© 2014. Published by The Company of Biologists Ltd | Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

RESEARCH ARTICLE

TECHNIQUES AND RESOURCES

Cell type-specific transcriptome analysis in the early *Arabidopsis thaliana* embryo

Daniel Slane[1,*], Jixiang Kong[1,2,*], Kenneth W. Berendzen[3], Joachim Kilian[3], Agnes Henschen[1], Martina Kolb[1], Markus Schmid[4], Klaus Harter[3], Ulrike Mayer[5], Ive De Smet[6,7,8], Martin Bayer[1] and Gerd Jürgens[1,2,‡]

ABSTRACT

In multicellular organisms, cellular differences in gene activity are a prerequisite for differentiation and establishment of cell types. In order to study transcriptome profiles, specific cell types have to be isolated from a given tissue or even the whole organism. However, whole-transcriptome analysis of early embryos in flowering plants has been hampered by their size and inaccessibility. Here, we describe the purification of nuclear RNA from early stage *Arabidopsis thaliana* embryos using fluorescence-activated nuclear sorting (FANS) to generate expression profiles of early stages of the whole embryo, the proembryo and the suspensor. We validated our datasets of differentially expressed candidate genes by promoter-reporter gene fusions and *in situ* hybridization. Our study revealed that different classes of genes with respect to biological processes and molecular functions are preferentially expressed either in the proembryo or in the suspensor. This method can be used especially for tissues with a limited cell population and inaccessible tissue types. Furthermore, we provide a valuable resource for research on *Arabidopsis* early embryogenesis.

KEY WORDS: Fluorescence-activated nuclear sorting, Proembryo, Suspensor, Transcriptome analysis

INTRODUCTION

Multicellular organisms are made up of various cell and tissue types consisting of differentiated cells that all derive from pluripotent, undifferentiated progenitor cells. As these cell and tissue types fulfill a plethora of different functions during the life cycle, progenitor cells have to undergo coordinated changes in spatial and temporal gene expression programs during differentiation. Comprehensive characterization of transcriptional profiles is therefore of great importance to understand the establishment and maintenance of specific cell types. In the case of embryogenesis in flowering plants with the embryos often being deeply embedded in the maternal seed tissue, however, the isolation of cells from specific cell types is already a very challenging task. In general, several existing methods have been employed to overcome such difficulties for different tissues and organisms, such as laser capture microdissection (LCM), fluorescence-activated cell sorting (FACS), translating ribosome affinity purification (TRAP) and isolation of nuclei tagged in specific cell types (INTACT) (Bonner et al., 1972; Emmert-Buck et al., 1996; Heiman et al., 2008; Deal and Henikoff, 2010). At present, TRAP and INTACT are still under optimization in order to be widely used for special tissues such as those in plant embryos (Palovaara et al., 2013). LCM has been used in different studies to isolate tissues from sectioned material without the need to generate transgenic plants (Kerk et al., 2003). Recently, parts of different tissues inside the *Arabidopsis thaliana* seed, including the embryo were isolated by LCM and the different expression profiles were analyzed (Spencer et al., 2007; Le et al., 2010). Nonetheless, LCM requires high precision during tissue excision in order to avoid contamination from adjoining cells. Additionally, as the used material originates from tissue sections, only parts of the cell can be effectively collected. Consequently, precise isolation of certain cell types, such as shoot apical meristem cells, which are deeply embedded within the embryo, is a considerable challenge. Evidently, FACS in combination with gene expression analysis has been broadly employed for many studies, such as purification of *Drosophila melanogaster* embryonic cell populations (Cumberledge and Krasnow, 1994; Shigenobu et al., 2006), clinical applications (Jayasinghe et al., 2006; Jaye et al., 2012) and isolation of different cell types in *Arabidopsis* root and shoot tissue (Birnbaum et al., 2003; De Smet et al., 2008; Yadav et al., 2014). Most of the FACS studies in plants were based on the generation of protoplasts from easily accessible tissues and therefore this method is very difficult to apply to *Arabidopsis* embryos, in particular in large amounts. By contrast, fluorescently labeled nuclei from the companion cells of phloem root tissue were isolated by fluorescence-activated nuclear sorting (FANS) for further transcriptome analysis (Zhang et al., 2008). Importantly, reports showed that the diversity of nuclear and total cellular RNA are comparable overall (Barthelson et al., 2007; Zhang et al., 2008).

In light of specific advantages and disadvantages of the different techniques mentioned above, we combined fluorescent-activated sorting of nuclei (FANS) with linear RNA amplification and microarray analysis to characterize the transcriptomes of two cell types – the proembryo (PE) and suspensor (SUS) – in the early *Arabidopsis* embryo originating from a single cell – the zygote – as well as the whole embryo (EMB). Our strategy was to label nuclei with nuclear-localized GFP (nGFP) driven by cell-type specific promoters only active either in the cells of the proembryo or the suspensor, or uniformly active in the whole embryo. GFP-positive nuclei were sorted by flow cytometry and afterwards standard ATH1 microarray chips were used for transcriptome analysis. Our analysis demonstrated that specific transcripts are differentially expressed

[1]Department of Cell Biology, Max Planck Institute for Developmental Biology, Tübingen 72076, Germany. [2]Department of Developmental Genetics, Center for Plant Molecular Biology, University of Tübingen, Tübingen 72076, Germany. [3]Department of Plant Physiology, Center for Plant Molecular Biology, University of Tübingen, Tübingen 72076, Germany. [4]Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen 72076, Germany. [5]Microscopy facility, Center for Plant Molecular Biology, University of Tübingen, Tübingen 72076, Germany. [6]Department of Plant Systems Biology, VIB, Technologiepark 927, Ghent B-9052, Belgium. [7]Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, Ghent B-9052, Belgium. [8]Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK.
*These authors contributed equally to this work

‡Author for correspondence (gerd.juergens@zmbp.uni-tuebingen.de)

Received 8 August 2014; Accepted 12 October 2014

Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

between the proembryo and suspensor at early stages of embryogenesis, including genes that were previously reported to be differentially expressed *in vivo* (Lau et al., 2012). The datasets were further validated by promoter-reporter fusion analysis and *in situ* hybridization for a subset of genes that were preferentially expressed in proembryo or suspensor cell types. Additionally, we also compared our nuclear whole-embryo transcriptional profile with that of manually isolated, early-stage whole embryos, as well as with publicly available data. In summary, we have developed a robust method that allows the generation of comprehensive expression profiles of specific cell types in *Arabidopsis* early embryos. In particular, this method can be widely used for characterizing gene expression of deeply embedded cell types with a limited number of cells. In addition, we provide a comprehensive resource for the earliest stages and tissues of *Arabidopsis* development.

RESULTS

Identification of embryo-specific marker lines

In order to obtain marker lines that show specific expression during the early stages of *Arabidopsis* embryogenesis in the proembryo, suspensor or whole embryo, we first screened the GAL4-GFP enhancer-trap collection from the Haseloff lab (Haseloff, 1998). Tracing back expression from microscopic analysis of seedling roots, one of the Haseloff lines (N9322) showed specific suspensor expression and the insertion locus was identified by TAIL-PCR to position 610 bp upstream of the *AT5G42203*-coding sequence (supplementary material Fig. S1). We then cloned about 2 kb of the upstream region, including 5′ untranslated region (5′ UTR) sequences for both the neighboring *AT5G42200* and *AT5G42203* genes fused to n3×GFP in order to check whether one or the other of the two promoters could recapitulate the expression pattern of the enhancer trap line. Regarding the expression pattern of the different transgenic lines, the promoter containing the upstream region of the *AT5G42200* gene showed specific expression only in the suspensor from the embryonic two-cell stage onwards (Fig. 1A).

Second, according to published data, the *DORNROESCHEN* (*DRN*) gene (*AT1G12980*) was shown to be expressed exclusively in the proembryo until the early globular stage (Chandler et al., 2007; Cole et al., 2009). Therefore, we cloned the upstream region of *DRN* together with its 3′UTR as described previously (Chandler et al., 2007). Indeed, the expression pattern for this construct in transgenic embryos fit the published data for *DRN* (Fig. 1B). Finally, as a whole embryo marker, we used a marker line available in the lab that drives GFP expression from the upstream region of the *AT3G10010* locus (Fig. 1C).

FANS analysis and microarray results

The individual fluorescent marker lines showing specific expression in proembryo, suspensor or whole-embryo nuclei were subsequently used to generate cell type-specific nuclear transcription profiles of the early *Arabidopsis* embryo. As we were not able to recover protoplasts from early embryonic stages due to the embryonic cell wall and cuticle being recalcitrant to enzymatic digestion, we developed a workflow that enabled us to extract nuclei efficiently from ovule tissue. For nuclear extraction, we isolated ovules from self-pollinated young siliques. The embryos from those ovules ranged from 1- to 16-cell embryonic stages that we checked microscopically from a number of siliques of the plants used prior to the start of the workflow. Afterwards, we fixed the ovules with 0.1% paraformaldehyde in order to maintain nuclear integrity. Additionally, by fixing the cellular contents we made sure that the transcriptional status of the nuclei did not change during the subsequent extraction and separation steps. After nuclear extraction, ~1000 GFP-positive nuclei from ovules of about 100 siliques were purified for the different marker lines on average by flow cytometry (supplementary material Fig. S2). Pools of ~3000 GFP-positive nuclei were used for RNA extraction, representing one biological replicate.

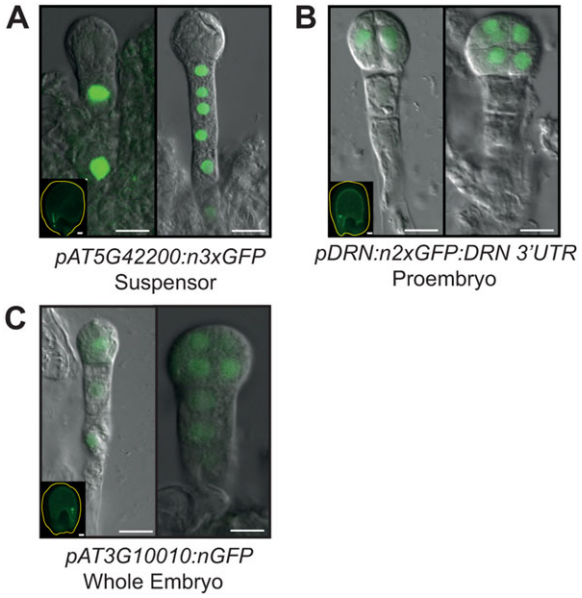

Fig. 1. Specific marker lines used for FANS. (A) Suspensor marker line at the two-cell (left) and early globular (right) stages. (B) Proembryo marker line at the two-cell (left) and eight-cell (right) stages. (C) Whole-embryo marker line at the one- (left) and four-cell (right) stages. Insets show overviews of seeds with embryo-specific GFP expression for each marker line. Scale bars: 10 µm; 20 µm in insets.

