
AT2G28400	None	Yes	Nuclear, endoplasmic reticulum
ATFXG1	Golgi apparatus, nucleus.	Yes	Not observed
CML37	None	Yes	Nucleus, endoplasmic reticulum.
ELO2	Cytoplasm	Yes	Nucleus cytoplasm
EMB1705	Cell membrane, cytoplasm	Yes	Endoplasmic reticulum
ERL2	Cell membrane, golgi apparatus.	Yes	Nucleus, endoplasmic reticulum
ERF5	None	Yes	Endoplasmic reticulum
GDU4	None	Yes	Not observed

Table 4: Subcellular localization of reporter proteins. Predicted localization of the cloned portion of the protein CDS was obtained using Plant-mPLoc (Kuo-Chen and Hong-Bin, 2010). Observed localization was determined by comparing with pictures of reporters of known localization. *: The reporter originally built for *MINU2* was replaced with a GFP-tagged version of the full protein, which was able to rescue the mutant phenotype of the *minu1/minu2* double mutant (Sang et al., 2012).

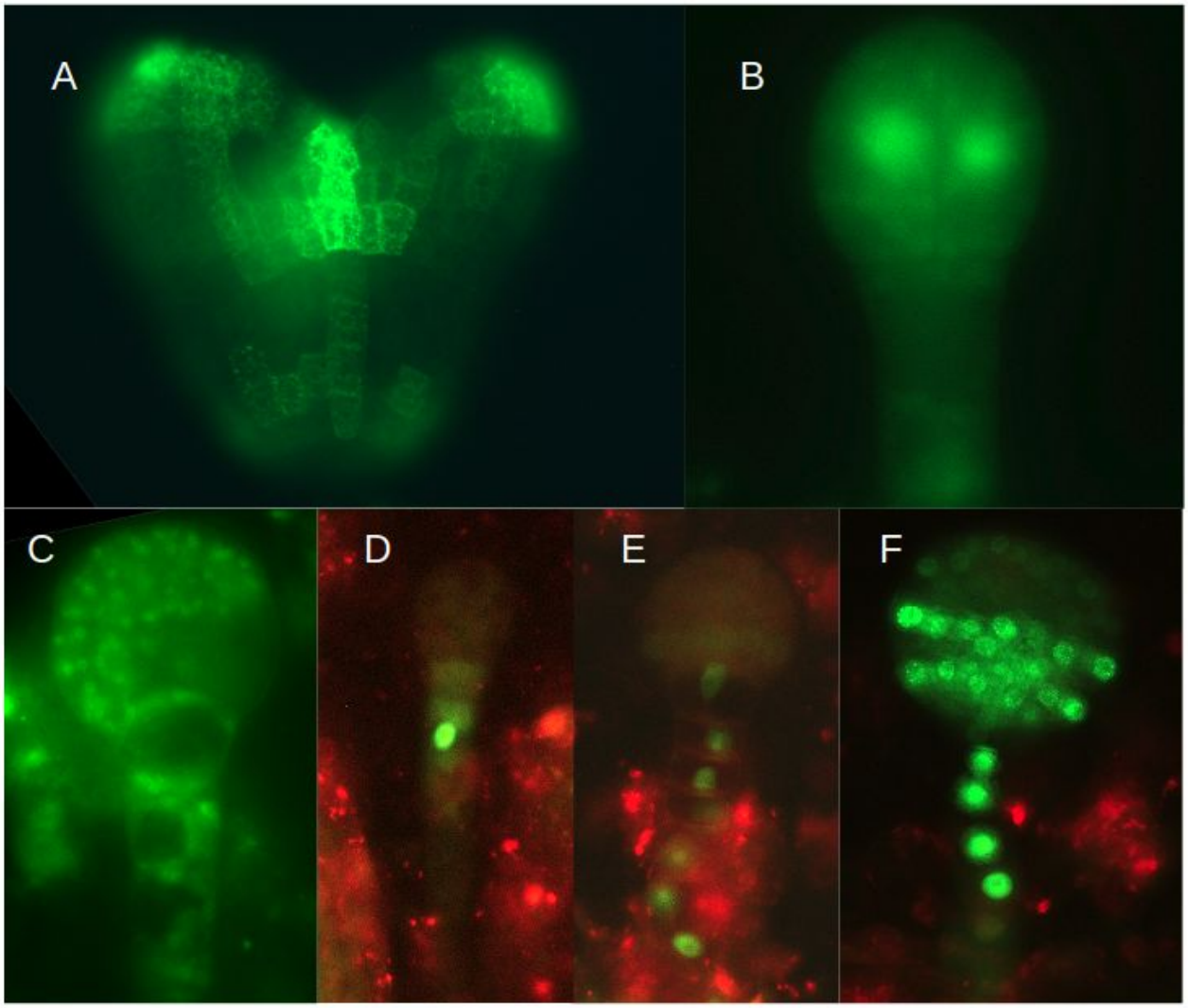


Figure 4: Sub-cellular localization and expression pattern of reporters. (A) pAT3G54940-3xYFP-N7 expressed at the heart stage embryo, showing what appears as endoplasmic reticulum localization. (B) pPLT1-GFP-GUS expressed at the 2-cell stage, showing nuclear localization. (C) pEMB1705-3xYFP-N7 expressed at the globular stage showing what appears to be endoplasmic reticulum localization. (D to F) Expression pattern of pTTN9-3xYFP-N7 trough development, nuclearly localized and stronger in the suspensor, with increasing strength in the lower tier of the embryo proper; this pattern was observed in most T1 lines for TTN9.

While screening T1 plants, multiple subcellular localizations were observed for both nuclearly-localized and endoplasmic reticulum-localized reporters (Figure 4, Table 4). This is likely due to coding sequences (CDS) cloned with the promoters (Table 2). Indeed, different localizations were predicted for the cloned sequences (Table 4), which probably conflicted with the N7 localization signal.

7.3- Strategy 3: Determine if there is a parent-of-origin-dependent effect on the expression of reporters, and compare the results with RNA-seq data.

Due to several delays in plant growth (death by pesticide, etc) and time restrictions, only 6 out of the 15 obtained lines were profiled. For these lines, penetrance was determined in embryos of reciprocal crosses between the reporter lines and WT Col-0 plants at 2, 3, and 5 DAP.

Reporter	Selection criteria	DAP	Expected	Observed	*p* value
L41	RNA-seq	2	Maternal	Maternal	1.0E-16
		3	Equal	Maternal	2.5E-15
		5	Equal	Equal	0.56
FUS3	RNA-seq	2	Paternal	Paternal	4.8E-6
		3	Equal	Paternal	0.03
		5	Equal	-	-
MINU2	Functional	2	Maternal	Equal	0.47
		3	Maternal	Equal	1
		5	Equal	Equal	0.43
CYP86A4	RNA-seq	2	Maternal	Equal	0.66
		3	Equal	Equal	0.19
		5	Equal	Equal	1
CAF1a	RNA-seq	2	Maternal	Maternal	4.1E-5
		3	Maternal	Equal	0.12
		5	Maternal	Equal	1
PLT1	RNA-seq	2	Equal	Equal	0.47
		3	Equal	Equal	1
		5	Equal	-	-

Table 5: Parent-of-origin expression of reporter constructs. The observed bias was determined by measuring the penetrance of the reporter when inherited from each parent. *p* value was obtained using a Fisher's test for the number of embryos expressing and not expressing the reporters (see Table S1). Expected bias is based on the corresponding selection criteria. Cells are highlighted in

green when the observed data matched the expectation for parental bias; when not, in yellow.

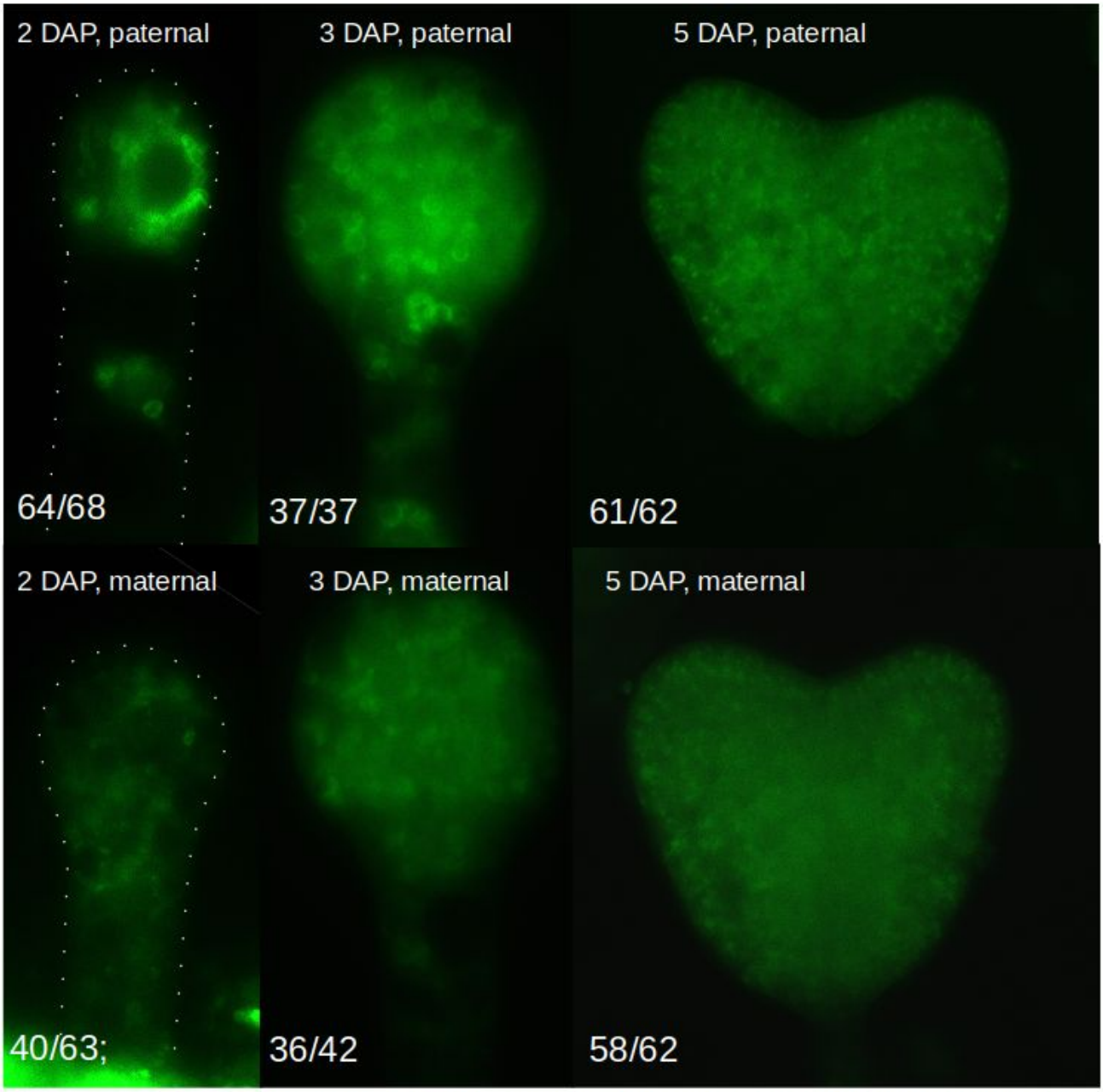


Figure 5: Parental bias in expression of the reporter for FUS3. Top row shows the expression of the reporter when inherited paternally at 2, 3, and 5 DAP, respectively; bottom row shows the analog when maternally inherited. Numbers at the bottom of each embryo show the penetrance of the GFP expression. Pictures were taken with the same settings.

For 4 out of the 6 reporters tested, the results were in agreement with the hybrid RNA-seq data (Xiang et al., unpublished; Table 5, Figure 5), validating the parental transcript-bias observed there, and showing that the parental bias observed in the Col x Tsu transcriptome also occurs in isogenic Col-0 plants.

7.3.1- Quantitative analysis of reporter expression

In order to get more quantitative and faster results, the intensity of the fluorescence was measured by analyzing the pictures of reciprocal crosses between reporters and wild type (WT) plants (Figure S3). Fluorescence intensity of embryos was measured with Image J software (Schneider, Rasband and Eliceiri, 2012). For this approach, 10 to 20 pictures from embryos extracted from 1 to 2 siliques were taken per treatment (direction of the cross, reporter and DAP). Upon analysis of the results, several characteristics that could significantly bias the interpretations were identified (Figure S2):

1. Different measured-area distributions were detected for maternal and paternal embryos for all the reporters for at least 1 stage sampled. Meaning that either different stages or picture sizes (see point 4) were sampled between directions of the cross.

2. Significant associations were detected for the measured area and intensities. Meaning that bigger embryos did have more signal, giving relevance to point 1.

3. Significant positive associations were detected for the measured area and mean intensities (intensity/area). Meaning that bigger embryos had more intense signal, so the association between intensity and area could not be normalized out by using mean intensities.

4. Leica application suite (LAS) would sometimes, for unknown reasons, vary the dimensions of the acquired images.

Thus, it is not possible to draw conclusions from these experiments. For the future, increasing the sample size, inserting size-scale bars when acquiring pictures, and grouping embryos according to the number of cells instead of DAP, would likely address many of the problems mentioned above.

