・**Corresponding authors**

Tsukaho Hattori

Graduate School of Bioagricultural Sciences

Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Phone: 052-789-5215. Fax: 052-789-5214

E-mail: hattori@agr.nagoya-u.ac.jp

Shin Takeda

Graduate School of Bioagricultural Sciences

Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Phone: 052-789-5205. Fax: 052-789-5214

E-mail: takeda@agr.nagoya-u.ac.jp

Yutaka Sato

National Institute of Genetics

1111 Yata, Mishima 411-8540, Japan

Phone: 055-981-6808. Fax: 055-981-6879

E-mail: yusato@nig.ac.jp

・**Subject area**

(4) Genetics/Developmental Biology

・**Number of tables: 1**

・**Black and white figures: 0**

・**Color figures: 6**

・**Names of authors: Yoshinori Takafuji[1], Sae Shimizu-Sato[2], Kim Nhung Ta[2], Toshiya Suzuki[2,3], Misuzu Nosaka-Takahashi[2,3], Tetsuro Oiwa[1], Wakana Kimura[1], Hirokazu Katoh[1], Mao Fukai[1], Shin Takeda[1], Yutaka Sato[2,3] and Tsukaho Hattori[1]**

・**Title: High-resolution Spatiotemporal Transcriptome Analyses during Cellularization of Rice Endosperm Unveil the Earliest Gene Regulation Critical for Aleurone and Starchy Endosperm Cell Fate Specification**

・**Authors' affiliations**

[1]Graduate school of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan

[2]National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

[3]Department of Genetics, School of Life Science, SOKENDAI (The Graduate University for Advanced Studies), 1111 Yata, Mishima, Shizuoka 411-8540, Japan

・**Corresponding authors' e-mail:**

hattori@agr.nagoya-u.ac.jp

takeda@agr.nagoya-u.ac.jp

yusato@nig.ac.jp

Abstract

The major tissues of the cereal endosperm are the starchy endosperm (SE) in the inner and the aleurone layer (AL) at the outer periphery. The fates of the cells that comprise these tissues are determined according to positional information; however, our understanding of the underlying molecular mechanisms remains limited. Here, we conducted a high-resolution spatiotemporal analysis of the rice endosperm transcriptome during early cellularization. In rice, endosperm cellularization proceeds in a concentric pattern from a primary alveolus cell layer, such that developmental progression can be defined by the number of cell layers. Using laser-capture microdissection to obtain precise tissue sections, transcriptomic changes were followed through five histologically defined stages of cellularization from the syncytial to 3-cell layer (3L) stage. In addition, transcriptomes were compared between the inner and the outermost peripheral cell layers. Large differences in the transcriptomes between stages and between the inner and the peripheral cells were found. SE attributes were expressed at the alveolus-cell-layer stage but were preferentially activated in the inner cell layers that resulted from periclinal division of the alveolus cell layer. Similarly, AL attributes started to be expressed only after the 2L stage and were localized to the outermost peripheral cell layer. These results indicate that the first periclinal division of the alveolus cell layer is asymmetric at the transcriptome level, and that the cell-fate-specifying positional cues and their perception system are already operating before the first periclinal division. Several genes related to epidermal identity (i.e., type IV homeodomain-leucine zipper genes and wax biosynthetic genes) were also found to be expressed at the syncytial stage, but their expression was localized to the outermost peripheral cell layer from the 2L stage onward. We believe that our findings significantly enhance our knowledge of the mechanisms underlying cell fate specification in rice endosperm.

Keywords: rice, aleurone, starchy endosperm, RNA-seq

Introduction

The development and germination of the plant embryo is supported by a tissue produced inside the seed called the endosperm. In cereals, the bulk of the endosperm is composed of starchy endosperm (SE), which is surrounded by a layer of cells called the aleurone layer (AL). Both of these tissues act as nutrient reserves for the embryo, with the SE storing starch and protein and the AL storing oils. AL cells remain alive even in the dry mature seed by acquiring desiccation tolerance through maturation programs similar to those that the embryo follows (Becraft 2007). In contrast, SE cells do not acquire desiccation tolerance and die via programed cell death and dehydration (Becraft 2007). In cereals, the AL can be a single layer of cells or it can contain multiple layers. Typically, the AL in maize and wheat is a single cell layered, in rice it is a single or double cell layered, and in barley it is a triple cell layered (Becraft 2007). Also, in cereals other than maize and related species, there is an exceptional region called modified AL in which AL-like cells constitute more cell layers (Royo et al. 2007). Modified AL is continuous with peripheral AL and is located in the vicinity of the major vein of the pericarp. On the basis of its location and cellular structures, modified AL is believed to be analogous to maize basal endosperm transfer layers, although maize basal endosperm transfer layers are recognized as distinct tissue from AL (Royo et al. 2007; Zhan et al. 2015).

In many plant species, including popular dicot and cereal crops as well as model plants such as those in genus Arabidopsis, the start of endosperm development is characterized by the primary triploid endosperm nucleus, which results from karyogamy of the two central nuclei and one of the sperm nuclei from the pollen tube, mitotically proliferating without cytokinesis to make a syncytium (Olsen 2001, 2004). This nuclear proliferation occurs at the periphery of the central vacuole and the nuclei align themselves in a single layer across the inside of the central cell wall. In rice, a syncytium comprising 4000 to 6000 nuclei is produced (Ishikawa et al. 2011; Hara et al. 2015) prior to the start of cellularization. Each nucleus in the syncytium constitutes an individual nuclear cytoplasmic domain that is delineated by a radial system of microtubules extending from the nucleus. Cellularization begins with the formation of cell walls that emanate from the wall of the original central cell between adjacent nuclear cytoplasmic domains. This cell

wall formation is mediated by cytoplasmic phragmoplasts and results in incompletely enclosed cells called alveoli that are open toward the central vacuole and have cytoplasmic continuity. The alveolar nuclei then undergo periclinal mitosis followed by regular cytokinesis, which results in the creation of a complete cell on the peripheral side and a second alveolus on the inner side. The second alveolus then divides similarly to produce another complete cell and alveolus. This process repeats until the central vacuole is entirely filled with complete endosperm cells. At the same time, the already-complete cells continue to divide in various planes, increasing the cell number (Hoshikawa 1967a; Brown et al. 1996b, c; Suzuki et al. 2000; Olsen 2001, 2004; Brown and Lemmon 2007).

The first periclinal division of the primary alveoli was recognized as a formative division that produced two layers of cells, each with a distinct cell fate; the inner alveoli and peripheral complete cells were thought to be SE and AL cell initials, respectively (Becraft 2001, 2007; Olsen 2001). This is because it was thought that the peripheral complete cells propagated only by anticlinal division to form a single-cell layered AL which is typical in maize, and that the inner alveoli only differentiated by periclinal division to form SE. However, the fates of these cells do not always bifurcate at the first periclinal division of the primary alveoli. The peripheral complete cells undergo further periclinal division and contribute to the internal SE cell mass (Hoshikawa 1967b; Brown et al. 1996c) and the inner daughter cells can differentiate into SE even after the fates of their parent cells have been specified or differentiated into AL (Morrison et al. 1975; Becraft and Asuncion-Crabb 2000; Gruis et al. 2006). That is, the peripheral complete cells resulting from periclinal division of the primary alveoli are the initials of both the SE and AL (Becraft 2007).

Several molecular genetic studies have examined the mechanism underlying the specification of cells for the AL lineage. In maize seeds harboring *defective kernel1* (*dek1*) mutation or in rice seeds harboring mutation in *DEK1* orthologue, *ADAXIALIZED LEAF 1,* SE instead of AL differentiates at the periphery of the SE (Becraft and Asuncion-Crabb 2000; Becraft et al. 2002; Lid et al. 2002; Hibara et al. 2009), suggesting that the proteins encoded by these genes are required for AL fate specification and maintenance.

Indeed, in a maize mutant harboring transposon-induced *dek1* mutant alleles, transposon excision resulted in reversion of the peripheral SE to AL, even at a late stage of endosperm development (Becraft and Asuncion-Crabb 2000). Conversely, induction of the *dek1* mutation in cells with a wild-type background caused conversion of AL to SE cells, even at a late stage of endosperm development (Becraft and Asuncion-Crabb 2000). Together, these observations indicate the plasticity of the endosperm cell fates and the existence of positional cues that specify the AL cell fate until the late developmental stage. These findings are supported by histological studies in wheat and endosperm culture studies in maize (Morrison et al. 1975; Gruis et al. 2006). Mutation of another maize gene, *crinkly4* (*cr4*), also results in the replacement of AL cells with SE cells, but in a patchy manner (Becraft et al. 1996). In contrast, mutation of the maize genes *SUPERNUMERARY ALEURONE LAYER 1* (*SAL1*), *NAKED ENDOSPERM* (*NKD*), AND *THICK ALEURONE1* (*THK1*) results in multiple cell layers at the periphery of the endosperm with full or partial AL characteristics (Shen et al. 2003; Becraft and Yi 2011; Yi et al. 2015). Interestingly, *thk1* is epistatic to *dek1* (Becraft et al. 1996). DEK1 and CR4 are a membrane-anchored plant calpain protein and a receptor kinase, respectively (Becraft et al. 1996; Lid et al. 2002), and SAL1 is a protein that resembles human CHARGED MULTIVESICULAR BODY PROTEIN 1A, which is possibly involved in vesicle trafficking (Shen et al. 2003). Taken together, these observations suggest that membrane-mediated signaling contributes to AL specification. THK1 is a NOT1 subunit of the CCR4–NOT complex that is involved in many aspects of regulation of gene expression, from the production of mRNAs to their degradation (Collart 2016; Wu et al. 2020). *NKD1* and *NKD2* are duplicate *INDETERMINATE DOMAIN* genes that encode transcription factors that regulate endosperm-specific gene expression, including activation of *OPAQUE2* and *VIVIPAROUS1 (VP1)* (Yi et al. 2015; Gontarek et al. 2016). Recently, it has been reported that in rice, dominant mutations in *THICK ALEURONE 2* (*TA2*) gene, *Ta2*, result in increased numbers of ALs. *TA2* has been shown to encode the DNA demethylase OsROS1 (Liu et al. 2018), indicating that epigenetic mechanisms are involved in endosperm cell fate specification and differentiation.

Several mutants that affect endosperm cell fate specification have also been reported, including *extra cell layers 1* (*xcl1*) (Kessler et al. 2002), *disorganized aleurone layer 1* (*dil1*) and *dil2* (Lid et al. 2004) in maize, and *elongation2 (elo2)* in barley (Lewis et al. 2009). The *xcl1* in maize results in a double-layered (instead of single-layered) AL and an increase in the number of epidermal layers. From their shared histological characteristics (e.g., surface position, single layer of cuboidal cells), a homologous relationship between AL and epidermis or protoderm has been suggested (Javelle et al. 2011). Furthermore, mutation not only of *xcl1* but also of *dek1*, *cr4*, *sal1*, and *thk1* results in abnormalities in leaf epidermis or embryo protoderm (Becraft et al. 1996, 2002; Shen et al. 2003; Hibara et al. 2009; Yi et al. 2011). This suggests that the cell fate specification mechanisms may be common to epidermis/protoderm and AL. However, direct mechanistic links between AL and epidermis/protoderm specification are yet to be explored.

Other studies examining mechanisms for endosperm cell fate specification have also been conducted. For example, Gruis et al. (2006) have suggested that it is surface position, not signals from the maternal tissues surrounding the endosperm (Jin et al. 2000; Geisler-Lee and Gallie 2005; Becraft and Yi 2011), that induces AL cell specification. Consistent with this, AL can be produced at the surface of internally formed cavity of cultured endosperm cells (Gruis et al. 2006). Consistent with this, when two independent endosperms are fused and their outermost layers are internalized, AL formation is suppressed along the fusion plane and instead SE differentiation is observed (Geisler-Lee and Gallie 2005). Hormonal regulation, especially by auxin and cytokinin, has also been suggested to be involved in AL and SE cell differentiation (Becraft and Yi 2011). Also, recent results of spaciotemporal analyses of transcriptomic changes during endosperm development have become available and are useful sources of data for understanding the mechanisms of endosperm cell differentiation (Drea et al. 2005; Nie et al. 2013; Chen et al. 2014; Ishimaru et al. 2015; Zhan et al. 2015; Qu et al. 2016). Despite these findings our understanding of cell fate specification and differentiation during endosperm development remains limited: the positional cues that induce the formation of SE and AL are unclear, little is known about how the first periclinal division of the

primary alveoli results in the initial cells of SE and AL, and it is unknown whether the first periclinal division of the primary alveoli is symmetric or asymmetric with respect to the transcriptome. Also, in most previous studies, cellularization has been regarded as a stage of endosperm development independent from differentiation of SE and AL cells and endosperm maturation. Therefore, the process of cellularization have not been studied from the view point of cellular differentiation. In this sense, the spatiotemporal resolution to fully elucidate the mechanisms of cell fate specification during endosperm differentiation including cellularization stage has not been enough.

Here, to obtain insights into endosperm cell specification for the development of AL and SE, we performed a detailed histological analysis of the early cellularization process in rice by using a confocal laser scanning microscope (CLSM). Then, we analyzed cell-layer-specific transcriptomic changes from the syncytial to 3-cell-layer stage by using a combination of laser-capture microdissection (LCM) and RNA-sequencing (RNA-seq) techniques. We took the advantages of the early cellularization phase in rice, where it proceeds in a concentric pattern, suggesting that developmental progression can be analyzed histologically by following the increase in the number of cell layers and cell division patterns. The data were analyzed by using novel AL- and SE-specific marker sets obtained by RNA-seq-based transcriptomic analysis using manually dissected AL and SE tissues at later developmental stage. Based on the high-resolution spatiotemporal transcriptomic and histological data obtained, we discuss the onset of cell specification and the differentiation of SE and AL in relation to the first periclinal division, as well as the relationship between AL and epidermal cell specification.

Materials and methods

Plant materials

Plants of wild-type rice (*Oryza sativa* L. cv. 'Nipponbare') were grown in a greenhouse. Transgenic rice

lines carrying the reporter genes *AGPL2 promoter-3x*(*NLS-Venus*) or *BT1-1 promoter-3x*(*NLS-Venus*) were grown in an isolated greenhouse at 30 °C under a 13-h light and 11-h dark photoperiod.

Histochemical analysis of rice endosperm

Ovary/caryopsis tissues were fixed with FAA, stained with propidium iodide, and cleared with methyl salicylate as described in (Hara et al. 2015). The prepared tissues were then cross-cut with a razor blade at the approximate median, and the cut surface was observed under CLSM (FV 3000 or FV 1000, Olympus, Tokyo, Japan; or LSM 5 PASCAL, Carl Zeiss, Heidelberg, Germany) at wavelengths of 559 nm/570–670 nm (excitation/emission).

For observation of the fluorescent reporter, the tissue samples were collected, cross-cut at the approximate median using a 76-μm razor blade (Nisshin EM, Tokyo, Japan), fixed in PBS containing 4 % paraformaldehyde, stained with -*Cellstain*-DAPI solution (Dojindo, Kumamoto, Japan), and cleared with TOMEI solution (Tokyo Chemical Industry, Tokyo, Japan). The cut surfaces were observed under CLSM at wavelengths of 488 nm/500–600 nm for Venus and 405 nm/430–470 nm for DAPI.

Laser-capture microdissection (LCM)

LCM was carried out by following the methods of Ishimoto et al. (2019). Ovary tissues were fixed in a solution of 75 % (v/v) ethanol and 25 % (v/v) glacial acetic acid, dehydrated in an ethanol series from 80 % (v/v) to absolute, and then embedded in paraffin. The paraffin-embedded tissues were screened by pilot sectioning for the desired developmental stages. Serial sections of the embedded tissues (12-μm thick) were collected on PEN membrane glass slides (Leica, Wetzlar, Germany) and dried at 42 °C. After drying, the slides were soaked twice in 100 % (v/v) Histo-Clear II solution (National Diagnostics, Charlotte, NC), deparaffinized, and dried at 4 °C. To allow for experimental replication, 6–10 serial sections of “whole” and “inner” tissue (for definition see later) were cut using an LMD 6000 Laser Microdissection system

(Leica, Wetzlar, Germany) and collected alternately into two separate tubes. Three (for the 2L(e) and 2L(l) stages) or four (for the 1L and 3L stages) independent ovary samples for each “whole” and “inner” tissue were used for RNA preparation. To evaluate the extent of nucellar RNA contamination in the 1L, 2L(e), and 2L(l) samples, whole endosperm tissue fragments intentionally contaminated with the inner edge of degenerating nucellus tissue (‘nucellus + whole’) were also prepared from approximately six serial sections. For comparison of transcripts at the syncytium or 1L stage, triplicate RNA samples were prepared from tissue fragments of the serial paraffin sections from the same ovary.

RNA extraction, cDNA synthesis, and transcriptome data analysis

SE and AL tissues were manually dissected from caryopses at 10 days after flowering (DAF) by using a razor blade (Kuwano et al. 2011); frozen in liquid nitrogen in a screw-capped, 2-mL tube with 3mm metal beads (As One, Osaka, Japan); and ground into a powder using a Mini-BeatBeader-S (Wakenyaku, Kyoto, Japan). Total RNA was then extracted by using an RNeasy Plant Mini Kit and QIAshredder (Qiagen, Valencia, CA), and used for cDNA synthesis with a QuantiTect Reverse Transcription Kit (Qiagen). The quality of the cDNA was verified by specific amplification of SE- and AL-specific genes by real-time reverse-transcription polymerase chain reaction using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA) and StepOne (Applied Biosystems, Waltham, MA). Total RNA extraction from the tissue fragments obtained by LCM, cDNA synthesis, and transcriptome data analyses were performed as described in Ishimoto et al. (2019).

In order to identify genes specifically expressed in AL or SE, we first enriched genes expressed in SE or AL by comparing RNA-seq data between these tissues at a false discovery rate with a threshold of $q < 0.05$ and log fold-change (logFC) > 2. Then, using the RiceXPro public database of gene expression profiles in rice (https://ricexpro.dna.affrc.go.jp/GGEP/), we selected genes specifically expressed in endosperm-containing tissues or embryo (ovary, endosperm, and embryo) compared with

other tissues were selected (logFC > 4, relative to the highest expression levels in other tissues); because some genes are known to be preferentially expressed in both embryo and AL (Zhan et al. 2015). SE- and AL- specific genes were identified by extracting the overlap between genes enriched in SE or AL, and genes specifically expressed in endosperm and in endosperm and embryo, respectively.

