

Transcriptomic profiles of non-embryogenic and embryogenic callus cells in a highly regenerative Upland Cotton line (*Gossypium hirsutum* L.)

Li Wen

Clemson University

Wei Li

Clemson University

Stephen Parris

Clemson University

Matthew West

Clemson University

John Lawson

Clemson University

Michael Smathers

Clemson University

Zhigang Li

Clemson University

Don Jones

Cotton Incorporated

Shuangxia Jin

Huazhong Agriculture University College of Science

Christopher Saski (✉ saski@clemson.edu)

Clemson University College of Agriculture Forestry and Life Sciences <https://orcid.org/0000-0002-2780-4274>

Research article

Keywords: *Gossypium hirsutum* L, Somatic embryogenesis, Transcriptomes

Posted Date: May 31st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28441/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at BMC Developmental Biology on December 1st, 2020. See the published version at <https://doi.org/10.1186/s12861-020-00230-4>.

Abstract

• Background

• Genotype independent transformation and whole plant regeneration through somatic embryogenesis relies heavily on the intrinsic ability of a genotype to regenerate.

• Results

• In this study, gene expression profiles of a highly regenerable *Gossypium hirsutum* L. cultivar, Jin668, were analyzed at two critical developmental stages during somatic embryogenesis, non-embryogenic callus (NEC) cells and embryogenic callus (EC) cells. The rate of EC formation in Jin668 is 96%. Differential gene expression analysis revealed a total of 5,333 differentially expressed genes (DEG) with 2,534 upregulated and 2,799 downregulated in EC. A total of 144 genes were unique to NEC cells and 174 genes unique to EC. Clustering and enrichment analysis identified genes upregulated in EC that function as transcription factors/DNA binding, phytohormone response, oxidative reduction, and regulators of transcription; while genes categorized in methylation pathways were downregulated. Four key transcription factors were identified based on their sharp upregulation in EC tissue; *LEAFY COTYLEDON 1* (LEC1), *BABY BOOM* (BBM), *FUSCA* (FUS3) and *AGAMOUS-LIKE15* with distinguishable subgenome expression bias.

• Conclusions

• This comparative analysis of NEC and EC transcriptomes gives new insights into the genetic underpinnings of somatic embryogenesis in cotton.

Background

The domain of plant transformation has enabled foundational discoveries in plant biology over the last 30 years. Plant genetic engineering has facilitated fundamental knowledge about gene function, trait genetics, and established critical linkages between the genome and the presentation of traits. Transgenic approaches have been widely used in the improvement and breeding of many crops, such as soybean, rice, and maize [1–4]. Furthermore, the rapid pace of technological advancements in genome sequencing has delivered high quality reference genomes of major domesticated crop plants and many non-model and wild species that have led to deep understandings of domestication and genetic diversity, gene function and functional genomics, and the development of tools for precise plant breeding [5]. However, the rate of genome characterization has far outpaced our ability to functionally profile genes and biochemical pathways in crops.

Genome editing through engineered nucleases [6] or CRISPR-Cas9 systems (spCas9-NG, base editing, xCas9, Cpf1, Cas13, Cas14) [7] are unprecedented technological breakthroughs that behold disruptive potential to precisely edit the genome of living organisms. These systems offer biologists the ability to ask more precise biological questions; which has led to a new capacity in understanding gene function through targeted knockouts. Furthermore, genome editing technology has the potential to drastically reduce plant breeding cycle times and endow tailored trait genetics of elite breeding lines which ultimately hold the potential to revolutionize commercial agriculture [5, 8]. However, major obstacles remain in deploying these frontier technologies for global crop improvement. A main bottleneck is in plant transformation and regeneration through somatic embryogenesis, and a secondary challenge is the delivery of the genome editing reagents to plant cells to produce the intended effects [5]. Agrobacterium-mediated transformation [9] and subsequent whole plant regeneration from a single somatic cell is perhaps the most historic and popular method to produce plants homoplastic for the transgene [9]. However, this type of whole plant regeneration is generally limited to a narrow range of genotypes within a species, usually with poor agronomic traits and with low efficiency [10–12].

Somatic embryogenesis (SE) is a unique process of embryo development that involves cycles of cellular de-differentiation and reprogramming events that are controlled through signaling networks and gene expression cascades that eventually lead to embryonic cells [13]. During the developmental reprogramming, somatic cells de-differentiate to generate non-embryogenic cells from various explant sources, such as hypocotyls, young leaves, and immature embryos [14]. NEC cells can further differentiate to generate EC cells which function in somatic embryo development [15], which can be described by four primary stages in dicots: globular embryos, heart-shaped embryos, torpedo embryos, and cotyledon embryos. Embryogenic cells are considered to be developmentally plastic and their programming of a particular developmental pathway is heavily influenced by environmental factors imposed by the tissue culture micro-environment, in addition to gene expression and regulation.

Manipulation of these programs *in vitro* has been achieved with some success through large Design of Experiments (DOE) that determine and adjust the choice and combination of growth regulators, mainly the balance of auxin and cytokinin, for each genotype [16–18]. Somatic embryogenesis is also sensitive to carbohydrate sources, inorganic salts, antibiotics, and amino acids [19–22], further adding to the complexity of the process and the difficulty of optimizing a protocol. In addition to medium components, genotype also plays a significant role in a species' ability to regenerate, exhibiting vastly different responses across closely related genotypes [23–27].

The molecular mechanisms that orchestrate somatic embryogenesis and endow biological totipotency are becoming better understood [28]. Previous studies have shown that genes expressed during the induction of SE can be divided into three primary categories; stress-related genes, plant growth regulator (PGR) related genes, and transcription factors [29–31]. A biological stress, such as senescence or an acute stress such as abiotic/biotic factors can trigger stem cell formation through altered chromatin conformation and promiscuous expression of transcription factors [32–34]. Transcription factors have been shown to function in an important role on the regulation of plant differentiation [35] and development [36]. In monocots, overexpression of various plant transcription factors such as *LEAFY COTYLEDON1* [37], *WUSCHEL* [38], and *BABY BOOM* [39] have been shown to improve embryo formation and enhanced regeneration

[12]. In dicots, overexpression of *BABY BOOM* to enhance regenerative capacity has been reported in tobacco [40], sweet pepper [41], and *Theobroma cacao* [42].

In upland cotton (*Gossypium hirsutum* L.), successful somatic embryogenesis was reported nearly 40 years ago [43], however mature plants were not obtained until 1983 in the genotype Coker 310 [44]. After another five years, a report for the optimal media formulation for the induction of somatic embryogenesis from mature and immature tissues from Coker 312 was published [45]. A follow-up study screened 38 cotton cultivars for embryogenic potential and found only Coker 312 to be responsive and reported a high degree of genotype specificity in embryogenic potential [46]. Another study substantiated that somatic embryogenesis is a heritable trait in cotton and suggested that it is polygenic [47]. Coker lines are not well-suited for transformation studies because of their general poor agronomic qualities. In addition, Coker lines suffer from slow growth in tissue culture, a decline in vigor, and low regenerative capacity of the cultures – which includes low potency of embryogenesis, difficulty of embryo germination, a low rate of rooting, low transplant success, and a progressive loss of totipotency in culture [48]. Recently, a line with a much higher embryogenic potential and better agronomic traits was released called Jin668 [31]. This genotype is a Coker relative and was developed through successive regeneration acclimation (SRA) that involves cycling a genotype through somatic embryogenesis multiple times [31]. This continuous cycling resulted in altered DNA methylation patterns that had an effect on gene expression profiles that persisted in the germ line with a greater regeneration efficiency [31].

In this study, we analyzed and compared the transcriptional landscape of NEC and EC cells derived from Jin668. We present and discuss the various classes of genes that are active during the transition from NEC to EC cells, and identify several new candidate genes that enhances our knowledge of somatic embryogenesis in upland cotton.

Results

Somatic embryogenesis efficiency of Jin668

The embryogenic potential of Jin668 is high. NEC cells can be distinguished from EC cell by visual discrimination. NEC is characterized as fluffy, non-differentiated or polarized cells that are actively dividing, Fig. 1A. Typical EC is characterized as small cell clusters, with condensed cytoplasm, that are rapidly dividing and proliferating, Fig. 1B. After inducing callus for 45 days, 4% of the explants were found to be in the NEC stage and subsequently did not develop into embryos (Fig. 1a and Table 1). The remaining 96% of the explants were found to be induced to form characteristic embryonic callus cells (Table 1 and Fig. 1b). In all experimental replications, both NEC and EC cells could be observed in the same piece of callus (Fig. 1a and 1b). Therefore, those calli that consisted of both NEC and EC cells were sampled for gene expression profiling between the NEC and EC (Fig. 1b).

