Wu et al. *Zool. Res.* 2024, 45(1): 201−214
https://doi.org/10.24272/j.issn.2095-8137.2023.021

ZR | Zoological Research

Article Open Access

Combined ATAC-seq, RNA-seq, and GWAS analysis reveals glycogen metabolism regulatory network in Jinjiang oyster (*Crassostrea ariakensis*)

Biao Wu[1,2,#], Xi Chen[1,2,#], Jie Hu[3,#], Zhen-Yuan Wang[1,4], Yan Wang[1], Da-You Xu[3], Hao-Bing Guo[3], Chang-Wei Shao[1,2], Li-Qing Zhou[1,2], Xiu-Jun Sun[1,2], Tao Yu[5], Xiao-Mei Wang[5], Yan-Xin Zheng[5], Guang-Yi Fan[3,6,*], Zhi-Hong Liu[1,2,*]

[1] *State Key Laboratory of Mariculture Biobreeding and Sustainable Goods*, *Yellow Sea Fisheries Research Institute*, *Chinese Academy of Fishery Sciences*, *Qingdao*, *Shandong* 266071, *China*
[2] *Laboratory for Marine Fisheries Science and Food Production Processes*, *Laoshan Laboratory*, *Qingdao*, *Shandong* 266071, *China*
[3] *BGI-Qingdao*, *BGI-Shenzhen*, *Qingdao*, *Shandong* 266426, *China*
[4] *National Demonstration Center for Experimental Fisheries Science Education*, *Shanghai Ocean University*, *Shanghai* 201306, *China*
[5] *Changdao Enhancement and Experiment Station*, *Chinese Academy of Fishery Sciences*, *Yantai*, *Shandong* 265800, *China*
[6] *State Key Laboratory of Agricultural Genomics*, *BGI- Shenzhen*, *Shenzhen*, *Guangdong* 518083, *China*

ABSTRACT

Glycogen serves as the principal energy reserve for metabolic processes in aquatic shellfish and substantially contributes to the flavor and quality of oysters. The Jinjiang oyster (*Crassostrea ariakensis*) is an economically and ecologically important species in China. In the present study, RNA sequencing (RNA-seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq) were performed to investigate gene expression and chromatin accessibility variations in oysters with different glycogen contents. Analysis identified 9 483 differentially expressed genes (DEGs) and 7 215 genes with significantly differential chromatin accessibility (DCAGs) were obtained, with an overlap of 2 600 genes between them. Notably, a significant proportion of these genes were enriched in pathways related to glycogen metabolism, including “Glycogen metabolic process” and “Starch and sucrose metabolism”. In addition, genome-wide association study (GWAS) identified 526 single nucleotide polymorphism (SNP) loci associated with glycogen content. These loci corresponded to 241 genes, 63 of which were categorized as both DEGs and DCAGs. This study enriches basic research data and provides insights into the molecular mechanisms underlying the regulation of glycogen metabolism in *C. ariakensis*.

Keywords: *Crassostrea ariakensis*; Glycogen; Transcriptome; ATAC; GWAS

INTRODUCTION

Glycogen, a branched glucose polysaccharide, is ubiquitous across various taxa, including animals and fungi (Liu et al., 2021). It serves multifaceted functions in biological processes, such as energy storage, cell differentiation, signal transduction, redox regulation, and stem cell formation (Zhang et al., 2021). As the primary form of carbohydrate storage in animals, glycogen has garnered considerable attention regarding its role in energy storage and supply. In shellfish, glycogen functions as the most direct and effective energy reserve for metabolism (Liu et al., 2020; Ojea et al., 2004; Smolders et al., 2004). Beyond its biological functions, glycogen content has also been found to affect the flavor, fatness, and color of oyster meat, thereby affecting consumer preferences (Liu et al., 2020). Glycogen content has also been implicated in stress tolerance in oysters (Li et al., 2017). As such, it is considered a critical quality trait in the selection of new oyster varieties, such as *Crassostrea gigas* (Zhang et al., 2020). Overall, glycogen is critical for both the biological mechanisms of oysters and the economic viability of oyster aquaculture.

In 2021, oyster production in China reached 5.82 million tons, accounting for 37.6% of total shellfish production (Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2022). The Jinjiang oyster (*Crassostrea*

Received: 10 May 2023; Accepted: 08 September 2023; Online: 09 September 2023

Foundation items: This work was supported by the National Key R&D Program of China (2022YFD2400105, 2018YFD0900104), Central Public-interest Scientific Institution Basal Research Fund, CAFS (2021XT0102, 2023TD30), Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (2021QNLM050103), Key Research and Development Project of Shandong Province (2021LZGC028), and National Marine Genetic Resource Center

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: fanguangyi@genomics.cn; liuzh@ysfri.ac.cn

ariakensis) exhibits a broad geographic distribution in China, extending from the northern boundary with North Korea to the southern estuarine regions bordering Vietnam. Demonstrating adaptability to a broad spectrum of salinities and temperatures, *C. ariakensis* is well-suited for the aquaculture industry (Li et al., 2021; Wu et al., 2022). It is one of the five major oyster varieties currently cultured in China and is expected to undergo larger-scale cultivation in the near future (Peng et al., 2021; Zhang et al., 2020). Glycogen content in oysters can be influenced by environmental conditions and is closely associated with gonadal development stages (Figure 1A, C). Recent evidence suggests that glycogen, protein, and amino acid levels are subject to genetic control and inheritance (Meng et al., 2019). Given the widely varied and elevated glycogen content in oysters compared to other aquatic species (Figure 1B; Supplementary Table S1), they represent an ideal model for studying glycogen metabolism. However, the molecular regulatory mechanisms underlying glycogen metabolism in oysters remain inadequately understood. As such, more comprehensive research is needed to elucidate the regulation of glycogen metabolism in *C. ariakensis* given its biological and economic importance.

Over the past decade, rapid advances in sequencing technologies have enabled large-scale identification of candidate genes associated with different phenotypes (Swain et al., 2021). Multi-omics techniques, such as whole-genome sequencing, simplified genome sequencing, transcriptomics, proteomics, and metabolomics, have been widely applied to study genetic regulatory mechanisms in shellfish, including preliminary studies on the regulation of glycogen metabolism in *C. gigas*. For example, using transcriptomics, Wang et al. (2021b) identified several non-coding RNAs (ncRNAs) in the gonads of *C. gigas* postulated to indirectly regulate the expression of genes involved in glycogen metabolism, including protein phosphatase 1 regulatory subunit 3b (*ppp1r3b*), isocitrate dehydrogenase 1 (*idh1*), and β-glucosidase. In addition, based on transcriptomics and metabolomics, Chen et al. (2021a) identified several genes related to glycogen metabolism, including glycogen synthase (*gys*), solute carrier family 2 member 4 (*slc2a4*), and glycerol-3-phosphate acyltransferase 4 (*gpat4*). These findings highlight the feasibility and efficiency of multi-omics technologies in screening genes related to quality traits of aquatic products.

Multi-omics analyses can enhance the accuracy and efficiency of gene screening. In this study, an integrated approach combining RNA sequencing (RNA-seq), assay for transposase-accessible chromatin using sequencing (ATAC-

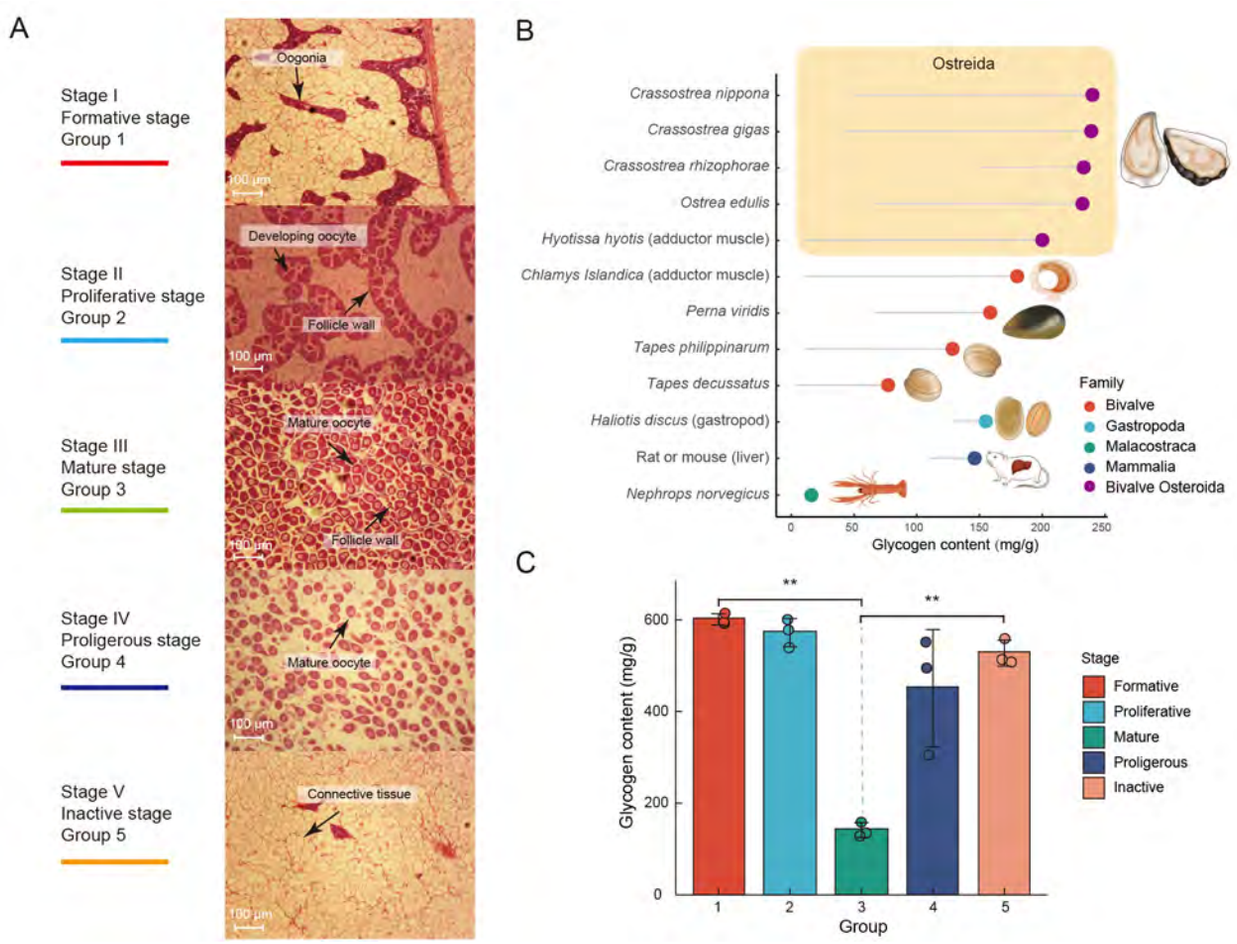

Figure 1 Developmental stages and glycogen content of *Crassostrea ariakensis* gonads

A: Sections of female *C. ariakensis* gonads at different developmental stages: Formative stage (Stage I), Proliferative stage (Stage II), Mature stage (Stage III), Proligerous stage (Stage IV), and Inactive stage (Stage V). B: Lollipop chart of glycogen content of different species (mg/g, dry weight). Head of lollipop represents highest value of glycogen content, and line represents range of glycogen content in certain species. Unless otherwise specified (e.g., specific tissue is mentioned in parentheses for some species), values indicate total glycogen content in all soft tissues of the species (source literature in Supplementary Table S1). C: Column chart of glycogen content in different oyster groups. **: $P<0.01$.

seq), and genome-wide association study (GWAS) was applied to screen for genes implicated in glycogen metabolism in *C. ariakensis*. To the best of our knowledge, this is the first study to employ ATAC-seq to examine the correlation between chromatin accessibility and gene expression related to glycogen metabolism in oysters. More importantly, the concurrent use of ATAC-seq and RNA-seq enabled analysis of whether differentially expressed genes (DEGs) exhibit significantly different chromatin accessibility around their transcription start sites (TSSs), and whether DEGs are regulated by transcription factors associated with specific motifs or footprints within open chromatin regions (Buenrostro et al., 2013). This multifaceted approach also revealed insights into the regulation of glycogen metabolism and cellular heterogeneity among different oyster groups from an epigenetic perspective (Yan et al., 2020). In addition, GWAS enabled the exploration and identification of genes and loci at the genome level, further facilitating analysis of the mechanisms underlying glycogen metabolism. These results contribute to the existing body of basic research data and provide novel insights into the molecular mechanisms governing glycogen metabolism regulation in *C. ariakensis*.