Differentially expressed genes (inner vs. whole) were selected with a false discovery rate threshold of $q < 0.05$ between expression levels in whole and inner tissues by pairwise *t*-test. Genes preferentially expressed in inner tissue (Inner DEGs) were selected using fold change > 1.5 (inner vs. whole); this corresponds to fold change > 3 (inner vs. outermost peripheral). Genes preferentially expressed in outermost peripheral tissue (Outermost peripheral DEGs) were selected using fold change < 0.5 (inner vs. whole); this corresponds to fold change < 3 (inner vs. outermost peripheral). Considering possible contamination by nucellus tissue in samples of whole tissues, genes with fold change > 2 ('nucellus + whole' vs. whole) for the 2L(l) stage data were excluded from the Outermost peripheral DEGs. For the 1L and 2L(e) stage data, genes with fold change > 2 ('nucellus + whole' vs. whole) and an expression value in 'whole' > 5 cpm (count per million) were excluded from the Outermost peripheral DEGs because the fold change of genes with low levels of expression can be overestimated. Transcript abundance (cpm) in outer tissue was estimated by using the formula:

$(peripheral) = 2 \times (whole) - (inner)$.

Gene Ontology (GO) term enrichment analysis was performed by using the Protein Analysis Through Evolutionary Relationships (PANTHER) Classification system (http://www.pantherdb.org/) with Fisher's Exact test (false discovery rate < 0.05) (Mi et al. 2020). The significance of enrichment of SE-specific genes in the Inner DEGs, and of AL-specific genes in the Outermost peripheral DEGs, was evaluated by using a hypergeometric distribution. The *P*-values calculated using the number of overlapping genes among SE- (or AL-) specific genes and the expected number of genes among randomly sampled gene sets were compared. Significance was then evaluated by calculating the ratio between the *P*-values. Hierarchical

clustering of RNA expression patterns based on Pearson's correlations by complete linkage clustering was performed by using Multiple Experiment Viewer software (v4.9.0; http://mev.tm4.org/) .

Plasmid construction of promoter–reporter fusion genes and transformation of rice

The promoter region of *AGPL2* was amplified from genomic DNA by PCR using appropriate primer sets (**Table S1**), digested with restriction enzymes, and inserted between the *Xba*I and *Sma*I sites of pPZP-2H-lac (Fuse et al. 2001). A reporter gene fragment carrying *3x*(*NLS-Venus*) was prepared from the MU110 plasmid (Ueda et al. 2011) by digestion with *Psh*AI and *Hind*III, and then inserted between the *Sma*I and *Hind*III sites of the pPZP-2H-lac binary vector carrying the *AGPL2* promoter. The promoter region of *BT1-1* was amplified by PCR using appropriate primer sets (**Table S1**), cloned into pENTR/D-TOPO (Thermo Fisher Scientific, Waltham, MA), and then inserted by LR reaction into a Gateway-based binary vector that had been prepared as follows: The *3x*(*NLS-Venus*) reporter fragment prepared from the MU110 plasmid by digestion with *Psh*AI and *Eco*RI was inserted into the *Aor*51HI site of pUGW2Δ35S (pUGW2-derived vector from which the 35S promoter region had been removed Nakagawa et al. 2007a). From the resulting plasmid, the reporter gene fragment containing site-specific recombination sites [*attR1-ccdB-attR2-3x*(*NLS-Venus*)] was prepared by digestion with *Xba*I and *Sac*I, and then inserted between *the Xba*I and *Sac*I sites of pGWB500 (Nakagawa et al. 2007b). The binary vectors were used for transformation of *Agrobacterium* (EHA101). *Agrobacterium*-mediated transformation of rice (cv. 'Nipponbare') was performed as described previously (Nishimura et al. 2007).

***In situ* hybridization**

In situ hybridization was performed as described previously (Kouchi and Hata 1993). DNA fragments for probe templates were cloned into pCR4-Blunt-TOPO (Invitrogen, Carlsbad, CA) vector after amplification by PCR from rice genomic DNA or cDNA derived from 10-DAF SE tissues using specific primer sets

(**Table S1**). Detailed methods are provided in the Supplementary Methods.

Results

Anatomy of the early cellularization stage in rice endosperm

To better understand the mechanism of cell fate specification and differentiation during rice endosperm development, we first analyzed the patterns of cell division and proliferation during the early stages of endosperm cellularization, paying special attention to cell layer formation at the periphery of the developing endosperm that is fated to become either AL or SE.

Using CLSM, we performed visual observation of cross-sections of ovaries collected at 68–84 h after flowering and stained with propidium iodide. The samples encompassed the syncytial to the completely acellularized stage. Syncytia containing both propagating and arrested nuclei was defined as the "growing syncytial" stage (Syn(g)) (**Fig. 1Aa**), and syncytia with the maximum number of free nuclei was defined as the "full syncytial" stage (Syn(f)) (**Fig. 1Ab**). The stage with a primary alveolus layer, which is the result of anticlinal cell wall formation after Syn(f), was defined as the single-layer stage (1L) (**Fig. 1Ac**). Periclinal division of the primary alveolus produced an inner alveolus and a peripheral complete cell, resulting in a double-layer stage (2L) (**Fig. 1Ad**, **e**). The periclinal wall was first formed at a position closer to the central vacuole so that the length (along the radial axis) of the peripheral daughter cell was longer than of the alveolus daughter (**Fig. 1Ad** and **E**). Both the peripheral cell and the inner alveolus elongated, but the rate of elongation in the alveolus was higher; thus, the size of the inner alveolus gradually exceeded that of the peripheral complete cell (**Fig. 1Ae** and **E**). Here, the stage where the length of the inner alveolus is equal to or less than that of the peripheral cell is defined as the early 2L stage (2L(e)) (**Fig. 1Ad** and **E**), and the stage where it is longer than the peripheral cell as the late 2L stage (2L(l)) (**Fig. 1Ae** and **E**). This division was well synchronized across the cell layer.

Late in, or following, the 2L(l) stage, synchronous periclinal nuclear divisions were usually observed in the vicinity of the open ends of the alveolus layer, as evidenced by synchronized mitotic chromosomes in the alveolus layer, indicating that a third cell-layer stage (3L) resulted from these periclinal divisions (**Fig. 1B** and **E**). However, in some cases, periclinal mitotic divisions at the peripheral cell layer preceded or occurred simultaneously with those in the inner alveoli following the 2L(l) stage (**Fig. 1Ae** and **E**). Periclinal mitotic nuclei were generally observed in the peripheral cells at the 3L stage to become the 4L stage (**Fig. 1Af** and **E**). Notably, the peripheral periclinal divisions appeared asymmetric in that the peripheral daughter nuclei were positioned very close to the peripheral wall, which resulted in longer inner and shorter peripheral daughter cells (**Fig. 1C** and **E**).

The fifth cell layer was found to be produced mainly by synchronous divisions of the alveoli, although periclinal nuclear divisions were occasionally found in the middle layers (**Fig. 1Ag** and **E**). Because of these occasional irregular patterns of divisions, as well as participation of anticlinal divisions, it became difficult to follow the putative lineage of cell layers after the stage with more than five cell layers (5L+) (**Fig. 1Ah** and **E**). Nonetheless, active periclinal division was observed to continue in both the peripheral and internal layers with comparable frequencies even up to 5 DAF, when 13–15 cell layers had formed and the central vacuole had been completely filled with layers of cells (**Fig. 1D**).

Together, these observations indicated that not only the periclinal divisions of the primary alveoli but also the peripheral periclinal divisions at the 2L stage through to 5 DAF contributed to the inner layers of cells that are fated to the SE lineage to an extent corresponding to the number of rounds of peripheral periclinal divisions. It is noteworthy that nuclear division, cell wall formation, cellular elongation, and periclinal division were observed to occur synchronously and concentrically within each layer, suggesting that endosperm development proceeds in a cell-layer-specific manner. However, developmental progression was obviously not synchronized among endosperm in relation to time after flowering (**Fig. S1**). Indeed, even at 84 h after flowering, around 25 % of the developing endosperm remained at the Syn(g)

stage while around 25 % of endosperm were at the 5L+ stage.

Identification of marker genes specifically expressed in AL or SE cells

A set of marker genes to discriminate between AL and SE cells would be useful for examining the differentiation of these two cell types. To produce a set of marker genes, we surgically isolated AL and SE from developing rice seeds at 10 DAF, and conducted an mRNA-seq analysis to identify genes specifically expressed in AL or SE cells (**Fig. 2**). In total, we detected expression of more than 20,000 genes in each of the two cell type (**Table S2**), among which 978 were preferentially expressed in SE and 3182 were preferentially expressed in AL. From these two sets of genes, we selected genes that were specifically expressed in endosperm-containing tissues or embryo by using the RiceXPro database (https://ricexpro.dna.affrc.go.jp) (Sato et al. 2013). As a result, we identified 105 putative SE-specific genes and 141 putative AL-specific genes (**Table S3**).

GO term enrichment analysis revealed that genes related to sugar reserves (glycogen biosynthetic process) and starch biosynthesis were significantly enriched among the 105 putative SE-specific genes (**Table S4**), whereas genes related to seed oil-body biosynthesis, lipid storage, response to abscisic acid (including response to freezing and water deprivation), and cold acclimation which is potentially mediated by abscisic acid signaling were enriched among the 141 putative AL-specific genes (**Table S4**). Based on the consistency of the results of the enrichment analysis with the known functions of AL and SE, and the fact that several known SE- or AL-specific genes (e.g., *AGPL2*, *Wx*, *BT1-1*, *NF-YB1*, and *OsVP1*) were included in the identified specifically expressed genes, we concluded that we had successfully isolated sets of tissue-specific genes.

mRNA-seq analysis of genes expressed in peripheral and inner cell layers of endosperm at the 2L(I) and 3L stages

Our present findings show that the early development of endosperm progresses largely synchronously in a cell-layer-specific manner. This suggested that isolation of total RNA from the peripheral or inner layers would allow us to examine the process of cellular differentiation to AL or SE as well as to elucidate the characteristics of the initial cells of these tissues. For this analysis, we collected samples of endosperm at the 2L(l) and 3L stages by LCM and used the samples for mRNA-seq analysis.

Even with LCM, it is not possible to isolate single layers of peripheral cells. Therefore, we used LCM to prepare two sets of tissue samples for the mRNA-seq analysis. The first set comprised whole endosperm at the 2L(l) and 3L stages and therefore included both the peripheral cell layer and the inner cell layer(s) (hereafter referred to as "whole"). The second set comprised only the inner cell layer(s) and was obtained by removing the peripheral cell layer (hereafter referred to as "inner"). We then isolated RNA from the tissue samples and conducted the mRNA-seq analysis. From the results of the analysis, we identified a set of differentially expressed genes (DEGs) between the "whole" and "inner" samples. From those genes, we selected genes that were preferentially expressed in the peripheral cell layer by choosing genes with an expression level at least two-fold higher in the "whole" samples than in the "inner" samples, and excluded genes potentially derived from nucellus RNA contamination; we refer to these genes as Outermost peripheral DEGs. For the genes expressed in the inner cell layer(s), we chose genes with an expression level at least 1.5 times higher in the "inner" samples than in the "whole" samples; we refer to these as Inner DEGs. These methods are summarized in **Fig. 2** and **Fig. S2A**.

In the analysis using endosperm at the 2L(l) and 3L stages, expression of more than 22,000 genes within each of the whole and inner samples was detected (**Table S2**). Of these genes, 334 were identified as Inner DEGs and 515 were identified as Outermost peripheral DEGs at the 2L(l) stage (**Fig. 2** and **Table S5**). GO term enrichment analysis of the Inner DEGs revealed significant enrichment of terms related to sugar reserve such as "glycogen biosynthetic process", "sucrose metabolic process", and "starch biosynthetic process" (**Table 1**). Of these three terms, two ("glycogen biosynthetic process" and "starch biosynthetic

process") were also significantly enriched in the SE-specific genes (**Table S4**), suggesting that at the 2L(l) stage the inner cell layers have a similar transcriptome to that of SE. In addition, many terms related to cell division ("regulation of microtubule cytoskeleton organization", "spindle assembly", "mitotic cell cycle phase transition", "regulation of cyclin-dependent protein serine/threonine kinase activity", "regulation of cell cycle process", and "cell division") were significantly enriched in the Inner DEGs but not in the Outermost peripheral DEGs (**Table 1**). This suggests that the mitotic state of the cells is different between the inner and outermost peripheral layers of the developing endosperm, and that the cell cycles of the inner and outermost peripheral cells are synchronized within each layer and they are at different phases. This is consistent with our anatomical observations that development of endosperm progresses in a cell-layer-specific manner. In contrast, GO term enrichment analysis of the Peripheral DEGs revealed significant enrichment of terms related to transportation of compounds ("organic hydroxy compound transport" and "transmembrane transport"), indicating that differentiation of the outermost peripheral cell layer is mediated by inter- and/or intra-cellular transport of metabolites/hormones. These terms could reflect the transportation function of AL, although differentiation to AL at this stage was not evident in the data obtained from the GO term enrichment analysis.

Next, to clarify whether the tissue-specific markers that we identified earlier from 10-DAF endosperm were expressed at the 2L(l) stage, we investigated the overlap between the identified sets of genes and found 9 genes common to both the SE-specific genes and Inner DEGs and 26 genes common to both the AL-specific genes and Outermost peripheral DEGs (**Fig. 3A** and **B**, **Table S3**, and **Table S5**). The numbers of common genes were 6- and 8-fold higher than those obtained by using randomly sampled gene sets ($p < 5.0 \times 10^{-5}$ and $p < 2.4 \times 10^{-16}$), respectively. These findings suggest that at the 2L(l) stage the inner and outermost peripheral cell layers share characteristics specific to SE and AL, respectively.

We then proceeded to check the expression profiles of several of these common genes (**Fig. 3C** and **D**). From the Inner DEGs, *AGPL2*, *Wx*, and *BT1-1*, which are known to be involved in starch biosynthesis

exclusively in SE (Jeon et al. 2010) were selected (**Table S3** and **Table S5**). As expected, these genes showed localized expression in the inner cell layers (**Fig. 3C**). From the Outermost peripheral DEGs, *NF-YB1*, *CYP71X14*, and *OsPRR12* were selected. *NF-YB1* is known to function and be expressed specifically in AL (**Table S3** and **Table S5**) (Xu et al. 2016). Although the functions of *CYP71X14* and *OsPRR12* in AL are unknown, they were two of the top six genes with highest transcript abundance among the Outermost peripheral DEGs at the 2L(l) stage (**Table S5**). As expected, these three genes showed localized expression in the outermost peripheral cell layers (**Fig. 3D**). In addition, the expression of *OsVP1*, a known key regulator of AL and embryo functions (Hattori et al. 1994, 1995; Miyoshi et al. 2002) , was also examined. Although this gene was not detected among the Outermost peripheral DEGs at the 2L(l) stage, probably due to a relatively low expression level at the 2L(l) stage and/or due to its function being exerted at later stages of AL development, we found that this gene was preferentially expressed in the outermost peripheral cell layer, albeit at a low level (**Fig. 3D**). These cell-layer expression profiles were also found at the 3L stage (**Fig. 3C** and **D**).

Next, we selected *AGPL2* and *BT1-1* as inner-cell-layer genes for spatial expression analysis. We constructed fluorescent protein reporter lines transformed with the promoters of *AGPL2* and *BT1-1* fused to *3x(NLS-Venus)* (Ueda et al. 2011) and observed the fluorescence signals of the reporter in developing endosperm at the 2L(l) stage. Nuclear-localized signals were detected preferentially in the cells of the inner cell layer, indicating that *AGPL2* and *BT1-1* are expressed predominantly in the inner cell layer (**Fig. 4A** and **B**). The inner-cell-layer-specific expression of *AGPL2* was further confirmed by *in situ* hybridization using endosperm at the 2L stage. (**Fig. 4C**). We also used *in situ* hybridization to examine the spatial expression of two genes preferentially expressed in the outermost peripheral cell layer, *CYP71X14* and *OsPRR12*, and confirmed their localized expression (**Fig. 4D** and **E**). As a control, we also examined the spatial expression of the gene *RFA2*, whose expression was detected both in the inner and outermost peripheral cell layers at comparable levels in the mRNA-seq analysis (**Fig. 4F**). As expected, *RFA2* mRNA

was detected in both cell layers (**Fig. 4G**). The cell-layer-specific expression of all of these genes was also found at the 3L stage (**Fig. S3**).

Estimation of the earliest timing of SE identity acquisition

To determine the timing of onset of SE cell attributes, we examined the expression of the earlier selected subset of SE-specific markers (i.e., *AGPL2*, *Wx*, and *BT1-1*) during early endosperm development (i.e., at stages Syn(f), 1L, 2L(e), 2L(l) and 3L) by means of mRNA-seq analysis using RNA prepared from tissues collected by LCM (**Fig. 5**, **Fig. S2**, and **Table S2**). Little to no expression of the three genes was detected at the Syn(f) stage. However, expression of these genes was clearly detected at the 1L stage, suggesting that at least part of the SE identity is already established in the primary alveolar cells. The expression levels of these genes markedly increased as endosperm development advanced. Also, whereas *AGPL2*, *Wx*, and *BT1-1* were expressed similarly in the inner and outermost peripheral cell layer(s) at the 2L(e) stage, the expression levels of these genes markedly increased in the inner layer(s) at the 2L(l) and 3L stages, with concomitant slight increases in the outermost peripheral layer.

Because AGPL2 protein functions by forming a complex with AGPS2 (Greene and Hannah 1998), we also investigated the expression of *AGPS2* in the mRNA-seq data. Although *AGPS2* is involved in starch synthesis, its expression is not unique to endosperm, so it was not included in the SE-specific genes (**Fig. S4A**). There are two mRNA products encoding two types of AGPS2, AGPS2a and AGPS2b, which differ in the presence or absence, respectively, of chloroplast transit peptide in their deduced amino acid sequences (Akihiro et al. 2005) (**Fig. S4B**). It is known that *AGPS2b* encodes a specific isoform of AGPS2 that functions only in SE, whereas *AGPS2a* encodes a different isoform that functions in tissues other than SE such as leaves and pollen (Lee et al. 2007). We found that the mRNA levels of *AGPS2* increased and expression was gradually localized to the inner layer(s) as the endosperm developed from the 1L to 3L stage, as was found for *AGPL2*, *Wx*, and *BT1-1* (**Fig. S4C**). However, *AGPS2* was obviously expressed at the

Syn(f) stage, whereas *AGPL2*, *Wx*, and *BT1-1* expression was very low or undetectable at this stage (**Fig. 5**).

To examine whether the *AGPS2* transcripts in the syncytium were *AGPS2a* or *AGPS2b*, we analyzed the mapping data obtained by mRNA-seq analysis (**Fig. S4D**). We found that the isoform that functions in tissues other than SE, *AGPS2a*, was uniquely expressed at the Syn(f) stage. In contrast, the isoform that functions in SE, *AGPS2b*, was detectable at the 1L and 2L(e) stages and expressed almost exclusively at the 2L(l) stage (**Fig. S4C** and **D**). This is a clear indication that SE-specific transcription and RNA processing has started by the 1L stage.

