the plant journal

SEB Society for Experimental Biology

The Plant Journal (2017) **91,** 1088–1107 doi: 10.1111/tpj.13621

RESOURCE

Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea

Rohini Garg[1,2], Vikash K. Singh[1], Mohan Singh Rajkumar[1], Vinay Kumar[1,†] and Mukesh Jain[1,3,*]

[1]*National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi 110067, India,*

[2]*Department of Life Sciences, School of Natural Sciences, Shiv Nadar University, Gautam Buddha Nagar, Uttar Pradesh 201314, India, and*

[3]*School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India*

Received 15 February 2017; revised 2 June 2017; accepted 9 June 2017; published online 22 June 2017.

*For correspondence (e-mail mjain@jnu.ac.in).

†Present address: School of Basic and Applied Sciences, Central University of Punjab, Bathinda, 151001, India.

SUMMARY

Seed development is an intricate process regulated via a complex transcriptional regulatory network. To understand the molecular mechanisms governing seed development and seed size/weight in chickpea, we performed a comprehensive analysis of transcriptome dynamics during seed development in two cultivars with contrasting seed size/weight (small-seeded, Himchana 1 and large-seeded, JGK 3). Our analysis identified stage-specific expression for a significant proportion (>13%) of the genes in each cultivar. About one half of the total genes exhibited significant differential expression in JGK 3 as compared with Himchana 1. We found that different seed development stages can be delineated by modules of coexpressed genes. A comparative analysis revealed differential developmental stage specificity of some modules between the two cultivars. Furthermore, we constructed transcriptional regulatory networks and identified key components determining seed size/weight. The results suggested that extended period of cell division during embryogenesis and higher level of endoreduplication along with more accumulation of storage compounds during maturation determine large seed size/weight. Further, we identified quantitative trait loci-associated candidate genes harboring single nucleotide polymorphisms in the promoter sequences that differentiate small- and large-seeded chickpea cultivars. The results provide a valuable resource to dissect the role of candidate genes governing seed development and seed size/weight in chickpea.

Keywords: *Cicer arietinum*, coexpression network, gene expression, GSE79719, GSE79720, seed development, seed size/weight, transcriptome dynamics, transcriptional modules.

INTRODUCTION

Seeds, besides being the progenitor of the next generation, provide human and animal nutrition worldwide. Legume seeds are rich source of proteins, carbohydrates and oils in the human diet. Seed size is an important agronomic trait related to yield potential and is an important trait in legumes. In legumes, seed development initiates with embryogenesis (cell division) after double fertilization followed by seed maturation (seed filling via accumulation of storage macromolecules) and desiccation. In dicots, cotyledons represent the major nutrient storage organ of the seed, as the endosperm is almost totally resorbed at the initiation of seed maturation. A comprehensive understanding of molecular mechanisms regulating various aspects of seed development is required to facilitate the development of new varieties.

Seed development is a complex trait controlled by multiple biological processes/pathways. High-resolution transcriptome studies in model/crop plants have provided

some molecular insights into pathways and networks along with their interactions that are involved in several aspects of seed development (Le *et al.*, 2010; Jones *et al.*, 2010; Belmonte *et al.*, 2013; Sreenivasulu and Wobus, 2013; Becker *et al.*, 2014). The molecular mechanism underlying seed size has also been elucidated to some extent in a few plants (Li and Li, 2016). It has been demonstrated that carbohydrate partitioning, supply of photoassimilates and rate of accumulation of storage compounds during early stages of seed development are very important in determining seed size/weight (Borras *et al.*, 2004, 2009; Weber *et al.*, 2005; Gambin and Borras, 2010; Zhang *et al.*, 2013). In addition, hormonal signal transduction and epigenetic mechanisms/imprinting have also been characterized as key regulatory mechanisms that determine seed size (Schruff *et al.*, 2006; Xiao *et al.*, 2006; Sun *et al.*, 2010; Jiang *et al.*, 2013; Locascio *et al.*, 2014). A positive correlation between cell number in cotyledon and mature seed size has been reported in legumes, such as pea and soybean (Egli *et al.*, 1981; Lemontey *et al.*, 2000).

Chickpea is a nutritionally and agriculturally important legume crop plant. The accessibility to transcriptome and draft genome sequences of chickpea (Garg *et al.*, 2011; Jain *et al.*, 2013; Varshney *et al.*, 2013) along with next-generation RNA sequencing (RNA-seq) technology provide an opportunity to reveal genetic diversity among various genotypes/cultivars, specifically those with important agronomic traits. A few transcriptome studies have been performed to understand the flower development and abiotic stress responses (Singh *et al.*, 2013; Kudapa *et al.*, 2014; Garg *et al.*, 2016). However, no such analysis has been performed to dissect the molecular mechanisms underlying seed development and/or seed size/weight trait in chickpea. Seed size is an important end-user quality parameter of chickpea. Large seed size of chickpea is highly desirable to fetch higher market price. Although huge variations in seed size have been observed among chickpea genotypes, this phenotypic variability could not be harnessed to improve seed size in important chickpea cultivars due to poor understanding of molecular mechanisms underlying this important trait. Recently, a few quantitative trait loci (QTLs) and candidate genes involved in regulation of seed size have been identified via marker-based, QTL mapping or association analysis studies (Kujur *et al.*, 2013, 2015; Saxena *et al.*, 2014; Bajaj *et al.*, 2015, 2016; Verma *et al.*, 2015; Singh *et al.*, 2016).

The availability of genotypes/cultivars with contrasting phenotypes for a specific trait offers an excellent opportunity to reveal the genetic factors controlling that trait. A comparative transcriptome analysis of the genotypes/cultivars with different seed size has not yet been performed in chickpea to the best of our knowledge. Here, we have used RNA-seq technology to analyze the transcriptomes of seeds of two chickpea cultivars (differing significantly in their seed size) at different stages of development. We dissected these data to reveal transcriptome dynamics and transcriptional network associated with seed development, and identified key differences that determine seed size/weight in chickpea. The transcripts and/or modules of coexpressed genes expressed predominantly/specifically at different stages of seed development and/or cultivars, were identified. The overlap of known QTLs with differential gene expression and discovery of single nucleotide polymorphisms (SNPs) identified candidate genes that might determine seed size/weight. This study provides insights into the molecular mechanisms underlying seed development and the factors determining seed size/weight in chickpea.

RESULTS

Global transcriptome analysis of whole seed in chickpea cultivars

The transcriptome analysis of different stages of seed development in the chickpea cultivars differing in their seed size/weight (Figures 1 and S1) can provide crucial systems-level insights into molecular mechanisms underlying seed development and seed size/weight. We selected two chickpea cultivars, Himchana 1 (small-seeded; average 100 seed weight 13.1 $\pm$ 0.15 g) and JGK 3 (large-seeded; average 100 seed weight 53.3 $\pm$ 1.48 g), which differ significantly in their seed size/weight. Seven stages of seed development representing major events occurring within the seed, including embryo development, storage reserve accumulation and maturation, were analyzed. Different stages of seed development were designated as S1–S7 (Figure 1). The stages S1–S3 represented stages of embryo development (embryogenesis), S4 and S5 stages corresponded to early and mid-maturation stages (grain filling), and S6 and S7 stages represented the late maturation stages (seed desiccation). The seed weight increased from S1 (average of 3.3 mg in Himchana 1 and 4.4 mg in JGK 3) to S5 (average of 236.2 mg in Himchana 1 and 1066.5 mg in JGK 3) stages and decreased thereafter (Figure 1b). Maximum weight gain was observed between S4 and S5 stages in both the cultivars. The weight increased 2.1 times in Himchana 1, whereas 4.7 times weight gain was observed in JGK 3 from S4 to S5 stage. The weight of seed corresponded well with the seed size, which also increased from S1 to S5 stages and decreased thereafter in both the cultivars (Figure 1c). However, there was a minor (11.3%) increase in seed size of Himchana 1 from S4 to S5, whereas it was 41.4% increase for JGK 3.

To investigate the transcriptome dynamics during seed development, we performed RNA-seq experiments using total RNA isolated from the seven stages of seed development and leaves of mature plants from the two chickpea cultivars, Himchana 1 and JGK 3 (eight tissues from each

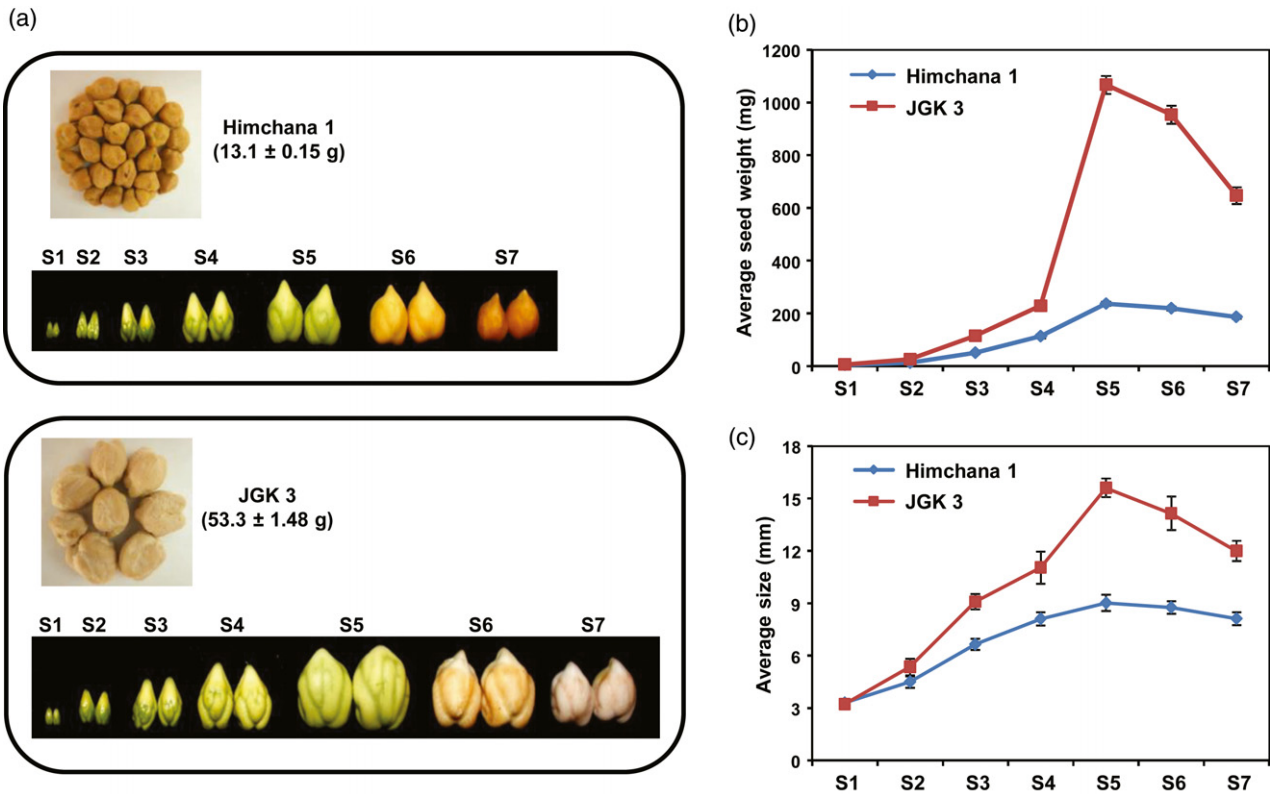


Figure 1. Seed phenotype at different stages of development in the chickpea cultivars.
(a) Phenotype of seed at different stages of development (S1–S7) and dry seeds in Himchana 1 and JGK 3. Average 100 seed weight (g) along with standard error (n = 30) has also been given.
(b,c) Physical measurements showing variation in seed weight (b) and seed size (c) of Himchana 1 and JGK 3 at different stages (S1–S7). An average weight (in mg) and size (in mm) of a single seed has been presented. Error bars indicate standard error (n = 40–60).

cultivar). All the tissues were analyzed in three independent biological replicates (48 samples in total). A total of more than 1.5 billion high-quality reads (average ~63 million reads from each sample) were generated for each of Himchana 1 and JGK 3 cultivars from different tissues (Table S1), and mapped to the chickpea genome (kabuli, v1.0) using TopHat. The mapped files were processed via Cufflinks and Cuffmerge, which generated a consensus transcriptome assembly with a total of 35 234 gene loci, including 28 269 known and 6965 novel gene loci. The uniquely mapped reads (28–68 million) for each sample (Table S1) were processed using Cufflinks to determine the normalized expression level as fragments per kilobase of transcript length per million mapped reads (FPKM) of each transcript. Spearman correlation coefficient (SCC) between the biological replicates of different tissues varied from 0.94 to 0.99 (except for one replicate of S5 stage of Himchana 1, which was removed from further analyses), indicating the high quality of the replicates (Figure S2).

Overall, a total of ~86% genes were identified as expressed in at least one of the 16 samples. The number of expressed genes in different tissue samples varied from 56.4% (S7) to 66.2% (S4) in Himchana 1, and 59.4% (S7) to 67.6% (S3) in JGK 3 (Figure S3a). About 12–22% of genes exhibited very high (FPKM ≥50) expression level in different tissues analyzed (Figure S3b). The number of genes showing high (10 ≤ FPKM ≥ 50), moderate (2 ≤ FPKM ≥ 10) and low (0.1 ≤ FPKM ≥ 2) expression was similar in all the tissues. Interestingly, the largest number of genes showed very high expression at S7 stage in both the cultivars. In general, a slightly larger fraction of genes showed high/very high expression in JGK 3 as compared with Himchana 1 (Figure S3b). Overall, these analyses showed sufficient coverage of the transcriptome during seed development in the chickpea cultivars.

