

# Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos

Michael D. Nodine<sup>1,2</sup> & David P. Bartel<sup>1,2</sup>

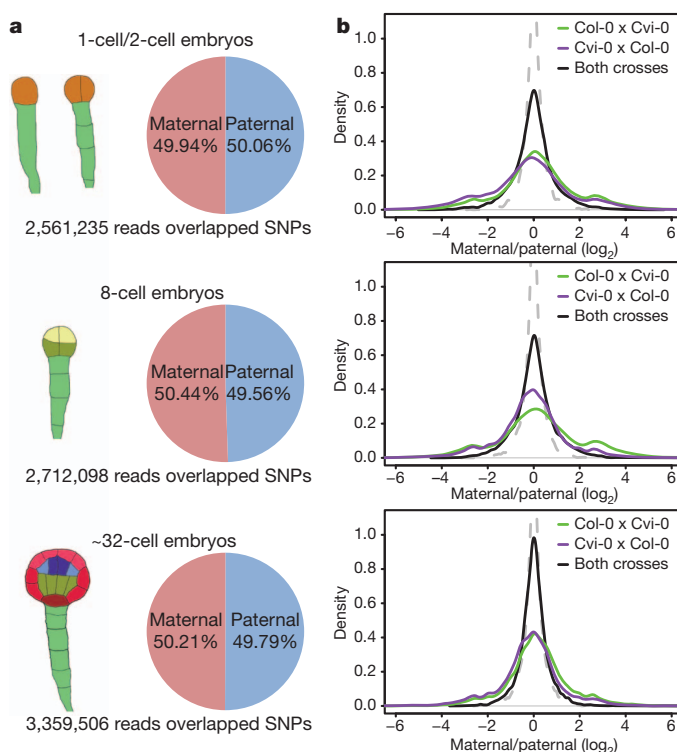
In animals, maternal gene products deposited into eggs regulate embryonic development before activation of the zygotic genome<sup>1</sup>. In plants, an analogous period of prolonged maternal control over embryogenesis is thought to occur based on some gene-expression studies<sup>2–6</sup>. However, other gene-expression studies and genetic analyses show that some transcripts must derive from the early zygotic genome<sup>7–14</sup>, implying that the prevailing model does not fully explain the nature of zygotic genome activation in plants. To determine the maternal, paternal and zygotic contributions to the early embryonic transcriptome, we sequenced the transcripts of hybrid embryos from crosses between two polymorphic inbred lines of *Arabidopsis thaliana* and used single-nucleotide polymorphisms diagnostic of each parental line to quantify parental contributions. Although some transcripts seemed to be either inherited from primarily one parent or transcribed from imprinted loci, the vast majority of transcripts were produced in near-equal amounts from both maternal and paternal alleles, even during the initial stages of embryogenesis. Results of reporter experiments and analyses of transcripts from genes that are not expressed in sperm and egg indicate early and widespread zygotic transcription. Thus, in contrast to early animal embryogenesis, early plant embryogenesis is mostly under zygotic control.

The prevailing model for the maternal-to-zygotic transition in plants proposes that most early embryonic messenger RNAs are maternally derived transcripts, resulting either from maternal inheritance or from higher transcriptional activity of maternally derived genes until the globular stages (in which the embryo proper has between ~32 to >100 cells)<sup>2,6,15</sup>. Because fundamental patterning events, including apical–basal and radial axis formation<sup>16,17</sup>, occur during the preglobular stages, this model implies that key cell-specification decisions are mostly under maternal control. However, this model is difficult to reconcile with other studies that report equivalent maternal and paternal expression of interrogated genes in preglobular stages<sup>8,9,11</sup> and zygotic-recessive behaviour of mutants with preglobular developmental phenotypes<sup>12–14</sup>.

To determine the origins of embryonic transcripts globally, we crossed polymorphic Col-0 and Cvi-0 *Arabidopsis thaliana* accessions and performed RNA-seq on poly(A)<sup>+</sup> RNA isolated from hybrid embryos with either one-to-two, eight, or ~32 cells in the embryo proper (hereafter referred to as 1-cell/2-cell, 8-cell and ~32-cell embryos). To control for inherent expression differences between Col-0 and Cvi-0 loci, the same procedure was performed using embryos derived from reciprocal crosses. Illumina sequencing of the six samples yielded 73,955,956 reads that both perfectly and uniquely matched the transcribed regions of 23,874 genes (Supplementary Table 1). Overall, transcript levels from the same stage but different reciprocal crosses were highly correlated ( $r \geq 0.96$ ; Supplementary Fig. 1). This reproducibility and sequencing depth indicated that our results would be informative for inferring the maternal, paternal and zygotic contributions to the early embryonic transcriptome.