Estimation of the timing of AL-identity acquisition

To determine the timing of onset of AL cell attributes, we examined the expression of the earlier selected subset of AL-specific markers (i.e., *NF-YB1*, *OsVP1*, *CYP71X14*, and *OsPRR12*) by using the mRNA-seq data. Similar to the timing of initiation of differential expression of the subset of SE-specific genes (**Fig. 5**), marked increases in the gene expression of the AL markers were detected at the 2L(l) and 3L stages exclusively in the outermost peripheral layer (**Fig. 6A**). Whereas increases in the expression of *OsVP1* from 1L to 2L(e) were not evident, the expression levels of *NF-YB1*, *CYP71X14*, and *OsPRR12* were markedly higher at the 2L(e) stage than at the 1L stage. These variable expression patterns were in contrast to the consistent pattern of SE gene expression at the earlier 1L and 2L(e) stages.

To further clarify the timing of the start of AL-identity acquisition, we conducted hierarchical clustering of the set of 2L(l) Outermost peripheral DEGs, projected for the Syn(f), 1L 2L(e), 2L(l), and 3L stages (**Fig. 6B**), and identified two clusters of gene expression patterns. The genes in Cluster 1 showed significantly higher expression in the outermost peripheral layer at the 2L(l) and 3L stages, whereas they were scarcely expressed at the earlier Syn(f) and 1L stages. In contrast, those in Cluster 2 were persistently expressed after the Syn(f) stage and only showed a localized expression pattern in the outermost peripheral layer at

stages 2L(l) and 3L. As expected, the Outermost peripheral DEGs at 2L(l), including *NF-YB1*, *CYP71X14*, and *OsPRR12*, belonged to Cluster 1. Overall, the localized expression of genes in the outermost peripheral cell layer, including AL-identity genes, was evident at the 2L(l) stage, suggesting that the acquisition of AL identity occurs before then.

Next, because the genes in Cluster 2 were expressed before the 2L(l) stage, we examined whether these genes could potentially have functions in AL specification. Of the 26 common genes between the Outermost peripheral DEGs and AL-specific genes, 7 were included in Cluster 2 (**Fig. 6B**). The expression of many of these genes was detected from the Syn(f) stage through to the 3L stage, with the expression becoming localized to the peripheral cell layer at the 2L(l) stage (**Fig. 6B** and **Fig. S5**). Although the function of those genes with respect to AL specification or differentiation is unknown, their early expression indicates that at least a part of AL identity is established as early as the Syn(f) stage.

In addition to these seven genes, we also examined the *RICE OUTER CELL LAYER SPECIFIC* (*ROC*) genes. *ROC* genes belong to the *type IV HD-ZIP* family of transcription factor genes, most of which are expressed in epidermis, protoderm, and specific cell types derived from epidermal cells (Ito et al. 2002, 2003; Qing and Aoyama 2012). Among the *ROC* genes, *ROC1* and *ROC7* are co-orthologues of *Arabidopsis ATML1* and *PDF2* (**Fig. S6**), which are central regulators of epidermal cell differentiation and maintenance (Abe et al. 2001). In our transcriptome data, five *ROC* genes, including *ROC1* and *ROC7*, were classified into Cluster 2 and were significantly expressed at all stages from Syn(f) to 3L, with their outermost peripheral localization becoming evident at the 2L(l) stage (**Fig. 6B, C, D** and **Fig. S7A**). This demonstrates that the outermost peripheral endosperm cells share the epidermal identity and that this identity is established before the endosperm cell layer is formed. At the 2L(e) stage, *ROC1* and *ROC7* were expressed in both layers, although their expression was slightly higher in the outermost peripheral layer; however, by the 2L(l) stage, their expression was predominantly in the outermost peripheral layer (**Fig. 6D**). This temporal change in the expression of these epidermal factors in the two daughter cells just after periclinal

division is reminiscent of the changes in *ATML1* expression in protoderm in *Arabidopsis* embryo (Iida et al. 2019), which further supports the epidermal identity of the outermost peripheral cells of developing endosperm. Two other *ROC* genes (*ROC4* and *TF1*) were classified in Cluster 1; no expression of these genes was evident at the Syn(f) stage but their outermost peripherally localized expression was evident at the 2L(l) stage (**Fig. S7B**). Considering that autoregulation among *type IV HD-ZIP* genes operates in determining and maintaining epidermal identity (Tanaka et al. 2007) these two *ROC* genes could be controlled by *ROC* genes classified in Cluster 2, whose expression was observed at the Syn(f) and later stages (**Fig. S7A**).

There is one further line of evidence supporting the acquisition of epidermal character at the syncytium stage and maintenance of that character in the outermost peripheral endosperm cells. Four genes involved in wax and cuticle formation were found to be preferentially expressed in the outermost peripheral cell layer at the 2L(l) stage, and the expression patterns of two of these genes were similar to those of the *type IV HD-ZIP* genes *ROC1* and *ROC7* (**Fig. S8**). The four genes, *OsPAS2*, *ONION1* (*ONI1*), *ONION3* (*ONI3*), and *OsKCR1*, encode enzymes essential for fatty acid synthesis (3-hydroxy acyl-CoA dehydratase, β-ketoacyl CoA synthase, ω-alcohol dehydrogenase, and β-ketoacyl CoA reductase, respectively) (Bach et al. 2008; Beaudoin et al. 2009; Ito et al. 2011; Akiba et al. 2014). *ONI1* and *ONI3* and their orthologues in *Arabidopsis* are expressed specifically in the outer surface of many organs (Yephremov et al. 1999; Kurdyukov et al. 2006; Ito et al. 2011; Akiba et al. 2014). Deposition of hydrophobic cuticle made of waxes produced by those enzymes on the outer organ surface, which forms an impermeable protective film, is a characteristic feature of epidermal cells (Javelle et al. 2011). The high expression of these genes indicates that the peripheral cells of endosperm, which eventually become AL, have epidermal identity.

Together, these data show that the timing of AL-identity acquisition is somewhere between stages 2L(e) and 2L(l), judging from the expression of AL-specific genes such as *NF-YB1*. The data also show that the peripheral cells of endosperm acquire an identity and character that is similar to that of epidermal cells and

that this epidermal identity is acquired at the syncytium stage. Thus, the AL identity could be established by specialization of the epidermal cell lineage into AL cells, which would resemble the differentiation of other epidermis-derived cell types such as stomata and root hair cells.

Discussion

The cellular composition and arrangement of differentiated cells in endosperm are relatively simple; thus, the initial process of endosperm development is an ideal system in which to analyze the mechanisms of cell and tissue differentiation in plants (Olsen 2004). Extensive efforts have been made to understand the initial process of endosperm development, such as how and when AL and SE cell fates are specified and the cells differentiate; however, the fundamental mechanisms of cell fate specification and differentiation are largely unknown. We tackled this issue by combining precise anatomical observations and cell-layer-specific gene expression profiling analyses to examine the early stages of rice endosperm development. We also developed two sets of cell-type-specific genetic markers for rice AL and SE.

Our CLSM observation of endosperm cellularization at the early developmental stages from Syn(g) to 5L+ with particular focus on elucidating the pattern of periclinal division by inferring the most recent periclinal walls based on the appearance of periclinal mitotic chromosomes, revealed that, in rice endosperm development, frequent and synchronized periclinal divisions contribute to the cellularization process, thereby generating a concentric well-layered cellular arrangement (**Fig. S9**) (Hoshikawa 1967a; Brown et al. 1996a; Wu et al. 2016). Such a cellular arrangement is rather unique among the major cereals. In maize, the orientation of cell divisions in the inner cell layers is more random (Kiesselbach 2001). In barley and wheat, the arrangement of cells in the central part of the endosperm is complex due to the cells' dorso–ventral bipartite and flat structure and the presence of the large crease at the major vein side (Bosnes et al. 1992; Milligan et al. 2013; Leroux et al. 2014). In most plant cells undergoing cell division, the

preprophase band marks the future position of cytokinesis; however, it has been demonstrated that the preprophase band does not develop in inner endosperm cells undergoing cytokinesis (Brown et al. 1994, 1999; Olsen et al. 1995; Brown and Lemmon 2001). Thus, the regular angles that characterize the repeated periclinal division of inner endosperm cells in rice, which results in the observed concentrically layered structure, suggest the presence of an unknown mechanism that regulates the cell division plane independent of the preprophase band.

It is now known that the periclinal division of the peripheral alveoli does not represent formative division, in that the inner and peripheral daughter cells do not necessarily behave as SE and AL initials, respectively (Becraft 2001; Olsen 2001); however, the nature of this periclinal division remains unknown. In particular, whether the division is asymmetric or symmetric at the transcriptome level has not been elucidated or otherwise addressed, except that it is known that the preprophase band is involved in the subsequent division of the peripheral, but not the inner, daughter cells. Our present results clearly show that the periclinal division of the peripheral alveoli results in daughter cells with asymmetric transcriptomes. Furthermore, we were able to detect signs of differentiation between the inner and peripheral layer cells: at the 2L(I) stage, the transcriptomes of the inner and peripheral daughter cells showed strong bias toward the SE and AL, respectively. Together, these results indicate that the fates of these cells are specified in a position-dependent manner upon the first periclinal division.

Our present results also demonstrate that the outermost peripheral periclinal division continued until fairly late in endosperm development (Hoshikawa 1967a; Brown et al. 1996c; Wu et al. 2016). Such periclinal division at the periphery has not previously been described in other cereals, except for in the multi-layered aleurone of barley. Thus, in rice, the daughter cells resulting from peripheral periclinal division directly represent a substantial portion of the inner SE cells, which means that AL/SE cell-fating events occur continuously at the periphery during endosperm development from the very first periclinal division of the alveoli. In maize, the cellular identity of the peripheral layer can be reversed bidirectionally

between AL and SE, even at a late stage of development, by switching the *dek1* genotype from wild-type to mutant or vice versa (Becraft and Asuncion-Crabb 2000). Position-dependent redifferentiation upon peripheral periclinal division is also histologically recognized during normal development in wheat or in cell culture systems in maize (Morrison et al. 1975; Gruis et al. 2006). Such plasticity in the differentiation of AL/SE cells also suggests that a continuous fating mechanism operates throughout endosperm development in other cereals (Morrison et al. 1975; Becraft et al. 2002).

We consider the spatiotemporal resolution afforded by the conventional staging approach to describe the progression of endosperm development (i.e., syncytium, cellularization, differentiation, and maturation) (Wu et al. 2016) to be inadequate to elucidate the mechanisms of cell fate specification in endosperm development. Therefore, in the present study we conducted a spatiotemporally high-resolution transcriptomic analysis using histologically staged tissues from independent seeds and found not only that the gene expression profile undergoes cell-layer-dependent changes during the very early stage of cellularization but also that the progression of endosperm development in the very early stages varied considerably among ovaries sampled at the same time point. These findings indicate that it is not appropriate to depict rice endosperm development using absolute time after flowering, even though pollination occurs at almost the same time as flowering. In the present study, at a given time point, a fairly large proportion of ovaries remained at the Syn(g) and Syn(f) stages while others had developed much faster. Therefore, a process that limits the rate of development may exist before the alveolar stage; such a process may correspond to the phenomenon described as “mitotic hiatus” that precedes alveolar mitosis in other cereals (Brown et al. 1994).

Based on the present findings, we propose that a checkpoint may exist that controls the developmental procession beyond the syncytial stage. Grass endosperm is an organ that ensures embryo and seedling growth by storing maternal resources such as sources of carbon. Thus, the availability to translocate sugars from maternal tissue into the endosperm is pivotal for endosperm development. Such a checkpoint would

be necessary to ensure endosperm development because the amount of translocating sugars could differ between filial grains depending on the position in the panicle (Ishimaru et al. 2003). Indeed, it is reported that a high concentration of sucrose is a signal that induces *in vitro* endosperm development in maize (Gruis et al. 2006). In the present study, we found that cell division or the increase in the number of cell layers beyond the syncytial stage proceeded at a steady rate. We estimated that at all stages of development it takes about 6 h to add an additional cell layer, based on the observation that the ratio of the number of endosperm cells at the 1L and 2L stages was reversed between 66 and 72 HAF (**Fig. S1**).

The synchronous developmental progression across each cell layer in the early cellularization stages in rice will be of great value for conducting cell-layer-specific transcriptome analysis using RNA extracted from tissues isolated by LCM. Our transcriptomic data clearly show that there is a discrete transition from non-SE to SE-type gene expression, leading to establishment of the SE cell fate beyond the syncytial stage. It would not be surprising if such a switch in gene expression was controlled by sugar-mediated signaling, because starch synthesis requires sugars supplied from the mother tissue (Olsen 2020). However, sugar-mediated signaling would not be the sole factor leading to the switch from non-SE to SE gene expression because many of the starch synthesis enzymes known to respond to sugars were found not to be expressed at the 1L stage. This suggests that additional layers of regulation exist (e.g., epigenetic regulation). Indeed, in rice, *ta2* mutation results in defective expression of the DNA demethylase *OsROS1*, which is necessary for releasing repression of gene expression by DNA methylation or imprinting of genes in the female gamete and early endosperm (Gehring and Satyaki 2017). The increased aleurone cell layers of *ta2* in rice is consistent with the involvement of DNA-demethylation-mediated gene expression in the specification of SE cells.

It is often the case that there is a cue to trigger differentiation to one cell type, whereas without the cue another cell type develops as if the latter is the default cellular fate. In this sense, our finding that SE attributes are expressed at the 1L stage before AL and SE cells are specified is in accordance with the

proposition that SE is the default fate of endosperm cells, as reported from genetic studies and cell culture studies in maize (Olsen 2020). In contrast, for AL, because we have not yet identified a set of genes that comprehensively represent the AL identity, like the SE-specific starch synthesis genes for SE identity, it is difficult to examine the onset of the AL identity. However, we found seven AL-specific genes that were also expressed in the peripheral tissue at the syncytial stage (**Fig. S5**). The expression of these seven genes was not detected in tissues other than endosperm, based on an examination of the RiceXPro database, suggesting their specific involvement in AL function. If this is the case, then at least part of the AL character is expressed as early as the syncytium stage.

Several lines of evidence, including our present observations, suggest that positional information guides the outermost peripheral cell layer to the AL fate (Becraft and Asuncion-Crabb 2000; Gruis et al. 2006). In addition, the continuing plasticity even after AL/SE differentiation (Becraft and Asuncion-Crabb 2000; Gruis et al. 2006), suggests that this positional information must persist throughout the cellularization process. Thus, this positional information and the system to perceive this information must rely on mechanisms intrinsic to cellular processes or fundamental to plant development. One possibility for such a mechanism for AL cell fate specification could be acquisition of the epidermal identity. We found that many members of the *type IV HD-ZIP* transcriptional factor family, including those that dictate epidermal identity, are persistently expressed from the Syn(f) stage but become asymmetrically expressed in the outermost peripheral cell layer. Furthermore, some genes encoding fatty acid biosynthetic enzymes such as *OsPAS2*, *ONI1*, *ONI3*, and *OsKCR1*, which are essential for wax production, a known epidermal attribute, were found to be coregulated with these *type IV HD-ZIP* genes. These findings support the idea that acquisition of epidermal identity is a prerequisite for the AL fate. Epidermal identity is also the basis for differentiation of many specialized cell types such as trichoblasts, root hair cells, and stomata lineage cells (Larkin et al. 2003). Indeed, differentiation of these cell types is affected by the expression of *type IV HD-ZIP* transcriptional factors (Qing and Aoyama 2012; Takada et al. 2013).

Although the *type IV HD-ZIP* transcriptional factors and some of the AL-specific genes were co-expressed at the Syn(f) stage, we did not detect any sign of gene expression linked to known AL attributes at this stage. This suggests that AL differentiation proceeds in a stepwise manner that requires additional developmental instructors. This is in accordance with the fact that the aleurone specific gene expression such as *VACUOLAR H^+-TRANSLOCATING INORGANIC PYROPHOSPHATASE* and several AL-characters are acquired in a stepwise manner in maize (Wisniewski and Rogowsky 2004). It is intriguing to speculate that signaling of such an additional instructor may require the activity of DEK1. The stepwise acquisition of cellular identity also operates in SE cell differentiation, because, the expression of SE-specific starch biosynthetic genes, which evoke SE identity, at earlier stages represent only a fraction of the known SE attributes. In fact, essentially no expression of genes encoding endosperm storage proteins such as glutelins and prolamins was detected during the analyzed stages. Thus, additional developmental factors may be required for SE differentiation that advance the degree of differentiation but allow the reversible state to persist. Taking these findings together, we propose that cellular differentiation mechanisms in rice endosperm operate through integration of both positional and developmental cues that allow for cell differentiation while retaining plasticity through to a certain developmental stage.

Acknowledgements

We thank Drs. M. Ueda and T. Nakagawa for their gifts of the plasmid carrying *3x*(*NLS-Venus*) (MU110) and the plasmids used for the preparation of the Gateway-based vectors (pGW2Δ35S and pGWB500), respectively. This work was supported by JSPS KAKENHI (19K05968, 17H06471, and 20H00424) and by the National Institute of Genetics (NIG)-JOINT (54A2018, 12A2019, and 41A2020).

Compliance with ethical standards

Conflict of interest: The authors have no conflict of interest to declare.

References

Abe M, Takahashi T, Komeda Y (2001) Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. Plant J 26:487–494. https://doi.org/10.1046/j.1365-313X.2001.01047.x

Akiba T, Hibara KI, Kimura F, et al (2014) Organ fusion and defective shoot development in oni3 mutants of rice. Plant Cell Physiol 55:42–51. https://doi.org/10.1093/pcp/pct154

Akihiro T, Mizuno K, Fujimura T (2005) Gene expression of ADP-glucose pyrophosphorylase and starch contents in rice cultured cells are cooperatively regulated by sucrose and ABA. Plant Cell Physiol 46:937–946. https://doi.org/10.1093/pcp/pci101

Bach L, Michaelson L V., Haslam R, et al (2008) The very-long-chain hydroxy fatty acyl-CoA dehydratase PASTICCINO2 is essential and limiting for plant development. Proc Natl Acad Sci U S A 105:14727–14731. https://doi.org/10.1073/pnas.0805089105

Beaudoin F, Wu X, Li F, et al (2009) Functional characterization of the Arabidopsis β-ketoacyl-coenzyme a reductase candidates of the fatty acid elongase. Plant Physiol 150:1174–1191. https://doi.org/10.1104/pp.109.137497

Becraft PW (2007) Aleurone cell development. Plant Cell Monogr 8:45–56. https://doi.org/10.1007/7089_2007_108

Becraft PW (2001) Cell fate specification in the cereal endosperm. Semin Cell Dev Biol 12:387–394. https://doi.org/10.1006/scdb.2001.0268

Becraft PW, Asuncion-Crabb Y (2000) Positional cues specify and maitain aleurone cell fate in maize endosperm development. Development 127:4039–4048. https://doi.org/10.1007/978-1-4020-8854-

4_27

Becraft PW, Li K, Dey N, Asuncion-Crabb Y (2002) The maize dek1 gene functions in embryonic pattern formation and cell fate specification. Development 129: 5217-5225.