Global comparison of transcriptomes revealed the relationship among seed stages

To investigate the global differences in the transcriptome dynamics during seed development in Himchana 1 and JGK 3 cultivars, we performed hierarchical clustering and principal component analysis (PCA) based on SCC analysis of average FPKM values for all the expressed genes in at least one of the 16 tissue samples (Figure 2). The tissues/stages showing higher correlation in these analyses are expected to have more similar transcriptomes and

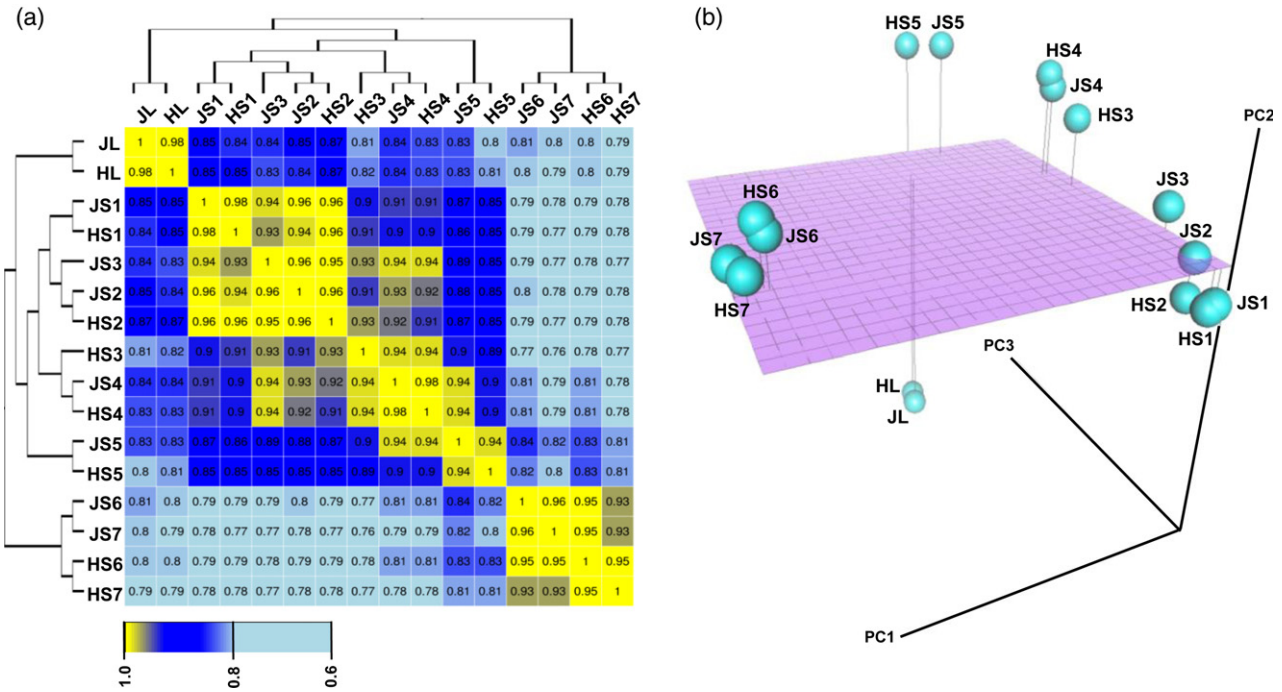


Figure 2. Correlation between the transcriptomes of different stages of seed development in the two chickpea cultivars.
(a) Spearman correlation coefficient (SCC) analysis of RNA-seq data from leaf and seven stages of seed development in Himchana 1 and JGK 3.
(b) Principal component analysis (PCA) plot showing clustering of transcriptomes of leaf and different stages of seed development in Himchana 1 (HL, HS1–HS7) and JGK 3 (JL, JS1–JS7).

functions/activities. These analyses showed a higher correlation of similar tissue/developmental stage between the two cultivars. As expected, leaf transcriptome of both the genotypes clustered together and showed substantial differences with seed development stages (Figure 2a and b). The two early stages of seed development (S1 and S2) showed closer correlation within and between the cultivar (s), indicating high similarity in their transcriptional programs. Interestingly, the clustering of S3 stage from JGK 3 and Himchana 1 was strikingly different; S3 stage of JGK 3 exhibited closer correlation with S2 stage, whereas S3 stage of Himchana 1 tended towards the later (S4) stage of seed development (Figure 2b). This observation indicates that Himchana 1 showed faster progression during the early stage of seed development than that of JGK 3, which is consistent with an earlier study on two maize populations with different seed sizes (Sekhon *et al.*, 2014). Seeds of both cultivars at S4 stage showed very tight clustering, indicating higher similarity in their transcriptional programs. However, S5 stage of JGK 3 and Himchana 1, though grouped closer, showed somewhat distinct transcriptional activity. Further, S6 and S7 stages from the two cultivars were grouped together, indicating a higher degree of similarity in their transcriptional programs. Taken together, these results suggested major differences in the transcriptional programs of early (S1–S3), mid (S4 and S5) and late (S6 and S7) stages of seed development within each cultivar. Further, the difference in the transcriptional programs at S3 and S5 stages may determine developmental specificities and/or seed size/weight of the two chickpea cultivars.

Differential gene expression during seed development

To investigate the transcriptional differences that characterize different stages of seed development in both the cultivars, we identified preferentially/specifically expressed genes in each stage of seed development in both the cultivars. We used the stage specificity (SS) algorithm with SS score ≥0.5 to identify the genes expressed at a particular stage of development for each cultivar. Based on this criterion, we identified a total of 5201 and 4664 genes in Himchana 1 and JGK 3, respectively, with specific expression at a particular stage of seed development. A significant fraction (6.1% in Himchana 1 and 5.6% in JGK 3) of stage-specific genes encoded for transcription factors (TFs). Although a high similarity in the overall transcriptome was detected among different stages of seed development within and across the cultivars, the number of stage-specific genes was significantly different among various stages. The number of stage-specific genes varied from 353 to 1257 for Himchana 1 and 456 to 1071 for JGK 3 (Figure 3a). The S2 stage showed the lowest number (353) of stage-specific genes in Himchana 1, whereas stage-specific genes were lowest at S6 stage (456) in JGK 3. The largest

numbers of stage-specific genes (1257 in Himchana 1 and 1071 in JGK 3) were identified at S5 stage in both the cultivars (Figure 3a). A significant proportion of stage-specific genes was common in both the cultivars (Figure 3a). A heatmap depicting the stage-specific expression of the genes in chickpea cultivars is shown in Figure 3b. The variable number and proportion of preferentially/specifically expressed genes suggest that each stage has their own independent developmental programs. Alternatively, the transcriptional complexity may simply reflect the intricacy of the captured seed stages, which contain more than one cell type.

The gene ontology (GO) enrichment analyses of all the stage-specific genes in Himchana 1 and JGK 3 showed representation of genes related to various reproductive processes, cell wall organization, cell cycle and cell division, carbohydrate metabolic processes, response to stress/hormone and regulation of transcription (Figure S4). These processes are well known to be involved during various aspects of seed development. At least 457 TF-encoding genes belonging to 67 families exhibited stage-specific expression in at least one of the cultivars. The members of AP2-EREBP, MYB, bHLH, CCAAT and HB TF families were highly represented in these cultivars. In general, the enrichment analysis of functional classes during seven stages showed that early stages (S1–S3) of seed development were marked by cell cycle and growth. Mid-stages of seed development were marked by cell wall, lipid metabolism, secondary metabolism and protein synthesis. During late stages, biological processes, including abiotic stress, transcription and protein folding, were overrepresented. Further, an overlap analysis of significantly enriched GO terms in each cultivar via generation of enrichment maps revealed several terms either specifically enriched in one

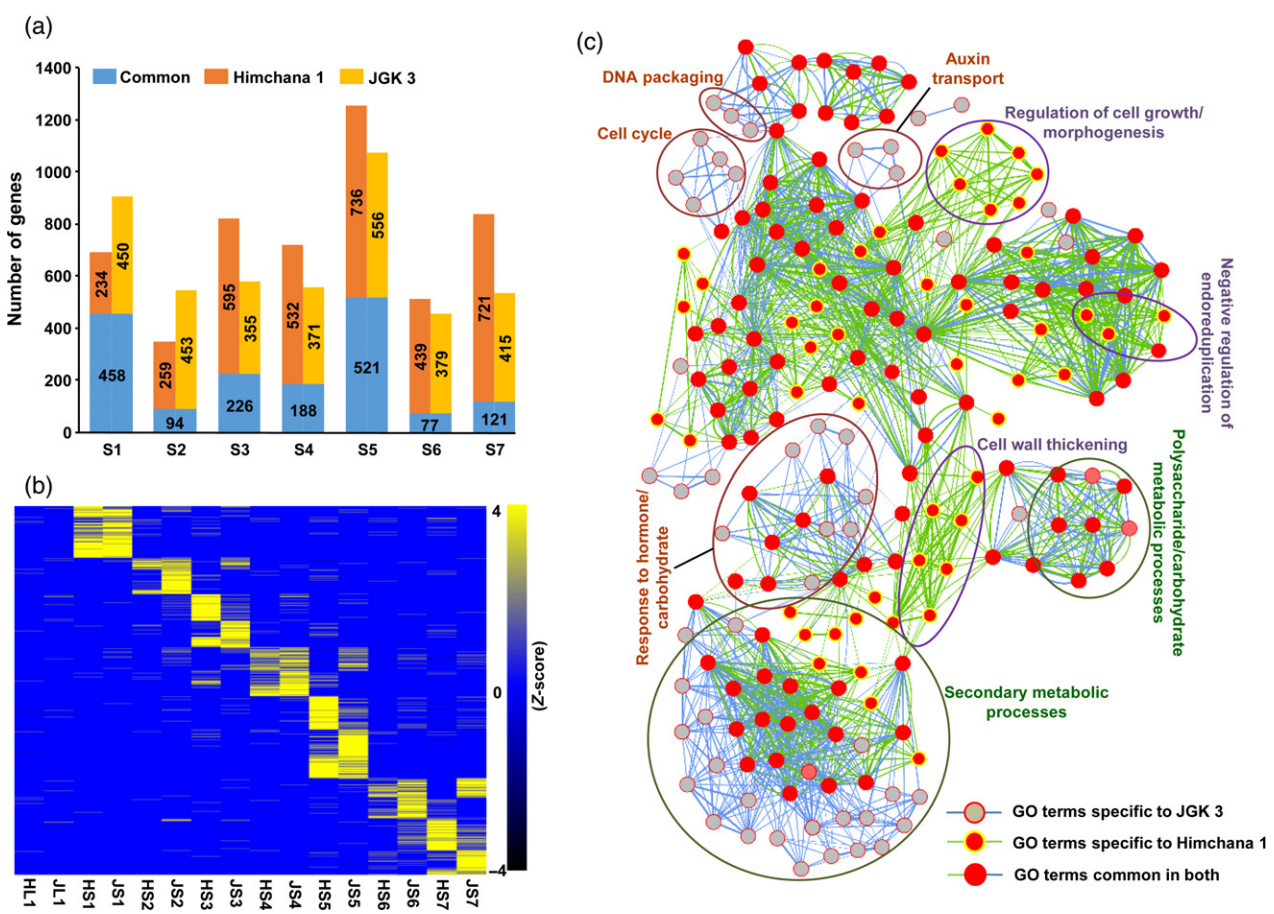


Figure 3. Preferential/stage-specific expression of genes during seed development stages in chickpea cultivars.
(a) Bar graph showing the number of preferentially expressed genes specifically and commonly in Himchana 1 and JGK 3 at each stage of seed development.
(b) Heatmap showing the expression profile of preferentially expressed genes in different tissues/stages in both chickpea cultivars. Color scale represents Z-score.
(c) Gene ontology (GO) enrichment map of preferentially expressed genes at all the stages of seed development in Himchana 1 and/or JGK 3. Significantly enriched GO terms (biological process) in Himchana 1 and/or JGK 3 were overlapped to highlight the terms specifically or commonly in Himchana 1 and/or JGK 3. Selected broad GO terms significantly enriched specifically or commonly in Himchana 1 and/or JGK 3 have been highlighted in colored ovals/circles (brown, JGK 3; purple, Himchana 1; green, common). An enlarged version with GO terms labeled is provided in Figure S4.

cultivar or commonly enriched in both the cultivars (Figures 3c and S5). For example, biological process GO terms related to cell cycle, DNA packaging and response to hormones were specific to JGK 3, whereas GO terms related to regulation of cell growth/morphogenesis, cell wall thickening and negative regulation of endoreduplication were specific to Himchana 1. GO terms related to polysaccharide/carbohydrate response pathways and secondary metabolic processes were found enriched in both the cultivars. Taken together, these results indicate that a set of genes, including those encoding for TFs, perform stage-specific functions during seed development in chickpea.

We performed reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analyses for 18 genes showing stage-specific expression in all the 16 tissue samples in Himchana 1 and/or JGK 3. The expression profiles and stage-specific expression of the tested genes revealed by RT-qPCR were similar to those observed in RNA-seq data (Figure S6), indicating accuracy of RNA-seq data to reflect the abundance of transcript levels. The correlation coefficient between the RNA-seq and real-time PCR-based analyses was $\geq$0.70 for most of the genes tested (Figure S6).

Gene sets differentially expressed between the chickpea cultivars

We identified gene sets showing significant differential expression between Himchana 1 and JGK 3 at each stage of seed development. In total, 8562 genes (including 497 TF-encoding genes) exhibited significant higher expression, and 9023 genes (including 474 TF-encoding genes) exhibited significant lower expression at different stages of seed development in JGK 3 as compared with Himchana 1. The largest number of genes (4783) exhibited differential expression at S7 stage followed by S3 stage (3893) between the two cultivars (Figure 4a). The least number of genes was differentially expressed at the S4 stage between the two cultivars. Overall, most of the TF families showed differential expression in JGK 3 and have diverse functions during seed development (Figure 4b; Table S2). Some of the TF families exhibited a significant difference in the number of members exhibiting significantly higher and lower expression in JGK 3 (Figure 4b). For example, the members of TF families involved in cell differentiation, such as ARF and HB, as well as those implicated in hormone signaling pathways, such as ARF and Aux/IAA (auxin signaling), and ARR-B (cytokinin signaling), and epigenetic regulation (SET) showed significantly higher expression in JGK 3. On the other hand, the majority of the TFs showing lower expression in JGK 3 were represented from the families, such as WRKY, Alfin-like, CCAAT, Myb-related, C_2C_2-CO and heat shock factors. The expression profiles of selected TF families during seed development in both the cultivars are given in Figure S7.