Table 1
Callus induction of cotton hypocotyls explants

No. Explants	No. NEC	No. EC (efficiency rate %)
300	12	288 (96)
320	12	308 (96)
330	13	317 (96)

Differential gene expression and GO enrichment of NEC and EC callus

For each callus cell type, an average of 36.8 and 42.7 million read pairs were produced for NEC and EC samples, respectively (BioProject ID PRJNA629328). A total of 39,105 genes were expressed in both NEC and EC callus cells, Fig. 2. Overall, a total of 5,333 genes were differentially expressed with 2,534 upregulated and 2,799 downregulated in EC (Additional file 1). Among these genes, a total of 144 were uniquely expressed in NEC cells, whereas, 174 genes were unique to EC (Fig. 2, Additional file 2: Table S1). Genes with unique expression to either EC or NEC cells were mostly genes of unknown function. Genes with annotations that were highly expressed in NEC include ROTUNDIFOLIA-LIKE 21, gibberellin-regulated, WRKY DNA binding, and others. Genes with high expression unique to EC callus with annotations include MYB family transcription factor, cytochrome C biogenesis, ROTUNDIFOLIA-LIKE 10, heat shock, and several others (Additional file 2: Table S1). Gene expression fold changes ranged from - 12 to 12 in EC when compared to NEC, (Additional file 2: Table S2). A slightly larger set of transcripts indicated transcriptional downregulation from NEC to EC, with a total of 6,786 genes ($\log_{2}FC \geq 1$) becoming downregulated in EC; while 6,538 ($\log_{2}FC \geq 1$) were upregulated (Fig. 3A and 3B, Additional file 2: Table S2). The top 10 genes with the sharpest fold changes in upregulation EC versus NEC is a gene with a nodulin late domain (*Gohir.D11G247300.1*), and 9 genes with unknown function (Table 2). The top 10 genes with the sharpest fold changes in downregulation are a gibberellin-regulate homolog (*Gohir.D04G048900.1*), 4 genes with unknown function, a gene with an S-adenosyl-L-methionine decarboxylase (AdoMetDC) leader peptide, a ROTUNDIFOLIA like gene (*Gohir.D07G179500.2*), and 60S ribosomal protein (Table 2). Gene ontology enrichment analysis of the upregulated genes in EC (Fig. 2A and B) discovered genes enriched in biological processes such as trehalose biosynthesis, transmembrane transporter, sugar transport, iron binding, growth factor activity, embryo development, development, auxin efflux, and response to stress among other processes related to cellular reprogramming and differentiation (Fig. 4A, Additional file 2: Table S3). The most pronounced enriched category was DNA binding and transcription factor activity with 309 genes combined, followed by regulation of transcription (155 genes), membrane (228 genes), oxidation reduction (105 genes), and transmembrane transport (80 genes) (Additional file 2: Table S3). Notable categories that were downregulated were categorized by assigned GO terms enriched for oxidation-reduction (524 genes), transmembrane transport (233 genes), response to hormones and auxin (48 genes), methyltransferase activity (112 genes), and peroxidase activity (58 genes) (Additional file 2: Table S4).

Table 2
Top 10 genes upregulated in EC relative to NEC

Gene	Control	Observed Effect	logFC	PValue	FDR	Functional Domains (descriptions)	Functional Domains (domain name)	Athaliana_
Gohir.D11G247300.1	NEC	EC	12.2179854	2.64E-60	2.44E-56	Nodulin_late	Nodulin_late	
Gohir.A09G068700.1	NEC	EC	11.0311639	2.26E-16	1.41E-14	DUF3963	DUF3963	
Gohir.A03G181200.1	NEC	EC	10.8541708	5.80E-67	1.34E-62	#N/A	#N/A	
Gohir.A05G291150.1	NEC	EC	10.7942425	2.50E-24	5.12E-22	DUF4912	DUF4912	
Gohir.A09G187300.1	NEC	EC	10.6509234	1.66E-19	1.72E-17	#N/A	#N/A	
Gohir.A11G045350.1	NEC	EC	10.6244362	0.00116229	0.00472295	#N/A	#N/A	
Gohir.D12G136400.1	NEC	EC	10.6014584	6.50E-05	0.0003752	#N/A	#N/A	
Gohir.A08G165050.1	NEC	EC	10.5716199	1.71E-34	1.16E-31	#N/A	#N/A	
Gohir.A07G162000.3	NEC	EC	10.488521	1.52E-25	3.68E-23	#N/A	#N/A	
Gohir.D11G067750.1	NEC	EC	10.3328265	7.12E-05	0.00040677	#N/A	#N/A	
Gohir.D04G048900.1	NEC	EC	-9.94544461036326	1.19E-12	3.79E-11	GASA	GASA	GASR7 - Gi regulated GASA/GAS family prot precursor, expressed; protein hor
Gohir.A09G040900.1	NEC	EC	-9.974189713011	0.0001288	0.0006863	#N/A	#N/A	
Gohir.D06G144900.2	NEC	EC	-9.98990989847144	1.08E-39	1.16E-36	#N/A	#N/A	
Gohir.D10G179800.1	NEC	EC	-10.4131140476177	5.93E-22	8.57E-20	VanY	VanY	
Gohir.D05G002150.1	NEC	EC	-10.4414002634445	0.00169419	0.00655082	#N/A	#N/A	
Gohir.A11G280900.1	NEC	EC	-10.675999492066	9.84E-05	0.00054239	Phage_holin_3_7	Phage_holin_3_7	
Gohir.D13G226251.1	NEC	EC	-10.6995332832501	0.00118551	0.00480505	#N/A	#N/A	
Gohir.D05G008300.1	NEC	EC	-11.6191438066039	7.96E-05	0.00044992	AdoMetDC_leader	AdoMetDC_leader	S-adenosyl methionine decarboxyl leader pept putative, ex conserved upstream c reading fra
Gohir.D07G179500.2	NEC	EC	-11.6654565629349	0.00012965	0.00069003	DVL	DVL	expressed ROTUNDIF 21
Gohir.D08G229750.1	NEC	EC	-12.7143630581706	1.18E-53	5.45E-50	Ribosomal_L18A, DUF1891	Ribosomal_L18A, DUF1891	60S riboso protein L18 putative, ex Ribosomal L18ae/LX 1 protein

Hierarchical clustering and GO enrichment of key clusters

Hierarchical clustering of statistically significant gene expression profiles identified 3 distinct sub-clusters when cut into 50% of the tree (Fig. 5A-C, Additional file 2: Table S5). Subcluster 1 contained 2,186 genes, and the expression trend of these genes ranged from moderate to low in NEC and EC callus, respectively (Fig. 5A, Additional file 2: Table S5). Processes enriched in this subcluster were assigned to oxidation reduction, methyltransferase activity, iron binding, and defense (Fig. 5A, Additional file 2: Table S6). Subcluster 2 contained 675 genes, and showed an upward trend in expression in EC callus and enrichment of processes related to transcription factor activity, membrane, growth factor, and developmental processes (Fig. 5B, Additional file 2: Table S7). Subcluster 3 contains 183 genes, and specifically showed a more discreet trend of higher expression in NEC and lower expression in EC callus, with enrichment in processes such as membrane and transport, peroxidase activity, carbohydrate metabolism, and fatty acid binding (Fig. 5C, Additional file 2: Table S8).

Phytohormone gene expression profiles in NEC and EC callus

Genes involved in phytohormone signaling and biosynthesis are crucial for the transition from NEC to EC. In total, there are 6,132 genes annotated as a phytohormone in the *Gossypium hirsutum* (CV TM1) reference genome assembly [49]. In general, we observed an abundance of phytohormone related genes being downregulated in EC callus relative to NEC (Table 3). Abscisic acid and auxin related genes were the categories with the most abundant differentially

expressed genes, including 1,508 and 1,332 genes, respectively (Table 3). Expression magnitudes ranged from -10 (logFC) in genes involved in gibberellic acid to 9 (logFC) in genes involved in abscisic acid (Additional file 2: Table S9 and Fig. 6). Genes annotated as cytokinin, gibberellic acid, jasmonic acid, salicylic acid, ethylene, and brassinosteroid were differentially expressed in NEC and EC (Table 3 and Additional file 2: Table S9).

