

SCIENTIFIC REPORTS



OPEN

Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development

Niraj Khemka¹, Vikash Kumar Singh², Rohini Garg² & Mukesh Jain^{1,2}

Received: 19 May 2016
 Accepted: 22 August 2016
 Published: 15 September 2016

Non-coding RNAs constitute a major portion of the transcriptome in most of eukaryotes. Long non-coding transcripts originating from the DNA segment present between the protein coding genes are termed as long intergenic non-coding RNAs (lincRNAs). Several evidences suggest the role of lincRNAs in regulation of various biological processes. In this study, we identified a total of 2248 lincRNAs in chickpea using RNA-seq data from eight successive stages of flower development and three vegetative tissues via an optimized pipeline. Different characteristic features of lincRNAs were studied and compared with those of predicted mRNAs in chickpea. Further, we utilized a method using network propagation algorithm to reveal the putative function of lincRNAs in plants. In total, at least 79% of the identified chickpea lincRNAs were assigned with a putative function. A comprehensive expression profiling revealed differential expression patterns and tissue specificity of lincRNAs in different stages of flower development in chickpea. In addition, potential lincRNAs-miRNA interactions were explored for the predicted lincRNAs in chickpea. These findings will pave the way for understanding the role of lincRNAs in the regulatory mechanism underlying flower development in chickpea and other legumes.

Recent study of transcriptional landscape in human revealed that most of the genome sequence can transcribe into RNAs and have some biological role¹. The identification and characterization of all non-coding transcripts along with coding transcripts is indispensable for the comprehensive understanding of biological processes and their mechanisms. Based on length, non-coding transcripts can be classified into small and long non-coding RNAs. Long intergenic/intervening non-coding (linc) RNAs represent a class of long non-coding transcripts transcribed from the intergenic regions of protein coding genes^{2,3} with minimum length of 200 bp and lack coding potential. Thousands of lincRNAs have been identified in animals and plants^{2,4-7}. A few studies have revealed involvement of lincRNAs in gene regulation by wide variety of mechanisms at transcriptional and post-transcriptional level, such as formation of chromatin modifying complexes, attachment with promoter binding complex and acting as miRNA sponges to protect mRNAs from degradation⁸⁻¹⁰. Several evidences suggest that lincRNA participates in different biological processes, such as cell cycle regulation, embryonic development, cell pluripotency and aging in animals^{2,8,11}.

In contrast to extensive investigation of lincRNAs in animals, only a few such studies have been performed in plants. Recently, lincRNAs have been predicted in different biological contexts in quite a few plant species^{6,7,12-14}. A few studies have provided evidence for the role of lincRNAs in stress response, male sterility and phosphate homeostasis in plants^{13,15,16}. For example, lincRNAs, *COOLAIR* and *COLDIAIR*, were found to repress the expression of *FLOWERING LOCUS C (FLC)* gene through an epigenetic silencing mechanism, thereby regulating flowering in *Arabidopsis*¹⁷. *Mt4* lincRNA in *Medicago truncatula* was induced by phosphate starvation and acted as decoy of miRNA-399 to allow the accumulation of *PHO2* mRNA, which is important for phosphate uptake¹⁶. *LDMAR (Long-Day Specific Male fertility Associated RNA)* was found to be important for normal pollen development in rice under long day conditions^{13,18}. Due to recent demonstration of their important biological roles in various aspects of development, lincRNAs have become the focus of plant biology research.

¹School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi-110067, India. ²National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi-110067, India. Correspondence and requests for materials should be addressed to M.J. (email: mjain@jnu.ac.in)

Chickpea (*Cicer arietinum*) is second most widely grown legume and is an important source of dietary proteins and fibers. High nutritional and economical value makes it extremely important for food security, especially in developing countries. The availability of draft genome and transcriptome sequences of chickpea provides an opportunity to perform advanced genomic research^{19–21}. Further, transcriptome analysis have been performed to understand the molecular basis of some agronomic traits, such as flower development and abiotic stress responses^{13,22,23}. Recently, small non-coding RNA component (miRNAs) in various tissues/organs have also been characterized in chickpea^{24,25}. However, the discovery and annotation of lincRNAs still remains obscure in chickpea as of now.

In this study, we identified lincRNAs from 11 different tissue samples using RNA-seq data in chickpea. Different properties of lincRNAs and their comparative analysis with coding RNAs were carried out. The differential expression patterns and tissue specificity of chickpea lincRNAs was established. The functional annotation of identified lincRNAs was performed using network propagation algorithm. Furthermore, the interactions of identified lincRNAs and chickpea miRNAs were analyzed to reveal the potential regulatory aspects of lincRNAs. Our study provides the first systematic study of lincRNAs in chickpea and lay the foundation for further studies to elucidate their precise functions.

Results and Discussion

Genome-wide discovery of lincRNAs in chickpea. High-throughput sequencing technologies are the powerful tools to detect novel transcripts quantitatively. These transcripts contain both protein coding and non-coding portion. LincRNAs represent a class of non-coding transcripts with minimum length of 200 nt originating from intergenic region. For comprehensive identification of lincRNAs in chickpea, RNA-seq data (>350 million reads) obtained from 11 different tissues samples²⁶. The samples including three vegetative tissues [germinating seedling (GS), young leaves (YL) and shoot apical meristem (SAM)] and eight stages of flower development [Flower bud (FB1–FB4) and Flower (FL1–FL4)], were utilized (Supplementary Table S1). The RNA-seq reads for each sample were processed via TopHat2 and Cufflinks programs using chickpea (Kabuli CDC frontier) genome as reference. The transcriptome assembly obtained after Cuffmerge was used to retain a set of unique and non-overlapping transcripts. This resulted in a total of 32984 transcripts with an average length of 1435 bp. From the total transcripts, putative intergenic transcripts (5782) were extracted and subjected to an optimized pipeline (Fig. 1) to identify lincRNAs. The pipeline included different filtering criteria for lincRNAs identification. The transcripts with length ≤ 200 bp and putative open reading frame (ORF) length of > 300 bp were removed as transcripts having ORF length > 300 bp might encode for a functional protein. This resulted in a total of 3707 chickpea transcripts, which were subjected to BLAST search in SwissProt and hmmscan in Pfam. At least, 633 and 23 transcripts shared sequence similarity with annotated proteins and protein domains, respectively, available in these databases and hence discarded. In the next step, coding potential of the remaining 3051 transcripts was predicted using Coding Potential Calculator (CPC). A total of 177 transcripts showing CPC score > 0 and additional 626 transcripts present within 500 bp flanking regions of the predicted coding genes were removed. Finally, a total of 2248 transcripts were retained as putative chickpea lincRNAs and designated as Ca_linc_0001 to Ca_linc_2248 (Supplementary Data 1).

