Comparative Transcriptome Analysis of SE initial dedifferentiation in cotton of different SE capability

Aiping Cao, Yinying Zheng, Yu Yu, Xuwen Wang, Dongnan Shao, Jie Sun & Baiming Cui

Somatic embryogenesis (SE) is a critical transition from vegetative to embryogenic growth in higher plants; however, few studies have investigated the mechanism that regulates SE initial differentiation. Most cotton varieties have not undergone regeneration by SE, so only a few varieties can be used in genetic engineering. Here, two varieties of cotton with different SE capabilities (HD, higher differentiation and LD, lower differentiation) were analyzed by high throughput RNA-Seq at the pre-induction stage (0h) and two induction stages (3h and 3d) under callus-induction medium (CIM). About 1150 million clean reads were obtained from 98.21% raw data. Transcriptomic analysis revealed that “protein kinase activity” and “oxidoreductase activity” were highly represented GO terms during the same and different treatment stages among HD and LD. Moreover, several stress-related transcription factors might play important roles in SE initiation. The SE-related regulation genes (SERKs) showed different expression patterns between HD and LD. Furthermore, the complex auxin and ethylene signaling pathway contributes to initiation of differentiation in SE. Thus, our RNA-sequencing of comparative transcriptome analysis will lay a foundation for future studies to better define early somatic formation in cotton with different SE capabilities.

Somatic embryogenesis (SE) resulting in regeneration of new plants is an important step in the Agrobacterium-mediated method. This process includes two stages. First, the cell reverts back to its own lineage to re-enter the cell cycle and transform into a dedifferentiated cell state. Callus formation is frequently considered the manifestation of the dedifferentiated cellular state. Second, calli can regenerate somatic embryos and new plants by redifferentiation. Dedifferentiation is frequently considered the manifestation of the stem cell-like state to switch fate preceding the commitment for proliferation. Following that, dedifferentiation is an important biological phenomenon in whole SE process.

Various molecular biological technologies have been described to investigate molecular mechanisms during SE in different plant species, including carrot, Arabidopsis, soybean and potato. These studies have revealed valuable information regarding the physiological and biochemical changes that occur during SE. In addition, many studies have investigated the initiation of SE in various plant species. Plant hormones and stress are the two most important factors involved in stimulation of SE initiation. However, the regulatory mechanism of vegetative-to-embryogenic triggered remains elusive. For example, numerous candidate genes associated with hormonal regulation, oxidative stress and response to stress were shown to be specifically activated or exhibit differential expression.

Cotton is a major raw material used by the textile industry and a source of oil. Transgenic technologies have been widely applied to cotton molecular breeding through Agrobacterium-mediated transformation via SE. However, few varieties have been successfully used in genetic engineering and regeneration in vitro. The majority of the desirable cotton cultivars are incapable of regeneration via SE, and many factors can effect SE, including culture conditions and tissue background. Furthermore, the key factor is a genotype-dependent response. Several recent studies have been conducted to investigate the molecular mechanism and identify genes critical for...
SE\textsuperscript{16}. More recently, RNA deep-sequencing technology has provided a platform for analysis of differences in gene expression\textsuperscript{17}. There are reports demonstrating that hormonal signaling pathways such as auxin and cytokinin\textsuperscript{16}, as well as stress-responsive pathways are critical during cotton SE development\textsuperscript{18}. However, the initial SE molecular mechanisms have not been thoroughly studied and are still unclear in different dedifferentiated cotton cultivars. Here, we investigated early events of SE during the without induction stage (0h) and induction stage (3h and 3 d) by comparing cotton of different SE capability (HD and LD) using RNA-seq profiling combined with differentially expressed genes (DEG) and GO analysis. This study provides a comprehensive analysis of gene expression of different cotton types during the initial stages of somatic embryogenesis. Furthermore, the data presented herein provides an abundant gene catalog that should be a useful resource for future studies of molecular and developmental mechanisms of SE.

Results

Differentiation efficiency of various cotton species. Four \textit{Gossypium hirsutum} \textit{L} cultivars (YZ1, R15, X33 and X42) were sampled at three stages, the pre-induction stage (0h) and two induction stages (3h and 3d) in the CIM. YZ1 and R15 (HD) were selected because they exhibited good SE potential and were the main transgenic materials. Although X33 and X42 (LD) are the major commercial cultivars in Xinjiang, China, they have a low rate of differentiation during SE relative to YZ1 and R15. The high differentiation lines and the recalcitrant lines showed diverse morphology during the callus induction stage. The early calli of HD were friable, loose, and primarily whitish (Fig. 1a–d and i–l), while those of LD were firmer and less friable than HD (Fig. 1e–g and m–p).
The significant differences in calli may be due to differences in gene expression during SE initiation; therefore, cDNA libraries were constructed from three time stages between HD and LD and two biological replicates were performed.

