Transcriptome dynamics revealed by a gene expression atlas of the early Arabidopsis embryo

Joakim Palovaara¹, Shunsuke Saiga¹, Jos R. Wendrich¹,³, Nicole van ‘t Wout Hofland¹, J. Paul van Schayck¹, Friederike Hater¹, Sumanth Mutte¹, Jouke Sjollema¹, Mark Boekschoten², Guido J. Hooiveld² and Dolf Weijers¹*

During early plant embryogenesis, precursors for all major tissues and stem cells are formed. While several components of the regulatory framework are known, how cell fates are instructed by genome-wide transcriptional activity remains unanswered—in part because of difficulties in capturing transcriptome changes at cellular resolution. Here, we have adapted a two-component transgenic labelling system to purify cell-type-specific nuclear RNA and generate a transcriptome atlas of early Arabidopsis embryo development, with a focus on root stem cell niche formation. We validated the dataset through gene expression analysis, and show that gene activity shifts in a spatio-temporal manner, probably signifying transcriptional reprogramming, to induce developmental processes reflecting cell states and state transitions. This atlas provides the most comprehensive tissue- and cell-specific description of genome-wide gene activity in the early plant embryo, and serves as a valuable resource for understanding the genetic control of early plant development.

And plants (Embryophytes) develop as a result of the continuous formation of tissue layers and organs from stem cells. This necessitates the coordination of oriented cell divisions, cell–cell communication and genetic regulatory mechanisms that together form a framework for cell fate determination and pattern formation. The framework is first established during early embryogenesis where it sets the basic body pattern from tissue precursors and stem cells. Thus, the formation of all plant structures, regardless of complexity, is the result of iteration of a developmental process first established in the early embryo. As such, plant embryogenesis can be considered a miniature model for plant development. Consequently, a central question in plant biology is how the framework is established and regulated in the embryo to allow for cell and tissue specification, and, subsequently, an ordered three-dimensional structure.

In seed plants, the mature embryo (seedling) displays the same basic body plan as the post-embryonic structure: an overlapping apical–basal and radial pattern with a shoot and root apical meristem along the apical–basal axis of polarity, and the basic tissues (epidermis, vascular tissue, ground tissue) forming concentric radial layers. One or more cotyledons flank the shoot apical meristem. The meristems are tissues that contain specialized microenvironments called niches, in which stem cells continuously divide to produce differentiated cells (for tissue/organ formation) and so-called organizer cells prevent stem cell differentiation. The niche organizer is termed the organizer centre in shoots and the quiescent centre (QC) in roots. In the flowering plant Arabidopsis thaliana, a highly invariant cell division pattern makes it possible to track the origin of the basic body plan to specific cell types in the early embryo. This, together with its small genome size and relative ease with which it can be genetically manipulated, has made Arabidopsis the de facto model organism to study the cell fate of stem cells, organizer cells and precursors of the major tissues. In past years, several important components of the regulatory framework underlying their establishment have been identified (reviewed previously); however, a key question of how these cell fates are instructed by cellular transcriptomes remains unanswered.

A major challenge has been to adapt cell-specific genome-wide approaches to the small plant embryo contained within seed and fruit tissues. Manual and laser capture microdissection have been used to isolate whole embryos and larger embryonic tissue domains from various plant species for transcriptional analysis. Notably, this has elucidated spatio-temporal gene expression patterns and the dynamics of zygotic genome activation and microRNA function. While important, these studies are limited when investigating developmental decisions since they, by definition, occur at a cellular scale. However, the recent development of several cell-specific approaches in other systems, such as fluorescence-activated cell/nuclei sorting (FACS/FANS) and transgenic labelling and affinity purification of nuclei/polysomes (INTACT/TRAP), has now made it possible to determine gene expression at the single-cell-type level in the plant embryo. Of these methods, INTACT holds the most promise for studying embryogenesis.

Although both FACS/FANS and TRAP have provided important insights regarding cell fate during Arabidopsis development (reviewed previously), a main concern is that promoters used to transgenically mark cells need to be exclusive to a specific cell or tissue type. This is an issue especially in embryos since many embryonically expressed genes are also found in equivalent tissues of the surrounding seed material. INTACT circumvents this by utilizing a two-component transgenic labelling system where biotin ligase (BirA) biotinylates a nuclear envelope-localized GFP protein (nuclear targeting fusion protein (NTF)) when co-expressed in the same cells. The fast, specific and high-affinity binding between

¹Laboratory of Biochemistry, Wageningen University, 6708 WE Wageningen, The Netherlands. ²Nutrition, Metabolism and Genomics Group, Division of Human Nutrition, Wageningen University, 6708 WE Wageningen, The Netherlands. Present address: ³Department of Plant Biotechnology and Bioinformatics and VIB Center for Plant Systems Biology, Ghent University, Technologiepark 927, 9052 Ghent, Belgium. Joakim Palovaara and Shunsuke Saiga contributed equally to this work. *e-mail: dolf.weijers@wur.nl

Corrected: Publisher correction
biotin and streptavidin is then exploited to efficiently isolate biotin-tagged nuclei from crude nuclear preparations using streptavidin-coated beads. To date, INTACT has been used to isolate cell-type-specific nuclei from Arabidopsis root\textsuperscript{26,34}, female central cell\textsuperscript{16} and endosperm\textsuperscript{16}, from tomato root\textsuperscript{37}, and from several animal models\textsuperscript{38–40}, for chromatin, proteomic and gene expression studies.