Characteristic features of chickpea lincRNAs. We investigated different features of identified lincRNAs (2248) and compared them with those of mRNAs in chickpea. Both, lincRNAs and mRNAs showed similar distribution pattern across different chromosomes on the chickpea genome (Fig. 2A). The length of lincRNAs varied from 201 bp to 5663 bp with majority ($> 88\%$) of them having length ≤ 1000 bp (Fig. 2B). The mean length of lincRNAs was 614 bp, which was quite less as compared to mean length of mRNAs (1495 bp). The lesser mean length of predicted chickpea lincRNAs was similar to that reported earlier in other plants, such as rice (800 bp), cucumber (322 bp) and *Ganoderma* (609 bp)^{6,7,27}. However, the mean exon length of lincRNAs (395 bp) was substantially higher as compared to mRNAs (244 bp) in chickpea. The presence of more number of exons might have resulted in lesser exon length in mRNAs. Similar observation has been reported in rice with mean exon length of lincRNAs and mRNAs being 322 and 159 bp, respectively⁶. The majority (81%) of lincRNAs in chickpea were single exonic (Fig. 2C), consistent with earlier reports in maize (81%) and cucumber (89%)^{7,14}. The average exon count of lincRNA was 1.22 as compared to 4.97 for mRNAs in chickpea. The chickpea lincRNAs transcripts were AU rich as compared to mRNAs (Fig. 2D) similar to the predicted lincRNAs in Arabidopsis, cucumber, rice and maize^{5–7}. Various characteristic features, including genomic location, length, number of exons and GC content of all the chickpea lincRNAs have been given in Supplementary Table S2.

Differential expression and tissue specificity of chickpea lincRNAs. To investigate the putative function of chickpea lincRNAs, their expression patterns were explored in all the 11 tissues of chickpea (GS, SAM, YL, FB1–FB4 and FL1–FL4) using RNA-seq data. For estimation of expression level, number of fragments per kilobase of transcripts per million mapped reads (FPKM) for each lincRNA was determined using Cufflinks (Supplementary Table S2). The number of lincRNAs expressed in each tissue varied from 1267 in young leaves (YL) to 1779 in flower bud (FB4) (Supplementary Fig. S1). The number of expressed lincRNAs was higher in tissues representing actively dividing cells (SAM and GS) and stages of flower development. Earlier report of lincRNAs in different plant species also showed that reproductive tissues contain higher number of lincRNAs as compared to other tissues^{5–7,14,28}. Based on the expression level, lincRNAs were divided into five classes, very low (< 2 FPKM) to very high (> 20 FPKM) (Supplementary Fig. S1). The expression profile showed that a large number of lincRNAs were expressed at very low level, while smaller number of lincRNAs expressed at higher expression level (> 5 FPKM). The expression pattern of chickpea lincRNAs showed similarity with that of lincRNAs identified in different plants, such as Arabidopsis, cucumber, maize and rice^{6,7,14}.

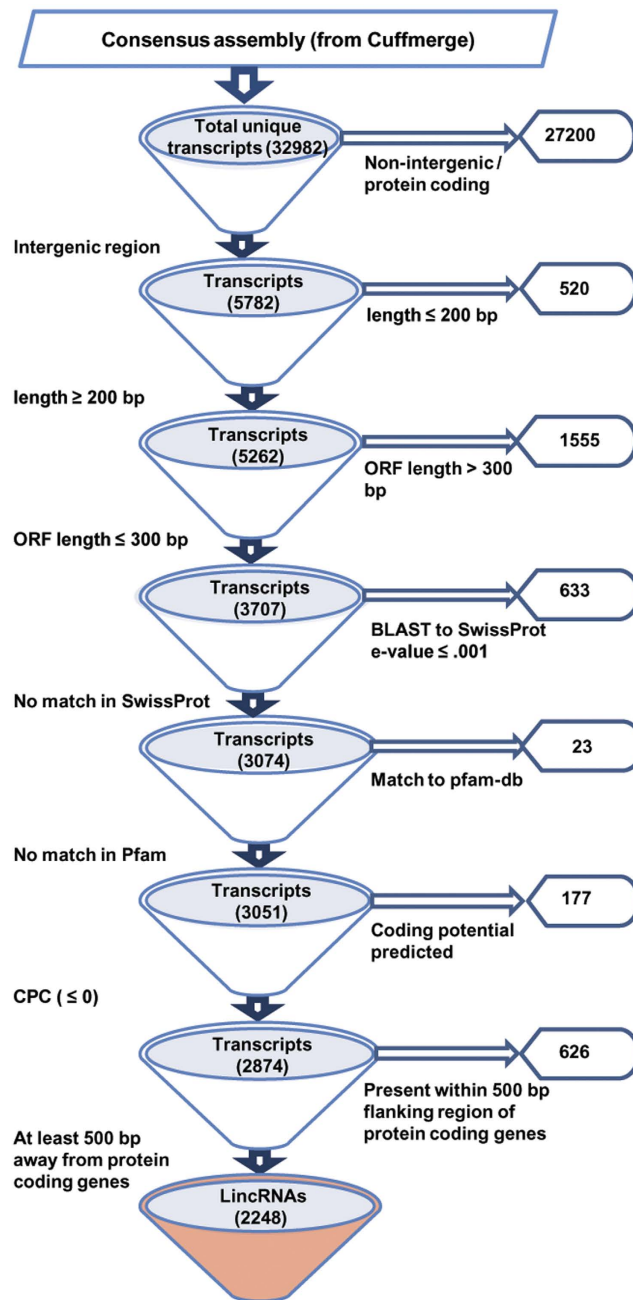


Figure 1. Strategy used for identification of lincRNAs in chickpea. All the steps followed for the identification of lincRNAs in chickpea are shown in the form of a flowchart. Numbers represent number of transcripts.

