Article

Comparative Transcriptomics of Non-Embryogenic and Embryogenic Callus in Semi-Recalcitrant and Non-Recalcitrant Upland Cotton Lines

Sonika Kumar [1], Ashleigh Ruggles [2], Sam Logan [2], Alora Mazarakis [2], Tyson Thomas [2], Matthew Bates [2], Clayton Grosse [2], Dave Reed [2], Zhigang Li [1], Jane Grimwood [3], Jeremy Schmutz [3], and Christopher Saski [1,*]

[1] Department of Plant and Environmental Sciences, Clemson University, Clemson, South Carolina; sonikak@clemson.edu
[2] Techshot Inc, Greenville, Indiana
[3] HudsonAlpha Institute for Biotechnology, Huntsville, Alabama
* Correspondence: saski@clemson.edu; Tel.: (864-656-6929, CS)

Abstract

Somatic embryogenesis-mediated plant regeneration is essential for genetic manipulation of agronomically important traits in upland cotton. Genotype specific recalcitrance to regeneration is a primary challenge in deploying genome editing and incorporating useful transgenes into elite cotton germplasm. In this study, transcriptomes of a semi-recalcitrant cotton (*Gossypium hirsutum* L.) genotype 'Coker312' were analyzed at two critical stages of Somatic Embryogenesis that includes non-embryogenic callus (NEC) and embryogenic callus (EC) cells, and the results compared to a non-recalcitrant genotype 'Jin668'. We discovered of 305 differentially expressed genes in Coker312, whereas, in Jin668, about 6-fold more genes (2,155) were differentially expressed. A total of 154 differentially expressed genes were common between the two genotypes. Gene enrichment analysis of upregulated genes identified functional categories such as lipid transport, embryo development, regulation of transcription, sugar transport, vitamin biosynthesis, among others. In Coker312 EC cells, five major transcription factors were highly upregulated: *LEAFY COTYLEDON 1 (LEC1)*, *WUS-related homeobox 5 (WOX5)*, *ABSCISIC ACID INSENSITIVE3 (ABI3)*, *FUSCA3 (FUS3)*, and *WRKY2*. In Jin668, *LEC1*, *BABY BOOM (BBM)*, *FUS3*, and *AGAMOUS-LIKE15 (AGL15)* were highly expressed in EC cells. We also found that gene expression of these embryogenesis genes was typically higher in Jin668 when compared to Coker312. We conclude that significant differences in expression of the above genes between Coker312 and Jin668 may be a critical factor affecting the regenerative ability of these genotypes.

Keywords: callus; genotype-specific recalcitrance; reprogramming; somatic embryogenesis; transcriptomics; upland cotton

1. Introduction

Plant somatic embryogenesis (SE) is a unique developmental process that ultimately leads to regeneration of a whole plant from a single somatic cell. This process involves sophisticated cellular reprogramming events that are controlled by gene expression programs and signaling pathways that direct callus cells to dedifferentiate, reprogram, and begin differentiation into polarized structures that eventually become a viable embryo [1-4]. SE is initiated by various factors such as culture medium conditions that includes high concentrations of plant growth regulators (PGRs), various stresses such as plant

wounding, temperature, and osmotic pressures [5-7]. Under SE initiation, somatic cells from various explant sources (e.g., hypocotyls, young leaves, and immature embryos) non-embryogenic callus cells (NEC) are formed that can be described as unorganized de-differentiated cell masses that are continuously dividing. These cells are responsive to the components in the growth medium and signals from the environment such as light and temperature. Eventually, cells with embryogenic potential will differentiate into embryogenic callus cells (EC) that are described as cells that are polarized and begin to form a ball-like structure and are the precursors to somatic embryos that ultimately lead to whole plant formation [8]. In most dicot plant species, the ability to achieve whole plant regeneration through SE is limited to only a few select genotypes with drastic differences in embryo formation frequencies and time to embryo formation.

Previous work has identified genes and transcription factors whose expression are required to achieve transition from NEC to EC cells [6,9,10]. In several monocot species such as rice and maize, studies have shown that ectopic expression of transcription factors with an inducible promoter such as *BABY BOOM (BBM)*, *WUSCHEL* (*WUS),* and *LEAFY COTYLEDON 1 (LEC1)* improve embryo formation frequency in semi-recalcitrant and recalcitrant genotypes [11-13]. In dicots, overexpression of *BBM* has been demonstrated to improve embryo formation frequency in tobacco [14], sweet pepper [15], and *Theobroma cacao* [16], but not to the same degree of efficiency as the monocot systems suggesting that we are still limited in our knowledge of the genes and transcriptional pathways that are involved in SE in dicot plants.

As with most plants, SE in Upland cotton (*Gossypium hirsutum* L.) is limited to only a few genotypes within the species. It has been nearly 4 decades since successful SE was reported. Various studies have screened genotypes for their ability to regenerate but only resulted in a handful of genotypes with regenerative capacity [17-19]. The most widely used public genotype for transformation and regeneration is Coker312 [20,21]. In another study the genotype YZ-1 was identified as a line with higher regeneration efficiency [22]. However, both Coker312 and the YZ-1 genotypes were developed 20-30 years ago and are without suitable agronomic traits for gene function studies in cotton. More recently, a non-recalcitrant genotype, named Jin668 was described that has a high frequency of embryo formation (~96%) with only a short duration of 45-60 days from callus initiation to the formation of ball stage structures [23]. Our group recently determined and characterized global gene expression profiles of this elite regeneration line at two primary developmental stages (NEC and EC) [24]. We discovered a significant transcriptome-wide difference between the two developmental stages and sharp upregulation of key transcription factors identified in EC cells that may have a primary role in reprogramming in this genotype during SE [24].

Whole plant regeneration through SE offers significant and biological advantages. For instance, genome editing through engineered nucleases [25] or the CRISPR-Cas9 technologies [26] offer approaches to directly modify the genome of living organisms. In plants, these powerful technologies provide a means for direct trait enhancement of elite breeding material offering unprecedented opportunities to improve crop breeding approaches such as reduced breeding cycle times, tailored trait genetics, and the potential for much larger genetic gains [27,28]. The promises of these technologies offer world changing outcomes, however major limitations such as delivery of the genome editing reagents and genotype specific recalcitrance to regeneration remain as primary obstacles preventing widespread use in crop improvement [27,29].