Here, we have optimized INTACT, both in regard to protocol set-up and the two-component labelling system, to generate a transcriptome atlas of early Arabidopsis embryo development at cell-type-specific resolution. Focus has been on the cell types essential for root stem cell niche formation. To accomplish this, we have established a wide array of INTACT lines driven by promoters expressed in discrete cell types of the early embryo. By describing and integrating spatial and temporal genome-wide gene activity at the cellular level, our work provides a resource to explore the developmental processes and the genetic networks that shape the first tissues of the plant.

Results
Adaptation of INTACT for early embryo cell types. To adapt the INTACT method for use on early Arabidopsis embryos, a codon-optimized version of BirA was synthesized to replace the original Escherichia coli-derived BirA fragment. In this mBirA, all codons were designed according to the Arabidopsis codon preference to facilitate translation (Supplementary Fig. 1). In addition, either a 3×Myc epitope tag (mBirA–3×Myc) or fluorescent mCherry (mBirA–mCherry) was fused to mBirA to allow determination of expression in transgenic plants.

INTACT is an intrinsic two-component labelling system\textsuperscript{26}. We exploited this capacity to isolate embryo-derived cell-type-specific nuclei without the need to dissect embryos from seeds (Supplementary Fig. 2a). First, mBirA–3×Myc was uniformly expressed in the entire embryo by using the WOX2 promoter\textsuperscript{41} (pWOX2–mBirA–3×Myc; Fig. 1a and Supplementary Fig. 2b–e). A more restricted and cell-specific localization of mBirA to the hypophysis and later to the QC and ground tissue precursors was achieved by expressing mCherry-tagged mBirA with the SCR promoter\textsuperscript{42} (pSCR–mBirA–mCherry; Fig. 1b). Second, a library of INTACT (INT) lines was generated as a resource for cell-type-specific nucleus isolation in the embryo. Thirty-eight promoters were selected on the basis of reported gene expression patterns (Supplementary Table 1), and used to express NTF in a genetic background containing either pWOX2–mBirA–3×Myc or pSCR–mBirA–mCherry. All INT lines were screened to obtain lines that show specific NTF expression during the early stages of embryo development (16-cell, early globular, late globular). Of the lines, 22 show cell-type-specific expression at these stages (Supplementary Table 1 and Supplementary Fig. 3). From this...
collection, we selected 'gold-standard' cell-type-specific INT lines that mark the whole embryo, suspensor, protoderm, hypophysis, and the QC, vascular and ground tissue precursors (Fig. 1c–j and Supplementary Fig. 2f–l). BirA transfers biotin onto the BLRP moiety of NTF43. Thus, we used immunostaining to determine whether NTF is efficiently biotinylated in vivo. Results confirmed local NTF biotinylation by mBirA in the embryo (Fig. 1a). Furthermore, since no visible phenotypes were observed in transgenic lines, neither NTF nor mBirA expression appears to affect early embryo development.

Arabidopsis embryos are small, contain few cells and are enclosed by a seed coat and surrounded by the micropylar endosperm4. This affected final yield and purity (data not shown) when INTACT was performed according to the original protocol, which was established for roots44, a larger and more accessible tissue. Therefore, several steps in the INTACT protocol were optimized (see Methods) to obtain a recovery efficiency of 20–50% with a purity (± s.e.m) of 86.2% ± 6.6% for all nuclei of interest, as based on 50 experiments (Supplementary Table 2). Since embryo development is a progressive process, with several cell specification events happening at specific moments for the first time in plant life, it is important to isolate embryos from defined stages to isolate nuclei from exact cell types. We performed manual pollination of emasculated pWOX2–mBirA–3xMyc flowers to control timing of fertilization, and harvested embryos at varying time intervals. A reasonable degree of temporal homogeneity in embryonic stages within each fruit was found, and we determined 72, 81 and 100 h after pollination as optimal intervals for 16-cell, early globular and late globular stages, respectively (Fig. 2a). This was subsequently verified in each INT line selected for transcriptomic profiling.

INTACT allows efficient isolation of targeted nuclei. Isolation of embryo nuclei through INTACT on whole seeds requires that the overlap of BirA and NTF expression is restricted to embryo cells. Due to chlorophyll autofluorescence in the seed coat and funiculus, we could not fully exclude the possibility of pWOX2 expression in this region, even though no difference in red fluorescence (outside of the embryo) was detected between pWOX2–nDsRed and wild-type seeds (Supplementary Fig. 2d,e). Several INT lines did not show any NTF expression outside the embryo, but some others did (for example, INT55, INT39; Supplementary Fig. 2f,j). Therefore, to ascertain that INTACT can be used on our INT lines to enrich cell-type-specific nuclei exclusive to the embryo, we investigated the expression of a panel of genes predominantly expressed in embryo, endosperm or seed coat/funiculus in nuclei derived from the gold-standard whole-embryo INT0 line (Figs. 1c and 2b and Supplementary Fig. 4 and Supplementary Table 3). In INT3, this line contains pWOX2–mBirA–3xMyc and NTF driven by the RPS5A promoter, which is expressed in all seed tissues45. When compared with total seed (NON-INTACT) nuclei, we found strong enrichment of embryo-expressed genes in the embryo-derived (INTACT) nuclei. In contrast, endosperm-expressed genes were greatly depleted in these nuclei. Seed coat/funiculus-expressed genes were also depleted, but to a lesser degree than that of endosperm-expressed genes. Thus, while contamination with seed coat/funiculus genes is observed, INTACT strongly enriched for embryo-expressed genes and, therefore, potential expression of the WOX2 and NTF promoters outside the embryo is not expected to affect expression profiles.