The prevailing model for the maternal-to-zygotic transition predicted that at the early embryonic stages transcripts would derive primarily

from maternal alleles, and then at the globular stage they would derive more evenly from both alleles because at this stage the zygotic genome would be active. In contrast to this expectation, we found equal amounts of paternally and maternally derived reads at all three stages, including the 1-cell/2-cell and 8-cell embryos (Fig. 1a). When examining transcripts with at least five reads overlapping single-nucleotide polymorphisms (SNPs) in each cross, most were expressed equally from the maternal and paternal alleles, even at the earliest stage (Fig. 1b and Supplementary Data 1; note that in the Col-0 × Cvi-0



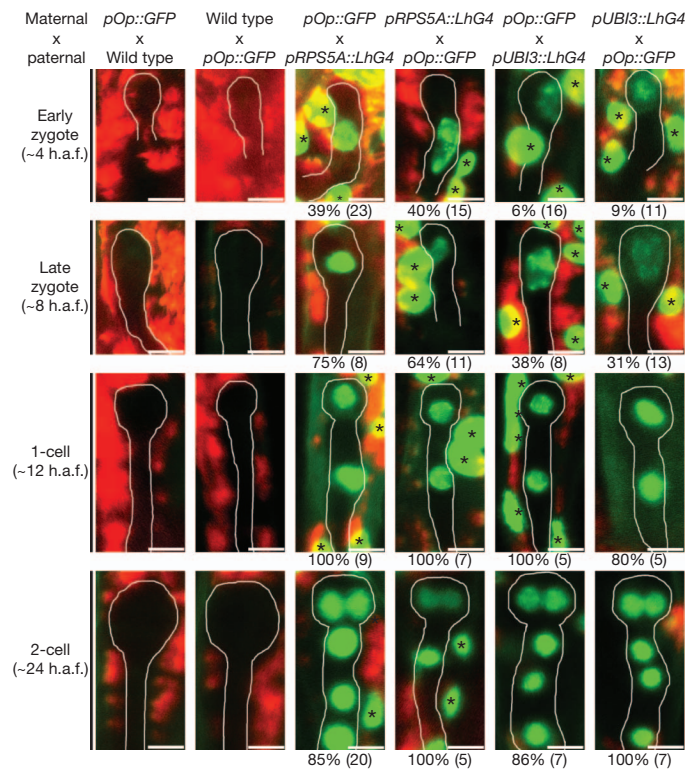
**Figure 1 | Maternal and paternal genomes contribute equally to the early embryonic transcriptome.** **a**, Proportion of RNA-seq reads overlapping maternal and paternal SNPs within transcribed regions of annotated genes. Drawings illustrate stages of isolated embryos, coloured to represent the diversity of cell lineages in each stage. Pie charts show proportions of reads overlapping maternal and paternal SNPs. **b**, Distributions of maternal-to-paternal ratios ( $\log_2$ ) in 1-cell/2-cell (top), 8-cell (middle) and ~32-cell (bottom) embryos. Maternal-to-paternal ratios were calculated for transcripts that had  $\geq 5$  SNP-overlapping reads from the Col-0 × Cvi-0 cross (green;  $n > 7,388$  for each stage) or the Cvi-0 × Col-0 cross (purple;  $n > 8,071$  for each stage), and transcripts that had  $\geq 5$  SNP-overlapping reads from both reciprocal crosses (black;  $n > 6,892$  for each stage). Binomial modelling determined the distributions expected for transcripts examined in both reciprocal crosses if there had been no parent-of-origin effects and no noise apart from stochastic counting variability (grey dashed lines).

