Genome-wide study of DNA methylation in the early development of the amphipod crustacean *Parhyale hawaiensis*



Wei Wei

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

Acknowledgements

First of all, I would like to express my deep gratitude to my supervisor Prof. Aziz Aboobaker. His extensive expertise and profound sense of responsibility have been crucial to my academic journey. I deeply appreciate his insightful guidance, consistent support, and the countless hours he has devoted to helping me navigate the complexities of my research. I would also like to thank my colleagues Manuel and Salha, with whom I collaborated on a fantastic project. Manuel, thank you very much for willingness to assist me whenever I encountered difficulties. We had many discussions, and you taught me many bioinformatic techniques, provided me with useful suggestions and research ideas. Salha thank you for helping me a lot in the lab. I also want to thank other colleagues in the laboratory, who provided me with help and a conducive environment for academic discussion, which enabled me to complete my research successfully.

I would like to thank Wenyuan, Tianyu, Liyu, Yian, Zhenhuan, as well as all my friends I met in Oxford and those who are in China. With your company, care, and help, I had the happiest time and created the most gorgeous memories during my doctoral studies, which made me feel less lonely despite being far away from home.

To my roommate PiPi, I want to say thank you for accompanying me during the two years of the pandemic. We chatted together, went shopping together, encouraged and supported each other. When I was anxious and depressed, you comforted me. Without your company, I could not have successfully passed through the most difficult days. I wish you a lifetime of happiness.

I also want to thank my parents for their unwavering love. They have given me unlimited material and spiritual encouragement and support, enabling me to pursue my studies without any concern and giving me the chance to become their pride.

Finally, to Hongbo, thank you for your love, support, patience, and understanding. You were always there to encourage and accompany me whenever I felt down. You prepared all of my favourite foods and took me to all of my desired destinations. Whenever I faced difficulties, you were the first to show up, always willing to make me laugh and bring joy into my life. Meeting someone as outstanding as you have been my greatest fortune.

Abstract

DNA methylation is a crucial epigenetic modification associated with the regulation of gene expression in vertebrates and has dynamic patterns associated with the reprogramming of parental methylomes during early embryogenesis. However, the dynamics and underlying mechanisms of DNA methylation during early invertebrate development remain inadequately understood. Maternal gene products initially control early animal development until the maternal genome hands over control to the zygotic genome, known as maternal to zygotic transition (MZT). In vertebrates, DNA methylation plays a key role in ensuring the successful transfer of developmental control across generations during MZT. It remains unclear whether and how DNA methylation is involved in the MZT of invertebrates. The amphipod crustacean *Parhyale hawaiensis* (*Parhyale*) genome encodes DNA methyltransferases (DNMT1 and DNMT3) and exhibits higher DNA methylation levels than many other invertebrates. Therefore, in this thesis, I used *Parhyale* as an animal model to investigate the potential function of DNA methylation in relation to the control of gene expression and the control of transposon activity, both of which are essential factors in early development.

First, I confirmed the presence of DNA methylation in the *Parhyale* genome by analysing the distribution of DNA methylation using Whole-Genome Bisulfite Sequencing (WGBS) and PacBio single molecule real-time (SMRT) data from distinct *Parhyale* tissues and mixed-stage embryos. Next, I applied Enzymatic Methyl-seq to three early developmental stages surrounding the onset of MZT during *Parhyale* embryogenesis. This revealed that 8% of CpGs in *Parhyale* were methylated, with a particular focus on repetitive elements within non-coding regions. These methylation data were then integrated with RNA-seq data. A positive correlation between gene body methylation and gene expression was discovered, particularly at the highly expressed genes, which were found to have high repetitive content and elevated intronic repeat methylation. The functional significance of DNA methylation was further investigated using the DNA methylation inhibitor, 5-aza-2'-

deoxycytidine (5AzaD), which was administered to pre-MZT embryos. The study found that 5AzaD treatment caused developmental delays and was associated with impaired transcription of genes with higher levels of methylation in early development.

Our findings indicate that DNA methylation may play a role in the early development of *Parhyale*, potentially acting as an epigenetic regulator of precise temporal dynamics of gene expression. DNA methylation deposition is necessary for the successful transfer of developmental control from maternally supplied gene products to those synthesized from the zygotic genome. These results contribute to our understanding of the evolutionary conservation of DNA methylation patterns and offer valuable insights into the epigenetic regulation of developmental processes in invertebrates.

Author contributions

Chapter 2: Aziz Aboobaker (AA) conceived the idea. Peter Sarkies's group provided the library constructions and performed the whole genome bisulfite sequencing. Salha Rawas (SR) extracted DNA for PacBio SMRT sequencing. Manuel Jara Espejo (MJE) extracted CpG methylation level from the raw PacBio SMRT sequencing data. Wei Wei (WW) performed WGBS and PacBio data analyses.

Chapter 3: WW, SR and AA conceived the idea. WW carried out embryo collections, DNA extractions, preparation of EM-seq libraries, library quantifications, sequencing, and EM-seq data analysis. SR performed RNA extraction, MJE and SR processed RNA-seq data, genome assembly and annotations.

Chapter 4: WW, MJE and AA conceived the idea. WW performed all laboratory processing including embryo collections, RNA extraction, live imaging, primer design and PCR cloning. WW processed RNA-seq data.

Contents

Chapter 1 General introduction 1

1.1 Introduction 2

1.1.1 Epigenetics 2

1.1.2 DNA methylation 3

1.1.3 DNA methylation toolkits 4

1.1.3.1 The writers of DNA methylation 4

1.1.3.2 The readers of DNA methylation 6

1.1.3.3 The erasers of DNA methylation 7

1.1.4 The DNA methylation landscape in animals 7

1.1.4.1 Vertebrates 7

1.1.4.2 Invertebrates 10

1.1.5 Conservation of methylation patterns 13

1.1.6 Cellular function of DNA methylation 15

1.1.6.1 Methylation at CGI promoters 15

1.1.6.2 Methylation at gene bodies 17

1.1.6.3 Methylation at repetitive elements 19

1.1.7 *Parhyale hawaiensis* 21

1.1.7.1 An overview of *Parhyale hawaiensis* 21

1.1.7.2 *Parhyale* as an animal model system 23

1.1.7.3 Embryonic development of *Parhyale* 24

1.2 Thesis aim and outline 26

Chapter 2 Genome-wide DNA methylation analysis of *Parhyale hawaiensis* 28

2.1 Introduction 29

2.1.1 DNA methylation in *Parhyale hawaiensis* 29

2.1.2 Genome-wide detection of cytosine methylation by PacBio single molecule real-time sequencing 29

2.1.3 Aim of this chapter 32
2.2 Materials and methods 34
2.2.1 Whole genome bisulfite sequencing read alignment 34
2.2.2 WGBS data statistical analysis 35
2.2.3 PacBio single molecule real-time sequencing and methylation analysis 35
2.2.4 Methylation level of different genomic features 36
2.2.5 Gene body methylation analysis 37
2.3 Results and discussion 38
2.3.1 Repetitive elements classification and annotation 38
2.3.1.1 Approaches introduction 38
2.3.1.2 Comparison of outputs from EarlGrey and RR 39
2.3.2 Summary of sequencing data 45
2.3.2.1 Whole genome bisulfite sequencing data 45
2.3.2.2 PacBio single molecule real-time sequencing data 48
2.3.3 Methylation targets in the *Parhyale* genome 50
2.3.4 Gene body methylation 52
2.3.4.1 Gene body methylation pattern 52
2.3.4.2 Comparison of gene body methylation across different samples 56
2.4 Conclusion 59
Chapter 3 DNA methylation analysis in the early development of *Parhyale hawaiensis* 61
3.1 Introduction 62
3.1.1 Early development of animals 62
3.1.2 DNA methylation during the MZT 65
3.1.3 Method to detect DNA methylation 66
3.1.4 Aim of this chapter 69
3.2 Materials and methods 71
3.2.1 *Parhyale* culture 71
3.2.2 Embryo collection 71

3.2.3 DNA extraction 74
3.2.4 Methyl-seq library preparation and Illumina sequencing 75
3.2.5 Methyl-seq read alignment and analysis 75
3.2.6 Methyl-seq data statistical analysis 76
3.2.7 Methylation level of different genomic features 76
3.2.8 Methylation analysis of repetitive elements 77
3.2.9 Correlation analysis of DNA methylation and gene expression 78
3.2.10 DNA methylation and ZGA gene expression analysis 78
3.2.11 Statistical analysis 79
3.3 Results and discussion 80
3.3.1 Summary of sequencing data 82
3.3.2 DNA methylation analysis 89
3.3.2.1 DNA methylation in different genomic regions 89
3.3.2.2 Gene body methylation 92
3.3.2.3 Methylation of repetitive elements 95
3.3.3 Integrative analysis of EM-seq and RNA-seq data 100
3.3.3.1 Correlation between gene body methylation and gene expression 100
3.3.3.2 Methylation of zygotically activated genes 106
3.4 Conclusion 111
Chapter 4 The effects of inhibiting DNA methylation in the early development of *Parhyale hawaiensis* 113
4.1 Introduction 114
4.1.1 DNA methylation inhibitors 114
4.1.2 Effects of decitabine on gene expression 117
4.1.3 DNA methylation inhibitors utilised in animal research 118
4.1.4 RNA sequencing technology 120
4.1.5 Aim of this chapter 123
4.2 Materials and methods 124
4.2.1 *Parhyale* culture and embryo collection 124

4.2.2 Phylogenetic analysis of DNMT1 from different species 124
4.2.3 Chemicals 124
4.2.4 Decitabine toxicity range-finding test 124
4.2.5 Test of decitabine exposure time window 125
4.2.6 Live-cell imaging with Hoechst dye 126
4.2.7 Decitabine exposure experiment 126
4.2.8 RNA extraction and Total RNA sequencing 126
4.2.9 RNA-seq read alignment and analysis 127
4.2.10 Expression of repetitive elements 128
4.2.11 Statistical analysis 129
4.3 Results and discussion 130
4.3.1 Selection of DNA methylation inhibitor 130
4.3.2 Decitabine exposure 131
4.3.3 Summary of sequencing data 134
4.3.4 Correlation with gene expression of wild-type embryos 136
4.3.5 Gene expression analysis 139
4.3.5.1 Differential expression analysis 139
4.3.5.2 DNA methylation of differentially expressed genes 141
4.3.5.3 Functional classification of DEGs 145
4.3.5.4 Correlation between DEGs and ZGA genes 150
4.3.6 Repeat expression analysis 154
4.3.6.1 Expressed repeats 154
4.3.6.2 Differential expression analysis 157
4.3.6.3 Methylation of differentially expressed repeats 159
4.3.7 Relationship between gene expression and repeat expression 161
4.3.8 Validation of demethylation effect induced by 5AzaD 167
4.4 Conclusion 169
Chapter 5 Thesis discussion and future directions 171
5.1 Discussion 172

5.1.1 Genome-wide methylation patterns in *Parhyale* 172
5.1.2 The relationship between DNA methylation and gene expression in *Parhyale* 177
5.1.3 The role of DNA methylation in the early development of *Parhyale* 178
5.1.4 DNA methylation of zygotically activated genes 181
5.1.5 DNA methylation inhibitor 182
5.1.6 The relationship between gene expression and repeat expression 182
5.1.7 Conclusion 184
5.2 Future directions 186
Reference 189

Abbreviations

5AzaD	Decitabine (5-Aza-2′-deoxycytidine)
5caC	5-carboxylcytosine
5hmC	5-hydroxymethylcytosines
5mC	5-methylcytosine
BS-seq	Bisulfite sequencing
CGIs	CpG islands
DEGs	Differentially expressed genes
DERs	Differentially expressed repeats
DMSO	Dimethyl Sulfoxide
DNMTs	DNA methyltransferases
EM-seq	Enzymatic methyl-seq
FASW	Filtered artificial seawater
GBM	Gene body methylation
LINEs	Long interspersed nuclear elements
lncRNAs	Long non-coding RNAs
LTRs	LTR retrotransposons
mCpGs	Methylated CpGs
MBD	Methyl CpG binding domain
MZT	Maternal to zygotic transition
ncRNA	Non-coding RNA
Parhyale	*Parhyale hawaiensis*
RC	Rolling-circle transposons
RR	RepeatModeler2 + RepeatMasker

rRNA	Ribosomal RNA
SINEs	Short interspersed nuclear elements
SMRT	Single molecule real-time
TEs	Transposable elements
TPM	Transcript per million
TSS	Transcription start site
UTR	Untranslated region
WGBS	Whole genome bisulfite sequencing
ZGA	Zygotic genome activation

Chapter 1

General introduction

1.1 Introduction

1.1.1 Epigenetics

Epigenetics is an intriguing and rapidly developing field that seeks to understand the expression and regulation of genetic information within cells. This field investigates heritable modifications in gene expression and phenotype that occur independently of changes in the DNA sequence (Holliday & Pugh, 1975; Sharma et al., 2010; Wolffe & Guschin, 2000). Epigenetic mechanisms act as dynamic genomic regulators, enabling the genome to adjust to external stimuli and respond to environmental changes. Additionally, these mechanisms facilitate the stable transmission of gene activity patterns from one generation of cells to the next, thereby ensuring the preservation of cell identity and function (Bollati & Baccarelli, 2010; Fisher, 2002).

Epigenetic information can be inherited through chromosomes in two primary forms. The first process is DNA methylation, which involves the addition of a methyl group to DNA molecules. The second form is modifications to chromatin proteins, typically caused by alterations to histone tails (Bernstein et al., 2007; Goldberg et al., 2007). Recently, non-coding RNAs and RNA editing have also been identified as crucial components of the epigenetic landscape. Both of these epigenetic marks can influence gene expression by changing chromatin structure and accessibility to transcriptional machinery. These modifications can be inherited through cell division to maintain stable gene expression patterns (Bernstein et al., 2007). Changes in epigenetic information affect the expression of genetic material during various developmental stages, across different body tissues, and in pathological states. Consequently, the study of epigenetics is essential for understanding the functionality of the genome in both normal and diseased states, and for developing innovative strategies for detecting and treating diseases (A. Bird, 2002; Goll & Bestor, 2005).

1.1.2 DNA methylation

DNA methylation is a process that occurs most frequently on the cytosine residue in eukaryotic genomes. It involves the attachment of a methyl group derived from the cofactor S-adenosyl-L-methionine to the fifth carbon of cytosine, resulting in the formation of 5-methylcytosine (5mC) (Moore et al., 2013). This modification was first discovered in 1948 and has been suggested to be involved in X-chromosome inactivation and gene regulation during development (Holliday & Pugh, 1975; Riggs, 1975). Recent sequencing studies and genetic approaches have significantly improved our understanding of the mechanisms and functions of DNA methylation.

DNA methylation can arise in three different contexts, CG, CHG, and CHH (H represents the nucleotides A, C, or T), which are present in all vertebrates, flowering plants, some invertebrates, fungi, and bacteria (Feng et al., 2010; Goll & Bestor, 2005). In animals, DNA methylation primarily takes place at CpG dinucleotides and plays a critical role in various biological processes, including development, transposon silencing, gene regulation, chromatin remodelling, genomic imprinting, and disease-related processes that involve changes in DNA methylation (Feng et al., 2010; Jones & Laird, 1999; Portela & Esteller, 2010).

In recent years, the study of DNA methylation has mainly focused on a limited number of species, including humans, mice, some insects, plants such as *Arabidopsis thaliana*, and fungi like *Neurospora crassa* (Schmitz et al., 2019; Standage et al., 2016; Tamaru & Selker, 2003; Zilberman et al., 2006). The functions of DNA methyltransferases (DNMTs), the patterns of methylomes, and the functional significance of methylation vary among eukaryotic genomes. To illustrate this point, I summarise some of the current understanding of the distribution and function of DNA methylation, concentrating on methylation in animal taxa and discussing similarities and differences between vertebrates and invertebrates regarding DNA methylation.

1.1.3 DNA methylation toolkits

In animals, three categories of enzymes and protein domains are responsible for the addition, removal, and recognition of DNA methylation: writers, erasers, and readers. Writers are responsible for adding methyl groups to cytosines, erasers modify or remove them, and readers recognise and respond to methyl groups, which can result in changes in gene expression (Biswas & Rao, 2018).

1.1.3.1 The writers of DNA methylation

The transfer of methyl groups to cytosine relies on DNA methyltransferases, a conserved group of proteins. This process occurs in three phases: establishment, maintenance, and demethylation, each catalysed by different members of the DNMT family and associated machinery (Moore et al., 2013). The DNMT family can be divided into two groups: *de novo* DNMTs, responsible for establishing new methylation on previously unmethylated cytosine sites (e.g., DNMT3), and maintenance DNMTs, that ensure the methylation mark is preserved during DNA replication to maintain the existing methylation pattern (e.g., DNMT1) (Biswas & Rao, 2018). Figure 1 provides a visual representation of this division.

The DNMT family is commonly divided into three subfamilies: DNMT1, DNMT2, and DNMT3(A/B/C/L). The most abundant DNMT in cells is DNMT1, which, in collaboration with the co-factor UHRF1, identifies and methylates newly synthesized, non-methylated DNA strands, thereby preserving CpG methylation patterns during cell division (Lyko, 2018; Robertson et al., 1999). In addition to its role in methylation maintenance, DNMT1 is also involved in repairing DNA methylation, further preserving the pattern of DNA methylation that is essential for appropriate cellular differentiation (Moore et al., 2013). In some invertebrates, DNMT1 is also capable of independently introducing genome-wide methylation (Lewis et al., 2020). Nonetheless, several studies have shown that the loss of this maintenance methyltransferase can result in loss of the methylation pattern throughout the vertebrate genome (Anderson et al., 2009; Li et al., 1992).

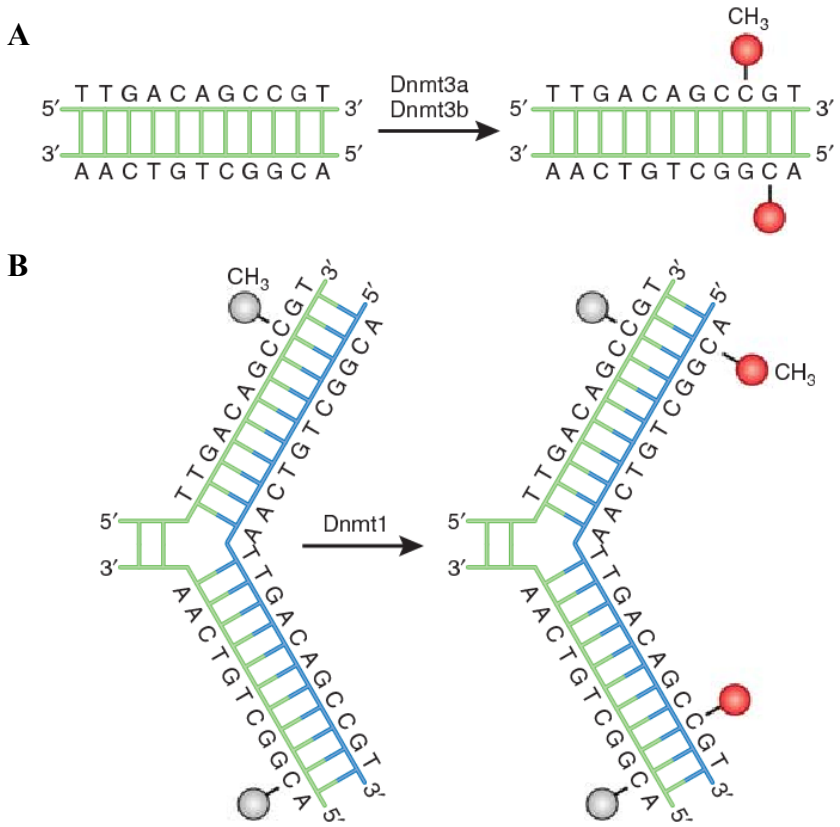


Figure 1. A conserved group of proteins DNMTs, that transfer a methyl-group to the fifth carbon of the cytosine to form 5-methylcytosine (5mC). **(A)** DNMT3 is known as *de novo* DNMT and has two homologues, DNMT3a and DNMT3b, which cooperate to establish new methylation to unmodified DNA. **(B)** DNMT1 is the maintenance DNMT, which maintain methylation pattern during DNA replication. Adapted from Moore et al, (2013).

In mammals and some invertebrates, the DNMT3 family of DNA methyltransferases is responsible for establishing genomic methylation patterns during development and reproduction (Xie et al., 1999). This family has three members: DNMT3A, DNMT3B, and DNMT3L. Unlike DNMT1, DNMT3 (DNMT3A, DNMT3B) is highly expressed in embryonic cells and methylate both hemi-methylated and unmethylated CpG sites on DNA (Yokochi & Robertson, 2002). DNMT3L is a catalytically inactive family member that functions as a *de novo* methylation regulatory factor (Gowher et al., 2005; Hata et al., 2002). Other species-specific DNMT3 enzymes have also been identified, such as DNMT3C in mice and DNMT4-8 in zebrafish (Barau et al., 2016; Shimoda et al., 2005). DNMT2, despite sharing significant sequence similarity with other DNMTs, exhibits little evidence of

cytosine methylation activity and is not considered a true DNA methyltransferase. In mammals, DNMT2 is responsible for methylating the 38th cytosine residue in the anticodon loop of tRNAs, making it the only methyltransferase responsible for RNA methylation (Goll et al., 2006; Kaiser et al., 2017).

1.1.3.2 The readers of DNA methylation

Methyl-CpG-binding proteins are specialized proteins that recognise methylated cytosines with high affinity and inhibit transcription factor binding, resulting in transcriptional repression. There are three major families of methyl-CpG-binding proteins: the methyl CpG binding domain (MBD) family, the SRA family, and the zinc finger family (Buck-Koehntop & Defossez, 2013).

The MBD proteins contain a highly conserved methyl-CpG-binding domain that allows them to interpret and respond to methylation patterns (Fatemi & Wade, 2006). The MBD family consists of seven proteins, including MeCP2 and MBDs 1-6, which are capable of recognising methylated cytosines and have a role in transcriptional repression. Among these, MBD4 is primarily involved in DNA repair (J. C. Hansen et al., 2010; Jørgensen et al., 2004; Laget et al., 2014; Lai & Wade, 2011; Menafra & Stunnenberg, 2014).

The UHRF protein family, which contains the SRA domain, interacts with DNMT1 and directs it to hemi-methylated DNA, thereby ensuring the maintenance of DNA methylation during replication (Avvakumov et al., 2008). Zinc finger domain-containing proteins, such as Kaiso, ZBTB33, ZBTB4, ZBTB38, and ZFP57, can also recognise methylated DNA regions and inhibit transcription. Kaiso, in particular, has been discovered to bind to methylated CpGs (Blattler et al., 2013; Filion et al., 2006; Quenneville et al., 2011; Sasai et al., 2010).

1.1.3.3 The erasers of DNA methylation

The epigenetic reprogramming of genes involves DNA demethylation, which can occur through passive or active processes. The passive process occurs during DNA replication in the absence of DNMT1, whereas the active process involves the enzymatic removal of the methyl group from methylated cytosines (Sadakierska-Chudy et al., 2015). The TET family of proteins, comprising TET1, TET2 and TET3, mediates the active DNA demethylation pathway. These proteins oxidize 5-methylcytosine to 5-hydroxymethylcytosine, which is further oxidized to 5-formylcytosine and 5-carboxylcytosine, ultimately leading to DNA demethylation (Ito et al., 2011; Kohli & Zhang, 2013; Wu & Zhang, 2017).

TET1 is crucial for the erasure of genomic imprinting, while TET3 is involved in the removal of paternal DNA methylation during mammalian reprogramming processes (Gu et al., 2011; Yamaguchi et al., 2013). TET2 acts as a tumour suppressor gene with pleiotropic functions during haematopoiesis (Solary et al., 2014). The presence of DNA demethylation implies that DNA methylation is not a permanent modification in mammalian cells and requires continuous regulation (Bhutani et al., 2011).

1.1.4 The DNA methylation landscape in animals

Cytosine methylation is a critical epigenetic mechanism involved in numerous biological processes in animals. However, the patterns, distribution, and functions of DNA methylation vary significantly between different animal taxa, particularly between vertebrates and invertebrates (Schübeler, 2015).

1.1.4.1 Vertebrates

Studies of various vertebrate species, including humans, chimpanzees, mice, and zebrafish, have revealed that their genomes are globally methylated, with CpG

methylation occurring predominantly in somatic tissues and ranging from 60% to 80% (Figure 2A) (Cedar & Bergman, 2009; Feng et al., 2010; Zeng et al., 2012). Non-CpG methylation has also been reported in pluripotent cells, female reproductive cells, and neurons, though it is less prevalent (J. Guo et al., 2014; He & Ecker, 2015; Ziller et al., 2011). In mammals, more than 80% of CpG sites are methylated, primarily in repetitive genomic regions such as satellite DNAs and transposable elements, non-repetitive intergenic DNA and exons (Li & Zhang, 2014; Yoder et al., 1997) (Figure 2B). DNA methylation of transposable elements in mammals serves as a protective mechanism against the potentially harmful consequences of transposon activity (Barau et al., 2016; Yoder et al., 1997).

CpG islands (CGIs), characterised by their high CpG density (>50%), are predominantly unmethylated in germ cells, early embryos, and most somatic tissues, representing a significant deviation from the typical methylation patterns seen in vertebrates (Bird, 1986). These regions are predominantly unmethylated, and the cytosines in CpG islands deaminate to uracil rather than thymine, allowing the mismatch repair system to precisely identify and correct deamination-induced mismatches (Coloundre et al., 1978). The high CpG density in CGIs is functionally significant, as roughly 60% of human gene promoters are associated with CpG islands (Bernstein et al., 2007; Li & Zhang, 2014). Actively transcribed genes that exhibit high levels of the H3K4me modification at lysine 4 are primarily associated with these promoters (Schübeler et al., 2004).

In addition to the consistent DNA methylation patterns, a small proportion of CpGs located outside CpG islands exhibit dynamic methylation patterns across various tissues and developmental stages. These CpGs are situated at tissue-specific differentially methylated regions and may overlap with gene regulatory sequences, such as enhancers that serve as binding sites for tissue-specific transcription factors (Schultz et al., 2015; Ziller et al., 2013). Furthermore, a large-scale bisulfite sequencing study of human chromosomes discovered varied methylation levels of individual CpG sites, which were

either hypomethylated or hypermethylated in cells derived from different tissues (Eckhardt et al., 2006). Similar variations in tissue-specific methylation have also been observed in many other vertebrates (Klughammer et al., 2023). Over evolutionary time, genome hypermethylation causes CpG depletion in vertebrate genomes, as methylated cytosines are prone to deamination, resulting in a C to T transition and a decrease in CpG content. However, sparsely methylated promoters are safeguarded against this methylated cytosine mutation, resulting in higher CpG density (Cooper & Krawczak, 1989; de Mendoza, Hatleberg, et al., 2019; Holliday & Grigg, 1993).

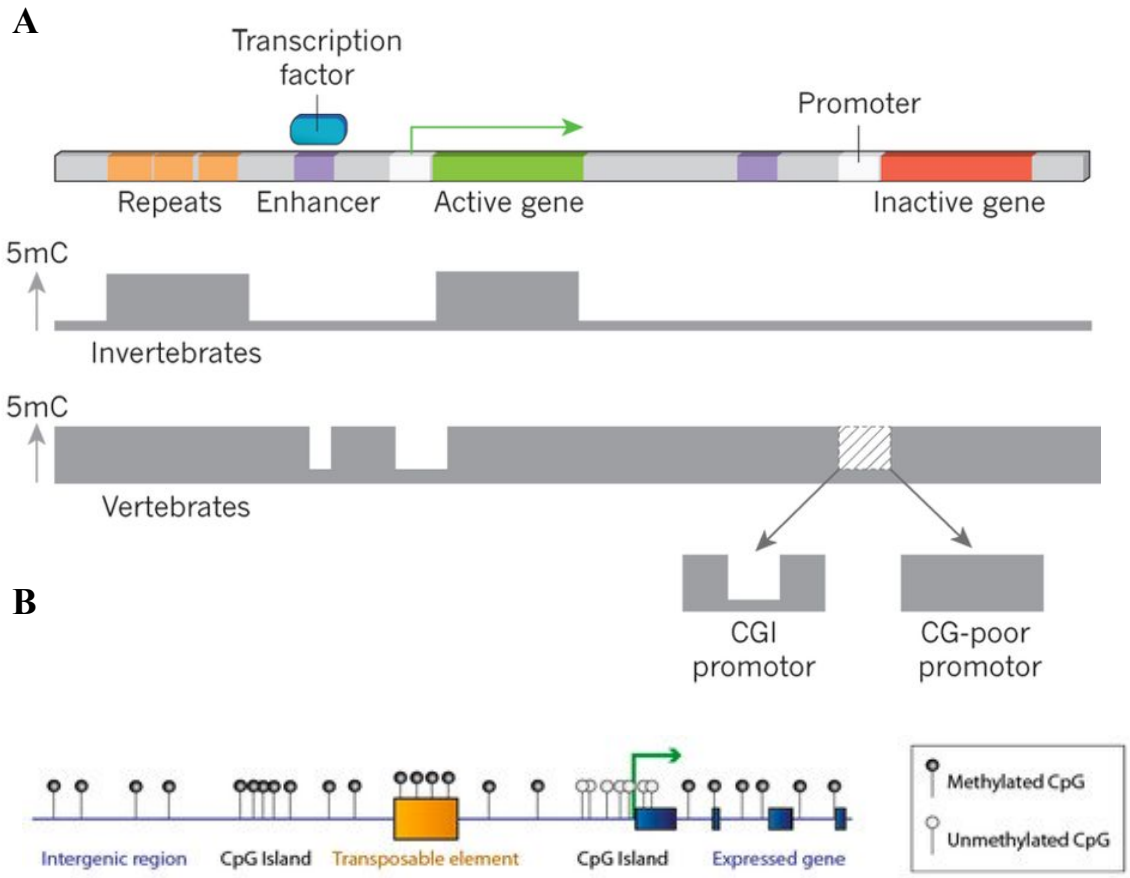


Figure 2. Methylation patterns in vertebrates and invertebrates. **(A)** A representative genomic region includes examples of an active and an inactive gene with regulatory regions, promoter, and enhancer. The height of the bar indicates the relative proportion of DNA methylation in each region. Adapted from Schübeler (2015) **(B)** DNA methylation pattern in mammals. 5mC: Cytosine methylation. Adapted from Laisné et al, (2018) .

Despite the conservation of high global methylation in vertebrates, differences in methylation patterns among species persist. For instance, birds exhibit lower genome methylation (53%) compared to mammals (70%). In addition, mice possess a higher proportion of unmethylated promoter CGIs, indicating that CpG-rich sequences in their genome are more protected. Nonetheless, a comprehensive examination of the conservation of DNA methylation patterns and functions among vertebrates has revealed numerous conserved properties of DNA methylation in gene and transposon regulation. (Adhami et al., 2022).

1.1.4.2 Invertebrates

In contrast to vertebrates, invertebrates display significant differences in their genome-wide DNA methylation patterns. Invertebrates exhibit a mosaic-type methylation where only specific genomic elements such as repetitive elements and some protein-coding genes are methylated (Figure 2A) (Feng et al., 2010; Suzuki & Bird, 2008; Zemach et al., 2010). DNA methylation in invertebrates is also highly variable, with some phylogenetically diverse invertebrate species, such as *Drosophila melanogaster* (insect), *Oikopleura dioica* (tunicate), *Caenorhabditis elegans* (nematode), and *Schmidtea Mediterranean* (planarian flatworm) likely lacking DNA methylation, which has been lost several times during metazoan evolution (Bewick et al., 2017; Feng et al., 2010). However, DNA methylation has been observed in other invertebrates such as *Trichinella spiralis* (nematode), *Crassostrea gigas* (oyster), *Chlamys farreri* (Japanese scallop), *Donax trunculus* (salt-water clam), and *Daphnia magna* (water flea) (Gao et al., 2012; Gavery & Roberts, 2013; Vandegehuchte et al., 2009). Data from various sources indicate that both CpG and non-CpG methylation levels in most invertebrates are typically much lower than in vertebrates, especially in insects (Bewick et al., 2017; Feng et al., 2010; Lewis et al., 2020; Zemach et al., 2010) (Table 1). Interestingly, a study has shown that the differences in genome-wide methylation patterns between vertebrates and invertebrates are not

distinct but rather gradual, indicating a continuum of variations. This suggests some degree of overlap between the two categories of species (Klughammer et al., 2023).

Studies have proposed that the evolutionary process of DNA methylation reprogramming played a pivotal role in the transition from invertebrates to vertebrates (Xu et al., 2019; Zhang et al., 2024). Vertebrate genomes undergo global methylation to facilitate the complex regulation of development, reproduction, and the emergence of novel genes. Additionally, it has been posited that genome duplication was a significant driver in vertebrate evolution (Lemons & McGinnis, 2006). This duplication event resulted in the replication of HOX genes and the emergence of numerous other new genes (S. Chen et al., 2013; Lemons & McGinnis, 2006). Unlike invertebrates, which typically maintain an unmethylated status and possess only one cluster of HOX genes, vertebrates harbour multiple sets of HOX genes. This multiplicity allows vertebrates to develop more complex body plans (Xu et al., 2019). Moreover, the presence of multiple HOX gene clusters in vertebrates exhibits dynamic patterns of DNA methylation during embryogenesis. Consequently, DNA methylation is implicated in the regulation of spatiotemporal expression of duplicated HOX genes, thereby facilitating the utilisation of duplicated HOX gene clusters vertebrates (Xu et al., 2019).

A fascinating discovery has been made regarding a sponge species, the Demosponge *Amphimedon queenslandica*, which exhibits DNA methylation levels comparable to those found in vertebrates. *Amphimedon* has a global CpG methylation level of 80%, similar to the pattern observed in vertebrates, with widespread hypermethylation throughout the genome, except for promoters (de Mendoza, Hatleberg, et al., 2019). Additionally, *Amphimedon* exhibits an enrichment of CpGs near the transcriptional start sites, a feature commonly found in vertebrates. Meanwhile, the depletion of CpG content, previously known to be a vertebrate characteristic, is also evident in *Amphimedon* (de Mendoza, Hatleberg, et al., 2019; Holliday & Grigg, 1993). The unmethylated regions in

Amphimedon largely coincide with promoters and contain motifs that match several methyl-sensitive transcription factors in vertebrates (de Mendoza, Hatleberg, et al., 2019).

Table 1. Genome-wide DNA methylation levels of selected species. Adapted from Feng et al, (2010); Zemach et al, (2010).

Species	Genome-wide methylation level
Apis mellifera	0.51%
Tribolium castaneum	0.11%
Bombys mori	0.71%
Drosophila melangaster	0.12%
Nematostella vectensis	9.40%
Ciona intestinalis	21.60%
Mus musculus	74.20%
Danio rerio	80.30%

However, the evolutionary significance of the vertebrate-like level of DNA methylation observed in sponges is difficult to interpret. The hypermethylation observed in vertebrates may be a result of whole-genome duplication to compensate for chromosome imbalance, but this process did not occur in *Amphimedon* (de Mendoza, Hatleberg, et al., 2019). Additionally, *Amphimedon* has a compact genome with a relatively high level of repetitive content. Hypermethylation might have evolved as a defensive mechanism against the expansion of transposons. However, other species with similar genome features do not exhibit hypermethylation (X. Wang et al., 2014; G. Zhang et al., 2012). Understanding how hypermethylation originated and evolved in the *Amphimedon* genome from the last common ancestor remains unclear. Given that the molecular machinery involved in methylation processes is conserved across animal taxa, it's improbable that observed

regulatory similarities between *Amphimedon* and vertebrates arose through convergent evolution. It's suggested that the evolution of hypermethylation in vertebrates and *Amphimedon* represents independent transitions from a "mosaic" methylation state to a hypermethylated state, leading to comparable methylation patterns in their genomes (de Mendoza, Hatleberg, et al., 2019).

The role of DNA methylation in both vertebrates and invertebrates remains incompletely understood. While the evolutionary costs linked to hypermethylation, such as heightened mutability of CpG sites and off-target damage, suggest that maintaining the hypermethylated state may have unknown evolutionary rationales and purposes (Rošić et al., 2018). it is speculated that hypermethylation might have initially emerged as a response to specific threats, like viral or transposon invasions, in ancestral species. However, it could have subsequently acquired new functions and persisted in descendant species. Consequently, speculating on the origins and evolution of hypermethylation is challenging. Nonetheless, it's likely that the evolution of hypermethylation is a complex process, shaped by various evolutionary pressures and potentially exhibiting lineage-specific patterns (de Mendoza, Hatleberg, et al., 2019).

1.1.5 Conservation of methylation patterns

The analysis of DNMTs in various vertebrates, insects, and arthropods with low levels of DNA methylation has shown that all three DNMTs were present in both the ancestor of vertebrates and invertebrates (Goll & Bestor, 2005; Kao et al., 2016; Lewis et al., 2020) (Figure 3). Mammals have all paralogs of the DNMT1 and DNMT3 families, while numerous invertebrates lack some DNMT homologues, which is correlated with the loss of DNA methylation (Lyko & Maleszka, 2011; Rošić et al., 2018). Insects show a diverse pattern of DNMT maintenance and loss, with some eusocial species maintaining a complete set of DNMTs, while other taxa within *Coleoptera*, *Hemiptera*, *Lepidoptera*, and

Hymenoptera appear to have lost either DNMT1 or DNMT3 (Bewick et al., 2017; Richards et al., 2008; Standage et al., 2016; Xiang et al., 2010).

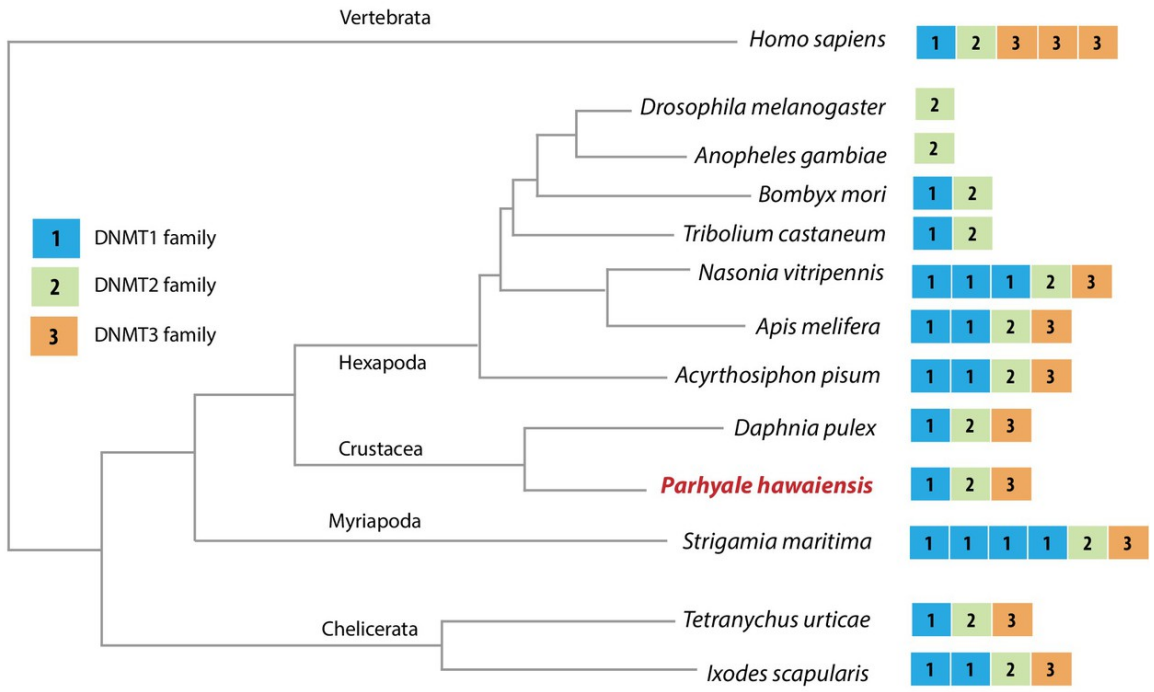


Figure 3. The phylogenetic distribution of DNMTs in vertebrates, insects, and other arthropods. Mammals possess proteins from DNMT1, DNMT2 and all three paralogs of the DNMT3 family, whereas no invertebrates possess all the DNMT homologues found in mammals. Adapted from Kao et al, (2016).

Studies on arthropods indicate that Diptera species lacking DNMT1 and DNMT3 genes have depleted methylation (Bewick et al., 2017). It is generally accepted that proper methylation patterns require both de novo (DNMT3) and maintenance (DNMT1) methyltransferases. However, cytosine methylation has been detected in some insects whose genomes contain only the DNMT1 gene (Bewick et al., 2017; Provataris et al., 2018; Xiang et al., 2013). Additionally, DNMT3 is not always accompanied by DNMT1 in invertebrates, whereas the absence of DNMT1 is associated with the absence of DNA methylation (Lewis et al., 2020). Another study of DNMT1 in T.castaneum and O. fasciatus revealed that DNMT1 suppression negatively affects reproduction and embryonic development (Bewick, Sanchez, et al., 2019; Schulz et al., 2018). These

findings suggest that the evolution of methylation depletion is associated with the simultaneous loss of DNMT1 and DNMT3. In insects, DNMT1 may have evolved to have both de novo and maintenance activity, or it may have non-catalytic functions unrelated to DNA methylation (Espada, 2012; Schulz et al., 2018).

1.1.6 Cellular function of DNA methylation

DNA methylation is a common phenomenon in the genomes of vertebrates; however, it is important to note that the function of DNA methylation may differ depending on the specific genetic sequences within various regions of the genome. Additionally, although DNA methylation is often associated with transcriptional silencing, the mechanisms involved in this process may not be the same at gene promoters, gene bodies, or repetitive elements.

1.1.6.1 Methylation at CGI promoters

Although CpG sites in mammalian genome are largely methylated, CpG islands often exhibit reduced or absent DNA methylation (Hon et al., 2013; Stadler et al., 2011; Ziller et al., 2013). Over 60% of the CGIs in the mammalian genome are located within promoters of housekeeping genes and a few developmentally related genes (Gardiner-Garden et al., 1987; Ku et al., 2008; Larsen et al., 1992). Methylation of these CGI promoter regions impact various biological processes, including DNA repair, genome imprinting, development, pluripotency of germ cells, and cell fate determination (Schübeler, 2015).

It's widely accepted that DNA methylation on CGIs within the promoter region is associated with gene repression (Bird, 1986; De Smet et al., 1999; Domcke et al., 2015; Futscher et al., 2002; Kaluscha et al., 2022; Kribelbauer et al., 2017; Lim & Maher, 2010; Song et al., 2005). However, evidence suggests that increased DNA methylation in promoter regions does not always lead to gene repression, although only a limited

number of *in vivo* studies have demonstrated this phenomenon (Domcke et al., 2015; Kribelbauer et al., 2017; Yin et al., 2017).

Several potential mechanisms have been proposed to explain how DNA methylation influences gene expression. Firstly, DNA methylation may directly affect gene expression by altering the binding affinity of transcription factors to DNA. This idea received initial support from *in vitro* biochemical assays (Domcke et al., 2015; Kribelbauer et al., 2017; Yin et al., 2017), and recent evidence *in cellulo* has further supported this notion (Kaluscha et al., 2022). Secondly, DNA methylation may act indirectly through proteins containing methyl-CpG-binding domains. These proteins can impede the access of transcription factors or serve as anchoring points for other effector proteins to regulate gene expression. For example, methyl-CpG domain-binding proteins can recognise the methylated form of CpG and recruit histone deacetylases, thereby modifying chromatin structure and contributing to methylation-mediated repression (Klose & Bird, 2006; Schübeler, 2015; Siegfried & Simon, 2010). Furthermore, previous studies have shown that DNA methylation could alter the flexibility of DNA, thereby affecting its potential to form nucleosomes with histones. This suggests that DNA methylation could also physically modulate chromatin configuration, even in the absence of protein machinery to read methylation signals (Buitrago et al., 2021; Pérez et al., 2012; Portella et al., 2013).

There exist three main groups of genes that require DNA methylation at their CGI promoters to maintain permanent silencing in somatic tissues. These groups include genes located on the inactive X chromosome, imprinted genes, and genes specific to the germline (Edwards & Ferguson-Smith, 2007; Greenberg & Bourc'his, 2019; Reik, 2007).

During early female mammalian development, X-chromosome inactivation occurs to silence transcription on one of the pair of X chromosomes, achieving dosage compensation with males who have a single X chromosome in addition to the Y chromosome. The inactive X chromosome is generally hypermethylated compared to the

active X chromosome (Prantera & Ferraro, 1990; Sharp et al., 2011; Viegas-Pequignot et al., 1988). Furthermore, methylation of X-linked CGI promoters has consistently been linked to the X-chromosome inactivation status, and individual gene studies have demonstrated that loss of DNA methylation can reactivate X-linked genes (Anderson & Brown, 2002; Brockdorff & Turner, 2015; Carrel et al., 1996; Cotton et al., 2011; R. S. Hansen & Gartler, 1990).

In mammals, genomic imprinting modifies a subset of genes, resulting in a distinctive pattern of parental-specific monoallelic expression. During early embryogenesis, differential methylation of the paternal and maternal genome is a crucial regulator of genomic imprinting (Barlow & Bartolomei, 2014). There is conclusive evidence demonstrating the essential role of DNA methylation in regulating imprinted gene expression. Knocking out either DNMT1 or DNMT3 gene results in reduced DNA methylation, consequently causing the abnormal expression of imprinted genes (Hata et al., 2002; Hirasawa et al., 2008; Kaneda et al., 2004; Li et al., 1993; Weaver et al., 2010). Surprisingly, the function of DNA methylation in regulating genomic imprints has also been observed in some social insects (Bonasio et al., 2012; Kronforst et al., 2008).

DNA methylation in vertebrates also plays a crucial role in silencing germline-specific genes, a function conserved across species (Adhami et al., 2022). In the early stages of differentiation, these genes are specifically regulated by DNA methylation via DNMT3B, which is activated by sequence-specific transcriptional repressors (De Smet & Loriot, 2012; Smith & Meissner, 2013). The establishment of correct DNA methylation patterns in the germline is essential for proper post-fertilization development, and failure to do so can have severe negative effects (Stewart et al., 2016).

1.1.6.2 Methylation at gene bodies

Research has shown that gene body methylation (GBM) is the primary form of DNA methylation in invertebrates, which is markedly distinct from the methylation patterns

predominant in mammalian genomes (Feng et al., 2010; Suzuki et al., 2007; Xu et al., 2019; Zemach et al., 2010). GBM is considered an ancestral form of DNA methylation found across various taxa, including invertebrates, vertebrates, and plants (Feng et al., 2010; Suzuki et al., 2007). In animals, GBM is associated with active transcription, in contrast to promoter methylation, which is commonly linked with the repression of gene expression (Adhami et al., 2022; Sarda et al., 2012; Zemach et al., 2010). Interestingly, in both vertebrates and invertebrates, methylation is more abundant in exons than in introns (Feng et al., 2010; Field, 2000; Lewis et al., 2020; Lister et al., 2009; Mandrioli & Borsatti, 2006; Ying et al., 2022). Throughout the process of invertebrate evolution, heavily methylated genes exhibit a higher degree of sequence conservation compared to genes with low or no methylation. Additionally, these heavily methylated genes are predominantly associated with housekeeping cellular functions, including translation, ribosome biogenesis, RNA splicing, and protein localization (Elango et al., 2009; B. G. Hunt et al., 2010; Sarda et al., 2012).

GBM plays various roles in gene regulation in vertebrates, such as aiding transcription elongation, regulating alternative splicing, inhibiting intragenic promoters, and preventing spurious transcription initiation (Carrozza et al., 2005; Gelfman et al., 2013; Lorincz et al., 2004; Neri et al., 2017; Yang et al., 2014; Yearim et al., 2015). However, research on the functional effects of GBM in invertebrates is limited, making it challenging to determine its precise role. Some studies suggested that GBM might be associated with gene transcription and could play a role in alternative splicing in invertebrates (Elango et al., 2009; Rivière, 2014), whereas others have found no association between GBM alterations and gene expression (Bewick, Zhang, et al., 2019; Dixon & Matz, 2022). The functional consequences of GBM in most invertebrate species remain unexplored, making it difficult to determine its exact function in these organisms.

1.1.6.3 Methylation at repetitive elements

Eukaryotic genomes, particularly mammalian genomes, consist mainly of repetitive sequences, such as transposable elements (TEs) and simple repeats, which together make up over half of the mammalian genome. In contrast, the coding sequences account for less than 1.5% of the sequence content in human and mouse genomes (Frith et al., 2005; Lander et al., 2001; Mattick & Makunin, 2006; Pheasant & Mattick, 2007). Repetitive elements can be classified into five categories: interspersed repeats (transposable elements), processed pseudogenes, simple sequence repeats, segmental duplications, and tandem repeats (satellite DNA sequences). However, the latter four categories combined account for only 10% of genomic DNA (Lander et al., 2001).

Transposable elements are the most abundant category of repetitive elements in the human genome, constituting about 45% of genomic DNA and comprising both retrotransposons and DNA transposons (Lander et al., 2001). These elements act like parasites, proliferating and inserting themselves into new genomic positions, thereby disrupting genome integrity and inducing insertional mutagenesis in genes. The transposition method of TEs varies by class. Class I TEs, also known as retrotransposons, transpose via an RNA intermediate using reverse transcriptase and are further divided into long terminal repeats (LTRs) and non-LTR sequences. Class II TEs include DNA transposons and rolling-circle (RC) transposons, which transpose via a DNA intermediate facilitated by transposases (Hancks & Kazazian, 2012; Klein & O'Neill, 2018; Makałowski et al., 2019; Sotero-Caio et al., 2017; Wicker et al., 2007). Despite the high number of TEs in some species, only a few are known to be active. For example, in humans, only non-LTR elements, classified as long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), are capable of retro-transposition (Lander et al., 2001; Levin & Moran, 2011). The expression of transposable elements can trigger genomic instability and rearrangement, potentially influencing gene function by

introducing incorrect promoters to new areas of the genome (Cordaux & Batzer, 2009; Hancks & Kazazian, 2016).

DNA methylation has evolved mainly to protect the genome against transposable elements, particularly retrotransposons, which are the primary targets of DNA methylation in mammalian genomes, rather than genes themselves (Barau et al., 2016; Yoder et al., 1997). Studies have shown that depletion of DNA methylation results in the transcription of transposable elements in several vertebrates, suggesting that maintaining their silencing is one of the primary functions of DNA methylation in vertebrates (Bestor & Bourc'his, 2004; Chernyavskaya et al., 2017; Liu et al., 1994; Walsh et al., 1998; Woodcock et al., 1997).

Methylation patterns of TEs display significant variation among invertebrates. Some invertebrate species, such as oysters, sea squirts, cnidarians, centipedes, and mealybugs, exhibit TE methylation (de Mendoza, Hatleberg, et al., 2019; Keller et al., 2016; Kenny et al., 2020; Lewis et al., 2020; Wang et al., 2014; Ying et al., 2022). However, the common ancestor of arthropods had low TE methylation, with only rare methylation observed in holometabolous insects and moderate methylation in chelicerates, crustaceans, and hemimetabolous insects (Lewis et al., 2020). There was no consistent relationship between genome size and transposon content across multiple invertebrate phyla. However, a strong positive correlation was observed between the levels of DNA methylation and the abundance of transposons within the genome (Ying et al., 2022). This relationship likely represents the ancestral condition among invertebrates, indicating that DNA methylation may have evolved as a mechanism to against the potentially deleterious effects of transposon activity across a broad range of invertebrate taxa (Ying et al., 2022). Nevertheless, deviations from this pattern are common. For instance, the invertebrate deuterostomes sea urchin and sea squirt exhibit unusually high DNA methylation levels relative to their transposon contents. This discrepancy may indicate the adoption of secondary, vertebrate-like roles for the DNA methylation system

in these organisms (de Mendoza, Hatleberg, et al., 2019; Xu et al., 2019; Ying et al., 2022).

1.1.7 *Parhyale hawaiensis*

1.1.7.1 An overview of *Parhyale hawaiensis*

Parhyale hawaiensis is a species of the marine amphipod crustacean belonging to the class Malacostraca. These amphipods were originally discovered in the Hawaiian Islands in 1853 and are now widely distributed in marine environments worldwide, especially in circumtropical, intertidal, and shallow waters (Barnard, 1965; Dana, 1853; Shoemaker, 1956). *Parhyale* prefer to inhabit rocky beaches and macroalgal fauna in intertidal areas, and they are often found in large populations (Barnard, 1965; Poovachiranon et al., 1986). They can tolerate a wide range of salinity (5 to 40ppt) and temperature variations. Additionally, they play a significant ecological role in environments such as mangrove forests (Poovachiranon et al., 1986).

In *Parhyale*, sexual dimorphism is evident, with males typically being larger than females of the same age and possessing larger chelipeds on the third thoracic segment, which are used to grasp females during mating (Figure 4). Mating pairs are formed for several days before females lay their eggs, with males depositing sperm into the female's brood pouch just prior to separation (Borowsky & Borowsky, 1987). Like other peracarid crustaceans, *Parhyale* females carry their fertilized eggs in a ventral brood pouch, which can be easily accessed without harming the mother or the developing embryos (Paris et al., 2022). Reproductive maturation in *Parhyale* occurs over a period of 6-7 weeks, after which females lay eggs every two weeks (Browne et al., 2005).

Parhyale produce comparatively large embryos, which are simple to collect and image. The embryos develop for approximately 10 days and hatch as miniature versions of adults, without passing through larval stages like many other crustaceans (Browne et al., 2005).

Upon hatching, the body of *Parhyale* is around 1 mm long, which grows to over 10 mm as an adult. *Parhyale* have a generation time of approximately 2 months and adults remain reproductive for 2-3 years. These features, including the body size and life cycle, make *Parhyale* an ideal candidate for laboratory culture and molecular genetic manipulation (Browne et al., 2005).

Parhyale share a close evolutionary relationship with economically significant animals like lobsters, crabs and shrimps. However, they are much easier to maintain in a laboratory setting and require less effort than other malacostracan crustaceans. *Parhyale* can survive in various culture systems, including those with poor water quality (Gerberding et al., 2002). To maintain *Parhyale* in a lab setting, tanks filled with artificial seawater are used, which are circulated with air stones and pumps. Crushed coral is added to the tanks to buffer the pH and provide concealment places for the animals, while their diet consists of fish flakes and carrots. *Parhyale* embryos can be maintained in petri dishes with filtered artificial seawater and antibiotics.

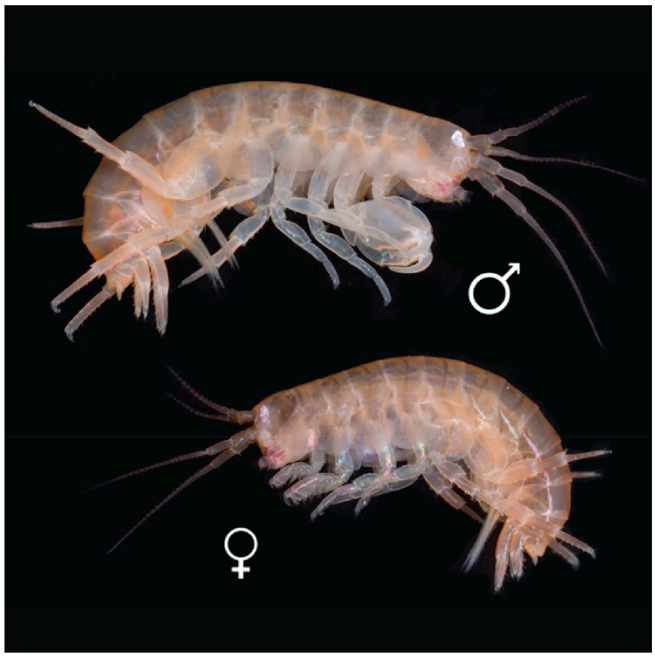


Figure 4. An adult male and female *Parhyale hawaiensis*. Adapted from Sun & Patel, (2019).

1.1.7.2 *Parhyale* as an animal model system

Due to its unique characteristics, *Parhyale* has become an attractive animal model system for studying a broad range of biological topics, including embryonic development, germ layer formation, body and limb patterning, regeneration, neurobiology, sensory systems and behaviour, circadian and tidal rhythms, lignocellulose digestion, and ecotoxicology. Several techniques have been developed over the past two decades, making it relatively easy to manipulate and study *Parhyale* (Alegretti et al., 2016; Alwes et al., 2016; Browne et al., 2005; Gerberding et al., 2002; Hunt et al., 2019; Kao et al., 2016; Liubicich et al., 2009; Price & Patel, 2008; Sun & Patel, 2019; Wittfoth et al., 2019).

Parhyale is a versatile model organism for research due to the availability of established protocols for techniques such as embryo injection, dissection, antibody staining, and *in situ* hybridization (Rehm et al., 2009a, 2009b, 2009c, 2009d). The development of genetic manipulation and genome-editing techniques has also facilitated genetic approaches for *Parhyale* research, including the generation of transgenic animals, gene knockdown using morpholino and RNAi technology, and the CRISPR-Cas9 knockout system (Bruce & Patel, 2020; Clark-Hachtel & Tomoyasu, 2020; Kao et al., 2016; Liubicich et al., 2009; Martin et al., 2016; Özhan-Kizil et al., 2009; Pavlopoulos & Averof, 2005; Vargas-Vila et al., 2010). In addition, an integrated embryogenesis staging system, a previously assembled and annotated reference genome, and diverse transcriptome data from various developmental stages and adult tissues permit the molecular study of development (Blythe et al., 2012; Browne et al., 2005; Calvo et al., 2022; Kao et al., 2016; Sun & Patel, 2019). All these techniques and resources make *Parhyale* a highly suitable model system for investigating a diverse range of topics.

1.1.7.3 Embryonic development of *Parhyale*

Parhyale embryogenesis lasts for approximately 250 hours and can be divided into 30 stages based on morphological traits. The initial cell divisions are holoblastic, and the first three cleavages take approximately eight hours to form eight-cell embryos, which consist of four macromeres and four micromeres. At this stage, each blastomere is already committed to a single germ layer and will subsequently develop into specific germ layers of the embryo, including germline, ectoderm, mesoderm, and endoderm (Gerberding et al., 2002). At around 12 hours of development, the embryo contains approximately 100 cells of equal size, and the cytoplasm of cells separates from the yolk. Cells then begin to move along the embryo's surface into two dominant clusters, one forming the ectoderm anlagen at the prospective ventral end, and the other forming a small round cluster of visceral mesoderm and germ cells, known as the "rosette". Gastrulation is initiated at approximately 20 hours, and the rosette migrates beneath the ectoderm anlagen, leading to the "germ disc" stage. The cells of the germ disc then move and proliferate, leading to the "germband" stage. The formation and elongation of the germband, along with the onset of engrailed expression indicating segmentation mechanisms, occur between 60 and 112 hours. During the fourth day of embryogenesis, the limb buds become clearly visible, while the fifth to seventh day (120-168 hours) is the period of appendage elongation and gut extension. Organogenesis is completed on day 9 (216 hours) and hatching occurs on day 10 (250 hours) (Figure 5) (Browne et al., 2005; Nestorov et al., 2013).

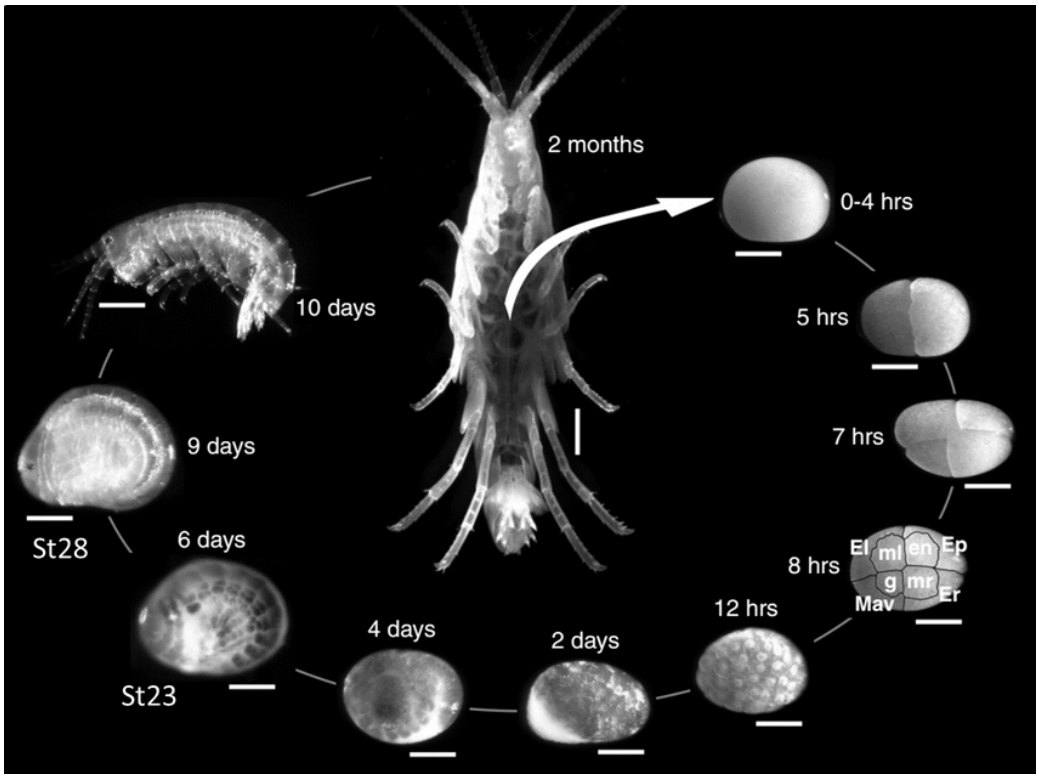


Figure 5. Life cycle of *Parhyale hawaiensis*. The embryogenesis of *Parhyale* spans approximately 10 days. Adapted from Ramos, (2017).

1.2 Thesis aim and outline

DNA methylation significantly impacts gene transcription in vertebrates. However, its function in invertebrates is variable and likely differs among invertebrate taxa. Considerable focus has recently been placed on the potential role of DNA methylation in regulating gene expression during the early stages of animal development. Previous studies have primarily focused on vertebrates, demonstrating that appropriate DNA methylation is crucial for maintaining precise temporal and spatial patterns of gene expression during early embryonic development. Despite this, little is known about DNA methylation and its mechanisms in invertebrates like *Parhyale hawaiensis*, which is an ideal model organism to study epigenetic modifications due to its methylated genome. Furthermore, characterising the DNA methylomes in these poorly studied taxa is crucial for understanding the evolution and biological significance of DNA methylation in eukaryotes. Consequently, this thesis aims to investigate the relationship between DNA methylation and gene expression, as well as the functional significance of DNA methylation in regulating the transcription of genes and transposable elements during the early development of *Parhyale*.

In Chapter 1, I provided a brief introduction to DNA methylation, including a description of the machinery involved in this process and the differences in DNA methylation patterns and functions between vertebrates and invertebrates. Furthermore, I also provided a detailed description of the model organism used in this thesis, *Parhyale hawaiensis.* The section began with a summary of *Parhyale'*s habitat and lifecycle, followed by a detailed description of genetic manipulation techniques and genomic analysis resources. Finally, I described the embryonic development of *Parhyale* in this Chapter.

In Chapter 2, I reclassified and annotated the repetitive elements in the *Parhyale* genome and compared the results to the previous classification generated by my colleague, Dr. Manuel Jara Espejo. Subsequently, I characterised the genome-wide DNA methylation

profile of four different tissues from *Parhyale hawaiensis* using WGBS and PacBio SMRT sequencing. These analyses confirmed the presence of DNA methylation in the *Parhyale* genome.

Building upon this knowledge, in Chapter 3, I analysed the methylation patterns of the *Parhyale* genome, as well as distinct genomic regions and repetitive elements, using Methyl-seq on samples from three distinct early developmental stages. In combination with RNA sequencing data, I investigated the correlation between DNA methylation of gene bodies and their expression, particularly for zygotically activated genes. These analyses revealed that DNA methylation may play a significant role in regulating gene expression during early embryonic development of *Parhyale*.

Chapter 4 delves deeper into the function of DNA methylation by utilising a methylation inhibitor in embryos before maternal to zygotic transition (MZT) to examine how this treatment affects gene expression during and after MZT and zygotic genome activation (ZGA), ultimately impacting embryonic development. By comparing the gene expression levels of treated and control groups, identifying differentially expressed genes and repeats, and comparing their corresponding methylation levels, the importance of DNA methylation in the regulation of early *Parhyale* development was elucidated.

In Chapter 5, a summary of the main results from the previous chapters is presented, along with an overview of current research progress and future directions for investigating the role of DNA methylation in early development.

Chapter 2

Genome-wide DNA methylation analysis of *Parhyale hawaiensis*

2.1 Introduction

2.1.1 DNA methylation in *Parhyale hawaiensis*

Parhyale hawaiensis is distinct from many other invertebrates in that its genome encodes genes for a full cytosine methyltransferase domain, including DNMT1, DNMT3, two orthologues of conserved methyl-CpG-binding proteins (MBD2/3, MBD4), and a single orthologue of Tet2 (Kao et al., 2016; Lewis et al., 2020). The *Parhyale* genome also exhibits moderate levels of CpG methylation, with patterns similar to those seen in vertebrates, where methylated cytosines are predominantly present in repetitive elements, near coding sequences, and within gene bodies (Kao et al., 2016; Lewis et al., 2020). These characteristics make *Parhyale* an ideal model system for investigating the dynamics of DNA methylation and exploring the functional role of DNA methylation, particularly during development.

2.1.2 Genome-wide detection of cytosine methylation by PacBio single molecule real-time sequencing

High-throughput DNA sequencing techniques that involve amplification of DNA during the sequencing process cannot directly detect modified bases in native DNA because the epigenetic marks are lost during amplification. However, bisulfite treatment of DNA samples converts unmodified cytosines to uracil, enabling the distinction between modified and unmodified cytosines on various sequencing platforms. In the era of democratized high-throughput sequencing, bisulfite sequencing has so far been the primary method for detecting cytosine methylation, despite its significant limitations. The harsh reaction conditions of bisulfite treatment can lead to DNA degradation, particularly in genomic regions enriched with unmethylated cytosines. This can result in an overestimation of global methylation and significant variations in specific genomic regions

(Grunau et al., 2001; Olova et al., 2018). The limitations imposed by bisulfite treatment make it particularly difficult to sequence long DNA molecules using this method.