In the present study, we analyzed the transcriptional profiles of NEC and EC cells harvested from the semi-recalcitrant genotype, Coker312 and compared these data to the non-recalcitrant genotype, Jin668 [24]. Our objectives were to determine gene expression profile differences and similarities between Jin668 and Coker312 that fall into the following categories: (i) genes uniquely expressed in each genotype (ii) genes with similar expression profiles and (iii) genes with different expression profiles between two genotypes. The results from this study will provide new opportunities to discover genes and

regulatory networks involved in somatic embryogenesis that can be leveraged to develop new strategies to avert genotype-specific recalcitrance to regeneration in dicot species.

2. Results

2.1. RNAseq of Coker312 at NEC and EC Developmental Stages

Differential gene expression analysis revealed a total of 196 genes upregulated and 109 genes downregulated, respectively, in Coker312 EC when compared to Coker312 NEC with at least 2-fold abundance difference and an adjusted p-value of 0.001, Supplemental Table 1.

Table 1. Top 20 up/down regulated genes in Coker312 EC cells.

Genes upregulated in EC cells					
Gene	**log2(EC)**	**log2(NEC)**	**logFC**	**Gene Function**	**Best Hit Arabidopsis**
Gohir.D13G121100.1	7.250014115	0.19660704	7.05		NA
Gohir.D13G121201.1	7.250014115	0.19660704	7.05		NA
Gohir.D13G121301.1	7.250014115	0.19660704	7.05		NA
Gohir.D13G121400.1	7.250014115	0.19660704	7.05		NA
Gohir.D13G121500.1	7.250014115	0.19660704	7.05		NA
Gohir.D13G121601.1	7.250014115	0.19660704	7.05		NA
Gohir.D13G121700.1	7.250014115	0.19660704	7.05		NA
Gohir.A02G027300.1	6.048672137	0.98477161	5.06	lipid transfer protein 1	AT2G38540
Gohir.D13G121800.1	4.873813198	0	4.87	NA	NA
Gohir.D11G255800.1	5.140655972	0.64431778	4.5	homeobox protein 31	NA
Gohir.D09G214700.1	4.210077099	0	4.21	lipid transfer protein 6	AT3G08770
Gohir.A13G117900.1	4.170726276	0	4.17		NA
Gohir.A05G157800.2	5.21680405	1.14990967	4.07	homeobox-3	AT2G33880
Gohir.A05G258900.2	3.996750279	0	4		NA
Gohir.D05G252700.1	4.069014678	0.14795788	3.92	sucrose-proton symporter 2	AT1G71880
Gohir.D02G178800.1	5.818876119	1.96458346	3.85	early nodulin-like protein 3	AT4G32490
Gohir.D01G170800.1	4.741520918	0.91838623	3.82	D-aminoacid aminotransferase-like PLP-dependent enzymes superfamily protein	AT1G50110
Gohir.D03G120300.1	6.743972672	2.93243919	3.81	Ctr copper transporter family	AT5G59030
Gohir.D06G172800.3	4.864136609	1.0765591	3.79		NA
Gohir.D05G160500.2	4.30560579	0.5685186	3.74	homeobox-3	AT2G33880
Genes downregulated in EC cells					
Gene	**log2(EC)**	**log2(NEC)**	**logFC**	**Gene Function**	**Best Hit Arabidopsis**
Gohir.A08G000012.1	0	2.61917822	-2.62	NA	NA
Gohir.A08G035700.1	0	2.65351867	-2.65	conserved peptide upstream open reading frame 9	AT3G25572
Gohir.D11G135400.1	0.532067552	3.19440229	-2.66	nodulin MtN21 /EamA-like transporter family protein	NA
Gohir.A08G163500.1	9.153073165	11.8621445	-2.71	expansin-like B1	AT4G17030
Gohir.A08G164800.1	2.183645305	4.89442976	-2.71	expansin-like B1	AT4G17030
Gohir.A11G128400.1	0	2.72421369	-2.72	Serine protease inhibitor, potato inhibitor I-type family protein	AT2G38870
Gohir.A08G221266.1	0	2.82048534	-2.82	NA	NA
Gohir.D08G183300.1	8.199755764	11.1391459	-2.94	expansin-like B1	AT4G17030

Gohir.A06G029900.1	1.270229907	4.25149188	-2.98	Phosphoglycerate mutase family protein	AT5G64460
Gohir.A10G027250.1	1.523561956	4.53362564	-3.01		NA
Gohir.D03G000201.1	0	3.07347751	-3.07		NA
Gohir.D04G021500.1	1.379066399	4.5088723	-3.13	NA	NA
Gohir.A08G053950.1	0	3.21396933	-3.21	Cellulose synthase family protein	AT4G32410
Gohir.A05G393450.1	1.019346089	4.39882918	-3.38	NA	NA
Gohir.A09G136000.1	0.298658316	3.73487217	-3.44		NA
Gohir.D12G006350.1	0	3.47235779	-3.47	NA	NA
Gohir.A05G393425.1	0.211635253	3.73595535	-3.52	NA	NA
Gohir.D05G275600.2	4.984908294	8.57295407	-3.59	NA	NA
Gohir.D08G110800.1	0	3.62690633	-3.63	small acidic protein 1	AT4G13520
Gohir.D04G021600.1	0.633198686	4.3772626	-3.74	NA	NA

Genes without known orthologs or predicted functional domains in model systems such as *Arabidopsis* were among the most differentially expressed (~7-fold), Table 1. Important genes with functional annotations that were largely upregulated in EC cells includes lipid transfer proteins, homeobox protein 31, homeobox-3 genes, early nodulin-like proteins, copper transporters, seed gene1, and *APETALA2 (AP2)* transcription factors, Table 1 and Supplemental Table 1. Genes with the largest downregulated expression profile (almost 4-fold) during the transition to EC cells include mevalonate diphosphate decarboxylase 1, conserved peptide upstream open reading frame 9, serine protease inhibitors, expansin-like proteins, and nodulin transporter family proteins, Table 1 and Supplemental Table 1. Functional enrichment of upregulated genes in Coker312 EC cells identified biological processes involved in biosynthesis of vitamins such as thiamine, transport of sucrose, copper, and lipids, and genes involved in embryo development, Figure 1 and Supplemental Table 2. In the molecular function category, the largest number of enriched genes categorized as transcription factors (14), lipid binding (8), growth factors (4), transporters (4), and hydrolase activity (3), Supplemental Table 2.

