Ollscoil na Gaillimhe
University of Galway

Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	DNA methylation and zygotic genome activation during early embryogenesis of animals
Author(s)	Febrimarsa, Febrimarsa
Publication Date	2020-10-25
Publisher	NUI Galway
Item record	http://hdl.handle.net/10379/16238

Downloaded 2024-04-27T00:18:35Z

Some rights reserved. For more information, please see the item record link above.

DNA METHYLATION AND ZYGOTIC GENOME ACTIVATION DURING EARLY EMBRYOGENESIS OF ANIMALS

Author:

Febrimarsa Febrimarsa

Supervisor:

Prof. Uri Frank

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at

DISCIPLINE OF BIOCHEMISTRY
CENTRE FOR CHROMOSOME BIOLOGY & SCHOOL OF NATURAL SCIENCES
NATIONAL UNIVERSITY OF IRELAND
GALWAY

OCTOBER 2020

Content

List of Figures

List of Tables

Glossary of Terms and Abbreviation

5mC	5-methyldeoxycytidine, a methylated form of deoxycytidine (dC) in DNA
6mA	N6-methyldeoxyadenosine, a methylated form of deoxyadenosine (dA) in DNA
ADD	A domain in DNMT3 that interact with H3K4
ALKBH	Oxidoreductase belongs to AlkB family that found to initiate demethylation of 6mA in bacteria.
ApT (or AT)	ApT is an abbreviation for adenosine and thymine separated by a phosphate, which links the two nucleotides together in DNA.
CpG (or CG)	CpG is an abbreviation for cytosine and guanine separated by a phosphate, which links the two nucleotides together in DNA.
CuAAC	Cu(I)-catalysed alkyne-azide chemistry
DE	Differentially express
DMAP1	A protein that interact with N-terminal domain of DNMT1 and facilitate the DNMT1 role in double strand break repair
DNMT	DNA (cytosine-5)-methyltransferase, enzymes that add methyl to deoxycytidine (dC).
dNTP	Monomer of DNA
Dpf	Days post fertilization
EdU	5-Ethynyl deoxyuridine, an analogue of deoxythymidine that can get incorporated into DNA during replication
EU	5-Ethynyl uridine, an analogue of uracil that can get incorporated into RNA during transcription
FDR	False discovery rate
fp	Gene encoding fluorescence protein
FP	Fluorescence protein
GFP	Green fluorescence protein
goi	Gene of interest
hpf	Hours post fertilization
HPLC-MS/MS	High performance liquid chromatography-mass spectrometry/mass spectrometry.
HU	Hydroxy urea
IF	Immunofluorescence

IPD	Interpulse duration, an analysis processing the data from SMRTseq output to generate the data of base modification map.
IVT	In vitro transcription
m6A	N6-methyl adenosine, a methylated form of adenosine (A) in RNA
METTL4	6mA methyltransferase 4, a candidate methyltransferase for depositing 6mA in DNA.
miRNA	microRNA
MSA	Multiple sequence alignment
MTase	Methyltransferase
MZT	Maternal to zygotic transition
N6AMT1	N6-deoxyAdenosine methyltransferase 1, a 6mA methyltransferase candidate
NLS	Nuclear localization signal
nt	Nucleotide
NTP	Monomer of RNA
PAGA-T	Polyethylene Glycol-Ethanol-Glycerol-Acetic Acid in Tris-HCl buffer.
PFA	Paraformaldehyde
PFA+GA	Paraformaldehyde and glutaraldehyde
PRC1	Polycomb repression complex 1, a protein complex that that recognize H3K27me3 and ubiquitinate H2AK119 to mediate repression.
PRC2	Polycomb repression complex 2, a protein complex that methylate H3K27 to mediate repression of genes.
PWWP	A domain in N-terminal of DNMT3 that interact with H3K36me3
QQQ	Triple Quadrupole, an MS/MS machine
RFD	A domain in N-terminal of DNMT1 protein that are essential for their role as CpG methylation maintainer post replication.
RNA pol II	RNA polymerase II, a polymerase enzyme that responsible for the transcription of most genes.
RNR	Ribonucleotide reductase, enzyme that convert NTP to dNTP
ROI	Region of interest
shRNA	Short hairpin RNA, a short (~50 nt) but with hairpin secondary structure which recognized by the RNA interference machineries and targeting the complementary mRNA for cutting thus silencing the transcripts.
SIDD	Stress-induced DNA double helix destabilization

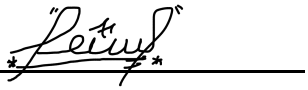
SMRTseq	Single molecule real time sequencing, a long-read sequencing technology that allow for detection of base modification such 6mA
TET	Ten-eleven translocation methyl-deoxycytidine dioxygenase, the enzyme that initiate demethylation of 5mC by adding hydroxyl onto the methyl of 5mC, thus converting it into 5hydroxy-mC (5hmC)
TF	Transcription factor, a protein that regulate transcription through their interaction with DNA.
UHRF1	Ubiquitin-like containing PHD and RING finger domains 1, a protein that interact with H3K9me2/3 and hemi methylated CpG near the replication fork and interact with DNMT1.
ZGA	Zygotic genome activation, an event where the for the first-time transcription occurred in embryos using the newly formed zygotic genome

Declaration

I, Febrimarsa Febrimarsa, declare that:

- this thesis has been composed solely by me under supervision of Prof. Uri Frank.
- the experimental work presented in this thesis is almost entirely my own work, while the collaborative contributions have been indicated clearly and acknowledged.
- the appropriate credit has been given within this thesis where reference has been made to the work of others and
- this thesis has not been submitted, in whole or in part, to any university for any other degree.

In conducting this research, I was funded by Hardiman Research Scholarship and Thomas Crawford Hayes Research Grant. Meanwhile, my supervisor Prof. Uri Frank was funded by the Wellcome Trust, National Science Foundation (NSF, USA), and Science Foundation Ireland (SFI).

Signed : 

Febrimarsa Febrimarsa, October 2020

Preface

Alhamdulillah, my sincere gratitude to Prof. Uri Frank who allowed me to conduct this research in his lab. His supervision, support, and guidance has been a tremendous factor for this thesis to come to existence. To Dr. Sebastian Gornik, who by his invitation and support has allowed me to be in contact with Prof. Uri Frank in the first instance, which then support me by his starting investigator research grant (from Science Foundation Ireland), I am thankful.

People in the lab, with whom I intensely discussed sciences, I debated fiercely in our lab meeting, I accidentally spill or touch your experiments, I give the “ugly” secret Santa gift, I enjoy coffee on the couches in between pipetting, and for all the laughs and sadness that we share, you all are a factor that kept me (and perhaps all of us) in sanity. Brian, Sebastian, Hakima, James, Liam, Emma, Anna, Ruth, Kate, Eleni, Miguel, Aine, Tim, Martin, Indu, Helen, Patricia to all of you (and to some extent Bertrand, Suad, and Mo), I do sincerely wish for you all the best in the future. I do hope we'll stay in contact in the next few decades and perhaps, share a moment of coffee again. Specifically, I have to mention my sense of gratitude to Dr. Timothy Dubuc who included my shRNA work in his *Science* article and to Dr. Anna Torok who included me in her latest publication on *Hydractinia* histone.

The collaborators from the USA have become a crucial part of this research. RNA-seq and SMRT-seq (PacBio) pipelines and analysis would hardly be done without such collaboration. Dr. Christine Schnitzler, Dr Paul Gonzalez, Dr. Sofia Barreira, and Prof. Andy Baxevanis are the most influential on this line of work, thus I express my thankfulness.

Along the way, I also spent time discussing the idea and direction of the project with Dr. Markus Mueller (TU Muenchen), Prof. Chuan He (Univ. Chicago), Dr. Chema Martin-Duran (QMUL), Dr. Paulyn Cartwright (Univ. Kansas) and Dr. Maxim Greenberg (IJM Paris), thus I would say thank you. The more casual science conversation with Steven Sanders, Miguel, Gonzalo, Taylor Skokan and Shounan He during cnidofest, which I enjoyed much thus thankful.

This thesis was part of 5-year long effort with inputs and guidance from GRC committees, Dr. Elaine Dunleavy, Dr. Gerhard Schlosser, Dr. Derrek Moris, Prof. Noel Lowndes, Prof. Heinz Peter, Dr. Eva Szengendi, Prof. Ciaran Morrison, Dr. Stepehen O’Rea, and Dr. Michael McCarthy, which finalized by the examiner committee Prof. Bob Lahue, Prof. Kevin Sullivan, and Prof. Yehu Moran. To all of them, I would express my gratefulness.

I would also like to add Mr. Brendan Harhen for his support of the Mass-Spectrometry work, Prof. Ciaran Morrison who kindly allowing me to use his UV-illuminatore for dot-blot experiment, Dr. Andrew Flaus who kindly allowing me to use his freeze-dryer, Dr. Peter Owens and Dr. Luis Gutierrez for their swift support on daily technical issue of confocal microscopy. Moreover, to PhD students, principal investigators and postdoctoral fellows at Centre for Chromosome Biology (CCB) who attend and gave CCB seminars thus creating an environment for me to expand my horizons in science of chromosome biology.

I would like to also extend my gratitude to my big family in Indonesia and Galway's friends outside the lab from Indonesian circles (Nurul, Yenni, Arya, Ady, Mira, Andi, Joula, Mega, Henri and Astri), Islamic Society circles (Ahmed Wanas, Alameen, Ahmad Suboor, Sameh, Hewa, Ammar, Mustofa, Mohanned, Moez, Amir, AbdurRazaq, Abu Bakr, Yusuf, Omar, Fauzah, Sabreena, Hira, Zunaira, Ilyani, Ala, Naemah, Hasnan and Gibran), and Literary and Debating Society circles (Chris, Sean, Kate, Eoghan, Michael, and Paschal) you guys all keep checking me and encouraging me to be a better person during this PhD.