Furthermore, short-read Illumina sequencing approaches also face inherent challenges in mapping reads to low-complexity regions of the genome, such as repetitive sequences. This issue is particularly problematic for DNA methylation analysis in regions with high GC content or repetitive elements, as these regions are difficult to accurately map using short-read sequencing technologies (Alkan et al., 2011; Roberts et al., 2013; Treangen & Salzberg, 2012).

Long-read sequencing technologies, such as Nanopore sequencing and PacBio single molecule real-time (SMRT) sequencing, enable the direct sequencing of native DNA without PCR amplification, addressing the limitations of short-read sequencing (Amarasinghe et al., 2020; Gouil & Keniry, 2019). Both Nanopore and SMRT sequencing methods are applicable to sequence DNA methylation due to their capacity to detect DNA base modifications, which alter polymerase kinetics in SMRT sequencing and affect electrical current signals near modified bases in Nanopore sequencing. Consequently, the primary advantage of long-read sequencing lies in its capability to identify modifications directly from native DNA reads, eliminating the need for enzymatic or chemical treatments of DNA molecules (Amarasinghe et al., 2020; Gouil & Keniry, 2019). This simplifies experimental procedures, expands the scope of detectable modifications, and reduces experimental complexity. Moreover, long-read sequencing avoids DNA degradation resulting from bisulfite conversion and minimises amplification biases, making it an attractive option for DNA methylation analysis (Gouil & Keniry, 2019).

Employing long-read sequencing to study DNA methylation offers a more thorough examination of methylation patterns by capturing longer DNA fragments. This enables the observation of multiple 5mCs in proximity and the correlation of methylation between adjacent sites, revealing single-molecule heterogeneity on a larger scale than traditional

WGBS data (Kerr et al., 2023). Additionally, long-read sequencing facilitates methylation profiling in complex regions with repetitive elements or structural variation (Chen et al., 2022). Consequently, long-read sequencing enhances our understanding of the spatial arrangement of methylated cytosines and their regulatory roles, which is crucial for studying complex genomic regions and heterochromatin where short-read sequencing might miss interactions between methylation sites.

The PacBio SMRT sequencing technique, which employs sequencing-by-synthesis, has the capability to identify modifications in DNA. To conduct SMRT sequencing, a SMRT cell is used which contains around one million wells, each with a DNA polymerase and a single DNA molecule. The sequencing process is initiated by introducing fluorescently labelled nucleotides into the well, where they are incorporated by the polymerase fixed at the bottom, resulting in a series of fluorescent pulses that indicate the sequence of the DNA molecule. Although base modifications do not alter the base-called sequence, they do affect the kinetics of the polymerase, such as the incorporation of the complementary T being delayed due to the presence of 6mA in the native DNA strand. Consequently, by analysing the differences in the duration of nucleotide incorporation between modified and unmodified bases, SMRT sequencing can identify the types and locations of DNA modifications (Clark et al., 2012; Eid et al., 2009; Flusberg et al., 2010).

Various DNA modifications, such as 5mC, m6A, and hm5C, can be detected using SMRT sequencing. The success of identifying a particular modification depends on the strength of the modification's impact on the kinetics of the polymerase. SMRT sequencing is particularly effective at detecting 4mC and 6mA, which are notably important in bacterial epigenomics. However, accurately detecting genome-wide cytosine methylation in native DNA molecules using SMRT sequencing is challenging due to the less noticeable impact of the methyl group of 5mC on synthesis kinetics, since it is not directly involved in base pairing (Clark et al., 2012; Flusberg et al., 2010). Clark et al, (2013) revealed that the detection rate of 5mCs in native DNA using SMRT sequencing is less than 5%, indicating

lower sensitivity for identifying methylated cytosines with this method. This subtler effect of 5mC on nucleotide incorporation necessitates sequencing with relatively high coverage.

Accurately identifying base modifications at a single-base, single-read level using PacBio SMRT sequencing poses challenges in reliability. Recommended minimum coverage thresholds of 25X for m6A and 250X for 5mC are advised for detection by the primary PacBio SMRT sequencing (Clark et al., 2013; PacBio, 2015). A recent advancement introduces a novel method that comprehensively detects kinetic signals and sequence context for each nucleotide within a measurement window using PacBio SMRT sequencing. This method significantly enhances 5mC detection rates and enables simultaneous genome-wide genetic and epigenetic analyses via PacBio SMRT sequencing (Tse et al., 2021). Additionally, whole-genome long-read TAPS (wglrTAPS), a novel approach combining TAPS (TET-assisted pyridine borane sequencing) with PacBio SMRT sequencing, offers promising prospects for whole-genome long-read methylome analysis. Despite relatively low sequencing depth (8.2X), wglrTAPS demonstrates strong correlation coefficients with short-read data at the individual CpG level, enabling direct long-range methylation profiling across the genome at base resolution. Furthermore, wglrTAPS holds potential for methylation profiling in complex regions with repetitive elements and allows for dissecting allele-specific methylation in imprinting genes (Chen et al., 2022). Thus, these recently developed methods, coupled with advancements such as the comprehensive detection of kinetic signals and sequence context, further enhance the capabilities of third-generation sequencing technologies for DNA methylation analysis.

2.1.3 Aim of this chapter

DNA methylation is a crucial epigenetic process that varies in genomic distribution and biological function across different taxa. Despite the paucity of information on DNA methylation in non-mammalian species, recent studies have shed light on the distinct

CpG methylation system in several arthropod species, which differs from that of mammals. These species also exhibit a wide range of DNA methylation patterns and functions.

Although DNMT1/3 have been identified in the *Parhyale* genome, a comprehensive genome-wide methylation analysis has yet to be conducted. This chapter aims to fill this gap by undertaking a thorough examination of DNA methylation across various tissues of *Parhyale*. The primary objective is to enhance our understanding of the significance and prevalence of methylation within this species. Utilising both whole-genome bisulfite sequencing and PacBio single molecule real-time sequencing data, this analysis establishes the methylome of *Parhyale*, yielding valuable insights into genome-wide methylation patterns across diverse tissues. Additionally, the investigation delves into the heterogeneity observed within the *Parhyale* methylome. Through the integration of whole genome methylation data from different tissues, this study not only aims to confirm methylation in *Parhyale* but also contributes to further exploration of methylation patterns in invertebrates.

2.2 Materials and methods

2.2.1 Whole genome bisulfite sequencing read alignment

In this experiment, three distinct samples were gathered: a) *Parhyale* embryos at different stages of development, ranging from 0 to 24 hours; b) tissues from limbs; c) body tissues excluding limbs, obtained from adult *Parhyale*. Following collection, genomic DNA extraction was conducted on each of these samples. The library construction and whole genome bisulfite sequencing (WGBS) were performed by Peter Sarkies's group in collaboration with our research team. TrimGalore (0.6.6, https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default parameters was used to remove low-quality bases and adaptor sequences from raw sequencing reads. Data quality was evaluated using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The clean reads were then mapped to the *Parhyale* Phaw5.1 genome reference using the bismark function of Bismark (0.22.3; https://www.bioinformatics.babraham.ac.uk/projects/bismark/ ; Krueger & Andrews, 2011) with the following parameters: -N 1 -L 22 --score_min L,0, -0.6. The resulting alignment was deduplicated using the deduplicate_bismark function. The Phaw5.1 genome reference represents an updated and improved assembly of the *Parhyale* genome, constructed by Dr. Manuel Jara Espejo, a postdoctoral researcher in our research group, using PacBio HiFi long reads. The bismark_methylation_extractor function was utilised to extract the methylation profile and generate a genome-wide cytosine report indicating the number of methylated and unmethylated CpG reads at each CpG site in the genome. The subsequent methylation analyses focused solely on the methylation profiles of the CpG sites.

2.2.2 WGBS data statistical analysis

The methylation value for each CpG was determined using the formula: CpG methylation value = mCpGs / (mCpGs + umCpGs) * 100, where mCpGs represents methylated CpGs and umCpG represents unmethylated CpGs. In order to accurately quantify methylation, the genome-wide methylation and subsequent analyses only included CpG sites with at least five reads coverage per site. Two methods were used to determine genome-wide DNA methylation. The first method determined the genome-wide methylation level which was calculated by dividing the number of methylated CpGs by the number of effectively covered CpGs (coverage ≥ 5). The second method determined the mean methylation value of all CpGs, calculated by averaging the CpG methylation value of all effectively covered CpGs in the genome. A methylated CpG was defined as CpG (with ≥ 5 coverage) having a methylation value of at least 20%. This criterion requires the presence of at least one unconverted C-containing read to classify a site as methylated. In different invertebrate species, the typical threshold for defining methylated cytosines ranges from 10% to 50%. The selection of this threshold aims to strike a balance between sensitivity, which ensures the detection of true positive methylated cytosines, and specificity, which aims to minimise false positive detections of methylated cytosines. (Gavery & Roberts, 2013; Olson & Roberts, 2014; Silliman et al., 2023; Trigg et al., 2022; Venkataraman et al., 2020; Wang et al., 2013).

2.2.3 PacBio single molecule real-time sequencing and methylation analysis

My colleague Dr. Salha Rawas extracted genomic DNA for PacBio SMRT sequencing from adult male and female *Parhyale* using a phenol-chloroform-based extraction protocol. Novogene company conducted the library construction and PacBio single molecule real-time sequencing. One DNA library was constructed and sequenced. Dr. Manuel Jara Espejo extracted the methylation status from the PacBio SMRT sequencing

data. To obtain methylation data, PacBio HiFi reads with kinetic information were generated using ccs (https://github.com/PacificBiosciences/ccs), which includes details on DNA polymerase activity essential for identifying methylation patterns. Primrose (https://github.com/mattoslmp/primrose) was then used to predict 5mC in these reads. The annotated HiFi reads were aligned to the reference genome to map the methylation sites. Finally, pb-CpG-tools (https://github.com/PacificBiosciences/pb-CpG-tools) was used to process the aligned reads and extract the methylation levels of each CpG site in the *Parhyale* genome.

2.2.4 Methylation level of different genomic features

To determine the methylation level of a genomic feature, I counted the total number of methylated CpGs (coverage ≥ 5 and methylation value ≥ 20%) and divided this by the total number of covered CpGs (coverage ≥ 5) in that region. Additionally, I calculated the fractional methylation of a genomic region by averaging the methylation values of all effectively covered CpGs (coverage ≥ 5) within that region. Bedtools map (Quinlan & Hall, 2010) was used to obtain the total number of CpGs, mCpGs, and fractional methylation of a specific region.

Using the Phaw 5.1 assembly as a reference, Dr. Manuel Jara Espejo conducted reference-guided transcriptome assembly using 79 RNA-seq libraries and produced a gene annotation. He also identified repetitive elements and created an annotation for these elements of the *Parhyale* genome. However, the major issue with this repeat annotation was the high proportion of unclassified repeats, making it challenging to study DNA methylation by repeat classes. Therefore, I reannotated the repetitive elements using a fully automated TE annotation pipeline named EarlGrey (Baril et al., 2024).

Using these gene annotations and repeat annotations (obtain from EarlGrey), I categorised CpGs as residing in gene-related or repetome-related regions. To profile DNA methylation patterns in the gene bodies, upstream (TSS - 2kb), downstream (TES

+ 2kb), and the exons and introns, I partitioned each region into a fixed number of windows (gene body: 100 windows, upstream/downstream: 20 windows, exon: 30 windows, intron: 80 windows) using Bedtools makewindows. I then used Bedtools map to obtain the fractional methylation of each window (the mean methylation values of all covered CpGs (coverage ≥ 5) within that window). The same method was utilised to analyse the fractional methylation distribution around TSS (TSS-2kb to TSS+4kb).

2.2.5 Gene body methylation analysis

The gene body methylation analysis excluded genes with less than three covered CpGs based on their coverage distribution (Figure 7A). Hypomethylated genes were characterised as having a fractional methylation value greater than 0 but less than 10%, while hypermethylated genes had a fractional methylation value greater than 10%.

2.3 Results and discussion

Prior to investigating the dynamics of DNA methylation of *Parhyale*, I first reclassified and annotated the repetitive elements in the *Parhyale* genome, comparing the results to the previous classification generated by my colleague Dr. Manuel Jara Espejo. Following this, I conducted a comprehensive analysis of the genome-wide DNA methylation patterns and the methylation status of various genomic features in the *Parhyale* genome. These analyses were performed using data from four distinct samples, generated through either Whole-Genome Bisulfite Sequencing (WGBS) or PacBio Single Molecule Real-Time (SMRT) sequencing.

2.3.1 Repetitive elements classification and annotation

2.3.1.1 Approaches introduction

After generating the updated *Parhyale* genome (Phaw5.1), Dr. Manuel Jara Espejo characterised its repeat landscape by constructing a *Parhyale*-specific repeat library of consensus sequences using RepeatModeler2 (Flynn et al., 2020). RepeatModeler2 identifies *de novo* repeats to create repeat library of consensus sequences without requiring a reference database. This repeat library was then employed by RepeatMasker (Smit et al., 2013), a widely used tool for masking repetitive sequences using similarity-based methods to produce a repeat annotation for the *Parhyale* genome. The combination of RepeatModeler2 followed by RepeatMasker (RR) is the currently most well-known and frequently employed TE annotation workflow (Loreto et al., 2023). However, the major issue with this annotation was the high proportion of unclassified repeats within the *Parhyale* genome, making it challenging to study DNA methylation by repeat classes.

Due to the complexity of repetitive elements, the scientific community recommends using a combination of multiple identification methods (*de novo*, similarity-based, and structure-

based approaches) for more accurate characterisation (Hoen et al., 2015; Loreto et al., 2023; Platt et al., 2016). EarlGrey, a recently developed TE annotation pipeline that combines these different identification methods, has been shown to outperform the RR workflow (Baril et al., 2024). Specifically, EarlGrey employs RepeatModeler2 for *de novo* repeat identification, followed by the "BLAST, Extract, Align, Trim" (BEAT) process by TEstrainer to generate maximum-length consensus sequences from the initial *de novo* TE consensus sequences. It then annotates repeats using RepeatMasker (similarity-based) and further analyses them with LTR_Finder (structure-based) (version 1.07; (Xu & Wang, 2007) to identify full-length LTR elements. Finally, the outputs from RepeatMasker and LTR_Finder are defragmented and merged to generate the final repeat annotation (Baril et al., 2024).

2.3.1.2 Comparison of outputs from EarlGrey and RR

Therefore, I reannotated the repetitive elements within the *Parhyale* genome using the EarlGrey workflow and compared the output with the RR workflow utilised by Dr. Manuel Jara Espejo. The comparison of the two consensus sequence libraries is presented in Table 1. EarlGrey generated a similar number of repetitive sequences as RR, but the total, average, maximum, minimum, and median lengths of the consensus library from EarlGrey are all higher than those from RR (Table 1). This improvement is due to the automated iterative BEAT process, which aims to create maximum-length sequences from the initial *de novo* repeat consensus sequences identified by RepeatModeler2. This process ensures the consensus sequences accurately reflect true repeat boundaries while minimising the inclusion of extraneous sequences (Baril et al., 2024).

Examining the distribution of sequences across different repeat classes in the consensus sequence libraries shows that EarlGrey offers significant improvements in repeat classification compared to RR. The proportion of unclassified repeats dropped from over

80% in RR to 60% in EarlGrey, with an increased number of classified consensuses across almost all repeat classes in EarlGrey (Table 2).

Table 1. Sequence metrics between EarlGrey and RR repeat libraries.

	EarlGrey	RR
Number of sequences	3,226	3,532
Total length	8,368,266	2,186,529
Average length	2,594	619
Maximum length	20,283	9,326
Minimum length	27	31
Median length	1,660	345

Table 2. Count of repeat consensus of different classes in the repeat consensus libraries.

Class	EarlGrey		RR	
	Count	%	Count	%
Unclassified	1,942	60.20	2,899	82.08
LINEs	583	18.07	279	7.90
DNA transposons	470	14.57	283	8.01
LTRs	137	4.25	42	1.19
Satellite DNAs	78	2.42	2	0.06
Rolling-circle transposons	9	0.28	6	0.17
SINEs	3	0.09	0	0
Simple repeats	4	0.12	21	0.59

Furthermore, the repeat masking results from the annotation process using both methods, which identify all repeat coordinates in the genome, indicate that repetitive elements cover approximately 60% of the entire Phaw5.1 genome (2,823,710,058 bp) (Table 3). EarlGrey reduced total masked unclassified repeats by 8.6% for the whole genome and 15% for all repeats compared to RR and classified more transposable elements (Table 3, Figure 1). The most abundant repetitive sequences in the *Parhyale* genome are LINEs, with DNA transposons following closely (Figure 1). LTRs, satellite DNAs, rolling-circle transposons, SINEs and simple repeats make up a small portion of the repeat dataset.

Table 3. The length of different repeat classes in *Parhyale* genome.

	EarlGrey length (bp)	**Percentage of *Parhyale* genome (%)**	**RR length (bp)**	**Percentage of *Parhyale* genome (%)**
All repeats	1,720,115,151	60.92	1,696,522,274	60.08
Unclassified	939,995,147	33.29	1,183,285,146	41.91
LINEs	437,173,152	15.48	351,163,777	12.44
DNA transposons	209,795,901	7.43	121,957,665	4.32
LTRs	55,662,339	1.97	27,090,032	0.96
Satellite DNAs	30,481,275	1.08	774,600	0.03
Rolling-circle transposons	18,573,220	0.66	7,377,767	0.26
Simple repeats	19,505,747	0.69	4,873,287	0.17
SINEs	8,928,370	0.32	0	0

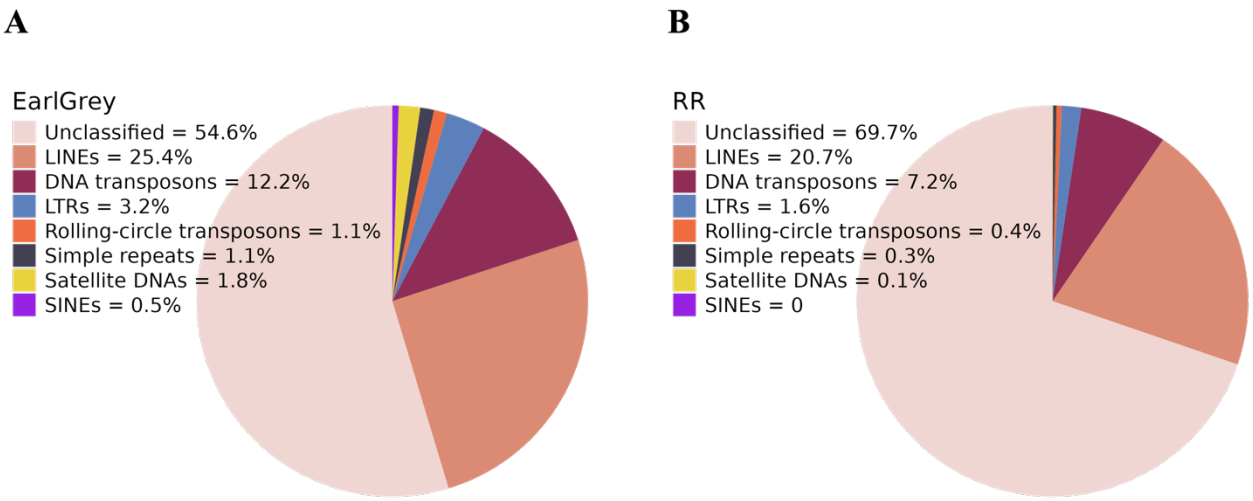


Figure 1. The proportion of different repeat classes among all repeats annotated from the genome by **(A)** EarlGrey and **(B)** RepeatModeler2 + RepeatMasker (RR). Unclassified: Unclassified repeats.

To determine the level of agreement between the two annotation methods, I calculated the length of overlapping sequences in all repeats as well as three major classes of transposable elements (LINE, DNA transposons, LTR) in the genome. The analysis revealed that the EarlGrey dataset incorporated almost all repeat loci identified by RR. Specifically, more than 90% of LINE and DNA transposon sequences identified by RR were also presented in the EarlGrey annotations (Figure 2). However, there was less overlap for LTR sequences, with approximately 26% of the LTRs in RR not annotated by EarlGrey. This discrepancy may be attributed to EarlGrey's use of LTR_Finder, a structure-based approach particularly effective in accurately identifying full-length LTR elements (Figure 2D) (Baril et al., 2024; Xu & Wang, 2007).

Given the high agreement between the repeat annotations of the *Parhyale* genome produced by RR and EarlGrey and considering EarlGrey's superior quality in generating repeat libraries and annotations, the EarlGrey's dataset was chosen for the subsequent methylation analysis of repetitive elements.

Identifying and characterising transposable elements using bioinformatics is particularly challenging, like to assemble a puzzle with multiple copies of the same or very similar pieces. Selecting the appropriate tools for TE identification and classification in genomes is a complex task, as no single tool can comprehensively fulfil this effort on its own. TE annotation software typically employs similarity-based methods, structure pattern-matching, *de novo* prediction, or a combination of these approaches. Each method involves a trade-off between strengths and weaknesses that must be balanced when choosing a program.

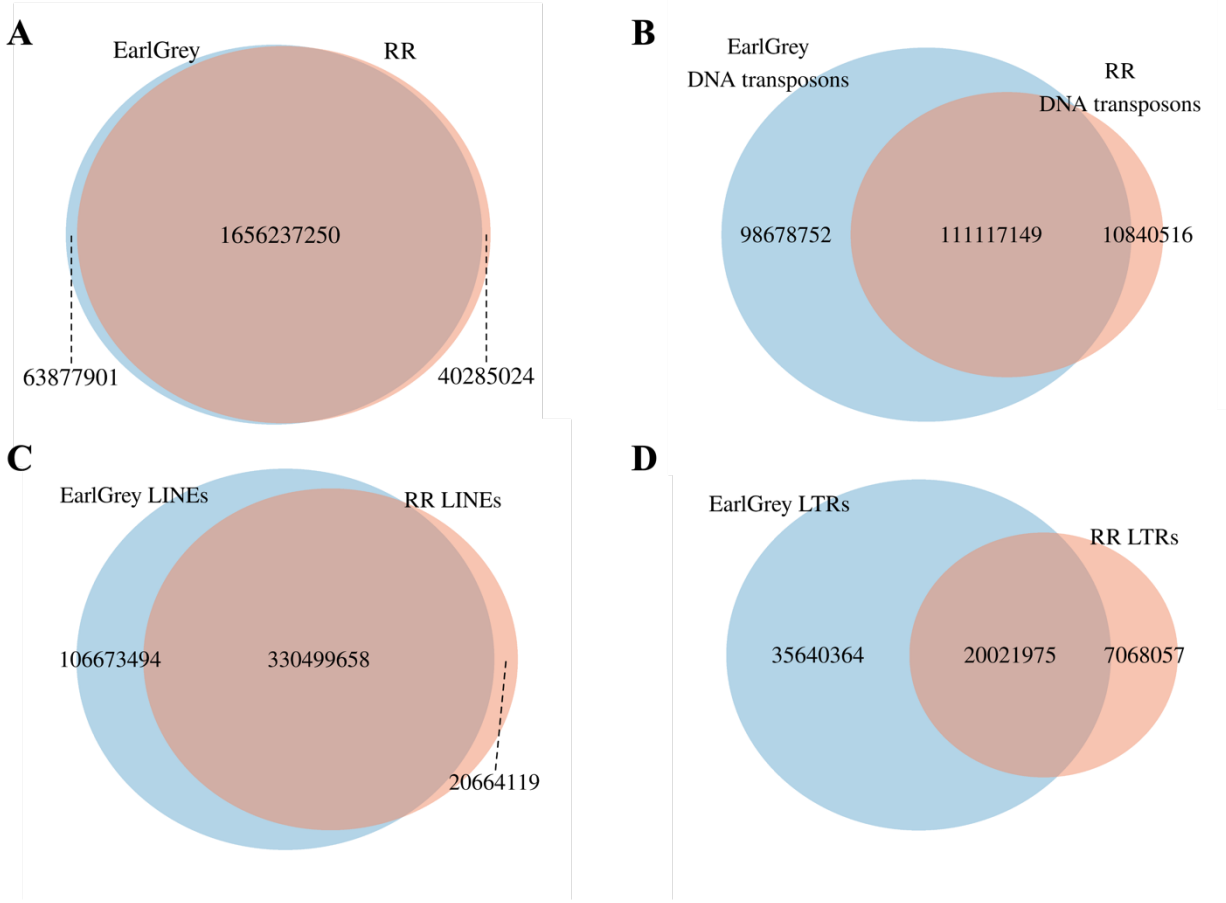


Figure 2. Agreement of repetitive sequences annotated by the two pipelines. **(A)** Agreement of all repetitive sequences in the *Parhyale* genome; **(B)** Agreement of DNA transposon sequences; **(C)** Agreement of LINE sequences; **(D)** Agreement of LTR sequencs. Number represents the sequence length (bp).

There were still a relatively high number of unclassified repeats in the annotation obtained from EarlGrey, yet this issue is common in many less-studied invertebrate species (Athanasouli & Rödelsperger, 2022; Gilbert et al., 2021; Petersen et al., 2021). A study on the transposable elements of insects showed that the unclassified TEs account for up to 90% of the annotated TEs in species that are phylogenetically distinct from *Drosophila* (Gilbert et al., 2021). Several factors contribute to this issue. First, accurately identifying repetitive sequences, especially in less-studied organisms, is computationally challenging. Most repeat classification tools rely on databases of known repetitive elements (e.g., Repbase, Dfam), which are heavily biased towards well-studied species, primarily vertebrates and model organisms like *Drosophila* and *C. elegans* (Hoen et al., 2015; Loreto et al., 2023). These tools may use sequence similarity thresholds or structural features that do not adequately capture the diversity of non-model invertebrate repeats. Consequently, many novel invertebrate-specific repeats might not be represented in these databases, leading to a higher proportion of unclassified repeats in invertebrate species (Hoen et al., 2015; Loreto et al., 2023).

Additionally, invertebrate genomes, particularly those of non-model organisms, often contain a high diversity of repetitive elements, many of which are complex or have undergone significant divergence. The distribution of repeat types in a database, as well as the divergence between the species under study and those present in the database, directly impacts classification quality. There may be a loss of repeat identification when highly divergent reference sequences are used (Platt et al., 2016). Nevertheless, the *Parhyale* genome, like other crustaceans, is repeat-rich and characterised by a higher content of retroelements compared to other transposable elements (Kawato et al., 2021).

2.3.2 Summary of sequencing data

2.3.2.1 Whole genome bisulfite sequencing data

To obtain high-resolution DNA methylation maps for *Parhyale*, we conducted Whole-Genome Bisulfite Sequencing (WGBS) on three DNA samples from various tissues, including mix-stage embryos (Embryo), limb tissues (Limb), and whole-body tissues (Body). In total, we obtained more than 218 million raw paired-end reads, with a total size of 108 Gb. The sequencing results are summarised in Table 4. For each sample, a single replicate was produced. These samples had genome coverages ranging from 1.8X to 2.8X, with a total mapping rate (including both uniquely and non-uniquely mapped reads) of approximately 50% for each sample. According to the reads mapping to the unmethylated spike-in DNA derived from the APC gene, the average bisulfite conversion efficiency for all three samples exceeded 99.9% (Table 4).

Table 4. Summary of WGBS data for each sample.

	Embryo	Body	Limb
Sequences	87,323,554	67,842,707	60,117,030
Uniquely mapped	33,845,426	22,105,311	22,387,190
Uniquely mapping rate	38.8%	32.6%	37.2%
Non-uniquely mapped	14,269,561	9,249,460	10,100,024
Non-uniquely mapping rate	16.3%	13.6%	16.8%
Total mapping rate	55.1%	46.2%	54.0%
Coverage	2.8X	1.8X	1.9X
Conversion rate (apc)	99.95%	99.97%	99.95%

The *Parhyale* genome contains a total of 195,470,590 CpGs, and on average, approximately 28% of these CpGs were covered by at least one read in the WGBS dataset. Table 5 and Figure 3A illustrate the CpG coverage for each sample. It is worth noting that a significant number of CpGs were not covered in the WGBS data.

Table 5. CpG coverage statistics for each sample.

	Embryo		Body		Limb	
	CpGs	**%**	**CpGs**	**%**	**CpGs**	**%**
Total	195,470,590	-	19,5470,590	-	195,470,590	-
Coverage =0	128,786,167	65.89	14,4941,021	74.15	143,376,049	73.35
Coverage ≥ 1	66,684,423	34.11	50,529,569	25.85	52,094,541	26.65
Coverage ≥ 2	31,868,137	16.30	1,8681,036	9.56	20,900,127	10.69
Coverage ≥ 5	5,883,410	3.01	2,084,342	1.07	2,854,904	1.46
Coverage ≥ 10	1,063,512	0.54	391,437	0.20	544,681	0.28

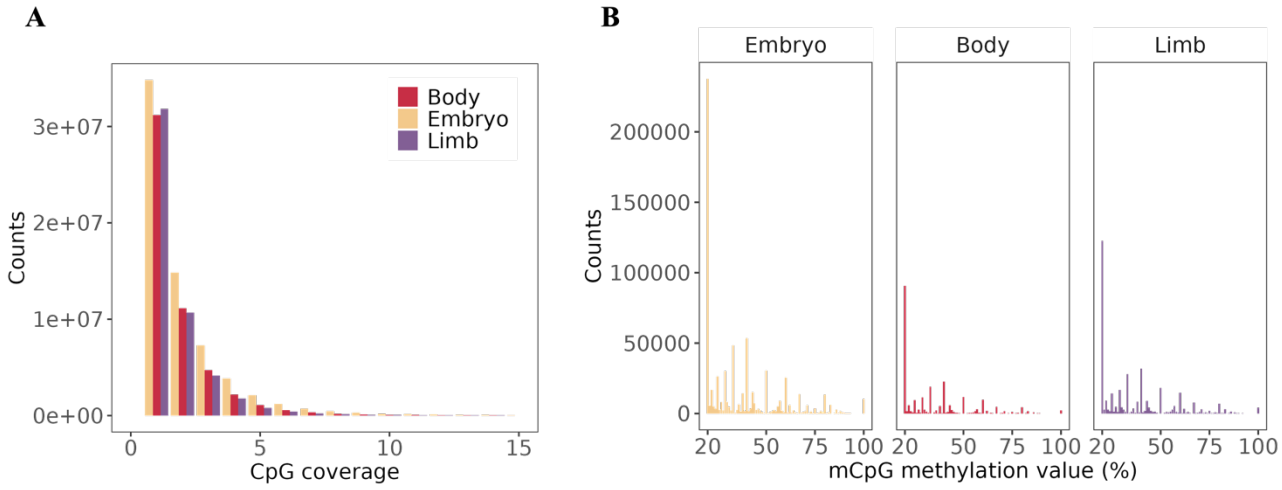


Figure 3. Genome-wide methylation statistics. Distribution of **(A)** CpG coverage and **(B)** methylation values of mCpG across the genome for the three WGBS samples.

To ensure reliable measurement of methylation levels, CpG sites with less than five read coverage were excluded prior to calculating genome-wide methylation levels. As a result, only a small proportion of CpGs (less than 3%) were retained for all three samples (Table 5). In comparison, the effective coverage of CpGs (defined as CpG coverage ≥ 5) in other bisulfite sequencing studies of invertebrates typically ranges from 7.4% to 77% (Gavery & Roberts, 2013; Olson & Roberts, 2014; Silliman et al., 2023; Venkataraman et al., 2020; Wang et al., 2013).

Methylated CpGs (mCpGs) were defined as having a methylation value of at least 20% and a minimum of five covered reads. The distribution of the methylation values of mCpGs across the genome was similar among three samples, however, the methylation value at mCpG sites was not evenly distributed, with only a small proportion of sites being fully methylated (Figure 3B). The vast majority of mCpG sites exhibited moderate methylation values, ranging from 20% to 50% (Figure 3B). Using the mCpG definition, I calculated the average methylation value and genome-wide methylation level for all three samples. Approximately 12% of the effectively covered CpGs (coverage ≥ 5) in the three samples were mCpGs. Meanwhile, the average methylation value of all effectively covered CpGs was around 5.5% (Table 6). However, it should be noted that these values may not be entirely reliable, as only a small fraction of all CpGs (less than 3%) in *Parhyale* genome were included in the analysis.

Table 6. Genome-wide methylation profile of the three WGBS samples.

%	Embryo	Body	Limb
Genome-wide methylation level (%)	11.52	11.74	12.93
Genome-wide average methylation value (%)	5.53	5.37	6.02

2.3.2.2 PacBio single molecule real-time sequencing data

In addition to the WGBS, a PacBio single molecule real-time sequencing sample was generated from a mixture of adult male and female *Parhyale*. In total, 904 Gb of PacBio continuous long reads (CCS) and 26.85 Gb of PacBio high-fidelity circular consensus sequencing (HiFi-CCS) reads were generated. The average length of the HiFi reads was 17,480bp. The methylation status of each CpG site was extracted from the PacBio SMRT sequencing data, the coverage and genome-wide methylation calculation results are displayed in Table 7.

Over 70% of CpGs in the PacBio SMRT sequencing data had coverage of at least five reads, which was notably higher compared to the WGBS data. However, less than 1% of CpGs exhibited coverage exceeding 250X. Of those effectively covered CpGs (coverage ≥ 5), around 8% were mCpGs and the average methylation value was 5%. The genome-wide methylation level and average methylation value were both lower in the PacBio data compared to the WGBS data (Table 7).

The relatively higher genome-wide methylation level observed in the WGBS data was likely due to bisulfite-induced DNA degradation, which preferentially targets genomic regions enriched for unmethylated cytosines, leading to an overestimation of global methylation (Olova et al., 2018). Prior to the subsequent analyses, CpG sites with less than five read coverage were also discarded from the PacBio dataset.

The genome-wide methylation levels were found to be comparable across the three WGBS samples, but the overlap of mCpG sites (coverage ≥ 5 & methylation value ≥ 20%) between samples was low, with only 7.4% of the total mCpGs overlapping in all three samples (Figure 4). This observation is intriguing as it could potentially indicate distinct methylation patterns. However, it is more likely that the limited overlap is a result of the low coverage typically associated with WGBS data. To achieve a higher degree of overlap, significantly greater coverage is required. Nonetheless, both WGBS and PacBio SMRT

sequencing revealed the presence of DNA methylation in the *Parhyale* genome, with a higher genome-wide methylation level than that observed in most invertebrates, especially insects (Lewis et al., 2020).

Table 7. CpG coverage and genome-wide methylation profile of the PacBio SMRT sequencing sample.

PacBio	CpGs	%
Total	195,470,590	-
Coverage ≥ 4	147,528,984	75.47
Coverage ≥ 5	137,267,278	70.22
Coverage ≥ 10	79,308,194	40.57
Coverage ≥ 250	255,902	0.13
Genome-wide methylation level (%)	8.18	
Genome-wide average methylation value (%)	4.99	

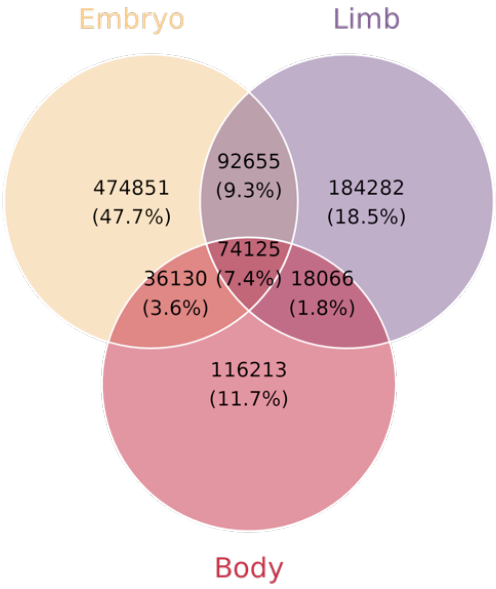


Figure 4. The overlap of mCpGs between WGBS samples. The percentage in each section represents the ratio of mCpG/total mCpG in that section.

2.3.3 Methylation targets in the *Parhyale* genome

The covered CpGs showed a similar distribution across all four samples (Figure 5A). In invertebrates, DNA methylation primarily occurs in transcribed regions, known as gene body methylation (Feng et al., 2010; Zemach et al., 2010). However, in the *Parhyale* genome, mCpGs were evenly distributed between gene bodies and intergenic regions. Within gene bodies, mCpGs were more frequently observed in introns than exons (Figure 5B). In terms of repetitive elements (Repeats), both CpGs and mCpGs were primarily present in LINEs, followed by LTRs and DNA transposons, with the exception of unclassified repeats (Figure 5A, B).

I then calculated both the methylation level (proportion of mCpGs to covered CpGs) and fractional methylation (mean methylation values of all covered CpGs) for various genomic regions and classes of repetitive elements. The genomic regions included gene body, exon, intron, 5'-UTR and 3'-UTR. For the methylation levels of gene regions, 5'-UTRs and introns demonstrated hypermethylation in compared to exons and 3'-UTRs of all the samples (Figure 5C). The methylation levels of repetitive elements in all four samples were higher compared to gene bodies, and the methylation level of various classes of repetitive elements were similar across all four samples, with the highest levels observed in LTRs and unclassified repeats (Figure 5C). The fractional methylation patterns were also consistent among the four samples, with intergenic regions having higher levels than all gene regions, and LTRs exhibiting higher levels than all other repeats classes (Figure 5D).

The elevated methylation levels observed in LTRs compared to other classes of TEs suggest differential targeting for methylation among various TEs. This distinctive methylation pattern also identified in oysters (Wang et al., 2014), contrasts with observations in vertebrate genomes (Feng et al., 2010; Lister et al., 2009; Zemach et al., 2010). Given that DNA methylation likely evolved as a defence mechanism against the

harmful effects of transposon activity in invertebrates, the variation in methylation levels among different TEs may indicate functional distinctions. Highly methylated TEs could be more efficiently silenced, potentially impacting genome stability and the regulation of gene expression (de Mendoza, Hatleberg, et al., 2019; de Mendoza, Pflueger, et al., 2019; Xu et al., 2019; Ying et al., 2022).

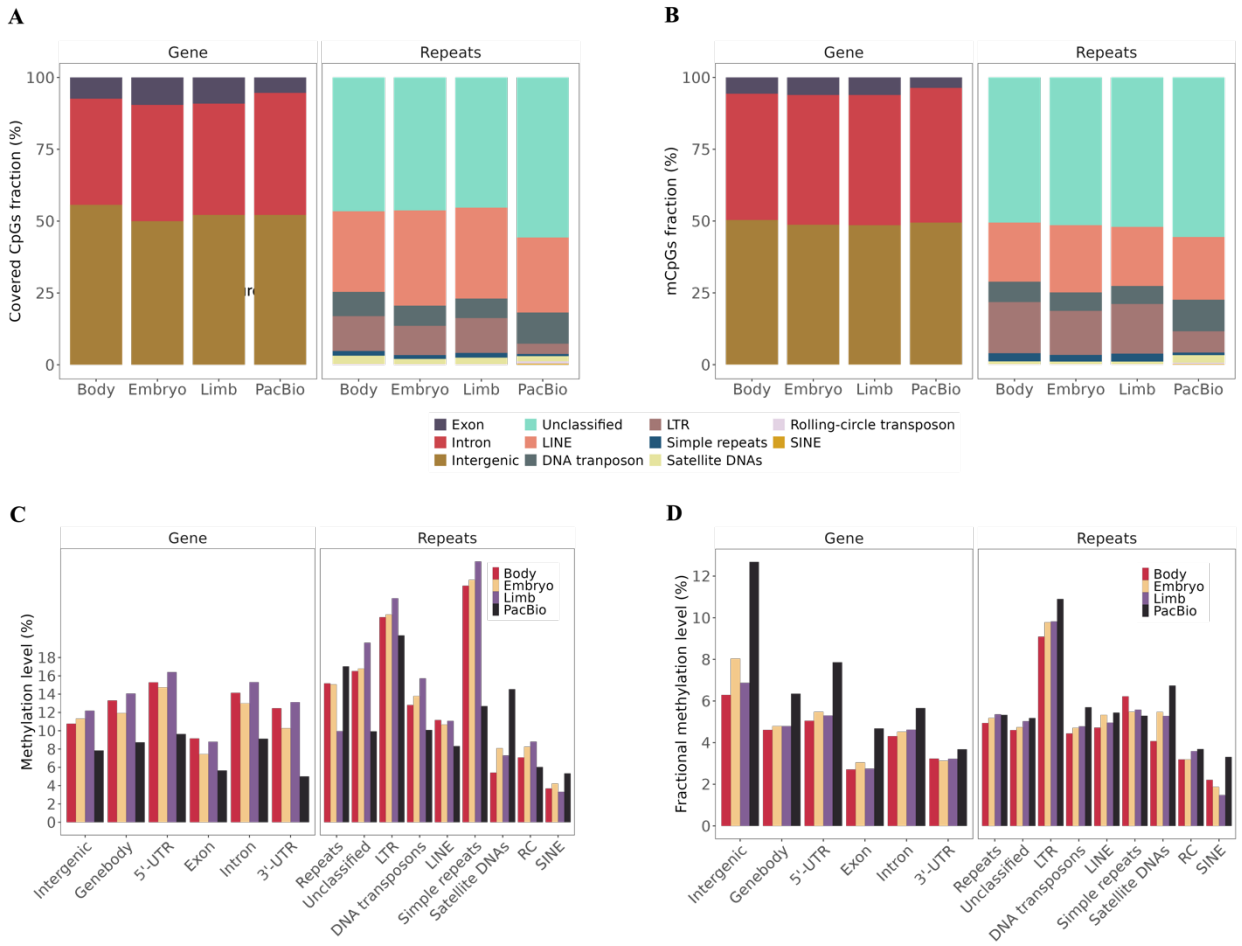


Figure 5. Characterisation of the *Parhyale* methylome. Distribution of **(A)** covered CpGs and **(B)** mCpGs across various genomic regions and repeat classes. **(C)** Methylation levels and **(D)** fractional methylation of gene regions and different repeat classes. Unclassified: unclassified repeats; RC : rolling-circle transposons.

According to the findings, DNA methylation in the *Parhyale* genome targeted both gene bodies and repetitive elements. Despite the substantially lower coverage of the three WGBS samples compared to the PacBio samples, the methylation levels and fractional

methylation of different genomic regions were comparable between the two types of sequencing data. Due to the insufficient data coverage, it is not feasible to interpret the data in the base-pair level. However, interpreting the methylation results at the genomic region level remains reasonable and may provide an overall methylation pattern of the *Parhyale* genome.

2.3.4 Gene body methylation

2.3.4.1 Gene body methylation pattern

To further examine the gene body methylation pattern across the four samples, I divided each gene body into several windows of varying sizes and calculated the fractional methylation for each window. The gene body methylation patterns in the three WGBS samples followed the same trend but differed marginally from the PacBio sample. Specifically, the gene body methylation pattern in the WGBS samples exhibited greater fluctuations, particularly in the gene upstream regions (Figure 6A). Additionally, the upstream 2kb of the transcription start site (TSS) in the WGBS samples demonstrated similar methylation levels to the downstream 2kb of TSS. This contrasted with the PacBio samples, where the upstream regions had lower methylation compared to the first 4000 bp of the gene bodies (Figure 6B).

Moreover, all four samples exhibited a slight decrease in methylation levels beginning from 500bp before the TSS, followed by an immediate increase after the TSS, and then a gradual decrease towards the 3'-end of the gene body (Figure 6A). Notably, the number of windows with effective values in the gene bodies, upstream, and downstream regions was lower in the WGBS samples compared to the PacBio samples (Table 8), indicating that the gene methylation pattern in the WGBS data may not accurately reflect the true pattern.

Furthermore, most genes contain fewer than 10 introns (Figure 6D). Therefore, I examined the methylation patterns of the first ten exons and introns and found that the patterns were similar across all four samples, with the PacBio sample exhibiting greater variability compared to the three WGBS samples. Notably, introns were found to be hypermethylated compared to exons in the *Parhyale* genome (Figure 6C). This pattern contrasts with that observed in many other arthropods, where methylation in their gene bodies tends to be enriched in exons rather than introns (Lewis et al., 2020).

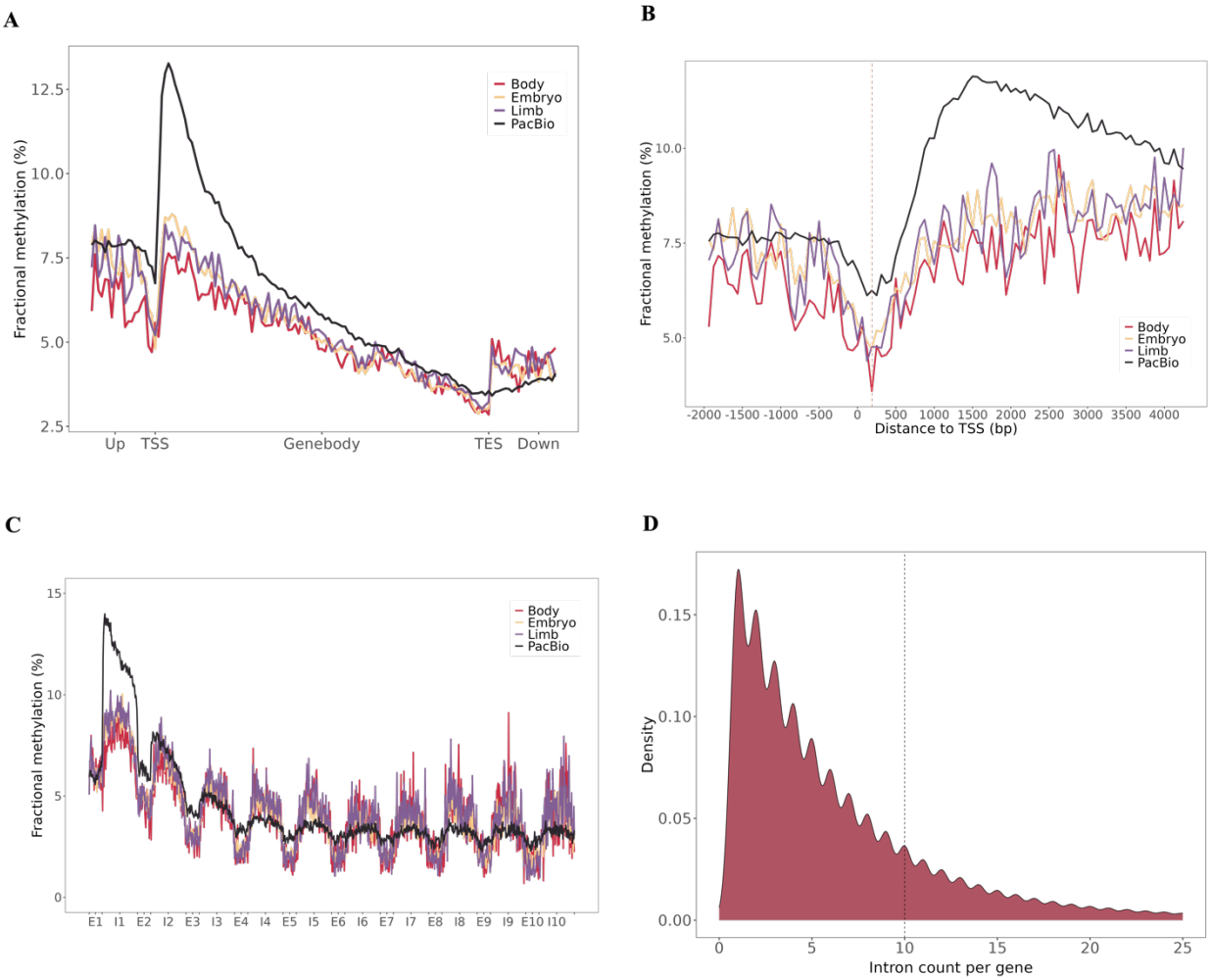


Figure 6. Gene body methylation patterns **(A)** Fractional methylation within the gene bodies and 2 kb flanking (upstream: up and downstream: down) regions for all the protein-coding genes of all samples. Each gene body was divided into 100 bins with varying bin sizes, while the upstream and downstream regions were divided into 20 bins with 100 bp bin sizes. **(B)** Fractional methylation pattern of the flanking TSS regions (TSS-2kb to TSS+4kb). Each region was divided into 100 bins of varying bin sizes **(C)** Pattern of fractional methylation in the first ten exons and introns of all protein-coding genes. Exon

(E) and intron (I) were divided into 30 and 80 fixed number of bins with varying sizes, respectively **(D)** Distribution of intron count per gene. The majority of genes contain fewer than 10 introns (the vertical dashed line).

Table 8. Percentage of windows with effective fractional methylation values in gene bodies, upstream and downstream regions of the four samples.

	Gene upstream	Gene body	Gene downstream
Embryo	6.9%	19%	7.6%
Body	2.5%	8.6%	2.6%
Limb	3.3%	11%	3.5%
PacBio	55%	56%	57%

The number of covered CpGs in each gene of the WGBS samples was considerably lower compared to the PacBio sample, with a large proportion of genes in the WGBS samples having fewer than 10 covered CpGs (Figure 7A). Therefore, genes with fewer than three covered CpGs were excluded from the subsequent gene body methylation analysis. All four samples exhibited a bimodal pattern in the distribution of gene body methylation, with a slight second peak observed (Figure 7B-E). This bimodal distribution, which can distinguish between lowly and highly methylated genes, has also been observed in the gene body methylation of other invertebrates (Sarda et al., 2012). Since the second peak occurs at around 15%, I chose 10% as the cut-off to differentiate between hypermethylated and hypomethylated genes in the following analysis (Figure 7B-E).

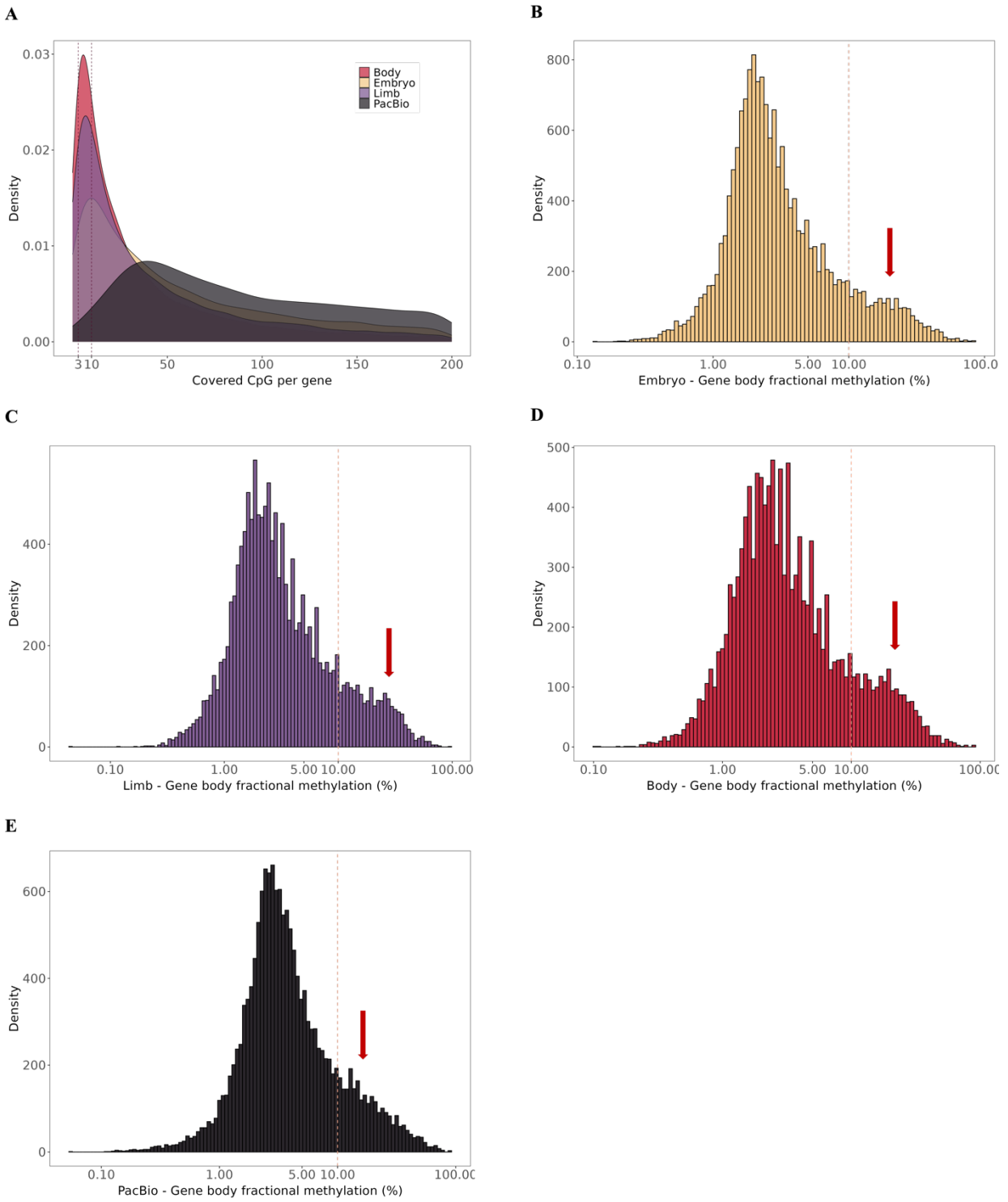


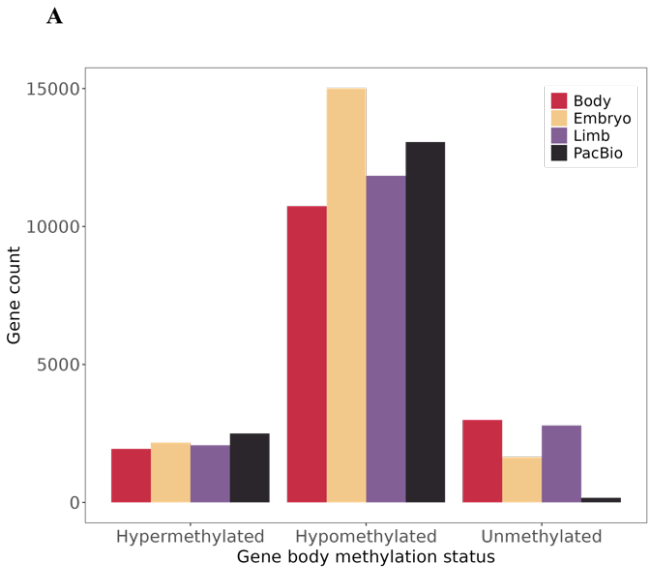
Figure 7. Distribution of gene body methylation **(A)** Distribution of covered CpG per gene. The majority of genes in the WGBS samples contain fewer than 10 covered CpGs in their gene bodies (the right vertical dashed line). The minimum gene retention threshold for gene body methylation analysis was set at 3 (the left vertical dashed line). **(B-E)** The distribution of fractional methylation for all the genes in the four samples. The vertical dashed line represents the methylation threshold (10%) used to categorise genes as hypermethylated or hypomethylated genes. The arrow points to the second peak.

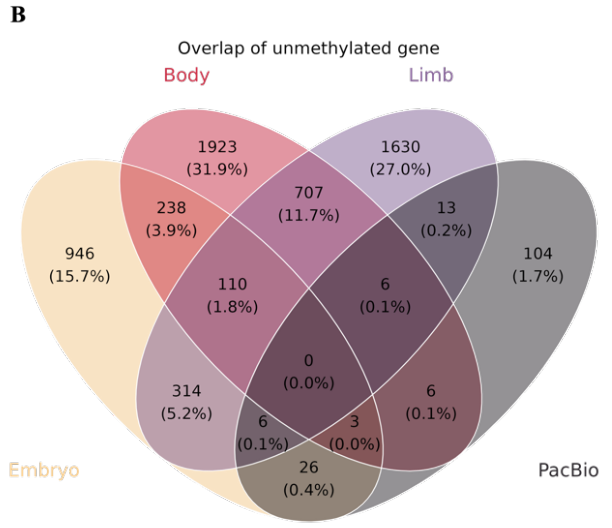
2.3.4.2 Comparison of gene body methylation across different samples

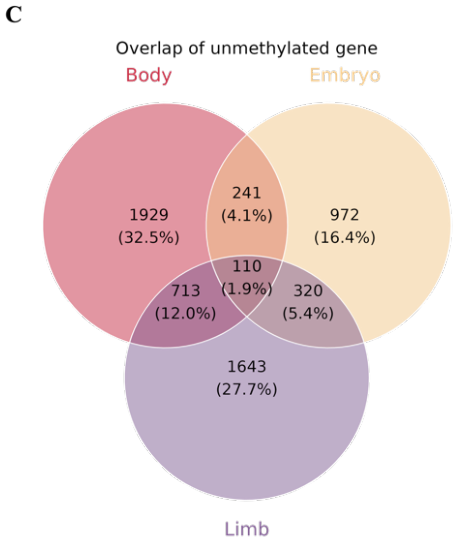
After categorising genes into unmethylated (methylation = 0), hypomethylated (0< methylation <10%), and hypermethylated (methylation ≥ 10%) groups based on their fractional methylation levels, I examined the degree of overlap between these three gene groups across the four samples. The hypomethylated gene group was the largest, while the other two groups contained comparable numbers of gene (Figure 8A). Interestingly, the PacBio sample had very few unmethylated genes, mainly due to the high CpG coverage in this dataset.

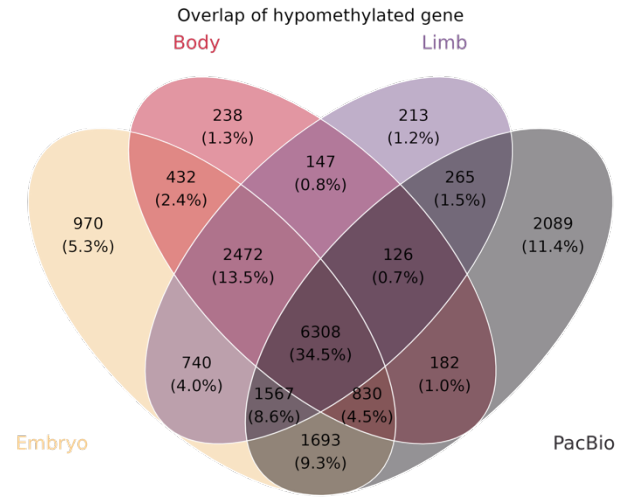
When comparing only the WGBS samples, there was a higher overlap of unmethylated genes between limb tissues and whole-body tissues, but less than 2% of the genes were unmethylated in all three samples (Figure 8B, C). The methylation status of hypomethylated genes was highly conserved across different samples, with 34.5% of the genes remaining hypomethylated in all four samples and an even higher proportion (54.3%) when comparing only the WGBS samples (Figure 8D, E). Although there were significantly fewer hypermethylated genes compared to hypomethylated genes, a relatively high proportion of these genes overlapped among all four samples, with a more pronounced proportion among the WGBS samples (Figure 8F, G). These results suggest that the gene body methylation status of *Parhyale* was conserved in a subset of genes across different tissues.

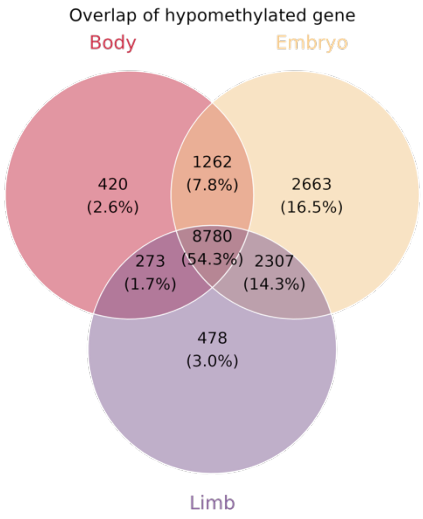
To further investigate the methylation patterns in *Parhyale*, I employed hierarchical clustering to generate a methylation heatmap of gene bodies for the four samples (Figure 9). The fractional methylation of each gene body was calculated and utilised in this analysis. The result indicates that the majority of genes might exhibit tissue-specific methylation patterns. However, it is also probable that the observed low similarity of gene body methylation status between the four samples could be attributed to low sequencing depth and insufficient coverage in the WGBS samples.


A
15000
10000
5000
0
Gene count
Body
Embryo
Limb
PacBio
Hypermethylated
Hypomethylated
Unmethylated
Gene body methylation status


B
Overlap of unmethylated gene
Body
Limb
1923 (31.9%)
1630 (27.0%)
238 (3.9%)
707 (11.7%)
13 (0.2%)
946 (15.7%)
104 (1.7%)
110 (1.8%)
6 (0.1%)
0 (0.0%)
314 (5.2%)
6 (0.1%)
6 (0.1%)
3 (0.0%)
26 (0.4%)
Embryo
PacBio


C
Overlap of unmethylated gene
Body
Embryo
1929 (32.5%)
241 (4.1%)
972 (16.4%)
110 (1.9%)
713 (12.0%)
320 (5.4%)
1643 (27.7%)
Limb


D
Overlap of hypomethylated gene
Body
Limb
238 (1.3%)
213 (1.2%)
147 (0.8%)
432 (2.4%)
265 (1.5%)
970 (5.3%)
2089 (11.4%)
2472 (13.5%)
126 (0.7%)
6308 (34.5%)
740 (4.0%)
182 (1.0%)
1567 (8.6%)
830 (4.5%)
1693 (9.3%)
Embryo
PacBio


E
Overlap of hypomethylated gene
Body
Embryo
420 (2.6%)
1262 (7.8%)
2663 (16.5%)
8780 (54.3%)
273 (1.7%)
2307 (14.3%)
478 (3.0%)
Limb

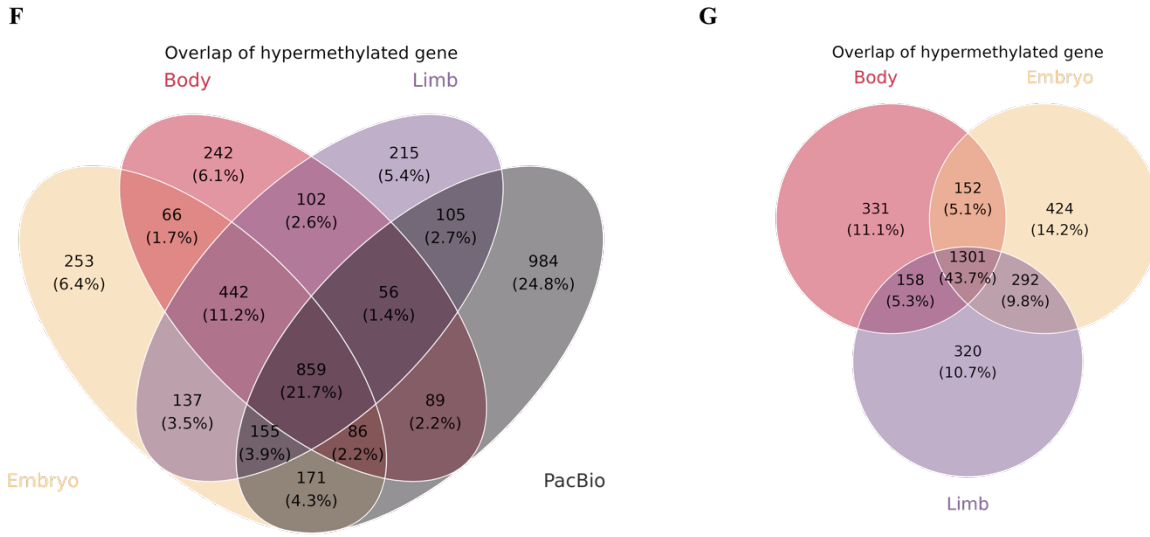


Figure 8. (A) The quantity of unmethylated, hypomethylated, and hypermethylated genes present in the four samples. Venn diagram showing the overlap of **(B,C)** unmethylated, **(D,E)** hypomethylated and **(F,G)** hypermethylated genes between the four sequenced samples (left) and only the WGBS samples (right).

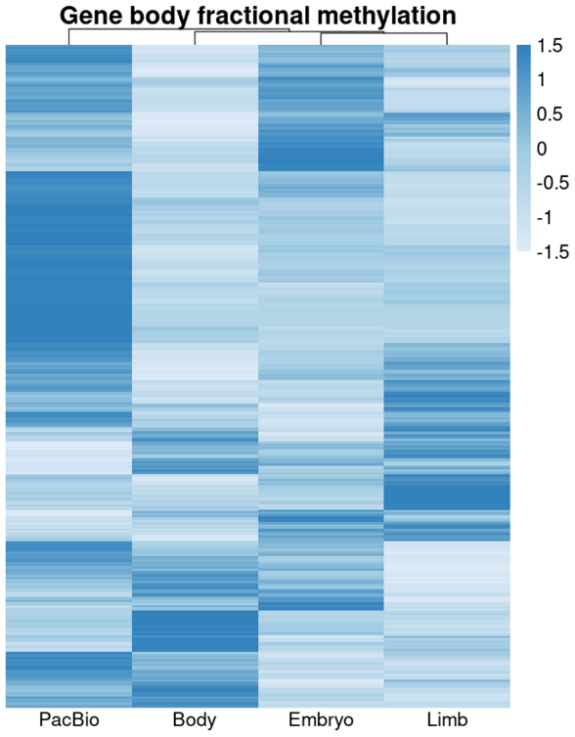


Figure 9. Comparative analysis of gene body methylation patterns shown in a heatmap with clustering for individual samples and tissues.

2.4 Conclusion

Analysis of four samples using both WGBS and PacBio SMRT sequencing has revealed the overall pattern of DNA methylation and the evidence of DNA methylation in gene bodies and repetitive elements of the *Parhyale* genome. However, technical limitations in coverage depth, mapping rate, and CpG coverage highlight the preliminary nature of these results and underscore the need for further validation.