The GO enrichment analysis of differentially expressed genes in JGK 3 identified several biological processes uniquely/commonly overrepresented at different stages of seed development. Various cell division-related terms, such as cell cycle, cell division, cell growth, as well as cell wall organization/modification-related terms, were significantly enriched in the genes with higher expression, particularly at S3 stage (Figure 4c). Likewise, GO terms, regulation of cell size and cellular component size were also highly enriched at S3 stage. These terms were enriched at S5 stage in both sets of genes exhibiting higher and lower expression. The G1/S phase of mitotic cell cycle-related genes exhibited higher expression at S5 stage, which indicated higher activity of DNA replication machinery in JGK 3. Interestingly, we observed more significant enrichment of GO term, regulation of DNA endoreduplication in the genes showing higher expression at the mature stages (JS5 and JS6) of seed development (Figure 4c). Brassinosteroid biosynthesis and carbohydrate metabolism-related genes seemed to be more active at S3 stage.

To investigate the metabolic pathways responsible for the difference in seed size/weight of JGK 3 and Himchana 1, the expression profiles of differentially expressed genes between the cultivars were overlaid onto the available metabolic pathways using MapMan tool. We observed differential activity of certain metabolic pathways at S3 stage, which was consistent with PCA results. Substantial differences in the transcriptional activity of the genes involved in starch biosynthesis were observed between the cultivars. The genes involved in starch metabolism and photosynthesis-related genes were more active in JGK 3, suggesting higher energy requirements for actively dividing cells in large-seeded JGK 3 (Figure 4d). Consistently, the genes involved in cell cycle and cell division also showed higher expression in JGK 3. A large number of genes involved in cell wall synthesis and modification also exhibited higher expression in JGK 3 at S3 stage (Figure 4d). In addition, the genes involved in DNA synthesis, regulation of transcription, RNA synthesis, protein synthesis, development and transport also exhibited higher transcriptional activity in JGK 3 as compared with Himchana 1 from S2 to S5 stages (Figure S8), indicating an extended period of mitotic activity. Together, these data demonstrated major transcriptional differences between Himchana 1 and JGK 3 at S3 stage.

Differential regulation of genes determining seed development and seed size/weight

The expression profiles of key gene families and individual genes involved in cell division, cell size determination, cell wall modification, carbohydrate metabolism and grain filling were analyzed. We observed higher expression of several members of these gene families in JGK 3 cultivar,

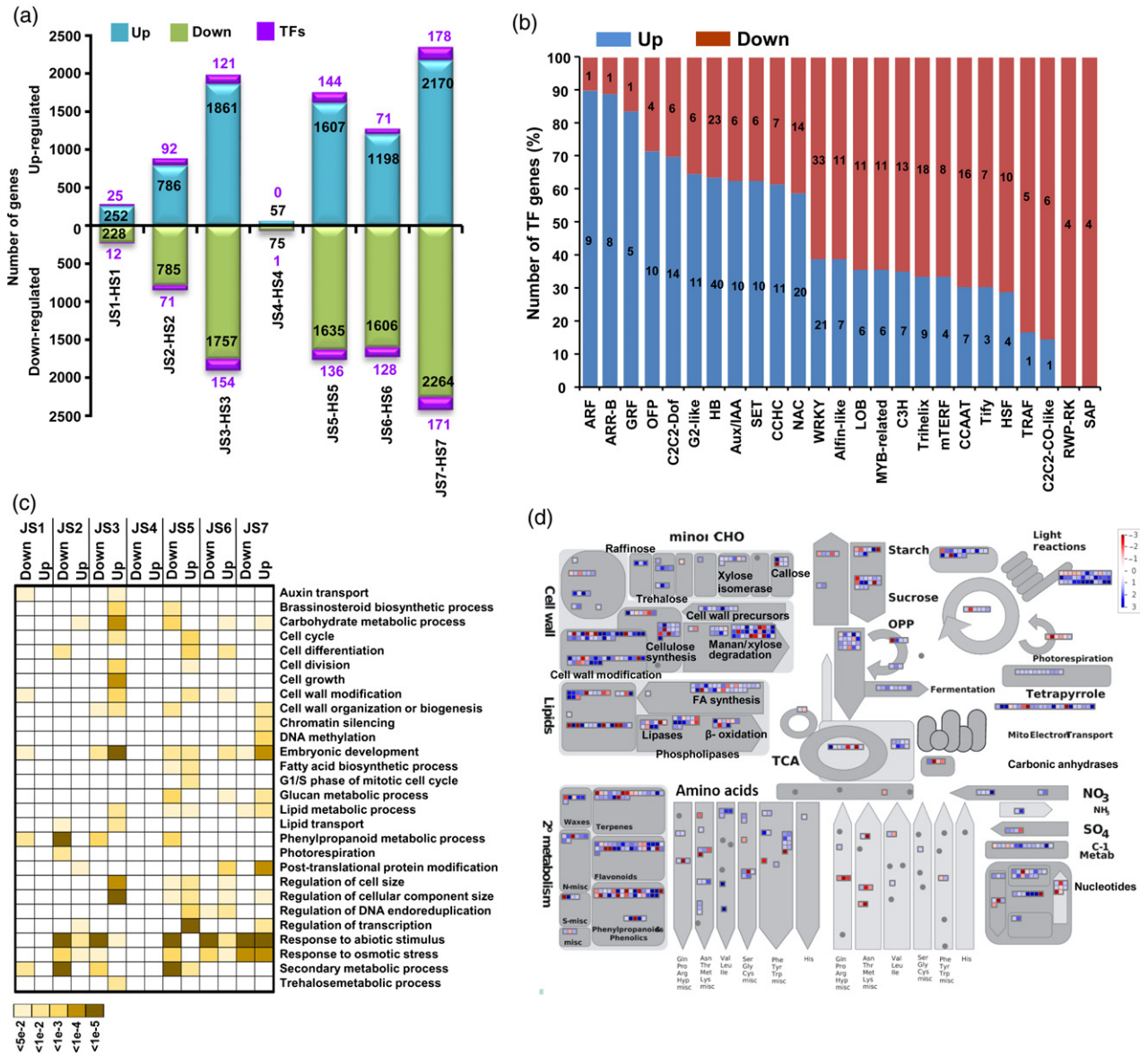


Figure 4. Differential gene expression in JGK 3 as compared with Himchana 1 at different stages of seed development.
(a) The number of upregulated (upper bars) and downregulated (lower bars) genes at each stage of seed development in JGK 3 as compared with Himchana 1 is given. The number of transcription factors (TFs) up- or downregulated at each stage of seed development in JGK 3 is also given.
(b) The number of genes from different TF families showing up- or downregulation in JGK 3 during seed development.
(c) Enriched gene ontology (GO) terms (biological process) at different stages of seed development in down- and upregulated genes in JGK 3. The color scale at the bottom represents significance (corrected *P*-value).
(d) Metabolic pathways with differential expression profile in JGK 3 as compared with Himchana 1 at S3 stage of seed development. Differentially expressed genes (fold change ≥2, *q*-value ≤0.05) between Himchana 1 and JGK 3 at S3 stage were loaded into MapMan to generate the overview. On the $\log_2$ scale, dark blue and dark red colors represent higher and lower expression, respectively, in JGK 3 as compared with Himchana 1.

particularly at S3–S5 stages of seed development (Figure S9). A higher transcriptional activity of A-type and B-type cyclin-encoding genes in JGK 3 till S5 stage indicated higher mitotic activity and an extended period of cell division in JGK 3. The expression of D-type cyclin-encoding genes was also higher in S3–S5 stages in JGK 3, which are involved in the control of G1/S transition (Dante *et al.*, 2014). The genes encoding glucan synthases, cellulose synthases and xyloglucan endotransglucosylases/hydrolases exhibited higher transcriptional activity in JGK 3 cultivar during embryogenesis and/or mid-maturation stages. These enzymes are the integral components of cell wall biosynthetic machinery and are involved in production of energy (Stork *et al.*, 2010; Mendu *et al.*, 2011; Redekar *et al.*, 2015). Further, the transcript abundance of genes involved in cell expansion (expansins), seed storage

proteins (vicilins) and lipid transfer proteins was also significantly higher in the large-seeded chickpea cultivar (Figure S9). These proteins have been implicated in various aspects of seed development, particularly during seed maturation (Chen *et al.*, 2001; Harrison *et al.*, 2001; Gallardo *et al.*, 2003; Milisavljević *et al.*, 2004; Liu *et al.*, 2015; Wang *et al.*, 2015).

The transcriptional activity of the genes encoding some of the bonafide components of cell cycle machinery, CYCD3;1 (*XLOC_013008*), APC2 (*XLOC_019430*), MYB3R-4 (*XLOC_011883*) and E2Fa/b (*XLOC_011436*), was significantly higher in JGK 3 cultivar during different stages of seed development (Figure 5). CYCD3;1 positively regulates G1/S phase transition during the cell cycle (Menges *et al.*, 2006; Dewitte *et al.*, 2007). APC2, a member of the cullin protein family, represents a subunit of APC/C (anaphase promoting complex or cyclosome) that regulates mitotic progression (Capron *et al.*, 2003). Transcriptional activity of the TF gene, *MYB3R-4*, was significantly higher in JGK 3 from S1 to S5 stages. MYB3R-4 regulated plant development via controlling the expression of G2/M-specific genes (Haga *et al.*, 2011). It has also been demonstrated that this TF is required for the establishment of endoreduplication at the site of fungal infection in Arabidopsis (Chandran *et al.*, 2010). E2Fa and E2Fb are the transcriptional regulators of genes involved in cell cycle progression and endoreduplication (De Veylder *et al.*, 2002; Rossignol *et al.*, 2002; Sozzani *et al.*, 2006). The expression of genes encoding components of DNA replication machinery, such as origin recognition complex (ORC1A, *XLOC_018689* and ORC6, *XLOC_004561*) and helicase (MCM8), was also higher in JGK 3 cultivar as compared with Himchana 1 (Figure 5). ORC1A is a PHD domain containing H3K4me3 effector protein, which represents a large subunit of ORC complex and is involved in defining the origin of DNA replication via acting as transcriptional activator of a subset of genes (de la Paz Sanchez and Gutierrez, 2009). ORC6 is involved in initiation of DNA replication with higher gene expression during G1/S phase and helps mitotic cell progression (Diaz-Trivino *et al.*, 2005; Vandepoele *et al.*, 2005). MCM8 showed significantly higher expression levels at S5 stage in JGK 3, which has been implicated in double-strand break repair in Arabidopsis (Crismani *et al.*, 2013). However, its helicase activity and role in assembly of pre-replicative complex in Xenopus and human, respectively, has also been demonstrated (Maiorano *et al.*, 2005; Volkening and Hoffmann, 2005).

Another gene *XLOC_008046* encoding for a putative CLASP protein, which is known to be associated with microtubules and is involved in both cell division and cell expansion (Ambrose *et al.*, 2007), is highly expressed during S3–S5 stages in JGK 3 (Figure 5). The transcript abundance of *NRP2* gene (*XLOC_023728*) encoding for a histone chaperone was also higher at S3 stage in JGK 3. NRP2 positively regulates cell cycle progression at G2/M in root (Zhu *et al.*, 2006). The expression of a brassinosteroid signaling gene, *TOP6B* (*BIN3/HYP6, XLOC_000845*), was high at S5 stage, which positively regulates cell expansion via controlling DNA replication and endoreduplication levels (Hartung *et al.*, 2002; Yin *et al.*, 2002). More recently, its role in carbon partitioning into secondary metabolites and cell wall has also been demonstrated (Mittal *et al.*, 2014).

In addition, the transcript levels of specific genes known to be involved in fatty acid biosynthesis during seed maturation, such as *WRINKLED1* (*WRI1, XLOC_026565*), *TRIACYLGLYCEROL BIOSYNTHESIS DEFECT1* (*TAG1, XLOC_004644*) and *STEROL METHYLTRANSFERASE2* (*SMT2, XLOC_015400*), were found to be significantly higher in JGK 3 cultivar from S1 to S5 stages of seed development (Figure 5). *WRI1* encodes an AP2-type TF, which acts as a master regulator of seed oil accumulation and seed maturation processes in several plants (Ruuska *et al.*, 2002; To *et al.*, 2012; Ma *et al.*, 2013; Adhikari *et al.*, 2016) and may perform a similar function in chickpea as well. The expression of *TAG1* has been found to be increased during seed maturation and might be responsible for accumulation of fatty acids. *SMT2* gene product catalyzes the addition of one or two methyl groups at C-24 in phytosterols (Husselstein *et al.*, 1996; Bouvier-Navé *et al.*, 1998). Mutation in *SMT2* causes altered sterol levels in the Arabidopsis plants (Carland *et al.*, 2010), which play a major role in cell division and cell expansion via transcriptional activation of associated genes (He *et al.*, 2003). The expression of *XLOC_015363*, an ortholog of Arabidopsis *unknown seed protein-like 1* (*USPL1*), was significantly higher in JGK 3 at S5 stage as compared with Himchana 1. *USPL1* encodes a BURP domain-containing protein, which is targeted to protein storage vacuoles, endoplasmic reticulum and Golgi bodies, and behaves like a storage protein (Van Son *et al.*, 2009). It is likely that these genes might be responsible for determining seed size/weight in chickpea. However, further in-depth functional analysis is required to establish their exact role in seed development and seed size/weight determination.