Then, and perhaps the most important one, I would like to express my deepest gratitude to my wife, Venny Fristy Anggraini, who without her support in handling our two beloved babies (Feargal and Miriam) during my PhD, on which I may often in absentia, I will never be able to complete this PhD. Thank you, my Love.

Finally, to my parent (Martini and Abdul Sari), whom without them I would not even exist. Their encouragement, guidance and inspiration are fundamental to my behaviour and decision on every step of the ladder so I can reach this point of my life and to any future milestone that I will reach. There are no words that can describe my gratitude.

ABSTRACT

Upon fertilization, after the fusion of two pronuclei, the newly formed zygotic genome is transcriptionally inactive. Hence, early animal development is driven by maternal RNAs and proteins. The zygotic genome becomes transcriptionally active several cell cycles later but the mechanisms that initiate zygotic genome activation (ZGA) remain largely unknown. Here, I propose a role for N6-methyldeoxyadenosine (6mA) in regulating ZGA during early development of animals.

Using ethyl-uridine incorporation in the cnidarian *Hydractinia*, I found that the major wave of ZGA occurs at the 64-128 cell stage. Then, using mass-spectrometry, I found that during early embryogenesis 6mA levels increase to 3-fold above background level, peaking at the 16-32 cell stage. At the 64-128 cell stage, concomitant with the major wave of ZGA, a marked decrease in 6mA methylation was observed. I propose that 6mA transiently halts zygotic transcription, which otherwise would commence at the 16-32cell stage, while keeping relevant genes poised for transcription. Once this temporal block is removed, transcription is free to commence. To confirm this proposal, I developed a shRNA mediated knockdown protocol in *Hydractinia* aiming to knock down genes encoding 6mA methyltransferases (Mettl4 and N6amt1) and oxidoreductase that initiates demethylation (Alkbh4 and Alkbh1)

At 64-128 cell embryos, downregulation of *Alkbh1*, a 6mA “demethylator”, maintained the level of 6mA high, and the EU incorporation (hence ZGA) low. This shows that high level of 6mA beyond 16-32 cell stage blocks transcription in *Hydractinia*. However, shRNA mediated knockdown of two putative 6mA methyltransferases failed to alter EU incorporation and ZGA of *Hydractinia* early embryos. I hypothesize that 6mA accumulation in the genome is a result of “accidental” incorporation from reduced m6A which derived from maternal RNA degradation.

I provide circumstantial evidence consistent with a scenario in which 6mA accumulates at the 16-32 cell stage randomly, resulting from recycling of methylated, zygotically degraded RNAs. Thus, it circumvents the need for functional 6mA methyltransferases to explain the 6mA accumulation at 16-32 cell embryos of *Hydractinia*. In the broader context, it suggests that 6mA is a forfuitous incorporation in animal genomes rather than an epigenetic mark. Alkbh1, therefore, purges contaminating 6mA, facilitating transcription.

1 AN INTRODUCTION ON DNA METHYLATION AND THEIR ROLE DURING EMBRYOGENESIS OF ANIMALS

1.1 DNA methylation in prokaryotes and in animals

DNA methylation is the covalent addition of a methyl group to the nucleobases of DNA. Historically, it was established in *Escherichia coli* that methylation of the exocyclic -NH_2 group at the sixth position of the purine ring of adenine (6mA; Figure 1.1;(Dunn and Smith 1958)) occurs in a palindromic 5'-ApT-3' context (Geier and Modrich 1979). Methylation on ApT context (Figure 1.1) is a factor initiating replication at the site of origin (Messer et al. 1985), which then interacts with sequestration machineries (i.e. SeqA) to deter multiple replication initiation at the Ori site (Campbell and Kleckner 1990; Slater et al. 1995). The 6mA also marks the parental strand distinctively from daughter strand, which helps to correct mismatches soon after DNA replication (Pukkila et al. 1983; Lahue et al. 1987). DNA methylation is also used by bacteria in the restriction-modification (R-M) system to distinguish foreign-pathogenic DNA from self DNA (Arber and Dussoix 1962; Boyer 1971). Furthermore, few specific cases showed that 6mA interacts with proteins, which regulates transcription in *E. coli* (Braaten et al. 1994; Hale et al. 1998; Haagmans and van Der Woude 2000).

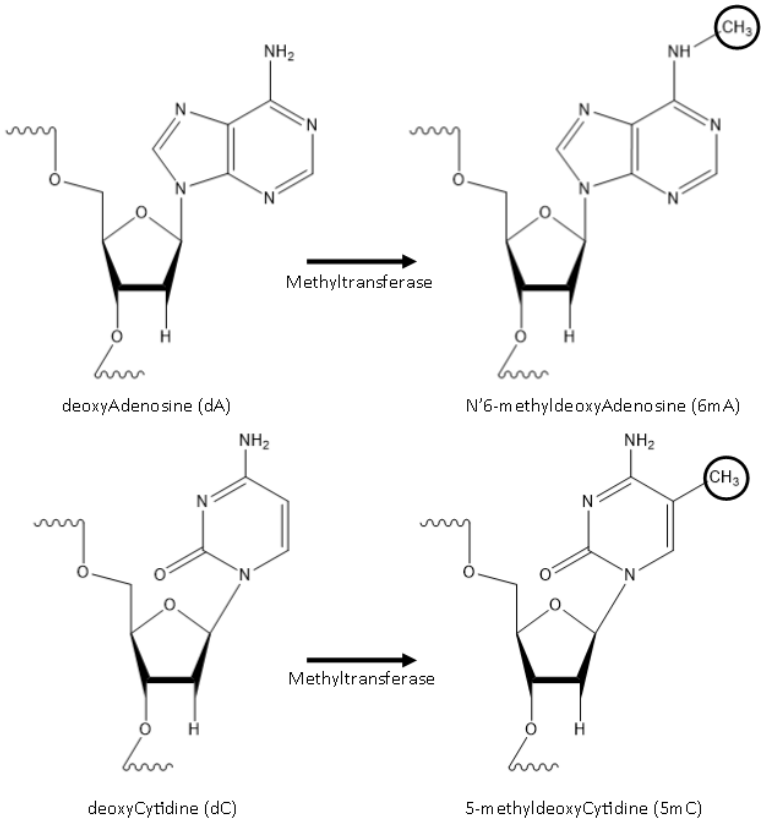


Figure 1.1. DNA methylation by methyltransferases on 6mA and 5mC

Attempts to find DNA methylation in animals indicated that methylation on the fifth position of the pyrimidine ring of cytosine (5mC; Figure 1.1) is more prevalent than 6mA (Vanyushin et al. 1970). In sea urchin embryos and to a greater extent also in the majority of vertebrate models, 5mC was found in palindromic 5'-CpG-3' dinucleotide (CpG methylation; (Grippo et al. 1968; Vanyushin et al. 1970; Bird and Southern 1978)). This CpG methylation setup resembles their bacterial counterpart, i.e., 6mA marked in ApT dinucleotide. Palindromic sequence is necessary for the methyl mark to be maintained after DNA replication on the daughter strands of newly synthesized DNA, thus CpG/ApT methylation is heritable (Figure 1.2; (Holliday and Pugh 1975; Bird and Southern 1978)). Without palindromic context, methylated dA (or dC) on one parent strand are very difficult to be copied to their respective daughter strands after replication (Figure 1.2;(Holliday and Pugh 1975; Bird and Southern 1978)). Therefore, non-CpG methylation (and non-ApT methylation) is difficult to maintain across cell cycles, hence very often overlooked. Nonetheless, relatively recent evidence indicates the potential roles of the non-CpG methylation in mammalian embryonic stem cells and brain development (Ramsahoye et al. 2000; Lister et al. 2013).