Despite these limitations, including both WGBS and PacBio data in this thesis is crucial for several reasons. Firstly, the data provide valuable preliminary insights into DNA methylation patterns in the *Parhyale* genome, laying the foundation for further investigations. Additionally, WGBS data offers a foundational layer of methylation context across the genome, serving as a benchmark for PacBio methylation data. Integrating data from both sequencing technologies allows for a more comprehensive analysis of the methylation landscape, leveraging the strengths of each platform to provide a better understanding of genomic methylation patterns.

Nevertheless, achieving a satisfactory coverage depth, mapping rate, and CpG coverage is essential for uncovering the full extent of DNA methylation patterns in *Parhyale*. Therefore, alternative methodologies need to be adapted for comprehensive methylome characterisation of the *Parhyale* genome. Beyond technical considerations, the lack of transcriptome data from corresponding tissues and embryonic developmental stages presents a significant hurdle for integrative analyses between DNA methylation and gene expression in *Parhyale*.

In the following chapter, I generated new Methyl-seq data using a recently developed enzymatic methyl-seq technology aiming to attain better quality genome-wide methylation data. This new approach will allow us to delve deeper into the biological significance of DNA methylation dynamics in *Parhyale*. By investigating the interplay between DNA

methylation, gene expression, and transposon activity during early development, my goal is to uncover the regulatory networks of DNA methylation in this species' early developmental stages.

Chapter 3

DNA methylation analysis in the early development of *Parhyale hawaiensis*

3.1 Introduction

3.1.1 Early development of animals

Animal early development is initially controlled by maternal mRNAs and proteins that are loaded into the egg during oogenesis. After fertilization, the embryo is transcriptionally silent; therefore, maternally inherited transcripts provide the embryo with gene products necessary for its earliest developmental processes. As development progresses, there is a specific time period when maternal products start to degrade gradually, and the transcription of the zygotic genome commences. This stage is known as the maternal-to-zygotic transition (MZT), which involves the transfer of developmental control from the maternal genome to the zygotic genome (M. T. Lee et al., 2014; Schultz, 2002; Tadros & Lipshitz, 2009; Vastenhouw et al., 2019; Wagner et al., 2004).

The maternal-to-zygotic transition (MZT) is a crucial process in animal development comprising two main steps: the clearance of maternally deposited products, such as mRNAs and proteins, and the activation of zygotic genome transcription, known as zygotic genome activation (ZGA) (M. T. Lee et al., 2014; Tadros & Lipshitz, 2009). Various studies of animal oocytes have revealed that they contain a considerable amount of mRNA products, representing a wide variety of the protein-coding transcriptome within their genome. In some animals, such as mice and *C. elegans*, maternal transcripts account for approximately one-third of the protein-coding transcriptome (Stoeckius et al., 2014; Q. T. Wang et al., 2004), while in others, such as *Drosophila*, and zebrafish, they account for as much as three-quarters (Harvey et al., 2013; Thomsen et al., 2010). Although the proportion of maternal products cleared from the embryo during MZT varies between species, genome-wide transcriptome analyses have demonstrated that maternal mRNAs are progressively degraded during the MZT period (Aanes et al., 2011; Baugh et al., 2003; Hamatani et al., 2004; Harvey et al., 2013; Thomsen et al., 2010). The initial phase of transcript degradation is regulated solely by maternal gene products, whereas

the subsequent phase requires a contribution of zygotic gene products. The timing and developmental stages of these two phases differ among species (Figure 1) (Tadros & Lipshitz, 2009).

ZGA is a coordinated process that occurs in two transcriptional bursts. During the first wave, known as the ‘minor wave’ of transcription, the silenced zygotic genome becomes active with a low level of transcription and relatively few loci. This is followed by a larger and more tightly regulated transcriptional activation, termed the ‘major wave’ of transcription. These two waves of transcription are gradually activated over a period of time during MZT, without clear boundaries between the minor and major waves (Aanes et al., 2011; Harvey et al., 2013; Heyn et al., 2014; Pauli et al., 2012; Sandler & Stathopoulos, 2016; Tan et al., 2013; White et al., 2017). Moreover, the timing and duration of ZGA vary widely among different species, with some initiating ZGA at the one-cell stage and others postponing it until after the first few cell divisions (Figure 1) (Abe et al., 2015; Paranjpe et al., 2013; Renzis et al., 2007; Tan et al., 2013; Yan et al., 2013).

ZGA entails the transcription of numerous genes that encode proteins responsible for regulating development. While the specific genes activated during ZGA vary across different species, they typically involve transcription factors and developmental regulators. Certain ZGA-produced proteins are essential for triggering the later waves of maternal clearance and promoting the transcription of subsequent genes (Aanes et al., 2011; Baugh et al., 2003; Collart et al., 2014; Hamatani et al., 2004; Harvey et al., 2013; Heyn et al., 2014; M. T. Lee et al., 2014; Renzis et al., 2007; Tan et al., 2013). There are two types of zygotically activated genes: a limited number of zygotic-only genes, which are not maternally provided, and the vast majority of genes, which are both maternally provided and zygotically expressed (M. T. Lee et al., 2014; Renzis et al., 2007; Yan et al., 2013).

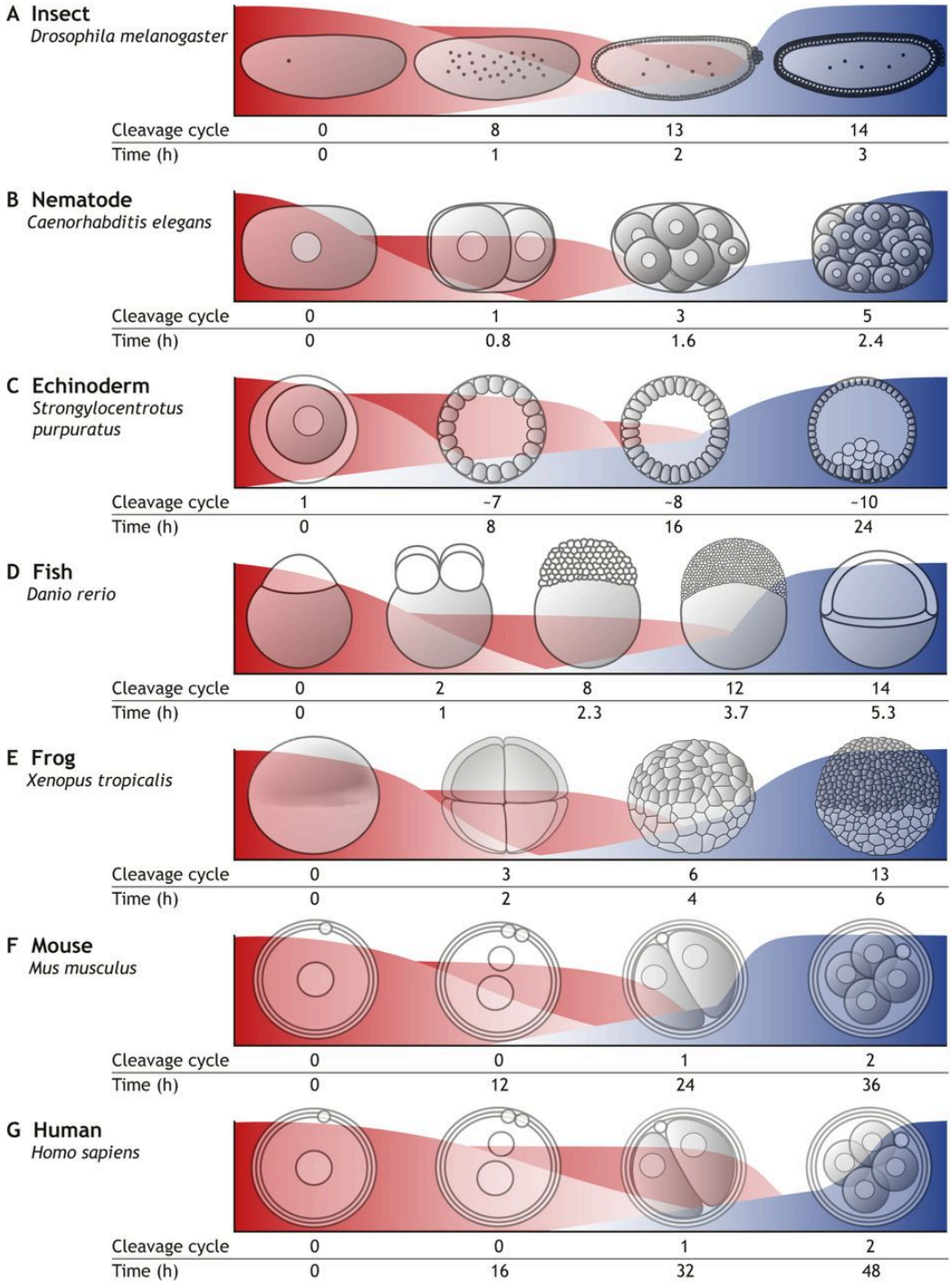


Figure 1. The maternal-to-zygotic transition of different organisms. **(A)** *Drosophila melanogaster*, **(B)** *Caenorhabditis elegans*, **(C)** *Strongylocentrotus purpuratus* (sea urchin), **(D)** *Danio rerio* (zebrafish), **(E)** *Xenopus tropicalis* (frog), **(F)** *Mus musculus* (mouse) and **(G)** *Homo sapiens* (human). Red curves represent degraded maternal transcripts, while red gradients represent the timing of the multiple phases of transcript degradation. Blue curves represent zygotic transcription, and blue gradients represent a gradual increase in zygotically transcribed genes during ZGA. Adapted from Vastenhouw et al, (2019).

3.1.2 DNA methylation during the MZT

During the MZT, various processes occur, including the regulation of maternal gene products and mRNA translation, the post-translational regulation of maternal proteins, the activation of the zygotic genome, and the initiation of zygotic transcription (Tadros & Lipshitz, 2009). However, invertebrate model systems such as *Drosophila melanogaster* and *C. elegans* do not possess DNA methylation in their genomes. Consequently, most studies on DNA methylation related to MZT and ZGA have been conducted in vertebrate species. These studies have revealed that the regulation of DNA methylation plays an important role in ensuring a successful transfer of developmental control from one generation to the next in vertebrates (Eckersley-Maslin et al., 2018; Vastenhouw et al., 2019).

Methylation levels in the genomes of vertebrates are consistently high, but the dynamics of methylation during the vertebrate ZGA vary substantially between species. The human and mouse methylomes are reprogramed in early embryos during preimplantation development. The zygotic genome undergoes substantial decreases in global DNA methylation, followed by re-establishment of gene-specific methylation during gastrulation (Hackett & Surani, 2012; Iurlaro et al., 2017; H. J. Lee et al., 2014; Messerschmidt et al., 2014). In contrast to mammals, zebrafish use a significantly distinct DNA methylation reprogramming strategy during early development. In zebrafish, the oocyte methylome is substantially hypomethylated relative to the sperm methylome. As a result, the pattern of paternal DNA methylation remains unchanged throughout early embryogenesis. While the oocyte methylome gradually undergoes elimination and reprogramming to eventually attain a pattern of methylation comparable to that of the sperm (Jiang et al., 2013). In addition, the *Xenopus* genome is constitutively hypermethylated during the early stages of its development (Veenstra & Wolffe, 2001). Despite these differences, research has shown that DNA methylation plays important role in regulating gene expression during ZGA in both mammals and zebrafish (Li et al., 2013; Potok et al., 2013).

Apart from vertebrates, a recent investigation analysed the reprogramming of parental DNA methylation in four invertebrate species. The analyses revealed that DNA methylation reprogramming occurs during early development in sea urchin and sea squirt but not in sea anemone and insects such as honeybees (Xu et al., 2019). Other invertebrates, such as *C. gigas,* share partial similarity with the mouse genome, which undergoes global demethylation shortly after fertilization and then becomes re-methylated at specific loci (Hershko et al., 2003; Riviere et al., 2013). Some studies have also highlighted the functional significance of DNA methylation during the early development of invertebrate species. For instance, DNA methylation has been shown to influence the transcriptional activity of developmental genes in the oyster *C. gigas* (Riviere et al., 2013). Moreover, knocking down dnmt1 in the wasp *Nasonia* genome reduces gene body methylation, which results in developmental failure during gastrulation (Arsala et al., 2022).

Taken together, different species use distinct DNA methylation reprogramming strategies during early development, and methylation reprogramming is not required for all species with DNA methylation in their genome, such as sea anemone (Xu et al., 2019). It is believed that DNA methylation plays a significant role in the early development of some invertebrates and may influence gene expression and several essential developmental processes, but its precise mechanisms and functions remain unknown (Arsala et al., 2022; Riviere et al., 2013; Xu et al., 2019).

3.1.3 Method to detect DNA methylation

The first step in DNA methylation research is determining whether a particular cytosine residue is methylated. This can be performed both directly and indirectly. Indirect techniques involve the use of enrichment-based methods such as antibodies, methyl-binding proteins, pull-down assays, and methylation-sensitive restriction enzyme digestion. The limitation of indirect methods is that they do not provide resolution down to

a single nucleotide. Direct techniques, such as sequencing, provide more precise and accurate information about DNA methylation in larger genomes (Chatterjee et al., 2017; Kurdyukov & Bullock, 2016).

Bisulfite sequencing (BS-seq) is the gold standard method for studying DNA methylation. By combining bisulfite treatment with high-throughput sequencing, the method can be extended to whole-genome DNA methylation analysis, referred to as whole-genome bisulfite sequencing (WGBS), which has become the standard technique for measuring genome-wide methylation over the past several decades. For bisulfite treatment, DNA is treated with sodium bisulfite, unmethylated cytosine is deaminated into uracil and then replaced with thymine during PCR (Figure 2). In contrast, methylated cytosines are resistant to the bisulfite reaction and continue to exist as cytosines. By comparing treated and untreated sequences, the position of methylated cytosines can be identified (Kurdyukov & Bullock, 2016). However, bisulfite sequencing has certain limitations. One major drawback is that bisulfite conversion of cytosine reduces genome complexity and causes substantial DNA degradation. Additionally, bisulfite treatment degrades DNA at unmethylated cytosines via depyrimidination more efficiently than at methylated cytosines. Moreover, the vertebrate genome contains a small amount of hydroxymethyl cytosines that are undetectable by BS-seq (Kurdyukov & Bullock, 2016; Yong et al., 2016). These factors make it difficult to conduct WGBS on samples with limited DNA input, and methylation estimates may be inflated as a consequence.

To address the limitations of bisulfite sequencing, two bisulfite-free and enzyme-based techniques have recently been developed. The first technique, TAPS-seq, uses ten-eleven translocation (TET) oxidation to convert 5-methylcytosines (5mC) and 5-hydroxymethylcytosines (5hmC) to 5-carboxylcytosine (5caC), which is then reduced to dihydrouracil using pyridine borane. PCR is then used to convert the resulting dihydrouracil to thymine, allowing a C-to-T transition of 5mC and 5hmC (Liu et al., 2019). The second method uses cytidine deaminase APOBEC3A to convert both 5mC and

unmethylated cytosines to uracil, which can then localize 5hmC, while minimising DNA damage. However, APOBEC3A alone is unable to distinguish between unmodified cytosine and 5mC, as both can be deaminated by this enzyme (Schutsky et al., 2018).

Enzymatic methyl-seq (EM-seq) is a recently developed non-bisulfite technique that uses TET2 and an oxidation enhancer to convert 5mC and 5hmC into 5caC and glycosylated 5hmC, following by deaminating cytosines into uracil by APOBEC. During the PCR amplification that follows, oxidized methyl cytosines and uracil form base pairs with guanines and adenines, respectively (Figure 2) (Williams et al., n.d.). The converted sequence of EM-seq is identical to the BS-seq sequence, enabling the use of the same alignment and downstream analysis tools.

Compared to BS-seq, multiple studies have shown that EM-seq is a preferable method for quantifying whole-genome methylation (Vaisvila et al., 2021). EM-seq libraries have larger insert sizes, more intact DNA, and produce higher quality libraries due to its gentler enzymatic treatment. Additionally, EM-seq libraries exhibit uniform GC coverage and identify a greater number of genomic features and CpGs at lower sequencing coverage depths than other methods. When compared to post-bisulfite adaptor tagging (PBAT), EM-seq displays greater consistency between replicates, better mapping quality, lower duplication rates, lower background noise, greater average coverage, and greater coverage of total cytosines. Overall, these findings suggest that EM-seq is a more cost-effective and efficient approach for studying whole-genome methylation, particularly in low-input samples (Feng et al., 2020; Han et al., 2022; Sun et al., 2021; Williams et al., n.d.).

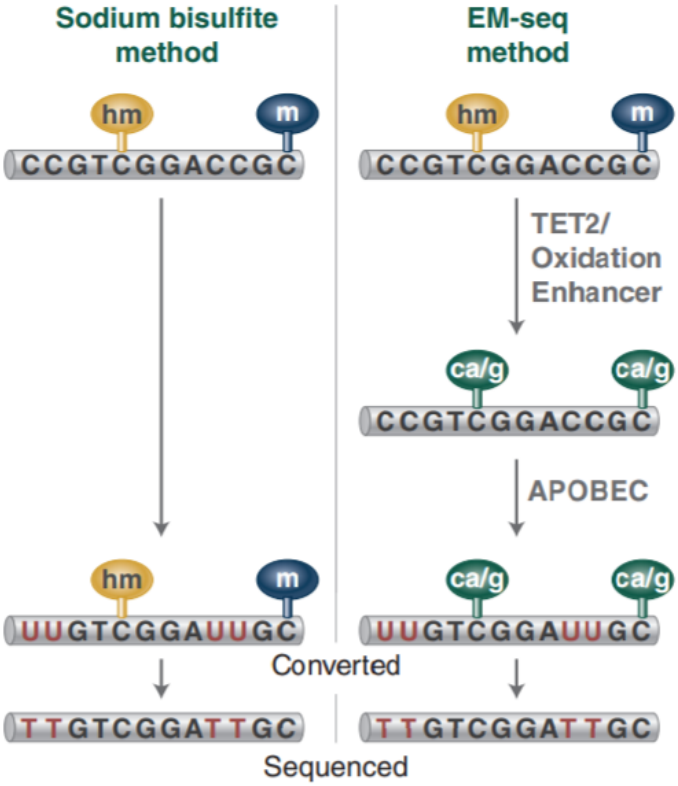


Figure 2. Enzymatic Methyl-seq and BS-seq methods. Adapted from https://uk.neb.com/products/e7120-nebnext-enzymatic-methyl-seq-kit#Product%20Information

3.1.4 Aim of this chapter

The reprogramming of parental methylomes in vertebrates during early embryogenesis has been extensively studied, revealing crucial insights into the role of DNA methylation in regulating developmental processes. However, little is known about DNA methylation dynamics in invertebrate early development. *Parhyale hawaiensis*, with its methylated genome, presents an ideal model organism to investigate the dynamics and potential functions of DNA methylation in this context.

This chapter employs EM-seq to analyse samples derived from *Parhyale* embryos at three distinct developmental stages. The primary aim of this chapter is to understand the complexity of DNA methylation changes and methylome dynamics during *Parhyale*'s early development, shedding light on its regulatory role in gene expression and transposon activity. Through genome-wide methylation analysis, I focused on

characterising the distribution and patterns of DNA methylation, particularly within intragenic and intergenic regions, with an emphasis on repetitive elements. Additionally, I aimed to explore the relationship between DNA methylation and gene expression during *Parhyale* early development. Drawing from previous research on invertebrates, I sought to elucidate the association between gene body methylation and gene expression, with a particular focus on the zygotic genome activation period. Overall, this chapter aims to gain valuable insights into the regulatory mechanisms underlying *Parhyale* early embryonic development, paving the way for further exploration into the functional significance of DNA methylation in subsequent chapters.

3.2 Materials and methods

3.2.1 *Parhyale* culture

The *Parhyale hawaiensis* population in the laboratory was initially obtained and isolate for cultured in the Patel lab. From this, an isofemale line was established, called the 'German line', and this was used in the *Parhyale* genome project (Kao et al., 2016). Currently, in our lab, these creatures are housed in plastic or glass containers lined with crushed coral substrate to buffer the pH and filled with artificial seawater having a salinity range of 28-32 PPT (Figure 3A). These tanks are situated in a climate-controlled room with a temperature of 25°C and are illuminated by fluorescent strip lights with a 12-hour day/night cycle. The aquariums are aerated using pumps and air stones, and the water is replaced weekly. The animals are fed a diet of fish flakes and carrots once per week.

3.2.2 Embryo collection

To obtain *Parhyale* embryos for experimentation, a group of mating pairs was collected in the afternoon and placed in a small container overnight. Most of the males and females in the container separate the following morning. This separation occurs when females moult after mating. Females then released mature oocytes via their oviducts, which were fertilized and deposited in the female's ventral brood pouch (Figure 3B-E).

Gravid females were collected in a petri dish containing artificial seawater and anaesthetised with clove oil (Sigma. C8392-100 ML) diluted at a ratio of 1:10,000 in artificial seawater (Figure 3E). The embryos were carefully extracted from the ventral brood pouch of paralytic females using forceps and a glass pipette with a pointed tip. These embryos were then staged using the bright filed microscopy using a Leica MZ16F Fluorescence Stereo. Only embryos with one to eight cells were retained and placed in a petri dish containing filtered artificial seawater (FASW) and antibiotics (penicillin-streptomycin: Gibco™ 15140122 and amphotericin B: Sigma 1397-89-3). These embryos

were then kept in a chamber with high humidity and a temperature of 28.3°C until they reached the appropriate developmental stage for further experimentation.

When the embryos reached the appropriate stages, they were inspected using a Leica MZ16F Fluorescence Stereo microscope, and those with developmental delays or that were dead were removed. The remaining embryos were collected and stored in 20 µl of lysis buffer (100 mM TrisHCl pH 8.0, 100 mM NaCl, 50 mM EDTA, 0.5% SDS and MiliQ H_2O) in a 1.5 ml Eppendorf tube. The embryos were then rapidly frozen with dry ice and kept at -80°C. The timing for embryo collection was based on the developmental staging of *Parhyale hawaiensis* generated by the Extavour lab (https://www.extavourlab.com/wp-content/uploads/2017/10/Parhyale-staging-table.pdf). Three developmental stages of *Parhyale* embryos were chosen for this study: 0-9 hour (Stage 1, S1), 11-12 hour (Stage 2, S2), 24-60 hour (Stage 3, S3). S1 included embryos at the one-cell, two-cell, four-cell and eight-cell stage, which can be identified visually. These embryos were immediately stored in lysis buffer and rapidly frozen after collection from the ventral brood pouch. Since embryos beyond the 16-cell stage (9-hour) cannot be visually recognised for specific developmental stages, time-based inference was used to identify embryos in the S2 and S3 stages.

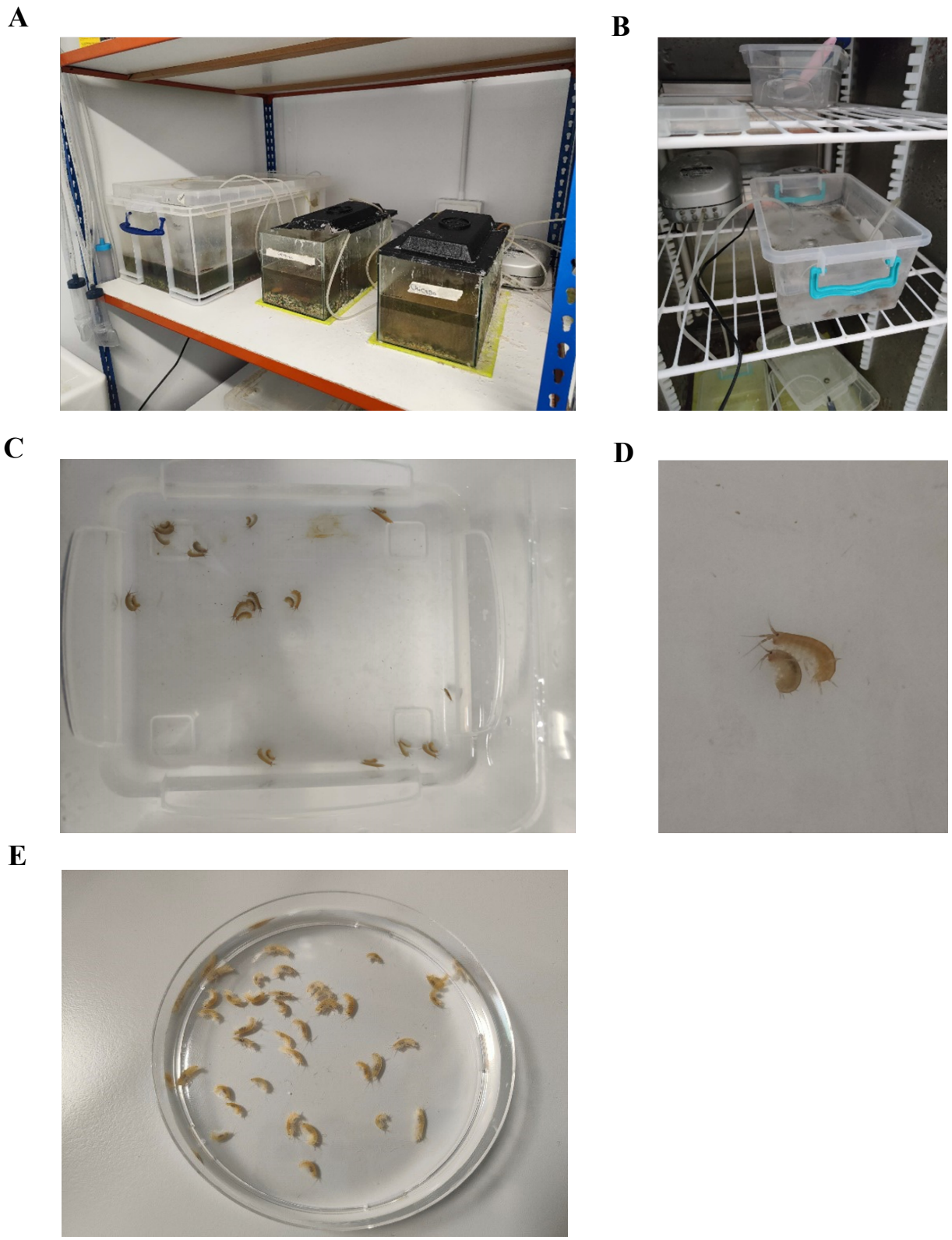


Figure 3. Images depicting various aspects of *Parhyale* culture and embryo collection: **(A)** Containers used for housing *Parhyale* with crushed coral substrate and artificial seawater; **(B)** Small box for keeping mating pairs; **(C) & (D)** Mating pairs of *Parhyale*; **(E)** Gravid female *Parhyale*.

3.2.3 DNA extraction

To conduct the experiment, six samples were collected, with each of the three developmental stages having two biological replicates. The S1 samples were a mix of 80% eight-cell embryos and 20% one to four-cell embryos. The S2 samples consisted of 33% of embryos at the 32-cell, 64-cell, and 132-cell stages, respectively. The S3 samples were a mix of embryos at the 24-hour, 36-hour, 48-hour, and 60-hour stages, with each stage contributing 25% to the samples.

When the collection of embryos was complete, embryos of the same developmental stage were combined in a single Eppendorf tube, and genomic DNA was extracted using a phenol-chloroform-based extraction protocol. In brief, 500 µl of lysis buffer (100 mM TrisHCl pH8, gibco, 15568-025; 100 mM NACl; 50 mM EDTA, corning, 46-034-CI; 0.5% SDS, Invitrogen, 15553-035) was added to the sample, and the sample was homogenised manually with a pestle for a few minutes. The sample was then treated with 10 µl Proteinase K (Sigma, P4850) and incubated at 55°C for about 3 hours. Following the addition of phenol (Sigma, P4557), the sample was gently shaken for 5-10 minutes at room temperature. The mixture was then centrifuged at 13,000 rpm for minutes, and the aqueous phase was transferred to a new tube.

After repeating this step with phenol:chloroform:isoamyl alcohol (Invitrogen, 15593-031), the sample was treated with 1 µl of RNaseA (Thermofisher, 12091-021) and incubated for one hour at 37°C. phenol:chloroform:isoamyl alcohol and chloroform: isoamyl alcohol (Sigma, C0549-1PT) were used to repeat the separation processes. Each time, the upper phase was transferred to a fresh tube and measured using a P1000 pipette. The sample was then precipitated overnight with 0.1 volume of 3M pH 5.2 sodium acetate (Panreac applichem, A3947,0100) and 2 volumes of 100% ice-cold ethanol (Fisher Scientific, BP2818-500). The DNA pellet was rinsed three times with 75% cold ethanol, air-dried, and dissolved in nuclease-free water. Qubit and Qubit™ dsDNA HS Assay Kit (Invitrogen,

Q32851) were utilised to measure the concentration of the DNA samples, while gel electrophoresis and Nanodrop were used to examine the purity of DNA samples. The genomic DNA was then stored at -20°C.

3.2.4 Methyl-seq library preparation and Illumina sequencing

The NEBNext Enzymatic Methyl-seq (EM-seq, New England BioLabs, E7120S) kits was utilised to construct Methyl-seq libraries. Before generating the libraries, purified genomic DNA and spike-in DNA were combined and sonicated using Covaris S2 to create 200–300 bp fragments. The EM-seq kit contains both unmethylated lambda phage DNA and methylated plasmid pUC19 DNA at CpG sites. These were used as negative and positive experiment controls, respectively. The libraries were prepared following the manufacturer's instructions, with 200ng of DNA inputs and 8 cycles of PCR amplification. The libraries were then sent to Novogene (Novogene UK Limited) and sequenced with the Illumina NovaSeq system (IIlumina, Inc. USA).

3.2.5 Methyl-seq read alignment and analysis

Raw sequencing reads were processed using TrimGalore (0.6.6) with default parameters to trim low-quality bases and adaptor sequences, and FastQC was used to evaluate the data quality. The clean reads were then mapped to the *Parhyale* Phaw5.1 genome reference using the bismark function of Bismark (0.22.3; Krueger & Andrews, 2011) with the following parameters: -N 1 -L 22 --score_min L,0, -0.6. Duplicate reads were removed using the deduplicate_bismark function, and bismark_methylation_extractor function was used to extract the methylation profile and to generate a genome-wide cytosine report indicating the number of methylated and unmethylated CpG reads at each CpG site in the genome. Only the methylation profiles of CpG sites were used in subsequent analyses.

3.2.6 Methyl-seq data statistical analysis

The calculation of CpG methylation values and genome-wide DNA methylation was identical to the method described in Chapter 2. Only CpG sites with at least five reads coverage were included in the genome-wide methylation and subsequent analyses. Effective CpGs (≥ 5 coverage) with methylation values of at least 20% were defined as methylated. After obtaining fundamental statistics for the six samples, the Methykit pool function (Akalin et al., 2012) was utilised to combine replicates within groups and produce a representative sample for each group.

3.2.7 Methylation level of different genomic features

The methylation level of a genomic feature was determined by dividing the total number of methylated CpGs (coverage ≥ 5 and methylation value ≥ 20%) by the total number of CpGs (coverage ≥ 5) in that region. The fractional methylation of a genomic region was calculated by averaging the methylation values of all CpGs (coverage ≥ 5) within that region. Bedtools map was utilised to acquire the total number of CpGs, mCpGs, and fractional methylation of a particular region. Using the annotations for protein-coding genes and repetitive elements, CpGs were classified as residing in gene-related or repeatome-related regions, respectively.

To profile DNA methylation patterns of the gene bodies, upstream (TSS - 2kb), downstream (TES + 2kb), and the exons and introns, each region was partitioned into a fixed number of windows (gene body: 100 windows, upstream/downstream: 20 windows, exon: 30 windows, intron: 80 windows) using Bedtools makewindows. The fractional methylation of each window was then calculated using Bedtools map. The same method was used to analyse the distribution of fractional methylation around TSS (TSS-2kb to TSS+4kb).

Similar to other invertebrate species, the fractional methylation of *Parhyale* gene bodies exhibits a bimodal distribution; consequently, a cut-off value for identifying methylated genes was manually selected based on the criteria used in methylation studies in other invertebrates. Based on the observed fractional methylation distribution and the CpG coverage of the gene bodies, methylated genes in each sample were defined as genes with at least five covered CpGs and a fractional methylation of at least 3% (Figure 11A, B).

3.2.8 Methylation analysis of repetitive elements

Exon, intron, and intergenic sequences were separated into repetitive sequences and non-repetitive sequences to examine the varying repeat methylation within the three genomic features. Bedtools intersect was used to extract repetitive sequences, and Bedtools subtract was used to extract non-repetitive sequences. The fractional methylation of the repetitive sequences and non-repetitive sequences of the three regions was then calculated using the same method described previously.

In addition, the entire repetitive elements were intersected with genes to define repeat-containing genes (Gene-Repeat) for the correlation analysis between expression and methylation. Genes that contain at least one repetitive element are referred to as Gene-Repeat.

To profile the DNA methylation patterns of several repetitive elements and the flanking upstream and downstream 2 kb regions, each element was partitioned into a fixed number of windows (repeat body: 40 windows, upstream/downstream: 10 windows) using Bedtools makewindows. The fractional methylation of each window was then calculated using Bedtools map.

3.2.9 Correlation analysis of DNA methylation and gene expression

In addition to this DNA methylation analysis, my colleague Dr. Salha Rawas extracted RNA from embryos at eight different stages between 0 and 60 hours and performed RNA-seq and total RNA-seq on these samples. These stages include 0-9 hour (9-hour), 11-hour, 12-hour, 14-hour, 16-hour, 20-hour, 24-hour, and 24-60 hour. Dr. Manuel Jara Espejo analysed these data and obtained gene expression results for these samples. Initially, the expression levels of methylated and unmethylated genes were compared. Genes were then classified into four groups according to their TPM (transcript per million) value (TPM = 0, TPM <1, TPM <5, TPM ≥ 5) for each RNA-seq time point. The relationship between DNA methylation and gene expression was examined by combining the expression results of the genes in the four groups with the methylation of gene bodies, introns, and exons at the same or adjacent stages of development (S1 methylation: expression level of 9-hour, 11-hour, 12-hour; S2 methylation: expression level of 11-hour, 12-hour, 14-hour, 16-hour, 20-hour; S3 methylation: expression level of 24-hour, 24-60-hour).

3.2.10 DNA methylation and ZGA gene expression analysis

Dr. Manuel Jara Espejo analysed the RNA-seq datasets and discovered the minor and major waves of zygotic gene activation (ZGA) of *Parhyale* embryos. The minor wave occurred between 11-hour and 14-hour, while the major wave began at 16-hour. The minor and major wave genes were subdivided further into zygotic-only and zygotic-maternal genes. Zygotic-only genes were those activated during minor or major waves but not expressed before the MZT period, whereas zygotic-maternal genes were those expressed maternally and also activated during minor/major waves. Maternal genes were defined as genes with a TPM value greater than three in the 9-hour RNA-seq dataset. The expression levels of genes in these three groups (maternal, zygotic-maternal,

zygotic-only) were then compared to the corresponding gene body methylation to investigate how DNA methylation affected the activation of zygotic genes.

3.2.11 Statistical analysis

The gene body/exon/intron fractional methylation of genes from different RNA-seq datasets is shown by box plot and the statistical difference was analysed using either the Wilcoxon test or Kruskal-Wallis's test. Correlation of gene body fractional methylation between different stages was tested at a 95% confidence interval using a one-tailed Pearson test. A p-value ≤ 0.05 was considered significant. Data were analysed and graphs were plotted using R.

3.3 Results and discussion

The analyses of methylation across the entire genome presented in Chapter 2 confirmed the presence of DNA methylation in the *Parhyale* genome. However, both WGBS and PacBio SMRT sequencing yielded low mapping rates. In this chapter, I performed EM-seq on samples from three different early developmental stages to investigate the dynamics of DNA methylation during the early development of *Parhyale*.

Specifically, total genomic DNA was extracted from stage 1 (S1), which includes 1-cell to 8-cell embryos (0-9 hour, representing the maternally driven embryonic development period); stage 2 (S2), which comprises 32-cell to 100-cell embryos (11-12 hour, the stage thought to be the time in which MZT occurs in *Parhyale*); and stage 3 (S3), which encompasses 24-60 hour embryos (covering the gastrulation to germ disc period) (Figure 4). The genomic DNA was enzymatically treated, libraries were prepared, sequenced, and aligned to the recently assembled *Parhyale* reference genome Phaw5.1. Each developmental stage was analysed in two biological replicates, resulting in a total of six sequencing samples.

The selection of these three stages was informed by the work of Nestorov et al, (2013), who used an antibody to detect the Ser2-phosphorylated form of RNA polymerase in early *Parhyale* development. Their research demonstrated that the MZT begins at the 32-cell stage and is followed by fully active transcription by the 100-cell stage.

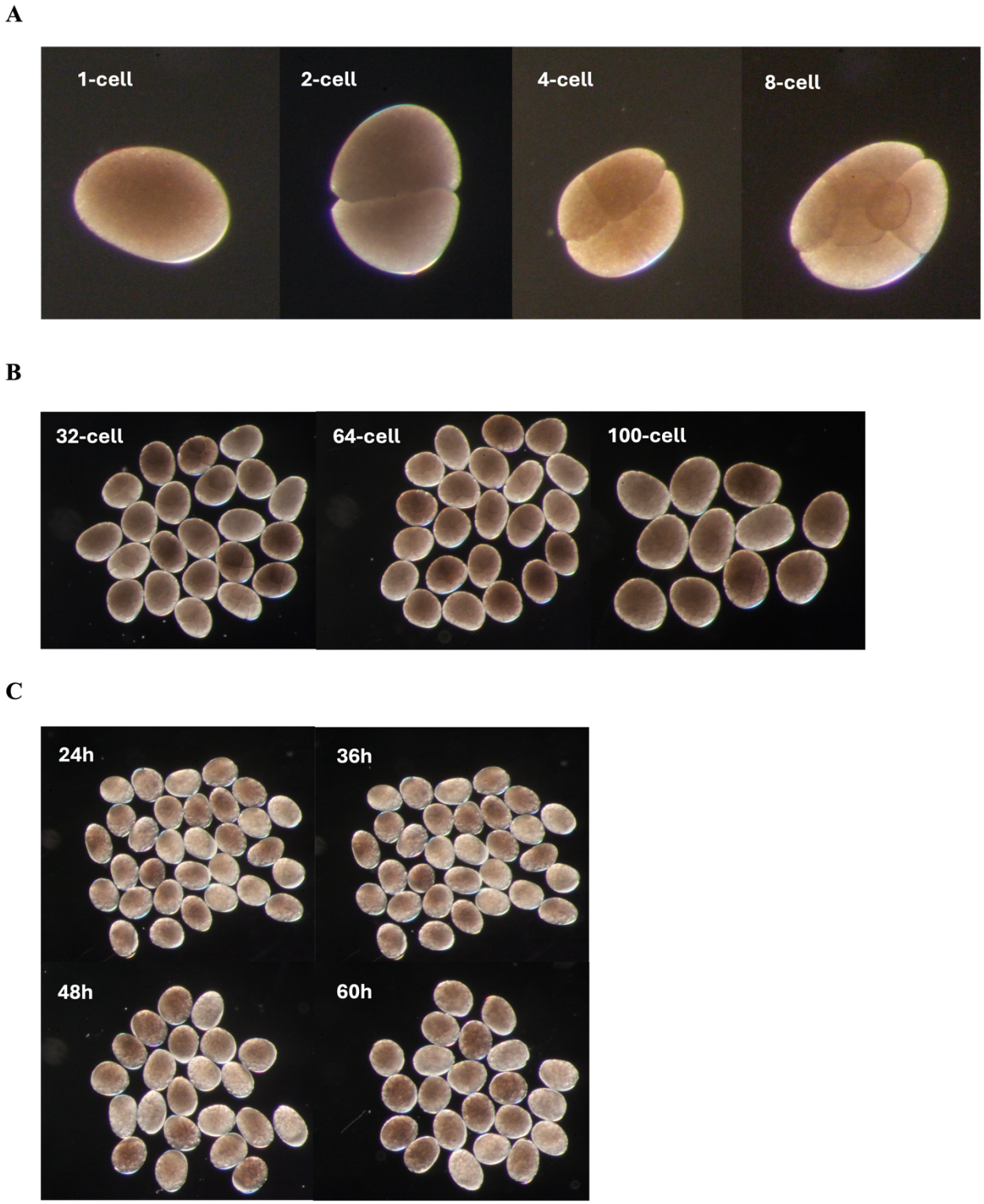


Figure 4. *Parhyale* embryos at the three chosen developmental stages. **(A)** S1 embryos **(B)** S2 embryos **(C)** S3 embryos

3.3.1 Summary of sequencing data

The *Parhyale* genome is approximately 3.6 Gb in size, and sequencing six samples generated a total of 154 Gb of sequence data. After trimming, approximately 99% of the clean reads were retained. Among these, about 66% of the clean reads from the six sequencing samples were uniquely aligned to the *Parhyale* genome, resulting in an average coverage of 11X. The mapping efficiency is 1.4 to 1.6 folds better than that observed with the previous WGBS data (Table 1).

Table 1. EM-seq results

Sample	Sequence pairs	Uniquely mapped	Non-uniquely mapped	Total mapping rate	Coverage
S1-1	184,864,864	73,476,138 (39.7%)	54,214,056 (29.3%)	69.0%	6.1X
S1-2	197,717,659	105,632,536 (53.4%)	61,587,612 (31.1%)	84.5%	8.8X
S2-1	179,010,729	104,250,426 (58.2%)	44,369,327 (24.8%)	83.0%	8.7X
S2-2	243,115,319	140,554,458 (57.8%)	58,193,406 (23.9%)	81.7%	11.7X
S3-1	206,532,452	130,582,277 (63.2%)	38,479,213 (18.6%)	81.8%	10.9X
S3-2	168,858,831	103,474,619 (61.3%)	30,319,825 (18%)	79.3%	8.6X

Generally, short-read methylation sequencing experiences low uniquely mapping rates, ranging from 40% to 60%, compared to regular short-read sequencing methods like RNA-seq, which usually achieve above 80% (Krueger & Andrews, 2011; Porter et al., 2015;

Tran et al., 2014). Therefore, the uniquely mapping rates of the six EM-seq samples are considered acceptable, especially in comparison to other bisulfite sequencing studies of invertebrates, which typically range from 28% to 53% (Olson & Roberts, 2014; Silliman et al., 2023; Trigg et al., 2022; Venkataraman et al., 2020). Notably, there were reads that did not map uniquely across all samples. When combining these reads with uniquely mapped ones, the overall mapping rate ranged from 69% to 84.5%. Subsequently, the uniquely mapped sequence pairs underwent deduplication for methylation analysis.

To assess the conversion efficiency of the enzymatic treatment, spikes of CpG-methylated pUC19 and unmethylated lambda DNA were used. The conversion rate for lambda DNA was greater than 99%, while the pUC19 conversion rate was less than 1.5% across all samples (Table 2). This demonstrated that the method had an exceptionally high conversion rate, validating the dataset's high quality.

Table 2. EM-seq spike-in DNA conversion rate

Sample	Conversion rate (lambda)	Conversion rate (pUC19)
S1-1	99.1%	0.5%
S1-2	99.0%	1.2%
S2-1	99.4%	1.4%
S2-2	99.4%	1.4%
S3-1	99.0%	0.1%
S3-2	99.3%	0.1%

There are 195,470,590 CpGs in the *Parhyale* genome, and approximately 74% of these CpGs had at least one read in the EM-seq dataset on average. The CpG coverage for each sample is shown in Table 3. The methylation value of each CpG was calculated by dividing the number of methylated reads by the total number of reads covering that CpG.

Table 3. CpG coverage of the six samples.

	CpGs	%	CpGs	%
Total	195,470,590	-	195,470,590	-
	S1-1		**S1-2**	
Coverage =0	85,316,253	43.65%	54,421,711	27.84%
Coverage ≥ 1	110,154,337	56.35%	141,048,879	72.16%
Coverage ≥ 2	55,702,631	28.50%	97,450,894	49.85%
Coverage ≥ 5	5,464,356	2.80%	21,134,131	10.81%
Coverage ≥ 10	988,774	0.51%	2,656,531	1.36%
	S2-1		**S2-2**	
Coverage =0	47,827,240	24.47%	37,569,048	19.22%
Coverage ≥ 1	147,643,350	75.53%	157,901,542	80.78%
Coverage ≥ 2	108,918,562	55.72%	127,503,468	65.23%
Coverage ≥ 5	30,056,851	15.38%	51,151,091	26.17%
Coverage ≥ 10	3,677,638	1.88%	7,205,961	3.69%
	S3-1		**S3-2**	
Coverage =0	37,082,659	18.97%	43,881,405	22.45%
Coverage ≥ 1	158,387,931	81.03%	151,589,185	77.55%
Coverage ≥ 2	128,850,561	65.92%	116,512,734	59.61%
Coverage ≥ 5	54,395,237	27.83%	38,689,821	19.79%
Coverage ≥ 10	8,013,979	4.10%	4,792,039	2.45%

To obtain accurate quantification of methylation, CpG sites with less than five read coverage were excluded before determining the genome-wide methylation level. This exclusion resulted in a range of 3% to 28% of all CpGs remaining for the six samples in

the genome, as shown in Table 3. Methylated CpGs were defined as having a methylation value at least 20% (coverage ≥ 5 & methylation value ≥ 20%). Using this formula and criteria, the genome-wide methylation level and average methylation value were calculated for all six samples. The results revealed that approximately 7% to 15% of the effectively covered CpGs (coverage ≥ 5) were methylated, with average methylation values ranging from 3.6% to 8.1% for the six samples (Table 4).

Table 4. Genome-wide methylation of the six samples.

%	Genome-wide methylation level (%)	Genome-wide average methylation value (%)
S1-1	15.2	8.1
S1-2	9.5	5.2
S2-1	8.4	4.7
S2-2	7.1	4.1
S3-1	6.7	3.6
S3-2	9.4	3.9

The EM-seq data provided better sequencing depth and CpG coverage than the previous WGBS data, enabling a more detailed investigation of DNA methylation changes and their correlation with gene expression during the early development of *Parhyale*. To assess the reproducibility and consistency of the replicates within each group, I performed Principal Component Analysis (PCA) on methylation values for effectively covered CpGs (coverage ≥ 5) across the six samples (Figure 5). Each pair of replicates from S2 and S3 are positioned close to each other, indicating good reproducibility within these groups. However, replicates S1-1 and S1-2 are spaced further apart on the PCA plot compared to other pairs, suggesting potential issues with reproducibility or consistency. This discrepancy is likely due to the quality of the input DNA. Given that each embryo from this stage contained less than eight cells, the amount of DNA available for extraction was limited. Additionally, the process of combining embryos into a single tube during DNA

extraction was time-consuming, potentially leading to DNA degradation. Moreover, since both S1 samples were distant from S2 and S3 on the PCA plot, it is challenging to determine which replicates can more accurately reflect true biological conditions. Therefore, the replicates in each stage were pooled to increase the overall sample size for each group to enhance the statistical power of the analysis.

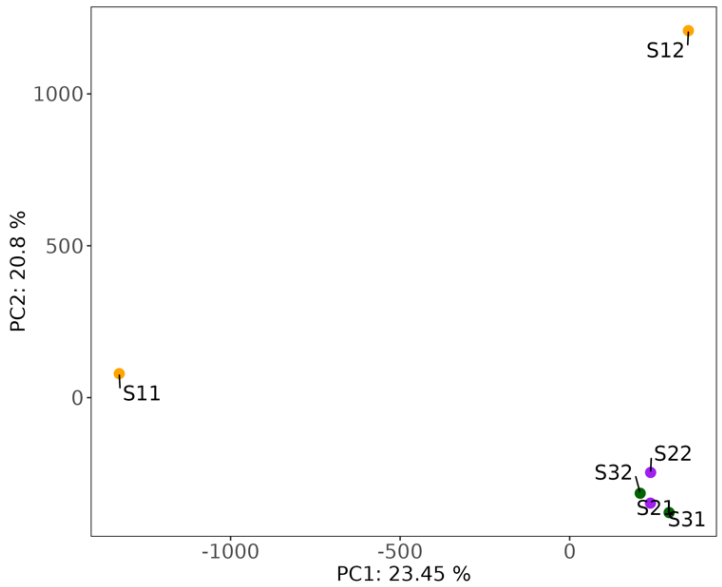


Figure 5. Principal component analysis of methylation values for CpGs with at least five coverage of the six samples. PC1, first principal component; PC2, second principal component. Each dot indicates a sample.

I utilised Methykit to combine replicates within each stage and generated a representative sample. This pooling increased the proportion of effective CpGs (coverage ≥ 5), from 17% to 45%, as indicated in Tables 3 and Table 5, making the number of effectively covered CpGs comparable with those covered in other bisulfite sequencing studies of invertebrates (Gavery & Roberts, 2013; Olson & Roberts, 2014; Silliman et al., 2023; Venkataraman et al., 2020; X. Wang et al., 2013). As a result of the increased coverage, fewer methylated CpGs and decreased average methylation values were observed in all three developmental stages (Table 6). For the downstream analyses, only CpGs with at least five reads coverage in each stage were retained.

The genome-wide methylation level observed in the EM-seq data is slightly lower than the level reported in a study on arthropod methylation. The arthropod study found a genome-wide methylation level of greater than 10% in *Parhyale* (Lewis et al., 2020). This may be because their samples were collected from adult *Parhyale*, and the previous chapter's analysis revealed that embryos have a lower methylation level than the whole-body tissue.

Table 5. CpG coverage after pooling replicates.

	S1		S2		S3	
	CpGs	**%**	**CpGs**	**%**	**CpGs**	**%**
Total	195,470,590	-	195,470,590	-	195,470,590	-
Coverage =0	26,113,479	13.36%	12,503,702	6.40%	11,818,240	6.05%
Coverage ≥ 1	169,357,111	86.64%	182,966,888	93.60%	183,652,350	93.95%
Coverage ≥ 2	136,042,651	69.60%	167,779,528	85.83%	169,738,225	86.84%
Coverage ≥ 5	46,646,777	23.86%	105,546,223	54.00%	112,404,751	57.50%
Coverage ≥ 10	5,768,227	3.41%	28,014,689	15.31%	34,683,345	18.89%

Notably, S1 samples exhibited the lowest coverage, with half the coverage of the other two stages (Table 5, Figure 6A). One possible reason for the lower coverage observed in the S1 samples could be attributed to the poor quality of the input DNA. However, the distribution of genome-wide methylated CpGs and the CpG coverage on various genomic regions of S1 were comparable to the other two stages; thus, the S1 sample is suitable for the subsequent analyses (Figure 6B).

The decision to pool the S1 data was pivotal in interpreting the results throughout this chapter. Pooling replicates from S1 increased the overall sample size for this group, thereby enhancing the statistical power of the analyses. This augmentation allowed for

more precise detection of subtle DNA methylation changes and their correlation with gene expression, potentially revealing insights that smaller sample sizes might have obscured. However, despite these benefits, pooling the S1 data introduces certain limitations that require careful consideration. Notably, the spacing between replicates S1-1 and S1-2 indicated in the PCA suggests reproducibility issues, which could compromise the reliability of the pooled dataset. Furthermore, including potentially inconsistent replicates might introduce noise or confounding factors, potentially obscuring true associations between DNA methylation and gene expression. Therefore, caution is necessary to account for potential sources of bias or variability that could impact the interpretation and conclusions of the study.

Table 6. Genome-wide methylation after pooling replicates.

%	S1	S2	S3
Genome-wide methylation level (%)	8.9	7.0	6.3
Genome-wide average methylation value (%)	4.5	3.7	3.4

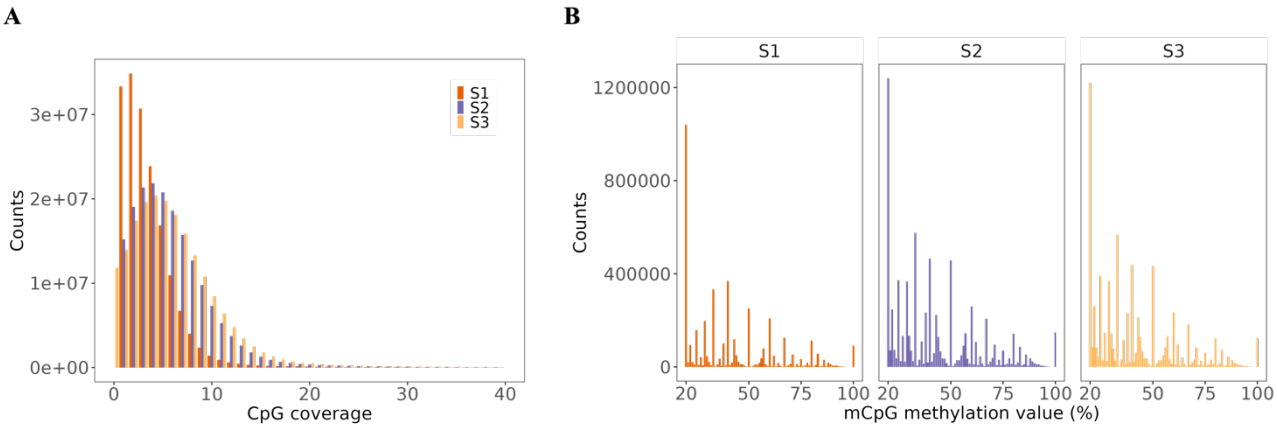


Figure 6. Genome-wide methylation patterns. **(A)** Genome-wide distribution of CpG coverage across the three developmental stages. **(B)** Distribution of the methylation value of mCpGs across the genome.

3.3.2 DNA methylation analysis

3.3.2.1 DNA methylation in different genomic regions

Consistent with the findings of Chapter 2, the three EM-seq embryo samples in this study displayed a comparable amount of methylated CpGs (mCpGs) across gene bodies and intergenic regions. The majority of the mCpGs within gene bodies were found in introns rather than exons (Figure 7A, C). To further investigate the methylation status of various genomic features in the *Parhyale* genome, I calculated the methylation level (the ratio of mCpGs to covered CpGs) and fractional methylation (the average methylation values of all covered CpGs) of selected genomic regions and classes of repetitive elements (Repeats). The selected genomic regions included gene body, exon, intron, 5'-UTR and 3'-UTR.

Similar patterns of methylation level and fractional methylation were observed in different genomic regions of the three samples. Methylation targeted both genes and repeats, with repeats exhibiting a slightly higher level of methylation than gene bodies. For the gene bodies, the methylation level of 5'-UTR and introns were found to be higher than exons and 3'-UTR. Furthermore, LTRs, although comprising only 4% of all covered CpGs within repeats (Figure 7B), exhibited the highest methylation level compared to other repeats. Notably, simple repeats and satellite DNAs, the only two classified non-TE repeats, also exhibited elevated methylation levels (Figure 7C). However, the reliability of this pattern may be questionable due to the very small representation of these two classes among all repeats. In addition, fractional methylation patterns differed slightly from methylation levels, with intergenic regions and LTRs exhibiting nearly twice the fractional methylation of other regions (Figure 7D). Overall, these methylation patterns were similar to those observed in the previously analysed PacBio sample.

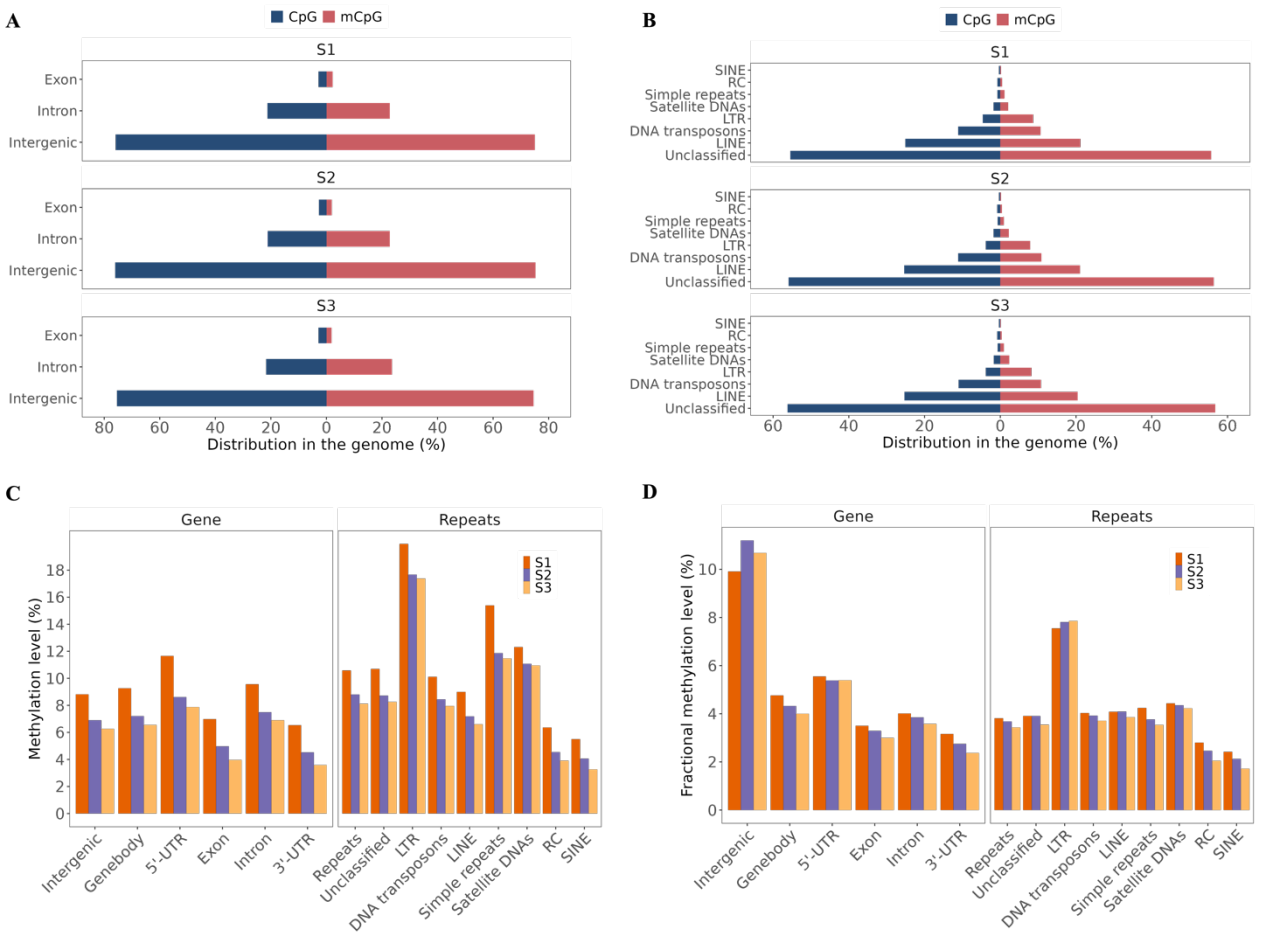


Figure 7. Methylation patterns of different genomic features. Comparison of the total CpG versus mCpG in the three stages by **(A)** genomic features and **(B)** repetitive elements (Repeats). **(C)** Methylation level and **(D)** fractional methylation over all covered CpGs mapped to several genomic features and repetitive elements of the three stages. Unclassified: unclassified repeats, RC: Rolling-circle transposons.

I then distinguished repetitive sequences from non-repetitive sequences for exons, introns, and intergenic regions to investigate the distribution of repeat methylation in these regions. I discovered that methylation level and fractional methylation varied between repetitive and non-repetitive sequences. In terms of methylation level, CpGs in repetitive sequences were more frequently methylated than CpGs in the respective non-repetitive background. Intronic repetitive sequences exhibited the highest methylation level, despite accounting for only 50% of total intron sequences. Repetitive sequences constituted 60% of all intergenic regions, and DNA methylation was observed to predominantly target these intergenic repetitive sequences (Figure 8).

Intriguingly, the methylation levels and fraction methylation are both higher in repetitive sequences, particularly those in non-coding regions, compared to the non-repetitive sequences across all samples (Figure 8A, B). These results indicate that repeats were the primary target of DNA methylation in the *Parhyale* genome, and that DNA methylation preferentially targeted repeats located in non-coding regions.

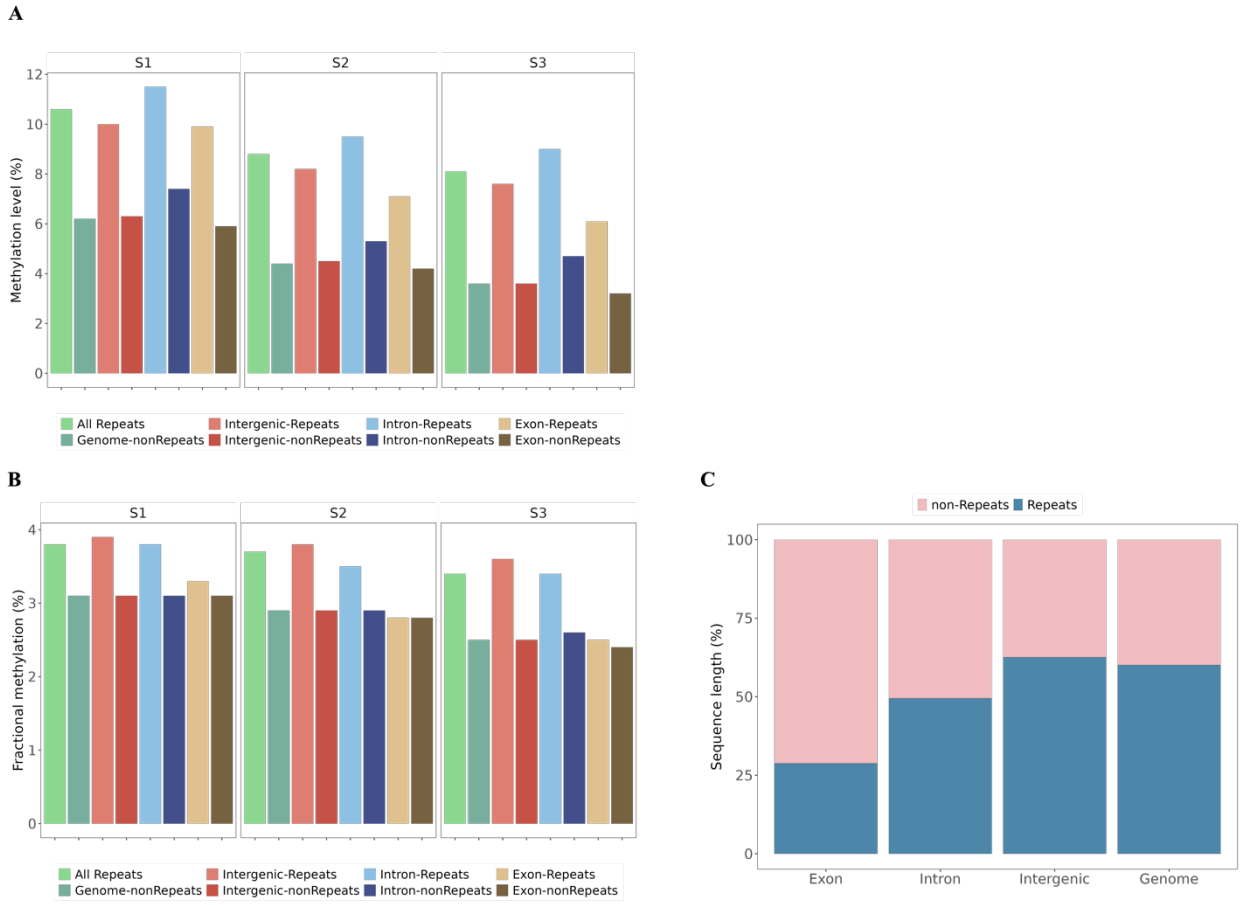


Figure 8. Methylation pattern of repetitive sequences **(A)** Methylation level and **(B)** fractional methylation of repetitive and non-repetitive sequences in different genomic regions. **'All Repeats':** all repetitive sequences in the genome; **'Genome: nonRepeats':** non-repetitive sequences in the genome; **'Intergenic-Repeats':** repetitive sequences in the intergenic regions; **'Intergenic-nonRepeats:** non-repetitive sequences in the intergenic regions; **'Intron-Repeats':** intronic repetitive sequences; **'Intron-nonRepeats':** intronic non-repetitive sequences; **'Exon-Repeats' :** exonic repetitive sequences; **'Exon-nonRepeats' :** exonic non-repetitive sequences **(C)** The proportion of the length of repetitive and non-repetitive sequences in the different genomic regions.

3.3.2.2 Gene body methylation

To examine the methylation patterns of gene bodies in *Parhyale*, I analysed the distribution of fractional methylation across the gene bodies of the three samples. Gene bodies were divided into windows of varying sizes, and fractional methylation was calculated for each window. The fractional methylation pattern remained consistent across the three samples, with a slight decrease occurring 500 bp prior to the transcription start site (TSS), followed by a sharp increase immediately after the TSS, peaking at 1500 bp within the gene body. Methylation levels then gradually decreased towards the 3'-end of the gene body (Figure 9A, B). The average length of the first exon of the gene body was less than 1000 bp (Figure 9D), indicating that the maximum level of methylation within the gene body occurs at the first intron. Additionally, a varying trend was identified in the fractional methylation of the exons and introns, where introns exhibited a greater fractional methylation than exons (Figure 9C). The methylation patterns of the gene bodies across the three developmental stages were consistent with those observed in the previously analysed PacBio adult sample but distinct from those seen in many other arthropods (Lewis et al., 2020). Nevertheless, the highest level of methylation was detected in the first intron of *Parhyale* genes.

Figure 10 depicts the relationship of gene body methylation between developmental stages. Overall, there were correlations between adjacent stages. The gene body methylation of S1 tended to be greater than that of S2, whereas the gene body methylation of S2 and S3 were comparable.