<sup>1</sup>Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>2</sup>Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

cross, the maternal parent is Col-0, as the convention is to write the maternal parent first). In each cross, small but discernable subpopulations of transcripts were derived predominantly from either maternal or paternal alleles. However, in the respective reciprocal cross those same transcripts tended to derive from the opposite parent, which showed that rather than arising from parent-of-origin effects, these subpopulations arose mostly from genotypic effects, that is, from preferential expression from either the Col-0 or Cvi-0 alleles. Thus, when considering results of both crosses together, no overall maternal (or paternal) bias was observed, and at each stage the distribution of maternal-to-paternal ratios resembled that predicted for equal contribution from both alleles, with most of the variability explained by stochastic counting statistics, as modelled by the binomial distribution (Fig. 1b). The remainder of the variability was attributed to both additional noise (both experimental and biological) and a small subset of transcripts with parent-of-origin effects, for which the number that were maternally enriched approximately equalled the number that were paternally enriched.

Because most transcripts were derived from the maternal and paternal genomes in near-equal amounts, we hypothesized that either equal amounts of each transcript were inherited upon fertilization or the zygotic genome was activated much earlier than previously proposed. Supporting the latter possibility, the fold-change distributions centred at 0.0 ( $\log_2$ ), showing no hint of the maternal bias that would be expected if the egg, with its larger cytoplasm, contributed more RNA than the sperm (Fig. 1b). In addition, 543 transcripts had at least four-fold higher reads per million (genome- and cDNA-matching reads) values at the 8-cell stage than at the 1-cell/2-cell stages, which suggested active transcription of the corresponding genes between the 1-cell/2-cell and 8-cell stages (as the alternative model of differential mRNA stability would require a large decrease in total mRNA between these two stages; Supplementary Data 1). Furthermore, 1,138 transcripts that were previously called undetectable in both egg and sperm microarray data sets<sup>18,19</sup> were among the top 50% most abundant transcripts in our 1-cell/2-cell data sets (Supplementary Data 1). These results, taken together with the observation that transcripts for RNA polymerase II subunits were among the most abundant in the early embryo (Supplementary Table 2), suggested that during the initial stages of embryogenesis many transcripts are transcribed from both maternal and paternal alleles, and that this transcription, combined with turnover of inherited transcripts, quickly overwrites most parent-of-origin biases present when the egg and sperm first fuse.

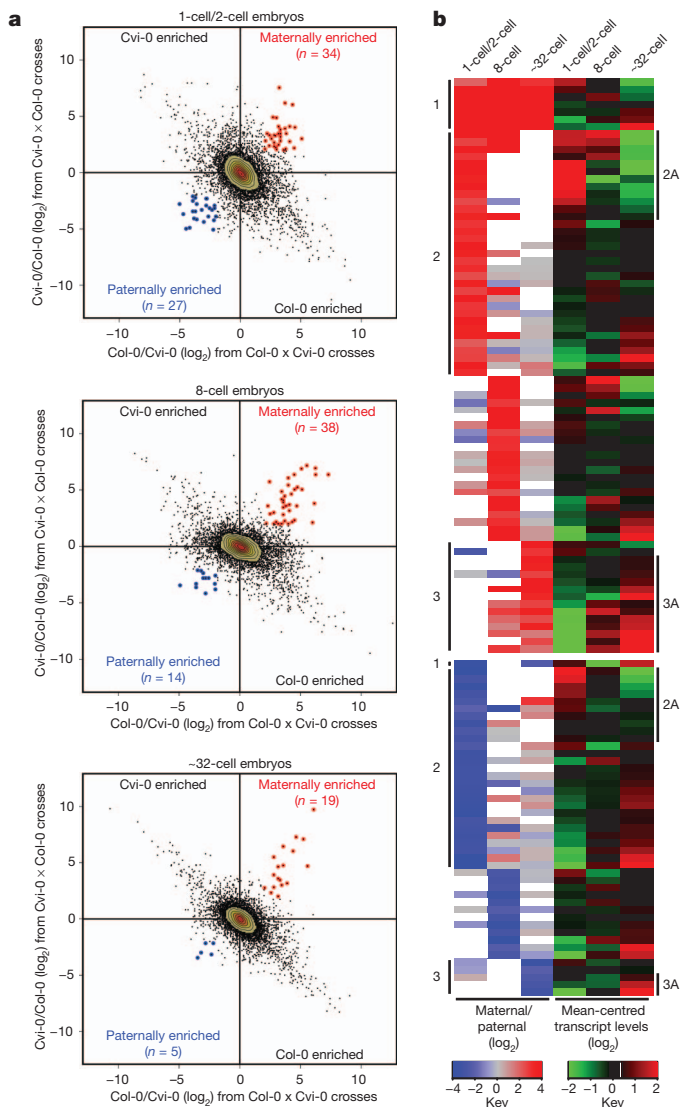
To test directly whether maternally and paternally inherited genes are transcribed in very early embryos, we used the LhG4/pOp transactivation system<sup>20</sup>. In this system, one parent contained a transgene encoding the LhG4 transcription factor under the control of either the *RPS5A* or the *UBI3* promoter (chosen because they generate products that can be found in the early embryo<sup>9,21</sup>), whereas the other parent harboured a nuclear-localized green fluorescent protein (GFP) reporter transgene under the control of the artificial pOp promoter to which LhG4 binds and transcriptionally activates (*pOp::GFP*). Because one gamete contributed the reporter gene and the other gamete contributed its activator, any reporter expression dependent on the activator could not occur until after the zygotic genome was transcriptionally active. No GFP signal was observed in embryos carrying the *pOp::GFP* reporter gene but no activator, which confirmed the absence of leaky expression in either the gametes or embryo (Fig. 2). GFP signal was detected in zygotes within 4–8 h after fertilization from crosses that brought the reporter together with its activator, regardless of whether the reporter was inherited through the egg or sperm (Fig. 2). The GFP signal in the early zygote was not as strong as that in the endosperm, which might indicate a slight delay in activation of the genome of the very early zygote or might result from more robust transcription in the endosperm. Nonetheless, when considered together our results demonstrate that both maternally and paternally inherited chromosomes are transcriptionally active at least in 1-cell embryos and most likely before.