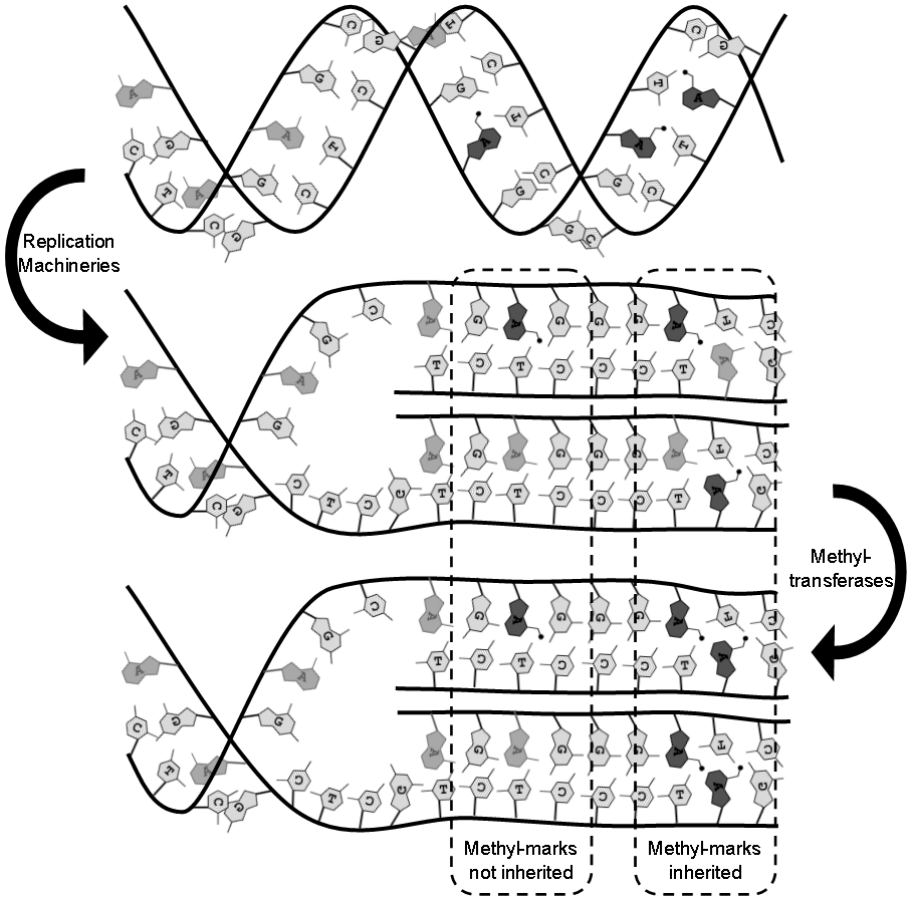


Figure 1.2. Inheritance of DNA methylation in palindromic context

1.2 DNA-5mC methylation is functional on developmental events

Around 90% of CpG dinucleotides are methylated in sea urchin and frog embryos (Grippo et al. 1968; Bird and Southern 1978), while mammals methylate 70-80% of their CpGs (Li and Zhang 2014). Interestingly, these ~25% unmethylated CpGs are mostly in the promoter region of genes and more than 60% of human genes have long stretched CpGs at their promoter region (CpG island) (Li and Zhang 2014). Many of these CpG islands are unmethylated in germ cells, early embryos and most somatic tissues (Bird et al. 1985). However, specific CpG islands are heavily methylated such as imprinted genes (Sapienza et al. 1987; Bartolomei et al. 1993; Ferguson-Smith et al. 1993; Stöger et al. 1993), germline-specific genes (Borgel et al. 2010; Auclair et al. 2014) and the inactive X chromosome (Mohandas et al. 1981; Lock et al. 1987). Mammalian genomes undergo two extensive waves of reprogramming (removal and redeposition) of CpG methylation patterns during early embryogenesis and lineages specification (Monk et al. 1987; Sanford et al. 1987). These waves indicate roles of DNA methylation (at least 5mC) in determining cell fate, hence also in gene expression regulation.

Early in vitro and in vivo evidence associated CpG island methylation with gene silencing of few specific genes (Stein et al. 1982b; Busslinger et al. 1983; Ben-Hattar and Jiricny 1988). Indeed, the heavily methylated CpG islands in imprinted genes, X chromosome inactivation, and germline-specific genes are associated with strong and lifelong silencing (Mohandas et al. 1981; Lock et al. 1987; Sapienza et al. 1987; Bartolomei et al. 1993; Ferguson-Smith et al. 1993; Stöger et al. 1993; Borgel et al. 2010; Auclair et al. 2014). While most of the unmethylated CpG islands are more associated with accessible genomic region, some are differentially methylated, hence the genes are silent in different somatic tissues (Borgel et al. 2010; Auclair et al. 2014). This indicates a specific role of CpG island methylation in gene repression during differentiation. However, how mechanistically CpG methylation in animals leads to transcription inhibition varies.

One explanation is that methylated promoters would inhibit the binding of transcription factors (TFs) hence repress adjacent genes (Iguchi-Ariga and Schaffner 1989; Comb and Goodman 1990; Gaston and Fried 1995). These studies were reiterated with the recent survey on transcription factors binding activity which shows that 117 TFs (out of 534 human transcription factors tested) did not or only weakly bind to the methylated versions of their respective recognition sequence (Yin et al. 2017). However, 34% (175) of the tested TFs preferentially bound to methylated recognition sequence. These methylated site-binding TFs include the homeodomain factors that specify the embryonic axis (e.g., HOXC11 and HOXB13), the NKX proteins that define cell lineages during development, and the pluripotency regulator POU5F1/OCT4 (Yin et al. 2017). Therefore,

DNA methylation has a context-dependent relationship with transcription factors or transcription in general during embryonic development.

1.3 Histone modification interplays surrounding DNA methylation

Eukaryotic DNA, *in vivo*, is wrapped around histones, a DNA associated protein, rich of lysine (K) and arginine (R) and evolutionarily conserved between Archaea and Eukaryota. Octamers of histones comprise of Histone 2A (H2A), H2B, H3 and H4 forming a bundle, which may disallow other protein like transcription machineries to access the DNA (Kornberg 1977; Mathis et al. 1980; McGhee and Felsenfeld 1980; Sandman et al. 1990; Kornberg and Lorch 1995). Some of these histones possess amino and/or carboxy terminal tails that are subject to post translational modifications, which mark specific areas of chromosomes (Reviewed in (Strahl and Allis 2000; Berger 2007)). One of the most studied histone modifications is the trimethylation (me3) of lysine (K) 4 of histone 3 (H3K4me3), which marks actively transcribed genes at their promoters (Schneider et al. 2004; Guenther et al. 2007). H3K4me3 repels DNMT3, a 5mC methyltransferase (Otani et al. 2009). The removal of these methyl groups generates H3K4 at promotors, which is known to attract the ADD domain of DNMT3(Zhang et al. 2010).

DNMT3 also contains a PWWP domain, which recognizes H3K36me3 (Chen et al. 2004). H3K36me3 are deposited immediately following RNA pol II during transcription elongation along gene bodies (Sun et al. 2005). Therefore, H3K36me3 interaction with DNMT3 is associated with gene body 5mC methylation. This interaction between H3K36me3 and DNMT3 also mediates gene body 5mC methylation at pericentromeric satellite repeats (Chen et al. 2004). The correlation between this interaction and the life-long repressed status of the pericentromeric repeats, however, is unclear.

Instead, the constitutive heterochromatin state of the pericentromeric satellite repeat is tightly associated with H3K9me3 (Peters et al. 2003), which interacts with DNMT1 (the methyltransferase that maintain 5mC mark post replication) either directly or through UHRF1 (Li et al. 2018a; Ren et al. 2020). Disruption of the H3K9 methylator gene, *Suv39h,* disrupts 5mC methylation specifically at pericentromeric loci of mouse (Lehnertz et al. 2003). Moreover, low level of CpG-methylation at pericentromeric loci is associated with massive transcription of the loci in tumour cells (Narayan et al. 1998).

The knockout of *Dnmt1* and *Dnmt3* mouse somatic cells and neural stem cells respectively, however, allows Polycomb repression complex 2 (PRC2) to deposit H3K27me3, which then occupies the supposedly methylated region, facilitating the repression of the region (Wu et al. 2010; Reddington et al. 2013). H3K27me3 is a repression marker like CpG methylation. However, both marks mutually exclusive

regions in the genome (Jermann et al. 2014). The areas, which are marked by two-repression markers (H3K27me3 and CpG methylation), are also mutually exclusive with most the nucleosome that contain H2A.Z and H3K4me3, the active genes markers (Barski et al. 2007; Jermann et al. 2014). However, there are plenty of exceptions to this general rule between repressive (H3K27me3 and CpG methylation) and active genes markers (H2A.Z and H3K4me3). For instance, in pluripotent cells, the bivalent and poised markers are particularly enriched at lineage-specific genes, which contain both active H3K4me3 and repressive H3K27me3 marks (Azuara et al. 2006).

H2A.Z is not a result of post translational modification on H2A tail, instead it is a histone variant of H2A that marks active genes (Barski et al. 2007). In ciliates, where 6mA ApT methylation was reported, 6mAs marks the area of active genes together with H2A.Z and H3K4me3 (Wang et al. 2017b). In mouse embryonic stem cell, H2A.X (another variant of H2A) marks nucleosomes that are enriched with 6mA and associated with the silencing of retrotransposons on the X chromosome (Wu et al. 2016). These two-contradicting utilization of 6mA (marking active gene in ciliates versus silenced transposon in mouse) display the adaptability of 6mA usage in distantly related organisms.

1.4 Epigenetic Regulation of Zygotic Genome Activation

During early embryonic development of vertebrates, DNA methylation (5mC) has an inverse relationship with transcription (section 1.2), while intertwined with the role of histone modification (section 1.3). The genome of early embryos is, however, inactive, hence no embryonic genes are transcribed. Upon fertilization, cellular processes in embryos are controlled by RNAs and proteins derived from the maternal genome. The zygotic genome is activated only after one to two cell division-cycles. lasting for more than 10 hours in slow developing animals such as human and mouse. In rapid developing animals (i.e. *Drosophila*, *C. elegans*, zebrafish, frogs, and most invertebrates), however, zygotic genome activation (ZGA) occurs within less than three hours but only after 6-8 division-cycles (reviewed in (Schulz and Harrison 2019)).

Because 5mC DNA methylation are strongly associated with a gene silencing role (Borgel et al. 2010; Auclair et al. 2014), the global 5mC-DNA demethylation in human and mouse genomes before ZGA indicates its temporal control over ZGA (Monk et al. 1987; Sanford et al. 1987). In the genome of the rapidly developing zebrafish, the unmethylated CpGs area inherited from the gametes are strongly associated with accessible promotors (hence the genes are actively transcribed) during ZGA (Liu et al. 2018). Meanwhile, the highly methylated distal regulatory regions in the genome, which are inherited from the sperm, facilitate the binding of transcription factors (e.g. Pou5f3

and Nanog) (Liu et al. 2018). These TFs then act as master regulators of the major ZGA wave, establishing accessible chromatin. OCT4, the human homolog of zebrafish Pou5f3, is found in human embryos and binds at regions of accessible chromatin during ZGA. Furthermore, knockdown of OCT4 downregulates ZGA (Gao et al. 2018). This indicates a conserved TF-ZGA association among vertebrates. However, the relationship between this TF-ZGA and 5mC DNA methylation require further investigation in mammals.