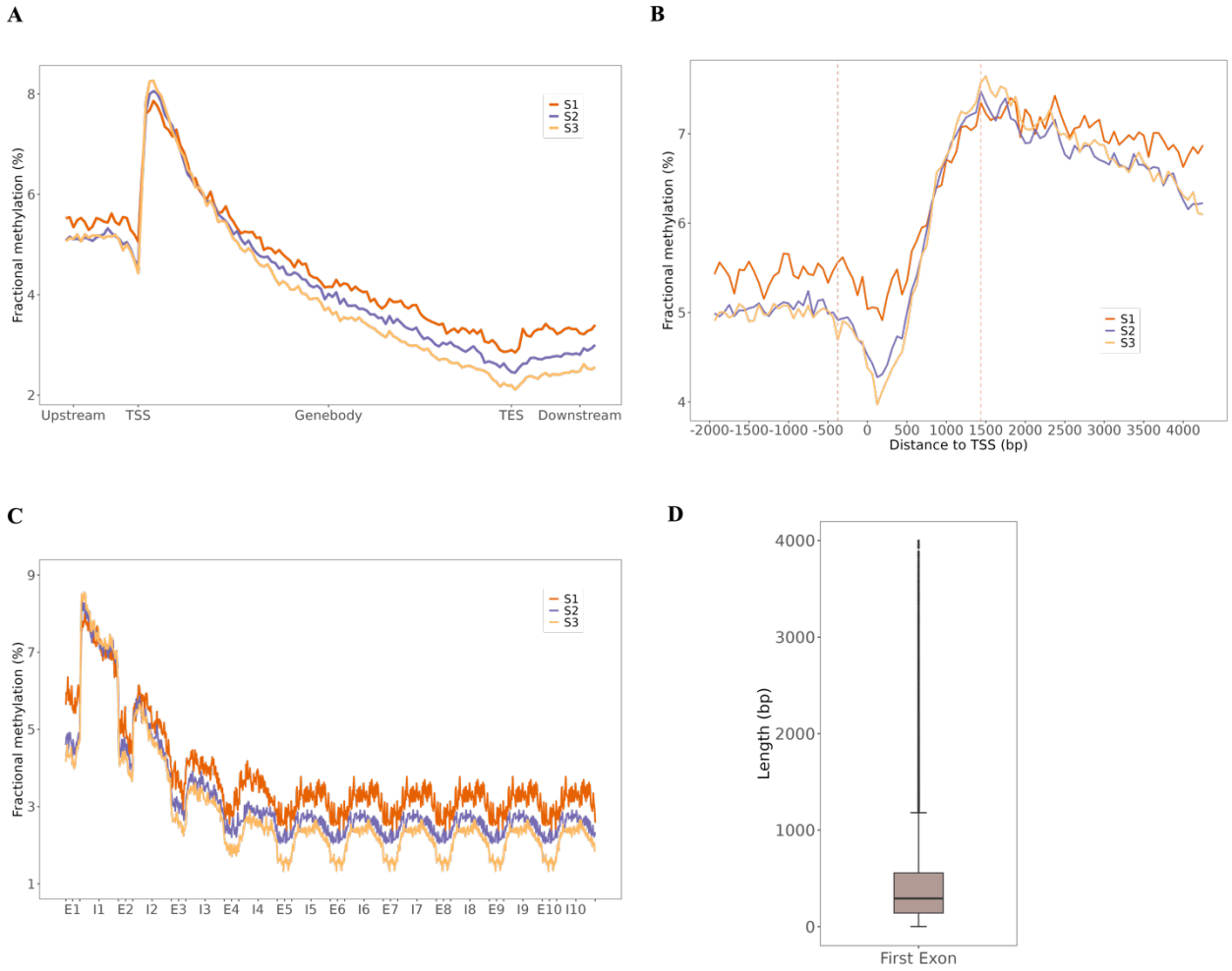


Figure 9. Gene body methylation pattern **(A)** Fractional methylation within gene body and 2 kb flanking (upstream and downstream) regions for all the protein-coding genes at the three stages. The gene body was divided into 100 bins with varied bin sizes, while both upstream and downstream were divided into 20 bins with bin sizes equal to 100 bp. **(B)** Pattern of fractional methylation of the flanking TSS regions (TSS-2kb to TSS+4kb). The entire region was divided into 100 bins of varying bin sizes **(C)** Pattern of fractional methylation in the first ten exons and introns of all protein-coding genes. Exons (E) and introns (I) were divided into 30 and 80 fixed number of bins with varied sizes, respectively. **(D)** The average length of the first exons in the gene body.

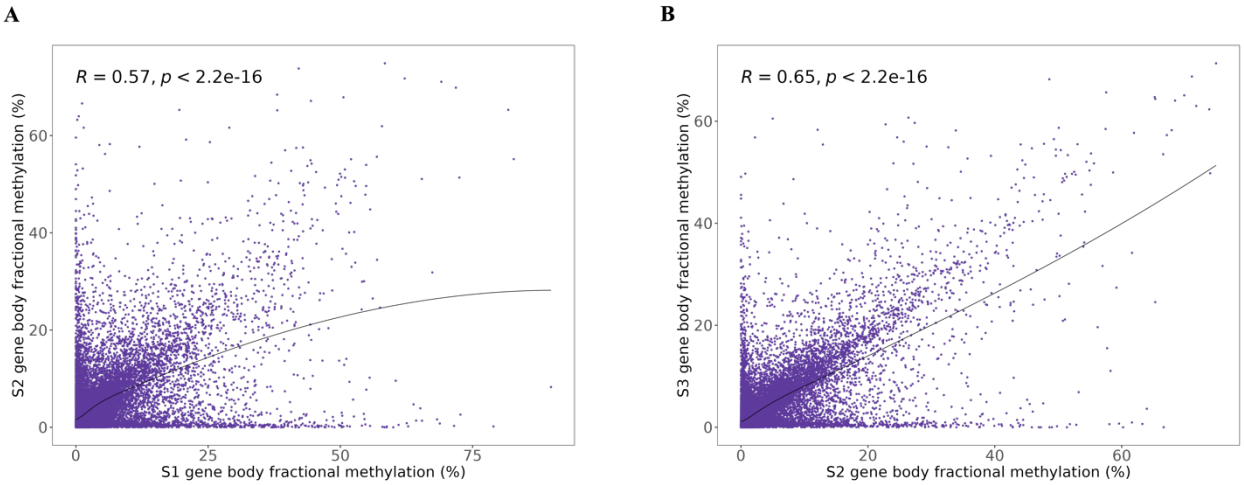


Figure 10. Correlation between **(A)** S1, S2 and **(B)** S2, S3 gene body fractional methylation.

The distribution of fractional methylation in the *Parhyale* gene bodies was found to be bimodal, with two peaks observed and the majority of genes exhibiting low levels of methylation (Figure 11A). Similar bimodal patterns have been observed in other invertebrates' genome-wide methylation study (Arsala et al., 2022; Sarda et al., 2012; Zemach et al., 2010). Interestingly, the bimodal pattern tended to strengthen as development progressed in *Parhyale* embryos. Approximately 10% of the genes exhibited no methylation, and most genes had fewer than 50 covered CpGs (Figure 11B). Based on the observed distribution pattern and CpG coverage of the gene body, as well as the criteria used in the methylation studies of other invertebrates, methylated genes were defined as those with at least five covered CpGs and 3% fractional methylation (represented by the dashed line in Figure 11A, B). Applying these criteria, approximately 7000 methylated genes were presented in each stage.

The analysis of overlapping methylated genes between the three stages revealed that approximately 21% (4,477) of *Parhyale*'s total protein-coding genes were consistently methylated in all stages. The number of overlapping methylated genes was higher between S2 and S3 than between S1 and S2. Moreover, the number of stage-specific methylated genes decreased during the course of development (Figure 11C). These

findings suggest that there was a general decline in gene body methylation during the early embryonic development of *Parhyale*.

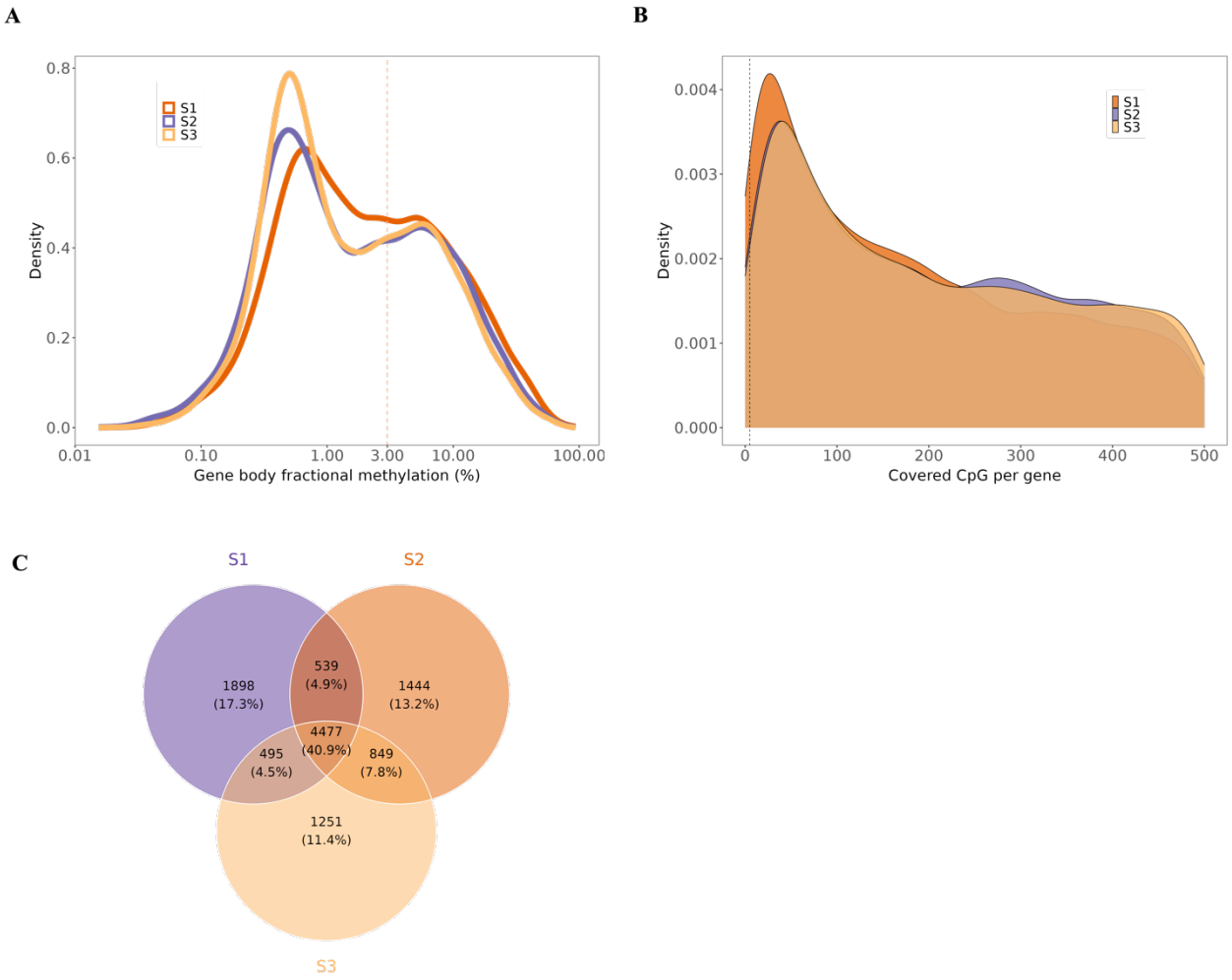


Figure 11. Gene body methylation distribution and methylated genes. **(A)** The fractional methylation distribution for all genes. The vertical dashed line represents the methylation threshold used to define methylated genes (3%). **(B)** Coverage distribution of CpGs per gene. The vertical dashed line represents the minimum gene retention threshold for methylated gene selection (Covered CpGs per gene ≥ 5). **(C)** Overlap of methylated genes (fractional methylation ≥ 3% & covered CpGs ≥ 5) of the three samples.

3.3.2.3 Methylation of repetitive elements

To assess DNA methylation levels and patterns across various classes of repetitive elements in the *Parhyale* genome, the distribution of methylation was examined across all repetitive elements and each class for all three stages (Figure 12), along with the

correlation of their fractional methylation between stages (Figure 13). For TE-specific class analysis, three classified TE classes (LTR, LINE, and DNA transposons) that were primarily enriched for mCpG were selected. Generally, repeats exhibited higher methylation levels than gene bodies. Repeats also show a general decline in methylation during development, which is comparable to gene body methylation (Figure 12,13). Of those classified TE classes, LTRs showed the highest fractional methylation (Figure 12E).

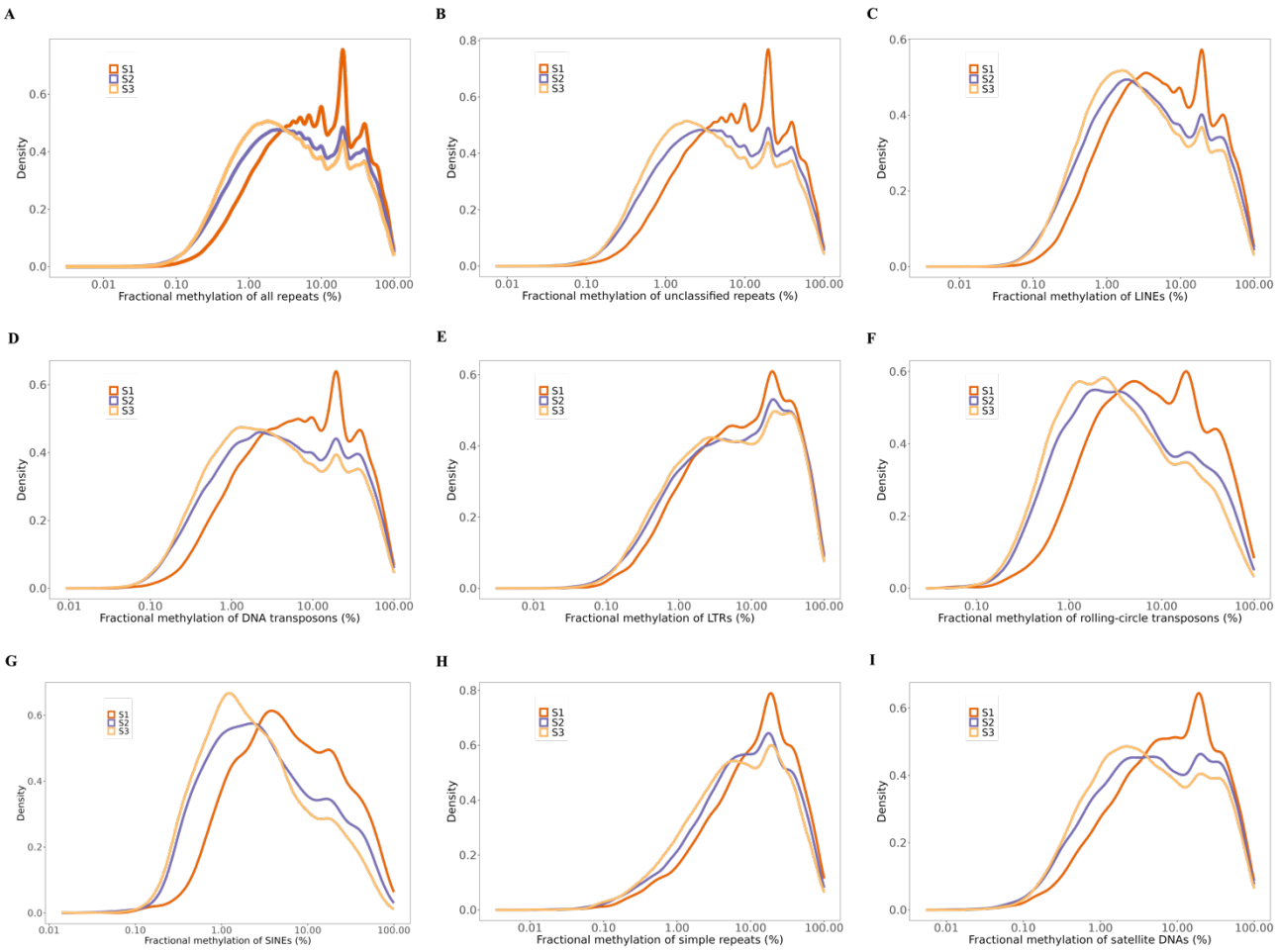
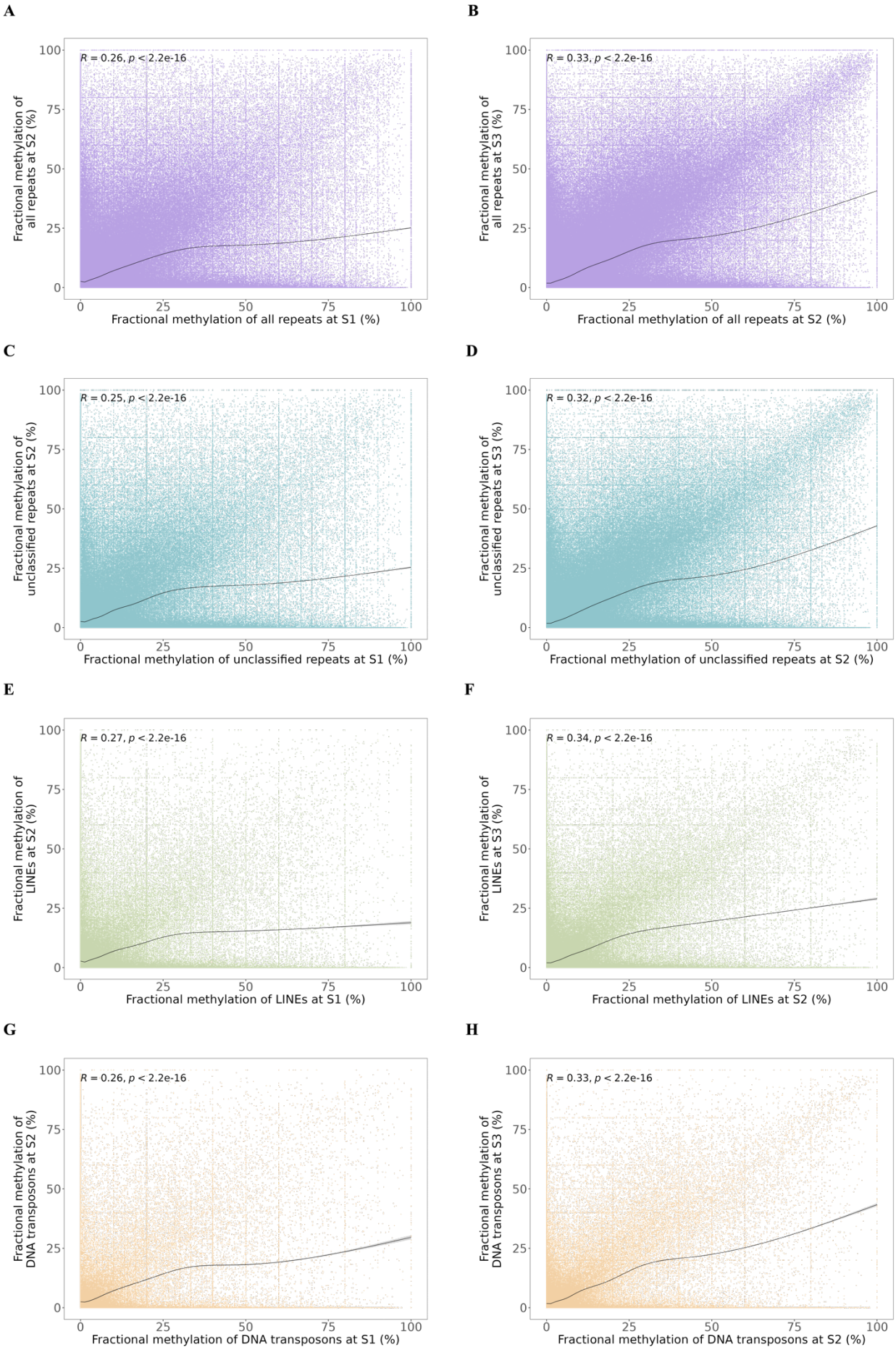


Figure 12. Distribution of fractional methylation for **(A)** all repeats, **(B)** unclassified repeats, **(C)** LINEs, **(D)** DNA transposons, **(E)** LTRs, **(F)** rolling-circle transposons, **(G)** SINEs, **(H)** simple repeats and **(I)** satellite DNAs.


A
R = 0.26, p < 2.2e-16
Fractional methylation of all repeats at S2 (%)
Fractional methylation of all repeats at S1 (%)
B
R = 0.33, p < 2.2e-16
Fractional methylation of all repeats at S3 (%)
Fractional methylation of all repeats at S2 (%)
C
R = 0.25, p < 2.2e-16
Fractional methylation of unclassified repeats at S2 (%)
Fractional methylation of unclassified repeats at S1 (%)
D
R = 0.32, p < 2.2e-16
Fractional methylation of unclassified repeats at S3 (%)
Fractional methylation of unclassified repeats at S2 (%)
E
R = 0.27, p < 2.2e-16
Fractional methylation of LINEs at S2 (%)
Fractional methylation of LINEs at S1 (%)
F
R = 0.34, p < 2.2e-16
Fractional methylation of LINEs at S3 (%)
Fractional methylation of LINEs at S2 (%)
G
R = 0.26, p < 2.2e-16
Fractional methylation of DNA transposons at S2 (%)
Fractional methylation of DNA transposons at S1 (%)
H
R = 0.33, p < 2.2e-16
Fractional methylation of DNA transposons at S3 (%)
Fractional methylation of DNA transposons at S2 (%)

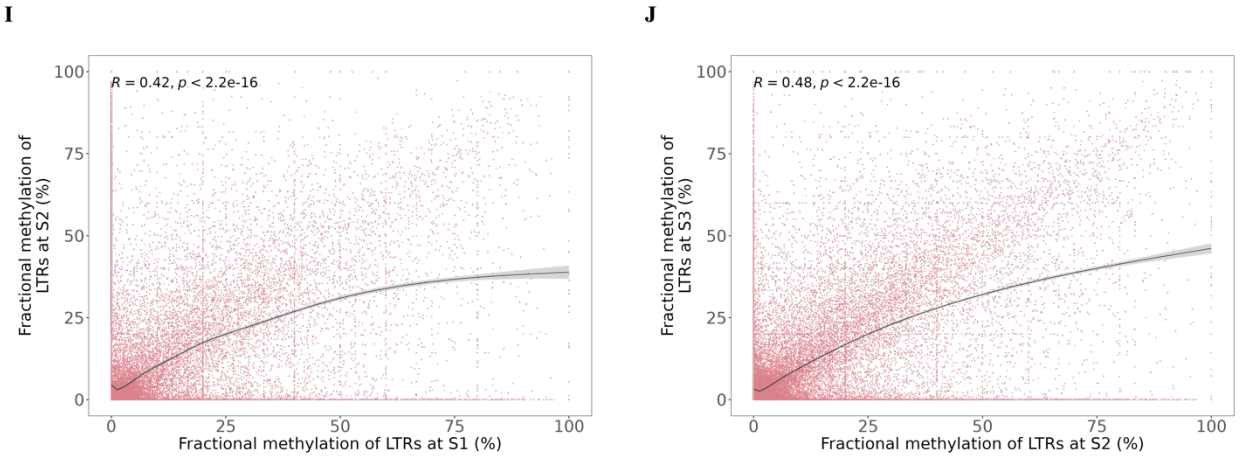


Figure 13. Correlation of fractional methylation of **(A,B)** all repeats, **(C,D)** unclassified repeats, **(E,F)** LINEs, **(G,H)** DNA transposons and **(I,J)** LTRs between the two adjacent stages.

To determine the fractional methylation patterns of repeats in the *Parhyale* genome, three major classes of TEs and unclassified repeats, along with their 1 kb flanking regions, were divided into a fixed number of windows of variable sizes. The fractional methylation of each window was then calculated. Generally, these repeats showed higher levels of methylation compared to adjacent regions (Figure 14). However, distinct differences were observed when comparing the methylation profiles across the different TE classes. Specifically, LTRs exhibited elevated methylation at the 5' end of TE bodies, which slightly decreased towards the 3' end. In contrast, LINEs showed relatively similar methylation levels between their bodies and flanking regions. The unclassified repeats and DNA transposons demonstrated a pattern similar to that of the overall repeats, with DNA transposons having slightly decreased methylation between 5' and 3' end of their bodies (Figure 14B-E). These findings suggest that the fractional methylation patterns of TEs are similar across different classes.

Over 90% of genes within the three sampled stages contain at least one repeat (Figure 15A). This prompted me to investigate whether there is a correlation between the methylation of repeats and the gene bodies they resided in. By comparing the fractional methylation levels of repeats and their host genes, I found a strong positive correlation,

suggesting that the methylation of genes is predominantly influenced by the methylation of the repeats resided within their gene bodies (Figure 15B-D).

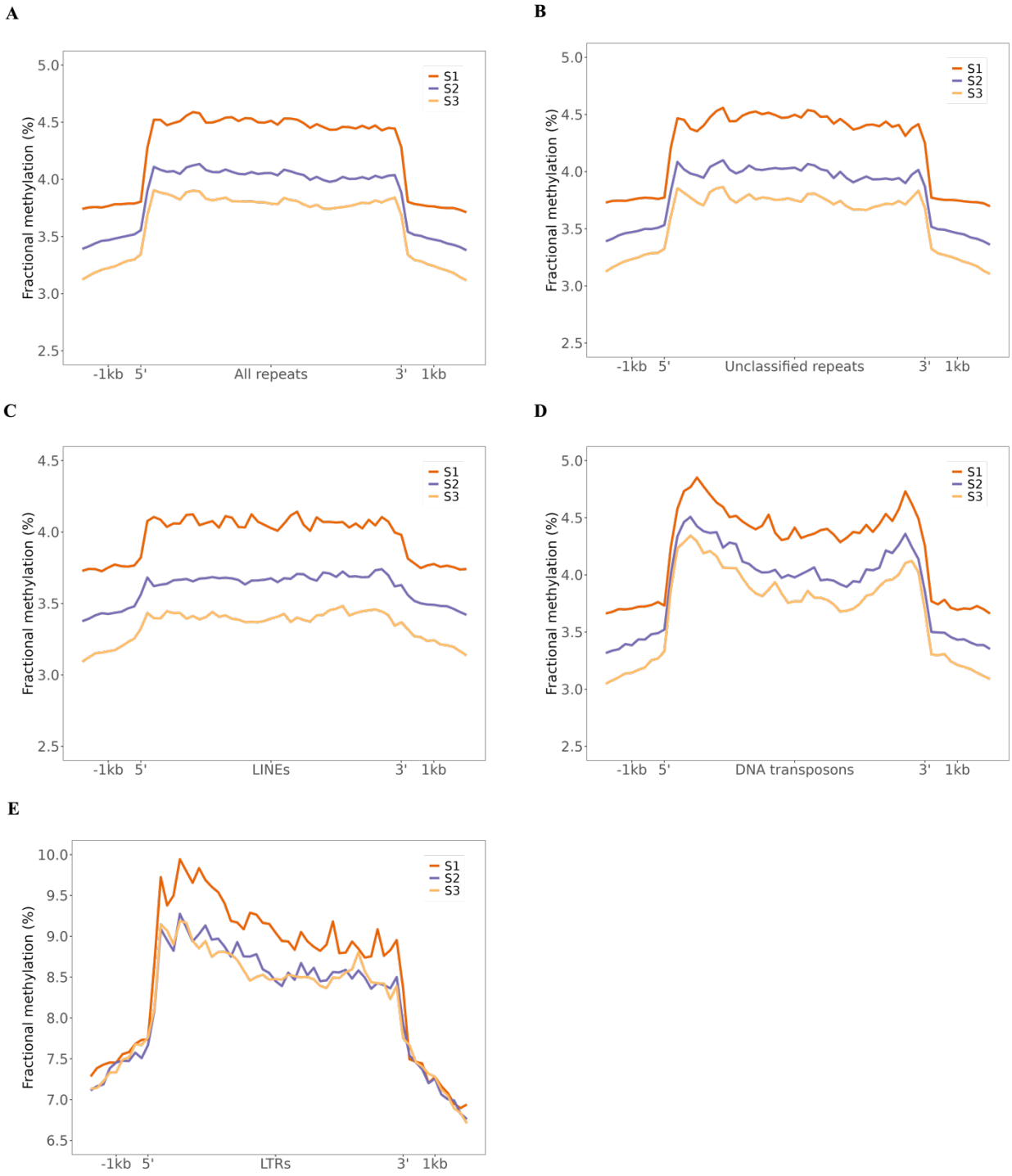


Figure 14. Methylation pattern of repeats. Distribution of fractional methylation within repeat bodies and the corresponding 1 kb adjacent (upstream and downstream) regions for **(A)** all covered repeats, **(B)** unclassified repeats, **(C)** LINEs, **(D)** DNA transposons, and **(E)** LTRs of the three developmental stages of *Parhyale*. Repeat body was divided

into 40 bins of varying sizes, whereas the upstream 1 kb and downstream 1 kb were both divided into 10 bins of the same size.

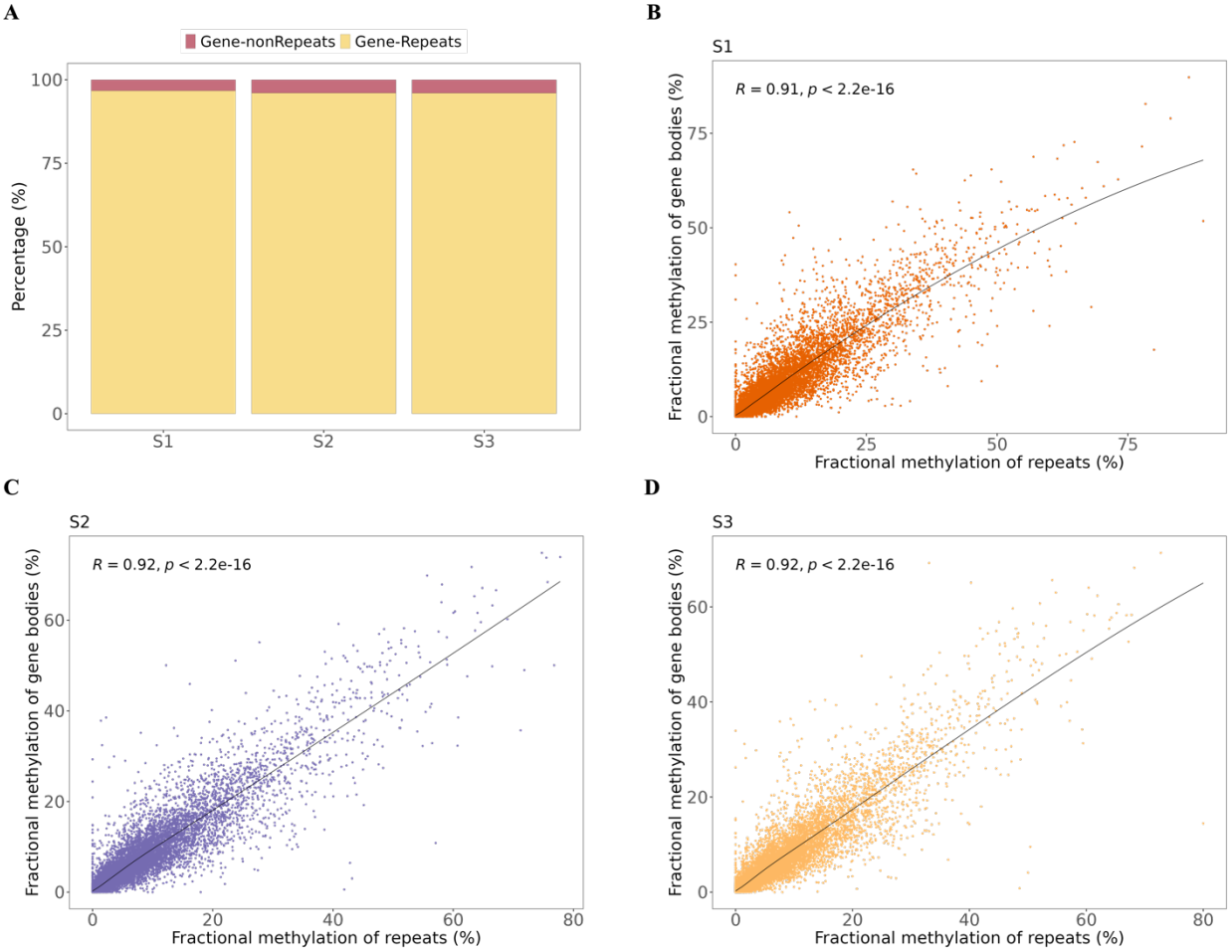


Figure 15. Methylation of repeats and the host genes **(A)** The proportion of repeat-containing and repeat-deficient genes at each of the three stages. **(B-D)** Correlation of methylation between repeats and their host genes at each stage. **Gene-nonRepeats:** genes that do not contain repeats in their gene bodies. **Gene-Repeats:** genes that contain repeats in their gene bodies.

3.3.3 Integrative analysis of EM-seq and RNA-seq data

3.3.3.1 Correlation between gene body methylation and gene expression

An integrated analysis of DNA methylation and gene expression at the temporally associated stages during early development was conducted to investigate their correlation. First, genes were categorised based on the definition of methylated genes

(CpG ≥ 5 and fractional methylation ≥ 3%). Then, methylation and gene expression data were merged from the same stage for each gene to compare the gene expression TPM values between methylated and unmethylated genes. The analysis revealed that moderately expressed genes are the main targets of gene body methylation, while low-expression genes are predominantly unmethylated. Additionally, methylated genes in all three stages exhibited higher expression levels than unmethylated genes (Figure 16).

In addition to the aforementioned three developmental stages, my colleague Dr. Salha Rawas also generated RNA-seq data from embryos at five additional developmental stages (11-hour, 14-hour, 16-hour, 20-hour, and 24-hour). After obtaining the gene expression results for the total eight stages, I divided the genes at each stage into four categories based on their TPM values: No (log2TPM = 0), Low (log2TPM<1), Median (log2TPM<5), and High (log2TPM ≥ 5), and correlated these expression levels with the fractional methylation of the corresponding genes at the same or adjacent stages to determine if there is a relationship between DNA methylation and gene expression. S1 methylation is correlated with gene TPM values at 0-9-hour, 11-hour and 12-hour, S2 methylation is correlated with gene TPM values at 11-hour, 12-hour, 14-hour, 16-hour, and 20-hour; and S3 methylation is correlated with gene TPM values at 24-hour and 24-60 hour.

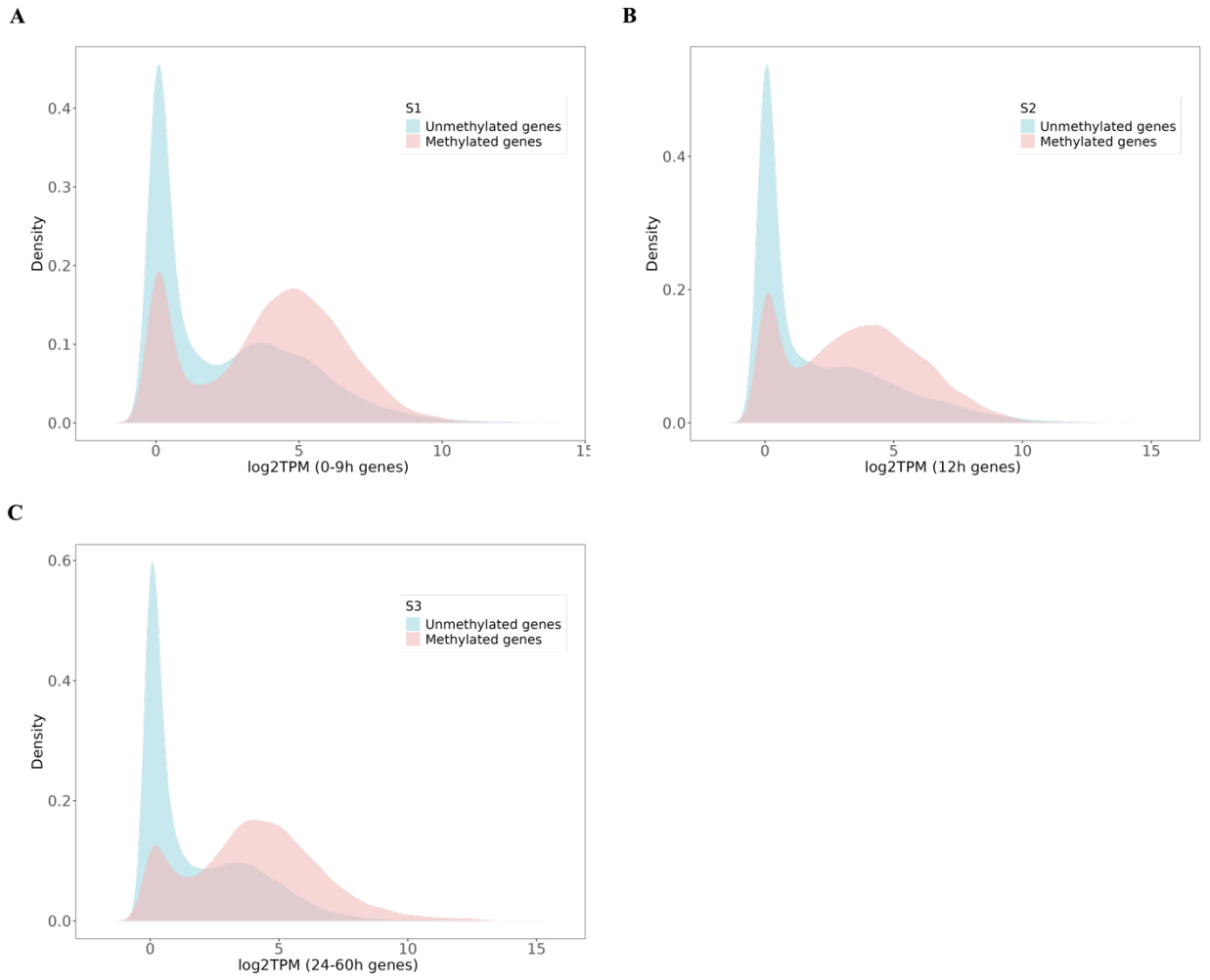


Figure 16. Gene expression level (log2TPM) for methylated and unmethylated genes in **(A)** S1, **(B)** S2, and **(C)** S3. TPM: transcript per million.

The correlations are shown in Figure 17; gene body methylation and gene expression were positively correlated across all time point comparisons, with highly expressed genes (TPM ≥ 5) having a tendency to be more methylated in the gene body than low or non-expressed genes (Figure 17A,C,E). In addition, the No and Low expression groups were predominantly composed of unmethylated genes, whereas over half of the highly expressed genes were methylated (fractional methylation ≥ 3%) (Figure 17G). These results indicate that unmethylated genes are typically not expressed, whereas increased transcriptional activity is generally accompanied by a higher methylation level in the gene body. This positive correlation between gene body methylation and gene expression has also been found in other invertebrates (Dimond & Roberts, 2015; Gavery & Roberts, 2013;

Ying et al., 2022; Zemach et al., 2010). These findings imply that DNA methylation could potentially play a role in regulating gene expression during the early embryonic development of *Parhyale*.

A recent study by Ying et al, (2022) identified a positive correlation between gene body methylation and the presence of TEs in invertebrates. This led to the hypothesis that the positive correlation between gene body methylation and gene expression level observed in this study was a consequence of relatively high TE/repeat contents within those genes. To explore this possibility, I analysed the relationship between the exon/intron methylation of each gene and the expression TPM values at the same or adjacent stages of development. The results showed that intron methylation was substantially greater than exon methylation of all the genes, while the overall intron methylation is higher in highly expressed genes compared to lowly expressed genes (Figure 18A-J). In addition, it was found that highly expressed genes contain a notably greater number of repeat-containing introns than lowly expressed genes (Figure 18K). Therefore, the difference in intron methylation between genes of different expression groups was attributed to the varying proportions of repeat-containing introns among these groups.

Comparing the methylation of those intron which contain repeats among the four expression groups at each RNA-seq time point. I observed that methylation of repeat-containing introns was consistently greater in the High expression group than in the No or Low expression group (Figure 19). These results suggest that the increased methylation of intronic repeats is a major contributor to the higher overall methylation observed in highly expressed genes.

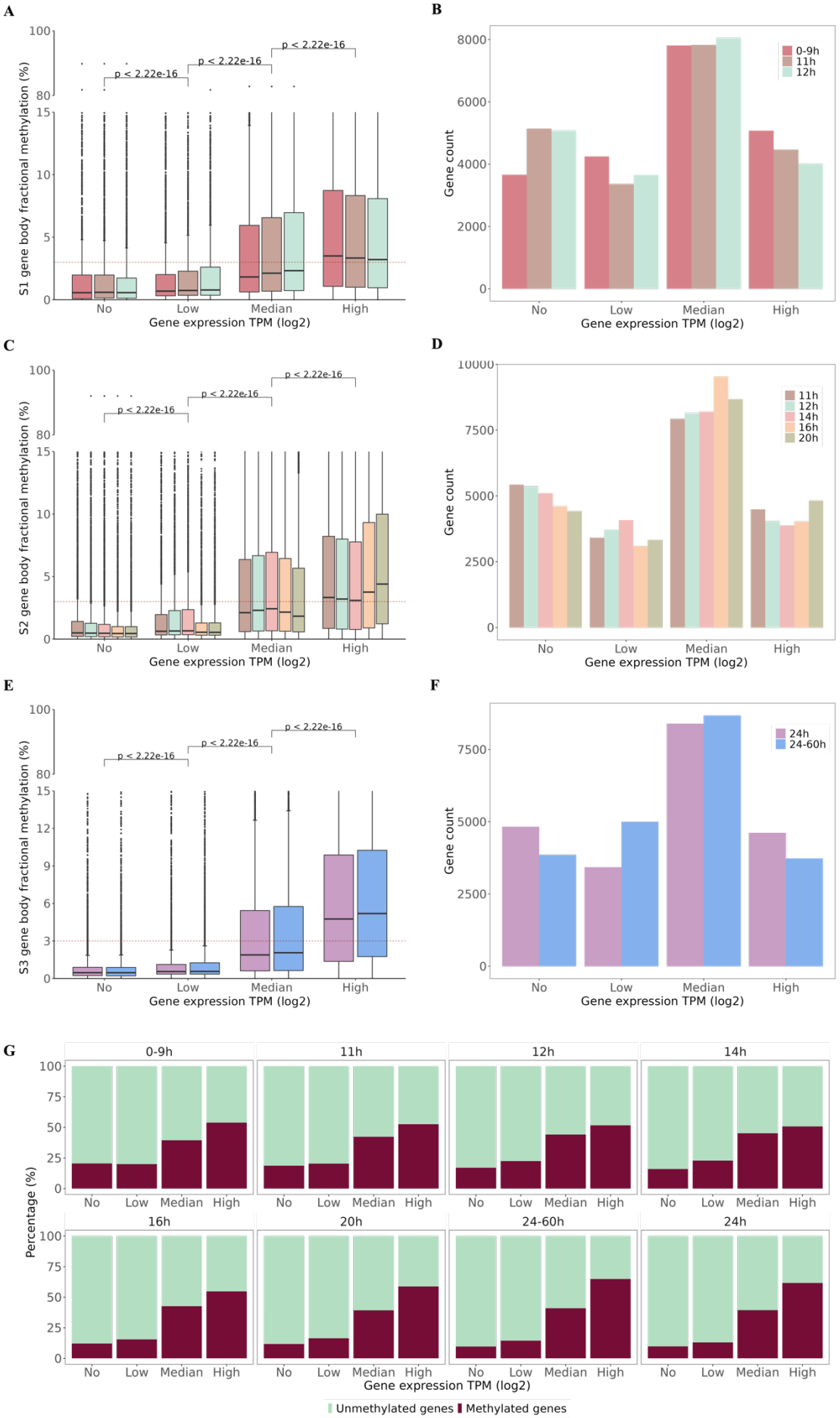

A
S1 gene body fractional methylation (%)
p < 2.22e-16
No
Low
Median
High
Gene expression TPM (log2)
B
Gene count
0-9h
11h
12h
C
S2 gene body fractional methylation (%)
D
11h
12h
14h
16h
20h
E
S3 gene body fractional methylation (%)
F
24h
24-60h
G
0-9h
11h
12h
14h
16h
20h
24-60h
24h
Percentage (%)
Unmethylated genes
Methylated genes

Figure 17. Correlation between gene body methylation and gene expression. **(ACE)** Methylation of genes expressed in four expression groups. No: log2TPM = 0, Low: log2TPM<1, Median: log2TPM<5, and High: log2TPM ≥ 5; Y-axis represents methylation for each EM-seq stage and X-axis represents temporally associated gene expression. **(BDF)** Quantity of genes within each expression group at each RNA-seq time point. **(G)** Percentage of unmethylated and methylated genes in the different expression groups at the eight RNA-seq time points.

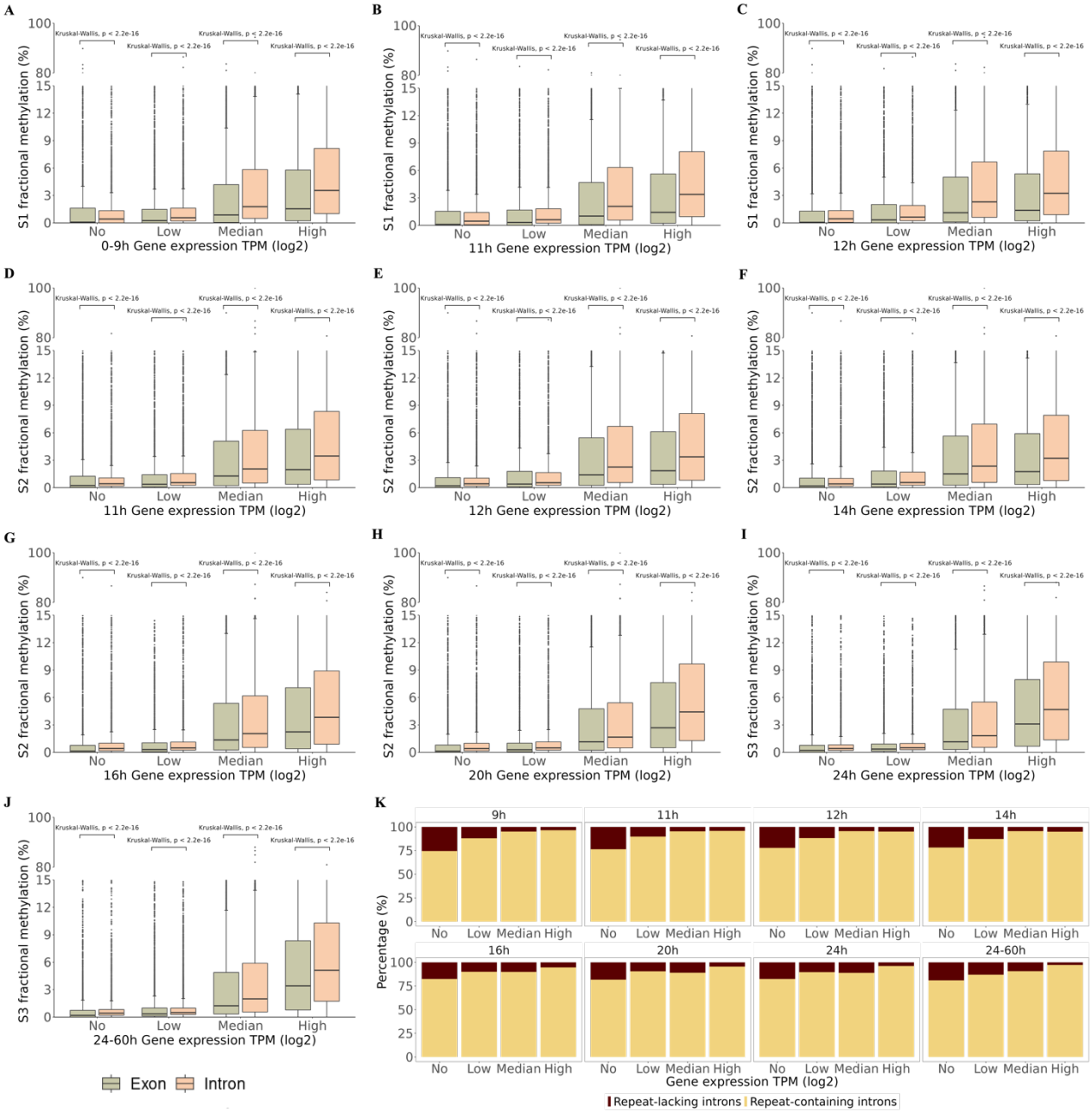


Figure 18. Correlation between intron/exon methylation and gene expression. **(A-J)** Intron and exon methylation of the four gene expression groups for each RNA-seq time point. **(K)** Proportion of repeat-containing introns and repeat-lacking introns in the four expression groups at each RNA-seq time point.

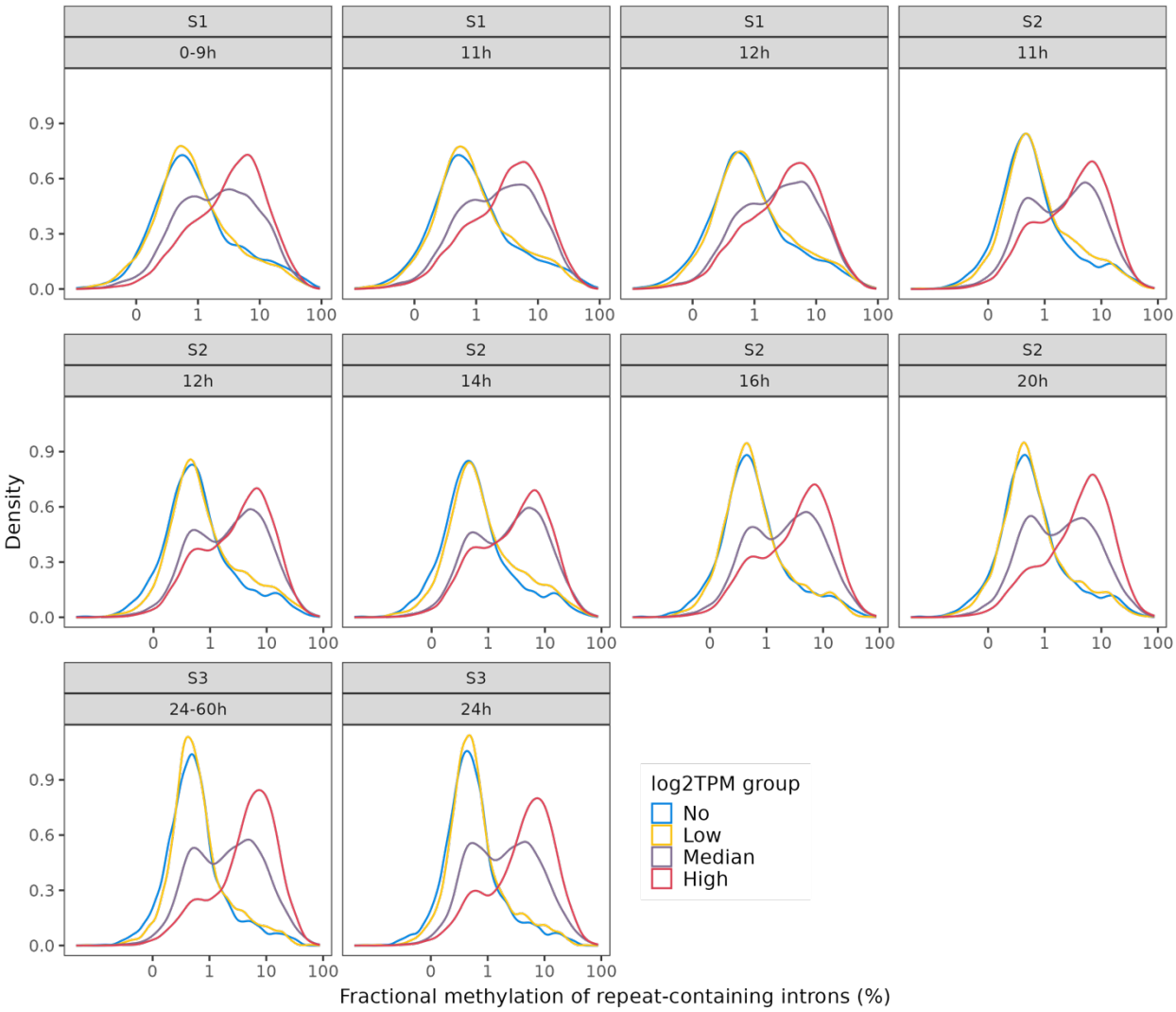


Figure 19. Distribution of fractional methylation of repeat-containing introns in different expression groups.

3.3.3.2 Methylation of zygotically activated genes

Based on the analysis of transcriptomic dynamics during *Parhyale* zygotic genome activation (ZGA), my colleague Dr. Manuel Jara Espejo identified two waves of transcription of zygotically activated genes: the minor wave occurring between 11-hour and 14-hour, and the major wave starting at 16-hour. To investigate the potential role of gene body methylation in the regulation of these genes, I compared the methylation of these two waves of genes and the maternally expressed genes before and after MZT. The results indicated that maternally expressed genes tended to lose methylation after

MZT, while minor and major wave genes maintained comparable levels of gene body methylation before and after the initiation of MZT (Figure 20). These findings suggest that gene body methylation may be involved in the regulation of zygotically activated genes during ZGA.

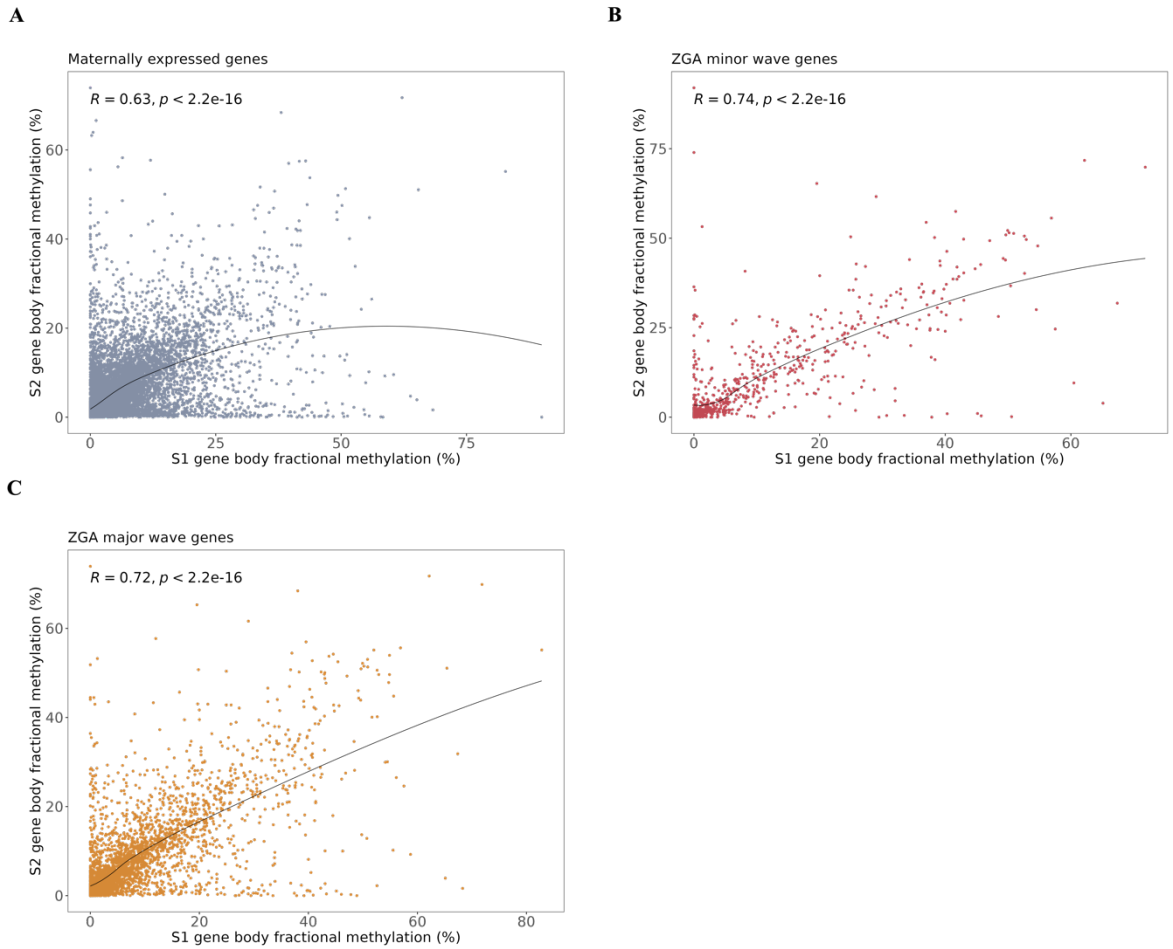


Figure 20. Comparison of methylation between pre-MZT (S1) and MZT (S2) for **(A)** maternally expressed genes, **(B)** minor wave genes, and **(C)** major wave genes

The levels of gene body methylation of zygotically activated genes in minor/major waves were compared with maternally expressed genes. Specifically, the zygotically activated genes in both waves exhibited greater gene body methylation than the maternally expressed genes (Figure 21A,C,E). To further investigate this, I examined the methylation levels of exons and introns separately. Minor wave zygotically activated genes showed

no significant difference in methylation between exons and introns, while major wave genes displayed higher methylation in introns than exons (Figure 21B,D,F).

The length of introns has been observed to be short in early transcribed zygotic genes in both vertebrates and invertebrates (Artieri & Fraser, 2014; Gildor et al., 2016; Heyn et al., 2014). Therefore, I compared the length of introns among the three gene sets and found that the introns of minor wave genes were the shortest, while major wave genes also had shorter introns compared to maternally expressed genes (Figure 21G). As gene body methylation mainly targets introns rather than exons in *Parhyale*, the shorter introns in zygotic genes may explain why there were no significant differences in exon and intron methylation of the minor wave genes.

Dr. Manuel Jara Espejo categorised the zygotically activated genes of both waves into two groups based on their expression pattern: zygotic-only genes and zygotic-maternal genes. The former group includes genes activated during the minor/major wave but not expressed prior to the MZT period, while the latter group includes genes expressed maternally and activated during the minor/major wave. Our analysis revealed that the fractional methylation of minor wave zygotic-only genes was significantly higher than that of both maternally expressed genes and zygotic-maternal genes (Figure 22A, B). Notably, significant differences in methylation levels were observed only between zygotic-maternal genes and maternally expressed genes for major wave genes, but not between zygotic-only genes and maternally expressed genes (Figure 22C).

Nonetheless, these results demonstrate a correlation between gene body methylation and the expression of zygotically activated gene before and after MZT. This implies that DNA methylation may have a significant role in activating gene expression, particularly for genes expressed during the ZGA of *Parhyale* embryonic development.

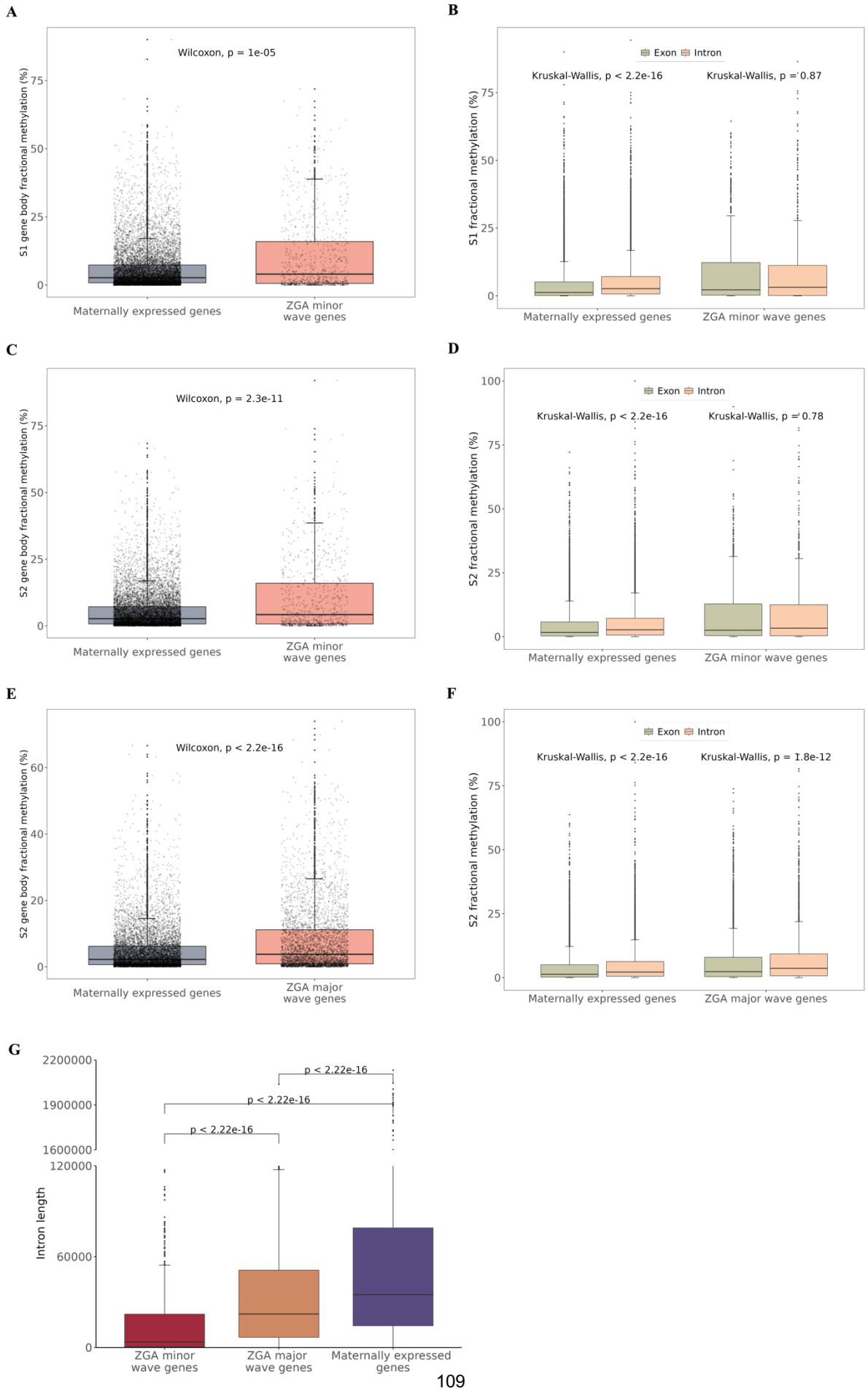

A
Wilcoxon, p = 1e-05
S1 gene body fractional methylation (%)
Maternally expressed genes
ZGA minor wave genes
B
Exon Intron
Kruskal-Wallis, p < 2.2e-16
Kruskal-Wallis, p = 0.87
S1 fractional methylation (%)
Maternally expressed genes
ZGA minor wave genes
C
Wilcoxon, p = 2.3e-11
S2 gene body fractional methylation (%)
Maternally expressed genes
ZGA minor wave genes
D
Exon Intron
Kruskal-Wallis, p < 2.2e-16
Kruskal-Wallis, p = 0.78
S2 fractional methylation (%)
Maternally expressed genes
ZGA minor wave genes
E
Wilcoxon, p < 2.2e-16
S2 gene body fractional methylation (%)
Maternally expressed genes
ZGA major wave genes
F
Exon Intron
Kruskal-Wallis, p < 2.2e-16
Kruskal-Wallis, p = 1.8e-12
S2 fractional methylation (%)
Maternally expressed genes
ZGA major wave genes
G
p < 2.22e-16
p < 2.22e-16
p < 2.22e-16
Intron length
ZGA minor wave genes
ZGA major wave genes
Maternally expressed genes

Figure 21. Gene body, exon and intron fractional methylation of maternally expressed and ZGA genes. **(A)** Gene body, **(B)** exon and intron methylation of minor wave genes at S1. **(C)** Gene body, **(D)** exon and intron methylation of minor wave genes at S2 **(E)** Gene body, **(F)** exon and intron methylation of major wave genes at S2. **(G)** Intron length of minor wave genes, major wave genes and maternally expressed genes.

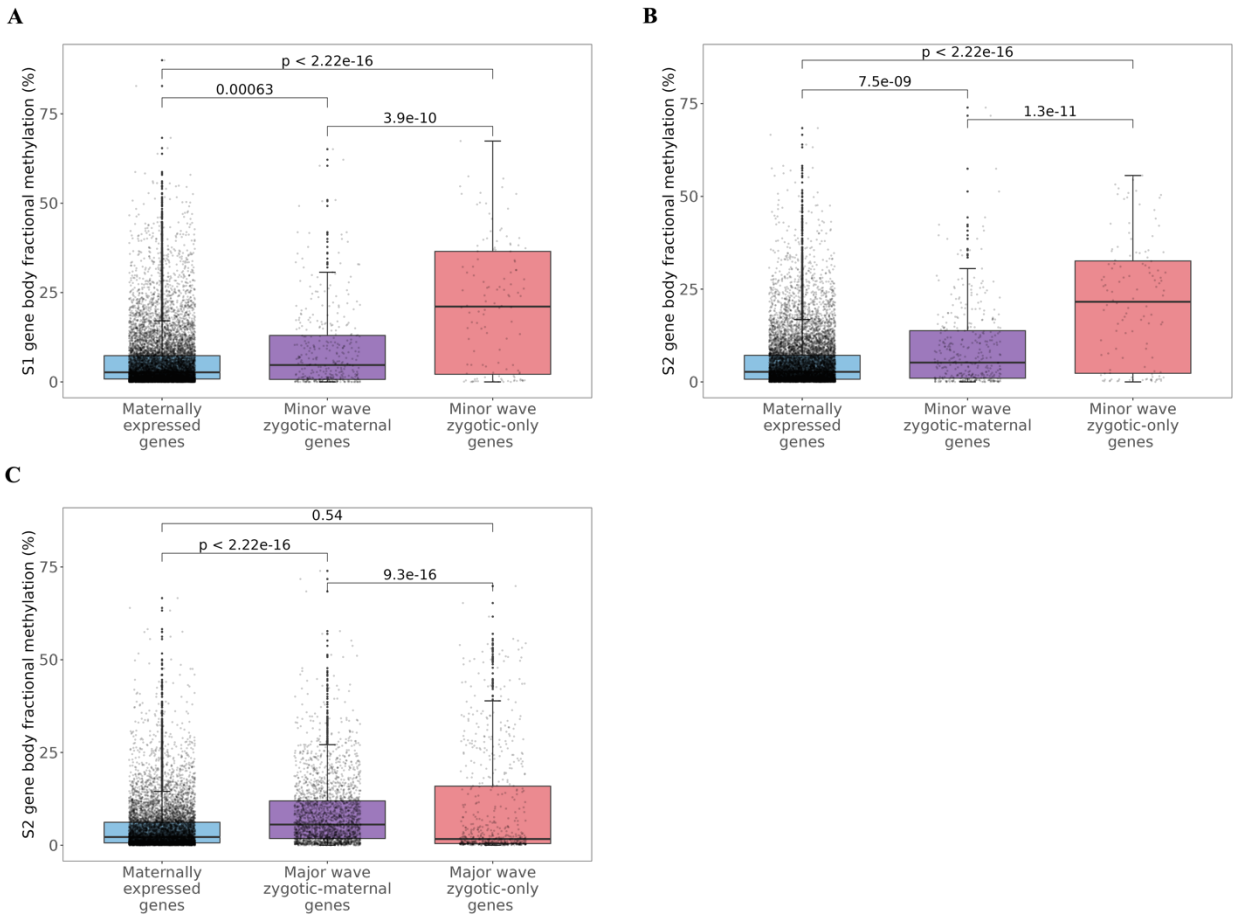


Figure 22. Gene body fractional methylation of maternally expressed genes and ZGA minor/major wave genes **(A)** Methylation of minor wave genes at S1. **(B)** Methylation of minor wave genes at S2. **(C)** Methylation of major wave genes at S2.