We next verified stage-specific nuclei isolation by using the INT3 line (Supplementary Fig. 3) to investigate the expression of genes predominantly expressed in the suspensor or ground tissue (Fig. 2c and Supplementary Table 3). In INT3, pOFP8–NTF expression is detected in the early globular suspensor, after which it shifts to the ground tissue in late globular stage. Reflecting the distribution of embryo stages at optimal time points (Fig. 2a), early globular-derived nuclei showed strong enrichment of suspensor-expressed genes but next to no expression of ground tissue genes (Fig. 2c). In contrast, these genes were strongly enriched in late globular-derived nuclei while suspensor genes were expressed at low levels. Together, these results demonstrate the efficient isolation of both cell-type- and stage-specific nuclei.
**A transcriptome atlas of the early Arabidopsis embryo.** With a collection of gold-standard INT lines and an optimized INTACT workflow, we set out to generate a transcriptome atlas of the early Arabidopsis embryo containing cell types important for root stem cell niche formation. INTACT relies on profiling of nuclear RNA, which consists of both nascent transcripts and mature messenger RNA not yet exported to the cytosol. To determine whether nuclear RNA is representative of total cellular mRNA, we compared whole-genome expression data from nuclear (INTACT) and cellular (manual whole-embryo isolation) RNA of the entire INT0 embryo at the 16-cell stage. We found a high correlation coefficient (\( r = 0.82 \)) in the composition of these RNA pools (Supplementary Table 4a), demonstrating that RNA isolated from INTACT-derived nuclei is a reasonable proxy for steady-state transcript levels and can thus be used for gene expression profiling. Interestingly, differences between nuclear and total mRNA do exist at the individual gene level. In addition to contribution of seed tissue contamination (Supplementary Table 4b), this may reflect the regulation, transient state and stability of mRNA (reviewed previously\(^{46,47}\)), as well as selective enrichment associated with cellular functions and nuclear retention\(^{48,49}\).

As a first step in generating the atlas, we isolated between 1,000–5,000 cell-type-specific nuclei from 100–200 manually pollinated siliques for each of 6–8 biological replicates of a selected INT line and embryonic stage (listed in Fig. 3). After RNA extraction, multiple biological replicates were first pooled and then all amplified simultaneously to reduce technical variance between samples and to avoid batch-to-batch effects\(^{50,51}\) (Fig. 4). Transcriptomic profiling was carried out using Affymetrix Arabidopsis Gene 1.1 ST arrays, which cover 84.2% (28,501) of the current total of annotated genes (33,602) (TAIR10)\(^{52}\). Using this approach, a spatio-temporal differential gene expression data set was generated (Supplementary Table 5a,b). As was shown in the quantitative RT-PCR analysis (Fig. 2b), manual and software-assisted\(^{53}\) comparison to previous seed tissue transcriptome data sets\(^{10,54}\) revealed that seed coat and funiculus contamination is present in the data set (Supplementary Table 5c), but given that its level is comparable across all samples, we do not expect this generic contamination to significantly affect the interpretation of differential gene expression between stages or cell types.

| INT line | Cell type | Development stage | Abbreviation | Representation |
|----------|-----------|-------------------|--------------|--------------|
| INT0     | Whole embryo | 16-cell | nEMB_16C     | ![Whole embryo](image1) |
| (pRPS5A–NTF, pWOX2–mBirA–3xMyc) | | Early globular | nEMB_EG | ![Early globular](image2) |
| | | Late globular | nEMB_LG | ![Late globular](image3) |
| INT4     | Vascular tissue precursor | Early globular | nVSC_EG | ![Vascular tissue precursor](image4) |
| (pQDIS–NTF, pWOX2–mBirA–3xMyc) | | Late globular | nVSC_LG | ![Late globular](image5) |
| INT7b    | Quiescent centre precursor | Late globular | nQC_LG | ![Quiescent centre precursor](image6) |
| (pWOX5–NTF, pSCR–mBirA–mCherry) | | | | |
| INT17    | Suspensor | Early globular | nSUS_EG | ![Suspensor](image7) |
| (pKMD2–NTF, pWOX2–mBirA–3xMyc) | | Late globular | nSUS_LG | ![Late globular](image8) |
| INT39    | Ground tissue precursor | Early globular | nGSC_EG | ![Ground tissue precursor](image9) |
| (pMGP–NTF, pWOX2–mBirA–3xMyc) | | Late globular | nGSC_LG | ![Late globular](image10) |
| INT55    | Inner lower tier | 16–cell | nILT_16C | ![Inner lower tier](image11) |
| (pSHR–NTF, pWOX2–mBirA–3xMyc) | | | | |

---

**Fig. 3 | Abbreviations and representations of selected INTACT (INT) lines and developmental stages used for transcriptomic profiling.**
To validate the cell-specific profiles, we first surveyed genes with well-known expression patterns and found that a majority of the genes showed enrichment above fold change (FC) 1.5 in the expected cell type when compared with the whole embryo (Fig. 5a and Supplementary Table 3). On the basis of this result, we then investigated the expression pattern of genes for which no independent expression pattern had been reported. We selected genes that were increasingly enriched (FC > 1.5, regardless of q value) in nVSC (vascular tissue precursor nuclei) and nQC (QC precursor nuclei) compared with late globular nEMB (whole embryo nuclei), or in early compared with late globular nVSC (Fig. 5b–d and Supplementary Fig. 5). RNA in situ results were largely in agreement with INTACT-based predictions. Most (≥78%) of the 22 spatially enriched genes were detected in, and concentrated to, the expected cell type (Fig. 5b,c and Supplementary Fig. 5a,b). No expression could be detected for the remaining genes. Likewise, both temporally enriched genes disappeared from the vascular tissue precursors after the early globular stage (Fig. 5d and Supplementary Fig. 5c). Promoter–reporter fusion analysis of a selected subset of these genes supported these findings. No clear correlation was observed between expression patterns and FC of expression enrichment, nor with the statistical q values. Several genes were more broadly expressed than only in the predicted cell type, which is a logical consequence of our (spatial) selection strategy in which enrichment in a single cell type relative to the entire embryo was considered. Such enrichment does not exclude expression in other cell types. This was especially evident for the nVSC selected genes because they, in general, were also enriched in other cell types (Fig. 5b and Supplementary Fig. 5a). Moreover, in situ hybridization mostly detects cytosolic mRNA, while INTACT captures nuclear RNA, and some investigated genes showed divergent expression in the nuclear versus whole cellular comparison (Supplementary Table 4c). Nevertheless, cell-specific enrichment is consistent with observed expression patterns, validating the applicability of the atlas.