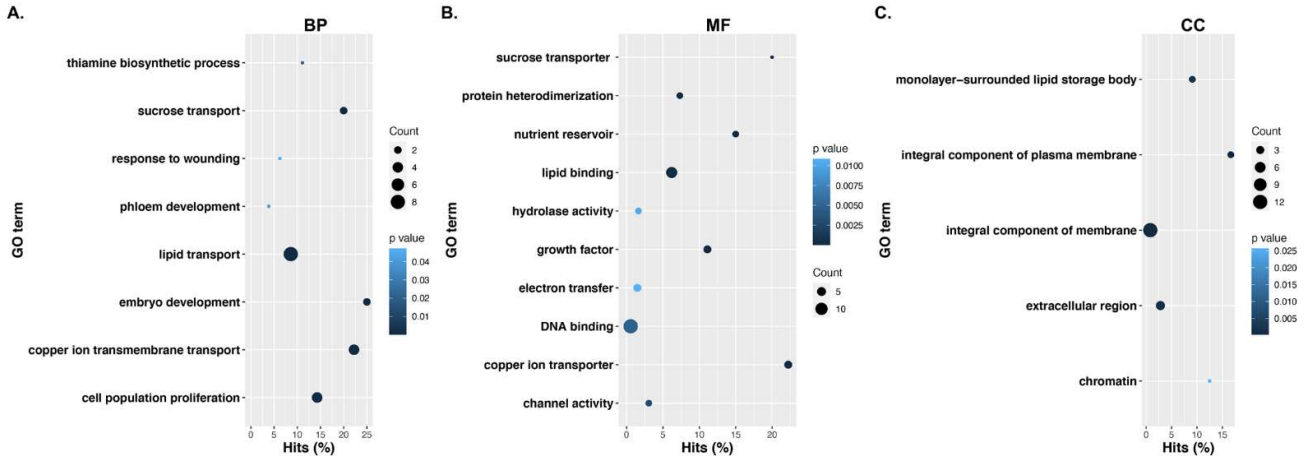

Figure 1. Functional enrichment of upregulated genes in Coker312 EC (A) in biological process (BP) (B) in molecular functions and (C) cellular components.

Gene enrichment analysis of downregulated genes in EC callus in Coker312 revealed only a handful of enriched functional categories (Supplemental Table 2). Downregulated enriched functional categories were only represented by a few genes (<4) and include

response to biotic stimulus, response to defense and wounding, hydrolase activity, and a few others (Supplemental Table 3).

2.2. Comparison of Coker312 to Jin668 at NEC and EC stages

We compared gene expression profiles of the semi-recalcitrant genotype, Coker312 to previously transcriptomic study of a non-recalcitrant genotype, Jin668 [24]. Comparative analysis of differentially expressed genes with at least a 2-fold change profile and an error corrected p-value of 0.001 identified 151 unique genes in Coker312 and 2,001 unique genes in Jin668, Figure 2. a total of 305 and 2,155 differentially expressed genes in Coker312 and Jin668, respectively. Figure 2.

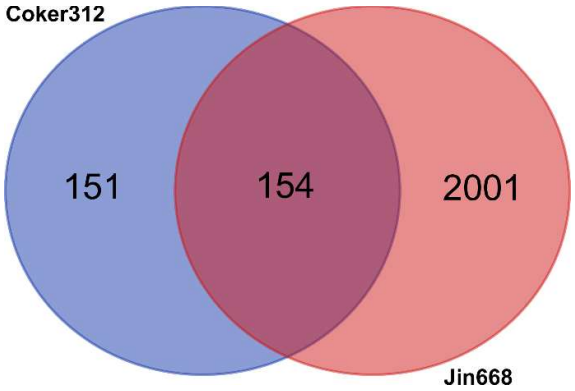

Figure 2. A VENN diagram representing gene expression that are common and unique in Coker312 and Jin668 at NEC and EC stages.

Of particular interest are the 154 genes that are differentially expressed in both genotypes. Interestingly, a duplicate gene pair (homeologous gene copies on both A and D subgenomes) with the highest expression values among each genotype were a bifunctional inhibitor of lipid transfer, followed by a tandemly duplicated gene on chromosome 13 of the d-subgenome with an unknown annotation, Figure 3 and Supplemental Table 4. Other genes with high expression in EC cells in both genotypes are lipid transfer proteins, genes with homeobox domains, and genes involved with histone proteins, Supplemental Table 4. A clustering analysis of the genes based on their expression profile showed aggregation by condition and not genotype, Figure 3.