**Illumina HiSeq mRNA sequencing.** High-throughput RNA-Seq sequencing generated 1171 million raw reads from 24 RNA samples using the Illumina HiSeqTM 2000 platform, approximately 48 million reads for each line. After discarding adapters, poly-N and low quality reads from raw data, a total of 1150 million clean reads were obtained from the total raw reads (98.26%). Two biological replicates were evaluated for each line, and the results indicated good agreement of the RNA-seq data (0.958 < R2 < 0.986). Overall, 86.31%–88.94% of the high-quality reads were mapped to the *Gossypium hirsutum* reference genome using TopHat29. Of the mapped reads, 84.3–89.4% were distributed in exon regions, while 1.7–2.4% were located in introns and 8.0–13.2% were in intergenic regions (Table 1). For all sequence data, the average Q20, Q30, and GC contents were 97.03%, 92.64%, and 43.82%, respectively. The error rate of all samples was 0.01%–0.02% (Table S1). Of which three categories of mapped data: 1) multiple mapped (7.63–9.11%) and unique mapped reads (77.75–81.06%), 2) forward mapped (38.83–40.53%) and reverse mapped reads (38.88–40.54%), and 3) non-splice reads (48.35–53.15%) and splice reads (26.87–29.64%) (Table S1).

**DEGs and GO terms during same-stage comparisons.** Significant differences in gene expression were examined in three same-stage comparisons of the two difference species (HD and LD). Using a P-value < 0.05 and |log2FoldChange| > 2 in one sampling point as the higher criteria. Our data shows a heatmap of DEGs and Venn diagrams showing details regarding the number of DEGs at each pairwise comparison (Fig. 2). The results revealed 140 DEGs at 0h (Fig. 2b), 143 DEGs at 3h (Fig. 2c) and 123 DEGs at 3d (Fig. 2d). Of the 140 DEGs at 0h, 143 DEGs at 3h, and 123 DEGs at 3d, significant differences in gene expression were examined (Table 2).

| Sample name | Raw reads | Clean reads | Clean bases | Total mapped | Exon (%) | Intron (%) | Intergenic (%) |
|-------------|-----------|-------------|-------------|--------------|----------|------------|----------------|
| R15_0h_1    | 45813740  | 44974414    | 6.75 G      | 38815533     | 84.3     | 2.4        | 13.2          |
| R15_0h_2    | 44436266  | 43634922    | 6.55 G      | 38509898     | 85.7     | 2.2        | 12.1          |
| R15_3h_1    | 46339512  | 45602268    | 6.84 G      | 40352472     | 87.6     | 1.8        | 10.6          |
| R15_3h_2    | 55664292  | 54795000    | 8.22 G      | 48396413     | 88.2     | 1.8        | 10.0          |
| R15_3d_1    | 58297408  | 57185824    | 8.58 G      | 49594358     | 87.2     | 1.9        | 10.9          |
| R15_3d_2    | 45041322  | 44718654    | 6.71 G      | 39092086     | 86.4     | 1.8        | 11.8          |
| YZ1_0h_1    | 54411121  | 53434048    | 8.02 G      | 46862756     | 87       | 2          | 11            |
| YZ1_0h_2    | 42332338  | 41472822    | 6.22 G      | 36244862     | 87.2     | 2          | 10.8          |
| YZ1_3h_1    | 48182224  | 47414878    | 7.11 G      | 41720626     | 88.2     | 1.8        | 10            |
| YZ1_3h_2    | 51910688  | 51103664    | 7.67 G      | 45108556     | 89.4     | 1.7        | 8             |
| YZ1_3d_1    | 41334482  | 40551474    | 6.08 G      | 35085465     | 88       | 1.8        | 10.2          |
| YZ1_3d_2    | 53111840  | 51931008    | 7.79 G      | 45087397     | 87.9     | 1.7        | 10.3          |
| X33_0h_1    | 55528574  | 54606234    | 8.19 G      | 48091726     | 85.1     | 2.3        | 12.5          |
| X33_0h_2    | 48595188  | 47799278    | 7.17 G      | 42052352     | 86.5     | 2          | 11.5          |
| X33_3h_1    | 51455830  | 50559176    | 7.58 G      | 44749580     | 87.9     | 1.8        | 10.2          |
| X33_3h_2    | 47675942  | 46858792    | 7.03 G      | 41408584     | 87.8     | 1.7        | 10.5          |
| X33_3d_1    | 44247472  | 43462830    | 6.52 G      | 38113846     | 87.7     | 1.7        | 10.6          |
| X33_3d_2    | 45522790  | 44759502    | 6.71 G      | 39043962     | 86.8     | 1.8        | 11.3          |
| X42_0h_1    | 40809328  | 40079772    | 6.01 G      | 35489334     | 86.9     | 1.9        | 11.2          |
| X42_0h_2    | 49641394  | 48808912    | 7.32 G      | 43159066     | 86.6     | 2.1        | 11.3          |
| X42_3h_1    | 49429384  | 48648396    | 7.3 G       | 43221078     | 88.5     | 1.9        | 9.7           |
| X42_3h_2    | 45973326  | 45219188    | 6.78 G      | 40215890     | 87.3     | 2          | 10.7          |
| X42_3d_1    | 55203894  | 54233944    | 8.13 G      | 47493700     | 86.3     | 2          | 11.7          |
| X42_3d_2    | 49845162  | 49014372    | 7.35 G      | 45171216     | 87       | 1.9        | 11.1          |

Table 1. Summary of transcriptome sequencing data.

The significant differences in calli may be due to differences in gene expression during SE initiation; therefore, cDNA libraries were constructed from three time stages between HD and LD and two biological replicates were performed.

