ORIGINAL RESEARCH

Long Non-coding RNA Derived from lncRNA–mRNA Co-expression Networks Modulates the Locust Phase Change

Ting Li [1,2], Bing Chen [3], Pengcheng Yang [1], Depin Wang [2], Baozhen Du [1]
Le Kang [1,2,3,4,*]

[1] *Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China*
[2] *CAS Center for Excellence in Biotic Interactions, University of Chinese Academy of Sciences, Beijing 100049, China*
[3] *College of Life Sciences, Hebei University, Baoding 071002, China*
[4] *State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China*

Abstract Long non-coding RNAs (lncRNAs) regulate various biological processes ranging from gene expression to animal **behavior**. Although protein-coding genes, microRNAs, and neuropeptides play important roles in the regulation of **phenotypic plasticity** in migratory locust, empirical studies on the function of lncRNAs in this process remain limited. Here, we applied high-throughput RNA-seq to compare the expression patterns of lncRNAs and mRNAs in the time course of locust phase change. We found that lncRNAs responded more rapidly at the early stages of phase transition. Functional annotations demonstrated that early changed lncRNAs employed different pathways in isolation and crowding phases to cope with changes in the population density. Two overlapping **hub lncRNA** loci in the crowding and isolation networks were screened for functional verification. One of them, LNC1010057, was validated as a potential regulator of locust phase change. This work offers insights into the molecular mechanism underlying locust phase change and expands the scope of lncRNA functions in animal behavior.

Introduction

Long non-coding RNAs (lncRNAs), which represent the largest class of ncRNAs, have a length of more than 200 nucleotides but lack coding potential [1]. Similar to mRNAs, lncRNAs are transcribed by RNA polymerase II, and can be capped, polyadenylated, and spliced. However, the average

* Corresponding author.
  E-mail: lkang@ioz.ac.cn (Kang L).

expression level and sequence conservation of lncRNAs are lower than those of mRNAs; as such, lncRNAs are more difficult to detect and annotate through conserved sequences [2,3]. With the development of high-throughput deep sequencing technologies, thousands of lncRNAs have been identified in many species and confirmed as crucial regulators of biological processes instead of byproducts of transcription [4–6]. The expression levels of protein-coding genes can be regulated by lncRNAs in a variety of biological processes, including mRNA transcription, stability, translation, and posttranslational modification [7]. lncRNAs modulate the expression of protein-coding genes through *cis*-acting on neighboring genes and *trans*-acting on distal genes [8]. Thus, the expression level of lncRNAs is correlated with that of potential target genes, and the function of lncRNAs can be predicted by the associated mRNAs in lncRNA–mRNA co-expression networks [9]. However, in different biological processes, the expression profiles of lncRNAs and mRNAs follow concordant or discordant patterns, suggesting diverse interaction patterns between lncRNAs and mRNAs [10].

lncRNA expression displays tissue specificity and is especially abundant in the nervous system [11]. Increasing lines of evidence suggest that lncRNAs play roles in the nervous system development and function [12,13]. lncRNAs can serve as vital regulators in modulating the behavior of mammals and insects. For example, upregulation of *MEG3* expression improves the spatial learning and memory capability of rats with Alzheimer's disease by inactivating the PI3K/Akt signaling pathway [14]. In addition, downregulation of *Gomafu* expression induces anxiety-like behavior in mice by regulating the expression of schizophrenia-related genes [15]. In *Drosophila*, lncRNA *CRG* maintains normal locomotor and climbing capabilities by regulating the transcription of the neighboring *CASK* gene, and *yar*, another lncRNA, regulates the night sleep time [16,17]. Nonetheless, roles of lncRNAs in regulating animal behavior remain a fascinating and growing research field for scientific investigation.

Although less studied than in mammals and *Drosophila*, lncRNAs have been identified in non-model insect species [18,19]. Previous studies on expression and functional annotation of lncRNAs revealed the involvement of insect lncRNAs in biological processes, such as insecticide resistance, fecundity, and gland apoptosis [19–21]. In particular, four lncRNAs (*i.e., Ks-1, AncR-1, kakusei,* and *Nb-1*) in *Apis mellifera* are expressed preferentially in the brain and are related to social behavior [22]. However, no specific lncRNA has been experimentally confirmed to regulate behavior in non-model insects.

The migratory locust, *Locusta migratoria*, is a worldwide agricultural pest and displays dramatic phenotypic plasticity. Morphological, physiological, and behavioral traits of this insect reversely change between solitarious and gregarious phases [23]. Under high-population density, gregarious locusts form large swarms, become attracted by conspecifics, exhibit active movement, and migrate long distances. In contrast, solitarious locusts live in an individual state, stay quiet, and are cryptically colored to blend with their surroundings. However, locusts at these two phases have the same genome, which is 6.5 Gb in size. Protein-coding genes account for a small portion of the locust genome, with many regions expressing ncRNAs [24]. The molecular regulatory mechanism underlying two-phase changes involves several protein-coding genes that are associated with olfaction [25], dopamine biosynthesis and release pathway [26], as well as neuropeptide F/ nitric oxide pathway [27,28]. Moreover, microRNAs modulate phase-related traits by regulating the expression of phase-related genes [29,30]. Phase-, habitat-, and gender-specific lncRNAs have been identified in migratory locusts, implying the possible roles of lncRNAs in locust phase change [31]. However, the expression and function of lncRNAs related to locust phase change remain unknown.

To identify phase-related lncRNAs, we systematically characterized the expression of locust lncRNAs and annotated their functions. In crowding and isolation phases, lncRNAs showed more temporally specific expression pattern and sensitive response to phase change than mRNAs. Finally, a hub lncRNA derived from the co-expression networks between early changed lncRNAs and mRNAs was experimentally validated to modulate phase-related behavior. Our findings reveal the important roles of lncRNAs in locust phase change and the interactions between protein-coding genes and ncRNAs in the phenotypic plasticity of locusts.

Results

Locust lncRNAs possess fewer but longer exons than mRNAs

To identify phase-related lncRNAs systematically, we constructed 24 rRNA-depleted RNA libraries with three biological replicates for each treatment (**Figure 1**A). The samples were obtained from brains of locusts undergoing time-course crowding and isolation treatments. In the crowding treatment (CS), solitarious locusts were crowded by keeping them in the same cage with gregarious locusts for 0, 4, 8, or 16 h to promote a behavioral transition toward the gregarious phase. In the isolation treatment (IG), gregarious locusts were isolated by individually keeping them in separate cages for 0, 4, 8, or 16 h. Strand-specific RNA-seq of these libraries was performed on an Illumina HiSeq 3000 platform.

After the low-quality reads were removed, 363.14 Gb clean data of all samples were obtained, with the Q30 higher than 91.1% as evaluated using FastQC [32]. These data indicate that all subsequent analyses were based on high-quality data (Table S1). A total of 1,597,688 transcripts from 1,431,906 loci were identified (Figure S1A). Sequencing saturation assessment indicated that the sequencing depth was sufficient to identify novel transcripts (Figure S1B). The expression levels of the transcripts were measured as fragments per kilobase of transcript per million mapped reads (FPKM) by using RSEM method [33].