After RNA amplification and biotinylation, the transcriptome analyses were carried out in biological triplicates with a standard Affymetrix ATH1 genome array, which covers roughly 71% of the to date presumed 33602 total *Arabidopsis* genes (Lamesch et al., 2012). For our analyses, we used MAS5 normalized probe set signals (see GEO GSE60242) as well as gcRMA (gene chip robust multi-array average) normalized and log2 transformed (supplementary material Table S1) values. When we compared microarray probe sets only detected as present (P) in the MAS5 normalization algorithm for raw values across all three replicates, they showed a chip coverage of 34, 32 and 25% for nEMB (nuclei from whole embryo), nPE (nuclei from proembryo) and nSUS (nuclei from suspensor), respectively. The lower coverage for the nSUS is due to the lower concordance in present (P), marginal (M) and absent (A) MAS5 calls between all three nSUS replicates compared with nPE or nEMB values (supplementary material Fig. S3). Nevertheless, there is substantial overlap of expressed genes designated as three times present (3×P) in the MAS5 calls between the three samples (supplementary material Fig. S4). The gcRMA values were used for correlation analysis of the biological replicates for nuclear transcriptomes from the whole embryo, proembryo and suspensor, as well as from data recently acquired from the shoot apex in adult plants (Yadav et al., 2014; GEO accession number GSE28109). This analysis showed high similarity between nuclear embryo replicates with Pearson correlation coefficients (PCC) ranging from 0.962 to 0.984. Interestingly, correlation was also high between nuclear replicates of the different embryonic tissues, whereas the correlation was low compared with the cellular shoot apex transcriptomes (supplementary material Table S2).

Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

Taken together, we detected a substantial number of genes that are active in the proembryo and/or the suspensor, as well as in the whole embryo during *Arabidopsis* early embryogenesis.

Differentially expressed candidate genes

In order to find significantly differentially expressed candidate genes between the nPE and nSUS samples, a rank product analysis was conducted with a percentage of false positives smaller than 0.1 and a change of greater than twofold. A total of 307 and 180 array elements corresponding to 335 and 181 locus identifiers were enriched for nPE and nSUS, respectively (supplementary material Tables S3 and S4). To gain insight into the function and complexity of both cell types during early embryogenesis, we analyzed Gene Ontology (GO) terms for differentially expressed genes between the proembryo and the suspensor. Our analysis showed that over-represented GO terms for proembryo-enriched genes are 'DNA or RNA metabolism'/'cell organization and biogenesis' and 'structural molecule activity'/'protein binding' in the categories 'Biological processes' and 'Molecular function', respectively (Fig. 2A; supplementary material Table S5). The multidimensional proembryo often changes its plane of cell division in contrast to the stereotypic suspensor division type and undergoes cellular differentiation during early embryogenesis, which requires dynamic cytoskeleton reorganization and the coordinated change of gene expression (Webb and Gunning, 1991; Lau et al., 2012). In the suspensor, preferentially expressed genes were associated with 'response to stress'/'transport' and 'receptor binding or activity'/'hydrolase' in the categories 'Biological processes' and 'Molecular function', respectively (Fig. 2B; supplementary material Table S6). Evidently, the suspensor has also been implicated in providing the proembryo with nutrients and plant hormones to be delivered by transporter proteins and it undergoes programmed cell death during late development that might share similar mechanisms with stress response signaling (Bozhkov et al., 2005; Kawashima and Goldberg, 2010). Overall, our results revealed that in addition to the morphological differences, the proembryo and the suspensor also appear distinct in gene expression profiles during early embryogenesis. Furthermore, our GO analysis indicated a distinct function and an increased complexity of cellular activities in the proembryo compared with the suspensor during embryogenesis.

When we had a closer look into the two gene lists, we could find genes that were previously shown to be differentially expressed and important for patterning and specification processes during embryogenesis, e.g. *PIN-FORMED 1* (*PIN1*, *AT1G73590*), *WUSCHEL-RELATED HOMEOBOX 2* (*WOX2*, *AT5G59340*), *HANABU TARANU* (*HAN*, *AT3G50870*), *OBF BINDING PROTEIN 1* (*OBP1*, *AT3G50410*) or *FUSCA3* (*FUS3*, *AT3G26790*) (Aida et al., 2002; Friml et al., 2003; Kroj et al., 2003; Haecker et al., 2004; Skirycz et al., 2008; Nawy et al., 2010). As *DRN* is not represented on the ATH1 chip, we tested mRNA levels by qRT-PCR. Indeed, *DRN* transcripts are highly abundant in the proembryo when compared with the suspensor (supplementary material Fig. S5A). Several other genes previously reported as being preferentially expressed in the suspensor (Friml et al., 2003; Haecker et al., 2004; Breuninger et al., 2008) either did not pass the stringent statistical analysis (*WOX9*/*AT2G33880*) or were not detectable with the microarray (*PIN7*/*AT1G23080*, *WOX8*/*AT5G45980*). However, even though, for example, *PIN7* was under the microarray detection limit, we were able to detect its mRNA slightly more abundant in the suspensor by qRT-PCR (supplementary material Fig. S5B), which in conclusion indicates the existence of false negatives in our data set due to sensitivity thresholds.

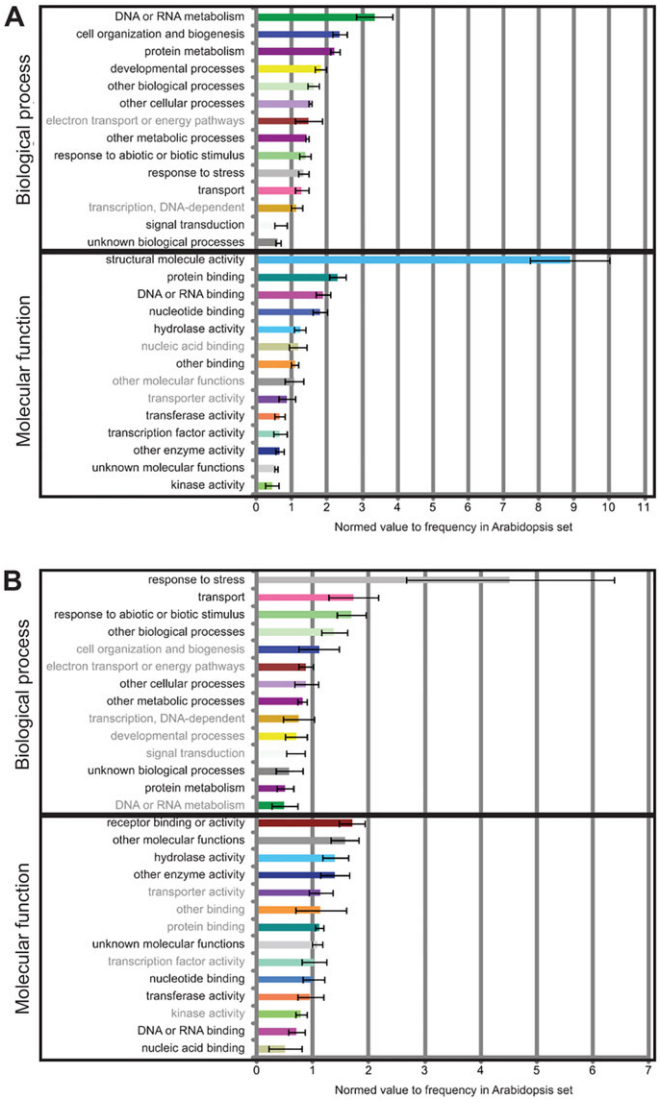

Fig. 2. Graphical representation of GO analysis for nuclear proembryo- and suspensor-enriched genes. (A) Enrichment of GO terms in categories 'biological process' and 'molecular function' for proembryo. (B) Enrichment of GO terms in categories 'biological process' and 'molecular function' for suspensor. Error bars indicate bootstrap s.d. Differences in the GO terms depicted in gray letters are not statistically significant (P>0.05).

Auxin was indirectly shown to accumulate in the cells of the proembryo at early embryonic stages (Friml et al., 2003). In addition to the already mentioned auxin efflux carrier PIN1, we were also able to detect the auxin biosynthesis genes *TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1* (*TAA1*, AT1G70560) and *YUCCA4* (*YUC4*, *AT5G11320*) as well as *SHORT INTERNODES* (*SHI*, *AT5G66350*) and *MYB77* (*AT3G50060*) as being proembryo enriched, which are involved in auxin biosynthesis and signal transduction, respectively (Cheng et al., 2007; Shin et al., 2007; Stepanova et al., 2008; Baylis et al., 2013). Interestingly, suspensor-specific promoter-reporter expression of *YUC4* was recently shown at 16-cell stage of embryogenesis (Robert et al., 2013). However, as previous results showed transcript accumulation at later stages only in the proembryo (Cheng et al., 2007) and LCM-derived data do not show expression in the suspensor at the globular stage (see GEO GSE11262), the observed suspensor-specific expression of *YUC4* might be due to the promoter-reporter

Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

construct likely not reflecting the *in vivo* mRNA distribution. All this is in accordance with the evidence demonstrating the importance of auxin signaling in cell division and differentiation during early embryogenesis (Lau et al., 2012).