7.3.2- Determine if the reporters are expressed in the egg cell

To test whether the parental bias observed for the reporters could come from the egg cell, I examined unfertilized ovules for reporter expression. For all 6 reporters examined, the Wuest et al (2010) dataset predicted no egg cell expression (Table 6).

Since the egg cell is within the female gametophyte, I used confocal microscopy to analyze the morphological location where the egg cell sits within the ovule. As a control to be able to detect the marker, I also analyzed the fertilized seed at 2 DAP with the same technique.

Interestingly, the reporters lines for *FUS3* and *CAF1a* showed considerable penetrance in the egg cell, despite not being detected in the microarray data. The reporter for *FUS3* showed a paternal bias in early embryos (Table 5) despite its strong presence in the egg-cell (Table 6, Figure 6).

The discrepancy between the published microarray egg cell data and my characterization of reporter constructs could reflect either poor reproduction of the transcription pattern of the gene by the reporters, or a poor reproduction of the posttranscriptional regulation of the native

transcript. Alternatively, some differences could be a consequence of slightly different periods after emasculation (48-72 h for the microarray experiment [Wuest et al., 2010]). As discussed above, the egg cell data available for Arabidopsis was enriched for poly(A) tailed transcripts, which would miss any non-polyadenylated transcripts deposited. Since the reporters constructs I used ended with the nopaline synthase terminator, I hypothesize that they could miss 3' regulatory regions controlling the native transcripts' translation, and would likely be translated whenever transcribed. These results show that carry over should be examined more deeply before determining whether there is a bias in transcription after fertilization.

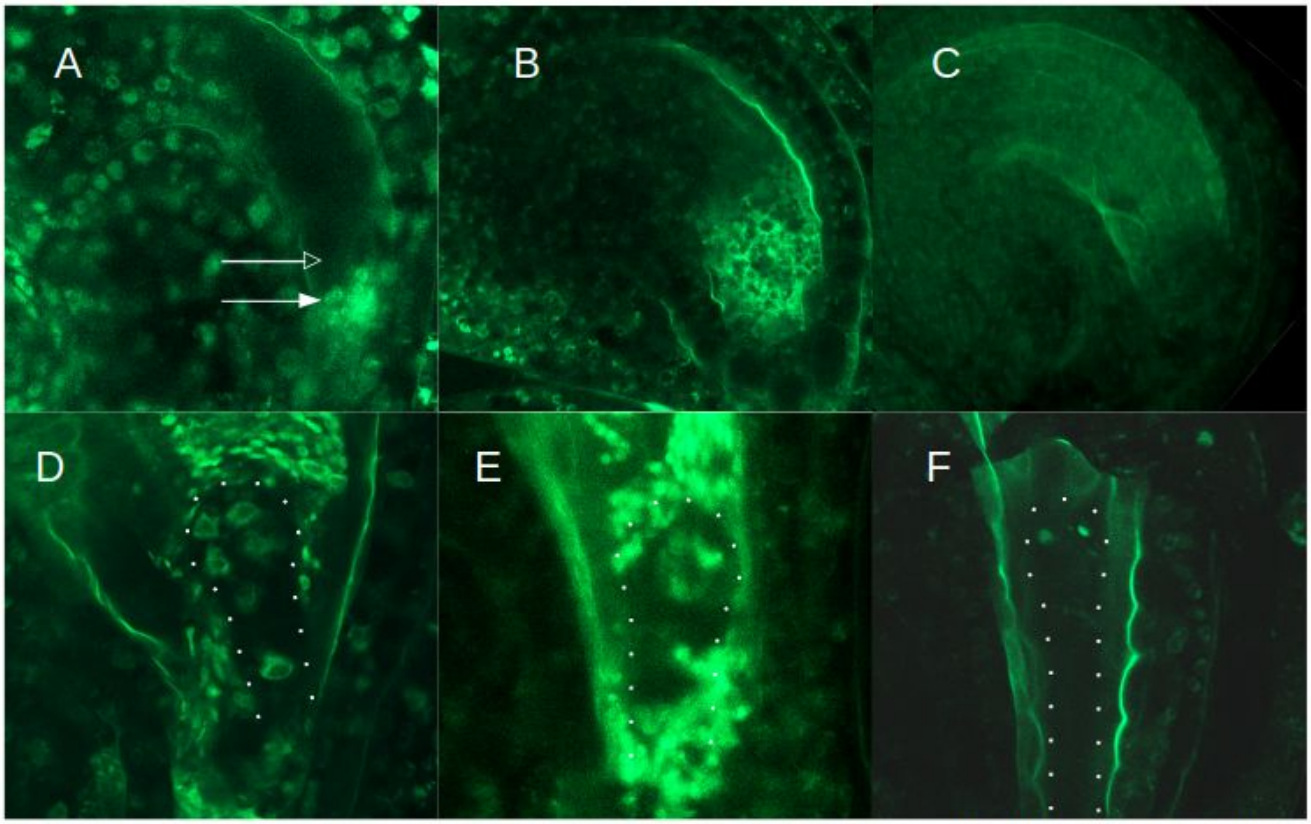


Figure 6: Egg-cell and young-embryo expression of reporters. (A) Unfertilized female gametophyte expressing the reporter for *MINU2*, filled arrow points to synergid nuclei, hollow arrow points to the region where egg cell nucleus should be. (B) Unfertilized female gametophyte expressing the reporter for *FUS3*, expression was detected in the whole egg apparatus. (C) Unfertilized female gametophyte carrying the reporter for *PLT1*, expression was not detected in these tissues. (D) 8-cell-stage embryo inside the seed, expressing the reporter for *MINU2*. (E) 1-2-cell stage embryo inside the seed, expressing the reporter for *FUS3*. (F) 2-cell stage embryo inside the seed, expressing the reporter for *PLT1* in the apical cells. Dotted lines were drawn surrounding the embryos.

Promoter	Presence calls in the egg-cell (Wuest et al., 2010)	Penetrance in egg-cell	Penetrance in embryo at 2 DAP (inside the seed)
MINU2 (GFP)	0/3	0/32	13/15
FUS3 (GFP)	0/3	24/32	14/14
AT2G28400 (3xYFP)	0/3	0/18	2/15
PLT1 (GFP)	0/3	0/15	6/11
AT3G54940 (3xYFP)	0/3	1/15	1/10
CAF1a (GFP)	0/3	6/16	3/13

Table 6: Egg-cell and young-embryo expression of reporters. To test whether the expression of reporters is in agreement with microarray data (Wuest et al., 2010), ovules at 2 days after emasculation were examined. Embryos at 2 DAP with the reporter-expressing plant used as mother were used as a sensibility control, since expression could be obscured within maternal tissues.

7.4- Strategy 4: Generate and validate reporter constructs capable of differentiating between carry over of egg cell transcripts and *de novo* expression in the zygote

As discussed above, the distinction between *de novo* and carryover products has not been made for Arabidopsis zygotes and early embryos. For this reason, Del Toro-De León (2017) fused a destruction-box (D-box) from cyclin B1 to the 3x YFP. This system is known to allow cell-specific analysis of reporter expression, by causing the turn-over of the reporter protein every cell cycle, or more precisely, every anaphase (Colón-Carmona et al., 1999). To test the applicability of this strategy in our system, we cloned the *EC1* egg cell specific promoter in front of the 3xYFP-N7 reporter, with and without the cyclin B1 D-box. The *EC1* promoter drives strong expression in the egg cell. The *EC1* mRNA is found at high levels in the egg cell, and is absent from the zygote, as determined by in situ hybridization (Sprunck et al., 2012). Nevertheless, the signal of a reporter protein driven by it, and presumably inherited from the egg cell, remains for up to 4 DAP (Del Toro-De León 2017). Therefore, if effective, the D-box fused 3xYFP-N7 reporter driven by the *EC1* promoter should be absent or greatly reduced after the first division of the zygote.

Constructs containing the promoter of *EC1* driving 3xYFP-N7, with and without the D-box, were transformed into plants and T2 homozygous individuals were identified by examining egg cell expression. While screening T1 and T2 plants a number of puzzling observations were made:

- Penetrance and expressivity for both constructs in the unfertilized ovule were highly variable, and depended on the time after emasculation or time after pollination (Table 7).
- Penetrance diminished to about 50% soon after fertilization even for the pEC1::3xYFP-N7 construct (without D-box), but went back up to around 100% after a couple days (Table 7). This observation was also made by Del Toro-De León (2017).

- No fully penetrant transformation event could be obtained for the D-box-fused construct (pEC1::D-box-3xYFP-N7).

As expected, I found a difference in the penetrance of fluorescence between the pEC1::3xYFP-N7 reporter, and the pEC1::D-box-3xYFP-N7. Similar penetrance was found before fertilization, but after fertilization, the penetrance of the pEC1::D-box-3xYFP-N7 construct was markedly lower than that of pEC1::3xYFP-N7. In addition, the signal for pEC1::D-box-3xYFP-N7 was always clearly weaker than for pEC1::3xYFP-N7. These results validate the purpose of the pEC1::D-box-3xYFP-N7 construct: inclusion of the D-box causes the 3xYFP to be degraded during anaphase of the cell cycle, so that YFP fluorescence is primarily due to newly transcribed 3xYFP-N7, and not YFP inherited from the mother cell.

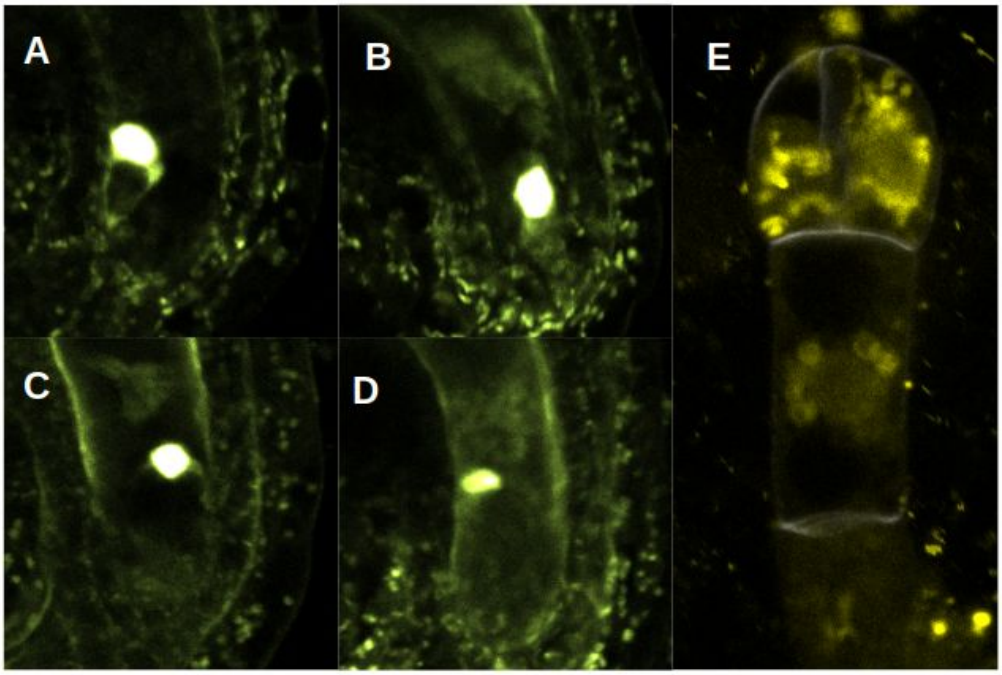


Figure 7: D-box-driven reporter degradation. (A and B) Expression of pEC1::3xYFP-N7 (stable) reporter before fertilization and after 12 hours, respectively. (C and D) Expression of pEC1::D-Box-3xYFP-N7 (unstable) reporter before fertilization and after 12 hours, respectively (E) Expression of pEC1::D-Box-3xYFP-N7 (unstable) reporter at the 2-cell stage; weak signal is observed around nuclei.

Stage	pEC1::3xYFP	pEC1::D-box-3xYFP
1 day after emasculation (egg)	17/45	13/63
2 days after emasculation (egg)	22/22	15/22
12 hours after pollination (zygote)	23/59	3/66
1 day after pollination (zygote)	66/141	11/94
2 days after pollination (embryo)	34/36	31/32

Table 7: Stable vs anaphase-destruction reporter expression in unfertilized egg-cells and early embryos.

Due to the high time-dependance of the expression of these reporters, and to eliminate inter-individual variation, an attempt was made to live image seeds during development. However, no egg cell fluorescence could be detected in those experiments, despite of the seeds appeared to be developing. Instead of the expected expression pattern, only seed-coat expression was observed (data not shown).

7.5- Strategy 5: Analyze publicly available data to search for insights into the nature of the parental bias in zygotic mRNA populations.