To study differential expression pattern and tissue specificity of lincRNAs, we calculated the tissue specificity index (TSI) for all the chickpea lincRNAs. The TSI differentiates between housekeeping genes and tissue-specific genes. The TSI value of 0 represents housekeeping genes due to similar expression in all tissues, whereas TSI value of 1 represents strictly tissue-specific expression. Chickpea lincRNAs having TSI score ≥ 0.9 were considered as tissue-specific. Applying this stringent criterion, at least 641 lincRNAs (29%) were identified as tissue-specific (Fig. 3A). A heatmap depicting expression pattern of tissue-specific lincRNAs has been shown (Fig. 3B). The largest number of tissue-specific lincRNAs were detected in SAM (196) followed by GS (117). This may be because, these tissues can act as primordial cells and various tissues originate from these tissues, which suggests putative role of lincRNAs in tissue specification^{29,30}. At least 13% of total lincRNAs showed tissue-specific expression during various stages of flower development (Fig. 3A). A comparative analysis revealed that a larger fraction (0.2–9%) of lincRNAs exhibited stage-specific expression (present study), as compared to the protein coding mRNAs²⁶ (0.04–1.5%). Different reports suggest that lincRNAs expressed in a specific tissue can perform specialized functions. For example, functional analysis of lincRNA, *XLOC_057324*, expressed specifically in panicle and pistils of rice showed decreased fertility and panicle development in mutant rice⁶. Likewise, lincRNA *Zm401*, which is

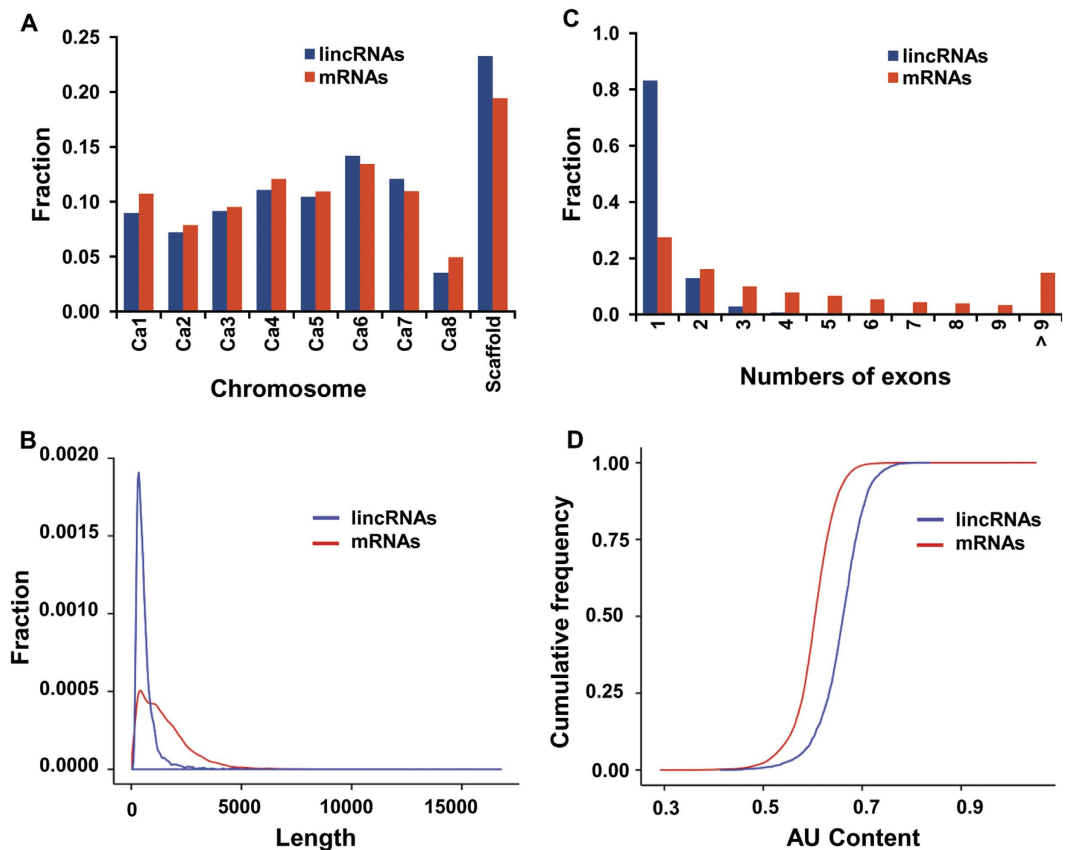


Figure 2. Features of chickpea lincRNAs and comparative analysis with mRNAs. (A) Chromosome-wise distribution of lincRNAs and mRNAs. (B) Length distribution of lincRNAs and mRNAs transcripts. (C) Distribution of numbers of exons in lincRNA and mRNA transcripts. (D) Cumulative frequency of AU in lincRNA and mRNA transcripts.

specifically expressed in anther, was found to be associated with male sterility in maize¹⁸. The higher expression of *IPS1* lincRNA in Arabidopsis shoot was found to be important for phosphate homeostasis¹⁶. A large portion of tissue-specific lincRNAs suggested their role in regulation of important characteristic processes in a particular tissue/developmental stage in chickpea.

To gain more insights into the role of chickpea lincRNAs in flower development and other biological processes, we identified soybean orthologs of chickpea mRNAs co-expressed with tissue/stage (FB, FL and SAM) specific lincRNAs. The top 10 co-expressed mRNAs were extracted from the mRNA-lincRNA co-expression matrix for each set of lincRNAs specific to FB, FL and SAM in chickpea. These mRNAs were searched for their orthologs (BLASTX, E-value < e^{-10}) against the differentially expressed transcripts (2951) reported in the time series RNA-seq data of SAM in soybean³¹. A total of 292, 184 and 117 chickpea mRNAs co-expressed with lincRNAs specifically expressed in FB, FL and SAM, respectively, shared significant similarity with differentially expressed transcripts in soybean (Supplementary Table S3). Many of these orthologs represented well known proteins (FT, AGL6, API, SVP and TCP1) associated with various aspects of flower development. These results strengthen the idea of the crucial role of lincRNAs in flower developmental process in chickpea.

Validation of differential expression patterns. We randomly selected 15 chickpea lincRNAs for validation of their expression patterns using quantitative real time polymerase chain reaction (qRT-PCR). We observed similar expression pattern in qRT-PCR analysis as that of RNA-seq data for the selected lincRNAs (Fig. 3C). Overall, the results of qRT-PCR analysis were in good agreement ($r = 0.72$) with RNA-seq (Fig. 3D). For example, based on RNA-seq data analysis, Ca_linc_0051 and Ca_linc_0139 showed specific expression in FB4 and SAM, respectively. Similar expression patterns were confirmed via qRT-PCR analysis as well (Fig. 3C,D).