3.4 Conclusion

In this chapter, I utilised EM-seq to obtain methylation data from three different developmental stages of *Parhyale*. Genome-wide methylation analysis revealed that approximately 8% of CpGs were methylated in *Parhyale* embryos. The analysis also revealed that DNA methylation targeted both intragenic and intergenic regions, with a particular focus on repetitive elements within these regions. The extent of repeats methylation was influenced by their location, with the maximum levels of repeat methylation observed in non-coding regions. Additionally, as development progressed, the *Parhyale* genome experienced a decline in global gene body methylation.

Previous studies on invertebrates have reported biased associations between gene body methylation and gene expression, with some studies showing positive correlations while others showing negative correlations (Dimond & Roberts, 2015; Dixon et al., 2018; Gavery & Roberts, 2013; Riviere et al., 2013; Suzuki et al., 2007; Ying et al., 2022; Zemach et al., 2010). However, in the case of *Parhyale* early development, a positive relationship was found between gene body methylation and gene expression, especially in highly expressed genes that had high repetitive content and increased intronic repeat methylation. These findings suggest that repeat methylation may play a crucial role in both global DNA methylation and the elevated levels of gene body methylation observed in highly expressed genes during *Parhyale* early development.

In addition to the overall correlation between gene expression and gene body methylation, I also performed a detailed analysis comparing the methylation patterns of maternally expressed genes and zygotically activated genes. The results demonstrated that zygotically activated genes exhibited significantly higher methylation levels compared to maternally expressed genes, with the most pronounced difference observed in minor wave zygotic-only genes.

In summary, this chapter sheds light on the relationship between gene body methylation and gene expression in *Parhyale*, particularly during the ZGA period. The results suggest that DNA methylation may play a significant role in regulating gene expression during the early embryonic development of *Parhyale*. However, the current conclusions are based on a predictive model that assumes methylation remains consistent between 14-hour and 20-hour. The actual changes in DNA methylation during this interval, and their relationship to gene expression, remain unclear. To validate this predictive model, additional DNA methylation data from the three RNA-seq time points (14-hour, 16-hour, 20-hour) should be collected and analysed.

In the next chapter, I further explore the function of DNA methylation in *Parhyale* early development by using a commonly utilized methylation inhibitor aiming to block DNA methylation and investigate its effect on gene expression during MZT and ZGA, and ultimately on embryonic development.

Chapter 4

The effects of inhibiting DNA methylation in the early development of *Parhyale hawaiensis*

4.1 Introduction

4.1.1 DNA methylation inhibitors

Utilising DNMT inhibitors to prevent DNA methylation is a commonly employed approach for studying the influence of DNA methylation on gene expression and biological processes. DNMTs are essential for catalysing DNA methylation, especially DNMT1 and DNMT3, which are the enzymes responsible for converting cytosines into methylated cytosines (Goll & Bestor, 2005). DNMT inhibitors fall into two categories. The first category consists of nucleotide analogues that bind to DNA and form a covalent complex to promote DNMT degradation. The second category comprises non-nucleotide analogues that directly interact with the active sites of DNMTs (Brueckner et al., 2005; Dan et al., 2019; Ghoshal et al., 2005).

Decitabine and azacytidine are the most commonly used nucleotide analogue type DNMT inhibitors. Both were first synthesized by Sorm and Pliml in 1964 and were initially used as anti-metabolite agents in leukaemia chemotherapy (Pliml & Sorm, 1964, Sorm & Vesely, 1968). Azacytidine and decitabine are analogues of cytidine and 2'-deoxycytidine, respectively, and contain nitrogen instead of carbon at the fifth position of the pyrimidine ring (Gnyszka et al., 2013). In the late 1970s, it was discovered that both drugs possess hypomethylating properties, and it was later reported that low doses of these drugs could effectively modify DNA methylation to target cancer cells with reduced cytotoxicity (Jones & Taylor, 1980; Momparler, Bouchard, et al., 1984; Taylor & Jones, 1979).

Decitabine and azacytidine are widely used in research and as therapeutics (Momparler, 2003). In cells, decitabine is phosphorylated and incorporated into nascent DNA by DNA polymerase upon cell entry. This triggers a series of chemical reactions, resulting in the formation of Aza-triphosphate (Aza-TP). During the S-phase of the cell cycle, Aza-TP is incorporated into DNA, replaces cytosine and forms a covalent bond with a cysteine

residue in the active site of DNMTs. Since decitabine's fifth position is replaced by nitrogen, it cannot undergo a methyl transfer reaction, causing the enzyme to become irreversibly trapped. Consequently, the number of functional DNMTs decreases progressively, leading to their degradation and the loss of DNA methylation (Patel et al., 2010; Seelan et al., 2018; Stresemann & Lyko, 2008). Therefore, decitabine functions as an S-phase-specific drug and irreversibly binds to DNMT1, leading to the loss of DNA methylation at CpG sites after cell replication (Hagemann et al., 2011; Schermelleh et al., 2005).

Decitabine can induce DNA demethylation in all regions of DNA, including promoters, gene bodies, and intergenic regions. The hypomethylating effect of decitabine is most pronounced at low doses and does not impair DNA synthesis. At these low doses, decitabine induces DNA demethylation, which can reactivate tumour suppressor genes and downregulate oncogenic processes. However, decitabine has a dual mechanism of action, in which it is cytotoxic at large doses and inhibits DNA polymerase function, causing growth arrest and cell death (Jabbour et al., 2008; Momparler, 2003; Qin et al., 2009; Silverman et al., 2002; Taby & Issa, 2010).

On the other hand, azacytidine primarily incorporates into RNA, thereby interfering with RNA metabolism, RNA methylation, and protein synthesis and may cause substantial cytotoxicity and adverse effects (Leer & Karon, 1976; Schaefer et al., 2009). Although azacytidine is less effective than decitabine at inducing DNA demethylation as only 10% of it incorporates into DNA, it still has significant effects on DNA methylation and is effective in treating certain types of cancer (Issa et al., 2005; Jabbour et al., 2008; Momparler, 2003; Momparler, Momparler, et al., 1984).

Due to the dose-limiting toxicity and restricted patient tolerance associated with cytidine analogues, coupled with their limited efficacy in treating solid tumours, there has been a continuous pursuit of alternative DNMT inhibitors (Sato et al., 2017; Stomper et al., 2021).

Apart from decitabine and azacytidine, a range of other DNA methylation inhibitors have been developed, including Zebularine, Guadecitabine, SGI-1027, MC3343, and N-Phthaloyl-L-tryptophan (RG108). However, none of these inhibitors manage to effectively balance potency, selectivity, and translatability across both *in vitro* and *in vivo* activities, and they are generally less effective than decitabine at inhibiting DNA methylation and reactivating silenced genes in cancer cells (Brueckner et al., 2005; Chuang et al., 2005; Dan et al., 2019; Gros et al., 2015; Hu et al., 2021; Manara et al., 2018).

In recent years, GlaxoSmithKline (GSK) introduced a new class of DNA methylation inhibitors that are not cytidine analogues, including GSK-3484862, GSK-3735967, GSK-3685032, and GSK-3830052 (Chen et al., 2023; Horton et al., 2022; Pappalardi et al., 2021). These compounds, containing the dicyanopyridine moiety, have the ability to selectively interact with DNA and the DNMT1 active-site loop of human and mice, halting DNMT1 activity without necessitating DNA incorporation (Horton et al., 2022; Pappalardi et al., 2021). Studies have demonstrated that these compounds induce significant loss of DNA methylation in human and mice, trigger transcriptional activation, and inhibits cancer cell growth both *in vitro* and *in vivo*. Notably, they exhibit lower cytotoxicity and enhanced *in vivo* tolerability compared to traditional cytidine analogues (Gilmartin et al., 2021; Haggerty et al., 2021; Pappalardi et al., 2021; Portilho et al., 2021). Additionally, GSK-3484862 has demonstrated the ability to induce DNA hypomethylation to a level comparable to that observed in dnmt1 knockout murine embryonic stem cells (Portilho et al., 2021). Furthermore, it exhibits potential in tumour regression that surpasses that of decitabine in a mouse model of acute myeloid leukaemia (Pappalardi et al., 2021). However, their inhibition efficiency relies heavily on the conserved active sites of DNMT1, making it challenging to apply these inhibitors to species that are evolutional distinct from mammals without careful examination of the DNMT1 active site and proper validation. To date, no studies have utilised GSK-3484862 in any invertebrate model system.

Therefore, despite the genotoxic effects associated with 5-azanucleosides and ongoing research aimed at enhancing the efficacy of DNMT inhibitors, cytosine analogues such as decitabine are still among the most commonly utilised and efficacious drugs for inducing DNA demethylation in both research and clinical practice.

4.1.2 Effects of decitabine on gene expression

Decitabine can affect gene expression through both direct and indirect processes. Direct effects arise when decitabine is incorporated into a gene, causing significant alterations in methylation and expression levels of that gene. In cancer therapy, most studies on decitabine have focused on its ability to induce demethylation of gene promoters. Promoters often contain CpG residues that are frequently methylated, leading to the silencing of genes associated with tumour suppression. Numerous experiments have shown that decitabine treatment of tumour cell lines containing silenced genes reduces CpG methylation within the promoter, thereby reactivating the expression of those tumour suppressor genes (Kim et al., 2015; Rawłuszko-Wieczorek et al., 2015; Sayar et al., 2015).

Decitabine can also affect gene expression through indirect mechanisms. These indirect effects are generated when decitabine modifies gene expression without significantly altering methylation within the gene itself. Demethylation of upstream genes, such as those encoding transcription factors or regulatory elements, may be affected by decitabine-induced demethylation, leading to changes in the expression of downstream gene (Seelan et al., 2018; Soengas et al., 2001). Another indirect effect of decitabine on gene expression is the modification of histones, which can increase gene expression without significantly altering DNA methylation (Hong et al., 2012; Nguyen et al., 2002; Poplineau et al., 2015; Zheng et al., 2012). Additionally, responses to DNA damage or repair mechanisms induced by decitabine toxicity are a third mechanism by which decitabine may trigger gene expression independently of promoter demethylation (Speidel, 2015; Zhu et al., 2004).

Hagemann et al, (2011) investigated the demethylation effects of decitabine in colon cancer and leukemic cell lines and suggested that certain loci were more susceptible to demethylation than others, indicating that not all regions within genes are equally affected by the demethylating drugs. Additionally, the study found that highly methylated CG dinucleotides located predominantly in non-CGI contexts, especially those in repetitive elements, were more likely to be demethylated by decitabine. As a result, the authors concluded that decitabine may not induce global demethylation but instead demethylates specific and repetitive sites in human cancer cell lines (Hagemann et al., 2011).

Decitabine treatment can result in both the upregulation and downregulation of genes, as well as the reactivation of genes with unmethylated promoters (Claus et al., 2005; Karpf et al., 2004; Liang et al., 2002; Y. W. Zhang et al., 2012). Decitabine tends to target two types of genes: those with high methylation in promoters and intermediate CpG content, and those with low levels of methylation and high CpG content. The CpG density of the promoter seems to be essential for decitabine to induce demethylation (Rubinstein et al., 2010). Although few studies have explored the effect of decitabine on gene body methylation in cancer therapy, evidence suggests that decitabine may have a complex effect on gene body methylation and expression. Further research is required to completely comprehend the underlying mechanisms (Shenker et al., 2015; Yang et al., 2014).

4.1.3 DNA methylation inhibitors utilised in animal research

In addition to their role in human cancer therapy, DNA methylation inhibitors like decitabine and azacytidine have been widely used in functional studies of DNA methylation in both vertebrates and invertebrates. For instance, decitabine treatment has been shown to decrease DNA methylation in a dose-dependent manner and altered gene expression in the wasp *P. xylostella*, leading to significant changes in its physiological processes (Kumar & Kim, 2017). Similarly, decitabine treatment in the silkworm *Bombyx*

mori pupa decreased DNA methylation at specific CpG sites in gene promoters, resulting in increased mRNA levels of those genes (Xu et al., 2018). In the marine annelid *Platynereis dumerilii*, decitabine treatment decreased global DNA methylation, resulting in developmental defects, delayed regeneration, and interference with post-regenerative growth (Planques et al., 2021). In oyster embryos, exposure to low concentrations of azacytidine caused a severe decrease in DNA methylation and altered development (Riviere et al., 2013), while exposure to azacytidine also modified the expression of Hox genes during embryogenesis (Saint-Carlier & Riviere, 2015). In the blood fluke *Schistosoma mansoni*, both azacytidine and decitabine were found to reduce DNA methylation, with azacytidine being more effective than decitabine (Geyer et al., 2011).

Cook et al, (2019) confirmed that decitabine disrupts methylation in the wasp *Nasonia*. They also found that methylation tended to increase in the first 24 hours of decitabine exposure and then diminish during the following 48 hours. The DNA methylation study of *Bombus terrestris* bumblebees revealed that decitabine treatment induced both hyper- and hypomethylation, altering methylation patterns in callow worker bees but not inducing genome-wide demethylation in bees over one day old; instead, significant changes in methylation were observed in specific loci in younger bees (Amarasinghe et al., 2014). Aigner et al, (2022) also found no global DNA demethylation in earthworm *Lumbricus terrestris* treated with decitabine, possibly due to a low proliferation rate, as decitabine requires cells division to exert its demethylating effect. However, the effect of decitabine treatment is highly dependent on variables such as exposure duration, dosage, tissue type, and collection time (Cook et al., 2019).

In addition to invertebrates, decitabine and azacytidine have been shown to affect global DNA methylation levels, alter the transcriptome, and influence development in multiple vertebrate animal models. This includes chicken (Chen et al., 2017; Zagris & Podimatas, 1994), zebrafish (Ribas et al., 2017), medaka fish (Iida et al., 2006), goldfish Carassius auratus (Zhang et al., 2014), Japanese rice fish (Dasmahapatra & Khan, 2015), porcine

embryos (Diao et al., 2013). According to the aforementioned studies, decitabine and azacytidine are reliable and stable DNA methylation inhibitors and are valuable tools for functional research on DNA methylation.

4.1.4 RNA sequencing technology

RNA sequencing, also known as RNA-seq, has revolutionized the study of transcriptomes and is an immensely valuable tool in biological research, drug discovery, and clinical medicine (Khatoon et al., 2014; Kukurba & Montgomery, 2015; Wang et al., 2009). It generates large amounts of transcriptome data and enables the simultaneous measurement of expression levels of thousands of genes, providing valuable insights into functional pathways and regulatory networks of biological systems. Furthermore, RNA-seq can also reveal new information on alternative splicing, unannotated exons, and novel transcripts (Khatoon et al., 2014; Wang et al., 2009).

In animal cells and tissues, various RNA molecules are present, including ribosomal RNA (rRNA), precursor messenger RNA (pre-mRNA), mRNA, and several types of non-coding RNA (ncRNA). However, rRNA is the most abundant component, making up the majority of molecules in a total RNA sample (Kukurba & Montgomery, 2015; Wilhelm & Landry, 2009). Therefore, efficient removal of rRNA is crucial for sequencing the remaining RNA molecules, which provide more comprehensive insights into gene expression and regulation.

Multiple techniques can be utilised to eliminate rRNA from total RNA, such as polyadenylated RNA (poly(A)) enrichment, selective degradation of abundant transcripts, amplification of non-ribosomal RNA, or direct depletion of rRNA using hybridisation capture followed by magnetic bead separation. Each of these techniques has unique advantages and disadvantages, and the choice of technique relies on the experimental design and the objectives of the study. For example, poly(A) enrichment is commonly employed for transcriptome studies focused on mature, polyadenylated transcripts, as it

requires less sequencing depth when analysing the protein-coding region of the transcriptome. Direct rRNA depletion is more efficient in removing all ribosomal RNA, including non-coding rRNA, and provides more information on poly(A) transcripts. Moreover, rRNA-depleted libraries are more effective than poly(A)-selected libraries in analysing degraded RNAs. Recent studies have highlighted the advantages of using rRNA-depleted RNA libraries for transcriptome analysis. Ultimately, the selection of the method depends on the specific research question, desired sequencing depth, and available resources (Alberti et al., 2014; Guo et al., 2015; A. Kumar et al., 2017; O'Neil et al., 2013; Petrova et al., 2017; Schuierer et al., 2017).

RNA sequencing technology can be divided into three subcategories based on the type of RNA being sequenced: messenger RNA (mRNA), micro-RNA (miRNA), and total RNA (Guo et al., 2015). Although the sequencing method is identical, the RNA molecules present during cDNA synthesis and library construction vary between subclasses. In mRNA sequencing, the cDNA library is constructed exclusively from poly(A) mRNA, and small RNAs are not captured during oligo-dT-based mRNA enrichment. To date, mRNA sequencing has been the most prevalent use of RNA-seq technology, particularly in studying gene expression and splicing (Y. Guo, Li, et al., 2013; Y. Guo, Sheng, et al., 2013; Wang et al., 2009). However, non-coding RNA lacking a poly(A) tail also contains valuable information that is not captured in mRNA-seq. These RNA transcripts are important as they provide insights into the regulatory and functional aspects of the transcriptome (Amaral & Mattick, 2008; Kukurba & Montgomery, 2015).

The discovery of novel RNAs has been greatly enhanced by the development and utilisation of new RNA sequencing techniques. In human, over 75% of the genome is transcribed into primary and processed transcripts, with only 1.2% of the human genome encoding proteins (Djebali et al., 2012). Non-coding RNAs (ncRNAs) can be classified into three main categories: housekeeping RNAs (rRNAs, tRNAs, snRNAs, and snoRNAs), short non-coding RNAs (<200 nucleotides), which consist of miRNAs, piRNAs, and

retrotransposon-derived ncRNAs, and long non-coding RNAs (lncRNAs) (Derrien et al., 2012; Leva & Garofalo, 2013). The majority of ncRNAs in the cell originate from non-coding regions, while others arise from protein-coding regions, such as antisense transcripts, intron regions, and scrambled coding sequences. Moreover, it has been discovered that ncRNAs can either target and degrade mRNAs or participate in the regulation of transcription, highlighting the growing importance of studying non-coding RNAs in the animal genome (Hadjiargyrou & Delihas, 2013; Memczak et al., 2013; Quinodoz & Guttman, 2014).

Long non-coding RNAs (lncRNAs), which encompass large intergenic non-coding RNAs (lincRNAs), intronic transcripts, antisense transcripts, and circular RNA transcripts, are defined as non-coding RNAs with a length exceeding 200 base pairs (Derrien et al., 2012; Leva & Garofalo, 2013). Some lncRNAs share structural similarities with messenger RNAs, which are also capped, spliced, and polyadenylated (Carninci et al., 2005; Ulitsky & Bartel, 2013; Zhao et al., 2018). These RNAs were once thought to be non-functional. However, recent studies have provided evidence of their involvement in various biological contexts, including chromatin reprogramming, embryonic stem cell differentiation, post-transcriptional modification, and pathological processes in both vertebrates and invertebrates (Amaral & Mattick, 2008; Dinger et al., 2008, 2009; Mercer et al., 2008, 2009; Schoeftner & Blasco, 2008). As a result, interest in the study of lncRNA has grown significantly over the past few years.

Total RNA-seq is a powerful method to investigate lncRNAs because it does not require the removal of non-polyadenylated RNAs, such as those lacking a poly(A) tail. Instead, total RNA-seq selectively removes ribosomal RNA, allowing for the sequencing of all types of lncRNAs (Guo et al., 2015). Although poly(A) RNA-seq and total RNA-seq generate comparable results when studying protein-coding RNAs, the latter method allows for the identification and characterisation of immature and nascent transcripts and

intronic sequences from pre-mature mRNAs, providing unique insights into transcriptional processes (Guo et al., 2015; Zhao et al., 2018).

4.1.5 Aim of this chapter

The incorporation of decitabine into DNA results in the irreversible binding to DNMT1, leading to DNA demethylation during cell division. Yet, there is a lack of research utilising demethylation agents in *Parhyale* to investigate the function of DNA methylation during early development.

The results presented in Chapter 3 indicate a positive correlation between DNA methylation and gene expression in *Parhyale* during the MZT and the ZGA periods. This chapter examines the effect of the DNA methylation inhibitor decitabine on the early development of *Parhyale* embryos. The aim is to determine whether decitabine treatment and the accompanying inhibition of DNA methylation impact gene expression and transposable element activity during the embryonic development of *Parhyale*. The hypothesis being tested is that proper regulation of DNA methylation is crucial for normal embryogenesis, particularly during MZT and ZGA. The findings of this study may serve as a valuable model for understanding the role of DNA methylation regulation in invertebrate species.

4.2 Materials and methods

4.2.1 *Parhyale* culture and embryo collection

The *Parhyale* culture and embryo collection procedures were conducted as outlined in Chapter 3.

4.2.2 Phylogenetic analysis of DNMT1 from different species

To validate the possibility of utilising the GSK dicyanopyridine moiety-containing DNA methylation inhibitors for *Parhyale*, I first extracted the coding sequences of *Parhyale* DNMT1 from the *Parhyale* gene annotation file generated by my colleague Dr. Manuel Jara Espejo. Using the translate tool from Expert Protein Analysis System (ExPASy, (Gasteiger et al., 2003)), I generated the protein sequence of *Parhyale* DNMT1. The Basic Local Alignment Search Tool (BLAST, (Altschul et al., 1990)) was then used to analyse the identity of *Parhyale* DNMT1 compared to human (UniProt: P26358) and mouse (UniProt: P13864) DNMT1. Multiple sequence alignment of DNMT1s from human, mouse and *Parhyale* was performed using MUSCLE (Edgar, 2004) with default settings and visualised with MEGA X (Kumar et al., 2018).

4.2.3 Chemicals

Decitabine (5-Aza-2'-deoxycytidine, 5AzaD) was purchased from Biosynth (NA02969). Stock solutions were prepared by dissolved 100 mg 5AzaD in 7 ml Dimethyl Sulfoxide (DMSO, Fisher Scientific, BP231-100) to obtain a stock concentration of 62.5 mM. The stock solution was then store at -80°C.

4.2.4 Decitabine toxicity range-finding test

To assess the potential toxicity of 5AzaD on *Parhyale* embryos, a range-finding toxicity test was performed as this drug had never been used in *Parhyale* embryos before. First,

50 one-cell embryos were collected and placed in a petri dish containing filtered artificial seawater They were then treated with 5AzaD from the one-cell stage until they reached 24 hours of development. The experiment employed eight different concentrations of 5AzaD ranging from 0.5 μM to 150 μM (0.5 μM, 1 μM, 5 μM, 10 μM, 25 μM, 50 μM, 100 μM, and 150 μM). Working solutions of 5AzaD were freshly prepared by adding different volumes of stock solutions into embryo medium to achieve the different concentrations. After 24 hours, the surviving embryos were counted, rinsed twice in fresh FASW, and then transferred to a new petri dish containing FASW. They were maintained in a chamber at 28.3°C, and the number of surviving embryos was counted every two days until they hatched.

4.2.5 Test of decitabine exposure time window

After determining the non-lethal dose, an experiment was conducted to identify the sensitive developmental periods during which *Parhyale* embryos could be exposed to 5AzaD. To accomplish this, *Parhyale* embryos (50 per replicate) were treated with either 5 μM or 10 μM of 5AzaD or DMSO as a vehicle control in three replicates per treatment. The volume of DMSO administered to embryos in the control groups was equivalent to the volume of 5AzaD administered to embryos in the 5AzaD groups. The following developmental exposure windows were chosen: a) from 1-cell to 6-hour (4-cell); b) from 1-cell to 12-hour (100-cell); c) from 6-hour (4-cell) to 12-hour (100-cell); d) from 12-hour (100-cell) to 18-hour; e) from 12-hour (100-cell) to 24-hour; and f) from 18-hour to 24-hour.

All embryos were treated in 25 ml of the treatment solution for the designated time period (e.g., 0-12 hour: 1-cell embryos were exposed to 5 μM 5AzaD or 10 μM 5AzaD, or DMSO until they reached the 12-hour developmental stage). After exposure, all surviving embryos were examined under a microscope for signs of developmental failure, then rinsed in fresh FASW and transferred to a new petri dish, which was then incubated at

28.3°C. For the first 72 hours, the embryos were examined every 24 hours, and then every 48 hours thereafter. The percentage of mortality was recorded each time the embryos were checked.

4.2.6 Live-cell imaging with Hoechst dye

For each replicate, a total of 12 embryos were placed in 10 ml of embryo medium containing 3 µg/ml of Hoechst 33342 (Sigma, 14533) and incubated in the dark for 6-7 hours. Following the incubation, the embryos were rinsed in fresh embryo medium for 30 seconds before being placed on a slide. The embryos were then viewed using Leica MZ16F Fluorescence Stereo microscope.

4.2.7 Decitabine exposure experiment

Based on the results obtained from the time window test, the chosen exposure time window for the formal experiment was 0-12 hours, and the selected treatment dose was 5 µM of 5AzaD. The experimental procedure was similar to the previous experiments. In brief, one-cell embryos were exposed to either 5 µM of 5AzaD or a vehicle control of 0.008% DMSO until they reached 12 hours of development. Surviving embryos were rinsed twice, and a third of them were immediately snap-frozen in Trizol Reagent (ambion, 15596026) for RNA extraction, while the remaining two-thirds were maintained in normal embryo medium until 24 hours of development. At 24 hours, half of the embryos were collected and snap frozen in Trizol for RNA extraction, while the remaining embryos were incubated until hatching to assess survival rates.

4.2.8 RNA extraction and Total RNA sequencing

After obtaining the desired quantity of embryos for RNA-seq, I performed Trizol-based RNA extraction. The concentration of the extracted RNA samples was determined using a Qubit and Qubit™ RNA HS Assay Kit (Invitrogen, Q32852), while their purity was

evaluated using gel electrophoresis and Nanodrop. The RNA samples were stored at -80°C for future use. There were three replicates for each treatment and time point. Total RNA libraries were prepared using each RNA sample, with ribosomal RNA removal method. Novogene (Novogene UK company Limited) was responsible for the library preparation and sequencing. The RNA-seq libraries were sequenced using paired-end sequencing on the Illumina platform (average length 150 bp).

4.2.9 RNA-seq read alignment and analysis

To evaluate sequencing quality, raw sequence reads were inspected with FastQC v.0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). TRIMMOMATIC (Bolger et al., 2014) (parameters: SLIDINGWINDOW:4:20 LEADING:10 TRAILING:10 MINLEN:36) was then used to remove adapter sequences, low-quality reads, and PCR duplication. The resulting clean reads were then mapped to the *Parhyale* Phaw 5.1 assembly using HISAT2 with default parameters (Kim et al., 2019). FeatureCounts (Liao et al., 2014) was employed to summarize the aligned reads based on exon annotation and generate exon counts. The raw exon counts were initially used for correlation and PCA analysis to assess the similarity among the replicates. Transcripts per million (TPM) were normalized based on exon counts, and differential expression analysis was conducted using the R package DESeq2. A significance threshold of padj < 0.01 and log2FoldChange > 1 was utilised to define differentially expressed genes. DESeq2's variance-stabilizing transformation (VST) was applied to the normalized counts to stabilize the variance of the data. The VST-transformed counts were then used as input for PCA to investigate the variability within both the 5AzaD/DMSO treated and wild-type RNA-seq datasets. DESeq2 is also used to identify differentially expressed genes. To determine the potential functions of the differentially expressed genes (DEGs), Gene Ontology (GO) enrichment analysis was conducted separately on upregulated and downregulated genes using the R package clusterProfiler.

4.2.10 Expression of repetitive elements

SalmonTE (Jeong et al., 2018) was utilised to identify and quantify the expression of repeats at specific genomic loci using the repetitive content (repeatome) library as the reference. A repeat locus was considered expressed if at least one read was mapped to it. Differential expression analysis of repeats was performed using SalmonTE. With padj < 0.01 and log2FoldChange > 1 as the thresholds to define differentially expressed repeats.

SalmonTE is a specialized tool designed for analysing transposable elements using RNA-seq data. It organizes TEs into distinct groups, called clades, based on their repeat class annotations. Once cDNA sequences are collected, SalmonTE constructs an index using these sequences. This process involves manual curation of clades and index creation, allowing SalmonTE to effectively manage highly repetitive TE families. This ensures precise measurement of TE expression despite their repetitive nature (Jeong et al., 2018).

To quantify TE abundance from RNA-seq reads, SalmonTE employs the Salmon algorithm (Patro et al., 2017). This algorithm utilises a quasi-mapping approach to assign each read to its target TE and determine its position and orientation. By employing a generalized suffix array, SalmonTE identifies possible mapping locations for each read. Additionally, SalmonTE utilises probabilistic models for read alignment and inference, considering sample-specific parameters and biases. These methods enable SalmonTE to accurately handle reads that can map to multiple locations, ensuring precise quantification of TE expression. SalmonTE calculates Transcripts Per Million (TPM) to measure the relative abundance of each TE in a sample. It utilises a stochastic collapsed variational Bayesian inference process to estimate read counts and nucleotide fractions. Although longer reads typically offer more information for precise mapping, SalmonTE's quasi-mapping procedure and probabilistic models aim to accurately estimate expression levels even with shorter reads. By employing these techniques, SalmonTE effectively

mitigates biases introduced by read length, ensuring reliable quantification of TE expression (Jeong et al., 2018).

4.2.11 Statistical analysis

All data are presented as mean ± SD (n = 3). The gene fractional methylation of differentially expressed genes and repeats are shown by box plot and the statistical difference was analysed using either the Wilcoxon test or Kruskal-Wallis's test. The two-tailed Fisher exact test was employed to examine the differences between the number of maternally expressed and ZGA genes in the upregulated and downregulated gene sets, as well as the relationship between gene expression and repeat expression. Correlation analyses were performed at a 95% confidence interval using a one-tailed Pearson's test. A p-value ≤ 0.05 was considered significant. All data were analyses and graph plotting were conducted using R.

4.3 Results and discussion

Previous chapters have suggested that DNA methylation may play a role in regulating gene expression in *Parhyale* during the early stages of embryonic development. I thereby further investigate the significance of DNA methylation during this phase. To achieve this, I administered a commonly used hypomethylating agent, Decitabine (5-aza-2'-deoxycytidine, 5AzaD), to the *Parhyale* embryos and conducted total RNA sequencing to generate transcriptomic data for two early developmental time points.

4.3.1 Selection of DNA methylation inhibitor

DNA methylation inhibitors containing the dicyanopyridine component, developed by GSK, typically exhibit low cellular toxicity and specifically target DNMT1. However, their inhibitory mechanism relies on binding to DNMT1, a mechanism not necessarily conserved across different species (Horton et al., 2022; Pappalardi et al., 2021). To evaluate the potential use of these inhibitors in *Parhyale*, I analysed the DNMT1 protein sequence of *Parhyale*, derived from its coding sequence. BLAST analysis revealed that *Parhyale* DNMT1 displayed only 32.74% and 31.85% sequence identity with human (UniProt: P26358) and mouse DNMT1 (UniProt: P13864), respectively. In contrast, human and mouse DNMT1 share 77.59% identity with each other.

Multiple sequence alignment of DNMT1 from human, mouse and *Parhyale* revealed that all three species share three conserved phenylalanine (F628, F631, and F632) located in the ‘activating helix’, which is essential for DNA methylation maintenance (Kikuchi et al., 2022). However, *Parhyale* DNMT1 differs from the other two species at five amino acid sites within the active site loop, which is crucial for DNA binding (Figure 1) (Horton et al., 2022; Pappalardi et al., 2021).

More importantly, crystallographic studies have shown that GSK drugs interact with the specific side chains of H1507, W1510, and K1535 in human DNMT1 (PDB: 7SFC)

(Horton et al., 2022). These amino acids are also present in mouse DNMT1, but *Parhyale* DNMT1 lacks H1507, and K1535 is substituted with glutamine (Q) (Figure 1). Mutagenesis of H1507 to Y1507 in human DNMT1 reduces the sensitivity to GSK drugs by more than 350-fold *in vitro* (Pappalardi et al., 2021). Given the lack of key amino acids for inhibitor binding and the absence of successful use of GSK dicyanopyridine moiety-containing DNA methylation inhibitors in invertebrates, this study opted to use the cytidine analogue decitabine for DNA methylation inhibition.



Figure 1. Multiple sequence alignment of *Parhyale* DNMT1 with DNMT1 from human (NCBI: P26358) and mouse (NCBI: P13864). Catalytically important amino acids are marked with green boxes, amino acids within the active site loop involved in DNA binding are highlighted with a blue box, and crucial amino acids involved in binding the GSK dicyanopyridine moiety-containing DNA methylation inhibitor are indicated with red boxes. The two critical mutations in *Parhyale* DNMT1 are highlighted with green triangles.

4.3.2 Decitabine exposure

To determine the non-lethal concentration of 5AzaD, I first conducted toxicity range-finding tests on *Parhyale* embryos. Starting at the 1-cell stage, I exposed 50 wild-type embryos per replicate (with three replicates per treatment) under static conditions to either vehicle (0.01% DMSO) or 5AzaD (0.5–150 µM) until 24 hours of embryogenesis. Embryos exposed to concentrations of 5AzaD higher than 25 µM resulted in 100% mortality before hatching (Figure 2A). While a 10 µM concentration was less detrimental, with some embryos surviving until hatching (Figure 2B). After determining the non-lethal dose (10 µM), embryos were treated with 5AzaD at concentration of 5 µM and 10 µM for six different time windows before or after the MZT (0–6 hours, 0-12 hours; 6–12 hours,

12-18 hours, 18-24 hours, 12-24 hours) to evaluate potential sensitive windows of embryonic development.

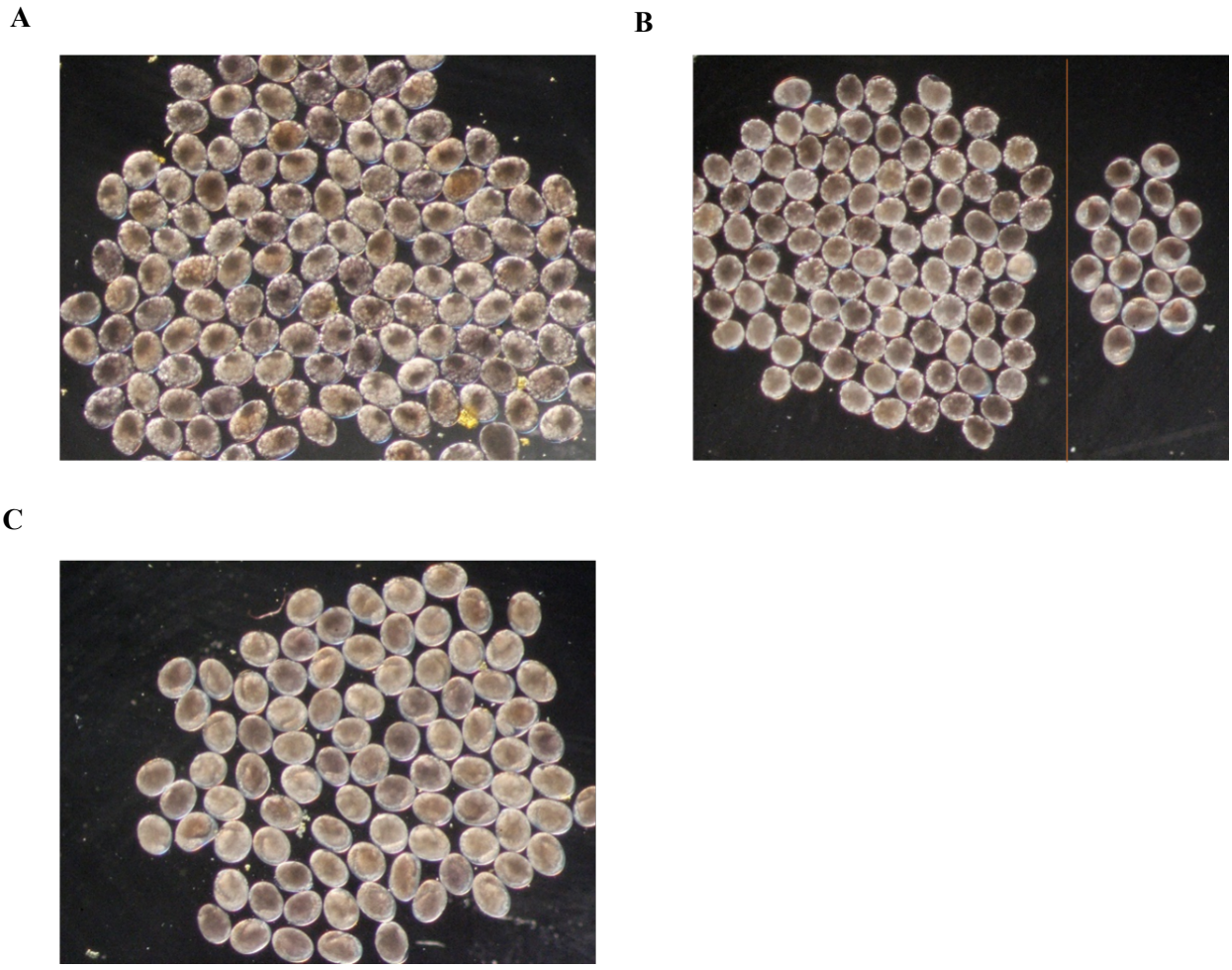


Figure 2. Images of embryos subjected to various doses of 5AzaD and DMSO on the fifth day of embryonic development. **(A)** Embryos treated with 25 µM 5AzaD. showing all dead embryos with cells displaying a bubble-like shape. **(B)** Embryos treated with 10 µM 5AzaD, divided between dead (left of the orange line) and alive (right of the orange line). **(C)** Embryos treated with DMSO, all of which are alive.

Compared to wild-type and DMSO-treated embryos, those exposed to 5AzaD at any of the six time periods showed a significant increase in mortality at hatching (Figure 3). Within the 5AzaD treatment group, embryos exposed before MZT (0–12-hours, 0-6 hours, 6-12 hours) had significant lower survival and hatching rates compared to embryos treated after MZT (12-18 hours, 18-24 hours, 12-24 hours). Additionally, pre-MZT treated embryos exhibited a higher mortality rate within the first 12 hours and decreased survival

rate by day 3. However, most embryos treated with 5AzaD after MZT retained a high survival rate until day 5. Furthermore, there were no significant differences observed between embryos treated with the two different 5AzaD doses (5 µM and 10 µM) (Figure 3). This experiment confirmed that administration of 5AzaD induced embryonic developmental failure at either 5 µM or 10 µM doses. It also highlighted a critical early time window (up to the beginning of MZT, 0-12 hours) during which the inhibition of DNA methylation strongly leads to later developmental defects.

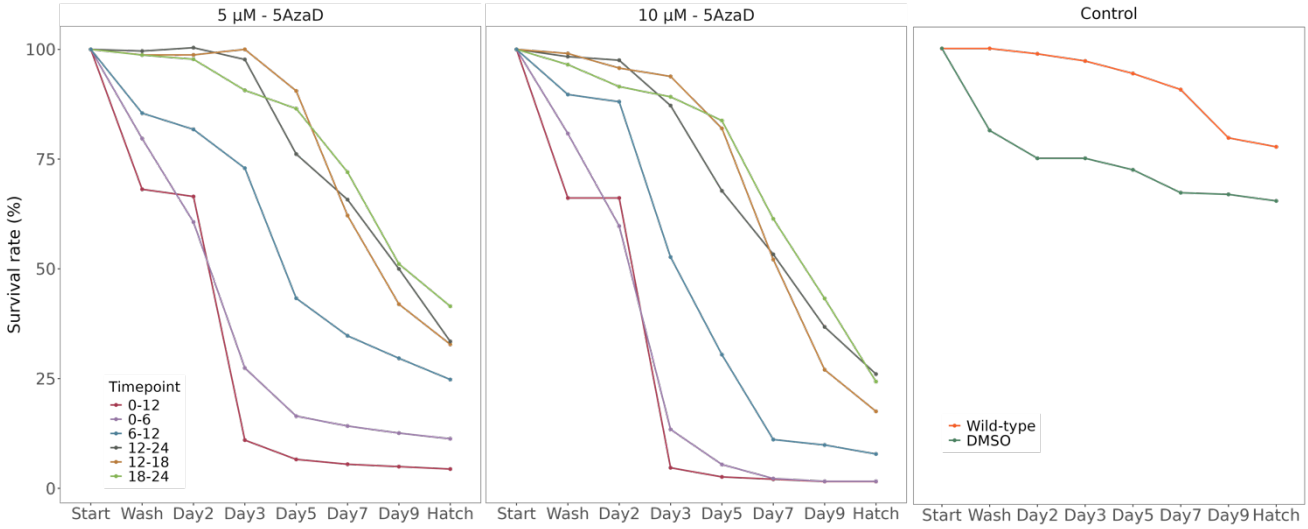


Figure 3. Survival rates of embryos treated under different conditions: **(Left)**: Embryos treated with 5 µM 5AzaD; **(Middle):** Embryos treated with 10 µM 5AzaD; **(Right):** Embryos treated with DMSO and the wild-type embryos. Start: 1-cell embryos, 1-4 hours of development; Wash: 12 hours of development.

The Hoechst 33342 staining technique is valuable for differentiating apoptotic cells from healthy or necrotic cells (Zhivotosky & Orrenius, 2001). To assess the effect of 5AzaD at the molecular level in *Parhyale* embryos, I utilised live-cell imaging with Hoechst 33342 staining on both wild-type and 5AzaD-treated embryos. Despite only a small percentage of 5AzaD-treated embryos dying before 72 hours of development, live-cell imaging revealed distinct differences in the appearance of 5AzaD-treated embryos compared to the wild-type embryos starting at 48 hours of development. At 24 hours, the nuclei of wild-type and 5AzaD-treated embryos appeared similar (Figure 4A, B), but by 48 hours, the

nuclei in healthy, untreated cells were round and evenly stained, while the 5AzaD-treated embryos exhibited fragmented nuclei that were more intensely stained due to DNA condensation, indicating apoptosis of the cells (Figure 4C, D).

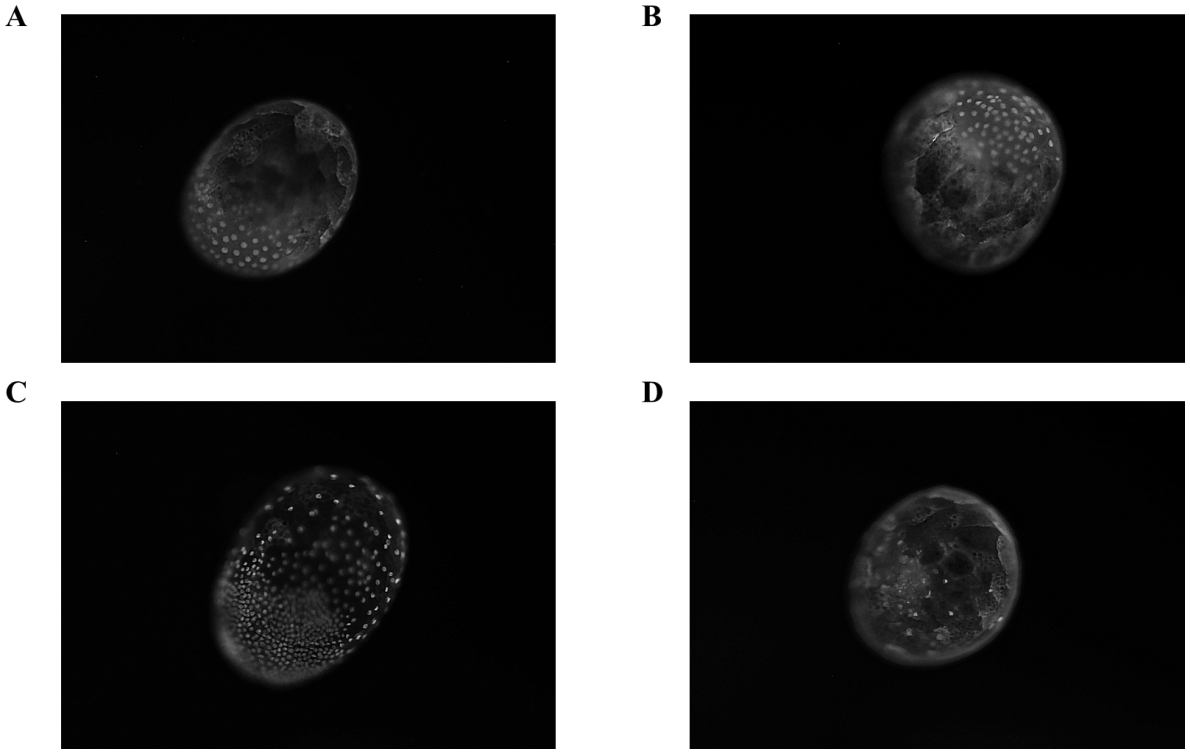


Figure 4. Images of embryos stained with Hoechst 33342. **(A)** Wild-type embryo at 24-hour. **(B)** 5AzaD-treated embryos at 24-hour. **(C)** wild-type embryo at 48-hour. **(D)** 5AzaD-treated embryos at 48-hour.

4.3.3 Summary of sequencing data

Based on the 5AzaD test results, I conducted total RNA-seq on 12-hour and 24-hour *Parhyale* embryos treated with either 5 μM 5AzaD or 0.0008% vehicle DMSO during the first 12 hours of development. Each group included three biological replicates, resulting in a total of 12 sequencing samples. The alignment information for total RNA-seq is shown in Table 1. All 12 samples exhibited an alignment rate above 80%, with more than 33 million reads mapped to the *Parhyale* transcriptome.

Table 1. Total RNA-seq results.

Time point	Treatment	Replicate	Clean reads	Mapped reads	Mapping rate (%)
12-hour	DMSO	1	51,468,297	45,441,359	88.29
		2	53,603,882	44,518,024	83.05
		3	86,052,518	71,122,406	82.65
	5AzaD	1	58,282,565	48,275,449	82.83
		2	40,232,711	33,356,941	82.91
		3	53,678,469	44,467,244	82.84
24-hour	DMSO	1	50,027,421	42,173,116	84.30
		2	40,878,666	34,493,418	84.38
		3	44,977,771	38,361,541	85.29
	5AzaD	1	42,440,165	35,819,499	84.40
		2	42,356,830	36,380,281	85.89
		3	43,057,949	37,240,820	86.49

Exon counts were obtained for each of the sequenced samples, a PCA plot and a correlation matrix of the raw counts between the 12 total RNA-seq datasets were generated (Figure 5). In general, samples from the same time point showed higher correlation regardless of the treatment, compared to samples from the other time point (Figure 5A). Furthermore, the PCA plot revealed that the 12-hour datasets of both 5AzaD and DMSO treatments clustered closely together, indicating similar transcriptional alterations at 12-hour of development. Notably, the third replicate of DMSO at 12-hour was slightly located outside the 12-hour cluster of developmental time (PC1, 54.64% of average variation), while still positioned within the 12-hour gene expression variation (PC2, 13.06% of average variation) and far from the two 24-hour clusters (Figure 5B).

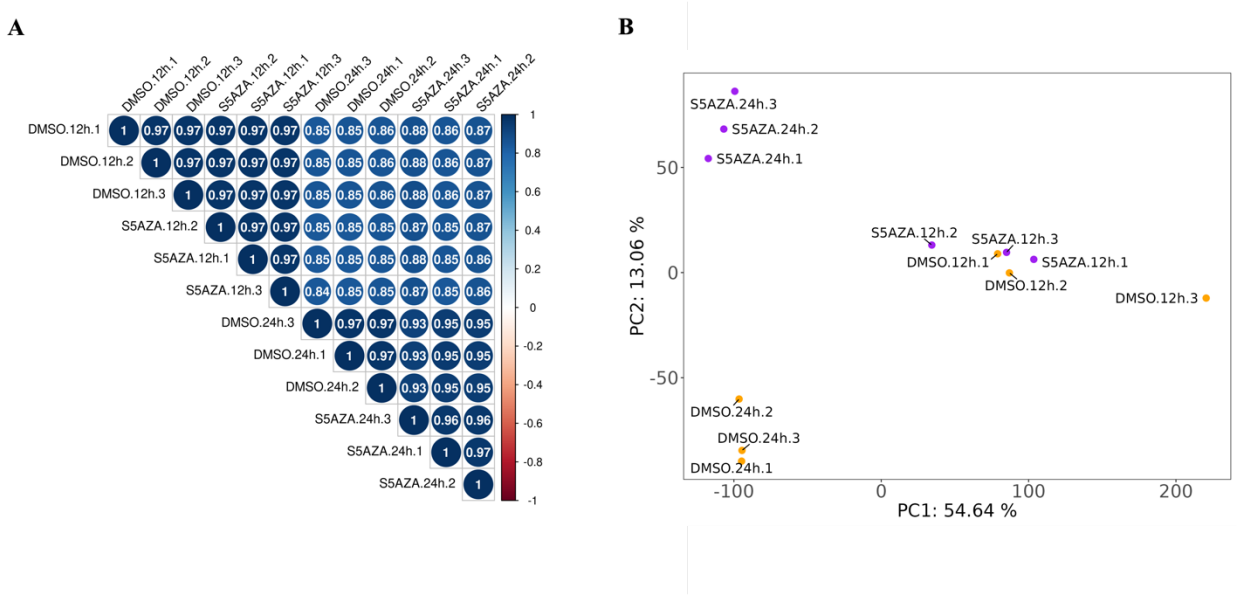


Figure 5. Systematic evaluation of RNA-seq data. **(A)** Correlogram showing the correlation score matrix among the 12 samples. **(B)** Principal component analysis of expression in the 12 samples. PC1, first principal component; PC2, second principal component. Each dot indicates a sample.

4.3.4 Correlation with gene expression of wild-type embryos

The embryos treated with the vehicle DMSO also showed reduced survival rates compared to the wild-type embryos, as determined by the dose test experiment. To further investigate the effect of 5AzaD treatment and the potential effect of DMSO on *Parhyale* embryos, I correlated the gene expression levels of wide-type embryos with those of 5AzaD-treated and DMSO-treated embryos, respectively. The results showed that the gene expression patterns in the 12-hour sampled embryos treated with 5AzaD and DMSO resembled those of wild-type embryos at 0-9 hours more closely than those at 12 hours (Figure 6A, B). In addition, the gene expression of 5AzaD-treated embryos at 24 hours was more closely related to wild-type embryos at 16 hours, while the expression of DMSO-treated embryos was more similar to wild-type 20-hour embryos than to wild-type 24-hour embryos (Figure 6C, D).

Moreover, employing PCA on the top 1000 most variable genes at the respective time points offered valuable insights into the variability within both the 5AzaD/DMSO treated

and the wild-type RNA-seq dataset. The results revealed that drug treatments represented the second source of gene expression variation (PC2, accounting for 18% of average variation), following developmental time (PC1, accounting for 70% of average variation) (Figure 7). Specifically, both 5AzaD and DMSO treatments induced transcriptional changes during the initial 12 hours of development, distinct from those of wild-type embryos (Figure 7A). Conversely, 5AzaD-treated 24-hour embryos exhibited the most pronounced transcriptional alterations compared to those exposed to DMSO or wild-type embryos (Figure 7B). Moreover, their expression variability clustered closely with that of wild-type embryos at 16 hours, consistent with the above correlation results (Figure 6).

These findings suggest that exposure to both 5AzaD and DMSO caused developmental delays in the embryos, with the delay being more pronounced in 5AzaD-exposed embryos. Similar delays in development due to exposure to 5AzaD have also been observed in several other species (Dasmahapatra & Khan, 2015; Kumar & Kim, 2017; Planques et al., 2021; Sasaki & Satoh, 2007).

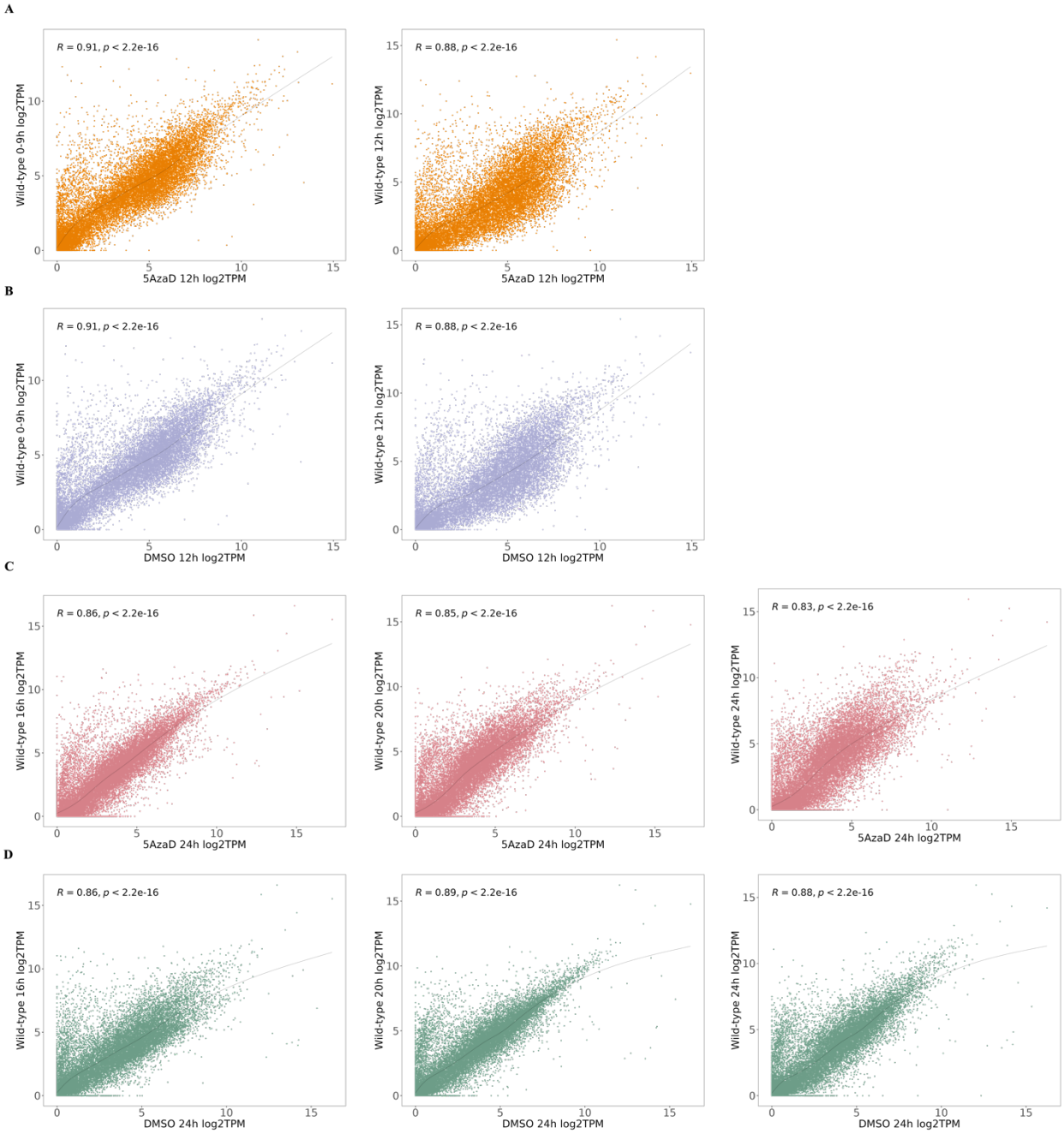


Figure 6. The correlation of gene expression between wild-type embryos and those treated with 5AzaD/DMSO at different time points. **(A)** Correlation between wide-type 0-9-hour (left) & 12-hour (right) and 5AzaD 12-hour embryos. **(B)** Correlation between wide-type 0-9-hour (left) & 12-hour (right) and DMSO 12-hour embryos **(C)** Correlation between wide-type 16-hour (left), 20-hour (middle) & 24-hour (right) and 5AzaD 24-hour embryos **(D)** Correlation between wide-type 16-hour (left), 20-hour (middle) & 24-hour (right) and DMSO 24-hour embryos.

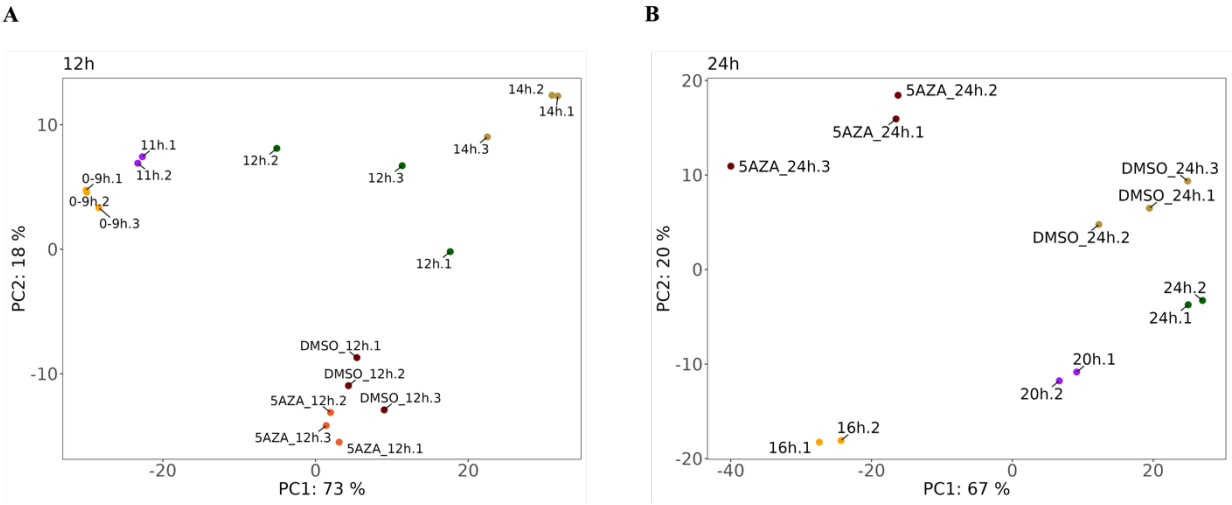


Figure 7. Principal component analysis generated using the top 1000 most variable genes of 5AzaD/DMSO treated embryos compare to the wild-type embryos for **(A)** 12-hour and **(B)** 24-hour treated embryos.

4.3.5 Gene expression analysis

4.3.5.1 Differential expression analysis

To explore how 5AzaD treatment impacts transcription during the first 24 hours, I conducted differential expression analysis to identify upregulated and downregulated genes between 5AzaD and vehicle-DMSO treated embryos at two time points. In the 12-hour gene set, only 24 genes showed differential expression, with 19 genes being downregulated and 5 genes being upregulated (Figure 8A). This small number of differentially expressed genes at 12 hours could be due to the delayed development of these embryos, resulting in their actual developmental stage being before the MZT and ZGA, thereby limiting the transcriptome activity of the zygotic genome. In contrast, in the 24-hour gene set, a substantial number of differentially expressed genes were detected, with 1988 genes significantly downregulated and 1698 genes significantly upregulated (Figure 8B).

In addition to comparing the 5AzaD-treated and DMSO-treated embryos, I also performed a differential expression analysis comparing DMSO-treated embryos with wild type

embryos. Surprisingly, there were genes showing differential expression at both 12-hour and 24-hour time points. Specifically, at 12 hours, 347 genes were significantly downregulated, and 165 genes were significantly upregulated. Moreover, at 24 hours, 710 genes were significantly downregulated, and 470 genes were significantly upregulated (Figure 8C, D). These results, combined with previous correlational findings suggest that both the 5AzaD and DMSO treatments may affect gene expression during *Parhyale* early development.

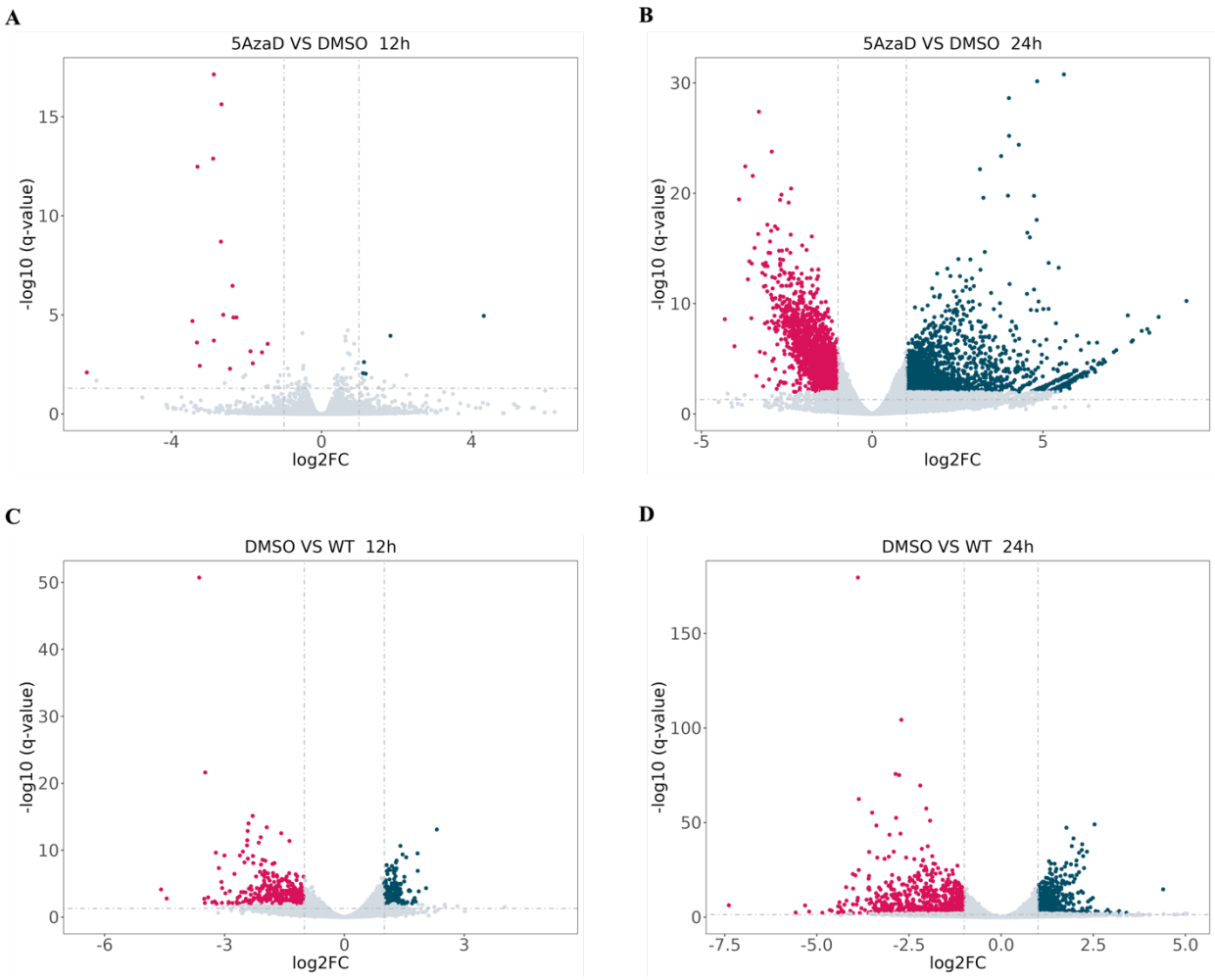


Figure 8. Differential expression analysis of **(A)** 5AzaD and DMSO embryos at the 12-hour **(B)** 5AzaD and DMSO embryos at 24-hour **(C)** DMSO and wild-type embryos at the 12-hour and **(D)** DMSO and wild-type embryos at 24-hour. WT: wild type.

4.3.5.2 DNA methylation of differentially expressed genes

Next, I integrated previous EM-seq data with the current total RNA-seq data to evaluate the methylation levels of the differentially expressed genes at both the 12-hour and 24-hour time points. Despite the limited changes in expression at 12-hour, the downregulated genes at this time point exhibited significantly higher methylation levels compared to upregulated and non-differentially expressed genes before and during MZT (Figure 9A). Meanwhile, the 24-hour downregulated genes showed a tendency to maintain elevated methylation levels both prior to and after MZT/ZGA (Figure 9B). I then further analysed the methylation patterns of the gene body by dividing it into exons and introns. At the 12-hour time point, there was no significant difference in exon/intron methylation levels between upregulated and downregulated genes (Figure 9C). However, at 24-hour time point, all DEGs showed higher intron methylation levels compared to exon methylation levels (Figure 9D).

Additionally, I calculated the proportion of methylated genes in the three groups of differentially expressed genes (downregulated, upregulated and non-differentially expressed) at the two time points. These methylated genes were identified from the three EM-seq datasets, which were defined as having gene body methylation of at least 3%. The results indicated that there were more methylated genes in the downregulated gene sets at both the 12-hour and 24-hour time points compared to the upregulated and non-differentially expressed gene sets (Figure 10A, B). Furthermore, after administering 5AzaD, both methylated and unmethylated genes exhibited changes in expression levels. Among the unmethylated genes, approximately 70% were upregulated and 30% were downregulated. However, the trend was opposite for the methylated genes, which displayed greater variability, with 80% of the methylated genes being downregulated and 20% being upregulated (Figure 10C).