Based on the annotation of locust genome sequences [24], 15,309 known protein-coding transcripts were identified from the assembled transcripts by using the Cuffcompare program in the Cufflinks package. The unknown transcripts were used for putative lncRNA screening, and an analysis pipeline was developed to identify *bona fide* lncRNAs (Methods and Figure S1A). Transcripts with length < 200 bp and protein-coding potential were discarded. Finally, 14,373 highly reliable putative lncRNAs (FPKM > 1.0) from 10,304 loci were identified (Figure S1C; Table S2). The expression levels of lncRNAs were verified by qRT-PCR and displayed a high correlation (Pearson's $r = 0.86$) with those from RNA-seq (Figure S2).
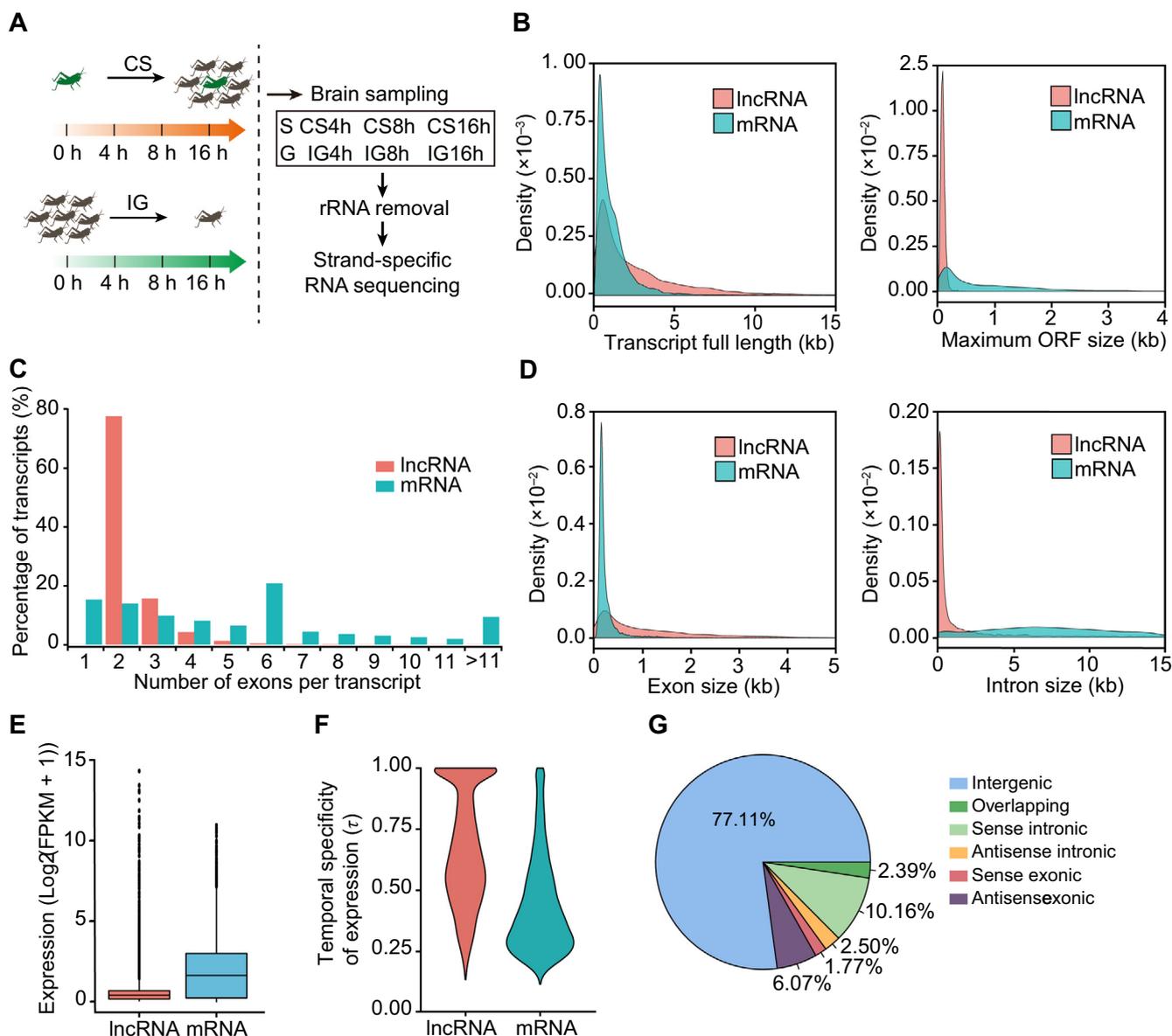
Figure 1   Different structures and expression between lncRNAs and mRNAs
A. Experimental pipeline for the crowding and isolation treatments, sampling, and sequencing. The black vertical lines represent sampling time points. **B.** Full-length and maximum ORF size distribution. **C.** Number of exons per transcript. **D.** Density distribution of exon size and intron size. **E.** Overall expression $\log_2(\text{FPKM} + 1)$ of lncRNAs compared with mRNAs in the locust brain. **F.** Temporally specific expression of lncRNAs and mRNAs in the time course of crowding and isolation. **G.** Locust lncRNA categories and their distribution. G, gregarious locust; S, solitarious locust. CS, crowding treatment of solitarious locusts; IG, isolation treatment of gregarious locusts.

We characterized the genomic features of locust lncRNAs by comparing them with protein-coding mRNAs assembled in this study. The lengths of lncRNAs ranged from 202 bp to 25,251 bp, with the median length of 1568 bp. As shown in Figure 1B (left panel), lncRNAs were significantly longer than mRNAs (median length of 786 bp) [Kolmogorov–Smirnov test (KS test), $D = 0.30002$, $P < 2.2\text{E}{-}16$]. However, lncRNAs were shorter than mRNAs in terms of the size of the maximum ORF predicted (median maximum ORF length of 78 bp for lncRNAs and 544 bp for mRNAs; KS test, $D = 0.71984$, $P < 2.2\text{E}{-}16$; Figure 1B, right). Approximately 78% of lncRNAs comprised 2 exons, whereas the number of exons contained in mRNAs ranged from 1 to 120 (Figure 1C).

Thus, exons of lncRNAs were significantly longer than those of mRNAs (average length: 1142 bp *vs*. 263 bp; KS test, $D = 0.53967$, $P < 2.2\text{E}{-}16$; Figure 1D, left). Meanwhile, introns of lncRNAs were significantly shorter than those of mRNAs (average length: 10,086 bp *vs*. 12,442 bp; KS test, $D = 0.61789$, $P < 2.2\text{E}{-}16$; Figure 1D, right). The expression level analysis indicated that the global expression level of lncRNAs was significantly lower than that of mRNAs (average of 0.6 *vs*. 1.9; Student's *t*-test, $t = 79.868$, $P < 2.2\text{E}{-}16$; Figure 1E). However, the expression specificity analysis demonstrated that lncRNAs exhibited more temporally restricted expression than mRNAs (average of 0.672 *vs*. 0.436; Student's *t*-test, $t = 79.868$, $P < 2.2\text{E}{-}16$; Figure 1F).

Based on lncRNA genomic locations relative to mRNAs, locust lncRNAs were classified as intergenic, overlapping, sense intronic, antisense intronic, sense exonic, and antisense exonic. More than 77% of locust lncRNAs were long intergenic ncRNAs (lincRNAs, Figure 1G). These results indicate that locust lncRNAs differ considerably from mRNAs in terms of structure and expression. Locust lncRNAs are longer, and possess fewer but longer exons and shorter introns. Moreover, the expression pattern of lncRNAs displays higher temporal-specificity than that of mRNAs.

More lncRNAs are expressed specifically in gregarious locusts than in solitarious locusts

A total of 9722 lncRNAs (73.9%) were commonly expressed in solitarious and gregarious locust brains, whereas 962 lncRNAs (7.3%) were expressed specifically in solitarious locusts and 2469 (18.8%) in gregarious locusts (Figure 2A, top). Only 479 and 1090 mRNAs (3.1% and 7.1%) were expressed specifically in solitarious and gregarious locusts, respectively (Figure 2A, bottom). These results indicate that a higher percentage of lncRNAs are expressed specifically at the two locust phases compared with mRNAs, and more lncRNAs are expressed in gregarious than in solitarious locusts. The expression levels of 335 lncRNAs and 779 mRNAs were downregulated, whereas those of 313 lncRNAs and 261 mRNAs were upregulated in gregarious locusts compared with those in solitarious locusts [fold change (FC) > 2 and $P < 0.05$; Figure 2B]. lincRNAs account for the highest percentages among both down- and upregulated lncRNAs (81.2% and 87.8%). The second most lncRNAs among down- and upregulated lncRNAs were antisense exonic and sense intronic lncRNAs, respectively (Figure 2B). These lncRNAs and their nearby protein-coding genes are shown in Table S3, and the top 10 lncRNA genes ranked by FC in expression are displayed in Figure 2C. Thus, the number of lncRNAs expressed in the brain and their expression levels significantly differ between solitarious and gregarious locusts.

lncRNAs act as signatures of phase change different from mRNAs

At different time points of CS, the percentages of expressed lncRNAs among total lncRNAs ranged from 64.3% to 68.9%, whereas those of mRNAs among total mRNAs ranged from 87.9% to 89.7% (Figure S3A, top left). At different time points of IG, the percentages of expressed lncRNAs among total lncRNAs ranged from 64.8% to 81.0%, whereas those of mRNAs among total mRNAs ranged from 87.2% to 91.9% (Figure S3A, top right). However, the proportions of specifically expressed lncRNAs in the same samples were higher than those of mRNAs during CS and IG (Figure S3A, bottom). The specific expression pattern of lncRNAs suggests its association with locust phase change.

Differentially expressed (DE) lncRNAs and mRNAs during CS and IG were identified by comparing the normalized expression of transcripts at each point of the time course with that at 0 h time point (*i.e.*, CS4h/S, CS8h/S, CS16h/S, IG4h/G, IG8h/G, and IG16h/G). A total of 1246 and 1871 transcripts were differentially expressed during CS and IG, respectively. Among these transcripts, the expression levels of 733 lncRNAs and 513 mRNAs changed during CS, whereas those of 837 lncRNAs and 1034 mRNAs changed during IG (Figure S3B). Thus, more lncRNAs than mRNAs were involved in CS, whereas more mRNAs than lncRNAs were engaged in IG. DE lncRNAs and mRNAs at CS 16 h compared with G and at IG 16 h compared with S were also identified (Figure S3C). The results show that more mRNAs than lncRNAs were involved in CS16h/G, whereas more lncRNAs than mRNAs were involved in IG16h/S. This result is contrary to the aforementioned finding but indicates the effectiveness of crowding and isolation treatments [25]. At each time point of CS and IG, the numbers of DE lncRNAs differed slightly, whereas those of mRNAs varied obviously (Figure 2D, left and 2E, left). The numbers of up-regulated and downregulated lncRNAs were comparable during CS and IG, but there are more upregulated mRNAs than downregulated mRNAs for both CS and IG stages (Figure 2D, right and 2E, right). The numbers of constantly changed transcripts which were differentially expressed at all time points were different for lncRNAs and mRNAs. More lncRNAs were constantly changed than mRNAs in CS but fewer than mRNAs in IG (lncRNAs *vs.* mRNA: 33 *vs.* 16 in CS and 85 *vs.* 103 in IG; Figure 2F and G). Additionally, the K-means clustering of the lncRNA expression differed with that of mRNAs during CS and IG (Figure 2H and I). All these data suggest that lncRNAs respond differently to crowding and isolation treatments in comparison with mRNAs. However, the hierarchical clustering analyses demonstrate that DE lncRNAs and mRNAs clustered samples collected at the same time point together in CS and IG. Hence, lncRNAs and mRNAs can be signatures of locust phase change (Figure S4).