Microarray data validation by promoter expression analysis and *in situ* hybridization

To further validate the microarray results, we randomly selected 12 genes statistically enriched for the proembryo and nine for the suspensor for global expression analysis. Promoters of differing lengths, including the 5′UTR region, were constructed to drive expression of n3×GFP or n3×RFP. In most cases, the expression patterns of the promoter fusion constructs correlated with the microarray results (Table 1, Fig. 3). In the one- or two-cell stage embryo, there was no exclusive expression detectable in either the suspensor or the proembryo, but rather a broad expression in all cells of the whole embryo with differences visible in expression strength between proembryo and suspensor. Interestingly, some genes (e.g. *AT3G62480* and *AT3G52780*) showed expression in the suspensor but not in the proembryo (Table 1, Fig. 3A,B). One gene (*AT5G46230*) showed global expression in the whole embryo at the earliest stages but expression was later predominantly visible in the suspensor (Fig. 3H). Moreover, reporter expression for several candidate genes (e.g. *AT2G32100*, *AT5G05940* and *AT3G17290*) remained universal in the whole embryo, which only later appeared stronger in one cell lineage and weaker in the other (Table 1, Fig. 3C,E,F). Three promoter fusion constructs did not confer any visible GFP expression in the embryo, which might be due to missing elements that are important for proper expression or a false-positive signal from the microarray (not shown). Taken together, the expression patterns of the promoter fusions are overall in concordance with the differences found in the statistical analysis of the microarray data. Minor discrepancies between the promoter fusion and the microarray data can most likely be attributed to the stability and low turnover rate of GFP protein inside the plant cell. In total, we tested 21 promoters fused to n3×GFP or n3×RFP, of which only three were not embryo expressed. Of the 18 embryo-expressed genes, 16 recapitulated the microarray results of differentially expressed transcripts (Table 1).

As promoter fusion constructs in some cases may not fully recapitulate true gene expression due to the possible lack of crucial regulatory elements, we performed *in situ* hybridization for some of the proembryo- and suspensor-enriched transcripts. Overall, the *in situ* hybridization results for the selected, differentially expressed candidate genes (*AT1G04645*, *AT1G28300*, *AT5G46230*, *AT5G61030* and *AT3G44750*) were consistent with the microarray analysis (Fig. 3D,G-J). However, we could not detect any signal in the early embryo for *AT2G46690* (data not shown). As the promoter-reporter lines also did not produce any signal, this is probably a false-positive signal on the microarray. Moreover, the promoter fusion analysis for two proembryo-enriched genes (*AT5G61030* and *AT3G44750*) did not correlate with our microarray analysis as the corresponding reporter-gene constructs indicated ubiquitous expression in all cells of the embryo (Fig. 3I,J). The *in situ* hybridization for these two genes, however, showed stronger signals in the proembryo at early stages of embryogenesis (Fig. 3I,J), indicating a possible lack of some regulatory elements in the respective promoter regions cloned or post-transcriptional regulation of the endogenous gene. Furthermore, the validation of the differentially

Table 1. Differentially expressed candidate genes used for *in vivo* validation

Locus	Probe set ID	Promoter expression analysis *in vivo*	*In situ*?	FC:(cl1/cl2)	Average MAS5 nPE	Average MAS5 nSUS	Embryo specific?
Proembryo-enriched transcripts tested							
AT5G26270	246888_at	EMB, stronger PE		35.21	5141.74	155.85	Yes
AT1G77580	259760_at	Globular stage PE		7.07	315.51	85.40	Yes
AT5G05940	250756_at	EMB, at late globular/early heart stage stronger PE		6.50	326.50	65.65	Yes
AT3G17290	258459_at	EMB, at 8/16-cell stage stronger PE	Yes	4.44	324.18	90.10	Yes
AT2G35605	266641_at	EMB, stronger PE		4.41	2616.68	850.30	
AT5G61030	247575_at	EMB, stronger PE	Yes	3.92	1181.90	365.75	
AT1G31400	262555_at	No expression		3.70	464.74	202.79	Yes
AT1G64220	262336_at	Globular stage PE		3.44	555.46	235.96	
AT1G28300	245669_at	Not available	Yes	3.41	448.10	244.37	Yes
AT5G22650	249901_at	Inconsistent expression		3.41	5074.09	2203.15	
AT5G43510	249157_at	PE early heart stage		3.15	362.94	159.05	Yes
AT3G44750	252625_at	Inconsistent expression	Yes	2.31	1970.12	885.23	
AT3G55660	251778_at	Late globular stage hypophysis and lower tier		2.22	584.78	366.01	Yes
Suspensor enriched transcripts tested							
AT2G46690	266322_at	No expression	Yes	8.60	119.51	618.06	
AT3G62480	251212_at	SUS		5.98	100.65	683.83	
AT1G48470	261305_at	No expression		5.67	139.09	603.22	Yes
AT1G04645	264610_at	Not available	Yes	4.92	2343.15	10,066.28	
AT1G54160	263158_at	EMB, stronger SUS		4.39	210.41	588.35	
AT3G52780	252004_at	SUS		4.17	366.63	1419.88	
AT5G46230	248889_at	EMB, stronger SUS	Yes	3.87	166.79	413.70	
AT5G07440	250580_at	EMB, stronger SUS		3.51	451.98	1162.80	
AT1G74190	260253_at	SUS		3.30	27.33	166.46	Yes
AT2G32100	265724_at	EMB, stronger SUS		2.84	252.65	613.39	Yes

For all constructs, a short description of the expression patterns in transgenic embryos is given. Gene expression tested by in situ hybridization is indicated by the 'yes'. Results of the RankProduct analysis for fold change (FC) are indicated. Additionally, average MAS5 expression values of the three replicates are given for nPE and nSUS samples, and the genes overlapping with the embryo-specific analysis results are designated with 'yes'. PE, proembryo; SUS, suspensor; EMB, whole embryo; FC, fold change.

Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

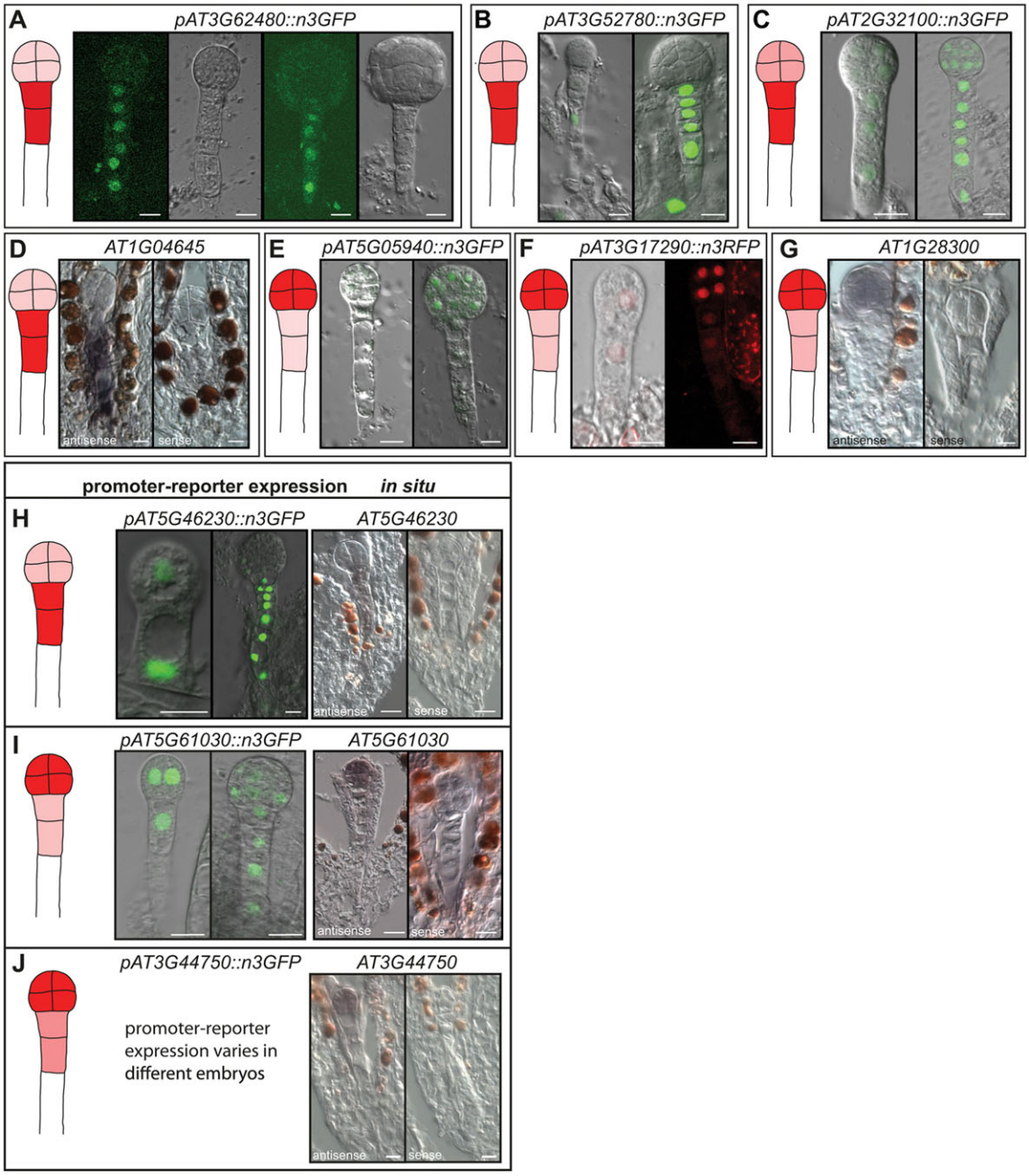

Fig. 3. Promoter fusion analysis and *in situ* hybridization for selected differentially expressed candidate genes in early embryos. (A-G) Temporal promoter-reporter expression and *in situ* hybridization of suspensor-enriched genes (A-D) and proembryo-enriched genes (E-G) during early embryogenesis. (H-J) Comparison of promoter-reporter expression and *in situ* hybridization for the same genes enriched in suspensor (H) and proembryo (I,J). Color shading in the schematic representation of the *Arabidopsis* embryo indicates the expression levels according to the microarray dataset (dark red, stronger expression; light red, weaker expression). Scale bars: 10 µm.

expressed genes by *in situ* hybridization was not only additive to but also complementary with the promoter fusion analysis. In summary, the promoter fusion studies and *in situ* hybridization results for 23 genes in total strongly correlated with the results of the microarray analysis, which emphasizes the high quality of the whole dataset.

Nuclear transcriptomic data as proxy for gene expression profiling

For comparability reasons and to demonstrate that the nuclear results are indeed useful for detection of tissue-specific transcripts, we manually isolated intact whole embryos at the 16- to 32-cell stages and directly extracted RNA without prior fixation. After amplification and microarray hybridization, samples were analyzed as mentioned above (cEMB; see GEO GSE60242 and supplementary material Table S1). When we compared MAS5 calls 3xP between nEMB and cEMB, we observed a strong 70% overlap (Fig. 4A). Additionally, the 30% genes not overlapping in the analysis showed weaker expression across the replicates on average (nEMB average value 239, cEMB 203) compared with the average expression of the overlapping 70% of 934 (supplementary material Tables S7-S9), indicating that the differences in detection calls might be due to sample/microarray noise.

Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

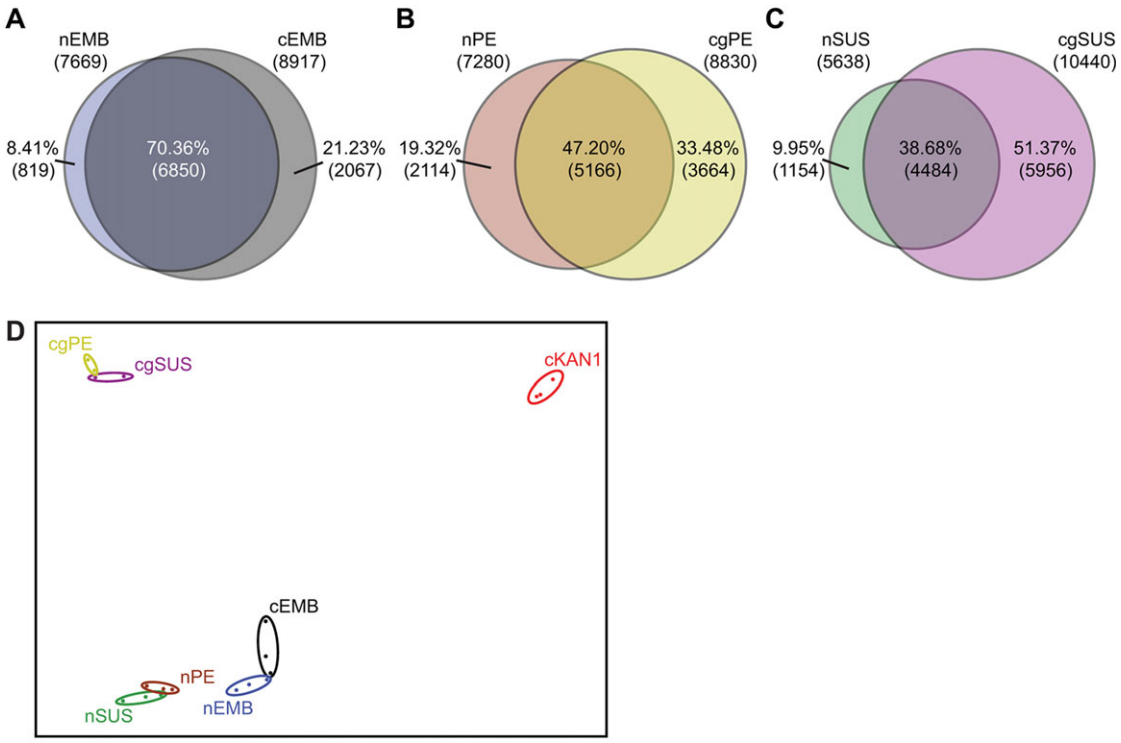

Fig. 4. Comparison of nuclear and cellular transcriptome data from different tissue types. (A-C) Venn diagrams showing overlap of MAS5 3× present calls between nEMB and cEMB (A), nPE and cgPE (B), and nSUS and cgSUS (C). (D) Principal component analysis of biological replicates from the different nuclear and cellular tissue types. nPE, nuclei from proembryo; nSUS, nuclei from suspensor; nEMB, nuclei from whole embryo; cEMB, cells from whole embryo; cgPE, cellular globular-stage proembryo; cgSUS, cellular globular-stage suspensor; cKAN1, cellular KANADI 1 expression domain in the shoot.

Recently, LCM was used in combination with microarrays to generate a very elaborate expression atlas of various seed compartments, including the embryo at different developmental stages of the ovule (Le et al., 2010; Belmonte et al., 2013). Among other tissue types, cellular expression profiles were created for the proembryo and suspensor at the globular embryo stage, which we here term cellular globular proembryo (cgPE) and cellular globular suspensor (cgSUS), respectively (Le et al., 2010). We MAS5-normalized and log2-transformed the raw values from the cgPE and cgSUS replicates as described above (see GEO GSE11262 and supplementary material Table S1). To determine whether the corresponding nuclear and cellular datasets were comparable, we first checked the overlap of MAS5 calls 3xP. However, unlike the higher overlap between nuclear samples (nPE/nSUS 64.5%, supplementary material Fig. S4), there are substantially fewer array elements shared between nPE/cgPE (47.2%) and nSUS/cgSUS (38.7%, Fig. 4B,C); furthermore, these percentages are very similar for nPE/cgSUS or nSUS/cgPE (data not shown).

Additionally, after testing the normalized and transformed values of all replicates for comparability by box-plot analysis (supplementary material Fig. S6A), we performed hierarchical cluster analysis to group the different expression profiles. In summary, all replicates of one specific experiment group together and there are two main clusters consisting of: (1) all nuclear (nPE, nSUS, nEMB) plus the cellular embryo sample (cEMB); and (2) the cellular globular-embryo samples (cgPE, cgSUS). In cluster (1) there are subgroups of nuclear samples and the cellular embryo sample (supplementary material Fig. S6B). These differences are further corroborated in a principal component analysis (PCA) plot where cgPE clusters with cgSUS, nPE with nSUS, and nEMB with cEMB (Fig. 4D). Published data of a KAN1 (KANADI 1) expression domain was used as an outgroup (Yadav et al., 2014; GEO accession number GSE28109). However, the cgPE and cgSUS cluster is farther away from the two other embryonic clusters. We can conclude that as the influence of fixation, nuclear RNA, and age of embryos on the observed expression profiles seems to be subtle, the main factor for these discrepancies between the LCM-derived and our datasets must be the different extraction techniques and RNA amplification protocols.

To compare the nuclear FANS and cellular LCM datasets beyond *in silico*, we compared the expression values of the LCM data for genes we tested with the promoter fusion constructs and *in situ* hybridization Table 1; see GEO accession number GSE11262). For three constructs not showing expression in the embryo, the LCM values were consistent with our microarray results, indicating the same false-positive results (*AT1G31400*, *AT2G46690* and *AT1G48470*). The LCM array element values for 17 genes showing expression in the embryo were consistent with our results (*AT1G77580*, *AT5G05940*, *AT2G35605*, *AT5G61030*, *AT1G64220*, *AT1G28300*, *AT5G22650*, *AT5G66940*, *AT3G44750*, *AT3G55660*, *AT3G62480*, *AT1G04645*, *AT1G54160*, *AT3G52780*, *AT5G07440*, *AT1G74190* and *AT2G32100*). Four genes (*AT5G26270*, *AT3G17290*, *AT5G43510* and *AT5G46230*) appeared as false negatives in the LCM dataset as the expression values were very low and often the MAS5 call was absent for both the proembryo and the suspensor replicates.

Recently, a report described the expression patterns of multiple auxin response factors (ARF) using promoter-reporter constructs during early *Arabidopsis* embryogenesis (Rademacher et al., 2011). Four of the tested ARF promoters (*ARF12*/*AT1G34310*, *ARF17*/*AT1G77850*, *ARF21*/*AT1G34410* and *ARF23*/*AT1G43950*) were designated as only being expressed in the endosperm but not the embryo itself (Rademacher et al., 2011); these genes were designated as absent and in essence not expressed in our dataset (supplementary

Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

material Table S10). In the LCM dataset, however, all four genes were called present and at least weakly expressed within the suspensor. We also compared the different suspensor data sets for presence of previously described endosperm-specific genes (Kinoshita et al., 1999, 2004; Luo et al., 2000; Kang et al., 2008; Li et al., 2013; Barthole et al., 2014). The LCM results show all six genes tested as present, whereas our data indicate that only two out of six are also at least weakly expressed in the suspensor (supplementary material Table S10) and those two were also detected in a RNA-seq transcriptome analysis (Nodine and Bartel, 2012). This suggests contamination of at least the suspensor samples with surrounding endosperm in the LCM dataset. As it is essentially impossible to dissect tissue accurately with LCM in the third dimension, the list of apparent suspensor genes is likely contaminated by endosperm-expressed genes. On the contrary, using our methodology, we could minimize contamination with cellular or nuclear material from embryo-surrounding cells. To substantiate this notion, we generated a list of putative embryo-specific genes for the *Arabidopsis* seed by comparing our statistically enriched candidates with publicly available seed transcriptome data. For this purpose, we combined our enriched nPE and nSUS gene lists, and differentiated it to the combined LCM data 3xP for different seed compartments at the globular stage, excluding the globular proembryo and suspensor (cgSEED). In total, we detected 95 genes (supplementary material Tables S11 and S12) that contained known genes such as the aforementioned *HAN*/*MNP*, *WOX2*, *OBP1* and *YUC4*, as well as the embryonic identity regulators *LEAFY COTYLEDON 2* (*LEC2*) and *BABYBOOM* (*BBM*) (Stone et al., 2001; Boutilier et al., 2002). Interestingly, 11 out of our 23 *in vivo* tested candidate genes were also present in this putative embryo-specific list (Table 1). In summary, this strongly indicates that with our approach we can detect tissue-specific genes.

DISCUSSION

In this study, we have described and validated a nuclear extraction and purification protocol for expression analysis of inaccessible cell types in the *Arabidopsis thaliana* seed. Given that the unequal distribution of some transcripts in the early embryo leads to distinguishable cell types and likewise the unequal distribution of specific transcripts was reported in the apical and basal cells of *Arabidopsis* and tobacco embryos (Breuninger et al., 2008; Hu et al., 2010; Ueda et al., 2011), we reasoned that these cell types might be a well-suited test field for our method and that the generation of expression profiles for the proembryo and suspensor of early *Arabidopsis* embryos will provide insights into better understanding of early embryo development. Several of the statistically enriched candidate genes for the proembryo were previously described to have important functions during early embryogenesis, some of which were shown as proembryo-enriched expressed genes in our dataset. For example, *HANABA TANARU* (*HAN*) was shown to be expressed in the apical cell of the embryo and plays a role in setting up the boundary between proembryo and suspensor (Nawy et al., 2010). *PIN-FORMED 1* (*PIN1*), known as an auxin efflux facilitator, is expressed in the proembryo cells, mediating auxin flow from apical cells to the hypophysis (Friml et al., 2003), which is in turn crucial for root initiation. Another apically expressed gene is the homeobox transcription factor *WUSCHEL RELATED HOMEOBOX 2* (*WOX2*), which plays a fundamental role in the establishment of the apical domain (Haecker et al., 2004). Moreover, the suspensor-expressed gene *FUSCA3* (*FUS3*) lacks apical expression due to repression by *DICER-LIKE1* (*DCL1*), and early matured embryos in the *dcl1* mutant show ectopic expression of *FUS3* in the proembryo (Willmann et al., 2011). All these examples initially supported our results as these genes were not only present in one or the other dataset but were also among the statistically most significant ones.