Identification of conserved and/or divergent gene coexpression modules

To investigate the gene regulatory network (GRN) during seed development, we identified coexpressed gene sets via weighted gene coexpression network analysis (WGCNA). The genes showing very low expression and/or low SCC value were not considered for this analysis to avoid inclusion of spurious edges in the networks. We performed GRN analysis for both the chickpea cultivars separately. This GRN analysis revealed several major subnetworks representing interaction among genes with similar expression profiles, which were referred to as coexpression modules. A total of 27 modules (comprised of

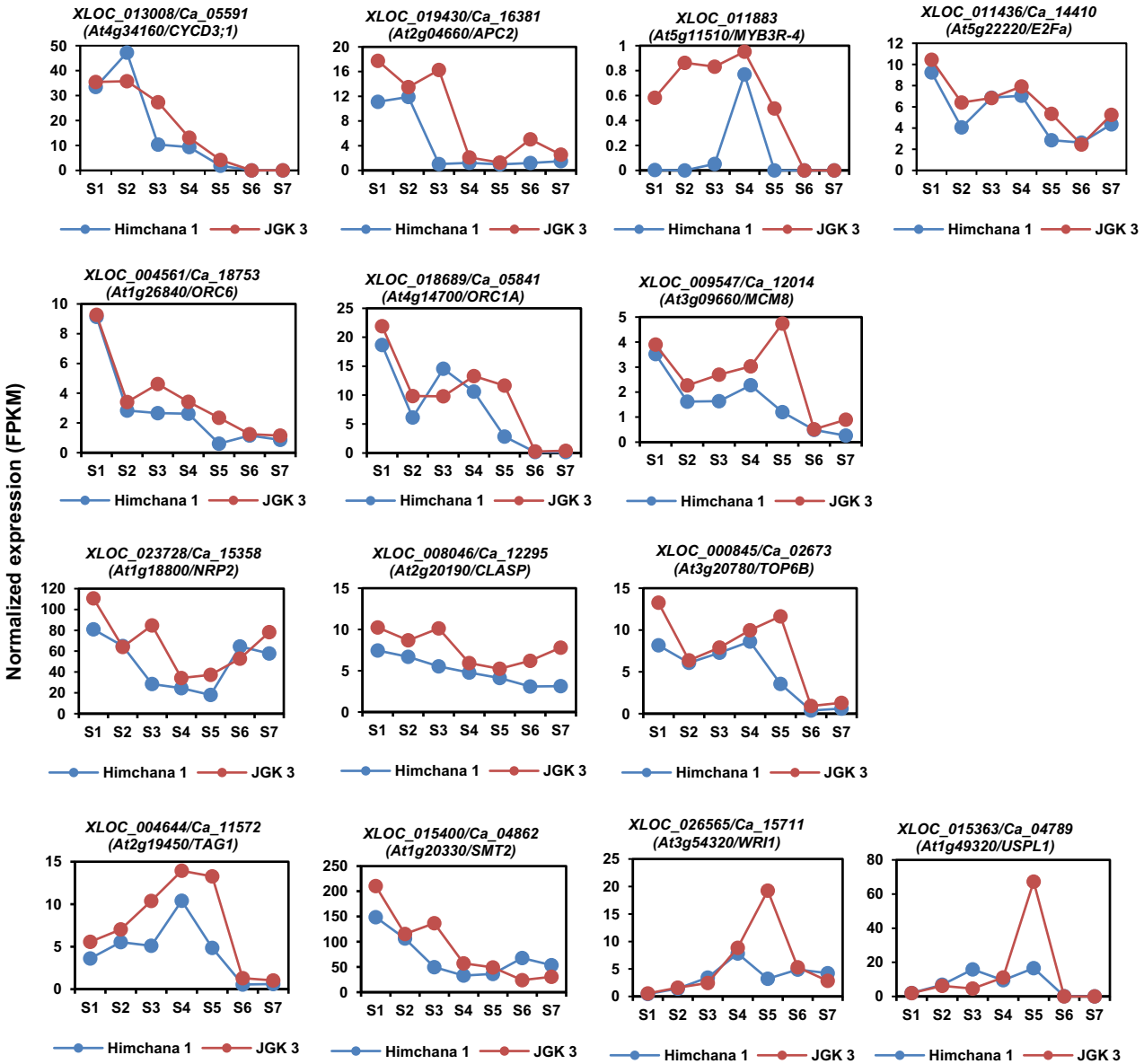


Figure 5. Differential expression of representative genes involved in various aspects of seed development.
Expression profiles of genes involved in cell division, DNA replication, endoreduplication and seed development during different stages of seed development in Himchana 1 (HS1–HS7) and JGK 3 (JS1–JS7) have been shown.

85–2221 genes) were identified in Himchana 1 (Figures 6a and S10Aa,c; Table S3), while 21 modules (comprised of 45–2873 genes) were recognized in JGK 3 (Figures 6b and S10b,d; Table S4). All the modules in both Himchana 1 and JGK 3 harbored TF-encoding genes varying from two (darkred) to 152 (turquoise) in Himchana 1, and one (lightyellow and royalblue) to 246 (turquoise) in JGK 3 (Figure S11). TF-encoding genes represented >10% of the total genes included in some of the modules, indicating tight regulation of the transcriptional activity. Further, we associated each of the coexpression modules with stages of seed development via Pearson correlation coefficient analysis. Interestingly, 13 coexpression modules of Himchana 1 and six coexpression modules of JGK 3 showed relatively higher correlation ($r \geq 0.60$) with seed developmental stages (Figure S10c,d). Many of the modules were correlated with more than one seed development stage; however, a few of them were correlated with a specific seed development stage only. For example, the darkred module of Himchana 1 was specifically correlated ($r \geq 0.93$) with S4 stage. In JGK 3, lightyellow and royalblue modules were specifically correlated with S4 (0.93) and S5 (0.94) stages of seed development, respectively (Figure S10c,d).

Gene ontology enrichment analysis of each module highlighted key biological processes represented by a set of coexpressed genes, and corroborated our results of

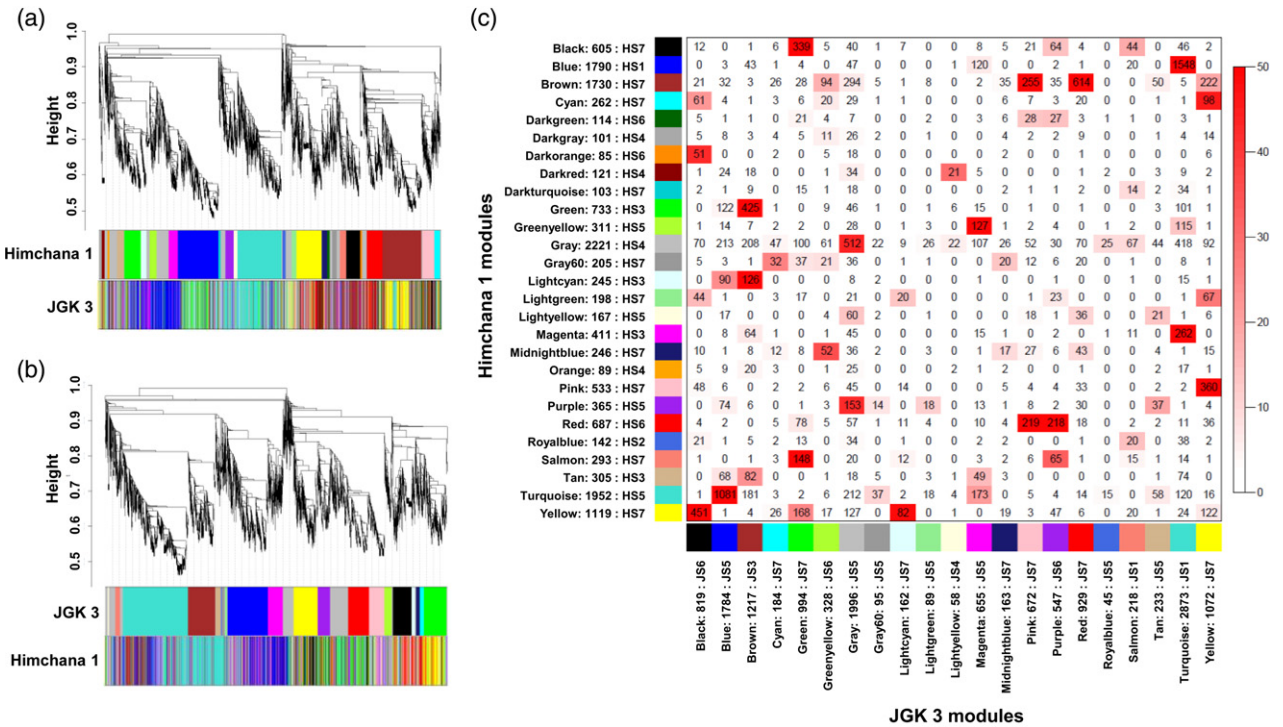


Figure 6. Coexpression network during seed development in chickpea cultivars.
(a,b) Hierarchical clustering tree (dendrogram) of genes based on coexpression network analysis in Himchana 1 (a) and JGK 3 (b). Each 'leaf' (short vertical line) corresponds to individual gene. The genes were clustered on the basis of dissimilarity measure (1-TOM). The branches correspond to modules of highly interconnected genes. The color rows below the dendrograms indicate module membership in Himchana 1 and their corresponding modules in JGK 3 (a), or module membership in JGK 3 and their corresponding modules in Himchana 1 (b).
(c) Comparison of Himchana 1 and JGK 3 modules based on cross-tabulation and *Z*-score calculation. Each Himchana 1 (row) and JGK 3 (column) module is labeled by the corresponding module color, and the total number of genes in the module is shown. The numbers of genes for each intersection of the corresponding row and column modules are given. The color in the boxes represents $-log(P)$, the Fisher exact test *P*-value, according to the color legend on the right.

differential gene expression analysis (Tables S3 and S4). For example, the modules associated with early stages of development showed enrichment of GO terms related to cell division, cell morphogenesis, cell wall organization and regulation of gene expression. Mid-maturation stages associated modules represented GO terms related to chromatin modification, protein/histone modification, fatty acid biosynthesis and secondary metabolism, etc. Likewise, the modules associated with late maturation exhibited overrepresentation of terms, such as transport, nitrogen compound metabolism, regulation of transcription, response to abiotic/osmotic stimulus and protein metabolic processes. The darkred module of Himchana 1 associated with S4 stage included most of the genes involved in epigenetic modifications. The lightyellow module of JGK 3 specific to S4 stage included genes related to RNA splicing, and the royalblue module specific to S5 stage included genes involved in response to amino acid and nitrogen.

Next, we studied the preservation of coexpression modules between Himchana 1 and JGK 3 at different stages of seed development via cross-tabulation approach by calculation of Z-summary (Figure 6). Most of the modules were found to be preserved across the cultivars with similar seed development SS (Figure 6c). Interestingly, a few modules preserved across the cultivars exhibited a difference in developmental SS/timing of transcriptional activity. Further, we identified three modules of Himchana 1, i.e. orange (HS4 and HS2), darkgray (HS4) and darkturquoise (HS7), and four modules of JGK 3, i.e. lightyellow (JS4), royalblue (JS5), gray60 (JS5 and JS4) and lightgreen (JS5 and JS4), which were cultivar-specific. The Himchana 1-specific modules harbored genes involved in regulation of transcription, programed cell death, senescence, chromatin silencing and nitrogen metabolic processes (Table S3). JGK 3-specific modules were found to be enriched in genes associated with GO terms, including DNA replication, cell division, gene expression, protein modification, amino acid stimulus and xylan biosynthetic process, etc. (Table S4). Taken together, these results indicated that each of the seed development stages was associated with one or more coexpression modules in each cultivar that reflected the gene regulatory processes specific to each stage, and are indicators of the differential programs functioning at each stage of seed development within or across the cultivars.

Transcription regulatory modules related to seed development and seed size/weight

We defined the GRNs that associate the TFs with their coexpressed target genes harboring their binding sites (significantly enriched motifs) at S3 and S5 stages of seed development in JGK 3. For this analysis, we used coexpressed genes included in the modules found associated with S3 and S5 stages of seed development in JGK 3 and Himchana 1 (Datasets 1, 2). This resulted in prediction of transcriptional modules linking the enriched regulatory motifs with known TFs involved in seed development and their association with specific GO terms represented significantly in the target genes. The transcriptional module of S3 stage of seed development in JGK 3 (brown module) included significantly enriched DNA sequence motifs, ATHB1, SORLIP3/5, JASE1 and L1-box mainly associated with TFs, like WOX9, PDF2, RLT2, ANL2, JAB and HDG11, and target genes associated with GO terms, regulation of gene expression, cell wall organization, regulation of cell size and post-translational protein modification (Figure S12). Many of these components were similar for the transcriptional module at S3 stage in Himchana 1 (magenta, green and lightcyan modules) with some specific components, including motifs, CARG3, PRHA and TELO-box, and TFs, AGL15, WOX2/3 and SEP3 (Figure S12). Some of these regulatory motifs and TFs are known to be associated with embryogenesis and act coordinately for transcriptional activation of genes. The MADS-box TF, AGL15, plays a crucial role in embryogenesis and binds to CArG motif (Perry *et al.*, 1999; Fernandez *et al.*, 2000; Tang and Perry, 2003).

At S5 stage of seed development in JGK 3 (gray60, lightgreen, blue, royalblue and tan modules), motifs like CCA1, ABRE, ABFS, PRHA, TELO-box and MYB1 were enriched (Figure S13). The DNA motifs, including ABRE and ABFS, were associated with bZIP family TFs (bZIP44 and DPBF2/3). Similarly, CCA1 motif was associated with TRB1, TRFL6 and PRZ1 TFs, and PRHA motif was associated with HB, PDF2 and WOX12 TFs. MYB1 motif was associated with several MYB family TFs. The transcriptional module included target genes involved in brassinosteroid-mediated signaling, epigenetic regulation of gene expression, glucan metabolic process, lipid metabolic process, regulation of cell size and response to hormones. An evident overlap of enriched DNA motifs and TFs was observed in the transcriptional module of Himchana 1 at S5 stage. However, some of the regulatory motifs, such as RAV1-A, TGA1, LFY and OBF, and TFs, LEC2, FUS3 and VAL1 with target genes involved in phenylpropanoid metabolic processes and response to abscisic acid were unique to the transcriptional module in Himchana 1 (Figure S13). Many of these DNA motifs have been found to be associated with promoters of genes involved in seed maturation (Gutierrez *et al.*, 2007; Penfield and Hall, 2009). The TFs LEC2 and FUS3 are well-known key transcriptional regulators of seed maturation (Santos-Mendoza *et al.*, 2008). Overall, these analyses identified key regulators of seed development and suggested that seed development processes are regulated similarly, but not identically, in the two chickpea cultivars.

