

RESEARCH

Open Access



# Cyclin-dependent kinases (CDKs) are key genes regulating early development of *Neptunea arthritica cumingii*: evidence from comparative transcriptome and proteome analyses

Fengxiao Lv<sup>1</sup>, Xinfan Ge<sup>1</sup>, Yaqing Chang<sup>1,2\*</sup> and Zhenlin Hao<sup>1,2\*</sup>

## Abstract

In this study, we applied comparative transcriptomics and proteomics techniques to systematically investigate the dynamic expression patterns of genes and proteins at various stages of early embryonic development of the gastropod *Neptunea arthritica cumingii*. Twelve cyclin-dependent kinase (CDKs) genes and five downstream proteins associated with these CDKs were identified. Through techniques such as qRT-PCR, our data elucidate for the first time the regulatory functions of CDK family genes and establish CDKs as a pivotal gene cluster in the early embryonic development of *N. cumingii*. These findings not only enhance the understanding of molecular developmental biology in *N. cumingii* and marine gastropods in general but also provide significant insights into the mechanisms involved in early embryonic development in *N. cumingii*. Furthermore, our results provide theoretical guidance for advancing artificial breeding technology for *N. cumingii*.

**Keywords** *N. cumingii*, CDKs, Transcriptome, Proteome, Early development

## Background

The gastropod *Neptunea arthritica cumingii* thrives in temperate seawaters and is primarily found in the Yellow and Bohai Seas of China and in the seas around Korea and Japan. Currently, commercially available *N. cumingii* are predominantly harvested from natural sources, but due to both human activities and natural factors, natural resources of this snail have been declining drastically. According to the Dalian Statistical Yearbook, the annual yield of *N. cumingii* has been only about 6000 tons in recent years, with a market price reaching as high as ¥200 per kilogram [1]. Because of the decreasing production and sustained price increase, market demands are not being met. Therefore, it is imperative to promptly

\*Correspondence:

Yaqing Chang  
yqchang@dlou.edu.cn  
Zhenlin Hao  
haozhenlin@dlou.edu.cn

<sup>1</sup>Key Laboratory of Mariculture and Stock Enhancement in North China's Sea (Dalian Ocean University), Ministry of Agriculture, Dalian 116023, China

<sup>2</sup>Present address: College of Fisheries and Life Science, Dalian Ocean University, 52, Heishijiao Street, Shahekou District, Dalian, Liaoning Province, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

advance artificial breeding technology for *N. cumingii* to increase its production.

Several studies have indicated that the embryonic development of *N. cumingii* is unique. In general, most gastropods undergo indirect development, transforming into veliger larvae and subsequently metamorphosing into juveniles [2]. However, *N. cumingii* exhibits direct development, which is characterized by internal fertilization, external hatching, and direct transformation into juveniles without metamorphosis [3]. The embryonic development period of *N. cumingii* is relatively prolonged, taking approximately 80 days from fertilized egg to juvenile.

During artificial breeding of *N. cumingii* in the laboratory, researchers have observed frequent stagnation of embryonic development. Stagnation often leads to embryo decay and prevents normal hatching, which significantly reduces hatching rates and affects seedling production. Therefore, a comprehensive analysis and clarification of the genetic and regulatory mechanisms underlying *N. cumingii* embryonic development are needed. Identifying the key genes or gene clusters that influence *N. cumingii* embryonic development is essential for developing effective artificial breeding techniques and establishing large-scale *N. cumingii* aquaculture. However, research focused on identifying the molecular dynamics of *N. cumingii* embryonic development and potential molecular markers for breeding has not yet been undertaken.

Transcriptomics is a scientific discipline dedicated to the comprehensive study of the occurrence and relative abundance of diverse RNA transcripts [4, 5]. Due to its well-established detection ability and relatively affordable sequencing costs, it has become the prevailing approach for investigating gene expression regulation in current scientific practice [6]. Proteins are crucial biomolecules that are fundamental to life processes, and proteomics has emerged as a scientific discipline dedicated to the comprehensive analysis of protein composition changes, expression levels, and protein interactions and interconnections at a holistic level [7].

In recent years, transcriptomic and proteomic studies focused on marine gastropod embryo development have been published. For instance, Heyland et al. [8] utilized transcriptomic techniques to assess gene expression patterns, larval stages, and metamorphosis in the sea hare *Aplysia californica* during embryonic development. Lambert et al. [9] studied the transcriptome of the embryonic development phase of the mud snail *Ilyanassa obsoleta* and described the gene expression patterns crucial for organ formation in this species. In another study, Franchini et al. [10] used Illumina high-throughput sequencing technology to sequence the transcriptome of the South African abalone *Haliotis midae*. Their study

yielded the most comprehensive transcriptomic data for this species, which provided essential insights for population and functional genomics research within the abalone genus. Song et al. [11] performed de novo sequencing of the early developmental stages of the whelk *Rapana venosa* using Hi-seq 2500 technology and acquired essential transcriptomic data vital for understanding the early developmental processes of this species. In a study of the protein expression profiles during embryonic development of the marine gastropod *Pomacea canaliculata*, Sun et al. [12] identified a series of candidate protein markers intricately involved in diverse physiological processes crucial for embryonic development. Using intensity-based absolute quantification to conduct proteomic analysis of tissues related to shell formation in the limpet *Lottia gigantea*, Manm and Edsinger [13] provided reliable data essential for subsequent research in cell biology, genetics, materials science, and related fields. Martínez-Fernández et al. [14] performed a comparative proteomic analysis on two ecological types of the same species of periwinkle *Littorina saxatilis*. Using mass spectrometry, they identified differentially expressed proteins associated with energy metabolism in diverse ecological environments. However, transcriptomic and proteomic studies of *N. cumingii* have not been reported to date.

To identify the molecular mechanisms underlying the embryonic development of *N. cumingii* and potential molecular markers for breeding, we applied comparative transcriptomic and proteomic techniques. We meticulously analyzed and compared differential expression of genes and proteins during key early developmental stages of *N. cumingii* to pinpoint candidate genes and proteins intricately linked to its embryonic development. This study will significantly enhance the understanding of molecular developmental biology in *N. cumingii* and marine mollusks and provide invaluable insights into the early embryonic development mechanisms of this species. Moreover, these findings serve as a theoretical groundwork to guide advancements in artificial breeding techniques of marine gastropods.

## Results

### Transcriptome sequencing and quantitation of protein

Twelve RNA sequencing (RNA-seq) libraries (three of each developmental stage), were constructed and then sequenced. Approximately 21.26–23.46 million raw reads were obtained, and about 19.70–21.66 million clean reads were obtained from each sample after trimming. The Q20 of clean reads range was 96.55–97.47%, and the Q30 range was 92.22–93.74%. The GC content range of the samples was 42.50–48.25%.

Clean reads from *N. cumingii* were assembled using Trinity (v2.6.6) software to generate reference sequences for subsequent analyses. The default setting for

min\_kmer\_cov was set to 3. Evaluation of completeness using BUSCO revealed that 87.5% of the unigenes were single-copy genes, while 9.7% were duplicated genes. A total of 314,734 transcripts were obtained, with an average length of 1,057 bp. The N50 and N90 lengths averaged 1,528 bp and 448 bp, respectively. Around 157,119 unigenes were also generated, with an average length of 949 bp and an N50 length of 1,283 bp. Following BLAST annotation, 100% of the unigenes were annotated across seven databases, respectively KEGG (<https://www.genome.jp/kegg/>), GO (<http://geneontology.org/>), NR (<https://www.ncbi.nlm.nih.gov/refseq/>), NT (<https://www.ncbi.nlm.nih.gov/nucleotide/>), SwissProt (<https://www.uniprot.org/>), Pfam (<http://pfam.xfam.org/>), KOG (<http://ekg.verailles.inra.fr/cgi-bin/index.cgi>). A total of 65,904 unigenes were annotated in at least one database, and 3,968 unigenes could be annotated in all databases. Pearson's correlation coefficients for the expected number of fragments per kilobase of transcript sequence per millions base pairs sequenced (FPKM) distributed among the three biological replicates of each treatment ranged from 0.735 to 0.753 (LS), 0.595 to 0.699 (PK), 0.625 to 0.781 (FD), 0.633 to 0.705 (ZL) indicating the reproducibility of our RNA-seq data (Fig. 1A).

Twelve treatment samples were selected for protein quantification, each labeled as follows: LS (pLS1, pLS2, pLS3), shell formation stage PK (pPK1, pPK2, pPK3), shell development stage FD (pFD1, pFD2, pFD3), and veliger stage ZL (pZL1, pZL2, pZL3). Mass spectrometry analysis was performed on the twelve protein samples,

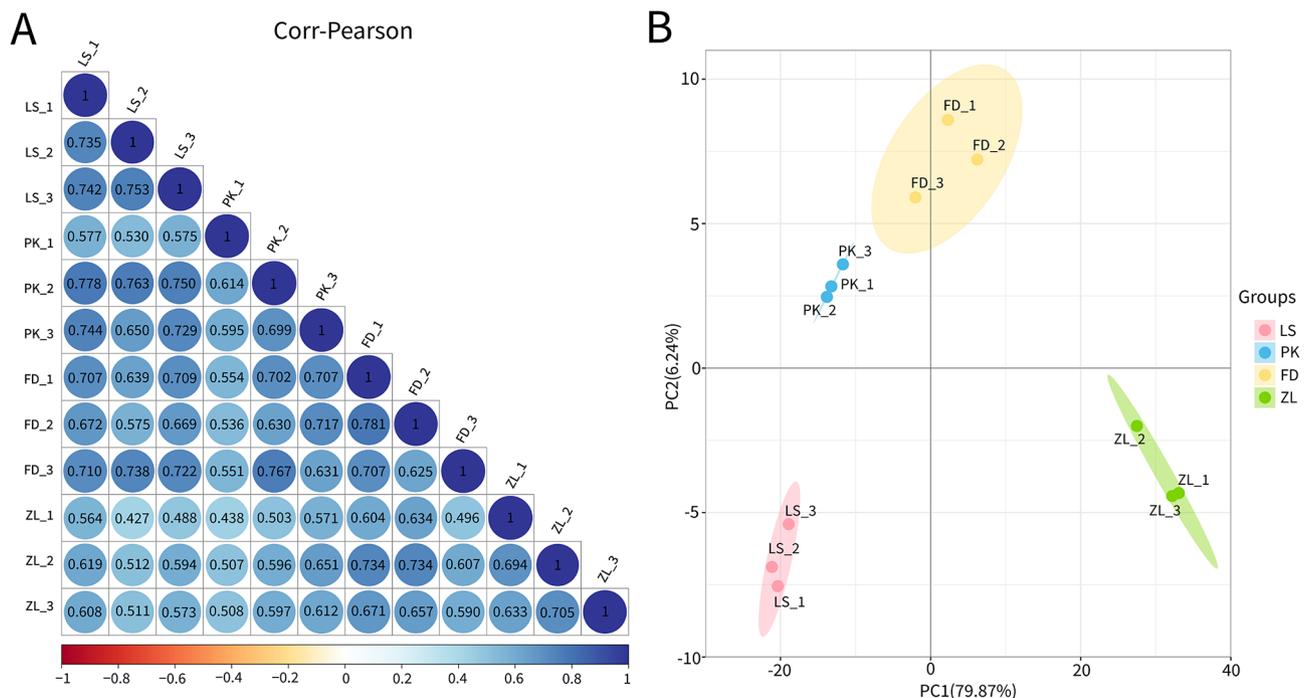
utilizing a foundation of 371,406 identified spectra. Among these, quantitative analysis was conducted on 4,011 proteins, revealing the presence of 48,502 distinct peptide segments.

