Article

Analysis of Spatio-Temporal Transcriptome Profiles of Soybean (*Glycine max*) Tissues during Early Seed Development

Shuo Sun [1], Changyu Yi [2], Jing Ma [1], Shoudong Wang [3], Marta Peirats-Llobet [2], Mathew G. Lewsey [4,5], James Whelan [1,2,5,*] and Huixia Shou [1,*]

[1] State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China; sunshuo2019@zju.edu.cn (S.S.); majing1002@zju.edu.cn (J.M.)

[2] Australian Research Council Centre of Excellence in Plant Energy Biology, Department of Animal, Plant and Soil Science, School of Life Science, La Trobe University, Bundoora, Victoria 3086, Australia; C.Yi@latrobe.edu.au (C.Y.); M.Peirats-Llobet@latrobe.edu.au (M.P.-L.)

[3] Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Changchun 130102, China; wsd@zju.edu.cn

[4] Department of Animal, Plant and Soil Science, AgriBio Building, La Trobe University, Bundoora, Victoria 3086, Australia; M.Lewsey@latrobe.edu.au

[5] Australian Research Council Research Hub for Medicinal Agriculture, AgriBio Building, La Trobe University, Bundoora, Victoria 3086, Australia

* Correspondence: J.Whelan@latrobe.edu.au (J.W.); huixia@zju.edu.cn (H.S.)

check for updates

Abstract: Soybean (*Glycine max*) is an important crop providing oil and protein for both human and animal consumption. Knowing which biological processes take place in specific tissues in a temporal manner will enable directed breeding or synthetic approaches to improve seed quantity and quality. We analyzed a genome-wide transcriptome dataset from embryo, endosperm, endothelium, epidermis, hilum, outer and inner integument and suspensor at the global, heart and cotyledon stages of soybean seed development. The tissue specificity of gene expression was greater than stage specificity, and only three genes were differentially expressed in all seed tissues. Tissues had both unique and shared enriched functional categories of tissue-specifically expressed genes associated with them. Strong spatio-temporal correlation in gene expression was identified using weighted gene co-expression network analysis, with the most co-expression occurring in one seed tissue. Transcription factors with distinct spatiotemporal gene expression programs in each seed tissue were identified as candidate regulators of expression within those tissues. Gene ontology (GO) enrichment of orthogroup clusters revealed the conserved functions and unique roles of orthogroups with similar and contrasting expression patterns in transcript abundance between soybean and Arabidopsis during embryo proper and endosperm development. Key regulators in each seed tissue and hub genes connecting those networks were characterized by constructing gene regulatory networks. Our findings provide an important resource for describing the structure and function of individual soybean seed compartments during early seed development.

Keywords: *Glycine max*; seed development; transcriptome; RNA-sequencing (RNA-seq); tissue specific; transcription factor (TF); orthologue; gene regulatory network (GRN)

1. Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the most economically important cultivated plants. Soybean seeds are an excellent source of vegetable oil and protein as well as a significant source of

human nutrition and animal feed (Soystats 2020 http://soystats.com). Soybean belongs to the Fabaceae family, one of the largest and most diverse families of flowering plants. Legumes, as this family is known colloquially, have been used extensively as model species in plant biology and historically, soybean and other close relatives have been used as model species to study seed and embryo development [1–6]. Plant seed development comprises a series of morphological, physiological, and biochemical changes and can be divided into three major phases: embryogenesis, including cell division and expansion, seed maturation, and desiccation. Seed development is triggered by the double fertilization of the egg cell and central cells that leads to the differentiation of the seed tissues, embryo, endosperm and seed coat [5,7]. These three compartments have different origins and distinct roles in seed formation; the embryo, which is diploid and is derived from one paternal and one maternal genome equivalent; the endosperm, a structure that provides nourishment for the developing embryo, is triploid and derives from one paternal and two maternal genome equivalent; and the seed coat, which provides protection to the embryo and endosperm, has a strictly maternal origin and is derived from the ovule integuments. As a result of the developmental process, the mature seed will consist of a number of different gene expression programs occurring at the same time in the different compartments (e.g., embryo, endosperm, seed coat) as well as within specific regions and tissues (e.g. embryo proper, suspensor, epidermis) [6].

While the genome assembly of soybean (variety Williams 82) was completed in 2010 [8], regulatory networks of gene expression during soybean seed development remain poorly understood. One of the first studies in soybean seed development used RNA-seq to develop an atlas of soybean during seed development [9]. Subsequent studies identified and annotated several seed-specific genes in early seed imbibition [10], as well as analyzing gene expression patterns at all seed developmental stages from post-fertilization to mature seed [11]. The focus of several studies has been on identifying genes controlling agronomically important seed traits such as size, weight, oil and protein contents [12–15]. For those genes related to seed development, the gene networks and developmental processes controlled by LEAFY COTYLEDON1 (LEC1) were identified by comparing the mRNA profiles of wild-type and lec1-null mutant seeds at several developmental stages [16]. Nonetheless, our understanding of how cellular processes in tissues and regions of tissues are integrated to achieve the coordinated development of soybean seeds is still poor.

Studies of gene expression in seeds of other species have been a useful tool to better understand the biological processes required for proper seed development. For example, such studies have been used to determine that embryo and endosperm have distinct roles during the early development of rice seeds. In the endosperm, expression of plant hormone, galactose metabolism, ribosome and carbon fixation related genes increased significantly from 3 to 6 days after pollination (DAP), whereas genes for defense against disease or response to stress as well as genes for starch/sucrose metabolism were strongly expressed at 10 DAP [17]. In the embryo, metabolism, transcriptional regulation, nucleic acid replication/processing, and signal transduction related genes were expressed predominantly at 3–7 DAP, whilst genes for starch/sucrose metabolism and protein modification were highly expressed at 7–14 DAP [18]. Similarly, a genome-wide analysis of gene expression in several tissues and tissue subregions of Arabidopsis seeds during development revealed several limitations in our knowledge [19]. For example, expression of flavonoid biosynthetic genes was found to be specific to the Arabidopsis seed coat, which may protect seeds against biotic and abiotic stresses, and that the key enzyme required for the synthesis of trehalose is located in chalazal seed coat. The developmentally unique chalazal endosperm expressed the largest number of seed-specific mRNAs. It was concluded this region may serve as a communication hub that integrates seed developmental processes, based upon the finding that it expressed enzymes for the biosynthesis of several hormones, including gibberellic acid, abscisic acid, and cytokinin. A similar study was performed in *Brassica napus,* demonstrating that transcriptional profiles and functional enrichment during the globular stage of *B. napus* seed development is similar to those in the Arabidopsis embryo proper. However, different species have their distinct features during seed development. For instance, the micropylar endosperm in *B. napus* globular seeds was

found to significantly accumulate transcripts associated with lipid storage, transport and seed oil body biogenesis during the globular stage, while these GO terms were significantly enriched during maturation rather than in morphogenesis in Arabidopsis. This indicated that *B. napus* may accumulate lipids much earlier than in Arabidopsis [20]. In addition to seed development, the germinating seed is also an excellent model to study genome regulation between cell-types. Recently laser-capture microdissection RNA-seq was applied to plumule, radicle tip and scutellum of germinating barley from which tissue-specific gene expression and storage of transcripts was characterized [21]. In maize, transcriptome analysis of embryo/endosperm interfaces revealed how the embryo and endosperm communicate in order to regulate the production the impermeable cuticle. The endosperm zone adjacent to scutellum was found to have a distinct transcriptional signature, which is specialized in nutrient transport and influenced by the presence of the neighboring growing embryo [22]. In contrast, uncovering this type of comprehensive analysis in soybean is also urgently needed.

Here, we analyzed changes in the transcriptome of tissues of soybean seeds from the globular stage of embryogenesis in early seed development to the cotyledon stage of mature seeds using public RNA-seq data from Gene Networks in Seed Development project (http://seedgenenetwork.net/) [16]. Tissue specific processes, conserved between species, were identified using GO analysis, weighted gene co-expression network analysis (WGCNA) and orthology comparison to and Arabidopsis seed developmental series. These findings enabled the prediction of a set of high-confidence regulators of spatiotemporal-specific gene expression. The resulting analysis provides a comprehensive overview of gene regulation and expression with high spatial and temporal resolution, providing researchers with a roadmap to modify seed characteristics in a tissue specific manner.

2. Results and Discussion

2.1. Source of Spatial and Temporal Transcriptome Data during Soybean Seed Early Development

We obtained RNA-seq data of soybean seeds (Williams 82) during early seed development from the Gene Networks in Seed Development project (http://seedgenenetwork.net/) [16]. These data were obtained from eight seed compartments (embryo proper, endosperm, endothelium, epidermis, hilum, inner integument, outer integument and suspensor) over three stages of early seed development (globular stage, heart stage and cotyledon stage) using laser capture microdissection (LCM) (Figure 1A). Three biological replicates for each tissue were provided by the resource, comprising approximately 21 million raw reads per library. We mapped raw reads to the soybean reference genome resulting in 79.6 to 85.2% reads successfully aligning (Table S1). We classified a gene as expressed if its TPM (transcripts per million reads) value was ≥ 5. In total, 40,739 genes, including 2750 transcription factors (TFs), were expressed in at least one of the 72 samples. For each sample, an average of ~15,397 genes and 800 TFs were detected (Table S2).