**Figure 2 | Maternally and paternally inherited transgenes are transcribed in the initial stages of embryogenesis.** Confocal scanning laser microscopy images show early embryos from the indicated crosses, which were between plants with a transgenic reporter (*pOp::GFP*) and either plants without an activator (wild type) or plants with a transgenic activator driven from either the *RPS5A* or the *UBI3* promoter (*pRPS5A::LhG4* or *pUBI3::LhG4*, respectively). Embryonic stages and the approximate number of hours after fertilization (h.a.f.) are shown to the left. GFP signal is indicated by green nuclei; background autofluorescence is shown in red. Embryos are outlined in white for clarity. Asterisks indicate endosperm nuclei. Of the seeds with GFP signal in endosperm ( $n$  in parentheses), the percentage that also had GFP signal in embryos is indicated below each image. In endosperm, expression from maternal alleles is expected to be twofold higher than that from paternal alleles, but this difference is not observed in these images because the GFP signal in the endosperm was beyond its dynamic range. Scale bars represent 10  $\mu$ m.

Having established that most early embryonic transcripts derived equally from the maternal and paternal genomes, we turned our attention to the few that might be preferentially inherited or preferentially expressed from the maternal or paternal alleles. For most early embryonic transcripts the genotype of the allele had a larger effect on transcript levels than did parent of origin (Figs 1b and 3a). Indeed, hundreds of transcripts were preferentially expressed from either the Col-0 or Cvi-0 allele irrespective of their parent of origin (Supplementary Data 2), presumably as direct consequences of either DNA-sequence or epigenetic differences between the Col-0 and Cvi-0 alleles. This prevalence of genotypic/epigenotypic effects over parent-of-origin effects led to significant negative correlations between the maternal-to-paternal ratios at each stage (Fig. 3a, Pearson's  $r < -0.43$ ,  $P < 10^{-15}$ ). Nonetheless, at each stage a small subset of transcripts passed our cutoffs for classification as maternally or paternally enriched, that is, fourfold maternally or paternally enriched in each cross, and not exceeding the false discovery rate (FDR) threshold of 0.05 (Pearson's  $\chi^2$  tests, Benjamini and Hochberg FDR). With these cutoffs, 77 and 45 transcripts were designated maternally and paternally enriched, respectively, in at least one stage (Fig. 3a and Supplementary Data 3). For five transcripts, an independent assay involving diagnostic cleavage of an amplified polymorphic sequence was performed, and for all five, the parent-of-origin effects were confirmed (Supplementary Fig. 2).