### DEGs and GO terms during same-stage comparisons

Significant differences in gene expression were examined in three same-stage comparisons of the two difference species (HD and LD). Using a P-value < 0.05 and |log2FoldChange| > 2 in one sampling point as the higher criteria. Our date shows a heatmap of DEGs and Venn diagrams showing details regarding the number of DEGs at each pairwise comparison (Fig. 2). The results revealed 140 DEGs at 0h (Fig. 2b), 143 DEGs at 3h (Fig. 2c) and 123 DEGs at 3d (Fig. 2d). Of the 140 DEGs at 0h, 143 DEGs at 3h, and 123 DEGs at 3d, significant differences in gene expression were examined (Table 2).

### Comparison of DEGs and GO terms at different-stages

DEGs and three main GO terms were identified by comparisons of two induction stages (3h and 3d) and the pre-induction stage (0h). Comparison of 3h vs 0h revealed 7175 and 10,718 DEGs in the HD and LD cultivars, respectively, and more than half (50.8%) of the DEGs were present in all cotton cultivars at 3h vs 0h (Fig. 3a and b). Upon GO classification based on biological processes, "oxidation-reduction process", "protein phosphorylation" and "regulation of transcription,"
DNA-dependent” were found to be associated with a higher number of DEGs in HD and LD (Fig. 3d). The six top molecular functions were “protein binding”, “ATP binding”, “protein kinase activity”, “protein tyrosine kinase activity”, “protein serine/threonine kinase activity” and “oxidoreductase activity” (Fig. 3d). Among cellular components, “membrane”, “integral component of membrane” and “nucleus” were highly represented in both cotton types (Fig. 3d).

Comparison of the 3d vs 0h samples revealed 9428 DEGs in the HD species and 8923 DEGs in the LD cultivars (Fig. 3a and c). “Oxidation-reduction process”, “translation” and “protein phosphorylation” were the main GO terms identified among differentially expressed genes corresponding to biological processes. Among molecular functions, the highly represented GO terms were “protein binding”, “ATP binding”, “structural constituent of ribosome”, “protein kinase activity”, “protein tyrosine kinase activity” and “DNA binding”. “Intracellular”, “ribosome” and “membrane” were grouped into cellular components in HD and LD (Fig. 3e).

The number of DEGs identified in LD was higher than in HD at 3h vs 0h, while the number of DEG detected at 3d vs 0h showed the opposite pattern. With the exception of the comparisons of LD at 3h vs 0h, the number of up-regulated DEGs was slightly higher than that of down-regulated DEGs in both comparisons (Fig. 3a). The cellular component was the most overrepresented GO term in different-stage comparisons between HD and LD.

Analysis of TF during SE initiation. Transcription factors (TFs) recognize DNA in a sequence-specific manner to regulate gene expression. Based on the \(|\text{log2FoldChange}| > 2\) at one sampling point, a total of 1590 specifically different transcription factors were annotated in 69 TF families between two cultivars. Most TFs were involved in response to pathogen challenge, including AP2–EREBP (211), MYB (155), bHLH (128), WRKY (91), NAC (89), HB (73), C2H2 (72), GRAS (47), bZIP (43), AUX/IAA (42), C2C2-Dof (38), CH3 (33) and CCAAT (24) (Fig. 4a). Putative homologs of SE differentiation related genes were identified in two cotton cultivars (Fig. 4b–d). UPBEAT1 (UPB1), belongs to bHLH TF family, three UPB1 homologues were same expression model and were up-regulated during at 3h, while then the expression level reduced to very low (FPKM < 1) at 3d. However, another two bHLH transcription factors, MYC1 and GL3, showed higher expression levels at 3d. The expression of MYC1 homologues increased with induction. GL3 decreased at 3h, then increased at 3d. The MYB transcription factors CPC and ETC1 showed different expression patterns during three time stages. GAI is a sub-family of the GRAS family, the value of FPKM was decreased with induction. bZIP transcription factors, bZIP1, bZIP44 and acid responsive elements-binding factor, ABF2 and ABF3, most of those TFs showed higher FPKM at the 3h, while bZIP44 (Gh_A06G1847) was generally opposite expression pattern.
inferred that these GO terms would be more likely to play a role in cells differentiating during the SE process. Therefore, we enriched analysis revealed that protein kinase activity, oxidation-reduction processes and protein/ATP binding were associated with a higher number of DEGs in different and same time stages between HD and LD. The DEGs and GO groups showed many differences in the 3h vs 0h and 3d vs 0h comparisons, indicating that 3h were associated with a higher number of DEGs in different and same time stages between HD and LD. Moreover, only a few cultivars can produce SE. The number of cotton totipotent cells differentiating into calli is partially expressed during early SE (|log2FoldChange| > 2 at one sampling point) were selectively analyzed. The expression patterns of SE-related genes are shown in Table 3 and Fig. 5. WUSCHEL-related homeobox genes were involved, including WOX5 (1) and WOX11 (2). Overall, WOX homeobox genes were significantly up-regulated in the two cotton cultivars at 3d vs 0h, while there was no significant difference in expression between 3h and 0h. WOX5 and WOX11 exhibited similar expression patterns at 3d vs 0h, which were just significant up-regulated in HD and LD. In the present study, SERK1 (somatic embryogenesis receptor kinase 1), SERK2 (somatic embryogenesis receptor kinase 2), BAK1 (SERK3/BR11-associated kinase 1), TDR/PXY (TDIF receptor/phloem intercalated with xylem) and BAK1 were identified in two cotton types. SERK1 showed a similar expression pattern at 3h vs 0h between HD and LD. Five homologous genes of BAK1 were found to be down-regulated at 3h vs 0h, while just one BAK1 (Gh_D12G0758) was down-regulated in 3d vs 0h among all cotton cultivars. The CLAVATA3/ESR-RELATED (CLE) family of seven genes was showed, these genes involving CLE1, CLE5, CLE27, CLE44 and TDIF. At 3h vs 0h, CLE1 and CLE27 showed the opposite expression in LD cotton, while CLE44 and two TDIF were down-regulated between HD and LD. At 3d vs 0h, CLE44 and one TDIF were down-regulated in both cultivars, while CLE5 (Gh_D01G0413) and TDIF (Gh_A07G1469) were up-regulated in LD and HD, respectively. Five homologous genes of EBE were significantly upregulated at 3h vs 0h in all cotton cultivars. With the exception of Gh_A11G0359 and Gh_A12G2541, the three EBE homologous genes were also up-regulated at 3d vs 0h in the two cotton cultivars, while one EBE (Gh_A11G0309) was up-regulated at 3d vs 0h in the HD. One YUC4, four YUC8 and one YUC10 were confirmed to show differential expression patterns between cotton types. Only YUC4 was up-regulated in the HD cultivars at 3h vs 0h. At 3d vs 0h, YUC8 and YUC10 showed the opposite expression patterns in the two types.