Rapid response of lncRNAs to population density change

With the time course of the crowding or isolation treatment, the numbers of DE lncRNAs and mRNAs increased (Figure 2D and E). However, the percentage of the regulated lncRNAs was two-fifths higher than that of mRNAs at the 4 h time point in CS (chi-square test, $P = 0.002$) and one-thirds higher than that of mRNA at the 4 h time point in IG (chi-square test, $P = 3.62E–4$; Figure 3A). The number of changed lncRNAs reached the highest level at the earlier stage (4 h) than that of mRNAs (8 h or later). The percentage of specifically changed lncRNAs at the 4 h time point was also higher than that of mRNAs during CS and IG (chi-square test, $P = 0.008$ in CS and $P = 3.28E–13$ in IG) and reached a relatively stable stage rapidly (Figure 3B). Similar observation is obtained for the upregulated lncRNAs and mRNAs during IG (chi-square test, $P = 0.002$), and the downregulated lncRNAs and mRNAs during CS and IG (chi-square test, $P = 3.30E–6$ in CS and $P = 0.02$ in IG; Figure 3C). These results indicate that lncRNAs respond more rapidly to crowding and isolation treatments than mRNAs.

To further analyze the expression patterns of lncRNAs and mRNAs, we used the Short Time-series Expression Miner (STEM) to cluster transcripts based on expression. In the STEM analysis, 26 model profiles were set to describe the expression patterns of the regulated lncRNAs and mRNAs during CS and IG, but only several significant expression profiles were detected ($P < 0.05$, Figure S5). During CS, 214 lncRNAs and 242 mRNAs were clustered into eight and nine significant expression profiles, respectively (Figure 3D).
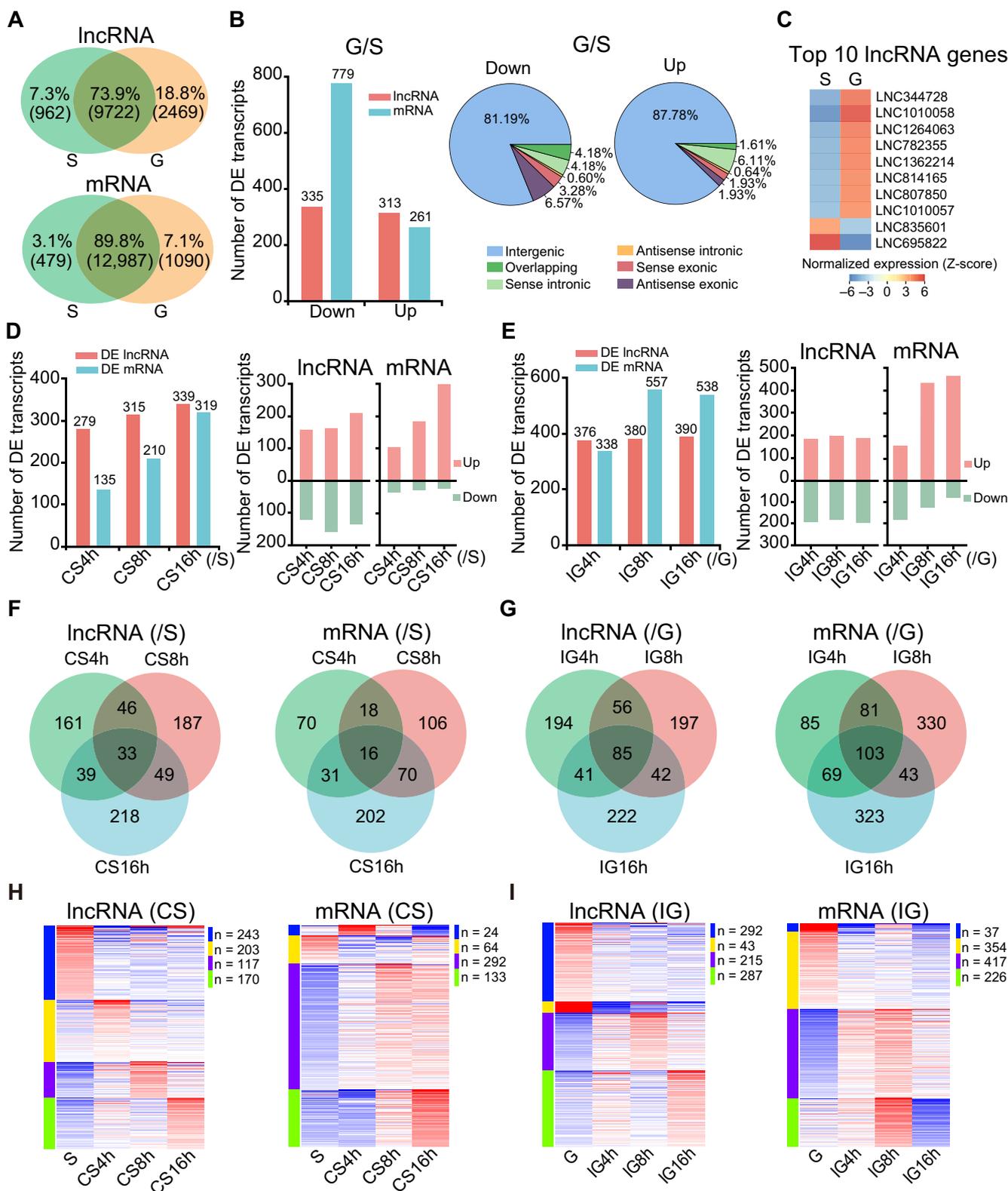
Figure 2   lncRNAs display distinct patterns of expression change in locust phase change
A. Venn diagram of expressed lncRNAs and mRNAs in gregarious and solitarious locusts. **B.** Number of downregulated and upregulated lncRNAs and mRNAs in gregarious locusts compared with those in solitarious locusts (left panel). The right panel shows the percentages of different types of DE lncRNAs between G and S. **C.** Top 10 DE lncRNAs between gregarious and solitarious locusts. **D.** Number of total regulated, upregulated, and downregulated lncRNAs and mRNAs at each time point relative to S. **E.** Number of total regulated, upregulated, and downregulated lncRNAs and mRNAs at each time point relative to G. **F.** Venn diagram of DE lncRNAs and mRNAs in CS. **G.** Venn diagram of DE lncRNAs and mRNAs in IG. **H.** K-mean clustering of DE lncRNAs and mRNAs in CS. **I.** K-mean clustering of DE lncRNAs and mRNAs in IG. n indicates the number of transcripts in each cluster. DE, differentially expressed.
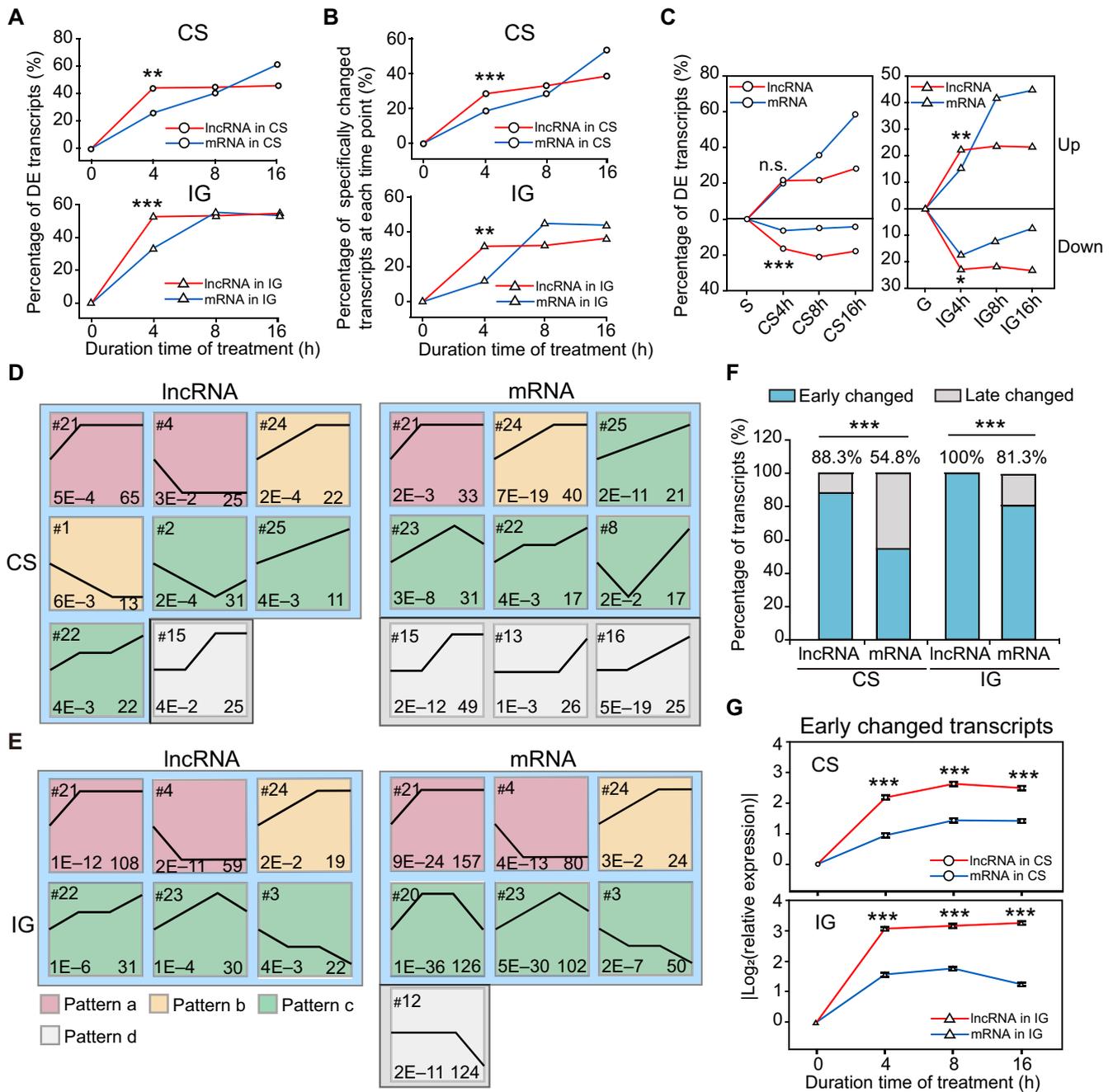
Figure 3 Rapid response of lncRNAs to crowding and isolation treatments
A. Percentages of the DE lncRNAs and mRNAs at each time point of time course relative to total DE lncRNAs and mRNAs in CS and IG. **B.** Percentages of specifically changed lncRNAs and mRNAs at each time point of time course relative to total specifically expressed lncRNAs and mRNAs in CS and IG. **C.** Percentages of upregulated and downregulated lncRNAs and mRNAs at each time point of time course in CS (left) and IG (right) relative to the total DE lncRNAs and mRNAs. **D.** Significant expression profiles ($P < 0.05$) of lncRNAs and mRNAs clustered via STEM software in CS. **E.** Significant expression profiles ($P < 0.05$) of lncRNAs and mRNAs clustered via STEM software in IG. The numbers in the top left corner of boxes are profile serial numbers, those in the bottom left corner are P values, and those in the bottom right corner are numbers of transcripts contained in profiles. Pattern a (early-changed) profiles, pattern b (early-middle-changed) profiles, pattern c (sustainably-changed) and pattern d (late-changed) profiles are shown with backgrounds in violet, orange, green, and gray, respectively. **F.** Percentages of the early and late changed transcripts (lncRNAs or mRNAs) of CS and IG. **G.** Expression patterns of early changed lncRNAs and mRNAs during CS and IG. Relative expression levels of all early changed lncRNAs or mRNAs are presented as $|\log_2(v(i)/v(0))|$. $v(i)$ and $v(0)$ represent the expression levels of the transcripts at i h (i = 4, 8, 16) and 0 h time points, respectively. Significant differences between lncRNAs and mRNAs in panels **A**, **B**, **C**, and **F** were determined by chi-squared test. Data in **G** are shown as mean ± SE. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Among these profiles, lncRNA expression profile 15, as well as mRNA expression profiles 15, 13, and 16, did not alter until 8 or 16 h after the crowding treatment. These profiles were termed late changed profiles (pattern d). In contrast to late changed profiles, early changed profiles featured transcript expression change starting at the 4 h time point. In accordance with the expression change after the 4 h time point, the early changed profiles were subdivided into pattern a (early-changed), pattern b (early-middle-changed), and pattern c (sustainably-changed). During the IG process, 269 lncRNAs and 639 mRNAs were clustered into six and seven significant expression profiles, respectively. All lncRNA profiles were early changed (Figure 3E). Among the mRNA expression profiles, six were early changed, while profile 12 was a late changed profile. The numbers of clustered profiles for lncRNAs did not significantly differ from those for mRNAs during CS and IG. However, calculation of transcript number in profiles indicated that the percentage of early changed lncRNAs was 61.1% higher than that of mRNAs in CS (88.3% vs. 54.8%; chi-square test, $P = 1.05E-7$; Figure 3F). Similarly, the percentage of early changed lncRNAs was higher than that of mRNAs in IG (100% vs. 81.3%; chi-square test, $P = 3.93E-12$; Figure 3F). Thus, lncRNAs respond sooner to the crowding and isolation treatments than mRNAs in general.