Table 3
Phytohormone genes involved in NEC and EC callus

Phytohormone	No. Genes	No. Upregulated (> 1 FC)	No. Downregulated (> 1 FC)	logFC range
Auxin	1332	129	188	-8 to 7
Cytokinin	355	42	47	-9 to 5
Gibberellic acid	461	48	71	-10 to 8
Abscisic acid	1508	150	213	-8 to 9
Jasmonic acid	781	82	108	-8 to 5
Salicylic acid	593	74	88	-5 to 5
Ethylene	782	82	129	-6 to 5
Brassinosteroid	320	47	40	-8 to 8

Transcription factors and signaling cascades

Transcription factors and cellular signaling genes that regulate reprogramming and differentiation during somatic embryogenesis were differentially expressed. We found 24 different classes of transcription factors that were significantly differentially expressed (Additional file 2: Table S10). The class with the most abundant genes was the GRAS (11 genes), followed by TGA like (9), MADS-box (9), CBF/NF-Y/archaeal histone domain (9), GATA (8), and others with 7 or fewer members (Additional file 2: Table S10). Specific genes with known involvement in embryogenesis were individually analyzed for expression profiles in NEC and EC callus in Jin668. Our data shows that noteworthy embryogenesis transcription factors are lowly expressed in NEC callus with several key genes that are sharply upregulated in EC (Table 4). For example, the gene with the highest upregulation is *LEAFY COTYLEDON 1* (*LEC1*), with the dominant copy residing in the D-subgenome (Table 4). The next highly expressed embryogenesis related gene was the morphogenic regulator *BABY BOOM* (*BBM*) with the dominant copy on the A-subgenome, followed by *FUSCA* (*FUS3*) and *AGAMOUS-LIKE15* (*AGL15*) both with the dominant copies on the D-subgenome. Interestingly, *WUSHEL* (*WUS*) had very little to no expression with no apparent subgenome bias (Table 4).

Table 4
Embryogenesis genes and their expression values in Jin668

GeneID	Gene Description	Arabidopsis Ortholog (GeneID)	A-Subgenome			D-Subgenome		
			GeneID	NEC TMM	EC TMM	GeneID	NEC TMM	EC TMM
<i>WOX5</i>	WUS-related homeobox 5	AT3G11260	Gohir.A10G233000	5.015	0.043	Gohir.D10G245300	2.261	0.234
<i>WUS</i>	<i>WUSHEL</i>	AT2G17950	Gohir.A10G098300	0.223	0.773	Gohir.D10G089500	0	0
			Gohir.A12G059800	0	0	Gohir.D12G060100	0	0
<i>WRKY2</i>	WRKY2	AT5G56270	Gohir.A08G026100	3.614	6.522	Gohir.D08G036600	5.855	6.839
			Gohir.A13G154100	3.503	6.941	Gohir.D13G158700	2.199	4.291
<i>GRD/RKD4</i>	GROUNDING	AT5G53040	Gohir.A03G051800	0.092	0.682	Gohir.D03G115300	0.027	0.351
<i>BBM</i>	BABY BOOM	AT5G17430	Gohir.A08G227000	4.889	26.81	Gohir.D08G247400	0.372	8.677
<i>LEC1</i>	LEAFY COTYLEDON1	AT1G21970	Gohir.A13G132600	4.363	85.08	Gohir.D13G136000	6.251	110.289
			Gohir.A08G025100	0.526	8.433	Gohir.D08G035600	2.069	15.03
<i>FUS3</i>	FUSCA	AT3G26790	Gohir.A07G230400	0.98	17.37	Gohir.D07G237600	1.353	23.266
<i>ABI3</i>	ABSCISIC ACID INSENSITIVE3	AT3G24650	Gohir.A07G154900	0.3	6.25	Gohir.D07G161100	0.047	1.09
<i>AGL15</i>	AGAMOUS-LIKE15	AT5G13790	Gohir.A08G141500	0.193	2.283	Gohir.D08G162600	0.485	4.502
			Gohir.A12G100400	0.948	10.92	Gohir.D12G103400	0.901	14.455

Regulation of DNA methylation in NEC and EC

DNA methylation has a critical role in governing gene expression. In general, we observed a larger downregulation of genes regulating various methylation pathways in EC. We identified 226 genes with at least one-fold change in EC (71 upregulated, 155 downregulated) relative to NEC (Additional file 2: Table S11). Of the downregulated methylation genes, 15 proteins belong to protein family methyltransferase, 26 belong to plant invertase/pectin methyltransferase inhibitors,

14 involved in S-adenosyl-L-methionine-dependent methyltransferase, 12 in the pectin methylesterase inhibitor family, and 35 in the O-methyltransferase superfamily (Additional file 2: Table S11). In contrast, very few methylation related protein families were upregulated in EC callus. Those that were include ribosomal protein L11 methyltransferase (3), tRNA methyltransferase (3), SAM dependent carboxyl methyltransferase (3), N6-adenine methyltransferase (3), and several others (Additional file 2: Table S11).

RT-qPCR and validation of RNA-seq

To validate the RNA-seq data, ten critical embryogenesis related genes were selected for RT-qPCR analysis. All the primer pairs used in this test amplify a unique band with a unique melting curve peak (data not shown). The results showed (Fig. 7 and Additional file 2: Table S12) that five of the predicted genes were up-regulated in EC and are as follows: *GhUGt73C5*, UDP-glycosyltransferase 73C5-like gene (*Gohir.A01G078300.1*) [50]; *GhPin2*, auxin transporters, encodes an auxin efflux carrier (*Gohir.A05G001400.1*) [51]; *GhNFYB6*, nuclear factor Y subunit B-6, acts as a key component regulating embryogenesis and seed maturation in *Arabidopsis thaliana* (*Gohir.A05G176400.1*) [52]; *GhORG2*, transcription factor ORG2-like gene with basic helix-loop-helix (*Gohir.A09G106500.1*) [53]; *GhBBM*, AP2-like ethylene-responsive baby boom transcription factor (*Gohir.D08G247400.1*) [54]. The five deduced genes that were down-regulated in EC, i.e., *GhARF4*, auxin response factor 4 like, encodes a member of the ARF family of transcription factors (*Gohir.A05G074600.1*) [55]; *GhACRY1/BLU1*, cryptochrome-1-like, a flavin-type blue-light photoreceptor with ATP binding and autophosphorylation activity (*Gohir.A05G226100.1*) [56] (Pooam et al., 2018); *GhOMT1*, encoding a reticuline 7-O-methyltransferase-like gene (*Gohir.D02G181400.1*); *GhCOMT1*, raimondii caffeic acid 3-O-methyltransferase-like (*Gohir.D12G246900.1*) gene [57]; *GhCKX3*, cytokinin dehydrogenase 3 (*Gohir.D07G011600.1*) [58]. The results showed that the expression level evaluated by RT-qPCR exhibited similar patterns to the RNA-seq results of the ten genes, confirmed the results of RNA-seq. (Fig. 7, Additional file 2: Table S12).

Discussion

Cotton is a major oilseed and fiber crop that could benefit from biotechnological improvement but is currently restricted to a few semi-recalcitrant genotypes that are amenable to low transformation frequencies and subsequent regeneration through somatic embryogenesis. Genotype specific recalcitrance to regeneration is common in most crop species and has proven to be a profound bottleneck in applying the latest advances in biotechnology to elite breeding material. In monocots, several advancements have been made in ectopically expressing key transcription factors (eg, *BBM/WUS2*) to improve or enhance the development of embryogenic cell formation [39]. However, this system does not directly translate to dicot system improvement. Here, we describe a gene expression changes during critical developmental time points in a non-recalcitrant genotype, Jin668 [31].

Our comparative results identified many genes of unknown function are active during the transition from NEC to EC in Jin668. Of those with annotations, we found the expression of many transcription factors increased in EC cells, such as TF NikR, AP2/ERF, MYB-CC, subunit A of Y, PWRKY, BREVIS RADIX, GTE1/GTE6, CBF/NF-Y, GATA, GRAS, IIF, MYC/MYB, TGA, k-box, MADS-box, NFYB/HAP3, TCP, GT3 and members of the heat shock TF family. Transcription factors are known to have important roles that guide the regulation of plant differentiation and development [36]. For example, BREVIS RADIX and MYB are involved in cell proliferation and elongation, GATA TFs function in cell differentiation and early embryos initiation and development, TCP are involved in cell spatial orientation and determination, Both GRAS and k-box regulate plant reprogramming, and MYC/MYB, MYB-CC, MADS-box, CBF/NF-Y, AP2/ERF, GT3 and TGA transcription factors function in the response to phytohormones and/or biotic or abiotic stresses [59, 60]. We also identified several genes that were upregulated in EC callus that are involved in phytohormone response, such as genes involved in carrier, efflux, response to auxin, cytokinin dehydrogenase, and homeobox domain. Plant cells use these TFs to transmit signals to respond to environmental stimuli or initiate developmental programs. Furthermore, studies have shown non-cell-autonomous transcription factors are crucial for successful somatic regeneration [61].