Validation of RNA-seq data by qRT-PCR. To further validate the RNA-seq data, 11 DEGs were randomly selected along with their specific primers for qRT-PCR analysis. The cotton ERFi10 (NCBI Reference Sequence: XM_016892582.1) was used for relative gene expression normalization. In addition, RT-qPCR data were analyzed using the Origin 8 software. As shown in Fig. 6, there were strong positive correlations (R² = 0.98–0.96) between RNA-seq data and qRT-PCR data, demonstrating the reliability of the RNA-seq data.

Discussion

Transcriptome analysis of SE initial dedifferentiation in two cotton cultivars. *Gossypium hirsutum* is one of the most recalcitrant cultivars for *in vitro* plant regeneration through somatic embryogenesis. Moreover, only a few cultivars can produce SE. The number of cotton totipotent cells differentiating into calli is key to manipulating the SE process. Despite a few studies conducted to determine the molecular mechanisms regulating SE initial dedifferentiation, different cultivars for SE initial dedifferentiation have been not well elucidated. In the present study, gene expression changes during SE initiation and formation were analyzed between the cultivars in which it was easy or difficult to produce SE. Our analyses showed that the number of DEGs and GO identified in similar types of cotton at different induction stages were generally higher than those observed during same-stage comparisons between HD and LD, indicating that HD and LD of different differentiation ability may be not only its varieties of differences, the more important be the response diversity for induction.

The DEGs and GO groups showed many differences in the 3h vs 0h and 3d vs 0h comparisons, indicating that 3h and 3d of induction may be the main regulation phase in the SE initiation. Previous studies have shown that hypocotyls cultured for 3h had no morphological changes compared to non-induction hypocotyls. In tobacco and Arabidopsis, cellular dedifferentiation and initiation of cell division by protoplasts occurs within 48–72h. Moreover, histological observations have shown that cotton somatic cells undergo initial differentiation within 72h. Furthermore, GO enrichment analysis revealed that protein kinase activity, oxidation-reduction processes and protein/ATP binding were associated with a higher number of DEGs in different and same time stages between HD and LD. Therefore, we inferred that these GO terms would be more likely to play a role in cells differentiating during the SE process.

| Gene id | log2 Fold Change | Gene description |
|---------|-----------------|------------------|
| Gh_A06G1786 | −3.6825 | G-type lectin S-receptor-like Serine/Threonine-kinase |
| Gh_A03G310 | 2.0274 | Protein kinase superfamily protein |
| Gh_A06G1617 | −3.2889 | G-type lectin S-receptor-like Serine/Threonine-kinase |
| Gh_A11G2003 | 1.3657 | Protein kinase superfamily protein |
| Gh_A02G1701 | 1.2525 | MAP kinase kinase 2 |
| Gh_A06G1786 | −2.4601 | G-type lectin S-receptor-like Serine/Threonine-kinase |
| Gh_D10G0162 | 2.2204 | Protein kinase superfamily protein |
| Gh_A06G1617 | −5.1736 | Protein kinase superfamily protein |
| Gh_A11G2002 | −2.3912 | Protein kinase superfamily protein |
| Gh_D07G2384 | −2.0518 | Protein kinase superfamily protein |
| Gh_A06G1786 | −2.8407 | G-type lectin S-receptor-like Serine/Threonine-kinase |
| Gh_A06G1617 | −2.5474 | G-type lectin S-receptor-like Serine/Threonine-kinase |
| Gh_D07G2384 | −2.7462 | G-type lectin S-receptor-like Serine/Threonine-kinase |