7.5.1 Egg cell carryover

As discussed in the introduction, there is evidence for the role of egg cell products in promoting early embryo development (Pillot et al., 2010; Grimanelli et al., 2005). Nevertheless, the extent of the effect of carry over on maternal bias in transcriptomes has not been assessed. This is of crucial importance for determining the source of such bias (i.e. whether it results from *de novo* transcription or only carry over). Attempts to profile zygotic transcripts not detected in the egg cell have given rather inconclusive results, due to the exceptionally similar transcript populations in these tissues (Anderson et al., 2017). Therefore, I decided to analyze influence of carry over on maternal bias by comparing the intensity with which a transcript is detected in the egg cell and the bias detected in the zygote.

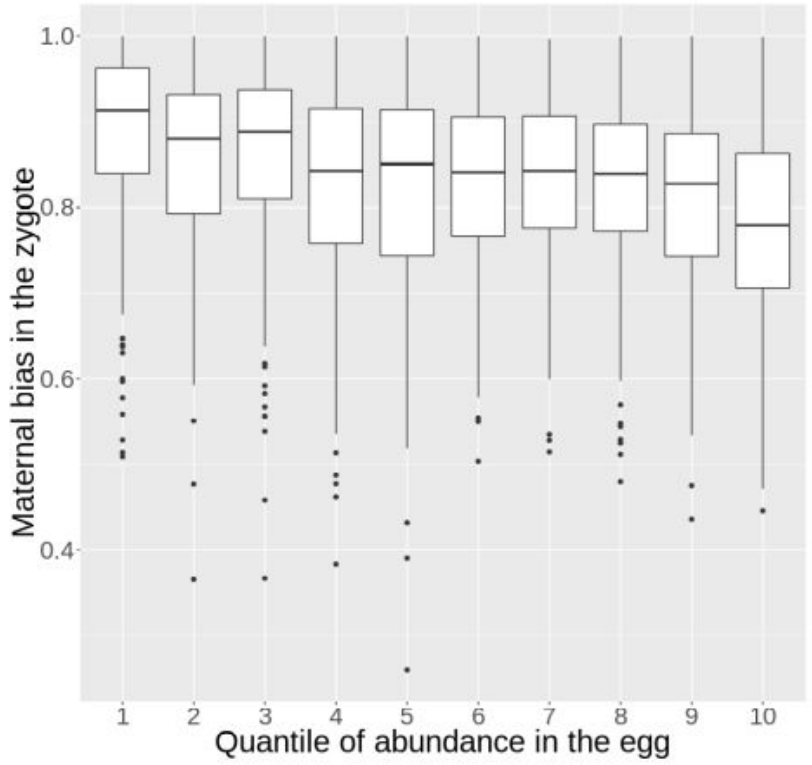


Figure 8: Effect of transcript abundance in the egg over maternal bias in the zygote. Expected association (more abundant in the egg = more maternal in the zygote) is not found. Insead, a significant negative correlation was found (Rho= -0.25, *p*= 2.8*10^-26). 1770 genes were used (177 per box, see materials and methods)

If the transcript bias is caused by carry over, a strong correlation between the abundance of a transcript in the egg cell and its maternal bias in the zygote would be expected, since a highly abundant maternal transcript would be much less likely to be matched by sufficient paternal

allele transcriptional activity. Intriguingly this is not the case, since there is no positive correlation between maternal bias and egg cell transcript abundance (Figure 8). And more puzzlingly, a significant negative correlation is observed (Rho= -0.25, $p= 2.8 * 10^{-26}$). Additionally, analyzing transcripts increasingly detected in the zygote compared to the egg cell yielded similar results (Figure S2).

This data could naively suggest that there is no carry over from the egg cell, which is challenging to reconcile with experimental information (Pillot et al., 2010; Grimanelli et al., 2005). Some possibilities could, however, resolve this conflict:

1. Certain modifications in the transcript, such as m6A (a de-stabilizing mark that promotes the degradation of maternally-deposited transcripts in mice [Ivanova et al., 2017]), or the binding of stabilizing proteins (such as Cup in *Drosophila* [Broyer et al., 2017]), could be controlling transcript carry over and/or degradation. Therefore the mere abundance of the transcript does not result in functional carry over, but instead the transcripts undergo a more complicated regulation.

2. Inherited transcripts could not be detected due to the poly(A) bias in both datasets. Thus breaking the association between detection and actual abundance in the egg cell.

3. The onset of zygotic transcription could trigger a quick degradation of maternally inherited transcripts. Carry over could be then interpreted as a backup mechanism in case zygotic transcription fail to initiate at the appropriate time.

4. Lack of reproducibility between the different technologies used (microarrays for egg-cell [Wuest et al., 2010], and RNA-seq for embryos [Xiang et al., unpublished]).

I conclude that carry over is likely to be regulated in a complex enough way to hinder its dissection by simple analysis of transcript abundances.

The interpretation of intronic reads as *de novo* transcription has been used to distinguish carried over transcripts from zygotic transcripts in zebrafish (Lee et al., 2013), and would likely help resolve in this system also. The rationale behind this is that the deposited transcripts would have been spliced during oocyte maturation, while only the new transcripts would have introns and the possibility to have reads aligned there. Unfortunately, due to time limitation, this analysis is not within the scope of this thesis. Other tools, like the analysis of intronic and exonic reads (Gaidatzis et al., 2015), would likely help to discern parental transcription rates without the influence of carryover. Gaidatzis et al. propose the change in intronic reads as a measure of change in transcription rate; and a model where the changes in exonic and intronic reads between two time points can be used to estimate changes in mRNA stability; both with good correlation with experimental Nascent-seq and RNA half life data (Gaidatzis et al., 2015).

7.5.2- DNA methylation

Since imprinting seems to occur in isogenic embryos of Arabidopsis (Raissig *et al.*, 2013; Del Toro-De León, García-Aguilar and Gillmor, 2014), some epigenetic dimorphism is likely to exist between parental genomes. The only primary imprinting marks (those that establish the

difference between parental alleles and may act as effectors or give rise to secondary imprinting marks) so far described are DNA methylation (in Arabidopsis endosperm, reviewed by Rodrigues and Zilberman, 2015), and H3K27me3 (in mice zygotes, Inoue *et al.*, 2017). I reasoned that the methylation status of the loci whose transcription is maternally-biased, and that of those that are not, should be different. Unfortunately there is no available methylome for Arabidopsis egg cells, so I used somatic and sperm cell data to analyze methylation patterns possibly driving parental bias.

Tissue	Gene region	Corr CG	p.val CG	Corr CHG	p.val CHG	Corr CHH	p.val CHH
Sperm	Intron	-0.146	8.3E-17	-0.019	0.58	-0.066	8.1E-05
Sperm	Gene	-0.210	2.3E-44	-0.035	0.04	0.033	0.03
Sperm	Exon	-0.219	8.4E-48	-0.011	0.56	0.025	0.10
Sperm	UTR5	-0.014	0.48	0.008	0.91	-0.001	0.96
Sperm	UTR3	-0.009	0.65	-0.136	0.03	-0.044	0.01
Somatic	Intron	-0.157	3.5E-19	0.031	0.37	-0.082	1.0E-06
Somatic	Gene	-0.223	6.2E-50	-0.045	0.01	0.012	0.43
Somatic	Exon	-0.218	1.3E-47	-0.025	0.18	0.009	0.55
Somatic	UTR5	-0.013	0.51	0.077	0.27	-0.057	1.4E-03
Somatic	UTR3	-0.044	0.03	-0.113	0.08	0.005	0.75
Difference	Intron	-0.092	1.7E-07	-0.019	0.58	-0.059	4.2E-04
Difference	Gene	-0.060	7.2E-05	-0.030	0.08	0.017	0.26
Difference	Exon	-0.075	9.8E-07	-0.011	0.55	0.017	0.25
Difference	UTR5	-0.003	0.87	0.008	0.91	0.009	0.61
Difference	UTR3	0.003	0.90	-0.145	0.02	-0.042	0.01

Table 8: Association of methylation with maternal bias. “Corr” columns correspond to the Pearson’s correlation coefficient of the mean methylation frequency (for each Cytosine in the specified context in the specified region) with the maternal bias detected in the Tsu and Col hybrids. “Difference” refers to the mean difference of methylation frequency between sperm and somatic tissues. Associations of particular interest (absolute correlation of at least 0.1 and p value <= 0.01) are highlighted in light blue.

A negative association was found between CG methylation and maternal bias for the gene body (Table 8, Figure 9), indicating a possible role of the CG methylation in driving maternal bias. I hypothesize that demethylation could be happening for certain loci in the egg cell, as

has been shown in the central cell of Arabidopsis and rice, and the egg cell of rice (Park *et al.*, 2016), but not in the sperm cell, thus establishing an epigenetic dimorphism over which other mechanisms could drive biased expression.

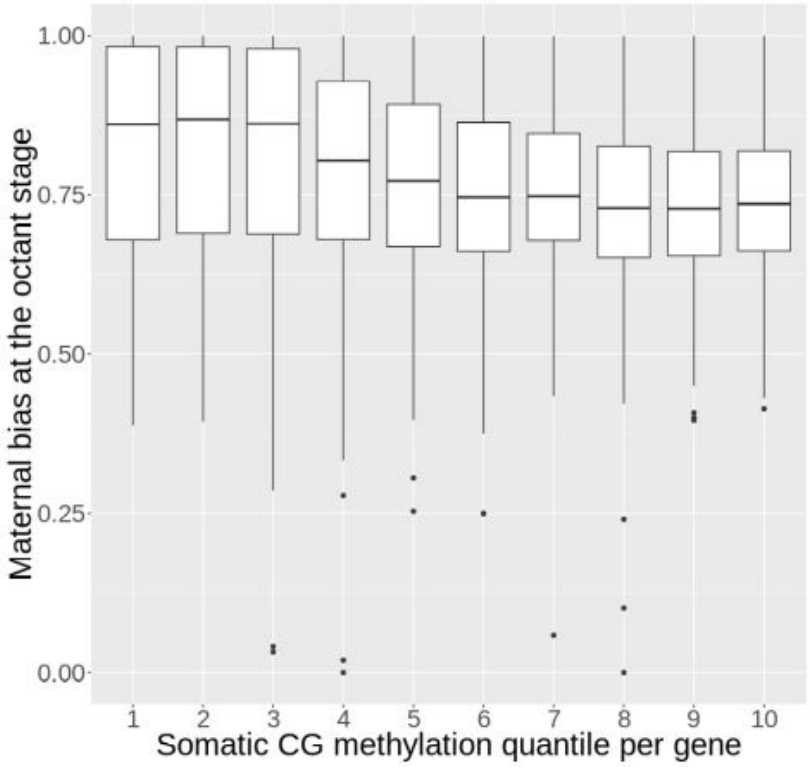


Figure 9: Association between CG gene-body-methylation and maternal bias in the arabidopsis zygote. There seems to be a link between CG gene-body methylation and maternal bias. Mean maternal bias at the octant stage and somatic methylation data were used. Rho= -0.23, P = 2.7*10^-52. 431 genes per box are represented (see materials and methods)

The maternal transcript bias in early embryogenesis seems to happen also in rice (Anderson et al., 2017). However, the association between the maternal bias and genic CG methylation (using egg cell and leaf methylation data [Park et al., 2016; Zemach, 2010; respectively]) could not be found in rice data (correlation of -0.004, p value 0.7 for leaf CG and maternal bias at 9 hours after pollination [HAP]), nor any association for the other 2 contexts with leaf nor with egg cell methylation data (data not shown). But this dataset is limited to the 9 HAP stage, and the maternal bias is almost absolute at that stage, so the effect of CG methylation may not be visible that early in development.

7.5.3- Histone modifications

Since the only other primary imprinting mark besides DNA methylation described to date is H3K27me3 (Inoue *et al.*, 2017), I reasoned that testing the association between such marks and imprinting in our dataset might give an insight into how this process is regulated. To broaden the analysis, I considered a comprehensive annotation of “chromatin states” (Liu et al., 2018), which defined 38 chromatin states based on the enrichment for several epigenetic marks and chromatin conformations (DNA accessibility by susceptibility to DNAses and others) using 216 datasets from various tissues.

For this analysis, I determined if the genes associated with certain chromatin state by Liu et al. had a different maternal bias than the rest of the genes. Also, the Liu dataset contains

information about the region of each gene with which the particular chromatin state was associated, which was taken into account.

While analyzing the distribution of maternal bias, I noticed two different behaviors or modes. One group of genes has around 75% maternal bias and the other has around 98% (Figure 10).