Subsequent to protein sequencing, principal component analysis (PCA) was employed using the expression levels of credible proteins (Fig. 1B). This analysis illustrated the relationships among the samples from different dimensions. The repeatability within the four sets of sequenced samples was good, while notable differences were observed between samples. This data could be utilized for subsequent data analysis.

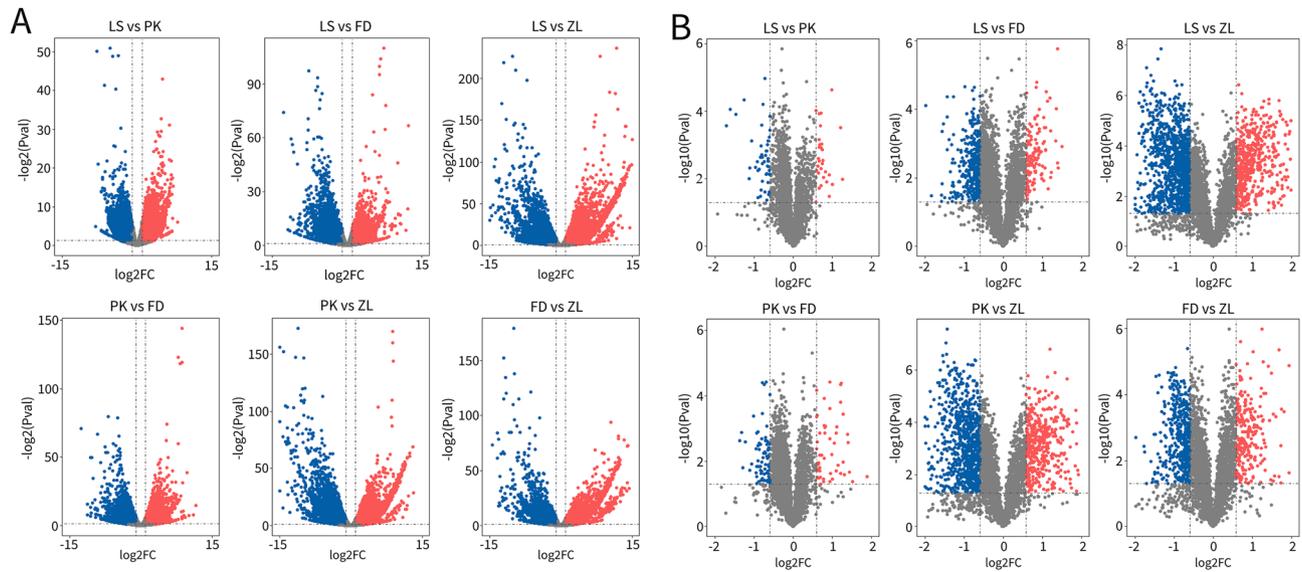
**Comparative transcriptomic and proteomic analysis of early embryonic developmental stages**

After pairwise comparison of the transcriptomic data from four developmental stages (LS, PK, FD, ZL), a comprehensive analysis revealed the following results: in the LS vs. PK comparison, 5549 DEGs (upregulated: 2841; downregulated: 2708) were identified; in the LS vs. FD comparison, 8824 DEGs (up: 3453; down: 5371) were identified; in the LS vs. ZL comparison, 21,944 DEGs (up: 10,562; down: 11,382) were identified; in the PK vs. FD comparison, 7470 DEGs (up: 3593; down: 3877) were identified; in the PK vs. ZL comparison, 14,063 DEGs (up: 5810; down: 8253) were identified; and in the FD vs. ZL comparison, 8929 DEGs (up: 4521; down: 4408) were identified (Fig. 2A).

Comparison of the proteomic data from the four developmental stages (LS, PK, FD, ZL) revealed that, as



**Fig. 1** Pearson correlation of RNA-seq (A) and PCA score of all quantitated proteins (B)



**Fig. 2** Analysis of differentially expressed genes (DEGs) and differentially expressed proteins (DEPs). **(A)** Volcano plot of DEGs. Red spots indicate upregulated DEGs; green spots indicate downregulated DEGs. **(B)** Volcano plot of DEPs. Red spots indicate upregulated DEPs; green spots indicate downregulated DEPs

embryonic development progressed, there was a consistent downregulation trend in the proteome. In the LS vs. PK comparison, 77 DEPs were identified (30 upregulated; 47 downregulated). In the LS vs. FD comparison, 482 DEPs were detected (127 up; 355 down), and in the LS vs. ZL comparison 1652 DEPs (536 up; 1116 down) were identified. In the PK vs. FD comparison, 132 DEPs were identified (53 up; 79 down). In the PK vs. ZL comparison, 1298 DEPs (446 up; 852 down) were found, and in the FD vs. ZL comparison, 690 DEPs (254 up; 436 down) were identified (Fig. 2B).

### GO enrichment analysis

GO enrichment analysis of all identified DEGs was performed to assess gene functions in early embryos across different developmental stages of *N. cumingii*. In the LS vs. PK comparison, DEGs were predominantly enriched in the nucleus (GO:0005654, GO:0005634) and were involved in processes related to metabolic breakdown (GO:0009056) (Fig. 3A). DEPs were primarily associated with L-phenylalanine metabolic processes (GO:0006559) and nucleic acid binding (GO:0003676) (Fig. 3B). Additionally, protein folding processes (GO:0006457) and functional enrichments related to DNA binding (GO:0003677) and RNA binding (GO:0003723) were identified at both transcriptomic and proteomic levels.

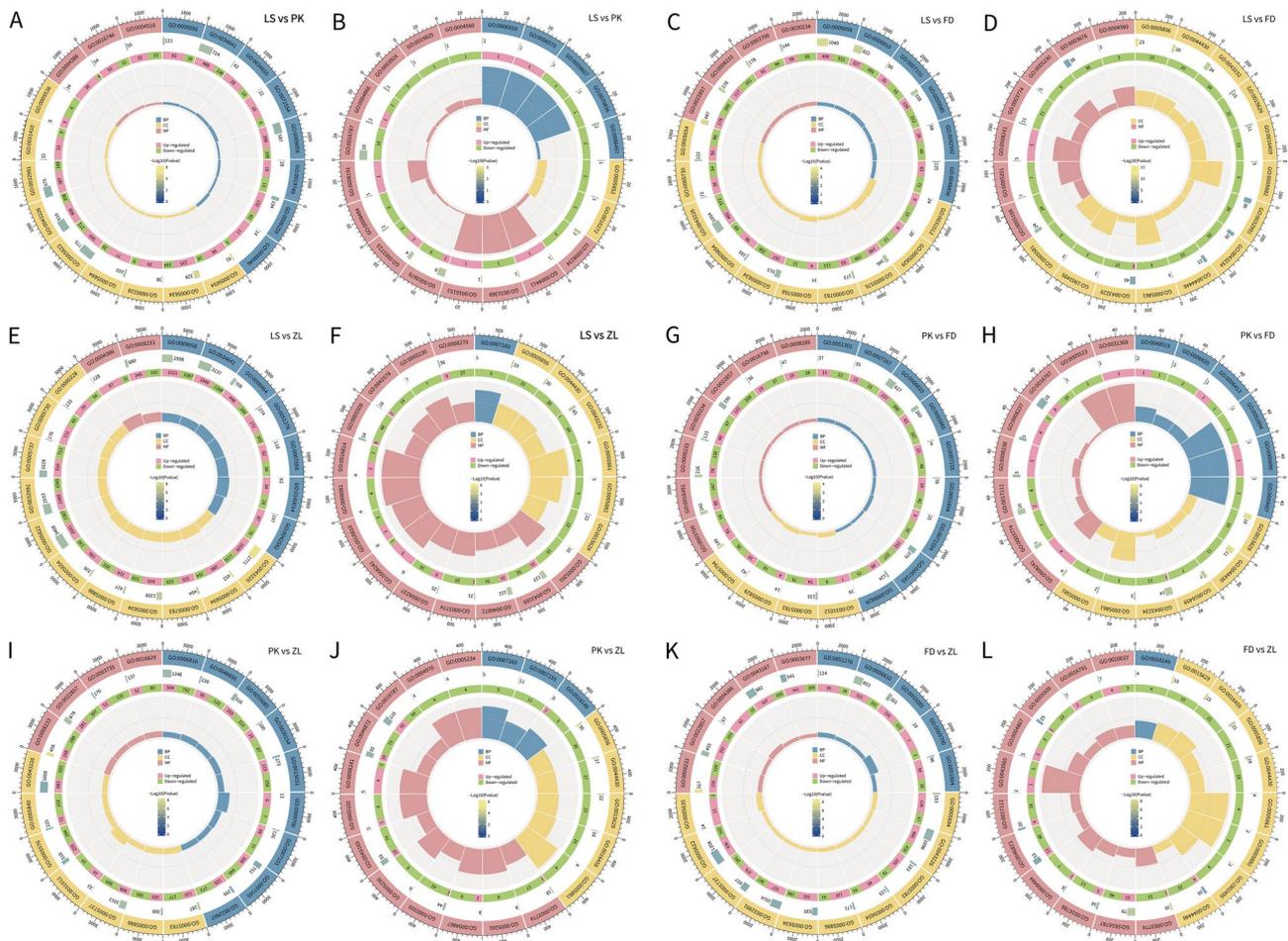
The DEGs in the LS vs. FD comparison were primarily enriched in the extracellular matrix (GO:0031012) and participated in biosynthetic processes (GO:0009058) as well as peptidase activity (GO:0008233) (Fig. 3C). The identified DEPs were mainly associated with extracellular ligand-gated ion channel activity (GO:0005230) and

nucleic acid binding (GO:0003676) (Fig. 3D). Processes related to lipid metabolism (GO:0006629) were enriched at both transcriptomic and proteomic levels. Additionally, functional enrichments such as hydrolase activity (GO:0016798) were detected at both levels.

In the LS vs. ZL comparison, DEGs were enriched in both the nucleus (GO:0043226, GO:0005694) and cytoplasm (GO:0005737, GO:0005783), participating in cellular protein modification processes (GO:0006464), chromosome segregation (GO:0007059), and activities related to helicase (GO:0004386) (Fig. 3E). DEPs were mainly involved in cell matrix adhesion processes (GO:0007160) and metal ion binding (GO:0046872) (Fig. 3F). Processes related to DNA metabolism (GO:0006259) and functional enrichments such as rRNA binding (GO:0019843) were detected at both transcriptomic and proteomic levels.

