

SCIENTIFIC REPORTS



Transcriptional profiling of long non-coding RNAs in mantle of *Crassostrea gigas* and their association with shell pigmentation

Dandan Feng¹, Qi Li^{1,2}, Hong Y¹, Lingfeng Kong¹ & Shaojin D³

Long non-coding RNAs (lncRNAs) play crucial roles in diverse biological processes and have drawn extensive attention in the past few years. However, lncRNAs remain poorly understood about expression and roles in *Crassostrea gigas*, a potential model organism for marine molluscan studies. Here, we systematically identified lncRNAs in the mantles of *C. gigas* from four full-sib families characterized by white, black, golden, and partially pigmented shell. Using poly(A)-independent and strand-specific RNA-seq, a total of 441,205,852 clean reads and 12,243 lncRNA transcripts were obtained. lncRNA transcripts were relatively short with few exons and low levels of expression in comparison to protein coding mRNA transcripts. A total of 427 lncRNAs and 349 mRNAs were identified to differentially express among six pairwise groups, mainly involving in biomineralization and pigmentation through functional enrichment. Furthermore, a total of 6 mRNAs and their *cis*-acting lncRNAs were predicted to involve in synthesis of melanin, carotenoid, tetrapyrrole, or ommochrome. Of them, chorion peroxidase and its *cis*-acting lincRNA TCONS_00951105 are implicated in playing an essential role in the melanin synthetic pathway. Our studies provided the first systematic characterization of lncRNAs catalog expressed in oyster mantle, which may facilitate understanding the molecular regulation of shell colour diversity and provide new insights into future selective breeding of *C. gigas* for aquaculture.

The large proportion of a eukaryotic genome is transcribed to produce a huge array of RNA molecules differing in protein-coding capability, size, and abundance¹. Over the past decade, with the development of next-generation sequencing techniques, genome-wide transcriptome analysis, it was discovered that the genomes of eukaryotes encode a vast range of non-protein coding RNAs (ncRNAs)^{2,3}. ncRNAs comprised of different types of small RNA (sRNA) and long noncoding RNAs (lncRNAs) that have been implicated in transcriptional and post-transcriptional regulation of gene expression or in guiding DNA modification⁴. Thousands of lncRNAs have been characterized in a limited number of eukaryotes. lncRNAs showed generally lower expression level, shorter length compared with counterpart mRNAs^{2,3,5,6}. As expected for regulatory molecules, lncRNAs display specific spatiotemporal expression patterns, high tissue specificity and can regulate expression of genes in close genomic proximity (*cis*-acting) or at distance (*trans*-acting)^{3,7}.

Many lncRNAs have been shown to play crucial roles in diverse biological processes^{5,7}. Emerging evidence indicated that lncRNAs may have important roles in pigmentation. For example, the whole transcriptome analysis of pigmented and non-pigmented skin suggests a possible functional relevance of lncRNA in the modulation of pigmentation processes both in bovine⁸ and goat⁹. In goat, the impact of lncRNAs on its target genes in *cis* and *trans* was investigated, indicating that these lncRNAs have a strict tissue specificity and functional conservation⁹. Study on lncRNAs and their *cis*-target genes in melanocytes suggested the role in the melanogenesis¹⁰.

The fabulous and diverse colours of molluscan shells are generally believed to be determined by presence of biological pigments. The widely recognized shell colour diversities have been appreciated for hundreds of years

¹Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao, 266003, China.

²Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, 266237, China. ³

Sample name	BSM	GSM	NSM	WSM
Raw reads	99,758,320	109,996,464	119,299,572	136,748,678
Clean reads	93,987,492	103,305,608	112,944,804	130,967,948
Clean bases	14.1 G	15.5 G	16.94 G	19.65 G
Error rate(%)	0.02	0.01	0.02	0.01
Q20(%)	96.91	97.1	96.96	97.6
Q30(%)	92.47	92.81	92.57	93.88
GC content(%)	44.08	46.57	44.08	45.16
Total mapped	70571864 (75.09%)	81660962 (79.05%)	84586510 (74.89%)	100807144 (76.97%)
Multiple mapped	7312694 (7.78%)	9367467 (9.07%)	7700828 (6.82%)	10709954 (8.18%)
Uniquely mapped	63259170 (67.31%)	72293495 (69.98%)	76885682 (68.07%)	90097190 (68.79%)
Read-1	32615939 (34.7%)	37203463 (36.01%)	39517852 (34.99%)	45957811 (35.09%)
Read-2	30643231 (32.6%)	35090032 (33.97%)	37367830 (33.09%)	44139379 (33.7%)
Reads map to '+'	31768580 (33.8%)	36272547 (35.11%)	38639356 (34.21%)	45154870 (34.48%)
Reads map to '-'	31490590 (33.51%)	36020948 (34.87%)	38246326 (33.86%)	44942320 (34.32%)
Non-splice reads	41421851 (44.07%)	50059318 (48.46%)	52748486 (46.7%)	58198638 (44.44%)
Splice reads	21837319 (23.23%)	22234177 (21.52%)	24137196 (21.37%)	31898552 (24.36%)

Table 1. Basic characteristic of reads in four libraries and data of sequencing reads mapping to the reference genome.

tRNA, rRNA, snoRNA, snRNA, pre-miRNA, and pseudogenes by cuffcompare. (v) The remaining transcripts were blasted with known mRNA and completed the preliminary screening. (vi) Classify candidate lncRNAs into three subtypes (lincRNA, intronic lncRNA, and antisense lncRNA) using information of *calss_code* of cuffcompare. (vii) The tools of CPC (Coding Potential Calculator), CPAT (Coding-Potential Assessment Tool), Pfamscan were used to detect putative protein encoding transcripts, potential lncRNA transcripts were retained which are not detected in any tool. (viii) Select putative lncRNAs which can be detected in at least three libraries.