To gain further insight into the transcriptome dynamics of early soybean seed development, all the seed tissues and stages (72 samples) were combined and compared using a principal component analysis (PCA) (Figure 1B). The three biological replicates for each sample clustered together, indicating high reproducibility of the data. Overall, the transcriptomes were more similar within the same tissue across the different stages rather than seed tissues clustering within the same stage, similar to gene expression in different tissues of germinating barley seeds [20]. The transcriptome profile in endosperm at the globular stage was distinct from those at the heart and cotyledon stages, which reflects the large changes in endosperm biology with growth occurring at the globular stage but degradation occurring during heart and cotyledon stage. The outer integument transcriptomes also showed differences from heart to cotyledon stage, indicating a change in gene expression patterns in late morphogenesis. In summary, the PCA analysis showed that gene expression in the embryo proper, endosperm and suspensor were more similar, probably because of their maternal and paternal origin, and more distant from the subregions of the seed coat that derived from maternal tissue only.
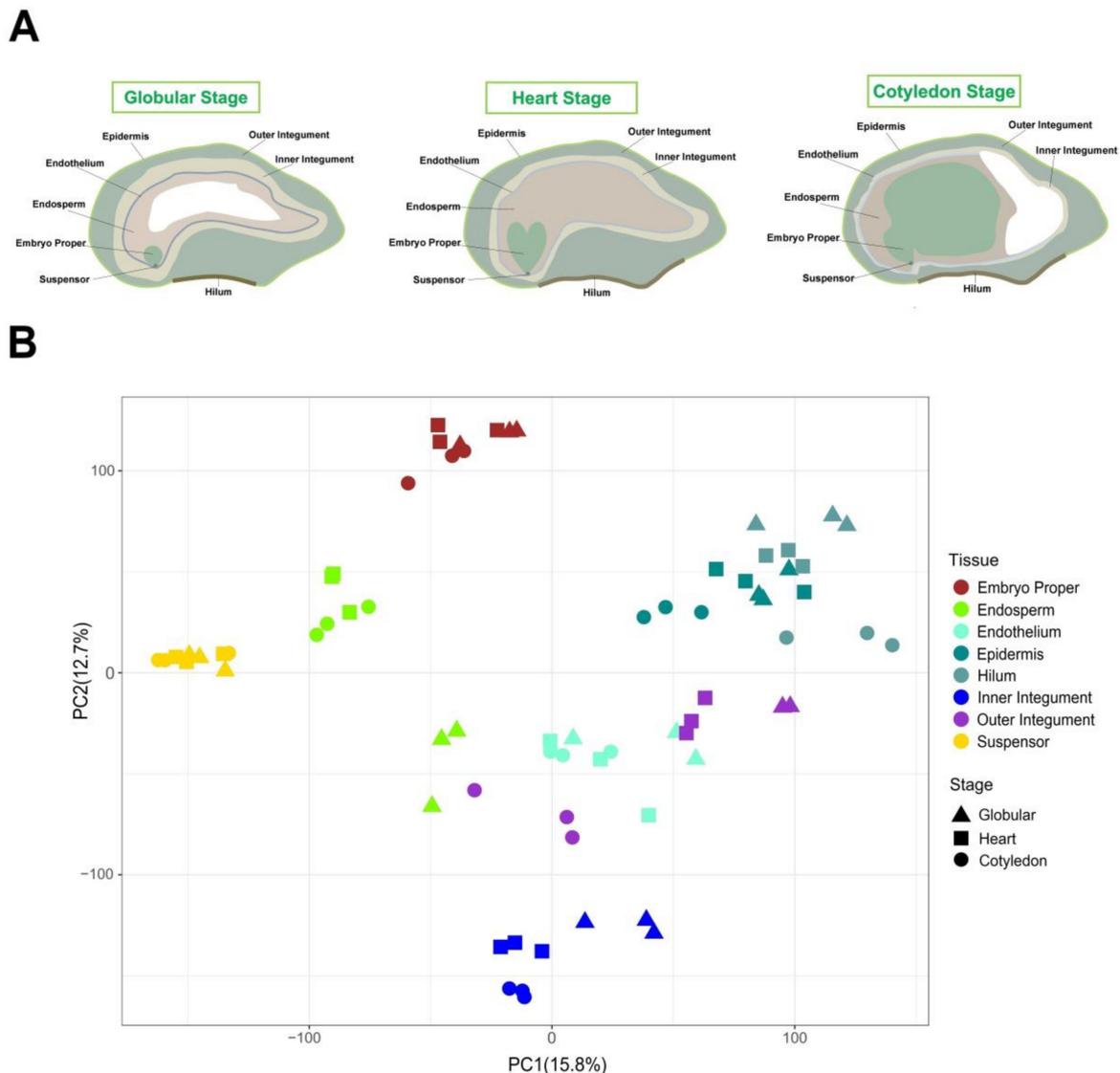
Figure 1. Transcriptome relationships amongst the 72 samples of early soybean seed development from the Gene Networks in Seed Development project. (**A**) Schematic diagram of the soybean seed over three developmental stages showing the various tissues dissected for analysis. (**B**) Principal Component Analysis (PCA) of the transcriptomes of three development stages and eight seed tissues.

We analyzed developmental progression in gene expression by comparing the globular stage datasets to the heart and cotyledon stage datasets, respectively. A total of 25,433 genes were differentially expressed (DEGs; FDR < 0.05; Dataset S1). The difference in gene expression was greatest in the endosperm, with 10,242 DEGs between globular and heart stages (S1) and 8999 DEGs between globular and cotyledon stages (S2) (Figure 2A). All seed tissues, except endosperm, had more DEGs in S2, indicating that gene expression in most regions changes dramatically during late morphogenesis when cell differentiation occurs. Nearly half (12,263 of 25,433) of the DEGs were differentially expressed in only one seed tissue, while the other half were differentially expressed over several tissues (Figure 2B). This is similar to previous findings from the seed development of Arabidopsis and barley [18,20]. Only three genes, Glyma.06G012600 (glycine-rich RNA-binding protein 2), Glyma.03G249000 (growth regulating factor-interacting factor1-like) and Glyma.04G122900 (Acyl-CoA-binding protein), were differentially expressed in all seed tissues (Figure 2B). Interestingly, the Arabidopsis glycine-rich RNA-binding protein 2 (soybean homolog Glyma.06G012600) has a role in RNA transcription and processing during stress and the rice homolog exhibited RNA chaperone activity

during cold and freezing adaptation to stress [23]. The growth regulating factor-interacting factors (GIF), AtGIF1/ANGUSTIFOLIA3, has a central role in the control of cell proliferation [24] and associates with the core SWI/SNF chromatin-remodeling complex to tightly regulate the transition between cell division and cell expansion in growing leaves [25]. Acyl-CoA-binding proteins (ACBP) had been found to participate in lipid trafficking and signaling. In Arabidopsis, the endoplasmic reticulum-localized AtACBP1 together with other proteins, regulated the generation of sterol signals, which are the putative ligands for TFs required for organ patterning and developmental gene expression [26]. AtACBP1 is localized at the plasma membrane of heart, torpedo, and cotyledon stage embryonic cells demonstrated its role in embryogenesis [27–29], while its homolog, AtACBP2, also accumulated in developing embryos at various stages [30].