**Figure 3 | Dozens of transcripts are preferentially inherited or preferentially expressed.** **a**, Maternal-to-paternal ratios for transcripts in 1-cell/2-cell, 8-cell and ~32-cell embryos. Ratios for transcripts that had  $\geq 8$  SNP-overlapping reads in both the Col-0  $\times$  Cvi-0 and Cvi-0  $\times$  Col-0 crosses are plotted. Points, as well as the numbers, that passed cutoffs for maternally or paternally enriched transcripts (see text) are indicated in red and blue, respectively. **b**, Heat maps of both maternal-to-paternal ratios and changes in transcript levels for transcripts passing cutoffs as maternally (top) or paternally (bottom) enriched in at least one of the three stages. Maternal-to-paternal ratios are colour-coded (key), unless the transcript had  $< 8$  SNP-overlapping reads in one cross, in which case the corresponding heat-map value is not coloured. Changes in the levels of a transcript (reads per million) relative to its mean are colour-coded (key). Numbers to the left indicate candidates for both preferential inheritance and preferential expression (1), preferential inheritance (2) or preferential expression (3). Numbers to the right highlight a subset of the class 2 transcripts, which had relatively high levels in 1-cell/2-cell embryos compared to ~32-cell embryos and thus potentially derived mainly from preferential inheritance (2A), or a subset of class 3 transcripts, which had relatively low levels in 1-cell/2-cell embryos compared to ~32-cell embryos and thus potentially derived mainly from parent-of-origin-specific transcription (3A).

Although the maternal enrichment of some transcripts might be due to their transport from maternal sporophytic tissues, such a mechanism cannot explain paternal enrichment. Some transcripts had strong parent-of-origin effects in 1-cell/2-cell embryos, suggesting that they were preferentially inherited from one parent (Fig. 3b). Confirmation that these biases are indeed due to inheritance would substantially add to the single previously documented example of an

inherited transcript in plants<sup>22</sup>. Other transcripts had stronger parent-of-origin effects at later stages, suggesting that they were preferentially expressed from either the maternally or the paternally inherited allele (Fig. 3b). This potential parent-of-origin-specific expression implied a form of imprinting in the *Arabidopsis* embryo. Genome-wide approaches similar to ours but looking much later after fertilization greatly expanded the list of genes with parent-of-origin-specific expression in the endosperm but did not identify such genes in embryos<sup>23,24</sup>. Thus, the imprinting-like phenomenon that we observed in early embryos is short lived, which suggests that it differs from the more persistent imprinting previously characterized in either endosperm or mammals<sup>25</sup>. For example, it might involve alternative chromatin states inherited from the egg and sperm, which after several cell divisions equilibrate between the two alleles. Such a mechanism would not necessarily require DNA methylation, although we note that in the only previous report of imprinting in plant embryos (at the *MEE1* locus of maize), methylation marks are lost from maternal alleles at the initial stages of embryogenesis, but then re-established to match the paternal alleles at later stages<sup>26</sup>.

Our results showing that during the initial stages of *Arabidopsis* embryogenesis both the maternal and paternal genomes are active and make essentially equivalent contributions to the embryonic transcriptome are in marked contrast to a recent report that ~88% of the 2-cell/4-cell embryonic transcriptome is derived from the maternal genome<sup>2</sup>. That study, published while our manuscript was in preparation, used SNPs in the transcriptomes of hybrid embryos derived from a cross between the Ler-1 (maternal parent) and Col-0 (paternal parent) accessions to estimate maternal and paternal genomic contributions. Because results from the reciprocal cross were not reported, the previous study could not distinguish parent-of-origin effects from genotypic effects. However, because an analysis unable to account for genotypic effects would be expected to misidentify transcripts as paternally derived as frequently as it misidentifies them as maternally derived, we considered other possibilities for the discrepancy between their results and ours. Our pilot studies had shown that early embryos must be extensively washed to prevent seed-coat RNA contamination, and indeed we found evidence that the embryo RNA samples of ref. 2 contained large amounts of seed-coat mRNA (Supplementary Fig. 3). Because the seed coat is a maternal tissue, this contamination explained why they observed such a large bias in maternal RNAs.

Genes zygotically required for the initial zygotic division have been identified in *Arabidopsis*<sup>9,12,13</sup>. Moreover, paternal gene expression has been detected for 24 endogenous genes and a transgene in maize zygotes<sup>8,27</sup>, and *de novo* transcription has been demonstrated recently in tobacco zygotes<sup>28</sup>. Rather than interpreting these findings as exceptions to the model or as differences between species, our transcriptome-wide analyses and reporter data strongly support the proposal of an alternative model for the maternal-to-zygotic (or, to put it more precisely, the maternal/paternal-to-zygotic) transition in plants. In this model, the zygotic genome is activated within the first few hours after fertilization, and equal transcription of both maternal and paternal alleles generates most of the early embryonic transcriptome at the 1-cell stage or before. Our model implies that although the maternal and paternal gene products are inherited in the zygote upon fertilization and contribute to its development<sup>22</sup>, the zygotic gene products are the ones that primarily control the initial cell-specification events of embryogenesis.