Table 2. Protein kinase-related genes differentially expressed in the same-stage.
Stress-related transcription factors regulate differentiation of cotton SE. Transcription factors play regulatory roles in the embryogenesis processes in different plant species. In the present study, 624 TFs were identified, among which the TF families involved in diverse stress responses as well as in SE processes such as MYB, bHLH, bZIP and GRAS are of particular interest. In Arabidopsis SE initiation, half of the stress-related transcription factors can be induced by 2, 4-D. Stress and differentiation are strongly interrelated, overlapping processes, with certain cells being competent to switch cell fate. In our study, most of the TF families displayed diverse expression profiles that have complex regulatory functions during SE in cotton. Furthermore, the most differentially expressed TFs were identified in the greater the difference differentiation rate between the varieties. In Arabidopsis, bHLH transcription factors play a pivotal role in cell fate determination. UPBEAT1 (UPB1), a bHLH transcription factor, regulated the balance between cell proliferation and differentiation by directly regulating the expression of a set of peroxidases. Three UPB1 homologues were expressed at higher levels at 3h, and the characterized UPB1 might indicate that cell fate was changed from cellular proliferation to initial differentiation at 3h. MYC1 and GL3 influence cell differentiation for root formation. MYB proteins work together with BHLH proteins in a variety of cellular processes including control of cell fate determination and regulation of the cell cycle. The MYB transcription factors, CPC and ETC1, are central regulators of cell differentiation in Arabidopsis thaliana, and ETC1 has enhancer functions with CPC. In the present study, the expression patterns of CPC and ETC1 were consistent with those during three time stages. Finally, lower expression levels were observed in the CPC gene than in the two ETC1 (Fig. 4b,c).

GA1 is a highly homologous DELLA protein. DELLA proteins are also GA signaling repressors that are responsive to GA-induced degradation by the ubiquitin-proteasome pathway, as well as activation of GA-mediated physiological and developmental processes. GA is a negative regulator of SE in Arabidopsis. As revealed in this study, GA1 decreased with increasing induction times. In plants, the bZIP transcription factors regulate diverse functions, such as plant development and stress response. In the present study, bZIP1, ABF2 and ABF3 showed the same expression pattern. GmbZIP1 was most closely related to AtABF2. Cross-talk between bZIP1 and ABA occurred, potentially leading to improved tolerance to abiotic stresses. In contrast, GmbZIP44 is a negative regulator of ABA signaling that showed the opposite expression model compared to bZIP1. In the present study, bZIP homologues showed diverse expression profiles, which suggest a complex regulation by bZIP proteins as an ABA-dependent transcription factor in cotton SE. In addition, low ABA:GA ratios...
acted synergistically to stimulate SE in *Medicago truncatula*34. Our results suggest that these TFs interactions and crosstalk regulate stress response and influence the initiation of SE differentiation. However, further research is required to determine how TFs affect differentiation of cotton SE.

**Reactive oxygen species (ROS) regulated differentiation of SE initiation.** According to our transcriptomic data, the oxidation-reduction process was significantly enriched in most comparisons. Redox homeostasis is essential to sustain metabolism and growth. In this study, some oxidative phosphorylation related genes also showed differential expression. The strength and duration of signaling regulated reactive oxygen species (ROS) via the redox-dependent signal transduction pathway38. NADPH oxidase (RBOH protein) is an important generator of ROS in the *M. truncatula* SE induction period34. The major NADPH oxidase catalytic subunits of homologues *AtRBOHD* and *AtRBOHF* are required for accumulation of ROS intermediates in the plant defense response in Arabidopsis39. In our study, the same expression model was observed in two *RBOHD* homologous genes, with relatively high expression occurring at 3h compared to 0h and 3d (Fig. 7a,b). In addition, *RBOHA*, *RBOHB* and *RBOHC* were found to have different expression patterns. Most studies of the interaction between ROS and other genes have been well documented. *AtRBOHD* and *AtRBOHF* mediate ABA-induced ROS production in Arabidopsis39. The *UPBEAT1* gene of the bHLH TF family controls the transition from cellular proliferation to differentiation by directly regulating the expression of ROS39. DELLA activity is regulated by environmental variability40, and DELLA then regulate plant growth and stress tolerance through modulation of
**Table 3.** The log2 fold change of SE-related genes in different-stage comparisons. (Note: Among the 3d vs 0h, Inf indicates that the FPKM of WOX5 was close to 0 at 3d).

ROS levels. Our results suggest that synergistic interactions between ROS and other genes may be essential for the differentiation of SE initialization.

**Crosstalk among auxin and ethylene contributes to initiation of SE differentiation.** Auxin and ethylene are critical plant growth regulators (PGRs) for initiating differentiation in the induction of SE. The differentiation of cotton is correlated with sharp changes in endogenous auxin levels, which may be one of the first steps regulating SE differentiation. Aux/IAAs and ARFs are the key transcription factors involved in regulating the expression of auxin-responsive genes. Aux/IAAs bind auxin response promoter elements, mediate transcriptional responses to auxin, and regulate auxin-mediated transcriptional activation/repression together with ARFs. ARFs bind auxin response promoter elements, mediate transcriptional responses to auxin-responsive genes.

ARFs and Aux/IAAs are down-regulated during two induction stages (Fig. 8a). The function of ARFs may be negatively regulated by IAA14 and other Aux/IAAs in Arabidopsis. In addition, auxin induces degradation of Aux/IAAs, which release ARFs to regulate transcription of their target genes, thus establishing a negative feedback loop. In the present study, two IAA14 (Fig. 8b) displayed different expression patterns, with one (Gh_Sca006835G01) showing the same pattern as ARF19. Gh_A09G1947 was up-regulated at 3h, then down-regulated at 3d. YUCCA (YUC) is an auxin biosynthetic gene, and YUC10 is essential for embryogenesis and leaf formation in Arabidopsis. In the present study, YUC10 increased continuously during the three time stages (Fig. 8a).