2.2. Tissue-Specific Regulation of Gene Expression Occurs over Developmental Stages

To further understand the dynamics of the global transcriptome during early seed development, we analyzed the expression pattern of all tissue-specific DEGs relative to globular stage. The vast majority of DEGs followed a consistent trajectory of either up-regulation or down-regulation throughout all the stages (Figure 3A). The tissues with highest number of DEGs were endosperm, suspensor and embryo proper, consistent with these tissues being the most biochemically active tissues in the seed. These results highlight extensive tissue-specific gene expression during early seed development, indicating that the expression of the soybean genome is regulated in a tissue-specific manner. The potential functions of the tissue-specific DEGs were investigated in order to gain insight into the biological processes unique to each tissue. To achieve this, we applied PageMan over-representation analysis (z-score, *p*-value < 0.05, Figure 3B, Dataset S2). Significantly over-represented, up-regulated genes in the endosperm were involved in lipid and nucleotide metabolism and chromatin and cytoskeleton organization (Figure 3B). After fertilization in the endosperm, endoreduplication (DNA replication occurs without cytokinesis) occurs and the endosperm cells enlarge using microtubules that emanate from the nuclear envelope and determine the size and shape of the future cells. Microtubule arrangement has been shown to be important in the cellularization of the endosperm [31,32]. The involvement of cytoskeletal elements in protein body formation has also been proposed [33]. Contrastingly, proteolysis was significantly enriched amongst the down-regulated endosperm specific DEGs, which is consistent with the function of protein storage in this tissue [34].
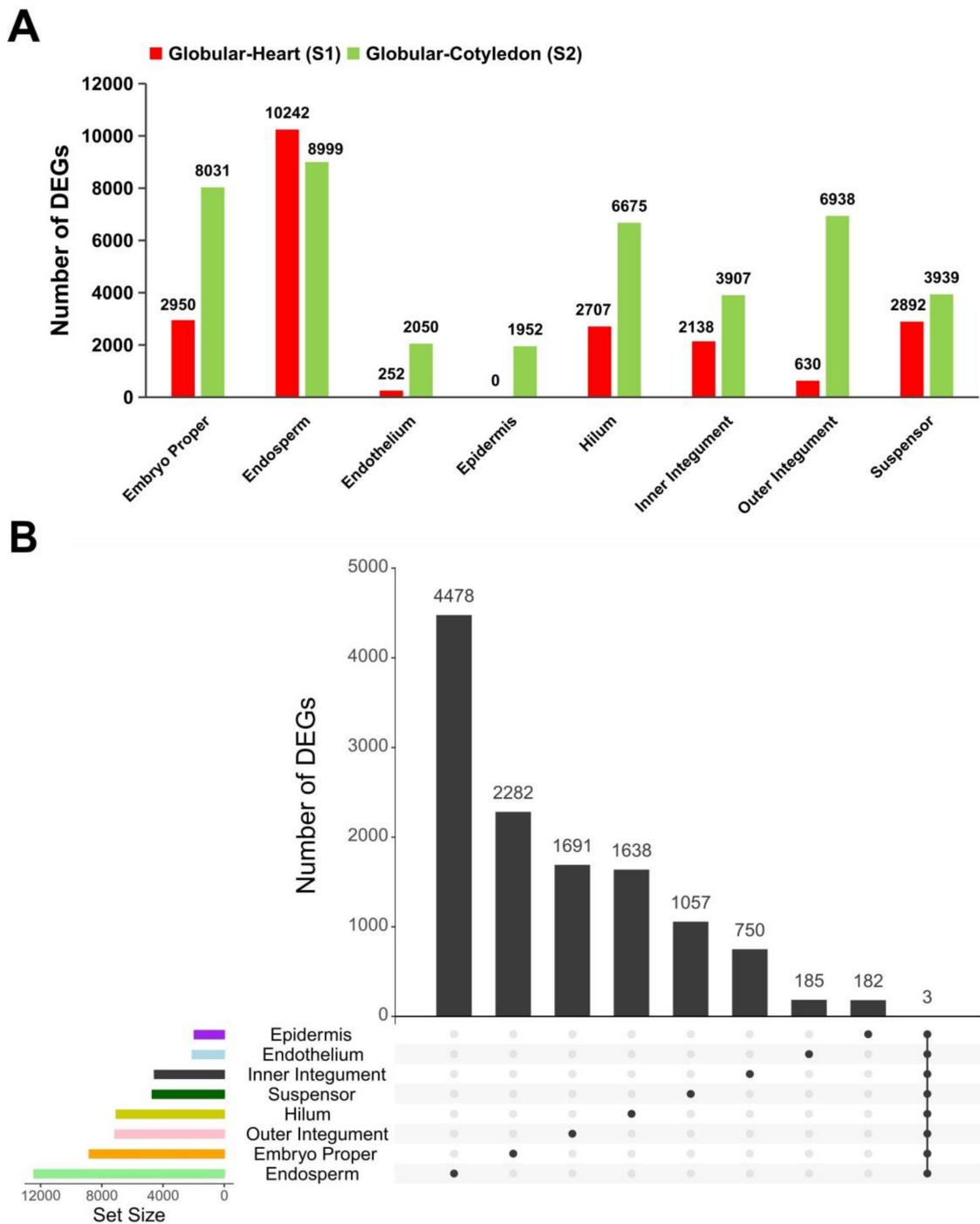
Figure 2. Overlapping and specific responses in soybean seed tissues. (**A**) Number of DEGs between globular-heart stage, and globular-cotyledon stage. (**B**) UpSetR plot shows overlapping and specific responses in seed tissues.
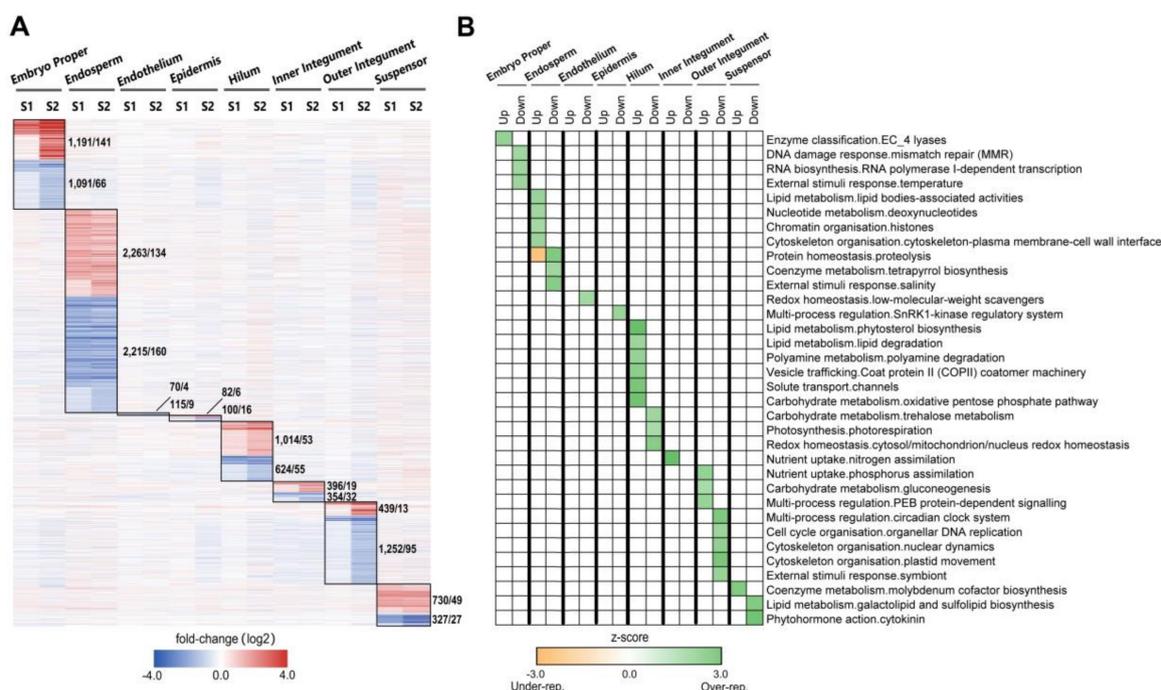
Figure 3. Functional analysis of tissue-specific DEGs. (**A**) Hierarchical clustering of fold-changes in transcript abundance for all specific DEGs in each seed tissue. S1, Heart stage relative to Globular stage; S2, Cotyledon stage relative to Globular stage. The number of genes and TFs in each box are listed on the right. (**B**) PageMan over-representation analysis (Fisher's test, *P* <0.05) of specific DEGs in each seed tissue. Over-represented and under-represented functional categories specific in one tissue are shown (full output in Dataset S2).

The hilum is the tissue that connects the seed to the pod and gene expression in this tissue is not well studied. Our analysis determined that highly up-regulated hilum-specific genes were involved in vesicle trafficking and solute transport. It is also evident from its role in nutrition, lipid and carbohydrates metabolism and polyamine degradation (Figure 3B). These results indicated the importance of this tissue in nutrition and pod-seed communication. Contrastingly, for a sub-region of seed coat, the integuments, tissue-specific regulation of genes involved in nutrient uptake (nitrogen and phosphorus), carbohydrate and lipid metabolism was evident. These results reflect the importance of nutrient allocation during early seed development. Our analysis demonstrated a special role of outer integument in controlling of cytoskeleton organization, specially, nuclear dynamics and plastid movement, which were significantly enriched amongst the down-regulated outer integument-specific DEGs.

2.3. Predicting Candidate Transcriptional Regulators Using Weighted Gene Correlation Network Analysis

We defined co-expressed cohorts of genes (termed modules) in order to identify potential spatiotemporal-specific regulators that may underpin the tissue specificity of the gene expression outlined above. To do so we applied weighted gene correlation network analysis (WGCNA). Genes were filtered prior to WGCNA to remove those with low variability (Coefficient of Variance < 0.5) or expression (TPM < 5) across samples because these would not contribute to the network, leaving 30,930 genes for analysis. After generating a summary profile (eigengene) for each module, we observed that 18 of the 21 modules were predominantly co-expressed in a single seed tissue (Figure 4). Three out of 18 co-expression modules, including the brown, grey60 and purple modules, were specifically expressed in embryo proper. Notably, the brown module, which contains 1289 genes, was mainly enriched in protein translation and ribosome biogenesis related processes. Ribosomal proteins were highly expressed in the globular stage and coordinated with embryogenesis in Arabidopsis [35]. Therefore, the brown module genes might be responsible for the initiation of embryogenesis. Moreover, the gradual reduction of

expression of brown module genes from globular to cotyledon stage implies the ribosome biogenesis activity was gradually decreasing during embryo proper establishment, with the ribosome machinery established for germination where translation is essential and a very early process [36]. It is known that photosynthesis occurs in the embryo, and embryonic photosynthesis affects post-germination plant growth [18,35,37,38]. In line with this, grey60 and purple modules were significantly enriched in photosynthesis related GO terms, indicating embryonic photosynthesis might occur during the heart stage and cotyledon stage. It is notable that, unlike the brown module, the expression of purple module genes increased from globular to cotyledon stage, suggesting a functional transition of the embryo from globular to cotyledon state (Figure 4B, Dataset S3). The integuments, including outer and inner integuments, are the interface of maternal tissues and endosperm. Interestingly, the genes from integument modules, such as light green, turquoise and greenyellow, were significantly enriched in transport related functions including transmembrane transport, protein transport and carbohydrate transport. These enriched functions suggest the integuments may play an important role in transporting metabolites such as protein and carbohydrates from maternal tissues to the endosperm and embryo for their growth. The epidermal cells of the seed coat protect seeds against environmental stresses, partly by accumulating toxic compounds such as flavonoids and terpenoids [39,40]. In agreement with this, the epidermis-specific yellow module was highly enriched in flavonoid biosynthesis processes, confirming the protective function of the seed coat. The suspensor is a supportive structure that pushes the embryo close to the nourishing endosperm and is involved in nutrient and hormone regulation, such as auxin transport [18,41,42]. Consistent with this, auxin-activated signaling pathway genes were enriched in the pink module, which was predominantly expressed in the suspensor. Two genes (Glyma.02G007300, Glyma.10G138500), which were both annotated as indole-3-acetic acid inducible 30 (IAA30), were strongly associated with the pink module (kMEs 0.99 and 0.98, respectively).
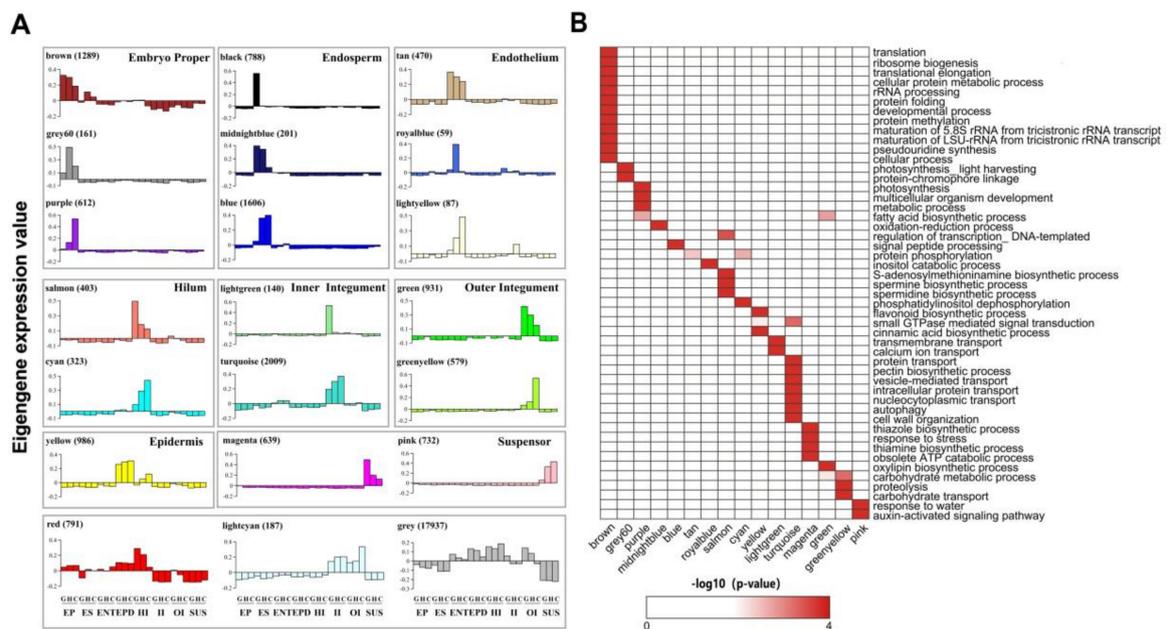
Figure 4. Eigengene expression patterns for each module eigengene. (**A**) Eigengenes represent the expression of their co-expression modules across all samples, in arbitrary units. We categorize the eigengenes according to the seed tissue(s) in which they are predominantly expressed. Numbers in parentheses correspond to the number of member genes in the module. G-globular stage, H-heart stage, C-cotyledon stage, EP-embryo proper, ES-endosperm, ENT-endothelium, EPD-epidermis, HI-hilum, II-inner integument, OI-outer integument, and SUS-suspensor. (**B**) Significant enriched GO terms (biological process) of tissue specific WGCNA co-expression modules (*P* < 0.001).