Developmental stage has a dominant impact on transcriptome dynamics. Since the atlas provides spatial and temporal distinct expression values for individual genes, we next estimated overall cell-type- and stage-specific diversity by comparing mRNA populations using principal component analysis (PCA) on the top 2,000 variable genes (Fig. 6a and Supplementary Fig. 6), and pairwise expression profile correlation on mean replicate values of the entire populations (Fig. 6b and Supplementary Table 5d). The cell-specific samples, excluding late globular nGSC (ground tissue precursor nuclei), formed stage-specific subgroups that clustered separately from (stage-separated) nEMB. Surprisingly, late globular nGSC associated closer with nEMB of the same stage than with any other samples. A closer study of the ground tissue line (INT39; Fig. 1j) revealed weak NTF expression also in neighbouring tissues (vascular, QC, protoderm) at this stage. This probably resulted in broader nuclei isolation and, thus, could explain the similarity between these populations. Moreover, the low overall variability seen in the PCA analysis (Fig. 6a) is a probable consequence of the pooling of RNA samples and the simultaneous RNA amplification. Nevertheless, all replicates except for nQC clustered according to cell type, although in some cases (such as nSUS (suspensor nuclei) and nILT (inner lower tier nuclei)) not very distinctly. For nQC, the result reflects the low correlation observed between the nQC replicates (Supplementary Table 5a). Hierarchical clustering confirmed the PCA results (Supplementary Fig. 7).

The analyses demonstrate that INTACT-derived transcriptomes are reproducible, as most biological replicates cluster together. Secondly, transcriptomes are different both between cell types, and between stages. However, embryo stage has a more profound impact on transcriptome differences than cell type does, as is particularly evident from pairwise correlation analysis (Fig. 6; white outline). This suggests that developmental stage greatly influence cell state.

Transcriptional reprogramming during embryogenesis. To explore how the transcriptomes of the atlas define cell states and state transitions during development, we investigated gene ontology (GO) terms and transcription factor (TF) families in dominant spatial and temporal gene expression patterns of unique and shared genes (Supplementary Fig. 8). Here, ‘spatial’ refers to comparison with all other cell types of the same stage and ‘temporal’ refers to comparison with the same cell type of a different stage. Of note, GO-term enrichment analysis revealed that terms such as ‘cell wall biogenesis’, ‘cell growth’, ‘epidermal cell differentiation’ and ‘auxin polar transport’ were spatially enriched in early globular nSUS (Fig. 7a and Supplementary Table 5e), but not or to a lesser degree in late globular stage. This developmental shift was also seen in the temporal analysis.
Since the vascular and ground tissue precursors are established following the periclinal division of the four inner lower tier cells during the 16-cell to early globular stage transition, the lack of developmental progression in nVSC suggests that vascular identity is already established in these cells at the 16-cell stage, while ground tissue cell fate is acquired through transcriptional reprogramming after their division. Supporting this, the vascular markers SHR and ZLL are both expressed in the inner lower tier cells at the 16-cell stage (INT55 (SHR)), INT56 (ZLL); Fig. 1h and Supplementary Fig. 3).
Resource

During the transition from the early to the late globular stage, asymmetric and anticlinal division of the hypophysis and the vascular and ground tissue precursors generates small apical cells that differentiate into the QC and the vascular and ground tissues, respectively. Results from the GO-term enrichment analysis reflect this development. Terms associated with cell division, differentiation and specification were spatially and/or temporally enriched in late globular nVSC and nGSC (Fig. 7). As previously noted, transcripts from the vascular tissue precursors are probably also present in the late globular nGSC samples. For late globular nQC, terms associated with cell morphology and its involvement in differentiation were enriched (Fig. 7a).

These results suggest that, as a reflection of the shifts in cell-specific gene expression (Fig. 6), developmental processes change in a corresponding manner to define cell fates during root stem cell niche formation in the early embryo. This conclusion is supported by TF family enrichment analysis that showed highly dynamic expression patterns of TF families crucial for early embryo patterning, including auxin response factor (ARF), basic-helix-loop-helix (bHLH) and Homeobox (HOX) (Supplementary Fig. 9, Table 5e,f).

**Discussion**

In this study, we adapted, optimized and validated a two-component transgenic labelling system (INTACT) to efficiently isolate cell-type-specific nuclei from the early seed-enclosed Arabidopsis embryo. With this approach, a transcriptome atlas was generated as a resource to explore how genome-wide gene activity shapes fates at the cellular level during the formation of the root stem cell niche.