The *in vivo* expression analyses using promoter-GFP fusion constructs as well as *in situ* hybridization strongly correlated with the microarray results for the candidate genes tested. This demonstrated the validity of the microarray results after stringent statistical analysis from expression data generated for specific tissues in the *Arabidopsis* embryo at the earliest developmental stages.

The high correlation of nuclear and cellular embryonic transcriptome data generated is also very encouraging for the use of this method in other studies. However, apart from a few similarities, comparison with published expression data generated from respective cellular embryonic tissues by laser capture microdissection (Le et al., 2010) revealed major differences in types of genes expressed in the given tissues. Potentially, there are many factors influencing the final transcriptomic data. These include: (1) the plant accession used; (2) the developmental stage of the tissue studied; (3) the RNA composition (cellular, cytoplasmic, nuclear); (4) different fixation approaches; (5) RNA extraction; and (6) RNA amplification method. By comparing our nuclear RNA transcriptome results with those from cellular RNA of non-fixed embryos – assuming the influence of accession and RNA extraction method as marginal – we conclude that either the RNA amplification or (probably) the tissue isolation approach has the greatest impact. This notion seems reasonable because we used a commercial kit and a polyT primer, whereas the LCM RNA was amplified with a polyT/random primer mixture (Le et al., 2010).

Crosschecking *in vivo* expression results, we did not see any disadvantages of our transcriptomic data except a certain proportion of false negatives in detection of low-expressed genes. On the contrary, we propose that our approach has certain advantages, the most important one being the possibility of studying any tissue of interest and the other being a decreased risk of contamination with embryo-surrounding cells compared with LCM. Furthermore, we demonstrate that this approach is able to detect tissue-specific genes with a very small expression domain. Even though transgenics are required in order to use our approach, it is nevertheless applicable to any other transformable plant or animal tissue where generating expression data from a given cell type is the goal. Importantly, the method described here not only enables expression studies to be performed but also has the potential to study DNA and histone modifications.

MATERIALS AND METHODS

Plant materials and growth conditions

All *Arabidopsis thaliana* lines used are *Col-0*. The GAL4-GFP enhancer-trap lines generated by the Haseloff lab were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC). For growth under sterile conditions, seeds were surface sterilized with 25% bleach, washed three times, and grown on half-strength Murashige and Skoog (MS)-containing 0.8% agar plates with 10 g/l sucrose. Seedlings were transferred to soil and grown at 22°C to 24°C in a growth chamber under a 16 h/8 h light/dark cycle.

Molecular cloning and genotyping

TAIL-PCR was performed as previously described (Liu and Chen, 2007). All genomic fragments (626 bp-2615 bp upstream of ATG) for the promoter-GFP fusions were PCR amplified and sub-cloned into *pGEM-T* vector (Promega). The n3×RFP was assembled from PCR-amplified monomers in *pGII Kan* vector. All fragments were finally introduced into *pGII Kan:n3×GFP* (Takada and Jürgens, 2007) or *pGII Kan:n3×RFP*. A *pAT3G10100* fragment was introduced into *pGII Kan:n3×GFP* that resulted in *pAT3G10100::nGFP*.

Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

The n2xGFP was amplified from *pGII Kan:n3×GFP* and introduced into *pGII Kan*. For generating *pDRN:n2×GFP:DRN 3′UTR*, a 1378 bp *DRN 3′ UTR* fragment was PCR amplified and sub-cloned into *pGEM-T*. Then the *DRN 3′UTR* was introduced into *pGII Kan:n2×GFP* generating *pGII Kan: n2×GFP:DRN 3′UTR*. A 4145 bp DRN promoter upstream of the start codon was PCR amplified and sub-cloned into *pGEM-T*. *pDRN* was finally introduced into *pGII Kan:n2×GFP:DRN 3′UTR* generating *pDRN: n2×GFP:DRN 3′UTR*. Oligonucleotides used for cloning can be found in supplementary material Table S13.

Nuclear isolation

Fresh *Arabidopsis* ovules were collected in RNAlater buffer (QIAGEN) and kept in fixation buffer (0.1% paraformaldehyde in RNAlater) for 5-10 min and ground thoroughly using a pestle in a 1.5 ml tube. The CelLytic PN kit (Sigma) was used for the following procedures.

FANS

Fluorescently labeled nuclei were identified by plotting peak GFP fluorescence (513/17) against autofluorescence (575/25) using a MoFlo Legacy (Beckman Coulter) FACS fitted with a 488 nm laser (100 mW) triggering off the FSC (forward scatter channel). Tests of co-staining with propidium iodide to label free nuclei identified the same GFP population; therefore, staining was deemed unnecessary. Flow cytometric analyses were carried out as follows: 1× PBS pH 7.0, 70 µM stream, ∼60.5/∼60.0 psi, ∼95 kHz, 1-2 single drop envelope.

Manual isolation of embryos

Isolation was performed essentially as previously described (Nodine and Bartel, 2012). In brief, early globular stage embryos were squeezed out from the ovules on a microscope slide and washed three times in water and subsequently collected in RNAlater. Forty to 50 embryos were pooled per biological replicate.

RNA extraction and amplification

The sorted positive nuclei were collected in RNA extraction buffer [10 mM Tris-HCl (pH 7.9), 50 mM EDTA (pH 7.9), 0.2 M NaCl, 0.5% SDS, 0.5 mg/ml RNase inhibitor (Fermentas), 600 µg/ml proteinase K] (Khodosevich et al., 2007). The buffer containing the GFP-positive nuclei was incubated at 55°C with vigorous shaking for 10-15 min. The total volume was adjusted to 600 µl RNase-free water and an equal volume of phenol (pH 4.2) was added. The solution was vortexed thoroughly and kept on ice for 5 min and afterwards centrifuged at 14,000 ***g*** for 10 min at 4°C. The aqueous phase was transferred into a new tube and an equal volume of phenol:chloroform (1:1) was added. The solution was mixed thoroughly and kept on ice for 5 min and centrifuged at 14,000 ***g*** for 10 min at 4°C. The aqueous phase was transferred into a new tube and equal volumes of isopropanol and 20 µg glycogen were added. The solution was then mixed thoroughly and kept at −20°C overnight and centrifuged at 16,100 ***g*** for 45 min at 4°C. Following the centrifugation, the resulting pellet was washed with 70% cold ethanol and dried at room temperature. The pellet was eventually dissolved in RNase-free water. For DNase treatment, a commercial kit (DNase I, Fermentas) was used and afterwards the RNeasy Micro Kit (QIAGEN) was used for RNA cleanup.

One to three ng of total RNA was used for cDNA synthesis and amplification (Arcturus RiboAmp HS PLUS RNA Amplification Kit) and the resulting cDNA was fragmented and labeled using the ENZO BioArray Single-round RNA amplification and biotin labeling system. Fragmented cDNA (12.5 µg) was hybridized on Affymetrix GeneChip ATH1 *Arabidopsis* Genome Array.

Microarray data analysis

Microarray datasets as .CEL files for LCM and shoot KAN1 were downloaded from the GEO DataSets on the NCBI (National Center for Biotechnology) website (https://www.ncbi.nlm.nih.gov/) (accession numbers GSE11262 and GSE28109). The globular-stage seed gene list excluding the embryo and suspensor was downloaded from Gene Networks in Seed Development website (http://seedgenenetwork.net/). Microarray data analyses were performed using diverse packages implemented in 'R' (v2.14.2; http://www.r-project.org). Log2-based expression estimates were obtained from .CEL files using 'gcRMA' (v2.26.0) (Wu et al., 2004). Differentially expressed genes were identified by 'RankProducts' (v2.26.0) using 100 permutations and a percentage false-positive (pfp) cut-off of 0.05 (Breitling et al., 2004). Present, marginal and absent calls were calculated using MAS5 as implemented in the 'affy' package (v1.32.1). Pearson correlation coefficients of gcRMA values were calculated with Microsoft Office Excel 2007. GO classification of proembryo- and suspensor-enriched genes was created with the Classification SuperViewer Tool from BAR (The Bio-Analytical Resource for Plant Biology) (http://bar.utoronto.ca/ntools/cgi-bin/ntools_classification_superviewer.cgi). All Venn diagrams were generated with a combination of BioVenn (http://www.cmbi.ru.nl/cdd/biovenn/), Venn diagram plotter (http://omics.pnl.gov/software/venn-diagram-plotter) and Adobe Illustrator. Quality control analyses (Box plot, hierarchical clustering of samples, principal component analysis) of all biological replicates were performed with CLC Main Workbench software (version 6.6.2). Microarray data have been deposited with GEO with accession number GSE60242.

Quantitative real-time PCR

Owing to limitations in RNA quantity, amplified cRNA was used for cDNA synthesis (RevertAid First Strand cDNA Synthesis Kit, Fermentas). Quantitative PCR was performed on a Chromo4 Real-Time Detector (Bio-Rad) with Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen). PCRs were carried out in triplicate with specific primer pairs (supplementary material Table S13) and transcript levels were normalized to *ACTIN2*.

RNA *in situ* hybridization

The primers for probe synthesis are listed in supplementary material Table S13. The fragments for the sense and antisense probes were PCR-amplified and inserted into *pBSK*⁻ or *pGEM-T* vectors. *In vitro* transcription was performed with T7 or SP6 primers and with Fermentas *in vitro* transcription kit. Both ends of young siliques were cut off and the middle part was fixed in cold fixation solution (4% paraformaldehyde in DEPC-treated water, 0.1% Tween-20). A conventional plastic syringe was used for vacuum infiltration and the samples were kept overnight in the fixation solution at 4°C. Following 1×PBS incubation for 2×30 min, the samples were dehydrated through a graded ethanol series (30%, 40%, 50%, 60%, 80%, 90%, 95%) for 1 h each and finally embedded in paraffin. Paraffin-embedded samples were sectioned on a microtome at 6 µm. The procedures of hybridization and staining were performed as described previously (Schlereth et al., 2010).