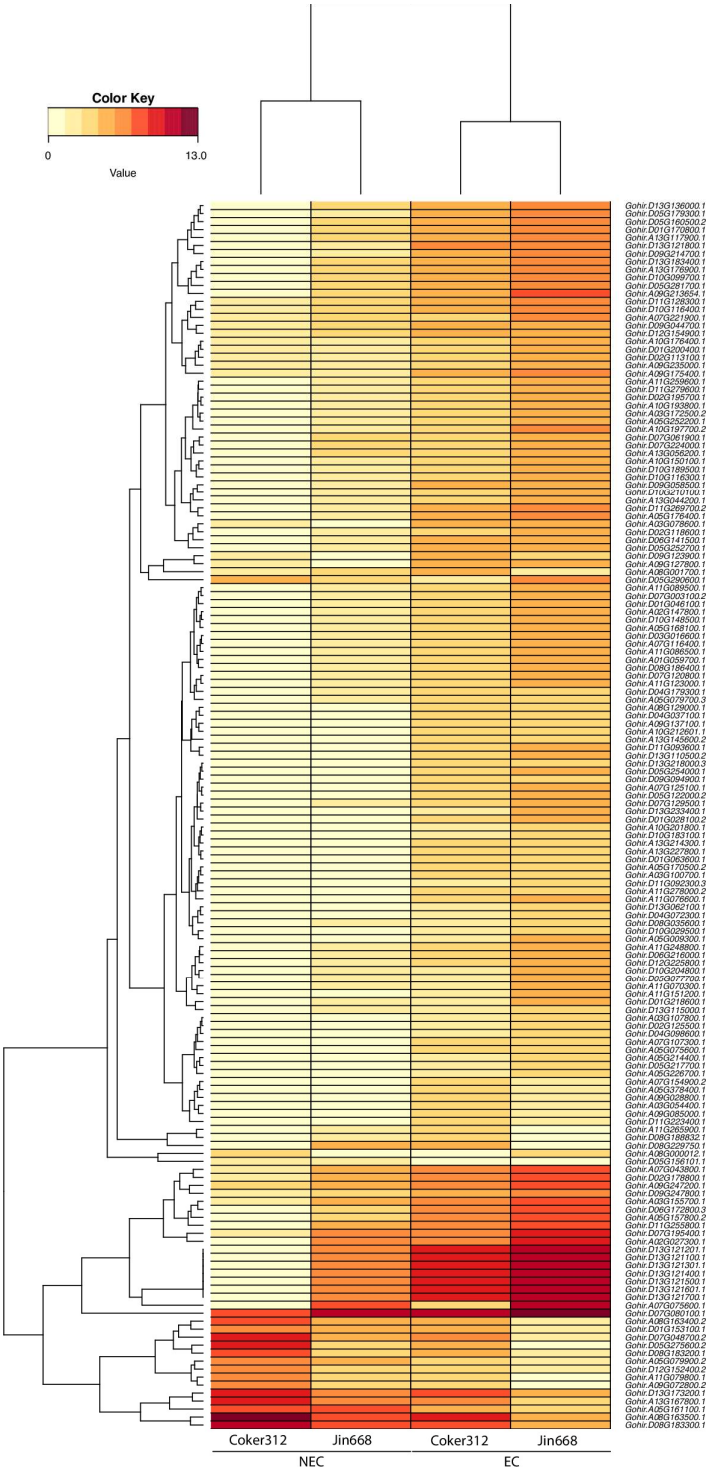

Figure 3. Heatmap representing the gene expression of Coker312 and Jin668 at NEC and EC callus cells.

Functional gene enrichment of the 154 overlapping differentially expressed genes identified genes in enriched categories such as lipid transport, embryo development, regulation of transcription, sugar transport, vitamin biosynthesis, and others, Supplemental Table 5 and Figure 4. We also observed that among the 154 overlapping genes, gene expression profiles were typically higher in Jin668 EC cells versus Coker312 EC cells,

Figure 3. For example, clusters of genes in Jin668 EC were more often expressed higher when compared to Coker EC, Figure 3.

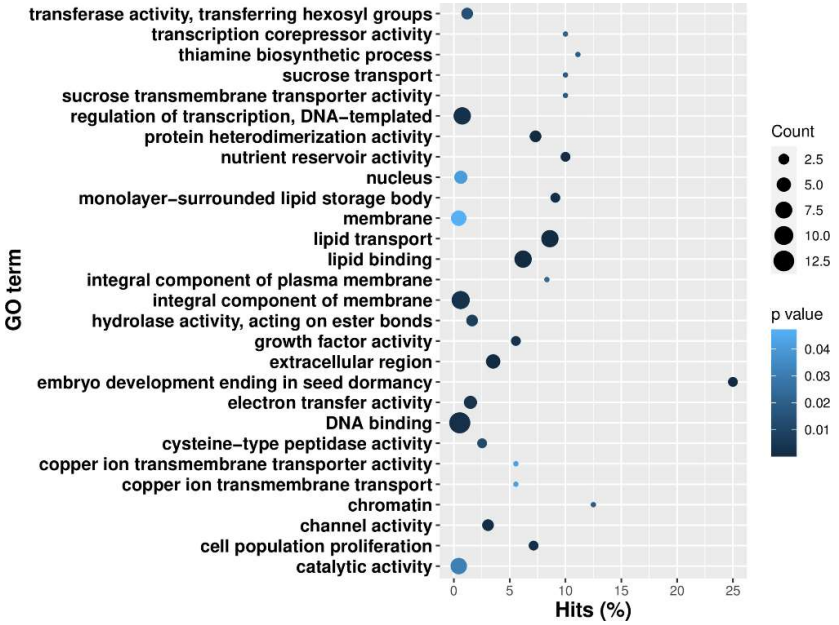

Figure 4. Gene ontology enrichment of significantly upregulated genes in Coker312 at EC cells.

We also examined expression profiles of genes known to have a role in somatic embryogenesis, such as *BBM, WUS, LEC1, WUSCHEL-RELATED HOMEOBOX 5 (WOX5), FUSCA3 (FUS3),* and several others [24]. Genes with the sharpest fold changes were Gohir.D13G136000.1*(LEC1−1)*, Gohir.A13G132600.1*(LEC1−2)*, Gohir.D07G237600.1*(FUS3−2)*, Gohir.D08G035600.1*(LEC1−3)* and Gohir.A07G230400.1*(FUS3_1)* in Coker312 EC and Jin668EC, although expression was higher in Jin668 EC in comparison of Coker312 EC, Figure 5. However, Gohir.A08G227000.1*(BBM−1)* showed high expression in Jin668 EC while much lower expression in Coker312 EC was observed. Two other interesting genes, Gohir.A10G233000.1*(WOX5−1)* and Gohir.D10G245300.1*(WOX5)*, were upregulated in Coker312 EC, Coker312 NEC, and Jin668 NEC while they were almost off in Jin668 EC, Figure 5. Several important transcription factors previously reported with a role in somatic embryogenesis, such as Gohir.D03G115300.1*(GRD/RKD)*, Gohir.D10G089500.1*(WUS−3)*, Gohir.A12G059800.1*(WUS−2)* and Gohir.D12G060100.1*(WUS−4)* showed very little to no expression in either genotype or developmental stage, Figure 5. In the Coker312 highest upregulated gene is *LEC1* followed by *WOX5*, *ABSCISIC ACID INSENSITIVE (ABI3)*, *FUS3* and *WRKY2* while in Jin668 *LEC1*, *BBM*, *FUS3* and *AGAMOUS-LIKE15* (*AGL15*) were the highly expressed genes, and their expression was several times higher in comparison to Coker312. Surprisingly, all the copies of *WUS* were either off or very lowly expressed in either genotype or developmental stage.

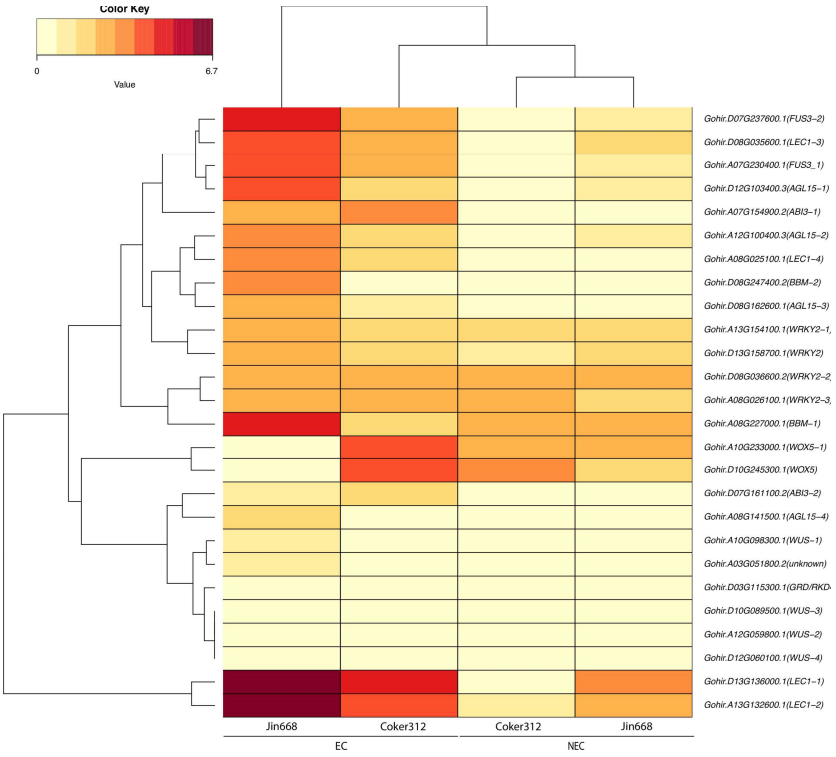

Figure 5. Heatmap representing the gene expression of Coker312 and Jin668 at NEC and EC callus cells.