Characterization and quantification of transcripts. RepeatMasker (<http://www.repeatmasker.org>) was used with default parameters to identify various TE components in oyster.

Cis role is lncRNA acting on neighboring target genes. For the *cis* action of lncRNAs, we searched for protein-coding genes 100 kb upstream and downstream of the lncRNAs, respectively. Cuffdiff (v2.1.1) was used to calculate FPKMs of both lncRNAs and coding genes transcripts in each sample²⁸. Transcripts with *P*-adjust < 0.05 and the absolute value of log₂ (Fold change) > 1 were described as differentially expressed between any two shell colours, which were profiled as differentially expressed transcripts (DETs). Differentially expressed mRNA assemblies (DEM)²⁰ were also independently analyzed and recorded in six pairwise groups, which were detected from the same four shell colours oyster lines. Shared differentially expressed genes from the same two pairwise groups were retained.

Functional enrichment analysis. Gene Ontology (GO) enrichment analysis of differentially expressed genes or lncRNA target genes was implemented by the Goseq R package, in which gene length bias was corrected. GO terms with corrected *P* value less than 0.05 were considered significantly enriched by differential expressed genes.

KEGG is a database resource for understanding high-level functions and utilities of the biological system. We used KOBAS software to test the statistical enrichment of differentially expressed genes or lncRNA target genes in KEGG pathways. Hypergeometric *P* value < 0.05 was considered significant.

Validation of gene expression by quantitative PCR analysis. To validate the RNA-seq data, 16 differentially expressed transcripts of interest were selected for quantitative real-time PCR (qPCR) analysis. Total RNA was extracted separately from the same 24 samples used for RNA sequencing. Then cDNA was synthesized from RNA, which was used for qPCR, using Prime Script™ RT reagent Kit with gDNA Eraser (TaKaRa, Dalian, China). Specific primers for qPCR were designed using Premier Primer 5 (Supplementary Table S1) and verified by NCBI primer-BLAST. Elongation Factor was used as an endogenous control²⁹. The amplification was performed on the LightCycler 480 real-time PCR instrument (Roche Diagnostics, Burgess Hill, UK) using SYBR® Premix Ex Taq™ (TaKaRa). Cycling parameters were 95 °C for 5 min, then 40 cycles of 95 °C for 5 s, 60 °C for 20 s. Melting curve analyses were performed following amplifications to verify specific amplification. Relative gene expression data was analyzed using the comparative threshold cycle (CT) method³⁰. Data were examined for homogeneity of variances (F text), and were analyzed by t test using software SPSS 13.0 with *P* < 0.05. Eight oysters, typically having black for the left shell and white for the right shell, were also picked up for qPCR.

Results

Identification and characterization of lncRNAs in oyster mantles. The number of RNA-seq reads, quality of the reads, and the mapping rate for each of the four libraries sequenced are summarized (Table 1). A total of 99,092 transcripts were assembled by the Cufflinks, which were used for subsequent analysis. Using the criteria shown in Fig. 1a, 12,243 lncRNAs transcripts expressed in mantle were identified from at least three of

the four samples analyzed. They consist of 8,226 lincRNAs, 387 antisense lincRNAs, and 3,630 intronic lincRNAs (Fig. 1b). These lincRNA transcripts correspond to 11,637 lincRNA gene loci. In addition, 45,393 protein-coding transcripts were also identified.

The exon number, sequence length, open reading frame length, and expression levels were characterized for the obtained 12,243 lincRNAs and 45,393 mRNAs. Our results indicated that most of lincRNAs contained fewer exons (one or two) than mRNAs (Fig. 2a). The distribution of transcript length was obviously different. The average length of lincRNAs was shorter than that of mRNAs (Fig. 2b), and the lincRNAs in our dataset tend to be shorter in open reading frame length than mRNAs (Fig. 2c). In addition, lincRNAs exhibited a lower level of expression than mRNA (Fig. 2d). With the absence of lincRNAs data in Mollusca, we failed to properly evaluate the conservation of lincRNAs in the oyster.