MATERIALS AND METHODS

Experimental oysters

The *C. ariakensis* samples were collected from five locations in China, including Binzhou and Dongying (Shandong Province), Luchaogang (Shanghai), Guangzhou (Guangdong Province), and Beihai (Guangxi Province). A total of 271 oysters were collected, with 250 used for genome resequencing, 15 used for RNA-seq and ATAC-seq, and six used for quantitative real-time polymerase chain reaction (qPCR) experiments. Six tissues, including the mantle, gill, lip, gonad, hepatopancreas and adductor muscle, were sequentially dissected from each oyster, rapidly frozen in liquid nitrogen, and stored at −80°C for DNA/RNA extraction and glycogen content determination. The gonads were also collected for tissue section analysis (see below).

Glycogen content in *C. ariakensis* is closely related to gonadal development (Liu et al., 2021). For RNA-seq analysis, a total of 15 gonads from Dongying oysters were sampled, corresponding to five distinct gonadal development stages characterized by varying glycogen levels. These stages included the formative stage (group 1), proliferative stage (group 2), mature stage (group 3), proligerous stage (group 4), and inactive stage (group 5), with each represented by three biological replicates. Groups 1, 2, and 3 were further selected for ATAC-seq analysis. Detailed sample information is provided in Supplementary Table S2. For GWAS analysis, a total of 250 oysters were collected from the above five locations. Resequencing data for these samples can be found in our previous study (Wu et al., 2022), with additional information shown in Supplementary Table S3. Oysters exhibiting significant differences in glycogen content were collected for qPCR to verify the expression of key candidate genes (Supplementary Table S2).

Determination of gonadal development stages and glycogen content

Tissue sectioning was used to accurately determine the developmental stages of the oyster gonads. Specifically, gonadal tissue (1 cm^3) was dissected and fixed in Bouin's solution for 24 h, then transferred to a 70% alcohol solution for cleaning and preservation. The samples were subsequently dehydrated using ethanol solutions at different concentrations (80%, 90%, 95%, 100%), immersed in wax, cut into sections (4 μm thick), and stained with hematoxylin and eosin (H&E). A light microscope LEICA DM4000 (Leica Microsystems GmbH, Germany) was used to observe the sex and gonad stage of each sample.

Glycogen content in all tissues from 250 oysters was detected using the improved anthrone method (Chen et al., 2021b). In brief, fresh tissues were first lyophilized for 48 h, then ground into powder under freezing conditions. Each sample (20±1 mg) was transferred to a 10 mL test tube and digested for 45 min with 2 mL of 30% potassium hydroxide solution (w/w) under boiling conditions. Total volume was adjusted to 5 mL with distilled water, with centrifugation at 4 000 ×*g* for 8 min at room temperature subsequently performed to collect the supernatant. After being diluted 10 times, 100 μL of the supernatant was mixed with 200 μL of 0.2% (w/v) anthrone (Tokyo Chemical Industry, Japan) sulfuric acid solution on ice. The mixture was then heated in a boiling water bath for 10 min, and finally cooled with tap water. After 30 min at room temperature, 200 μL of the mixture was transferred to a microplate for the determination of optical density (OD) at 620 nm using a microplate reader.

A standard curve was established using glucose solutions with serial concentrations ranging from 0.01 mg/mL to 0.2 mg/mL. Glycogen content in each sample was then calculated using the formula $C=C_0\times(5\ mL/W)\times10\times0.9$, where C is the glycogen content of the sample (mg/g), C_0 is the glucose content in the test solution calculated according to the standard curve (mg/mL), W is the weight of the sample to be tested (g), 5 mL is the constant volume, 10 is the dilution factor, and 0.9 is the conversion factor between glucose and glycogen. The glycogen content in each oyster is shown in Supplementary Table S3.

RNA-seq analysis

Comparative transcriptome analysis was carried out to explore differential gene expression in oysters characterized by different gonadal development stages and glycogen levels. Total RNA was extracted using TRIzol reagent (Thermo Fisher Scientific, USA) following the manufacturer's instructions. Total RNA quality was assessed using an Agilent 2100 Bioanalyzer; subsequently, Oligo (dT) magnetic beads were employed for the enrichment of eukaryotic mRNA. The RNA integrity number of the 15 samples ranged from 8.0 to 9.2 (Supplementary Table S2). The obtained mRNA was fragmented and used as a template for first-strand cDNA synthesis. The cDNA was then used as a template for the generation of double-stranded cDNA, which was subsequently purified and recovered. Following sticky-end repair, an "A" base was added to the 3' end of the cDNA for subsequent adapter ligation. The cDNA was then used for library construction and short-read sequencing. A 2×150 bp paired-end library was successfully constructed and sequenced on the BGISEQ-500 platform (MGI Tech, China).

SOAPnuke v1.4.0 (Chen et al., 2018) was used to remove low-quality reads, joint contaminants, and reads with a high proportion of unknown base N from the initial dataset. After obtaining clean reads, HISAT v2.1.0 (Kim et al., 2019) was used to compare clean reads with the *C. ariakensis* genome (GenBank assembly accession: GCA_020458035.1) to predict

new transcripts, while Bowtie2 v2.2.5 (Langmead & Salzberg, 2012) was used to compare clean reads to the reference sequence for gene alignment. RSEM v1.3.3 (Li & Dewey, 2011) was then applied to calculate the expression levels of genes and transcripts to obtain fragments per kilobase of transcript per million mapped reads (FPKM) (Supplementary Table S4). Principal component analysis (PCA) and hierarchical clustering analysis were respectively conducted using the princomp and hclust functions (Müllner, 2013) in R. DESeq2 v1.30.1 (Love et al., 2014) was used to detect DEGs among different samples, screened based on fold-change>=2 and corrected *P*<= 0.05 (DESeq2 parameters: fold-change>=2.00 and adjusted *P*<=0.05).

GSEA v4.0.0.0 (Subramanian et al., 2005) was used for gene set enrichment analysis (GSEA) of the MSigDB "c2.cp.kegg.7.5.1.symbols.gmt", "c2.cp.reactome.v7.5.1.symbols.gmt", and "c5.go.7.5.1.symbols.gmt" gene sets (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp). To facilitate analysis, the *C. ariakensis* genes were named according to the HGNC database nomenclature (https://www.genenames.org/) (Tweedie et al., 2020).

ATAC-seq analysis

Gonadal cells were prepared according to previously published protocols (Xu et al., 2021), with subsequent transposition reaction, purification, library construction, and quantification conducted using established methods (Buenrostro et al., 2015). Nine samples were sequenced using 50 bp paired-end reads on the BGISEQ-500 platform (MGI Tech, China). The quality of the sequencing data was assessed using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Contaminating primer and adapter sequences were removed, followed by the filtration of low-quality reads to obtain high-quality clean reads. Bowtie2 v2.2.5 (Langmead & Salzberg, 2012) was subsequently used to align the clean reads to the *C. ariakensis* genome (GenBank assembly accession: GCA_020458035.1). Read depth distributions within gene regions, as well as regions located 3 kb upstream and downstream in the genome, were then calculated.

Peak calling was performed using MACS2 v2.1.0 (https://pypi.python.org/pypi/MACS2/) to obtain peak summit, peak cumulative density, standardized *P*- values, and fold enrichment information. A region was considered a peak at a *P*-value<1×10^{-5}. Peaks were classified into different gene functional elements, including TTS, exon, intron, intergenic, and promoter elements, and the proportions of peaks in those regions were calculated according to peak location information. All peaks were associated with genes according to the positional relationship between peak and genes (*C. ariakensis* gene annotation files are available in the CNGBdb with the accession number CNA0022698.). Using DESeq2 v1.30.1 (Love et al., 2014), different peaks among samples were screened, and related genes were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis.

GWAS

GWAS was conducted to explore the single nucleotide polymorphisms (SNPs) associated with the regulation of glycogen metabolism. Data related to genome resequencing and SNP loci of 250 individuals were obtained from our previous study (Wu et al., 2022), and glycogen content was detected using the abovementioned method. To reduce false positive associations, SNP quality control was carried out using VCFtools with parameters "--max- missing 0.99 --maf 0.05 -- minDP 4 -- minGQ 10 -- minQ 30 --min- meanDP 10". GWAS analysis was carried out using EMMAX (Kang et al., 2010). PLINK v1.9 was used for population stratification assessment and PCA (Purcell et al., 2007). Considering the factors that influence glycogen content, PCA values associated with varying sample collection dates and gonadal developmental stages in oysters were used as covariates to adjust for population structure and the kinship matrix. Considering tissue-specific regulation of glycogen metabolism, GWAS was conducted for each tissue separately. The genomic inflation value (λ) was calculated using the R package "qqman".

Dimensionality reduction via PCA was performed on the glycogen content data across six different tissues, with the first principal component (PC1) used as phenotypic data for GWAS analysis (Ning et al., 2019). Prior to that, missing data were completed using the missMDA package in R (Supplementary Table S5). To screen as many target candidate genes related to glycogen content as possible, the threshold for GWAS was set to *P*<1×10^{-5} (He et al., 2021; Li et al., 2019). Upon identifying significant SNPs, standard practice dictates their automatic mapping to proximal genes (Brodie et al., 2016).

Prediction of potential transcription factors of target genes was carried out using ChEA3 (https://maayanlab.cloud/chea3/) (Keenan et al., 2019). Gene functional enrichment analysis in the context of GWAS was carried out using Metascape (https://metascape.org/gp/index.html) (Zhou et al., 2019). Prior to this, the annotated gene names of *C. ariakensis* were standardized using HGNC database nomenclature (Tweedie et al., 2020).