The DEGs in the PK vs. FD comparison were partially enriched in the endoplasmic reticulum (GO:0005783) and partially in the cytoplasm (GO:0005829), and they participate in the cell division process (GO:0051301) and activities related to DNA-binding transcription factors (GO:0003700) (Fig. 3G). At the protein level, the DEPs were primarily involved in negative regulation processes of biological activities (GO:0048519) and were associated with activities such as nucleoside-triphosphatase (GO:0017111) (Fig. 3H). Functional enrichments related to GTPase activity (GO:0003924) and kinase activity (GO:0016301) were observed at both transcriptomic and proteomic levels.

The differentially expressed genes in the PK vs. ZL comparison were predominantly enriched in the



**Fig. 3** GO analysis of differentially expressed genes (DEGs) and differentially expressed proteins (DEPs) in LS, PK, FD, and ZL (BP: biological process, CC: cellular component, MF: molecular function). The enriched GO terms of the DEGs (**A, C, E, G, I, K**). The enriched GO terms of the DEPs (**B, D, F, H, J, L**)

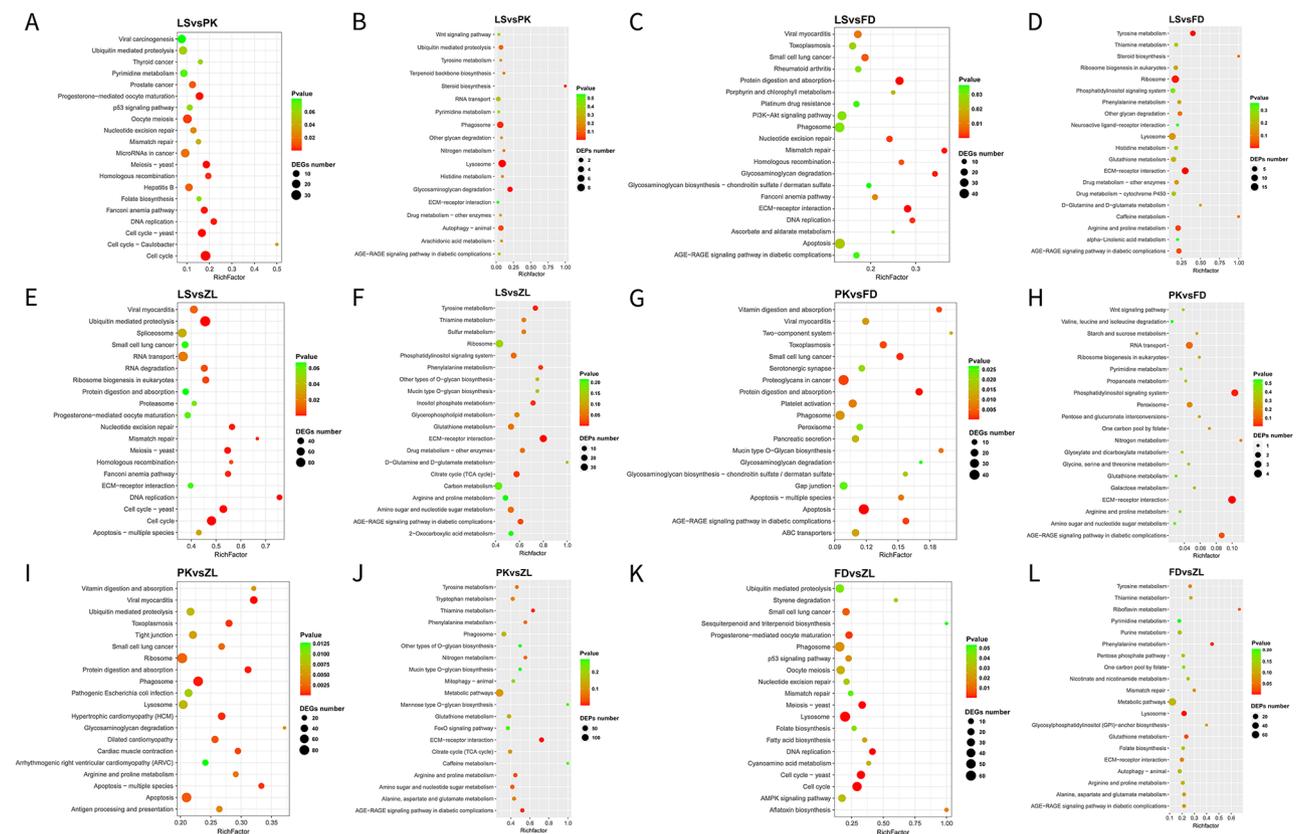
cytoplasm (GO:0005737, GO:0031012), participating in processes related to substance transport (GO:0006810, GO:0048856, GO:0055085), and activities associated with transmembrane transport proteins (GO:0022857) (Fig. 3I). At the protein level, the identified differentially expressed proteins were mainly involved in cell adhesion processes (GO:0007160, GO:0007155), and these proteins were associated with calcium ion binding (GO:0043169) (Fig. 3J). Cell adhesion processes (GO:0007155) and lipid metabolism processes (GO:0006629) were enriched at both transcriptomic and proteomic levels. Additionally, functional enrichments such as enzyme regulatory activity (GO:0030234) were detected at both transcriptomic and proteomic levels.

In the FD vs. ZL comparison, DEGs were found throughout the entire cell (GO:0043226) and were predominantly enriched in the chromosome (GO:0005694), participating in processes related to chromosome organization (GO:0051276) and activities associated with helicase (GO:0004386) (Fig. 3K). DEPs were primarily engaged in peptide cross-linking processes (GO:0018149)

(Fig. 3L). Processes related to DNA metabolism (GO:0006259) were enriched at both transcriptomic and proteomic levels.

#### KEGG enrichment analysis

KEGG analysis revealed that in the LS vs. PK comparison, 1464 genes were enriched in 277 pathways. Among these, the cell cycle pathway (ko04110) exhibited the highest enrichment, followed by the progesterone-mediated oocyte maturation pathway (ko04914) and DNA replication pathway (ko03030). However, these pathways did not exhibit enrichment at the protein level (Fig. 4A). At the protein level, 35 DEPs were identified and were enriched in 18 pathways. Among these, the lysosome pathway (map04142) displayed the highest enrichment, followed by the glycosaminoglycan degradation pathway (map00531) and steroid biosynthesis pathway (map00100) (Fig. 4B). The KEGG enrichment results at the protein level were highly consistent with those at the transcriptomic level. Among the 18 pathways, 15 corresponded with the transcriptomic data. Notably,



**Fig. 4** KEGG analysis of differentially expressed genes (DEGs) and differentially expressed proteins (DEPs) in LS, PK, FD, and ZL. The enriched KEGG terms of the DEGs (A, C, E, G, I, K). The enriched KEGG terms of the DEPs (B, D, F, H, J, L)

the ubiquitin-mediated proteolysis pathway (ko04120, map04120) exhibited substantial enrichment, with 18 DEGs and 2 DEPs.

In the LS vs. FD comparison, 2963 genes were found to be enriched in 316 pathways. Among these, the protein digestion and absorption pathway (ko04974) exhibited the highest level of enrichment, followed by the extracellular matrix (ECM)-receptor interaction pathway (ko04512) and the mismatch repair pathway (ko03430) (Fig. 4C). Additionally, 124 DEPs were found to be enriched across 36 pathways. The ECM-receptor interaction pathway (map04512) displayed the highest enrichment, which was consistent with the transcriptomic data (Fig. 4D). The KEGG enrichment results at the protein level demonstrated a high degree of consistency with the transcriptomic findings, with 32 of 36 pathways in concordance with the transcriptomic data. Notably, the ECM-receptor interaction pathway (ko04512, map04512) exhibited substantial enrichment, with 22 DEGs and 12 DEPs.

In the LS vs. ZL comparison, 7763 genes were enriched in 344 pathways. The DNA replication pathway (ko03030) showed the highest level of enrichment, followed by the ubiquitin-mediated proteolysis pathway (ko04120). However, these pathways did not exhibit enrichment at the

protein level (Fig. 4E). At the protein level, 822 DEPs were identified, and they were enriched across 64 pathways. The lysosome pathway (map04142) displayed the highest enrichment (Fig. 4F). The KEGG enrichment results at the protein level were highly consistent with the transcriptomic data. Of the 64 pathways, 56 corresponded with the transcriptomic data. Notably, the nucleotide excision repair pathway (ko03420, map03420) exhibited significant enrichment, with 35 DEGs and 6 DEPs.

In the PK vs. FD comparison, 2206 genes were enriched in 298 pathways. Among these, the apoptosis pathway (ko04210) exhibited a relatively high level of enrichment. However, these pathways did not show enrichment at the protein level (Fig. 4G). At the protein level, 35 DEPs were enriched across 24 pathways. The ECM-receptor interaction pathway (map04512) displayed the highest enrichment, which aligned with the transcriptomic data (Fig. 4H). Additionally, 23 of the 35 pathways corresponded with the transcriptomic data. Notably, the ECM-receptor interaction pathway exhibited the highest enrichment, with seven DEGs and four DEPs.

In the PK vs. ZL comparison, 4619 genes were enriched in 325 pathways, with pathways such as the apoptosis - multiple species pathway (ko04215) showing a relatively high level of enrichment. However, these pathways did

not exhibit enrichment at the protein level (Fig. 4I). At the protein level, 646 DEPs were identified, and they were enriched across 67 pathways. The arginine and proline metabolism pathway (map00330) were also enriched in the transcriptomic data, and it displayed a notably high level of enrichment (Fig. 4J). Twenty-three DEGs and 13 DEPs were enriched in this pathway. The KEGG enrichment results at the protein level demonstrated a high degree of consistency with the transcriptomic data, with 58 out of 67 pathways corresponding with the transcriptomic findings.

In the FD vs. ZL comparison, 3353 genes were enriched in 321 pathways, with pathways such as the cell cycle pathway (ko04110) exhibiting a high level of enrichment. Notably, the DNA replication pathway and lysosome pathway also showed enrichment at the protein level (Fig. 4K). A total of 281 DEPs were identified and found to be enriched across 59 pathways. Among these pathways, the lysosome pathway (map04142) exhibited substantial enrichment at both transcriptomic and proteomic levels (Fig. 4L). Fifty-one out of 59 pathways corresponded with the transcriptomic findings.

#### Selection and expression trends of key candidate genes in early embryonic development of *N. cumingii*

Pathways associated with embryonic development were identified through comparative analysis of KEGG

pathways across four developmental stages. Notably, the CDK gene family emerged as a consistent set of DEGs. The LS vs. PK, LS vs. FD, LS vs. ZL, PK vs. FD, PK vs. ZL, and FD vs. ZL comparisons revealed 12 DEGs within the CDK family: CDK1, CDK2, CDK5, CDK6, CDK8, CDK11, CDK13, CDK16, and CDK20, with two variants each for CDK1, CDK8, and CDK13 (Table 1; Fig. 5). When subjected to qRT-PCR analysis, the expression profiles of all 12 target genes displayed a robust correlation with those derived from RNA-seq analysis. This result illustrates the accuracy and reliability of the RNA-seq data (Fig. 5).