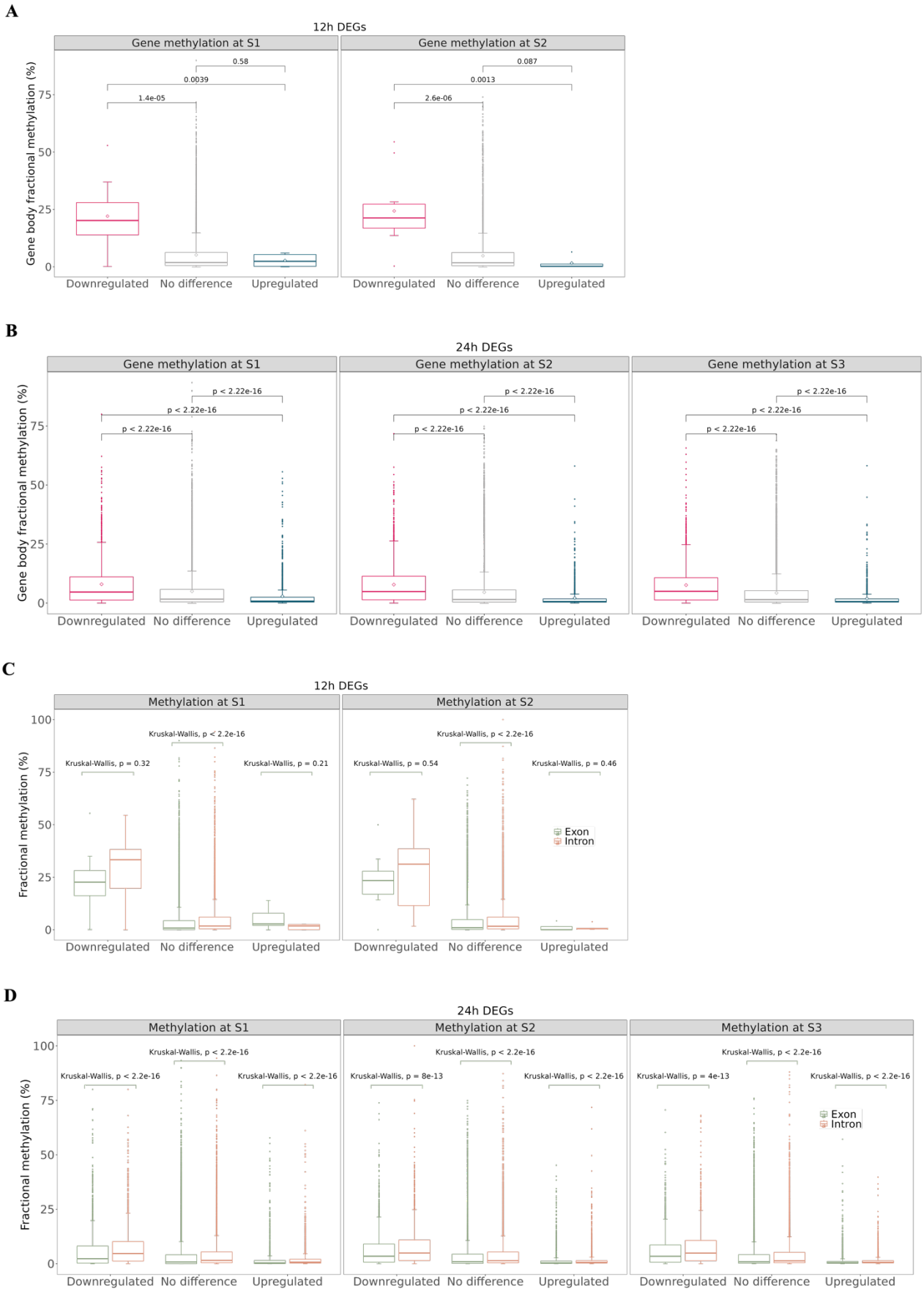

A
12h DEGs
Gene methylation at S1
Gene methylation at S2
Gene body fractional methylation (%)
0.58
0.0039
1.4e-05
0.087
0.0013
2.6e-06
Downregulated
No difference
Upregulated
B
24h DEGs
Gene methylation at S1
Gene methylation at S2
Gene methylation at S3
p < 2.22e-16
C
12h DEGs
Methylation at S1
Methylation at S2
Fractional methylation (%)
Kruskal-Wallis, p < 2.2e-16
Kruskal-Wallis, p = 0.32
Kruskal-Wallis, p = 0.21
Kruskal-Wallis, p = 0.54
Kruskal-Wallis, p = 0.46
Exon
Intron
D
24h DEGs
Methylation at S1
Methylation at S2
Methylation at S3
Kruskal-Wallis, p < 2.2e-16
Kruskal-Wallis, p = 8e-13
Kruskal-Wallis, p = 4e-13

Figure 9. Gene body fractional methylation of differentially expressed genes at the corresponding and previous developmental stages for embryos treated with 5AzaD and DMSO at **(A)** 12-hour time point and **(B)** 24-hour time point. The fractional methylation of exon and intron of DEGs at **(C)**12-hour and **(D)** 24-hour time point. No difference: non-differentially expressed genes.

By analysing the methylation changes of the genes that were not only downregulated at 24-hour time point but also classified as methylated genes in one of the three EM-seq stages (S1, S2, S3), I found that more than 70% of these genes reached their highest methylation level during the pre-MZT and MZT period, corresponding to the S1 or S2 stage, respectively (Figure 10D).

These findings suggest that during early development, downregulated genes tend to have higher methylation levels compared to upregulated genes, with the highest methylation levels occurring before or during MZT. Additionally, the 5AzaD treatment affects gene expression in two different ways, with methylated genes displaying decreased transcription and unmethylated genes exhibiting increased transcription. Therefore, it is expected that exposing *Parhyale* embryos to 5AzaD results in a loss of DNA methylation, which may then impact the regulation of gene expression during MZT, particularly in activating heavily methylated genes. Furthermore, exposure to 5AzaD during the initial 12 hours of *Parhyale* development primarily impacts transcriptional activity in the subsequent 12 hours.

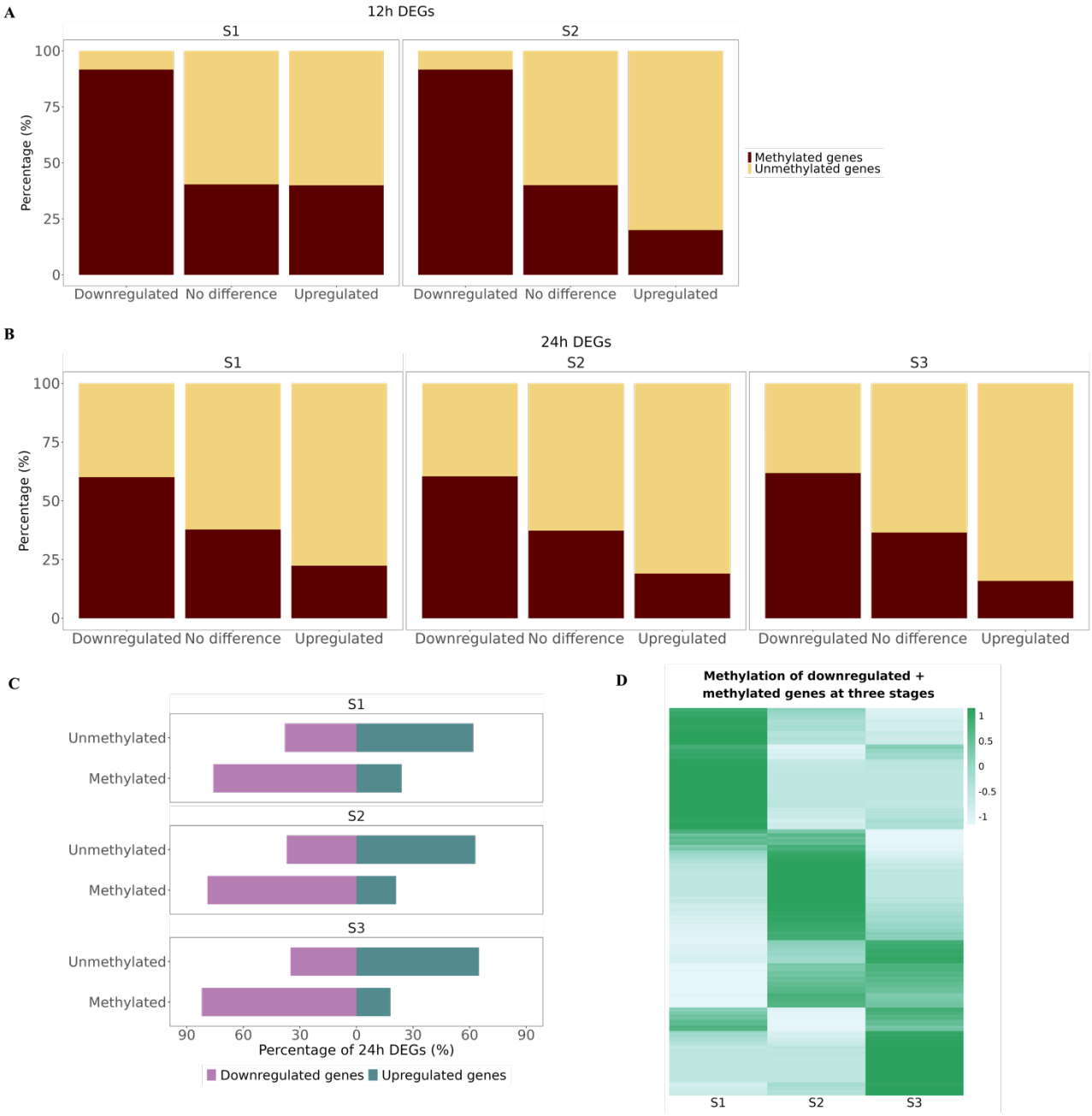


Figure 10. Percentage of methylated genes in **(A)** 12-hour DEGs and **(B)** 24-hour DEGs. **(C)** Changes in the expression of methylated and unmethylated genes within the two DEG sets for the 24-hour time point. **(D)** Heatmap illustrating the methylation dynamics during early development of genes that were both downregulated at 24 hours and identified as methylated genes in one of the three EM-seq stages. No difference: non-differentially expressed genes.

4.3.5.3 Functional classification of DEGs

To identify the characteristic biological attributes of differentially expressed genes at two distinct time points, I performed GO enrichment analyses on these DEGs. Using Orthofinder (Emms & Kelly, 2019), Dr. Manuel Jara Espejo determined the orthologs for *Parhyale* in comparison to *Drosophila melanogaster*, *Caenorhabditis elegans*, *Xenopus tropicalis*, and human. By merging the DEGs with the *Parhyale* orthologs, I identified 1577 24-hour DEGs with clear orthologs and UniProt IDs (Table 2). These DEGs were then subjected to GO enrichment analyses, which were categorised into biological process, molecular function, and cellular component. Out of the 1577 profiled 24-hour DEGs, 775 of these genes were assigned to a total of 324 GO annotations from *Drosophila melanogaster*. The top 10 most enriched GO annotations for the three GO categories of downregulated and upregulated genes are shown in Figure 11. The enriched GO terms were different between the up and downregulated genes.

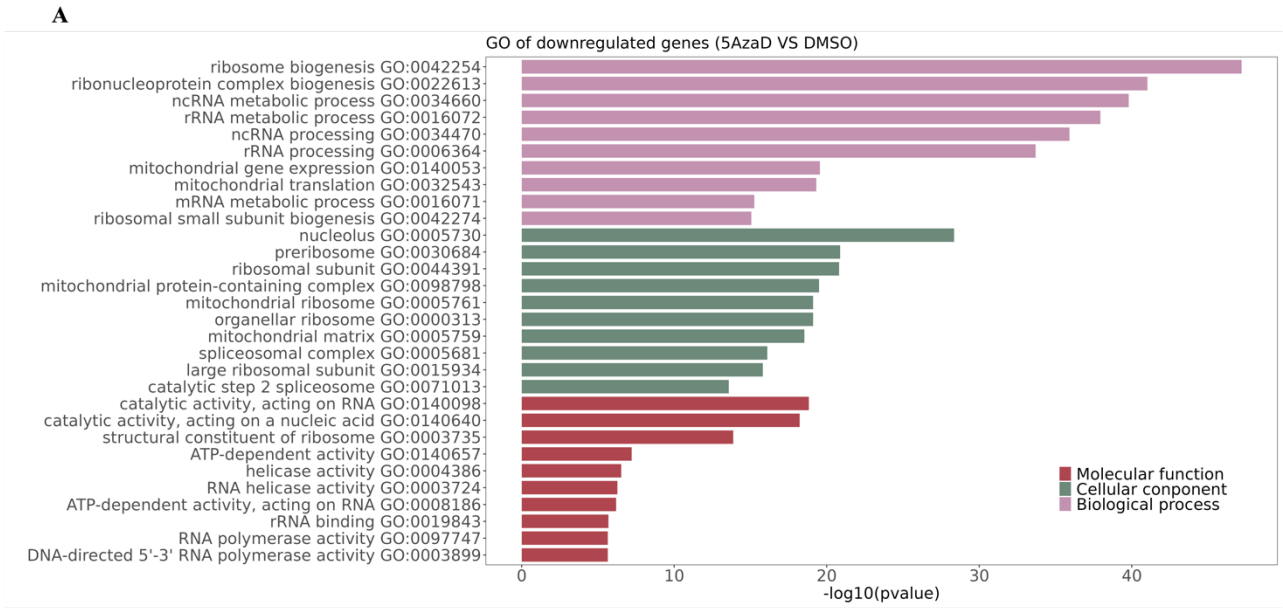
Table 2. Summary of DEGs along with their associated Gene Ontology terms.

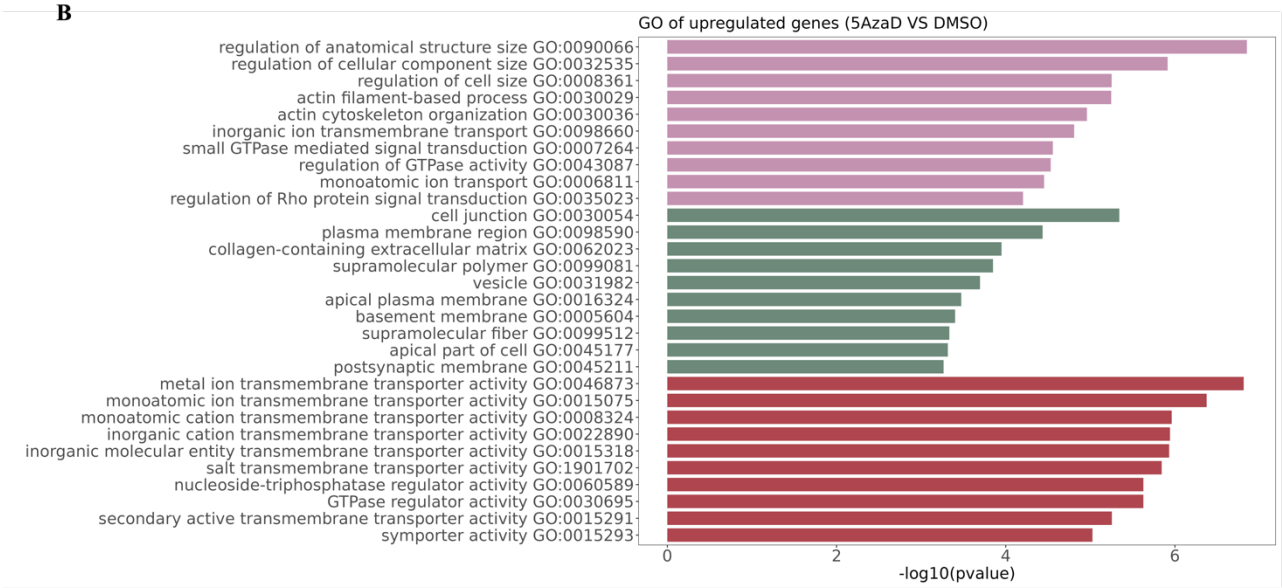
		Down	%	Up	%	All	%
	Total	1988	-	1698	-	3686	-
24-hour	**Orthologous+**	1022	51.4	564	33.2	1557	42.2
	TF +	105	9.5	47	8.2	152	9.1
	Total	19	-	5	-	24	-
12-hour	**Orthologous+**	0	0	3	60	3	12.5
	TF+	0	0	1	33.3	1	33.3

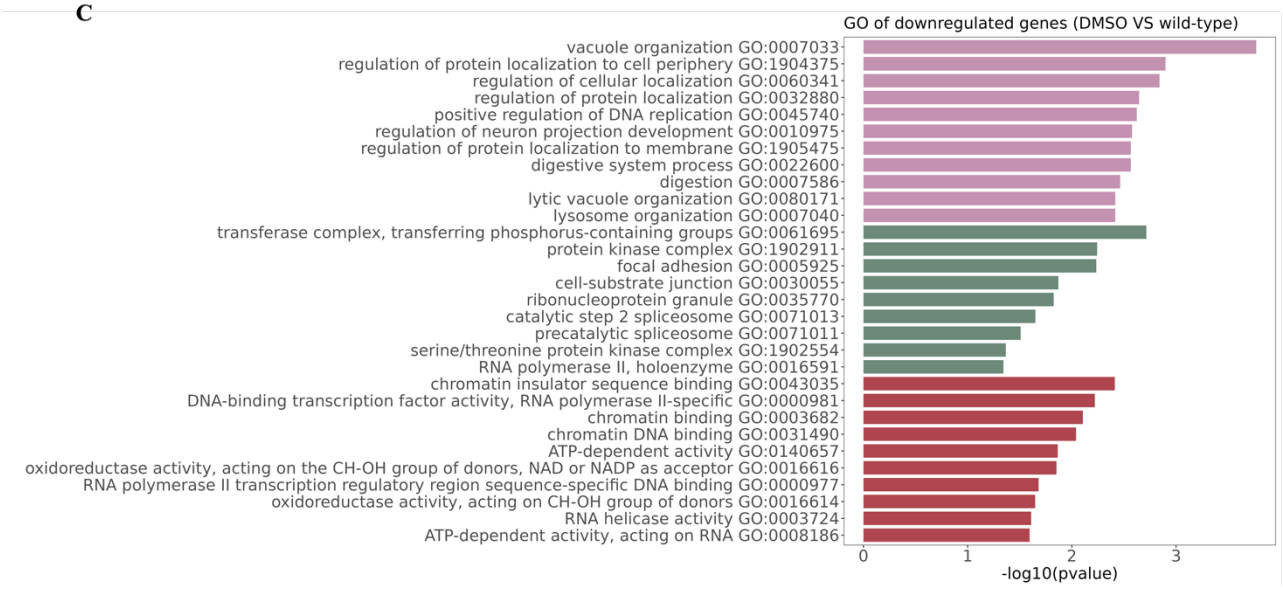
Concerning biological processes, downregulated genes showed significant enrichment of terms related to metabolism and universal housekeeping functions of gene expression

and translation (e.g. GO:0042254, GO:0034660, GO:0034470, GO:0140053). In contrast, upregulated genes demonstrated significant enrichment of terms associated with cytoskeletal organization and signalling (e.g. GO:0090066, GO:0032535, GO:0098660, GO:0007264). Moreover, the GO cellular components that were most enriched for downregulated genes were nucleus and mitochondria (e.g. GO:0005730, GO:0098798), whereas membrane-related structures and extracellular matrix components were most abundant for upregulated genes (e.g. GO:0030054, GO:0098590, GO:0062023). Additionally, downregulated genes were highly associated with RNA helicase, catalytic and polymerase activity (e.g. GO:0140098, GO:0004386, GO:0097747), emphasising their role in transcription and RNA processing. While upregulated genes were enriched in transporter activity and ion transport across membranes (e.g. GO:0046873, GO:0008324, GO:0022890). (Figure 11A, B)

Furthermore, I also conducted GO analysis on DEGs between DMSO-treated and wild-type embryos. The downregulated genes showed enrichment in chromatin structure, RNA metabolism related functions and transcriptional regulation processes (e.g. GO:0043035, GO:0000981, GO:0003682, GO:0003724). Moreover, there was enrichment in cellular components related to RNA processing and protein complexes (e.g. GO:0061695, GO:1902911, GO:0035770, GO:0071013). Biologically, the downregulated genes are involved in cellular organization and metabolism (e.g. GO:0007033, GO:0060341, GO:0022600). In contrast, the upregulated genes display enrichment in distinct molecular functions, such as GTPase binding, transporter activity and enzyme binding (e.g. GO:0022804, GO:0019899, GO:0031267). Additionally, there was enrichment in cellular components, intracellular compartments and actin-based structures (e.g. GO:0005938, GO:0032153, GO:0000138, GO:0098862). These upregulated genes also play roles in diverse processes, including homeostatic mechanisms, metabolism and cellular responses (e.g. GO:0048878, GO:0055088, GO:0046890, GO:0097306). (Figure 11C, D)


A
GO of downregulated genes (5AzaD VS DMSO)
ribosome biogenesis GO:0042254
ribonucleoprotein complex biogenesis GO:0022613
ncRNA metabolic process GO:0034660
rRNA metabolic process GO:0016072
ncRNA processing GO:0034470
rRNA processing GO:0006364
mitochondrial gene expression GO:0140053
mitochondrial translation GO:0032543
mRNA metabolic process GO:0016071
ribosomal small subunit biogenesis GO:0042274
nucleolus GO:0005730
preribosome GO:0030684
ribosomal subunit GO:0044391
mitochondrial protein-containing complex GO:0098798
mitochondrial ribosome GO:0005761
organellar ribosome GO:0000313
mitochondrial matrix GO:0005759
spliceosomal complex GO:0005681
large ribosomal subunit GO:0015934
catalytic step 2 spliceosome GO:0071013
catalytic activity, acting on RNA GO:0140098
catalytic activity, acting on a nucleic acid GO:0140640
structural constituent of ribosome GO:0003735
ATP-dependent activity GO:0140657
helicase activity GO:0004386
RNA helicase activity GO:0003724
ATP-dependent activity, acting on RNA GO:0008186
rRNA binding GO:0019843
RNA polymerase activity GO:0097747
DNA-directed 5'-3' RNA polymerase activity GO:0003899
Molecular function
Cellular conponent
Biological process
0
10
20
30
40
-log10(pvalue)


B
GO of upregulated genes (5AzaD VS DMSO)
regulation of anatomical structure size GO:0090066
regulation of cellular component size GO:0032535
regulation of cell size GO:0008361
actin filament-based process GO:0030029
actin cytoskeleton organization GO:0030036
inorganic ion transmembrane transport GO:0098660
small GTPase mediated signal transduction GO:0007264
regulation of GTPase activity GO:0043087
monoatomic ion transport GO:0006811
regulation of Rho protein signal transduction GO:0035023
cell junction GO:0030054
plasma membrane region GO:0098590
collagen-containing extracellular matrix GO:0062023
supramolecular polymer GO:0099081
vesicle GO:0031982
apical plasma membrane GO:0016324
basement membrane GO:0005604
supramolecular fiber GO:0099512
apical part of cell GO:0045177
postsynaptic membrane GO:0045211
metal ion transmembrane transporter activity GO:0046873
monoatomic ion transmembrane transporter activity GO:0015075
monoatomic cation transmembrane transporter activity GO:0008324
inorganic cation transmembrane transporter activity GO:0022890
inorganic molecular entity transmembrane transporter activity GO:0015318
salt transmembrane transporter activity GO:1901702
nucleoside-triphosphatase regulator activity GO:0060589
GTPase regulator activity GO:0030695
secondary active transmembrane transporter activity GO:0015291
symporter activity GO:0015293
0
2
4
6
-log10(pvalue)


C
GO of downregulated genes (DMSO VS wild-type)
vacuole organization GO:0007033
regulation of protein localization to cell periphery GO:1904375
regulation of cellular localization GO:0060341
regulation of protein localization GO:0032880
positive regulation of DNA replication GO:0045740
regulation of neuron projection development GO:0010975
regulation of protein localization to membrane GO:1905475
digestive system process GO:0022600
digestion GO:0007586
lytic vacuole organization GO:0080171
lysosome organization GO:0007040
transferase complex, transferring phosphorus-containing groups GO:0061695
protein kinase complex GO:1902911
focal adhesion GO:0005925
cell-substrate junction GO:0030055
ribonucleoprotein granule GO:0035770
catalytic step 2 spliceosome GO:0071013
precatalytic spliceosome GO:0071011
serine/threonine protein kinase complex GO:1902554
RNA polymerase II, holoenzyme GO:0016591
chromatin insulator sequence binding GO:0043035
DNA-binding transcription factor activity, RNA polymerase II-specific GO:0000981
chromatin binding GO:0003682
chromatin DNA binding GO:0031490
ATP-dependent activity GO:0140657
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor GO:0016616
RNA polymerase II transcription regulatory region sequence-specific DNA binding GO:0000977
oxidoreductase activity, acting on CH-OH group of donors GO:0016614
RNA helicase activity GO:0003724
ATP-dependent activity, acting on RNA GO:0008186
0
1
2
3
-log10(pvalue)

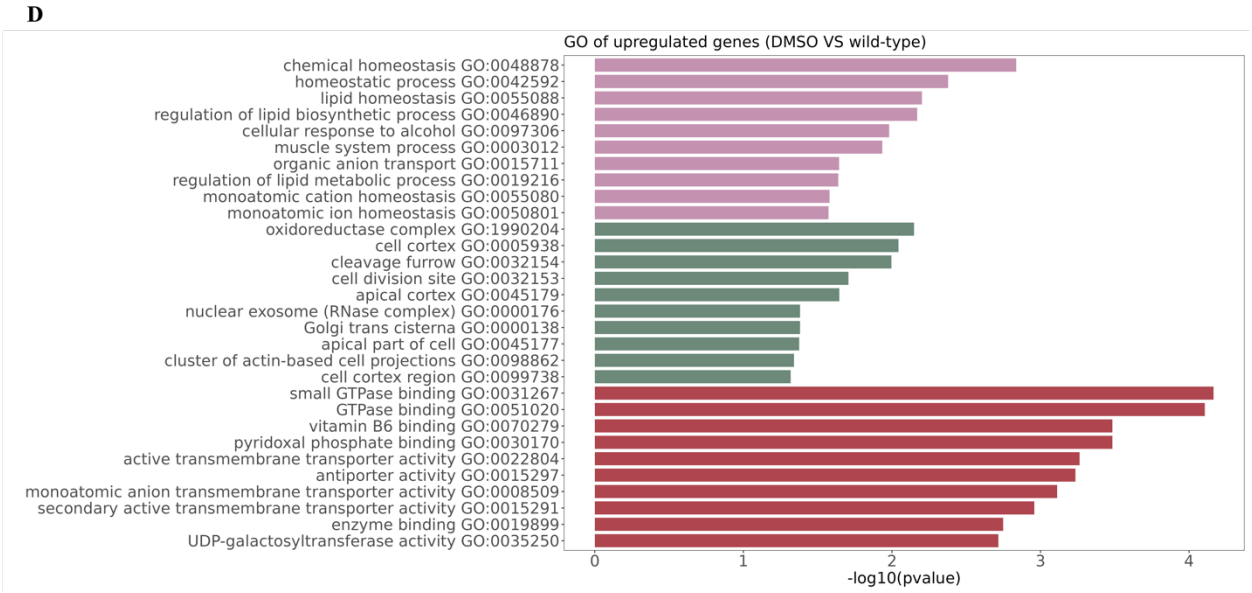


Figure 11. Gene Ontology (GO) enrichment analysis for **(A)** downregulated and **(B)** upregulated genes of 24-hour time point between 5AzaD and DMSO treated embryos, **(C)** downregulated and **(D)** upregulated genes between DMSO and wild-type embryos of 24-hour time point. GO terms categories: Biological Processes, Cellular Components, and Molecular Functions. Significantly enriched GO terms were selected based on a FDR < 0.05.

Based on the comparison, there are some commonalities between the GO terms enriched in both comparisons, such as the involvement of downregulated genes in transcriptional regulation processes (GO:0003724). DMSO may confound the identification of 5AzaD-treated DEGs by influencing gene expression and cellular processes in a manner that overlaps with 5AzaD-induced changes.

However, each comparison also highlights distinct biological processes and molecular functions associated with the respective treatments. For examples, DMSO treatment alone induces changes in cellular processes such as ion homeostasis, lipid metabolism, and cellular organization. These effects are independent of 5AzaD treatment and reflect the inherent biological response to DMSO. Furthermore, 5AzaD treatment leads to alterations in ribosome biogenesis and translation, indicating a specific impact on gene expression and protein synthesis machinery. These effects are distinct from those induced by DMSO alone. Nevertheless, while DMSO may influence some gene

expression changes, it does not appear to significantly confound or define the identified DEGs between 5AzaD and DMSO-treated embryos.

Checking the *Parhyale* orthologs of DEGs revealed that 153 genes, representing over 9% of the DEGs (48 upregulated and 105 downregulated) at both time points were found to be TF domain-containing genes (Table 2). In addition, regardless of the DEG sets, 58 TF domain-containing genes have been assigned a putative biological function, with over half of these genes participating in developmental processes (Table 3).

Table 3. The list of TF domain-containing genes within the DEGs set involved in developmental processes. Genes were named based on *Drosophila* orthologs.

Downregulated		**Upregulated**	
Gene id	Gene name	Gene id	Gene name
MSTRG.62248	fkh	MSTRG.31629	Mad
MSTRG.61821	mor	MSTRG.39411	ct
MSTRG.30536	aop	MSTRG.33248	dati
MSTRG.6007	lolal	MSTRG.21192	pan
MSTRG.44991	Acf	MSTRG.52171	Hr4
MSTRG.29254	btd	MSTRG.67078	abd-A
MSTRG.43156	pdm2	MSTRG.41578	lbe
MSTRG.42806	Kdm2	MSTRG.17588	Hand
MSTRG.41381	mTerf3	MSTRG.61833	Fer2
MSTRG.12352	slbo	MSTRG.38271	Blimp-1
MSTRG.20441	eve	MSTRG.67067	Abd-B
MSTRG.67091	Dfd	MSTRG.29989	ERR
MSTRG.67099	lab	MSTRG.58064	Ptx1
MSTRG.33167	Ets98B	MSTRG.67084	Scr
MSTRG.20408	crp	MSTRG.51345	al
		MSTRG.56994	Sp1
		MSTRG.14471	Sox102F

According to the findings, 5AzaD treatment may have a direct role in repressing the transcription of downregulated genes, which are primarily involved in housekeeping functions. As a result, the mis-regulation of these genes may indirectly affect upregulated genes that were mostly involved in signal transduction within cells. Overall, should 5AzaD treatment indeed induce DNA methylation loss, DNA methylation appears pivotal for the precise temporal dynamics of gene expression. Furthermore, this loss of DNA methylation seems to have a significant impact on genes involved in transcription, cellular processes, and development, resulting in delays and defects in development.

4.3.5.4 Correlation between DEGs and ZGA genes

To further explore the influence of 5AzaD treatment on the activation of genes during zygotic transcription, I tallied the number of genes that were maternally expressed or zygotically activated (zygotic-only and zygotic-maternal) within the two sets of differentially expressed genes at the 12-hour and 24-hour time points. I then employed the Fisher's exact test to determine if there was a statistically significant association between maternal/ZGA genes within the two DEG sets.

The results of this investigation revealed intriguing patterns. At the 12-hour time point, out of the 19 downregulated genes, 15 were zygotic-only genes. This proportion was significantly higher than the proportion of zygotic-only genes in the upregulated gene set at 12 hours (Table 4). Similarly, at the 24-hour time point, the downregulated gene set displayed a noteworthy enrichment of both zygotic-only and zygotic-maternal genes when compared to the upregulated gene set. Furthermore, the majority of upregulated genes at 24 hours were non-ZGA genes, while the proportion of maternally expressed genes did not exhibit a significant difference between the downregulated and upregulated gene sets at this time point (Table 4 and Figure 12). Based on these findings, it can be inferred that 5AzaD treatment might impact the activation of zygotically activated genes from the early MZT period.

Table 4. Summary of the number of maternally expressed and ZGA genes presented in the DEGs of 12- and 24-hour time points.

Time point	Gene type	Downregulated	Upregulated	Fisher test between DEGs
12-hour	Maternally expressed	0	3	p = 0.004941**
	Zygotic-only	15	1	p = 0.02767*
	Zygotic-Maternal	0	0	-
	Other	4	1	p = 1
24-hour	Maternally expressed	1451	937	p = 0.1511
	Zygotic-only	285	24	p < 2.2e-16***
	Zygotic-Maternal	729	65	p < 2.2e-16***
	Other	117	711	p < 2.2e-16***

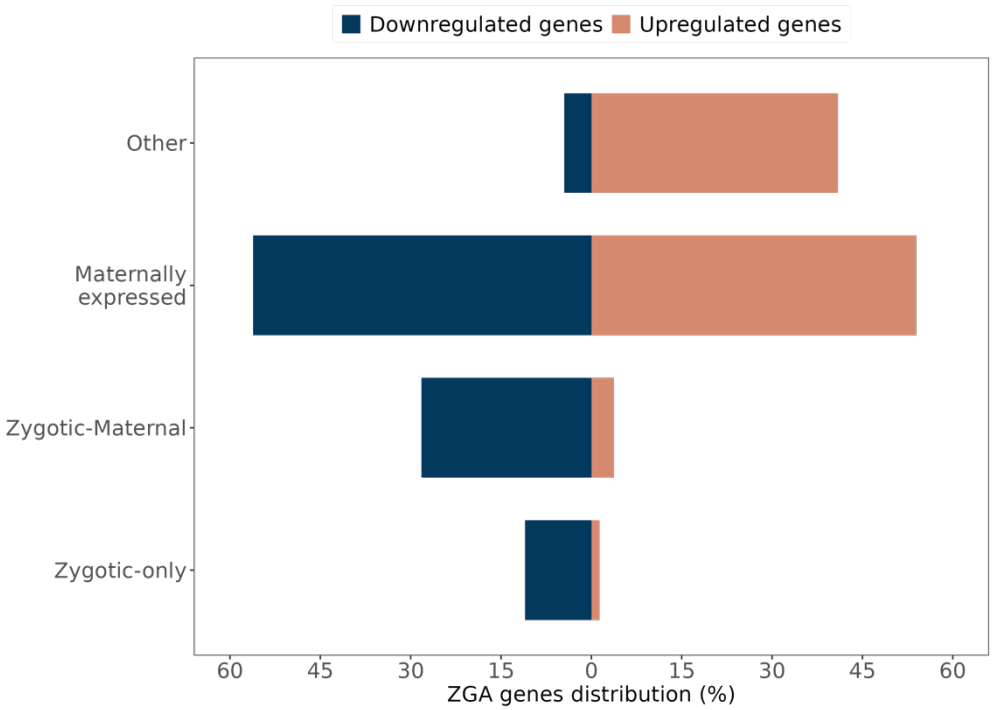


Figure 12. The distribution of maternally expressed and ZGA genes within the upregulated and downregulated gene sets at 24-hour time point.

Since downregulated genes were generally more methylated compared to upregulated and non-differentially expressed genes, it is worthwhile to explore the methylation patterns of maternally expressed and ZGA genes within the up and downregulated gene sets, respectively. Surprisingly, zygotic-maternal genes in the 24-hour downregulated gene set exhibited significantly higher methylation levels than other downregulated genes (Figure 13A). However, no significant differences were observed in methylation levels between maternally expressed and ZGA genes within the 24-hour upregulated gene set, or between zygotic-only and other genes within the 12-hour downregulated gene set (Figure 13B, C). These findings suggest that if 5AzaD treatment does induce DNA methylation loss, methylation could be important in triggering the activation of zygotic-maternal genes. Moreover, the methylation deposition on these genes appears crucial for accurately regulating gene expression during MZT/ZGA.

I then examined the wild-type expression levels and the expression dynamics of those zygotic-maternal genes that were either downregulated or upregulated at the 24-hour time point following 5AzaD administration. Interestingly, I observed that zygotic-maternal genes that were downregulated in 5AzaD-treated embryos exhibited a decrease in expression before MZT and during the minor wave of zygotic activation in wild-type embryos, followed by an increase during the major wave (Figure 14A). Conversely, the expression of upregulated zygotic-maternal genes in 5AzaD-treated embryos exhibited the opposite trend in wild-type embryos, with expression remaining stable until the major wave and then decreasing (Figure 14B). These findings suggest that zygotic-maternal genes in the two DEG sets exhibit distinct expression dynamics during MZT and ZGA in wide-type embryos.

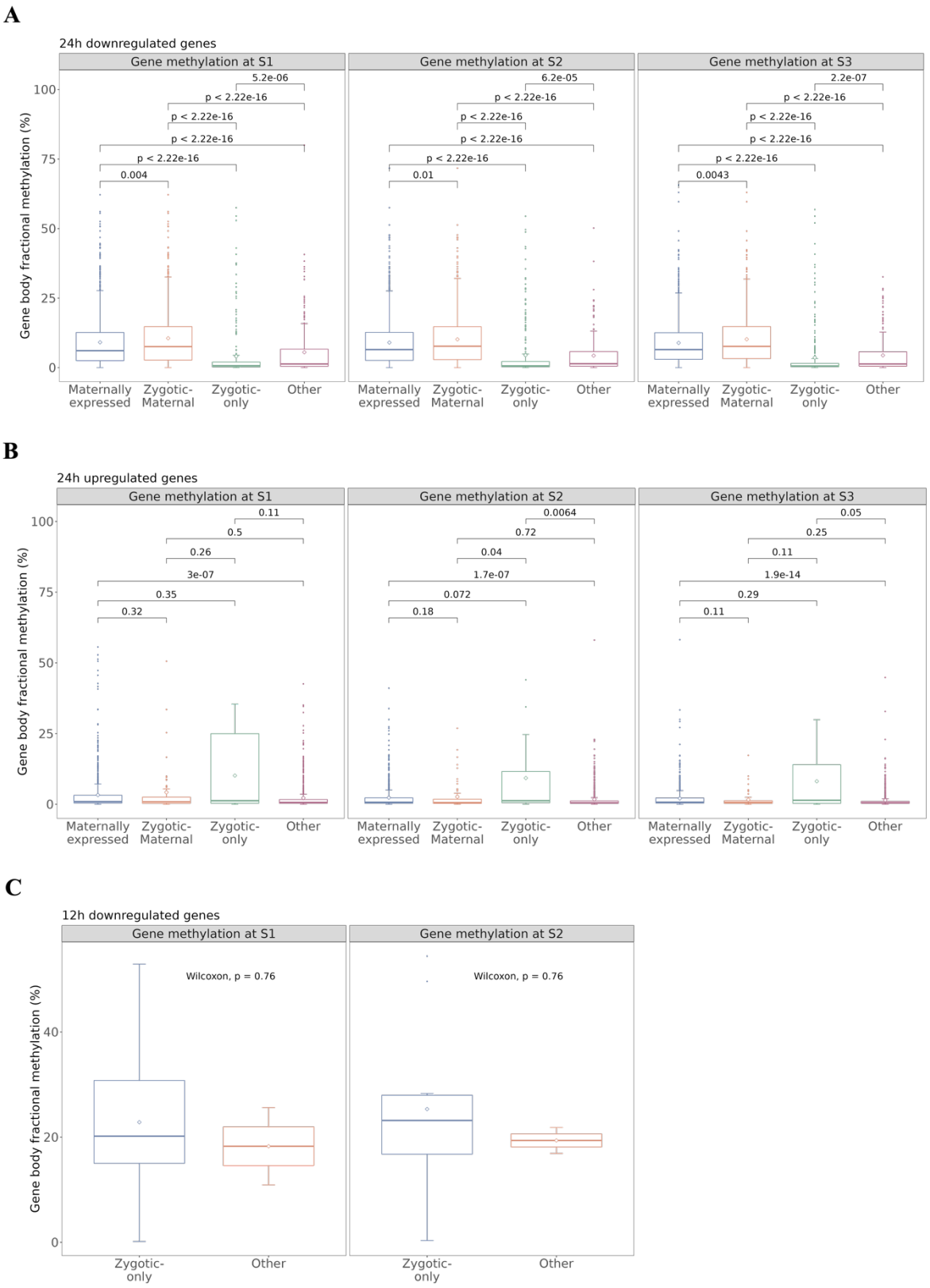


Figure 13. Gene body methylation at different EM-seq stages of maternally expressed genes, zygotic-maternal genes, zygotic-only genes and other genes within **(A)** 24-hour downregulated gene set, **(B)** 24-hour upregulated gene set, and **(C)** 12-hour downregulated gene set.

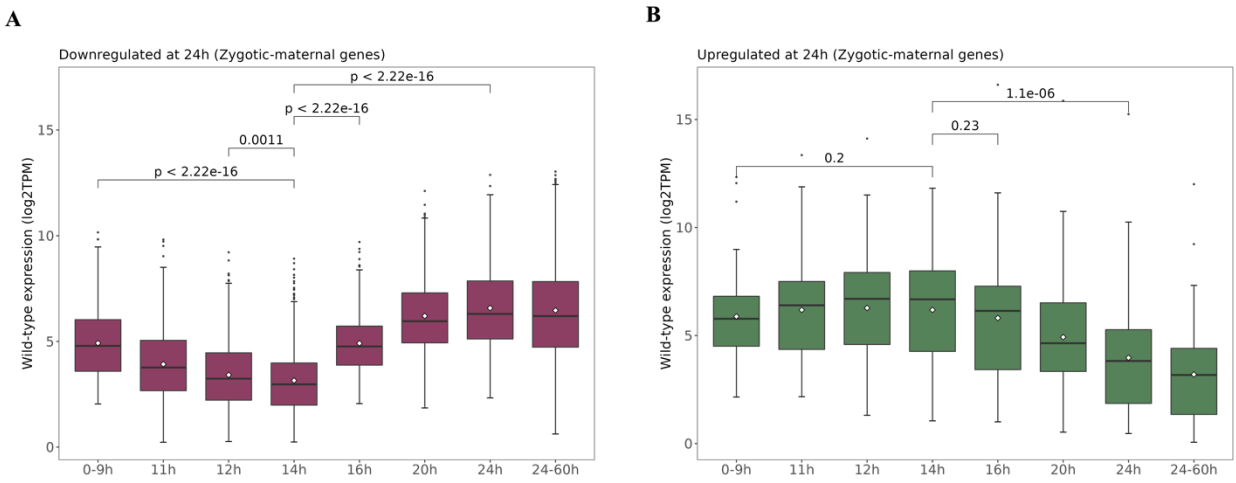


Figure 14. The wild-type normal expression dynamics of zygotic-maternal genes that were **(A)** downregulated and **(B)** upregulated at 24-hour time point following 5AzaD treatment.

Overall, the downregulated genes affected by 5AzaD treatment were mostly ZGA genes that had highly methylated CpGs. By exposing embryos to 5AzaD prior to MZT, the expression of these ZGA genes was reduced from the major wave of zygotic genome activation. If 5AzaD treatment does indeed lead to DNA methylation loss, these findings then suggest that DNA methylation may have a direct regulatory effect on a subset of ZGA genes, and that this regulation occurs before MZT, which is essential for the activation of these genes during zygotic genome activation. Failure to activate these ZGA genes may then influence the repression of other low methylation genes, resulting in long-term harm to embryonic development in *Parhyale*.

4.3.6 Repeat expression analysis

4.3.6.1 Expressed repeats

In addition to gene expression analysis, I used SalmonTE to investigate differences in the transcription of repetitive elements between 5AzaD and DMSO treated embryos at the two time points. By aligning reads to repetitive sequences, I found that, on average, 23% of the total RNA-seq reads were mapped to repetitive sequences across all the samples

(Table 5). I then analysed the total read counts of the expressed repeats, defined as repeats with at least one covered read. In total, 791,671 and 778,805 repeat loci were detected in at least one treatment of the 12-hour and 24-hour datasets, respectively. The unclassified repeats were the most highly expressed, accounting for over 64% of all expressed repeats. Of those repeats with a clear class assignment, LINEs were the most highly expressed, comprising over 15% of expressed repeats, followed by DNA transposons, which accounted for over 9%. However, LTRs only accounted for less than 3% of the expressed repeats. Other repeat classes that were expressed in the datasets, such as rolling-circle (RC) transposons, simple repeats, satellite DNAs and SINEs, constituted less than 3% of all the expressed repeats (Table 6).

Table 5. Alignment results of repeat read.

Time point	Treatment	Replicate	Clean reads	Mapped reads	%
12-hour	5AzaD	1	9,815,080	51,468,297	19.07
		2	10,298,455	53,603,882	19.21
		3	17,029,851	86,052,518	19.79
	DMSO	1	11,622,312	58,282,565	19.94
		2	10,298,455	53,603,882	19.21
		3	17,029,851	86,052,518	19.79
24-hour	5AzaD	1	13,318,433	50,027,421	26.62
		2	10,718,037	40,878,666	26.22
		3	11,845,757	44,977,771	26.34
	DMSO	1	12,132,073	42,440,165	28.59
		2	10,846,555	42,356,830	25.61
		3	11,975,521	43,057,949	27.81

Comparing the total normalized read counts of all expressed repeats at the 12-hour time point, only unclassified repeats and DNA transposons exhibited significant differences

between the two treatment groups (Figure 15A). However, at the 24-hour time point, embryos treated with DMSO showed significantly higher expression levels of unclassified repeats, DNA transposons, LINEs, LTRs, simple repeats, and satellite DNAs compared to embryos treated with 5AzaD (Figure 15B).

Table 6. The number of expressed repeats present in each class.

Family	DMSO-12-hour	5AzaD-12-hour	DMSO-24-hour	5AzaD-24-hour
All repeats	221,359	192,165	570,312	586,640
Unclassified	149,275 (67.4%)	130,454 (67.9%)	374,835 (65.7%)	377,119 (64.3%)
LINEs	34,294 (15.5%)	29,314 (15.3%)	96,519 (16.9%)	101,972 (17.4%)
DNA transposons	20,740 (9.4%)	17,720 (9.2%)	57,665 10.1%)	64,477 (11%)
LTRs	4,000 (1.8%)	2,987 (1.6%)	14,837 (2.6%)	14,502 (2.5%)
Simple repeats	7,359 (3.3%)	6,815 (3.6%)	11,000 (2%)	11,499 (2%)
Satellite DNAs	3,364 (1.5%)	2,923 (1.4%)	8,628 (1.5%)	9,012 (1.5%)
RC transposons	1,741 (0.8%)	1,493 (0.8%)	5,092 (0.9%)	6,038 (1%)
SINEs	586 (0.3%)	459 (0.2%)	1736 (0.3%)	2021 (0.3%)

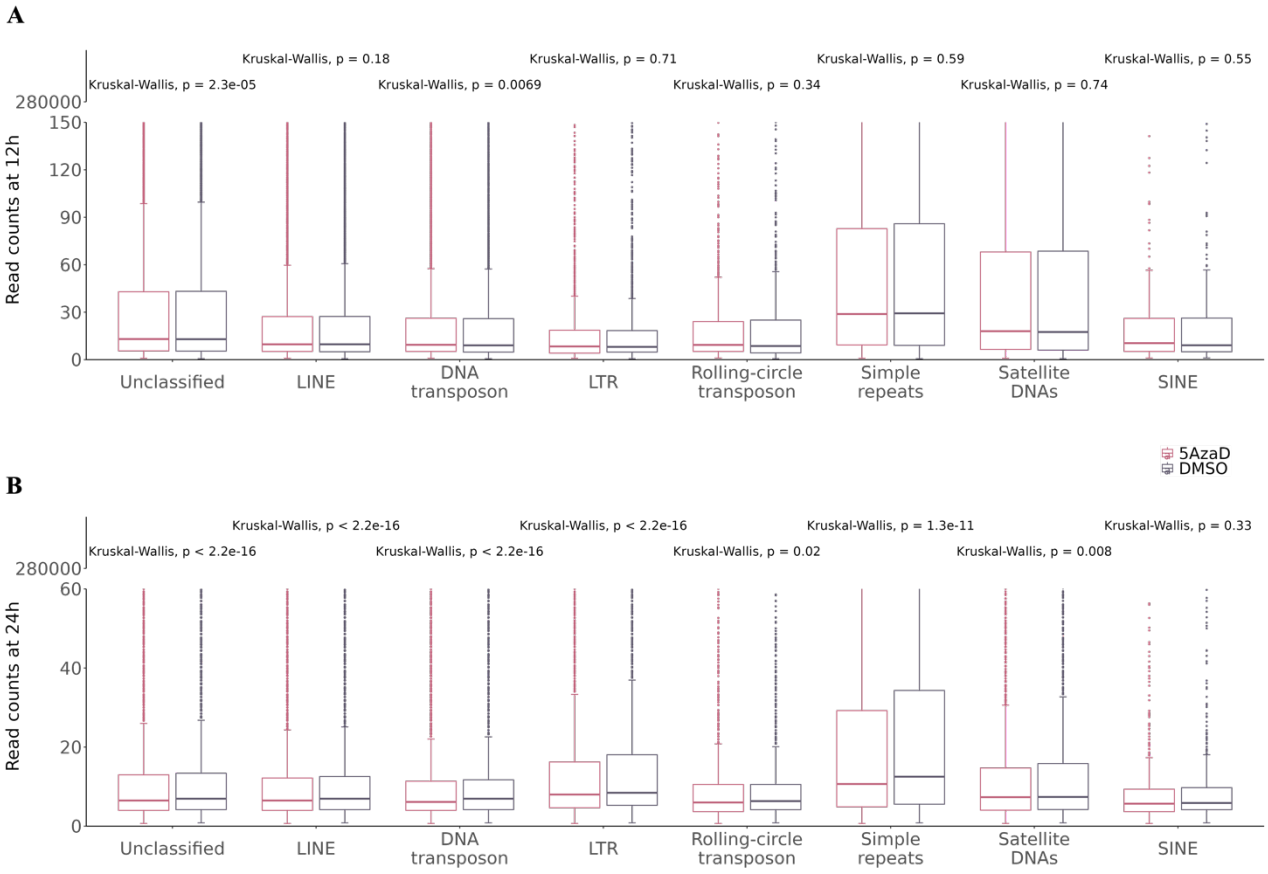


Figure 15. Box plots of all DESeq normalized read counts of different repeat classes in 5AzaD and DMSO treated embryos at **(A)** 12-hour and **(B)** 24-hour time point. Unclassified: unclassified repeats.

4.3.6.2 Differential expression analysis

Following the global characterisation of expressed repeats, I performed differential expression analysis on all expressed repeats in embryos treated with 5AzaD and DMSO. The pattern of differentially expressed repeats (DERs) at each time point was comparable to that of differentially expressed genes, with few repeats showing up or downregulation at 12-hour, but a substantial number of DERs detected at 24-hour (Figure 16, Table 7). Upon examining the number of DERs in each repeat class, I observed a relatively higher proportion of LINEs and DNA transposons, excluding unclassified repeats, showing both up and downregulation at 24-hour (Table 7). These results suggest that 5AzaD treatment prior to MZT influenced subsequent transcription of some classes of transposable elements, especially LINEs and DNA transposons.

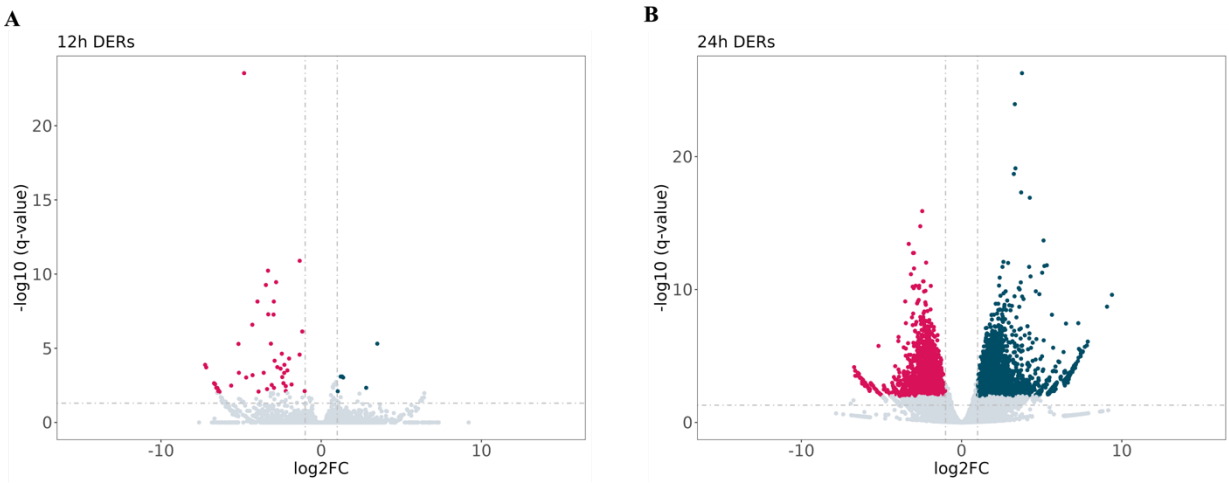


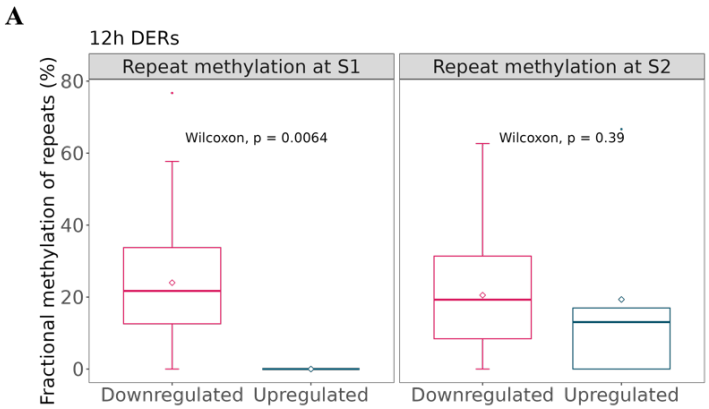
Figure 16. Differential expression analysis for repeats expressed at **(A)** 12-hour and **(B)** 24-hour.

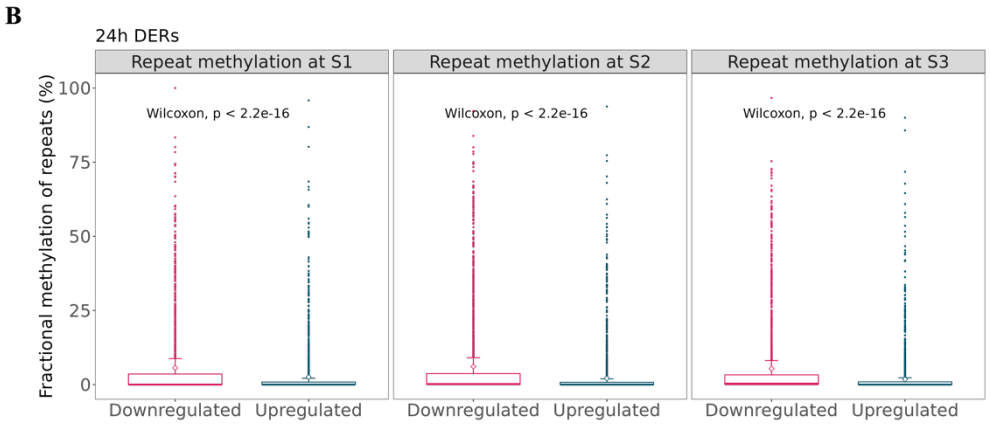
Table 7. Number of differentially expressed repeats in each repeat class at the two time points.

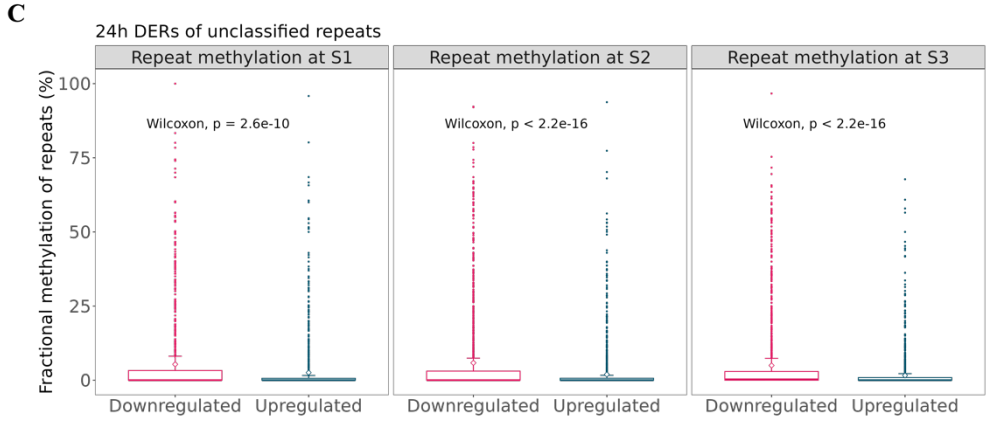
Time point	Class	Upregulated	%	Downregulated	%
12-hour	**All**	6	-	48	-
	Unclassified	6	100	31	64.58
	LINEs	0	0	7	14.58
	DNA transposons	0	0	5	10.42
	LTRs	0	0	2	4.17
	Simple repeats	0	0	3	6.25
24-hour	**All**	3114	-	2233	-
	Unclassified	2085	66.96	1442	64.58
	LINEs	331	10.63	319	22.12
	DNA transposons	261	8.38	219	9.81
	LTRs	61	2.00	87	3.90
	Rolling-circle transposons	8	0.26	23	1.60
	Simple repeats	276	8.86	90	4.03
	Satellite DNAs	85	2.73	42	1.88
	SINEs	7	0.22	11	0.49

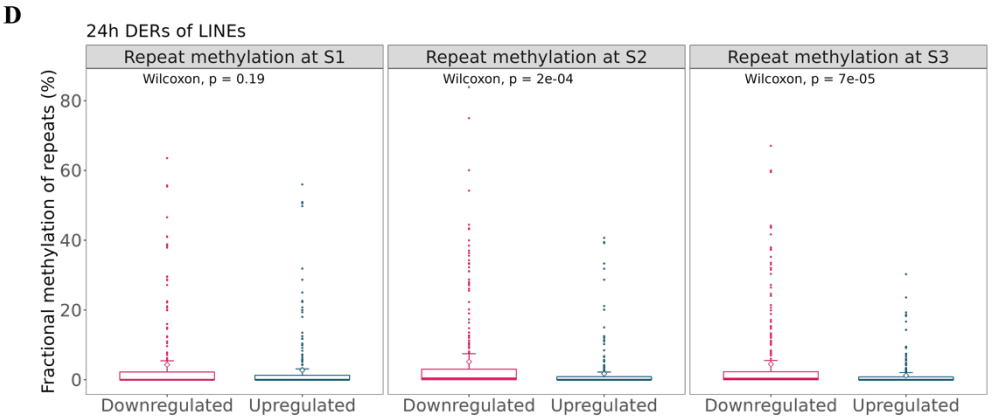
4.3.6.3 Methylation of differentially expressed repeats

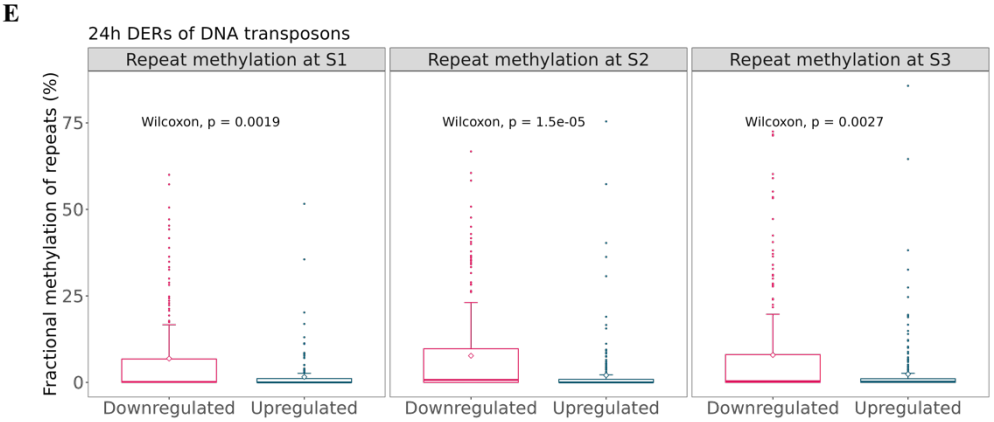
To gain further insights into the effect of 5AzaD treatment on repeat expression, I merged DERs with the methylation level of the corresponding and preceding developmental stage from the EM-seq datasets. The methylation levels of 24-hour downregulated repeats were consistently higher than those of upregulated repeats at the same time point (Figure 17A, B). However, the 12-hour downregulated repeats only showed higher methylation at S1 compared to the 12-hour upregulated repeats. When analysing the methylation between DERs for different repeat classes, I found that differentially expressed unclassified repeats and DNA transposons followed the same methylation pattern as the entire repeat dataset, with downregulated repeats of these two classes having higher methylation levels than upregulated repeats within their respective classes. In contrast, downregulated LINEs only showed significantly higher methylation than upregulated LINEs at S2 and S3 but exhibited similar methylation levels between upregulated and downregulated LINEs at S1. Lastly, no differences were found in the methylation levels between upregulated and downregulated LTRs at S1 and S2, but downregulated LTRs exhibited higher methylation at S3 (Figure 17C-F). Nevertheless, downregulated repeats generally have higher methylation levels compared to upregulated repeats.




B
24h DERs
Repeat methylation at S1
Repeat methylation at S2
Repeat methylation at S3
Fractional methylation of repeats (%)
Wilcoxon, p < 2.2e-16
Wilcoxon, p < 2.2e-16
Wilcoxon, p < 2.2e-16
Downregulated
Upregulated


C
24h DERs of unclassified repeats
Repeat methylation at S1
Repeat methylation at S2
Repeat methylation at S3
Fractional methylation of repeats (%)
Wilcoxon, p = 2.6e-10
Wilcoxon, p < 2.2e-16
Wilcoxon, p < 2.2e-16
Downregulated
Upregulated


D
24h DERs of LINEs
Repeat methylation at S1
Repeat methylation at S2
Repeat methylation at S3
Fractional methylation of repeats (%)
Wilcoxon, p = 0.19
Wilcoxon, p = 2e-04
Wilcoxon, p = 7e-05
Downregulated
Upregulated


E
24h DERs of DNA transposons
Repeat methylation at S1
Repeat methylation at S2
Repeat methylation at S3
Fractional methylation of repeats (%)
Wilcoxon, p = 0.0019
Wilcoxon, p = 1.5e-05
Wilcoxon, p = 0.0027
Downregulated
Upregulated

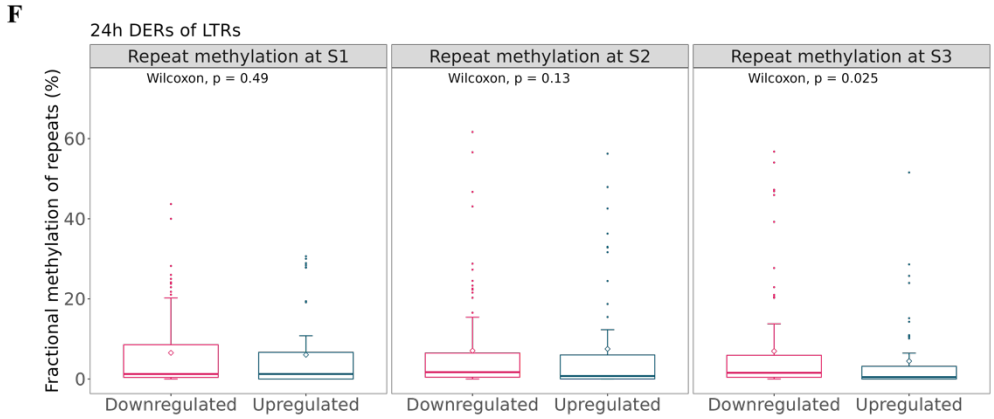


Figure 17. The fractional methylation of differentially expressed repeats in various repeat classes at the corresponding and preceding developmental stages. Methylation of **(A)** 12-hour and **(B)** 24-hour DERs. Methylation of 24-hour DERs of **(C)** unclassified repeats, **(D)** LINEs, **(E)** DNA transposons and **(F)** LTRs.

4.3.7 Relationship between gene expression and repeat expression

DNA methylation is widely recognised as a mechanism for silencing transposon expression (Bestor & Bourc'his, 2004; Chernyavskaya et al., 2017; W. man Liu et al., 1994; Walsh et al., 1998; Woodcock et al., 1997). The expectation for the 5AzaD treatment of *Parhyale* embryo is to inhibit DNA methylation. However, previous corelations showing higher methylation levels in downregulated repeats compared to upregulated repeats after the drug treatment challenge this notion. Additionally, downregulated genes exhibited higher methylation levels than upregulated genes. Hence, it is worth to investigate the correlation between gene expression, repeat expression, and the methylation regulation of their expression. Due to the limited number of differentially expressed genes and repeats at the 12-hour time point, the subsequent analysis focuses on the 24-hour time point dataset.

Firstly, I compared the enrichment of repeats in DEGs versus non-differentially expressed genes (non-DEGs), as well as in upregulated versus downregulated genes after drug treatment. The analysis revealed a significant enrichment of repeats in the gene bodies

of DEGs compared to non-DEGs (Table 8). Additionally, repeats were more enriched in downregulated genes compared to upregulated genes (Table 9). These findings suggest that genes affected by the 5AzaD treatment are more likely to contain repeats in their gene bodies. Furthermore, the higher methylation levels observed in downregulated genes, coupled with the generally high methylation of repeats, indicate that the increased number of repeats in these downregulated genes likely contributes to their elevated methylation.

Table 8. Comparison of repeats enrichment between DEGs and non-DEGs.

	DEGs	%	non-DEGs	%	Fisher test
With repeats	3,567	96.77	14,188	93.27	p < 2.20e-16
Without repeats	119	3.23	1,023	6.73	

Table 9. Comparison of repeats enrichment between upregulated and downregulated genes.