Becraft PW, Stinard PS, McCarty DR (1996) CRINKLY4: A TNFR-Like Receptor Kinase involved in maize epidermal differentiation. Science (80-) 273:1406–1409. https://doi.org/10.1126/science.273.5280.1406

Becraft PW, Yi G (2011) Regulation of aleurone development in cereal grains. J Exp Bot 62:1669–1675. https://doi.org/10.1093/jxb/erq372

Bosnes M, Weideman F, Olsen OA (1992) Endosperm differentiation in barley wild-type and sex mutants. Plant J 2:661–674. https://doi.org/10.1111/j.1365-313X.1992.tb00135.x

Brown R, Lemmon B, Olsen O (1996a) Development of the endosperm in rice (Oryza sativa L.): Cellularization. J Plant Res 36:301–313. https://doi.org/10.1007/BF02344477

Brown RC, Lemmon BE (2007) The developmental biology of cereal endosperm. Plant Cell Monogr 8:1–20. https://doi.org/10.1007/7089_2007_106

Brown RC, Lemmon BE (2001) The cytoskeleton and spatial control of cytokinesis in the plant life cycle. Protoplasma 215:35–49. https://doi.org/10.1007/BF01280302

Brown RC, Lemmon BE, Nguyen H, Olsen OA (1999) Development of endosperm in Arabidopsis thaliana. Sex Plant Reprod 12:32–42. https://doi.org/10.1007/s004970050169

Brown RC, Lemmon BE, Olsen OA (1996b) Polarization predicts the pattern of cellularization in cereal endosperm. Protoplasma 192:168–177. https://doi.org/10.1007/BF01273889

Brown RC, Lemmon BE, Olsen OA (1994) Endosperm development in barley: Microtubule involvement in the morphogenetic pathway. Plant Cell 6:1241–1252. https://doi.org/10.1105/tpc.6.9.1241

Brown RC, Lemmon BE, Olsen OA (1996c) Development of the endosperm in rice (Oryza sativa L.): Cellularization. J Plant Res 109:301–313. https://doi.org/10.1007/bf02344477

Chen J, Zeng B, Zhang M, et al (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. Plant Physiol 166:252–264. https://doi.org/10.1104/pp.114.240689

Collart MA (2016) The Ccr4-Not complex is a key regulator of eukaryotic gene expression. Wiley Interdiscip Rev RNA 7:438–454. https://doi.org/10.1002/wrna.1332

Drea S, Leader DJ, Arnold BC, et al (2005) Systematic spatial analysis of gene expression during wheat caryopsis development. Plant Cell 17:2172–2185. https://doi.org/10.1105/tpc.105.034058

Fuse T, Sasaki T, Yano M (2001) Ti-plasmid vectors useful for functional analysis of rice genes. Plant Biotechnol. 18:219–222. http://dx.doi.org/10.5511/plantbiotechnology.18.219

Gehring M, Satyaki PR (2017) Endosperm and imprinting, inextricably linked. Plant Physiol 173:143–154. https://doi.org/10.1104/pp.16.01353

Geisler-Lee J, Gallie DR (2005) Aleurone cell identity is suppressed following connation in maize kernels. Plant Physiol 139:204–212. https://doi.org/10.1104/pp.105.064295

Gontarek BC, Neelakandan AK, Wu H, Becraft PW (2016) NKD transcription factors are central regulators of maize endosperm development. Plant Cell 28:2916–2936. https://doi.org/10.1105/tpc.16.00609

Greene TW, Hannah LC (1998) Maize endosperm ADP-glucose pyrophosphorylase SHRUNKEN2 and

BRITTLE2 subunit interactions. Plant Cell 10:1295–1306. https://doi.org/10.1105/tpc.10.8.1295

Gruis D, Guo H, Selinger D, et al (2006) Surface position, not signaling from surrounding maternal tissues, specifies aleurone epidermal cell fate in maize. Plant Physiol 141:898–909. https://doi.org/10.1104/pp.106.080945

Hara T, Katoh H, Ogawa D, et al (2015) Rice SNF2 family helicase ENL1 is essential for syncytial endosperm development. Plant J 81:1–12. https://doi.org/10.1111/tpj.12705

Hattori T, Terada T, Hamasuna S (1995) Regulation of the Osem gene by abscisic acid and the transcriptional activator VP1: analysis of cis-acting promoter elements required for regulation by abscisic acid and VP1. Plant J. 7:913–925

Hattori T, Terada T, Hamasuna ST (1994) Sequence and functional analyses of the rice gene homologous to the maize Vp1. Plant Mol Biol 24:805–810. https://doi.org/10.1007/BF00029862

Hibara K, Obara M, Hayashida E, et al (2009) The ADAXIALIZED LEAF1 gene functions in leaf and embryonic pattern formation in rice. Dev Biol 334:345–354. https://doi.org/10.1016/j.ydbio.2009.07.042

Hoshikawa K (1967a) Studies on the development of endosperm in rice : 1. Process of endosperm tissue formation. Japanese J Crop Sci 36:151–161. https://doi.org/10.1626/jcs.36.151

Hoshikawa K (1967b) Studies on the Development of Endosperm in Rice : 4. Differentiation and development of the aleurone lyaer. Jap J Crop Sci 216–220

Iida H, Yoshida A, Takada S (2019) ATML1 activity is restricted to the outermost cells of the embryo through post-transcriptional repressions. Dev 146:. https://doi.org/10.1242/dev.169300

Ishikawa R, Ohnishi T, Kinoshita Y, et al (2011) Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. Plant J 65:798–806. https://doi.org/10.1111/j.1365-313X.2010.04466.x

Ishimaru T, Ida M, Hirose S, et al (2015) Laser microdissection-based gene expression analysis in the aleurone layer and starchy endosperm of developing rice caryopses in the early storage phase. Rice 8:1–15. https://doi.org/10.1186/s12284-015-0057-2

Ishimaru T, Matsuda T, Ohsugi R, Yamagishi T (2003) Morphological development of rice caryopses located at the different positions in a panicle from early to middle stage of grain filling. Funct Plant Biol 30:1139–1149. https://doi.org/10.1071/FP03122

Ishimoto K, Sohonahra S, Kishi-Kaboshi M, et al (2019) Specification of basal region identity after asymmetric zygotic division requires mitogen-activated protein kinase 6 in rice. Dev 146:. https://doi.org/10.1242/dev.176305

Ito M, Sentoku N, Nishimura A, et al (2002) Position dependent expression of gl2-type homeobox gene, roc1: Significance for protoderm differentiation and radial pattern formation in early rice embryogenesis. Plant J 29:497–507. https://doi.org/10.1046/j.1365-313x.2002.01234.x

Ito M, Sentoku N, Nishimura A, et al (2003) Roles of rice GL2-type homeobox genes in epidermis differentiation. Breed Sci 53:245–253. https://doi.org/10.1270/jsbbs.53.245

Ito Y, Kimura F, Hirakata K, et al (2011) Fatty acid elongase is required for shoot development in rice. Plant J 66:680–688. https://doi.org/10.1111/j.1365-313X.2011.04530.x

Javelle M, Vernoud V, Rogowsky PM, Ingram GC (2011) Epidermis: the formation and functions of a fundamental plant tissue. New Phytol 189:17–39. https://doi.org/10.1111/j.1469-8137.2010.03514.x

Jeon JS, Ryoo N, Hahn TR, et al (2010) Starch biosynthesis in cereal endosperm. Plant Physiol Biochem 48:383–392. https://doi.org/10.1016/j.plaphy.2010.03.006

Jin P, Guo T, Becraft PW (2000) The maize CR4 receptor-like kinase mediates a growth factor-like differentiation response. genesis 27:104–116. https://doi.org/10.1002/1526-968X(200007)27:3<104::AID-GENE30>3.0.CO;2-I

Kessler S, Seiki S, Sinha N (2002) Xcl1 causes delayed oblique periclinal cell divisions in developing maize leaves, leading to cellular differentiation by lineage instead of position. Development 129:1859–1869

Kiesselbach T (2001) The structure and reproduction of corn. Biol Plant 44:238–238. https://doi.org/10.1023/a:1010248820060

Kouchi H, Hata S (1993) Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. MGG Mol Gen Genet 238:106–119. https://doi.org/10.1007/BF00279537

Kurdyukov S, Faust A, Trenkamp S, et al (2006) Genetic and biochemical evidence for involvement of HOTHEAD in the biosynthesis of long-chain α-,ω-dicarboxylic fatty acids and formation of extracellular matrix. Planta 224:315–329. https://doi.org/10.1007/s00425-005-0215-7

Kuwano M, Masumura T, Yoshida KT (2011) A novel endosperm transfer cell-containing region-specific gene and its promoter in rice. Plant Mol Biol 76:47–56. https://doi.org/10.1007/s11103-011-9765-1

Larkin JC, Brown ML, Schiefelbein J (2003) How do cells know what they want to be when they grow up? Lessons from epidermal patterning in arabidopsis. Annu Rev Plant Biol 54:403–430. https://doi.org/10.1146/annurev.arplant.54.031902.134823

Lee SK, Hwang SK, Han M, et al (2007) Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (Oryza sativa L.). Plant Mol Biol 65:531–546. https://doi.org/10.1007/s11103-007-9153-z

Leroux BM, Goodyke AJ, Schumacher KI, et al (2014) Maize early endosperm growth and development: Rom fertilization through cell type. Am J Bot 101:1259–1274. https://doi.org/10.3732/ajb.1400083

Lewis D, Bacic A, Chandler PM, Newbigin EJ (2009) Aberrant cell expansion in the elongation mutants of barley. Plant Cell Physiol 50:554–571. https://doi.org/10.1093/pcp/pcp015

Lid SE, Al RH, Krekling T, et al (2004) The maize disorganized aleurone layer 1 and 2 [dil1, dil2] mutants lack control of the mitotic division plane in the aleurone layer of developing endosperm. Planta 218:370–378. https://doi.org/10.1007/s00425-003-1116-2

Lid SE, Gruis D, Jung R, et al (2002) The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. Proc Natl Acad Sci 99:5460–5465. https://doi.org/10.1073/pnas.042098799

Liu J, Wu X, Yao X, et al (2018) Mutations in the DNA demethylase OsROS1 result in a thickened aleurone and improved nutritional value in rice grains. Proc Natl Acad Sci U S A 115:11327–11332. https://doi.org/10.1073/pnas.1806304115

Mi H, Ebert D, Muruganujan A, et al (2020) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive. Nucleic Acids Res 1–10. https://doi.org/10.1093/nar/gkaa1106

Milligan AS, Lopato S, Kovalchuk N, Langridge P (2013) Functional Genomics of Seed Development in Cereals. In Gupta P and Varshney R (ed) Cereal Genomics II. Springer Netherlands, Dordrecht,

pp215-245. https://doi.org/10.1007/978-94-007-6401-9

Miyoshi K, Kagaya Y, Ogawa Y, et al (2002) Temporal and spatial expression pattern of the OSVP1 and OSEM genes during seed development in rice. Plant Cell Physiol 43:307–313. https://doi.org/10.1093/pcp/pcf040

Morrison IN, Kuo J, O'Brien TP (1975) Histochemistry and fine structure of developing wheat aleurone cells. Planta 123:105–116. https://doi.org/10.1007/BF00383859

Nakagawa T, Kurose T, Hino T, et al (2007a) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng 104:34–41. https://doi.org/10.1263/jbb.104.34

Nakagawa T, Suzuki T, Murata S, et al (2007b) Improved gateway binary vectors: High-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci Biotechnol Biochem 71:2095–2100. https://doi.org/10.1271/bbb.70216

Nie DM, Ouyang YD, Wang X, et al (2013) Genome-wide analysis of endosperm-specific genes in rice. Gene 530:236–247. https://doi.org/10.1016/j.gene.2013.07.088

Nishimura A, Aichi I, Matsuoka M (2007) A protocol for Agrobacterium-mediated transformation in rice. Nat Protoc 1:2796–2802. https://doi.org/10.1038/nprot.2006.469

Olsen OA (2001) Endosperm development: Cellularization and cell fate specification. Annu Rev Plant Biol 52:233–267. https://doi.org/10.1146/annurev.arplant.52.1.233

Olsen OA (2004) Nuclear endosperm development in cereals and Arabidopsis thaliana. Plant Cell 16:214–228. https://doi.org/10.1105/tpc.017111

Olsen OA (2020) The Modular Control of Cereal Endosperm Development. Trends Plant Sci 25:279–290. https://doi.org/10.1016/j.tplants.2019.12.003

Olsen OA, Brown RC, Lemmon BE (1995) Pattern and process of wall formation in developing endosperm. BioEssays 17:803–812. https://doi.org/10.1002/bies.950170910

Qing L, Aoyama T (2012) Pathways for epidermal cell differentiation via the homeobox gene GLABRA2: update on the roles of the classic egulator. J Integr Plant Biol 54:729–737. https://doi.org/10.1111/j.1744-7909.2012.01159.x

Qu J, Ma C, Feng J, et al (2016) Transcriptome dynamics during maize endosperm development. PLoS One 11:1–22. https://doi.org/10.1371/journal.pone.0163814

Royo J, Gómez E, Hueros G (2007) Transfer Cells. In: Olsen OA (ed) Endosperm. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 73–89. http://doi.org/10.1007/7089_2007_110

Sato Y, Takehisa H, Kamatsuki K, et al (2013) RiceXPro Version 3.0: Expanding the informatics resource for rice transcriptome. Nucleic Acids Res 41:1206–1213. https://doi.org/10.1093/nar/gks1125

Shen B, Li C, Min Z, et al (2003) sal1 determines the number of aleurone cell layers in maize endosperm and encodes a class E vacuolar sorting protein. Proc Natl Acad Sci U S A 100:6552–7. https://doi.org/10.1073/pnas.0732023100

Suzuki K, Miyake H, Taniguchi T, Maeda E (2000) Cellularization of the free nuclear endosperm in rice caryopsis revealed by light and electron microscopy. Plant Prod Sci 3:446–458. https://doi.org/10.1626/pps.3.446

Takada S, Takada N, Yoshida A (2013) ATML1 promotes epidermal cell differentiation in Arabidopsis

shoots. Dev 140:1919–1923. https://doi.org/10.1242/dev.094417

Tanaka H, Watanabe M, Sasabe M, et al (2007) Novel receptor-like kinase ALE2 controls shoot development by specifying epidermis in Arabidopsis. Development 134:1643–1652. https://doi.org/10.1242/dev.003533

Ueda M, Zhang Z, Laux T (2011) Transcriptional activation of Arabidopsis axis patterning genes WOX8/9 Links zygote polarity to embryo development. Dev Cell 20:264–270. https://doi.org/10.1016/j.devcel.2011.01.009

Wisniewski JP, Rogowsky PM (2004) Vacuolar H+-translocating inorganic pyrophosphatase (Vpp1) marks partial aleurone cell fate in cereal endosperm development. Plant Mol Biol 56:325–337. https://doi.org/10.1007/s11103-004-3414-x

Wu H, Gontarek BC, Yi G, et al (2020) The thick aleurone1 gene encodes a NOT1 subunit of the CCR4-NOT complex and regulates cell patterning in endosperm. Plant Physiol 184:960–972. https://doi.org/10.1104/pp.20.00703

Wu X, Liu J, Li D, Liu CM (2016) Rice caryopsis development II: Dynamic changes in the endosperm. J Integr Plant Biol 58:786–798. https://doi.org/10.1111/jipb.12488

Xu JJ, Zhang XF, Xue HW (2016) Rice aleurone layer specific OsNF-YB1 regulates grain filling and endosperm development by interacting with an ERF transcription factor. J Exp Bot 67:6399–6411. https://doi.org/10.1093/jxb/erw409

Yephremov A, Wisman E, Huijser P, et al (1999) Characterization of the FIDDLEHEAD gene of Arabidopsis reveals a link between adhesion response and cell differentiation in the epidermis. Plant Cell 11:2187–2201. https://doi.org/10.1105/tpc.11.11.2187

Yi G, Lauter AM, Paul Scott M, Becraft PW (2011) The thick aleurone1 mutant defines a negative regulation of maize aleurone cell fate that functions downstream of defective kernel. Plant Physiol 156:1826–1836. https://doi.org/10.1104/pp.111.177725

Yi G, Neelakandan AK, Gontarek BC, et al (2015) The naked endosperm genes encode duplicate INDETERMINATE domain transcription factors required for maize endosperm cell patterning and differentiation. Plant Physiol 167:443–456. https://doi.org/10.1104/pp.114.251413

Zhan J, Thakare D, Ma C, et al (2015) RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. Plant Cell 27:513–531. https://doi.org/10.1105/tpc.114.135657

Table 1. Gene Ontology **(**GO) term enrichment analysis of Inner or Outermost peripheral DEGs

Inner DEGs[a] at 2L(l)

GO biological process complete	Reference (43659)	Input (334)	Fold enrichment	FDR[b]
glycogen biosynthetic process (GO:0005978)	10	4	52.44	1.06E-03
plastid transcription (GO:0042793)	9	3	43.7	8.77E-03
regulation of microtubule cytoskeleton organization (GO:0070507)	11	3	35.76	1.35E-02
starch biosynthetic process (GO:0019252)	33	7	27.81	1.09E-05
mitotic spindle organization (GO:0007052)	20	3	19.67	4.88E-02
sucrose metabolic process (GO:0005985)	39	5	16.81	3.34E-03
spindle assembly (GO:0051225)	39	5	16.81	3.24E-03
mitotic cell cycle phase transition (GO:0044772)	56	6	14.05	1.64E-03
regulation of anatomical structure morphogenesis (GO:0022603)	40	4	13.11	2.71E-02
regulation of cell cycle process (GO:0010564)	85	6	9.25	7.25E-03
regulation of cyclin-dependent protein serine/threonine kinase activity (GO:0000079)	71	5	9.23	2.37E-02
cell division (GO:0051301)	197	9	5.99	3.74E-03
response to alcohol (GO:0097305)	158	7	5.81	2.35E-02
response to lipid (GO:0033993)	222	8	4.72	3.04E-02
response to hormone (GO:0009725)	516	13	3.3	2.03E-02
Unclassified (UNCLASSIFIED)	26790	146	0.71	8.15E-07

Outermost DEGs[a] at 2L(l) stage

GO biological process complete	Reference (43659)	Input (515)	Fold enrichment	FDR[b]
organic hydroxy compound transport (GO:0015850)	14	4	24.56	2.52E-02
negative regulation of catalytic activity (GO:0043086)	229	11	4.13	3.24E-02
hormone-mediated signaling pathway (GO:0009755)	321	13	3.48	3.74E-02
anion transport (GO:0006820)	325	13	3.44	3.82E-02
transmembrane transport (GO:0055085)	1312	35	2.29	1.04E-02
oxidation-reduction process (GO:0055114)	1983	44	1.91	2.60E-02
Unclassified (UNCLASSIFIED)	26790	214	0.69	2.12E-14

[a]DEGs, differentially expressed genes.