3. Discussion

Upland cotton is one of the most important economic crops worldwide that produces the largest source of renewable textile fiber. However, cotton is highly restricted to genetic improvement via transformation and whole plant regeneration through somatic embryogenesis mainly because of somaclonal variation in tissue culture, long culture times, decline in vigor, and low potency of embryogenesis [22]. Moreover, regenerative capacity is highly genotype dependent, and previous investigations for regenerable genotypes in cotton have not yielded many significant advancements [21,30]. In several monocot species, such as rice and maize, somatic embryogenesis has been examined at the transcriptional levels in both recalcitrant and semi-recalcitrant species and these studies have identified several key transcription factors such as *BBM*, *WUS2*, *LEC1* and *LEC2*, that when ectopically expressed in recalcitrant genotypes have initiated somatic embryo formation, although frequencies and time to embryo formation still remain low and slow, respectively [13,16,31,32].

A recent study revealed an Upland cotton genotype, Jin668 with elite somatic regeneration properties, such as high frequency of embryo formation (~96%) and rapid time to cellular differentiation (45-60 days) that was developed through successive regeneration acclimation (SRA) [23]. The authors hypothesize that the regenerative potential (totipotency) is a trait that is encoded in genome, but is epigenetically suppressed in most genotypes, hence, genotype-specific recalcitrance [23]. As a follow up, we compared global gene expression at two key developmental stages (NEC and EC) in the Jin668 genotype to identify genes necessary for cellular reprogramming and the transition to EC cells [24]. We identified sharp upregulation of several transcription factors that likely have a

major role in regulating the shift from NEC to EC with subgenome bias in this allotetraploid species [24].

In this study, we collected transcriptome data from the semi-recalcitrant genotype, Coker312. Coker312 is considered semi-recalcitrant because of its long time to embryo formation (90-120 days) and low frequency of embryo formation (<15%). The most up-regulated genes in Coker312 EC cells are annotated as lipid transfer proteins (LTPs), homeobox-3 genes, *AP2* transcription factors, early nodulin-like proteins, and copper transporters. Previous studies have demonstrated that the LTPs are involved in the formation of a protective layer of cutin in the cell wall, surrounding the young embryo, and are implicated in the initiation of somatic embryogenesis [33]. LTPs are also abundantly expressed in the epidermis of developing tissues and play an important role in fiber elongation [34]. Earlier studies also demonstrated the role of *AP2*, a super-family transcription factors, may have a role in callus formation [35], and also contribute to biotic and abiotic stress resistance in cotton [36]. Additionally, some APETALA 2/ethylene-responsive element binding factor *(AP2/ERFs)* are implicated in growth and developmental processes mediated by growth hormones like gibberellins (GAs), cytokinins (CTK), and brassinosteroids (BRs) [37], and may also have a role in hormone sensing and signaling pathways important to cellular reprogramming during the transition from NEC to EC in Upland cotton as demonstrated by the results presented here.

Our previous data identified several thousand differentially expressed genes during the transition from NEC to EC in the non-recalcitrant genotype, Jin668 [24]. Comparative transcriptome analysis between Jin668 and the semi-recalcitrant genotype, Coker312 could provide a much smaller 'candidate set' of key genes by examining the overlap between the two genotypes. In EC calli of both Coker312 and jin668, the d-subgenome encoded homolog of *LEC1* has the highest expression, indicating subgenome expression bias and a primary role in somatic embryogenesis in Upland cotton. *LEC1* is described as a master regulator that shapes embryo development in Arabidopsis [38], and in other studies it is described as a central regulator controlling different parts of embryo morphogenesis and photosynthesis as well as seed development [39]. Another gene of interest, *WOX5*, is a transcription factor that has been shown to be a regulator of a pool of pluripotent stem cells in the apical meristem [40], and has endowed gain-of-function mutants with somatic embryo formation in Arabidopsis [12]. It expressed in different stages of embryogenesis and post germination growth stages [41,42]. In both Coker312 and Jin668, *WOX5* is moderately expressed in NEC cells and is downregulated in EC cells, suggesting that it may be an upstream regulator of cellular reprogramming and that early expression of this gene is necessary for the transition to embryo formation. Earlier studies have reported that *BBM* play important role in transcription of *LEC1, LEC2, ABI3,* and *FUS3* [31]. In Jin668 *BBM* is highly expressed while in Coker312 it shows less expression. Critical difference in expression of morphogenic regulator BBM may be one of the reasons of different regenerative ability of both genotypes. The genes *ABI3* and *FUS3* are also expressed in EC callus of both Upland cotton genotypes. These genes are transcription factors of *LAFL* genes [31] and ectopic expression of *ABI3* did not result in successful embryo development, but it has a reported role in embryo programming being activated by *BBM* [31]. Ectopic expression of *FUS3* has resulted into cotyledon likes leaves, while *LEC1* and *LEC2* over expression results into spontaneous development of somatic embryos turning into plantlets [43,44].

Comparative analysis of common expressed genes in Coker312 and Jin668 has discovered few new genes like growth-regulating factor 2 *(GRF2)*, late embryogenesis Abundant 4-5 *(LEA4-5)* and Late Embryogenesis Abundant protein *(LEA)* family protein. Previous work has shown that *GRF2* expressed strongly in developing tissues of shoot apical meristem, upper stems, and root tips and required for the coordination of cell

division and differentiation during leaf development in Arabidopsis [45]. *LEA* a large group of hydrophilic proteins that provides a major role in drought stress tolerance in upland cotton and are required for normal growth and development. These proteins are mostly expressed during abiotic stresses like cold, drought and high salinity [46,47], but may be functioning in callus cells in response to the tissue-culture microenvironment.