Furthermore, we identified differences in Transposable element (TE) components between lincRNAs and mRNAs, as well as among the three lincRNA subtypes. Our analysis revealed TE component characteristics that distinguished the four transcripts subtypes (Fig. S1; Supplementary Table S2). At the global level, Class II DNA transposons, minisatellite, and rollingcircle helitron (RC/Helitron) were the three most abundant known repetitive elements to overlap with oyster transcripts. Significant differences in the percentage of TE components were observed between mRNAs and the individual subtype of lincRNAs. A total of 23,396 TEs were found in 8,281 lincRNAs, which account for 67.64% (8,281/12,243) of the total number of lincRNAs. A total of 521,392 TEs were found in 40,120 mRNAs, which account for 88.38% (40,120/521,390) of the total number of mRNAs. The results revealed that TE percentage in oyster is considerably lower 510.999927base lncmRNNo0000033.20000004(w)7.9000001(o) t-

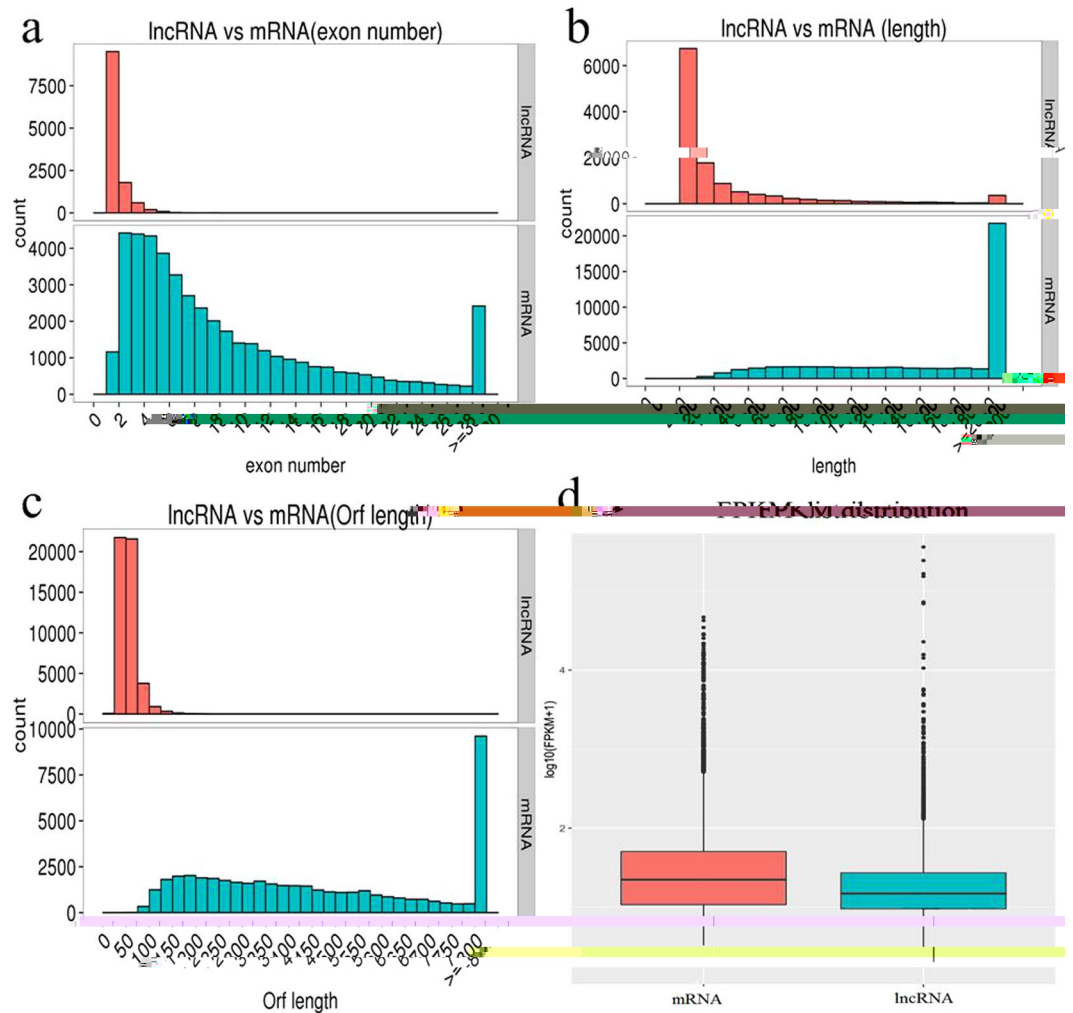


Figure 2. Comparison of the identified lncRNAs and mRNAs. **(a)** Distribution of the number of exons in the mRNAs and lncRNAs. **(b)** Distribution of transcript lengths in the mRNAs and lncRNAs. **(c)** Distribution of open reading frame lengths in the mRNAs and lncRNAs. **(d)** Expression level analysis in the mRNAs and lncRNAs.

in a total of 349 significantly DEMs (Supplementary Tables S3 and S6). A total of 6 mRNAs were confirmed by qPCR (Supplementary Table S1). These 6 genes are known to play essential roles in pigment biosynthesis of melanin, tetrapyrrole, carotenoid, and ommochrome (Table 2; Fig. 3).

GO and KEGG analyses were also performed on 349 significantly differentially expressed mRNA. As the results, we derived 13 highly enriched GO terms (Supplementary Table S7) and 10 significantly enriched KEGG pathways (Supplementary Table S8). Importantly, we also observed several pigment biosynthesis related terms, such as “tyrosine metabolism”, “tryptophan metabolism”, and “retinol metabolism” in the data from KEGG analyses (Fig. 4).