Functional annotation of lincRNAs. Several studies have demonstrated that lincRNAs are involved in various biological processes³. Therefore, it is important to annotate the putative function of lincRNAs. Function prediction of lincRNAs using sequence homology based approach is trivial due to non-conservative nature of lincRNAs across different species. Recently, a few methods have been proposed for functional annotation of lincRNAs^{32,33}. To infer the putative function of lincRNAs in chickpea, we adopted a recently proposed method for annotation of lincRNAs using network propagation algorithm³³. This annotation method uses transcriptome and protein-protein interaction (PPI) data for the prediction of putative function of lincRNAs and showed improved accuracy compared to other available methods. We used this network propagation based method to annotate the

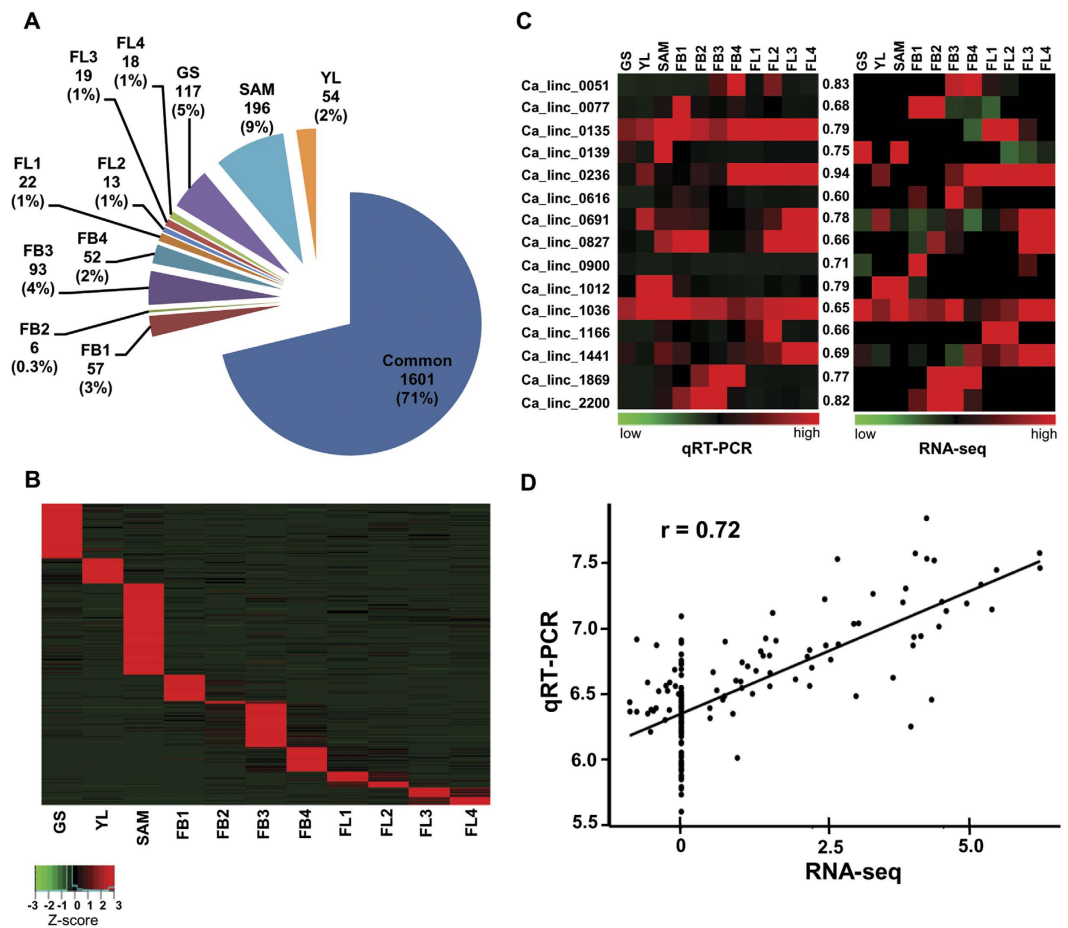


Figure 3. Tissue-specific expression of lincRNAs. (A) Number and percentage of lincRNAs showing specific expression in different tissues. (B) Heatmap of lincRNAs showing specific expression pattern in different tissues. Color scale representing Z-score is given at the bottom. (C) Heatmaps showing expression patterns in qRT-PCR and RNA-seq experiments for 15 lincRNAs. The correlation between the expression pattern obtained by qRT-PCR and RNA-seq for each lincRNA is given in the middle. (D) Overall correlation between qRT-PCR and RNA-seq data. Each dot represents one data point (normalized expression level of a lincRNA in a particular tissue obtained via RNA-seq and qRT-PCR).

putative function of predicted lincRNAs in chickpea. We integrated RNA-seq based expression and PPI data to form the bicolor network. On this bicolor network, global propagation algorithm was applied to infer the putative function (gene ontology) of lincRNAs.

Using the network propagation based annotation method, putative function could be assigned to at least 1790 (79%) of chickpea lincRNAs (Supplementary Table S4). In general, the functional annotation revealed the predominance of GO terms related to different biological processes, such as, photorespiration, regulation of cell cycle, cell morphogenesis, RNA splicing, mRNA processing and histone modification along with a variety of metabolic processes among chickpea lincRNAs (Supplementary Fig. S2). Further, we explored the functional annotation of lincRNAs specifically expressed in flower bud (FB), flower (FL) and shoot apical meristem (SAM) in detail. The enrichment analyses showed that majority of lincRNAs are involved in biological processes in a stage-specific manner. Flower bud specific lincRNAs were found to be enriched in GO terms associated with homeostasis processes, developmental processes and floral whorl development with $p\text{-value} < e^{-10}$ (Fig. 4A,C, Supplementary Fig. S3). The lincRNAs expressed specifically during flower development stages were enriched in GO terms, such as chemical homeostasis, different signaling pathways, developmental process, response to external stimuli and localization with $p\text{-value} < e^{-10}$ (Fig. 4B,D, Supplementary Fig. S3). The SAM-specific lincRNAs were found enriched in GO terms associated with anatomical structure development, secondary metabolite process, regulation of timing of transition from vegetative to reproductive phase, regulation of timing of meristematic phase transition along with gibberellins biosynthetic process and auxin influx with $p\text{-value} < e^{-9}$ (Supplementary Fig. S4). Previous report of coding genes showed that GO terms associated with the reproductive structure and floral whorl development were highly enriched at flower bud (FB1-FB4) stages, whereas GO terms associated with different signaling pathways along with reproduction and post-embryonic development were enriched at flower development stages (FL1-FL4)²⁶. Likewise, reports of SAM specific protein-coding genes in chickpea and soybean also showed that GO terms associated with anatomical structure development, cell differentiation, cell proliferation, floral meristem determinacy, reproductive development process along with GO terms associated