	Upregulated genes	%	Downregulated genes	%	Fisher test
With repeats	1,624	95.64	1,943	97.74	p = 3.60e-04
Without repeats	74	4.36	45	2.26	

I then proceeded to explore the association between differentially expressed repeats and their host genes. Firstly, I investigated whether repeats expressed during the drug treatment were preferentially located within gene bodies. Table 10 highlights a consistent trend across treatment groups and time points, revealing a significantly higher presence of expressed repeats within gene bodies compared to non-expressed repeats. Additionally, the distribution analysis (Table 11) shows that DERs are notably more concentrated inside gene bodies compared to non-DERs. However, when specifically

examining upregulated and downregulated DERs (Table 12), no significant differences were found in their localization within gene bodies.

Table 10. Comparison of expressed and non-expressed repeats in gene bodies.

Class	5azaD-12h	Dmso-12h	5azaD-24h	dmso-24h
Expressed repeats	192,165	221,359	586,640	570,312
Number of intragenic expressed repeats	110,959 (58%)	127,359 (58%)	328,349 (56%)	353,936 (62%)
Non-expressed repeats	5,560,800	5,531,606	5,166,325	5,182,653
Number of intragenic non-expressed repeats	1,731,992 (31%)	1,715,592 (31%)	1,514,602 (29%)	1,489,015 (29%)
Fisher test	p < 2.20e-16	p < 2.20e-16	p < 2.20e-16	p < 2.20e-16

Table 11. Distribution of DERs and non-DERs in intragenic and intergenic regions.

	DERs	%	non-DERs	%	Fisher test
Intragenic	3,775	70.60	211,540	65.16	p < 2.20e-16
Intergenic	1,572	29.40	113,120	34.84	

Table 12. Distribution of upregulated and downregulated repeats in intragenic and intergenic regions.

	Upregulated repeats	%	Downregulated repeats	%	Fisher test
Intragenic	2214	71.10	1561	69.91	p = 0.3457
Intergenic	900	28.90	672	30.09	

Moreover, I also investigated the relationship between DERs and DEGs. The results, presented in Table 13, indicate a significantly higher proportion of DERs within the gene bodies of DEGs compared to non-DERs. Additionally, downregulated repeats were significantly enriched within the gene bodies of DEGs compared to upregulated repeats (Table 14). Furthermore, Table 15 shows that over 95% of DERs within the gene bodies of DEGs exhibited the same direction of expression change as their host genes. The analysis of repeat expression indicates that repeats within gene bodies tend to be expressed more compared to the repeats within intergenic regions during *Parhyale* early development. Additionally, these repeats which reside in gene bodies are also prone to changing their expression status after 5AzaD treatment.

Table 13. Distribution of DERs and non-DERs within DEGs and non-DEG regions.

	DERs	%	non-DERs	%	Fisher test
Inside DEGs	2,077	38.84	53,750	16.56	p < 2.20e-16
Outside DEGs	3,270	61.16	270,910	83.44	

Table 14. Distribution of upregulated and downregulated repeats within DEGs and non-DEG regions.

	Upregulated repeats	%	Downregulated repeats	%	Fisher test
Inside DEGs	904	29.03	1173	52.53	p < 2.20e-16
Outside DEGs	2210	70.97	1060	47.47	

Table 15. Correlation between the direction of expression change between DERs and DEGs.

	Upregulated genes	%	Downregulated genes	%
Downregulated Repeats	23	2.4	1224	95.6
Upregulated Repeats	932	97.6	56	4.4

After observing that DERs and their host DEGs exhibited the same direction of change, I examined the correlation between the fractional methylation of those DERs and their host DEGs that shared this common direction at three EM-seq stages. The overall correlation was weak, with DEGs generally displayed higher methylation levels compared to the DERs within their gene bodies. Notably, the methylation of downregulated genes and repeats exhibited a stronger correlation compared to the upregulated ones (Figure 18).

In summary, 5AzaD treatment influences both the expression of genes and the repeats within them. Repeats that change their expression status after drug treatment are more often reside within gene bodies. Furthermore, these differentially expressed repeats tend to follow the same expression change pattern as their host genes. These correlation analyses indicate a potential association between the alterations in gene expression and the changes in repeat expression following 5AzaD treatment.

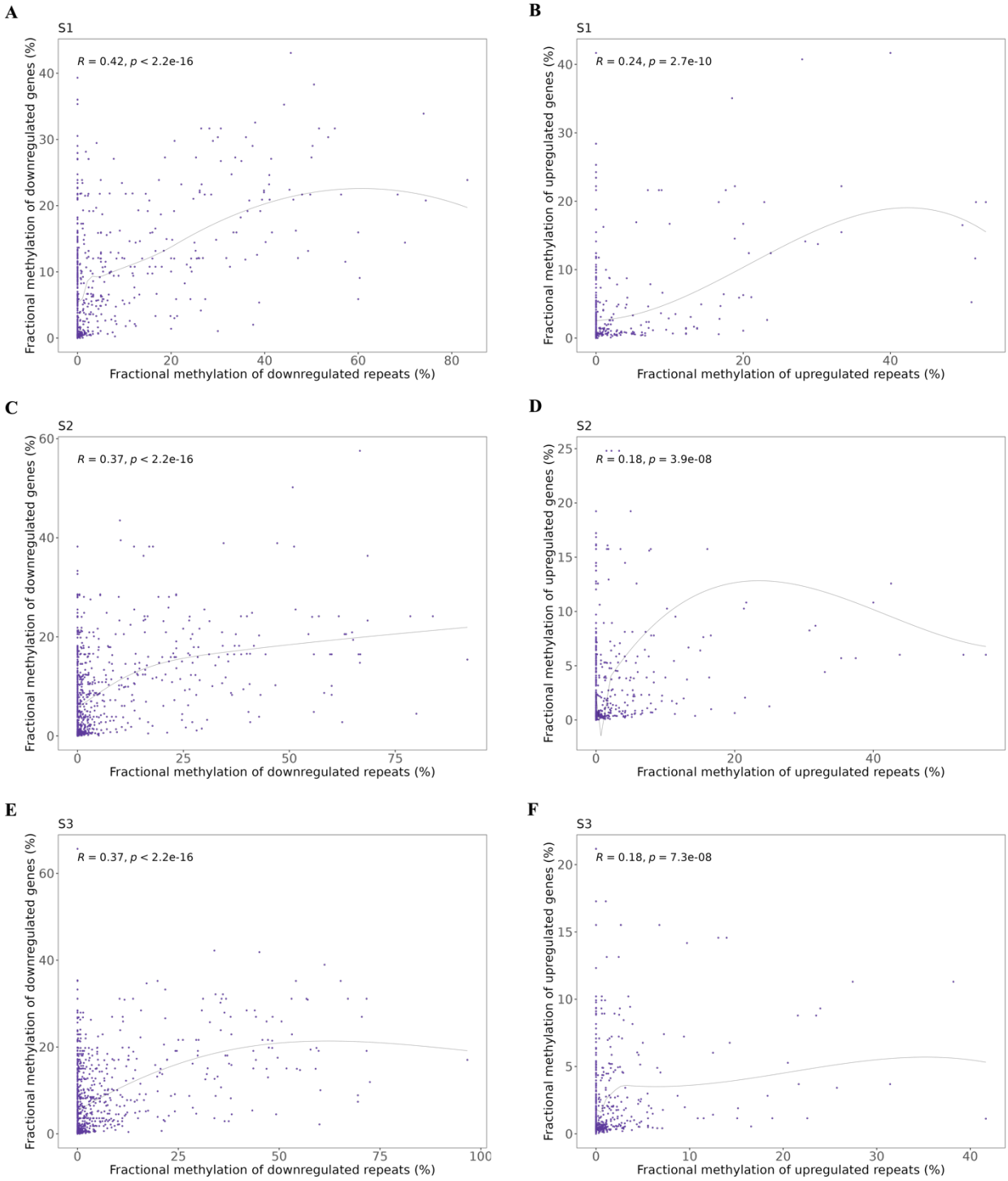


Figure 18. The relationship between fractional methylation in DERs and their corresponding host DEGs at three EM-seq stages: Correlation between **(A)** downregulated and **(B)** upregulated repeats and their host DEGs at S1. Correlation between **(C)** downregulated and **(D)** upregulated repeats and their host DEGs at S2. Correlation between **(E)** downregulated and **(F)** upregulated repeats and their host DEGs at S3.

4.3.8 Validation of demethylation effect induced by 5AzaD

To assess the methylation status of the 5AzaD-treated embryos, I conducted targeted gene bisulfite sequencing on selected gene regions. Bisulfite treatment was carried out on genomic DNA extracted from the 24-hour 5AzaD and DMSO treated embryos. Based on previous results, gene body methylation mainly occurred on introns. Therefore, I decided to select several introns as targets for methylation status analysis. First, I separated the introns using the sliding window method with a window size of 900 bp and steps of 50 bp. Next, I assessed the CpG coverage and fractional methylation for each intron sub-region at S2 and S3 of Methyl-seq data and revealed that the majority of them had fewer than 20 covered CpG sites (Figure 19). I selected intron sub-regions containing at least 20 CpGs and a fraction methylation of more than 80% in both S2 and S3. I randomly chose five regions for PCR amplification; these five regions were amplified with primers designed by MethPrimer (Li & Dahiya, 2002); three of them were successfully amplified and sent for sanger sequencing.

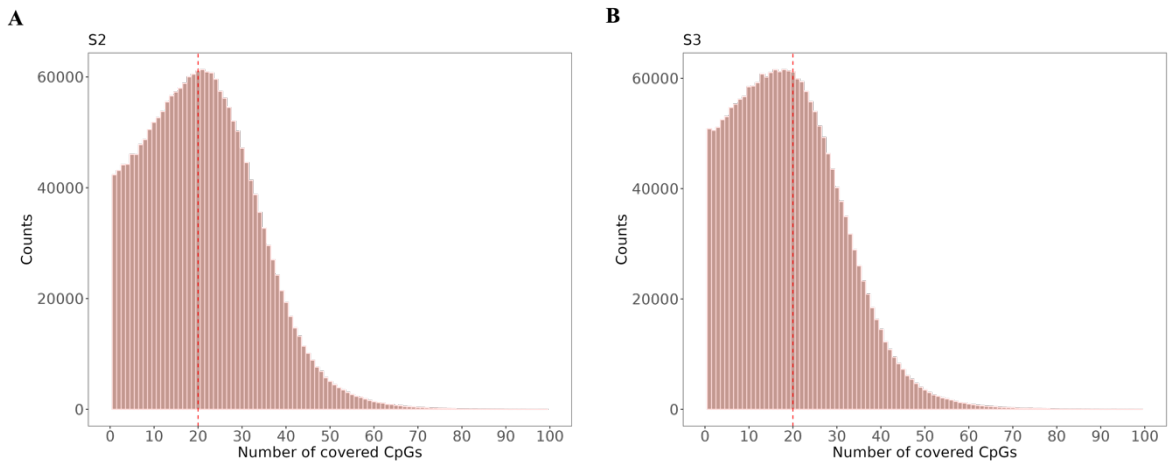


Figure 19. The number of covered CpG sites at **(A)** S2 and **(B)** S3 in each sub-region of the intron. The vertical dashed line represents the minimum CpG coverage threshold for intron sub-region selection (Covered CpGs per region ≥ 20).

The methylation status of the selected intron sub-regions was found to be consistent between the 5AzaD and DMSO treated embryos (Figure 20). By blasting the reads to the reference sequences, it was determined that all cytosines in the CpG context of the sequenced region remained cytosines after bisulfite conversion, indicating that these CpGs were highly methylated in the embryos of both treatment groups and were not converted after bisulfite conversion. One possible explanation is that methylation was already restored during the following 12-hour period after exposure ended. If this is the case, it provides additional support for the notion that DNA demethylation prior to MZT has lasting impact on embryogenesis, even if the methylation status recovers in later stages.

In addition to this validation experiment, no further experiments were conducted to directly analyse the effects of 5AzaD on the methylomes of *Parhyale* embryos. Therefore, a limitation of the research presented in this chapter is the lack of determination of how DNA methylation is altered on a genome-wide scale in response to 5AzaD treatment of *Parhyale*.

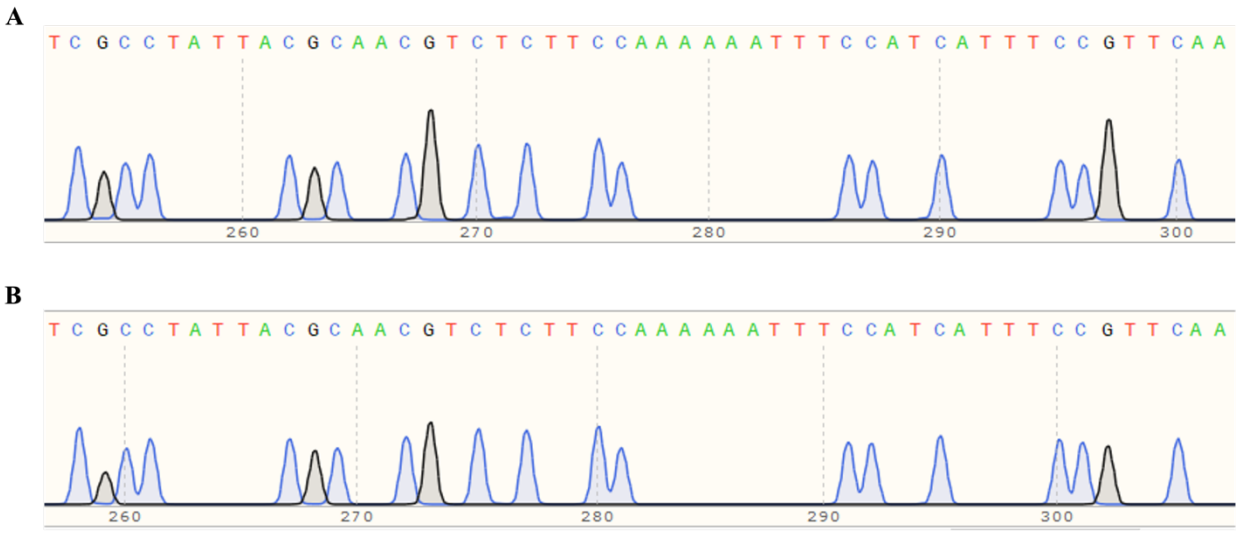


Figure 20. A portion of the sequencing chromatogram for one of the chosen intron sub-regions. **(A)** Sequence of 5AzaD-treated embryos **(B)** Sequence of DMSO-treated embryos

4.4 Conclusion

This chapter investigated the functional importance of DNA methylation in the early embryonic development of *Parhyale hawaiensis*. To achieve this, embryos were treated during the first 12-hour of development with 5AzaD, a hypomethylating agent that inhibits DNMT1 activity and results in cell-division-dependent DNA demethylation (Christman, 2002). The findings indicated that 5AzaD treatment led to a significant delay in embryonic development and a high mortality rate among the embryos. However, it is important to note that the observed phenotypes may be partly attributed to the toxicity of 5AzaD, which may be unrelated to methylation.

Analysis of differentially expressed genes and their biological functions in 12-hour and 24-hour *Parhyale* embryos following different treatment revealed that DNA methylation deposition is crucial for maintaining accurate temporal expression dynamics. This was demonstrated by the delayed development and altered transcription of genes involved in cellular and developmental processes in response to exposure to 5AzaD before or during MZT.

Further investigation revealed that 5AzaD treatment may has a significant effect on zygotic transcriptional activation. Specifically, genes that were downregulated following exposure to 5AzaD were found to have higher levels of DNA methylation than upregulated or non-differentially expressed genes. These downregulated genes were enriched in zygotic-only and zygotic-maternal genes, indicating that DNA methylation likely plays a regulatory role in the activation of a subset of ZGA genes. These findings suggest that precise DNA methylation deposition is required for the proper transition of developmental control from maternally supplied gene products to those synthesized from the zygotic genome.

Finally, transcription analysis of repeats was performed to investigate the impact of 5AzaD treatment on repeat expression at two time points. The results revealed that 5AzaD treatment also influenced repeat expression. Specifically, intragenic repeats tend to be expressed more than intergenic repeats during *Parhyale* early development. Furthermore, intragenic repeats were also prone to changing their expression status and follow the same expression change pattern as their host genes after 5AzaD treatment. Therefore, it is anticipated that there is a potential association between the alterations in gene expression and the changes in repeat expression following 5AzaD treatment.

The study aims to shed light on the potential role of DNA methylation in the early embryonic development of *Parhyale*. However, it's essential to stress that these conclusions remain tentative. No definitive causality can be inferred from the above findings due to the unsuccessful validation experiment and the lack of evidence for DNA methylation loss after 5AzaD treatment. Further validation using alternative methods to detect methylation reduction is essential. This will help determine whether the observed effects result from DNA demethylation due 5AzaD treatment or from toxicity caused by the drug or its solvent, DMSO. Such investigations are crucial to deepen our understanding of the precise impact of DNA methylation on the early embryonic processes in *Parhyale*.

Chapter 5

Thesis discussion and future directions

5.1 Discussion

DNA methylation is a modification that is largely conserved across the animal kingdom. Although it has been suggested that invertebrates possess a system of DNA methylation distinct from that of vertebrates, the potential function of DNA methylation in invertebrate species remains largely unknown (Feng et al., 2010; Keller et al., 2016; Wang et al., 2006; Zemach et al., 2010). Here, I present a comprehensive analysis of DNA methylation in *Parhyale hawaiensis*, a marine amphipod crustacean that serves as an excellent model organism due to its accessibility to developmental stages, established genetics, transgenic tools, genomic resources, and imaging techniques (Paris et al., 2022). The aim of this research is to understand the relationship between DNA methylation and the transcription of genes and transposable elements, as well as the potential function of DNA methylation in the early embryonic development of *Parhyale*. To accomplish this, the study utilised methyl-seq data from WGBS, PacBio SMRT sequencing, and EM-seq to compare methylation patterns across various tissues and embryonic stages. Additionally, RNA-seq data was integrated with EM-seq data to investigate the correlations between DNA methylation and gene expression.

5.1.1 Genome-wide methylation patterns in *Parhyale*

Through analysing the DNA methylation data obtained from WGBS and PacBio SMRT sequencing samples, I was able to confirm the presence of DNA methylation in the *Parhyale* genome. Further analysis using EM-seq data from embryos at three early developmental stages revealed several notable features of the *Parhyale* methylomes. Firstly, approximately 8% of all CpGs in the *Parhyale* genome were methylated across different tissues and embryonic stages. Methylation accumulation was observed in both gene bodies and non-coding regions, with a higher number of methylated CpGs and elevated methylation levels in introns compared to exons within the gene bodies. In

addition, repetitive elements were found to be the primary targets of DNA methylation in the *Parhyale* genome, particularly those located in introns and intergenic regions.

In the WGBS data presented in this thesis, the uniquely mapping rate was relatively low, and a large portion of the sequences exhibit non-uniquely mapping to the reference genome. These multiply mapping reads cannot be used in downstream analysis, as they compromise the accuracy of methylation state interpretations (Liu & Xu, 2021). Despite ongoing efforts to enhance or develop new bisulfite read mappers, short-read methylation sequencing consistently encounters low mapping rates, typically ranging from 40% to 60%, significantly lower than the mapping rates observed for regular short reads, such as RNA-seq data (Krueger & Andrews, 2011; Porter et al., 2015; Tran et al., 2014).

The relative low mapping efficiency observed in the methylation data can be attributed to several potential factors. Firstly, the expansive size of *Parhyale* genome, which totals 3.6 Gb and includes an abundance of repetitive elements constituting over 60% of the genome, poses challenges for precise read alignment. The presence of these repetitive elements complicates the alignment process, introducing uncertainties and potentially contributing to increased rates of non-uniquely mappings, particularly evident in short-read sequencing (Treangen & Salzberg, 2012). Furthermore, the complex nature of the genome may hinder the effectiveness of alignment algorithms, resulting in reduced overall mapping rates.

It's noteworthy that the methylation data in this study were mapped to both the previous *Parhyale* reference assembly (Phaw 5.0, GCA_001587735.2) and the updated assembly, Phaw 5.1. The mapping results presented here were aligned with Phaw 5.1 (2.82 Gb), which exhibited a slight increase in size compared to Phaw 5.0 (2.76 Gb) and a marginal improvement in N50 from 53.69 Mb to 54.78 Mb (+2.03%). A comparative analysis of the mapping results between the two assemblies reveals a 2% increase in the average counts of CpGs covered by at least five reads across all samples. The updated *Parhyale* genome

assembly has improved by filling missing regions and closing some gaps from Phaw 5.0, but the genome remains imperfect, and further efforts are needed to generate a chromosome-level assembly by integrating additional long reads and genome-wide chromatin interaction data. Moreover, inherent imperfections in genome assemblies exacerbate the complexity of the mapping process. These inaccuracies can impede mapping, particularly in regions where the assembly is incomplete or contains discrepancies. This issue is compounded by the reduced complexity of bisulfite-converted DNA, where unmethylated cytosines are converted to uracil and subsequently to thymine during sequencing, which further complicates the alignment of sequences and result in low mapping rate (Liu & Xu, 2021; Silliman et al., 2023).

Similar to the results indicated in a genome-wide methylation study across arthropods, *Parhyale* exhibits moderate methylation in its genome. This level of methylation is significantly higher than that observed in most insect species but lower than that found in the centipede *Strigamia maritima* (Bewick et al., 2017; Lewis et al., 2020). It is worth noting that the extent of genome-wide DNA methylation varies widely across arthropods, with insects displaying extremely low levels of methylation compared to other arthropod clades that typically exhibit moderate levels of methylation (Bewick et al., 2017; Feng et al., 2010; Lewis et al., 2020; Zemach et al., 2010).

Comparative genomic data analysed by Lewis et al, (2020) reveal that the most extensively conserved gene methylation characteristic of arthropod methylomes is the enrichment of methylation at exons relative to introns. However, this conserved feature was not observed for *Parhyale hawaiensis* and *Armadillidium vulgare*. Instead, these two species show a marked preference for targeting methylation within introns rather than exons. Specifically, *Parhyale* exhibits a distinct pattern where exons are less methylated than introns and the genome-wide background, a finding corroborated by the current study. *Armadillidium vulgare* exons display only slightly higher methylation levels than the background (Lewis et al., 2020). The reduced exon methylation in the two species,

contrasting with the conserved feature of exon-enriched methylation in other arthropods, indicates that this loss may be an evolutionary adaptation. These changes could have made gene body methylation redundant or disadvantageous under the specific environmental or biological conditions faced by the ancestors of these species (Dimond et al., 2017; Gatzmann et al., 2018; Gavery & Roberts, 2010; Lewis et al., 2020; Putnam et al., 2016; Suarez-Ulloa et al., 2019; Trigg et al., 2022).

Our findings indicate that in *Parhyale*, the methylation levels in the 5’ UTR of the genes are higher compared to the global background, whereas methylation levels in the 3' end of genes were relatively lower. This distribution pattern of gene body methylation in *Parhyale* was similar to that observed in other invertebrate species, where methylation levels rapidly elevated from the first exon and decrease towards the 3' end of the genes (Bonasio et al., 2012; Patalano et al., 2015; Wang et al., 2013).

In the *Parhyale*, repetitive elements occupy over 60% of the genome, a content comparable to many mammalian genomes and significantly higher than many invertebrates, for example *C. elegans* and *Drosophila* (Almeida et al., 2007; Biémont & Vieira, 2006; Lander et al., 2001; Levy et al., 2008; Mouse Genome Sequencing Consortium, 2002; Zhou et al., 2020). Despite the general trend of low or absent TE methylation in many arthropods and especially insects (Feng et al., 2010; Lewis et al., 2020; Zemach et al., 2010; Zemach & Zilberman, 2010), *Parhyale* shows a moderate level of methylation in repetitive elements.

In animals, plants, and fungi, DNA methylation of transposable elements is associated with the suppression of their transposition activities. This suppression is crucial for protecting genomes from the potentially harmful effects of transposon activity (Barau et al., 2016; Regev et al., 1998; Ying et al., 2022; Yoder et al., 1997; Zemach & Zilberman, 2010). Consequently, methylation in repetitive elements likely regulates transposable

element activity and provides tolerance for TE invasions in *Parhyale* (Regev et al., 1998; Walsh et al., 1998; Zemach et al., 2010).

Our study found that the extent of repeat methylation in *Parhyale* was strongly influenced by its genomic location. Similar to many plants and animals, transposable elements in *Parhyale* primarily integrate into the non-coding regions of the genome. Intragenic regions and introns are more enriched with repetitive sequences than exons (Almeida et al., 2007; Le et al., 2015; Sela et al., 2010; Wang et al., 2012; Zhou et al., 2020). Furthermore, these repetitive sequences within introns exhibit higher methylation levels compared to exons, leading to higher intron methylation in *Parhyale*.

In vertebrates, despite the high TE content in intron regions, introns usually exhibit lower methylation compared to exons (Jiang et al., 2014; Singer et al., 2015; Zhang et al., 2024). Studies have found that TE inserted into intron may introduce novel splice sites, leading to alternative splicing events that can compromise transcriptional integrity (Cordaux & Batzer, 2009; Lev-Maor et al., 2008; Treiber & Waddell, 2020).

DNA methylation has been found to preferentially target young transposons, which are presumed to be the most active transposon group in the genome (Deniz et al., 2019; Ying et al., 2022; Zhou et al., 2020). However, studies indicate that mammalian genomes tend to have more ancient, divergent, and fractionated TE copies, while the number of recent TE copies in the human genome is lower compared to many other vertebrates and invertebrates (Chalopin et al., 2015; Smit, 1999). In the human genome, old L1 transposons gradually lose methylation over the lifespan, whereas the loss of methylation in young L1 transposons is associated with cancer (Morandini et al., 2024). Additionally, non-insect arthropods show a high coverage of varying TE types with low divergence (Petersen et al., 2021).

Therefore, the higher methylation of intronic repetitive sequences in *Parhyale* compared to mammals may be due to introns being enriched with younger TEs, necessitating

repression of these transposons by DNA methylation. Future studies should estimate the divergence (Kimura distance) of individual transposon sequences compared with the inferred consensus for each transposon family in the *Parhyale* genome, particularly those residing in introns, to examine the age of those transposons and gain further insights into this phenomenon.

5.1.2 The relationship between DNA methylation and gene expression in *Parhyale*

Gene body methylation in *Parhyale* embryos exhibits a bimodal distribution similar to other invertebrates (Sarda et al., 2012). RNA-seq analysis from *Parhyale* embryos at the same developmental stages used for methylation analysis shows that genes with moderate to high expression levels are more likely to be methylated compared to genes with low expression. This finding aligns with previous studies in invertebrates and plants, where methylation accumulates on genes with intermediate to high expression levels (Feng et al., 2010; Lewis et al., 2020; Zemach et al., 2010). Additionally, our study identified a positive correlation between gene expression and DNA methylation, with higher transcriptional activity generally associated with higher methylation levels. However, this pattern contrasts with many other invertebrates, where methylation typically shows a positive correlation with expression levels except for highly expressed genes (Bonasio et al., 2012; Flores et al., 2012; Kvist et al., 2018; X. Wang et al., 2014; Zemach et al., 2010; Zilberman et al., 2006). This variance might indicate an additional regulatory layer, potentially involving alternative splicing (Bonasio et al., 2012; Kim et al., 2007).

Extensive research has shown that DNA methylation regulates alternative splicing by interacting with RNA polymerase II elongation and modifying its rate, thereby influencing splice site recognition (Flores et al., 2012; Lev Maor et al., 2015; Wang et al., 2024). Additionally, it recruits splicing factors through mediator proteins (Lev Maor et al., 2015;

Wang et al., 2024). DNA methylation also targets genes encoding splicing factors, altering their expression and causing changes in splicing patterns (Ashok et al., 2021; Maunakea et al., 2013). In species that do not show an absolute correlation between methylation levels and gene expression, there is typically higher methylation in exons than in introns (Lewis et al., 2020; X. Wang et al., 2014; Zemach et al., 2010). This suggests that intragenic methylation in this species may correlate with splicing events, influencing the inclusion or exclusion of certain exons.

In contrast, our observations in *Parhyale*, which display higher intron methylation, and has a consistent positive correlation between methylation levels and gene expression. This unusual pattern compared to other invertebrates may suggest a unique role of gene body methylation in *Parhyale*, beyond alternative splicing. Future research could investigate whether increased gene body methylation or specifically intron methylation affects alternative splicing or has other biological roles in *Parhyale*. Such research could yield deeper insights into the regulatory complexities of gene expression in *Parhyale* and potentially other species.

5.1.3 The role of DNA methylation in the early development of *Parhyale*

Previous studies have extensively investigated the crucial role of DNA methylation in regulating gene expression and its significance in embryonic development in mammals. However, there is a significant paucity of research on the mechanisms and biological functions of this epigenetic modification in invertebrates, mainly due to the absence of DNA methylation in conventional invertebrate models such as *Drosophila melanogaster* and *Caenorhabditis elegans*. In the final section of this thesis, I addressed this gap by exploring the functional significance of DNA methylation on gene transcription during early development in *Parhyale*. Specifically, I treated pre-MZT *Parhyale* embryos with Decitabine (5-aza-2'-deoxycytidine, 5AzaD), a hypomethylating agent that induces DNA demethylation by inhibiting the activity of DNMT1 during cell division (Christman, 2002).

Although 5AzaD has been commonly used in vertebrates, especially in human, its use has been limited in invertebrates, including *Parhyale*. To the best of our knowledge, our study is the first to utilise this agent in *Parhyale* (Kumar & Kim, 2017; Pozo et al., 2021; Xu et al., 2018).

I observed several effects of 5AzaD treatment on *Parhyale* embryos. Firstly, embryos treated with 5AzaD during the initial 24-hour period of development exhibited a significantly higher mortality rate at 72 hours compared to the vehicle control (DMSO-treated) and wild-type embryos. Similar deleterious effects were observed in annelids treated with 5AzaD during regeneration, which had long-lasting effects and influenced subsequent post-regenerative posterior growth (Planques et al., 2021). Therefore, it is possible that the 5AzaD treatment may not affect developmental processes that occur soon after treatment, but instead may strongly affect processes that begin later in development.

Comparing the gene expression profiles of 5AzaD-treated and wild-type embryos revealed that 5AzaD-treated embryos are significantly delayed in development, as evidenced by the transcriptome of 5AzaD-treated embryos being more similar to the transcriptome of the preceding developmental stage than the transcriptome of the corresponding stage of wild-type embryos. This developmental delay is consistent with findings regarding the development of oysters, where demethylation led to aberrant gene expression and delayed cell differentiation, resulting in developmental defects (Riviere et al., 2013).

In addition, our results indicated that 5AzaD treatment before MZT caused a disturbance in gene expression during the early stages of embryogenesis, with the most pronounced changes occurring at 24-hour of development. The failure to establish the appropriate transcriptome is likely responsible for the death of 5AzaD-treated embryos. This finding is similar to a study on wasps, which observed a significant decrease in global gene body

methylation and a large number of differentially expressed genes during the major wave of zygotic genome activation following knockdown of dnmt1 (Arsala et al., 2022).

Importantly, we found that there was a strong correlation between the methylation state of the genes and the direction of expression change after 5AzaD treatment. Downregulated genes in 5AzaD-treated *Parhyale* embryos had higher levels of gene body methylation and were more likely to be methylated genes, while upregulated genes were more commonly unmethylated. The high proportion of methylated genes in the downregulated gene set is consistent with previous findings in some invertebrates, such as oysters, cockroaches, and wasps, which suggest a positive association between DNA methylation and stable gene expression (Arsala et al., 2022; Riviere et al., 2017; Vento´s Alfonso et al., 2020; Wang et al., 2013). For example, in wasps, the degree of expression reduction is related to the level of demethylation induced by knockdown of dnmt1 (Arsala et al., 2022). Similarly, in oysters, low gene body methylation was associated with transcription repression, whereas methylated DNA was positively correlated with gene expression levels (Riviere et al., 2017). In contrast, in some plants and insects, gene body methylation has a less pronounced impact on transcription, and the loss of gene body methylation does not significantly affect expression levels. These divergent findings may indicate that the function of gene body methylation is variable and plastic over evolutionary time, and its precise function may vary from species to species (Arsala et al., 2022; Bewick, Sanchez, et al., 2019; Bewick, Zhang, et al., 2019).

By analysing the functional role of DEGs, I revealed that downregulated genes that showed hypermethylation were primarily involved in housekeeping functions such as gene expression, translation, and RNA processing, based on Gene Ontology analysis. Previous research on invertebrates demonstrated that genes involved in housekeeping functions that exhibit consistent expression across developmental time points and tissues tend to have high levels of DNA methylation (Elango et al., 2009; Lewis et al., 2020; Sarda et al., 2012). Nevertheless, the results of this thesis support the functional significance of

DNA methylation and suggest that gene body methylation plays a crucial role in maintaining high expression levels of methylated genes during *Parhyale* early development. The potential effect of reducing gene body methylation after 5AzaD treatment may lead to a significant decrease in transcription of these genes. Moreover, the observed increase in expression levels of unmethylated genes following 5AzaD treatment may be a consequence of the initial reduction of those methylated differentially expressed genes.

5.1.4 DNA methylation of zygotically activated genes

This study sampled two time points, 12-hour and 24-hour, to investigate the functional role of DNA methylation. Although a substantial number of genes were differentially expressed at 24 hours, very few genes were differentially expressed at 12 hours, likely due to the developmental delay caused by 5AzaD treatment (i.e., ZGA had not progressed). However, the lack of transcriptome data for the treated embryos at stages between the minor and major waves of ZGA makes it difficult to determine the exact stages at which the effects of the treatment were initiated. In fact, I discovered that zygotically activated genes, which were enriched in the set of downregulated genes, have a significantly higher methylation level than non-zygotically activated genes.

During the maternal-to-zygotic transition, zygotic gene expression supplements the maternal contribution by providing the embryo with crucial factors that prepare it for gastrulation and cellular differentiation. These genes generally encode transcription factors and components of signalling pathways and perform basic cellular functions that are essential for all cell types and tissues (Aanes et al., 2011; Hamatani et al., 2004; M. T. Lee et al., 2014; Paranjpe et al., 2013; Xue et al., 2013). According to these findings, it is suggested that DNA methylation has a direct regulatory effect on the expression of zygotic genes and is crucial for the transfer of developmental control from maternally supplied gene products to those synthesized from the zygotic genome.

5.1.5 DNA methylation inhibitor

Despite demonstrating that 5AzaD is a preferable choice over other non-cytidine analogue GSK drugs, it cannot be denied that 5AzaD may introduce cytotoxicity to the system. Additionally, there has been debate over the solvents used to dissolve 5AzaD. 5AzaD exhibits solubility in organic solvents like DMSO and dimethyl formamide, with some reports suggesting water solubility of up to 25 mg/mL (Mahalingam et al., 2007; Neupane et al., 2014). However, its stability in aqueous solutions is a concern; prolonged exposure to water and non-neutral pH can lead to potency loss and impurity formation due to product hydrolysis (Balakishore et al., 2021; Kurtin & Demakos, 2010; Lim et al., 2011). Consequently, researchers now commonly dissolve 5AzaD in DMSO and use the corresponding DMSO solution as a vehicle control (Cook et al., 2019; Guynes et al., 2023; Kumar & Kim, 2017; Planques et al., 2021). However, these studies have not investigated the potential adverse effects of DMSO on the transcriptome.

In this thesis, in addition to the DMSO-containing vehicle control, wild-type controls were also included. DMSO-treated embryos exhibited slightly delayed development and alterations in gene expression compared to wild-type embryos, suggesting that the toxic effects of DMSO cannot be disregarded. Future studies should explore alternative solvents for dissolving 5AzaD. Additionally, alternative methods for demethylating the *Parhyale* genome are discussed in the further directions section.

5.1.6 The relationship between gene expression and repeat expression

The analysis of gene and repeat expression conducted in this study revealed a notable relationship between gene expression and repeat expression, particularly under the influence of 5AzaD treatment. Our findings indicated that DNA methylation, traditionally recognised for silencing transposon expression, affects both gene and repeat expression. Specifically, downregulated genes exhibited higher methylation levels than upregulated genes, suggesting a differential impact of DNA methylation on gene regulation based on

their expression status. This observation was further supported by enrichment analysis, which showed a significant presence of repeats in downregulated genes compared to upregulated ones.

The impact of transposable element insertions on host genes can be either beneficial or harmful. Regardless of their ability to transpose, TE sequences can be co-opted to regulate host genes (Gebrie, 2023). Mechanistically, TEs can impact gene expression at various stages, including transcriptionally, post-transcriptionally, or translation through their encoded products (Chuong et al., 2017; Drongitis et al., 2019; Fort et al., 2021; Fueyo et al., 2022; Jangam et al., 2017; Sundaram & Wysocka, 2020). The mechanisms by which TEs affect host gene-regulatory networks include functioning as cis-regulatory elements, modifying 3D chromatin architecture, producing novel regulatory genes and ncRNAs, and the collateral effects of TE silencing mechanisms (Fueyo et al., 2022).

In this study, the enrichment of repeats in differentially expressed genes highlights the significant role of repeats in gene regulation. DEGs, especially those downregulated, showed a higher enrichment of repeats within their gene bodies compared to non-DEGs, suggesting that repeats may act as regulatory elements influencing gene expression.

The analysis also revealed a strong association between differentially expressed repeats (DERs) and their host genes. DERs were more likely to be located within gene bodies, and this localization was consistent across different treatment groups (5AzaD & DMSO) and time points (12-hour & 24-hour). Interestingly, while DERs were more concentrated within gene bodies compared to non-DERs, there was no significant difference in the localization of upregulated versus downregulated DERs within gene bodies. This suggested that the positional preference of repeats within gene bodies may be a general feature rather than being influenced by their expression status.

Moreover, the expression patterns of DERs tended to mirror those of their host genes. Over 95% of DERs within the gene bodies of DEGs exhibited the same direction of

expression change as their host genes. This indicated a coordinated regulation between repeats and their host genes, implying that the expression of these repeats may rely on the transcriptional activity of host genes.

Similar gene and transposon co-expression has been observed in a transcriptome study of the *Drosophila* brain. By combining genomic DNA sequencing data with single-cell mRNA sequencing data to map transposon expression, it was found that most transposon-containing transcripts are chimeric, indicating that TE expression was non-autonomous and driven by neighbouring genes. The same study also suggested that TE expression in specific neurons is linked to alternative splicing of genes containing those TEs in their introns (Treiber & Waddell, 2020).

Intronic repeats account for a significant proportion of the *Parhyale* genome. This may be related to the fact that transposable elements can introduce splicing sites into introns, leading to the generation of chimeric transcripts. Therefore, the expression activity of these intragenic repeats appears to be driven by the host gene, which potentially explains the correlated expression pattern between genes and the residing repeats. However, future studies should examine if these intragenic repeats are frequently expressed as parts of chimeric mRNAs with genes and if TE expression is non-autonomous from host genes to further support this idea.

5.1.7 Conclusion

The regulatory role of DNA methylation in crustaceans remains largely unexplored. A comprehensive genome-wide analysis of this epigenetic marker, combined with the temporal expression patterns during the maternal-to-zygotic transition and zygotic genome activation, provides significant insights. These findings illuminate the potential for disruptions in DNA methylation to lead to failures in downstream developmental regulatory networks in *Parhyale*. Additionally, this knowledge is crucial for understanding

the function and evolution of DNA methylation, a ubiquitous yet poorly understood epigenetic modification in invertebrates.

In conclusion, our study investigated the relationship between DNA methylation and gene transcription, as well as repetitive elements. Our findings highlighted the crucial role of DNA methylation in the early development of *Parhyale*. Moreover, we identified the association between gene body methylation and gene expression, proposing that DNA methylation serves as a potentially key epigenetic regulator. This regulation is essential for managing the precise temporal dynamics of gene expression during the early developmental stages of *Parhyale*. By shedding light on the mechanisms that regulate gene expression, these insights significantly advance our understanding of the molecular frameworks that underpin embryonic development in this species.

5.2 Future directions

In chapter 4, the initial hypothesis posited that 5AzaD treatment would result in a decrease in methylated CpG sites compared to controls. This expectation was founded on 5AzaD's established mechanism of action, which includes its integration into DNA and its inherent chemical properties that inhibit methylation(Qin et al., 2009). To test this hypothesis, I performed targeted gene bisulfite sequencing on three specific gene regions for embryos collected at the 24-hour time point, after removing 5AzaD at the 12-hour time point. Contrary to expectations, no significant differences in methylation were observed between the 5AzaD-treated embryos and the vehicle control group across the selected regions at the 24-hour time point. Therefore, it is essential to acknowledge the lack of definitive confirmation regarding the demethylating effect of 5AzaD on DNA methylation in this thesis. Consequently, drawing firm conclusions regarding the functional significance of DNA methylation alterations induced by 5AzaD treatment remains tentative.

Several factors could explain the unchanged methylation status observed in the targeted gene bisulfite sequencing experiment. First, the validation method employed, Sanger sequencing of PCR products from bisulfite-treated DNA, may not be sufficiently reliable due to excessive background noise and over-scaled cytosine signals (Brisotto et al., 2015). Second, it's possible that 5AzaD does not induce global demethylation but instead targets specific CpG sites. The regions selected for validation might contain genes that are resistant to demethylation and thus less likely to be affected by 5AzaD treatment (Cook et al., 2019; Hagemann et al., 2011). This phenomenon has also been observed by other researchers who noted that 5AzaD did not alter the DNA methylation level of a few selected genes but led to overall demethylation across the whole genome, confirmed by WGBS (Cook et al., 2019). Moreover, after the removal of 5AzaD, DNA continues to replicate and incorporate normal cytosine, gradually reducing the proportion of incorporated 5AzaD, eventually restoring methylation patterns over time. This is likely

more pronounced in cells with high division rates, such as early developing embryos. Therefore, despite the potential long-lasting effects of 5AzaD treatment, it is possible that methylation could have been re-established by the 24-hour time point in *Parhyale* embryos, which had developed for an additional 12 hours after the removal of 5AzaD at the 12-hour time point.

To enhance the robustness and reliability of our findings and further our understanding of the functional consequences of DNA methylation dynamics in *Parhyale* early development, alternative strategies need to be employed. These include comprehensive assessment methods like whole-genome bisulfite sequencing or high-throughput bisulfite amplicon sequencing (BSAS) (Cook et al., 2019; Masser et al., 2015). In addition to sequencing method, alternative techniques for detecting changes in DNA methylation, such as immunofluorescence staining (Chowdhury et al., 2015; Kumar & Kim, 2017; Liyanage et al., 2013) and the LUminometric Methylation Assay (LUMA) (Karimi et al., 2006), could provide complementary insights into the potential spatial and temporal distribution of DNA methylation alterations induced by 5AzaD treatment. Nonetheless, it is crucial to validate the demethylating effect of 5AzaD immediately after treatment to avoid the potential restoration of the methylation pattern.

After confirming the demethylating effect of 5AzaD, it would be valuable to collect additional samples from various stages, particularly those spanning the MZT and ZGA period (12-16 hours). These specific time points would enable us to capture the complex and dynamic alterations in gene expression that occur during these crucial developmental stages. Expanding the sample collection to include these stages will provide valuable insights into the intricate processes occurring during early development of *Parhyale*.

Given the cytotoxicity associated with cytidine analogues, it is crucial to consider alternative strategies for reducing DNA methylation levels in future experiments. Among the available techniques capable of modifying DNA methylation, CRISPR/Cas9 and

morpholino injections have shown promise (Kang et al., 2019; Taghbalout et al., 2019; Tang et al., 2022). The advent of CRISPR-based technologies has revolutionized the precision of DNA methylation editing. CRISPR has also been effectively used in *Parhyale* for somatic mutagenesis and the functional analysis of targeted genes (Bruce & Patel, 2020; Kao et al., 2016; Martin et al., 2016; Serano et al., 2016). CRISPR/Cas9 knockout experiments on methyltransferase genes were conducted by my colleague Dr. Salha Rawas, revealing lethality in embryos (Rawas, 2023). Importantly, the CRISPR knockout of dnmt genes results in permanent alterations, precluding the possibility of specifically targeting only the maternal-to-zygotic transition (MZT) period, whereas drug treatments can achieve this temporal specificity.

Furthermore, morpholinos offer a distinctive approach to modulate the activity of DNA methyltransferases or their associated factors, as well as other components of the DNA methylation machinery. By specifically targeting mRNA transcripts that encode DNA methylation enzymes or proteins involved in their regulation, morpholinos can inhibit their expression or disrupt their function, resulting in alterations in DNA methylation patterns (Anderson et al., 2009; Gao et al., 2017; Shimoda et al., 2005; Tang et al., 2022).

Reference

Aanes, H., Winata, C. L., Lin, C. H., Chen, J. P., Srinivasan, K. G., Lee, S. G. P. P., Lim, A. Y. M. M., Hajan, H. S., Collas, P., Bourque, G., Gong, Z., Korzh, V., Aleström, P., & Mathavan, S. (2011). Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Research*, *21*(8), 1328–1338.

Abe, K., Yamamoto, R., Franke, V., Cao, M., Suzuki, Y., Suzuki, M. G., Vlahovicek, K., Svoboda, P., Schultz, R. M., & Aoki, F. (2015). The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3′ processing. *The EMBO Journal*, *34*(11), 1523–1537.

Adhami, H. Al, Bardet, A. F., Dumas, M., Cleroux, E., Guibert, S., Fauque, P., Acloque, H., & Weber, M. (2022). A comparative methylome analysis reveals conservation and divergence of DNA methylation patterns and functions in vertebrates. *BMC Biology*, *20*(1), 1–18.

Aigner, G. P., Nenning, P., Fiechtner, B., Šrut, M., & Höckner, M. (2022). DNA methylation and detoxification in the earthworm *Lumbricus terrestris* exposed to cadmium and the DNA demethylation agent 5-aza-2′-deoxycytidine. *Toxics*, *10*(100), 1–13.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, *13*(10).

Alberti, A., Belser, C., Engelen, S., Bertrand, L., Orvain, C., Brinas, L., Cruaud, C., Giraut, L., Da Silva, C., Firmo, C., Aury, J. M., & Wincker, P. (2014). Comparison of library preparation methods reveals their impact on interpretation of metatranscriptomic data. *BMC Genomics*, *15*(912), 1–13.

Alegretti, L., Umbuzeiro, G. D. A., & Flynn, M. N. (2016). Population dynamics of Parhyale hawaiensis (Dana, 1853) (Amphipoda: Hyalidae) associated with an intertidal algal belt in southeastern Brazil. *Journal of Crustacean Biology*, *36*(6), 785–791.

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, *12*(5), 363–376.

Almeida, L. M., Silva, I. T., Silva, W. A., Castro, J. P., Riggs, P. K., Carareto, C. M., & Amaral, M. E. J. (2007). The contribution of transposable elements to *Bos taurus* gene structure. *Gene*, *390*(1–2), 180–189.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. In *Journal of Molecular Biology* (Vol. 215, pp. 403–410).

Alwes, F., Enjolras, C., & Averof, M. (2016). Live imaging reveals the progenitors and cell dynamics of limb regeneration. *ELife*, *5*(e19766), 1–18.

Amaral, P. P., & Mattick, J. S. (2008). Noncoding RNA in development. *Mammalian Genome*, *19*, 454–492.

Amarasinghe, H. E., Clayton, C. I., & Mallon, E. B. (2014). Methylation and worker reproduction in the bumble-bee (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences*, *281*(1780), 20132502.

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*(1), 1–16.

Anderson, C. L., & Brown, C. J. (2002). Variability of X chromosome inactivation: effect on levels of TIMP1 RNA and role of DNA methylation. *Human Genetics*, *110*(3), 271–278.

Anderson, R. M., Bosch, J. A., Goll, M. G., Hesselson, D., Dong, P. D. S., Shin, D., Chi, N. C., Shin, C. H., Schlegel, A., Halpern, M., & Stainier, D. Y. R. (2009). Loss of Dnmt1 catalytic activity reveals multiple roles for DNA methylation during pancreas development and regeneration. *Developmental Biology*, *334*(1), 213–223.

Arsala, D., Wu, X., Yi, S. V., & Lynch, J. A. (2022). Dnmt1a is essential for gene body methylation and the regulation of the zygotic genome in a wasp. *PLoS Genetics*, *18*(5), e1010181.

Artieri, C. G., & Fraser, H. B. (2014). Transcript length mediates developmental timing of gene expression across *Drosophila*. *Molecular Biology and Evolution*, *31*(11), 2879–2889.

Ashok, C., Ahuja, N., Natua, S., Mishra, J., Samaiya, A., & Shukla, S. (2021). E2F1 and epigenetic modifiers orchestrate breast cancer progression by regulating oxygen-dependent ESRP1 expression. *Oncogenesis*, *10*(8), 1–10.

Athanasouli, M., & Rödelsperger, C. (2022). Analysis of repeat elements in the *Pristionchus pacificus* genome reveals an ancient invasion by horizontally transferred transposons. *BMC Genomics*, *23*(1), 1–14.

Avvakumov, G. V, Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C. H., & Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, *455*(7214), 822–825.

Balakishore, G., Thejaswini, H. B., & Puranik, S. B. (2021). Nonaqueous formulation and evaluation of decitabine. *Journal of Current Pharma Research*, *11*(2), 22–28.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., & Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, *354*(6314), 909–912.

Baril, T., Galbraith, J., & Hayward, A. (2024). Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. *Molecular Biology and Evolution*, *41*(4).

Barlow, D. P., & Bartolomei, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harbor Perspectives in Biology*, *6*(a018382), 143–149.

Barnard, J. L. L. (1965). Marine amphipoda of atolls in micronesia. *Proceedings of the United States National Museum*, *117*(3516), 459–552.

Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L., & Hunter, C. P. (2003). Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. *Development*, *130*(5), 889–900.

Bernstein, B. E., Meissner, A., & Lander, E. S. (2007). The mammalian epigenome. *Cell*, *128*(4), 669–681.

Bestor, T. H., & Bourc'his, D. (2004). Transposon silencing and imprint establishment in mammalian germ cells. *Cold Spring Harbor Symposia on Quantitative Biology*, *69*, 381–387.

Bewick, A. J., Sanchez, Z., McKinney, E. C., Moore, A. J., Moore, P. J., & Schmitz, R. J. (2019). Dnmt1 is essential for egg production and embryo viability in the large milkweed bug, *Oncopeltus fasciatus*. *Epigenetics and Chromatin*, *12*(6), 1–14.

Bewick, A. J., Vogel, K. J., Moore, A. J., & Schmitz, R. J. (2017). Evolution of DNA methylation across insects. *Molecular Biology and Evolution*, *34*(3), 654–665.

Bewick, A. J., Zhang, Y., Wendte, J. M., Zhang, X., & Schmitz, R. J. (2019). Evolutionary and experimental loss of gene body methylation and its consequence to gene expression. *G3: Genes, Genomes, Genetics*, *9*(8), 2441–2445.

Bhutani, N., Burns, D. M., & Blau, H. M. (2011). DNA demethylation dynamics. *Cell*, *146*(6), 866–872.

Biémont, C., & Vieira, C. (2006). Junk DNA as an evolutionary force. *Nature*, *443*(7111), 521–524.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes and Development*, *16*(1), 6–21.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature*, *15*, 1–5.

Biswas, S., & Rao, C. M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *European Journal of Pharmacology*, *837*, 8–24.

Blattler, A., Yao, L., Wang, Y., Ye, Z., Jin, V. X., & Farnham, P. J. (2013). ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes. *Epigenetics and Chromatin*, *6*(13), 1–18.

Blythe, M. J., Malla, S., Everall, R., huan Shih, Y., Lemay, V., Moreton, J., Wilson, R., & Aboobaker, A. A. (2012). High through-put sequencing of the *Parhyale hawaiensis* mRNAs and microRNAs to aid comparative developmental studies. *PLoS ONE*, *7*(3), e33784.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bollati, V., & Baccarelli, A. (2010). Environmental epigenetics. *Heredity*, *105*(1), 105–112.

Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y., Jin, Z., Shen, S. S., Wang, Z., Wang, W., Wang, J., Berger, S. L., Liebig, J., Zhang, G., & Reinberg, D. (2012). Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Current Biology*, *22*(19), 1755–1764.

Borowsky, B., & Borowsky, R. (1987). The reproductive behaviors of the amphipod crustacean *Gammarus palustris* (Bousfield) and some insights into the nature of their stimuli. *Journal of Experimental Marine Biology and Ecology*, *107*(2), 131–144.

Brisotto, G., Gennaro, A., Damiano, V., Armellin, M., & Perin, T. (2015). An improved sequencing-based strategy to estimate locus-specific DNA methylation. *BMC Cancer*, *15*(639), 1–10.

Brockdorff, N., & Turner, B. M. (2015). Dosage compensation in mammals. *Cold Spring Harbor Perspectives in Biology*, *7*(3), a019406.

Browne, W. E., Price, A. L., Gerberding, M., & Patel, N. H. (2005). Stages of embryonic development in the amphipod crustacean, *Parhyale hawaiensis*. *Genesis*, *42*(3), 124–149.

Bruce, H. S., & Patel, N. H. (2020). Knockout of crustacean leg patterning genes suggests that insect wings and body walls evolved from ancient leg segments. *Nature Ecology and Evolution*, *4*(12), 1703–1712.

Brueckner, B., Boy, R. G., Siedlecki, P., Musch, T., Kliem, H. C., Zielenkiewicz, P., Suhai, S., Wiessler, M., & Lyko, F. (2005). Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. *Cancer Research*, *65*(14), 6305–6311.

Buck-Koehntop, B. A., & Defossez, P. A. (2013). On how mammalian transcription factors recognize methylated DNA. *Epigenetics*, *8*(2), 131–137.

Buitrago, D., Labrador, M., Arcon, J. P., Lema, R., Flores, O., Esteve-Codina, A., Blanc, J., Villegas, N., Bellido, D., Gut, M., Dans, P. D., Heath, S. C., Gut, I. G., Brun Heath, I., & Orozco, M. (2021). Impact of DNA methylation on 3D genome structure. *Nature Communications*, *12*(1).

Calvo, L., Birgaoanu, M., Pettini, T., Ronshaugen, M., & Griffiths-Jones, S. (2022). The embryonic transcriptome of *Parhyale hawaiensis* reveals different dynamics of microRNAs and mRNAs during the maternal-zygotic transition. *Scientific Reports*, *12*(174), 1–15.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S.

E., Batalov, S., Forrest, A. R. R., Zavolan, M., Davis, M. J., Wilming, L. G., … Wang, J. (2005). The transcriptional landscape of the mammalian genome. *Science*, *309*(5740), 1559–1563.

Carrel, L., Clemson, C. M., Dunn, J. M., Miller, A. P., Hunt, P. A., Lawrence, J. B., & Willard, H. F. (1996). X inactivation analysis and DNA methylation studies of the ubiquitin activating enzyme E1 and PCTAIRE-1 genes in human and mouse. *Human Molecular Genetics*, *5*(3), 391–401.

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., & Workman, J. L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, *123*(4), 581–592.

Cedar, H., & Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics*, *10*(5), 295–304.

Chalopin, D., Naville, M., Plard, F., Galiana, D., & Volff, J. N. (2015). Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biology and Evolution*, *7*(2), 567–580.

Chatterjee, A., Rodger, E. J., Morison, I. M., Eccles, M. R., Stockwell, P. A., Seymour, G. J., Cullinan, M. P., & Editors, N. C. K. H. (2017). Tools and strategies for analysis of genome-wide and gene-specific DNA methylation patterns. *Methods Mol Biol,* 1537:249-277.

Chen, J., Cheng, J., Chen, X., Inoue, M., Liu, Y., & Song, C. X. (2022). Whole-genome long-read TAPS deciphers DNA methylation patterns at base resolution using PacBio SMRT sequencing technology. *Nucleic Acids Research*, *50*(18), E104.

Chen, Q., Liu, B., Zeng, Y., Hwang, J. W., Dai, N., Correa, I. R., Estecio, M. R., Zhang, X., Santos, M. A., Chen, T., & Cheng, X. (2023). GSK-3484862 targets DNMT1 for degradation in cells. *NAR Cancer*, *5*(2), 1–13.

Chen, S., Hu, X., Zhu, W., Jia, C., Han, X., Yuan, J., Sun, Z., Yang, Z., Geng, T., & Cui, H. (2017). Reactivation of development-related genes by the DNA methylation inhibitor 5-Aza-2′-deoxycytidine in chicken embryo fibroblasts. *Poultry Science*, *96*(4), 1007–1014.

Chen, S., Krinsky, B. H., & Long, M. (2013). New genes as drivers of phenotypic evolution. *Nature Reviews Genetics*, *14*(9), 645–660.

Chernyavskaya, Y., Mudbhary, R., Zhang, C., Tokarz, D., Jacob, V., Gopinath, S., Sun, X., Wang, S., Magnani, E., Madakashira, B. P., Yoder, J. A., Hoshida, Y., & Sadler, K. C. (2017). Loss of DNA methylation in zebrafish embryos activates retrotransposons to trigger antiviral signaling. *Development*, *144*(16), 2925–2939.

Chowdhury, B., McGovern, A., Cui, Y., Choudhury, S. R., Cho, I. H., Cooper, B., Chevassut, T., Lossie, A. C., & Irudayaraj, J. (2015). The hypomethylating agent Decitabine causes a paradoxical increase in 5-hydroxymethylcytosine in human leukemia cells. *Scientific Reports*, *5*(Ii), 1–8.

Christman, J. K. (2002). 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: Mechanistic studies and their implications for cancer therapy. *Oncogene*, *21*(35), 5483–5495.

Chuang, J. C., Yoo, C. B., Kwan, J. M., Li, T. W. H., Liang, G., Yang, A. S., & Jones, P. A. (2005). Comparison of biological effects of non-nucleoside DNA methylation inhibitors versus 5-aza-2'-deoxycytidine. *Molecular Cancer Therapeutics*, *4*(10), 1515–1520.

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics*, *18*(2), 71–86.

Clark, T. A., Lu, X., Luong, K., Dai, Q., Boitano, M., Turner, S. W., He, C., & Korlach, J. (2013). Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biology*, *11*(4), 1–10.

Clark, T. A., Murray, I. A., Morgan, R. D., Kislyuk, A. O., Spittle, K. E., Boitano, M., Fomenkov, A., Roberts, R. J., & Korlach, J. (2012). Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Research*, *40*(4), e29.

Clark-Hachtel, C. M., & Tomoyasu, Y. (2020). Two sets of candidate crustacean wing homologues and their implication for the origin of insect wings. *Nature Ecology and Evolution*, *4*(12), 1694–1702.

Claus, R., Almstedt, M., & Lübbert, M. (2005). Epigenetic treatment of hematopoietic malignancies: *in vivo* targets of demethylating agents. *Seminars in Oncology*, *32*(5), 511–520.

Collart, C., Owens, N. D. L., Bhaw-Rosun, L., Cooper, B., De Domenico, E., Patrushev, I., Sesay, A. K., Smith, J. N., Smith, J. C., & Gilchrist, M. J. (2014). High-resolution analysis of gene activity during the *Xenopus* mid-blastula transition. *Development*, *141*(9), 1927–1939.

Coloundre, C., Weber, K., Files, G., Platt, T., Coulondre, C., Miller, J. H., Farabaugh, P. J., & Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli. Nature*, *274*(24), 775–780.

Cook, N., Parker, D. J., Tauber, E., Pannebakker, B. A., & Shuker, D. M. (2019). Validating the demethylating effects of 5-aza-2′-deoxycytidine in insects requires a whole-genome approach: (A reply to ellers et al.). *American Naturalist*, *194*(3), 432–438.

Cooper, D. N., & Krawczak, M. (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Human Genetics*, *83*(2), 181–188.

Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics*, *10*(10), 691–703.

Cotton, A. M., Lam, L., Affleck, J. G., Wilson, I. M., Peñaherrera, M. S., McFadden, D. E., Kobor, M. S., Lam, W. L., Robinson, W. P., & Brown, C. J. (2011). Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Human Genetics*, *130*(2), 187–201.

Dan, H., Zhang, S., Zhou, Y., & Guan, Q. (2019). DNA methyltransferase inhibitors: catalysts for antitumour immune responses. *OncoTargets and Therapy*, *12*, 10903–10916.

Dana, J. D. (1853). Crustacea. Part II. In *United States Exploring Expedition v.14*.

Dasmahapatra, A. K., & Khan, I. A. (2015). DNA methyltransferase expressions in Japanese rice fish (*Oryzias latipes*) embryogenesis is developmentally regulated and modulated by ethanol and 5-azacytidine. *Comparative Biochemistry and Physiology Part - C: Toxicology and Pharmacology*, *176–177*, 1–9.

de Mendoza, A., Hatleberg, W. L., Pang, K., Leininger, S., Bogdanovic, O., Pflueger, J., Buckberry, S., Technau, U., Hejnol, A., Adamska, M., Degnan, B. M., Degnan, S. M., & Lister, R. (2019). Convergent evolution of a vertebrate-like methylome in a marine sponge. *Nature Ecology and Evolution*, *3*(10), 1464–1473.

de Mendoza, A., Pflueger, J., & Lister, R. (2019). Capture of a functionally active methyl-CpG binding domain by an arthropod retrotransposon family. *Genome Research*, *29*(8), 1277–1286.

De Smet, C., & Loriot, A. (2012). DNA Hypomethylation and Activation of Germline-Specific Genes in Cancer. In *Epigenetic Alterations in Oncogenesis* (Vol. 754, pp. 149–166). Springer Science+Business Media.

De Smet, C., Lurquin, C., Lethé, B., Martelange, V., & Boon, T. (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Molecular and Cellular Biology*, *19*(11), 7327–7335.

Deniz, Ö., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics*, *20*(7), 417–431.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., … Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789.

Diao, Y. F., Naruse, K., Li, X. X., Han, R. X., Kim, D. K., Lin, T., & Jin, D. I. I. (2013). Effects of 5-aza-2'-deoxyctidine on the development of porcine parthenogenetic and nuclear transfer embryos. *Natural Science*, *05*(7A), 31–37.

Dimond, J. L., Gamblewood, S. K., & Roberts, S. B. (2017). Genetic and epigenetic insight into morphospecies in a reef coral. *Molecular Ecology*, *26*(19), 5031–5042.

Dimond, J. L., & Roberts, S. B. (2015). Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. *Molecular Ecology*, *25*(8), 1895–1904.

Dinger, M. E., Amara, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., Askarian-Amiri, M. E., Ru, K., Soldà, G., Simons, C., Sunkin, S. M., Crowe, M. L., Grimmond, S. M., Perkins, A. C., & Mattick, J. S. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research*, *18*(9), 1433–1445.

Dinger, M. E., Amaral, P. P., Mercer, T. R., & Mattick, J. S. (2009). Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. *Briefings in Functional Genomics and Proteomics*, *8*(6), 407–423.

Dixon, G., Liao, Y., Bay, L. K., & Matz, M. V. (2018). Role of gene body methylation in acclimatization and adaptation in a basal metazoan. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(52), 13342–13346.

Dixon, G., & Matz, M. (2022). Changes in gene body methylation do not correlate with changes in gene expression in Anthozoa or Hexapoda. *BMC Genomics*, *23*(234), 1–11.

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108.

Domcke, S., Bardet, A. F., Adrian Ginno, P., Hartl, D., Burger, L., & Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*, *528*(7583), 575–579.

Drongitis, D., Aniello, F., Fucci, L., & Donizetti, A. (2019). Roles of transposable elements in the different layers of gene expression regulation. *International Journal of Molecular Sciences*, *20*(22).

Eckersley-Maslin, M. A., Alda-Catalinas, C., & Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nature Reviews Molecular Cell Biology*, *19*(7), 436–450.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V, Davies, R., Down, T. A., Haefliger, C., Horton, R., Howe, K., Jackson, D. K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., … Beck, S. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genetics*, *38*(12), 1378–1385.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Edwards, C. A., & Ferguson-Smith, A. C. (2007). Mechanisms regulating imprinted genes in clusters. *Current Opinion in Cell Biology*, *19*(3), 281–289.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., … Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*(5910), 133–138.

Elango, N., Hunt, B. G., Goodisman, M. A. D., & Yi, S. V. (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(27), 11206–11211.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 1–14.

Espada, J. (2012). Non-catalytic functions of DNMT1. *Epigenetics*, *7*(2), 115–118.

Fatemi, M., & Wade, P. A. (2006). MDB family proteins: reading the epigenetic code. *Journal of Cell Science*, *119*(15), 3033–3037.

Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E., Ukomadu, C., Sadler, K. C., Pradhan, S., Pellegrini, M., & Jacobsen, S. E. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8689–8694.

Feng, S., Zhong, Z., Wang, M., & Jacobsen, S. E. (2020). Efficient and accurate determination of genome-wide DNA methylation patterns in *Arabidopsis thaliana* with enzymatic methyl sequencing. *Epigenetics and Chromatin*, *13*(24), 1–17.

Field, L. M. (2000). Methylation and expression of amplified esterase genes in the aphid *Myzus persicae* (Sulzer). *Biochemical Journal*, *349*(3), 863–868.

Filion, G. J. P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., & Defossez, P.-A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and Cellular Biology*, *26*(1), 169–181.

Fisher, A. G. (2002). Cellular identity and lineage choice. *Nature Reviews Immunology*, *2*, 977–982.

Flores, K., Wolschin, F., Corneveaux, J. J., Allen, A. N., Huentelman, M. J., & Amdam, G. V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*, *13*(1).

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., & Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, *7*(6), 461–465.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451–9457.

Fort, V., Khelifi, G., & Hussein, S. M. I. I. (2021). Long non-coding RNAs and transposable elements: A functional relationship. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1868*(1), 118837.

Frith, M. C., Pheasant, M., & Mattick, J. S. (2005). The amazing complexity of the human transcriptome. *European Journal of Human Genetics*, *13*(8), 894–897.

Fueyo, R., Judd, J., Feschotte, C., & Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. *Nature Reviews Molecular Cell Biology*, *23*(7), 481–497.

Futscher, B. W., Oshiro, M. M., Wozniak, R. J., Holtan, N., Hanigan, C. L., Duan, H., & Domann, F. E. (2002). Role for DNA methylation in the control of cell type-specific maspin expression. *Nature Genetics*, *31*(2), 175–179.