Among the significantly differentially expressed lncRNA and mRNA transcripts, a total of 16 transcripts, including 6 lncRNAs and 10 mRNAs, were selected to confirm the utility of RNA-seq for quantitative analyses using quantitative polymerase chain reaction (qPCR). The results show that these transcripts were differentially expressed among different shell colours oysters and generally exhibited consistent with RNA sequencing data (Supplementary Table S1). Using oysters of asymmetric shell pigment pattern, we found that peroxidase, TCON_00924022, and TCON_00951105 showed a higher expression level in left mantle representing black shell than in right mantle representing white shell. Other six transcripts showed no significant difference.

The *cis* role of lncRNAs in target genes. To investigate the function of lncRNAs, we performed bioinformatics analysis identifying the potential targets of lncRNAs *in cis*. Our analysis included 11,157 lncRNAs that are associated with 24,057 protein-coding genes within a range of 100 kb. We identified 427 differentially expressed lncRNA genes potentially targeting to 2,088 protein coding genes. GO analysis of the lncRNA *cis*-acting mRNA targets revealed two over-represented terms including RNA methyltransferase activity and tRNA methyltransferase activity that are primarily involved in regulation of gene expression (Supplementary Table S7). Pathway analysis showed that these lncRNAs *cis*-acting target genes were mainly enriched in six KEGG pathways involved

Gene_id	Gene description from Swiss prot	Related pigments	LncRNAs in <i>cis</i> -acting			GO_molecular_function_description	
LOC105344040	Tyrosinase-like protein 2	melanin	TCONS_00117832 TCONS_00117906	TCONS_00117563	TCONS_00117745	GO:0016491	oxidoreductase activity
LOC105324831	Tyrosinase-like protein 3	melanin	TCONS_00821172 TCONS_00820688	TCONS_00820679 TCONS_00820896	TCONS_00820814	GO:0016491	oxidoreductase activity
LOC105334556	Dopamine beta-monoxygenase	melanin	TCONS_00909452 TCONS_00907826	TCONS_00909160 TCONS_00909199	TCONS_00909168	GO:0016491	oxidoreductase activity
LOC105324712	Chorion peroxidase	melanin, tetrapyrrole	TCONS_00951105	TCONS_00950402	TCONS_00950769	GO:0046906// GO:0016491	tetrapyrrole binding// oxidoreductase activity
LOC105336634	Cytochrome P450 2U1	carotenoid, melanin tetrapyrrole,	TCONS_00454937 TCONS_00455648	TCONS_00454715	TCONS_00454720	GO:0046906// GO:0016491	tetrapyrrole binding// oxidoreductase activity
LOC105326901	Kynurenine 3-Monooxygenase	melanin ommochrome,	TCONS_00119630 TCONS_00120197	TCONS_00120204 TCONS_00119738	TCONS_00119485	GO:0016491	oxidoreductase activity

Table 2. LncRNAs and their potential *cis*-acting genes that are involved in pigmentation.

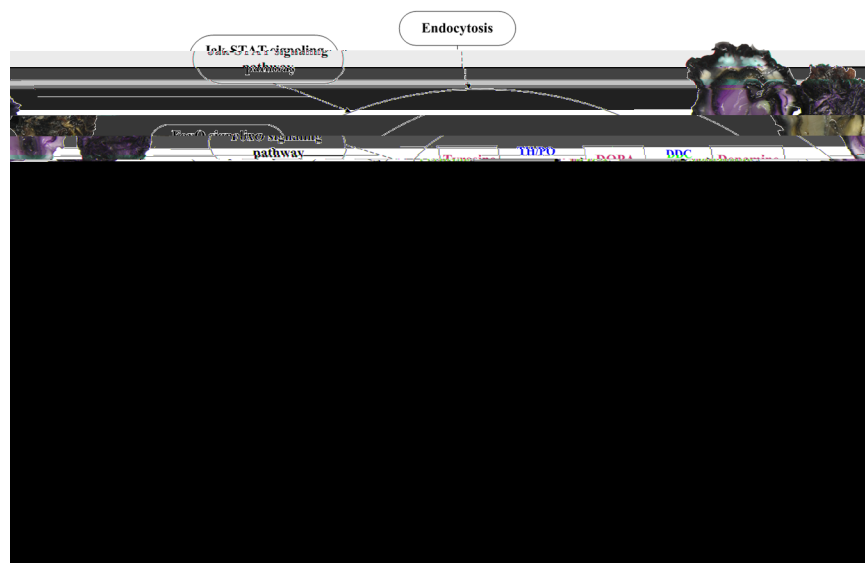


Figure 3. Common pathways of melanin biosynthesis in animals and its related pathways. Enzymes are abbreviated as follows: tyrosine hydroxylase (TH), phenoloxidase (PO), DOPA decarboxylase (DDC), dopachrome tautomerase (DCT), dopachrome conversion enzyme (DCE), dopachrome rearranging enzymes (DRE). Pigment precursors are shown in blue, enzyme are shown in red. Some signaling pathways are indicated around melanin biosynthesis, which were significantly identified in this study and reported to regulate melanin biosynthesis in animals. Some metabolism pathways are also indicated around melanin biosynthesis, which are significantly identified here and reported to involve in other pigments biosynthesis.

in ECM-receptor interaction, Ubiquitin-mediated proteolysis, Jak-STAT signaling pathway, Notch signaling pathway, Homologous recombination and other types of O-glycan biosynthesis (Supplementary Table S8).