Combination and visualization of multi-omics analysis results

Visualization of gene structure, RNA-seq, ATAC-seq, and SNPs was performed using IGV v2.14.0 (Thorvaldsdóttir et al., 2013). Subsequent color matching and integration were carried out using Adobe Illustrator (https://adobe.com/products/illustrator).

Motif enrichment analysis of peak sequences within the ATAC-seq data was performed using the SEA module of the MEME suite v5.4.1 (Bailey & Grant, 2021). Based on the enrichment results and annotations in the HOCOMOCO human (v11 FULLI) motif database with default parameters, potential transcription factor-binding sites and transcription factor types in the peak sequences were identified.

Heatmaps depicting gene expression and box plots representing glycogen content were generated using Hiplot (Li et al., 2022). Linkage disequilibrium (LD) blocks for significant genes were plotted using LDBlockShow (Dong et al., 2021). The glycogen metabolism pathway diagram was created using Servier Medical Art image resources (https://smart.servier.com/).

Gene validation using qPCR

Five genes screened through joint RNA-seq and ATAC-seq analysis and four genes identified from GWAS were selected for qPCR validation. The oysters were divided into two groups (high and low glycogen content, respectively) for analysis. Three biological replicates for each group (Supplementary Table S2) and three technical replicates for each biological

replicate were utilized. In brief, cDNA was synthesized from 1 μg of total RNA using HiScript III RT Super Mix for qPCR (Vazyme, China), with qPCR then performed using an ABI StepOnePlus Real-Time PCR System with Tower (Applied Biosystems, USA) and ChamQ SYBR Color qPCR Master Mix (Vazyme, China) following the manufacturers' instructions. Melting curve analysis was performed at the end of the total number of cycles to confirm amplification specificity. The β-actin (GenBank Number: AF026063) gene served as the internal control and the primers used for qPCR analysis are listed in Supplementary Table S6.

RESULTS

Oysters with different glycogen content showed significant transcriptomic differences

A total of 15 individuals, categorized into five groups based on glycogen levels (group 1: 601 mg/g, group 2: 572 mg/g, group 3: 141.3 mg/g, group 4: 450.6 mg/g, group 5: 527.3 mg/g), were subjected to sequencing on the BGISEQ-500 platform, yielding an average of 6.11 Gb of data per sample. The average alignment rate of the sample genome was 73.58%, indicative of high-quality sequencing (Supplementary Table S7). Overall, 27 266 genes, including 25 327 known genes and 1 950 predicted new genes, were identified. Among these genes, 22 589 were present in all samples (Figure 2C; Supplementary Table S8).

The PCA results revealed that the 15 samples were relatively closely related based on PC1 (abscissa) but could be divided into two large categories based on PC2 (ordinate): groups 3 and 4 belonged to one category, while groups 1, 2, and 5 belonged to the other (Figure 2A). This division was consistent with the observed patterns of oyster gonadal development.

The DEG analysis results were consistent with the PCA findings (Figure 2B, left side of dotted line). While a small number of DEGs were observed between groups 1 and 2, 1 and 5, and 3 and 4, with counts ranging from 45 to 341, a large number of DEGs were identified between all other pairwise comparisons, ranging from 3 578 to 7 296 (Figure 2B, right side of dotted line). Notably, groups 3 and 5, which exhibited significant differences in glycogen content, had the highest number of DEGs (Figure 2C). Relative similarity in gene expression was found among groups 1, 2, and 5, and

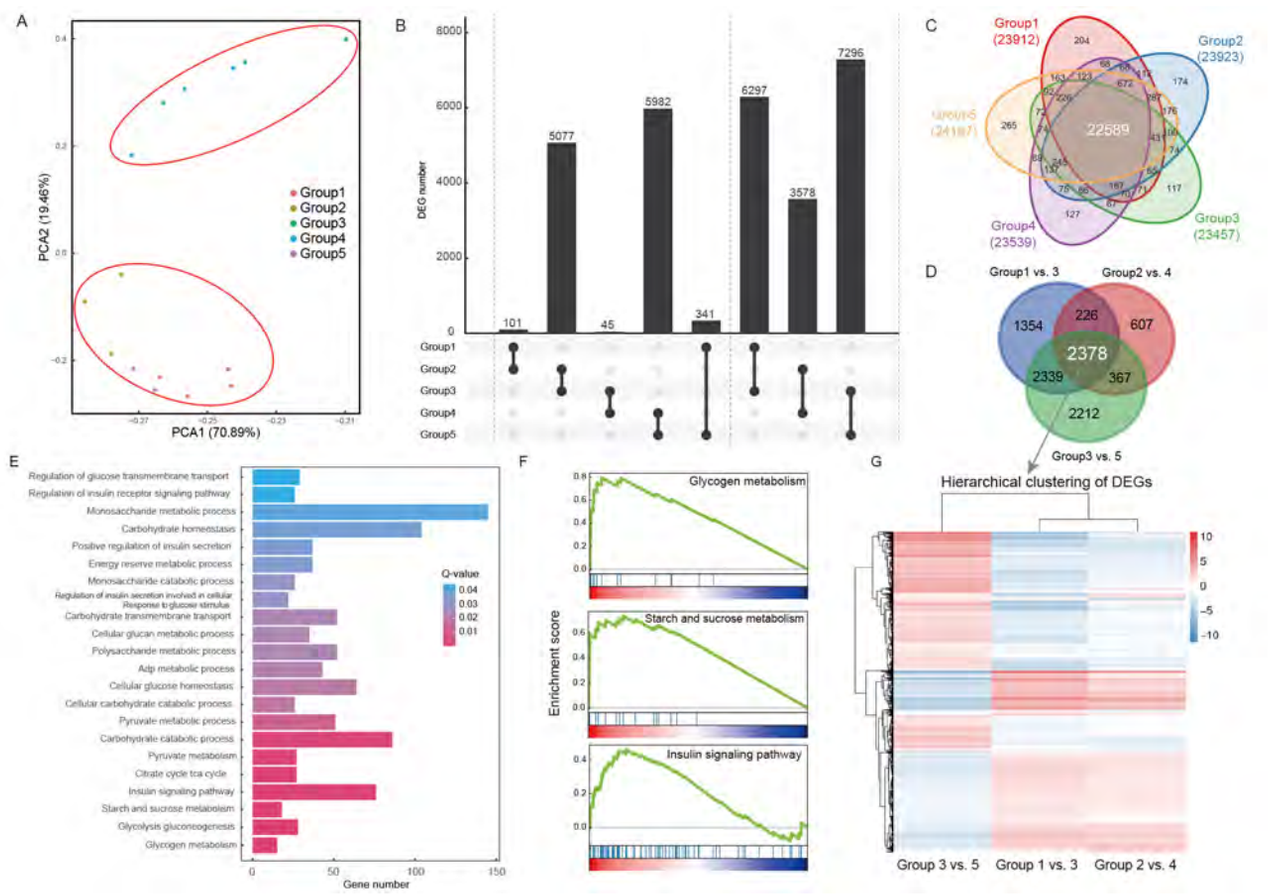

Figure 2 Transcriptomic analysis of *Crassostrea ariakensis* gonad tissue at different developmental stages

A: PCA of *C. ariakensis*. B: DEG number between different groups of *C. ariakensis*. C: Venn diagram of DEGs among five groups of *C. ariakensis*. D: Venn diagram of DEGs between groups 1 and 3, groups 2 and 4, and groups 3 and 5. E: Summary of pathways and entries closely related to glycogen metabolism in GSEA. F: Enrichment plot: Glycogen metabolism, Insulin signaling pathway, and Starch and sucrose metabolism. Top part of each graph shows enrichment score (ES) of each gene. In the middle part of each graph, each vertical line represents a gene in the gene set and its rank position. Bottom part of each graph shows gene-phenotype association matrix. Red indicates positive correlation with first phenotype (stage I or group 1) and up-regulation of gene expression in group 1, while blue indicates positive correlation with second phenotype (stage III or group 3) and up-regulation of gene expression in group 3. G: Hierarchical cluster analysis of DEGs between groups 1 and 3, groups 2 and 4, and groups 3 and 5.

between groups 3 and 4. Comparative analysis between groups 1 and 3, 2 and 4, and 3 and 5 yielded a total of 9 483 DEGs, 2 378 of which were shared across these group comparisons (Figure 2D). Moreover, these shared DEGs demonstrated analogous expression patterns in groups 1 and 3 and in groups 2 and 4, in contrast to the patterns observed in groups 3 and 5 (Figure 2G). This observation is in alignment with the distinct phases of oyster gonadal development, notably diverging in the mature and proligerous stages relative to other developmental stages (Figure 1A), implying that these DEGs are intricately linked to the process of gonadal development. As gonadal development has been shown to correlate with glycogen content in oysters (Chen et al., 2021a; Liu et al., 2020; Qin et al., 2021), the observed expression patterns also correlated with the glycogen phenotypes of these groups, whereby glycogen content was higher in groups 1, 2, and 5 than in groups 3 and 4. Consequently, the identified DEGs are likely associated with glycogen content in oysters.

GSEA identified pathways closely related to glycogen metabolism

GSEA was performed on the entire set of 27 266 identified genes to ascertain differential pathways. Given that groups 1 and 3 exhibited a considerable number of DEGs (6 297) (Figure 2C), along with marked differences in glycogen content (Figure 1C), these two groups were selected for GSEA as well as subsequent ATAC-seq analysis. Among the differential pathways identified, 22 were selected as potential pathways related to glycogen metabolism (Figure 2E; Supplementary Table S9).

Enrichment analysis based on the GO gene set revealed that all genes could be classified into 3 633 gene sets, 1 875 of which were up-regulated in group 1 (Supplementary Table S10) and 1 758 of which were up-regulated in group 3 (Supplementary Table S11). Based on GO enrichment analysis, two gene sets were enriched in glycogen metabolism-related terms, including "Energy Reserve Metabolic Process" and "Cellular Glucan Metabolic Process". Within the Reactome gene set, 15 genes were found to be involved in the oyster-specific "Glycogen Metabolism Pathway". Notably, within this pathway, 12 genes were up-regulated in group 1, characterized by high glycogen content (Supplementary Figure S1). Additional enrichment analyses identified other gene sets enriched in glycogen metabolism-related terms, including "Regulation of Glucose Transmembrane Transport" (Figure 2F; Supplementary Figure S2).

Upon conducting KEGG enrichment analysis, 142 gene sets were identified, with 92 gene sets up-regulated in group 1 and 50 in group 3 (Supplementary Tables S12–S13). Several of these gene sets were enriched in glycogen metabolism-related pathways, including "Starch and Sucrose Metabolism" and "Insulin Signaling Pathway" (Figure 2F). Notably, 18 genes were enriched in the "Starch and Sucrose Metabolism" pathway, 17 of which exhibited high expression in group 1, while only one showed high expression in group 3 (Supplementary Figure S1). These findings suggest the potential activation of the "Starch and Sucrose Metabolism" and "Insulin Signaling Pathway" in group 1.