[b]FDR, false discovery rate.

Figure legends

Fig. 1. Progression of cellularization and periclinal mitotic division during early endosperm development.

(A–C) Z-stack images of transverse sections of propidium iodide–stained developing rice endosperm (68–84 h after flowering unless otherwise specified) were obtained by CLSM. Red and purple arrowheads mark syncytial and alveolar nuclei layers, respectively. Yellow arrowheads mark examples of typical periclinal divisions of peripheral cells. Green arrowheads mark examples of typical periclinal divisions of inner alveoli. White arrowheads mark uncommon divisions at each stage. CV, central vacuole. Scale bars, 50 μm.

(A) Representative cross-sections of endosperm at the five stages of cellularization defined in this study; magnified images are shown beneath each panel. (a) Stage Syn(g): Endosperm nuclei are positioned in the peripheral cytoplasm, and the number of nuclei is below 40. (b) Stage Syn(f): After several rounds of nuclear division, a maximum number of nuclei are arranged along the inner circumference of the syncytium. (c) Stage 1L: Endosperm nuclei form one cell layer after anticlinal walls are formed between the syncytial nuclei, resulting in open-ended compartments called alveoli. (d) Stage 2L(e): Early developmental phase of stage 2L, where endosperm nuclei form two cell layers, with the peripheral cells forming complete cells and the inner cells forming alveoli. (e) Stage 2L(l): Late developmental phase of stage 2L. The inner alveoli become longer than the peripheral cells. (f and g) Stages 3L and 4L, where three and four cell layers are formed, respectively. (h) Stage 5L+: Endosperm constitutes five or more cell layers that fill most of the central vacuole.

(B) Periclinal division of alveoli at the 2L(l) stage. Periclinal divisions occur at the inner alveoli. The green arrowhead marks an example of typical periclinal division of the inner alveoli.

(C) Periclinal division in peripheral cells during early endosperm development. (a) Peripheral periclinal division during stage 3L to 4L. (b) Periclinal division during stage 5L to 6L. Periclinal division occurs at the outermost peripheral cell layer as well as among the inner cells. (c) Endosperm corresponding to about the 13- to 14-cell-layer stage at 5 days after flowering (DAF). Periclinal division by the peripheral cells is still observed. Yellow arrowheads mark examples of typical periclinal division by the peripheral cells.

(D) Frequencies of cells undergoing periclinal nuclear divisions within the three cell layers in endosperm at 5 DAF. 1st, outermost peripheral cell layer; 2nd, second outermost cell layer; 3rd, third outermost cell layer. n = 414 (1st), n = 367 (2nd), and n = 317 (3rd).

(E) Schematic representation of cell division and proliferation during rice endosperm cellularization. Black dots and lines represent nuclei and cell walls, respectively. Solid red lines represent periclinal cell walls formed by the most recent cell division at each stage. Dashed red lines represent cell walls formed by the

uncommon type of cell division at each stage. Dashed blue lines connect parental and daughter cells before and after cell division and before and after cell enlargement at the 2L(e) and 2L(l) stages. Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage.

Fig. 2. Overviews of the transcriptome analysis approaches used to examine endosperm cell-type- and cell-layer-specific gene expression during early endosperm development.

(Left) Approach used to identify cell-type-specific marker genes to discriminate between starchy endosperm (SE) and aleurone layer (AL). SE and AL were manually isolated according to the methods of Kuwano et al. (2011). After enrichment of genes preferentially expressed in SE or AL by comparative transcriptome analysis based on mRNA-seq expression data, genes highly expressed in endosperm but not in other tissues (see materials and methods) were selected as SE-specific or AL-specific genes by using the data published in the RiceXPro database. En, endosperm. Em, Embryo. DEG, differentially expressed genes.

(Right) Approach used to analyze cell-layer-specific gene expression at the early stage of endosperm cellularization. Whole or inner tissues at the 2L(l) stage were collected by LCM and subjected to mRNA-seq. After removal of genes likely resulting from nucellus contamination, genes that were preferentially expressed in whole but not inner tissues were assumed to be outermost peripheral-cell-layer-specific genes. At the 2L(l) stage, 399 and 654 genes were detected as Inner and Outermost peripheral DEGs, respectively. Among them, 334 Inner DEGs and 515 Peripheral DEGs whose expressions are profiled in RiceXPro were used for further analysis.

Fig. 3. Transcriptome analyses of cell-layer-specific gene expressions during early endosperm development.

(**A**, **B**). Venn diagrams showing the overlap between Inner DEGs and starchy endosperm (SE)-specific genes (A), and between Outermost peripheral DEGs and aleurone layer (AL)-specific genes (B). (**C**) Expression patterns of three genes selected from among the Inner DEGs; *AGPL2* and *Wx*, which encode enzymes involved in starch synthesis in SE, and *BT1-1*, which encodes a transporter of precursor of starch. (**D**) Expression patterns of three genes selected from among the Outermost peripheral DEGs, as well as that of *OsVP1*, a gene known to be expressed in AL. Estimated transcript abundance (count per million, cpm)

in inner (in) and outermost peripheral (out) cell layers of the 2L(l) and 3L stages. Data are presented as mean ± SD (n ≥ 3). 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage.

Fig. 4. Spatial expression of cell-layer-specific genes during early endosperm development.
(A, **B)** Expression of *3x(NLS-Venus)* reporter genes driven by *AGPL2* promoter (A) and *BT1-1* promoter (B) in transverse sections of developing endosperm at the 2L(l) stage. Counter-staining with DAPI revealed the nucleus and cell wall. Magnified images are shown beneath each panel. In each magnified image, the upper yellow line indicates the boundary between the inner (in) and outermost peripheral (out) cell layers, and the lower yellow line represents the outer boundary of the endosperm. Scale bars, 50 μm. **(C–E** and **G)** *In situ* mRNA localization of *AGPL2* (C), *CYP71X14* (D), *OsPRR12* (E), and *RFA2* (G) in transverse sections of developing rice endosperm at the 2L stage. Signals obtained with sense or antisense RNA probes are shown. Scale bars, 50 μm. **(F)** Expression pattern of *RFA2*. Estimated transcript abundances (count per million, cpm) in the inner (in) and outermost peripheral (out) cell layers of the 2L(l) and 3L stages. Data are shown as mean ± SD (n ≥ 3). cv, central vacuole; nu: nucellus. The dorsal side of the endosperm, which is where the nucellar projection resides, is oriented to the upper right in panels A and B (top), C–E, and G. 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage.

Fig. 5. Expression patterns of starchy endosperm–specific genes during early endosperm development.
Estimated transcript abundances (count per million, cpm) of *AGPL2* **(A)**, *Wx* **(B)**, and *BT1-1* **(C)** at the syncytium and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), 2L(l), and 3L stages. Data are shown as mean ± SD (n ≥ 3). Insets are magnified views of the graphs for the 1L and 2L(e) stages. Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage. For each gene, the comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments.

Fig. 6. Expression patterns of aleurone-layer-specific genes during early endosperm development.
(A) Estimated transcript abundances (count per million, cpm) of *NF-YB1*, *OsVP1*, *CYP71X14*, and *OsPRR12* at the 1L stage and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e) and 2L(l) and 3L stages. Data are presented as mean ± SD (n ≥ 3). **(B)** Hierarchical clustering of the expression patterns of 515 Peripheral DEGs at the 2L(l) stage visualized as a heatmap. The 515 outer

DEGs were classified into two clusters by complete linkage clustering. Type IV, *type IV HD-ZIP* genes. **(C**, **D)** Estimated transcript abundances (count per million, cpm) of *ROC1* and *ROC7* at the syncytium and 1L stages and in the inner and outermost peripheral cell layers of the 2L(e), 2L(l), and 3L stages. Data are presented as mean ± SD ($n \geq 3$). Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage.

Fig. S1. Frequency of appearance of the different stages of endosperm development, as defined by cell-layer number, at the indicated time points.

Ovary tissues were fixed at 66–84 h after flowering (HAF) and used for histological analysis using CLSM. Images were obtained for all samples, the number of samples at each developmental stage was counted, and the number of samples at each stage among all of the samples examined for that time point was calculated. Three independent collections of ovary samples from each time point were combined to calculate the frequencies. Syn(g), growing syncytium stage; Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage; 4L, 4-cell-layer stage; 5L+, 5 or more -cell-layer stage.

Fig. S2. Collection of tissue fragments from cross-sections of endosperm in the early developmental stages by LCM.

Images of whole and inner tissues **(A)** and endosperm tissue with adjacent nucellus (Nucellus + Whole) **(B)** were obtained before (left) and after (right) LCM. The colored lines represent the path of the laser irradiation. **(C)** Images of the syncytium Syn(f) and 1L stages before dissection. Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage.

Fig. S3. Spatial expression of layer-specific genes in the inner and outermost peripheral cell layers of developing endosperm.

In situ mRNA localization of *AGPL2* (A), *CYP71X14* (B), *OsPRR12* (C), and *RFA2* (D) in transverse sections of developing endosperm at the 3L stage. Signals obtained by using sense or antisense RNA probes are shown. Scale bars, 50 μm. cv, central vacuole. In the panels, the dorsal side of the endosperm, where the nucellar projection resides, is oriented to the upper right.

Fig. S4. **Expression of *AGPS2* gene during early endosperm development.**

(A) Expression of *AGPS2* in various tissues and at various time points. Redrawn from RiceXPro database. **(B)** Structures of the alternative RNA products of *AGPS2*: *AGPS2a* and *AGPS2b.* The boxes to the right show the structure of AGPS2a and AGPS2b. The orange box shows the position of chloroplast transit

peptide. **(C)** Estimated transcript abundances (count per million, cpm) of *AGPS2* at the syncytium (Syn(f)) and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), 2L(l), and 3L stages. Data are presented as mean ± SD (n ≥ 3). The comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments. **(D)** Histograms showing the coverage of mapped reads around exon 1 though 3 of *AGPS2*, as determined by mRNA-seq analysis. Arrowheads indicate expression of *AGPS2b*-specific transcripts. Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage.

Fig. S5. Expressions of seven aleurone-layer-specific genes selected from Cluster 2 of the hierarchical clustering analysis shown in Fig. 6B.

Estimated transcript abundances (count per million, cpm) at the syncytium and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), 2L(l), and 3L stages. Data are presented as mean ± SD (n ≥ 3). Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage. For each gene, the comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments.

Fig. S6. Phylogenic analysis of type IV HD-ZIP transcription factors in rice and *Arabidopsis thaliana*.

The phylogenetic tree was made by means of the neighbor-joining method (Saitou and Nei 1987) using alignment of the full-length amino acid sequences of type IV HD-ZIP proteins in rice and *A. thaliana*.

Fig. S7. Expressions of *type IV HD-ZIP* genes other than *ROC1* and *ROC7* during early endosperm development.

Estimated transcript abundances (count per million, cpm) of *type IV HD-ZIP* genes included in Cluster 2 (A) and Cluster 1 (B) of the hierarchical clustering analysis shown in Fig. 6B, as well as those genes not included among the Outermost peripheral DEGs (C), at the syncytium and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), and 2L(l), and 3L stages. Data are presented as mean ± SD (n ≥ 3). Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage. For each gene, the comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments.

Fig. S8. Expression of fatty acid synthesis genes during early endosperm development.

Estimated transcript abundances (count per million, cpm) of *OsPAS2* (A), *ONION1* (B), *ONION3* (C), and *OsKCR1* (D) at the syncytium and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), 2L(l), and 3L stages. Data are presented as mean ± SD (n ≥ 3). Syn(f), full syncytium stage;

1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage. For each gene, the comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments.

Fig. S9. CLSM images of a transverse section of mature rice caryopsis stained with propidium iodide. (A) Tiled image showing the entire section. The white blank in the center is a missing part of the tiled image. (B) enlarged images of the region within the yellow rectangles in panel (A). Scale bars, 200 μm. AL, aleurone layer; SE, starchy endosperm.

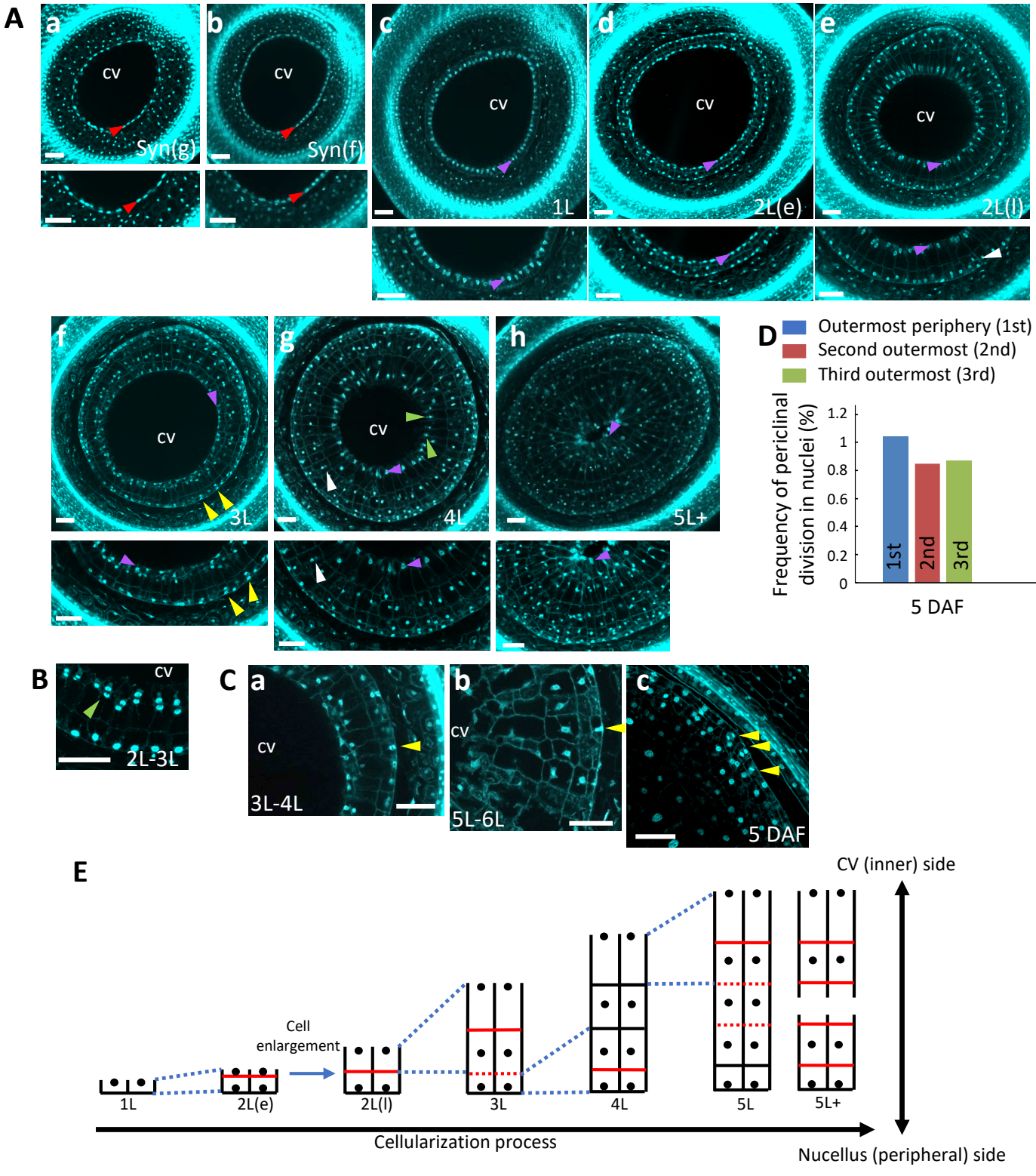
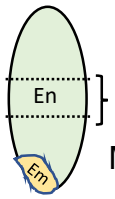

A
a
cv
Syn(g)
b
cv
Syn(f)
c
cv
1L
d
cv
2L(e)
e
cv
2L(l)
f
cv
3L
g
cv
4L
h
5L+
D
Outermost periphery (1st)
Second outermost (2nd)
Third outermost (3rd)
Frequency of periclinal division in nuclei (%)
1.2
1
0.8
0.6
0.4
0.2
0
1st
2nd
3rd
5 DAF
B
cv
2L-3L
C
a
cv
3L-4L
b
cv
5L-6L
c
5' DAF
E
CV (inner) side
Cell enlargement
1L
2L(e)
2L(l)
3L
4L
5L
5L+
Cellularization process
Nucellus (peripheral) side

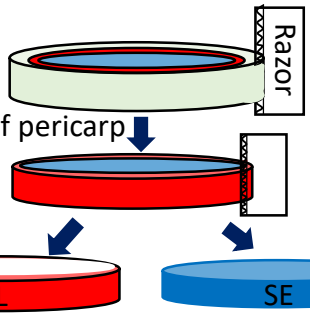
Fig. 1

Cell-type-specific transcriptome



10 DAF endosperm

Manual isolation of SE and AL



RNA extraction, RNA-sequencing

DEG analysis (AL vs. SE)

978 SE preferential genes

3182 AL preferential genes

Selection of endosperm-specific genes using the data in RiceXPro

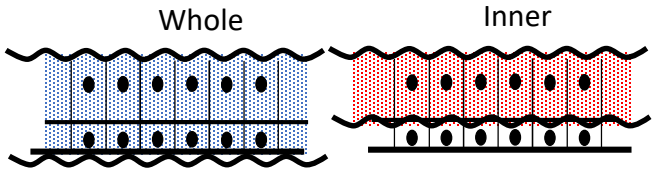
105 SE-specific genes

141 AL-specific genes

Cell-layer-specific transcriptome

2L(I) endosperm cross-sections

Collection of “whole” or “inner” tissue fragments from serial sections by laser-capture microdissection