We identified major shifts in unique cell-type-specific gene sets in the early embryo that, while associated with clonal origin, was more dependent on the developmental stage of the embryo. Dominant expression patterns, probably signifying transcriptional reprogramming, revealed enrichment of transcription factor families (ARF, bHLH, HOX) and biological processes important for cell fate determination that correspond to known activities of the cell types they encompass and reflect developmental transitions. Indeed, our results support earlier data showing that the pre- but not post-globular suspensor is necessary for development, and reveal that vascular identity is probably established prior to ground tissue cell fate specification. A substantial coordination between dynamic gene activity, reprogramming and biological processes in defining developmental transitions has earlier been shown in seed tissues (including the embryo) and during stomatal lineage progression. This, together with the expression of known tissue markers, provides support for previously published results suggesting that cell types that will form the root apical stem cell niche acquire their fates already at this early stage of development, and that this is, at least partially, mediated by auxin signalling through ARFs and their downstream targets, including bHLHs and WOX5. From this, it is tempting to speculate that self-regulating auxin response provides the spatio-temporal cues that initiate the observed shifts in cell-type-specific gene expression. Future investigations using our atlas together with other published data sets may help to dissect the source and functional outcome of these shifting patterns.

To explore the atlas, it is important to have confidence that the microarray results represent true in vivo gene expression differences. By surveying expression profiles of well-known genes and using in vivo expression analysis of selected genes, we could show that atlas predictions correlate with observed expression patterns, even for nQC where, presumably, small cell numbers caused high variation between replicates. Seed tissue contamination is common in the atlas; however, this is expected since the INTACT method, as applied here, isolates between 10–20% of non-biotinylated and, thus, total seed nuclei. Nevertheless, this is not expected to greatly affect overall expression profile comparisons since contamination levels are similar across all samples. Residual values from the nuclear versus whole cellular comparison can be consulted to identify putative contamination or other nuclear/cellular-derived RNA differences that may affect predictions for specific genes. When compared, our atlas performs favourably to and, by providing more pattern information, supplements already available cellular and nuclear transcriptomes (Supplementary Fig. 10 and Supplementary Table 5g), even though the data sets are different (Supplementary Fig. 11). It also appears to contain similar or even lower levels of seed tissue contamination compared with the FANS-generated nuclear transcriptomes of comparable cell types and developmental stages (Supplementary Table 6). It should be noted, that unlike the other data sets, the atlas contains both pre-processed and mature mRNA as a consequence of the RNA amplification kit used. Furthermore, the higher differential gene expression observed for the cellular transcriptome suggests that small initial differences in transcription level between cells are amplified by post-transcriptional processes.

Much remains to be learned about cell fate in the early embryo. We anticipate that this atlas, in providing the most comprehensive description of plant embryonic tissue- and cell-specific gene activity to date, will be a valuable resource, not only to investigate patterning events involved in cell fate specification and differentiation during root apical
stem cell formation, but also together with other genetic resources to understand the regulatory mechanisms underlying plant development. For this purpose, an interactive graphical representation of the atlas has been established at http://www.albertodb.org (ALBERTO).

Methods

Plant material. The plants used in all experiments were Columbia (Col-0) ecotype. The pWOX2-nDsRed reporter line was kindly provided by T. Laux and has been described previously41.

After seed sterilization, transgene selection was performed on solid Murashige and Skoog medium supplemented with 50 mg l\(^{-1}\) kanamycin or 15 mg l\(^{-1}\) phosphinothricin. Resistant seedlings were transferred to soil and further grown at a constant temperature of 22 °C under long-day conditions (16 h light/8 h dark). Floral dip was used for plant transformations62.

Generation of transgenic constructs. All transgenic constructs were generated using the LIC cloning system and vectors described previously63,64. To generate promoter fusions for nuclei purification, a nuclear targeting fusion protein (NTF) was cloned into pPLV1 (LIC–NTF). The E. coli biotin ligase (mBirA) was synthesized with codon usage according to Arabidopsis preference (Mr. Gene; codon choice according to the manufacturer’s specifications) and cloned into pPLV2 tagged with either 3\(\times\)Myc (LIC–mBirA–3\(\times\)Myc) or mCherry (LIC–mBirA–mCherry). The original NTF and BirA fragments were kindly provided by R. B. Deal and have been previously described26. For promoter–NTF, pWOX2–mBirA–3\(\times\)Myc and pSCR–BirA–mCherry constructs, promoter fragments corresponding to a region approximately three kilobase pairs upstream from the start codon of the genes were cloned into the corresponding LIC vectors. To generate promoter–n3GFP reporter constructs, promoter fragments were cloned into pPLV4. Amplification primers of each promoter fragment are listed in Supplementary Table 7.

Western blotting and immunostaining. Protein was extracted by grinding siliques in liquid nitrogen, adding 200 μl lysis buffer (50 mM Tris–HCl (pH 7.4); 150 mM NaCl; 1% Triton X-100; 1% (v/v) protease inhibitor cocktail (Roche Diagnostics GmbH) and then centrifugation (18,000 g). The supernatant was incubated with 50 μl 1× sample buffer (62.5 mM Tris–HCl (pH 6.8); 10% (v/v) glycerol, 2% (w/v) SDS, 1% (v/v) β-mercaptoethanol, 0.0005% (w/v) bromophenol blue) for 5 min

Fig. 7 | Spatio-temporal shifts in developmental processes define early embryonic cell fate. a, b, Heat maps show enriched biological gene ontology (GO) terms in dominant spatial (a) and temporal (b) gene expression patterns of the atlas. High P value is in white, and low P value is in red. Enrichment patterns generally reflect cell state and state transitions in the early embryo. Representations are described in Fig. 3.
ISOlation of Embryos. Embryo isolation was performed according to a method described earlier57, using seeds collected and inoculated, as described above, from ~60 siliques after manual pollination. A Zeiss Confocor 1 inverted microscope (Zeiss) was used for the visualization of the embryos. The isolated embryos were immediately incubated with 500 μl of TRizol reagent (Ambion) for 5 min at room temperature (vortexing) and 30 min at 60 °C, respectively, and then purified according to the TRizol reagent protocol for small sample quantities. Glycoblue (2 μl of 15 mg ml−1; Ambion) was used as a co-precipitant. The RNA was treated with DNase I (~14 kunitz units; QIAGEN GmbH) for 5 min at room temperature and then purified (RNeasy Micro Kit; QIAGEN GmbH) and concentrated to 5 μl (Concentrator plus; Eppendorf AG) before being stored at ~80 °C until further processing.