We observed that some modules displayed opposite expression patterns between the two chickpea cultivars at S3 and S5 stages. For example, orange module of Himchana 1, and royalblue and salmon modules of JGK 3 showed low transcript abundance in Himchana 1, but high in JGK 3 cultivar at S3 stage. Likewise, lightyellow and purple modules of Himchana 1, and gray60 and lightgreen modules of JGK 3 exhibited high transcript abundance in Himchana 1, but low in JGK 3 at S3 stage. Tan module of Himchana 1 included genes with high expression in JGK 3 and low expression in Himchana 1 at S5 stage. Considering that these modules might be responsible for the divergence in seed development program of Himchana 1 and JGK 3, we analyzed the GRNs for these modules (Figure 7; Data S3). The GRN of genes showing lower expression in Himchana 1 and higher expression in JGK 3 at S3 stage included significantly enriched regulatory elements, AtbZIP53/44, ARF, CBF, AGL, CARG1/3, MYB2/3/4 and HB with TFs, such as bZIP34, ORE14, ERF13, FEM111 and TRB1 and target genes involved in regulation of cell cycle, chromatin organization, histone modifications, signal transduction, regulation of transcription and gluconeogenesis. The expression pattern of top hub genes at S3 stage reflected the expression pattern of all the genes included in different modules (Figures 7 and S14). The transcriptional module of genes showing high expression in Himchana 1 and low expression in JGK 3 at S3 stage was comprised of TFs, bZIP5/44, AGL30, GBF3/6, DPBF2, WRKY22/27, WAM1, MYB33/98 and ARF17, significantly enriched cis regulatory motifs, AGL15, CARG3, ABFS, ABRE, G-box, AtMYC, PRHA, RAV1-B and HB with plausible target genes involved in mitotic cell cycle check point, cell wall modification, regulation of cell size, gene expression, response to abscisic acid stimulus and lipid metabolic process, etc. The GRN at S5 stage (high expression in JGK 3 and low expression in Himchana 1) included enriched regulatory motifs, EIL1, ERF1, CARG3, TEF-box, ARF and RY-repeat with TFs, CRF10, DDF2, WOX1, HDG11 and MYB3R-4, and their target genes involved in epigenetic regulation of gene expression, regulation of cell size and post-translational protein modification. Many of these regulatory motifs and TFs are known to be involved in determination of seed size/weight. Overall, these analyses identified transcriptional modules as key regulators and their role in controlling seed development and determination of seed size/weight.

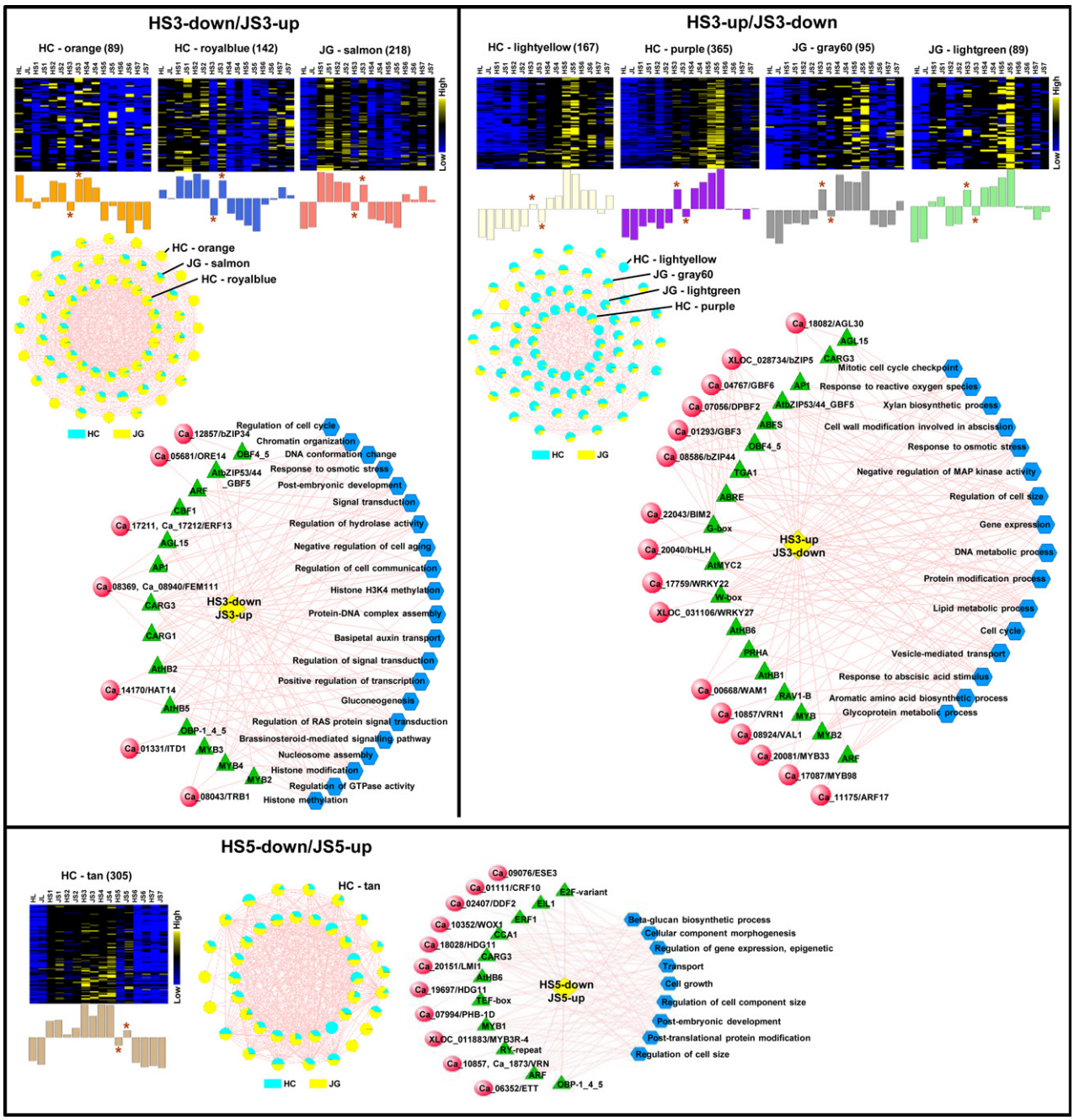


Figure 7. Expression profile and transcriptional regulatory network associated with the modules showing opposite expression patterns in Himchana 1 and JGK 3 at S3 or S5 stages.

Heatmaps show the expression profile of all the coexpressed genes (number given on the top) in the modules (labeled on top). The color scale represents *Z*-score. Bar graphs (below the heat maps) show the consensus expression pattern of the coexpressed genes in each module. The bars showing opposite expression patterns in the chickpea cultivars at S3 or S5 stage have been marked with asterisks. The network of top hub genes is shown in circular form below the bar graphs. For the modules showing HS3-down/JS3-up and HS3-up/JS3-down expression, the network of top 20 hub genes and for the module showing HS5-down/JS5-up expression network of top 40 hub genes is shown. The names of the modules have been indicated. Yellow and blue colors indicate expression in JGK 3 and Himchana 1, respectively, at the S3 or S5 stage. HC, Himchana 1; JG, JGK 3. A larger version of these networks with gene identifier labels is given in Figure S14. The predicted transcriptional regulatory network [significantly enriched transcription factor (TF)-binding sites along with the associated TFs and enriched gene ontology (GO) terms] associated with the gene sets showing opposite expression patterns at S3 or S5 stage of seed development are given. The significantly enriched cis-regulatory motifs (green triangles) and GO terms (blue hexagons) within the given set of genes. The TFs are represented by magenta circles. Edges represent known interactions between the cis-regulatory motifs and TFs. The transcriptional regulatory networks were predicted for the module sets with opposite expression patterns in Himchana 1 and JGK 3 at S3 or S5 stages (HS3-down_JS3-up, HS3-up_JS3-down and HS5-down_JS5-up).

Candidate genes and SNPs underlying QTLs determining seed size/weight

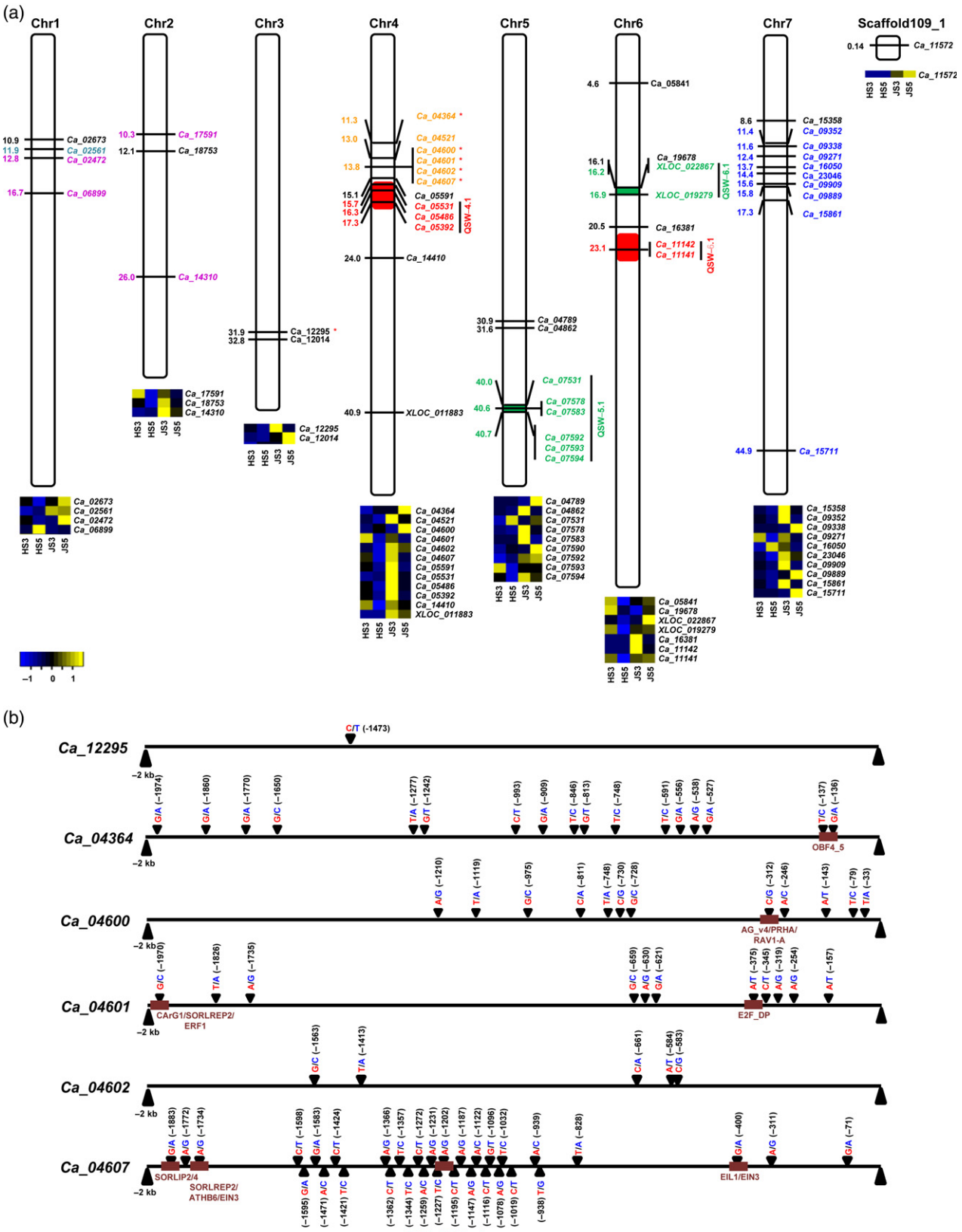
Further, we sought to identify genetic variations among the small- and large-seeded cultivars that might be responsible for differential gene expression and determine seed weight/size in chickpea. For this analysis, we integrated the known QTLs (based on genomic coordinates) and candidate genes (based on gene identifiers) associated with seed size/weight reported in previous studies (Kujur *et al.*, 2013, 2015; Saxena *et al.*, 2014; Bajaj *et al.*, 2015, 2016; Verma *et al.*, 2015; Singh *et al.*, 2016) with our differential gene expression results. A total of 34 genes, including those located in the QTLs, genes associated with seed weight/size, involved in cell cycle/cell division, endoreduplication and seed storage protein accumulation and/or showing higher transcript abundance in JGK 3 at S3 and/or S5 stages as compared with Himchana 1, were identified (Figure 8a; Dataset 4). In addition, 14 candidate genes involved in different processes that are likely to determine seed size/weight in chickpea as described in Figure 5 were selected. Firstly, we validated the differential expression pattern of 34 (including 14 genes from Figure 5) of these 48 genes between JGK 3 and Himchana 1 cultivars at S3 and S5 stages via RT-qPCR (Figure S15). This analysis confirmed the higher transcript abundance of most of these genes at S3 and/or S5 stage(s) in JGK 3 cultivar. Next, we analyzed 2-kb upstream (promoter) sequences of the 48 genes in 10 chickpea cultivars/genotypes, including four small-seeded (Himchana 1, JG 62, ICC 1882 and ICC 4567 with average 100 seed weight 13.1–16.7 g) and six large-seeded (JGK 3, PG 0515, ICC 8261, LBeG7, ICC 4958 and SBD 377 with average 100 seed weight 28.5–56.9 g) cultivars/genotypes, for identification of SNPs among them. A total of 107 SNPs among the 10 chickpea cultivars/genotypes located in the promoter sequences of 23 genes were identified (Data S5). Among these, at least 79 SNPs located in the promoter sequences of six genes (*Ca_12295* on chromosome 3, and *Ca_04364*, *Ca_04600*, *Ca_04601*, *Ca_04602* and *Ca_04607* on chromosome 4) were identified, which clearly differentiated small-seeded and large-seeded chickpea cultivars/genotypes (Figure 8b; Dataset 5). The number of the identified SNPs ranged from one (*Ca_12295*) to 33 (*Ca_04607*) per gene. The genes harboring these SNPs encoded for a putative CLASP protein (*Ca_12295*), protein kinase family protein (*Ca_04364*), aminotransferase-like protein (*Ca_04600*), SurE-like survival protein (*Ca_04601*), phosphatidylinositol transfer protein (*Ca_04602*) and SNARE-associated Golgi protein (*Ca_04607*). RNA-seq and/or RT-qPCR analyses revealed higher transcript abundance of these genes at S3 and/or S5 stages of seed development in JGK 3 as compared with Himchana 1 (Figures 8a and S15). Some of these SNPs were located within the cis-regulatory (TF binding) motifs that constituted the transcriptional modules determining seed development and seed size/weight (Figures 7 and S13). The genotyping of identified SNPs in a large set of chickpea cultivars/genotypes can help developing haplotype(s) associated with seed weight/size determination in chickpea, and the genes harboring these SNPs are the most promising targets for engineering seed weight/size in chickpea.