RNA extraction, RNA-sequencing

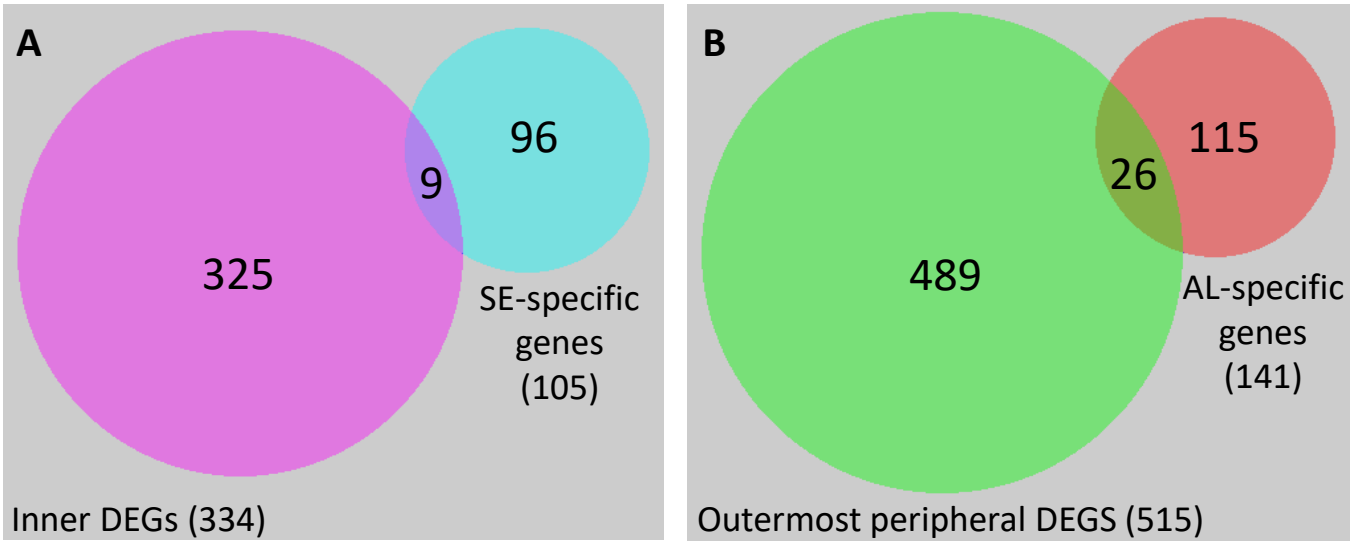
1. DEG analysis (“whole” vs. “Inner”)
2. Removal of genes due to nucellar contamination in “whole“ sample
3. Estimation of transcript abundance in the outermost peripheral layer

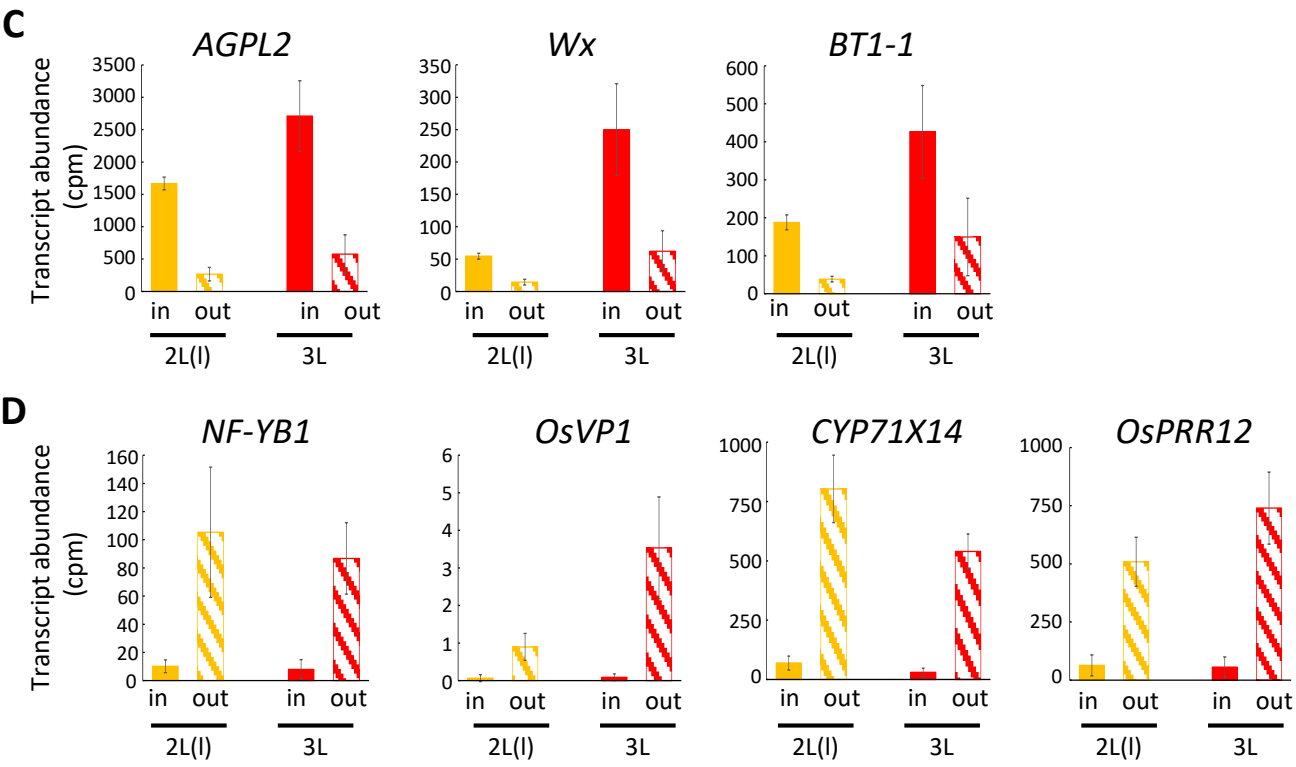
399* Inner DEGs
Preferentially expressed genes in the Inner cell layer

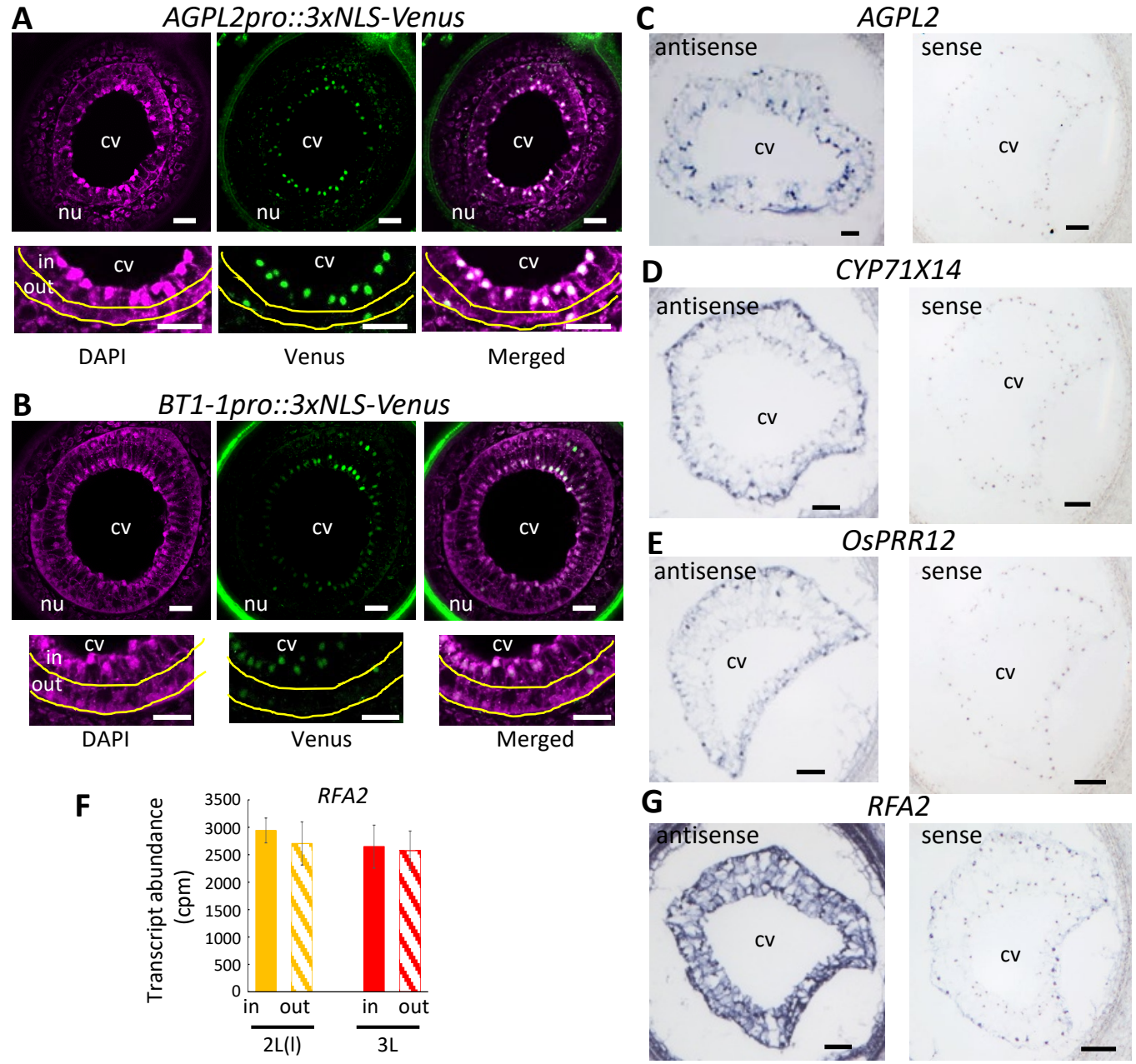
* of which **334** were profiled in RiceXPro

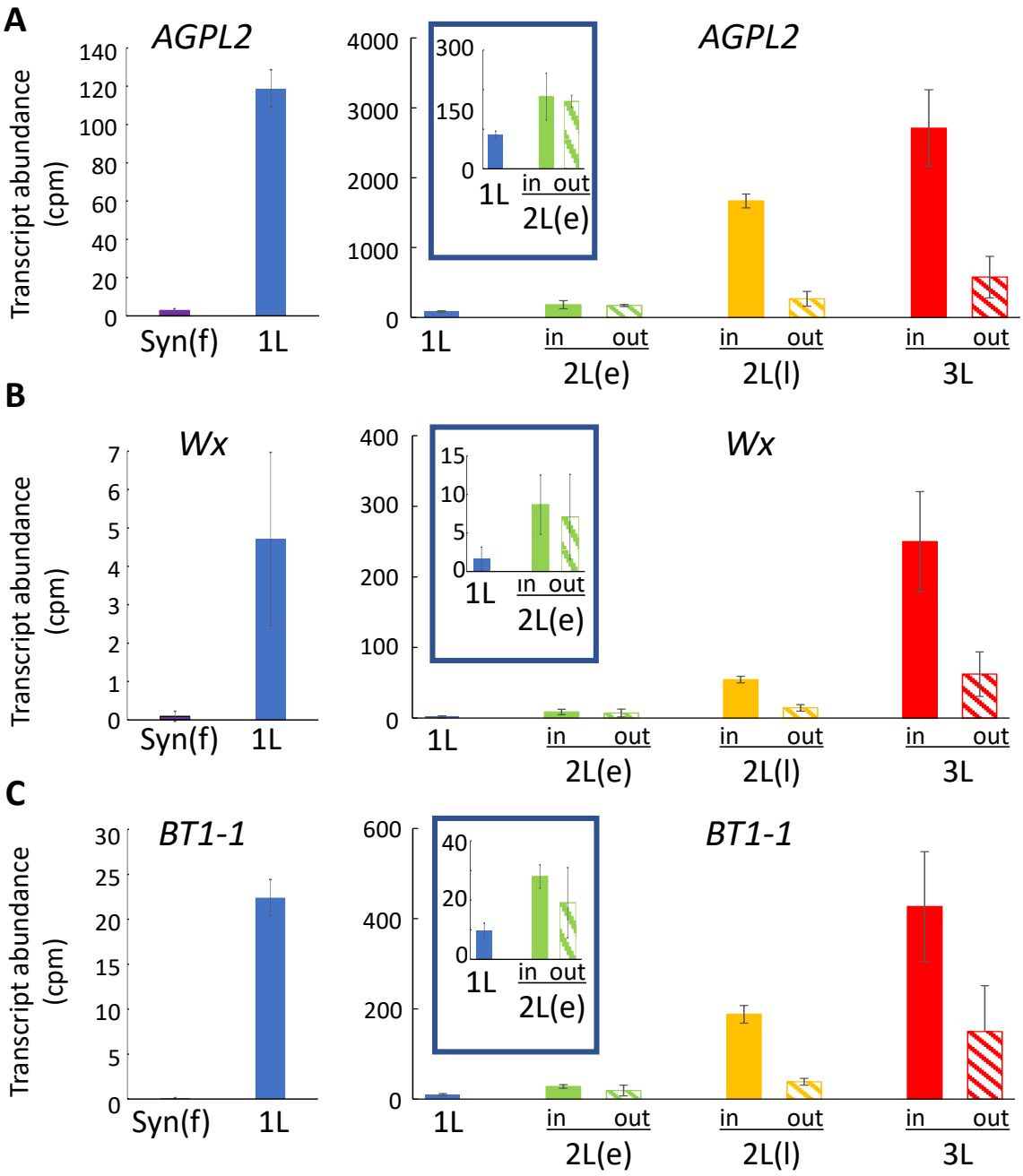
654 Outermost peripheral DEGs**
Preferentially expressed genes in the outermost peripheral cell layer

** of which **515** were profiled in RiceXPro


A
325
9
96
SE-specific genes (105)
Inner DEGs (334)
B
489
26
115
AL-specific genes (141)
Outermost peripheral DEGS (515)


C
AGPL2
Wx
BT1-1
Transcript abundance (cpm)
in
out
2L(I)
3L
D
NF-YB1
OsVP1
CYP71X14
OsPRR12


A
AGPL2pro::3xNLS-Venus
cv
nu
in
out
DAPI
Venus
Merged
B
BT1-1pro::3xNLS-Venus
C
AGPL2
antisense
sense
D
CYP71X14
E
OsPRR12
F
RFA2
Transcript abundance (cpm)
3500
3000
2500
2000
1500
1000
500
0
2L(I)
3L
G


A
AGPL2
Transcript abundance (cpm)
140
120
100
80
60
40
20
0
Syn(f)
1L
AGPL2
4000
3000
2000
1000
0
300
150
0
1L
in out
2L(e)
1L
in out
2L(e)
in out
2L(l)
in out
3L
B
Wx
Transcript abundance (cpm)
7
6
5
4
3
2
1
0
Syn(f)
1L
Wx
400
300
200
100
0
15
10
5
0
1L
in out
2L(e)
1L
in out
2L(e)
in out
2L(l)
in out
3L
C
BT1-1
Transcript abundance (cpm)
30
25
20
15
10
5
0
Syn(f)
1L
BT1-1
600
400
200
0
40
20
0
1L
in out
2L(e)
1L
in out
2L(e)
in out
2L(l)
in out
3L

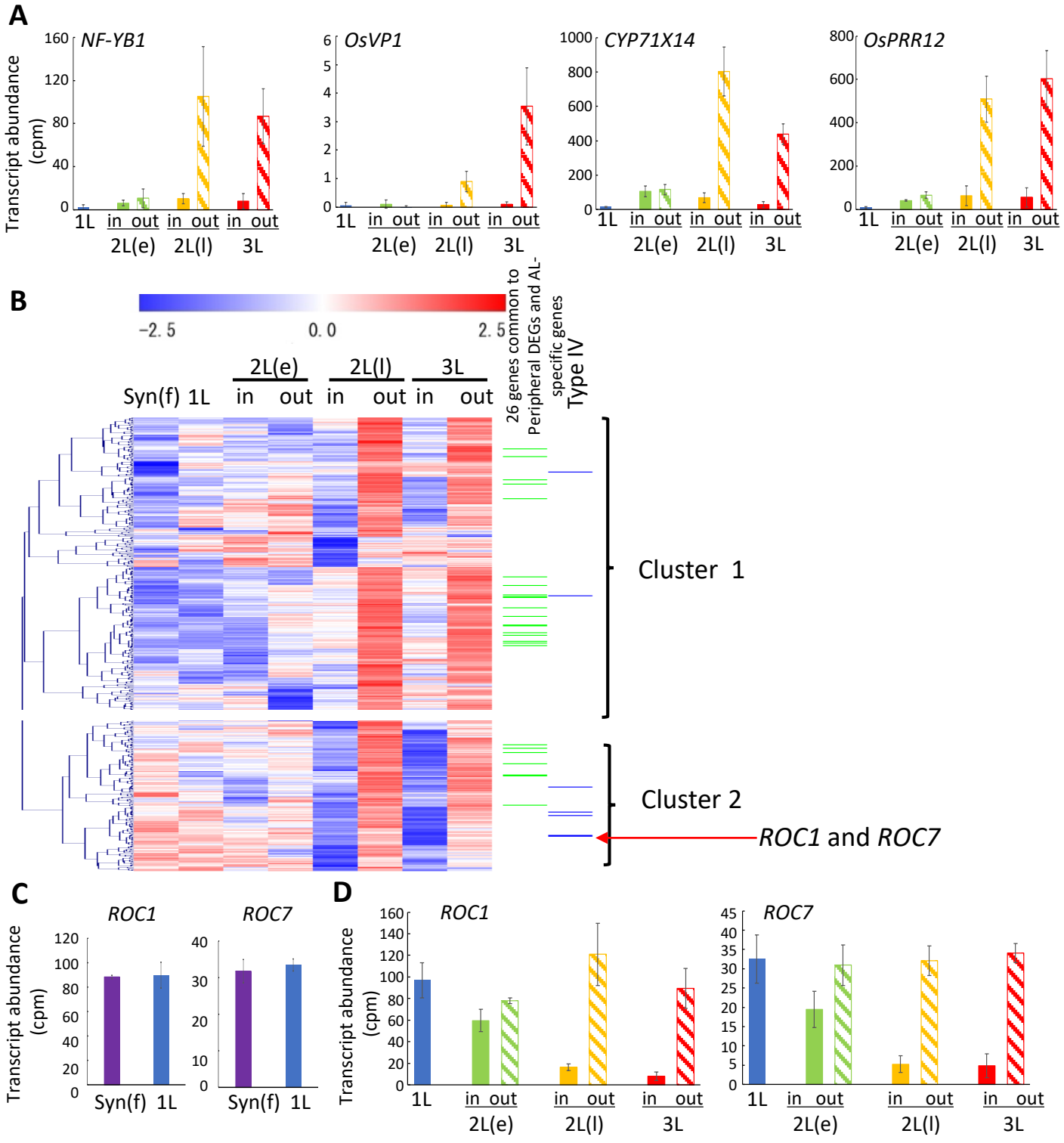

A
NF-YB1
OsVP1
CYP71X14
OsPRR12
Transcript abundance (cpm)
1L
in out
2L(e)
2L(I)
3L
B
−2.5
0.0
2.5
Syn(f) 1L
26 genes common to Peripheral DEGs and AL-specific genes
Type IV
Cluster 1
Cluster 2
ROC1 and ROC7
C
ROC1
ROC7
Syn(f) 1L
D
ROC1
ROC7

Fig. 6

Supplementary information

・**Title:**

High-resolution Spatiotemporal Transcriptome Analyses during Cellularization of Rice Endosperm Unveil the Earliest Gene Regulation Critical for Aleurone and Starchy Endosperm Cell Fate Specification

・**Journal name:** ***Journal of Plant Research***

・**Names of authors:**

Yoshinori Takafuji[1], Sae Shimizu-Sato[2], Kim Nhung Ta[2], Toshiya Suzuki[2,3], Misuzu Nosaka-Takahashi[2,3], Tetsuro Oiwa[1], Wakana Kimura[1], Hirokazu Katoh[1], Mao Fukai[1], Shin Takeda[1], Yutaka Sato[2,3] and Tsukaho Hattori[1]

・**Corresponding Authors' affiliations**

[1]Graduate school of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya 464-8601, Japan

[2]National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

[3]Department of Genetics, School of Life Science, SOKENDAI (The Graduate University for Advanced Studies), 1111 Yata, Mishima, Shizuoka 411-8540, Japan

・**Corresponding authors' e-mail:**

hattori@agr.nagoya-u.ac.jp

takeda@agr.nagoya-u.ac.jp

yusato@nig.ac.jp

Supplementary Methods
Supplementary Figures (Fig. S1-S9)

Supplementary Methods

***In situ* hybridization**

Digoxigenin (DIG)-labeled RNA probes were prepared by using a MAXIscriptT7 Kit (Thermo Fisher, Waltham, MA). DNA fragments were amplified by polymerase chain reaction using appropriate primer sets (Table S1), subjected to agarose gel electrophoresis, and then purified with a QIAEX II Gel Extraction Kit (Qiagen, Valencia, CA). DIG-labeled RNA was precipitated with LiCl, washed with 70 % ethanol, dried at room temperature, dissolved in nuclease-free water, and stored at −80 °C. The length and concentration of the RNA probes were determined by MOPS-formaldehyde gel electrophoresis.