Cellular proliferation and elongation are also very important for callus development during somatic regeneration. In Jin668 at the EC stage, we found upregulation of the TFs BREVIS RADIX (BRX) and MYC/MYB when compared to the NEC stage. A previous study demonstrated that the *BRX* gene controlled the extent of cell proliferation and elongation in the growth zone of root tips in *Arabidopsis* [62]. Thereafter, BRX was found to be involved in regulation of brassinosteroid biosynthesis and maintenance to keep brassinosteroid biosynthesis above a critical threshold, which ultimately affects lateral root development in *Arabidopsis* [63].

We also found upregulation of 12 *GATA* genes which indicates that they play an important role in embryo initiation in cotton. GATA TFs are widely distributed in fungi, animals, and plants consist of a protein family containing a DNA-binding domain recognizing the DNA consensus sequence (A/T) GATA(A/G) and one or two highly conserved type-IV zinc fingers (C-X2-C-X17 - 20-C-X2-C) [64]. In plants, GATA TF expression is involved in light-mediated processes such as flowering, maturation, petal differentiation and expansion, and embryo development. [65]. In the *Gossypium* genus, 30 GATA TFs have been identified [64]. Moreover, Nawy et al reported that a GATA transcription factor, HANABA TARANU (HAN) was required to position the inductive *Arabidopsis* pro-embryo boundary and revealed that HAN regulated transcription in the basal pro-embryo [66].

Previous studies suggest that GRAS transcription factors act as essential regulators, not only in polar development [67] and reprogramming [68], but also in response to biotic and abiotic stresses [69, 70], and auxin response [71, 72]. Stuurman et al. reported that *HAM* (mutant hairy meristem) is a gene encoding a putative transcription factor of the *GRAS* family and mediates the signal from differentiating cells. This mechanism controls signals from differentiating tissues that extrinsically control stem cell fate in the shoot apex [59]. We found 11 *GRAS* genes upregulated in EC, suggesting that the upregulation of these genes is associated with embryo initiation and formation during somatic regeneration in cotton.

Previous work has shown that transient activation or ectopic expression of one or several transcription factors can trigger the transition from NEC to EC or increase the frequency of EC formation. These include *WOX5* [73], *WUSHEL* [74], *WRKY2* [75], *GRD/RKD4* [76], *BBM* [77], *LEC1* [54], *FUS3* [78], *ABI3* [79], and *AGL15* [80]. For example, the AP2/ERF TF *BBM* in plants induced plentiful cell regeneration and was used to produce a large number of somatic embryos that could developed to seedlings [81]. Furthermore, ectopic expression of *BBM* derived from *Brassica napus* was transformed into pepper and obtained efficiently

regenerated transgenic plants from recalcitrant sweet pepper (*C. annuum*) varieties [41]. Our data revealed dramatic upregulation of *GRD/RKD4*, *BBM*, *LEC1*, *FUS3*, *ABI3* and *AGL15* genes, both in A- and D-Subgenome in EC during the somatic embryogenesis. We also found four genes annotated as *WRKY2* that were upregulated in EC with subtle fold changes (1.18 to 1.95 logFC). Interestingly, only one of the four copies of *WUS* (*Gohir.A10G0998300*), was found up-regulated in EC callus, while the other three genes were not expressed, suggesting this homeologous subgenome specific gene is essential for somatic embryogenesis in cotton. Moreover, unlike the gene mentioned above, two *WOX5* genes, *Gohir.A10G233000* and *Gohir.D10G245300*, were found down-regulated in EC callus (Table 4). We hypothesize that this *WOX5* transcription factor may be critical in triggering the embryo initiation, while the other transcription factors contribute as regulatory or maintenance factors that are critical to embryo formation and development. Furthermore, this homeobox gene was reported to contribute to signaling spiral differentiation of stem cells in *Arabidopsis thaliana*; while loss of *WOX5* function in the root meristem stem cell niche caused terminal differentiation in distal stem cells [82]. Our data also indicates that a series of related transcription regulators are involved in signaling of embryo initiation and formation in cotton.

Phytohormone sensing is critical to cell fate reprogramming during somatic embryogenesis. Our results showed that during conversion from NEC to EC, a series of genes involved in the biosynthesis and transportation of auxin and cytokinin are upregulated, which indicates that these two phytohormones play important roles during the initiation of the embryos.

Auxins are key components to regeneration, and their exogenous application has shown to recuperate the embryogenic potential of mitotically quiescent somatic cells [83]. In *Triticum aestivum*, regeneration efficiency was associated with auxin exposure time and catalase metabolism during somatic embryogenesis. Induction of pluripotent cells, termed, “callus,” by auxin represents a typical cell fate change required for plant in vitro regeneration [84]. Studies have shown that relative signaling through MONOPTEROS (MP)/ARF 5 was required for shoot formation in Arabidopsis callus. Moreover, variants of MP expression revealed that this gene can promote *de novo* shoot formation in tissues that are normally recalcitrant [85]. Cytokinin also plays a fundamental role in cell fate reprogramming and is implicated in the process of shoot organogenesis [86]. Cytokinin promotes shoot regeneration through the up-regulation of *WUS*, and recent studies demonstrated that key cytokinin signaling components, namely type-B ARABIDOPSIS RESPONSE REGULATORS (ARRs), directly bind the *WUS* promoter and regulate transcription of this gene [87]. These studies also suggest that the transcriptional regulation of early embryo patterning to hormonal control of plant callus cell initiation is comparable to similar strategies for stem cell formation in the animal kingdom [87]. Furthermore, it was found that the *WUS* expression is controlled by the ratio of cytokinin with auxin [88]. Our results show approximately 3 times more auxin related genes being upregulated in EC than cytokinin genes (129 vs. 42), suggesting a critical role for these genes in somatic embryogenesis in cotton.

Conclusions

Understanding endogenous changes in developmental biology at the molecular level is paramount to developing strategies toward genotype independent transformation and subsequent whole plant regeneration. Here we describe the critical transcriptional landscape of a highly regeneratable elite line, Jin668, and report genes and new targets that are important in somatic embryogenesis. Optimizing transformation and regeneration systems is a large stride toward realizing the promises of synthetic biology and genome editing. We anticipate that this work will lead to new understandings of somatic embryogenesis in upland cotton and other dicots.

Methods

Genotype and plant material source, callus, and sampling

The non-recalcitrant plant genotype used for tissue culture was Jin668 (*Gossypium hirsutum* L.), which was developed through successive regeneration acclimation (SRA) [89]. Jin668 seeds were provided by one of the co-authors, Dr. Shuangxia Jin from Huazhong Agricultural University, in 2017 December and then germinated in April of 2018 at the Clemson University Systems Genetics lab. Seeds were sacrificed by sulfuric acid for 1 min, then rinsed under running deionized water for 2 h. The seeds were washed with 70% ethanol for 1 min and then washed three times with sterile deionized water. The seeds then were sterilized by shaking them in a flask containing 100 mL of 10% commercial bleach (+ 2 drops of Tween-20) for 5 min and rinsed three times with sterile DW. Finally, they were cultured in Magenta boxes on germination media (4.33 g/L Murashige and Skoog (MS) salts (MS basal salts mixture; PhytoTechnology Laboratories, cat. no. M524), 2% glucose (PhytoTechnology Laboratories, USA), pH 6.0, 0.26% Phytigel (PhytoTechnology Laboratories, USA) at 28 °C, under kept in dark conditions for 7 days. The 7-day hypocotyls were cut to 0.5-1.0 cm segments and placed on callus induction media [90], with media changes every 4 weeks. The plates were cultivated in a growth chamber with controlled conditions of 28 ± 1 °C, 16 h (day)/8 h (night) photoperiod, with light provided by cool-white fluorescent lamps at an irradiation of 135 μmolm⁻² s⁻¹, and 50% relative humidity. The regeneration rates were calculated by the number of explants producing embryogenic callus per 100 explants (hypocotyl segments) within the 1.5-month tissue culture process, with 3 replicates. The embryonic and non-embryonic calli were harvested and pooled as samples. Both the embryonic and non-embryonic calli were collected as 3 replicates.