## METHODS SUMMARY

Col-0 and Cvi-0 seed stocks were obtained from the *Arabidopsis* Biological Resource Center. Flowers were emasculated 1 day before crossing, and pools of 20 1-cell/2-cell, 8-cell and ~32-cell embryos were hand-dissected approximately 40, 64 and 78 h after pollination, respectively. RNA isolation, linear amplification of poly(A) RNA and strand-specific RNA-seq were as described<sup>21</sup>, except that embryos were extensively washed before RNA isolation. After removing adaptor sequences, reads were mapped to both the *A. thaliana* genome and transcript models (Col-0 genome reference, TAIR10) with the Bowtie short-read aligner<sup>29</sup>.

Reads were also mapped to a 'pseudo' Cvi-0 genome and Cvi-0 transcript models, in which SNPs in the Col-0 genome and transcript models were replaced with Cvi-0 variants ([ftp://ftp.arabidopsis.org/home/tair/Sequences/Polymorphism\\_datasets/Ecker\\_Cvi\\_snps.txt](ftp://ftp.arabidopsis.org/home/tair/Sequences/Polymorphism_datasets/Ecker_Cvi_snps.txt)) generated by the 1001 Genomes Project. Reads that both perfectly and uniquely matched either the genomes or transcript models were retained, and those overlapping transcribed regions of annotated genes were evaluated for overlap with SNPs. Those diagnostic of one genome or the other were tallied after normalizing for differences in sequencing yield between samples. Tallies for multiple SNPs within the same gene were combined before calculating maternal-to-paternal ratios for each transcript.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 5 September; accepted 1 December 2011.**