Ovary tissues were fixed in Paraformaldehyde (PFA) solution (50 mM Na phosphate buffer [pH 7.2] containing 4 % paraformaldehyde, 0.25 % glutaraldehyde, and 0.5 % Triton X-100]) on ice under vacuum for 1 h, and for an additional 1 h in fresh PFA solution. After changing PFA solution twice, the tissues were incubated at 4 °C overnight, transferred to 100 mM Na phosphate buffer (pH 7.2), and incubated at room temperature for 30 min. The tissues were dehydrated in an ethanol series (30, 50, 70, 80, 90, 100 %) for 1 h at each concentration, twice in absolute ethanol for 1 h, and then in absolute ethanol containing safranin overnight. The tissues were then treated with a series of absolute ethanol and Histo-Clear II (Cosmo Bio, Tokyo, Japan) solutions (2:1, 1:1, and 1:2 [v/v]) each for 1 h and then twice with 100 % Histo-Clear II for 1 h, before finally being immersed in Paraplast Plus (Leica, Wetzlar, Germany) at 60 °C for 4 days, after which the tissues were stored at 4 °C.

For *in situ* hybridization, 8-μm-thick sections were prepared on MAS-coated glass slides (Matsunami, Osaka, Japan). Tissue sections were soaked twice in 100 % Histo-Clear II for 10 min, once in Histo-Clear II and ethanol solution (1:1 [v/v]) for 5 min, twice in 100 % ethanol for 5 min, and in an ethanol series (90, 70, 50, 30 %) for 2 min at each concentration. After soaking twice in sterilized water for 5 min, sections were treated with Proteinase K (Wako, Osaka, Japan) (final concentration of 5 μg/mL) at 37 °C for 30 min, washed twice with sterilized water for 5 min, and fixed with 4 % paraformaldehyde in 10 mM Na phosphate buffer (pH 7.2) for 10 min. After washing twice with sterilized water for 5 min, sections were treated with 0.1 M triethanolamine (pH 8.0) containing freshly added acetic anhydride (1/200 dilution) for 10 min, and with

2× SSPE (1× SSPE = 150 mM NaCl, 10m M NaH_2PO_4, 1 mM EDTA [pH 7.4]) twice for 5 min. The tissue sections were then dehydrated in an ethanol series (30, 50, 70, 90, 100, 100 %) for 2 min at each concentration and then dried under a vacuum for 1 h. Hybridization mixture (50 % formamide, 0.3 M NaCl, 10 mM Tris–HCl, 1 mM EDTA, 1× Denhardt's solution, 10 % dextran sulfate, 60 mM Dithiothreitol (DTT), 1 mg/mL tRNA, 0.5 mg/mL poly(A)) containing the required RNA probe was added to the sections (300 μL/glass slide) before incubation for at least 16 h at 52 °C in a chambered box containing formamide diluted by half with sterilized water. After hybridization, tissue sections were washed four times with 4× SSC (1× SSC = 15 mM sodium citrate, 150 mM NaCl) for 5 min with gentle agitation, and treated with RNase A (50 μg/mL final concentration; Sigma, St. Louis, MO) in RNase buffer (0.5 M NaCl, 10 mM Tris–HCl [pH 7.5], 5 mM EDTA) at 37 °C for 30 min. The sections were washed three times with RNase buffer at 37 °C for 5 min, and twice with 0.5× SSC at 52 °C for 20 min, and then stained with alkaline phosphatase-conjugated anti-DIG-antibody (Roche, Basel, Switzerland) and color substrate solutions using standard protocols. After washing with TE buffer, the sections were dehydrated with a gradient ethanol series (30, 50, 70, 90 %, absolute, absolute), 50 % Lemosol (Wako, Osaka, Japan) in absolute ethanol, and 100 % Lemosol for each 10 sec, and then mounted with EUKITT neo (Cosmo Bio, Tokyo, Japan). Hybridization signals within tissues were observed under a microscope (AXIOPLAN2, ZEISS, Germany).

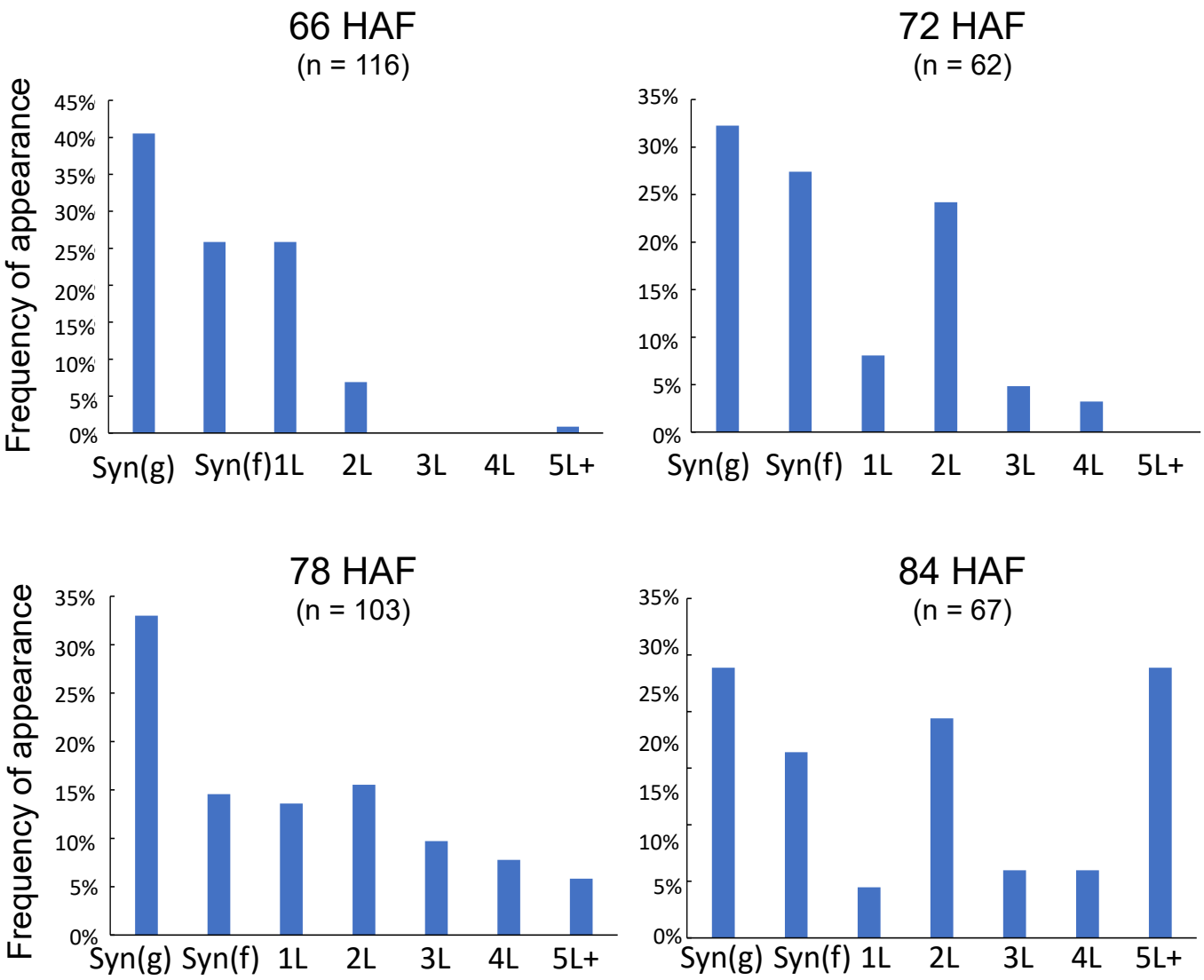


Fig. S1. Frequency of appearance of the different stages of endosperm development, as defined by cell-layer number, at the indicated time points.

Ovary tissues were fixed at 66–84 h after flowering (HAF) and used for histological analysis using CLSM. Images were obtained for all samples, the number of samples at each developmental stage was counted, and the number of samples at each stage among all of the samples examined for that time point was calculated. Three independent collections of ovary samples from each time point were combined to calculate the frequencies. Syn(g), growing syncytium stage; Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage; 4L, 4-cell-layer stage; 5L+, 5 or more -cell-layer stage.

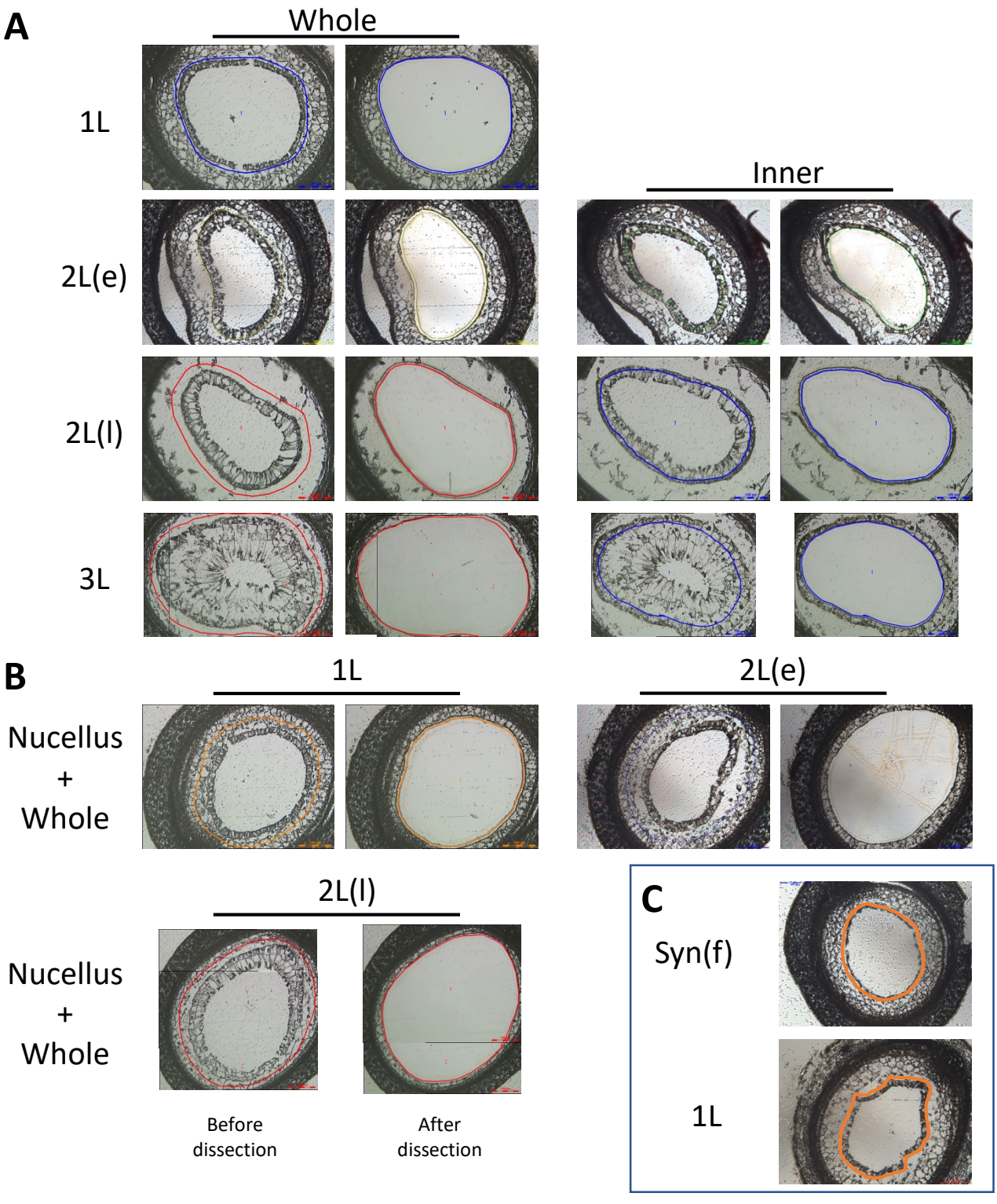


Fig. S2. Collection of tissue fragments from cross-sections of endosperm in the early developmental stages by LCM.

Images of whole and inner tissues **(A)** and endosperm tissue with adjacent nucellus (Nucellus + Whole) **(B)** were obtained before (left) and after (right) LCM. The colored lines represent the path of the laser irradiation. **(C)** Images of the syncytium Syn(f) and 1L stages before dissection. Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage.

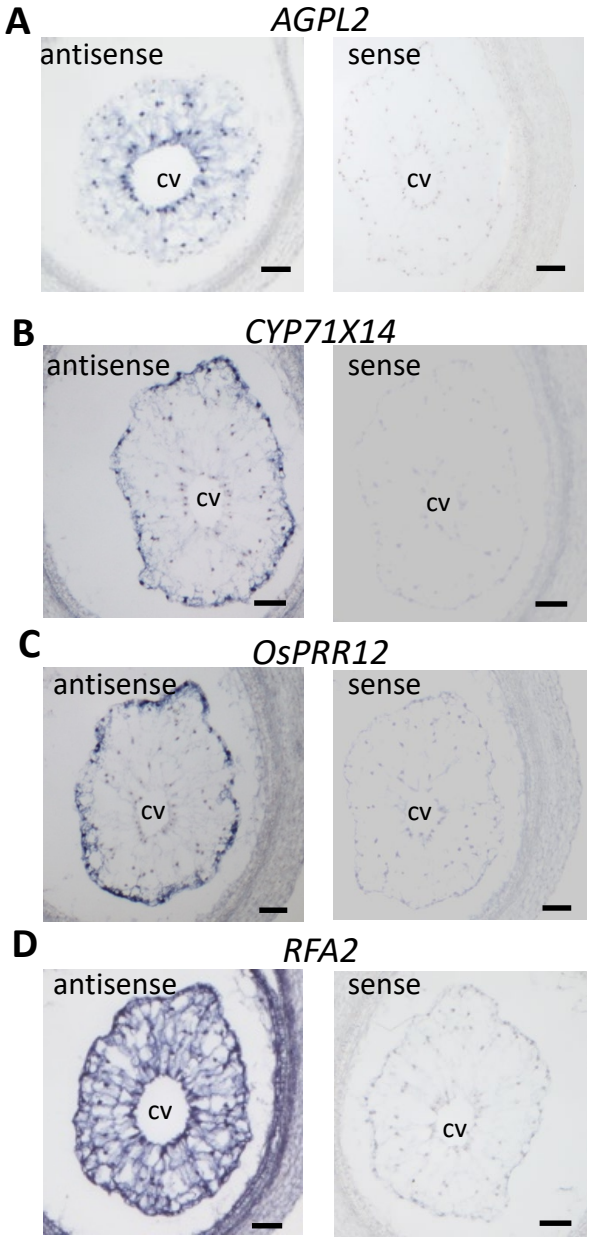


Fig. S3. Spatial expression of layer-specific genes in the inner and outermost peripheral cell layers of developing endosperm.
In situ mRNA localization of *AGPL2* (A), *CYP71X14* (B), *OsPRR12* (C), and *RFA2* (D) in transverse sections of developing endosperm at the 3L stage. Signals obtained by using sense or antisense RNA probes are shown. Scale bars, 50 μm. cv, central vacuole. In the panels, the dorsal side of the endosperm, where the nucellar projection resides, is oriented to the upper right.

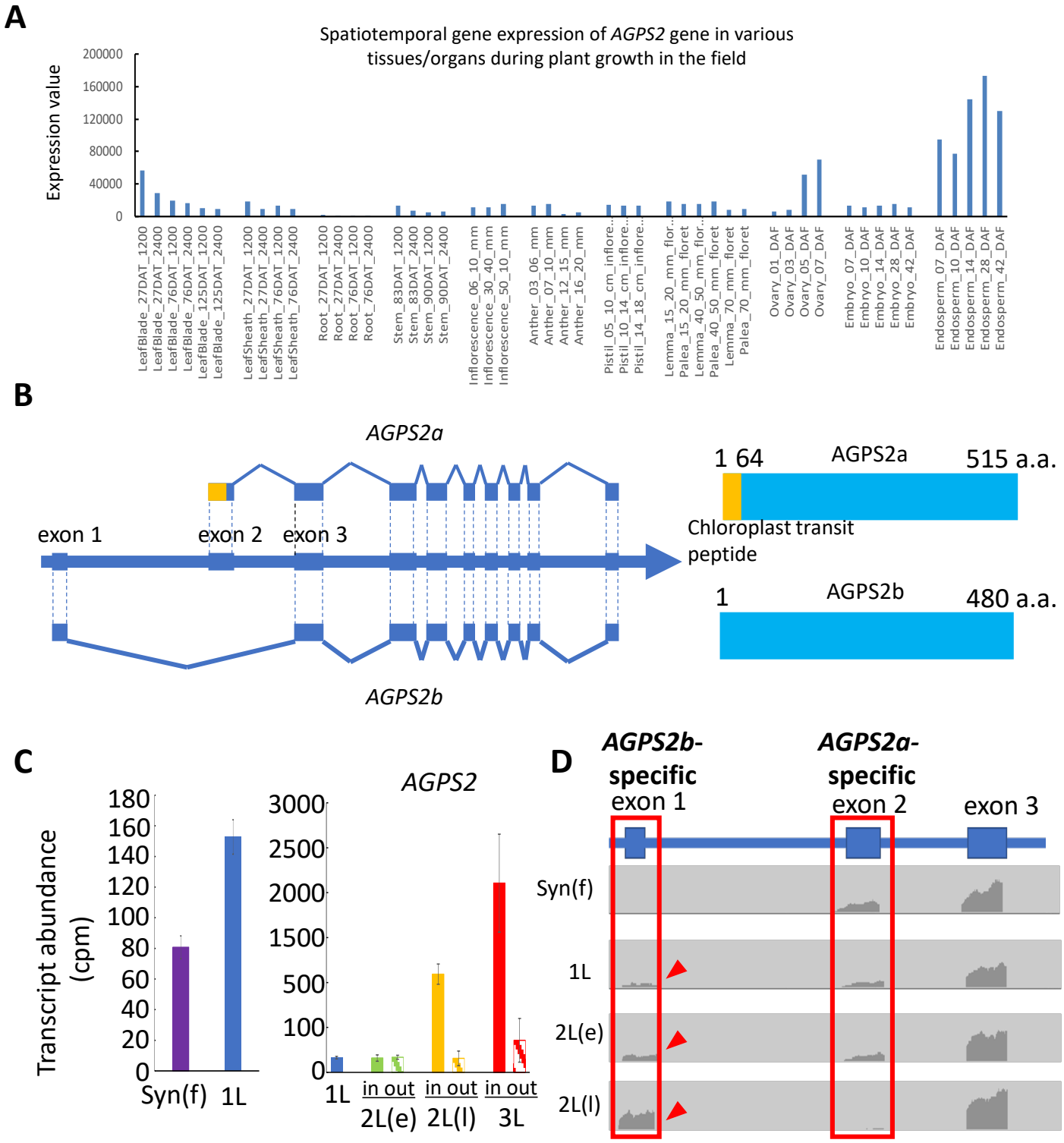


Fig. S4. **Expression of *AGPS2* gene during early endosperm development.**
(A) Expression of *AGPS2* in various tissues and at various time points. Redrawn from RiceXPro database. **(B)** Structures of the alternative RNA products of *AGPS2*: *AGPS2a* and *AGPS2b.* The boxes to the right show the structure of AGPS2a and AGPS2b. The orange box shows the position of chloroplast transit peptide. **(C)** Estimated transcript abundances (count per million, cpm) of *AGPS2* at the syncytium (Syn(f)) and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), 2L(l), and 3L stages. Data are presented as mean ± SD (n ≥ 3). The comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments. **(D)** Histograms showing the coverage of mapped reads around exon 1 though 3 of *AGPS2*, as determined by mRNA-seq analysis. Arrowheads indicate expression of *AGPS2b*-specific transcripts. Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage.