AP2/EREBP is crucial for ethylene-regulated developmental processes, and ethylene may play a positive role in SE. ERFs contain an AP2 DNA-binding domain of transcriptional factors. **ERF1 (ETHYLENE RESPONSE FACTOR1)** is a key integrator of the ethylene responses in developmental processes. In our study, two ERF1 and two ERF1 were identified (Fig. 8c), with the highest expression level being observed in 3h. Ethylene also interferes with auxin during the initiation of embryogenesis. Previous research showed that auxin could increase ethylene levels via activation of the expression of ACC (1-aminocyclopropane-1-carboxylic acid) synthase genes, leading to induction of the expression of ARF19 to regulate downstream gene expression. **ARF19 serves as a cross**
In the present study, the FPKM of ACS1 and ACS8 showed the lowest level at 3d (Fig. 8d,e). Both auxin related genes and ethylene related genes were differentially expressed in the two cotton cultivars. Overall, auxin related genes and ethylene related genes showed complex expression profiles, and the results indicate that they may crosstalk with each other during SE dedifferentiation in different cultivars.

Complex regulation of LRR-RLKs in cotton SE initiation. In the present study, many genes were classified into "protein kinase" and "protein/ATP binding" terms, such as LRR-RLKs (leucine-rich receptor-like protein...
kinase family protein), which have been shown to play important roles in cell differentiation and embryo pattern formation. SERKs, TDR/PXY and BAM1 belong to the LRR-RLK family of proteins. In Arabidopsis, BAM1 plays an important regulatory role in stem cell maintenance. In the present study, BAM1 showed significant differences at 3h vs 0h and 3d vs 0h (Table 3). SERK genes play an essential role in determining embryogenic competence. In Arabidopsis, SERK proteins included two distinct groups, one group containing SERK1 and SERK2, BAK1, SERK4, and SERK5 are clustered to another group. Our results showed that SERK1, SERK2, BAK1 and SERK1 displayed similar expression patterns between HD and LD, while SERK2 and BAK1 expression differed significantly in LD at 3h vs 0h. BAK1 cannot replace SERK1 in male fertility. TDR/PXY also belongs to the LRR-RLK subclass. SERKs serve as co-receptors in TDIF that interact with PXY to regulate plant development. CLE (CLAVATA3/EMBRYO SURROUNDING REGION-RELATED) contains a conserved CLE domain at the C-terminal region and a hydrophobic signal peptide at the N-terminal region (CLAVATA3/EMBRYO). Based on bioactivity and receptor specificity, R-type CLE and H-type CLE can be recognized in the CLE peptide family. CLV3 (CLAVATA3) belongs to R-type CLE, while the H-type CLE includes TDIF, and CLE peptides play important roles in regulating differentiation of stem cells and maintenance of shoot apical meristems in Arabidopsis. When compared to HD and LD, CLE proteins showed more differences in LD species at different time stages. When combined with the results of previous studies, our results indicate that LRR-RLKs play a role in the SE process and regulate the SE differentiation rate.

Materials and Methods

Plant materials and tissue culture conditions. Four cultivars of Gossypium hirsutum, YZ1, R15, X33 and X42, were investigated, among which YZ1 and R15 have a relatively high SE differentiation rate and therefore comprise the main transgenic material. Although X33 and X42 are the major commercial cultivars in Xinjiang, China, they have a low rate of differentiation during SE compared with YZ1 and R15. Specimens of all cultivars have been conserved in our laboratory.

Four cotton seeds were sterilized with 0.1% (w/v) Hg₂Cl₂ for 10 min, then rinsed five times with sterilized distilled water. The sterilized seeds were germinated on ½ MS for hypocotyl induction, then incubated at 28 °C in the dark for 6 d. Next, hypocotyls were excised from sterile seedlings, cut into 1 cm segments and placed on callus-induction medium (CIM; MS medium plus B5 vitamins, supplemented with 0.05 mg l⁻¹ IAA, 0.05 mg l⁻¹ kinetin, 0.05 mg l⁻¹ 2,4-D, pH 5.8) for 2 weeks. Induction cultures were conducted under 16 h light: 8 h dark conditions at 28 °C. Different stages of explants at 0h, 3h and 3d were used for transcription analysis.

RNA and library preparation for transcriptome sequencing. Total RNA was extracted from each sample using the Purelink™ RNA Mini Kit (Life Technologies, Carlsbad, CA, USA) following the manufacturer’s protocol. We collected 24 samples from two biological replicates of each sample and monitored RNA degradation and contamination on 1% agarose gels. The RNA purity was checked using a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA), while the RNA concentration and integrity were measured using a Qubit® RNA Assay Kit and a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). RNA samples were stored at −80 °C until further processing. For each sample, a total of 3 μg RNA was used as the input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kits for Illumina® (NEB, USA) according to the manufacturer’s instructions, and index codes were added to attribute sequences to each sample. Clustering of the index-coded samples was performed on a cBot Cluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer’s recommendations. Finally, paired-end reads of 125 bp/150 bp were generated via an Illumina HiSeq platform.
Bioinformatics analysis of RNA-Seq data. Quality control is first step, clean reads were obtained by removing adaptor tags, reads containing poly-N and low quality reads from the raw reads of fastq format. In the addition, the clean reads of Q20, Q30 and GC were calculated, and high quality clean reads were used in all downstream analyses. For annotation, all reads were mapped to the reference sequences. An index of the reference genome was constructed using Bowtie v2.2.3 and paired-end clean reads were aligned to the reference genome using TopHat v2.0.12. To quantify the gene expression level, we employed HTSeq v0.6.1 to calculate the numbers of mapped reads and then normalized the results to the FPKM (expected number of Fragments Per Kilobase of transcript sequence per millions of base pairs sequenced), which is the most commonly used normalized method for estimating gene expression levels that considers the effects of sequencing depth and gene length for the reads at the same time. We used the DESeq R package (1.18.0) to confirm the differential expression analysis. The resulting P-values were adjusted using Benjamini and Hochberg’s approach for controlling the false discovery rate. We used a threshold value of an adjusted P-value (padj) ≤ 0.05 and a log2FoldChange ≥ 2 in at least one of these stages.