Through GO survey using lncRNAs *cis*-acting target genes in six pairwise groups, only one GO term of “Cysteine-type endopeptidase inhibitor” was significantly enriched when comparing BS with WS (Supplementary Table S7). In addition, 16 enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were detected (Supplementary Table S8). Of these pathways, four pathways have been reported to be implicated in biomineralization^{31–33}, including ECM-receptor interaction, Other types of O-glycan biosynthesis, Ubiquitin mediated proteolysis, and Pantothenate and CoA biosynthesis. Five pathways have been previously reported to regulate pigmentation, including the Jak-STAT signaling pathway^{34,35}, Endocytosis^{20,36}, FoxO signaling pathway^{37,38}, and Notch signaling pathway²² (Fig. 3). It is noteworthy that several pigment biosynthesis related terms, such as “Tryptophan metabolism”, and “Porphyrin and chlorophyll II metabolism” were also identified (Figs 3 and 4). Taken together, data from these functional enrichment analyses showed that these *cis*-acting genes of lncRNAs mainly involved in regulation of biomineralization and pigmentation.

Association study. To further ascertain that lncRNA-protein-coding gene pairs exhibited DNA co-localization (for *cis*-acting) and expression correlation relationships, detailed examination was conducted. To deepen our understanding of the relationship between lncRNAs and pigmentation, first, we selectively analyzed

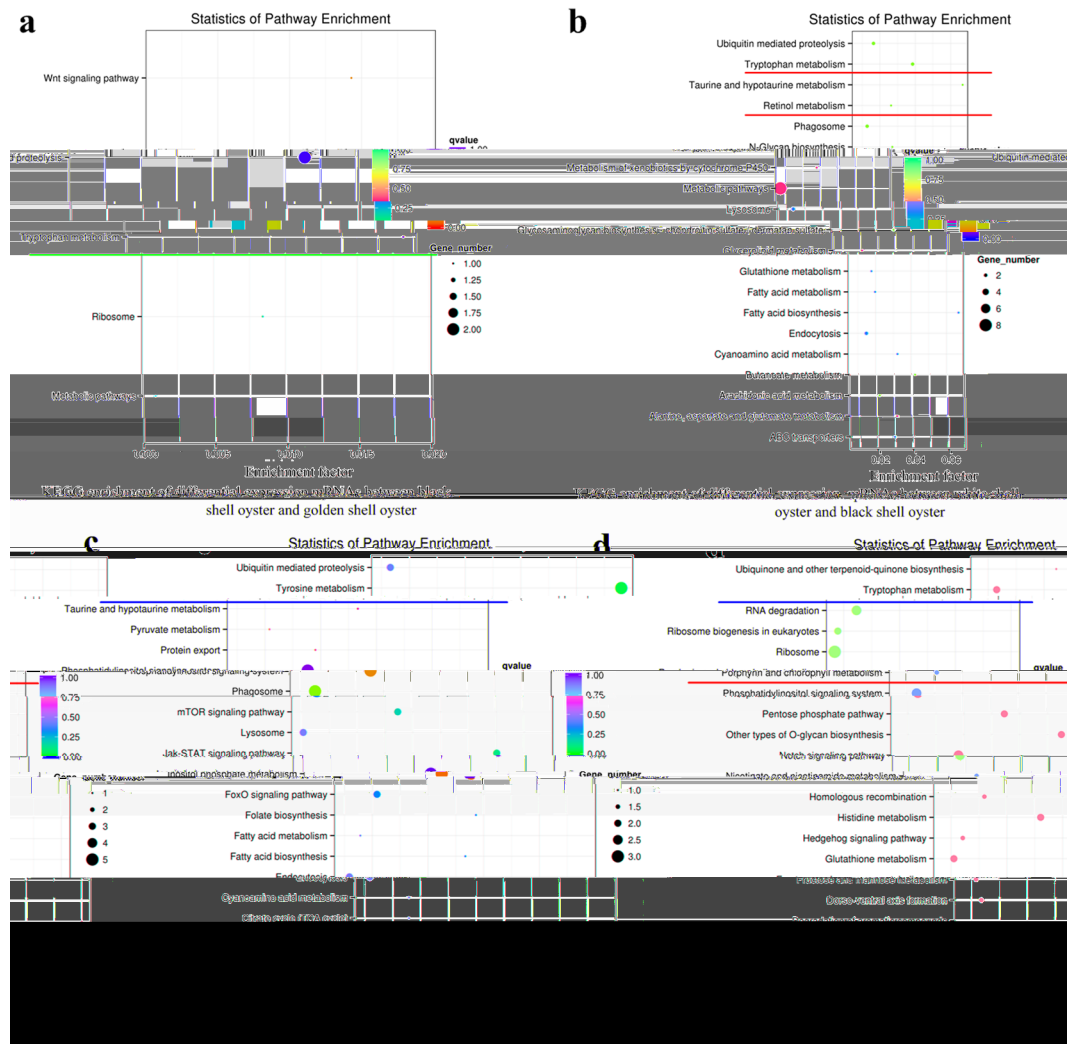


Figure 4. KEGG enrichment analysis of differentially expressed transcripts. The y-axis represented the KEGG enriched pathways, the x-axis represented the enrichment factor, which was calculated by ratio of the number of differentially expressed transcripts divided by the number of annotated transcripts in this pathway. The potential pigmentation-related pathways were underlined by red line.