The genes screened by RNA-seq analysis were taken as important candidate genes related to glycogen metabolism in further studies.

Analysis of ATAC-seq data demonstrated high-sequencing quality

ATAC-seq analysis performed on groups 1, 2, and 3 generated between 7.64 and 12.72 Gb of clean reads per sample, with an average size of 9.55 Gb (Supplementary Table S14). Comparison of these clean reads to the reference genome revealed read mapping rates ranging from 85.47% to 92.92%, indicative of high-quality sequencing (Supplementary Tables S15–S16).

The distribution of read depths across gene regions, as well as 3 kb upstream and downstream, is shown in Figure 3A and Supplementary Figure S3. Average sequencing depths in those regions were lower than those near the TSS and transcription end site (TES) of each gene. Peak proportions in the TTS and promoter regions were lower than those in the intron intergenic and exon regions (Figure 3B; Supplementary Tables S16, S17). These results align with the expected characteristics of the ATAC-seq data. The fragment size distribution of the acquired DNA spanned both the nucleosome and nucleosome-free regions (NFR). The curve patterns observed across the nine samples aligned with the characteristics of the ATAC-seq data (Figure 3C; Supplementary Figure S4), thereby reaffirming the high quality of the sequencing results.

As the peak regions could be either protein-binding domains or genomic regions regulated subject to histone modification, correlations between peak regions and genomic genes were analyzed. Among the three groups analyzed, 21 133 genes were identified as being associated with peaks, 18 612 of which were shared by all three groups (Figure 3D).

Genes associated with differential peaks were enriched in glycogen metabolism-related pathways

Due to the substantial number of DEGs found in groups 1, 2, and 3, ATAC-seq analysis was conducted on these three groups. Analysis identified many differential peaks corresponding to 7 215 genes. Groups 1 and 3 exhibited the largest number of differential peaks, associated with 6 841 genes. In contrast, groups 1 and 2 displayed the fewest differential peaks, associated with only 129 genes (Figure 3E). The differential peaks observed between groups 2 and 3 were related to 3 021 genes, 90.8% (2 743) of which were also observed between groups 1 and 3 (Figure 3D). In addition, a total of 2 600 genes overlapped with those identified in RNA-seq analysis (Figure 3D). Based on KEGG enrichment analysis, 21 genes between groups 1 and 3 were enriched in the "Starch and sucrose metabolism" pathway, suggesting direct involvement in oyster glycogen metabolism (Figure 3F; Supplementary Table S18).

Combined RNA-seq and ATAC-seq analysis

Based on GSEA and ATAC-seq data, five KEGG pathways, one Reactome pathway, and 16 GO terms potentially relevant to glycogen metabolism were identified (Figure 2E). Specifically, the "Glycogen metabolic process" and "Starch and sucrose metabolism" pathways are directly implicated to glycogen metabolism. These pathways included 67 genes covering 54 functional categories, highlighting their possible importance in glycogen metabolism (Figure 4A; Supplementary Table S19). Of these 67 genes, 49 were identified either through RNA-seq or ATAC-seq analyses, with 21 detected in both. Subsequent Metascape enrichment

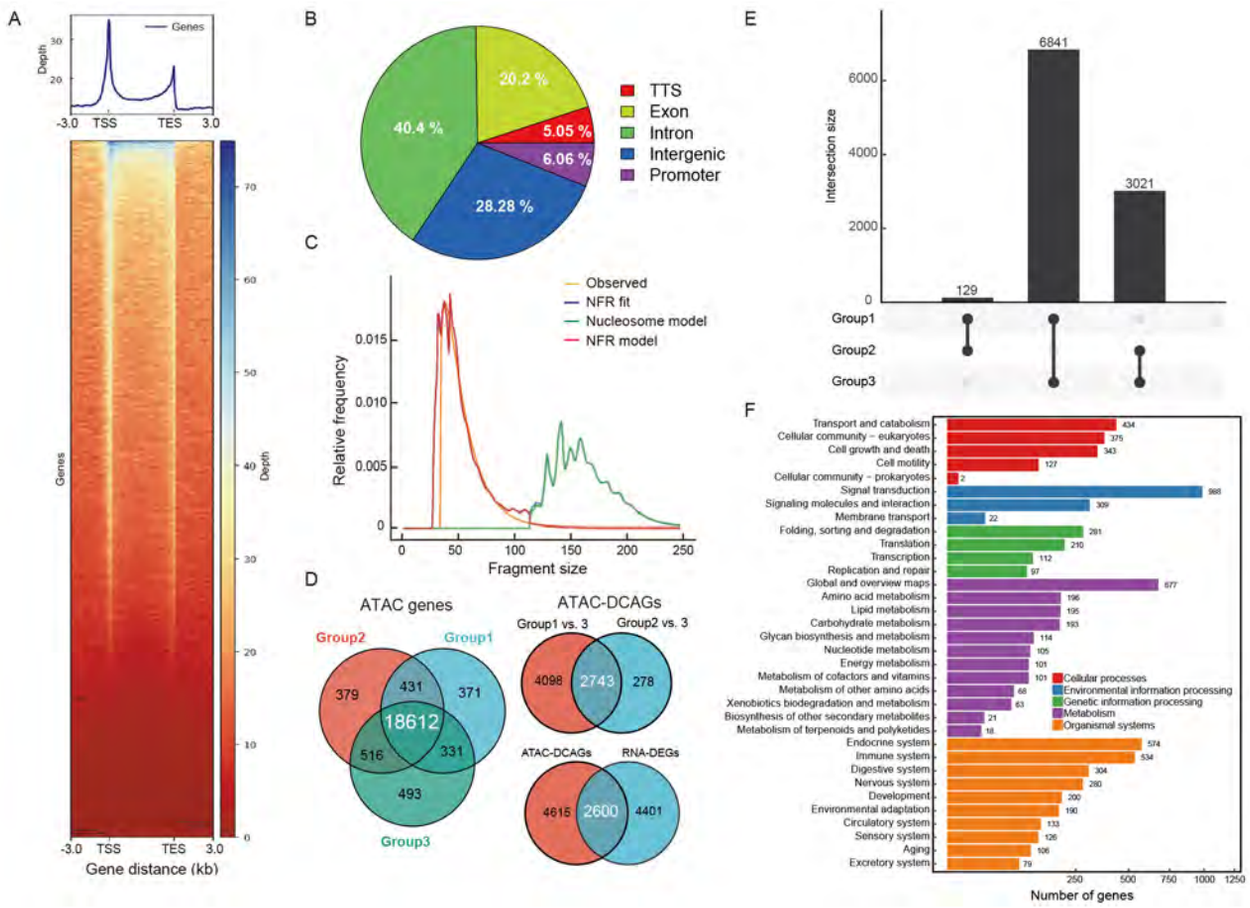

Figure 3 Comprehensive analysis of ATAC-seq data

A: Sequencing depth distribution map of promoter (up 3 kb) region of the sample gene. Abscissa represents sequence length and ordinate represents sequencing depth. B: Distribution map of peaks in gene functional elements. C: Results of nucleosome localization prediction. D: Venn diagram of genes related to peaks in different groups revealed by ATAC-seq (left), Venn diagram of genes associated with differential peaks between different groups revealed by ATAC-seq (right upper), and Venn diagram of DEGs revealed by ATAC-seq and RNA-seq (right lower). E: Number of genes related to differential peaks between different groups. F: KEGG enrichment analysis of DEGs between groups 1 and 3.

analysis of these 67 genes revealed the top six enriched pathways, including “Glycogen metabolism process”, “Starch and sucrose metabolism”, “Insulin signaling pathway”, “Glycogen synthesis and degradation”, “Galactose metabolic”, and “Glucagon signaling pathway” (Figure 4C; Supplementary Table S20). Furthermore, these terms and pathways were closely related (Figure 4D), suggesting that the identified genes may be involved in multiple metabolic pathways potentially relevant to oyster glycogen metabolism.

To verify the accuracy of our multi-omics analysis, five well-known genes closely related to glycogen metabolism were selected for further analysis, including glycogen debranching enzyme (*gde*), glycogen synthase (*gys*), glycogen phosphorylase (*pyg*), UTP-glucose-1-phosphate uridylyltransferase (*ugpa*), and phosphoglucomutase-1 (*pgm1*). RNA-seq analysis indicated significant disparities in gene expression among the various oyster groups, while distinct differences in chromatin accessibility were observed for *pyg*, *ugpa* , and *pgm1* across these groups. The qPCR results confirmed that the five genes were highly expressed in the high-glycogen group, corroborating the activation of glycogen metabolism-related pathways in this group as indicated by the multi-omics data (Figure 4B).

GWAS of glycogen phenotypes yielded 394 SNPs

Glycogen content (dry weight) was measured in six different tissue types across 250 individuals collected from five geographic locations (Supplementary Table S3). Results showed that glycogen content varied widely, from 5.6 mg/g to 128.9 mg/g in the adductor muscle, 21.0 mg/g to 464.2 mg/g in the gill, 28.31 mg/g to 706.9 mg/g in the mantle, 62.8 mg/g to 626.5 mg/g in the hepatopancreas, 24.72 mg/g to 674.36 mg/g in the gonads, and 103.62 mg/g to 749.64 mg/g in the labial palps (Figure 5A). Average glycogen content in the gonads (366.3 mg/g) and labial palps (399.0 mg/g) was significantly higher than that in the other four tissues (P<0.01). Compared to other tissues, glycogen content in the adductor muscle was relatively stable, consistent with previous studies (Liu et al., 2020; Qin et al., 2021). Glycogen content in the hepatopancreas, labial palps, and gonads showed a relatively normal distribution pattern, whereas that in the other tissues showed a skewed distribution (Figure 5B), potentially due to the small sampling size.

A total of 4.8 billion paired-end reads were generated from the 250 individuals, yielding an average sequencing depth of approximately 22.43× per oyster (Supplementary Table S3). Following reference genome alignment and SNP calling, a total of 9 653 216 SNPs were identified. Post-filtration, 1 918 741 high-quality SNPs were retained for subsequent GWAS analysis.