Microarray analysis. RNA was amplified using the Ovation Pico WTA System V2 (NuGEN), labelled with the ENCORE Biotin Module (NuGEN) and hybridized to Arabidopsis Gene 1.1 ST 24-Array plates (atlas) or single-array strips (nuclear versus cellular RNA) (Affymetrix) according to the manufacturers protocol. Microarray analysis was performed using the MADMAX pipeline and a custom CDF file (CustomCDF version 19.0.0)76. Here, all expression values were (quantile) normalized by the Robust multi-array average algorithm (RMA)76 using the median polish (nuclear versus cellular RNA) or M-estimator (atlas) algorithm for probes to probe set summarization. Probe sets were redefined using current genome information and reorganized according to TAIR10 gene definitions. Linear models and an intensity-based moderated t statistic approach70,71 were used to identify differentially expressed genes (probe sets). P values were corrected for multiple testing using an optimized false discovery rate (FDR) approach72.

For software-assisted76 comparison of the atlas data set to previously published seed tissue transcriptomic data32, a minimum fold of 2 and P value of 0.05 was considered ‘enriched’ and temporal changes were not buffered. Genes not considered ‘expressed’ (mean expression value <9.85; see Supplementary Fig. 12) in each tested sample were removed before analysis. Dominant expression patterns were determined after first filtering the data set by removing non-expressed genes in any of the compared samples, and selecting the top 50% (4,796 to 6,368) varying genes across the data set. Then, either manual selection (2 data sets; fold change >1.5) or fuzzy K-means and hierarchical clustering (>2 data sets)32, with K = 1.105 and a cluster membership probability (m values) at or above 0.4, was performed. Biological significance of co-expressed genes was explored using agrigo32, where significance of gene ontology (GO) term enrichment was relative to TAIR locus IDs covered by the Gene 1.1 ST array. P values from Fisher distributions were adjusted via the Benjamini–Yekutieli method to assess the FDR. GO terms with P < 0.05 or FDR < 0.05 were significantly enriched. Transcription factors (TFs) belonging to a family were identified using the Arabidopsis TF database at AGRIS76.

RNA in situ hybridization. In situ hybridization was performed as described earlier22 and successfully repeated at least two times for each gene. Target gene primers are listed in Supplementary Table 7.

Microscopy. Differential interference contrast microscopy, fluorescence microscopy and confocal microscopy were performed as previously described46 with the following modifications: confocal imaging was performed on a Leica SP5-II system (Hyd detector) and 0.1% (v/v) SR2200 (Renaissance Chemicals) was used in the mounting solution (0.4% (v/v) dimethylsulfoxide, 5% (v/v) glycerol, 4% (v/v) paraformaldehyde in PBS (pH 8.0)) for counterstaining of transgenic plant embryos carrying a reporter construct.