We identified 332 transcriptional regulators (Table S3) that were candidate regulators of the modules using the following criteria; (a) TFs having high correlation with the expression profile of the module to which they belonged (kME > 0.9, termed hub genes), and (b) TFs themselves significantly differentially expressed (FDR < 0.05, Dataset S1). Twenty of the 21 modules contained candidates, which encompassed a wide range of transcriptional regulator families. Most TFs from the same family were predicted to co-regulate modules from one or two tissues. For instance, all 15 plant-specific transcription factor YABBY family members were only identified as hub regulators in embryo proper modules. The YABBY family TFs were only found in seed plants and are involved in the regulation of adaxial–abaxial polarity and development of lateral organs [43–46]. The enrichment of YABBY family genes in the embryo properly indicated that these TFs might also have an important role in embryo development. Interestingly, there were six TFs belonging to four different families specifically expressed in the black, endosperm-specific module, including ETHYLENE RESPONSE FACTOR 1 (ERF1, Glyma.13G122700), CYTOKININ RESPONSE FACTOR 4 (Glyma.14G123900), MADS-box transcription factor family protein (Glyma.05G140000 and Glyma.18G053400) and TELOMERASE ACTIVATOR 1 (TAC1, Glyma.10G184500 and Glyma.20G206000). Genetic variation of *ERF1* was associated with thermotolerance in lettuce seed germination [47]. Our results imply that ERF1 might not only play a role in seed germination, but also seed development, specifically, endosperm development. Telomerase expression is mainly observed in reproductive organs and rapidly dividing cells [48,49]. After fertilization, endosperm undergoes rapid proliferation without the synthesis of cytoplasm, cell membranes and cell walls, and cell division in endosperm is faster than in embryo [50,51]. Given the high expression of TAC1 in endosperm, it is reasonable to think that TAC1 is critical during the endosperm development. However, there were some transcription factor families predicted to regulate more than one seed tissue, such as AP2, bHLH, BLH, C2H2, homeodomain, MYB, NAC and WRKY. For example, one soybean homologue of AtNAC87 in Arabidopsis (Glyma.19G109100), which has been reported to be involved in seed development [52], was identified as a candidate regulator of the inner integument. However, several other NAC family genes, such as NAC2, NAC44 and NAC47, were mainly expressed in endosperm, suspensor and endothelium (Table S3). The Arabidopsis *AGAMOUS* (AG) encodes a MADS box transcription factor, which is involved in flower development [53,54]. Several homologues of AG in *Brassica napus* and strawberry (*Fragaria vesca*) were predicted as putative regulators of seed coat development [19,55]. We observed 15 AGAMOUS-like (AGL) genes that were highly expressed in the endosperm and the endothelium, indicating AGL genes might also regulate endosperm and endothelium development in soybean. The identification of many distinct modules of co-expressed genes with distinct expression profiles that contain a large number of candidate regulators, illustrates the complexity of the tissue-specific regulation in the soybean seed.

2.4. Integration of Orthology and Expression Information Identifies Conserved Orthologues between Soybean and Arabidopsis

Soybean and *Arabidopsis thaliana* are two important model organisms for dicots, both of which progress through globular, heart and cotyledon stages during early seed development. To carry out an unbiased comparison of seed development transcriptomes between these species, public Arabidopsis seed development datasets were obtained from Gene Expression Omnibus (GEO) database (GSE11262, GSE15160, and GSE12403) [19,56]. OrthoFinder was used to define sets of orthologous genes in the two species [57]. A total of 12,344 orthogroups were defined between soybean and Arabidopsis representing 18,110 genes in soybean and 26,975 genes in Arabidopsis (Table S4). Orthogroups containing genes in soybean and Arabidopsis were subsequently analyzed with Clust to define sets of co-expressed orthologous genes within and across species [58]. The embryo proper and micropylar endosperm data were available in the public Arabidopsis data; thus, we only compared these two tissues here.
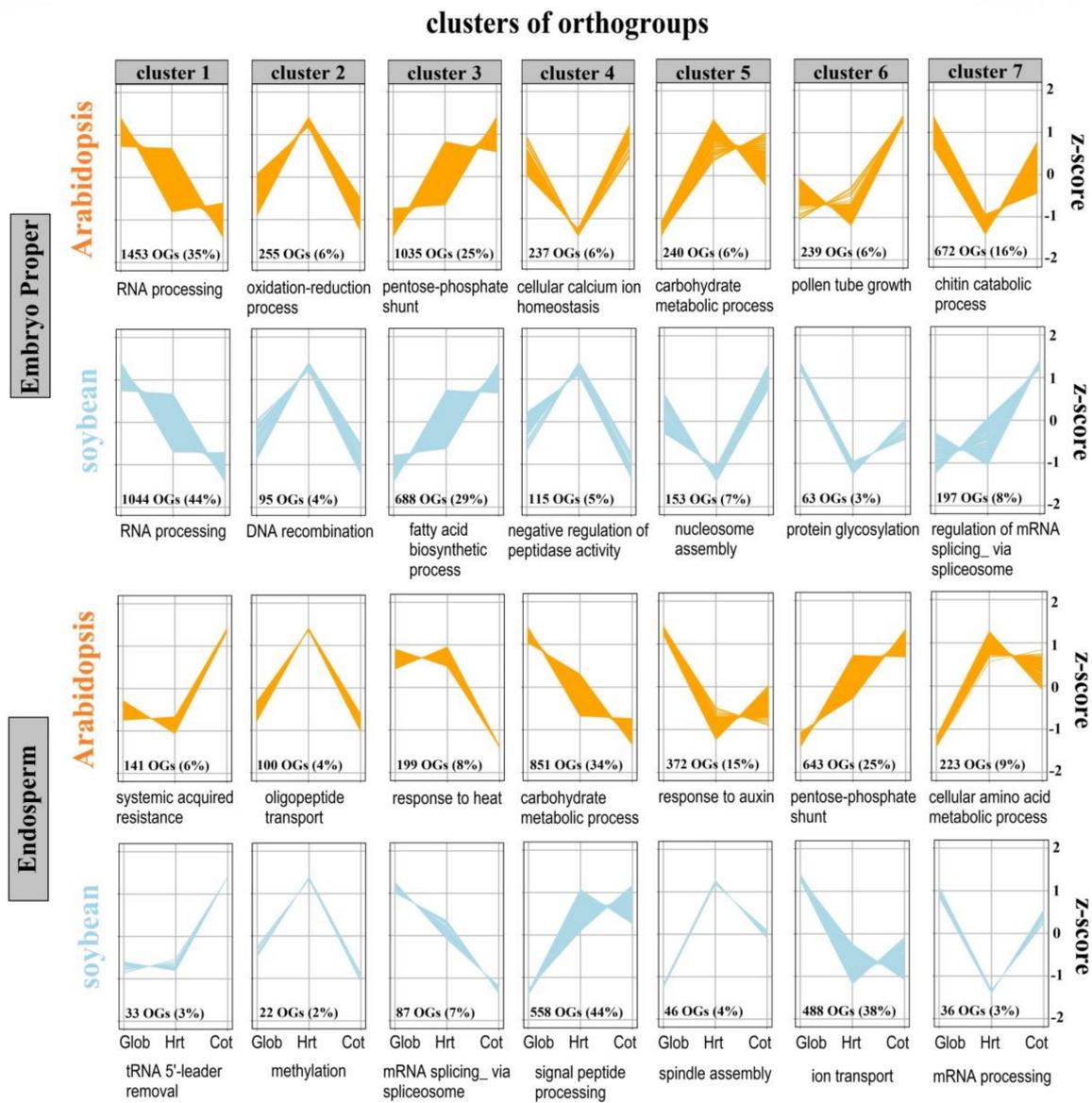
For soybean and Arabidopsis embryo proper, a total of 2,355 and 4,131 orthogroups were distributed in seven clusters, respectively (Figure 5, Table S5A). The gene expression pattern in clusters 1, 2 and 3 was highly conserved between soybean and Arabidopsis. Subsequently, a GO term enrichment analysis

was performed for the genes in the orthogroups included in each of the clusters (Table S6). Expression of genes in cluster 1, which represents 44% and 35% of the orthogroups in soybean and Arabidopsis, respectively, declined during seed early development. Genes in this cluster were associated with enriched GO terms related to RNA processing, mRNA processing and embryo sac egg cell differentiation. These likely represented the breakdown of RNA with modified functions reserved in embryo proper during seed development. By contrast, the orthogroups in cluster 3 contained genes for which transcript abundance increased and the GO term enrichment was associated with biosynthetic cysteine and fatty acid biosynthetic processes. These indicated the requirements for the synthesis of amino acids during seed development in soybean and Arabidopsis embryo proper. Another conserved cluster, cluster 2, included those orthogroups whose expression peaked at heart stage. The genes of orthogroups in this cluster were associated with DNA recombination in soybean and oxidation-reduction and steroid metabolism in Arabidopsis. Other clusters (4, 5, 6 and 7) contained genes with opposing co-expression profiles between the two species. This highlights that, whilst processes may be conserved overall, the spatiotemporal expression of some genes will have diverged between species, and consequently translational comparative studies from one species to another are essential.