Functional classification of differentially expressed genes (DEGs). Functional analysis of DEGs including Gene Ontology (GO) and KEGG was performed. GO enrichment analysis of DEGs was implemented by the GOSeq R package. We used the KOBAS software to test the statistical enrichment of differential expression genes in the KEGG database. GO terms and KEGG pathways with corrected P-values ≤ 0.05 were considered the thresholds to determine significant enrichment by differentially expressed genes.

Figure 8. Expression of auxin (a–c) and ethylene (d–h) related genes at three time stages, 0h, 3h and 3d, in the two cultivars, HD and LD.
qRT-PCR analysis. Eleven DEGs were selected to estimate the validation of RNA-seq data by qRT-PCR. Gene primers (Table S2) were designed using NCBI/Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome) and synthesized by BGITECH. The cDNA was synthesized from 1 μL of total RNA (200–500 ng) using the PrimerScript RT reagent Kit (Takara, Dalian, China) in a 10 μL reaction mixture according to the manufacturer's instructions. qRT-PCR was performed in 10 μL reactions on a LightCycler® 480 (Roche) using 1 μL of first-strand cDNA as the template, 5 μL of 2 × SYBR Premix Ex Taq II (Tli RanseH Plus) (Takara, Dalian, China), 0.4 μL each of 10 μM forward and reverse gene-specific primers and 3.2 μL of ddH2O in 96-well plates. The cotton ERF1α (NCBI Reference Sequence: XM_016892582.1) was used as an internal standard. The qRT-PCR conditions were as follows: pre-incubation at 95 °C for 30 s, followed by amplification by 45 cycles at 95 °C for 10 s, 60 °C for 10 s and 72 °C for 10 s. qRT-PCR analysis was conducted in the three biological replicates. Data were analyzed using the Origin 8 software.