The expression change rate of early changed lncRNAs was faster than that of mRNAs in CS (0.55 vs. 0.24; Student's t-test, $t = 7.88$, $P = 9.32E-14$) and IG (0.77 vs. 0.39; Student's t-test, $t = 10.42$, $P = 2.32E-22$; Figure 3G). Therefore, the ratio and expression change rate of early changed lncRNAs are higher than those of early changed mRNAs. Thus, locust lncRNAs are more sensitive to changes in population density than mRNAs.

lncRNAs are involved in different pathways in the early-period changes in CS and IG

To determine the functions of the early changed lncRNAs and hub lncRNAs in locust phase change, we constructed lncRNA–mRNA co-expression networks by analyzing their expression correlations. Pearson's correlation coefficient between early changed lncRNAs and all detectable mRNAs was evaluated in CS and IG. Only lncRNA–mRNA pairs with strong correlation ($|r| > 0.9$) were used for network construction (**Figure 4**A and B; Table S4). In the co-expression networks, the nodes were divided into three modules based on lncRNA expression patterns, namely patterns a, b, and c. The protein-coding genes correlated with lncRNAs were separated into relevant modules accordingly. KEGG enrichment of the mRNAs in each module was analyzed to predict biological pathways where lncRNAs might be involved.

In CS, autophagy regulation, spliceosome, and inositol phosphate metabolism pathways were overrepresented in the top rank and in at least two modules (Figure 4C). Thus, early changed lncRNAs in CS may play important roles in regulating autophagy, RNA splicing, and signal transduction pathways. Regulation of lncRNAs on autophagy was involved in the initial and middle stage of gregarization, whereas regulation on the spliceosome and inositol phosphate metabolism was involved in the entire process. The synaptic vesicle cycle pathway, which is related to neurotransmitter release, was only

overrepresented for lncRNAs in pattern a (early-changed) module (Figure 4C). This result implies that early-changed lncRNAs may have important roles in regulating the nervous system during CS. In addition, several lncRNAs were strongly associated with known phase-related genes. For instance, early-changed lncRNAs LNC531328.2, LNC1425451.2, LNC1088763.7, and LNC492755.1 were associated with genes *NPYR*, *NPF1a*, and *Vat1* [28] ($r = -0.90, -0.91, -0.98, -0.95$; Figure 4A, Table S5). Sustainably-changed lncRNAs LNC494161.1 and LNC1065048.14 were associated with genes *Ebony* and *Vat1* in the dopamine pathway [26] ($r = -0.92$ and 0.98, respectively).

In the lncRNA–mRNA networks, an lncRNA is considered important if the network degree is high. In the CS network, the top 5% lncRNAs in terms of the degree were regarded as hub lncRNAs (Figure 4A). The degree values of lncRNA loci were calculated and ranked to identify hub lncRNAs at the "gene" level because several lncRNAs are derived from the same lncRNA locus and lncRNA genes can be functional units [34]. In CS, 10 lncRNA loci that were ranked in the top 5% based on degree values were identified as hub lncRNA genes. Among them, four genes exhibited upregulated expression and the other six genes exhibited downregulated expression (Figure 4D).

During IG, synapse-related pathways including glutamatergic synapse, dopaminergic synapse, and cholinergic synapse pathways were enriched. Most of these pathways were enriched in pattern a (early-changed) module. Meanwhile, the dopaminergic synapse and cholinergic synapse pathways were only involved in pattern a (early-changed) module (Figure 4E). Signal transduction pathways, such as calcium signaling, Ras signaling, insulin signaling, and MAPK signaling pathways, were also enriched in IG (Figure 4E). The results of functional annotations suggest that early changed lncRNAs participate more in synapse-related and signal processing pathways in IG than in CS.

The top 5% lncRNAs ranked based on the degree values were regarded as hub lncRNAs in the IG network (Figure 4B), and the largest number of hub lncRNAs were found in pattern a (early-changed) module. Thus, early-changed lncRNAs may play important roles in locust phase change. In total 13 hub lncRNA loci that ranked in the top 5% based on degree values were identified in IG (Figure 4F). The expression of two hub lncRNAs was upregulated, and that of 11 others was downregulated. Interestingly, two overlaps were observed in hub lncRNA loci identified in CS and IG (Figure 4D and F). Moreover, the expression levels of these loci increased in CS and decreased in IG, indicating that they sensitively respond to crowding and isolation treatments in different directions. Thus, LNC1010057 and LNC992414 are regarded as valuable candidate regulators in locust phase change.