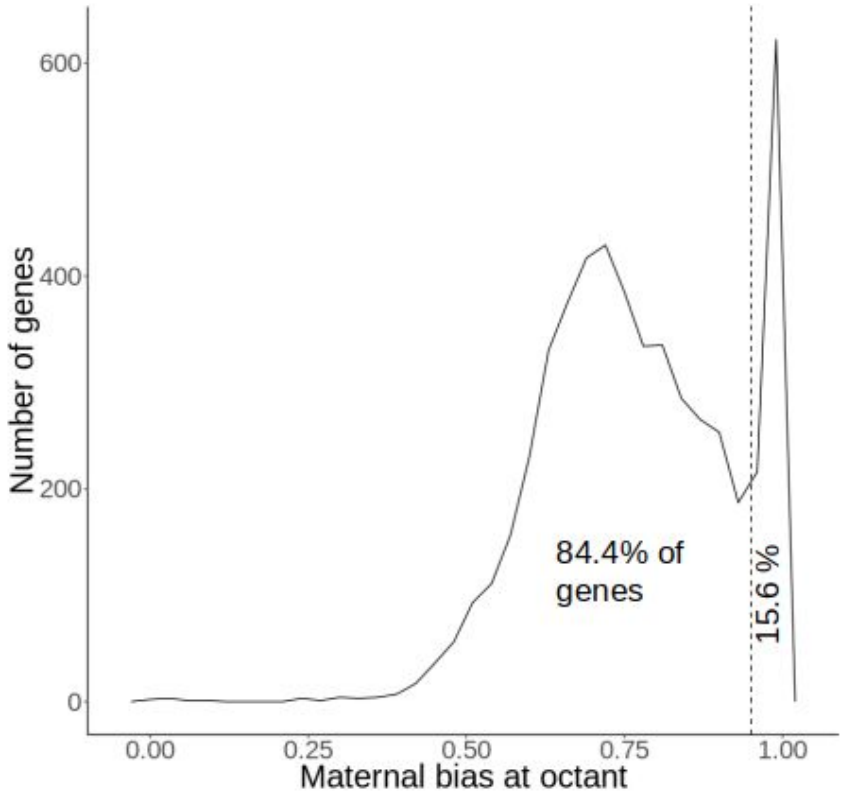


Figure 10: Maternal bias distribution at the octant stage. Figure shows a histogram-like representation of the distribution of maternal bias. The average maternal bias between directions of the cross (Col0 and Tsu0) is represented for the 5163 genes that passed the filters (see materials and methods). A bimodal distribution is evident, with a maxima around 73% maternal bias and another around 99%; 84.4% and 15.5% of the genes are below each of the curves, respectively.

The main effect of the chromatin states in maternal bias was to change the balance between these two distributions (Figure 11). Therefore, I used the proportion of genes under the second curve as a response variable for the effect of the chromatin state (Table 9).

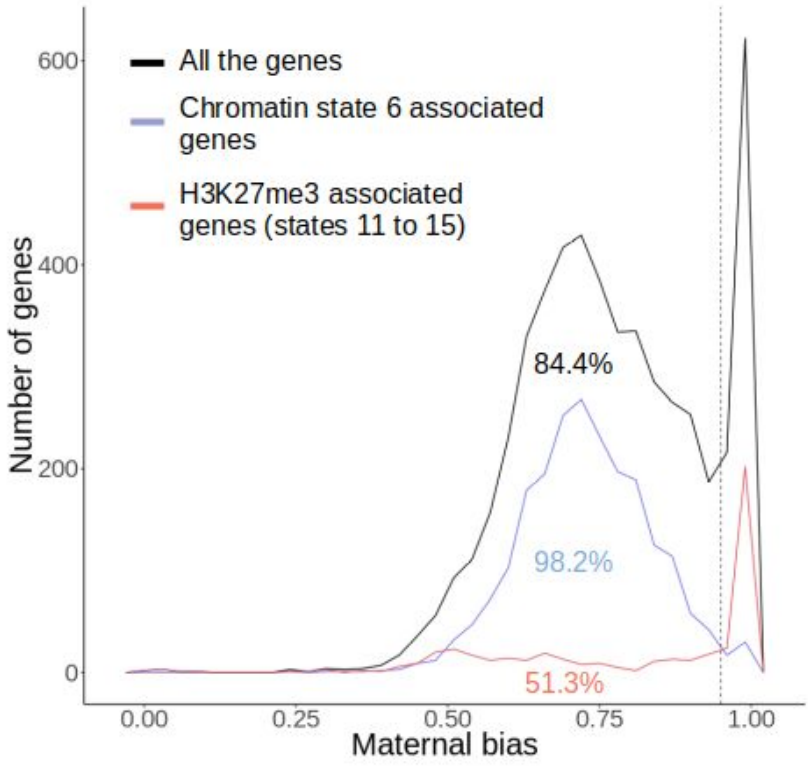


Figure 11: Genes associated with certain chromatin states follow different distributions. Numbers under the curves refer to the percent of genes in that group with under 95% maternal bias. All 3 groups are statistically different from each other (G test for the above/under 95% proportions, $p < 0.001$). 2182, and 464 genes are represented for the chromatin-state-6-associated and H3K27me3-associated genes, respectively. Only genes associated with promoter, introns or exons were considered for CS6, and 5' UTR, exons and introns for H3K27me3 group.

Significant associations were found for several chromatin states and maternal bias (Table 9, Figure 10). Interestingly, the effects of chromatin states seem to group into 2 categories with opposite effects:

I. Activating marks, such as H3K4me1/2/3 or H3K36me3. The genes associated with these marks have a more balanced parental contribution and seem fall almost totally under the more parentally-balanced distribution (Figure 11).

II. Repressive marks, such as H3K27me3 or H3K9me2. Genes associated with these marks fall preferentially under the more maternal distribution (more than 95% maternal, Figure 11).

Not surprisingly, some of these activating and repressing marks have been found to be mutually exclusive in Arabidopsis, with the loss or gain of one type of marks causing the opposite effect on the other type (Deleris et al., 2012; Inagaki et al., 2017; reviewed by Xiao et al, 2016). This could explain the bimodal nature of the maternal bias distribution, with depletion of genes in between the 2; and suggests a dimorphism that could cause this phenomenon: chromatin-silencing pathways enhanced in the sperm and relaxed in the egg (or the opposite for chromatin-activating pathways). However, no data regarding the status of these modifications in plant gametes exist.

State	Preferential marks	Fraction of genes with 95% or more maternal bias when the state is associated with the region					
		Whole	Promoter	5' UTR	Exons	Introns	Downstream
Any	Any	0.156					
1	H3.3	0.08	0.10	0.22	0.07	0.07	0.08
2	H3.3, histone acetylation,H3K4me2,H2A.Z	0.15	0.12	0.48	0.17	0.17	0.12
3	H3K4me1,H3.3,H3.1	0.09	0.09	0.00	0.08	0.08	0.09
4	H3K4me1,H3.3	0.05	0.05	0.08	0.04	0.03	0.08
5	H3K4me1,H3K36me3,H3.3,H3.1	0.03	0.03	0.00	0.02	0.02	0.10
6	H3K4me1,H3K36me3	0.03	0.02	0.00	0.02	0.02	0.08
7	H3K4me1,H3K36me3,H3K4me2	0.05	0.05	0.03	0.05	0.04	0.09
8	H3K4me1,H3K4me2,H2A.Z	0.17	0.12	0.36	0.24	0.23	0.14
9	H3K4me1	0.08	0.09	0.21	0.07	0.06	0.10
10	H2A.Z	0.21	0.18	0.51	0.40	0.47	0.17
11	H3K27me3,H2A.Z,H3K4me2	0.28	0.12	0.52	0.52	0.59	0.15
12	H3K27me3,H2A.Z	0.24	0.20	0.65	0.49	0.54	0.21
13	H3K27me3,H2A.Z	0.38	0.40	0.75	0.58	0.62	0.37
14	H3K27me3	0.27	0.25	0.51	0.46	0.57	0.27
15	H3K27me3,accessible DNA	0.34	0.33	0.50	0.54	0.57	0.36
16	accessible DNA	0.21	0.23	0.45	0.29	0.33	0.24
17	accessible DNA	0.21	0.21	0.27	0.30	0.44	0.27
18	accessible DNA	0.18	0.16	0.21	0.23	0.39	0.23
19	accessible DNA	0.17	0.15	0.12	0.17	0.41	0.19
20	accessible DNA	0.20	0.20	0.27	0.25	0.42	0.16
21	accessible DNA	0.17	0.16	0.27	0.20	0.43	0.17
22	histone acetylation,H3K4me2	0.21	0.14	0.23	0.27	0.35	0.17
23	accessible	0.10	0.07	0.08	0.10	0.13	0.15

	DNA,H3K36ac,H3K56ac,H4K16ac,H3K4me3						
24	accessible DNA,histone acetylation,H3K4me3	0.11	0.06	0.09	0.11	0.11	0.12
25	histone acetylation,H3K4me3,H3K4me2,H2A.Z	0.07	0.04	0.05	0.06	0.06	0.12
26	histone acetylation,H3K4me3,H3K4me2,H2A.Z	0.13	0.06	0.08	0.14	0.10	0.11
27	H3K4me2,histone acetylation,H3K4me3,H2A.Z	0.22	0.10	0.27	0.28	0.28	0.12
28	H3K4me3,H3K4me2,H2A.Z	0.06	0.04	0.08	0.06	0.05	0.09
29	weak signal	0.19	0.20	0.34	0.22	0.40	0.20
30	rare signal	0.16	0.15	0.26	0.18	0.29	0.16
31	DNA methylation,H3K9me2,H3K27me3	0.14	0.08	0.14	0.36	0.56	0.20
32	DNA methylation,H3K9me2	0.13	0.11	NA	NA	NA	0.18
34	H3K9me2,DNA methylation,H3K27me1	0.18	0.03	NA	NA	NA	0.50

Table 9: Correlation of chromatin states and maternal bias. When genes associated with certain chromatin state were heavily biased towards one of the distributions in comparison with the whole set of genes, they were highlighted. If they were biased to the first distribution they are highlighted in blue, biased to the second one are highlighted in red. 'Heavily biased' was called if the ratio of second to first distribution was at least 50% more or 50% less than with the whole set of genes. Chromatin states, associated genes and regions, and preferential marks, were defined by Liu et al., 2018. For the table with *p* values see Supplementary Table S1.

6- Conclusions

I. Reporter results in Col-0 isogenic embryos are in general agreement with RNA-seq data, demonstrating that the maternal transcript bias observed in the Col x Tsu transcriptome is not solely an effect of hybridization, nor maternal tissue contamination.

II. Reporter results cannot so far distinguish between maternal transcriptional bias and carry-over.

III. A *de novo* paternal expression paternal bias was detected for *FUS3*.

IV. The D-box reporter system allows the detection of *de novo* expression for strong promoters.

V. Bioinformatic analyses on egg and zygote transcriptomes suggest that carry-over is finely regulated and examination of the state of transcripts in the egg and early embryos is required to dissect this process.

VI. Association analyses suggest that DNA methylation and histone marks are linked to maternal bias, but not enough experimental data exists to confirm this.

8- Discussion

8.1 Parentally biased reporter expression

As mentioned in the results section, 4 out of 6 reporters were in agreement with the hybrid RNA-seq data. Regarding the 2 whose expression bias didn't match the expected one, the following interpretations can be mentioned:

1. The promoter was strong enough for the expression of the least active allele to be detected with the same frequency (my assay for expression of reporter only determined with the reporter could be seen or not; quantification of expression was not successful).
2. The cloned regulatory region driving expression did not include regulatory regions responsible for parent-of-origin differences.
3. The transcript bias seen in RNA-seq data was an effect of hybridization and therefore not reproducible in isogenic Columbia plants.
4. The constructs in the transgenic lines were inserted in chromatin regions that disrupted the normal parental ratio of expression.
5. Transcript bias observed in RNA-seq data was due to maternal tissue contamination.

Although I do not have sufficient evidence to know which of these phenomena caused the disagreement between expected and obtained results, I consider number 4 and 5 to be unlikely, since imprinted genes are not clustered in Arabidopsis (Ghering et al., 2011), so the region of insertion shouldn't have an effect. There is also no reason to believe contamination had only biased the 2 reporters that didn't behave as expected. I think the RNA-seq results for *CYP86A4* are likely an effect of hybridization, since transcripts for the gene are not detected in isogenic Col datasets (Belmonte et al., 2013; Palovaara et al., 2017).

For the GFP-tagged MINU2, a maternal bias was mistakenly expected, since the maternal phenotype shown in early embryos was actually for the *minu 1/2* double mutant (Del Toro-De León et al., 2014). And parent-of-origin-assigned reads exist only in the Col x Tsu direction of the cross (28 maternal reads, 0 paternal, at zygote stage). While this does suggest maternal bias in transcription, I cannot discard the possibility that the maternal bias seen in double mutants is driven only by *MINU1*, since single mutants exhibit no mutant phenotype (Sang et al., 2012). A significant bias was detected in the quantitative experiments for the *MINU2* reporter, but those results require more extensive sampling (see section "results of quantitative reporter expression").

8.2 *De novo* expression and egg cell expression of reporters

Although parental bias was found in the expression of reporters, generally matching with the RNA-seq results, it is not possible to know whether the observed parental bias was a consequence from carry-over, since penetrance was similar between egg cell and early embryos for all the parentally-biased reporters, despite them not being detected in the egg cell (Table 6). The egg cell expression shows that a careful examination should be done before interpreting the results of the crosses as a maternal bias in *de novo* expression, for which I have so far no evidence. Testing of the remaining reporters is still needed.