In *Drosophila*, Zelda is the transcription factor that regulates ZGA (Liang et al. 2008). However, *Drosophila* contains no 5mC, thus the association between 5mC-DNA methylation with pioneer transcription factors and ZGA is, probably, not conserved among animals. In *Drosophila*, however, one study provides evidence for an association between 6mA-DNA methylation, a transcription factor, and ZGA (He et al. 2019). This indicate that, perhaps, different animals utilize different DNA methylation marks to control TF-ZGA association. Nonetheless, the relationship between 6mA and TF or ZGA in vertebrate embryos remains unknown, likewise in invertebrates apart from *Drosophila*.

1.5 DNA methylation in invertebrate and early diverging animals

The studies that dissect the role of CpG methylation in animals have mainly come from mammalian cell cultures, and mouse and frog embryos (Bird and Southern 1978; Mohandas et al. 1981; Stein et al. 1982a). Despites the early contribution of studies on invertebrate embryos, such as sea urchin (Grippo et al. 1968), the best established invertebrate model organisms, *Drosophila melanogaster* and *Caenorhabditis elegans* lack 5mC nearly entirely (Urieli-Shoval et al. 1982; Simpson et al. 1986), thus cannot provide any meaningful insights into 5mC function in invertebrates. However, many invertebrates, apart from these two ecdysozoans, evidently contain 5mC as it has been established from sea urchin embryos (Grippo et al. 1968), honey bees, tunicates, silk moth and early diverging animals modelled by the sea anemone *Nematostella vectensis* (Zemach et al. 2010; Schwaiger et al. 2014) and the sponge *Amphimedon queenslandica* (Figure 1.3, (de Mendoza et al. 2019). The sponge (Porifera) and sea anemone (Cnidaria) are two out of four phyla outside of the bilaterian cluster, which generally accepted as early diverging animals (Dunn et al. 2008; Pick et al. 2010). The other two, Ctenophora and Placozoa, also have scarce reports on DNA methylation (Dabe et al. 2015; de Mendoza et al. 2019). Further, sister to the metazoan cluster are the unicellular holozoans such as *Salpingoeca rosetta* and *Capsaspora owczarzaki*. DNA methylation are understudied in these organisms (Figure 1.3).

The bilaterian clade is divided into two superphyla: protostomes and deuterostomes. The former is usually modelled by the two ecdysozoans mentioned above (*C. elegans* and

D. melanogaster), while deuterostomes are mostly represented by vertebrates like mouse, human, frog and zebrafish. Given the lack of CpG methylation in the two ecdysozoans (Urieli-Shoval et al. 1982; Simpson et al. 1986), the urge to understand the evolution of CpG methylation outside the vertebrates became apparent. Most invertebrates studied (including early diverging animals) exhibit a distinct pattern of 5mC methylation. The unmethylated CpG islands at the transcription start sites were thought to be a vertebrate specific traits and gene body methylation to be more widespread (de Mendoza et al. 2019; de Mendoza et al. 2020). The function of gene body CpG methylation, however, remains elusive. As reverse genetic tools are not available for most invertebrates apart from the two established ecdysozoans, which lack any CpG methylation, thus functional investigation on this evolutionary conserved gene body methylation pattern in invertebrates is still scarce. Nonetheless, the genomes of sponges display CpG methylation patterns resembling the mammalian patterns (de Mendoza et al. 2019). However, functional mechanistic studies in sponge are limited (Adamska 2016), thus the questions on how mechanistically DNA methylation plays a role in cell fate specification in invertebrates remain unanswered. Hence, early diverging animals with available genetic manipulation tools would play important role to uncover it.

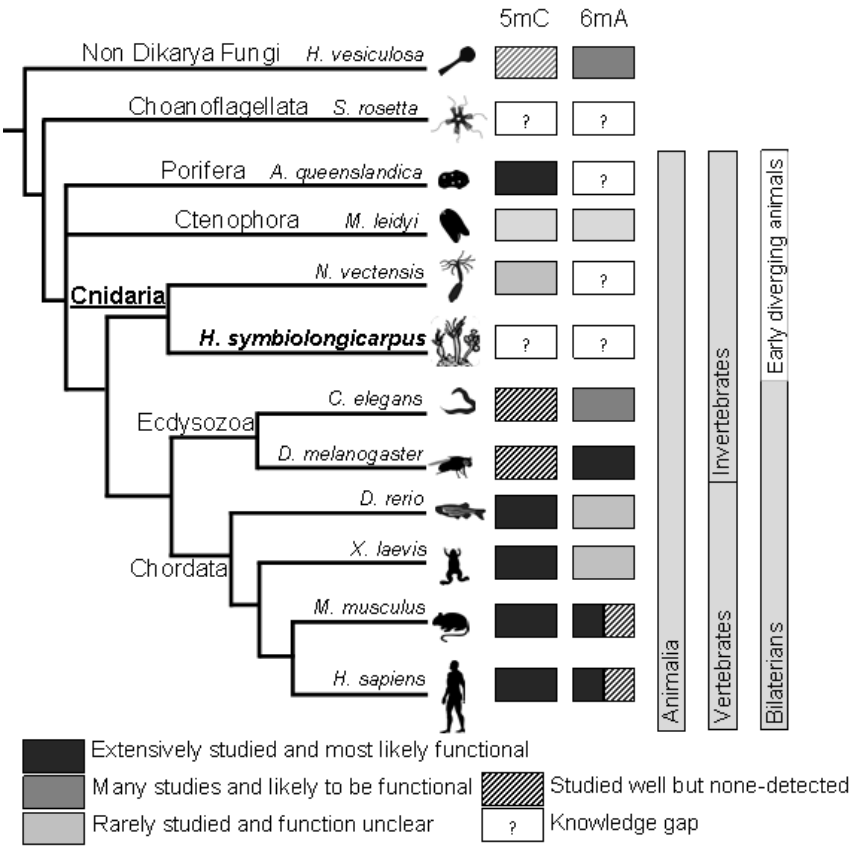


Figure 1.3. Studies of DNA methylation in animal models

Efforts to find DNA methylation in animals, instead of focusing only on 5mC, have begun by open searches for 6mA and other modes of DNA modification. Thus, it was indicated that *Drosophila melanogaster*, despite containing no 5mC, might have 6mA in its genome (Achwal et al. 1983). In 2015, when I began my PhD project, two independent groups published reports on 6mA in the two main invertebrate model animals, *D. melanogaster* (Zhang et al. 2015) and *C. elegans* (Greer et al. 2015) (Figure 1.3). Hence, it was opening new avenues for research on DNA methylation function in invertebrates. These studies were then followed by 6mA reports in frog, fish, pig, mouse, and cultured mammalian cells (Dunn and Smith 1958; Koziol et al. 2016; Liu et al. 2016b; Wu et al. 2016; Mondo et al. 2017; Xiao et al. 2018; Xie et al. 2018) (Figure 1.3). While 6mA reports from mammals are conflicting with each other (Figure 1.3,(Schiffers et al. 2017; O'Brown et al. 2019; Hao et al. ; Tian et al. 2020), those that reported 6mA in zebrafish and *Drosophila* implied potential role of 6mA during early embryonic development (Zhang et al. 2015; Liu et al. 2016b). Therefore, I argue that studying the function of 6mA, in addition to the 5mC investigation, in early embryo of an early diverging animal would be insightful.

1.6 Why *Hydractinia symbiolongicarpus*?

An attempt to understand 5mC and 6mA in early diverging animals has also been made in a ctenophore but without mechanistic functional investigations (Figure 1.3, (Dabe et al. 2015). Even after the description of 5mC methylation in *Nematostella vectensis* in 2010, so far, reports of 6mA in cnidarians have been scarce (Figure 1.3). Therefore, I decided to aim for not only describing DNA methylation (5mC and 6mA) in early diverging animals, but also investigating the function of 6mA in embryonic development of *Hydractinia symbiolongicarpus*.

To investigate the function of DNA methylation during early embryogenesis of early diverging animals, the animal model used for the study must meet few criteria. Among the non-bilaterian animals, there are five animals that can be considered as model organism or emerging model organism, where genome sequence is available and with protocols for genes expression manipulation. The options are *Hydra* (Klimovich et al. 2019), *Nematostella vectensis* (Ikmi et al. 2014; Renfer and Technau 2017; Karabulut et al. 2019), *Exaiptasia pallida* (Jones et al. 2018), *Hydractinia symbiolongicarpus* (Sanders et al. 2018; Quiroga-Artigas et al. 2020), and *Clytia hemisphaerica* (Momose et al. 2018; Leclère et al. 2019). However, *Hydra* embryos, after early cleavages, are covered with thick cuticle thus difficult to access for immunofluorescence or in situ hybridization (Martin et al. 1997), and *Exaiptasia pallida* larva are as yet impossible to metamorphose in a laboratory set up (Bucher et al. 2016). Thus, these reasonings narrowed my options

to only three model animals *Hydractinia*, *Clytia* and *Nematostella*. Evolutionarily, the anthozoan *Nematostella* is sister to the hydrozoans *Hydractinia* and *Clytia* (Figure 1.4). As either *Clytia* or *Hydractinia* would simply represent hydrozoans, out of the three, I narrow my options down to *Nematostella* or *Hydractinia.*

One aspect of studying 6mA is the necessity to collect a quarter or a half million of embryos for DNA extraction and mass spectrometry. Comparing to the availability of collecting only dozens of thousands of eggs from *Nematostella* lab culture weekly (Stefanik et al. 2013), 20 pairs of *Hydractinia* sexual polyps can produce ~10,000 eggs weekly (Frank et al. 2001). Moreover, *Hydractinia* is a colonial animal which spread on glass slides, hence a single pair of glass slides that contain hundreds of sexual polyps that can easily produce ~100.000 eggs weekly if not daily. Thus rationally, *Hydractinia* is more suitable comparing to other cnidarians to investigate DNA methylation during their early embryogenesis.