Microscopy

For differential interference contrast (DIC) microscopy and fluorescence analysis, ovules were mounted on slides containing clearing solution [chloral hydrate, water, and glycerol (ratio w/v/v: 8:3:1)]. For fluorescence analysis, embryos were gently squeezed out from ovules and mounted in 10% glycerol (v/v). An Olympus IX81 confocal laser scanning microscope (image acquisition software: FV10-ASW; objectives: UPlanSApo ×40) was used for confocal microscopic analysis. Images were further processed using Adobe Photoshop software. Zeiss Axio Imager (image acquisition software: AxioVision; camera: AxioCam HRc; objectives: Plan-APOCHROMAT ×20 and ×40) was used for wide-field and DIC images and images were further processed with AxioVision SE64 Rel. 4.9.1 software.

Acknowledgements

We thank Steffen Lau for critical reading of the manuscript and Arvid Herrmann and Ole Herud for assistance. We thank the Nottingham *Arabidopsis* Stock Centre (NASC) for providing enhancer-trap lines.

Competing interests

The authors declare no competing financial interests.

Author contributions

D.S., J. Kong, U.M., I.D.S., M.B. and G.J. were involved in the conception and design of the experiments. D.S. and J. Kong co-wrote the manuscript and performed most of the experiments. K.H., K.W.B. and J. Kilian performed the flow cytometry analysis. A.H. and M.B. generated the transcriptome datasets for manually isolated embryos.

M.S. performed microarray analysis for MAS5, gcRMA and RankProducts. M.K. performed the qRT-PCR analysis.

Funding
This work was funded by the SIREN network [European Commission under FP7-PEOPLE-2007-1-1-ITN], the European Molecular Biology Organization [EMBO-ALTF 108-2006 to I.D.S.], the Marie Curie Intra-European Fellowship scheme [postdoctoral fellowship FP6 MEIF-CT-2007-041375 to I.D.S.], the DFG [Deutsche Forschungsgemeinschaft, BA3356/2-1 to M.B.] and the Max Planck Society.

Supplementary material
Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.116459/-/DC1

References
Aida, M., Vernoux, T., Furutani, M., Traas, J. and Tasaka, M. (2002). Roles of PIN-FORMED1 and MONOPTEROS in pattern formation of the apical region of the Arabidopsis embryo. *Development* **129**, 3965-3974.
Barthelson, R. A., Lambert, G. M., Vanier, C., Lynch, R. M. and Galbraith, D. W. (2007). Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. *BMC Genomics* **8**, 340.
Barthole, G., To, A., Marchive, C., Brunaud, V., Soubigou-Taconnat, L., Berger, N., Dubreucq, B., Lepiniec, L. and Baud, S. (2014). MYB118 represses endosperm maturation in seeds of arabidopsis. *Plant Cell*.
Baylis, T., Cierlik, I., Sundberg, E. and Mattsson, J. (2013). SHORT INTERNODES/STYLISH genes, regulators of auxin biosynthesis, are involved in leaf vein development in Arabidopsis thaliana. *New. Phytol.* **197**, 737-750.
Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., Hashimoto, M., Fei, J., Harada, C. M., Munoz, M. D. et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl. Acad. Sci. USA* **110**, E435-E444.
Birnbaum, K., Shasha, D. E., Wang, J. Y., Jung, J. W., Lambert, G. M., Galbraith, D. W. and Benfey, P. N. (2003). A gene expression map of the Arabidopsis root. *Science* **302**, 1956-1960.
Bonner, W. A., Hulett, H. R., Sweet, R. G. and Herzenberg, L. A. (1972). Fluorescence activated cell sorting. *Rev. Sci. Instrum.* **43**, 404.
Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C. M., van Lammeren, A. A., Miki, B. L. et al. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **14**, 1737-1749.
Bozhkov, P. V., Filonova, L. H. and Suarez, M. F. (2005). Programmed cell death in plant embryogenesis. *Curr. Top. Dev. Biol.* **67**, 135-179.
Breitling, R., Armengaud, P., Amtmann, A. and Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* **573**, 83-92.
Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M. and Laux, T. (2008). Differential expression of WOX genes mediates apical-basal axis formation in the Arabidopsis embryo. *Dev. Cell* **14**, 867-876.
Chandler, J. W., Cole, M., Flier, A., Grewe, B. and Werr, W. (2007). The AP2 transcription factors DORNRÖSCHEN and DORNRÖSCHEN-LIKE redundantly control *Arabidopsis* embryo patterning via interaction with PHAVOLUTA. *Development* **134**, 1653-1662.
Cheng, Y., Dai, X. and Zhao, Y. (2007). Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in Arabidopsis. *Plant Cell* **19**, 2430-2439.
Cole, M., Chandler, J., Weijers, D., Jacobs, B., Comelli, P. and Werr, W. (2009). *DORNRÖSCHEN* is a direct target of the auxin response factor MONOPTEROS in the *Arabidopsis* embryo. *Development* **136**, 1643-1651.
Cumberledge, S. and Krasnow, M. A. (1994). Preparation and analysis of pure cell populations from Drosophila. *Methods Cell Biol.* **44**, 143-159.
De Smet, I., Vassileva, V., De Rybel, B., Levesque, M. P., Grunewald, W., Van Damme, D., Van Noorden, G., Naudts, M., Van Isterdael, G., De Clercq, R. et al. (2008). Receptor-like kinase ACR4 restricts formative cell divisions in the Arabidopsis root. *Science* **322**, 594-597.
Deal, R. B. and Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev. Cell* **18**, 1030-1040.
Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S. R., Weiss, R. A. and Liotta, L. A. (1996). Laser capture microdissection. *Science* **274**, 998-1001.
Friml, J., Vieten, A., Sauer, M., Weijers, D., Schwarz, H., Hamann, T., Offringa, R. and Jürgens, G. (2003). Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. *Nature* **426**, 147-153.
Haecker, A., Gross-Hardt, R., Geiges, B., Sarkar, A., Breuninger, H., Herrmann, M. and Laux, T. (2004). Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. *Development* **131**, 657-668.
Haseloff, J. (1998). GFP variants for multispectral imaging of living cells. *Methods Cell Biol.* **58**, 139-151.
Heiman, M., Schaefer, A., Gong, S., Peterson, J. D., Day, M., Ramsey, K. E., Suárez-Fariñas, M., Schwarz, C., Stephan, D. A., Surmeier, D. J. et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738-748.
Hu, T.-X., Yu, M. and Zhao, J. (2010). Comparative transcriptional profiling analysis of the two daughter cells from tobacco zygote reveals the transcriptome differences in the apical and basal cells. *BMC Plant Biol.* **10**, 167.
Jayasinghe, S. M., Wunderlich, J., McKee, A., Newkirk, H., Pope, S., Zhang, J., Staehling-Hampton, K., Li, L. and Haug, J. S. (2006). Sterile and disposable fluidic subsystem suitable for clinical high speed fluorescence-activated cell sorting. *Cytometry B Clin. Cytom.* **70B**, 344-354.
Jaye, D. L., Bray, R. A., Gebel, H. M., Harris, W. A. C. and Waller, E. K. (2012). Translational applications of flow cytometry in clinical practice. *J. Immunol.* **188**, 4715-4719.
Kang, I.-H., Steffen, J. G., Portereiko, M. F., Lloyd, A. and Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant Cell* **20**, 635-647.
Kawashima, T. and Goldberg, R. B. (2010). The suspensor: not just suspending the embryo. *Trends. Plant Sci.* **15**, 23-30.
Kerk, N. M., Ceserani, T., Tausta, S. L., Sussex, I. M. and Nelson, T. M. (2003). Laser capture microdissection of cells from plant tissues. *Plant Physiol.* **132**, 27-35.
Khodosevich, K., Inta, D., Seeburg, P. H. and Monyer, H. (2007). Gene expression analysis of in vivo fluorescent cells. *PLoS ONE* **2**, e1151.
Kinoshita, T., Yadegari, R., Harada, J. J., Goldberg, R. B. and Fischer, R. L. (1999). Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. *Plant Cell* **11**, 1945-1952.
Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Fischer, R. L. and Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* **303**, 521-523.
Kroj, T., Savino, G., Valon, C., Giraudat, J. and Parcy, F. (2003). Regulation of storage protein gene expression in Arabidopsis. *Development* **130**, 6065-6073.
Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M. et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202-D1210.
Lau, S., Slane, D., Herud, O., Kong, J. and Jürgens, G. (2012). Early embryogenesis in flowering plants: setting up the basic body pattern. *Annu. Rev. Plant Biol.* **63**, 483-506.
Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S. et al. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. USA* **107**, 8063-8070.
Li, J., Nie, X., Tan, J. L. H. and Berger, F. (2013). Integration of epigenetic and genetic controls of seed size by cytokinin in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **110**, 15479-15484.
Liu, Y.-G. and Chen, Y. (2007). High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *Biotechniques* **43**, 649-656, 652, 654 passim.
Luo, M., Bilodeau, P., Dennis, E. S., Peacock, W. J. and Chaudhury, A. (2000). Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proc. Natl. Acad. Sci. USA* **97**, 10637-10642.
Nawy, T., Bayer, M., Mravec, J., Friml, J., Birnbaum, K. D. and Lukowitz, W. (2010). The GATA factor HANABA TARANU is required to position the proembryo boundary in the early Arabidopsis embryo. *Dev. Cell* **19**, 103-113.
Nodine, M. D. and Bartel, D. P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**, 94-97.
Palovaara, J., Saiga, S. and Weijers, D. (2013). Transcriptomics approaches in the early Arabidopsis embryo. *Trends Plant Sci.* **18**, 514-521.
Rademacher, E. H., Möller, B., Lokerse, A. S., Llavata-Peris, C. I., van den Berg, W. and Weijers, D. (2011). A cellular expression map of the Arabidopsis AUXIN RESPONSE FACTOR gene family. *Plant J.* **68**, 597-606.
Robert, H. S., Grones, P., Stepanova, A. N., Robles, L. M., Lokerse, A. S., Alonso, J. M., Weijers, D. and Friml, J. (2013). Local auxin sources orient the apical-basal axis in Arabidopsis embryos. *Curr. Biol.* **23**, 2506-2512.
Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E. H., Schmid, M., Jürgens, G. and Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* **464**, 913-916.
Shigenobu, S., Arita, K., Kitadate, Y., Noda, C. and Kobayashi, S. (2006). Isolation of germline cells from Drosophila embryos by flow cytometry. *Dev. Growth Differ.* **48**, 49-57.
Shin, R., Burch, A. Y., Huppert, K. A., Tiwari, S. B., Murphy, A. S., Guilfoyle, T. J. and Schachtman, D. P. (2007). The Arabidopsis transcription factor MYB77 modulates auxin signal transduction. *Plant Cell* **19**, 2440-2453.
Skirycz, A., Radziejwoski, A., Busch, W., Hannah, M. A., Czeszejko, J., Kwaśniewski, M., Zanor, M. I., Lohmann, J. U., De Veylder, L., Witt, I. et al. (2008). The DOF transcription factor OBP1 is involved in cell cycle regulation in Arabidopsis thaliana. *Plant J.* **56**, 779-792.