DISCUSSION

The molecular mechanisms underlying seed development are poorly understood in legumes. We used RNA-seq approach to detect the transcriptome dynamics in two chickpea cultivars (with contrasting seed size/weight) at different stages of development, and investigated the molecular mechanism underlying seed development and difference in seed size/weight. More than 80% of the chickpea genes were found to be expressed in at least one stage of seed development. Deep RNA-seq facilitated the discovery of novel genes and their expression profiles. The expression data across the seven seed development stages showed high reproducibility in both the chickpea cultivars and each stage was clearly distinguished in the PCA plot, suggesting that significant changes in gene expression occur from one stage to another. The dataset provides a comprehensive description of transcriptional activity during different stages of seed development in chickpea. The comprehensive transcriptome analyses with the inference of coexpression networks and transcriptional modules identified several coregulated and specific transcriptional programs within and across the cultivars associated with seed development and seed size/weight determination.

It has been demonstrated that differential allocation of resources during seed set determines seed size/weight across plant species (Gambin and Borras, 2010). Further, an increase in seed size/weight has been associated with an increase in other traits, including seed weight, stem

Figure 8. Location of selected candidate genes on chickpea genome, their differential expression and single nucleotide polymorphisms (SNPs) differentiating small- and large-seeded chickpea genotypes.
(a) Genomic location of the selected 48 genes on the known quantitative trait loci (QTLs; if any, shown in colored boxes) and chickpea genome are shown. QTLs/candidate genes identified from different studies have been shown in different colored fonts. Heatmaps show the differential expression of the selected genes (present on the above chromosome). The color scale represents *Z*-score.
(b) The position of identified SNPs that differentiate small- and large-seeded chickpea cultivars are shown on the upstream (promoter) regions (2 kb) of the six genes. The alleles in small- (blue) and large-seeded (red) cultivars are given in different colored font. The *cis*-regulatory motifs overlapping the SNP position(s) have been shown in brown boxes.



thickness and vigor in preflowers for flowering plants (Elmore and Abendroth, 2005; Gambin and Borras, 2010). The existence of common genetic components regulating size of seeds and other organs via affecting cell cycle and duration of development has been hypothesized (Linkies *et al.*, 2010). In contrast, many other studies did not support correlation between seed size and plant biomass (Graven and Carter, 1990; Elmore and Abendroth, 2005). A prolonged phase of cell division resulting in a higher number of cells in the seed coat, cotyledon or endosperm is considered to be responsible for larger seed size/weight (Weber *et al.*, 1996; Sekhon *et al.*, 2014). The increase in seed weight between S3 and S5 stages in both the chickpea cultivars suggested that this period represents transition between cell division to maturation. The differential rate of increase in seed weight suggested that the number of cell divisions occurring in the two chickpea cultivars are different. The GO enrichment analyses revealed a prolonged period of cell division and mitotic activity with higher expression of genes involved in cell division and cell cycle in JGK 3. The extended period of mitotic activity has been found accompanied with higher gluconeogenesis, which plays a key role in seed size determination in maize (Sekhon *et al.*, 2014). A larger number of cells in JGK 3 may lead to more accumulation of storage compounds/proteins. Further, endoreduplication is considered as a progressive phenomenon in reserve accumulating plant organs during the transition between cell division and maturation, and a positive correlation between the endoreduplication (ploidy) level in mature cotyledon and seed size has been observed in previous studies (Galbraith *et al.*, 1991; Lemontey *et al.*, 2000; Dante *et al.*, 2014). We observed higher transcriptional activity of genes involved in G1/S transition and enrichment of GO term, regulation of endoreduplication, during S5 stage in JGK 3, suggesting higher endoreduplication in JGK 3. The higher transcript abundance of *BIN3*/*TOP6B* at S5 stage in large-seeded JGK 3 indicated the role of brassinosteroid signaling in determining seed size/weight in chickpea. Earlier studies have reported a crucial role of brassinosteroids in determining seed development and seed size (Sun *et al.*, 2010; Jiang and Lin, 2013; Locascio *et al.*, 2014). Brassinosteroid-insensitive mutants resulted in production of smaller seeds (Choe *et al.*, 2000; Jiang *et al.*, 2013; Jiang and Lin, 2013; Li and Li, 2016). In addition, a higher transcriptional activity of several known genes associated with cell expansion, fatty acid biosynthesis and storage compounds was found to be associated with seed maturation in the large-seeded chickpea cultivar. Altogether, an extended period of cell division/mitotic activity along with higher endoreduplication level and more accumulation of storage compounds in JGK 3 might explain its larger seed size/weight.

Many of the TF families have been implicated in seed development (Le *et al.*, 2010; Agarwal *et al.*, 2011; Verdier *et al.*, 2013); however, only a few TFs have been implicated in determination of seed size. In our dataset, about 50% of the total TFs were found to be differentially expressed between Himchana 1 and JGK 3, suggesting a differential transcriptional regulatory network between the chickpea cultivars. Some of the known families were represented among the differentially expressed TF-encoding genes; however, the exact function of most of these genes remains to be uncovered. The role of some members of TF families, such as ARF, homeobox, NAC and WRKY, which showed differential regulation across the chickpea cultivars, is well known in seed development (Le *et al.*, 2010; Agarwal *et al.*, 2011; Verdier *et al.*, 2013; Li and Li, 2016). The differential regulation of members of the same gene family in different chickpea cultivars may lead to different regulatory networks, which can determine cultivar-specific seed development and seed size/weight. To better understand the seed development process, we performed coexpression network analysis and identified common and unique gene sets/modules including a significant number of TFs associated with developmental stages from embryo to fully-developed seed for both the cultivars. The TFs can act as regulatory components to coordinate the activity of coexpressed genes within each module. GO enrichment analyses of modules highlighted the role of several biological processes/pathways in seed development and suggested that determination of seed size/weight occurs via processes occurring mainly at late embryogenesis (S3) and maturation (S5) stages of seed development.

In Arabidopsis, the regulatory gene circuitry of the seed maturation process has been elucidated based on the comprehensive transcriptome profiling of different seed parts (Belmonte *et al.*, 2013). The discovery of transcriptional modules can identify GRNs that control biological processes associated with seed development (Sreenivasulu and Wobus, 2013; Becker *et al.*, 2014). Therefore, we constructed the transcriptional modules linking TFs with their potential binding motifs and coexpressed target genes (TF–binding motifs–target genes) for the two (S3 and S5) crucial stages of seed development that are supposed to determine differential seed development and seed size/weight in the two chickpea cultivars. Although there was an extensive overlap in the transcriptional programs during seed development in Himchana 1 and JGK 3 cultivars, several components exhibited specificity of their transcript accumulation, which defined uniqueness of the transcriptional modules operating in the two chickpea cultivars. Further, the transcriptional modules defined from the genes showing opposite gene expression patterns during S3 and S5 stages of seed development in the small- and large-seeded chickpea cultivars included regulatory gene circuits (TFs–regulatory motifs–target genes) involved in cell cycle, chromatin organization, epigenetic/

histone modifications, cell size/growth and energy production. These genes were not found differentially expressed in leaves of Himchana 1 (desi) and JGK 3 (kabuli) cultivars, suggesting that these differences are not due to the background effects of the desi and kabuli chickpea and might determine the differential seed size/weight phenotype. Many components of these transcriptional modules have been implicated in various aspects of seed development processes and/or determination of seed size/weight in different plants (Belmonte *et al.*, 2013; Sreenivasulu and Wobus, 2013; Becker *et al.*, 2014; Dante *et al.*, 2014; Li and Li, 2016). Our results highlighted that construction of transcriptional modules in addition to the coexpression network analysis can act as a powerful tool to understand the molecular mechanisms underlying agronomic traits, such as seed development and seed size/weight. However, further functional studies on individual members of the network are required to elucidate a complete picture of GRNs.

Quite a few studies have identified QTLs and/or candidate genes associated with seed size/weight in chickpea using different biparental populations or diverse germplasm via various approaches (Kujur *et al.*, 2013, 2015; Saxena *et al.*, 2014; Verma *et al.*, 2015; Bajaj *et al.*, 2015, 2016; Singh *et al.*, 2016). In most of these studies, large genomic regions harboring several genes have been identified as QTLs. It has been suggested that integration of transcriptome data with the molecular genetics approaches is a better method to identify candidate genes controlling a particular trait. The integration of our transcriptome analysis with known QTLs and/or candidate genes associated with seed size/weight followed by discovery of SNPs in the promoter sequences identified six candidate genes that might determine seed size/weight in chickpea. Interestingly, five of these genes were located within one QTL on CaLG04 associated with seed size/weight, which was identified recently via QTL-seq approach (Singh *et al.*, 2016). Two of these genes, *XLOC_008046* encoding for a putative CLASP protein, and *Ca_04607* encoding for SNARE-associated Golgi transmembrane protein, are involved in cell division/cell expansion and vesicular membrane trafficking, respectively, which are important processes that regulate seed development and size/weight (Ambrose *et al.*, 2007; Bassham and Blatt, 2008; Ebine *et al.*, 2008; Dante *et al.*, 2014). The presence of SNPs within the TF binding motifs can result in differential expression of these genes in chickpea cultivars with contrasting seed size/weight phenotype. Based on these results, it is possible to target these genes for engineering seed size/weight in chickpea.

In conclusion, RNA-seq data generated from two chickpea cultivars with contrasting seed size/weight in this study present a robust resource to study seed biology, including development and seed size/weight in legumes. We identified gene sets that are expressed in specific stages, determined their enriched biological processes/pathways and defined coexpressed gene sets with high temporal resolution. Further, we identified GRNs that operate during seed development and govern seed size/weight via identifying TFs and enriched cis-regulatory motifs that accumulate in a stage-specific manner. Our results revealed that an extended period of mitotic activity during embryogenesis, and a higher level of endoreduplication and storage compound accumulation during seed maturation are majorly responsible for determining seed size/weight in chickpea. The correlation of known QTLs and seed size/weight-associated genes with the transcriptome data followed by discovery of SNPs in the promoter regions identified six potential target genes determining seed size/weight. Overall, the present study suggests that transcriptional profiles and deduced GRNs along with molecular genetic approaches can help identification of the most promising candidate genes and establish their precise role in seed size/weight determination. We anticipate that comprehensive information presented here will serve as a crucial resource to understand seed biology, in particular leveraging the seed size/weight in legumes.

EXPERIMENTAL PROCEDURES

Plant materials and sampling

Two chickpea (*Cicer arietinum* L.) cultivars with contrasting phenotype for seed size/weight, Himchana 1 (small-seeded) and JGK 3 (large-seeded), were used in this study. Chickpea plants were grown in the field during the winter season of 2014–2015. At least 50 flower buds for each cultivar were analyzed at different stages to determine the day of pollination. Individual flower buds were tagged at 1 day after pollination (DAP) for both the cultivars. Seeds were collected in three biological replicates at 5, 9, 12, 19, 25, 30 and 40 DAP, representing S1, S2, S3, S4, S5, S6 and S7 stages, respectively. At the 5-DAP, 9-DAP and 12-DAP stages, at least 30–40 seeds were harvested for each biological replicate. For later stages, 18–20 seeds were collected for each biological replicate. At least 50 seeds from each stage were used for measurement of weight and size (length). In addition, leaves (subtending the flower bud/pod) from the chickpea plants at S5 stage of seed development were collected in three biological replicates for both cultivars. The tissue samples were collected on dry ice and snap-frozen in liquid nitrogen.

Illumina sequencing, read mapping and differential gene expression analyses

For RNA-seq, total RNA extraction and library preparation for each sample were performed as described previously (Garg *et al.*, 2016). All the 48 libraries (16 samples in three biological replicates) were sequenced on Illumina platform (HiSeq 2000) to generate 100-nucleotide-long paired-end sequence reads. The raw sequence data were assessed for various quality parameters, and high-quality reads were filtered using NGS QC Toolkit (v2.3; Patel and Jain, 2012) as described earlier (Garg *et al.*, 2016). The filtered high-quality reads were mapped on the kabuli chickpea genome

(v1.0; Varshney *et al.*, 2013) using TopHat (v2.0.0) with default parameters. The mapped output was processed via Cufflinks (v2.0.2) to obtain FPKM for all the chickpea genes in each sample. Correlation between the biological replicates was determined via calculating SCC. Hierarchical clustering and PCA analysis were performed using corrplot and prcomp utilities in R package. The differential expression between different samples was determined by Cuffdiff. The genes exhibiting a difference of at least twofold change with corrected *P*-value after adjusting with false discovery rate (*q*-value) ≤ 0.05 were considered to be significantly differentially expressed. The stage-specific/preferential genes in both the cultivars were identified via SS scoring algorithm, which compares the expression of a gene in a given stage with its maximum expression level in other stages as described previously (Zhan *et al.*, 2015). The higher value of SS score of a gene in a particular stage signifies its more specific expression at that stage. For a given set of genes, row-wise *Z*-scores were determined and heatmaps were generated using heatmap2 of R package.