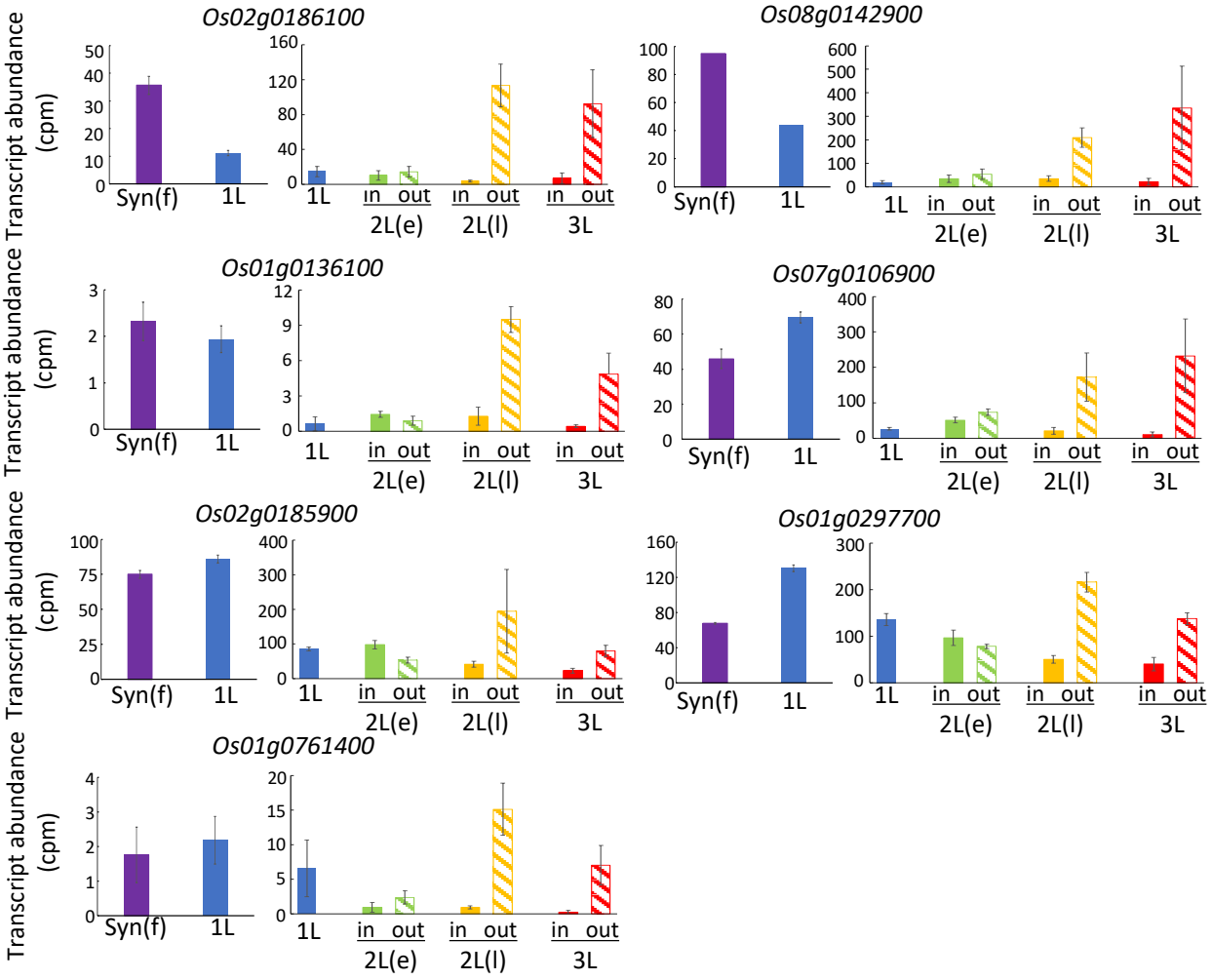


Fig. S5. Expressions of seven aleurone-layer-specific genes selected from Cluster 2 of the hierarchical clustering analysis shown in Fig. 6B.

Estimated transcript abundances (count per million, cpm) at the syncytium and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), 2L(l), and 3L stages. Data are presented as mean ± SD ($n \geq 3$). Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage. For each gene, the comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments.

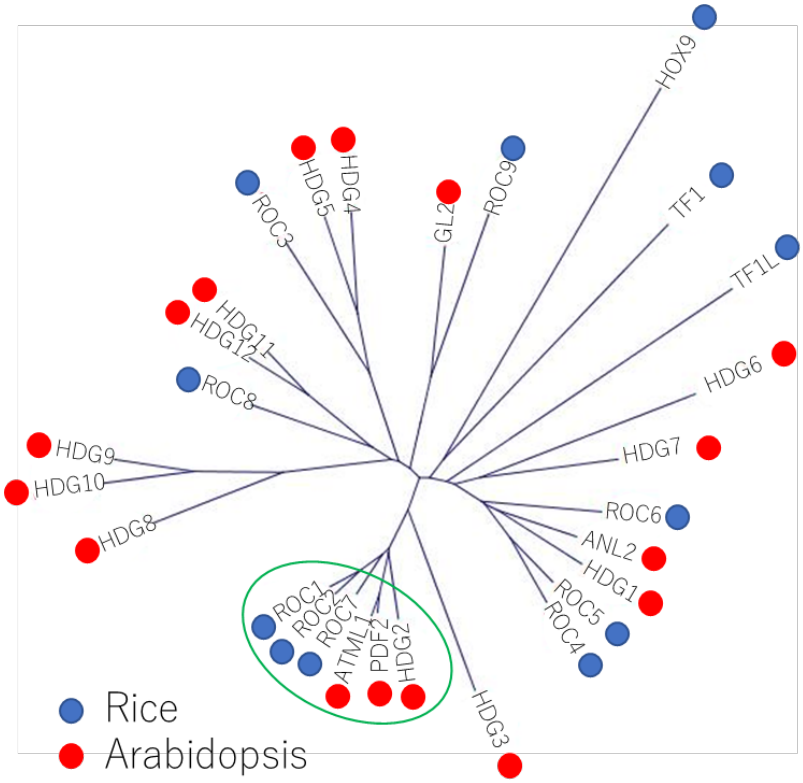


Fig. S6. Phylogenic analysis of type IV HD-ZIP transcription factors in rice and *Arabidopsis thaliana*.

The phylogenetic tree was made by means of the neighbor-joining method (Saitou and Nei 1987) using alignment of the full-length amino acid sequences of type IV HD-ZIP proteins in rice and *A. thaliana*.

Reference list
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

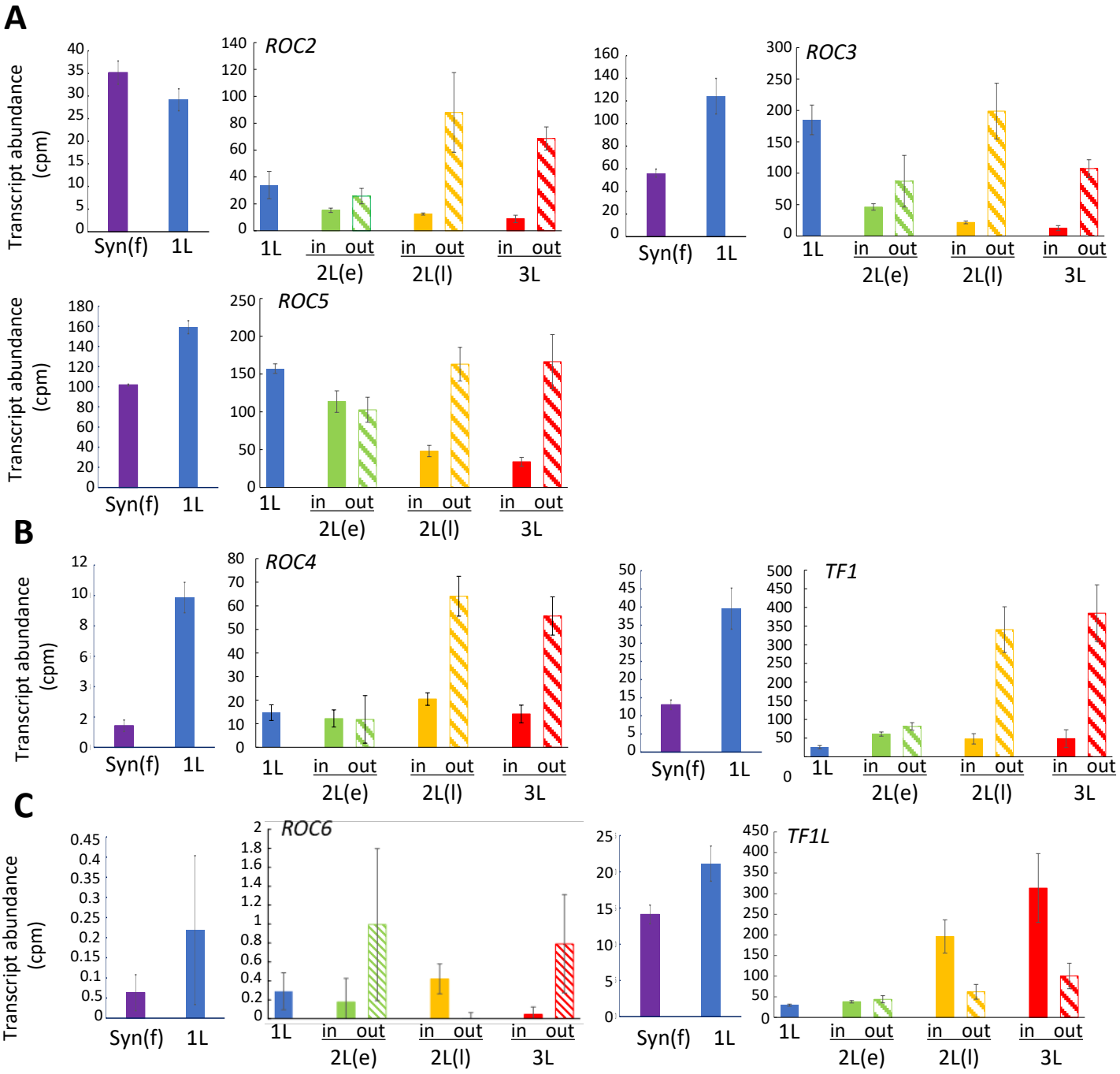


Fig. S7. Expressions of *type IV HD-ZIP* genes other than *ROC1* and *ROC7* during early endosperm development.

Estimated transcript abundances (count per million, cpm) of *type IV HD-ZIP* genes included in Cluster 2 (A) and Cluster 1 (B) of the hierarchical clustering analysis shown in Fig. 6B, as well as those genes not included among the Outermost peripheral DEGs (C), at the syncytium and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), and 2L(l), and 3L stages. Data are presented as mean ± SD ($n \geq 3$). Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage. For each gene, the comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments.

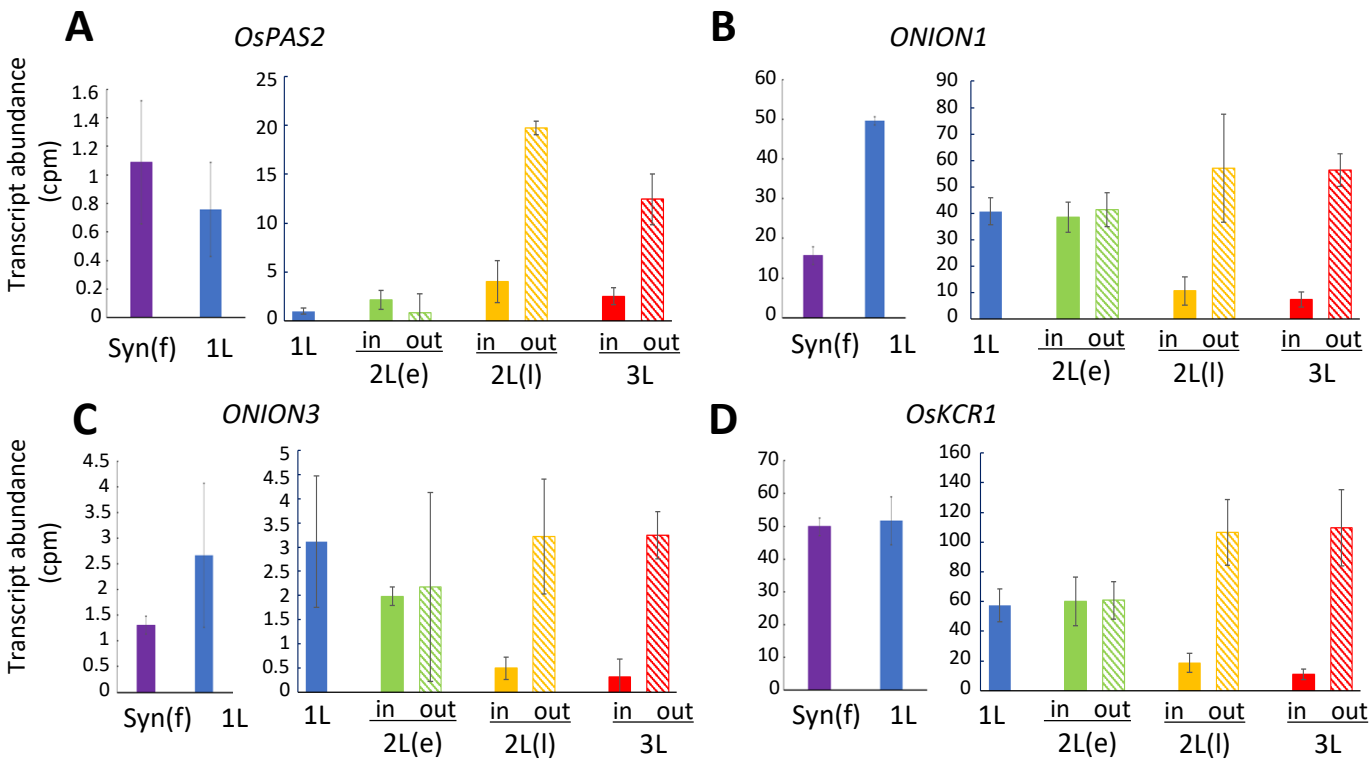


Fig. S8. Expression of fatty acid synthesis genes during early endosperm development.

Estimated transcript abundances (count per million, cpm) of *OsPAS2* (A), *ONION1* (B), *ONION3* (C), and *OsKCR1* (D) at the syncytium and 1L stages, and in the inner (in) and outermost peripheral (out) cell layers of the 2L(e), 2L(l), and 3L stages. Data are presented as mean ± SD ($n \geq 3$). Syn(f), full syncytium stage; 1L, single-cell-layer stage; 2L(e), early 2-cell-layer stage; 2L(l), late 2-cell-layer stage; 3L, 3-cell-layer stage. For each gene, the comparison of Syn(f) to 1L (graph on the left) and that of 1L to the later stages (graph on the right) was done in separate experiments.

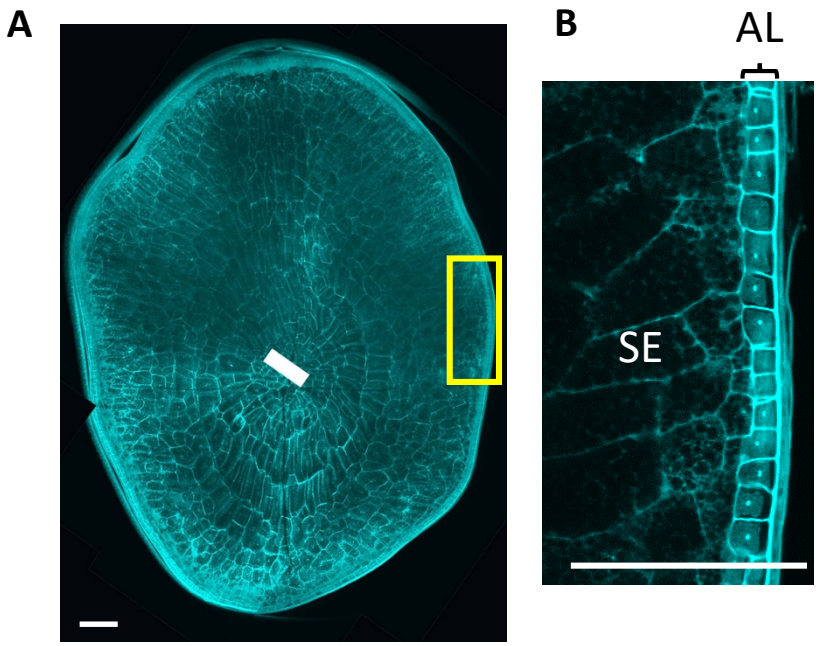


Fig. S9. CLSM images of a transverse section of mature rice caryopsis stained with propidium iodide.
(A) Tiled image showing the entire section. The white blank in the center is a missing part of the tiled image. (B) enlarged images of the region within the yellow rectangles in panel (A). Scale bars, 200 μm. AL, aleurone layer; SE, starchy endosperm.