LNC1010057 potentially regulates locust phase change

To verify the function of LNC1010057 and LNC992414 in locust phase change, we performed a series of molecular biology experiments. First, the longest transcripts identified in LNC1010057 and LNC992414 loci were cloned. LNC1010057 consisted of repeat A, B, and C elements, which were sequentially distributed and repeated four times, except for the C element (repeated three times, **Figure 5**A). The
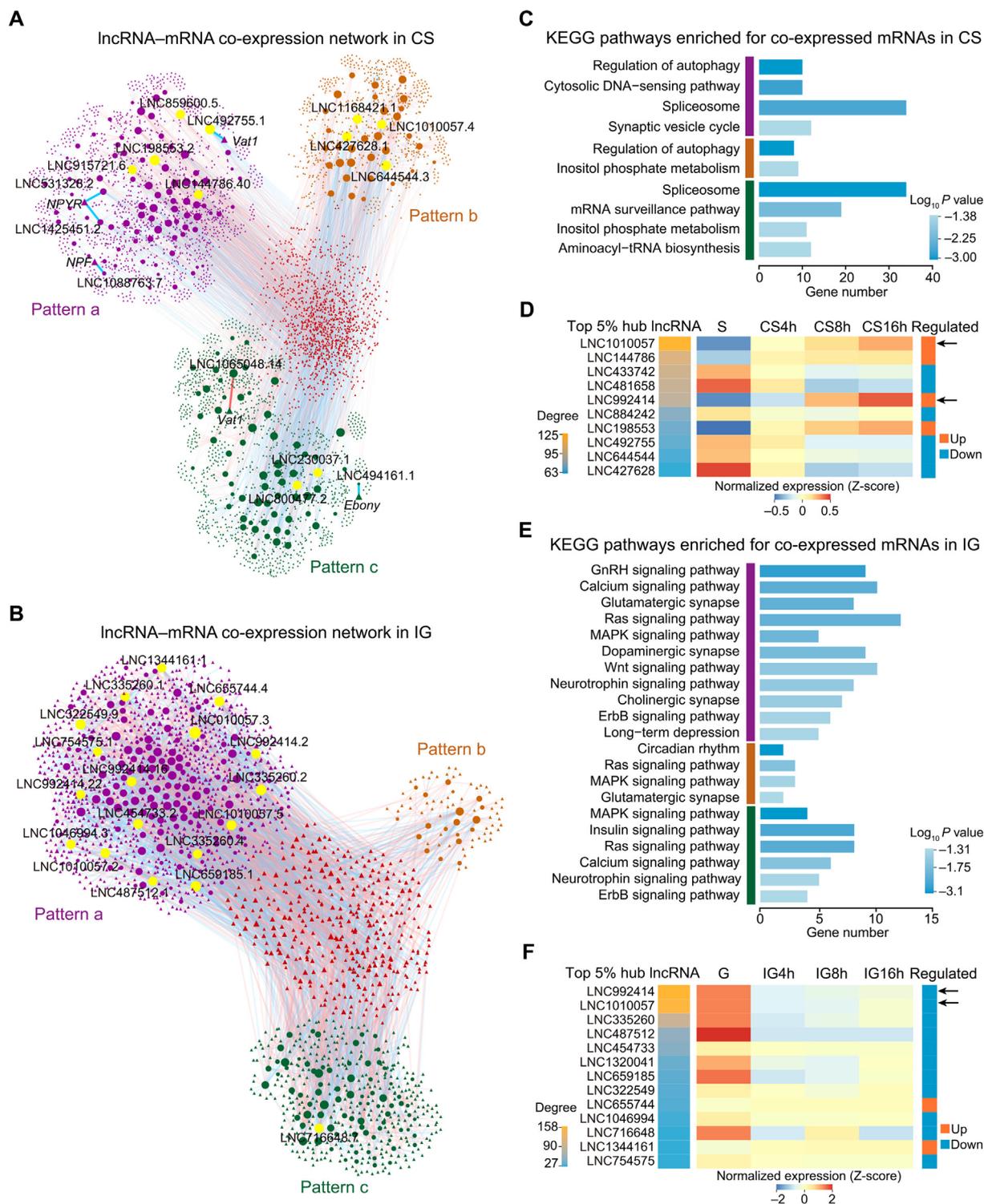
Figure 4   Early changed lncRNAs are involved in different pathways in CS and IG
A. Co-expression networks between lncRNA and protein-coding genes constructed in CS using Pearson correlation coefficient. **B.** Co-expression networks between lncRNA and protein-coding genes constructed in IG using Pearson correlation coefficient. The circular and squared nodes represent lncRNAs and mRNAs, respectively. The red and blue lines denote positive and negative correlations, respectively. The nodes in the network are divided into three modules, including pattern a (violet), pattern b (orange), and pattern c (green) modules, based on the lncRNA expression patterns. The lncRNAs with top-ranked degrees are colored yellow. **C.** KEGG pathways enriched for each module in the CS network. Colored vertical bars represent the different expression modules. **D.** Top 5% of hub lncRNA loci ranked by degree in CS. The heatmap represents the expression changes of hub lncRNA loci in CS. **E.** KEGG pathways enriched for each module in the IG network. **F.** Top 5% of hub lncRNA loci ranked by degree in IG. The heatmap represents the expression changes of hub lncRNA loci in IG. The black arrows indicate the overlapping hub lncRNAs in CS and IG networks.

sequence alignments demonstrate that LNC992414 shares similar repeat sequences with LNC1010057 at the 5′ end but they were transcribed from different genome loci (Figures 5A and S6). Although the quantitative expression level of LNC992414 can be detected by specific primers, the expression level of LNC1010057 represents the total expression level of repetitive elements. As predicted, the expression patterns of LNC1010057 and LNC992414 were extremely similar. During CS, the expression of both lncRNAs increased consistently and achieved a nearly fourfold increase at 16 h time point (one-way ANOVA, $F = 5.22$, $P = 0.009$; Figure 5B). During IG, their expression levels were significantly decreased after 4 h of isola-

tion and remained relatively stable afterward (one-way ANOVA, $F = 33.47$, $P = 5.45E–8$; Figure 5B). The similarities of the sequence structures and expression patterns between LNC1010057 and LNC992414 suggest that they are homologous lncRNAs. However, absolute real-time qPCR analyses demonstrate that the LNC1010057 expression level was approximately 1000-fold higher than that of LNC992414 in the brains of gregarious and solitarious locusts (Figure 5C).

Second, to test whether LNC1010057 and LNC992414 are involved in locust phase change regulation, we conducted behavioral assay after suppressing their expression in locust brains by RNA interference (RNAi). Double-stranded RNA
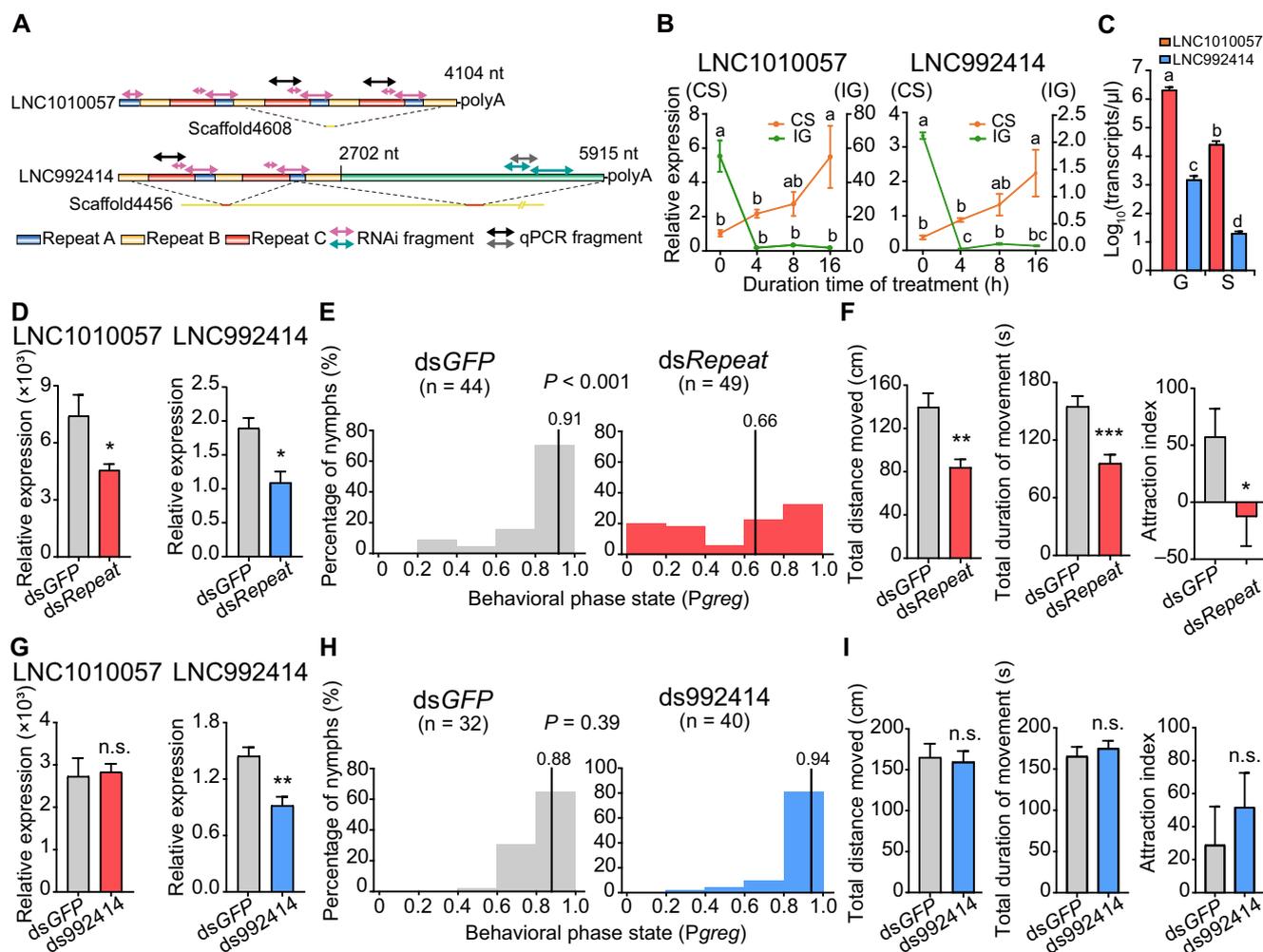
Figure 5   LNC1010057 potentially regulates the phase change of gregarious locusts
A. Full-length sequence structures and genomic locations of LNC1010057 and LNC992414. **B.** Expression changes of LNC1010057 and LNC992414 in the time course of CS and IG. Left and right axes show lncRNA expression level normalized to that of the reference gene *RP49* in CS and IG, respectively. **C.** Absolute expression levels (transcripts/μl) of LNC1010057 and LNC992414 in gregarious and solitarious locusts. **D.** Efficiency of RNAi of LNC1010057 and LNC992414 in the locust brain after ds*Repeat* injections with ds*GFP* injection as the control. The RNAi of ds*Repeat* targeted all repetitive elements of LNC1010057 and LNC992414. **E.** Phase changes in gregarious locusts when the expression of LNC1010057 and LNC992414 was knockdown in the brain. P_{greg}, probabilistic metric of gregariousness. Vertical lines indicate median P_{greg} values. **F.** Changes in total distance moved, duration of movement, and attraction index of gregarious locusts after knocking down the expression of LNC1010057 and LNC992414. **G.** Efficiency of LNC992414 RNAi in locust brains after ds992414 or ds*GFP* injection. **H.** Phase changes in gregarious locusts when the expression of LNC992414 was knocked down in the brain. **I.** Changes in total distance moved, duration of movement, and attraction index of gregarious locusts after the expression of LNC992414 was knocked down. Measurements are shown as mean ± SE. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ (Student's *t*-test).