In plant transformation and regeneration systems, both frequency and time to embryo formation are critical factors. In agrobacterium-mediated transformation, it is important to note that agrobacterium stress and selection pressure results in a reduction in embryo formation efficiency and duration to form embryo when compared to simply regenerating a whole plant without transformation. In Jin668 and Coker312, this stress typically adds ~8 weeks. However, embryo formation frequency in Jin668 still remains high (>80%) while in Coker312, embryo formation drops to less than 15% and the probability of transformed plants thorough SE is very low. This may in part be due to differences in expression levels of key genes between the two genotypes. When analyzing global trends in gene expression, a general upregulation of nearly the same transcription factors were found, but with a much higher expression profile in Jin668.

4. Materials and Methods

4.1. Plant material and callus growth conditions

The semi recalcitrant genotype, Coker312, used in this study was obtained from the USDA Crop Germplasm Collection, College Station, TX. The seed sterilization, seedling germination and callus cultures were prepared as previously described in Wen et al. [24] at Clemson University. These callus cultures are part of a larger spaceflight experiment and were transferred to the Space Life Sciences Lab (SLSL: https://www.spaceflorida.gov/facilities/space-life-sciences-lab/) by automobile for growth, observation, and harvesting during the science verification testing (SVT) component of the larger experimental evaluation. The SVT evaluates growing conditions in a growth chamber that mimics conditions of the Advanced Plant Habitat (APH) on orbit in the International Space Station (ISS). The conditions used for these callus cultures include: 28 ± 1 °C, 16 h (day)/8 h (night) photoperiod, 1,000 ppm CO_2 , with light provided by cool-white fluorescent lamps at an irradiation of 60 $\mu molm^{-2} s^{-1}$, and 50% relative humidity. Plates were rotated on alternate day for the equal light distribution.

.

4.2. RNAseq

Whole calli were harvested at two different stages, i.e., NEC (45-days old calli) and EC (90-days old calli with ball-stage embryo structures) into RNA later and kept at room temperature for 24 hours and moved to -80C for longer storage. A total of four biological replications for each stage were used for mRNA sequencing. Total RNA was extracted from Coker312 callus material following the guanidine thiocyanate method described in [48]. RNA integrity and concentration was assayed on an BioAnalyzer2100 (Agilent) and was considered high quality with RNA Integrity Number (RIN) values ≥ 7 and total masses (≥ 2.0 μg total RNA) for all biological replicates. Sequencing libraries were prepared following standard protocols of the Illumina TruSeq Stranded RNA kit. Transcriptome sequences were collected on an Illumina NovaSeq to a depth of at least 40 million read pairs per replicate sample. Raw sequences were preprocessed to remove adapter and low quality bases with Trimmomatic software v.0.38 [48]. Cleaned reads were mapped to the *Gossypium hirsutum* (TM1 v.2.0) reference assembly [49] using the Bowtie2 short-read aligner [50]. Transcript abundance was quantified with RSEM [51]

and differentially expressed transcripts were determined with edgeR [52]. Because cotton is an allotetraploid species with subgenomes that are highly identical, genes are expected to be in multiple copies and perhaps, be expressed with bias at the subgenome level [49]. It is important to note that genes described in Figure 5 are assigned a gene name and the annotated gene and its expression value are derived from primary transcripts of the V2.0 assembly described in Chen et al., 2020 [49].

5. Conclusions

In this study, comparative transcriptome profiling of two Upland cotton genotypes that differ in regenerative capacity and developmental timing revealed a short list of candidate genes whose expression and expression abundance are critical for somatic embryogenesis in Upland cotton. The genes *LEC1*, *BBM*, *FUS3* and *AGL15*, *ABI)*, and *WOX5* were commonly expressed in Coker312 and Jin668, EC cells. These results provide a foundation for candidate gene testing (in various combinations) for their role to initiate somatic embryogenesis in recalcitrant cotton genotypes.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: Differentially expressed genes in Coker312 in NEC compared to EC cells with at least 2-fold change and error corrected p-values less than or equal to 0.001., Table S2: Functional enrichment categories of genes upregulated in EC callus cells in Coker312., Table S3: Functional enrichment categories of genes downregulated in EC callus cells in Coker312., Table S4: Gene expression matrix (Log TMM+1) of the 154 overlapping differentially expressed genes in Coker312 and Jin668., Table S5: Functional enrichment categories of the 154 genes that are commonly differntially expressed among Coker312 and Jin668.

Author Contributions: Conceptualization, C.S. and S.K.; methodology; formal analysis, S.K. and C.S.; investigation, A.R., S.K., S.L., A.M., D.R., J.S., J.G., and C.S.; data curation, S.K., C.S., J.S., J.G.; writing—original draft preparation, S.K.; writing—review and editing, A.R., S.K., S.L., A.M., D.R., J.G., J.S., and C.S; funding acquisition, C.S., D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NASA/CASIS/TARGET, grant number:GA-2019-007

Data Availability Statement: All sequence data reported in this manuscript can be found in the NCBI short read archive under BioProject study PRJNA747913 (Samples: SRR15186777-ARR15186784).

Acknowledgments: Not Applicable

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sato, Y.; Sentoku, N.; Matsuoka, M. [Expression patterns and functions of plant homeobox genes during embryogenesis]. *Tanpakushitsu Kakusan Koso* **1997**, *42*, 1866-1874.
2. DeLille, J.M.; Sehnke, P.C.; Ferl, R.J. The arabidopsis 14-3-3 family of signaling regulators. *Plant Physiol* **2001**, *126*, 35-38, doi:10.1104/pp.126.1.35.
3. Winkelmann, T. Somatic Versus Zygotic Embryogenesis: Learning from Seeds. *Methods Mol Biol* **2016**, *1359*, 25-46, doi:10.1007/978-1-4939-3061-6_2.
4. Goldberg, R.B.; de Paiva, G.; Yadegari, R. Plant embryogenesis: zygote to seed. *Science* **1994**, *266*, 605-614, doi:10.1126/science.266.5185.605.