Quantitative RT-PCR analysis. Poly(DT) cDNA, prepared with an iScript cDNA Synthesis Kit (Bio-Rad Laboratories) from nuclear RNA, was analysed on a Bio-Rad CFX384 Real-Time PCR detection system with iQ SYBR Green Supermix (Bio-Rad Laboratories) according to the manufacturer’s instructions. Targets were quantified with primers designed with Beacon Designer 7.0 (PREMIER Biosoft) and listed in Supplementary Table 7. All individual reactions were done in triplicate. Data were analysed with qBase+. Expression levels were normalized to those of the reference genes ELONGATION FACTOR 1-ALPHA (EFA1) and GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C SUBUNIT (GAPC).
4. Mansfield, S. G. & Briarty, L. G. Early embryogenesis in Arabidopsis thaliana. II. The developing embryo. Can. J. Bot. 69, 461–476 (1991).
5. Jurgens, G. & Mayer, U. In A Colour Atlas of Developing Embryos (ed. Bard, J. B. L.) 7–21 (Wolfe Publishing, London, 1994).
6. Yoshida, S. et al. Genetic control of plant development by overriding a geometric division rule. Dev. Cell 29, 75–87 (2014).
7. Gooh, K. et al. Live-cell imaging and optical manipulation of Arabidopsis early embryogenesis. Dev. Cell 34, 242–251 (2015).
8. Palovaara, J., de Zeeuw, T. & Weijers, D. Tissue and organ initiation in the plant embryo: a first try for everything. Annu. Rev. Cell Dev. Biol. 32, 47–75 (2016).
9. Kerk, N. M., Ceserani, T., Tausta, S. L., Sussex, I. M. & Nelson, T. M. Laser capture microdissection of cells from plant tissues. Plant Physiol 132, 27–35 (2003).
10. Belmonte, M. E. et al. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc. Natl Acad. Sci. USA 110, E435–E444 (2013).
11. de Vega-Bartol, J. J. et al. Transcriptomic analysis highlights epigenetic and transcriptional regulation during zygotic embryo development of Pinus pinaster. BMC Plant Biol. 13, 123 (2013).
12. Venglat, P. et al. Gene expression profiles during embryo development in Brassica napus. Plant Breeding 132, 514–522 (2013).
13. Chen, J. et al. Dynamic transcriptome landscape of maize embryo and endosperm development. Plant Physiol. 166, 252–264 (2014).
14. Itoh, J.-I. et al. Genome-wide analysis of spatiotemporal gene expression patterns during early embryogenesis in rice. Development 143, 1217–1227 (2016).
15. Casson, S., Spencer, M., Walker, K. & Lindsay, K. Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. Plant J. 42, 111–123 (2005).
16. Spencer, M. W. B., Casson, S. A. & Lindsay, K. Transcriptional profiling of the Arabidopsis embryo. Plant Physiol. 143, 924–940 (2007).
17. Le, B. H. et al. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proc. Natl Acad. Sci. USA 107, 8063–8070 (2010).
18. Xiang, D. et al. Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in Arabidopsis. Plant Physiol. 156, 346–356 (2011).
19. Autran, D. et al. Maternal epigenetic pathways control parental contributions to Arabidopsis early embryogenesis. Cell 145, 707–719 (2011).
20. Nodine, M. D. & Bartel, D. P. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature 482, 94–97 (2012).
21. Nodine, M. D. & Bartel, D. P. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev. 24, 2678–2692 (2010).
22. Birnbaum, K. et al. A gene expression map of the Arabidopsis root. Science 302, 1956–1960 (2003).
23. Zhang, C., Barthelson, R. A., Lambert, G. M. & Galbraith, D. W. Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. Plant Physiol. 147, 30–40 (2008).
24. Zanetti, M. E., Chang, I.-F., Gong, F., Galbraith, D. W. & Bailey-Serres, J. Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot. Development 127, 595–603 (2000).
25. Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoholoenzyme synthetase-catalyzed biotinylation. Protein Sci. 8, 921–929 (1999).
26. Deal, R. B. & Henikoff, S. The intact method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat. Protoc. 6, 56–68 (2011).
27. Weijers, D. et al. An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. Development 138, 4289–4299 (2011).
28. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 136, 688–700 (2009).
29. Hocine, S., Singer, R. H. & Grünwald, D. RNA processing and export. Cold Spring Harb. Perspect. Biol. 2, a00752 (2010).
30. Barthelson, R. A., Lambert, G. M., Vainier, C., Lynch, R. M. & Galbraith, D. W. Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. BMC Genom. 8, 340 (2007).
31. Grindberg, R. V. et al. RNA-sequencing from single nuclei. Proc. Natl Acad. Sci. USA 110, 19802–19807 (2013).
32. Clément-Ziza, M. et al. Evaluation of methods for amplification of picogram amounts of total RNA for whole genome expression profiling. BMC Genom. 10, 246 (2009).
33. Morse, A. M., Carballo, V., Baldwin, D. A., Taylor, C. G. & McIntyre, L. M. Comparison between NuGEN’s WT-OVATION PICO and One-Direct Amplification Systems. J. Biomol. Tech. 21, 141–147 (2010).
34. Lamesch, P. et al. Using the Arabidopsis information resource (TAIR) to find information about Arabidopsis genes. Curr. Protoc. Bioinform. 30, 1.11.1–1.11.51 (2002).
35. Schon, M. A. & Nodine, M. D. Widespread contamination of Arabidopsis embryo and endosperm transcriptome data sets. Plant Cell 29, 608–617 (2017).
36. Khan, D. et al. Transcriptome atlas of the Arabidopsis thaliana funiculus – a study of maternal seed subregions. Plant J. 82, 41–53 (2015).
37. Liu, Y. et al. Direct evidence that suspensor cells have embryogenic potential that is suppressed by the embryo proper during normal embryogenesis. Proc. Natl Acad. Sci. USA 112, 12432–12437 (2015).
38. Radoeva, T., ten Hove, C. A., Saiga, S. & Weijers, D. Molecular characterization of Arabidopsis GAI4/UAS enhancer trap lines identifies novel cell type-specific promoters. Plant Physiol. 171, 1169–1181 (2016).
39. Sarkar, A. K. et al. Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. Nature 446, 811–814 (2007).
40. Schereth, A. et al. MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature 464, 913–916 (2010).
41. De Rybel, B. et al. A BHLH complex controls embryonic vascular tissue establishment and indeterminate growth in Arabidopsis. Dev. Cell 24, 426–437 (2013).
42. De Rybel, B. et al. Integration of growth and patterning during vascular tissue formation in Arabidopsis. Science 345, 1255215 (2014).
43. Crawford, B. C. W. et al. Genetisk control of distal stem cell fate within root and embryonic meristems. Science 347, 655–659 (2015).
44. Clough, S. J. & Bent, A. F. Floral dip: A simplified method for agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16, 735–743 (1998).
63. De Rybel, B. et al. A versatile set of ligation-independent cloning vectors for functional studies in plants. *Plant Physiol.* **156**, 1292–1299 (2011).
64. Wendrich, J. R., Liao, C. Y., van den Berg, W. A., De Rybel, B. & Weijers, D. Ligation-independent cloning for plant research. *Methods Mol. Biol.* **1284**, 421–431 (2015).
65. Llavata-Peris, C., Lokerse, A., Möller, B., De Rybel, B. & Weijers, D. in *Plant Organogenesis: Methods and Protocols* (ed. De Smet, I.) 137–148 (Humana Press, Totowa, New Jersey, 2013).
66. Raissig, M. T., Gagliardini, V., Jaenisch, J., Grossniklaus, U. & Baroux, C. Efficient and rapid isolation of early-stage embryos from *Arabidopsis thaliana* seeds. *J. Vis. Exp.* **76**, e50371 (2013).
67. Lin, K. et al. MADMAX – Management and analysis database for multiple –omics experiments. *J Integr. Bioinform.* **8**, 160 (2011).
68. Dai, M. et al. Evolving gene/transcript definitions significantly alter the interpretation of genechip data. *Nucleic Acids Res.* **33**, e175 (2005).
69. Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
70. Sartor, M. A. et al. Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments. *BMC Bioinformatics* **7**, 1–17 (2006).
71. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. App. Stat.* **10**, 946–963 (2016).
72. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
73. Orlando, D. A., Brady, S. M., Koch, J. D., Dinneny, J. R., Benfey, P. N. in *Plant Systems Biology* (ed. Belostotsky, A. D.) 57–77 (Humana Press, Totowa, New Jersey, 2009).
74. Palaniswamy, S. K. et al. AGRIS and AtRegNet: a platform to link cis-regulatory elements and transcription factors into regulatory networks. *Plant Physiol.* **140**, 818–829 (2006).
75. Saiga, S. et al. The *Arabidopsis* OBERON1 and OBERON2 genes encode plant homeodomain finger proteins and are required for apical meristem maintenance. *Development* **135**, 1751–1759 (2008).
76. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, 1–14 (2007).