(ds*Repeat*) was injected into brains of gregarious locusts to target all the repetitive elements and significantly decreased the expression levels of LNC1010057 (7470 *vs.* 4610; Student's *t*-test, $t = 2.93$, $P = 0.02$) and LNC992414 (1.9 *vs.* 1.0; Student's *t*-test, $t = 3.88$, $P = 0.004$; Figure 5D). The injection of ds*GFP* was as the control. The behavioral assay demonstrates that gregarious locusts significantly changed their behavior toward the solitarious state with the decreased expression levels of LNC1010057 and LNC992414 (Mann–Whitney *U* test, $U = 599$, $P < 0.001$, ds*Repeat vs.* ds*GFP*; Figure 5E). Furthermore, multiple phase-related behavior parameters changed after the injection of repetitive element dsRNA. The total distance moved (TDM) and total duration of movement (TDMV) were significantly decreased (Student's *t*-test, $t = 4.12$, $P < 0.001$; $t = 4.64$, $P < 0.001$; Figure 5F). However, the movement speed did not differ with the reduced repetitive element expression (Student's *t*-test, $t = 0.008$, $P = 0.93$; Figure S7A). The conspecific preferred behavior of gregarious locusts significantly decreased when the expression of repetitive elements was knocked down (58.3 *vs.* −13.5; Student's *t*-test, $t = 2.07$, $P = 0.04$; Figure 5F). This finding could be mainly due to the fact that locusts spent a short duration in the attraction zone (Student's *t*-test, $t = 2.34$, $P = 0.02$) but not a long duration in the repulsion zone (Student's *t*-test, $t = 1.14$, $P = 0.26$; Figure S7B). The movement duration in the border zone was significantly decreased (Student's *t*-test, $t = 3.44$, $P < 0.001$). However, the difference in the central zone was insignificant (Student's *t*-test, $t = 1.76$, $P = 0.08$) after the repetitive element expression decreased (Figure S7C). The number of turns decreased (Student's *t*-test, $t = 3.06$, $P = 0.003$), and the mean turn angle increased (Student's *t*-test, $t = 2.96$, $P = 0.004$), compared with those of the *GFP* control groups (Figure S7D). However, the specific knockdown of LNC992414 expression did not cause the shift from the gregarious phase to the solitarious state (Figure 5G and H). The phase-related behavior parameters including the movement and the preferred behavior for the conspecifics were not different, compared with the ds*GFP* control (Figures 5I and S7E–H). The RNAi experiments for LNC992414 only did not cause the behavioral changes, so the possibility of LNC992414 in regulating locust phase change was excluded. However, no LNC1010057-specific RNAi was performed because the sequence of LNC1010057 was almost identical to the 5′ end of LNC992414. These results indicate that LNC1010057 but not LNC992414 potentially regulates the locust phase change.

Discussion

Here, we provided the time-series lncRNA expression profiling in locusts. Thus, we could identify DE lncRNAs between gregarious and solitarious phases as well as during locust phase change. We found that lncRNAs responded sooner to the change in population density than mRNAs. Based on the lncRNA–mRNA co-expression networks, the functions of the early changed lncRNAs were predicted and hub lncRNAs were selected. Finally, one hub lncRNA was experimentally confirmed to regulate the locust phase change.

Longer locust lncRNAs

lncRNAs of locusts possess different features compared with those of other insect species. One of the most significant feature is the longer transcript length of locust lncRNAs than mRNAs. In other insect species, such as *Nilaparvata lugens* [20], *Apis mellifera* [18], and *Plutella xylostella*, the average transcript length of lncRNAs is shorter than that of mRNAs [19]. This is possibly due to the shorter length of locust mRNAs compared with those of other species. However, the average or median transcript length of locust lncRNAs is also longer than that of these species. The longer transcript length of locust lncRNAs compared with that of other species should be further studied. The large portion of non-coding regions in the locust genome possibly contributes to this phenomenon to a certain extent [24]. Meanwhile, locust lncRNAs differ significantly from mRNAs in terms of structure and expression. Locust lncRNAs possess longer exons, fewer exons per transcript, and lower global expression levels than mRNAs. A long exon may influence the evolution, alternative splicing, and expression of locust lncRNAs [35]. The highest number of lncRNAs derived from different genomic regions in the locust are lincRNAs, which is similar to that in other species, such as flies [5,36], zebrafish [3], and humans [1]. These results indicate the unique and common structural features of lncRNAs in locusts and other species.

Initial response of lncRNAs to locust phase change

The interaction analysis of lncRNAs and mRNAs in time course treatments demonstrates that more lncRNAs are expressed in gregarious locusts than in solitarious ones, and a higher percentage of lncRNAs are specifically expressed in the two locust phases than mRNAs. lncRNAs respond more rapidly to crowding and isolation than mRNAs. The quicker response of the former than the latter is reflected by the higher ratio of lncRNAs regulated at the 4 h time point and their faster expression change rate in the first 4 h compared to those of mRNAs. Therefore, both the ratio and the expression change rate of early changed lncRNAs are notably higher than those of early changed mRNAs; thus, locust lncRNAs are more sensitive to population density changes than mRNAs. The initial role of lncRNAs in the transcriptional process is extremely crucial in the response to changes in population density, similar to the initial role of *CSP* gene in the locust phase change [25]. Given that lncRNAs can regulate transcription, translation, and mRNA stability [7], it is reasonable that lncRNAs respond first as regulators. This finding may be explained by the low stability and rapid turnover rate of lncRNAs [37]. However, the transcription speed of lncRNAs may be more important than the other explanations above [38]. The density-cued expression of LNC1010057 could be regulated by several transcription factors, such as ZNF300, lolal, and Gtf2e2, whose coding genes are co-expressed with LNC1010057 (Table S6). The lncRNA expression in the time series of the development and other biological processes has been fully studied in other species [5,39]; however, the dynamic expression of lncRNAs in time-course treatments on behavior

has been rarely described in insects. The temporal analysis uncovers for the first time the initial roles of lncRNAs in response to the change in population density. The rapid response of ncRNAs may help organisms rapidly adapt to changes in environmental conditions.

Hub lncRNA identified in networks regulates locust phase change

Given that functions of lncRNAs can be inferred by protein-coding genes in co-expression networks and nearby genes in the genome [9], we analyzed co-expressed genes instead of nearby genes of lncRNAs to characterize the potential functions of early changed lncRNAs. Nearby genes, whose expression changed, were extremely few to analyze (Table S7). Several lncRNAs are strongly associated with known phase-related genes, such as *NPYR* and *NPF1a* genes in the neuropeptide F/nitric oxide pathway [27,28], and *Ebony* and *Vat1* genes in the dopamine pathway [26]. This result reveals a new non-coding layer that is connected to the primary mechanism, in which protein-coding genes participate in the locust phase change. Functional annotation of lncRNAs indicates that the locust phase change is involved in multiple regulatory mechanisms and is concentrated on common pathways that have been reported in our previous studies [25,26,28].

In this study, we predicted two hub and homologous lncRNAs as potential regulators because they are conversely expressed in crowding and isolation. LNC1010057 but not LNC992414 was validated to primarily participate in the locust phase change by experiments *in vivo*, although they share several common domains. The LNC992414 expression knockdown was insufficient to cause the shift from the gregarious phase to the solitarious phase. Thus, LNC1010057, a repetitive element-containing lncRNA, regulates the locust behavior, although it is not conserved among other species. Repetitive element-containing lncRNAs have been shown to function in other species. In primates, an Alu element-inserted lncRNA *5S-OT* regulates the alternative splicing of target genes through complementary base pairing [40]. Another short interspersed elements (SINE)-containing lncRNA regulates myogenesis in rodents [41]. In the nervous system, lncRNAs and repetitive elements were predicted *in silico* to function in synaptic plasticity, but there lacks experimental evidence [42]. In migratory locusts, repetitive elements constitute more than half of the genome [24]. The present study offers evidence about their important functions in phase-related behavioral change in locusts. The different roles of homologous lncRNAs may be due to the extremely high expression level of LNC1010057, offsetting the decrease in the expression level of LNC992414. Another explanation for this observation is that the secondary structure of the longer lncRNA LNC992414 may affect the target binding sites that is necessary for the function [43,44]. The diverse functions of homologous lncRNAs are also identified for lncRNAs SNHG10 and SCARNA13 in hepatocarcinogenesis [45]. Finally, given the absence of neighboring genes within 100 kb for LNC1010057 in the locust genome, LNC1010057 shows the least possibility for *cis*-regulation. Therefore, LNC1010057 may regulate distal target genes through *trans*-acting mechanisms [46]. The genes co-expressed with LNC1010057 in CS and IG networks may be the *trans*-targets that warrant further validation in the future

(Table S4). However, we cannot rule out the possibility that the neighboring genes located more than 100 kb away may interact with LNC1010057. With the improvement of sequencing quality, neighboring genes may also be identified as candidates for further studies.

In conclusion, this work discovers the relationships between protein-coding genes and lncRNAs in locust phase change, proves the regulatory role of lncRNAs in phase-related behavior in insects, and provides insights into additional mechanisms underlying the phenotype plasticity of animals.

Materials and methods

Insects

Gregarious and solitarious locusts were obtained from the same locust colony which were collected from Hebei Province, China and maintained at the laboratory. Gregarious locusts were cultured in well-ventilated cages (40 cm × 40 cm × 40 cm) at a density of 500–1000 locusts per cage for eight generations. Solitarious locusts were cultured individually in white metal boxes (10 cm × 10 cm × 25 cm) supplied with fresh air for more than 10 generations. Both colonies were maintained under the same photocycle regime and temperature conditions, and fed with the same food as described in a previous work [29].