GO and pathway enrichment analysis

Gene ontology enrichment analysis for differentially expressed gene sets was performed using Cytoscape (BiNGO plug-in; Maere *et al.*, 2005). *P*-value for enrichment was calculated for each represented GO term and corrected via Benjamini Hoschberg error correction method. The GO terms exhibiting a corrected (after adjusting with false discovery rate) *P*-value of ≤ 0.05 were considered to be significantly enriched. Further, pathway enrichment analysis of different sets of genes was performed using MapMan (v3.5.1R2) categories (significance value ≤ 0.05) for the best Arabidopsis (TAIR10) homolog.

RT-qPCR analysis

Results of RNA-seq were validated via RT-qPCR experiments. Real-time PCR analyses were performed as described earlier (Garg *et al.*, 2016). The gene-specific primers designed using Primer Express (v3.0) software are listed in Table S5. The real-time PCR analysis was performed using three biological replicates for each tissue sample and at least three technical replicates of each biological replicate. The normalization of transcript level of each gene was done with the transcript level of the most suitable internal control gene, *Elongation factor 1 alpha* or *UBIQUITIN5* (Garg *et al.*, 2010) for each sample, and fold change was calculated using the $2^{-\Delta\Delta CT}$ method.

Coexpression network analysis for construction of modules

For coexpression network analysis, the WGCNA (Zhang and Horvath, 2005; Langfelder and Horvath, 2008) package was used. Based on $\log_2$ (1 + FPKM) values, a matrix of pairwise SCCs between all pairs of genes was generated and transformed into an adjacency matrix (a matrix of connection strengths) using the formula: connection strength (adjacency value) = $|(1 + \text{correlation})/2|^{\beta}$. Here, β represents soft threshold for the correlation matrix, which gives greater weight to strongest correlations while maintaining the gene–gene connectivity. A β value of 12 was selected based on the scale-free topology criterion described by Zhang and Horvath (2008). The resulting adjacency matrix was converted to a topological overlap (TO) matrix via TOM similarity algorithm, and the genes were hierarchically clustered based on TO similarity. The dynamic tree-cutting algorithm was used to cut the hierarchal clustering dendrogram and modules were defined after decomposing/combining branches to reach a stable number of clusters (Langfelder and Horvath, 2008). For each module, a summary profile (module eigengene, ME) was calculated via PCA. Further, the modules with a higher TO value (average TO for all the genes in a given module) than the TO values of modules comprised of randomly selected genes were retained. GO enrichment analysis of each module was performed using Cytoscape as described above.

Identification of stage-specific modules and preservation across the cultivars

To determine the association of module with stage-specific expression for each cultivar, we determined the correlation between each ME with the binary indicator (tissue/stage = 1 and all other samples = 0) as described (Downs *et al.*, 2013). A positive correlation indicated that genes in a module have higher/preferential expression in a particular stage relative to all other samples. Further, we followed a cross-tabulation approach to make a contingency table that reports the number of genes that fall into modules of Himchana 1 versus modules of JGK 3. We used *Z*-summary to assess the preservation of modules of Himchana 1 and JGK 3 in each other.

Prediction of transcriptional modules

We followed the approach described by Belmonte *et al.* (2013) for prediction of transcription modules with some modifications. Firstly, we analyzed 1-kb upstream sequences of all the genes in each module/gene set and identified significantly enriched DNA sequence motifs via HOMER (v4.8.3). Further, their association/binding with the TFs included in the same module/gene set was analyzed via STAMP using AGRIS database. These data along with enriched GO terms for each module were used to generate transcriptional modules via software package ChipEnrich (Orlando *et al.*, 2009).

Identification of SNPs

The 2-kb upstream (promoter) sequences of the selected genes were retrieved from the chickpea genome sequence, and multiple primer (both forward and reverse) combinations were designed. The upstream regions were PCR amplified after optimization using genomic DNA of four small-seeded (Himchana 1, JG 62, ICC 1882 and ICC 4567) and six large-seeded (JGK 3, PG 0515, ICC 8261, LBeG7, ICC 4958 and SBD 377) chickpea cultivars. The amplified PCR products were sequenced using Sanger sequencing platform. The high-quality sequences of each region from the 10 cultivars were aligned and SNPs were detected.

ACCESSION NUMBERS

The entire sequencing data generated in the study are available in the Gene Expression Omnibus (GEO) public repository under the accession numbers GSE79719 and GSE79720.

ACKNOWLEDGEMENTS

This work was supported by the Department of Biotechnology, Government of India, New Delhi (BT/AB/NIPGR/SEED BIOLOGY/2012). RG acknowledges INSPIRE Faculty Award from the Department of Science and Technology, Government of India, New Delhi. VKS acknowledges research fellowship from the Department of Biotechnology, Government of India, New Delhi.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Overview of experiment design and analysis strategy.

Figure S2. Heatmaps showing correlation between transcriptomes of three biological replicates of each tissue sample from Himchana 1 and JGK 3.

Figure S3. Gene expression in Himchana 1 and JGK 3.

Figure S4. Gene ontology (GO) enrichment of preferentially expressed genes in Himchana 1 and/or JGK 3.

Figure S5. Gene ontology (GO) enrichment map of preferentially expressed genes at all the stages of seed development in Himchana 1 and/or JGK 3.

Figure S6. Correlation between expression profiles of selected genes obtained from RNA-seq and reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis.

Figure S7. Heatmaps showing expression profiles of members of selected transcription factor (TF) families with preferential expression in JGK 3 or Himchana 1 cultivar during seed development.

Figure S8. Cellular response pathways showing differential expression between Himchana 1 and JGK 3 at S2–S5 stages of seed development.

Figure S9. Differential expression of members of selected gene families in chickpea cultivars.

Figure S10. Dendrogram and correlation heatmap of module eigengenes (MEs) with seed developmental stages in Himchana 1 and JGK 3.

Figure S11. Distribution of number of genes and transcription factors (TFs) in different modules in Himchana 1 (a) and JGK 3 (b).

Figure S12. Expression profile and transcriptional regulatory network of the modules associated with S3 stage of seed development in Himchana 1 and JGK 3.

Figure S13. Expression profile and transcriptional regulatory network of the modules associated with S5 stage of seed development in Himchana 1 and JGK 3.

Figure S14. Network of top hub genes in the modules exhibiting opposite expression patterns in Himchana 1 and JGK 3 at S3 or S5 stages.

Figure S15. Correlation between expression profiles of the selected genes obtained from RNA-seq and reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis.

Table S1. Summary of read data generated, quality control and mapping on the chickpea genome for different samples for Himchana 1 and JGK 3 cultivars.

Table S2. Number of members of different TF families showing upregulation in the chickpea cultivars at different stages of seed development.

Table S3. GO enrichment of the coexpressed genes in each module in Himchana 1.

Table S4. GO enrichment of the coexpressed genes in each module in JGK 3.

Table S5. List of primers used for RT-qPCR analysis.

Data S1. List of genes included in modules associated with S3 stage used for construction of GRNs in the chickpea cultivars.

Data S2. List of genes included in modules associated with S5 stage used for construction of GRNs in the chickpea cultivars.

Data S3. List of genes included in modules showing opposite expression patterns at S3 or S5 stage used for construction of GRNs in the chickpea cultivars.

Data S4. List of genes selected based on their location within QTL, association with seed weight/size and differential expression at S3 and/or S5 stages in chickpea cultivars.

Data S5. List of SNPs among small-seeded and large-seeded chickpea cultivars identified in the 2-kb upstream (promoter) region of the selected genes.

REFERENCES

Adhikari, N.D., Bates, P.D. and Browse, J. (2016) WRINKLED1 rescues feedback inhibition of fatty acid synthesis in hydroxylase-expressing seeds. *Plant Physiol.* **171**, 179–191.

Agarwal, P., Kapoor, S. and Tyagi, A.K. (2011) Transcription factors regulating the progression of monocot and dicot seed development. *BioEssays*, **33**, 189–202.

Ambrose, J.C., Shoji, T., Kotzer, A.M., Pighin, J.A. and Wasteneys, G.O. (2007) The Arabidopsis *CLASP* gene encodes a microtubule-associated protein involved in cell expansion and division. *Plant Cell*, **19**, 2763–2675.

Bajaj, D., Upadhyaya, H.D., Khan, Y., *et al.* (2015) A combinatorial approach of comprehensive QTL-based comparative genome mapping and transcript profiling identified a seed weight-regulating candidate gene in chickpea. *Sci. Rep.* **5**, 9264.

Bajaj, D., Srivastava, R., Nath, M., Tripathi, S., Bharadwaj, C., Upadhyaya, H.D., Tyagi, A.K. and Parida, S.K. (2016) EcoTILLING-based association mapping efficiently delineates functionally relevant natural allelic variants of candidate genes governing agronomic traits in chickpea. *Front. Plant Sci.* **7**, 450.

Bassham, D.C. and Blatt, M.R. (2008) SNAREs: cogs and coordinators in signaling and development. *Plant Physiol.* **147**, 1504–1515.

Becker, M.G., Hsu, S.W., Harada, J.J. and Belmonte, M.F. (2014) Genomic dissection of the seed. *Front. Plant Sci.* **5**, 464.

Belmonte, M.F., Kirkbride, R.C., Stone, S.L., *et al.* (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl Acad. Sci. USA*, **110**, E435–E444.

Borras, L., Slafer, G.A. and Otegui Ma, E. (2004) Seed dry weight response to source-sink manipulations in wheat, maize and soybean: a quantitative reappraisal. *Field Crops Res.* **86**, 131–146.

Borras, L., Zinselmeier, C., Senior, M.L., Westgate, M.E. and Muszynski, M.G. (2009) Characterization of grain-filling patterns in diverse maize germplasm. *Crop Sci.* **49**, 999–1009.

Bouvier-Navé, P., Husselstein, T. and Benveniste, P. (1998) Two families of sterol methyltransferases are involved in the first and the second methylation steps of plant sterol biosynthesis. *Eur. J. Biochem.* **256**, 88–96.

Capron, A., Serralbo, O., Fülöp, K., *et al.* (2003) The Arabidopsis anaphase-promoting complex or cyclosome: molecular and genetic characterization of the APC2 subunit. *Plant Cell*, **15**, 2370–2382.

Carland, F., Fujioka, S. and Nelson, T. (2010) The sterol methyltransferases SMT1, SMT2, and SMT3 influence Arabidopsis development through nonbrassinosteroid products. *Plant Physiol.* **153**, 741–756.

Chandran, D., Inada, N., Hather, G., Kleindt, C.K. and Wildermuth, M.C. (2010) Laser microdissection of Arabidopsis cells at the powdery mildew infection site reveals site-specific processes and regulators. *Proc. Natl Acad. Sci. USA*, **107**, 460–465.

Chen, F., Dahal, P. and Bradford, K.J. (2001) Two tomato expansin genes show divergent expression and localization in embryos during seed development and germination. *Plant Physiol.* **127**, 928–936.

Choe, S., Tanaka, A., Noguchi, T., Fujioka, S., Takatsuto, S., Ross, A.S., Tax, F.E., Yoshida, S. and Feldmann, K.A. (2000) Lesions in the sterol delta reductase gene of Arabidopsis cause dwarfism due to a block in brassinosteroid biosynthesis. *Plant J.* **21**, 431–443.

Crismani, W., Portemer, V., Froger, N., Chelysheva, L., Horlow, C., Vrielynck, N. and Mercier, R. (2013) MCM8 is required for a pathway of meiotic double-strand break repair independent of DMC1 in *Arabidopsis thaliana*. *PLoS Genet.* **9**, e1003165.

Dante, R.A., Larkins, B.A. and Sabelli, P.A. (2014) Cell cycle control and seed development. *Front. Plant Sci.* **5**, 493.

De Veylder, L., Beeckman, T., Beemster, G.T., *et al.* (2002) Control of proliferation, endoreduplication and differentiation by the Arabidopsis E2Fa-DPa transcription factor. *EMBO J.* **21**, 1360–1368.

Dewitte, W., Scofield, S., Alcasabas, A.A., *et al.* (2007) Arabidopsis CYCD3 D-type cyclins link cell proliferation and endocycles and are rate-limiting for cytokinin responses. *Proc. Natl Acad. Sci. USA*, **104**, 14537–14542.

Diaz-Trivino, S., del Mar Castellano, M., de la Paz Sanchez, M., Ramirez-Parra, E., Desvoyes, B. and Gutierrez, C. (2005) The genes encoding Arabidopsis ORC subunits are E2F targets and the two ORC1 genes are differently expressed in proliferating and endoreplicating cells. *Nucl. Acids Res.* **33**, 5404–5414.

Downs, G.S., Bi, Y.M., Colasanti, J., Wu, W., Chen, X., Zhu, T., Rothstein, S.J. and Lukens, L.N. (2013) A developmental transcriptional network for maize defines coexpression modules. *Plant Physiol.* **161**, 1830–1843.

Ebine, K., Okatani, Y., Uemura, T., *et al.* (2008) A SNARE complex unique to seed plants is required for protein storage vacuole biogenesis and seed development of Arabidopsis thaliana. *Plant Cell*, **20**, 3006–3021.

Egli, D.B., Fraser, J., Leggett, J.E. and Poneleit, C.G. (1981) Control of seed growth in soybeans. *Ann. Bot.* **48**, 171–176.

Elmore, R. and Abendroth, L. (2005) Do corn kernel size and shape matter? In *Newsletter*. (Watch, C., ed). Lincoln: University of Nebraska.

Fernandez, D.E., Heck, G.R., Perry, S.E., Patterson, S.E., Bleecker, A.B. and Fang, S.C. (2000) The embryo MADS domain factor AGL15 acts postembryonically. Inhibition of perianth senescence and abscission via constitutive expression. *Plant Cell*, **12**, 183–198.