Manhattan and quantile-quantile (QQ) plots are shown in

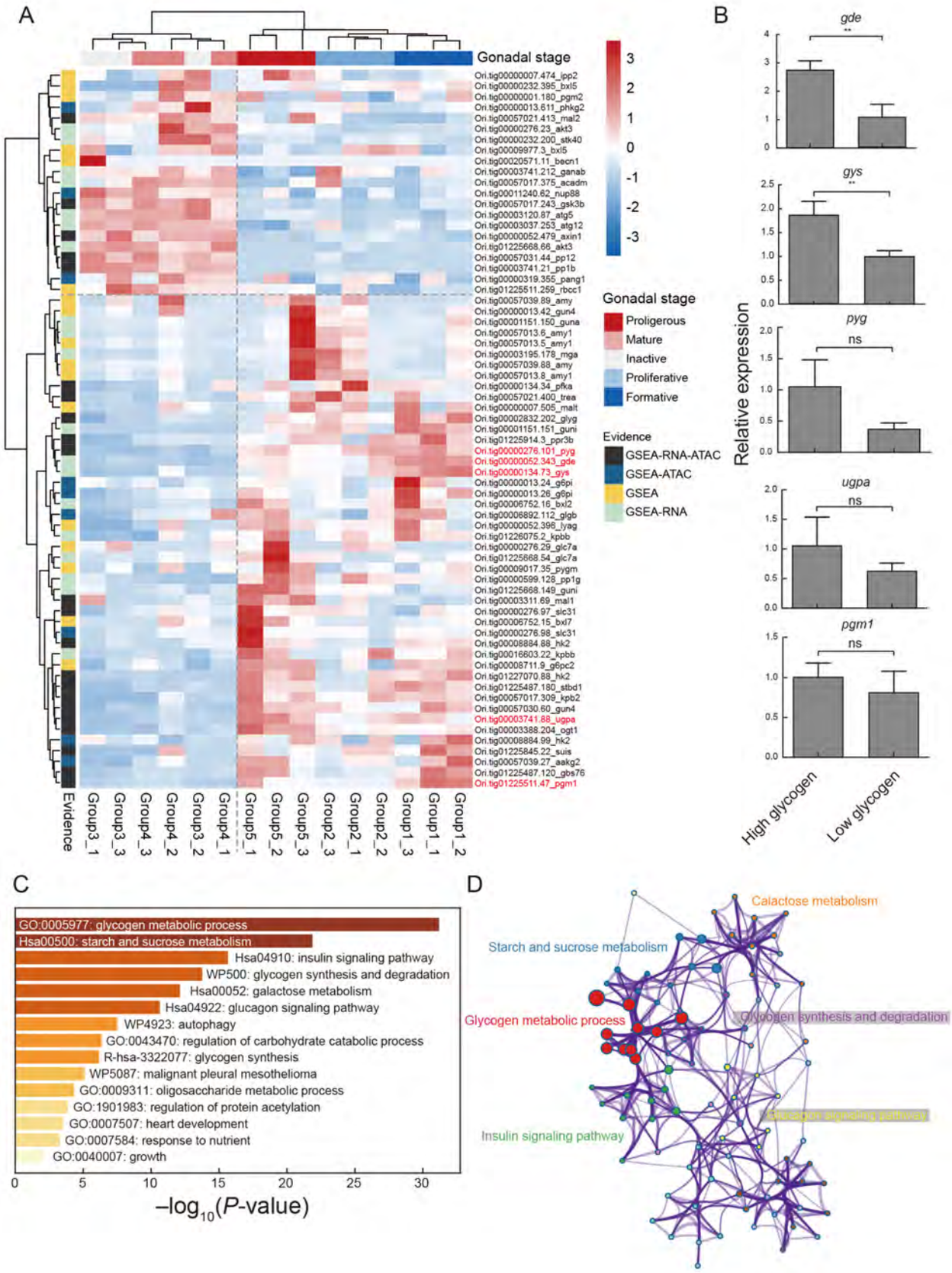

Figure 4 Summary analysis of genes related to glycogen metabolism

A: Heatmap of expression of genes related to glycogen metabolism. “Evidence” bar means that different analyses, including GSEA, ATAC-seq and RNA-seq, were used to reveal genes related to glycogen metabolism. Genes in red are validated by qPCR in B. B: qPCR verification of five genes, including *gde*, *gys*, *pyg*, *ugpa* , and *pgm1* . Vertical axis of bar graph represents relative expression level. ns: Not significant; **: P <0.01. C: Metascape enrichment analysis of genes in A. D: Correlation diagram among individual pathways in C.

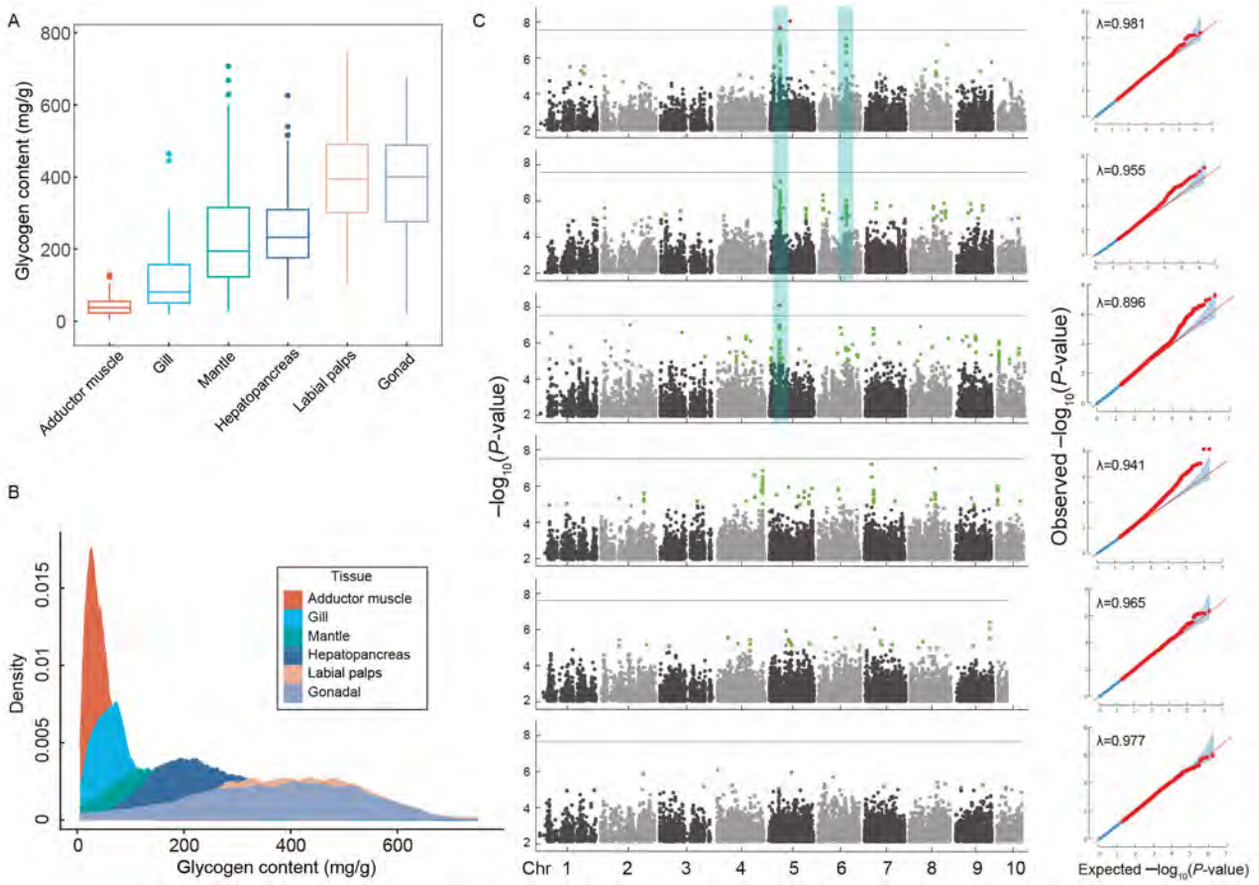

Figure 5 Glycogen contents in six different tissues of *Crassostrea ariakensis* and GWAS analysis

A: Box plot of glycogen content in six tissues from 250 oysters. B: Distribution of glycogen content in different tissues among individuals. C: GWAS analysis of glycogen content in six tissues in *C. ariakensis*. For each tissue, a Manhattan plot and corresponding QQ-plot (upper right) are shown. λ represents associated genomic inflation value. Dotted line corresponds to an ordinate value of 5 ($-\log_{10}P$) and solid line corresponds to an ordinate value of 7.58, calculated from $-\log_{10}$ (0.05/1 918 741).

Figure 5C. In total, 394 SNPs potentially associated with glycogen content were screened (Supplementary Table S21). Of these, 15 SNP loci, each located within a distinct gene, were detected in at least two tissues. Moreover, several significant loci (highlighted in blue boxes) located in two common chromosome regions (chromosomes 5 and 6) were detected in the gonad, mantle, and gill tissues. These genes warrant further study. The standard QQ plot confirmed the reliability of these SNPs across tissues.

GWAS after PCA dimensionality reduction of glycogen data obtained 132 significant SNPs

PCA of glycogen contents showed that PC1 accounted for 41.86% of the total phenotypic variance, indicating the feasibility of PC1 representing the measured glycogen content trait (Supplementary Table S22). Therefore, GWAS was further carried out using PC1 as a compound glycogen content trait to obtain more comprehensive SNPs (Figure 6A). Results yielded 132 SNP loci associated with the target trait, more than those yielded on a tissue basis (Supplementary Table S23). The *P*-values of nine loci met the Bonferroni test threshold criteria ($P<2.61\times10^{-8}$), as shown by the red dots in Figure 6A, and they were directly related to six genes. Among these six genes, growth factor independent 1 (*gfi1*) and Brachyury protein homolog A (T-box, *tbx5*), corresponding to the two SNPs, chr5_15152293 and chr5_15219936 (15 152 293rd and 15 219 936th base loci on chromosome 5, respectively), were identified as transcription factors ($-\log_{10}P>7.58$), which are involved in the regulation of many genes (Kazanjian et al., 2006; Smith, 1997) (Figure 6D). In addition, two SNPs associated with hexokinase type 2 (*hk2*) and glucose transporter type 1 (*glut1*) were also found on chromosomes 1 and 9, respectively (Figure 6A), which may be involved in glycogen metabolism (Bramer et al., 2017; Fajardo et al., 2021). SNPs falling within the *hk2* (chr9_37174195) gene were divided into three genotypes, containing CC, AC, and AA. Glycogen content corresponding to the three genotypes (represented by the first column of data after PCA dimensionality reduction) is shown in a violin diagram (Figure 7C), which suggested that the three genotypes were characterized by significantly different glycogen contents ($P<0.01$). These results further verified that this SNP was correlated with glycogen content.

The qPCR results shown in Figure 6C indicated that *hk2*, Ori.tig00057017.233, and *glut1* were differentially expressed between the high- and low-glycogen oysters. Notably, *hk2* was highly expressed in the high-glycogen oysters, while Ori.tig00057017.233 and *glut1* were highly expressed in the low-glycogen oysters. The *gfi1* gene was slightly more up-regulated in the low-glycogen group compared to the high-glycogen group. These expression patterns were consistent with the transcriptome data (Supplementary Figure S5).