RNA isolation and RNA sequencing

Total RNA was isolated from Jin668 callus material at different stages using a modified guanidine thiocyanate method [91]. RNA-Seq libraries were constructed using the Illumina TruSeq Stranded RNA kit (San Diego CA, USA) following the manufacturer's recommendations with three biological replicates per condition. Paired-end sequences were collected on an Illumina NovaSeq through a third-party vendor. Raw sequence files were trimmed of low quality bases and adapter sequences with the trimmomatic software package [92].

Identification of differentially expressed genes and enrichment analysis

Preprocessed reads were aligned to the *Gossypium hirsutum* v2.0 reference assembly (Chen et al, 2020) with the Bowtie2 short read aligner [93], and alignment files were coordinate sorted and indexed with samtools v1.3.1 [94]. Raw counts, FPKM and TMM values were generated with RSEM [95]. Counts for

each replicate were collected in tabular format and normalized according to biological replicates to remove sample-specific effects with EdgeR [96]. Relative pairwise changes in gene level expression were made using a generalized linear model (glm) and classic pairwise comparison methods with the EdgeR package [96]. Genes abounding a p-value of ≤ 0.01 were corrected by the false discovery rate (FDR) of 0.05 [97]. Gene level fold-change values were output in tabular format and candidate genes abounding assigned thresholds were clustered into functional pathways with the GOSeq software tool [98] and filtered by error corrected P-values ≤ 0.05 .

Gene validation through RT-qPCR

Quantitative real-time PCR was performed to verify the RNA-Seq data [55]. Three biological replicates of total RNA of cotton callus for RNA-seq were used for RT-qPCR. Briefly, 0.5-1 μg of total RNAs were reverse-transcribed to first strand cDNAs using M-MuLV reverse transcriptase (New England Biolabs, GA, USA) and primed by d(T)25-VN following manufacturer's instructions. RT-qPCR was conducted on an iCycler iQ system (Bio-Rad, Hercules, CA, USA) in 20 μl of PCR reaction solution (SsoAdvanced™ Universal SYBR® Green Supermix, Bio-Rad, USA) with 100 nM of each primer; the input amount of the first strand cDNA per reaction is around equivalent to 0.6–1.2 ng of total RNA. Four technical repeats were used for each of the three biological replicates. PCR was conducted with the following program: an initial denature at 95 °C for 60 s, followed by 40 cycles of 95 °C for 20 s, 62 °C for 20 s, and 72 °C for 20 s. Finally, a melting curve was performed from 55.0 °C to 95.0 °C at 0.5 °C increments. The amplicons were resolved on 2% agarose gel in 0.5 × TBE buffer with ethidium bromide and purified for sequencing. The $\Delta\Delta\text{Ct}$ method was used for real-time PCR analysis. Two reference genes, *GhPP2A1* [99] and *GhUBQ7* [89] were used as endogenous controls. Relative expression level was calculated using the $2^{-\Delta\Delta\text{Ct}}$ formula. All primer pairs except for the reference genes used for examining the expression levels of cotton endogenous genes were designed based on the Jin668 RNAseq sequences and listed in Additional file 2: Table SY.

Abbreviations

NEC: Non-embryogenic callus; EC: Embryogenic callus; DEG: Differentially expressed genes; RT-qPCR: Quantitative reverse transcription polymerase chain reaction; SRA: Successive regeneration acclimation; PGR: Plant growth regulator; LEC1: LEAFY COTYLEDON 1; BBM: BABY BOOM; TF: Transcription factor

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Additional file 1. The transcriptomic datasets generated during the current study are available in the BioProject database of NCBI under the accession number PRJNA629328, (<http://www.ncbi.nlm.nih.gov/bioproject/629328>) and in the supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by NIFA/USDA hatch funds under project number SC-1700530 and Technical Contribution No. 6867 of the Clemson University Experiment Station.

Authors' contributions

LW, SJ, DJ, and CS conceived and designed the study. LW and WL prepared callus conducted tissue culture. WL, SP, ZL, MW, MS, JL, and LW collected and analyzed the RNAseq data. LW, CS, SP, ZL, MW, MS, JL, DJ, SJ wrote and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

N/A

References

1. Dunwell JM. Transgenic approaches to crop improvement. J Exp Bot. 2000; 51 Spec No:487–496.

2. McCabe DE, Swain WF, Martinell BJ, Christou P. Stable Transformation of Soybean (Glycine-Max) by Particle-Acceleration. *Bio-Technology*. 1988;6(8):923–6.
3. Cho MJ, Yano H, Okamoto D, Kim HK, Jung HR, Newcomb K, Le VK, Yoo HS, Langham R, Buchanan BB, et al. Stable transformation of rice (*Oryza sativa* L.) via microprojectile bombardment of highly regenerative, green tissues derived from mature seed. *Plant Cell Rep*. 2004;22(7):483–9.
4. Gordon-Kamm WJ, Spencer TM, Mangano ML, Adams TR, Daines RJ, Start WG, O'Brien JV, Chambers SA, Adams WR Jr, Willetts NG, et al. Transformation of Maize Cells and Regeneration of Fertile Transgenic Plants. *Plant Cell*. 1990;2(7):603–18.
5. Altpeter F, Springer NM, Bartley LE, Blechl AE, Brutnell TP, Citovsky V, Conrad LJ, Gelvin SB, Jackson DP, Kausch AP, et al. Advancing Crop Transformation in the Era of Genome Editing. *Plant Cell*. 2016;28(7):1510–20.
6. Method of the Year. 2011. *Nat Methods*. 2012; 9(1):1.
7. Manghwar H, Lindsey K, Zhang X, Jin S. CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing. *Trends Plant Sci*. 2019;24(12):1102–25.
8. Bernardo R. Prospective Targeted Recombination and Genetic Gains for Quantitative Traits in Maize. *Plant Genome*. 2017; 10(2).
9. Leelavathi S, Sunnichan VG, Kumria R, Vijaykanth GP, Bhatnagar RK, Reddy VS. A simple and rapid *Agrobacterium*-mediated transformation protocol for cotton (*Gossypium hirsutum* L.): embryogenic calli as a source to generate large numbers of transgenic plants. *Plant Cell Rep*. 2004;22(7):465–70.
10. Nam J, Matthyse AG, Gelvin SB. Differences in susceptibility of *Arabidopsis* ecotypes to crown gall disease may result from a deficiency in T-DNA integration. *Plant Cell*. 1997;9(3):317–33.
11. Liu W, Parrott WA, Hildebrand DF, Collins GB, Williams EG. *Agrobacterium* induced gall formation in bell pepper (*Capsicum annuum* L.) and formation of shoot-like structures expressing introduced genes. *Plant Cell Rep*. 1990;9(7):360–4.
12. Lowe K, Wu E, Wang N, Hoerster G, Hastings C, Cho MJ, Scelonge C, Lenderts B, Chamberlin M, Cushatt J, et al. Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation. *Plant Cell*. 2016;28(9):1998–2015.
13. Mordhorst AP, Toonen MAJ, deVries SC. Plant embryogenesis. *Crit Rev Plant Sci*. 1997;16(6):535–76.
14. Leelavathi S, Sunnichan VG, Kumria R, Vijaykanth GP, Bhatnagar RK, Reddy VS. A simple and rapid *Agrobacterium*-mediated transformation protocol for cotton (*Gossypium hirsutum* L.): Embryogenic calli as a source to generate large numbers of transgenic plants. *Plant Cell Rep*. 2004;22(7):465–70.
15. Yang XY, Zhang XL. Regulation of Somatic Embryogenesis in Higher Plants. *Crit Rev Plant Sci*. 2010;29(1):36–57.
16. Carman JG. Embryogenic Cells in Plant-Tissue Cultures - Occurrence and Behavior. *In Vitro Cellular Developmental Biology*. 1990;26(8):746–53.
17. Schmidt EDL, Guzzo F, Toonen MAJ, deVries SC. A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development*. 1997;124(10):2049–62.
18. Ikeuchi M, Shibata M, Rymer B, Iwase A, Bagman AM, Watt L, Coleman D, Favero DS, Takahashi T, Ahnert SE, et al. A Gene Regulatory Network for Cellular Reprogramming in Plant Regeneration. *Plant Cell Physiol*. 2018;59(4):765–77.
19. Eapen S, George L. Influence of phytohormones, carbohydrates, aminoacids, growth supplements and antibiotics on somatic embryogenesis and plant differentiation in finger millet. *Plant Cell Tissue Organ Cult*. 1990;22(2):87–93.
20. Mashayekhi K, Neumann KH. Effects of boron on somatic embryogenesis of *Daucus carota*. *Plant Cell Tissue Organ Cult*. 2006;84(3):279–83.
21. Shekhawat GS, Mathur S, Batra A. Role of phytohormones and nitrogen in somatic embryogenesis induction in cell culture derived from leaflets of *Azadirachta indica*. *Biol Plant*. 2009;53(4):707.
22. Yu T-A, Yeh S-D, Yang J-S. Effects of carbenicillin and cefotaxime on callus growth and somatic embryogenesis from adventitious roots of papaya. *Botanical Bulletin - Academia Sinica Taipei*. 2001;42:281–6.
23. Komatsuda T, Kaneko K, Oka S. Genotype Sucrose Interactions for Somatic Embryogenesis in Soybean. *Crop Science*. 1991; 31(2):cropsci1991.0011183 × 003100020023x.
24. McKentley AH. Effect of genotype on somatic embryogenesis from axes of mature peanut embryos. *Plant Cell Tissue Organ Cult*. 1995;42(3):251–4.
25. Niskanen A-M, Lu J, Seitz S, Keinonen K, Von Weissenberg K, Pappinen A. Effect of parent genotype on somatic embryogenesis in Scots pine (*Pinus sylvestris*). *Tree Physiol*. 2004;24(11):1259–65.
26. Parrott WA, Williams EG, Hildebrand DF, Collins GB. Effect of genotype on somatic embryogenesis from immature cotyledons of soybean. *Plant Cell Tissue Organ Cult*. 1989;16(1):15–21.
27. Trolinder NL, Xhixian C. Genotype specificity of the somatic embryogenesis response in cotton. *Plant Cell Rep*. 1989;8(3):133–6.
28. Condit ML. Totipotency. What It Is and What It Isn't. *Stem Cells Development*. 2014;23(8):796–812.
29. Chu Z, Chen J, Sun J, Dong Z, Yang X, Wang Y, Xu H, Zhang X, Chen F, Cui D. De novo assembly and comparative analysis of the transcriptome of embryogenic callus formation in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol*. 2017;17(1):244.
30. Xu ZZ, Zhang CJ, Zhang XY, Liu CL, Wu ZX, Yang ZR, Zhou KH, Yang XJ, Li FG. Transcriptome Profiling Reveals Auxin and Cytokinin Regulating Somatic Embryogenesis in Different Sister Lines of Cotton Cultivar CCR124. *J Integr Plant Biol*. 2013;55(7):631–42.
31. Li JY, Wang MJ, Li YJ, Zhang QH, Lindsey K, Daniell H, Jin SX, Zhang XL. Multi-omics analyses reveal epigenomics basis for cotton somatic embryogenesis through successive regeneration acclimation process. *Plant Biotechnol J*. 2019;17(2):435–50.
32. Florentin A, Damri M, Grafi G. Stress Induces Plant Somatic Cells to Acquire Some Features of Stem Cells Accompanied by Selective Chromatin Reorganization. *Dev Dyn*. 2013;242(10):1121–33.
33. Ikeuchi M, Sugimoto K, Iwase A. Plant callus: mechanisms of induction and repression. *Plant Cell*. 2013;25(9):3159–73.
34. Grafi G, Barak S. Stress induces cell dedifferentiation in plants. *Biochim Biophys Acta*. 2015;1849(4):378–84.