pairs, in which the lncRNAs and their target genes were significantly differentially expressed between any two shell colours (Table S9). At the same time, gene annotation was used to identify the lncRNA-protein-coding gene pairs associated with pigment biosynthesis. According to these selective criteria, we found that chorion peroxidase, a potential pigment synthesis gene, and its *cis*-acting lincRNA TCONS_00951105 were detected to higher expression levels in pigmented oysters compared to white shell oysters, which were also confirmed in asymmetric oysters (Supplementary Table S1). We predicted that this lincRNA was probably involved in shell pigmentation. However, uncovering the definitive function of the predicted lincRNA requires additional verification studies.

Discussion

In this study, we represented the first long non-coding transcripts catalog expressed in *C. gigas* mantle and analyzed their association with shell pigmentation. Our study not only enriched the knowledge of lncRNAs in marine invertebrate, but also provided new insights into potential functions of lncRNAs in molluscs. These RNA-seq data might provide molecular targets assisting the selective breeding of *C. gigas*.

These lncRNAs share many characteristics of their eukaryotic counterparts: such as shorter length, fewer exons, lower levels of expression compared with mRNAs. Conservation is missing because the lncRNA catalog in other molluscs is inaccessible. A previous study has estimated the relatively low conservation of lincRNA in *C. gigas*². The same characteristics were also detected in lncRNAs found in sponge, goat, and other eukaryotes^{1,3,9}. These common factors of lncRNAs in eukaryotes perhaps indicate their essential regulation during development. In addition to the preliminary examination of lncRNAs, we performed an extensive characterization to reveal major differences in transposon element (TE) components among mRNAs, lincRNAs, intronic lncRNAs, and anti-sense lncRNAs, which may be responsible for the observed differences in their evolution and function. TEs are mobile genetic elements that are capable of movement and proliferation within the genome. TEs are also considered as one of three evolutionary scenarios involved in the origin of lncRNAs³⁹. TE coverage in *C. gigas* is

considerably lower for lncRNAs than for protein-coding mRNAs. Thus, although little is known about repetitive elements in oyster, our findings are consistent with TEs being the origin of protein-coding genes than lncRNAs in *C. gigas*, which has also been proposed in *Amphimedon*³.

Unlike mRNA sequences that could provide potential information regarding their function, the sequence motifs of lncRNA are usually uninformative for predicting lncRNA function given that lncRNA functions are highly complex and diverse⁴⁰. We predicted the potential function of lncRNAs in oyster mantle by analyzing their *cis*-acting protein-coding gene targets. Although this may not be the most appropriate model to explain the function of lncRNAs, GO analysis of all differentially expressed lncRNA *cis*-acting mRNAs identified two overrepresented terms of RNA methyltransferase activity and tRNA methyltransferase activity. RNA methylation has been reported to play a vital role in post-transcriptional regulation of gene expression^{41,42}. Our studies were focused on the characterization of differentially expressed lncRNAs and their *cis*-acting mRNAs, and uncovering their potential functions by GO and KEGG analyses.

It has been reported in mollusk that shell colour is regulated by shell matrix proteins (SMPs) expressed in different shell layers¹¹. While some of these proteins may have a role in shell colour determination, it is possible that these genes may play other roles in shell construction²⁴. Notably, GO functional annotation analysis showed that only one GO term, namely Cysteine-type endopeptidase inhibitor, was significantly enriched, which were extensively characterized in SMPs^{11,32}. It has been suggested that cysteine-type endopeptidase inhibitor might inhibit cysteine-type endopeptidase to degrade shell matrix proteins^{43,44}. Our studies also revealed several enriched pathways that have been implicated in biomineralization, including ECM-receptor interaction, other types of O-glycan biosynthesis and ubiquitin mediated proteolysis^{31,32,44}. Thus, a close relationship between the differentially expressed lncRNAs and biomineralization was observed.

Although pigmentation is a multifactorial phenotypic traits, only a small numbers of pathways regulating pigmentation have been validated to date⁹. Of that, Jak-STAT signaling pathway^{34,35}, Endocytosis^{20,36}, and Notch signaling pathway^{20,22} have been identified in our study. It is worth noting that several pigment biosynthesis related terms, such as tryptophan metabolism and porphyrin and chlorophyll2 metabolism were identified in the GO analysis. Tryptophan is used to synthesize ommochrome and substitute for tyrosine as an oxidizable substrate for melanin^{16,45,46}. Porphyrin, a tetrapyrroles, was one of the shell pigments found in Mollusca. Therefore, data from our functional enrichment analysis suggest a close correlation between the differentially expressed lncRNAs and pigmentation.