Multi-omics analysis

Combining these 132 SNPs with the above 394 SNPs, we obtained 460 SNP loci belonging to 241 genes

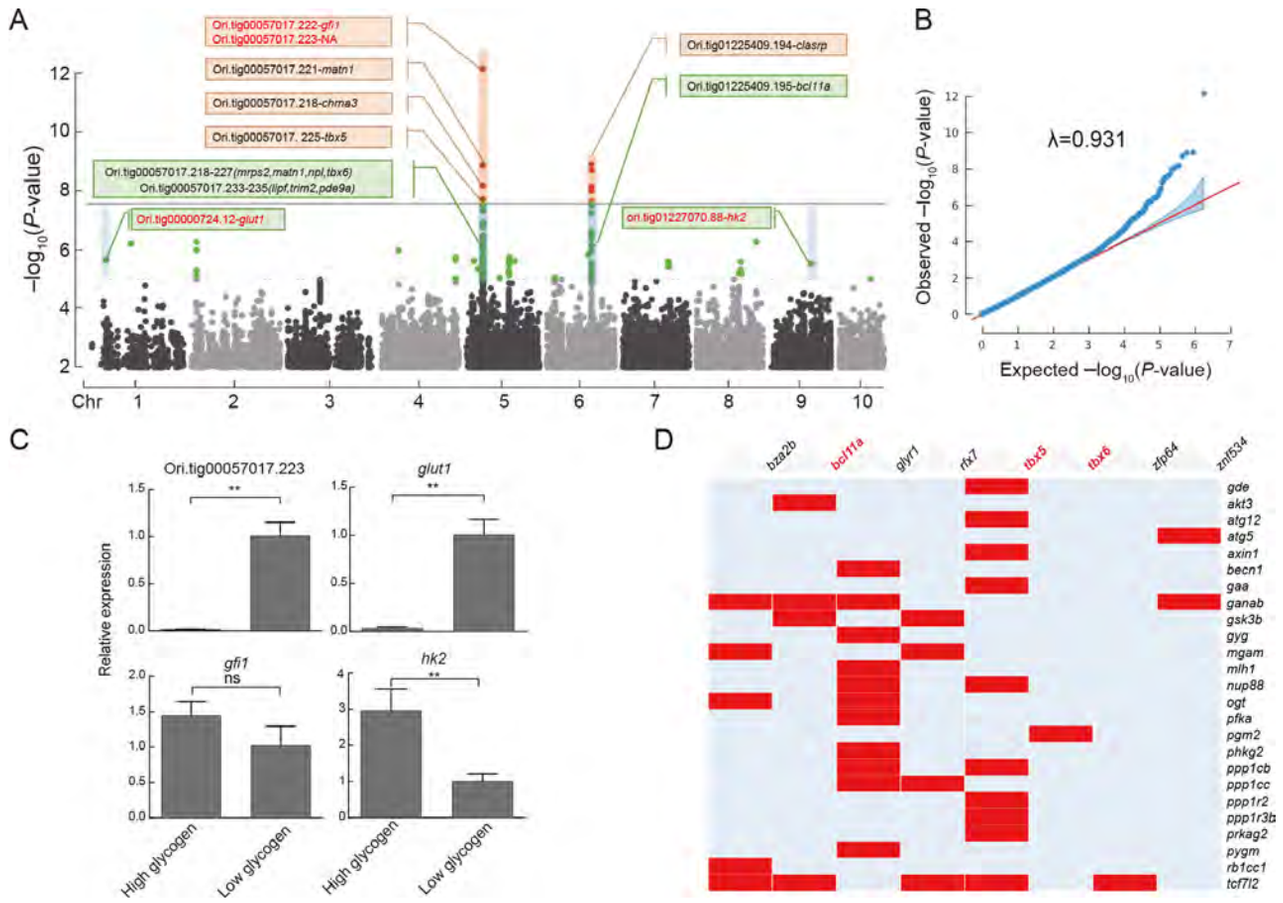

Figure 6 Analysis of genes associated with significant SNPs

A: Manhattan plot obtained by GWAS analysis of glycogen content data after dimensionality reduction using GEMMA software. Dotted line corresponds to an ordinate value of 5 ($-\log_{10}P$) and solid line corresponds to an ordinate value of 7.58, calculated from $-\log_{10}$ (0.05/1 918 741). B: QQ-plot corresponding to Manhattan plot, λ represents associated genomic inflation value. C: qPCR verification of four genes in oysters, including *gfi1*, Ori.tig00057017.233, *glut1*, and *hk2*. Vertical axis of bar graph represents relative expression level. ns: Not significant; **: P<0.01. D: Bar graph of enriched terms across input gene lists, colored by P- values. E: Glycogen metabolism-related genes corresponding to transcription factors obtained by GWAS.

(Supplementary Tables S21–S23). Further combining with RNA-seq results, 120 of the 241 genes showed significantly different expression among oysters with different glycogen contents (Supplementary Figure S6), while 63 of the 120 genes were characterized by different chromatin accessibility in oysters with different glycogen contents. We determined that these genes were probably related to oyster glycogen metabolism. Notably, 52 of the 241 genes were transcription factors, which could be classified into 26 transcription factor types (Supplementary Table S24). Through target gene prediction, eight transcription factors were involved in regulating glycogen metabolism, including BAF chromatin remodeling complex subunit Bcl11A (*bcl11a*), *tbx5*, and T-box transcription factor 6 (*tbx6*) (Figure 6D; Supplementary Table S25).

Through RNA-seq, ATAC-seq, and GWAS analysis, several important genes were identified, including *hk2*. The *hk2* gene plays an important role in glycogen metabolism, as it can convert glucose into glucose 6 phosphate, a precursor of glycogen synthesis (Blaha et al., 2022). We also found that higher chromatin accessibility of the *hk2* transcription initiation region was associated with higher gene expression (Figure 8A). The LD-block of all SNP loci in this gene showed strong linkage imbalance among them (Figure 7B). The SNP (chr9_37174195) was exactly localized to the peak (Chr9:37173892~37174226) of the *hk2* gene intron. This peak region showed higher chromatin accessibility in groups 1 and 2 than in group 3 (Figure 7A), suggesting this region may be closely related to *hk2* transcription. Further analysis predicted a motif binding to the forkhead box O1 (*foxo1*) transcription factor in this peak region (Supplementary Tables S26, S27). Therefore, this SNP site may affect the binding of transcription factor *foxo1* in this region, thereby affecting *hk2* gene expression.

We finally integrated all multi-omics results and selected 12 key genes to explore the regulation of glycogen metabolism in *C. ariakensis*, with differences in transcriptional expression and chromatin accessibility observed between the high- and low-glycogen groups (Figure 8). Specifically, seven genes (*glut1*, *hk2*, *pgm1*, *ugp*, *glyg*, *pyg* , and *foxo1*) exhibited significant differences in both mRNA expression and chromatin accessibility, while two genes (*gys* and *agl*) exhibited significant differences in mRNA expression. Excluding *glut1* , the remaining eight genes were highly expressed in the high-glycogen group. Additionally, insulin (*ins*) showed differences in chromatin accessibility but not in mRNA expression. Similar to the seven DCAGs described above, *ins* showed high chromatin accessibility in the high-glycogen group.

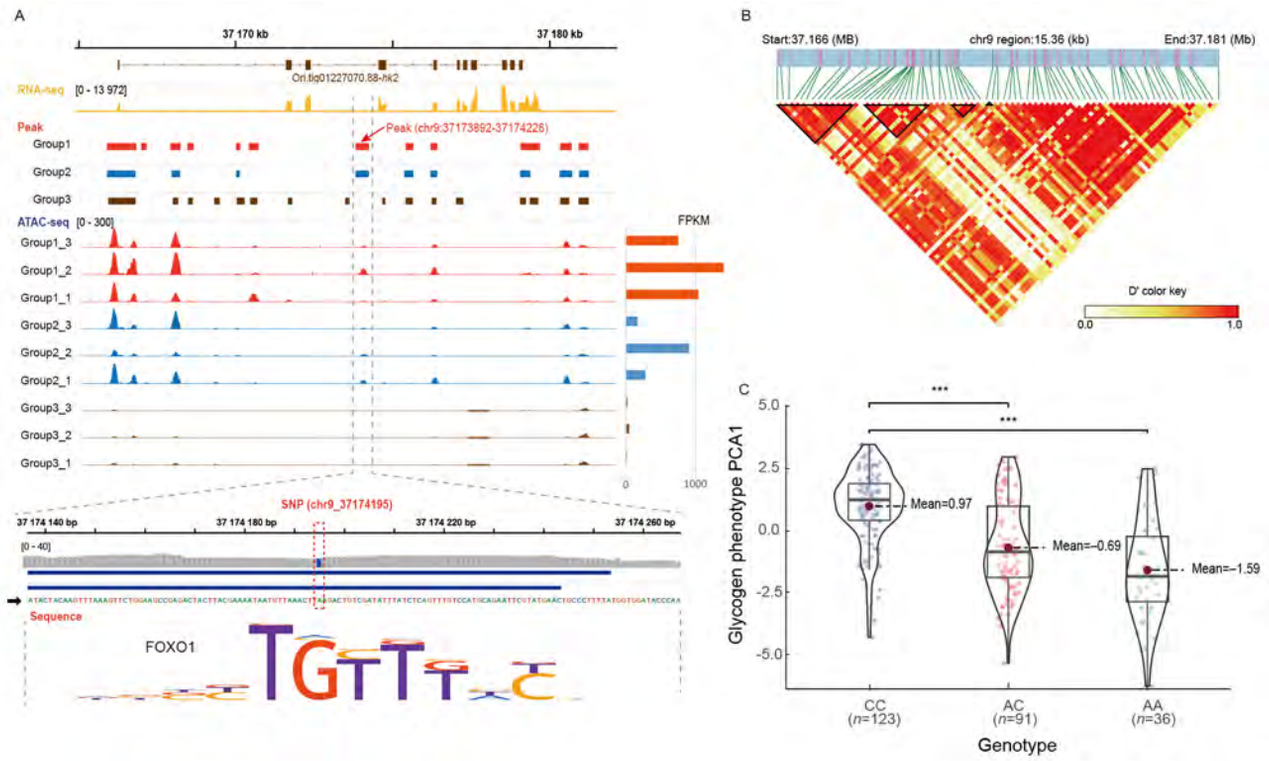

Figure 7 Multi-omics analysis of *hk2*

A: Visualization of ATAC-seq and RNA-seq data of *hk2*. Column chart is expression level of *hk2* in oyster gonads at different developmental stages, expressed in FPKM; bottom is predicted binding transcription factor and motif in this peak. B: LD-block diagram of *hk2* gene. C: Violin diagram showing glycogen content in oysters with different genotypes. Glycogen content was obtained after PCA dimensionality reduction. ***: P<0.001.