Development (2014) 141, 4831-4840 doi:10.1242/dev.116459

Spencer, M. W. B., Casson, S. A. and Lindsey, K. (2007). Transcriptional profiling of the Arabidopsis embryo. *Plant Physiol.* **143**, 924-940.

Stepanova, A. N., Robertson-Hoyt, J., Yun, J., Benavente, L. M., Xie, D.-Y., Doležal, K., Schlereth, A., Jürgens, G. and Alonso, J. M. (2008). TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* **133**, 177-191.

Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., Goldberg, R. B. and Harada, J. J. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci. USA* **98**, 11806-11811.

Takada, S. and Jürgens, G. (2007). Transcriptional regulation of epidermal cell fate in the *Arabidopsis* embryo. *Development* **134**, 1141-1150.

Ueda, M., Zhang, Z. and Laux, T. (2011). Transcriptional activation of Arabidopsis axis patterning genes WOX8/9 links zygote polarity to embryo development. *Dev. Cell* **20**, 264-270.

Webb, M. C. and Gunning, B. E. S. (1991). The microtubular cytoskeleton during development of the zygote, proembryo and free-nuclear endosperm in Arabidopsis thaliana (L.) Heynh. *Planta* **184**, 187-195.

Willmann, M. R., Mehalick, A. J., Packer, R. L. and Jenik, P. D. (2011). MicroRNAs regulate the timing of embryo maturation in Arabidopsis. *Plant Physiol.* **155**, 1871-1884.

Wu, Z., Irizarry, R. A., Robert, G., Martinez-Murillo, F. and Spencer, F. (2004). A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *J. Am. Stat. Assoc.* **99**, 909-917.

Yadav, R. K., Tavakkoli, M., Xie, M., Girke, T. and Reddy, G. V. (2014). A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. *Development* **141**, 2735-2744.

Zhang, C., Barthelson, R. A., Lambert, G. M. and Galbraith, D. W. (2008). Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. *Plant Physiol.* **147**, 30-40.

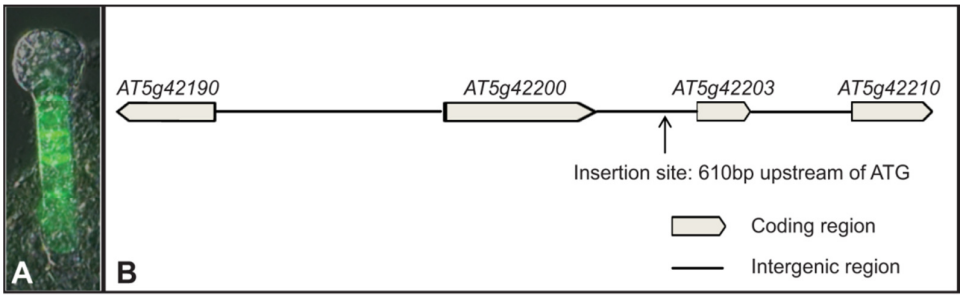

Figure S1. Enhancer-trap line N9322 and identification of genomic insertion site. (A) Suspensor and hypophysis expression at globular stage. (B) Insertion site of T-DNA determined by TAIL-PCR.

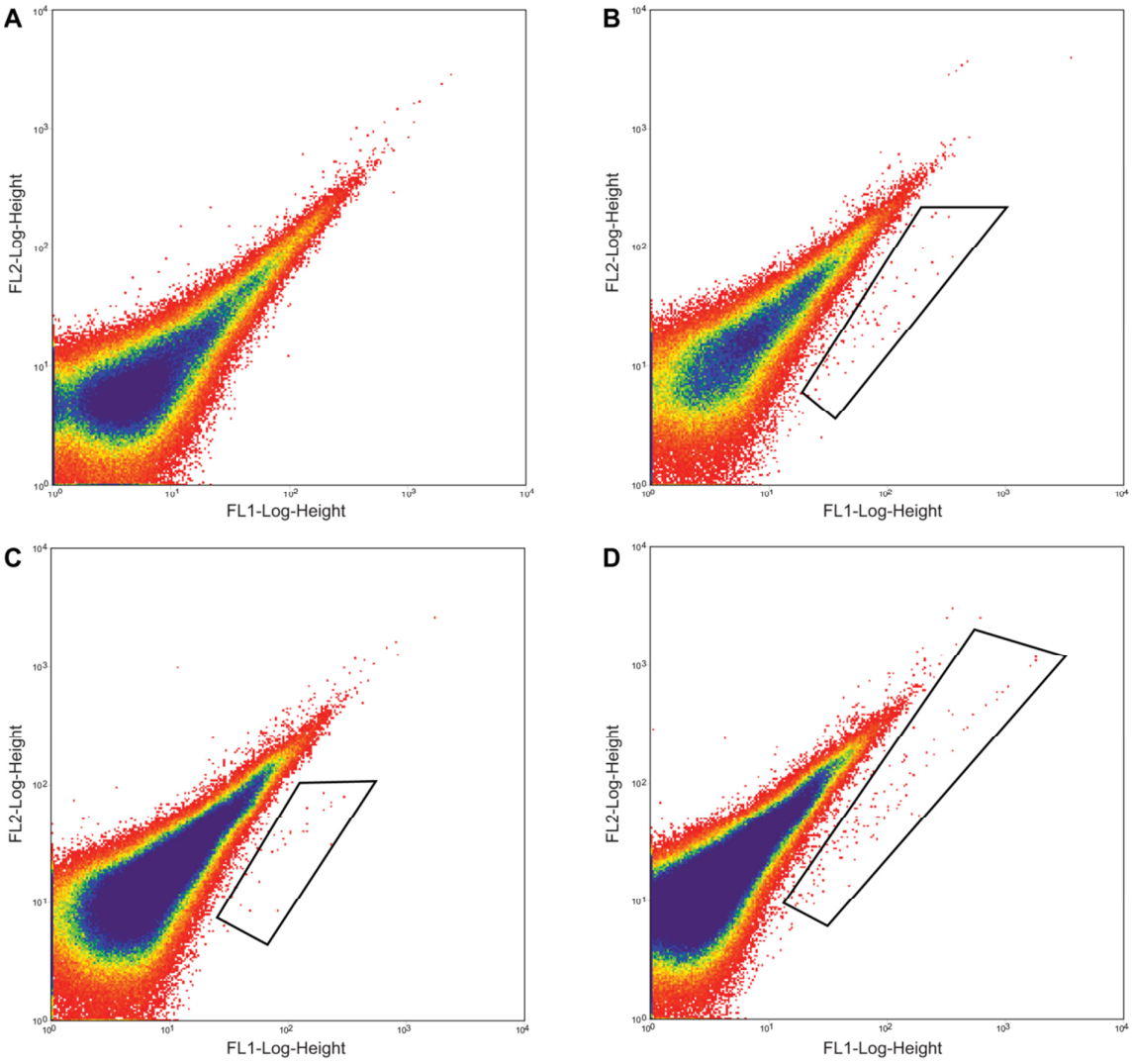

Figure S2. Scatter plots of FANS for GFP-tagged nuclear samples. (A) Mock sample. (B) Suspensor marker line *pAT5G42200:n3xGFP*. (C) Proembryo marker line *pDRN:n2xGFP:DRN 3'UTR*. (D) Whole embryo marker line *pAT3G10010:nGFP*. Fluorescent nuclei were detected by plotting the GFP channel (FL1, log, 513/17, x-axis) against auto-fluorescence (FL2, log, 575/25, y-axis) and drawing a gate around the GFP-positive events.

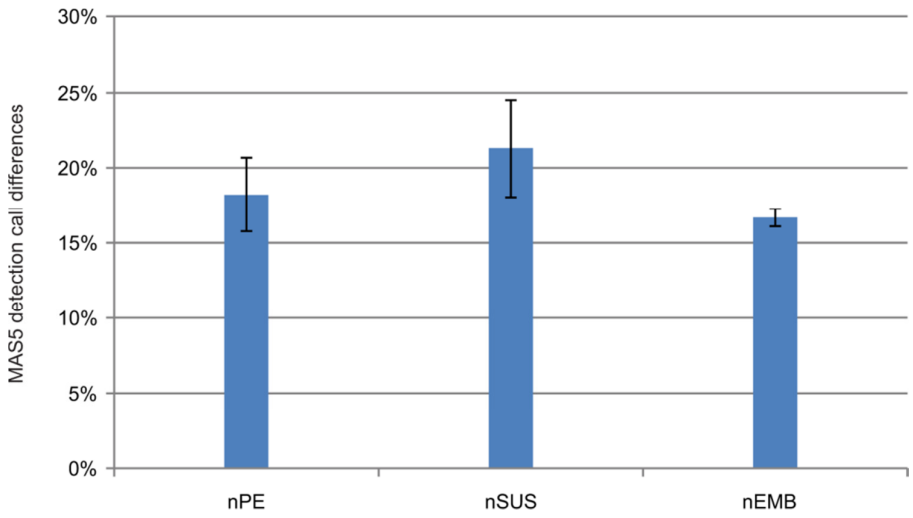

Figure S3. Percentage and standard deviation of MAS5 calls not correlating across three biological replicates. Replicates were compared to each other and the average percentage of calls not matching was calculated.

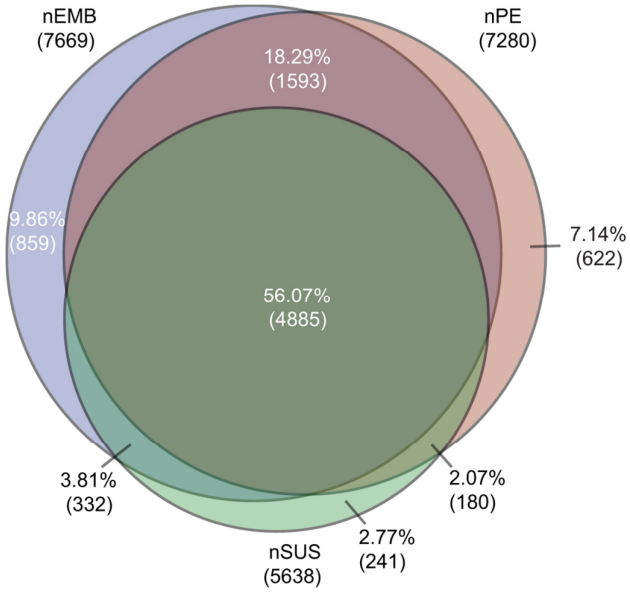

Figure S4. Venn diagram showing overlap of genes expressed in nuclei of the proembryo, suspensor, and whole embryo. For the analysis, only array elements with calls of 3x present (P) across all three biological replicates were used.