Isolation and crowding treatments

For isolation treatment, gregarious locusts were separately raised as solitarious locusts as described above. After 4, 8, or 16 h of isolation, the brains of the 3-day-old fourth-instar gregarious locusts were collected at 2 PM. The brains of gregarious locusts were collected as the 0 h time point sample for the isolation course. For crowding treatment, 10 fourth-instar solitarious locusts were reared together with a stimulating group of 20 fourth-instar gregarious locusts in a small cage (10 cm × 10 cm × 10 cm). After 4, 8, or 16 h of crowding, the brains of the 3-day-old fourth-instar solitarious locusts were collected at 2 PM. The brains of solitarious locusts were collected as the 0 h time point sample for the crowding course. Sixty cages containing 1 gregarious locust per cage and six cages containing 10 solitarious locusts per cage were prepared for each time point of isolation and crowding treatments, respectively. All locust samples from the isolation and crowding treatments were immediately frozen in liquid nitrogen for storage. Three biological replicates of the samples were collected at the same time point. Each replicate contained 20 insects, with equal numbers of males and females.

RNA extraction, cDNA library construction, and RNA-seq

Total RNA was extracted from the collected locust brains by using TRIzol reagent (Catalog No.15596-026, Invitrogen, Carlsbad, CA) in accordance with the protocol. The quantity and purity of the total RNA extracted were determined with the RIN value of >7 by using an ND-1000 Nanodrop (Thermo Fisher Scientific, Wilmington, DE) and an Agilent 2200 Tapestation (Agilent, Waldbronn, Germany). RNA integrity was examined by agarose gel electrophoresis. rRNA

was removed from total RNA by using an Epicentre Ribo-Zero rRNA removal kit (Catalog No. MRZH11124, Illumina, San Diego, CA). rRNA-depleted RNA was randomly fragmented into short segments and instantly used as templates to synthesize first-strand cDNA by random hexamers. The second-strand cDNA was synthesized, and RNA was digested using RNase H. The double-stranded DNA was purified and added with base A and adaptor in the 3′ end. After selection via AMPureXP beads, the second-strand cDNA containing U was degraded by the enzyme USER (Catalog No. M5505, New England Biolabs, Ipswich, MA). Finally, the cDNA library for sequencing was constructed by PCR amplification. The quality of the cDNA library was also inspected, and sequencing was performed on an Illumina Hiseq3000. Finally, the sequencing reads were produced with the length of pair-ended to be 2 × 150 bp (PE150).

Transcriptome assembly and lncRNA identification

Prior to assembly, the quality of raw reads was evaluated by FastQC and cleaned by removing adaptor sequences and poor-quality reads. The clean data were mapped to the *L. migratoria* genome (version 2.4.1, available at http://159.226.67.243/) by using the read aligner HISAT2 (version 2.1.0). Next, the transcriptome was assembled by the StringTie (version 1.3.1) on the basis of the reads mapped to the locust genome. The assembled transcripts were annotated by using Cuffcompare from the Cufflinks package. lncRNA screening pipeline was developed as follows. (1) The known protein-coding transcripts according to the annotation of the reference genome were discarded. (2) The transcripts with length of < 200 bp, single exon, and FPKM of < 0.1 were filtered out. (3) The transcripts with class code " = ", "j", "p", and "s" were excluded. (4) The transcripts with protein-coding potential [Coding-Potential Assessment Tool (CPAT) score > 0.38, Coding-Non-Coding Index (CNCI) score > 0, and Coding Potential Calculator (CPC) score > 0] were removed [47–49]. (5) The transcripts with similarity to known protein-coding domains in the Pfam database (version 31.0, E-value < 1E−3) were eliminated. (6) The transcripts with FPKM of < 1.0 were discarded. Finally, the remaining transcripts were classified as lncRNAs.

Differential expression analysis of lncRNAs and mRNAs

The FPKMs of all transcripts in each sample were calculated by using StringTie (version 1.3.1). The differential expression analysis of lncRNAs and mRNAs between two conditions was performed using the DESeq2 (version 1.10.1) package in the R project. The transcripts with $P < 0.05$ and absolute FC > 2.0 were considered DE.

qRT-PCR

qRT-PCR was performed to validate the lncRNA expression in samples. For each sample, 2 μg of total RNA was reversely transcribed to cDNA with random primers by using Moloney murine leukemia virus reverse transcriptase (Catalog No. M1705, Promega, Madison, WI). Then, qPCR was performed using the SYBR Green I Master (Catalog No. 4887352001, Roche, Mannheim, Germany) on a Light Cycler 480 instru-

ment (Roche) to evaluate the expression level. Six independent biological replicates were used for each condition in expression analysis. Then, the relative expression levels of transcripts were calculated using the $2^{-\Delta\Delta Ct}$ method with RP49 as the endogenous reference gene. The melting curve was used to validate unique amplification, and qPCR amplicons were sequenced to confirm the specific products. The absolute expression levels of lncRNAs were calculated in accordance with the standard curves derived via absolute qPCR. The standard curves presented the correlation between the Ct values and real RNA concentrations. Table S8 lists the primers used.

Expression specificity analysis

Index τ was used to measure specific expression [50] as follows:

$$\tau_i = \frac{\sum_{j=1}^{n}\left\{1 - \frac{log_2\left[S_{(i,j)}+1\right]}{log_2\left[S_{(i,\max)}+1\right]}\right\}}{n-1} \tag{1}$$

where n is the number of samples tested, $S_{(i,j)}$ is the FPKM of transcript i in sample j, and $S_{(i,\max)}$ is the highest FPKM of the transcript i in n samples.

Time-series expression profile analysis

The DE transcripts within crowding and isolation treatments for different time points were used to perform the expression pattern analysis by STEM (version 1.3.11). A set of 26 distinct temporal expression profiles independent of the data were selected to summarize the transcript expression change. The expression levels of transcripts were normalized [log_2(0 h/0 h), log_2(4 h/0 h), log_2(8 h/0 h), and log_2(16 h/0 h)] before being clustered. lncRNAs or mRNAs were assigned to the model profiles that most closely represented their expression patterns as determined by the correlation coefficient ($r > 0.7$). The significance levels at which the number of transcripts assigned to a model profile compared with the expected number of transcripts assigned were calculated for each model profile. The profiles with $P < 0.05$ were identified as significant temporal expression profiles, which obviously responded to the treatment.

lncRNA–mRNA co-expression network construction

The correlation coefficients of lncRNAs and mRNAs were evaluated on the basis of normalized expression (FPKM) by using the Pearson correlation method in R software. The expression levels of the transcripts at each time point were computed by averaging the FPKMs in three biological replications. The protein-coding genes with correlation coefficients of > 0.9 (positive) or < −0.9 (negative) with the corresponding lncRNAs were considered co-expressed genes. These correlations between lncRNAs and mRNAs were put into Cytoscape software (v3.6.1) to construct co-expression networks.

Pathway analysis

KEGG enrichment analyses were performed using the Enrichment widget implemented in the Locust Mine database (http://www.locustmine.org:8080/locustmine) [51]. Pathways and biological processes with $P < 0.05$ were regarded as enriched.

Rapid amplification of cDNA ends

The 5′ and 3′ Rapid amplification of cDNA ends (RACE) experiments were performed following the instruction manual (Catalog No. 634923, SMARTer RACE cDNA amplification kit, Clontech, Mountain View, CA). The nested RACE was performed on the 50-fold diluted templates. The touchdown PCR programs were used to amplify specific target sequences. Table S8 shows the primers used.

RNAi

The dsRNAs of *GFP* and lncRNAs for RNAi were prepared using T7 RiboMAX express RNAi system (Catalog No. P1700, Promega, Madison, WI). Then, 69 nl of dsRNA for each treatment was microinjected into the middle of the brain of each locust. Then, the gregarious locusts were returned to the cage for rearing. Every treatment included 36–42 locusts. After 3 days, the brains of the locusts were sampled as six biological replicates with 6–7 locusts in each one, followed by RNA purification and qRT-PCR to confirm RNAi effectivity.

Behavior test in the arena

Behavioral assay was conducted in a rectangular Perspex arena (40 cm × 30 cm × 10 cm) with opaque walls in accordance with a previous study [25]. Two similar campers (7.5 cm × 30.0 cm × 10.0 cm) were on each side of the arena, with one containing 30 fourth-instar gregarious nymphs and the other being empty. The locusts were gently placed into the assay arena through a cylinder beneath it and recorded for 300 s after visual appearance. The video recording and behavioral data analysis were performed using an EthoVision video tracking system (Noldus, Wageningen, The Netherlands).

Five different behavioral parameters including attraction index (AI), duration in the attraction zone, duration in the repulsion zone, TDMV, and TDM were extracted and regarded as categorical behavior markers. The behavioral phenotypes of locusts were analyzed by applying the binary logistic regression model described in our previous study [25]. The regression model was described as follows.

$$P_{greg} = e^{\eta}/(1 + e^{\eta}); \eta$$
$$= -2.11 + 0.005 \times AI + 0.012 \times TDM$$
$$+ 0.015 \times TDMV \qquad (2)$$

where P_{greg} represents the probability that a locust is in the gregarious phase ($P_{greg} = 1$ indicates that the locust is fully gregarious, whereas $P_{greg} = 0$ indicates fully solitarious behavior), AI indicates the extent to which the tested locusts are attracted by the gregarious locusts. AI = total duration in attraction zone (near the camper containing locusts) − total duration in repulsion zone (near the empty camper); TDMV and TDM indicate the locomotor activity levels. In the behavioral assay, more than 29 locust individuals were tested for each treatment.