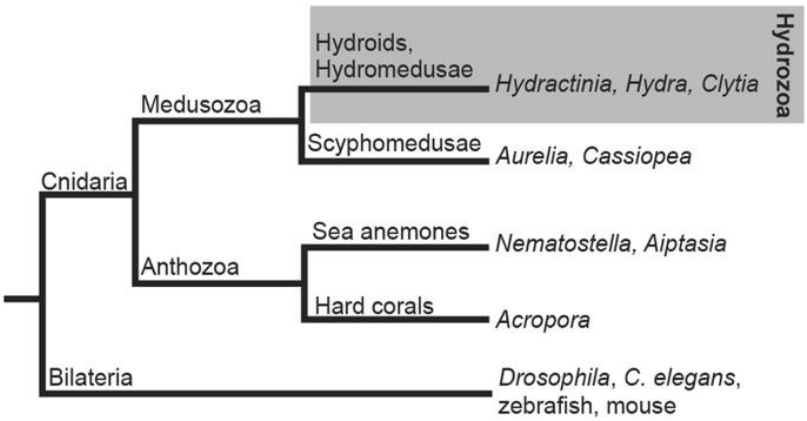


Figure 1.4 Simple cladogram showing *Hydractinia* evolutionary relationships with other animal models. Source:(Frank et al. 2020)

1.7 The aim and hypotheses

In this PhD project, I aim to describe DNA methylation (5mC and 6mA) in early diverging animals and to investigate the function of 6mA in embryonic development of *Hydractinia symbiolongicarpus*. To achieve this aim, I formulated three hypotheses: (1) The genomes of early diverging animals contain 5mC and 6mA; (2) The levels of DNA methylation during embryonic development are dynamic, thus the expression of genes that control methylation and demethylation of DNA will follow the dynamics of their respective target of DNA methylation; (3) Mis-expressing genes that methylate and demethylate DNA will alter DNA methylation level, thus changing global gene expression pattern during early embryonic development. To test these three hypotheses, I have used the cnidarian *Hydractinia symbiolongicarpus* as model animal.

The specific aims of this thesis are:

1. To measure the level of 5mC and 6mA in the genome of early diverging animals
2. To measure the level of 5mC and 6mA during embryogenesis of *Hydractinia*.
3. To investigate the global gene expression profile of *Hydractinia* during embryogenesis.
4. To manipulate the expression of genes predicted to control 6mA and study the effect of it on *Hydractinia* embryogenesis.

1.8 References

Achwal, Chhaya W, Chitra A Iyer, and H Sharat Chandra. 1983. 'Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, *Drosophila* and mealybug DNA', *FEBS letters*, 158: 353-358.

Adamska, Maja. 2016. 'Sponges as models to study emergence of complex animals', *Current Opinion in Genetics & Development*, 39: 21-28.

Arber, Werner, and Daisy Dussoix. 1962. 'Host specificity of DNA produced by *Escherichia coli*: I. Host controlled modification of bacteriophage λ', *Journal of Molecular Biology*, 5: 18-36.

Auclair, Ghislain, Sylvain Guibert, Ambre Bender, and Michael Weber. 2014. 'Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse', *Genome biology*, 15: 545.

Azuara, Véronique, Pascale Perry, Stephan Sauer, Mikhail Spivakov, Helle F. Jørgensen, Rosalind M. John, . . . Amanda G. Fisher. 2006. 'Chromatin signatures of pluripotent cell lines', *Nature Cell Biology*, 8: 532-538.

Barski, Artem, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, Zhibin Wang, . . . Keji Zhao. 2007. 'High-resolution profiling of histone methylations in the human genome', *Cell*, 129: 823-837.

Bartolomei, Marisa S, Andrea L Webber, Mary E Brunkow, and Shirley Marie Tilghman. 1993. 'Epigenetic mechanisms underlying the imprinting of the mouse H19 gene', *Genes & development*, 7: 1663-1673.

Ben-Hattar, Jean, and Josef Jiricny. 1988. 'Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo', *Gene*, 65: 219-227.

Berger, Shelley L. 2007. 'The complex language of chromatin regulation during transcription', *Nature*, 447: 407-412.

Bird, Adrian P, and Edwin M Southern. 1978. 'Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from *Xenopus laevis*', *Journal of Molecular Biology*, 118: 27-47.

Bird, Adrian, Mary Taggart, Marianne Frommer, Orlando J Miller, and Donald Macleod. 1985. 'A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA', *Cell*, 40: 91-99.

Borgel, Julie, Sylvain Guibert, Yufeng Li, Hatsune Chiba, Dirk Schübeler, Hiroyuki Sasaki, . . . Michael Weber. 2010. 'Targets and dynamics of promoter DNA methylation during early mouse development', *Nature genetics*, 42: 1093.

Boyer, Herbert W. 1971. 'DNA restriction and modification mechanisms in bacteria', *Annual Reviews in Microbiology*, 25: 153-176.

Braaten, Bruce A, Xiangwu Nou, Linda S Kaltenbach, and David A Low. 1994. 'Methylation patterns in pap regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*', *Cell*, 76: 577-588.

Bucher, Madeline, Iliona Wolfowicz, Philipp A Voss, Elizabeth A Hambleton, and Annika Guse. 2016. 'Development and symbiosis establishment in the cnidarian endosymbiosis model *Aiptasia sp*', *Scientific reports*, 6: 19867.

Busslinger, Meinrad, Jacky Hurst, and Richard A Flavell. 1983. 'DNA methylation and the regulation of globin gene expression', *Cell*, 34: 197-206.

Campbell, Joseph L, and Nancy Kleckner. 1990. '*E. coli* oriC and the *dna*A gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork', *Cell*, 62: 967-979.

Chen, Taiping, Naomi Tsujimoto, and En Li. 2004. 'The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin', *Molecular and Cellular Biology*, 24: 9048-9058.

Comb, Michael, and Howard M Goodman. 1990. 'CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2', *Nucleic Acids Research*, 18: 3975-3982.

Dabe, Emily C, Rachel S Sanford, Andrea B Kohn, Yelena Bobkova, and Leonid L Moroz. 2015. 'DNA methylation in basal metazoans: Insights from ctenophores', *Integrative and comparative biology*, 55: 1096-1110.

de Mendoza, Alex, William L. Hatleberg, Kevin Pang, Sven Leininger, Ozren Bogdanovic, Jahnvi Pflueger, . . . Ryan Lister. 2019. 'Convergent evolution of a vertebrate-like methylome in a marine sponge', *Nature Ecology & Evolution*, 3: 1464-1473.

de Mendoza, Alex, Ryan Lister, and Ozren Bogdanovic. 2020. 'Evolution of DNA methylome diversity in eukaryotes', *Journal of Molecular Biology*, 432: 1687-1705.

Dunn, Casey W, Andreas Hejnol, David Q Matus, Kevin Pang, William E Browne, Stephen A Smith, . . . Gonzalo Giribet. 2008. 'Broad phylogenomic sampling improves resolution of the animal tree of life', *Nature*, 452: 745-749.

Dunn, D. B., and J. D. Smith. 1958. 'The occurrence of 6-methylaminopurine in deoxyribonucleic acids', *The Biochemical journal*, 68: 627-636.

Ferguson-Smith, Anne C, Hiroyuki Sasaki, Bruce M Cattanach, and M Azim Surani. 1993. 'Parental-origin-specific epigenetic modification of the mouse H19 gene', *Nature*, 362: 751-755.

Frank, Uri, Thomas Leitz, and Werner A Müller. 2001. 'The hydroid *Hydractinia*: a versatile, informative cnidarian representative', *BioEssays*, 23: 963-971.

Frank, Uri, Matthew L Nicotra, and Christine E Schnitzler. 2020. 'The colonial cnidarian *Hydractinia*', *EvoDevo*, 11: 1-6.

Gao, Lei, Keliang Wu, Zhenbo Liu, Xuelong Yao, Shenli Yuan, Wenrong Tao, . . . Dongdong Fan. 2018. 'Chromatin accessibility landscape in human early embryos and its association with evolution', *Cell*, 173: 248-259. e215.

Gaston, Kevin, and Mike Fried. 1995. 'CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes', *Nucleic Acids Research*, 23: 901-909.

Geier, Gail E, and Paul Modrich. 1979. 'Recognition sequence of the dam methylase of *Escherichia coli* K12 and mode of cleavage of Dpn I endonuclease', *Journal of Biological Chemistry*, 254: 1408-1413.

Greer, Eric Lieberman, Mario Andres Blanco, Lei Gu, Erdem Sendinc, Jianzhao Liu, David Aristizábal-Corrales, . . . Yang Shi. 2015. 'DNA methylation on N6-adenine in *C. elegans*', *Cell*, 161: 868-878.

Grippo, P., M. Iaccarino, E. Parisi, and E. Scarano. 1968. 'Methylation of DNA in developing sea urchin embryos', *Journal of Molecular Biology*, 36: 195-208.

Guenther, Matthew G., Stuart S. Levine, Laurie A. Boyer, Rudolf Jaenisch, and Richard A. Young. 2007. 'A chromatin landmark and transcription initiation at most promoters in human cells', *Cell*, 130: 77-88.

Haagmans, W, and Marjan van Der Woude. 2000. 'Phase variation of Ag43 in *Escherichia coli*: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription', *Molecular microbiology*, 35: 877-887.

Hale, WB, MW Van Der Woude, BA Braaten, and DA Low. 1998. 'Regulation of uropathogenic *Escherichia coli* adhesin expression by DNA methylation', *Molecular genetics and metabolism*, 65: 191-196.