35. Mendez-Hernandez HA, Ledezma-Rodriguez M, Avilez-Montalvo RN, Juarez-Gomez YL, Skeete A, Avilez-Montalvo J, De-la-Pena C, Loyola-Vargas VM. Signaling Overview of Plant Somatic Embryogenesis. *Front Plant Sci.* 2019;10:77.
36. 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.
37. Lotan T, Ohto M, Yee KM, West MAL, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell.* 1998;93(7):195–205.
38. Zuo J, Niu QW, Frugis G, Chua NH. The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant J.* 2002;30(3):349–59.
39. Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang LM, Hattori J, Liu CM, van Lammeren AAM, Miki BLA, et al. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell.* 2002;14(8):1737–49.
40. Srinivasan C, Liu Z, Heidmann I, Supena ED, Fukuoka H, Joosen R, Lambalk J, Angenent G, Scorza R, Custers JB, et al. Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (*Nicotiana tabacum* L.). *Planta.* 2007;225(2):341–51.
41. Heidmann I, de Lange B, Lambalk J, Angenent GC, Boutilier K. Efficient sweet pepper transformation mediated by the BABY BOOM transcription factor. *Plant Cell Rep.* 2011;30(6):1107–15.
42. Florez SL, Erwin RL, Maximova SN, Guiltinan MJ, Curtis WR. Enhanced somatic embryogenesis in *Theobroma cacao* using the homologous BABY BOOM transcription factor. *Bmc Plant Biology.* 2015; 15.
43. Mitten DH. Somatic embryogenesis in *Gossypium hirsutum* L. 1985; v. p. 57–58.
44. Davidonis GH, Hamilton RH. Plant-Regeneration from Callus-Tissue of *Gossypium-Hirsutum*-L. *Plant Science Letters.* 1983;32(1–2):89–93.
45. Trolinder NL, Goodin JR. Somatic Embryogenesis in Cotton (*Gossypium*).1. Effects of Source of Explant and Hormone Regime. *Plant Cell Tissue Organ Culture.* 1988;12(1):31–42.
46. Trolinder NL, Xhixian C. Genotype specificity of the somatic embryogenesis response in cotton. *Plant Cell Rep.* 1989;8(3):133–6.
47. Gawel NJ, Robacker CD. Genetic-Control of Somatic Embryogenesis in Cotton Petiole Callus-Cultures. *Euphytica.* 1990;49(3):249–53.
48. Lambe P, Mutambel HSN, Fouche JG, Deltour R, Foidart JM, Gaspar T. DNA methylation as a key process in regulation of organogenic totipotency and plant neoplastic progression? In *Vitro Cellular Developmental Biology-Plant.* 1997;33(3):155–62.
49. Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J, et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat Genet.* 2020.
50. Shin S, Torres-Acosta JA, Heinen SJ, McCormick S, Lemmens M, Paris MP, Berthiller F, Adam G, Muehlbauer GJ. Transgenic Arabidopsis thaliana expressing a barley UDP-glucosyltransferase exhibit resistance to the mycotoxin deoxynivalenol. *J Exp Bot.* 2012;63(13):4731–40.
51. Wang MY, Zhao PM, Cheng HQ, Han LB, Wu XM, Gao P, Wang HY, Yang CL, Zhong NQ, Zuo JR, et al. The cotton transcription factor TCP14 functions in auxin-mediated epidermal cell differentiation and elongation. *Plant Physiol.* 2013;162(3):1669–80.
52. Kim HU, Jung SJ, Lee KR, Kim EH, Lee SM, Roh KH, Kim JB. Ectopic overexpression of castor bean LEAFY COTYLEDON2 (LEC2) in Arabidopsis triggers the expression of genes that encode regulators of seed maturation and oil body proteins in vegetative tissues. *Febs Open Bio.* 2014;4:25–32.
53. Bailey PC, Martin C, Toledo-Ortiz G, Quail PH, Huq E, Heim MA, Jakoby M, Werber M, Weisshaar B. Update on the basic helix-loop-helix transcription factor gene family in Arabidopsis thaliana. *Plant Cell.* 2003;15(11):2497–501.
54. Horstman A, Li M, Heidmann I, Weemen M, Chen B, Muino JM, Angenent GC, Boutilier K. The BABY BOOM Transcription Factor Activates the LEC1-ABI3-FUS3-LEC2 Network to Induce Somatic Embryogenesis. *Plant Physiol.* 2017;175(2):848–57.
55. Li J, Tao X, Li L, Mao L, Luo Z, Khan ZU, Ying T. Comprehensive RNA-Seq Analysis on the Regulation of Tomato Ripening by Exogenous Auxin. *PLoS One.* 2016;11(5):e0156453.
56. Pooam M, Arthaut LD, Burdick D, Link J, Martino CF, Ahmad M. Magnetic sensitivity mediated by the Arabidopsis blue-light receptor cryptochrome occurs during flavin reoxidation in the dark. *Planta.* 2018.
57. Ho-Yue-Kuang S, Alvarado C, Antelme S, Bouchet B, Cezard L, Le Bris P, Legee F, Maia-Grondard A, Yoshinaga A, Saulnier L, et al. Mutation in *Brachypodium caffer* caffeic acid O-methyltransferase 6 alters stem and grain lignins and improves straw saccharification without deteriorating grain quality. *J Exp Bot.* 2016;67(1):227–37.
58. Gigli-Bisceglia N, Engelsdorf T, Strnad M, Vaahtera L, Khan GA, Yamoune A, Alipanah L, Novak O, Persson S, Hejatko J, et al. Cell wall integrity modulates Arabidopsis thaliana cell cycle gene expression in a cytokinin- and nitrate reductase-dependent manner. *Development.* 2018; 145(19).
59. Stuurman J, Jaggi F, Kuhlemeier C. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. *Genes Dev.* 2002;16(17):2213–8.
60. Aguilar-Martinez JA, Sinha N. Analysis of the role of Arabidopsis class I TCP genes AtTCP7, AtTCP8, AtTCP22, and AtTCP23 in leaf development. *Front Plant Sci.* 2013;4:406.
61. Han X, Kumar D, Chen H, Wu S, Kim JY. Transcription factor-mediated cell-to-cell signalling in plants. *J Exp Bot.* 2014;65(7):1737–49.
62. Mouchel CF, Briggs GC, Hardtke CS. Natural genetic variation in Arabidopsis identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. *Genes Dev.* 2004;18(6):700–14.
63. Li J, Mo X, Wang J, Chen N, Fan H, Dai C, Wu P. BREVIS RADIX is involved in cytokinin-mediated inhibition of lateral root initiation in Arabidopsis. *Planta.* 2009;229(3):593–603.
64. Zhang Z, Zou X, Huang Z, Fan S, Qun G, Liu A, Gong J, Li J, Gong W, Shi Y, et al. Genome-wide identification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of *Gossypium* genus. *Gene.* 2019;680:72–83.