2.5. Tissue-Specific Expression of Transcriptional Regulators during Early Seed Development
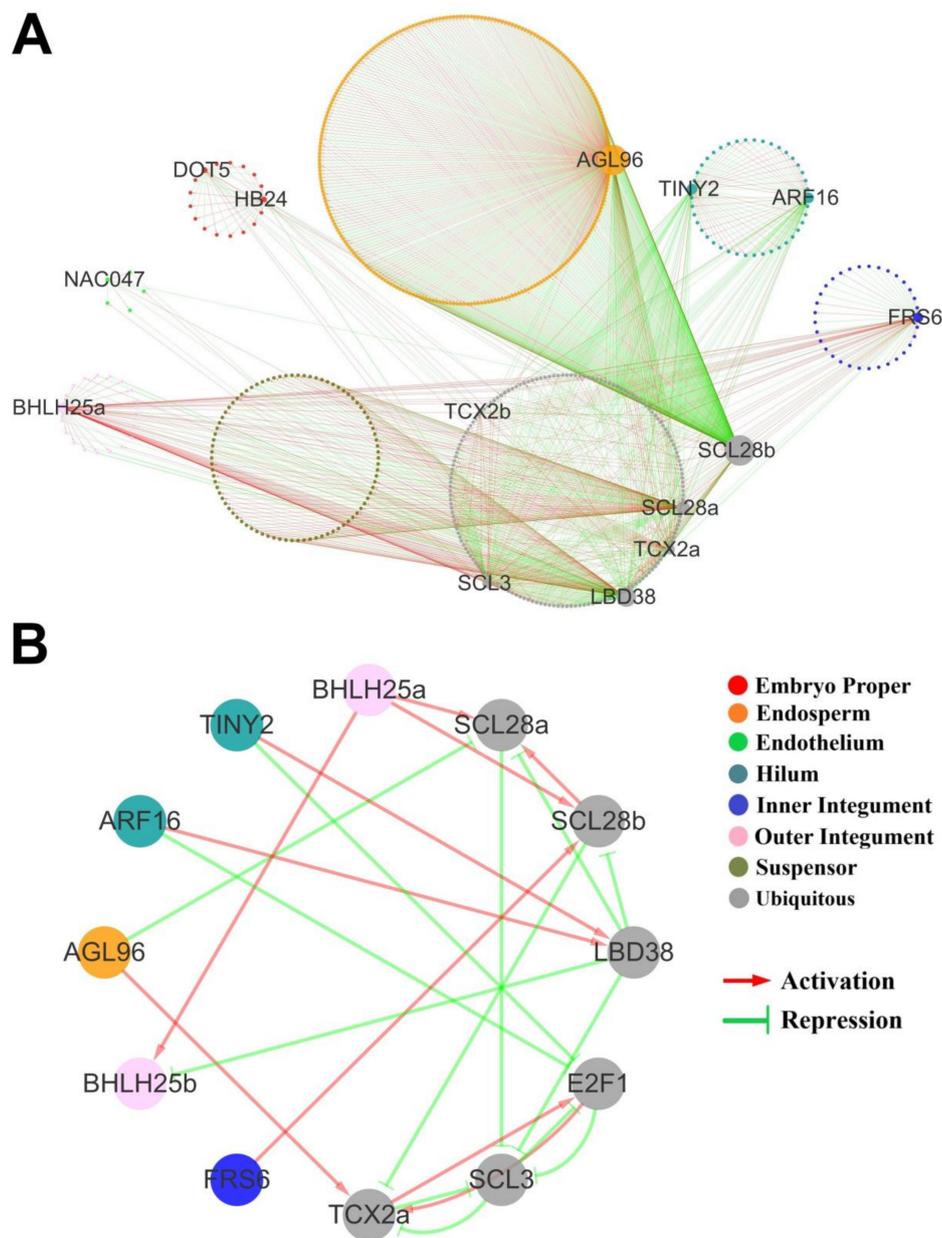
To understand how tissue-specific genome regulatory programs are established we analyzed the differential expression of TFs between soybean tissues. In total, 879 tissue-specific TFs were identified (Dataset S4). We reasoned that these were strong candidates to control tissue-specific genome regulatory programs and wanted to determine which families of TFs were operating in each seed tissue. A total of nine out of ten families were significantly enriched amongst the DEGs specific to one seed tissue (Figure 6A; *P* < 0.005; embryo proper - YABBY, TALE; endosperm - M-type_MADS, MIKC_MADS, B3, BBR-BPC; epidermis - EIL; hilum - MYB; suspensor - NAC). The ERF TF family was the only enriched family shared between tissues (endothelium and epidermis). There was no enrichment of any transcriptional regulator family in inner or outer integument specific DEGs.

Binding of TFs to regulatory regions of the genome usually requires the presence of DNA motifs that TFs recognize. We examined DNA motifs enriched in the promoter regions of our tissue-specific DEGs in order to identify the families of TFs likely to be active in each seed tissue (Figure 6B). Several motifs significantly associated with promoters of DEGs were identified de novo in each tissue (E-value < 1e-10). The BASIC PENTACYSTEINE (BPC) family (BPC1, BPC5 and BPC6) had the most enriched motifs in the promoters of DEGs in embryo proper, endosperm, endothelium, epidermis and hilum. The BPC TF family is integral in a wide range of processes that support plant normal growth and development [59]. Another universally enriched motif, associated with TF OBP3, was identified in embryo proper, endosperm and suspensor, but not significantly enriched in any sub-region of seed coat. Mis-expression of this TF correlates with growth defects [60]. Our results also supported the tissue-specific regulation of OBP3. Similarly, RAMOSA1, NUC and MGP were only significantly enriched in embryo proper and endosperm. In contrast, Adof1 was only enriched in the sub-regions of seed coat (hilum and outer integument). The most enriched motif in hilum had no significant similarity to motifs from previously characterized plant TFs, demonstrating that there are uncharacterized transcriptional regulators active in hilum. These analyses considered together indicated that a diverse range of TFs were likely to contribute to genome regulation and have tissue-specific and common features during early seed development. Our analysis provided a better understanding of the spatiotemporal expression of known transcriptional regulators.

Figure 5. Comparison of embryo proper and endosperm transcript abundance profiles during seed development in soybean and Arabidopsis. For cross-species comparison of seed development, changes in transcript abundance in soybean were compared to published data in Arabidopsis [19]. The numbers in each panel correspond to the number of orthogroups in each cluster and the percentage of all orthogroups per species. The most significantly enriched GO term in each cluster is given below each panel, and the full list is available in Table S6.

Figure 6. Genes encoding transcriptional regulators are tissue-specifically expressed during soybean seed development. (**A**) Transcription factor family enrichment of tissue-specific differentially expressed transcription factors in each seed tissue ($P < 0.005$). (**B**) Motifs significantly enriched ($E \leq 1e\text{-}10$) in promoters of genes specific expressed in each seed tissue were identified de novo. Enriched motifs were compared to known motifs and families of significant matches to plant transcription factors are indicated.

2.6. Gene Regulatory Network (GRN) of the Seed Tissue-Specific Genes

GRNs are a useful tool to better understand the relationships between regulators and their targets. In this study, we inferred the GRN for each seed tissue separately (eight networks for each tissue) using genist algorithm from TUXNET [61] and then combined them to generate the final network. To discover the tissue-specific regulatory relationship among the key regulators for each network, genes from two sets were used for network inference. The first gene set included genes that overlapped between tissue-specific DEGs and hub genes (kME > 0.9) of this tissue specific WGCNA module, and the second set contained genes (ubiquitous genes) that were differentially expressed between multiple tissues (including this tissue) and also identified as hub genes in the non-tissue-specific WGCNA modules (red, grey, and lightcyan). In order to obtain robust final networks, only the most reliable regulatory relationship was retained. Interestingly, almost all genes in each tissue were regulated by one or two genes, which helped us find the core genes (Figure 7, Table S7).

Figure 7. Predicted gene regulatory network (GRN) of early developmental soybean seed. (**A**) Network hubs regulating genes in each seed tissue. Node size represents the degree and key regulators with high degree are highlighted. Node colors represent the tissues which the DEG belong to. Edge colors represent the regulatory relationships is inferred. Red: activation; Green: repression. (**B**) The regulatory relationship between the key regulators in each tissue and ubiquitous regulators with high degree.

The endosperm network was the largest, with 600 interactions including 259 endosperm-specific and 43 ubiquitous genes. Interestingly, most endosperm-specific genes (205/258) in this network, including a set of AGAMOUS-LIKE (AGL) genes (AGL6, AGL20, AGL61, AGL62, AGL80), were activated by an endosperm-specific gene AGL96 (Figure 7A). Although the function of AGL96 is still unknown, six AGLs (AGL36, AGL37, AGL38, AGL40, AGL62, AGL90) were co-expressed in the endosperm and absent in the embryo in Arabidopsis [62]. As the reported central hub of the AGL cluster in Arabidopsis, AGL62 was revealed to be essential for the normal development of seeds and avoid premature death [63]. Although most members of Arabidopsis AGL co-expression cluster were not consistent with the endosperm-specific network in soybean, AGL96, the ortholog of the core

regulator of the Arabidopsis AGL co-expression cluster AGL62, was identified in a core position in the soybean network. This indicated that similar AGL clusters may also play a critical role in maintaining seed activity during soybean endosperm development.

Unlike the AGL genes that mainly functioned in the endosperm, three SCARECROW-LIKE (SCL) genes were found as ubiquitous genes. For example, two paralogue genes, SCL28a and SCL28b, regulated all genes in soybean suspensor and endosperm networks, respectively. In Arabidopsis root, SCL28 was found to orchestrate cell division and elongation under various environmental conditions [64]. This suggests SCL28 may promote soybean seed growth by regulating cell division in the suspensor and endosperm. In addition, another SCL ubiquitous gene, SCL3, which is involved in gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis seeds [65], was predicted to regulate nearly all ubiquitous genes in epidermis and endothelium. Given SCL3 was predicted to be repressed by SCL28a, SCL28 may promote soybean seed development by indirectly regulating gibberellin homeostasis.

In the hilum-specific network, all genes were inferred to be regulated by AUXIN RESPONSE FACTOR16 (ARF16) and TINY2 (Figure 7A). This was consistent with their roles of promoting and inhibiting cell growth and proliferation by responding to auxin and brassinosteroid in Arabidopsis [66,67]. Moreover, a ubiquitous gene E2F1, which was reported to regulate cell proliferation by acting as a target of auxin in Arabidopsis [68], was predicted to be repressed by ARF16 and TINY2 in soybean (Figure 7B). In addition, E2F1 was also inferred to be negatively regulated by SCL3, whose potential involvement in gibberellin signaling in epidermis and endothelium has been described in the previous paragraph. These indicated that in the soybean seed coat the above core regulators may form a previously uncharacterized pathway, coordinating with plant hormones to regulate seed development. In summary, our analysis provided a solid foundation for the identification of key regulators active in seed tissues during early seed development.