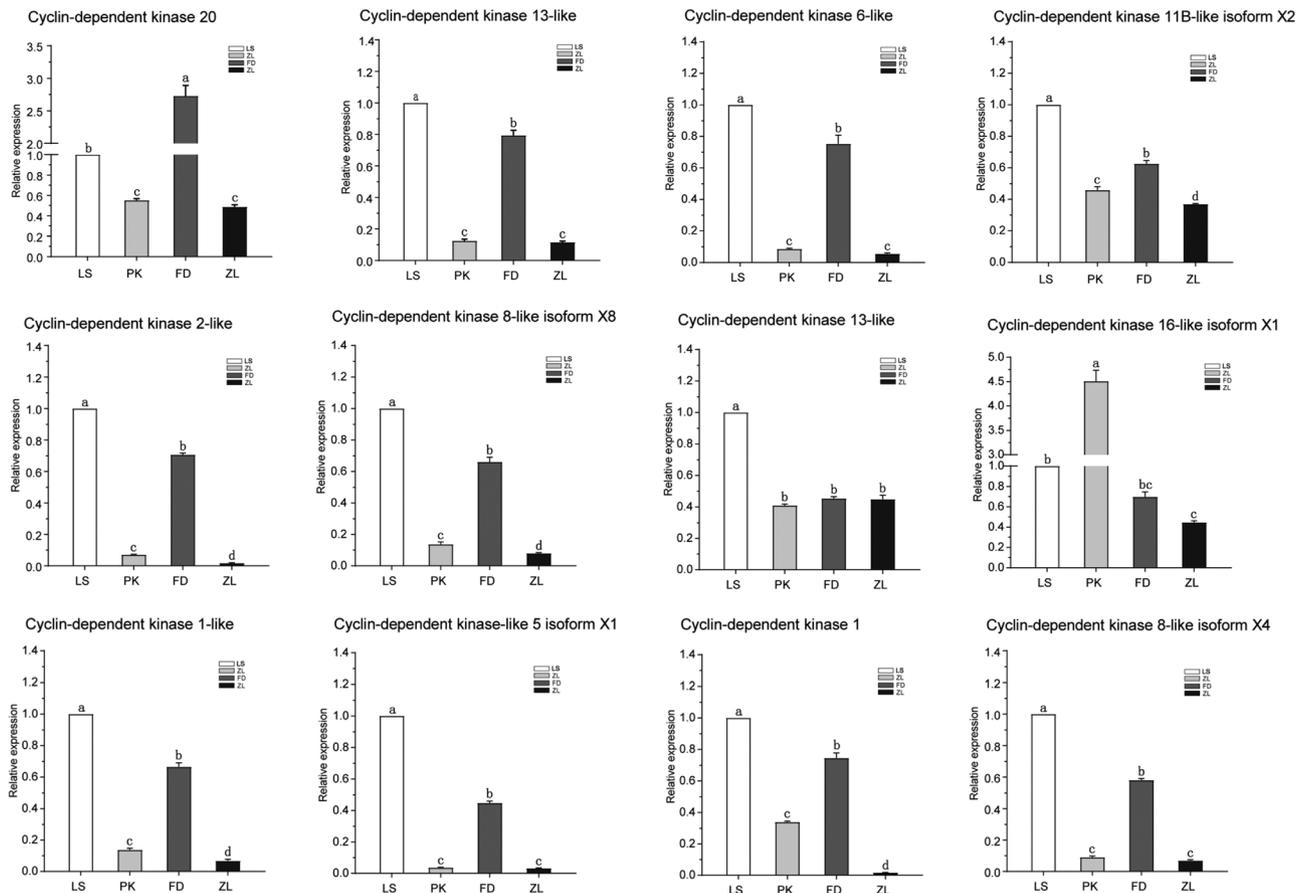
Using CDKs family genes selected from the transcriptome, corresponding analyses were conducted on proteomic data from four developmental stages (LS, PK, FD, ZL). Differentially expressed proteins closely associated with CDKs family-regulated signaling pathways were identified, including Mediator of RNA polymerase II transcription subunit 4-like, microtubule-associated protein futsch-like isoform X2, tumor protein p53-inducible protein 3, P25-alpha, and Serine/threonine-protein kinase PLK1-like.

#### Effect of flavopiridol treatment on early embryonic development of *N. cumingii*

To further validate the pivotal role of CDK family members in the early embryonic development of *N. cumingii*,

**Table 1** Primers used for mRNA-seq verification. Tm: primer-specific annealing temperature

Category	Description	Unigene ID	Primer sequences (5'→3')	Tm(°C)
Reference gene	18 S rRNA		F: TCTTGATTCCGGTGGGTGGTG R: CCCGGACATCTAAGGGCATC	60
mRNA	Cyclin-dependent kinase 20	Cluster-26337.54349	F: GACGGAGGCCCAAGGTCAAATCGTACA R: AGCAGCCACAGGTGCACTCCTTCAT	61
	Cyclin-dependent kinase 13-like	Cluster-26337.77830	F: CAACAACAGGGGTGAGATCAAGCTGGG R: GGTGCTGTCTTTGGGCTTGAAGGCA	61
	Cyclin-dependent kinase 6-like	Cluster-26337.23585	F: TACAAGGCTCGGGACCTGCGTAACCA R: GAGGCCAGGTCATGCTCGATGACTC	61
	Cyclin-dependent kinase 11B-like isoform X2	Cluster-26337.71332	F: CCGTATAAGAGAACGCCCCGTAGAG R: GTGTTCTTCTCATGGTGACAGGCG	61
	Cyclin-dependent kinase 2-like	Cluster-26337.55501	F: CCAGAACCTCTCATCGACTCCAAGAT R: GGGGTGCCCAAAGTTCTGAAGATTCGG	61
	Cyclin-dependent kinase 8-like isoform X8	Cluster-26337.26275	F: CGGGTTTCTGTCCCATACAGACCG R: GAGGCTTGAGTGGGAGTTGAAGAGAC	61
	Cyclin-dependent kinase 13-like	Cluster-26337.85551	F: CGGCCTGCACTACTGCCACCAGAAAA R: GCAAACCTCCTTGGCCCGAAGAT	61
	Cyclin-dependent kinase 16-like isoform X1	Cluster-26337.84189	F: ACATCATGGACCAGCAGTCCCCAAC R: TGTTGGGTGACGGCAGAGAAGAGGA	61
	Cyclin-dependent kinase 1-like	Cluster-26337.70194	F: GGCCGCCACAGGAGAACAATCAGCTT R: GTTCTGAGGCTTGAGATCCCGATGACAG	61
	Cyclin-dependent kinase-like 5 isoform X1	Cluster-26337.90933	F: GGACCAATGAGAGAGCTGACCATGC R: CCCAGTTGTGCCGTTGCTGATGTTTCG	61
	Cyclin-dependent kinase 1	Cluster-26337.69678	F: TTGACTCCTCCAGAAGACGACA R: TGGCTTGAGATCACGGTGC	61
	Cyclin-dependent kinase 8-like isoform X4	Cluster-26337.5567	F: TATCATCAAGTTCCACCGTTCA R: AGAGGCGAGCAAACCA	61



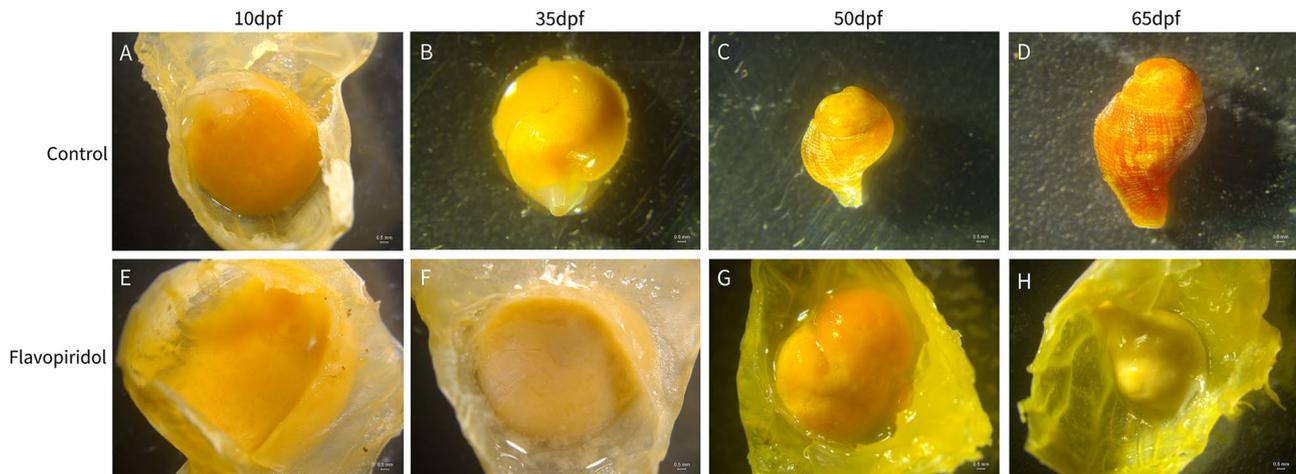
**Fig. 5** Analysis of gene expression of the CDK family genes involved in early embryonic development of *N. cumingii*. Different letters represent significant difference between different groups

we treated embryos with the CDK inhibitor flavopiridol. We found that the development of embryos not treated with flavopiridol progressed normally to the veliger stage. In contrast, embryos treated with flavopiridol exhibited slowed development arrest at the trochophore stage, and eventual natural death following developmental arrest (Fig. 6).

## Discussion

In this study, we applied comparative transcriptomic and proteomic analyses to identify the molecular mechanisms involved in *N. cumingii* embryonic development and explore potential molecular markers for selective breeding. We evaluated differential gene and protein expression during the crucial early developmental stages of *N. cumingii*, identified key CDK family genes associated with early embryonic development, and elucidated the related cellular signaling pathways. This research is the first to explore the molecular intricacies of early embryonic development in *N. cumingii* from a dynamic molecular expression perspective, and our results offer insights into promising molecular markers for breeding purposes.