Galbraith, D.W., Harkins, K.R. and Knapp, S. (1991) Systemic endopolyploidy in *Arabidopsis thaliana*. *Plant Physiol.* **96**, 985–989.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D. and Burstin, J. (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.* **133**, 664–682.

Gambin, B.L. and Borras, L. (2010) Resource distribution and the trade-off between seed number and seed weight: a comparison across crop species. *Ann. Appl. Biol.* **156**, 91–102.

Garg, R., Sahoo, A., Tyagi, A.K. and Jain, M. (2010) Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). *Biochem. Biophys. Res. Commun.* **396**, 283–288.

Garg, R., Patel, R.K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., Bhatia, S., Chattopadhyay, D., Tyagi, A.K. and Jain, M. (2011) Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiol.* **156**, 1661–1678.

Garg, R., Shankar, R., Thakkar, B., Kudapa, H., Krishnamurthy, L., Mantri, N., Varshney, R.K., Bhatia, S. and Jain, M. (2016) Transcriptome analyses reveal genotype- and developmental stage-specific molecular responses to drought and salinity stresses in chickpea. *Sci. Rep.* **6**, 19228.

Graven, L.M. and Carter, P.R. (1990) Seed size/shape and tillage system effect on corn growth and grain yield. *J. Prod. Agric.* **3**, 445–452.

Gutierrez, L., Van Wuytswinkel, O., Castelain, M. and Bellini, C. (2007) Combined networks regulating seed maturation. *Trends Plant Sci.* **12**, 294–300.

Haga, N., Kobayashi, K., Suzuki, T., *et al.* (2011) Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in Arabidopsis. *Plant Physiol.* **157**, 706–717.

Harrison, E.P., McQueen-Mason, S.J. and Manning, K. (2001) Expression of six expansin genes in relation to extension activity in developing strawberry fruit. *J. Exp. Bot.* **52**, 1437–1446.

Hartung, F., Angelis, K.J., Meister, A., Schubert, I., Melzer, M. and Puchta, H. (2002) An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Curr. Biol.* **12**, 1787–1791.

He, J.X., Fujioka, S., Li, T.C., Kang, S.G., Seto, H., Takatsuto, S., Yoshida, S. and Jang, J.C. (2003) Sterols regulate development and gene expression in Arabidopsis. *Plant Physiol.* **131**, 1258–1269.

Husselstein, T., Gachotte, D., Desprez, T., Bard, M. and Benveniste, P. (1996) Transformation of Saccharomyces cerevisiae with a cDNA encoding a sterol C-methyltransferase from *Arabidopsis thaliana* results in the synthesis of 24-ethyl sterols. *FEBS Lett.* **381**, 87–92.

Jain, M., Misra, G., Patel, R.K., *et al.* (2013) A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* **74**, 715–729.

Jiang, W.B. and Lin, W.H. (2013) Brassinosteroid functions in Arabidopsis seed development. *Plant Signal. Behav.* **8**, e25928.

Jiang, W.B., Huang, H.Y., Hu, Y.W., Zhu, S.W., Wang, Z.Y. and Lin, W.H. (2013) Brassinosteroid regulates seed size and shape in Arabidopsis. *Plant Physiol.* **162**, 1965–1977.

Jones, S.I., Gonzalez, D.O. and Vodkin, L.O. (2010) Flux of transcript patterns during soybean seed development. *BMC Genomics*, **11**, 136.

Kudapa, H., Azam, S., Sharpe, A.G., Taran, B., Li, R., Deonovic, B., Cameron, C., Farmer, A.D., Cannon, S.B. and Varshney, R.K. (2014) Comprehensive transcriptome assembly of chickpea (*Cicer arietinum* L.) using sanger and next generation sequencing platforms: development and applications. *PLoS ONE*, **9**, e86039.

Kujur, A., Bajaj, D., Saxena, M.S., Tripathi, S., Upadhyaya, H.D., Gowda, C.L., Singh, S., Jain, M., Tyagi, A.K. and Parida, S.K. (2013) Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* **20**, 355–374.

Kujur, A., Upadhyaya, H.D., Shree, T., *et al.* (2015) Ultra-high density intraspecific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. *Sci. Rep.* **5**, 9468.

Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Le, B.H., Cheng, C., Bui, A.Q., *et al.* (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc. Natl Acad. Sci. USA*, **107**, 8063–8070.

Lemontey, C., Mousset-Déclas, C., Munier-Jolain, N. and Boutin, J.P. (2000) Maternal genotype influences pea seed size by controlling both mitotic activity during early embryogenesis and final endoreduplication level/cotyledon cell size in mature seed. *J. Exp. Bot.* **51**, 167–175.

Li, N. and Li, Y. (2016) Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* **33**, 23–32.

Linkies, A., Graeber, K., Knight, C. and Leubner-Metzger, G. (2010) The evolution of seeds. *New Phytol.* **186**, 817–831.

Liu, F., Zhang, X., Lu, C., Zeng, X., Li, Y., Fu, D. and Wu, G. (2015) Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. *J. Exp. Bot.* **66**, 5663–5681.

Locascio, A., Roig-Villanova, I., Bernardi, J. and Varotto, S. (2014) Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. *Front Plant Sci.* **5**, 412.

Ma, W., Kong, Q., Arondel, V., Kilaru, A., Bates, P.D., Thrower, N.A., Benning, C. and Ohlrogge, J.B. (2013) Wrinkled1, a ubiquitous regulator in oil accumulating tissues from Arabidopsis embryos to oil palm mesocarp. *PLoS ONE*, **8**, e68887.

Maere, S., Heymens, K. and Kuiper, M. (2005) BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448–3449.

Maiorano, D., Cuvier, O., Danis, E. and Méchali, M. (2005) MCM8 is an MCM2-7-related protein that functions as a DNA helicase during replication elongation and not initiation. *Cell*, **120**, 315–328.

Mendu, V., Griffiths, J.S., Persson, S., Stork, J., Downie, A.B., Voiniciuc, C., Haughn, G.W. and DeBolt, S. (2011) Subfunctionalization of cellulose synthases in seed coat epidermal cells mediates secondary radial wall synthesis and mucilage attachment. *Plant Physiol.* **157**, 441–453.

Menges, M., Samland, A.K., Planchais, S. and Murray, J.A. (2006) The D-type cyclin CYCD3;1 is limiting for the G1-to-S-phase transition in Arabidopsis. *Plant Cell*, **18**, 893–906.

Milisavljević, M.D.J., Timotijević, G.S., Radović, S.R., Brkljacić, J.M., Konstantinović, M.M. and Maksimović, V.R. (2004) Vicilin-like storage globulin from buckwheat (*Fagopyrum esculentum* Moench) seeds. *J. Agric. Food Chem.* **52**, 5258–5262.

Mittal, A., Balasubramanian, R., Cao, J., *et al.* (2014) TOPOISOMERASE 6B is involved in chromatin remodelling associated with control of carbon partitioning into secondary metabolites and cell walls, and epidermal morphogenesis in Arabidopsis. *J. Exp. Bot.* **65**, 4217–4239.

Orlando, D.A., Brady, S.M., Koch, J.D., Dinneny, J.R. and Benfey, P.N. (2009) Manipulating large-scale Arabidopsis microarray expression data: identifying dominant expression patterns and biological process enrichment. *Methods Mol. Biol.* **553**, 57–77.

Patel, R.K. and Jain, M. (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE*, **7**, e30619.

de la Paz Sanchez, M. and Gutierrez, C. (2009) Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. *Proc. Natl Acad. Sci. USA*, **106**, 2065–2070.

Penfield, S. and Hall, A. (2009) A role for multiple circadian clock genes in the response to signals that break seed dormancy in Arabidopsis. *Plant Cell*, **21**, 1722–1732.

Perry, S.E., Lehti, M.D. and Fernandez, D.E. (1999) The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. *Plant Physiol.* **120**, 121–130.

Redekar, N.R., Biyashev, R.M., Jensen, R.V., Helm, R.F., Grabau, E.A. and Maroof, M.A. (2015) Genome-wide transcriptome analyses of developing seeds from low and normal phytic acid soybean lines. *BMC Genomics*, **16**, 1074.

Rossignol, P., Stevens, R., Perennes, C., Jasinski, S., Cella, R., Tremousaygue, D. and Bergounioux, C. (2002) AtE2F-a and AtDP-a, members of the E2F family of transcription factors, induce Arabidopsis leaf cells to re-enter S phase. *Mol. Genet. Genomics*, **266**, 995–1003.

Ruuska, S.A., Girke, T., Benning, C. and Ohlrogge, J.B. (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell*, **14**, 1191–1206.

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M. and Lepiniec, L. (2008) Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant J.* **54**, 608–620.

Saxena, M.S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., Bansal, K.C., Tyagi, A.K. and Parida, S.K. (2014) An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* **21**, 695–710.

Schruff, M.C., Spielman, M., Tiwari, S., Adams, S., Fenby, N. and Scott, R.J. (2006) The *AUXIN RESPONSE FACTOR 2* gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development*, **133**, 251–261.

Sekhon, R.S., Hirsch, C.N., Childs, K.L., Breitzman, M.W., Kell, P., Duvick, S., Spalding, E.P., Buell, C.R., de Leon, N. and Kaeppler, S.M. (2014) Phenotypic and transcriptional analysis of divergently selected maize populations reveals the role of developmental timing in seed size determination. *Plant Physiol.* **165**, 658–669.

Singh, V.K., Garg, R. and Jain, M. (2013) A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. *Plant Biotechnol. J.* **11**, 691–701.

Singh, V.K., Khan, A.W., Jaganathan, D., *et al.* (2016) QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea. *Plant Biotechnol. J.* **14**, 2110–2119.

Sozzani, R., Maggio, C., Varotto, S., Canova, S., Bergounioux, C., Albani, D. and Cella, R. (2006) Interplay between Arabidopsis activating factors E2Fb and E2Fa in cell cycle progression and development. *Plant Physiol.* **140**, 1355–1366.

Sreenivasulu, N. and Wobus, U. (2013) Seed-development programs: a systems biology-based comparison between dicots and monocots. *Annu. Rev. Plant Biol.* **64**, 189–217.

Stork, J., Harris, D., Griffiths, J., Williams, B., Beisson, F., Li-Beisson, Y., Mendu, V., Haughn, G. and Debolt, S. (2010) CELLULOSE SYNTHASE9 serves a nonredundant role in secondary cell wall synthesis in Arabidopsis epidermal testa cells. *Plant Physiol.* **153**, 580–589.

Sun, X., Shantharaj, D., Kang, X. and Ni, M. (2010) Transcriptional and hormonal signaling control of Arabidopsis seed development. *Curr. Opin. Plant Biol.* **13**, 611–620.

Tang, W. and Perry, S.E. (2003) Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. *J. Biol. Chem.* **278**, 28154–28159.

To, A., Joubès, J., Barthole, G., Lécureuil, A., Scagnelli, A., Jasinski, S., Lepiniec, L. and Baud, S. (2012) WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in Arabidopsis. *Plant Cell*, **12**, 5007–5023.

Van Son, L., Tiedemann, J., Rutten, T., Hillmer, S., Hinz, G., Zank, T., Manteuffel, R. and Bäumlein, H. (2009) The BURP domain protein AtUSPL1 of *Arabidopsis thaliana* is destined to the protein storage vacuoles and overexpression of the cognate gene distorts seed development. *Plant Mol. Biol.* **71**, 319–329.

Vandepoele, K., Vlieghe, K., Florquin, K., Hennig, L., Beemster, G.T., Gruissem, W., Van de Peer, Y., Inzé, D. and De Veylder, L. (2005) Genome-wide identification of potential plant E2F target genes. *Plant Physiol.* **139**, 316–328.

Varshney, R.K., Song, C., Saxena, R.K., *et al.* (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* **31**, 240–246.

Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I. and Righetti, K. (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiol.* **163**, 757–774.

Verma, S., Gupta, S., Bandhiwal, N., Kumar, T., Bharadwaj, C. and Bhatia, S. (2015) High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using Genotyping-by-Sequencing (GBS). *Sci. Rep.* **5**, 17512.

Volkening, M. and Hoffmann, I. (2005) Involvement of human MCM8 in prereplication complex assembly by recruiting hcdc6 to chromatin. *Mol. Cell. Biol.* **25**, 1560–1568.

Wang, X., Zhou, W., Lu, Z., Ouyang, Y., O, C.S. and Yao, J. (2015) A lipid transfer protein, OsLTPL36, is essential for seed development and seed quality in rice. *Plant Sci.* **239**, 200–208.

Weber, H., Borisjuk, L. and Wobus, U. (1996) Controlling seed development and seed size in *Vicia faba*: a role for seed coat-associated invertases and carbohydrate state. *Plant J.* **10**, 823–834.

Weber, H., Borisjuk, L. and Wobus, U. (2005) Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* **56**, 253–279.

Xiao, W., Brown, R.C., Lemmon, B.E., Harada, J.J., Goldberg, R.B. and Fischer, R.L. (2006) Regulation of seed size by hypomethylation of maternal and paternal genomes. *Plant Physiol.* **142**, 1160–1168.

Yin, Y., Cheong, H., Friedrichsen, D., Zhao, Y., Hu, J., Mora-Garcia, S. and Chory, J. (2002) A crucial role for the putative Arabidopsis topoisomerase VI in plant growth and development. *Proc. Natl Acad. Sci. USA*, **99**, 10191–10196.

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

Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, e17.

Zhang, K., Wang, J., Zhang, L., *et al.* (2013) Association analysis of genomic loci important for grain weight control in elite common wheat varieties cultivated with variable water and fertiliser supply. *PLoS ONE*, **8**, e57853.

Zhu, Y., Dong, A., Meyer, D., Pichon, O. and Renou, J.P. (2006) Arabidopsis NRP1 and NRP2 encode histone chaperones and are required for maintaining postembryonic root growth. *Plant Cell*, **18**, 2879–2892.