3. Summary

In this study, we analyzed a transcriptomic dataset from 8 seed tissues (embryo proper, endosperm, suspensor, endothelium, epidermis, hilum, inner integument and outer integument) at 3 developmental stages (globular stage, heart stage and cotyledon stage) during soybean seed morphogenesis (Figure 1). Our analyses determined that tissue specificity of gene expression was stronger than developmental stage specificity, as reported in other plant species [19,21]. Nearly half of the DEGs detected were specific to an individual seed tissue.

Many of the predicted gene expression regulators identified by our GRN functioned tissue-specifically, which is consistent with the large proportion of tissue-specific gene expression during seed development. However, not all regulators were predicted to have tissue-specific function. This likely indicates that establishing the unique functions of individual tissues requires the activity of regulators specific to each tissue, but that there is also a set of core cellular functions conserved between a number of tissues. Many predicted soybean regulators had Arabidopsis orthologs with known functions in seed development. The GRN approach predicted a previously uncharacterized role for the transcription factor AGL96, proposing that it is a central regulator of endosperm gene expression. This may be comparable to its ortholog AGL62, which is co-expressed with a group of AGL transcription factors in the Arabidopsis endosperm but not expressed in the embryo of that species. Overall, the GRN analyses indicated that there is relatively high connectivity between tissue-specific and ubiquitous regulators during seed development, including repression, activation and feedback loops.

A large proportion of the genes differentially expressed in the embryo proper and endosperm during seed development was also conserved between soybean and Arabidopsis. Using a strict definition of orthology, ~65% of orthologous genes exhibited the same trend in transcript abundance between species in the embryo proper (Clusters 1, 2 and 3, Figure 5). However, the remaining ~35% exhibited opposite trends in the expression between species. The differing annotation quality of the soybean and Arabidopsis genomes makes direct comparison of functional enrichment between

species challenging. However, based on the clear conservation in gene expression trends alone, we can conclude that there are likely conserved units comprised of orthologous genes that perform base seed developmental processes. Contrastingly, the cases of opposite expression trends likely indicate processes that vary between the seeds of the two species, due to their different morphology and biochemistry. These differences are important to consider when designing translational applications of GRNs that use data from another species.

Our study as a whole provides a detailed spatiotemporal view of gene expression and regulation during early soybean seed development. Our work provides the research community with a better understanding of the expression characteristics and functions of different seed compartments. Insight into the regulators and their targets during seed development is an essential tool when designing strategies for the biotechnological improvement of seed performance.

4. Materials and Methods

4.1. RNA-Seq Data Quality Control Analysis

FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used for initial reads quality control (QC) metrics (base quality distribution). NGS QC toolkits were used to trim adaptors and low-quality reads [69]. The *G. max* reference genome (Wm82.a4.v1) and annotation files were downloaded from Phytozome V13 (https://phytozome.jgi.doe.gov) [70]. RNA-seq reads were mapped to the *G. max* reference genome using hisat2 V2.1.0 with default settings [71]. The bam files of uniquely mapped reads were used as inputs for the Stringtie [72], and TPM (transcripts per million reads) values were calculated to measure the expression levels of genes. To reduce the data noise, we defined a gene as expressed if its TPM was ≥ 5. To determine the variation of the RNA-seq samples, we performed principle component analysis (PCA) using plotPCA function in DESeq2 R package [73].

4.2. PageMan Analysis of DEGs

Lowly expressed genes (average TPM < 5) were filtered out and differential expression analyses was carried out using edgeR [74]. DEGs were identified as those with a $P < 0.05$ and FDR < 0.05. UpsetR was used to show overlapping and specific responses in tissues and stages [75]. The significant DEGs were analyzed for overrepresented functional categories using PageMan (Mapman v. 3.5.1 R2) [76]. To do this, a custom mapping file for soybean was generated using Mercator, with all available query databases. Using this mapping file and the DEGs, over-represented functional categories were identified among the DEG sets using Fisher's test ($P < 0.05$). Given the highly conserved changes in differential expression over stages, a single DEG set was generated for each tissue using the maximum (for up-regulated genes) or minimum (for down-regulated genes) fold-changes. The outputs from these PageMan analysis were then visualized as heatmaps indicating the z-scores (> 1.96 indicates $P < 0.05$).

4.3. Identification of Co-Expression Modules and Hub Genes

Prior to analyze co-expressed genes, we calculated the coefficient of variation (CV) of each gene based on their TPM values and filtered the genes with low CVs (< 0.5), removing low expression and/or low variability of expression across samples. Next, the R package WGCNA (Weighted Correlation Network Analysis) was used to identify modules of highly correlated genes across all samples [77]. The parameters of WGCNA used default settings, except for the power was 12, network type was signed, minModuleSize was 30. Eigengenes and clusters were calculated based on the correlations to quantify the co-expression similarity of entire modules, using a strict cut-off of 0.25, corresponding to correlation of 0.75. The eigengene expression values were calculated for each sample in each module. Then, kME, known as the module membership value, was calculated using SignedKME algorithm to represent the correlation between a gene and the module eigengene value. Intramodular hub genes were identified if their kME > 0.9.

4.4. Orthologues Identification between Soybean and Arabidopsis

Orthologues and corresponding orthogroups between Arabidopsis and soybean were inferred via OrthoFinder v. 2.3.12 (https://github.com/davidemms/OrthoFinder) with default parameters and MMseqs2 for sequence similarity searches [78]. Protein sequences used here correspond to the longest transcripts of all genes of soybean (Wm82.a4.v1 release) and Arabidopsis (TAIR10 release), and they were acquired from phytozome v13 (https://phytozome-next.jgi.doe.gov/). The expression data of Arabidopsis seed were acquired from the Gene Expression Omnibus (GEO) database (accession no. GSE11262, GSE15160, GSE12403). Co-expression analysis of orthogroups was performed using the Clust software v. 1.12.0 with default parameters [58].

4.5. Transcription Factor and Promoter Motif Enrichment

All soybean TFs were annotated using PlantTFDB v. 5.0 (http://planttfdb.cbi.pku.edu.cn) [79]. Significantly enriched TF families were identified by R package clusterProfiler (*p*-value < 0.005) [80]. All TFs expressed in this study were defined as background genes, those tissue-specific DE-TFs were used for enrichment. The promoter regions (1000 bases upstream the transcription start site) of tissue-specific DEGs were examined to identify enriched DNA motifs that may correspond to TF-binding sites. These were analyzed using MEME-ChIP (http://meme-suite.org/tools/meme-chip) with default parameters against the built-in motif library and a background of all soybean promoter regions [81]. Enriched motifs were reported if E-value < 1e-10, as suggested by the MEME-ChIP documentation. Significant similarities to known plant motif families were reported. Similarity to non-plant motifs was not reported.

4.6. Gene Regulatory Network Inference

Networks were inferred for each seed tissue separately (resulting in eight networks, one for each tissue) using genist algorithm from TUXNET [61] and then combined to form the final network. For each network, only a) tissue-specific DEGs and overlapped with hub genes of tissue-specific WGCNA modules (kME > 0.9) and b) DEG differentially expressed in multiple tissues (ubiquitous) and overlapped with hub genes of non-tissue-specific WGCNA modules were used in the network inference. To get the regulatory relationship of the genes both in the same time point or in the following time point, the time lapse value was set as (0,1). Other parameters of GENIST were set with default values.

Supplementary Materials: The following are available online at http://www.mdpi.com/1422-0067/21/20/7603/s1. Figure S1: Gene co-expression modules identified using weighted gene correlation network analysis (WGCNA). The figure illustrates module-sample relationships for each sample from each organ at each developmental stage. Rows correspond to a module eigengene identified during WGCNA. Module eigengenes are proxies that represent the expression profile of the genes within the module and are named by the unique colours far left. The correlation coefficient between the module and the sample is given at the top in each cell and is also represented by the colour of the cell. The p-value for the correlation is also given in each cell (bottom). Supports Figure 4. Table S1: Summary statistics of RNA-Seq reads and mapping. Table S2: Number of genes and transcription factors expressed in this study. Table S3: Candidate transcriptional regulators identified amongst the hub genes (kME (module membership value) > 0.9 and differentially expressed). Table S4: Orthologous genes in soybean and Arabidopsis. Orthologous genes of soybean and Arabidopsis were grouped (orthogroups) using the Orthofinder algorithm as described in the Methods section. Table S5: Arabidopsis and soybean genes grouped into orthoclusters in embryo proper (**A**) and endosperm (**B**) as shown in Figure 5. Given are normalised expression values using the clust algorithm. Shared orthogroups (OG) between Arabidopsis and soybean in each cluster were highlighted in orange and same OGs were in same row. For each OG, the function of one gene of this OG were given. Table S6: GO (biological process) term enrichment for Arabidopsis and soybean genes in embryo proper (**A**) and endosperm (**B**) orthogroup clusters (*p* value < 0.05). Table S6: High confidence gene regulatory networks (GRN) in different tissues of early developmental soybean. Dataset S1: Table of 25,433 significantly DEGs identified when samples were compared with the globular stage sample from the same tissue (*p* < 0.05, FDR < 0.05). Numerical values are log2 fold-change vs. the reference sample. S1-heart stage vs globular stage, S2-cotyledon stage vs globular stage. Supports Figure 2. Dataset S2: Over-representation analysis output for gene ontology terms of all DEGs in each organ, determined using PageMan. Significantly represented functional categories are indicated by z-score (Fisher's exact test) - positive scores (green) indicate over-representation, negative scores (yellow) indicate under-representation, and absence of score indicates not-significant. The maximum fold-change (for up-regulated genes) and minimum fold-change (for down-regulated genes) were used as the input fold-changes for the analyses.

Supports Figure 3. Dataset S3: Information on highly interconnected genes within each co-expression module. The module membership is defined by the kME value (eigengene-based connectivity measure). Values close to 0 mean that a gene is not part of a given module. Values close to 1 or -1 indicate highly connected module genes. The sign of the kME value indicates whether a module has a positive or negative relationship with the module eigengene (representative of gene expression profiles in a module). This value is also highly related to intramodular connectivity between genes, with hub genes generally having high kME values. Each individual worksheet that follows gives the hub genes (kME > 0.9) and significantly enriched GO terms for each module (biological_process, FDR < 0.05). Supports Figure 4. Dataset S4: Table of 879 tissue-specific differentially expressed transcription factors identified when samples were compared with the globular stage sample from the same tissue ($p < 0.05$, FDR < 0.05). Numerical values of log2 fold-change vs. the reference sample are shown for each comparison. S1 - heart stage vs globular stage, S2 - cotyledon stage vs globular stage. Supports Figure 6.