Surprisingly, egg cell and 2 DAP (inside the seed, as a mother) expression was clearly detected for the reporter line for *FUS3*, which also showed a paternal bias at 2 DAP. This suggests that there is carry over for the maternal allele, and that quick degradation of the product does not occur, but rather that the paternal allele becomes so active after fertilization that it is able to surpass the maternally inherited products at 2 DAP. So far, the evidence regarding the control of parental contributions points mostly towards silencing pathways repressing the transcription of one of the alleles (Autran et al., 2011, Raissig et al., 2013 , Del Toro-De León et al., 2017). But this results suggests that activating pathways can also cause a bias in parental expression, which is in agreement with previous results for this gene (Raissig et al., 2013). M. Raissig found the transcription of *FUS3* to be undetectable, rather than equalized, when the father was a *met1* mutant.

8.3 D-Box fused 3X-YFP reporter for EC1

Unexpectedly, signal was found for the pEC1::D-box-3xYFP-N7 reporter line. Considering that the expression of the stable reporter (without D-box) was also low shortly after fertilization but recovered a couple of days afterwards, this is most likely due to the reporter not exactly reproducing the expression of EC1. It is worth mentioning that Gerardo del Toro observed the same tendency for a stable GFP reporter driven by the same promoter (Del Toro-De León, 2017 [Doctoral dissertation]).

8.4 Methylation and histone marks

Due to technical difficulties and probably due to prioritization of generalized results, the studies of histone and DNA modifications in plants has been mostly done using broad categories of tissues, such as three week old leaves (Stroud et al., 2013), torpedo stage embryo (Pignatta et al., 2014), 12 days old whole seedlings (Ignaki et al., 2017) or unopened floral buds (Harris et al., 2018); each of these samples consist of several tissues. More focused studies have found significant differences in histone modifications and DNA methylation in the different root tissues (Braszewska-Zalewska et al., 2013; Kawakatsu et al., 2016), between milky-stage embryo, endosperm and aerial tissues (Hsieh et al., 2009)

Due to the lack of tissue specific data, association analyses between epigenetic modifications and maternal bias were done with data from different tissues. Liu chromatin states (Liu et al., 2017) use data for 33 different tissues, which mostly consist of seedlings (102 out of the 216 datasets) and leaves (57 out of 216 datasets), including some with more specific sampling (etiolated seedlings, young leaves, oldest 4 to 6 leaves etc).

The results of these analyses suggest that CG methylation is linked to paternal allele activation. This is reminiscent of a previously described phenomenon in the endosperm. Köler's group found that paternally expressed genes in the endosperm had reduced CG methylation and increased H3K27me3 on the maternal allele, and proposes that the CG demethylation in the central cell (which also happens in the egg cell of rice [Park et al., 2016]) makes the maternal alleles targets for silencing by the FIS-PRC2 (fertilization independent seed - Polycomb Repressive Complex 2), they also showed that in the *fie* mutant (fertilization independen endosperm, a core subunit of the PRC) the imprinting of those genes was lost (Moreno-Romero et al., 2016). Interestingly, they also found that the maternally expressed genes do not depend on Polycomb-deposited marks for their imprinting status, since they also contained H3K27me3 on the maternal allele, and were still imprinted in the *fie* mutant (Moreno-Romero et al., 2016).

Even though Köler's results indicate that the Polycomb Repressive Complex activity promotes paternal dominance for certain genes in the endosperm and that is not involved in maintaining maternally-biased genes, it also shows that this mark is not sufficient to imprint genes, and that the mechanism maintaining maternally biased genes must be different. Contrastingly, the results of association analyses in this study suggest that this mark is involved in maintaining paternal alleles repressed in the early embryo; whether this reflects different mechanisms regulating parental contributions in the embryo and endosperm, or is some sort of artifact, could be experimentally tested in the future.

A simpler, but less

9- Perspectives

More experimental data is needed to better understand the parental contributions in early plant embryogenesis. These studies would most likely generate useful information for other fields also:

1) Post-transcriptional modifications and profiling of RNA binding proteins for gamete transcripts would reveal how the transcript inheritance from the egg is regulated.

2) Histone modifications and DNA methylation profiling in egg, sperm, and early embryo would help elucidate the role of these modifications in regulating parental contributions. It would also provide information about the inheritance and or resetting of such marks.

3) Small RNA profiling of the Arabidopsis egg cell would help confirming the role of RdDM in silencing paternal alleles (Autran et al 2011). Additionally, it would likely provide information about the role of small RNAs in regulating early embryogenesis, for which only limited information is available (Armenta-Medina et al., 2017). Small RNAs have been shown to contribute to balance parental contributions in the endosperm (Erdman et al., 2017).

4) Protocols like GRO-seq (Gardini, 2017), that capture nascent transcripts, would clarify differences in transcriptional activities from both parental alleles.

Additionally, analysis of the intronic and exonic reads in our datasets would likely help to distinguish carry-over from *de novo* transcription, and possibly degradation and transcription rates (Gaidatzis et al., 2017). I consider that all of the above experiments could also provide useful information about the regulation of parthenogenesis and egg activation if compared before and after fertilization.

10- References

1. Aanes, H. *et al.* (2011) 'Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition', *Genome Research*, 21(8), pp. 1328–1338. doi: 10.1101/gr.116012.110.

2. 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), p. 349–358.e4. doi: 10.1016/j.devcel.2017.10.005.

3. Aoki, F., Worrad, D. M. and Schultz, R. M. (1997) 'Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo', Developmental Biology, 181(2), pp. 296–307. doi: 10.1006/dbio.1996.8466.

4. Armenta-Medina, A. *et al.* (2017) 'Arabidopsis thaliana miRNAs promote embryo pattern formation beginning in the zygote', *Developmental Biology*. Elsevier Inc., 431(2), pp. 145–151. doi: 10.1016/j.ydbio.2017.09.009.

5. Autran, D. *et al.* (2011) 'Maternal epigenetic pathways control parental contributions to arabidopsis early embryogenesis', *Cell*, 145(5), pp. 707–719. doi: 10.1016/j.cell.2011.04.014.

6. Avilés-Pagán, E. E. and Orr-Weaver, T. L. (2018) 'Activating embryonic development in Drosophila', *Seminars in Cell and Developmental Biology*. Elsevier Ltd. doi: 10.1016/j.semcdb.2018.02.019.

7. 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. doi: 10.1101/sqb.2008.73.053.

8. Bartolomei, M. S. (2009) 'Genomic imprinting: Employing and avoiding epigenetic processes', *Genes and Development*, 23(18), pp. 2124–2133. doi: 10.1101/gad.1841409.

9. Bar-Zvi, D. *et al.* (2017) 'Hybrid vigor: The best of both parents, or a genomic clash?', *Current Opinion in Systems Biology*. Elsevier Ltd, 6, pp. 22–27. doi: 10.1016/j.coisb.2017.08.004.

10. Bayer, M. *et al.* (2009) ‘Paternal control of embryonic patterning in Arabidopsis thaliana.’, *Science (New York, N.Y.)*, 323(5920), pp. 1485–8. doi: 10.1126/science.1167784.

11. Bazzini, A. A., Lee, M. T. and Giraldez, A. J. (2012) ‘Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in Zebrafish’, *Science*, 336(6078), pp. 233–237. doi: 10.1126/science.1215704.

12. 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. doi: 10.1073/pnas.1222061110.

13. Bushati, N. *et al.* (2008) ‘Temporal Reciprocity of miRNAs and Their Targets during the Maternal-to-Zygotic Transition in Drosophila’, *Current Biology*, 18(7), pp. 501–506. doi: 10.1016/j.cub.2008.02.081.

14. Braszewska-Zalewska, A. J. *et al.* (2013) ‘Tissue-Specific Epigenetic Modifications in Root Apical Meristem Cells of Hordeum vulgare’, *PLoS ONE*, 8(7), pp. 3–10. doi: 10.1371/journal.pone.0069204.

15. Broyer, R. M., Monfort, E. and Wilhelm, J. E. (2017) ‘Cup regulates oskar mRNA stability during oogenesis’, *Developmental Biology*. Elsevier, 421(1), pp. 77–85. doi: 10.1016/j.ydbio.2016.06.040.

16. Chen, J. *et al.* (2017) ‘Zygotic Genome Activation Occurs Shortly after Fertilization in Maize’, *The Plant Cell*, 29(9), pp. 2106–2125. doi: 10.1105/tpc.17.00099.

17. Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., et al. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell*, *110*(1), 33–42.

18. Colón-Carmona, A. *et al.* (1999) ‘Spatio-temporal analysis of mitotic activity with a labile cyclin-GUS fusion protein’, *Plant Journal*, 20(4), pp. 503–508. doi: 10.1046/j.1365-313X.1999.00620.x.

19. Deleris, A. *et al.* (2012) ‘Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana’, *PLoS Genetics*, 8(11). doi: 10.1371/journal.pgen.1003062.

20. Del Toro-De León, G. (2017) 'Análisis genético y genómico de la expresión de origen parental en el desarrollo embrionario de Arabidopsis thaliana', *PhD thesis*, Posgrado en Biotecnología de Plantas, CINVESTAV Irapuato.

21. 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. doi: 10.1038/nature13620.

22. Dowle M. and Srinivasan A. (2018) 'data.table: Extension of `data.frame`'. R package version 1.11.4. https://CRAN.R-project.org/package=data.table

23. Erdmann, R. M. *et al.* (2017) 'A Small RNA Pathway Mediates Allelic Dosage in Endosperm', *Cell Reports*. ElsevierCompany., 21(12), pp. 3364–3372. doi: 10.1016/j.celrep.2017.11.078.

24. Freitas, N. and Cunha, C. (2009) 'Mechanisms and Signals for the Nuclear Import of Proteins', *Current Genomics*, 10(8), pp. 550–557. doi: 10.2174/138920209789503941.

25. FUNAYA, S. and AOKI, F. (2017) 'Regulation of zygotic gene activation by chromatin structure and epigenetic factors', *Journal of Reproduction and Development*, 63(4), pp. 359–363. doi: 10.1262/jrd.2017-058.

26. Gallegos, J. E. and Rose, A. B. (2017) 'Intron DNA Sequences Can Be More Important Than the Proximal Promoter in Determining the Site of Transcript Initiation', *The Plant Cell*, p. tpc.00020.2017. doi: 10.1105/tpc.17.00020.

27. Gagolewski M. et al. (2018). 'R package stringi: Character string processing facilities'. doi:10.5281/zenodo.32557.

28. Gaidatzis, D. *et al.* (2015) 'Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation', *Nature Biotechnology*. Nature Publishing Group, 33(7), pp. 722–729. doi: 10.1038/nbt.3269.

29. 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. doi: 10.1016/j.pbi.2015.05.020.

30. Gardini, A. (2017) 'Global Run-On Sequencing (GRO-Seq)', in *Cell Cycle*, pp. 111–120. doi: 10.1007/978-1-4939-4035-6_9.

31. Gehring, M., Missirian, V. and Henikoff, S. (2011) 'Genomic analysis of parent-of-origin allelic expression in arabidopsis thaliana seeds', *PLoS ONE*, 6(8). doi: 10.1371/journal.pone.0023687.

32. Giraldez, A. J. *et al.* (2006) ‘Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs’, *Science*, 312(5770), pp. 75–79. doi: 10.1126/science.1122689.

33. Grimanelli, D. *et al.* (2005) ‘Timing of the maternal-to-zygotic transition during early seed development in maize.’, *The Plant cell*, 17(4), pp. 1061–1072. doi: 10.1105/tpc.104.029819.

34. Haaf, T. *et al.* (2004) ‘Differential Demethylation of Paternal and Maternal Genomes in the Preimplantation Mouse Embryo : Implications for Mammalian Development’, *Chromosomes Today*, 14, pp. 207–214.

35. Harris, C. J. *et al.* (2018) ‘A DNA methylation reader complex that enhances gene transcription’, *Science*, 362(6419), pp. 1182–1186. doi: 10.1126/science.aar7854.

36. Harvey, E. B. (1936) ‘Parthenogenetic merogony or cleavage without nuclei in arbacia punctulata’, *Biological Bulletin*, 75(1), pp. 170–188. doi: 10.2307/1537682.

37. Harvey, S. A. *et al.* (2013) ‘Identification of the zebrafish maternal and paternal transcriptomes’, *Development*, 140(13), pp. 2703–2710. doi: 10.1242/dev.095091.

38. Heyn, P. *et al.* (2014) ‘The earliest transcribed zygotic genes are short, newly evolved, and different across species’, *Cell Reports*. The Authors, 6(2), pp. 285–292. doi: 10.1016/j.celrep.2013.12.030

39. Hofmeister, B. T. *et al.* (2017) ‘Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation’, *Genome Biology*. Genome Biology, 18(1), pp. 1–16. doi: 10.1186/s13059-017-1288-x.

40. Hsieh, T. *et al.* (2009) ‘Genome-Wide Demethylation of’, *Science*, (June), pp. 3–6. doi: 10.1126/science.1172417.

41. Inagaki, S. *et al.* (2017) ‘Gene-body chromatin modification dynamics mediate epigenome differentiation in *Arabidopsis*’, *The EMBO Journal*, 36(8), pp. 970–980. doi: 10.15252/embj.201694983.

42. Ingouff, M. *et al.* (2010) ‘Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in arabidopsis’, *Current Biology*, 20(23), pp. 2137–2143. doi: 10.1016/j.cub.2010.11.012.