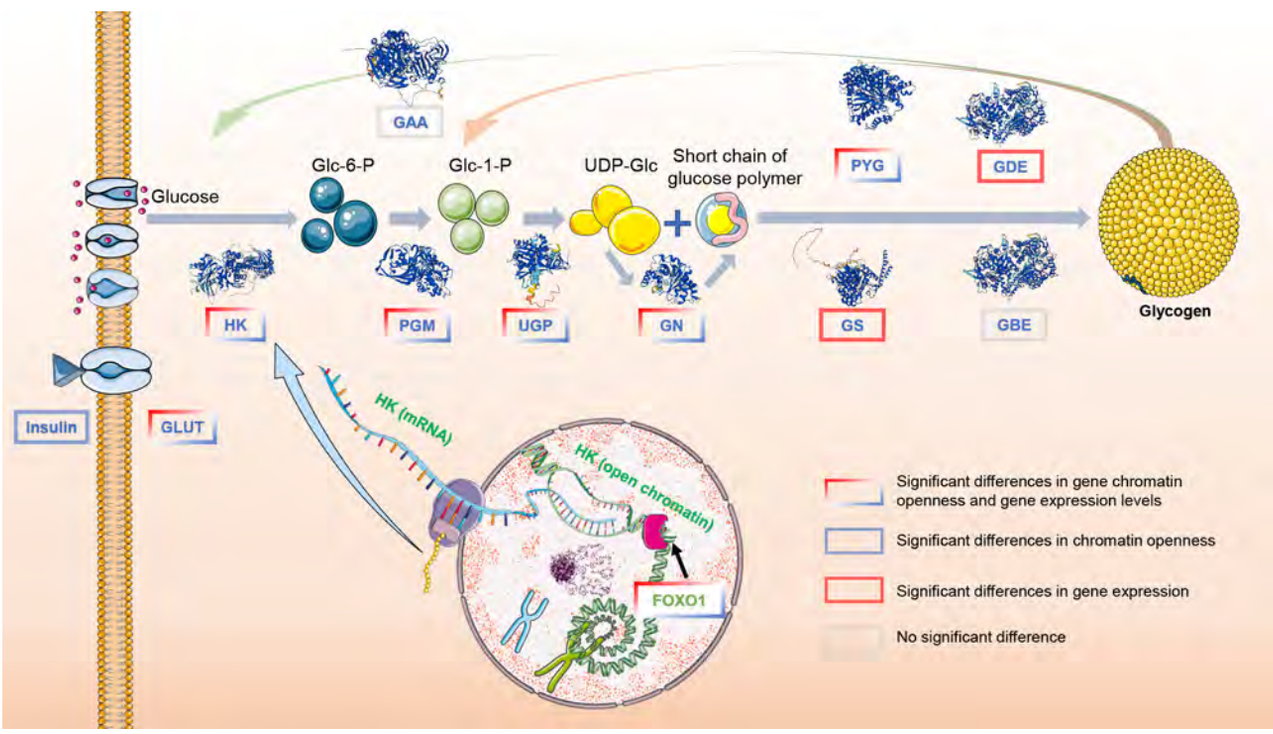

Figure 8 Diagram of glycogen metabolism pathway

Glycogen metabolism-related enzymes are shown in blue, and substance molecules involved in glycogen metabolism process are in black. Arrows indicate reaction direction. Three-dimensional structure of enzyme is quoted from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/). GLUT, glucose transporter; HK, hexokinase; PGM, phosphoglucomutase; UGP, UDP-glucose pyrophosphatase; GN, glycogen protein; GS, glycogen synthase; GBE, glycogen branching enzyme; PYG, glycogen phosphorylase; GDE, glycogen debranching enzyme; GAA, lysosomal alpha-glucosidase; FOXO1, forkhead box protein O1; Glc-6-P, Glucose-6-Phosphate; Glc-1-P, Glucose-1-Phosphate; UDP-Glc, Uridine Diphosphate Glucose.

DISCUSSION

GSEA revealed expression patterns of oyster glycogen metabolism-related gene sets

Various investigations have explored oyster glycogen metabolism based on transcriptomics (Chen et al., 2021a; Li et al., 2017; Wang et al., 2021a, 2021b). While certain DEGs and pathways related to glycogen metabolism have been identified, the comprehensive expression patterns of specific pathways have not yet been fully elucidated. In the current study, a multitude of gene sets were identified, including 22 directly related to glycogen metabolism. These important gene sets were notably up-regulated in oysters exhibiting high-glycogen content, a finding not reported in prior studies. Our approach facilitated the identification of diverse gene sets, thereby providing a more in-depth understanding of the regulatory network underlying oyster glycogen metabolism.

ATAC-seq revealed relationship between chromatin accessibility and expression of key genes

Despite a decade since its initial introduction (Buenrostro et al., 2013), ATAC-seq technology has only been applied in a single mollusk-based study to date (Xu et al., 2021). Notably, using both ATAC-seq and ChIP-seq, Xu et al. (2021) demonstrated that the upstream regulatory element of the oyster insulin gene (*cgILP*) can bind to the Pdx transcription factor, a finding congruent with the interaction between the vertebrate insulin-like gene and Pdx1 transcription factor. In vertebrate systems, the insulin gene plays a pivotal role in glycogen synthesis by inhibiting glycogen phosphorylase and concurrently activating glycogen synthase (Adeva-Andany et al., 2016). Consequently, we hypothesize that insulin-like genes in oysters may similarly be integral to glycogen synthesis.

In this study, ATAC-seq was employed for the first time to investigate changes in chromatin accessibility in *C. ariakensis* with varying levels of glycogen content. Contrary to traditional transcriptomic analysis, ATAC-seq not only yielded genome-wide data on chromatin accessibility but also enabled a more nuanced understanding of the epigenetic mechanisms regulating oyster glycogen metabolism. Among the identified chromatin open regions with significant accessibility differences, many genes involved in oyster glycogen metabolism were identified, including the insulin-like gene, *hk2*, *gys*, and *pyg*. Beyond the direct identification of genes related to differential chromatin accessibility, we also used the open chromatin sequences associated with 67 glycogen metabolism genes to predict 91 motifs and the transcription factors that may bind to them.

GWAS identified multiple genes and transcription factors associated with oyster glycogen content

In oysters, glycogen content exhibits marked variability among individuals and tissues (Meng et al., 2019; Shi et al., 2020), influenced by both genetic and environmental factors, such as season, nutrients, salinity, and temperature (Liu et al., 2020). To identify the principal factors affecting glycogen content, we utilized wild *C. ariakensis* oysters sourced from an extensive geographical range for GWAS analysis. Population structure, sampling time, and gonadal development stage were incorporated as the first, second, and third covariates, respectively, to eliminate potential environmental impact.

Limited GWAS analyses have been conducted on glycogen phenotypic traits in oysters. Using GWAS, Shi et al. (2020) identified the citrate lyase subunit beta-like (*clybl*) mitochondrial protein, which may enhance glycogen synthesis by promoting the accumulation of precursor substances. Similarly, Meng et al. (2019) mapped 17 SNP loci associated with 24 genes based on GWAS; notably, cytochrome P450 17A1 (*cyp17a1*) contained a non-synonymous SNP, displayed elevated expression in individuals with higher glycogen content, and potentially enhanced gluconeogenesis by transcriptionally regulating the expression of phosphoenolpyruvate carboxykinase (*pepck*) and glucose 6-phosphatase (*g6p*).

Previous GWAS studies on oyster glycogen content have focused on whole soft tissue analysis. Nevertheless, mammalian studies suggest that genes regulating glycogen metabolism may vary across different tissues, such as the brain, liver, and muscle (Adeva-Andany et al., 2016). Therefore, we conducted GWAS analysis of glycogen content in six individual tissues. The resultant Manhattan diagrams for the mantle, gill, and gonad were very similar, while that for the adductor muscle differed markedly. This divergence aligns with findings showing that genes regulating glycogen metabolism in mammalian muscle are distinct from those in other tissues (Adeva-Andany et al., 2016). However, further evidence is needed to ascertain whether glycogen metabolism regulation in oyster muscle differs from that in other tissues.

To better understand the relationship between overall glycogen content and SNPs, PCA was employed to reduce the dimensionality of the glycogen content data, with PC1 used as the phenotype for GWAS analysis. This approach yielded more significant SNP loci, the regions of which were consistent with our additional GWAS analysis, thereby validating the reliability of this method.

As an important enzyme in the glycogen synthesis pathway, hk2 mediates the initial step of glycolysis by catalyzing the phosphorylation of D-glucose to D-glucose 6-phosphate. Bramer et al. (2017) suggested that *hk2* serves as a key regulator of glycogen accumulation during equine diestrum and pregnancy. Glucose functions as an essential energy source for almost all life and is the only raw material for glycogen synthesis. *Glut1* is a member of the glucose transporter protein family, a diverse group of membrane proteins. Fajardo et al. (2021) showed that overexpression of *glut1* in mouse hearts can lead to increased glucose uptake and glycogen accumulation, highlighting the regulatory role of *glut1* in glycogen metabolism.

Multi-omics analysis revealed oyster glycogen regulatory network

Obtaining a complete regulatory network through single omics analysis is a challenge. While transcriptomic analysis can effectively identify DEGs related to glycogen metabolism, it is restricted to the RNA level (Wang et al., 2021b). Similarly, a single GWAS can only isolate SNPs associated with genes implicated in glycogen metabolism based on SNP locations. Therefore, for the first time, we combined RNA-seq, ATAC-seq, and GWAS analyses for a comprehensive exploration of glycogen metabolism regulation. We analyzed the open chromatin sequences in oyster gonads to elucidate the regulatory mechanisms of gene expression after transcription factor binding, thus establishing a regulatory network from DNA to RNA. The GWAS findings further clarified the important roles of candidate genes in glycogen metabolism at the DNA level. Notably, the multi-omics data could be cross-

validated, and further analysis on gene function and experimental phenotype provided a clear depiction of the biological process of epigenetic regulation-gene expression-gene function-phenotype.

In the present study, *hk2* , a critical gene implicated in glycogen metabolism, was identified through RNA-seq, ATAC-seq, and GWAS analysis. We hypothesized that the SNP site identified by GWAS was associated with the binding domain of the transcription factor *foxo1* (Sekine et al., 2007; Singh et al., 2017). Furthermore, co-expression between *foxo1* (Ori. tig00008711.64) and *hk2* (Ori. tig01227070.88) was observed in the transcriptomic data, suggesting that *foxo1* may serve as an enhancer regulating expression of the *hk2* gene.

Based on multi-omics data analysis, we identified several key genes related to glycogen metabolism. Notably, many key genes implicated in vertebrate glycogen metabolism were also identified in the oyster genome. These findings suggest that glycogen metabolism regulation in oysters and vertebrates may be evolutionarily conserved, although further evidence is needed to substantiate this claim. Future research should clarify the transcription factors regulating glycogen metabolism-related genes in oysters and explore the differences and similarities between mollusk and mammalian glycogen metabolic networks through multi-dimensional comparisons.

In conclusion, the present study employed multi-omics approaches to comprehensively investigate the relationship between chromatin accessibility and expression levels of key genes related to glycogen metabolism in *C. ariakensis*, and further revealed the underlying regulatory network for glycogen metabolism. This study provides valuable data and establishes a robust foundation for subsequent research.