5. Pencik, A.; Tureckova, V.; Paulisic, S.; Rolcik, J.; Strnad, M.; Mihaljevic, S. Ammonium regulates embryogenic potential in Cucurbita pepo through pH-mediated changes in endogenous auxin and abscisic acid. *Plant Cell Tiss Org* **2015**, *122*, 89-100, doi:10.1007/s11240-015-0752-0.
6. Mendez-Hernandez, H.A.; Ledezma-Rodriguez, M.; Avilez-Montalvo, R.N.; Juarez-Gomez, Y.L.; Skeete, A.; Avilez-Montalvo, J.; De-la-Pena, C.; Loyola-Vargas, V.M. Signaling Overview of Plant Somatic Embryogenesis. *Front Plant Sci* **2019**, *10*, doi:ARTN 77
10.3389/fpls.2019.00077.
7. Kadokura, S.; Sugimoto, K.; Tarr, P.; Suzuki, T.; Matsunaga, S. Characterization of somatic embryogenesis initiated from the Arabidopsis shoot apex. *Dev Biol* **2018**, *442*, 13-27, doi:10.1016/j.ydbio.2018.04.023.
8. Souter, M.; Lindsey, K. Polarity and signalling in plant embryogenesis. *J Exp Bot* **2000**, *51*, 971-983, doi:10.1093/jexbot/51.347.971.
9. Gruel, J.; Deichmann, J.; Landrein, B.; Hitchcock, T.; Jonsson, H. The interaction of transcription factors controls the spatial layout of plant aerial stem cell niches. *NPJ Syst Biol Appl* **2018**, *4*, 36, doi:10.1038/s41540-018-0072-1.
10. Riechmann, J.L.; Heard, J.; Martin, G.; Reuber, L.; Jiang, C.Z.; Keddie, J.; Adam, L.; Pineda, O.; Ratcliffe, O.J.; Samaha, R.R.; et al. Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **2000**, *290*, 2105-2110, doi:DOI 10.1126/science.290.5499.2105.
11. Lotan, T.; Ohto, M.; Yee, K.M.; West, M.A.L.; Lo, R.; Kwong, R.W.; Yamagishi, K.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **1998**, *93*, 1195-1205, doi:Doi 10.1016/S0092-8674(00)81463-4.
12. Zuo, J.; Niu, Q.W.; Frugis, G.; Chua, N.H. The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant J* **2002**, *30*, 349-359, doi:10.1046/j.1365-313x.2002.01289.x.
13. Boutilier, K.; Offringa, R.; Sharma, V.K.; Kieft, H.; Ouellet, T.; Zhang, L.; Hattori, J.; Liu, C.M.; van Lammeren, A.A.; Miki, B.L.; et al. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **2002**, *14*, 1737-1749, doi:10.1105/tpc.001941.
14. Srinivasan, C.; Liu, Z.; Heidmann, I.; Supena, E.D.; Fukuoka, H.; Joosen, R.; Lambalk, J.; Angenent, G.; Scorza, R.; Custers, J.B.; et al. Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (Nicotiana tabacum L.). *Planta* **2007**, *225*, 341-351, doi:10.1007/s00425-006-0358-1.
15. Heidmann, I.; de Lange, B.; Lambalk, J.; Angenent, G.C.; Boutilier, K. Efficient sweet pepper transformation mediated by the BABY BOOM transcription factor. *Plant Cell Rep* **2011**, *30*, 1107-1115, doi:10.1007/s00299-011-1018-x.
16. Florez, S.L.; Erwin, R.L.; Maximova, S.N.; Guiltinan, M.J.; Curtis, W.R. Enhanced somatic embryogenesis in Theobroma cacao using the homologous BABY BOOM transcription factor. *Bmc Plant Biol* **2015**, *15*, doi:ARTN 121
10.1186/s12870-015-0479-4.
17. Shoemaker, R.C.; Couche, L.J.; Galbraith, D.W. Characterization of somatic embryogenesis and plant regeneration in cotton (Gossypium hirsutum L.). *Plant Cell Rep* **1986**, *5*, 178-181, doi:10.1007/BF00269112.
18. Trolinder, N.L.; Goodin, J.R. Somatic Embryogenesis and Plant-Regeneration in Cotton (Gossypium-Hirsutum-L). *Plant Cell Reports* **1987**, *6*, 231-234, doi:Doi 10.1007/Bf00268487.
19. Trolinder, N.L.; Chen, X.X. Genotype Specificity of the Somatic Embryogenesis Response in Cotton. *Plant Cell Reports* **1989**, *8*, 133-136, doi:Doi 10.1007/Bf00716824.
20. Tohidfar, M.; Mohammadi, M.; Ghareyazie, B. Agrobacterium-mediated transformation of cotton (Gossypium hirsutum) using a heterologous bean chitinase gene. *Plant Cell Tiss Org* **2005**, *83*, 83-96, doi:10.1007/s11240-004-6155-2.