43. Inoue, A. *et al.* (2017) ‘Maternal H3K27me3 controls DNA methylation-independent imprinting’, *Nature*. Nature Publishing Group, 547(7664), pp. 419–424. doi: 10.1038/nature23262.

44. Ivanova, I. *et al.* (2017) 'The RNA m6A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence', *Molecular Cell*, 67(6), p. 1059–1067.e4. doi: 10.1016/j.molcel.2017.08.003.

45. Jian, Y. *et al.* (2012) 'Primer- BLAST: A tool to design target-specific primers for polymerase chain reaction', *BMC Bioinformatics*, 13(1), p. 134. doi: 10.1186/1471-2105-13-134.

46. Jukam, D., Shariati, S. A. M. and Skotheim, J. M. (2017) 'Zygotic Genome Activation in Vertebrates', *Developmental Cell*. Elsevier Inc., 42(4), pp. 316–332. doi: 10.1016/j.devcel.2017.07.026.

47. Karimi, M., Inzé, D. and Depicker, A. (2002) 'GATEWAY™ vectors for Agrobacterium-mediated plant transformation', *Trends in Plant Science*, 7(5), pp. 193–195. doi: 10.1016/S1360-1385(02)02251-3.

48. Kawakatsu, T. *et al.* (2016) 'Unique cell-type-specific patterns of DNA methylation in the root meristem', *Nature Plants*, 2(5). doi: 10.1038/NPLANTS.2016.58.

49. Kumar, M. *et al.* (2013) 'Novel insights into the genetic and epigenetic paternal contribution to the human embryo', *Clinics*, 68(S1), pp. 5–14. doi: 10.6061/clinics/2013(Sup01)02.

50. Kuo-Chen Chou and Hong-Bin Shen, "Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization", PLoS ONE, 2010, 5: e11335.

51. 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, pp. 581–613. doi: 10.1146/annurev-cellbio-100913-013027.

52. Lee, M. T. *et al.* (2013) 'Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition', *Nature* 503(7476), pp. 360–364. doi: 10.1038/nature12632.

53. Liang, H. L. *et al.* (2008) 'The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila', *Nature*, 456(7220), pp. 400–403. doi: 10.1038/nature07388.

54. Liu, C. (2013) 'Strategies for Designing Transgenic DNA Constructs', in *Methods Mol Biol*, pp. 183–201. doi: 10.1007/978-1-60327-369-5_8.

55. Liu, Y. *et al.* (2018) 'PCSD: A plant chromatin state database', *Nucleic Acids Research*. Oxford University Press, 46(D1), pp. D1157–D1167. doi: 10.1093/nar/gkx919.

56. Logemann, E. *et al.* (2006) 'An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol', *Plant Methods*, 2(1), pp. 1–5. doi: 10.1186/1746-4811-2-16.

57. Lukowitz, W. *et al.* (2004) 'A MAPKK Kinase Gene Regulates Extra-Embryonic Cell Fate in Arabidopsis', *Cell*, 116(1), pp. 109–119. doi: 10.1016/S0092-8674(03)01067-5.

58. Luo, A. *et al.* (2014) 'The expression and roles of parent-of-origin genes in early embryogenesis of angiosperms.', *Frontiers in plant science*, 5(December), p. 729. doi: 10.3389/fpls.2014.00729.

59. Matthews-Rochon, E. *et al.* (1994) 'Isolation and microinjection of active sperm nuclei into egg cells and central cells of isolated maize embryo sacs', *Zygote*, 2(1), pp. 29–35. doi: 10.1017/S0967199400001738.

60. Mayer, U., Büttner, G. and Jürgens, G. (1993) 'Apical-basal pattern formation in the Arabidopsis embryo', *Development*, 117, pp. 149–162.

61. Moreno-Romero, J. *et al.* (2016) 'Parental epigenetic asymmetry of PRC2-mediated histone modifications in the *Arabidopsis* endosperm', *The EMBO Journal*, 35(12), pp. 1298–1311. doi: 10.15252/embj.201593534.

62. Ning, J. *et al.* (2006) 'Differential gene expression in egg cells and zygotes suggests that the transcriptome is restructed before the first zygotic division in tobacco', *FEBS Letters*, 580(7), pp. 1747–1752. doi: 10.1016/j.febslet.2006.02.028.

63. 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. doi: 10.1038/nature10756.

64. Park, K. *et al.* (2016) 'DNA demethylation is initiated in the central cells of *Arabidopsis* and rice', *Proceedings of the National Academy of Sciences*, 113(52), pp. 15138–15143. doi: 10.1073/pnas.1619047114.

65. 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. doi: 10.1105/tpc.109.071647.

66. Pires, N. D. and Grossniklaus, U. (2014) 'Different yet similar: evolution of imprinting in flowering plants and mammals', *F1000Prime Reports*, 6(August), pp. 12–15. doi: 10.12703/P6-63.

67. Poccia, D. *et al.* (1985) 'RNA synthesis in male pronuclei of the sea urchin', *BBA - Gene Structure and Expression*, 824(4), pp. 349–356. doi: 10.1016/0167-4781(85)90042-9.

68. Radhakrishnan, A. *et al.* (2016) 'The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality', *Cell*. Elsevier Inc., 167(1), p. 122–132.e9. doi: 10.1016/j.cell.2016.08.053.

69. Raissig, M. T. *et al.* (2013) 'Genomic Imprinting in the Arabidopsis Embryo Is Partly Regulated by PRC2', *PLoS Genetics*, 9(12), pp. 21–23. doi: 10.1371/journal.pgen.1003862.

70. R Core Team (2013). 'R: A language and environment for statistical computing. R Foundation for Statistical Computing'.

71. Rodrigues, J. A. and Zilberman, D. (2015) 'Evolution and function of genomic imprinting in plants', *Genes and Development*, 29(24), pp. 2517–2531. doi: 10.1101/gad.269902.115.

72. Sang, Y. *et al.* (2012) 'Mutations in two non-canonical Arabidopsis SWI2/SNF2 chromatin remodeling ATPases cause embryogenesis and stem cell maintenance defects', *Plant Journal*, 72(6), pp. 1000–1014. doi: 10.1111/tpj.12009.

73. Schmitz, R. J. *et al.* (2013) 'Patterns of population epigenomic diversity', *Nature*, 495(7440), pp. 193–198. doi: 10.1038/nature11968.

74. Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*. Nature Publishing Group, 9(7), pp. 671–675. doi: 10.1038/nmeth.2089.

75. 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. doi: 10.1105/tpc.16.00845.

76. Shermoen, A. W. and O'Farrell, P. H. (1991) 'Progression of the cell cycle through mitosis leads to abortion of nascent transcripts', *Cell*, 67(2), pp. 303–310. doi: 10.1016/0092-8674(91)90182-X.

77. Spencer, H. G. and Clark, A. G. (2014) ‘Non-conflict theories for the evolution of genomic imprinting’, *Heredity*. Nature Publishing Group, 113(2), pp. 112–118. doi: 10.1038/hdy.2013.129.

78. Sprunck, S. *et al.* (2012) ‘Egg Cell-Secreted EC1 Triggers Sperm Cell Activation During Double Fertilization’, *Science*, 338(6110), pp. 1093–1097. doi: 10.1126/science.1223944.

79. Stroud, H. *et al.* (2013) ‘Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome’, *Cell*. Elsevier Inc., 152(1–2), pp. 352–364. doi: 10.1016/j.cell.2012.10.054.

80. Tadros, W. *et al.* (2007) ‘SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase’, *Developmental Cell*, 12(1), pp. 143–155. doi: 10.1016/j.devcel.2006.10.005.

81. Tadros, W. and Lipshitz, H. D. (2009) ‘The maternal-to-zygotic transition : a play in two acts’, 3042, pp. 3033–3042. doi: 10.1242/dev.033183.

82. Telford, N., Watson, A. and Schultz, G. A. (1990) ‘Transition from maternal to embryonic control in early mammalian development: a comparison of several species’, *Molecular reproduction and development*, 26(1), pp. 90–100. Available at: http://onlinelibrary.wiley.com/doi/10.1002/mrd.1080260113/full.

83. Toda, E., Ohnishi, Y. and Okamoto, T. (2018) ‘An imbalanced parental genome ratio affects the development of rice zygotes’, *Journal of Experimental Botany*, 69(10), pp. 2609–2619. doi: 10.1093/jxb/ery094.

84. Tzafrir, I. *et al.* (2004) ‘Identification of Genes Required for Embryo Development in Arabidopsis 1 [w]’, *Plant physiology*, 135(July), pp. 1206–1220. doi: 10.1104/pp.104.045179.published.

85. Ueda, M., Zhang, Z. and Laux, T. (2011) *Transcriptional Activation of Arabidopsis Axis Patterning Genes WOX8/9 Links Zygote Polarity to Embryo Development*, *Developmental Cell*. doi: 10.1016/j.devcel.2011.01.009.

86. 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. doi: 10.1038/35003595.

87. Wang, F. and Perry, S. E. (2013) ‘Identification of Direct Targets of FUSCA3, a Key Regulator of Arabidopsis Seed Development’, *Plant Physiology*, 161(3), pp. 1251–1264. doi: 10.1104/pp.112.212282.

88. Wang, M. *et al.* (2017) 'ME31B globally represses maternal mRNAs by two distinct mechanisms during the Drosophila maternal-to-zygotic transition', *eLife*, 6, pp. 1–22. doi: 10.7554/eLife.27891.

89. Wickham H. (2016). 'ggplot2: Elegant Graphics for Data Analysis'. Springer-Verlag. ISBN: 978-3-319-24277-4.

90. Wickham H. (2018). 'stringr: Simple, Consistent Wrappers for Common String Operations'. R package version 1.3.1. https://CRAN.R-project.org/package=stringr

91. Wiekowski, M., Miranda, M. and DePamphilis, M. L. (1993) 'Requirements for promoter activity in mouse oocytes and embryos distinguish paternal pronuclei from maternal and zygotic nuclei', *Developmental Biology*, pp. 366–378. doi: 10.1006/dbio.1993.1248.

92. Wuest, S. E. *et al.* (2010) 'Arabidopsis Female Gametophyte Gene Expression Map Reveals Similarities between Plant and Animal Gametes', *Current Biology*, 20(6), pp. 506–512. doi: 10.1016/j.cub.2010.01.051.

93. Xiang, D., Del Toro De León, G., Alaniz Fabián, J., Abreu Goodger, C., Gillmor, C.S., and Datla, R. 'Maternal transcript bias during early *Arabidopsis thaliana* embryogenesis'. *Manuscript in preparation.*

94. Xiang, D. *et al.* (2011) 'Genome-Wide Analysis Reveals Gene Expression and Metabolic Network Dynamics during Embryo Development in Arabidopsis', *Plant Physiology*, 156(1), pp. 346–356. doi: 10.1104/pp.110.171702.

95. Xiao, J., Lee, U. S. and Wagner, D. (2016) 'Tug of war: adding and removing histone lysine methylation in Arabidopsis', *Current Opinion in Plant Biology*. Elsevier Ltd, 34, pp. 41–53. doi: 10.1016/j.pbi.2016.08.002.

96. Xiao, Y. L. *et al.* (2010) 'High throughput generation of promoter reporter (GFP) transgenic lines of low expressing genes in Arabidopsis and analysis of their expression patterns', *Plant Methods*, 6(1), pp. 1–13. doi: 10.1186/1746-4811-6-18.

97. Xin, H. P. *et al.* (2011) 'Expressed sequence-tag analysis of tobacco sperm cells reveals a unique transcriptional profile and selective persistence of paternal transcripts after fertilization', *Sexual Plant Reproduction*, 24(1), pp. 37–46. doi: 10.1007/s00497-010-0151-y.

98. Zemach, A. *et al.* (2010) 'Genome-wide evolutionary analysis of eukaryotic DNA methylation', *Science*, 328(5980), pp. 916–919. doi: 10.1126/science.1186366.

99. Zhou, M. *et al.* (2014) 'Mediator: A key regulator of plant development', *Cell*. Nature Publishing Group, 15(1), pp. 7–18. doi: 10.1016/j.ydbio.2016.06.009.