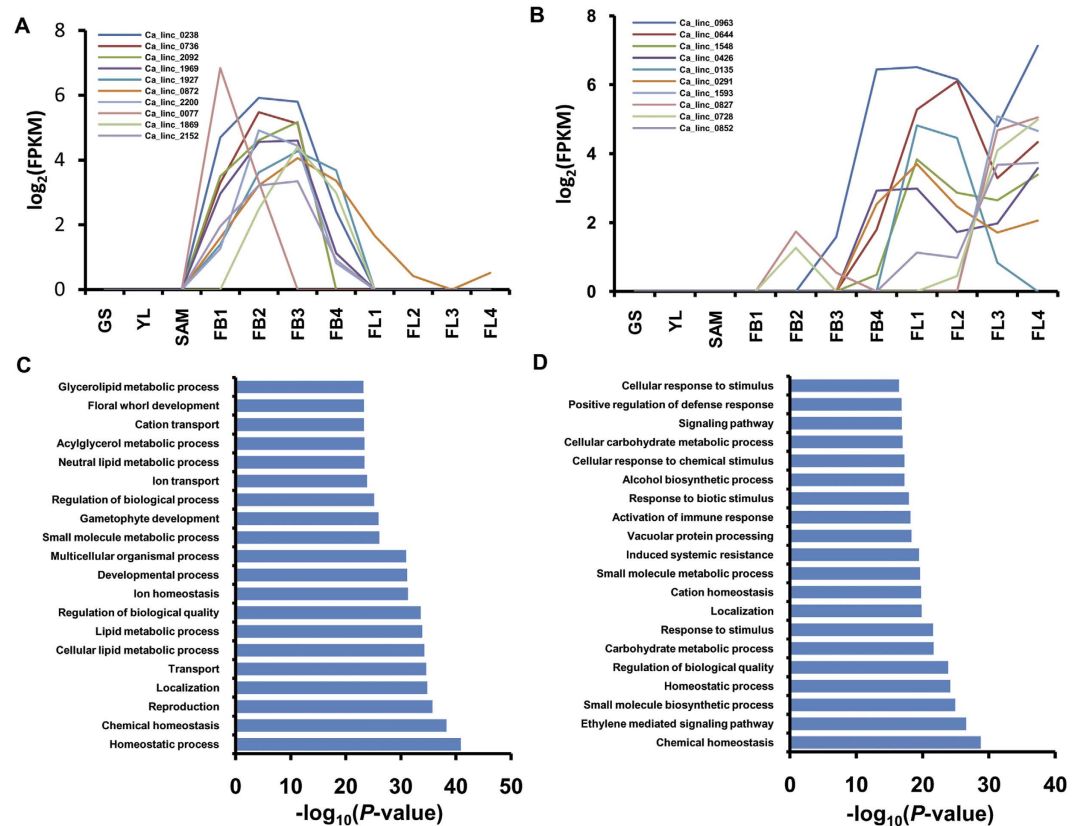


Figure 4. Flower bud and flower specific expression patterns of lincRNAs and GO enrichment analysis. (A,B) Expression pattern of 10 highly expressed lincRNAs in flower bud (A) and flower (B) stages of chickpea. Normalized expression (FPKM) of lincRNAs in different tissues has been presented in the line graph. Expression pattern of all the highly expressed lincRNAs in flower bud and flower stages of chickpea is given in Supplementary Fig S3. (C,D) Gene ontology enrichment for lincRNAs expressed only in flower bud (C) and flower (D) tissues. Top GO terms showing very high enrichment (P -value $< e^{-10}$) are shown.

with gibberellins and auxin hormone were highly enriched^{31,34}. The correlation of functional enrichment between co-expressed coding transcripts and lincRNAs indicates accuracy of our functional annotation method. These results also revealed that lincRNAs are involved in several biological processes occurring at different stages of flower development in chickpea. A few lincRNAs have been characterized for their biological functions in plants. The regulatory role of lincRNAs, such as *COOLAIR*, *LDMAR* and *IPS1* in flowering timing during cold stress, fertility and phosphate homeostasis, respectively, have been validated in different plants^{13,16,17}. The enrichment of GO terms associated with the specific processes with different tissues/stages of development can provide significant enhancement in knowledge and functioning of lincRNAs in plants.

LincRNAs as potential regulator of miRNAs. Recent studies have shown that lincRNAs can act as targets or decoy of miRNAs to mimic mRNAs and therefore regulate gene expression^{35–38}. The lincRNAs interact with miRNAs to regulate biological processes, such as plant growth, differentiation and reproduction³⁶. In this study, predicted lincRNAs of chickpea were used to identify the potential targets of miRNAs using psRNATarget^{24,39}. A total of 144 lincRNAs were identified as putative targets of 236 chickpea miRNAs (Supplementary Table S5). Multiple interaction patterns, such as one-to-one, one-to-many and many-with-many, were recognized. This resulted in a total of 349 lincRNA-miRNA pairs (Fig. 5, Supplementary Fig. S6). Majority of miRNAs involved in interaction with lincRNAs were from miR-156, miR-159, miR-166, miR-167, miR-170, miR-171, miR-172 and miR-319 families. Several reports have suggested regulatory roles of these miRNAs families in flower development in plants^{40–42}. Many of the chickpea lincRNAs were found to be associated with GO terms as that of known function of their interacting miRNAs. For example, miR-159 and miR-319 are highly conserved miRNA families that have important role in plant growth, morphogenesis and reproduction⁴³. Members of these families, miR-159c and miR-319c, were found to target *Ca_linc_1666* (Fig. 6A). The lincRNA, *Ca_linc_1666*, was associated with GO terms, such as specification of petal number (GO:0048834) and male meiosis cytokinesis (GO:0007112). Similarly, *Ca_linc_0964* harbor binding sites of miR-159-a/b and miR-408b (Fig. 6B). Earlier studies revealed that miR-408 family miRNAs are associated with stress response in chickpea⁴⁴. The GO terms associated with lincRNA, *Ca_linc_0964*, included reproduction related GO terms (seed development (GO:0048316), DNA replication associated terms (GO:0006260, GO:0006261 and GO:0006275)) and stress related GO terms (response to