Hao, Ziyang, Tong Wu, Xiaolong Cui, Pingping Zhu, Caiping Tan, Xiaoyang Dou, . . . Chuan He. 2020. 'N6-deoxyadenosine methylation in mammalian mitochondrial DNA', *Molecular cell*.

He, Shunmin, Guoqiang Zhang, Jiajia Wang, Yajie Gao, Ruidi Sun, Zhijie Cao, . . . Dahua Chen. 2019. '6mA-DNA-binding factor Jumu controls maternal-to-zygotic transition upstream of Zelda', *Nature communications*, 10: 2219.

Holliday, Robin, and John E Pugh. 1975. 'DNA modification mechanisms and gene activity during development', *Science*, 187: 226-232.

Iguchi-Ariga, SM, and Walter Schaffner. 1989. 'CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation', *Genes & development*, 3: 612-619.

Ikmi, Aissam, Sean A McKinney, Kym M Delventhal, and Matthew C Gibson. 2014. 'TALEN and CRISPR/Cas9-mediated genome editing in the early-branching metazoan *Nematostella vectensis*', *Nature communications*, 5: 5486.

Jermann, Philip, Leslie Hoerner, Lukas Burger, and Dirk Schübeler. 2014. 'Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation', *Proceedings of the National Academy of Sciences*, 111: E3415-E3421.

Jones, Victor AS, Madeline Bucher, Elizabeth A Hambleton, and Annika Guse. 2018. 'Microinjection to deliver protein, mRNA, and DNA into zygotes of the cnidarian endosymbiosis model *Aiptasia sp*', *Scientific reports*, 8: 1-11.

Karabulut, Ahmet, Shuonan He, Cheng-Yi Chen, Sean A McKinney, and Matthew C Gibson. 2019. 'Electroporation of short hairpin RNAs for rapid and efficient gene knockdown in the starlet sea anemone, *Nematostella vectensis*', *Developmental Biology*, 448: 7-15.

Klimovich, Alexander, Jörg Wittlieb, and Thomas CG Bosch. 2019. 'Transgenesis in *Hydra* to characterize gene function and visualize cell behavior', *Nature protocols*, 14: 2069.

Kornberg, Roger D. 1977. 'Structure of chromatin', *Annual review of biochemistry*, 46: 931-954.

Kornberg, Roger D, and Yahli Lorch. 1995. 'Interplay between chromatin structure and transcription', *Current opinion in cell biology*, 7: 371-375.

Koziol, Magdalena J., Charles R. Bradshaw, George E. Allen, Ana S. H. Costa, Christian Frezza, and John B. Gurdon. 2016. 'Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications', *Nature Structural & Molecular Biology*, 23: 24-30.

Lahue, Robert S, Shin-San Su, and Paul Modrich. 1987. 'Requirement for d(GATC) sequences in *Escherichia coli* mutHLS mismatch correction', *Proceedings of the National Academy of Sciences*, 84: 1482-1486.

Leclère, Lucas, Coralie Horin, Sandra Chevalier, Pascal Lapébie, Philippe Dru, Sophie Péron, . . . Séverine Romano. 2019. 'The genome of the jellyfish *Clytia hemisphaerica* and the evolution of the cnidarian life-cycle', *Nature Ecology & Evolution*, 3: 801-810.

Lehnertz, Bernhard, Yoshihide Ueda, Alwin AHA Derijck, Ulrich Braunschweig, Laura Perez-Burgos, Stefan Kubicek, . . . Antoine HFM Peters. 2003. 'Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin', *Current Biology*, 13: 1192-1200.

Li, En, and Yi Zhang. 2014. 'DNA methylation in mammals', *Cold Spring Harbor perspectives in biology*, 6: a019133.

Li, Tao, Linsheng Wang, Yongming Du, Si Xie, Xi Yang, Fuming Lian, . . . Chengmin Qian. 2018a. 'Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation', *Nucleic Acids Research*, 46: 3218-3231.

Liang, Hsiao-Lan, Chung-Yi Nien, Hsiao-Yun Liu, Mark M Metzstein, Nikolai Kirov, and Christine Rushlow. 2008. 'The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*', *Nature*, 456: 400-403.

Lister, Ryan, Eran A Mukamel, Joseph R Nery, Mark Urich, Clare A Puddifoot, Nicholas D Johnson, . . . Matthew D Schultz. 2013. 'Global epigenomic reconfiguration during mammalian brain development', *Science*, 341: 1237905.

Liu, Guifen, Wen Wang, Shengen Hu, Xiangxiu Wang, and Yong Zhang. 2018. 'Inherited DNA methylation primes the establishment of accessible chromatin during genome activation', *Genome research*, 28: 998-1007.

Liu, Jianzhao, Yuanxiang Zhu, Guan-Zheng Luo, Xinxia Wang, Yanan Yue, Xiaona Wang, . . . Ye Fu. 2016b. 'Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig', *Nature communications*, 7: 1-7.

Lock, Leslie F, Nobuo Takagi, and Gail R Martin. 1987. 'Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation', *Cell*, 48: 39-46.

Martin, Vicki J, C Lynne Littlefield, William E Archer, and Hans R Bode. 1997. 'Embryogenesis in *Hydra*', *The Biological Bulletin*, 192: 345-363.

Mathis, Diane, Pierre Oudet, and Pierre Chambon. 1980. 'Structure of transcribing chromatin.' in, *Progress in Nucleic Acid Research and Molecular Biology* (Elsevier).

McGhee, James D, and Gary Felsenfeld. 1980. 'Nucleosome structure', *Annual review of biochemistry*, 49: 1115-1156.

Messer, Walter, Ute Bellekes, and Heinz Lother. 1985. 'Effect of dam methylation on the activity of the *E. coli* replication origin, oriC', *The EMBO journal*, 4: 1327-1332.

Mohandas, T, RS Sparkes, and LJ Shapiro. 1981. 'Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation', *Science*, 211: 393-396.

Momose, Tsuyoshi, Anne De Cian, Kogiku Shiba, Kazuo Inaba, Carine Giovannangeli, and Jean-Paul Concordet. 2018. 'High doses of CRISPR/Cas9 ribonucleoprotein efficiently induce gene knockout with low mosaicism in the hydrozoan *Clytia hemisphaerica* through microhomology-mediated deletion', *Scientific reports*, 8: 1-12.

Mondo, Stephen J, Richard O Dannebaum, Rita C Kuo, Katherine B Louie, Adam J Bewick, Kurt LaButti, . . . Steven R Ahrendt. 2017. 'Widespread adenine N6-methylation of active genes in fungi', *Nature genetics*, 49: 964-968.

Monk, Marilyn, Michael Boubelik, and Sigrid Lehnert. 1987. 'Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development', *Development*, 99: 371-382.

Narayan, Ajita, Weizhen Ji, Xian-Yang Zhang, Aizen Marrogi, Jeremy R. Graff, Stephen B. Baylin, and Melanie Ehrlich. 1998. 'Hypomethylation of pericentromeric DNA in breast adenocarcinomas', *International Journal of Cancer*, 77: 833-838.

O'Brown, Zach K, Konstantinos Boulias, Jie Wang, Simon Yuan Wang, Natasha M O'Brown, Ziyang Hao, . . . Eric Lieberman Greer. 2019. 'Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA', *BMC genomics*, 20: 445.

Otani, Junji, Toshiyuki Nankumo, Kyohei Arita, Susumu Inamoto, Mariko Ariyoshi, and Masahiro Shirakawa. 2009. 'Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain', *EMBO reports*, 10: 1235-1241.

Peters, Antoine H. F. M., Stefan Kubicek, Karl Mechtler, Roderick J. O'Sullivan, Alwin A. H. A. Derijck, Laura Perez-Burgos, . . . Thomas Jenuwein. 2003. 'Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin', *Molecular cell*, 12: 1577-1589.

Pick, KS, Hervé Philippe, F Schreiber, D Erpenbeck, DJ Jackson, P Wrede, . . . G Woerheide. 2010. 'Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships', *Molecular Biology and Evolution*, 27: 1983-1987.

Pukkila, Patricia J, Janet Peterson, Gail Herman, Paul Modrich, and Matthew Meselson. 1983. 'Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*', *Genetics*, 104: 571-582.

Quiroga-Artigas, Gonzalo, Alexandrea Duscher, Katelyn Lundquist, Justin Waletich, and Christine E Schnitzler. 2020. 'Gene knockdown via electroporation of short hairpin RNAs in embryos of the marine hydroid *Hydractinia symbiolongicarpus*', *bioRxiv*.

Ramsahoye, Bernard H, Detlev Biniszkiewicz, Frank Lyko, Victoria Clark, Adrian P Bird, and Rudolf Jaenisch. 2000. 'Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a', *Proceedings of the National Academy of Sciences*, 97: 5237-5242.

Reddington, James P, Sara M Perricone, Colm E Nestor, Judith Reichmann, Neil A Youngson, Masako Suzuki, . . . Heidi Mjoseng. 2013. 'Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes', *Genome biology*, 14: R25.

Ren, Wendan, Huitao Fan, Sara A Grimm, Yiran Guo, Jae Jin Kim, Linhui Li, . . . John P Coan. 2020. 'Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation', *bioRxiv*.

Renfer, Eduard, and Ulrich Technau. 2017. 'Meganuclease-assisted generation of stable transgenics in the sea anemone *Nematostella vectensis*', *Nature protocols*, 12: 1844.

Sanders, Steven M, Zhiwei Ma, Julia M Hughes, Brooke M Riscoe, Gregory A Gibson, Alan M Watson, . . . Andreas D Baxevanis. 2018. 'CRISPR/Cas9-mediated gene knockin in the hydroid *Hydractinia symbiolongicarpus*', *BMC genomics*, 19: 649.

Sandman, Kathleen, Joseph A Krzycki, Beate Dobrinski, Rudi Lurz, and John N Reeve. 1990. 'HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones', *Proceedings of the National Academy of Sciences*, 87: 5788-5791.