11- Supplementary data

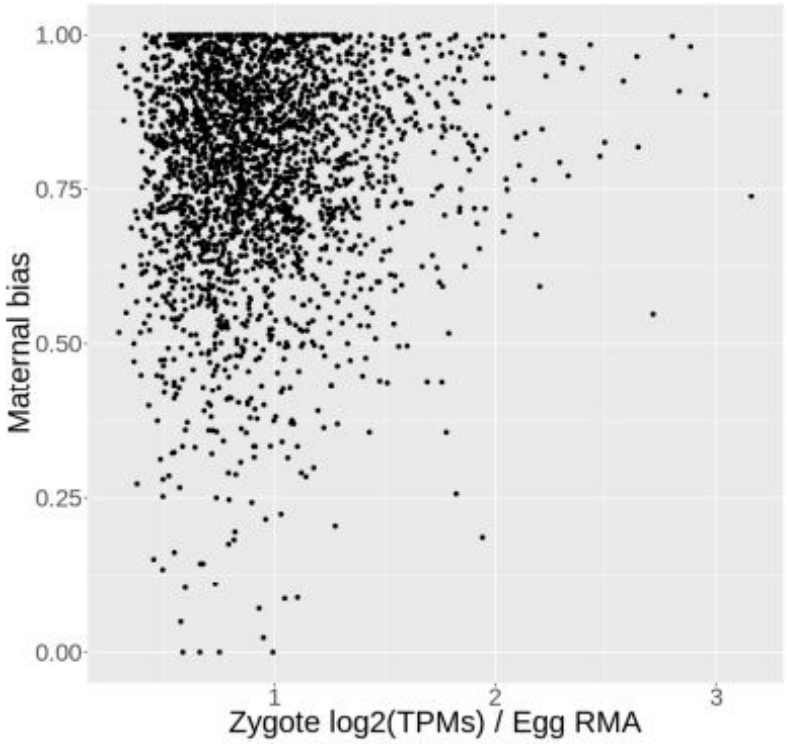
Reporter for:	Days after pollination	With maternal signal	Examined maternal	With paternal signal	Examined paternal	Maternal bias	*p* (Fisher's exact test)
L41	2	33	51	0	68	1	1.0E-16
	3	19	30	0	98	1	2.5E-15
	5	39	40	26	28	0.51	0.56
FUS3	2	40	65	64	68	0.40	4.8E-6
	3	36	42	37	37	0.46	0.03
	5	-	-	61	62	-	-
MINU2	2	48	51	41	46	0.51	0.47
	3	48	48	38	38	0.5	1
	5	28	29	38	38	0.49	0.43
CYP86A4	2	48	63	32	45	0.52	0.66
	3	45	47	59	59	0.49	0.19
	5	36	36	55	55	0.5	1
CAF1a	2	57	62	37	62	0.61	4.1E-5
	3	53	53	59	63	0.52	0.12
	5	62	62	41	41	0.5	1
PLT1	2	41	41	36	37	0.51	0.47
	3	41	41	43	43	0.5	1
	5	38	38	-	-	-	-

Table S1: Penetrance counts for reporters along development when inherited from each parent.

State	Whole	Prom.	5' UTR	Exon	Intron	Dwnstrm	whole p val	prom. p val	5'UTR p val	exon p val	intron p val	down p val
1	0.08	0.10	0.22	0.07	0.07	0.08	2E-21	5E-03	2E-01	2E-29	8E-12	1E-15
2	0.15	0.12	0.48	0.17	0.17	0.12	6E-02	4E-01	1E-05	2E-05	2E-02	2E-01
3	0.09	0.09	0.00	0.08	0.08	0.09	3E-08	3E-02	NA	3E-08	5E-05	2E-03
4	0.05	0.05	0.08	0.04	0.03	0.08	1E-41	3E-09	6E-01	4E-52	2E-47	2E-06
5	0.03	0.03	0.00	0.02	0.02	0.10	8E-46	1E-02	NA	1E-51	1E-45	2E-01
6	0.03	0.02	0.00	0.02	0.02	0.08	4E-71	9E-08	NA	9E-77	7E-79	2E-03
7	0.05	0.05	0.03	0.05	0.04	0.09	1E-41	1E-06	1E-05	2E-40	4E-39	2E-03
8	0.17	0.12	0.36	0.24	0.23	0.14	6E-04	4E-01	9E-10	3E-11	2E-07	5E-01
9	0.08	0.09	0.21	0.07	0.06	0.10	2E-11	2E-01	4E-01	1E-10	4E-12	6E-02
10	0.21	0.18	0.51	0.40	0.47	0.17	6E-14	1E-02	4E-11	4E-42	3E-39	5E-03
11	0.28	0.12	0.52	0.52	0.59	0.15	3E-27	6E-01	8E-15	8E-67	1E-45	3E-01
12	0.24	0.20	0.65	0.49	0.54	0.21	3E-14	4E-04	9E-27	1E-42	2E-40	7E-04
13	0.38	0.40	0.75	0.58	0.62	0.37	6E-44	2E-30	6E-35	5E-45	3E-31	1E-15
14	0.27	0.25	0.51	0.46	0.57	0.27	1E-23	1E-11	5E-09	5E-31	8E-37	8E-15
15	0.34	0.33	0.50	0.54	0.57	0.36	3E-12	1E-08	3E-06	6E-12	2E-08	4E-05
16	0.21	0.23	0.45	0.29	0.33	0.24	1E-16	6E-15	5E-08	8E-10	4E-06	1E-11
17	0.21	0.21	0.27	0.30	0.44	0.27	7E-21	4E-15	5E-08	1E-19	1E-17	5E-21
18	0.18	0.16	0.21	0.23	0.39	0.23	1E-09	4E-04	6E-05	5E-12	3E-12	2E-13
19	0.17	0.15	0.12	0.17	0.41	0.19	7E-06	9E-02	4E-01	4E-03	1E-08	3E-04
20	0.20	0.20	0.27	0.25	0.42	0.16	8E-14	3E-09	2E-08	4E-15	4E-16	2E-02
21	0.17	0.16	0.27	0.20	0.43	0.17	2E-08	1E-03	2E-07	2E-09	7E-26	6E-05
22	0.21	0.14	0.23	0.27	0.35	0.17	2E-21	5E-01	1E-11	6E-36	2E-41	1E-02
23	0.10	0.07	0.08	0.10	0.13	0.15	4E-05	3E-16	7E-11	2E-05	9E-01	4E-01
24	0.11	0.06	0.09	0.11	0.11	0.12	7E-03	3E-17	2E-08	1E-02	4E-02	6E-01
25	0.07	0.04	0.05	0.06	0.06	0.12	8E-27	6E-17	5E-23	7E-36	5E-38	4E-01

26	0.13	0.06	0.08	0.14	0.10	0.11	6E-01	1E-05	2E-03	3E-01	2E-02	1E-01
27	0.22	0.10	0.27	0.28	0.28	0.12	4E-21	1E-01	1E-05	2E-36	7E-14	6E-01
28	0.06	0.04	0.08	0.06	0.05	0.09	3E-37	2E-12	4E-04	3E-30	2E-40	8E-04
29	0.19	0.20	0.34	0.22	0.40	0.20	5E-18	2E-15	2E-08	1E-13	3E-24	8E-15
30	0.16	0.15	0.26	0.18	0.29	0.16	6E-04	1E-01	7E-02	2E-03	1E-07	1E-02
31	0.14	0.08	0.14	0.36	0.56	0.20	7E-01	1E-03	9E-01	3E-04	2E-05	3E-03
32	0.13	0.11	NA	NA	NA	0.18	1E+00	7E-01	NA	NA	NA	6E-01
34	0.18	0.03	NA	NA	NA	0.50	3E-01	2E-02	NA	NA	NA	2E-04

Table S2: Chromatin states in each gene region and association with maternal bias.



Supplementary figure S1: Maternal bias in the zygote vs the zygote to egg ratio of transcript abundance. R=0.10, P= 2*10-7. Transcripts with low zygote to egg ratio are expected to be more sensitive to carry-over bias from the egg, however, this is not the case. Since experimental evidence has shown carry-over (Pillot et al., 2010), this suggest that a process more complex than mere abundance is regulating which transcripts get carried over (See discussion).

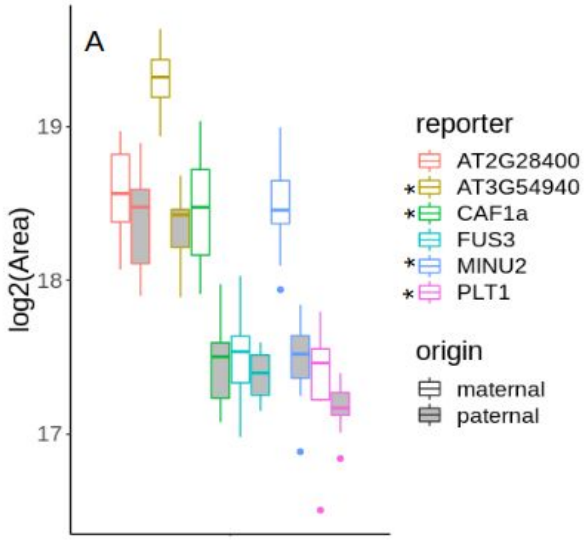
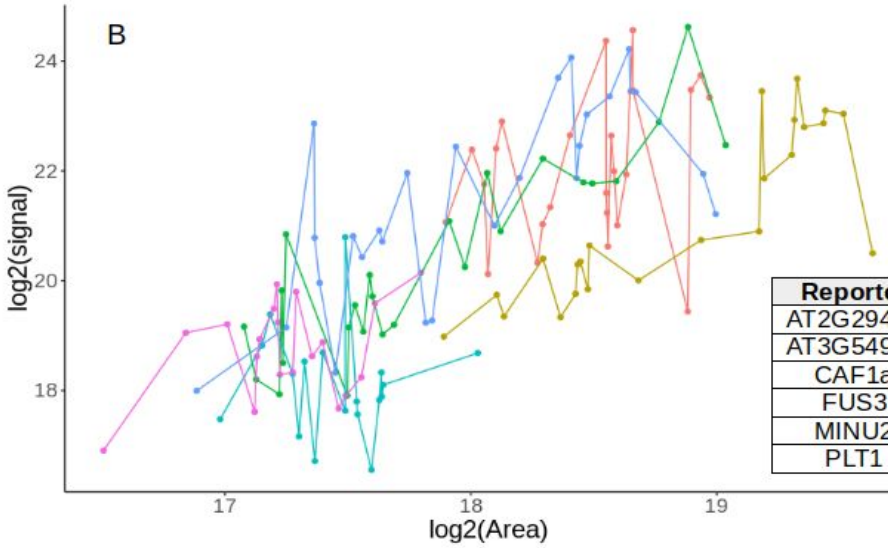
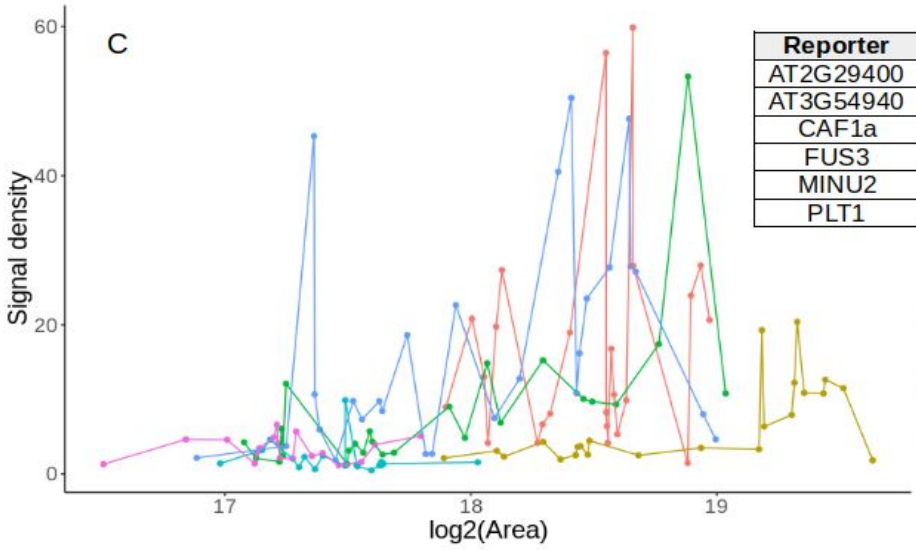


Figure S2: Data distribution in the quantification of reporter expression.
A: Areas in maternally-inherited and paternally-inherited embryos are different for some reporters, data at 2 DAP. * : P < 0.05 (Mann-Whitney test).
B: The detected signal is dependent upon the area, linear correlations (Pearson's) are shown, correlations and p values were calculated with non logarithmic data.
C: Signal density (signal / area) is also dependent upon the area, p values and correlations calculated as for B.