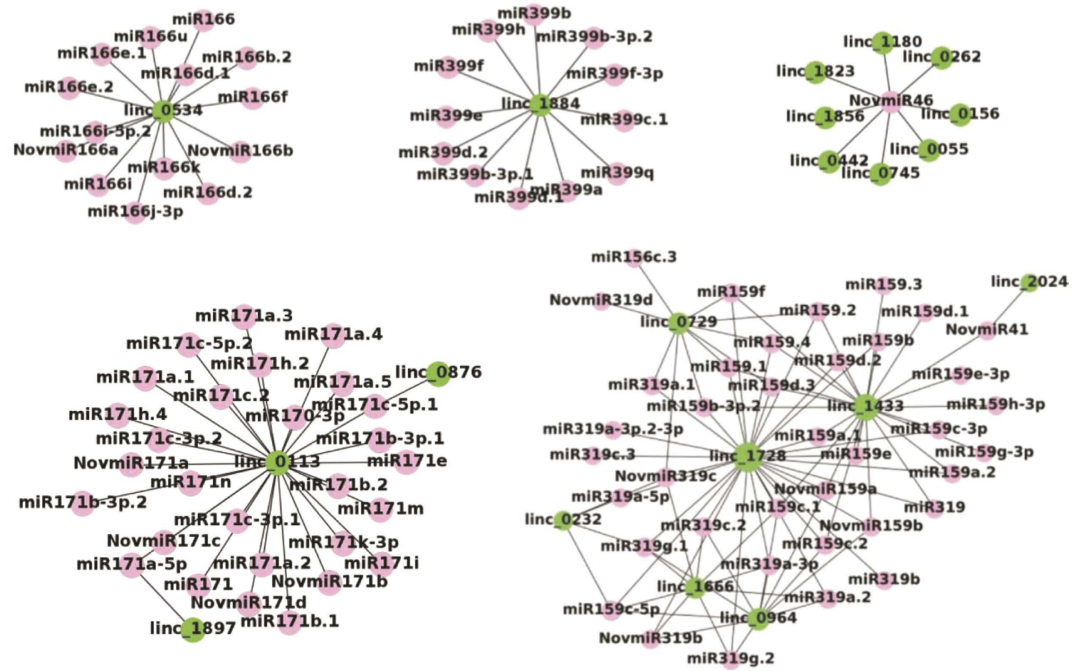


Figure 5. LincRNA - miRNA interaction network. The network includes lincRNAs and miRNAs. Red circle nodes represent miRNAs and green circle nodes represent lincRNAs. Examples of one lincRNA to many miRNAs, many lincRNAs to one miRNA, and many lincRNAs to many miRNAs interactions are shown. The complete network of chickpea lincRNAs and miRNAs has been given in Supplementary Fig. S5.

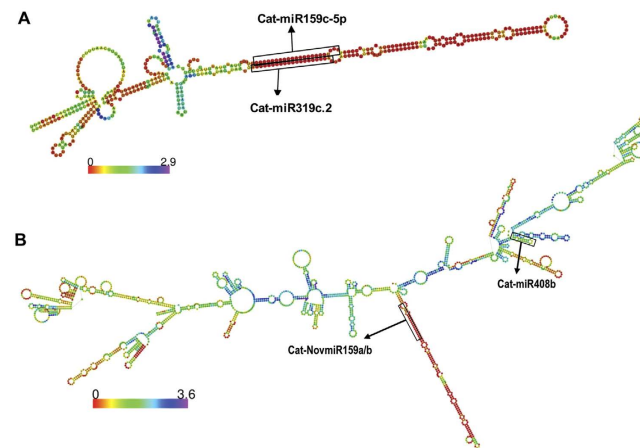


Figure 6. LincRNA – miRNA binding site prediction. Representative examples showing the secondary structure of lincRNAs with binding sites of miRNAs are given. Color coding shows the positional entropy of sequence (A) Secondary structure of Ca_linc_1666 with binding sites of Cat-miR159c-5p and Cat-miR319c.2. (B) Secondary structure of Ca_linc_0964 with binding sites of Cat-miR159a/b and Cat-miR408b.

UV-B (GO:0010224) and defense response (GO:0009814)). These examples suggested that function of lincRNAs can also be assigned by knowing their interacting miRNAs and their function.

Some of the miRNAs that target lincRNAs were found to be conserved in soybean. The orthologs of at least 12 SAM related soybean miRNAs reported earlier⁴² were found to target 44 lincRNAs in chickpea (Supplementary Table S5). For example, miR319a and miR159a targeted SAM-specific chickpea lincRNA, Ca_linc_1728.

Earlier reports also showed role of lincRNA-miRNA interactions in different biological functions. In Arabidopsis, it has been demonstrated that miR-399 targets *IPS1* lincRNA instead of *PSO4* mRNA to increase phosphate uptake during phosphate starvation¹⁶. Likewise, miR-160 and miR-164 were found to cleave lincRNAs associated with reproductive processes in rice⁶. Genome-wide analysis of degradome data in maize identified 166 lincRNAs as targets of 165 miRNAs, while 34 lincRNAs acted as decoy of 33 miRNAs³⁵. The interaction of miRNAs with their target lincRNAs suggests that the lincRNAs can disrupt miRNA-mRNA interactions to regulate gene

expression and biological processes. Overall, our analysis provides evidence for interaction of lincRNAs and miRNAs in chickpea and suggests their putative roles in various aspects of flower development.

Conclusions

In this study, a total of 2248 lincRNAs were identified in chickpea using RNA-seq data following stringent criteria. At least 1790 lincRNAs (79%) could be assigned with a putative function. The expression profiling revealed tissue-specific expression of a large number of lincRNAs in distinct tissues. Several chickpea lincRNAs were identified as targets of miRNAs and implicated them in the regulatory network of various developmental and reproductive processes. This study will provide an extensive resource for lincRNAs in chickpea and allow further study to gain insights into regulatory aspects of flower development in chickpea and other legumes. Further experiments focused on individual lincRNA are needed to elucidate their exact function.