According to the transcriptomic sequencing results, the DEGs identified in the LS vs. PK and FD vs. ZL comparisons exhibited an upregulation trend. In contrast, the DEGs in the other four comparisons showed a downregulation trend. However, in the comparative proteomic analysis, the relative expression of DEPs displayed a consistent downregulation trend. This result is inconsistent with the patterns of differential gene expression during the embryonic development process of *R. venosa*. This difference might be due to the different development pattern of the two species, as *R. venosa* undergoes indirect development whereas *N. cumingii* undergoes direct development. Thus, the relative expression levels of genes begin to significantly increase at the onset of metamorphosis in *R. venosa*, and the expression levels of DEGs continue to increase as development progresses [15]. Hence, we hypothesize that distinctive species-specific differences exist in gene expression patterns during the embryonic development of gastropods. In silk moths (*Bombyx mori*), more genes participate in late embryonic development compared to the early embryonic stage [16], which aligns with our study results and provides supporting evidence for the hypothesis that gene expression



**Fig. 6** Comparison of embryonic development between normal conditions and upon injection of CDK inhibitors. **(A)** The early embryonic development of *N. cumingii* 10 days post fertilization. (LS). **(B)** The early embryonic development of *N. cumingii* 35 days post fertilization. (PK). **(C)** The early embryonic development of *N. cumingii* 50 days post fertilization. (FD). **(D)** The early embryonic development of *N. cumingii* 65 days post fertilization. (ZL). **(E)** The early embryonic development of *N. cumingii* 10 days post fertilization and injection of CDK inhibitors. (LS). **(F)** The early embryonic development of *N. cumingii* 35 days post fertilization and injection of CDK inhibitors. (LS). **(G)** The early embryonic development of *N. cumingii* 50 days post fertilization and injection of CDK inhibitors. (LS). **(H)** The early embryonic development of *N. cumingii* 65 days post fertilization and injection of CDK inhibitors, and the embryo stop developing and begin to die in this stage. (LS)

gradually increases as embryonic development progresses [17].

KEGG analysis revealed pathways linked to *N. cumingii* embryonic development, including the cell cycle pathway, progesterone-mediated oocyte maturation pathway, DNA replication pathway, p53 signaling pathway, and AMP-activated protein kinase (AMPK) signaling pathway. Genes enriched in these pathways are primarily related to transcription factors, DNA replication, DNA polymerase, RNA polymerase, and related processes, which implies that DNA replication, gene expression, and the regulation of target genes by transcription factors play crucial roles in the embryonic development of *N. cumingii*. It is noteworthy that previous studies listed transforming growth factor beta as a factor influencing the growth and development of mollusks [18]. However, we did not detect differential expression of this factor in the early embryonic transcriptome of *N. cumingii*. Instead, we identified differential expression of multiple CDKs. Therefore, we hypothesize that in *N. cumingii* embryos, CDKs, rather than transforming growth factor beta, may play a crucial role as a key cluster driving embryonic development.

CDKs are threonine-directed serine/threonine-specific protein kinases, and their activity relies on the regulatory subunits cyclins. Based on the sequence of the kinase domains, CDKs belong to the CMGC class of kinases, along with mitogen-activated protein kinases, glycogen synthase kinase 3 $\beta$ , dual-specificity tyrosine-regulated kinases family members, and CDK-like kinases19/20. They drive the eukaryotic cell cycle by binding with cyclins. Control of the eukaryotic cell cycle is achieved

through the regulation of protein phosphorylation. CDKs serve as the initiation point for this phosphorylation regulation, and changes in activities of CDKs lead to extensive protein phosphorylation, which alters the activation state of substrates [21].

In this study, we identified nine differentially expressed CDKs: CDK1, CDK2, CDK5, CDK6, CDK8, CDK11, CDK13, CDK16, and CDK20. Among these, CDK1 and CDK2 have been shown to play essential roles in promoting DNA replication, regulating the cell cycle, and repairing DNA damage. These kinases are intricately involved in vital cellular processes crucial for cell survival. Specifically, in the normal cell cycle, cyclin E forms a complex with CDK2, facilitating the transition from the G1 to the S phase [22, 23]. In *N. cumingii* early embryos, CDK1 and CDK2 expression levels peaked during the LS stage, decrease during the PK stage, increased in the FD stage, and subsequently decreased again in the ZL stage. This pattern could be associated with the developmental traits of early *N. cumingii* embryos. Following the cleavage stage, *N. cumingii* embryos progressively transition into juveniles. During the LS stage, there is a rapid peak in DNA replication. Santamaría et al. [24] showed that mouse embryos lacking CDK1 are unable to progress beyond the morula and blastocyst stages. Hartley et al. [25] demonstrated that in clawed frog (*Xenopus*) embryonic development, Cyclin E/CDK2 is directly linked to an intrinsic maternal timer that drives the early embryonic cell cycle until the mid-blastula transition. These findings suggest that CDK1 and CDK2 are indispensable during early embryonic development and that these proteins

may exhibit similar expression patterns in both invertebrate and vertebrate embryo development.

CDK8 has been found to play a crucial role in regulating the transcription process. Galbraith et al. [26] conducted microarray experiments on cell lines with stable expression of short hairpin RNAs targeting CDK8 and found that CDK8 primarily serves as a positive regulator of transcription within this network. Westerling et al. [27] used mouse embryonic stem cell lines and sequencing based on fused cDNA and inserted a gene trap into the intron of CDK8. Genotyping of the hybrid offspring illustrated the indispensable role of CDK8 in normal mammalian development. In addition to CDK8, Aldridge [28] demonstrated that the absence of CDK11 in a mouse model resulted in early embryonic lethality, emphasizing the important role of CDK11 in normal development.

Chen et al. [29–31] provided evidence that the over-expression of CDK13 and an E1a mini-gene reporter construct in HEK293T cells led to dose-dependent alterations in the splicing pattern of E1a transcripts. They established that CDK13/CDC2L5 interacts with L-type cell cycle proteins and modulates alternative splicing. Alternative splicing serves as a fundamental mechanism that enhances the diversity of the cellular proteome, consequently influencing various cellular processes such as cell growth, differentiation, and apoptosis. In our research, we observed dynamic changes in the expression levels of CDK8, CDK11, and CDK13 during the embryonic development of *N. cumingii*. Specifically, their expression peaked during the LS stage, which is a critical period marked by cell proliferation and preparation for differentiation. Subsequently, as the embryos progressed to the PK stage, the expression levels of CDK8, CDK11, and CDK13 decreased. During the FD stage, when organ formation and significant cellular differentiation occurred, the expression levels of these genes increased. Upon the completion of organogenesis and the transition to the ZL stage, the expression levels of CDK8, CDK11, and CDK13 declined once more. Based on these observations, we hypothesize that CDK8, CDK11, and CDK13 play pivotal roles as developmental promoters in the early embryos of *N. cumingii*.

CDK5 plays a vital role in the development, functioning, and disorders of the central nervous system. It is involved in processes such as normal neuronal migration and axon extension. Previous studies have demonstrated a strong temporal correlation between CDK5 activation in the developing rat brain, p53 activation, and axon formation [32–36]. Zhang et al. [37, 38] reported that in primary neuronal cultures from the rat hippocampus, the absence of CDK5 activity inhibited dendritic growth induced by brain-derived neurotrophic factor. In the current study, we detected the presence of p53 in *N. cumingii* embryonic development, which suggested its potential

for activating CDK5. Throughout the embryonic development of *N. cumingii*, CDK5 showed the highest expression levels during the LS stage, followed by the FD stage. Notably, the formation of the head in *N. cumingii* occurs during the LS stage, suggesting a possible link between CDK5 and the development of the neuronal system in *N. cumingii*.

CDK6 plays an important role as a cell cycle regulatory factor. Apart from its conventional function in cell cycle regulation, Hu et al. [39] described the proliferative impact of CDK6 on adipocyte precursors in a CDK6 mouse model. Their findings also highlighted the indispensable role of CDK6 kinase activity in stem cell proliferation and survival. Biggs et al. [40, 41] reported that the regulatory role of CDKs in adipocyte precursors is contingent on the activity of one of its downstream factors, RUNX1. Previous studies demonstrated the critical role of RUNX1 in hematopoietic stem cell formation within the major vascular system of mouse embryos. Notably, RUNX1 is present in early *N. cumingii* embryos, implying a potential connection between CDK6 and the development of the circulatory system in *N. cumingii*. Furthermore, it is likely that CDK6 is involved in the differentiation of *N. cumingii* embryonic cells and the formation of adipose tissue.

CDK16 plays a pivotal role in a wide range of cellular processes that primarily govern the cell cycle. Ćwiek et al. [42, 43] described the role of CDK16 in regulating the G1 phase, and Yanagi et al. [44] elucidated the function of CDK16 in inhibiting the S, M, and G2/M phases. Of particular significance, CDK16 governs cell proliferation across various contexts, including cancer cells, neurite growth, neuronal cells, wound healing processes, and myoblast formation [45]. In *N. cumingii* embryonic development, CDK16 expression was highest during the PK stage, followed by the LS stage. During the PK stage, certain organs and muscles begin to form, and muscle development continues into the FD and ZL stages. As a result, CDK16 is also expressed in the FD and ZL stages, indicating its role in muscle cell formation during the embryonic development of *N. cumingii*.

CDK20 has recently been recognized as a key regulator of the cell cycle checkpoint. It controls cell growth and proliferation and is implicated in the development of numerous malignant tumors [46]. CDK20 activates CDK2 in human cells, and the decrease in CDK2 activity within the anti-mitotic signaling pathway results in cell cycle arrest [47]. Consequently, inhibiting CDK20 could potentially enhance the overall effect of CDK2 activation [48, 49]. Liu et al. [47] showed that CDK20 exhibits CDK-activating kinase activity, thereby influencing the proliferation of HeLa cells. Their study revealed that depletion of CDK20 induces G1 phase cell cycle arrest, reduces pCDK2 levels, inhibits CDK2 kinase activity, and

consequently hampers cell growth. In *N. cumingii* embryonic development, expression of CDK20 was highest during the FD stage, followed by the LS stage. During the FD stage, various organs in *N. cumingii* gradually mature and undergo rapid proliferation after cellular differentiation. Conversely, expression of CDK2 was highest during the LS stage. These results suggest that CDK20 primarily promotes cell proliferation during the FD stage and acts as a complementary factor by activating CDK2 during the LS stage.

In tumor cells, flavopiridol at concentrations between 200 and 300 nM can cause cell cycle arrest at both the G1 and G2 phases, thereby inhibiting the activities of CDK1, CDK2, CDK4, CDK6, and other proteins and causing cell death after cell cycle arrest [50]. To further validate the hypothesis that CDKs are crucial genes in the embryonic development of *N. cumingii*, we used flavopiridol as an inhibitor and observed its effects on embryonic development. As expected, *N. cumingii* embryos treated with flavopiridol displayed decelerated development, which stopped at the LS stage and eventually resulted in death. These observations are consistent with the research conducted by Zocchi et al. [51], who utilized flavopiridol to treat osteosarcoma cell lines both in vitro and in vivo.