21. Wu, J.H.; Zhang, X.L.; Nie, Y.C.; Jin, S.X.; Liang, S.G. Factors affecting somatic embryogenesis and plant regeneration from a range of recalcitrant genotypes of Chinese cottons (Gossypium hirsutum L.). *In Vitro Cell Dev-Pl* **2004**, *40*, 371-375, doi:10.1079/Ivp2004535.
22. Jin, S.; Zhang, X.; Nie, Y.; Guo, X.; Liang, S.; Zhu, H. Identification of a novel elite genotype for in vitro culture and genetic transformation of cotton. *Biol Plantarum* **2006**, *50*, 519-524, doi:10.1007/s10535-006-0082-5.
23. Li, J.Y.; Wang, M.J.; Li, Y.J.; Zhang, Q.H.; Lindsey, K.; Daniell, H.; Jin, S.X.; Zhang, X.L. Multi-omics analyses reveal epigenomics basis for cotton somatic embryogenesis through successive regeneration acclimation process. *Plant Biotechnol J* **2019**, *17*, 435-450, doi:10.1111/pbi.12988.
24. Wen, L.; Li, W.; Parris, S.; West, M.; Lawson, J.; Smathers, M.; Li, Z.; Jones, D.; Jin, S.; Saski, C.A. Transcriptomic profiles of non-embryogenic and embryogenic callus cells in a highly regenerative upland cotton line (Gossypium hirsutum L.). *BMC Dev Biol* **2020**, *20*, 25, doi:10.1186/s12861-020-00230-4.
25. Method of the Year 2011. *Nat Methods* **2012**, *9*, 1, doi:10.1038/nmeth.1852.
26. Manghwar, H.; Lindsey, K.; Zhang, X.; Jin, S. CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing. *Trends Plant Sci* **2019**, *24*, 1102-1125, doi:10.1016/j.tplants.2019.09.006.
27. Altpeter, F.; Springer, N.M.; Bartley, L.E.; Blechl, A.E.; Brutnell, T.P.; Citovsky, V.; Conrad, L.J.; Gelvin, S.B.; Jackson, D.P.; Kausch, A.P.; et al. Advancing Crop Transformation in the Era of Genome Editing. *Plant Cell* **2016**, *28*, 1510-1520, doi:10.1105/tpc.16.00196.
28. Bernardo, R. Prospective Targeted Recombination and Genetic Gains for Quantitative Traits in Maize. *Plant Genome* **2017**, *10*, doi:10.3835/plantgenome2016.11.0118.
29. Nam, J.; Matthysse, A.G.; Gelvin, S.B. Differences in susceptibility of Arabidopsis ecotypes to crown gall disease may result from a deficiency in T-DNA integration. *Plant Cell* **1997**, *9*, 317-333, doi:DOI 10.1105/tpc.9.3.317.
30. Jin, S.X.; Liu, G.Z.; Zhu, H.G.; Yang, X.Y.; Zhang, X.L. Transformation of Upland Cotton (Gossypium hirsutum L.) with gfp Gene as a Visual Marker (vol 11, pg 910, 2012). *J Integr Agr* **2012**, *11*, 1433-1433.
31. Horstman, A.; Li, M.F.; Heidmann, I.; Weemen, M.; Chen, B.J.; Muino, J.M.; Angenent, G.C.; Boutilier, K. The BABY BOOM Transcription Factor Activates the LEC1-ABI3-FUS3-LEC2 Network to Induce Somatic Embryogenesis. *Plant Physiology* **2017**, *175*, 848-857, doi:10.1104/pp.17.00232.
32. Lowe, K.; Wu, E.; Wang, N.; Hoerster, G.; Hastings, C.; Cho, M.J.; Scelonge, C.; Lenderts, B.; Chamberlin, M.; Cushatt, J.; et al. Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation. *Plant Cell* **2016**, *28*, 1998-2015, doi:10.1105/tpc.16.00124.
33. Pedroso, M.C.; Pais, M.S. Factors controlling somatic embryogenesis - Cell wall changes as an in vivo marker of embryogenic competence. *Plant Cell Tiss Org* **1995**, *43*, 147-154, doi:Doi 10.1007/Bf00052170.
34. Deng, T.; Yao, H.; Wang, J.; Wang, J.; Xue, H.; Zuo, K. GhLTPG1, a cotton GPI-anchored lipid transfer protein, regulates the transport of phosphatidylinositol monophosphates and cotton fiber elongation. *Sci Rep* **2016**, *6*, 26829, doi:10.1038/srep26829.
35. Ikeuchi, M.; Sugimoto, K.; Iwase, A. Plant callus: mechanisms of induction and repression. *Plant Cell* **2013**, *25*, 3159-3173, doi:10.1105/tpc.113.116053.
36. Liu, C.; Zhang, T. Expansion and stress responses of the AP2/EREBP superfamily in cotton. *BMC Genomics* **2017**, *18*, 118, doi:10.1186/s12864-017-3517-9.
37. Xie, Z.; Nolan, T.M.; Jiang, H.; Yin, Y. AP2/ERF Transcription Factor Regulatory Networks in Hormone and Abiotic Stress Responses in Arabidopsis. *Front Plant Sci* **2019**, *10*, 228, doi:10.3389/fpls.2019.00228.
38. Huang, M.; Hu, Y.; Liu, X.; Li, Y.; Hou, X. Arabidopsis LEAFY COTYLEDON1 controls cell fate determination during post-embryonic development. *Front Plant Sci* **2015**, *6*, 955, doi:10.3389/fpls.2015.00955.

39. Jo, L.; Pelletier, J.M.; Harada, J.J. Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *J Integr Plant Biol* **2019**, *61*, 564-580, doi:10.1111/jipb.12806.
40. Shani, E.; Yanai, O.; Ori, N. The role of hormones in shoot apical meristem function. *Curr Opin Plant Biol* **2006**, *9*, 484-489, doi:10.1016/j.pbi.2006.07.008.
41. Sarkar, A.K.; Luijten, M.; Miyashima, S.; Lenhard, M.; Hashimoto, T.; Nakajima, K.; Scheres, B.; Heidstra, R.; Laux, T. Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. *Nature* **2007**, *446*, 811-814, doi:10.1038/nature05703.
42. Haecker, A.; Gross-Hardt, R.; Geiges, B.; Sarkar, A.; Breuninger, H.; Herrmann, M.; Laux, T. Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. *Development* **2004**, *131*, 657-668, doi:10.1242/dev.00963.
43. Meinke, D.W.; Franzmann, L.H.; Nickle, T.C.; Yeung, E.C. Leafy Cotyledon Mutants of Arabidopsis. *Plant Cell* **1994**, *6*, 1049-1064, doi:10.1105/tpc.6.8.1049.
44. Braybrook, S.A.; Stone, S.L.; Park, S.; Bui, A.Q.; Le, B.H.; Fischer, R.L.; Goldberg, R.B.; Harada, J.J. Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *P Natl Acad Sci USA* **2006**, *103*, 3468-3473, doi:10.1073/pnas.0511331103.
45. Wang, L.; Gu, X.; Xu, D.; Wang, W.; Wang, H.; Zeng, M.; Chang, Z.; Huang, H.; Cui, X. miR396-targeted AtGRF transcription factors are required for coordination of cell division and differentiation during leaf development in Arabidopsis. *J Exp Bot* **2011**, *62*, 761-773, doi:10.1093/jxb/erq307.
46. Magwanga, R.O.; Lu, P.; Kirungu, J.N.; Lu, H.J.; Wang, X.X.; Cai, X.Y.; Zhou, Z.L.; Zhang, Z.M.; Salih, H.R.; Wang, K.B.; et al. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. *Bmc Genet* **2018**, *19*, doi:ARTN 6 10.1186/s12863-017-0596-1.
47. Lan, T.; Gao, J.; Zeng, Q.Y. Genome-wide analysis of the LEA (late embryogenesis abundant) protein gene family in Populus trichocarpa. *Tree Genet Genomes* **2013**, *9*, 253-264, doi:10.1007/s11295-012-0551-2.
48. Suzuki, Y.; Mae, T.; Makino, A. RNA extraction from various recalcitrant plant tissues with a cethyltrimethylammonium bromide-containing buffer followed by an acid guanidium thiocyanate-phenol-chloroform treatment. *Biosci Biotechnol Biochem* **2008**, *72*, 1951-1953, doi:10.1271/bbb.80084.
49. Chen, Z.J.; Sreedasyam, A.; Ando, A.; Song, Q.; De Santiago, L.M.; Hulse-Kemp, A.M.; Ding, M.; Ye, W.; Kirkbride, R.C.; Jenkins, J.; et al. Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. *Nat Genet* **2020**, *52*, 525-533, doi:10.1038/s41588-020-0614-5.
50. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **2012**, *9*, 357-359, doi:10.1038/nmeth.1923.
51. Li, B.; Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **2011**, *12*, 323, doi:10.1186/1471-2105-12-323.
52. 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, doi:10.1093/bioinformatics/btp616.