Methods

Transcriptome data processing and lincRNA identification pipeline. RNA-seq data (fastq files) of 11 tissues samples, were downloaded from the Gene Expression Omnibus (GEO) database under the accession number GSE42679²⁶. These samples included three vegetative tissues, germinating seedling (GS), young leaves (YL) and shoot apical meristem (SAM), and eight successive stages of flower tissues from closed flower bud (FB1) to drooped flower (FL4) as described²⁶. The RNA-seq data consisted of >350 million high-quality reads (Supplementary Table S1). The reads from each sample were filtered to remove adapters and low-quality reads using NGSQC Tool Kit⁴⁵ followed by mapping on the chickpea (Kabuli CDC frontier) genome¹⁹ using splice read aligner TopHat2²⁹. The transcriptome of each tissue was assembled using Cufflinks and merged together using Cuffmerge to generate the consensus transcriptome assembly.

The pipeline employed for the identification of lincRNAs from RNA-seq data has been described in Fig. 1. Briefly, from total transcripts, intergenic transcripts were extracted using class code “u”. The transcripts having length shorter than 200 bp and overlapping with known mRNAs were discarded from the intergenic transcripts. The ORF length of filtered transcripts was calculated using OrfPredictor webserver⁴⁶ (<http://proteomics.yzu.edu/tools/OrfPredictor.html>). The transcripts having ORF length >300 bp were discarded. The SwissProt⁴⁷ (BLASTX, E-value $\leq e^{0.001}$) and Pfam⁴⁸ databases (hmmscan, E-value $\leq e^{0.001}$) were used to filter transcripts with matched protein and protein domain family, respectively. The remaining transcripts were used for coding potential calculation using Coding Potential Calculator (CPC), (<http://cpc.cbi.pku.edu.cn>)⁴⁹. The transcripts having CPC score of >0 was considered as coding transcript and hence discarded. Transcripts present within the flanking region (500 bp) of coding genes were also removed from the remaining intergenic transcripts, as there may be chances of extension of transcription start site (TSS) for protein coding genes in improved annotation. Different in-house perl scripts were used for analysis.

Tissue specificity index. Tissue specificity of the putative lincRNAs was evaluated according to the tissue-specific index (TSI) as described previously^{7,50}. The TSI was calculated based on the expression value of each lincRNAs in different tissues. The value of TSI of lincRNAs varies from 0 for housekeeping genes to 1 for strict tissue-specific genes. The lincRNAs showing a TSI ≥ 0.9 were considered as tissue-specific lincRNAs.

qRT-PCR analysis. To validate the expression profile of predicted lincRNAs, total RNA was isolated from germinating seedling (GS), shoot apical meristem (SAM), young leaf (YL), flower buds (FB1–FB4; where FB1, FB2, FB3, and FB4 were flower buds of different sizes), and flowers (FL1–FL4; where FL1, young flower having closed petals; FL2, flower with partially opened petals; FL3, mature flower with faded petals; FL4, drooped flower with senescing petals), as described²⁶. Total RNA samples were checked for quality and quantity as described previously²⁶. Gene-specific primers were designed using Primer Express (Version 3.0) software, and real-time PCR analysis was performed using 7500 Fast real time PCR system (Applied Bio-systems) using at least two biological replicates and three technical replicates (for each biological replicate) of each sample as described⁵¹. The expression of *Elongation Factor-1 Alpha (EF1 α)* was used as internal control gene for normalization⁵¹. The primer sequences used for the analysis are provided in Supplementary Table S6. The correlation between expression profile of lincRNAs measured by RNA-seq and qRT-PCR were analyzed using R statistical package.

Functional annotation of lincRNAs. Functional annotation of lincRNAs was performed using network propagation based annotation method³³. In this method, gene co-expression and protein-protein interaction (PPI) data were used. The gene co-expression data and PPI data were integrated for bi-color network formation. Network propagation algorithm was applied on the bi-color network to obtain the annotation. Expression data (FPKM) of coding RNAs and lincRNAs were obtained from the chickpea RNA-seq data using Cufflinks. PPI data and GO annotations (biological process) for chickpea proteins were obtained for their best orthologs in Arabidopsis using String⁵² and TAIR⁵³ databases, respectively. The Spearman correlation coefficient was calculated based on FPKM values for all pair of lincRNA/mRNA. Top 0.5% percentile of the absolute value of correlation coefficient was used to make the co-expression bicolor network. GO terms for all annotated mRNAs were transferred to the bi-colored network. The lincRNA nodes connected with mRNAs were annotated with their first degree neighbor annotation. Further, the PPI score between chickpea proteins were normalized for further processing as described³³. To generate a weighted matrix, transcriptome based correlation matrix was combined with the PPI matrix and scores were normalized using min-max linear normalization algorithm⁵⁴. The global propagation algorithm⁵⁴ was applied to the bi-colored network using weighted matrix for inferring the putative function of lincRNAs as described³³. This process ranked associations between mRNA/lincRNA and all GO terms. All the lincRNAs were ranked according to the association scores with given functional category. The lincRNAs ranked in top 100 for a GO term were annotated with the same GO term for functional annotation (Supplementary Table S4).

GO enrichment analysis. Gene ontology enrichment analysis of lincRNAs was done using Bingo plugin of Cytoscape (version 3.2)⁵⁵. The GO terms with p -value $< e^{-10}$ were considered significant. The lincRNAs which expressed only in FB and FL tissues with FPKM ≥ 3 and no expression in other tissues (FPKM ≤ 0.1) were considered for the GO enrichment. The lincRNAs expressed specifically in SAM (TSI ≥ 0.9) was considered for GO enrichment.

Identification of lincRNAs as target of miRNAs. All the identified lincRNAs were analyzed for miRNA target sites using psRNATarget webserver³⁹. The chickpea miRNAs sequences were taken from Jain *et al.*²⁴. The miRNAs and lincRNAs were submitted to the psRNATarget and interaction were calculated using default parameters (maximum expectation = 3, allowed maximum energy to unpair the target site (UPE) = 25). The interaction network of lincRNAs and miRNAs in chickpea based on the miRNAs binding sites on lincRNAs was drawn using Gephi⁵⁶ and R statistical package⁵⁷.