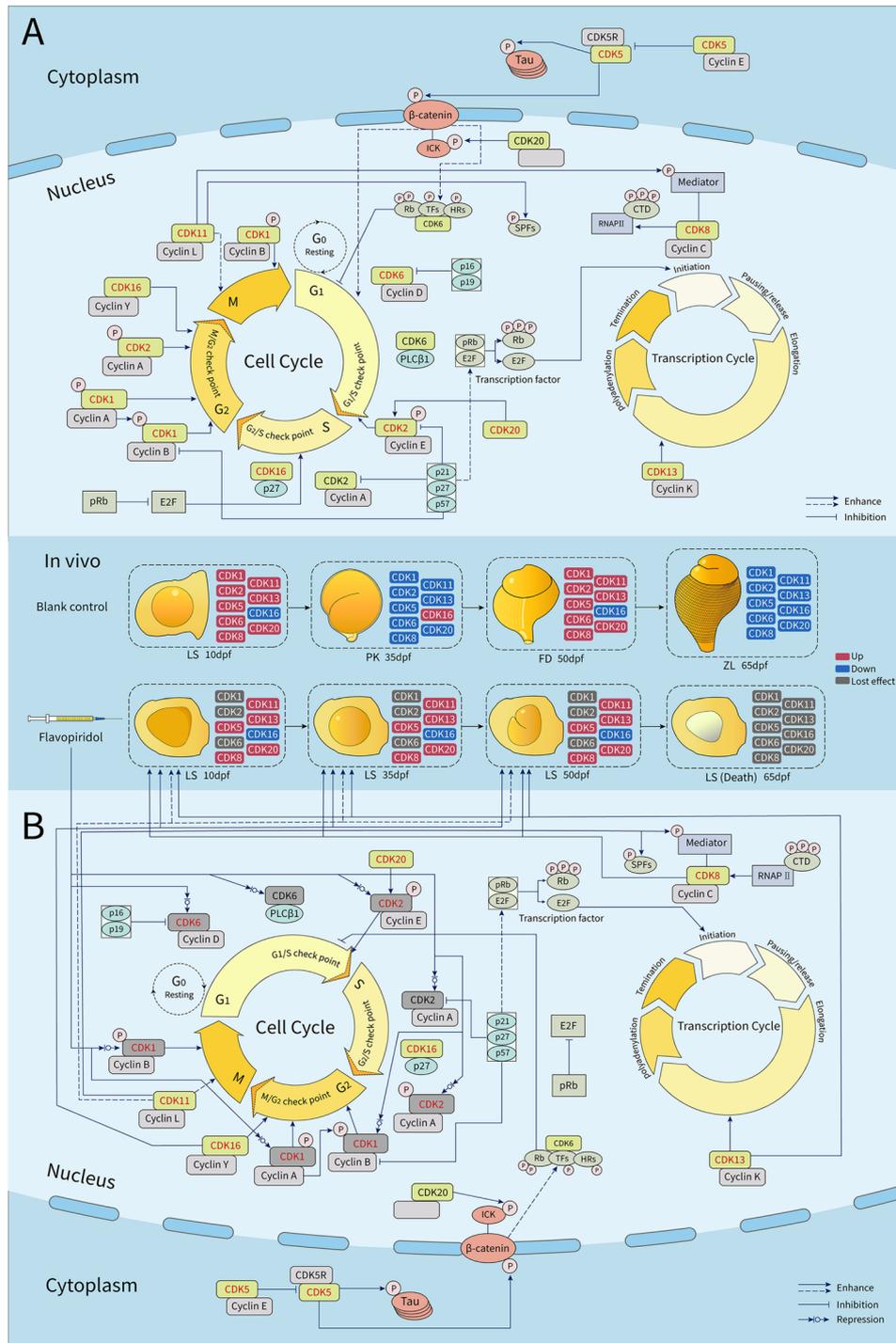
Integrating the data from our previous transcriptomic-proteomic analysis with the results of flavopiridol treatment led us to propose the following hypothesis (Fig. 7): the primary functions of CDK1 and CDK2, coupled with the supportive roles of CDK8 and CDK11, drive a peak in DNA replication during the LS stage of *N. cumingii* embryos. Simultaneously, CDK13 and CDK16 induce apoptosis, facilitating the gradual absorption of nutrients within the egg capsule by the viable eggs. Guided by the influence of CDK5 and CDK6, these viable eggs continue their development, forming the head and transitioning into the PK stage. Thus, the LS stage is a period marked by profound developmental changes in early *N. cumingii* embryos. Consequently, the expression levels of these eight CDKs are at their peak during the LS stage and rapidly decrease upon entering the PK stage. In the PK stage, *N. cumingii* embryos develop the shell layer, eyes, and other organs. CDK16 likely plays a role in promoting cell proliferation and muscle formation during this stage, leading to its highest expression levels during the development process. As the organism matures into the juvenile stage and progresses into the FD stage, the expression levels of CDK16 gradually decrease. During the FD stage, the shell of *N. cumingii* strengthens continuously, and the juvenile's color gradually changes from white to yellow-brown. All organs reach maturity during this period. The expression of CDK20 is highest in the FD stage, and expression levels of the other CDKs (except CDK16) are also upregulated. CDK20 likely activates CDK2 in the FD stage, allowing it to perform its inherent

function. CDK20 collaborates with other CDKs to facilitate the continued development of *N. cumingii* embryos as they transition into the ZL stage. During the ZL stage, *N. cumingii* hatches and gains the ability to live independently. At this point, CDKs have fulfilled their role in promoting embryonic development and are subsequently downregulated (Fig. 7).

After injecting CDK inhibitors into embryos during the LS stage, the activities of CDK1, CDK2, and CDK6 were suppressed, which prevented downstream genes of CDKs from exerting their regulatory effects. This blockade led to cell cycle arrest at the G1 and G2 phases, causing related cells to gradually cease development (Fig. 7). However, other CDKs were unaffected and continued to perform their respective functions. Therefore, *N. cumingii* embryos treated with CDK inhibitors exhibited slowed development during the LS stage. The inhibition of specific cells by CDK inhibitors hampered development and led to cell death. Consequently, other cells were unable to undergo typical differentiation and eventually died. As a result, early embryonic development in *N. cumingii* ceased during the LS stage. Therefore, we posit that CDKs constitute a pivotal gene cluster essential for promoting the early embryonic development of *N. cumingii*.

## Conclusions

In this study, we utilized transcriptome sequencing, proteomic analysis, and integrated transcriptome-proteome analysis techniques to obtain the temporal expression profiles of genes and proteins during the early development of *N. cumingii*. We elucidated the molecular expression patterns underlying the early development of *N. cumingii*, providing a comprehensive understanding of the molecular mechanisms governing embryonic development at the level of dynamic molecular expression. Furthermore, we identified and validated CDKs as a critical gene cluster regulating early embryonic development in *N. cumingii*. This study represents the first elucidation of the developmental regulatory functions of the CDK gene family in marine mollusks. Our findings provide valuable theoretical guidance for advancing artificial cultivation techniques for *N. cumingii*. Moreover, we identified essential molecular markers for facilitating molecular-assisted breeding of *N. cumingii*. This study enriches the molecular biology data pertaining to the early embryonic development not only within the Buccinidae family but also across marine mollusks. These results offer novel insights and perspectives for in-depth exploration of the molecular mechanisms governing organ differentiation and formation in gastropods.



**Fig. 7** The regulatory of cyclin-dependent kinase (CDK) family genes in early embryonic development of *N. cumingii*. **(A)** The regulatory of CDK family genes in early embryonic development of *N. cumingii* in normal conditions. **(B)** Changes in CDK family genes in early embryonic development of *N. cumingii* after injection of flavopiridol

**Materials and methods**

**Animals**

Two hundred *N. cumingii* specimens were collected in Zhangzidao, Dalian, Liaoning Province (39°1'42.2"N, 122°43'48.5"E). The shells of the captured *N. cumingii*

measured between 100 and 130 mm in height, and their body weight ranged from 100 to 128 g. All specimens were intact and undamaged. They were subsequently transferred to the Key Laboratory of Mariculture,

Ministry of Agriculture and Rural Affairs, Dalian Ocean University for temporary captivity.

### Culture conditions

The 200 parent snails were randomly divided into four groups, each consisting of 50 snails, and temporarily housed in four 300 L tanks. The water in the tanks was refreshed by half every day, and the tank bottoms were cleaned every other day. *N. cumingii* were provided with bay scallops *Argopecten irradians* as their feed. The shells of *Argopecten irradians* measure between 65 and 85 mm in height, while their body weight varies from 40 to 50 g. Spawning water temperature was kept at 11–18 °C, and dissolved oxygen levels, salinity, and pH were maintained at 9.5–10.7 mg/L, 31–35‰, and 8.1, respectively. After the parent snails laid eggs, the egg capsules were carefully removed from the tank walls and transferred to the same hatching tanks. Natural hatching conditions were maintained with water temperature 16–22 °C, dissolved oxygen 9.2–10.4 mg/L, salinity 31–35‰, and pH 8.0–8.1. The half of the water was changed daily and no cleaning of the tank bottoms. Periodic samples were collected from the egg capsules laid on the same day to observe the embryonic development process of *N. cumingii* under the stereomicroscope of Leica (LEICA M205 FA).

### Evaluation of CDK functions in embryonic development

To investigate the role of CDKs in the embryonic development of *N. cumingii*, snail embryos were treated with a CDK inhibitor. Flavopiridol, a novel flavone derivative, competes directly with ATP substrates and inhibits various CDKs, including CDK1, CDK2, CDK4, CDK7, and CDK9. It acts as a broad-spectrum CDK inhibitor and has been demonstrated to directly suppress kinase activity in immunoprecipitated CDKs extracted from exponentially growing cells, displaying an IC<sub>50</sub> value ranging from 100 to 400 nM [52]. Therefore, flavopiridol was selected for this study as the CDK inhibitor for treating early embryos of *N. cumingii*.

Sixty embryos at the identical developmental stage (Egg-swallowing stage) were chosen for both the injection group and the blank control group, with 30 embryos allocated to each of the injection group and the blank control group. Considering the thickness of *N. cumingii* egg capsules, flavopiridol was directly injected into the capsules at a concentration of 5 mg/kg. After injection, the eggs from both the injection group and the blank control group were placed separately into two new tanks. The embryos were incubated under the condition like the natural hatching conditions.

### RNA-seq sample preparation and library construction

Embryos were carefully selected at four distinct developmental stages: the Egg-swallowing stage (referred to

as LS), Protoconch forming stage (referred to as PK), Shell development stage (referred to as FD), and Juvenile stage (referred to as ZL). At each stage, ten embryos were meticulously collected, and the experiment was organized into six replicate groups. For every developmental stage, three sets of samples were meticulously prepared for both RNA-seq and proteomic analyses. Specifically, three sets were utilized for constructing mRNA libraries (LS1, LS2, LS3; PK1, PK2, PK3; FD1, FD2, FD3; ZL1, ZL2, ZL3). Additionally, another three sets of samples were designated for protein TMT labeling detection (pLS1, pLS2, pLS3; pPK1, pPK2, pPK3; pFD1, pFD2, pFD3; pZL1, pZL2, pZL3). All samples were meticulously stored at -80 °C to maintain their integrity and viability for subsequent analyses.