Comparison of two RNA-seq datasets identified 349 protein-coding transcripts that were shared between DET and DEM. These mRNAs could be used as auxiliary materials to further investigate the pigmentation associated lncRNAs. A total of 6 mRNAs are selected to function in pigments biosynthesis involving in melanin, tetrapyrrole, carotenoid and ommochrome (Table 2). Melanin is the end-product of complex multistep transformation of tyrosine¹⁵, extensively existing in the organism kingdom. Tyrosinase is the key enzyme in pigment synthesis, initiating a cascade of reactions converting tyrosine to the melanin biopolymer⁴⁷. In insects, multiple enzymes are identified to directly involve in melanogenesis including peroxidase, phenoloxidase (PO), dopachrome conversion enzyme (DCE)^{26,48}. Furthermore, two types of insect POs have been identified in some insect species, that can be identified as tyrosinase-like and laccase-type²⁶. In cephalopod, the melanin-producing pathway in the ink gland includes three main enzymes of tyrosinases, peroxidase and dopachrome rearranging enzymes⁴⁹. Porphyrins, termed as cyclic structure tetrapyrroles, are found in bacteria, plants, and animals and are synthesized via the haem pathway¹¹. Carotenoids can be transformed to apocarotenoids such as retinoids, whose metabolic were reported to be mediated by Cytochrome P450s^{50–53}. Kynurenine 3-monooxygenase catalyses the hydroxylation of kynurenine to 3-hydroxykynurenine, which has a key role in tryptophan catabolism and synthesis of ommochrome pigments^{54,55}.

Several studies indicate that the intricate mechanisms of pigmentation require a coordinated posttranscriptional regulatory network of genes expression. However, our knowledge on the role of lncRNAs in pigmentation is very limited^{40,56}. Our studies demonstrated that chorion peroxidase and its *cis*-acting lincRNA TCONS_00951105 showed the highest expression level in black shell oyster. Chorion peroxidase was initially identified from *Drosophila melanogaster* and reported to relate to eggshell chorion harden involving protein crosslinking and melanization in insects^{57,58}. Moreover, higher levels of expression were found in left black mantle relative to right white mantle of oysters with asymmetric pattern of shell colour. Peroxidase has been suggested to serve in an alternative melanogenic pathway in insect and cephalopod. Peroxidase is associated with melanosomes in the ink gland, where it is thought to catalyze the formation of eumelanin^{49,59}. Peroxidase has been identified in many DGE datasets^{20,24}, strongly suggesting its role in shell pigmentation. Phylogenetic tree using 26 peroxidase in *C. gigas* showed the chorion peroxidase LOC105324712 (CGI 10011763) clustered with peroxidases in insect and cephalopod, which have been implicated in melanin biosynthesis³³. Altogether, our studies suggest that chorion peroxidase and its *cis*-acting lincRNA TCONS_00951105 may play an important role in melanin synthesis and shell colour regulation (Fig. 3).

Conclusion

This study provided a catalog of lncRNAs in mantle of five-month-old Pacific oysters and profiled their expression in four shell colours variants. We identified a total of 12,443 lncRNAs, encoded by 11,637 gene loci, consisting of 8,226 lincRNAs, 387 antisense lncRNAs, and 3,630 intronic lncRNAs. lncRNA transcripts showed a relatively short length with fewer exons and low expression relative to their counterpart protein coding RNA (mRNA) transcripts. We identified 427 lncRNA transcripts that are differentially expressed among six pairwise groups based on one replicate per sib family. Functional enrichment of differentially expressed lncRNA and mRNA transcripts showed that they are potentially associated with biomineralization and shell pigmentation. And a total of 6 mRNAs are identified to influence pigment biosynthesis including melanin, carotenoid, tetrapyrrole, and ommochrome. Finally, we selectively analyzed lncRNAs and target gene pairs, in which the lncRNAs and

their target genes were differentially expressed between any two shell colours variants. Chorion peroxidase, a pigmentation associated gene, was found to be the *cis*-acting target of lincRNA (TCONS_00951105) simultaneously. Collectively, our studies of *C. gigas* mantle lincRNAs and their association with pigmentation might facilitate the selection of elite oyster lines with desired coloration patterns.