Reporter	Corr	P val
AT2G29400	0.4	0.06
AT3G54940	0.7	1.6*10^-4
CAF1a	0.69	1.1*10^-4
FUS3	-0.02	0.93
MINU2	0.55	0.003
PLT1	0.4	0.12



Reporter	Corr	P val
AT2G29400	0.2	0.37
AT3G54940	0.61	0.002
CAF1a	0.65	5*10^-4
FUS3	-0.16	0.51
MINU2	0.34	0.08
PLT1	0.03	0.91

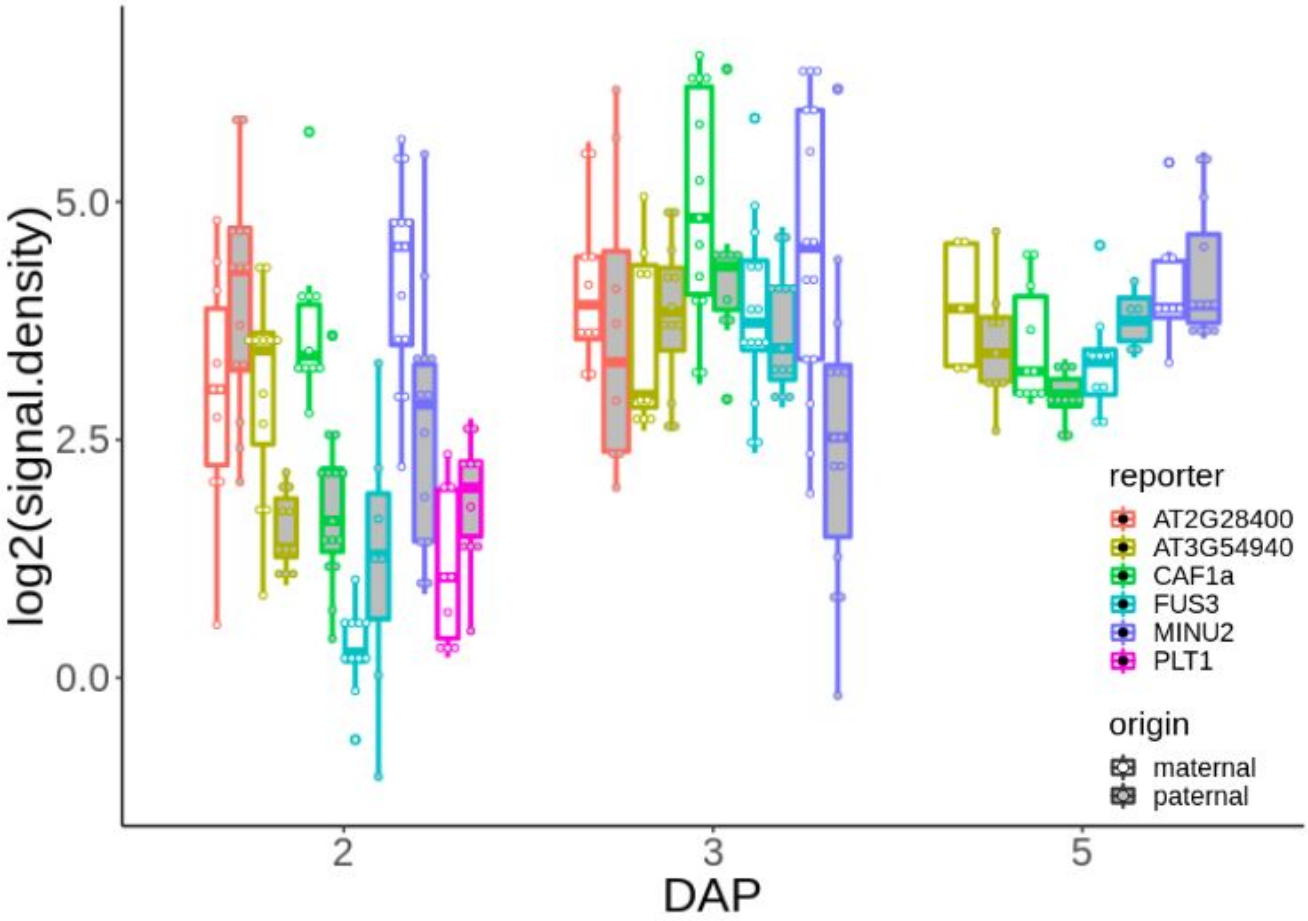


Figure S3: Signal intensity of the reporters inherited maternally and paternally in embryos. Signal was measured using Image J, differences should be interpreted with consideration of the possible bias (see figure S2).

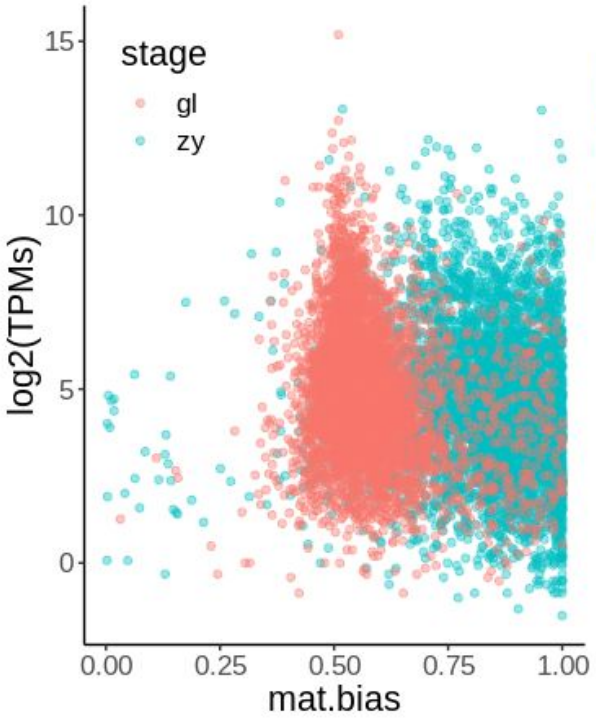


Figure S4: Transcript abundance vs maternal bias. No strong correlation is observed. Filters for at least 20 parent-of-origin-assigned reads and maximum 0.3 difference in maternal bias between directions of the cross were applied.

Stage	Pearson's correlation	P value
Zygote	-0.063	2.4*10^-5
Octant	-0.034	1.3*10^-2
Globular	-0.043	2.7*10^-3
Heart	-0.015	0.29

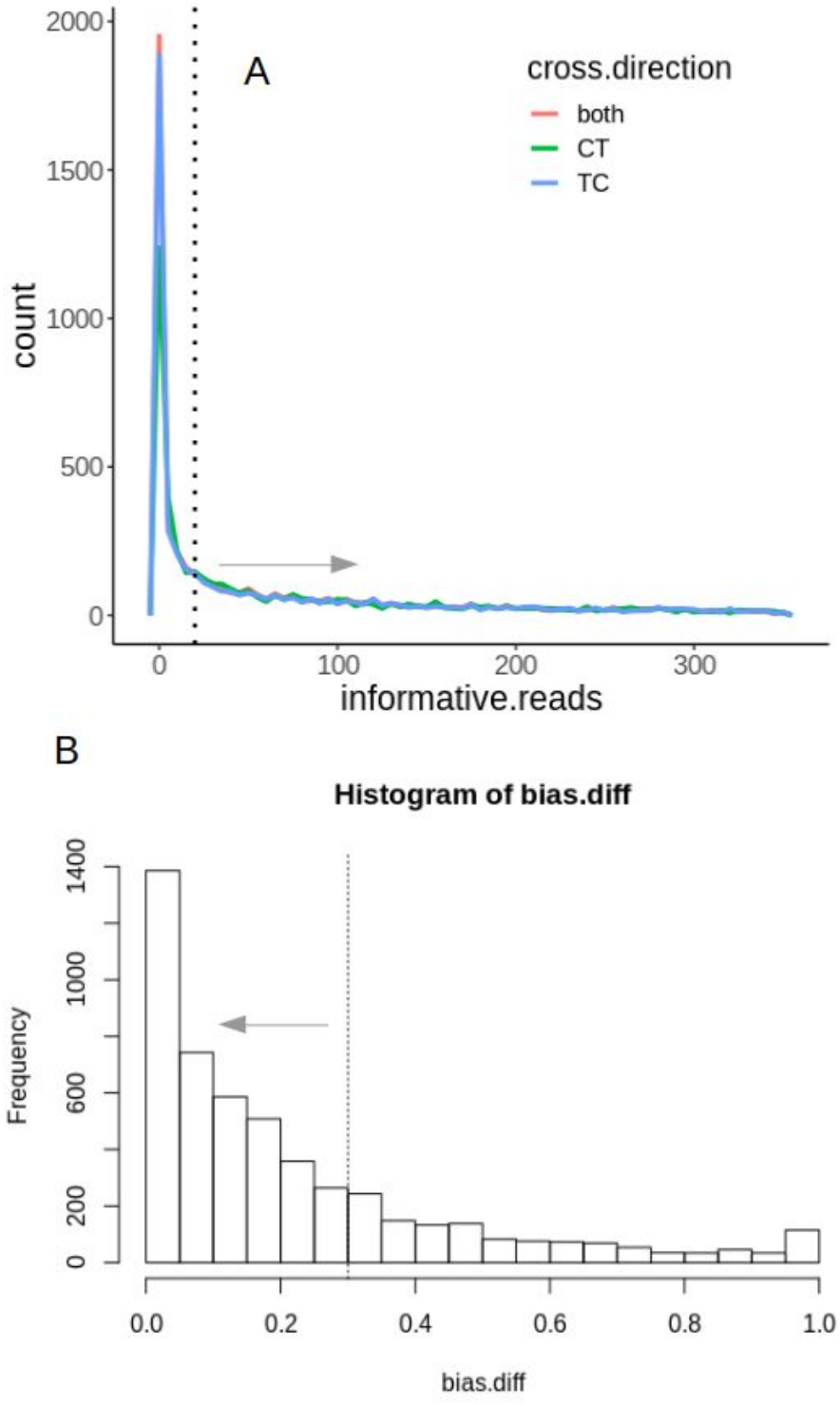


Figure S5: distributions of informative reads and difference in maternal bias. (A) 4167 genes (60.5% of possibly useful genes) comply with requirement for at least 20 parent-of-origin-assigned reads in both directions of the cross. (B) 3837 genes (56.1% of possibly useful genes) comply with the requirement for a maximum of 0.3 bias difference between directions of the cross (1709 genes are not represented in the plot due to lack of informative reads in both directions). Possible useful genes were designated as so if they had transcript differences between ecotypes and were detected with at least 3 TPMs at the zygote. 3247 genes passed both filters. Dotted lines were drawn at the cutoff values.

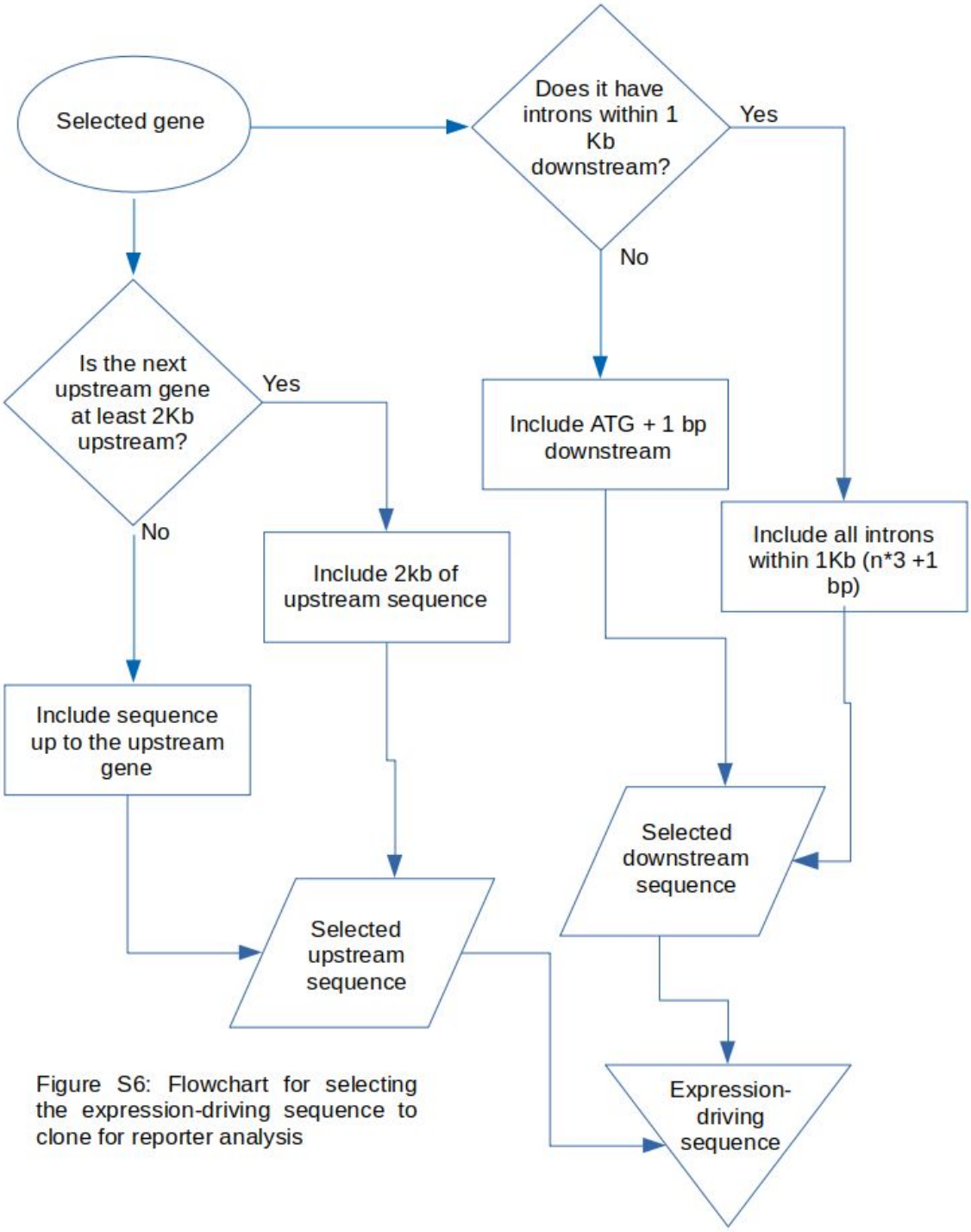


Figure S6: Flowchart for selecting the expression-driving sequence to clone for reporter analysis