### RNA sequencing and protein TMT labeling of peptides

The preparation and sequencing of the transcriptome library were conducted by NovoGene Bioinformatics Technology Co., Ltd. Total RNA was extracted using the TRIzol method (Ambion, USA) as per the instructions provided by the commissioning party. The RNA concentration of the extracted samples was assessed using Agilent 5400 (Agilent Technologies, California, USA) and Nanodrop. Furthermore, the integrity and purity of the RNA were evaluated using a 1% agarose gel electrophoresis method and Agilent 5400, respectively, in accordance with the guidelines provided by the commissioning party. The 12 individual samples from the same treatment group were pooled for library preparation and sequencing. Initially, mRNA with PolyA tails was enriched using Oligo(dT) magnetic beads. The enriched mRNA was fragmented using Fragmentation Buffer, and the resulting mRNA fragments were utilized as templates for the synthesis of the first cDNA strand using random oligonucleotide primers. Subsequently, the RNA strand was degraded using RNaseH, and the second cDNA strand was synthesized in the presence of dNTPs using DNA polymerase I. The purified double-stranded cDNA underwent end repair, A-tailing, and sequencing adapter ligation. AMPure XP beads were employed to select cDNA fragments of approximately 370–420 bp in size for PCR amplification and purification, thereby establishing the library. Upon completion of library construction, an initial quantification was performed using a Qubit 2.0 Fluorometer. The library was diluted to a concentration of 1.5 ng/μL and assessed for insert size using an Agilent 2100 bioanalyzer. Following successful validation of the insert size, qRT-PCR was conducted to accurately quantify the effective library concentration, ensuring library quality.

The samples were cryogenically ground into powder, swiftly transferred to pre-chilled centrifuge tubes in liquid nitrogen, and an appropriate volume of PASP

protein lysis buffer was added. The mixture was vigorously shaken and thoroughly lysed through 5 min of ultrasonication in an ice-water bath. After centrifugation at 4 °C and 12,000g for 15 min, the supernatant was collected and reacted with 10mM DTT at 56 °C for 1 h. Following this, an ample amount of IAM was added, and the reaction was shielded from light and allowed to proceed for 1 h at room temperature. Subsequently, the mixture was precipitated at -20 °C for a minimum of 2 h using 4-fold volume of pre-chilled acetone at -20 °C. The precipitate was then collected by centrifugation at 4 °C and 12,000g for 15 min. This process was repeated thrice, and the precipitate was air-dried. The protein precipitate was dissolved in protein solubilization solution [53–55].

For the ensuing steps, the protein sample was treated with DB protein solubilization solution, trypsin, and TEAB buffer. The mixture was incubated at 37 °C for 4 h for enzymatic digestion, followed by the addition of trypsin and CaCl<sub>2</sub> for overnight digestion. The pH was lowered to less than 3 using formic acid, and after thorough mixing, centrifugation was performed at room temperature and 12,000g for 5 min. The supernatant was then slowly passed through a C18 desalting column. Subsequent washes were performed using washing solution (0.1% formic acid, 3% acetonitrile) thrice, followed by elution with an appropriate volume of elution solution (0.1% formic acid, 70% acetonitrile). The filtrate was collected and freeze-dried.

The resultant material was re-dissolved in 0.1 M TEAB buffer and treated with TMT labeling reagent dissolved in acetonitrile. The mixture was gently agitated at room temperature for 2 h, followed by termination of the reaction with 8% ammonia solution. Equal volumes of labeled samples were combined, desalted, and subsequently freeze-dried [56].

#### Data processing and differential expression analysis

All RNA clean data were submitted to the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) Sequence Database (Accession No. PRJNA1025855). Q20 (percent bases with a “Phred value>20”), Q30 (percent bases with a “Phred value>30”), and GC-content were analyzed. The mass spectrometry proteomics data of *N. cumingii*. have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the iProX partner repository with the dataset identifier PXD046073.

After obtaining clean reads, the Trinity software (v2.4.0) was employed to perform transcript assembly. Clean reads from each sample were mapped to a reference genome (Ref), and subsequently, reads with alignment quality scores below 10, unpaired alignments, and reads aligning to multiple genomic regions were filtered

out. The RSEM software (v1.2.15) [57] was then utilized to calculate gene expression levels.

To assess the integrity of the assembled transcripts, the BUSCO (Benchmarking Universal Single-Copy Orthologs) software was employed [58]. This evaluation included a comprehensive assessment of transcript completeness using tools such as tblastn, augustus, and hmmer.

Gene functions were annotated based on seven databases: Nr, Nt, Pfam, COG, Swiss-Prot, KEGG, and GO. The annotation process utilized specific tools for each database: Diamond (v0.8.22) [59] for Nr, COG, and Swiss-Prot; NCBI blast2.2.28+ (v2.2.28+) [60] for Nt; Hmmscan (HMMER3) [61] for Pfam; Blast2GO (b2g4pipe\_v2.5) [62] for GO; and KAAS (r140224) [63] for KEGG. Transcription factors were identified using the animalTFDB 2.0 software. Differential expression analysis of genes in various comparison groups was performed using DESeq2 (1.6.3) [64]. The threshold for significant differential expression was set at  $P_{adj} < 0.05$  and  $|\log_2(\text{foldchange})| > 1$ . Additionally, GO functional enrichment analysis and KEGG pathway enrichment analysis of the differentially expressed genes were carried out using Goseq (1.10.0) [65] and KOBAS (v2.0.12) [66] software, respectively.

The Bradford protein quantification assay was conducted following these steps: BSA standard protein solution was prepared according to the manual, with a concentration gradient ranging from 0 to 0.5 µg/µL. Different concentrations of BSA standard protein solutions and varying dilutions of the test sample solutions were added to a 96-well plate, with a final volume adjusted to 20 µL. Each gradient was repeated three times. Next, 180 µL of G250 staining solution was rapidly added, and the plate was left at room temperature for 5 min, followed by measuring the absorbance at 595 nm.

A standard curve was generated using the absorbance values of the standard protein solution, and this curve was used to calculate the protein concentration of the test samples. Furthermore, 20 µg of protein from each test sample was subjected to 12% SDS-PAGE gel electrophoresis. Concentration gel electrophoresis was performed at 80 V for 20 min, followed by separation gel electrophoresis at 120 V for 90 min. After electrophoresis, the gel was stained with Coomassie Brilliant Blue R-250 until the protein bands became clearly visible, and then the gel was destained to achieve clear bands.

Using the Proteome Discoverer 2.2 software (PD2.2, Thermo), spectra from each run were searched against the protein database. Search parameters included a precursor ion mass tolerance of 10 ppm and a fragment ion mass tolerance of 0.02 Da. Fixed modification involved cysteine alkylation, while variable modifications encompassed methionine oxidation and TMT labeling.

N-terminal modifications included acetylation and TMT labeling, with allowance for up to 2 missed cleavage sites.

For result refinement, PD2.2 software employed additional filters: Peptide-spectrum matches (PSMs) with a confidence level exceeding 99% were deemed reliable, and proteins with at least one unique peptide segment were considered credible. Only reliable PSMs and proteins were retained, and a false discovery rate (FDR) validation was conducted to exclude peptides and proteins exceeding an FDR of 1%. Statistical analysis of protein quantification employed a T-test. Proteins showing significant expression differences between experimental and control groups were categorized as follows: For upregulated proteins, those with a P-value  $\leq 0.05$  and a fold change (FC)  $\geq 1.5$  were selected; for downregulated proteins, those with a P-value  $\leq 0.05$  and a FC  $\leq 0.67$  were chosen.

The interproscan software was employed for GO and IPR functional annotation, utilizing databases such as Pfam, PRINTS, ProDom, SMART, ProSite, and PANTHER. Additionally, COG and KEGG analyses were conducted to determine the functional protein families and pathways for the identified proteins [67]. Specifically focusing on DPE, a volcano plot analysis was performed, along with cluster heatmap analysis. Furthermore, enrichment analyses were carried out for GO, IPR, and KEGG pathways [68].

#### Quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR) validation

The annotation of Differentially Expressed Genes (DEGs) was validated using quantitative Real-Time Polymerase Chain Reaction (qRT-PCR), where 12 selected DEGs were examined. The first-strand cDNA of the DEGs was synthesized using the Fastking gDNA Dispelling RT SuperMix kit from TIANGEN (China). Primers were designed using the online primer design tool available at <http://biotools.nubic.northwestern.edu/OligoCalc.html> (Table 1), with 18S rRNA as the reference gene. For each sample, three technical replicates were set up, and qRT-PCR was conducted on the Light Cycler 96 system.

The total reaction volume was 20  $\mu$ L, consisting of 2  $\mu$ L cDNA, 10  $\mu$ L 2 $\times$ FastStart Essential DNA Green Master (Germany), 6.4  $\mu$ L DEPC-treated water, and 0.8  $\mu$ L each of upstream and downstream primers (10 mM). The cycling program was as follows: an initial denaturation at 95  $^{\circ}$ C for 600 s, followed by 45 cycles of denaturation at 95  $^{\circ}$ C for 10 s and annealing/extension at 60  $^{\circ}$ C for 60 s. After amplification, a PCR melt curve analysis was performed to confirm the presence of a single PCR product. The relative expression levels of each candidate DEG were determined using the  $2^{-\Delta\Delta C_t}$  method [69].

#### Acknowledgements

Thanks for the help of all members in Key Laboratory of Mariculture and Stock Enhancement in North China's Sea. In addition, the authors would like to thank the International Science Editing Company for helping to improve the language of this article.

#### Author contributions

Authors' contributions: Fengxiao Lv: Data curation, Writing-original draft, Writing-review & editing, Data visualization. Xinfan Ge: Performed the experiments, Data curation, Designed the experiments Yaqing Chang: Writing-review & editing, Resources, Supervision, Funding acquisition. Zhenlin Hao: Writing-review & editing, Polish the original draft, Resources, Supervision, Funding acquisition. All authors have read and approved the final manuscript.

#### Funding

This study was supported by funds from National Natural Science Foundation of China (42076101), General Project of Education Department of Liaoning Province (LJKZZZ0700), Project of Liaoning Provincial Natural Resources Department, Marine Economy Development Special Project of Liaoning Province Department of Natural Resources, the Central Government Subsidy Project for Liaoning Fisheries (2023).

#### Data availability

The RNA clean datasets generated during the current study are available in the National Center for Biotechnology Information (NCBI) Shrot Read Archive (SRA) repository, persistent accession number to datasets PRJNA1025855. The mass spectrometry proteomics datasets generated during the current study are available in the ProteomeXchange repository (<http://proteomecentral.proteomexchange.org>), persistent accession number to datasets PXD046073.