Abbreviations

AGL	AGAMOUS-like
ARF	AUXIN RESPONSE FACTOR
CV	Coefficient of variation
DAP	Days after pollination
DE	Differentially expressed
DEG	Differentially expressed gene
ERF	ETHYLENE RESPONSE FACTOR
GEO	Gene Expression Omnibus
GIF	Growth regulating factor-interacting factor
GO	Gene Ontology
GRN	Gene regulatory network
LCM	Laser capture microdissection
PCA	Principle component analysis
SCL	SCARECROW-LIKE
TF	Transcription factor
TPM	Transcripts per million
WGCNA	Weighted gene co-expression network analysis

References

1. Goldberg, R.B.; Barker, S.J.; Perez-Grau, L. Regulation of gene expression during plant embryogenesis. *Cell* **1989**, *56*, 149–160. [CrossRef]
2. Johnson, S.; Liu, C.M.; Hedley, C.L.; Wang, T.L. An analysis of seed development in pisum-sativum. The isolation of mutants defective in embryo development. *J. Exp. Bot.* **1994**, *45*, 1503–1511. [CrossRef]
3. Coste, F.; Ney, B.; Crozat, Y. Seed development and seed physiological quality of field grown beans (*Phaseolus vulgaris* L.). *Seed Sci. Technol.* **2001**, *29*, 121–136.
4. Weterings, K.; Apuya, N.R.; Bi, Y.P.; Fischer, R.L.; Harada, J.J.; Goldberg, R.B. Regional localization of suspensor mRNAs during early embryo development. *Plant Cell* **2001**, *13*, 2409–2425. [CrossRef]
5. Weber, H.; Borisjuk, L.; Wobus, U. Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* **2005**, *56*, 253–279. [CrossRef] [PubMed]
6. Le, B.H.; Wagmaister, J.A.; Kawashima, T.; Bui, A.Q.; Harada, J.J.; Goldberg, R.B. Using genomics to study legume seed development. *Plant Physiol.* **2007**, *144*, 562–574. [CrossRef] [PubMed]
7. Goldberg, R.B.; Depaiva, G.; Yadegari, R. Plant embryogenesis—Zygote to seed. *Science* **1994**, *266*, 605–614. [CrossRef]

8. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.X.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.J.; Thelen, J.J.; Cheng, J.L.; et al. Genome sequence of the palaeopolyploid soybean. *Nature* **2010**, *463*, 178–183. [CrossRef]

9. Severin, A.J.; Woody, J.L.; Bolon, Y.T.; Joseph, B.; Diers, B.W.; Farmer, A.D.; Muehlbauer, G.J.; Nelson, R.T.; Grant, D.; Specht, J.E.; et al. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biol.* **2010**, *10*, 16. [CrossRef]

10. Jones, S.I.; Gonzalez, D.O.; Vodkin, L.O. Flux of transcript patterns during soybean seed development. *BMC Genom.* **2010**, *11*, 15. [CrossRef]

11. Jones, S.I.; Vodkin, L.O. Using RNA-Seq to Profile Soybean Seed Development from Fertilization to Maturity. *PLoS ONE* **2013**, *8*, e59270. [CrossRef]

12. Jang, Y.E.; Kim, M.Y.; Shim, S.; Lee, J.; Lee, S.-H. Gene expression profiling for seed protein and oil synthesis during early seed development in soybean. *Genes Genom.* **2015**, *37*, 409–418. [CrossRef]

13. Lu, X.; Li, Q.-T.; Xiong, Q.; Li, W.; Bi, Y.-D.; Lai, Y.-C.; Liu, X.-L.; Man, W.-Q.; Zhang, W.-K.; Ma, B.; et al. The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *Plant J.* **2016**, *86*, 530–544. [CrossRef] [PubMed]

14. Du, J.; Wang, S.; He, C.; Zhou, B.; Ruan, Y.L.; Shou, H. Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J. Exp. Bot.* **2017**, *68*, 1955–1972. [CrossRef] [PubMed]

15. Qi, Z.; Zhang, Z.; Wang, Z.; Yu, J.; Qin, H.; Mao, X.; Jiang, H.; Xin, D.; Yin, Z.; Zhu, R.; et al. Meta-analysis and transcriptome profiling reveal hub genes for soybean seed storage composition during seed development. *Plant Cell Environ.* **2018**, *41*, 2109–2127. [CrossRef] [PubMed]

16. Pelletier, J.M.; Kwong, R.W.; Park, S.; Le, B.H.; Baden, R.; Cagliari, A.; Hashimoto, M.; Munoz, M.D.; Fischer, R.L.; Goldberg, R.B.; et al. LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E6710–E6719. [CrossRef]

17. Gao, Y.; Xu, H.; Shen, Y.; Wang, J. Transcriptomic analysis of rice (Oryza sativa) endosperm using the RNA-Seq technique. *Plant Mol. Biol.* **2013**, *81*, 363–378. [CrossRef]

18. Xu, H.; Gao, Y.; Wang, J. Transcriptomic analysis of rice (Oryza sativa) developing embryos using the RNA-Seq technique. *PLoS ONE* **2012**, *7*, e30646. [CrossRef]

19. Belmonte, M.F.; Kirkbride, R.C.; Stone, S.L.; Pelletier, J.M.; Bui, A.Q.; Yeung, E.C.; Hashimoto, M.; Fei, J.; Harada, C.M.; Munoz, M.D.; et al. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E435–E444. [CrossRef]

20. Ziegler, D.J.; Khan, D.; Kalichuk, J.L.; Becker, M.G.; Belmonte, M.F. Transcriptome landscape of the early Brassica napus seed. *J. Integr. Plant Biol.* **2019**, *61*, 639–650. [CrossRef]

21. Liew, L.C.; Narsai, R.; Wang, Y.; Berkowitz, O.; Whelan, J.; Lewsey, M.G. Temporal tissue-specific regulation of transcriptomes during barley (Hordeum vulgare) seed germination. *Plant J.* **2020**, *101*, 700–715. [CrossRef] [PubMed]

22. Doll, N.M.; Just, J.; Brunaud, V.; Caius, J.; Grimault, A.; Depege-Fargeix, N.; Esteban, E.; Pasha, A.; Provart, N.J.; Ingram, G.C.; et al. Transcriptomics at Maize Embryo/Endosperm Interfaces Identifies a Transcriptionally Distinct Endosperm Subdomain Adjacent to the Embryo Scutellum (OPEN). *Plant Cell* **2020**, *32*, 833–852. [CrossRef] [PubMed]

23. Kim, J.Y.; Kim, W.Y.; Kwak, K.J.; Oh, S.H.; Han, Y.S.; Kang, H. Glycine-rich RNA-binding proteins are functionally conserved in Arabidopsis thaliana and Oryza sativa during cold adaptation process. *J. Exp. Bot.* **2010**, *61*, 2317–2325. [CrossRef] [PubMed]

24. Fujikura, U.; Horiguchi, G.; Ponce, M.R.; Micol, J.L.; Tsukaya, H. Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of Arabidopsis thaliana. *Plant J.* **2009**, *59*, 499–508. [CrossRef]

25. Nelissen, H.; Eeckhout, D.; Demuynck, K.; Persiau, G.; Walton, A.; van Bel, M.; Vervoort, M.; Candaele, J.; De Block, J.; Aesaert, S.; et al. Dynamic Changes in ANGUSTIFOLIA3 Complex Composition Reveal a Growth Regulatory Mechanism in the Maize Leaf. *Plant Cell* **2015**, *27*, 1605–1619. [CrossRef]

26. Schrick, K.; Bruno, M.; Khosla, A.; Cox, P.N.; Marlatt, S.A.; Roque, R.A.; Nguyen, H.C.; He, C.W.; Snyder, M.P.; Singh, D.; et al. Shared functions of plant and mammalian StAR-related lipid transfer (START) domains in modulating transcription factor activity. *BMC Biol.* **2014**, *12*, 20. [CrossRef]

27. Chye, M.L.; Huang, B.Q.; Zee, S.Y. Isolation of a gene encoding Arabidopsis membrane-associated acyl-CoA binding protein and immunolocalization of its gene product. *Plant J.* **1999**, *18*, 205–214. [CrossRef]

28. Lung, S.C.; Liao, P.; Yeung, E.C.; Hsiao, A.S.; Xue, Y.; Chye, M.L. Acyl-CoA-Binding Protein ACBP1 Modulates Sterol Synthesis during Embryogenesis. *Plant Physiol.* **2017**, *174*, 1420–1435. [CrossRef]

29. Lung, S.C.; Liao, P.; Yeung, E.C.; Hsiao, A.S.; Xue, Y.; Chye, M.L. Arabidopsis ACYL-COA-Binding Protein1 interacts with Sterol C4-Methyl Oxidase1-2 to modulate gene expression of homeodomain-leucine zipper IV transcription factors. *New Phytol.* **2018**, *218*, 183–200. [CrossRef]

30. Chen, Q.F.; Xiao, S.; Qi, W.Q.; Mishra, G.; Ma, J.Y.; Wang, M.F.; Chye, M.L. The Arabidopsis acbp1acbp2 double mutant lacking acyl-CoA-binding proteins ACBP1 and ACBP2 is embryo lethal. *New Phytol.* **2010**, *186*, 843–855. [CrossRef]

31. Pignocchi, C.; Minns, G.E.; Nesi, N.; Koumproglou, R.; Kitsios, G.; Benning, C.; Lloyd, C.W.; Doonan, J.H.; Hills, M.J. Endosperm Defective1 Is a Novel Microtubule-Associated Protein Essential for Seed Development in Arabidopsis. *Plant Cell* **2009**, *21*, 90–105. [CrossRef] [PubMed]

32. Hehenberger, E.; Kradolfer, D.; Kohler, C. Endosperm cellularization defines an important developmental transition for embryo development. *Development* **2012**, *139*, 2031–2039. [CrossRef] [PubMed]

33. Komis, G.; Luptovciak, I.; Doskocilova, A.; Samaj, J. Biotechnological aspects of cytoskeletal regulation in plants. *Biotechnol. Adv.* **2015**, *33*, 1043–1062. [CrossRef] [PubMed]

34. Shewry, P.R.; Halford, N.G. Cereal seed storage proteins: Structures, properties and role in grain utilization. *J. Exp. Bot.* **2002**, *53*, 947–958. [CrossRef] [PubMed]

35. Xiang, D.; Venglat, P.; Tibiche, C.; Yang, H.; Risseeuw, E.; Cao, Y.; Babic, V.; Cloutier, M.; Keller, W.; Wang, E.; et al. Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in Arabidopsis. *Plant Physiol.* **2011**, *156*, 346–356. [CrossRef] [PubMed]

36. Chen, D.; Schultz, G.; Katchalski, E. Early ribosomal rna transcription and appearance of cytoplasmic ribosomes during germination of wheat embryo. *Nat.-New Biol.* **1971**, *231*, 69. [CrossRef]

37. Sela, A.; Piskurewicz, U.; Megies, C.; Mene-Saffrane, L.; Finazzi, G.; Lopez-Molina, L. Embryonic Photosynthesis Affects Post-Germination Plant Growth. *Plant Physiol.* **2020**, *182*, 2166–2181. [CrossRef]

38. Tejos, R.I.; Mercado, A.V.; Meisel, L.A. Analysis of chlorophyll fluorescence reveals stage specific patterns of chloroplast-containing cells during Arabidopsis embryogenesis. *Biol. Res.* **2010**, *43*, 99–111. [CrossRef]

39. Singh, D. Histopathology of some seed-borne infections—A review of recent investigations. *Seed Sci. Technol.* **1983**, *11*, 651–663.