Gao, F., Liu, X., Wu, X. P., Wang, X. L., Gong, D., Lu, H., Xia, Y., Song, Y., Wang, J., Du, J., Liu, S., Han, X., Tang, Y., Yang, H., Jin, Q., Zhang, X., & Liu, M. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. *Genome Biology*, *13*(10), R100.

Gao, F., Niu, Y., Sun, Y. E., Lu, H., Chen, Y., Li, S., Kang, Y., Luo, Y., Si, C., Yu, J., Li, C., Sun, N., Si, W., Wang, H., Ji, W., & Tan, T. (2017). *De novo* DNA methylation during monkey pre-implantation embryogenesis. *Cell Research*, *27*(4), 526–539.

Gardiner-Garden, M., Frommer, M., Majumdar, K. C., & Samanta, S. K. (1987). CpG Islands in vertebrate genomes. *Journal of Molecular Biology*, *196*(2), 261–282.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784–3788.

Gatzmann, F., Falckenhayn, C., Gutekunst, J., Hanna, K., Raddatz, G., Carneiro, V. C., & Lyko, F. (2018). The methylome of the marbled crayfish links gene body methylation to stable expression of poorly accessible genes. *Epigenetics and Chromatin*, *11*(57), 1–12.

Gavery, M. R., & Roberts, S. B. (2010). DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics*, *11*(1).

Gavery, M. R., & Roberts, S. B. (2013). Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc. *PeerJ*, *1*(e215), 1–15.

Gebrie, A. (2023). Transposable elements as essential elements in the control of gene expression. *Mobile DNA*, *14*(1), 1–17.

Gelfman, S., Cohen, N., Yearim, A., & Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Research*, *23*(5), 789–799.

Gerberding, M., Browne, W. E., & Patel, N. H. (2002). Cell lineage analysis of the amphipod crustacean *Parhyale hawaiensis* reveals an early restriction of cell fates. *Development*, *129*(24), 5789–5801.

Geyer, K. K., Rodríguez López, C. M., Chalmers, I. W., Munshi, S. E., Truscott, M., Heald, J., Wilkinson, M. J., & Hoffmann, K. F. (2011). Cytosine methylation regulates oviposition in the pathogenic blood fluke *Schistosoma mansoni*. *Nature Communications*, *2*(424), 1–10.

Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T., & Jacob, S. T. (2005). 5-Aza-Deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN Box, bromo-adjacent homology domain, and nuclear localization signal. *Molecular and Cellular Biology*, *25*(11), 4727–4741.

Gilbert, C., Peccoud, J., & Cordaux, R. (2021). Transposable Elements and the Evolution of Insects. *Annual Review of Entomology*, *66*, 355–372.

Gildor, T., Malik, A., Sher, N., & Ben-Tabou de-Leon, S. (2016). Mature maternal mRNAs are longer than zygotic ones and have complex degradation kinetics in sea urchin. *Developmental Biology*, *414*(1), 121–131.

Gilmartin, A. G., Groy, A., Gore, E. R., Atkins, C., Long, E. R., Montoute, M. N., Wu, Z., Halsey, W., Mcnulty, D. E., Ennulat, D., Rueda, L., Pappalardi, M. B., Kruger, R. G., Mccabe, M. T., Raoof, A., Butlin, R., Stowell, A., Cockerill, M., Waddell, I., … Benowitz, A. B. (2021). *In vitro* and *in vivo* induction of fetal hemoglobin with a reversible and selective DNMT1 inhibitor. *Haematologica*, *106*(7), 1979–1987.

Gnyszka, A., Jastrzębski, Z., & Flis, S. (2013). DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. *Anticancer Research*, *33*, 2989–2996.

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell*, *128*(4), 635–638.

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, *74*(1), 481–514.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., Golic, K. G., Jacobsen, S. E., & Bestor, T. H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science*, *311*(5759), 395–398.

Gouil, Q., & Keniry, A. (2019). Latest techniques to study DNA methylation. *Essays in Biochemistry*, *63*(6), 639–648.

Gowher, H., Liebert, K., Hermann, A., Xu, G., & Jeltsch, A. (2005). Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *Journal of Biological Chemistry*, *280*(14), 13341–13348.

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, *20*(10), 590–607.

Gros, C., Fleury, L., Nahoum, V., Faux, C., Valente, S., Labella, D., Cantagrel, F., Rilova, E., Bouhlel, M. A., David-Cordonnier, M. H., Dufau, I., Ausseil, F., Mai, A., Mourey, L., Lacroix, L., & Arimondo, P. B. (2015). New insights on the mechanism of quinoline-based DNA methyltransferase inhibitors. *Journal of Biological Chemistry*, *290*(10), 6293–6302.

Grunau, C., Clark, S. J., & Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research*, *29*(13), 65.

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F. G. L., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J., & Xu, G. F. G. L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, *477*(7366), 606–612.

Guo, J. U., Su, Y., Shin, J. J. H., Shin, J. J. H., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G., Song, H., Sciences, M., & Angeles, L. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature Neuroscience*, *17*(2), 215–222.

Guo, Y., Li, C. I., Ye, F., & Shyr, Y. (2013). Evaluation of read count based RNAseq analysis methods. *BMC Genomics*, *14*(SUPP 8), S2.

Guo, Y., Sheng, Q., Li, J., Ye, F., Samuels, D. C., & Shyr, Y. (2013). Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data. *PLoS ONE*, *8*(8), e71462.

Guo, Y., Zhao, S., Sheng, Q., Guo, M., Lehmann, B., Pietenpol, J., Samuels, D. C., & Shyr, Y. (2015). RNAseq by total RNA library identifies additional RNAs compared to Poly(A) RNA library. *BioMed Research International*, *2015*, 1–9.

Guynes, K., Sarre, L. A., Carrillo-Baltodano, A. M., Davies, B. E., Xu, L., Liang, Y., Martín-Zamora, F. M., Hurd, P. J., De Mendoza, A., & Martín-Durán, J. M. (2023). Annelid methylomes reveal ancestral developmental and ageing-associated epigenetic erosion across Bilateria. *BioRxiv*.

Hackett, J. A., & Surani, M. A. (2012). DNA methylation dynamics during the mammalian life cycle. *Philosophical Transactions of the Royal Society B*, *368*(1609), 20110328.

Hadjiargyrou, M., & Delihas, N. (2013). The intertwining of transposable elements and non-coding RNAs. *International Journal of Molecular Sciences*, *14*(7), 13307–13328.

Hagemann, S., Heil, O., Lyko, F., & Brueckner, B. (2011). Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. *PLoS ONE*, *6*(3), e17388.

Haggerty, C., Kretzmer, H., Riemenschneider, C., Kumar, A. S., Mattei, A. L., Bailly, N., Gottfreund, J., Giesselmann, P., Weigert, R., Brändl, B., Giehr, P., Buschow, R., Galonska, C., von Meyenn, F., Pappalardi, M. B., McCabe, M. T., Wittler, L., Giesecke-Thiel, C., Mielke, T., … Meissner, A. (2021). Dnmt1 has *de novo* activity targeted to transposable elements. In *Nature Structural and Molecular Biology* (Vol. 28, Issue 7). Springer US.

Hamatani, T., Carter, M. G., Sharov, A. A., & Ko, M. S. H. (2004). Dynamics of global gene expression changes during mouse preimplantation development. *Developmental Cell*, *6*, 117–131.

Han, Y., Zheleznyakova, G. Y., Marincevic-Zuniga, Y., Kakhki, M. P., Raine, A., Needhamsen, M., & Jagodic, M. (2022). Comparison of EM-seq and PBAT methylome library methods for low-input DNA. *Epigenetics*, *17*(10), 1195–1204.

Hancks, D. C., & Kazazian, H. H. (2012). Active human retrotransposons: variation and disease. *Current Opinion in Genetics and Development*, *22*(3), 191–203.

Hancks, D. C., & Kazazian, H. H. (2016). Roles for retrotransposon insertions in human disease. *Mobile DNA*, *7*(9), 1–28.

Hansen, J. C., Ghosh, R. P., & Woodcock, C. L. (2010). Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin. *IUBMB Life*, *62*(10), 732–738.

Hansen, R. S., & Gartler, S. M. (1990). 5-Azacytidine-induced reactivation of the human X chromosome-linked PGK1 gene is associated with a large region of cytosine demethylation in the 5′ CpG island. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(11), 4174–4178.

Harvey, S. A., Sealy, I., Kettleborough, R., Fenyes, F., White, R., Stemple, D., & Smith, J. C. (2013). Identification of the zebrafish maternal and paternal transcriptomes. *Development*, *140*(13), 2703–2710.

Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice. *Development*, *129*, 1983–1993.

He, Y., & Ecker, J. R. (2015). Non-CG methylation in the human genome. *Annual Review of Genomics and Human Genetics*, *16*(1), 55–77.

Hershko, A. Y., Kafri, T., Fainsod, A., & Razin, A. (2003). Methylation of HoxA5 and HoxB5 and its relevance to expression during mouse development. *Gene*, *302*(1–2), 65–72.

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A. T., & Neugebauer, K. M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. *Cell Reports*, *6*(2), 285–292.

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., & Sasaki, H. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes and Development*, *22*(12), 1607–1616.

Hoen, D. R., Hickey, G., Bourque, G., Casacuberta, J., Cordaux, R., Feschotte, C., Fiston-Lavier, A. S., Hua-Van, A., Hubley, R., Kapusta, A., Lerat, E., Maumus, F., Pollock, D. D., Quesneville, H., Smit, A., Wheeler, T. J., Bureau, T. E., & Blanchette, M. (2015). A call for benchmarking transposable element annotation methods. *Mobile DNA*, *6*(1).

Holliday, R., & Grigg, G. W. (1993). DNA methylation and mutation. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *285*(1), 61–67.

Holliday, R., & Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science*, *187*(4173), 226–232.

Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., & Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genetics*, *45*(10), 1198–1206.

Hong, X., Nelson, K., Lemke, N., & Kalkanis, S. N. (2012). Heparanase expression is associated with histone modifications in glioblastoma. *International Journal of Oncology*, *40*(2), 494–500.

Horton, J. R., Pathuri, S., Wong, K., Ren, R., Rueda, L., Fosbenner, D. T., Heerding, D. A., McCabe, M. T., Pappalardi, M. B., Zhang, X., King, B. W., & Cheng, X. (2022). Structural characterization of dicyanopyridine containing DNMT1-selective, non-nucleoside inhibitors. *Structure*, *30*(6), 793-802.e5.

Hu, C., Liu, X., Zeng, Y., Liu, J., & Wu, F. (2021). DNA methyltransferase inhibitors combination therapy for the treatment of solid tumor: mechanism and clinical application. *Clinical Epigenetics*, *13*(166), 1–16.

Hunt, B. G., Brisson, J. A., Yi, S. V, & Goodisman, M. A. D. (2010). Functional conservation of DNA methylation in the pea aphid and the honeybee. *Genome Biology and Evolution*, *2*(1), 719–728.

Hunt, B. J., Mallon, E. B., & Rosato, E. (2019). In silico Identification of a mlecular circadian system with novel features in the crustacean model organism *Parhyale hawaiensis*. *Frontiers in Physiology*, *10*, 1–9.

Iida, A., Shimada, A., Shima, A., Takamatsu, N., Hori, H., Takeuchi, K., & Koga, A. (2006). Targeted reduction of the DNA methylation level with 5-azacytidine promotes excision of the medaka fish Tol2 transposable element. *Genetical Research*, *87*(3), 187–193.

Issa, J. P. J., Kantarjian, H. M., & Kirkpatrick, P. (2005). Azacitidine. *Nature Reviews Drug Discovery*, *4*(4), 275–276.

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*, 1300–1303.

Iurlaro, M., von Meyenn, F., & Reik, W. (2017). DNA methylation homeostasis in human and mouse development. *Current Opinion in Genetics and Development*, *43*, 101–109.

Jabbour, E., Issa, J. P., Garcia-Manero, G., & Kantarjian, H. (2008). Evolution of decitabine development: Accomplishments, ongoing investigations, and future strategies. *Cancer*, *112*(11), 2341–2351.

Jangam, D., Feschotte, C., & Betrán, E. (2017). Transposable element domestication as an adaptation to evolutionary conflicts. *Trends in Genetics*, *33*(11), 817–831.

Jeong, H. H., Yalamanchili, H. K., Guo, C., Shulman, J. M., & Liu, Z. (2018). An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pacific Symposium on Biocomputing*, *0*(212669), 168–179.

Jiang, L., Zhang, J. J., Wang, J. J., Wang, L., Zhang, L., Li, G., Yang, X., Ma, X., Sun, X., Cai, J., Zhang, J. J., Huang, X., Yu, M., Wang, X., Liu, F., Wu, C. I., He, C., Zhang, B., Ci, W.,

& Liu, J. (2013). Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, *153*(4), 773–784.

Jiang, N., Wang, L., Chen, J., Wang, L., Leach, L., & Luo, Z. (2014). Conserved and divergent patterns of DNA methylation in higher vertebrates. *Genome Biology and Evolution*, *6*(11), 2998–3014.

Jones, P. A., & Laird, P. W. (1999). Cancer epigenetics comes of age. *Nature Genetics*, *21*(2), 163–167.

Jones, P. A., & Taylor, S. M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. *Cell*, *20*(1), 85–93.

Jørgensen, H. F., Ben-Porath, I., & Bird, A. P. (2004). Mbd1 Is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Molecular and Cellular Biology*, *24*(8), 3387–3395.

Kaiser, S., Jurkowski, T. P., Kellner, S., Schneider, D., Jeltsch, A., & Helm, M. (2017). The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. *RNA Biology*, *14*(9), 1241–1251.

Kaluscha, S., Domcke, S., Wirbelauer, C., Stadler, M. B., Durdu, S., Burger, L., & Schübeler, D. (2022). Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nature Genetics*, *54*(12), 1895–1906.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, H., Li, E., & Sasaki, H. (2004). Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, *429*(6994), 900–903.

Kang, J. G., Park, J. S., Ko, J. H., & Kim, Y. S. (2019). Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system. *Scientific Reports*, *9*(1), 1–12.

Kao, D., Lai, A. G., Stamataki, E., Rosic, S., Konstantinides, N., Jarvis, E., Di Donfrancesco, A., Pouchkina-Stancheva, N., Semon, M., Grillo, M., Bruce, H., Kumar, S., Siwanowicz, I., Le, A., Lemire, A., Eisen, M. B., Extavour, C., Browne, W. E., Wolff, C., … Aboobaker, A. (2016). The genome of the crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *ELife*, *5*, e20062.

Karimi, M., Johansson, S., Stach, D., Corcoran, M., Grandér, D., Schalling, M., Bakalkin, G., Lyko, F., Larsson, C., & Ekström, T. J. (2006). LUMA (LUminometric Methylation Assay)-A high throughput method to the analysis of genomic DNA methylation. *Experimental Cell Research*, *312*(11), 1989–1995.

Karpf, A. R., Lasek, A. W., Ririe, T. O., Hanks, A. N., Grossman, D., & Jones, D. A. (2004). Limited gene activation in tumor and normal epithelial cells treated with the DNA methyltransferase inhibitor 5-Aza-2′-deoxycytidine. *Molecular Pharmacology*, *65*(1), 18–27.

Kawato, S., Nishitsuji, K., Arimoto, A., Hisata, K., Kawamitsu, M., Nozaki, R., Kondo, H., Shinzato, C., Ohira, T., Satoh, N., Shoguchi, E., & Hirono, I. (2021). Genome and transcriptome assemblies of the kuruma shrimp, *Marsupenaeus japonicus*. *G3: Genes, Genomes, Genetics*, *11*(11).

Keller, T. E., Han, P., & Yi, S. V. (2016). Evolutionary transition of promoter and gene body DNA methylation across invertebrate-vertebrate boundary. *Molecular Biology and Evolution*, *33*(4), 1019–1028.

Kenny, N. J., Francis, W. R., Rivera-Vicéns, R. E., Juravel, K., de Mendoza, A., Díez-Vives, C., Lister, R., Bezares-Calderón, L. A., Grombacher, L., Roller, M., Barlow, L. D., Camilli, S., Ryan, J. F., Wörheide, G., Hill, A. L., Riesgo, A., & Leys, S. P. (2020). Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. *Nature Communications*, *11*(1), 1–11.

Kerr, L., Kafetzopoulos, I., Grima, R., & Sproul, D. (2023). Genome-wide single-molecule analysis of long-read DNA methylation reveals heterogeneous patterns at heterochromatin that reflect nucleosome organisation. In *PLoS Genetics* (Vol. 19, Issue 10 October).

Khatoon, Z., Figler, B., Zhang, H., & Cheng, F. (2014). Introduction to RNA-seq and its applications to drug discovery and development. *Drug Development Research*, *75*(5), 324–330.

Kikuchi, A., Onoda, H., Yamaguchi, K., Kori, S., Matsuzawa, S., Chiba, Y., Tanimoto, S., Yoshimi, S., Sato, H., Yamagata, A., Shirouzu, M., Adachi, N., Sharif, J., Koseki, H., Nishiyama, A., Nakanishi, M., Defossez, P. A., & Arita, K. (2022). Structural basis for activation of DNMT1. *Nature Communications*, *13*(1), 1–7.

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915.

Kim, E., Magen, A., & Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Research*, *35*(1), 125–131.

Kim, J. T., Li, J., Song, J., Lee, E. Y., Weiss, H. L., Townsend, C. M., & Evers, B. M. (2015). Differential expression and tumorigenic function of neurotensin receptor 1 in neuroendocrine tumor cells. *Oncotarget*, *6*(29), 26960–26970.

Klein, S. J., & O'Neill, R. J. (2018). Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Research*, *26*, 5–23.

Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, *31*(2), 89–97.

Klughammer, J., Romanovskaia, D., Nemc, A., Posautz, A., Seid, C. A., Schuster, L. C., Keinath, M. C., Lugo Ramos, J. S., Kosack, L., Evankow, A., Printz, D., Kirchberger, S., Ergüner, B., Datlinger, P., Fortelny, N., Schmidl, C., Farlik, M., Skjærven, K., Bergthaler, A., … Bock, C. (2023). Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. *Nature Communications*, *14*(232), 1–23.

Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, *502*(7472), 472–479.

Kribelbauer, J. F., Laptenko, O., Chen, S., Martini, G. D., Freed-Pastor, W. A., Prives, C., Mann, R. S., & Bussemaker, H. J. (2017). Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. *Cell Reports*, *19*(11), 2383–2395.

Kronforst, M. R., Gilley, D. C., Strassmann, J. E., & Queller, D. C. (2008). DNA methylation is widespread across social Hymenoptera. *Current Biology*, *18*(7), 287–288.

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571–1572.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., Adli, M., Kasif, S., Ptaszek, L. M., Cowan, C. A., Lander, E. S., Koseki, H., & Bernstein, B. E. (2008). Genome wide analysis of PRC1 and

PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genetics*, *4*(10), e1000242.

Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, *11*(11), 951–969.

Kumar, A., Kankainen, M., Parsons, A., Kallioniemi, O., Mattila, P., & Heckman, C. A. (2017). The impact of RNA sequence library construction protocols on transcriptomic profiling of leukemia. *BMC Genomics*, *18*(629), 1–13.

Kumar, S., & Kim, Y. (2017). An endoparasitoid wasp influences host DNA methylation. *Scientific Reports*, *7*, 43287.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549.

Kurdyukov, S., & Bullock, M. (2016). DNA methylation analysis: choosing the right method. *Biology*, *5*(3), 1–21.

Kurtin, S. E., & Demakos, E. P. (2010). An update on the treatment of myelodysplastic syndromes. *Clinical Journal of Oncology Nursing*, *14*(3).

Kvist, J., Gonçalves Athanàsio, C., Shams Solari, O., Brown, J. B., Colbourne, J. K., Pfrender, M. E., & Mirbahai, L. (2018). Pattern of DNA methylation in Daphnia: evolutionary perspective. *Genome Biology and Evolution*, *10*(8), 1988–2007.

Laget, S., Miotto, B., Chin, H. G., Estève, P. O., Roberts, R. J., Pradhan, S., & Defossez, P. A. (2014). MBD4 cooperates with DNMT1 to mediate methyl-DNA repression and protects mammalian cells from oxidative stress. *Epigenetics*, *9*(4), 546–556.

Lai, A. Y., & Wade, P. A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nature Reviews Cancer*, *11*(8), 588–596.

Laisné, M., Gupta, N., Kirsh, O., Pradhan, S., & Defossez, P. A. (2018). Mechanisms of DNA methyltransferase recruitment in mammals. *Genes*, *9*(617), 1–18.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland,

J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., … Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6846), 860–921.

Larsen, F., Gundersen, G., Lopez, R., & Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics*, *13*(4), 1095–1107.

Le, T. N., Miyazaki, Y., Takuno, S., & Saze, H. (2015). Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Research*, *43*(8), 3911–3921.

Lee, H. J., Hore, T. A., & Reik, W. (2014). Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell*, *14*(6), 710–719.

Lee, M. T., Bonneau, A. R., & Giraldez, A. J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. *Annual Review of Cell and Developmental Biology*, *30*, 581–613.

Leer, T. T., & Karon, M. R. (1976). Inhibition of protein synthesis in 5-azacytidine-treated HeLa cells. *Biochemical Pharmacology*, *25*, 1737–1742.

Lemons, D., & McGinnis, W. (2006). Genomic evolution of Hox gene clusters. *Science*, *313*, 1918–1922.

Lev Maor, G., Yearim, A., & Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends in Genetics*, *31*(5), 274–280.

Levin, H. L., & Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, *12*(9), 615–627.

Lev-Maor, G., Ram, O., Kim, E., Sela, N., Goren, A., Levanon, E. Y., & Ast, G. (2008). Intronic Alus influence alternative splicing. *PLoS Genetics*, *4*(9), 1–12.

Levy, A., Sela, N., & Ast, G. (2008). TranspoGene and microTranspoGene: Transposed elements influence on the transcriptome of seven vertebrates and invertebrates. *Nucleic Acids Research*, *36*(SUPPL. 1), 47–52.

Lewis, S., Ross, L., Bain, S. A., Pahita, E., Smith, S. A., Cordaux, R., Miska, E. M., Lenhard, B., Jiggins, F. M., & Sarkies, P. (2020). Widespread conservation and lineage-specific

diversification of genome-wide DNA methylation patterns across arthropods. *PLoS Genetics*, *16*(6), e1008864.

Li, E., Beard, C., & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *NATURE*, *366*, 362–365.

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, *69*, 915–926.

Li, E., & Zhang, Y. (2014). DNA methylation in mammals. *Cold Spring Harbor Perspectives in Biology*, *6*, a019133.

Li, L. C., & Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics*, *18*(11), 1427–1431.

Li, L., Lu, X., & Dean, J. (2013). The maternal to zygotic transition in mammals. *Molecular Aspects of Medicine*, *34*(5), 919–938.

Liang, G., Gonzales, F. A., Jones, P. A., Orntoft, T. F., & Thykjaer, T. (2002). Analysis of gene induction in human fibroblasts and bladder cancer cells exposed to the methylation inhibitor 5-aza-2′-deoxycytidine. *Cancer Research*, *62*(4), 961–966.

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.

Lim, D. H. K., & Maher, E. R. E. R. (2010). DNA methylation: a form of epigenetic control of gene expression. *The Obstetrician & Gynaecologist*, *12*, 37–42.

Lim, S. P., Neilsen, P., Kumar, R., Abell, A., & Callen, D. F. (2011). The application of delivery systems for DNA methyltransferase inhibitors. *BioDrugs*, *25*(4), 227–242.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322.

Liu, M., & Xu, Y. (2021). An effective method to resolve ambiguous bisulfite-treated reads. *BMC Bioinformatics*, *22*(1), 1–17.

Liu, W. man, Maraia, R. J., Rubin, C. M., Schmid, C. W., man Liu, W., Maraia, R. J., Rubin, C. M., & Schmid, C. W. (1994). Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Research*, *22*(6), 1087–1095.

Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., & Song, C. X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature Biotechnology*, *37*(4), 424–429.

Liubicich, D. M., Serano, J. M., Pavlopoulos, A., Kontarakis, Z., Protas, M. E., Kwan, E., Chatterjee, S., Tran, K. D., Averof, M., & Patel, N. H. (2009). Knockdown of *Parhyale* Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(33), 13892–13896.

Liyanage, V. R. B., Zachariah, R. M., & Rastegar, M. (2013). Decitabine alters the expression of Mecp2 isoforms via dynamic DNA methylation at the Mecp2 regulatory elements in neural stem cells. *Molecular Autism*, *4*(1), 1–21.

Loreto, E. L. S., Melo, E. S. de, Wallau, G. L., & Gomes, T. M. F. F. (2023). The good, the bad and the ugly of transposable elements annotation tools. *Genetics and Molecular Biology*, *46*(3 suppl 1), 1–10.

Lorincz, M. C., Dickerson, D. R., Schmitt, M., & Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature Structural and Molecular Biology*, *11*(11), 1068–1075.

Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, *19*(2), 81–92.

Lyko, F., & Maleszka, R. (2011). Insects as innovative models for functional studies of DNA methylation. *Trends in Genetics*, *27*(4), 127–131.

Mahalingam, R., Ravivarapu, H., Redkar, S., Li, X., & Jasti, B. R. (2007). Transbuccal delivery of 5-aza-2′-deoxycytidine: Effects of drug concentration, buffer solution, and bile salts on permeation. *AAPS PharmSciTech*, *8*(3).

Makałowski, W., Gotea, V., Pande, A., & Makałowska, I. (2019). Transposable Elements: classification, identification, and their use as a tool for comparative genomics. In *Evolutionary Genomics: Statistical and Computational Methods* (Vol. 1910).

Manara, M. C., Valente, S., Cristalli, C., Nicoletti, G., Landuzzi, L., Zwergel, C., Mazzone, R., Stazi, G., Arimondo, P. B., Pasello, M., Guerzoni, C., Picci, P., Nanni, P., Lollini, P. L., Mai, A., & Scotlandi, K. (2018). A quinoline-based DNA methyltransferase inhibitor as a possible adjuvant in osteosarcoma therapy. *Molecular Cancer Therapeutics*, *17*(9), 1881–1892.

Mandrioli, M., & Borsatti, F. (2006). DNA methylation of fly genes and transposons. *Cellular and Molecular Life Sciences*, *63*(17), 1933–1936.

Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C., & Patel, N. H. (2016). CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. *Current Biology*, *26*(1), 14–26.

Masser, D. R., Stanford, D. R., & Freeman, W. M. (2015). Targeted DNA methylation analysis by next-generation sequencing. *Journal of Visualized Experiments*, *96*, e52488.

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, *15*(1), 17–29.

Maunakea, A. K., Chepelev, I., Cui, K., & Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Research*, *23*(11), 1256–1269.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., Le Noble, F., & Rajewsky, N. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, *495*(7441), 333–338.

Menafra, R., & Stunnenberg, H. G. (2014). MBD2 and MBD3: elusive functions and mechanisms. *Frontiers in Genetics*, *5*(428), 1–7.

Mercer, T. R., Dinger, M. E., Mattick, J. S., Rnas, L., Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*, *10*(3), 155–159.

Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F., & Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(2), 716–721.

Messerschmidt, D. M., Knowles, B. B., & Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes and Development*, *28*(8), 812–828.

Momparler, R. L. (2003). Cancer epigenetics. *Oncogene*, *22*(43), 6479–6483.

Momparler, R. L., Bouchard, J., Onetto, N., & Rivard, G. E. (1984). 5-AZA-2′-deoxycytidine therapy in patients with acute leukemia inhibits DNA methylation. *Leukemia Research*, *8*(2), 181–185.

Momparler, R. L., Momparler, L. F., & Samson, J. (1984). Comparison of the antileukemic activity of 5-aza-2′-deoxycytidine, 1-β-d-arabinofuranosylcytosine and 5-azacytidine against L1210 leukemia. *Leukemia Research*, *8*(6), 1043–1049.

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, *38*(1), 23–38.

Morandini, F., Lu, J. Y., Rechsteiner, C., Shadyab, A. H., Casanova, R., Snively, B., Seluanov, A., & Gorbunova, V. (2024). Transposable element methylation state predicts age and disease. *BioRxiv*, 2024.03.15.585206.

Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–562.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*(7643), 72–77.

Nestorov, P., Battke, F., Levesque, M. P., & Gerberding, M. (2013). The maternal transcriptome of the crustacean *Parhyale hawaiensis* is inherited asymmetrically to invariant cell lineages of the ectoderm and mesoderm. *PLoS ONE*, *8*(2), e56049.

Neupane, Y. R., Srivastava, M., Ahmad, N., Kumar, N., Bhatnagar, A., & Kohli, K. (2014). Lipid based nanocarrier system for the potential oral delivery of decitabine: Formulation design,

characterization, *ex vivo,* and *in vivo* assessment. *International Journal of Pharmaceutics*, *477*(1–2), 601–612.

Nguyen, C. T., Weisenberger, D. J., Velicescu, M., Gonzales, F. A., Lin, J. C. Y. Y., Liang, G., & Jones, P. A. (2002). Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2′-deoxycytidine. *Cancer Research*, *62*(22), 6456–6461.

Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V, Branco, M. R., & Reik, W. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biology*, *19*(33), 1–19.

Olson, C. E., & Roberts, S. B. (2014). Genome-wide profiling of DNA methylation and gene expression in *Crassostrea gigas* male gametes. *Frontiers in Physiology*, *5 JUN*(June), 1–7.

O'Neil, D., Glowatz, H., & Schlumpberge, M. (2013). Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Current Protocols in Molecular Biology*, *4.19.1-4.1*(SUPPL.103), 1–8.

Özhan-Kizil, G., Havemann, J., & Gerberding, M. (2009). Germ cells in the crustacean *Parhyale hawaiensis* depend on Vasa protein for their maintenance but not for their formation. *Developmental Biology*, *327*(1), 230–239.

PacBio. (2015). *Detecting DNA base modifications using single molecule, real-time sequencing*. Pacific Bioscience. www.pacb.com/basemods%0Apapers3://publication/uuid/0630A5C0-9BE8-471D-B0EE-34A85DFC498F

Pappalardi, M. B., Keenan, K., Cockerill, M., Kellner, W. A., Stowell, A., Sherk, C., Wong, K., Pathuri, S., Briand, J., Steidel, M., Chapman, P., Groy, A., Wiseman, A. K., McHugh, C. F., Campobasso, N., Graves, A. P., Fairweather, E., Werner, T., Raoof, A., … McCabe, M. T. (2021). Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia. *Nature Cancer*, *2*(10), 1002–1017.

Paranjpe, S. S., Jacobi, U. G., van Heeringen, S. J., & Veenstra, G. J. C. C. (2013). A genome-wide survey of maternal and embryonic transcripts during *Xenopus tropicalis* development. *BMC Genomics*, *14*(762), 1–17.

Paris, M., Wolff, C., Patel, N. H., & Averof, M. (2022). The crustacean model *Parhyale hawaiensis.* In *Current Topics in Developmental Biology* (1st ed., Vol. 147). Elsevier Inc.

Patalano, S., Vlasova, A., Wyatt, C., Ewels, P., Camara, F., Ferreira, P. G., Asher, C. L., Jurkowski, T. P., Segonds-Pichon, A., Bachman, M., González-Navarrete, I., Minoche, A. E., Krueger, F., Lowy, E., Marcet-Houben, M., Rodriguez-Ales, J. L., Nascimento, F. S., Balasubramanian, S., Gabaldon, T., … Sumner, S. (2015). Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(45), 13970–13975.

Patel, K., Dickson, J., Din, S., Macleod, K., Jodrell, D., & Ramsahoye, B. (2010). Targeting of 5-aza-2'-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme. *Nucleic Acids Research*, *38*(13), 4313–4324.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419.

Pauli, A., Valen, E., Lin, M. F., Garber, M., Vastenhouw, N. L., Levin, J. Z., Fan, L., Sandelin, A., Rinn, J. L., Regev, A., & Schier, A. F. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Research*, *22*(3), 577–591.

Pavlopoulos, A., & Averof, M. (2005). Establishing genetic transformation for comparative developmental studies in the crustacean *Parhyale hawaiensis*. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(22), 7888–7893.

Pérez, A., Castellazzi, C. L., Battistini, F., Collinet, K., Flores, O., Deniz, O., Ruiz, M. L., Torrents, D., Eritja, R., Soler-López, M., & Orozco, M. (2012). Impact of methylation on the physical properties of DNA. *Biophysical Journal*, *102*(9), 2140–2148.

Petersen, M., Armisén, D., Gibbs, R. A., Hering, L., Khila, A., Mayer, G., Richards, S., Niehuis, O., & Misof, B. (2021). Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Ecology and Evolution*, *21*(1), 1–15.

Petrova, O. E., Garcia-Alcalde, F., Zampaloni, C., & Sauer, K. (2017). Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. *Scientific Reports*, *7*(January), 1–15.

Pheasant, M., & Mattick, J. S. (2007). Raising the estimate of functional human sequences. *Genome Research*, *17*(9), 1245–1253.

Planques, A., Kerner, P., Ferry, L., Grunau, C., Gazave, E., & Vervoort, M. (2021). DNA methylation atlas and machinery in the developing and regenerating annelid *Platynereis dumerilii*. *BMC Biology*, *19*(148), 1–26.

Platt, R. N., Blanco-Berdugo, L., & Ray, D. A. (2016). Accurate transposable element annotation is vital when analyzing new genome assemblies. *Genome Biology and Evolution*, *8*(2), 403–410.

Pliml J, Sorm F: Synthesis of 2`-deoxy-D-ribofuranosyl-5-azacytosine. *Coll Czech Chem* Commun 1964, 29:2576–2577.

Poovachiranon, S., Boto, K., & Duke, N. (1986). Food preference studies and ingestion rate measurements of the mangrove amphipod *Parhyale hawaiensis* (Dana). *Journal of Experimental Marine Biology and Ecology*, *98*(1–2), 129–140.

Poplineau, M., Schnekenburger, M., Dufer, J., Kosciarz, A., Brassart-Pasco, S., Antonicelli, F., Diederich, M., & Trussardi-Régnier, A. (2015). The DNA hypomethylating agent, 5-aza-2'-deoxycytidine, enhances tumor cell invasion through a transcription-dependent modulation of MMP-1 expression in human fibrosarcoma cells. *Molecular Carcinogenesis*, *54*(1), 24–34.

Portela, A., & Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology*, *28*(10), 1057–1068.

Portella, G., Battistini, F., & Orozco, M. (2013). Understanding the connection between Epigenetic DNA methylation and nucleosome positioning from computer simulations. *PLoS Computational Biology*, *9*(11).

Porter, J., Sun, M. an, Xie, H., & Zhang, L. (2015). Investigating bisulfite short-read mapping failure with hairpin bisulfite sequencing data. *BMC Genomics*, *16*(11), 1–9.

Portilho, N. A., Saini, D., Hossain, I., Sirois, J., Moraes, C., & Pastor, W. A. (2021). The DNMT1 inhibitor GSK-3484862 mediates global demethylation in murine embryonic stem cells. *Epigenetics and Chromatin*, *14*(1), 1–12.

Potok, M. E., Nix, D. A., Parnell, T. J., & Cairns, B. R. (2013). Germline epigenetics, and reprogramming in zebrafish early embryos. *Epigenetics & Chromatin*, *6*(Suppl 1), o23.

Pozo, M. I., Hunt, B. J., Van Kemenade, G., Guerra-Sanz, J. M., Wäckers, F., Mallon, E. B., & Jacquemyn, H. (2021). The effect of DNA methylation on bumblebee colony development. *BMC Genomics*, *22*(73), 1–11.

Prantera, G., & Ferraro, M. (1990). Analysis of methylation and distribution of CpG sequences on human active and inactive X chromosomes by *in situ* nick translation. *Chromosoma*, *99*(1), 18–23.

Price, A. L., & Patel, N. H. (2008). Investigating divergent mechanisms of mesoderm development in arthropods: the expression of Ph-twist and Ph-mef2 in *Parhyale hawaiensis. Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, *310*(B), 24–40.

Provataris, P., Meusemann, K., Niehuis, O., Grath, S., & Misof, B. (2018). Signatures of DNA methylation across insects suggest reduced DNA methylation levels in holometabola. *Genome Biology and Evolution*, *10*(4), 1185–1197.

Putnam, H. M., Davidson, J. M., & Gates, R. D. (2016). Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. *Evolutionary Applications*, *9*(9), 1165–1178.

Qin, T., Jelinek, J., Si, J., Shu, J., & Issa, J. P. J. (2009). Mechanisms of resistance to 5-aza-2'-deoxycytidine in human cancer cell lines. *Blood*, *113*(3), 659–667.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V, Grimaldi, G., Riccio, A., & Trono, D. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Molecular Cell*, *44*(3), 361–372.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.

Quinodoz, S., & Guttman, M. (2014). Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends in Cell Biology*, *24*(11), 651–663.

Ramos, A. P. P. F. (2017). Exploring sensory function and evolution in the crustacean visual system. *Neurobiology*. Université de Lyon,

Rawas, S. E. (2023). *Investigating the regulation of early development of the Amphipod Crustacean Parhyale hawaiensis*. University of Oxford.

Rawłuszko-Wieczorek, A. A., Horst, N., Horbacka, K., Bandura, A. S., Świderska, M., Krokowicz, P., & Jagodziński, P. P. (2015). Effect of DNA methylation profile on OATP3A1 and OATP4A1 transcript levels in colorectal cancer. *Biomedicine and Pharmacotherapy*, *74*, 233–242.

Regev, A., Lamb, M. J., & Jablonka, E. (1998). The role of DNA methylation in invertebrates: Developmental regulation or genome defense? *Molecular Biology and Evolution*, *15*(7), 880–891.

Rehm, E. J., Hannibal, R. L., Chaw, C. R., Vargas-Vila, M. A., & Patel, N. H. (2009a). Injection of *Parhyale hawaiensis* blastomeres with fluorescently labeled tracers. *Cold Spring Harbor Protocols*, *4*(1), 1–6.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A., & Patel, N. H. (2009b). Antibody staining of *Parhyale hawaiensis* embryos. *Cold Spring Harbor Protocols*, *4*(1), 1–5.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A., & Patel, N. H. (2009c). Fixation and dissection of *Parhyale hawaiensis* embryos. *Cold Spring Harbor Protocols*, *4*(1), 1–6.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A., & Patel, N. H. (2009d). *In situ* hybridization of labeled RNA probes to fixed *Parhyale Hawaiensis* embryos. *Cold Spring Harbor Protocols*, *4*(1), 1–6.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, *447*(7143), 425–432.

Renzis, S. De, Elemento, O., Tavazoie, S., & Wieschaus, E. F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. *PLoS Biology*, *5*(5), e117.

Ribas, L., Vanezis, K., Imués, M. A., & Piferrer, F. (2017). Treatment with a DNA methyltransferase inhibitor feminizes zebrafish and induces long-term expression changes in the gonads. *Epigenetics and Chromatin*, *10*(59), 1–16.

Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S., Denell, R. E., Beeman, R. W., Gibbs, R. A., Bucher, G., Friedrich, M., Grimmelikhuijzen, C. J. P. P., Klingler, M., Lorenzen, M., Roth, S., Schröder, R., Tautz, D., Zdobnov, E. M., Muzny, D., Attaway, T., Bell, S., … Grossmann, D. (2008). The genome of the model beetle and pest *Tribolium castaneum. Nature*, *452*(7190), 949–955.

Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenetics and Cell Genetics*, *14*, 9–25.

Rivière, G. (2014). Epigenetic features in the oyster *Crassostrea gigas* suggestive of functionally relevant promoter DNA methylation in invertebrates. *Frontiers in Physiology*, *5*(129), 1–7.

Riviere, G., He, Y., Tecchio, S., Crowell, E., Sourdaine, P., Guo, X., & Favrel, P. (2017). Dynamics of DNA methylomes underlie oyster development. *PLoS Genetics*, *13*(6), e1006807.

Riviere, G., Wu, G. C., Fellous, A., Goux, D., Sourdaine, P., & Favrel, P. (2013). DNA Methylation is crucial for the early development in the oyster *C. gigas*. *Marine Biotechnology*, *15*(6), 739–753.

Roberts, R. J., Carneiro, M. O., & Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome Biology*, *14*(405), 1–4.

Robertson, K. D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F. A., & Jones, P. A. (1999). The human DNA methyltransferases (DNMTs) 1, 3a and 3b: Coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Research*, *27*(11), 2291–2298.

Rošić, S., Amouroux, R., Requena, C. E., Gomes, A., Emperle, M., Beltran, T., Rane, J. K., Linnett, S., Selkirk, M. E., Schiffer, P. H., Bancroft, A. J., Grencis, R. K., Jeltsch, A., Hajkova, P., & Sarkies, P. (2018). Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nature Genetics*, *50*(3), 452–459.

Rubinstein, J. C., Tran, N., Ma, S., Halaban, R., & Krauthammer, M. (2010). Genome-wide methylation and expression profiling identifies promoter characteristics affecting demethylation-induced gene up-regulation in melanoma. *BMC Medical Genomics*, *3*(4), 1–9.

Sadakierska-Chudy, A., Kostrzewa, R. M., & Filip, M. (2015). A comprehensive view of the epigenetic landscape part I: DNA methylation, passive and active DNA demethylation pathways and histone variants. *Neurotoxicity Research*, *27*(1), 84–97.

Saint-Carlier, E., & Riviere, G. (2015). Regulation of Hox orthologues in the oyster *Crassostrea gigas* evidences a functional role for promoter DNA methylation in an invertebrate. *FEBS Letters*, *589*(13), 1459–1466.

Sandler, J. E., & Stathopoulos, A. (2016). Quantitative single-embryo profile of *Drosophila* genome activation and the dorsal–ventral patterning network. *Genetics*, *202*(4), 1575–1584.

Sarda, S., Zeng, J., Hunt, B. G., & Yi, S. V. (2012). The evolution of invertebrate gene body methylation. *Molecular Biology and Evolution*, *29*(8), 1907–1916.

Sasai, N., Nakao, M., & Defossez, P. A. (2010). Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Research*, *38*(15), 5015–5022.

Sasaki, A., & Satoh, N. (2007). Effects of 5-aza-2′-deoxycytidine on the gene expression profile during embryogenesis of the ascidian *Ciona intestinalis*: A microarray analysis. *Zoological Science*, *24*(7), 648–655.

Sato, T., Issa, J. P. J., & Kropf, P. (2017). DNA hypomethylating drugs in cancer therapy. *Cold Spring Harbor Perspectives in Medicine*, *7*(5), 1–14.

Sayar, N., Karahan, G., Konu, O., Bozkurt, B., Bozdogan, O., & Yulug, I. G. (2015). Transgelin gene is frequently downregulated by promoter DNA hypermethylation in breast cancer. *Clinical Epigenetics*, *7*(104), 1–16.

Schaefer, M., Hagemann, S., Hanna, K., & Lyko, F. (2009). Azacytidine inhibits RNA methylation at DNMT2 target sites in human cancer cell lines. *Cancer Research*, *69*(20), 8127–8132.

Schermelleh, L., Spada, F., Easwaran, H. P., Zolghadr, K., Margot, J. B., Cardoso, M. C., & Leonhardt, H. (2005). Trapped in action: Direct visualization of DNA methyltransferase activity in living cells. *Nature Methods*, *2*(10), 751–756.

Schmitz, R. J., Lewis, Z. A., & Goll, M. G. (2019). DNA methylation: shared and divergent features across eukaryotes. *Trends in Genetics*, *35*(11), 818–827.

Schoeftner, S., & Blasco, M. A. (2008). Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nature Cell Biology*, *10*(2), 228–236.

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, *517*(7534), 321–326.

Schübeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., Van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., Bell, S. P., & Groudine, M. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes and Development*, *18*(11), 1263–1271.

Schuierer, S., Carbone, W., Knehr, J., Petitjean, V., Fernandez, A., Sultan, M., & Roma, G. (2017). A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. *BMC Genomics*, *18*(442), 1–13.

Schultz, M. D., He, Y., Whitaker, J. W., Hariharan, M., Mukamel, E. A., Leung, D., Rajagopal, N., Nery, J. R., Urich, M. A., Chen, H., Lin, S., Lin, Y., Jung, I., Schmitt, A. D., Selvaraj, S., Ren, B., Sejnowski, T. J., Wang, W., & Ecker, J. R. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*, *523*(7559), 212–216.

Schultz, R. M. (2002). The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Human Reproduction Update*, *8*(4), 323–331.

Schulz, N. K. E., Wagner, C. I., Ebeling, J., Raddatz, G., Diddens-de Buhr, M. F., Lyko, F., & Kurtz, J. (2018). Dnmt1 has an essential function despite the absence of CpG DNA methylation in the red flour beetle *Tribolium castaneum. Scientific Reports*, *8*(16462), 1–10.

Schutsky, E. K., Denizio, J. E., Hu, P., Liu, M. Y., Nabel, C. S., Fabyanic, E. B., Hwang, Y., Bushman, F. D., Wu, H., & Kohli, R. M. (2018). Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nature Biotechnology*, *36*(11), 1083–1090.

Seelan, R. S., Mukhopadhyay, P., Pisano, M. M., & Greene, R. M. (2018). Effects of 5-Aza-2′-deoxycytidine (decitabine) on gene expression. *Drug Metabolism Reviews*, *50*(2), 193–207.

Sela, N., Kim, E., & Ast, G. (2010). The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biology*, *11*(6).

Serano, J. M., Martin, A., Liubicich, D. M., Jarvis, E., Bruce, H. S., La, K., Browne, W. E., Grimwood, J., & Patel, N. H. (2016). Comprehensive analysis of Hox gene expression in the amphipod crustacean *Parhyale hawaiensis. Developmental Biology*, *409*(1), 297–309.

Sharma, S., Kelly, T. K., & Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis*, *31*(1), 27–36.

Sharp, A. J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S. B., Dupre, Y., & Antonarakis, S. E. (2011). DNA methylation profiles of human active and inactive X chromosomes. *Genome Research*, *21*(10), 1592–1600.

Shenker, N. S., Flower, K. J., Wilhelm-Benartzi, C. S., Dai, W., Bell, E., Gore, E., Bahrawy, M., Weaver, G., Brown, R., & Flanagan, J. M. (2015). Transcriptional implications of intragenic DNA methylation in the oestrogen receptor alpha gene in breast cancer cells and tissues. *BMC Cancer*, *15*(337), 1–12.

Shimoda, N., Yamakoshi, K., Miyake, A., & Takeda, H. (2005). Identification of a gene required for *de novo* DNA methylation of the zebrafish no tail gene. *Developmental Dynamics*, *233*(4), 1509–1516.

Shoemaker, C. R. (1956). Observations on the amphipod genus *Parhyale*. *Proceedings of the United States National Museum*, *106*(3372), 345–358.

Siegfried, Z., & Simon, I. (2010). DNA methylation and gene expression. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *2*(3), 362–371.

Silliman, K., Spencer, L. H., White, S. J., & Roberts, S. B. (2023). Epigenetic and genetic population structure is coupled in a marine invertebrate. *Genome Biology and Evolution*, *15*(2), 1–18.

Silverman, L. R., Demakos, E. P., Peterson, B. L., Kornblith, A. B., Holland, J. C., Odchimar-Reissig, R., Stone, R. M., Nelson, D., Powell, B. L., DeCastro, C. M., Ellerton, J., Larson,

R. A., Schiffer, C. A., & Holland, J. F. (2002). Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: A study of the cancer and leukemia group B. *Journal of Clinical Oncology*, *20*(10), 2429–2440.

Singer, M., Kosti, I., Pachter, L., & Mandel-Gutfreund, Y. (2015). A diverse epigenetic landscape at human exons with implication for expression. *Nucleic Acids Research*, *43*(7), 3498–3508.

Smit, A. F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Current Opinion in Genetics and Development*, *9*(6), 657–663.

Smit A, Hubley R and Green P (2013) RepeatMasker Open-4.0., http://www.repeatmasker.org/

Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204–220.

Soengas, M. S., Capodieci, P., Polsky, D., Mora, J., Esteller, M., Opitz-Araya, X., McCombie, R., Herman, J. G., Gerald, W. L., Lazebnik, Y. A., Cordón-Cardó, C., & Lowe, S. W. (2001). Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature*, *409*(6817), 207–211.

Solary, E., Bernard, O. A., Tefferi, A., Fuks, F., & Vainchenker, W. (2014). The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia*, *28*(3), 485–496.

Song, F., Smith, J. F., Kimura, M. T., Morrow, A. D., Matsuyama, T., Nagase, H., & Held, W. A. (2005). Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(9), 3336–3341.

Sorm F, Veselý J. Effect of 5-aza-2'-deoxycytidine against leukemic and hemopoietic tissues in AKR mice. *Neoplasma*. 1968;15(4):339-43.

Sotero-Caio, C. G., Platt, R. N., Suh, A., & Ray, D. A. (2017). Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biology and Evolution*, *9*(1), 161–177.

Speidel, D. (2015). The role of DNA damage responses in p53 biology. *Archives of Toxicology*, *89*(4), 501–517.

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schübeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, *480*(7378), 490–495.

Standage, D. S., Berens, A. J., Glastad, K. M., Severin, A. J., Brendel, V. P., & Toth, A. L. (2016). Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. *Molecular Ecology*, *25*(8), 1769–1784.

Stewart, K. R., Veselovska, L., & Kelsey, G. (2016). Establishment and functions of DNA methylation in the germline. *Epigenomics*, *8*(10), 1399–1413.

Stoeckius, M., Grün, D., Kirchner, M., Ayoub, S., Torti, F., Piano, F., Herzog, M., Selbach, M., & Rajewsky, N. (2014). Global characterization of the oocyte-to-embryo transition in *C aenorhabditis elegans* uncovers a novel m RNA clearance mechanism. *The EMBO Journal*, *33*(16), 1751–1766.

Stomper, J., Rotondo, J. C., Greve, G., & Lübbert, M. (2021). Hypomethylating agents (HMA) for the treatment of acute myeloid leukemia and myelodysplastic syndromes: mechanisms of resistance and novel HMA-based therapies. *Leukemia*, *35*(7), 1873–1889.

Stresemann, C., & Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *International Journal of Cancer*, *123*(1), 8–13.

Suarez-Ulloa, V., Rivera-Casas, C., Michel, M., & Eirin-Lopez, J. M. (2019). Seasonal DNA methylation variation in the flat tree oyster *Isognomon alatus* from a mangrove ecosystem in north biscayne bay, Florida. *Journal of Shellfish Research*, *38*(1), 79–88.

Sun, D. A., & Patel, N. H. (2019). The amphipod crustacean *Parhyale hawaiensis*: An emerging comparative model of arthropod development, evolution, and regeneration. *Wiley Interdisciplinary Reviews: Developmental Biology*, *8*(5), e355.

Sun, Z., Vaisvila, R., Hussong, L. M., Yan, B., Baum, C., Saleh, L., Samaranayake, M., Guan, S., Dai, N., Corrêa, I. R., Pradhan, S., Davis, T. B., Evans, T. C., & Ettwiller, L. M. (2021). Nondestructive enzymatic deamination enables single-molecule long-read amplicon

sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Genome Research*, *31*, 291–300.

Sundaram, V., & Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1795).

Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, *9*(6), 465–476.

Suzuki, M. M., Kerr, A. R. W. W., De Sousa, D., & Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Research*, *17*(5), 625–631.

Taby, R., & Issa, J.-P. J. (2010). Cancer epigenetics. *A Cancer Journal for Clinicians*, *60*, 376–392.

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development*, *136*(18), 3033–3042.

Taghbalout, A., Du, M., Jillette, N., Rosikiewicz, W., Rath, A., Heinen, C. D., Li, S., & Cheng, A. W. (2019). Enhanced CRISPR-based DNA demethylation by Casilio-ME-mediated RNA-guided coupling of methylcytosine oxidation and DNA repair pathways. *Nature Communications*, *10*(1), 1–12.

Tamaru, H., & Selker, E. U. (2003). Synthesis of signals for *de novo* DNA methylation in *Neurospora crassa*. *Molecular and Cellular Biology*, *23*(7), 2379–2394.

Tan, M. H., Au, K. F., Yablonovitch, A. L., Wills, A. E., Chuang, J., Baker, J. C., Wong, W. H., & Li, J. B. (2013). RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. *Genome Research*, *23*(1), 201–216.

Tang, D., Zheng, S., Zheng, Z., Liu, C., Zhang, J., Yan, R., Wu, C., Zuo, N., Wu, L., Xu, H., Liu, S., & He, Y. (2022). Dnmt1 is required for the development of auditory organs via cell cycle arrest and Fgf signalling. *Cell Proliferation*, *55*(5), 1–15.

Taylor, S. M., & Jones, P. A. (1979). Multiple new phenotypes induced in 10T 1/2 and 3T3 cells treated with 5-azacytidine. *Cell*, *17*(4), 771–779.

Thomsen, S., Anders, S., Janga, S. C., Huber, W., & Alonso, C. R. (2010). Genome-wide analysis of mRNA decay patterns during early Drosophila development. *Genome Biology*, *11*(R:93), 1–27.

Tran, H., Porter, J., Sun, M. A., Xie, H., & Zhang, L. (2014). Objective and comprehensive evaluation of bisulfite short read mapping tools. *Advances in Bioinformatics*, *2014*.

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, *13*(1), 36–46.

Treiber, C. D., & Waddell, S. (2020). Transposon expression in the *Drosophila* brain is driven by neighboring genes and diversifies the neural transcriptome. *Genome Research*, *30*(11), 1559–1569.

Trigg, S. A., Venkataraman, Y. R., Gavery, M. R., Roberts, S. B., Bhattacharya, D., Downey-Wall, A., Eirin-Lopez, J. M., Johnson, K. M., Lotterhos, K. E., Puritz, J. B., & Putnam, H. M. (2022). Invertebrate methylomes provide insight into mechanisms of environmental tolerance and reveal methodological biases. *Molecular Ecology Resources*, *22*(4), 1247–1261.

Tse, O. Y. O., Jiang, P., Cheng, S. H., Peng, W., Shang, H., Wong, J., Chan, S. L., Poon, L. C. Y., Leung, T. Y., Chan, K. C. A., Chiu, R. W. K., & Lo, Y. M. D. (2021). Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(5), e2019768118.

Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell*, *154*(1), 26–46.

Vaisvila, R., Ponnaluri, V. K. C., Sun, Z., Langhorst, B. W., Saleh, L., Guan, S., Dai, N., Campbell, M. A., Sexton, B. S., Marks, K., Samaranayake, M., Samuelson, J. C., Church, H. E., Tamanaha, E., Corrêa, I. R., Pradhan, S., Dimalanta, E. T., Evans, T. C., Williams, L., & Davis, T. B. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research*, *31*(7), 1280–1289.

Vandegehuchte, M. B., Lemière, F., & Janssen, C. R. (2009). Quantitative DNA-methylation in *Daphnia magna* and effects of multigeneration Zn exposure. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*, *150*(3), 343–348.

Vargas-Vila, M. A., Hannibal, R. L., Parchem, R. J., Liu, P. Z., & Patel, N. H. (2010). A prominent requirement for single-minded and the ventral midline in patterning the dorsoventral axis of the crustacean *Parhyale hawaiensis. Development*, *137*(20), 3469–3476.

Vastenhouw, N. L., Cao, W. X., & Lipshitz, H. D. (2019). The maternal-to-zygotic transition revisited. *Development*, *146*(11), 1–20.

Veenstra, G. J. C., & Wolffe, A. P. (2001). Constitutive genomic methylation during embryonic development of *Xenopus*. *Biochimica et Biophysica Acta - Gene Structure and Expression*, *1521*(1–3), 39–44.

Venkataraman, Y. R., Downey-Wall, A. M., Ries, J., Westfield, I., White, S. J., Roberts, S. B., & Lotterhos, K. E. (2020). General DNA methylation patterns and environmentally-induced differential methylation in the eastern oyster (*Crassostrea virginica*). *Frontiers in Marine Science*, *7*(April), 1–14.

Vento´s Alfonso, A., Ylla, G., Montan˜es, J.-C., Belles, X., Ventós-Alfonso, A., Ylla, G., Montañes, J. C., & Belles, X. (2020). DNMT1 promotes genome methylation and early embryo development in cockroaches. *IScience*, *23*(12), 101778.

Viegas-Pequignot, E., Dutrillaux, B., & Thomas, G. (1988). Inactive X chromosome has the highest concentration of unmethylated Hha I sites. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(20), 7657–7660.

Wagner, D. S., Dosch, R., Mintzer, K. A., Wiemelt, A. P., & Mullins, M. C. (2004). Maternal control of development at the midblastula transition and beyond: mutants from the zebrafish II. *Developmental Cell*, *6*(6), 781–790.

Walsh, C. P., Chaillet, R., & Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics*, *20*, 116–117.

Wang, D., Su, Y., Wang, X., Lei, H., & Yu, J. (2012). Transposon-derived and satellite-derived repetitive sequences play distinct functional roles in mammalian intron size expansion. *Evolutionary Bioinformatics*, *2012*(8), 301–319.

Wang, N., Hu, Y., & Wang, Z. (2024). Regulation of alternative splicing: Functional interplay with epigenetic modifications and its implication to cancer. *Wiley Interdisciplinary Reviews: RNA*, *15*(1), 1–18.

Wang, Q. T., Piotrowska, K., Ciemerych, M. A., Milenkovic, L., Scott, M. P., Davis, R. W., & Zernicka-Goetz, M. (2004). A Genome-Wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Developmental Cell*, *6*, 133–144.

Wang, X., Li, Q., Lian, J., Li, L., Jin, L., Cai, H., Xu, F., Qi, H., Zhang, L., Wu, F., Meng, J., Que, H., Fang, X., Guo, X., & Zhang, G. (2014). Genome-wide and single-base resolution DNA methylomes of the Pacific oyster *Crassostrea gigas* provide insight into the evolution of invertebrate CpG methylation. *BMC Genomics*, *15*(1119), 1–11.

Wang, X., Wheeler, D., Avery, A., Rago, A., Choi, J. H., Colbourne, J. K., Clark, A. G., & Werren, J. H. (2013). Function and evolution of DNA methylation in Nasonia vitripennis. *PLoS Genetics*, *9*(10), e1003872.

Wang, Y., Jorda, M., Jones, P. L., Maleszka, R., Ling, X., Robertson, H. M., Mizzen, C. A., Peinado, M. A., Robinson, G. E., Miguel A. Peinado, Robinson, G. E., & Gene E. Robinson. (2006). Functional CpG methylation system in a social insect. *Science*, *314*(5799), 645–647.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*, 57–63.

Weaver, J. R., Sarkisian, G., Krapp, C., Mager, J., Mann, M. R. W., & Bartolomei, M. S. (2010). Domain-specific response of imprinted genes to reduced DNMT1. *Molecular and Cellular Biology*, *30*(16), 3916–3928.

White, R. J., Collins, J. E., Sealy, I. M., Wali, N., Dooley, C. M., Digby, Z., Stemple, D. L., Murphy, D. N., Billis, K., Hourlier, T., Füllgrabe, A., Davis, M. P., Enright, A. J., & Busch-Nentwich, E. M. (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. *ELife*, *6*, e30860.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982.

Wilhelm, B. T., & Landry, J. R. (2009). RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, *48*(3), 249–257.

Williams, L., Bei, Y., Church, H. E., Dai, N., Dimalanta, E. T., Ettwiller, L. M., Evans, T. C., Langhorst, B. W., Borgaro, J. G., Guan, S., Marks, K., Menin, J. F., Nichols, N. M., Chaithanya Ponnaluri, V. K., Saleh, L., Samaranayake, M., Sexton, B. S., Sun, Z., Tamanaha, E., … Davis, T. B. (n.d.). Enzymatic methyl-seq: the next generation of methylome analysis. In *New England Biolabs*.

Wittfoth, C., Harzsch, S., Wolff, C., & Sombke, A. (2019). The 'amphi'-brains of amphipods: New insights from the neuroanatomy of *Parhyale hawaiensis* (Dana, 1853). *Frontiers in Zoology*, *16*(30), 1–19.

Wolffe, A. P., & Guschin, D. (2000). Review: Chromatin structural features and targets that regulate transcription. *Journal of Structural Biology*, *129*, 102–122.

Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P., & Warren, W. D. (1997). Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *Journal of Biological Chemistry*, *272*(12), 7810–7816.

Wu, X., & Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics*, *18*(9), 517–534.

Xiang, H., Li, X., Dai, F., Xu, X., Tan, A., Chen, L., Zhang, G., Ding, Y., Li, Q., Lian, J., Willden, A., Guo, Q., Xia, Q., Wang, J., & Wang, W. (2013). Comparative methylomics between domesticated and wild silkworms implies possible epigenetic influences on silkworm domestication. *BMC Genomics*, *14*(646), 1–11.

Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., Zhang, H., Zhang, G., Li, D., Dong, Y., Zhao, L., Lin, Y., Cheng, D., Yu, J., Sun, J., Zhou, X., Ma, K., He, Y., Zhao, Y., … Wang, J. (2010). Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology*, *28*(5), 516–520.

Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K., & Li, E. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, *236*(1), 87–95.

Xu, G., Zhang, J., Lyu, H., Song, Q., Feng, Q., Xiang, H., & Zheng, S. (2018). DNA methylation mediates BmDeaf1-regulated tissue- and stage-specific expression of BmCHSA-2b in the silkworm, *Bombyx mori*. *Epigenetics and Chromatin*, *11*(32), 1–14.

Xu, X., Li, G., Li, C., Zhang, J., Wang, Q. F., Simmons, D. K., Chen, X., Wijesena, N., Zhu, W., Wang, Z. Z. Z., Wang, Z. Z. Z., Ju, B., Ci, W., Lu, X., Yu, D., Wang, Q. F., Aluru, N., Oliveri, P., Zhang, Y. E., … Liu, J. (2019). Evolutionary transition between invertebrates and vertebrates via methylation reprogramming in embryogenesis. *National Science Review*, *6*(5), 993–1003.

Xu, Z., & Wang, H. (2007). LTR-FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, *35*(SUPPL.2), 265–268.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E., Liu, J. Y., Horvath, S., & Fan, G. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, *500*(7464), 593–597.

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., & Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature*, *504*(7480), 460–464.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., Huang, J., Li, M., Wu, X., Wen, L., Lao, K., Li, R., Qiao, J., & Tang, F. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Structural and Molecular Biology*, *20*(9), 1131–1139.

Yang, X., Han, H., DeCarvalho, D. D., Lay, F. D., Jones, P. A., & Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell*, *26*(4), 577–590.

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J. P., Nissim-Rafinia, M., Cohen, A. H. S., Rippe, K., Meshorer, E., & Ast, G. (2015). HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Reports*, *10*(7), 1122–1134.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(502), eaaj2239.

Ying, H., Hayward, D. C., Klimovich, A., Bosch, T. C. G., Baldassarre, L., Neeman, T., Forêt, S., Huttley, G., Reitzel, A. M., Fraune, S., Ball, E. E., & Miller, D. J. (2022). The role of DNA methylation in genome defense in cnidaria and other invertebrates. *Molecular Biology and Evolution*, *39*(2), 1–14.

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, *13*(8), 335–340.

Yokochi, T., & Robertson, K. D. (2002). Preferential methylation of unmethylated DNA by mammalian *de novo* DNA methyltransferase Dnmt3a. *Journal of Biological Chemistry*, *277*(14), 11735–11745.

Yong, W. S., Hsu, F. M., & Chen, P. Y. (2016). Profiling genome-wide DNA methylation. *Epigenetics and Chromatin*, *9*(26), 1–16.

Zagris, N., & Podimatas, T. (1994). 5-Azacytidine changes gene expression and causes developmental arrest of early chick embryo. *International Journal of Developmental Biology*, *38*(4), 741–744.

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, *328*(5980), 916–919.

Zemach, A., & Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Current Biology*, *20*(17), R780--R785.

Zeng, J., Konopka, G., Hunt, B. G., Preuss, T. M., Geschwind, D., & Yi, S. V. (2012). Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *American Journal of Human Genetics*, *91*(3), 455–465.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P. W. H., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., … Wang, J. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, *490*(7418), 49–54.

Zhang, X., Li, H., Qiu, Q., Qi, Y., Huang, D., & Zhang, Y. (2014). 2,4-Dichlorophenol induces global DNA hypermethylation through the increase of S-adenosylmethionine and the upregulation of DNMTs mRNA in the liver of goldfish *Carassius auratus. Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*, *160*(1), 54–59.

Zhang, Y. W., Staal, B., Dykema, K. J., Furge, K. A., & Vande Woude, G. F. (2012). Cancer-type regulation of MIG-6 expression by inhibitors of methylation and histone deacetylation. *PLoS ONE*, *7*(6), e38955.

Zhang, Z., Liu, G., Zhou, Z., Su, Z., & Gu, X. (2024). Global level of methylation in the sea lamprey (jawless vertebrate) genome is intermediate between invertebrate and jawed vertebrate genomes. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 1–7.

Zhao, S., Zhang, Y., Gamini, R., Zhang, B., & Von Schack, D. (2018). Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Scientific Reports*, *8*(4781), 1–12.

Zheng, Z., Li, L., Liu, X., Wang, D., Tu, B., Wang, L., Wang, H., & Zhu, W. (2012). 5-Aza-2'-deoxycytidine reactivates gene expression via degradation of pRb pocket proteins. *The FASEB Journal*, *26*(1), 449–459.

Zhivotosky, B., & Orrenius, S. (2001). Assessment of apoptosis and necrosis by DNA fragmentation and morphological criteria. *Current Protocols in Cell Biology / Editorial Board, Juan S. Bonifacino ... [et Al.]*, *Chapter 18*, 18.3.1--18.3.23.

Zhou, W., Liang, G., Molloy, P. L., & Jones, P. A. (2020). DNA methylation enables transposable element-driven genome expansion. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(32), 19359–19366.

Zhu, W. G., Hileman, T., Ke, Y., Wang, P., Lu, S., Duan, W., Dai, Z., Tong, T., Villalona-Calero, M. A., Plass, C., & Otterson, G. A. (2004). 5-Aza-2′-deoxycytidine activates the p53/p21 Waf1/Cip1 pathway to inhibit cell proliferation. *Journal of Biological Chemistry*, *279*(15), 15161–15166.

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., & Henikoff, S. (2006). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, *39*(1), 61–69.

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T. Y. Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., Gnirke, A., & Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, *500*(7463), 477–481.

Ziller, M. J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C. B., Bernstein, B. E., Lengauer, T., Gnirke, A., & Meissner, A. (2011). Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genetics*, *7*(12), e1002389.