#### Declarations

##### Ethics approval and consent to participate

The animal study was approved by the Ethics Committee of Dalian Ocean University. The study was conducted in accordance with the local legislation and institutional requirements.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

Received: 14 March 2024 / Accepted: 28 October 2024

Published online: 19 December 2024

#### References

1. Statistics Bureau of Dalian City of People's Republic of China, Dalian Statistical Yearbook (2022). China Statistical Publishing House.
2. Switzer-Dunlap M, Hadfield MG. Observations on development, larval growth and metamorphosis of four species of Aplysiidae (Gastropoda: Opisthobranchia) in laboratory culture. *J Exp Mar Biol Ecol.* 1977;29(3):245–61.
3. Hao Z, Liu H, Yu Y, et al. Reproductive characteristics and variations in the biochemical composition of *Neptunea Arthritica Cumingii* *crosse* through embryonic development. *Aquac Res.* 2021;52(1):1–11.
4. Joyce AR, Palsson BØ. The model organism as a system: integrating omics' data sets. *Nat Rev Mol Cell Biol.* 2006;7(3):198–210.
5. Li B, Predel R, Neupert S, et al. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle *Tribolium castaneum*. *Genome Res.* 2008;18(1):113–22.
6. Tian Y, Zhang X, Zhou J. Advances and applications in Transcriptomics Research. *Middle School Biology Teach.* 2013;1(12):29–31.
7. Qian X, He F. *Proteomics: theory and methods*. Beijing: Science; 2003. (In Chinese).
8. Heyland A, Vuc Z, Voolstra CR, et al. Developmental transcriptome of *Aplysia californica*. *J Experimental Zool Part B: Mol Dev Evol.* 2011;316(2):113–34.
9. Lambert JD, Chan XY, Spiecker B, et al. Characterizing the embryonic transcriptome of the snail *Ilyanassa*. *Integr Comp Biol.* 2010;50(5):768–77.

10. Franchini P, Van der Merwe M, Roodt-Wilding R. Transcriptome characterization of the South African abalone *Haliotis midae* using sequencing-by-synthesis. *BMC Res Notes*. 2011;4:1–11.
11. Song H, Yu ZL, Sun LN, et al. De novo transcriptome sequencing and analysis of *Rapana Venosa* from six different developmental stages using Hi-seq 2500. *Comp Biochem Physiol D: Genomics Proteomics*. 2016;17:48–57.
12. Sun J, Zhang Y, Thiyagarajan V, et al. Protein expression during the embryonic development of a gastropod. *Proteomics*. 2010;10(14):2701–11.
13. Mann K, Edsinger E. The *Lotia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. *Proteome Sci*. 2014;12:1–12.
14. Martinez-Fernandez M, Rodriguez-Pineiro AM, Oliveira E, et al. Proteomic comparison between two marine snail ecotypes reveals details about the biochemistry of adaptation. *J Proteome Res*. 2008;7(11):4926–34.
15. Song H, Yu ZL, Sun LN et al. (2016) Transcriptomic analysis of differentially expressed genes during larval development of *Rapana venosa* by digital gene expression profiling. *G3: Genes, Genomes, Genetics*. 6(7): 2181–2193.
16. Xu GF, Gong C, Lyu H, et al. Dynamic transcriptome analysis of *Bombyx mori* embryonic development. *Insect Sci*. 2022;29(2):344–62.
17. Fan XB, Pang R, Li WX, et al. An overview of embryogenesis: external morphology and transcriptome profiling in the Hemipteran Insect *Nilaparvata lugens*. *Front Physiol*. 2020;11:106.
18. Zheng Z, Hao R, Xiong X, et al. Developmental characteristics of pearl oyster *Pinctada fucata martensii*: insight into key molecular events related to shell formation, settlement and metamorphosis. *BMC Genomics*. 2019;20:1–11.
19. Wood DJ, Endicott JA. Structural insights into the functional diversity of the CDK–cyclin family. *Open Biol*. 2018;8(9):180112.
20. Malumbres M. Cyclin-dependent kinases. *Genome Biol*. 2014;15(6):1–10.
21. MacKenzie AM, Laceyfield S. CDK Regulation of Meiosis: lessons from *S. Cerevisiae* and *S. Pombe*. *Genes*. 2020;11(7):723.
22. Liao H, Ji F, Ying S. CDK1: beyond cell cycle regulation. *Aging*. 2017;9(12):2465.
23. Fagundes R, Teixeira LK. Cyclin E/CDK2: DNA replication, replication stress and genomic instability. *Front Cell Dev Biol*. 2021;9:774845.
24. Santamaría D, Barrière C, Cerqueira A, et al. Cdk1 is sufficient to drive the mammalian cell cycle. *Nature*. 2007;448(7155):811–5.
25. Hartley RS, Sible JC, Lewellyn AL, et al. A role for cyclin E/Cdk2 in the timing of the Midblastula Transition in *Xenopus* embryos. *Dev Biol*. 1997;188(2):312–21.
26. Galbraith MD, Donner AJ, Espinosa JM. CDK8: a positive regulator of transcription. *Transcription*. 2010;1(1):4–12.
27. Westerling T, Kuuluvainen E, Maäkelaä TP. Cdk8 is essential for preimplantation mouse development. *Mol Cell Biol*. 2007;27(17):6177–82.
28. Aldridge, R.C. Investigating the Influence of CDK11 in developmental and cancer phenotypes. *Edinburgh Research Archive*. 2018-11-30.
29. Chen H, Wong YH, Geneviere AM, et al. CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. *Biochem Biophys Res Commun*. 2007;354(3):735–40.
30. Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*. 2002;418(6894):236–43.
31. Graveley BR. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet*. 2001;17(2):100–7.
32. Cheung ZH, Ip NY. Cdk5: a multifaceted kinase in neurodegenerative diseases. *Trends Cell Biol*. 2012;22(3):169–75.
33. Shukla V, Skuntz S, Pant HC. Deregulated Cdk5 activity is involved in inducing Alzheimer's disease. *Arch Med Res*. 2012;43(8):655–62.
34. Smith DS, Tsai LH. Cdk5 behind the wheel: a role in trafficking and transport? *Trends Cell Biol*. 2002;12(1):28–36.
35. Paglini G, Pigino G, Kunda P, et al. Evidence for the participation of the neuron-specific CDK5 activator P35 during laminin-enhanced axonal growth. *J Neurosci*. 1998;18(23):9858–69.
36. Dhariwala FA, Rajadhyaksha MS. An unusual member of the cdk family: Cdk5. *Cell Mol Neurobiol*. 2008;28:351–69.
37. Zhang R, Liu C, Ji Y, et al. Neuregulin-1 $\beta$  plays a neuroprotective role by inhibiting the Cdk5 signaling pathway after cerebral ischemia-reperfusion injury in rats. *J Mol Neurosci*. 2018;66:261–72.
38. Cheung ZH, Chin WH, Chen Y, et al. Cdk5 is involved in BDNF-stimulated dendritic growth in hippocampal neurons. *PLoS Biol*. 2007;5(4):e63.
39. Hu AJ, Li W, Pathak A et al. (2023). CDK6 is essential for mesenchymal stem cell proliferation and adipocyte differentiation. *Front Mol Biosci*. 10.
40. Biggs JR, Peterson LF, Zhang Y, et al. AML1/RUNX1 phosphorylation by cyclin-dependent kinases regulates the degradation of AML1/RUNX1 by the anaphase-promoting complex. *Mol Cell Biol*. 2006;26(20):7420–9.
41. Nottingham WT, Jarratt A, Burgess M, et al. Runx1-mediated hematopoietic stem-cell emergence is controlled by a gata Ets/SCL-regulated enhancer. *Blood J Am Soc Hematol*. 2007;110(13):4188–97.
42. Cwiek P, Leni Z, Salm F, et al. RNA interference screening identifies a novel role for PCTK1/CDK16 in medulloblastoma with c-Myc amplification. *Oncotarget*. 2015;6(1):116.
43. Yanagi T, Krajewska M, Matsuzawa S, et al. PCTAIRE1 phosphorylates p27 and regulates mitosis in cancer cells. *Cancer Res*. 2014;74(20):5795–807.
44. Yanagi T, Hata H, Mizuno E, et al. PCTAIRE1/CDK16/PCTK1 is overexpressed in cutaneous squamous cell carcinoma and regulates p27 stability and cell cycle. *J Dermatol Sci*. 2017;86(2):149–57.
45. Wang X, Liu R, Li S, et al. The roles, molecular interactions, and therapeutic value of CDK16 in human cancers. *Biomed Pharmacother*. 2023;164:114929.
46. Chivukula S, Malkhed V. The role of CDK20 protein in carcinogenesis. *Curr Drug Targets*. 2023;24(10):790–6.
47. Liu Y, Wu C. Galaktionov K. p42, a novel cyclin-dependent kinase-activating kinase in mammalian cells. *J Biol Chem*. 2004;279(6):4507–14.
48. Caligiuri M, Becker F, Murthi K, et al. A proteome-wide CDK/CRK-specific kinase inhibitor promotes tumor cell death in the absence of cell cycle progression. *Chem Biol*. 2005;12(10):1103–15.
49. Kaldis P, Solomon MJ. Analysis of CAK activities from human cells. *Eur J Biochem*. 2000;267(13):4213–21.
50. Sedlacek H. Mechanisms of action of flavopiridol. *Crit Rev Oncol/Hematol*. 2001;38(2):139–70.
51. Zocchi L, Wu SC, Wu J, et al. The cyclin-dependent kinase inhibitor flavopiridol (alvocidib) inhibits metastasis of human osteosarcoma cells. *Oncotarget*. 2018;9(34):23505.
52. Shapiro GI. Preclinical and clinical development of the cyclin-dependent kinase inhibitor flavopiridol. *Clin Cancer Res*. 2004;10(12):s4270–5.
53. Wiśniewski JR, Zougman A, Nagaraj N, et al. Universal sample preparation method for proteome analysis. *Nat Methods*. 2009;6(5):359–62.
54. Gillette MA, Satpathy S, Cao S, et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell*. 2020;182(1):200–25. e35.
55. Kachuk C, Stephen K, Doucette A. Comparison of sodium dodecyl sulfate depletion techniques for proteome analysis by mass spectrometry. *J Chromatogr A*. 2015;1418:158–66.
56. Zhang H, Liu T, Zhang Z, et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell*. 2016;166(3):755–65.
57. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:1–16.
58. Manni M, Berkeley MR, Seppey M, et al. BUSCO: assessing genomic data quality and beyond. *Curr Protocols*. 2021;1(12):e323.
59. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
60. Altschul SF, Madden TL, Schäffer A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402.
61. Punta M, Coghill PC, Eberhardt RY, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012;40(D1):D290–301.
62. Götz S, García-Gómez JM, Terol J, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36(10):3420–35.
63. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007;35(Web Server issue):W182–W185.
64. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1–21.
65. Young MD, Wakefield MJ, Smyth GK, et al. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. 2010;11(2):1–12.
66. Mao X, Cai T, Olyarchuk JG, et al. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*. 2005;21(19):3787–93.
67. Jones P, Binns D, Chang HY, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40.

68. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37(1):1–13.
69. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods.* 2001;25(4):402–8.

### **Publisher's note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.