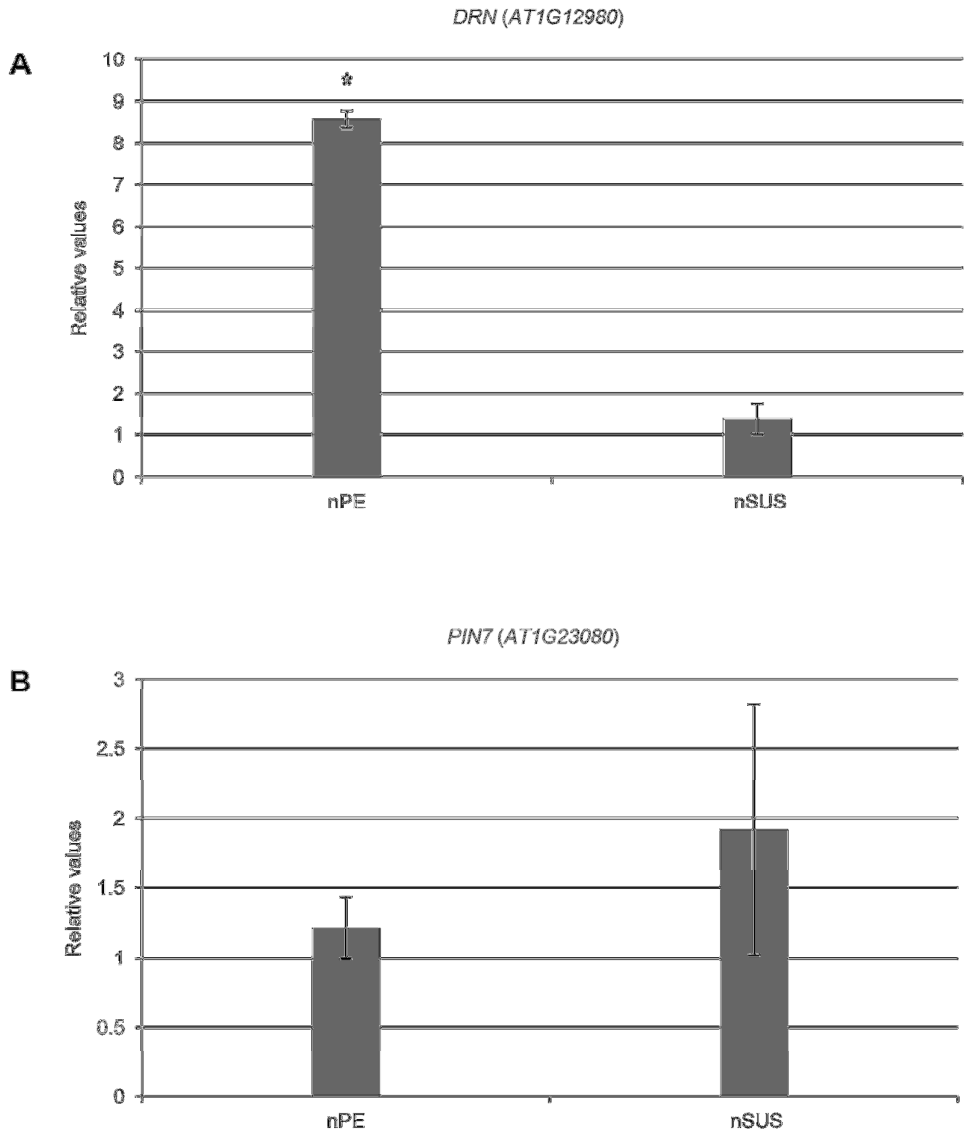

Figure S5. Relative mRNA levels detected by qRT-PCR analysis. (A) *PIN7* relative transcript levels are more abundant in nSUS compared to nPE. (B) *DRN* relative transcript levels are more abundant in nPE compared to nSUS. Average values and standard error are given for two biological replicates for nuclear RNA from both proembryo (nPE) and suspensor (nSUS). **P*<0.01 (Student's *t*-test).

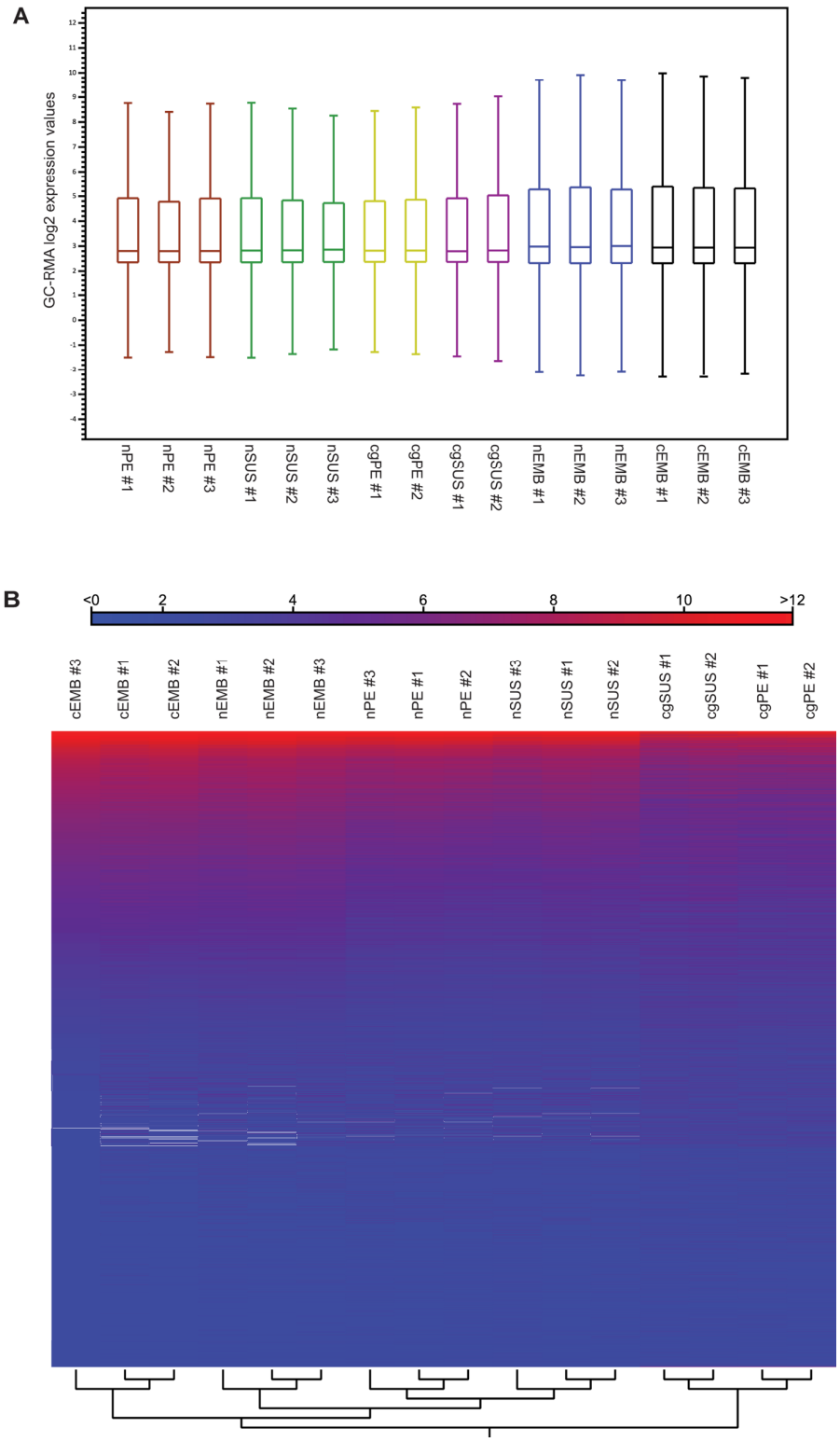

Figure S6. Qualitiy analysis of biological replicates from different nuclear and cellular tissue types. (A) Box-plot analysis. (B) Hierarchical clustering analysis (Pearson correlation, complete linkage).

Table S1. gcRMA normalized, log2 transformed values for nuclear embryo, cellular embryo, and LCM cellular samples.

Table S2. Pearson correlation coefficient (PCC) analysis of nSUS, nPE, nEMB and cKAN1.

Table S3. List of nPE statistically enriched genes. FC, fold change. Pfp, percentage of false positive prediction.

Table S4. List of nSUS statistically enriched genes. FC, fold change. Pfp, percentage of false positive prediction.

Table S5. GO term enrichment analysis of nPE-enriched genes. Color code corresponding to color code in bar graph Figure 2A.

Table S6. GO term enrichment analysis of nSUS-enriched genes. Color code corresponding to color code in bar graph Figure 2B.

Click here to Download Tables S1-S6

Table S7. Identification of transcripts expressed in both nEMB and cEMB based on MAS5 3xP detection calls. Average MAS5 expression values of the three replicates are given (decreasing values from red to blue).

Table S8. Identification of transcripts only expressed in nEMB compared to cEMB based on MAS5 3xP detection calls. Average MAS5 expression values of the three replicates are given (decreasing values from red to blue).

Table S9. Identification of transcripts only expressed in cEMB compared to nEMB based on MAS5 3xP detection calls. Average MAS5 expression values of the three replicates are given (decreasing values from red to blue).

Table S10. Expression levels of reportedly endosperm-specific genes in nSUS and cgSUS. gcRMA normalized and log2 transformed values for all replicates are given (decreasing values from red to blue).

Table S11. Identification of proembryo-specific genes by comparing nPE and cgSEED.

Table S12. Identification of suspensor-specific genes by comparing nSUS and cgSEED.

Table S13. Oligonucleotides used in this study and fragment length for each promoter, *in situ* probe, and qRT-PCR product.

Click here to Download Tables S7-S13