**Acknowledgements**
The authors thank T. Laux, R. Deal and S. Henikoff for sharing materials. Further, the authors thank the students M. Geerlings, D. van der Plaat, E. Asamoah Gyimah, M. Goranova, A. Kuhn, L. van der Bent, I. Papakosta, S. Thomaidou, S. Hedžet and K. Heidemeyer for their contributions. This study was supported by the Federation of European Biochemical Societies (FEBS) (J.P.), ERA-CAPS project EURO-PEC (grant number 849.13.006) and the European Research Council (ERC Starting Grant ‘CELLPATTERN’; contract number 281573 to D.W.).

**Author contributions**
D.W. conceived the study; J.P., S.S., J.R.W. and D.W. designed research; S.S. generated INTACT lines; J.P. and N.v.W.H. optimized and adapted experimental set-up; J.P., S.S. and J.R.W. performed INTACT; F.H. and J.P. performed nuclear versus cellular RNA comparison; M.B., G.J.H. and J.P. performed transcriptomic profiling and analysed data with support from S.M.; J.P., S.S., J.R.W., J.P.v.S. and J.S. validated expression patterns; J.P.v.S. and J.P. designed and developed the AlBERTO browser; J.P. and D.W. wrote the paper with input from all other authors.

**Competing interests**
The authors declare no competing financial interests.

**Additional information**
Supplementary information is available for this paper at https://doi.org/10.1038/s41477-017-0035-3.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.W.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

### Experimental design

1. **Sample size**
   - Describe how sample size was determined.
   - Sample sizes were chosen to be sufficient for the statistical analyses performed.

2. **Data exclusions**
   - Describe any data exclusions.
   - In total, 4 replicates were excluded from the microarray analysis due to poor quality control results. This exclusion criteria was pre-established.

3. **Replication**
   - Describe whether the experimental findings were reliably reproduced.
   - All attempts at replication were successful.

4. **Randomization**
   - Describe how samples/organisms/participants were allocated into experimental groups.
   - This is not relevant to our study. Samples were allocated into groups based on the cell type and developmental stage from which the nuclei were isolated. For each cell type and biological replicate, new plants spread randomly over several trays were used to isolate the nuclei.

5. **Blinding**
   - Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   - As group allocations was already defined prior to all analyses, blinding was not relevant for our study.

6. **Statistical parameters**
   - For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   - **n/a** Confirmed
     - The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
     - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
     - A statement indicating how many times each experiment was replicated
     - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
     - A description of any assumptions or corrections, such as an adjustment for multiple comparisons
     - The test results (e.g. $P$ values) given as exact values whenever possible and with confidence intervals noted
     - A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
     - Clearly defined error bars

*See the web collection on statistics for biologists for further resources and guidance.*
Software

Describe the software used to analyze the data in this study.

Primers for quantitative RT-PCR (qPCR) analysis were designed using Beacon Designer 7.0 (PREMIER Biosoft, CA, USA). QPCR data were analyzed with qBase (Hellermans et al. 2007). Microarray analysis was performed using the MADMAX pipeline (Lin et al. 2011) using a custom CDF files (MBNI CustomCDF version 19.0.0) (Dai et al. 2005). For software-assisted contamination determination, we used the R program code detailed in Schon and Nodine (2017). Any additional information regarding software is detailed in the Material and Methods section of the paper or in the Supplementary material.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All unique materials used in the study is readily available from the authors.

Four antibodies were used. Information for each antibody in regards to target, vendor, catalogue number, references and validation statements can be accesses using the Antibody ID (provided in the paper) in the Resource Identification Portal (https://scicrunch.org/resources). The four antibodies are:

- Anti-Myc Tag antibody (clone 9E10) (RRID:AB_309725) (Merck Millipore, CA, USA)
- Rabbit Anti-Mouse HRP-conjugated antibody (RRID:AB_92531 (Merck Millipore)
- Anti-Mouse Alexa Fluor 588 (RRID:AB_141514) (Thermo Fisher Scientific, OR, USA)
- Strepavidin conjugated with Alexa Fluor 647 antibody (RRID:AB_2336066) (Thermo Fisher Scientific)

No eukaryotic cell lines were used.

No eukaryotic cell lines were used.

No eukaryotic cell lines were used.

No commonly misidentified cell lines were used.

The study did not involve human research participants.