References

- Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
- Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
- Ulitsky, I. & Bartel, D. P. LincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
- Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
- Liu, J. *et al.* Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell* **24**, 4333–4345 (2012).
- Zhang, Y.-C. *et al.* Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* **15**, 512 (2014).
- Hao, Z. *et al.* Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in cucumber. *PLoS One* **10**, e0121800 (2015).
- Cabili, M. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
- Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
- Imig, J. *et al.* miR-CLIP capture of a miRNA targetome uncovers a lincRNA H19-miR-106a interaction. *Nat. Chem. Biol.* **11**, 107–114 (2015).
- Essers, P. B. *et al.* A long noncoding RNA on the ribosome is required for lifespan extension. *Cell Rep.* **10**, 339–345 (2015).
- Shuai, P. *et al.* Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J. Exp. Bot.* **65**, 4975–4983 (2014).
- Ding, J. *et al.* A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci.* **109**, 2654–2659 (2012).
- Li, L. *et al.* Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* **15**, R40 (2014).
- Wunderlich, M., Gross-Hardt, R. & Schöffl, F. Heat shock factor HSF2a involved in gametophyte development of Arabidopsis thaliana and its expression is controlled by a heat-inducible long non-coding antisense RNA. *Plant Mol. Biol.* **85**, 541–550 (2014).
- Franco-Zorrilla, J. M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**, 1033–1037 (2007).
- Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**, 76–79 (2011).
- Ma, J. *et al.* Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize. *J. Cell. Biochem.* **105**, 136–146 (2008).
- Varshney, R. K. *et al.* Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* **31**, 240–246 (2013).
- Jain, M. *et al.* A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* **74**, 715–729 (2013).
- Garg, R. *et al.* Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiol.* **156**, 1661–1678 (2011).
- Yao, Y. *et al.* Non-coding small RNAs responsive to abiotic stress in wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics* **10**, 187–190 (2010).
- Wang, H. *et al.* Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lincRNAs associated with agriculture traits. *Plant J.* **84**, 404–416 (2015).
- Jain, M., Chevala, V. V. S. N. & Garg, R. Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. *J. Exp. Bot.* **65**, 5945–5958 (2014).
- Srivastava, S. *et al.* High throughput sequencing of small RNA component of leaves and inflorescence revealed conserved and novel miRNAs as well as phasiRNA loci in chickpea. *Plant Sci.* **235**, 46–57 (2015).
- Singh, V. K., Garg, R. & Jain, M. A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. *Plant Biotechnol. J.* **11**, 691–701 (2013).
- Li, J., Wu, B., Xu, J. & Liu, C. Genome-wide identification and characterization of long intergenic non-coding RNAs in *Ganoderma lucidum*. *PLoS One* **9**, e99442 (2014).
- Kang, C. & Liu, Z. Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. *BMC Genomics* **16**, 815 (2015).
- Heisler, M. G. *et al.* Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the Arabidopsis inflorescence meristem. *Curr. Biol.* **15**, 1899–1911 (2005).
- Fatica, A. & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* **15**, 7–21 (2014).
- Wong, C. E., Singh, M. B. & Bhalla, P. L. The Dynamics of soybean leaf and shoot apical meristem Transcriptome undergoing floral initiation process. *PLoS One* **8**, e65319 (2013).
- Mattick, J. S. & Rinn, J. L. Discovery and annotation of long noncoding RNAs. *Nat. Struct. Mol. Biol.* **22**, 5–7 (2015).
- Guo, X. *et al.* Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic Acids Res.* **41**, e35 (2013).
- Singh, V. K. & Jain, M. Transcriptome profiling for discovery of genes involved in shoot apical meristem and flower development. *Genomics data* **2**, 135–138 (2014).
- Fan, C., Hao, Z., Yan, J. & Li, G. Genome-wide identification and functional analysis of lincRNAs acting as miRNA targets or decoys in maize. *BMC Genomics* **16**, 793 (2015).
- Gupta, P. K. Competing endogenous RNA (ceRNA): a new class of RNA working as miRNA sponges. *Current Science* 823–830 (2014).

37. Johnsson, P. *et al.* A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat. Struct. Mol. Biol.* **20**, 440–446 (2013).
38. Paraskevopoulou, M. D. *et al.* DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res.* **41**, D239–D245 (2013).
39. Dai, X. & Zhao, P. X. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* **39**, W155–W159 (2011).
40. Schwab, R. Roles of miR156 and miR172 in reproductive development in *MicroRNAs in Plant Development and Stress Responses* (ed. Sunkar, R.) 69–81 (Springer Berlin Heidelberg, 2012).
41. Huijser, P. & Schmid, M. The control of developmental phase transitions in plants. *Development* **138**, 4117–4129 (2011).
42. Wong, C. E. *et al.* MicroRNAs in the shoot apical meristem of soybean. *J. Exp. Bot.* **62**, 2495–2506 (2011).
43. Palatnik, J. F. *et al.* Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. *Dev. Cell* **13**, 115–125 (2007).
44. Hajyzadeh, M., Turktas, M., Khawar, K. M. & Unver, T. miR408 overexpression causes increased drought tolerance in chickpea. *Gene* **555**, 186–193 (2015).
45. Patel, R. K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7**, e30619 (2012).
46. Min, X. J., Butler, G., Storms, R. & Tsang, A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* **33**, W677–W680 (2005).
47. Magrane, M. & Consortium, U. UniProt Knowledgebase: a hub of integrated protein data. *Database* **2011** bar009 (2011).
48. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
49. Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349 (2007).
50. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
51. Garg, R., Sahoo, A., Tyagi, A. K. & Jain, M. Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). *Biochem. Biophys. Res. Commun.* **396**, 283–288 (2010).
52. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568 (2011).
53. Lamesch, P. *et al.* The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).
54. Vanunu, O., Mager, O., Rupp, E., Shlomi, T. & Sharan, R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.* **6**, e1000641 (2010).
55. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).
56. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An open source software for exploring and manipulating networks. *Third Int. AAAI Conf. Weblogs Soc. Media* 361–362 (2009).
57. Ihaka, R. & Gentleman, R. R. A language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314 (1996).

Acknowledgements

This work was financially supported by the Department of Biotechnology, Government of India, New Delhi, under the Challenge Programme on Chickpea Functional Genomics (BT/AGR/CG-PhaseII/01/2014) and the Department of Science & Technology, Government of India, under the Promotion of University Research and Scientific Excellence (PURSE) grant (Phase II) scheme.

Author Contributions

M.J. conceived and designed the study. N.K. and V.K.S. collected the tissue material, performed data analysis and participated in compiling and interpretation of data. M.J., N.K., V.K.S. and R.G. compiled and interpreted the data, and wrote the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Khemka, N. *et al.* Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development. *Sci. Rep.* **6**, 33297; doi: 10.1038/srep33297 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016