References

1. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* **136**, 629–641 (2009).
2. Yu, H., Zhao, X. & Li, Q. Genome-wide identification and characterization of long intergenic noncoding RNAs and their potential association with larval development in the Pacific oyster. *Scientific Reports* **6** (2016).
3. Gaiti, F. *et al.* Dynamic and widespread lincRNA expression in a sponge and the origin of animal complexity. *Molecular biology and evolution* **32**, 2367–2382 (2015).
4. Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics* **2**, 919–929 (2001).
5. Dong, X. *et al.* Comprehensive Identification of Long Non-coding RNAs in Purified Cell Types from the Brain Reveals Functional lincRNA in OPC Fate Determination. *PLOS Genetics* **11** (2015).
6. Karlic, R. *et al.* Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. *DNA research: an international journal for rapid publication of reports on genes and genomes* (2017).
7. Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics* **17**, 47–62 (2016).
8. Weikard, R., Hadlich, F. & Kuehn, C. Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. *BMC genomics* **14**, 789 (2013).
9. Ren, H. *et al.* Genome-wide analysis of long non-coding RNAs at early stage of skin pigmentation in goats (*Capra hircus*). *BMC genomics* **17**, 67 (2016).
10. Zeng, Q. *et al.* Analysis of lincRNAs expression in UVB-induced stress responses of melanocytes. *Journal of Dermatological Science* **81**, 53–60 (2016).
11. Williams, S. T. Molluscan shell colour. *Biological Reviews* (2016).
12. Nijhout, H. F. Molecular and physiological basis of colour pattern formation. *Advances in insect physiology* **38**, 219–265 (2010).
13. Braasch, I., Scharl, M. & Volf, J.-N. Evolution of pigment synthesis pathways by gene and genome duplication in fish. *BMC Evolutionary Biology* **7**, 74 (2007).
14. Grotewold, E. The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* **57**, 761–780 (2006).
15. Slominski, A., Tobin, D. J., Shibahara, S. & Wortsman, J. Melanin pigmentation in mammalian skin and its hormonal regulation. *Physiological Reviews* **84**, 1155–1228 (2004).
16. Christensen, B. M., Li, J., Chen, C. & Nappi, A. J. Melanization immune responses in mosquito vectors. *Trends in Parasitology*

40. Zhang, S. *et al.* Systematic Analysis of Long Noncoding RNAs in the Senescence-accelerated Mouse Prone 8 Brain Using RNA Sequencing. *Molecular Therapy—Nucleic Acids* **5**, e343 (2016).
41. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nature Reviews Genetics* **16**, 71–84 (2015).
42. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m6A RNA methylation. *Nature Reviews Genetics* **15**, 293–306 (2014).
43. Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Research* **38**, 325–331 (2010).
44. Liu, C. *et al.* In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Scientific Reports* **5** (2015).
45. Chauhan, P. *et al.* De novo transcriptome of *Ischnura elegans* provides insights into sensory biology, colour and vision genes. *BMC Genomics* **15**, 808–808 (2014).
46. Wright, T. R. The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster*. *Advances in Genetics* **24**, 127–222 (1987).
47. Marmol, V. D. & Beermann, F. Tyrosinase and related proteins in mammalian pigmentation. *FEBS Letters* **381**, 165–168 (1996).
48. Kronforst, M. R. *et al.* Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Pigment Cell & Melanoma Research* **25**, 411–433 (2012).
49. Derby, C. D. Cephalopod Ink: Production, Chemistry, Functions and Applications. *Marine Drugs* **12**, 2700–2730 (2014).
50. Zheng, H. *et al.* Total carotenoid differences in scallop tissues of *Chlamys nobilis* (Bivalve: Pectinidae) with regard to gender and shell colour. *Food Chemistry* **122**, 1164–1167 (2010).
51. Von Lintig, J. Colors with functions: elucidating the biochemical and molecular basis of carotenoid metabolism. *Annual Review of Nutrition* **30**, 35–56 (2010).
52. White, J. A. *et al.* Identification of the human cytochrome P450, P450RAI-2, which is predominantly expressed in the adult cerebellum and is responsible for all-trans-retinoic acid metabolism. *Proceedings of the National Academy of Sciences* **97**, 6403–6408 (2000).
53. Zhang, Q., Dunbar, D. & Kaminsky, L. S. Human Cytochrome P-450 Metabolism of Retinals to Retinoic Acids. *Drug Metabolism and Disposition* **28**, 292–297 (2000).
54. Linzen, B. The Tryptophan → Ommochrome Pathway in Insects. *Advances in Insect Physiology* **10**, 117–246 (1974).
55. Han, Q. *et al.* Analysis of the wild-type and mutant genes encoding the enzyme kynurenine monooxygenase of the yellow fever mosquito, *Aedes aegypti*. *Insect Molecular Biology* **12**, 483–490 (2003).
56. Koch, L. Functional genomics: Screening for lncRNA function. *Nature Reviews Genetics* (2017).
57. Mindrinos, M. N., Petri, W. H., Galanopoulos, V. K., Lombard, M. F. & Margaritis, L. H. Crosslinking of the *Drosophila* chorion involves a peroxidase. *Wilhelm Roux's archives of developmental biology* **189**, 187–196 (1980).
58. Li, J., Hodgeman, B. A. & Christensen, B. M. Involvement of peroxidase in chorion hardening in *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* **26**, 309–317 (1996).
59. Palumbo, A. Melanogenesis in the Ink Gland of *Sepia officinalis*. *Pigment Cell Research* **16**, 517–522 (2003).

Acknowledgements

This study was supported by the grants from National Natural Science Foundation of China (31772843), Shandong Province (2016ZDJS06A06), and Industrial Development Project of Qingdao City (17-3-3-64-nsh).

Author Contributions

D.D.F. did the experiment, analyzed the data and wrote the paper. Q.L., H.Y., L.F.K. and S.J.D. conceived and designed the study.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-19950-6>.

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

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



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

© The Author(s) 2018