Sanford, Janet P, Heather J Clark, Verne M Chapman, and Janet Rossant. 1987. 'Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse', *Genes & development*, 1: 1039-1046.

Sapienza, Carmen, Alan C Peterson, Janet Rossant, and Rudi Balling. 1987. 'Degree of methylation of transgenes is dependent on gamete of origin', *Nature*, 328: 251-254.

Schiffers, Sarah, Charlotte Ebert, René Rahimoff, Olesea Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, . . . Thomas Carell. 2017. 'Quantitative LC-MS provides no evidence for m6dA or m4dC in the genome of mouse embryonic stem cells and tissues', *Angewandte Chemie International Edition*, 56: 11268-11271.

Schneider, Robert, Andrew J. Bannister, Fiona A. Myers, Alan W. Thorne, Colyn Crane-Robinson, and Tony Kouzarides. 2004. 'Histone H3 lysine 4 methylation patterns in higher eukaryotic genes', *Nature Cell Biology*, 6: 73-77.

Schulz, Katharine N., and Melissa M. Harrison. 2019. 'Mechanisms regulating zygotic genome activation', *Nature Reviews Genetics*, 20: 221-234.

Schwaiger, Michaela, Anna Schönauer, André F Rendeiro, Carina Pribitzer, Alexandra Schauer, Anna F Gilles, . . . Ulrich Technau. 2014. 'Evolutionary conservation of the eumetazoan gene regulatory landscape', *Genome research*, 24: 639-650.

Simpson, Victoria J., Thomas E. Johnson, and Richard F. Hammen. 1986. '*Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging', *Nucleic Acids Research*, 14: 6711-6719.

Slater, Steven, Sture Wold, Min Lu, Erik Boye, Kirsten Skarstad, and Nancy Kleckner. 1995. '*E. coli* SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration', *Cell*, 82: 927-936.

Stefanik, Derek J, Lauren E Friedman, and John R Finnerty. 2013. 'Collecting, rearing, spawning and inducing regeneration of the starlet sea anemone, *Nematostella vectensis*', *Nature protocols*, 8: 916-923.

Stein, R, Y Gruenbaum, Y Pollack, A Razin, and H Cedar. 1982a. 'Clonal inheritance of the pattern of DNA methylation in mouse cells', *Proceedings of the National Academy of Sciences*, 79: 61-65.

Stein, R, Aharon Razin, and Howard Cedar. 1982b. 'In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells', *Proceedings of the National Academy of Sciences*, 79: 3418-3422.

Stöger, R, P Kubička, C-G Liu, T Kafri, A Razin, H Cedar, and DP Barlow. 1993. 'Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal', *Cell*, 73: 61-71.

Strahl, Brian D., and C. David Allis. 2000. 'The language of covalent histone modifications', *Nature*, 403: 41-45.

Sun, Xiao-Jian, Ju Wei, Xin-Yan Wu, Ming Hu, Lan Wang, Hai-Hong Wang, . . . Zhu Chen. 2005. 'Identification and characterization of a novel human histone H3

lysine 36-specific methyltransferase', *Journal of Biological Chemistry*, 280: 35261-35271.

Tian, Li-Fei, Yan-Ping Liu, Lianqi Chen, Qun Tang, Wei Wu, Wei Sun, . . . Xiao-Xue Yan. 2020. 'Structural basis of nucleic acid recognition and 6mA demethylation by human ALKBH1', *Cell Research*, 30: 272-275.

Urieli-Shoval, Simcha, Yosef Gruenbaum, John Sedat, and Aharon Razin. 1982. 'The absence of detectable methylated bases in *Drosophila melanogaster* DNA', *FEBS letters*, 146: 148-152.

Vanyushin, BF, SG Tkacheva, and AN Belozersky. 1970. 'Rare bases in animal DNA', *Nature*, 225: 948-949.

Wang, Yuanyuan, Xiao Chen, Yalan Sheng, Yifan Liu, and Shan Gao. 2017b. 'N6-adenine DNA methylation is associated with the linker DNA of H2A. Z-containing well-positioned nucleosomes in Pol II-transcribed genes in *Tetrahymena*', *Nucleic Acids Research*, 45: 11594-11606.

Wu, Hao, Volkan Coskun, Jifang Tao, Wei Xie, Weihong Ge, Kazuaki Yoshikawa, . . . Yi Eve Sun. 2010. 'Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes', *Science*, 329: 444-448.

Wu, Tao P, Tao Wang, Matthew G Seetin, Yongquan Lai, Shijia Zhu, Kaixuan Lin, . . . Mei Zhong. 2016. 'DNA methylation on N 6-adenine in mammalian embryonic stem cells', *Nature*, 532: 329-333.

Xiao, Chuan-Le, Song Zhu, Minghui He, De Chen, Qian Zhang, Ying Chen, . . . Feng Luo. 2018. 'N6-methyladenine DNA modification in the human genome', *Molecular cell*, 71: 306-318.

Xie, Qi, Tao P. Wu, Ryan C. Gimple, Zheng Li, Briana C. Prager, Qiulian Wu, . . . Jeremy N. Rich. 2018. 'N6-methyladenine DNA Modification in Glioblastoma', *Cell*, 175: 1228-1243.

Yin, Yimeng, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, . . . Jussi Taipale. 2017. 'Impact of cytosine methylation on DNA binding specificities of human transcription factors', *Science*, 356: eaaj2239.

Zemach, Assaf, Ivy E. McDaniel, Pedro Silva, and Daniel Zilberman. 2010. 'Genome-wide evolutionary analysis of eukaryotic DNA methylation', *Science*, 328: 916.

Zhang, Guoqiang, Hua Huang, Di Liu, Ying Cheng, Xiaoling Liu, Wenxin Zhang, . . . Jianzhao Liu. 2015. 'N6-methyladenine DNA modification in *Drosophila*', *Cell*, 161: 893-906.

Zhang, Yingying, Renata Jurkowska, Szabolcs Soeroes, Arumugam Rajavelu, Arunkumar Dhayalan, Ina Bock, . . . Wolfgang Fischle. 2010. 'Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail', *Nucleic Acids Research*, 38: 4246-4253.

2 DNA METHYLATION IN EARLY DIVERGING ANIMALS

DNA methylation studies among early diverging animals were mainly represented by exploiting bisulphite sequencing on the cnidarian *Nematostella vectensis* (Zemach et al. 2010) and very recently on the sponge *Amphimedon queenslandica* (de Mendoza et al. 2019), which overlooked 6mA. This chapter, and the thesis in general, will fill the knowledge gap on 6mA DNA methylation among early diverging animals using *Hydractinia* as model animal as well as the knowledge gap on 5mC DNA methylation among hydrozoans.

2.1 The presence of methylated DNA and the methyltransferases.

DNA methylation was classically detected by exploiting methylation sensitive endonuclease, which cleaved DNA in a sequence specific manner (Arber and Dussoix 1962; Bird and Southern 1978; Waalwijk and Flavell 1978; Geier and Modrich 1979; Stein et al. 1982a) or by a chemical approach, i.e. chromatography on sampled DNA (Dunn and Smith 1958; Grippo et al. 1968). Efforts to utilized immunohistochemistry or bisulphite sequencing were introduced later (Achwal et al. 1983; Frommer et al. 1992; Susan et al. 1994). Currently, bisulphite sequencing that specifically detect 5mC dominates the studies of DNA methylation on animals; hence 6mA studies were often overlooked.

The discovery of DNA Methyltransferases 1 (DNMT1) in animals (Shied et al. 1968), which specifically recognizes the hemi-methylated CpG area of DNA (Bird and Southern 1978; Stein et al. 1982a), is consistent with a proposal that CpG methylation has a non-genetic, but heritable, role in animals. The discovery of ten-eleven translocation (TET) as the enzyme that initiates the demethylation (Tahiliani et al. 2009) alongside the association between genomic imprinting with CpG methylation (Bartolomei et al. 1993; Stöger et al. 1993) has set the epigenetic role of CpG methylation in stone. These series of events that has encouraged the excitement of 5mC investigation in vertebrates were shadowed by the lack of detectable levels of 5mC in the two prominent and established model invertebrates, *Drosophila melanogaster* (Urieli-Shoval et al. 1982) and *Caenorhabditis elegans* (Simpson et al. 1986). Furthermore, any attempts to find DNMT or TET homologs acting on DNA in these two animals have failed. *Drosophila* Tet functionally initiates demethylation of m5C in RNA instead of oxidizing 5mC of DNA (Fu et al. 2014; Delatte et al. 2016).

Lack of 5mC in *Drosophila* and *C. elegans* encouraged two developments: first, establishment of more sensitive methods for detecting extremely scarce 5mC and, second, efforts to find other types of DNA methylation with potential epigenetic functions,

like 6mA. Thus, very sensitive HPLC-MS/MS have been developed to detect trace amounts of 5mC in *C. elegans* (Hu et al. 2014) and *Drosophila melanogaster* (Capuano et al. 2014). Despites the failed earlier attempts to detect 6mA by classical methods in *Drosophila* (Urieli-Shoval et al. 1982), studies using HPLC-MS/MS detected both 5mC and 6mA in this animal (Capuano et al. 2014; Zhang et al. 2015) and in *C. elegans* (Hu et al. 2014; Greer et al. 2015). However, these discoveries demanded an explanation for the lack of DNMT or TET acting on DNA in *Drosophila* and *C. elegans*.