CONCLUSIONS

Through GSEA and transcriptomic sequencing of oysters with different glycogen phenotypes, we observed the activation of specific pathways, including “Starch and sucrose metabolism”, “Insulin signaling pathway”, and “Glycogen metabolic process”, in oysters with high-glycogen content. Furthermore, genes in these pathways showed higher expression levels. Concurrently, ATAC-seq analysis revealed that genes with differential chromatin accessibility were also enriched in these pathways, thereby suggesting a direct relationship between chromatin accessibility and expression in genes. Furthermore, GWAS analysis indicated that transcription factors play a crucial role in the regulation of glycogen metabolism. In addition to SNPs directly related to transcription factors, several SNPs were found in key genes implicated in glycogen metabolism, such as *hk2*. Within open chromatin regions, we hypothesized that these SNP loci may influence the binding affinity of transcription factors, thereby modulating gene expression. Subsequently, a glycogen metabolism network in oysters was delineated based on these pathways. Remarkably, most of these key genes are also present in mammals, suggesting a high degree of evolutionary conservation in glycogen metabolism networks across diverse species.

DATA AVAILABILITY

The *C. ariakensis* RNA- seq data and ATAC-seq data generated in this study were submitted to the China National GeneBank Database (CNGBdb; https://db.cngb.org/; accession number CNP0003045), National Center for Biotechnology Information (NCBI SRA; https://www.ncbi.nlm.nih.gov/sra; accession number PRJNA999955), Genome Sequence Archive (GSA; https://ngdc.cncb.ac.cn/gsa; accession number PRJCA018671), and Science Data Bank (SDB; https://www.scidb.cn/en; Data DOI: 10.57760/sciencedb.09809). The *C. ariakensis* assembled genome and whole-genome resequencing datasets that support the findings of this study are available at CNGBdb (doi.org/10.1111/1755-0998.13556, accession number CNP0001149).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS’ CONTRIBUTIONS

B.W., X.C., and G.Y.F. designed the research. X.C., J.H., D.Y.X., and H.B.G. analyzed the data. B.W. and X.C. wrote the manuscript. Y.W. and Z.Y.W. performed histological sectioning and validation experiments. C.W.S., L.Q.Z., X.J.S., T.Y., X.M.W., and Y.X.Z. revised the manuscript. Z.H.L. guided revision of the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We would like to thank the China National GeneBank, National Center for Biotechnology Information, Genome Sequence Archive, and Science Data Bank for technical support in data service. Parts of Figure 8 were drawn using images from Servier Medical Art, licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).

REFERENCES

Adeva-Andany MM, González-Lucán M, Donapetry-García C, et al. 2016. Glycogen metabolism in humans. *BBA Clinical*, **5**: 85−100.

Bailey TL, Grant CE. 2021. SEA: simple enrichment analysis of motifs. *bioRxiv*, doi:https://doi.org/10.1101/2021.08.23.457422.

Blaha CS, Ramakrishnan G, Jeon SM, et al. 2022. A non-catalytic scaffolding activity of hexokinase 2 contributes to EMT and metastasis. *Nature Communications*, **13**(1): 899.

Bramer SA, Macedo A, Klein C. 2017. Hexokinase 2 drives glycogen accumulation in equine endometrium at day 12 of diestrus and pregnancy. *Reproductive Biology and Endocrinology*, **15**(1): 4.

Brodie A, Azaria JR, Ofran Y. 2016. How far from the SNP may the causative genes be. *Nucleic Acids Research*, **44**(13): 6046−6054.

Buenrostro JD, Giresi PG, Zaba LC, et al. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, **10**(12): 1213−1218.

Buenrostro JD, Wu BJ, Chang HY, et al. 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, **109**: 21.29.1−21.29.9.

Chen C, Yu H, Li Q. 2021a. Integrated proteomic and transcriptomic analysis of gonads reveal disruption of germ cell proliferation and division, and energy storage in glycogen in sterile triploid pacific oysters (*Crassostrea gigas*). *Cells*, **10**(10): 2668.

Chen X, Wu B, Wang Y, et al. 2021b. Establishment and optimization of micro-reaction system for determination of oyster glycogen content. *South China Fisheries Science*, **17**(4): 126−132. (in Chinese)

Chen YX, Chen YS, Shi CM, et al. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience*, **7**(1): 1−6.

Dong SS, He WM, Ji JJ, et al. 2021. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on

variant call format files. *Briefings in Bioinformatics*, **22**(4): bbaa227.
Fajardo VM, Feng I, Chen BY, et al. 2021. GLUT1 overexpression enhances glucose metabolism and promotes neonatal heart regeneration. *Scientific Reports*, **11**(1): 8669.
He X, Li CY, Qi HG, et al. 2021. A genome-wide association study to identify the genes associated with shell growth and shape-related traits in Crassostrea gigas. *Aquaculture*, **543**: 736926.
Kang HM, Sul JH, Service SK, et al. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, **42**(4): 348−354.
Kazanjian A, Gross EA, Grimes HL. 2006. The growth factor independence-1 transcription factor: New functions and new insights. *Critical Reviews in Oncology/Hematology*, **59**(2): 85−97.
Keenan AB, Torre D, Lachmann A, et al. 2019. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research*, **47**(W1): W212−W224.
Kim D, Paggi JM, Park C, et al. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**(8): 907−915.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357−359.
Li A, Dai H, Guo XM, et al. 2021. Genome of the estuarine oyster provides insights into climate impact and adaptive plasticity. *Communications Biology*, **4**(1): 1287.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**(1): 323.
Li BS, Song K, Meng J, et al. 2017. Integrated application of transcriptomics and metabolomics provides insights into glycogen content regulation in the Pacific oyster Crassostrea gigas. *BMC Genomics*, **18**(1): 713.
Li JF, Miao BB, Wang SX, et al. 2022. Hiplot: a comprehensive and easy-to-use web service for boosting publication-ready biomedical data visualization. *Briefings in Bioinformatics*, **23**(4): bbac261.
Li Y, Cao K, Zhu GR, et al. 2019. Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biology*, **20**(1): 36.
Liu QH, Tang JW, Wen PB, et al. 2021. From prokaryotes to eukaryotes: insights into the molecular structure of glycogen particles. *Frontiers in Molecular Biosciences*, **8**: 673315.
Liu S, Li L, Wang W, et al. 2020. Characterization, fluctuation and tissue differences in nutrient content in the Pacific oyster (*Crassostrea gigas*) in Qingdao, northern China. *Aquaculture Research*, **51**(4): 1353−1364.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**(12): 550.
Meng J, Song K, Li CY, et al. 2019. Genome-wide association analysis of nutrient traits in the oyster *Crassostrea gigas*: genetic effect and interaction network. *BMC Genomics*, **20**(1): 625.
Ministry of Agriculture and Rural Affairs of the People's Republic of China, National Fisheries Technology Extension Center, China Society of Fisheries. 2022. China Fishery Statistical Yearbook. Beijing: China Agriculture Press. (in Chinese)
Müllner D. 2013. fastcluster: fast hierarchical, agglomerative clustering routines for R and python. *Journal of Statistical Software*, **53**(9): 1−18.
Ning XH, Li X, Wang J, et al. 2019. Genome-wide association study reveals *E2F3* as the candidate gene for scallop growth. *Aquaculture*, **511**: 734216.
Ojea J, Pazos AJ, Martínez D, et al. 2004. Seasonal variation in weight and biochemical composition of the tissues of Ruditapes decussatus in relation to the gametogenic cycle. *Aquaculture*, **238**(1-4): 451−468.
Peng DM, Zhang SC, Zhang HZ, et al. 2021. The oyster fishery in China: trend, concerns and solutions. *Marine Policy*, **129**: 104524.
Purcell S, Neale B, Todd-Brown K, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**(3): 559−575.
Qin YP, Li XY, Li J, et al. 2021. Seasonal variations in biochemical composition and nutritional quality of Crassostrea hongkongensis, in relation to the gametogenic cycle. *Food Chemistry*, **356**: 129736.
Sekine K, Chen YR, Kojima N, et al. 2007. Foxo1 links insulin signaling to C/EBPα and regulates gluconeogenesis during liver development. *The EMBO Journal*, **26**(15): 3607−3615.
Shi RH, Li CY, Qi HG, et al. 2020. Construction of a high-resolution genetic map of Crassostrea gigas: QTL mapping and GWAS applications revealed candidate genes controlling nutritional traits. *Aquaculture*, **527**: 735427.
Singh P, Han EH, Endrizzi JA, et al. 2017. Crystal structures reveal a new and novel FoxO1 binding site within the human glucose-6-phosphatase catalytic subunit 1 gene promoter. *Journal of Structural Biology*, **198**(1): 54−64.
Smith J. 1997. Brachyury and the T-box genes. *Current Opinion in Genetics & Development*, **7**(4): 474−480.
Smolders R, Bervoets L, De Coen W, et al. 2004. Cellular energy allocation in zebra mussels exposed along a pollution gradient: linking cellular effects to higher levels of biological organization. *Environmental Pollution*, **129**(1): 99−112.
Subramanian A, Tamayo P, Mootha VK, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43): 15545−15550.
Swain P P, Sahoo L, Kumar R, et al. 2021. Applications of next-generation sequencing in aquaculture and fisheries. *In*: Pandey PK, Parhi J. Advances in Fisheries Biotechnology. Singapore: Springer, 41–64.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, **14**(2): 178−192.
Tweedie S, Braschi B, Gray K, et al. 2020. Genenames. org: the HGNC and VGNC resources in 2021. *Nucleic Acids Research*, **49**(D1): D939−D946.
Wang CG, Li A, Wang W, et al. 2021a. Integrated application of transcriptomics and metabolomics reveals the energy allocation-mediated mechanisms of growth-defense trade-offs in *Crassostrea gigas* and *Crassostrea angulata*. *Frontiers in Marine Science*, **8**: 744626.
Wang X, Wang WJ, Li Z, et al. 2021b. Comprehensive analysis of differentially expressed ncRNA, mRNA, and their ceRNA networks in the regulation of glycogen content in the Pacific oyster. *Crassostrea gigas. Aquaculture*, **531**: 735895.
Wu B, Chen X, Yu MJ, et al. 2022. Chromosome-level genome and population genomic analysis provide insights into the evolution and environmental adaptation of Jinjiang oyster Crassostrea ariakensis. *Molecular Ecology Resources*, **22**(4): 1529−1544.
Xu F, Marlétaz F, Gavriouchkina D, et al. 2021. Evidence from oyster suggests an ancient role for Pdx in regulating insulin gene expression in animals. *Nature Communications*, **12**(1): 3117.
Yan F, Powell DR, Curtis DJ, et al. 2020. From reads to insight: a hitchhiker' s guide to ATAC-seq data analysis. *Genome Biology*, **21**(1): 22.
Zhang GF, Li L, Que HY. 2020. An evolution of oyster mariculture industry in China: new knowledge, variety and product. *Oceanologia et Limnologia Sinica*, **51**(4): 740−749. (in Chinese)
Zhang HF, Ma JW, Tang K, et al. 2021. Beyond energy storage: roles of glycogen metabolism in health and disease. *The FEBS Journal*, **288**(12): 3772−3783.
Zhou YY, Zhou B, Pache L, et al. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, **10**(1): 1523.