References
1. Jiménez, V. M. Involvement of Plant Hormones and Plant Growth Regulators on in vitro Somatic Embryogenesis. Plant Growth Regulation 47, 91–110 (2005).
2. Kurczyńska, E. U. & Mazur, E. Histological analysis of direct somatic embryogenesis in Arabidopsis thaliana (L.) Heynh. Planta 226, 619–628 (2007).
3. Fehler, A., Pasternak, T. P. & Dudits, D. Transition of somatic plant cells to an embryogenic state. Plant Cell, Tissue and Organ Culture (PCTOC) 74, 201–228 (2003).
4. Grafi, G. How cells dedifferentiate: a lesson from plants. Developmental Biology 268, 1–6 (2004).
5. Toonen, M. A. J. et al. Description of somatic-embryo-forming single cells in carrot suspension cultures employing video cell tracking. Planta 194, 565–572 (1994).
6. Tuhaud-Nissen, F., Shealy, R. T., Khanna, A. & Vodkin, L. O. Clustering of Microarray Data Reveals Transcript Patterns Associated with Somatic Embryogenesis in Soybean. Plant Physiology 132, 118–136 (2003).
7. Sharma, S. K., Müssig, S., Hedley, P. E., Mcnicol, J. & Bryan, G. J. Molecular regulation of somatic embryogenesis in potato: an auxin led perspective. Plant Molecular Biology 68, 185–201 (2008).
8. Zhen, X. & C. Z. et al. Transcriptome Profiling Reveals Auxin and Cytokinin Regulating Somatic Embryogenesis in Different Sister Lines of Cotton Cultivar CCR214 T. Journal of Integrative Plant Biology 55, 631–642 (2013).
9. Karami, O. & Saidi, A. The molecular basis for stress-induced acquisition of somatic embryogenesis. Molecular Biology Reports 37, 2493–2507 (2010).
10. Karami, O., Aghavaisi, B. & Mahmoudi, P. A. Molecular aspects of somatic-to-embryogenic transition in plants. Journal of Chemical Biology 2, 177–190 (2009).
11. Zeng, F. et al. Isolation and characterization of genes associated to cotton somatic embryogenesis by suppression subtractive hybridization and macroarray. Plant Molecular Biology 60, 167–183 (2006).
12. Zhang, X. Regulation of Somatic Embryogenesis in Higher Plants. Critical Reviews in Plant Sciences 29, 36–57 (2010).
13. Jin, S. et al. Identification of a novel elite genotype for in vitro culture and genetic transformation of cotton. Biologia Plantarum 50, 519–524 (2006).
14. Zheng, W. et al. AtWuschel Promotes Formation of the Embryogenic Callus in Gossypium hirsutum. Plos One 9, e87502–e87502 (2014).
15. Kumria, R. et al. High-frequency somatic embryo production and maturation into normal plants in cotton (Gossypium hirsutum) through metabolic stress. Plant Cell Reports 21, 635–639 (2003).
16. Yang, X. et al. Transcript profiling reveals complex auxin signalling pathway and transcription regulation involved in dedifferentiation and redifferentiation during somatic embryogenesis in cotton. BMC Plant Biology 12, 1–19 (2012).
17. Wang, X. W. et al. De novo characterization of a whitely transcriptome and analysis of its gene expression during development. BMC Genomics 11, 326–333 (2010).
18. Cheng, W. H. et al. De novo transcriptome analysis reveals insights into dynamic homeostasis regulation of somatic embryogenesis in upland cotton (G. hirsutum L.). Plant Molecular Biology 92, 279–292 (2016).
19. Zhu, H. G. et al. Analysis of genes differentially expressed during initial cellular dedifferentiation in cotton. Science Bulletin 53, 3666–3676 (2008).
20. El, O. S. et al. Control of somatic embryogenesis and embryo development by AP2 transcription factors. Plant Molecular Biology 74, 313–326(S14) (2010).
21. Ilorca, C. M. M. P., Utrik Zentgraf. hZIPs and WRKYs: two large transcription factor families executing two different functional strategies. Frontiers in Plant Science 5, 169–169 (2014).
22. Zhao, H., Li, X. & Ma, L. Basic helix-loop-helix transcription factors and epidermal cell fate determination in Arabidopsis. Plant Signaling & Behavior 7, 1556–1560 (2012).
23. Gilwrick, M., Nowak, K., Balazadeh, S., Mullerrooer, B. & Gaj, M. D. Extensive Modulation of the Transcription Factor Genome during Somatic Embryogenesis in Arabidopsis thaliana. Plos One 8, e92613 (2013).
24. Grafi, G. et al. Plant response to stress meets dedifferentiation. Planta 233, 433–438 (2011).
25. Grafi, G., Florentin, A., Ransbotyn, V. & Morgenstern, Y. The stem cell state in plant development and in response to stress. Frontiers in Plant Science 2, 1085–1091 (2011).
26. Tsukagoshi, H., Busch, W. & Benley, P. N. Transcriptional regulation of ROS conversion from proliferation to differentiation in the root. Cell 143, 606–616 (2011).
27. Bernhardt, C., Lee, M. A., Zhang, F., Lloyd, A. & Schiefelbein, J. The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. Development 130, 6431–6439 (2004).
28. Zhang, F., Gonzalez, A., Zhao, M., Payne, C. T. & Lloyd, A. A network of redundant bHLH proteins functions in all TTTG1-dependent pathways of Arabidopsis. Development 130, 4859–4869 (1991).
29. Kirk, V., Simon, M., Wester, K., Schiefelbein, J. & Hulskamp, M. ENHANCER of TRY and CPC 2 (ETC2) regulates redunancy in the region-specific control of trichome development of Arabidopsis. Plant Molecular Biology 55, 389–398 (2004).
30. Kirk, V., Simon, M., Hulskamp, M. et al. ENHANCER OF TRY AND CPC1 gene acts redundantly with TRIPYCHON and CAPRICE in trichome and root hair cell patterning in Arabidopsis. Developmental Biology 268, 506–513 (2004).
31. Pien, J. et al. The Arabidopsis Ga1 gene defines a signaling pathway that negatively regulates gibberellin responses. Genes & Development 11, 3194–3205 (1998).
32. Itoh, H., Ueguchitana, M., Sato, Y., Ashikari, M. & Matsuoka, M. The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. Plant Cell 14, 57–70 (2002).
33. Richards, D. E., King, K. E., Tahar Aitali, A. & Harberd, N. P. How Gibberellin Regulates Plant Growth and Development: A Molecular Genetic Analysis of Gibberellin Signaling. Plant Biology 52, 67–88 (2001).
34. Nolan, KimE., Siyang Liao, Y. S., Saeed, NaziaA., Zhang, Xiyi & Rose, Rayl. An Unusual Abscisic Acid and Gibberellin Acid Synergism Increases Somatic Embryogenesis Efficiency, Facilitates Its Genetic Analysis and Improves Transformation in Medicago truncatula. Plos One 9, e99908 (2014).
The sequence flanking the N-terminus of the CLV3 peptide is critical for its cleavage and freezing tolerance in transgenic Arabidopsis. Plantas 228, 225–240 (2008).

58. Murashige, T. & Skoog, F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiologia Plantarum 15, 473–497 (1962).

59. C. T. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28, 511–515 (2010).

Acknowledgements
This work was supported by Breeding of Transgenic Short-season Cotton New Varieties in Exceptional Early Mature Cotton Region [2014ZX08005002-004].

Author Contributions
A.C., Y.Z., B.C., and Y.Y. conceived and designed the experiments. A.C., D.S., X.W. and L.Y. performed the experiments. A.C. and B.M. analyzed the data. A.C., Y.Z. and Y.Y. contributed reagents/materials/analysis tools. A.C. wrote the paper. All authors read and approved the final manuscript.

Additional Information
Supplementary information accompanies this paper at doi:10.1038/s41598-017-08763-8

Competing Interests: The authors declare that they have no competing interests.

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

Open Access
This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017