Statistical analysis

Kolmogorov–Smirnov test was conducted to analyze differences between lncRNAs and mRNAs in length, and the numbers of exons and introns. Pearson's chi-squared test was performed to compare the percentages of differentially changed lncRNAs and mRNAs. Student's *t*-test was used to analyze the effectiveness of RNAi and the changes in specific parameters in the arena assay. The data of behavioral phase change in the arena assay were analyzed by Mann–Whitney *U* test for their non-normal distribution characteristics.

Data availability

The Fastq files of the strand-specific transcriptome sequence for 24 samples in time course have been deposited in the Genome Sequence Archive [52] at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation (GSA: CRA002379), and are publicly accessible at http://bigd.big.ac.cn/gsa.

CRediT author statement

Ting Li: Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Bing Chen:** Writing - original draft, Writing - review & editing. **Pengcheng Yang:** Software, Validation, Data curation, Resources. **Depin Wang:** Formal analysis, Software. **Baozhen Du:** Software, Resources. **Le Kang:** Conceptualization, Supervision, Writing - review & editing. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gpb.2020.05.001.

ORCID

0000-0002-5503-5396 (Ting Li)
0000-0002-1238-3948 (Bing Chen)
0000-0001-5496-8357 (Pengcheng Yang)
0000-0003-2764-5477 (Depin Wang)
0000-0003-4197-818X (Baozhen Du)
0000-0003-4262-2329 (Le Kang)

References

[1] Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012;22:1775–89.

[2] Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell 2013;154:26–46.

[3] Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res 2012;22:577–91.

[4] Marques AC, Ponting CP. Intergenic lncRNAs and the evolution of gene expression. Curr Opin Genet Dev 2014;27:48–53.

[5] Graveley BR, Brooks AN, Carlson J, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome of *Drosophila melanogaster*. Nature 2011;471:473–9.

[6] Luo H, Sun S, Li P, Bu D, Cao H, Zhao Y. Comprehensive characterization of 10,571 mouse large intergenic noncoding RNAs from whole transcriptome sequencing. PLoS One 2013;8: e70835.

[7] Yao RW, Wang Y, Chen LL. Cellular functions of long noncoding RNAs. Nat Cell Biol 2019;21:542–51.

[8] Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904–14.

[9] Liao Q, Liu CN, Yuan XY, Kang SL, Miao RY, Xiao H, et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res 2011;39:3864–78.

[10] Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 2008;18:1433–45.

[11] Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature 2014;505:635–40.

[12] Qureshi IA, Mehler MF. Non-coding RNA networks underlying cognitive disorders across the lifespan. Trends Mol Med 2011;17:337–46.

[13] Briggs JA, Wolvetang EJ, Mattick JS, Rinn JL, Barry G. Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. Neuron 2015;88:861–77.

[14] Yi J, Chen B, Yao X, Lei Y, Ou F, Huang F. Upregulation of the lncRNA MEG3 improves cognitive impairment, alleviates neuronal damage, and inhibits activation of astrocytes in hippocampus tissues in Alzheimer's disease through inactivating the PI3K/Akt signaling pathway. J Cell Biochem 2019;120:18053–65.

[15] Spadaro PA, Flavell CR, Widagdo J, Ratnu VS, Troup M, Ragan C, et al. Long noncoding RNA-directed epigenetic regulation of gene expression is associated with anxiety-like behavior in mice. Biol Psychiatry 2015;78:848–59.

[16] Li M, Wen S, Guo X, Bai B, Gong Z, Liu X, et al. The novel long non-coding RNA *CRG* regulates *Drosophila* locomotor behavior. Nucleic Acids Res 2012;40:11714–27.

[17] Soshnev AA, Ishimoto H, McAllister BF, Li X, Wehling MD, Kitamoto T, et al. A conserved long noncoding RNA affects sleep behavior in *Drosophila*. Genetics 2011;189:455–68.

[18] Humann FC, Tiberio GJ, Hartfelder K. Sequence and expression characteristics of long noncoding RNAs in honey bee caste development – potential novel regulators for transgressive ovary size. PLoS One 2013;8:e78915.

[19] Liu F, Guo D, Yuan Z, Chen C, Xiao H. Genome-wide identification of long non-coding RNA genes and their association with insecticide resistance and metamorphosis in diamondback moth, *Plutella xylostella*. Sci Rep 2017;7:15870.

[20] Xiao H, Yuan Z, Guo D, Hou B, Yin C, Zhang W, et al. Genome-wide identification of long noncoding RNA genes and their potential association with fecundity and virulence in rice brown planthopper, *Nilaparvata lugens*. BMC Genomics 2015;16:749.

[21] Chen RT, Xiao Y, Liu Z, Li LL, Lu Y, Jiao P, et al. Three vital RNA functions and interactions in the process of silk gland apoptosis in silkworm *Bombyx mori*. Arch Insect Biochem Physiol 2019;100:e21511.

[22] Maleszka R. Epigenetic code and insect behavioural plasticity. Curr Opin Insect Sci 2016;15:45–52.

[23] Pener MP, Simpson SJ. Locust phase polyphenism: an update. Adv In Insect Phys 2009;36:1–272.

[24] Wang XH, Fang XD, Yang PC, Jiang XT, Jiang F, Zhao DJ, et al. The locust genome provides insight into swarm formation and long-distance flight. Nat Commun 2014;5:2957.

[25] Guo W, Wang X, Ma Z, Xue L, Han J, Yu D, et al. *CSP* and *takeout* genes modulate the switch between attraction and repulsion during behavioral phase change in the migratory locust. PLoS Genet 2011;7:e1001291.

[26] Ma Z, Guo W, Guo X, Wang X, Kang L. Modulation of behavioral phase changes of the migratory locust by the catecholamine metabolic pathway. Proc Natl Acad Sci U S A 2011;108:3882–7.

[27] Hou L, Li B, Ding D, Kang L, Wang X. CREB-B acts as a key mediator of NPF/NO pathway involved in phase-related locomotor plasticity in locusts. PLoS Genet 2019;15:e1008176.

[28] Hou L, Yang P, Jiang F, Liu Q, Wang X, Kang L. The neuropeptide F/nitric oxide pathway is essential for shaping locomotor plasticity underlying locust phase transition. Elife 2017;6:e22526.

[29] Yang M, Wei Y, Jiang F, Wang Y, Guo X, He J, et al. MicroRNA-133 inhibits behavioral aggregation by controlling dopamine synthesis in locusts. PLoS Genet 2014;10:e1004206.

[30] He J, Chen QQ, Wei YY, Jiang F, Yang ML, Hao SG, et al. MicroRNA-276 promotes egg-hatching synchrony by up-regulating *brm* in locusts. Proc Natl Acad Sci U S A 2016;113:584–9.

[31] Lee GS, Nguyen P, Choi AY, Kim AY, Yu Y, Jung JK, et al. Coding and long non-coding RNAs regulating adult migratory locust (*Locusta migratoria*) brain polyphenism revealed via whole transcriptome analyses. J Asia Pac Entomol 2018;21:58–68.

[32] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using *phred*. I. accuracy assessment. Genome Res 1998;8:175–85.

[33] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323.

[34] Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. Cell 2018;172:393–407.

[35] Dominski Z, Kole R. Selection of splice sites in pre-mRNAs with short internal exons. Mol Cell Biol 1991;11:6075–83.

[36] Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, et al. Diversity and dynamics of the *Drosophila* transcriptome. Nature 2014;512:393–9.

[37] Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, et al. Genome-wide analysis of long noncoding RNA stability. Genome Res 2012;22:885–98.

[38] Liu SL, Wang ZB, Chen D, Zhang BW, Tian RR, Wu J, et al. Annotation and cluster analysis of spatiotemporal- and sex-related lncRNA expression in rhesus macaque brain. Genome Res 2017;27:1608–20.

[39] Zhou J, Fan Y, Chen H. Analyses of long non-coding RNA and mRNA profiles in the spinal cord of rats using RNA sequencing during the progression of neuropathic pain in an SNI model. RNA Biol 2017;14:1810–26.

[40] Hu S, Wang X, Shan G. Insertion of an Alu element in a lncRNA leads to primate-specific modulation of alternative splicing. Nat Struct Mol Biol 2016;23:1011–9.

[41] Wang J, Gong C, Maquat LE. Control of myogenesis by rodent SINE-containing lncRNAs. Genes Dev 2013;27:793–804.

[42] Maag JL, Panja D, Sporild I, Patil S, Kaczorowski DC, Bramham CR, et al. Dynamic expression of long noncoding RNAs and repeat elements in synaptic plasticity. Front Neurosci 2015;9:351.

[43] Smith MA, Mattick JS. Structural and functional annotation of long noncoding RNAs. Methods Mol Biol 2017;1526:65–85.

[44] Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 2011;470:284–8.

[45] Lan T, Yuan K, Yan X, Xu L, Liao H, Hao X, et al. lncRNA SNHG10 facilitates hepatocarcinogenesis and metastasis by modulating its homolog SCARNA13 via a positive feedback loop. Cancer Res 2019;79:3220–34.

[46] Carmona S, Lin B, Chou T, Arroyo K, Sun S. lncRNA *Jpx* induces *Xist* expression in mice using both *trans* and *cis* mechanisms. PLoS Genet 2018;14:e1007378.

[47] Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 2013;41:e74.

[48] Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 2007;35: W345–9.

[49] Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res 2013;41:e166.

[50] Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 2005;21:650–9.

[51] Yang P, Hou L, Wang X, Kang L. Core transcriptional signatures of phase change in the migratory locust. Protein Cell 2019;10:883–901.

[52] Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome Sequence Archive. Genomics Proteomics Bioinformatics 2017;15:14–8.