65. Behringer C, Schwechheimer C. B-GATA transcription factors - insights into their structure, regulation, and role in plant development. *Front Plant Sci.* 2015;6:90.
66. Nawy T, Bayer M, Mravec J, Friml J, Birnbaum KD, Lukowitz W. The GATA factor HANABA TARANU is required to position the proembryo boundary in the early Arabidopsis embryo. *Dev Cell.* 2010;19(1):103–13.
67. Grimplet J, Agudelo-Romero P, Teixeira RT, Martinez-Zapater JM, Fortes AM. Structural and Functional Analysis of the GRAS Gene Family in Grapevine Indicates a Role of GRAS Proteins in the Control of Development and Stress Responses. *Front Plant Sci.* 2016;7:353.
68. Gallagher KL, Benfey PN. Both the conserved GRAS domain and nuclear localization are required for SHORT-ROOT movement. *Plant J.* 2009;57(5):785–97.
69. D'Alessandro S, Ksas B, Havaux M. Decoding beta-Cyclocitral-Mediated Retrograde Signaling Reveals the Role of a Detoxification Response in Plant Tolerance to Photooxidative Stress. *Plant Cell.* 2018;30(10):2495–511.
70. Zhou S, Hu Z, Li F, Yu X, Naeem M, Zhang Y, Chen G. Manipulation of plant architecture and flowering time by down-regulation of the GRAS transcription factor SGRAS26 in *Solanum lycopersicum*. *Plant Sci.* 2018;271:81–93.
71. Qiao L, Zhang W, Li X, Zhang L, Zhang X, Li X, Guo H, Ren Y, Zheng J, Chang Z. Characterization and Expression Patterns of Auxin Response Factors in Wheat. *Front Plant Sci.* 2018;9:1395.
72. Hakoshima T. Structural basis of the specific interactions of GRAS family proteins. *FEBS Lett.* 2018;592(4):489–501.
73. Richards S, Wink RH, Simon R. Mathematical modelling of WOX5- and CLE40-mediated columella stem cell homeostasis in Arabidopsis. *J Exp Bot.* 2015;66(17):5375–84.
74. Shinohara H, Matsubayashi Y. Reevaluation of the CLV3-receptor interaction in the shoot apical meristem: dissection of the CLV3 signaling pathway from a direct ligand-binding point of view. *Plant J.* 2015;82(2):328–36.
75. Guan Y, Meng X, Khanna R, LaMontagne E, Liu Y, Zhang S. Phosphorylation of a WRKY transcription factor by MAPKs is required for pollen development and function in Arabidopsis. *PLoS Genet.* 2014;10(5):e1004384.
76. Jeong S, Eilbert E, Bolbol A, Lukowitz W. Going mainstream: How is the body axis of plants first initiated in the embryo? *Dev Biol.* 2016;419(1):78–84.
77. Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AA, Miki BL, et al. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell.* 2002;14(8):1737–49.
78. Roscoe TJ, Vaissayre V, Paszkiewicz G, Clavijo F, Kelemen Z, Michaud C, Lepiniec LC, Dubreucq B, Zhou DX, Devic M. Regulation of FUSCA3 Expression During Seed Development in Arabidopsis. *Plant Cell Physiol.* 2019;60(2):476–87.
79. Yotsui I, Saruhashi M, Kawato T, Taji T, Hayashi T, Quatrano RS, Sakata Y. ABSCISIC ACID INSENSITIVE3 regulates abscisic acid-responsive gene expression with the nuclear factor Y complex through the ACTT-core element in *Physcomitrella patens*. *New Phytol.* 2013;199(1):101–9.
80. Perry SE, Zheng Q, Zheng Y. Transcriptome analysis indicates that GmAGAMOUS-Like 15 may enhance somatic embryogenesis by promoting a dedifferentiated state. *Plant Signal Behav.* 2016;11(7):e1197463.
81. Lowe K, La Rota M, Hoerster G, Hastings C, Wang N, Chamberlin M, Wu E, Jones T, Gordon-Kamm W. Rapid genotype "independent" *Zea mays* L. (maize) transformation via direct somatic embryogenesis. *In Vitro Cell Dev Biol Plant.* 2018;54(3):240–52.
82. Sarkar AK, 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(7137):811–4.
83. Singla B, Tyagi AK, Khurana JP, Khurana P. Analysis of expression profile of selected genes expressed during auxin-induced somatic embryogenesis in leaf base system of wheat (*Triticum aestivum*) and their possible interactions. *Plant Mol Biol.* 2007;65(5):677–92.
84. Xu C, Cao H, Zhang Q, Wang H, Xin W, Xu E, Zhang S, Yu R, Yu D, Hu Y. Control of auxin-induced callus formation by bZIP59-LBD complex in Arabidopsis regeneration. *Nat Plants.* 2018;4(2):108–15.
85. Kukurshumova W, Berleth T. Overcoming recalcitrance - Auxin response factor functions in plant regeneration. *Plant Signal Behav.* 2015;10(7):e993293.
86. Hill K, Schaller GE. Enhancing plant regeneration in tissue culture: a molecular approach through manipulation of cytokinin sensitivity. *Plant Signal Behav.* 2013; 8(10):doi: 10.4161/psb.25709.
87. Zhang Z, Tucker E, Hermann M, Laux T. A Molecular Framework for the Embryonic Initiation of Shoot Meristem Stem Cells. *Dev Cell.* 2017;40(3):264–77. e264.
88. Cheng ZJ, Zhu SS, Gao XQ, Zhang XS. Cytokinin and auxin regulates WUS induction and inflorescence regeneration in vitro in Arabidopsis. *Plant Cell Rep.* 2010;29(8):927–33.
89. Li J, Wang M, Li Y, Zhang Q, Lindsey K, Daniell H, Jin S, Zhang X. Multi-omics analyses reveal epigenomics basis for cotton somatic embryogenesis through successive regeneration acclimation process. *Plant Biotechnol J.* 2018.
90. Jin SX, Zhang XL, Liang SG, Nie YC, Guo XP, Huang C. Factors affecting transformation efficiency of embryogenic callus of Upland cotton (*Gossypium hirsutum*) with *Agrobacterium tumefaciens*. *Plant Cell Tissue Organ Culture.* 2005;81(2):229–37.
91. Suzuki Y, Mae T, Makino A. RNA extraction from various recalcitrant plant tissues with a cetyltrimethylammonium bromide-containing buffer followed by an acid guanidium thiocyanate-phenol-chloroform treatment. *Bioscience Biotechnology Biochemistry.* 2008;72(7):1951–3.
92. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
93. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.

94. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
95. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
96. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
97. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*. 1995;57(1):289–300.
98. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology*. 2010;11(2):R14.
99. Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biol*. 2010;10:49.
100. Liu W, Wang C, Shen X, Liang H, Wang Y, He Z, Zhang D, Chen F. Comparative transcriptome analysis highlights the hormone effects on somatic embryogenesis in *Catalpa bungei*. *Plant Reprod*. 2019;32(2):141–51.

Figures

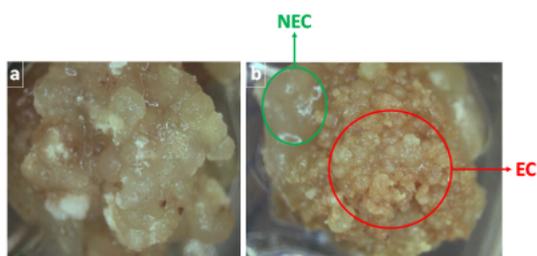


Figure 1

(a) 4% of the explants stay on NEC stage; (b) 96% of the explants can be induced to form EC, surrounded by NEC. The examples of NEC (green circle) and EC (red circle) and the section of the callus used for RNA extraction and sequencing were circled.

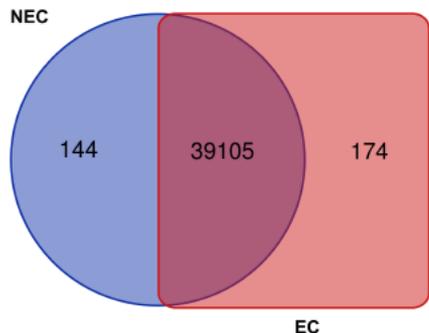


Figure 2

Genes expressed in NEC and EC callus

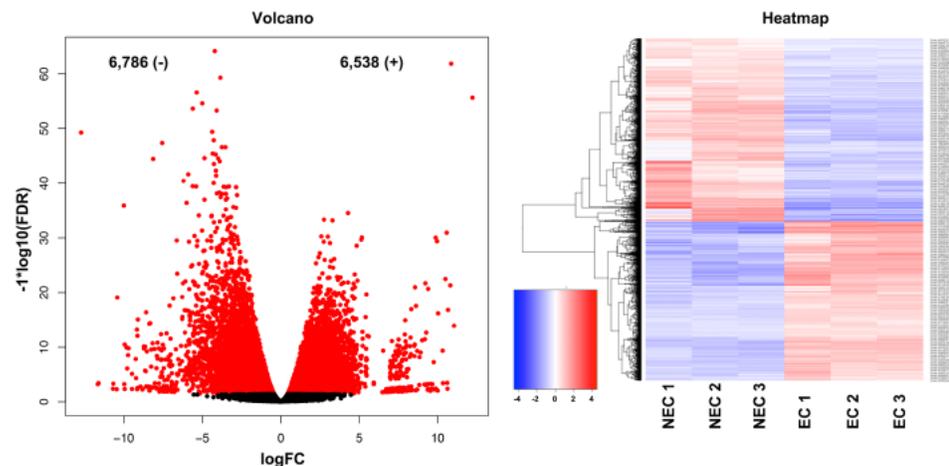


Figure 3

A. Volcano plot of significant genes with FDR corrected p-values ($\leq .001$). B. Heatmap of gene expression profiles of NEC and EC callus.

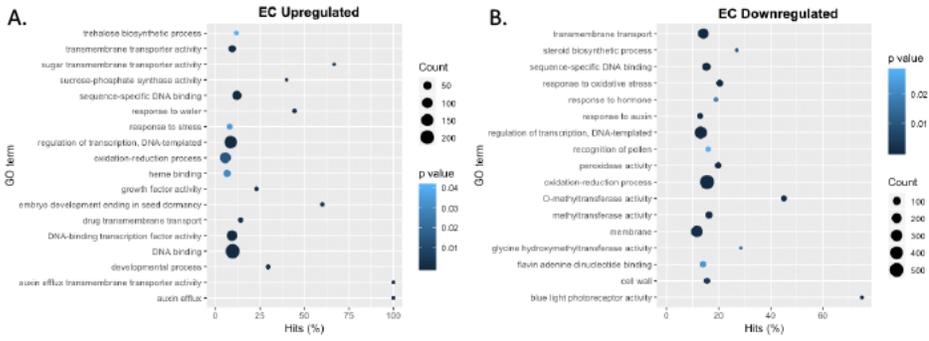


Figure 4

Gene ontology enrichment of A. Significant upregulated genes (6,786) and B. Significantly downregulated genes (6,538).

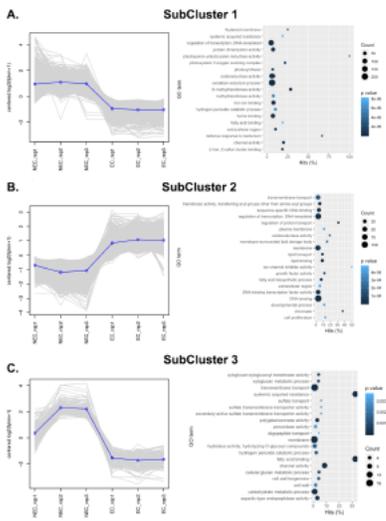


Figure 5

Subcluster gene expression profiles determined by hierarchical clustering and cutting the tree at 50%.

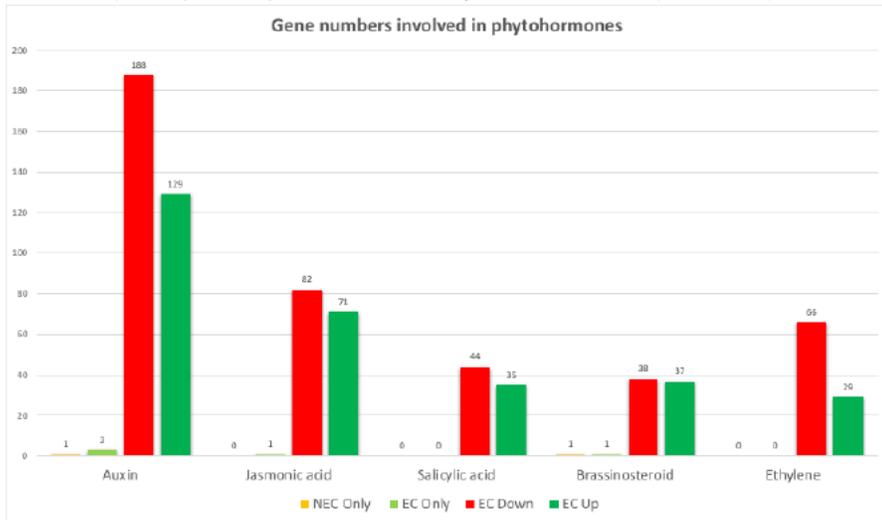


Figure 6

Genes involved in phytohormones

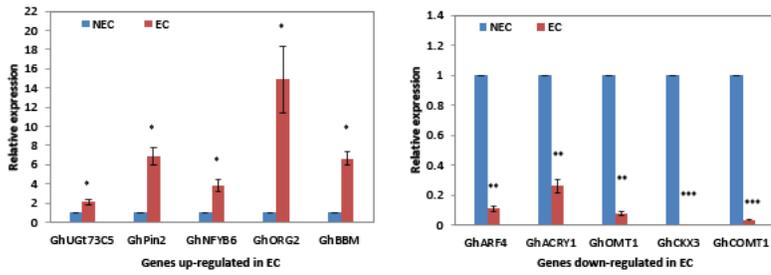


Figure 7

Evaluation of gene expression in embryogenic cells (EC) and non-embryogenic cells (NEC) by RT-qPCR analysis. Five of the selected genes were up-regulated (left); and the other five were down-regulated in embryogenic cells (right). GhPP2A1 and GhUb7 were used as the internal controls. Three biological replicates and three technical replicates were used for statistical analysis. Error bars indicate \pm SE (n=3). $\Delta\Delta$ Ct method was used for qPCR analysis. Asterisks (*, ** or ***) indicate a significant difference between embryogenic cells (EC) and non-embryogenic cells (NEC) at $P < 0.05$, 0.01, or 0.001, respectively, by student's t-test.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS6.xlsx](#)
- [TableS8.xlsx](#)
- [TableS7.xlsx](#)
- [TableS5.xlsx](#)
- [TableS9.xlsx](#)
- [TableS1.xlsx](#)
- [TableS10.xlsx](#)
- [TableS11.xlsx](#)
- [TableS12.xlsx](#)