40. Radchuk, V.; Borisjuk, L. Physical, metabolic and developmental functions of the seed coat. *Front. Plant Sci.* **2014**, *5*, 510. [CrossRef]

41. Babu, Y.; Musielak, T.; Henschen, A.; Bayer, M. Suspensor Length Determines Developmental Progression of the Embryo in Arabidopsis. *Plant Physiol.* **2013**, *162*, 1448–1458. [CrossRef] [PubMed]

42. Kawashima, T.; Goldberg, R.B. The suspensor: Not just suspending the embryo. *Trends Plant Sci.* **2010**, *15*, 23–30. [CrossRef] [PubMed]

43. Sarojam, R.; Sappl, P.G.; Goldshmidt, A.; Efroni, I.; Floyd, S.K.; Eshed, Y.; Bowman, J.L. Differentiating Arabidopsis Shoots from Leaves by Combined YABBY Activities. *Plant Cell* **2010**, *22*, 2113–2130. [CrossRef] [PubMed]

44. Kumaran, M.K.; Bowman, J.L.; Sundaresan, V. YABBY polarity genes mediate the repression of KNOX homeobox genes in Arabidopsis. *Plant Cell* **2002**, *14*, 2761–2770. [CrossRef] [PubMed]

45. Villanneva, J.M.; Broadhvest, J.; Hauser, B.A.; Meister, R.J.; Schneitz, K.; Gasser, C.S. Inner noouter regulates abaxial-adaxial patterning in Arabidopsis ovules. *Genes Dev.* **1999**, *13*, 3160–3169. [CrossRef]

46. Finet, C.; Floyd, S.K.; Conway, S.J.; Zhong, B.J.; Scutt, C.P.; Bowmanb, J.L. Evolution of the YABBY gene family in seed plants. *Evol. Dev.* **2016**, *18*, 116–126. [CrossRef]

47. Yoong, F.Y.; O'Brien, L.K.; Truco, M.J.; Huo, H.Q.; Sideman, R.; Hayes, R.; Michelmore, R.W.; Bradford, K.J. Genetic Variation for Thermotolerance in Lettuce Seed Germination Is Associated with Temperature-Sensitive Regulation of Ethylene Response Factor1 (ERF1). *Plant Physiol.* **2016**, *170*, 472–488. [CrossRef]

48. Fitzgerald, M.S.; McKnight, T.D.; Shippen, D.E. Characterization and developmental patterns of telomerase expression in plants. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14422–14427. [CrossRef] [PubMed]

49. Heller, K.; Kilian, A.; Piatyszek, M.A.; Kleinhofs, A. Telomerase activity in plant extracts. *Mol. Gen. Genet.* **1996**, *252*, 342–345. [CrossRef]

50. Sabelli, P.A.; Larkins, B.A. The Development of Endosperm in Grasses. *Plant Physiol.* **2009**, *149*, 14–26. [CrossRef]

51. Deshpande, P.K. Development of embryo and endosperm in eragrostis-unioloides (poaceae). *Plant Syst. Evol.* **1976**, *125*, 253–259. [CrossRef]

52. Huysmans, M.; Buono, R.A.; Skorzinski, N.; Radio, M.C.; De Winter, F.; Parizot, B.; Mertens, J.; Karimi, M.; Fendrych, M.; Nowack, M.K. NAC Transcription Factors ANAC087 and ANAC046 Control Distinct Aspects of Programmed Cell Death in the Arabidopsis Columella and Lateral Root Cap. *Plant Cell* **2018**, *30*, 2197–2213. [CrossRef]

53. Yanofsky, M.F.; Ma, H.; Bowman, J.L.; Drews, G.N.; Feldmann, K.A.; Meyerowitz, E.M. The protein encoded by the arabidopsis homeotic gene agamous resembles transcription factors. *Nature* **1990**, *346*, 35–39. [CrossRef]

54. Bowman, J.L.; Smyth, D.R.; Meyerowitz, E.M. Genetic interactions among floral homeotic genes of arabidopsis. *Development* **1991**, *112*, 1–20. [PubMed]

55. Shahan, R.; Zawora, C.; Wight, H.; Sittmann, J.; Wang, W.P.; Mount, S.M.; Liu, Z.C. Consensus Coexpression Network Analysis Identifies Key Regulators of Flower and Fruit Development in Wild Strawberry. *Plant Physiol.* **2018**, *178*, 202–216. [CrossRef] [PubMed]

56. Le, B.H.; Cheng, C.; Bui, A.Q.; Wagmaister, J.A.; Henry, K.F.; Pelletier, J.; Kwong, L.; Belmonte, M.; Kirkbride, R.; Horvath, S.; et al. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8063–8070. [CrossRef]

57. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 14. [CrossRef]

58. Abu-Jamous, B.; Kelly, S. Clust: Automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biol.* **2018**, *19*, 172. [CrossRef]

59. Monfared, M.M.; Simon, M.K.; Meister, R.J.; Roig-Villanova, I.; Kooiker, M.; Colombo, L.; Fletcher, J.C.; Gasser, C.S. Overlapping and antagonistic activities of BASIC PENTACYSTEINE genes affect a range of developmental processes in Arabidopsis. *Plant J.* **2011**, *66*, 1020–1031. [CrossRef]

60. Kang, H.G.; Singh, K.B. Characterization of salicylic acid-responsive, Arabidopsis Dof domain proteins: Overexpression of OBP3 leads to growth defects. *Plant J.* **2000**, *21*, 329–339. [CrossRef]

61. Spurney, R.J.; Van den Broeck, L.; Clark, N.M.; Fisher, A.P.; Balaguer, M.A.D.; Sozzani, R. Tuxnet: A simple interface to process RNA sequencing data and infer gene regulatory networks. *Plant J.* **2020**, *101*, 716–730. [CrossRef] [PubMed]

62. Walia, H.; Josefsson, C.; Dilkes, B.; Kirkbride, R.; Harada, J.; Comai, L. Dosage-Dependent Deregulation of an AGAMOUS-LIKE Gene Cluster Contributes to Interspecific Incompatibility. *Curr. Biol.* **2009**, *19*, 1128–1132. [CrossRef] [PubMed]

63. Kang, I.H.; Steffen, J.G.; Portereiko, M.F.; Lloyd, A.; Drews, G.N. The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant Cell* **2008**, *20*, 635–647. [CrossRef]

64. Choe, J.E.; Kim, B.; Yoon, E.K.; Jang, S.; Kim, G.; Dhar, S.; Lee, S.A.; Lim, J. Characterization of the GRAS transcription factor SCARECROW-LIKE 28's role in Arabidopsis root growth. *J. Plant Biol.* **2017**, *60*, 462–471. [CrossRef]

65. Zhang, Z.L.; Ogawa, M.; Fleet, C.M.; Zentella, R.; Hu, J.; Heo, J.O.; Lim, J.; Kamiya, Y.; Yamaguchi, S.; Sun, T.P. Scarecrow-like 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2160–2165. [CrossRef] [PubMed]

66. Ding, Z.; Friml, J. Auxin regulates distal stem cell differentiation in Arabidopsis roots. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12046–12051. [CrossRef] [PubMed]

67. Xie, Z.L.; Nolan, R.O.R.; Jiang, H.; Tang, B.Y.; Zhang, M.C.; Li, Z.H.; Yin, Y.H. The AP2/ERF Transcription Factor TINY Modulates Brassinosteroid-Regulated Plant Growth and Drought Responses in Arabidopsis. *Plant Cell* **2019**, *31*, 1788–1806. [CrossRef]

68. Magyar, Z.; De Veylder, L.; Atanassova, A.; Bako, L.; Inze, D.; Bogre, L. The role of the Arabidopsis E2FB transcription factor in regulating auxin-dependent cell division. *Plant Cell* **2005**, *17*, 2527–2541. [CrossRef]

69. Patel, R.K.; Jain, M. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLoS ONE* **2012**, *7*, e30619. [CrossRef]

70. Goodstein, D.M.; Shu, S.Q.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [CrossRef]

71. Kim, D.; Landmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef] [PubMed]

72. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290. [CrossRef] [PubMed]

73. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 38. [CrossRef] [PubMed]

74. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef]

75. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33*, 2938–2940. [CrossRef]

76. Usadel, B.; Nagel, A.; Steinhauser, D.; Gibon, Y.; Blasing, O.E.; Redestig, H.; Sreenivasulu, N.; Krall, L.; Hannah, M.A.; Poree, F.; et al. PageMan: An interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC Bioinform.* **2006**, *7*, 8. [CrossRef]

77. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 13. [CrossRef]

78. Steinegger, M.; Soding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [CrossRef]

79. Tian, F.; Yang, D.C.; Meng, Y.Q.; Jin, J.P.; Gao, G. PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Res.* **2020**, *48*, D1104–D1113. [CrossRef]

80. Yu, G.C.; Wang, L.G.; Han, Y.Y.; He, Q.Y. Clusterprofiler: An R Package for Comparing Biological Themes among Gene Clusters. *Omics J. Integr. Biol.* **2012**, *16*, 284–287. [CrossRef]

81. Machanick, P.; Bailey, T.L. MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **2011**, *27*, 1696–1697. [CrossRef] [PubMed]