**Published online 22 January 2012.**

- Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042 (2009).
- Autran, D. *et al.* Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* **145**, 707–719 (2011).
- Vielle-Calzada, J. P., Baskar, R. & Grossniklaus, U. Delayed activation of the paternal genome during seed development. *Nature* **404**, 91–94 (2000).
- Baroux, C., Blanvillain, R. & Gallois, P. Paternally inherited transgenes are down-regulated but retain low activity during early embryogenesis in *Arabidopsis*. *FEBS Lett.* **509**, 11–16 (2001).
- Grimanelli, D., Perotti, E., Ramirez, J. & Leblanc, O. Timing of the maternal-to-zygotic transition during early seed development in maize. *Plant Cell* **17**, 1061–1072 (2005).
- Baroux, C., Autran, D., Gillmor, C. S., Grimanelli, D. & Grossniklaus, U. The maternal to zygotic transition in animals and plants. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 89–100 (2008).
- Springer, P. S., Holding, D. R., Groover, A., Yordan, C. & Martienssen, R. A. The essential Mcm7 protein PROLIFERA is localized to the nucleus of dividing cells during the G<sub>1</sub> phase and is required maternally for early *Arabidopsis* development. *Development* **127**, 1815–1822 (2000).
- Meyer, S. & Scholten, S. Equivalent parental contribution to early plant zygotic development. *Curr. Biol.* **17**, 1686–1691 (2007).
- Weijers, D., Geldner, N., Offringa, R. & Jurgens, G. Seed development: Early paternal gene activity in *Arabidopsis*. *Nature* **414**, 709–710 (2001).
- Sprunck, S., Baumann, U., Edwards, K., Langridge, P. & Dresselhaus, T. The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). *Plant J.* **41**, 660–672 (2005).
- Ingouff, M. *et al.* Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in *Arabidopsis*. *Curr. Biol.* **20**, 2137–2143 (2010).
- Xu, J. *et al.* EMBRYONIC FACTOR 1 encodes an AMP deaminase and is essential for the zygote to embryo transition in *Arabidopsis*. *Plant J.* **42**, 743–758 (2005).
- Ronceret, A. *et al.* The first zygotic division in *Arabidopsis* requires *de novo* transcription of thymidylate kinase. *Plant J.* **53**, 776–789 (2008).
- Muralla, R., Lloyd, J. & Meinke, D. Molecular foundations of reproductive lethality in *Arabidopsis thaliana*. *PLoS ONE* **6**, e28398 (2011).
- Pillot, M. *et al.* Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in *Arabidopsis*. *Plant Cell* **22**, 307–320 (2010).
- Mayer, U., Torres Ruiz, R. A., Berleth, T., Misera, S. & Jurgens, G. Mutations affecting body organization in the *Arabidopsis* embryo. *Nature* **353**, 402–407 (1991).
- Laux, T., Wurschum, T. & Breuninger, H. Genetic regulation of embryonic pattern formation. *Plant Cell* **16** (Suppl. 1), S190–S202 (2004).
- Wuest, S. E. *et al.* *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol.* **20**, 506–512 (2010).
- Borges, F. *et al.* Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiol.* **148**, 1168–1181 (2008).
- Moore, I., Galweiler, L., Grosskopf, D., Schell, J. & Palme, K. A transcription activation system for regulated gene expression in transgenic plants. *Proc. Natl Acad. Sci. USA* **95**, 376–381 (1998).
- Nodine, M. D. & Bartel, D. P. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev.* **24**, 2678–2692 (2010).
- Bayer, M. *et al.* Paternal control of embryonic patterning in *Arabidopsis thaliana*. *Science* **323**, 1485–1488 (2009).
- Gehring, M., Missirian, V. & Henikoff, S. Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS ONE* **6**, e23687 (2011).
- Hsieh, T. F. *et al.* Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc. Natl Acad. Sci. USA* **108**, 1755–1762 (2011).
- Feil, R. & Berger, F. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet.* **23**, 192–199 (2007).
- Jahnke, S. & Scholten, S. Epigenetic resetting of a gene imprinted in plant embryos. *Curr. Biol.* **19**, 1677–1681 (2009).
- Scholten, S., Lorz, H. & Kranz, E. Paternal mRNA and protein synthesis coincides with male chromatin decondensation in maize zygotes. *Plant J.* **32**, 221–231 (2002).
- Zhao, J. *et al.* Dynamic changes of transcript profiles after fertilization are associated with *de novo* transcription and maternal elimination in tobacco zygote, and mark the onset of the maternal-to-zygotic transition. *Plant J.* **65**, 131–145 (2011).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Ecker and the 1001 Genomes Project for generating the list of Cvi-0 SNPs; the J. Harada and R. Goldberg laboratories for making the *Arabidopsis* seed development LCM microarray data sets publicly available; I. Moore for the transactivation vectors; J. Long for the *UBI3* promoter fragment; the Whitehead Genome Technology Core for sequencing; and D. Meinke for providing a curated list of preglobular zygotic-recessive mutants before publication. This work used the W. M. Keck Biological Imaging Facility at the Whitehead Institute and was supported by NIH grant GM067031 (D.P.B.) and NIH Postdoctoral Fellowship GM084656 (M.D.N.). D.P.B. is an Investigator of the Howard Hughes Medical Institute.

**Author Contributions** M.D.N. designed and performed the experiments. M.D.N. and D.P.B. interpreted the results and wrote the manuscript.

**Author Information** Raw and processed RNA-seq data sets have been deposited into NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE33713. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.P.B. ([dbartel@wi.mit.edu](mailto:dbartel@wi.mit.edu)).

## METHODS

**Plant growth and embryo dissections.** Col-0 and Cvi-0 seed stocks were obtained from the *Arabidopsis* Biological Resource Center. Plants were grown at 22 °C in a Conviron growth chamber with a 16-h light/8-h dark cycle. Flowers were emasculated 1 day before crossing, and pools of 20 1-cell/2-cell, 8-cell and ~32-cell embryos were hand-dissected approximately 40, 64 and 78 h after pollination, respectively.

Embryo dissections were performed in water on glass slides using a SZH stereo microscope (Olympus) with 128× magnification. Tungsten needles (Fine Science Tools) were used to cut open the seeds and remove the embryos from the rest of the seed material. For each embryo, a heat-pulled borosilicate glass microcapillary pipette (Harvard Apparatus) was used to move it to a depression slide containing 50 µl water. The embryo was then washed two additional times by serial transfer to slides containing 50 µl water, using a new pipette and fresh water for each transfer. Washed embryos were collected in 30 µl of RNAlater (Ambion).

**RNA-seq and analyses.** RNA isolation, linear amplification of poly(A) RNA and strand-specific RNA-seq were as described<sup>21</sup>, except that embryos were extensively washed before RNA isolation. After removing adaptor sequences, reads were mapped to both the *A. thaliana* genome and transcript models (Col-0 genome reference, TAIR10) with the Bowtie short-read aligner<sup>29</sup>. Reads were also mapped to a 'pseudo' Cvi-0 genome and Cvi-0 transcript models, in which SNPs in the Col-0 genome and transcript models were replaced with Cvi-0 variants ([ftp://ftp.arabidopsis.org/home/tair/Sequences/Polymorphism\\_datasets/Ecker\\_Cvi\\_snps.txt](ftp://ftp.arabidopsis.org/home/tair/Sequences/Polymorphism_datasets/Ecker_Cvi_snps.txt)) generated by the 1001 Genomes Project. Reads that both perfectly and uniquely matched either the genomes or transcript models were retained, and those overlapping transcribed regions of annotated genes were evaluated for overlap with SNPs. To normalize for differences in library sizes, the numbers of reads representing each transcript were divided by the total number of reads matching the genome and transcript models. SNP-overlapping reads were assigned to one of the parental genomes and tallied for each transcript, combining tallies for multiple SNPs within the same transcript.

**Statistical analyses and graphics.** Statistical analyses were performed and associated graphics were generated with the R statistical computing base package<sup>30</sup>. The SciPy tools for Python were used to calculate  $\chi^2$  test statistics and associated probabilities.

**Transgenic lines.** The pBIN+LhG4-GW vector was generated by digesting both the pBIN+LhG4 vector and the attR1/attR2-containing vector GWRFa::pET42a

with KpnI and ligating together the appropriate digestion fragments. *RPS5A* and *UBI3* promoter fragments were amplified with appropriate primer pairs, cloned into pENTR-D/TOPO (Invitrogen) and then recombined with pBIN+LhG4-GW to generate pRPS5A::LhG4-GW and pUBI3::LhG4-GW, respectively. The pV-TOP-GFP vector was generated by double-digesting pV-TOP(E3) and NLSGFP1R1::pCR8 with HindIII and PmeI, and ligating together the appropriate digestion fragments to replace the GUS reporter with a nuclear-localized GFP reporter. NLSGFP1R1::pCR8 was created by amplifying pCGTAG with appropriate primers and cloning into pCR8/GW-TOPO (Invitrogen). pOp::GFP was then generated by digesting pV-TOP-GFP and GWF1R3::pENTR with Sall, and ligating together the appropriate digestion fragments. GWF1R3::pENTR was created by amplifying the attR1/attR2 Gateway cassette from GWRFa::pET42a with a series of overlap extension PCRs to remove internal Sall sites and add Sall sites on the ends of amplicons, which were then cloned into pENTR/D-TOPO (Invitrogen). Oligonucleotides used to generate the above constructs are listed in Supplementary Table 3. All constructs were transformed into Col-0 by *Agrobacterium*-mediated transformation. Because the progeny of T<sub>1</sub> lines gave more robust GFP signal than did progeny of established lines, T<sub>1</sub> lines were used for all reporter crosses. Embryos from crosses using at least 13 *pOp::GFP*, four *pRPS5A::LhG4* and six *pUBI3::LhG4* independent T<sub>1</sub> lines were examined.

**Microscopy.** For confocal scanning laser microscopy, developing seeds were mounted in 50 mM potassium phosphate buffer, pH 7.2, with 5% glycerol. A 488-nm laser on a Zeiss LSM 510 confocal microscope was used to excite GFP and seed autofluorescence, and images were collected at 505–530 nm and 644–719 nm, respectively.

**Analyses of cleaved amplified polymorphic sequences.** Transcripts containing a SNP that created or disrupted a restriction site in the corresponding cDNA were selected for analysis. Random hexamers (Invitrogen) were used for reverse transcription with Superscript III (Invitrogen) and primers flanking the SNPs were used for amplification (Supplementary Table 3). The amplified DNA was digested with the appropriate restriction enzyme (New England Biolabs; Supplementary Table 4). Digestion products were resolved on 2% agarose gels stained with ethidium bromide, and bands were quantified with Quantity One 1-D analysis software (Bio-Rad).

30. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Austria, 2010).