Following the studies that detected 6mA in *Drosophila* and *C. elegans*, others provided evidence for 6mA methyltransferase (METTL4) and oxidoreductases (DMAD/TET and ALKBH4) (Greer et al. 2015; Zhang et al. 2015). These findings of 6mA and the enzymes that modify it from *Drosophila* and *C. elegans* provided 6mA a “second chance” as the DNA methylation type that may play an epigenetic role in animals. Very rapidly afterwards, 6mA has been detected in fungi (Mondo et al. 2017), zebrafish (Liu et al. 2016b), frog (Koziol et al. 2016), mouse (Wu et al. 2016) and human cells (Xiao et al. 2018). Moreover, additional enzymes, N6AMT1 and ALKBH1, were proposed to act as methylator or demethylation initiator of 6mA in mammals, respectively (Wu et al. 2016; Xiao et al. 2018).

These discoveries of 6mA encouraged a quick attempt to detect 6mA in ctenophores as early diverging animals’ representatives (Dabe et al. 2015). This study, however, relied on immunochemistry despite the general standard in the community to perform HPLC-MS/MS. An immunochemistry alone approach to detect 6mA is equivocal, requiring additional controls and testing to provide sufficient evidence for 6mA detection. However, early attempt to detect 6mA in *Drosophila* by immunochemistry that had been initially dismissed (Achwal et al. 1983) were confirmed by HPLC-MS/MS three decades after (Zhang et al. 2015). Thus, antibody-based methods may serve as good indication for the presence of 6mA but require confirmation by additional, independent methods.

In this chapter, I developed both approaches, immunochemistry, and HPLC-MS/MS, to detect 5mC and 6mA in early diverging animals. Moreover, I conducted surveys by sequence similarity and molecular phylogeny to provide evidence for the presence of genes associated with DNA methylation in early diverging animals. Together, I conclude that 5mC and 6mA are present and possibly actively deposited and removed by their respective specific enzymes among early diverging animals, with the exception of placozoans, to which I have gained no access to biological materials nor exhaustive data sources.

2.2 Methods

2.2.1 DNA extraction

Embryos or Polyps of *Hydractinia symbiolongicarpus* and *Nematostella vectensis*[1] were collected and washed at least 15 times with sterile filtered (0.2 µm, Thermo Scientific #723-2520) artificial sea water (Coral Pro Salt (Red Sea #Pro-22) to minimize the possibility of microbial contamination. In case of polyps, the last wash was done in 4% $MgCl_2$ (dissolved in 50% seawater) to anaesthetize the animals. I then let the embryos/polyps settled down at the bottom of a 2 ml tube.

Next, seawater (or 4% $MgCl_2$) was removed and replaced by 200-400 µl lysis-d1 buffer (100 mM Tris-HCl pH 8.0; 5 mM EDTA; 1%SDS; in MilliQ-water). Polyps were homogenized with a pestle and the embryos by vortexing for a few seconds. Then, 10-20 µl of proteinase K at 20 mg/ml (Macherey-Nagel #10655875) were added to the suspension and incubated at 55°C for a minimum 2 hours (maximum 18 hours). Next, 1 volume of Phenol:Chloroform:Isoamylalcohol (PCI) at 25:24:1 ratio was added to the suspension and mixed rigorously, followed by centrifugation at maximum speed at 4°C, for 10 minutes.

Aqueous phase was transferred to a new 2 ml tube and 1/50 volume of RNaseA (ThermoScientific #EN0531) and RNaseT1 (ThermoScientific #EN0541) were added and incubated at 37°C for 2 hours. Then, 1 volume of chloroform was added, mixed rigorously. This suspension was then centrifuged at maximum speed, 4°C, for 10 minutes. The aqueous phase was transferred to a new 2 ml tube then 1/20 volume of glycogen 20 mg/ml (ThermoScientific #R0551), 1/10 volume of Na-Acetate 4 M pH 5.2, and 2 volume of cold ethanol absolute were added. The tubes were carefully inverted after each addition. The suspension was then kept at -20°C overnight, then centrifuged at maximum speed 4°C for 10 minutes. Next, the supernatants were taken out and the pellets were washed with cold 70% ethanol twice then air dried at room temperature for at least 2 hours.

Next, the pellets were dissolved with 100 µl of pre-heated (70°C) TE buffer (10 mM Tris-HCl pH 8.0; 1 mM EDTA pH 8.0). This DNA solutions were then quickly assessed on Nanodrop simply to assure the presence of any DNA (despite dirty) in the solution. Then, 400 µl of diluted (1:4 v/v) QG buffer (Qiagen #19063) and 10 µl of Na-Acetate 4 M pH 5.2 were added into the solution and mixed thoroughly. Then, the solutions were transferred into DNA column (blue column from EpochLifeScience #1920) and spun to

[1] Specifically, for experiments involving *Nematostella vectensis* sea water were replaced with 1/3 sea water.

bind to DNA into the membrane on 6000 *x g* for 2 minutes. The flow-through were discarded, the membranes were washed with Wash-d2 buffer (10mM Tris-HCl pH 7.5; 60 mM potassium acetate; 80% Ethanol) twice, then dried with max speed centrifugation for 2 minutes followed by carefully moving the column to new 1.5/2 ml tube. The DNA was eluted with 30 µl pre-heated (70°C) nuclease free water.

2.2.2 DNA solution assessment

Mnemiopsis leidyi DNA was kindly provided by Dr. Miguel Salinas-Saavedra and Prof. Mark Martindale, University of Florida, USA. All DNA solutions were assessed using Nanodrop to get the nucleotide concentration in ng/µl, λ 260/280 ratio and absorbance at 340 nm. This method, however, measures all forms of nucleotides not only the polymers. To measure the intact double stranded polynucleotide DNA, I used the Qubit double stranded DNA broad range assay kit (dsDNA BR kit, ThermoScientific # Q32850) on Qubit fluorometer 2.0. RNA contamination was excluded by Qubit RNA high sensitivity (RNA HS kit, ThermoScientific # Q32852). Lastly, I electrophoresed the DNA solution on 0.8-1.0 % agarose gel to confirm the size and state of the DNA.

2.2.3 Dot-Blot

The extracted genomic DNA and plasmids were RNAse treated, purified, and reassessed with Qubit RNA-HS assay kit. Only samples with no detectable RNA used. The standard oligonucleotides were prepared with nuclease free water and mixed down to 0.1% of modifiedA/dA and labelled as MO1 and MO2 (see Table 2.3).

The DNA were diluted in 2x saline sodium citrate solution (SSC, 300 mM NaCl and 30 mM Trisodium Citrate; pH 7.0) to a final concentration of 100 ng/µl in a total volume of 15 µl. The DNAs were denatured in 95° for 5 minutes, then left on the bench for a short period, then 2 µl (200 ng, or otherwise stated) were pipetted onto the membrane (Amersham Hybond-N+, GE #RPN119B). The spots were air dried for 5-10 minutes before exposing them to UV (302 nm) irradiation for 3 minutes upside down on a Whatman paper twice with 1-minute interval. The membrane then was rinsed in block-DB1 solution (2.5% Skim-Milk, 0.1% Tween-20, dissolved in 1x phosphate-buffered saline (PBS, Sigma-Aldrich #11666789001)). The membrane was then transferred to wet chamber and incubated in primary antibody (see Table 2.1) solution (diluted in block-DB1 solution), which were pipetted on top of the membrane. The membrane was left in the wet chamber overnight at room temperature. The membrane was then washed three times by PBS-Tween (0.1% Tween-20 dissolved in 1x PBS). The membrane was then transferred back to the wet chamber and rinsed in block-DB2 solution (5% goat serum, 2.5% skim-milk, 0.1% tween-20, dissolved in 1x PBS) for one hour. Then, the horseradish peroxidase (HRP)-conjugated secondary antibody (see Table 2.1) solution

(diluted in block-DB2) was pipetted on top of the membrane and left in the wet chamber for one hour. The membrane was then washed with PBS-Tween four times. The membrane was incubated in horseradish peroxidase substrate (Thermo Scientific #32106) and was viewed using a gel documentation system (Alpha Innotech FluorChem FC2). Image were taken with FluorView 2.0 software.

Table 2.1. List of antibodies used for dot-blot experiments

Antibody	Sources (code)	Host	Antigen	Dilution
Polyclonal anti-m6A	Synaptic System (202003)	Rabbit	m6A/6mA	1:1000
Monoclonal anti-m6A	Synaptic System (202111)	Mouse	m6A/6mA	1:1000
Anti-m6A G1-11D11	Regensburg[2]	Mouse	m6A/6mA	1:5
Anti-m6A 2A-13G2	Regensburg	Rat	m6A/6mA	1:5
Monoclonal anti-5mC	Abcam (ab10805)	Mouse	5mC	1: 200
HRP-anti IgG	Abcam (ab6721)	Goat	Rabbit	1:1000

2.2.4 DNA digestion and preparation for HPLC-MS/MS

Total DNA was calculated then diluted (with nuclease free water) to made up a 28 µl of 1.5-2.5 µg DNA solution. I took 2 µl out this DNA solution for Nanodrop, Qubit dsDNA BR and Qubit RNA HS measurement. The DNA was then denatured at 100°C for 5 minutes, chilled on ice for 2 minutes, added 3 µl of 100 mM NH_4OAc pH 5.3 then digested by 1 µl DNaseI (NEB #M0303S) at 42°C overnight. It was followed by addition of 3.4 µl of 1 M NH_4HCO_3 and 1 µl of phosphodiesterase I from *Crotalus adamanteus* venom (an exonuclease, Sigma Aldrich # P3243-1VL) and incubated at 37 °C for at least 3 hours. Finally, 1 U Calf Intestinal Phosphatase (NEB # M0290S) was added to the solution and further incubated at 37 °C for at least 4 hours. The digested DNA solutions were then diluted up with nuclease free water to 200 µl and filtered by prepared centrifugal ultra-filtration tubes (MW cut-off 3 KDa, Amicon, Millipore #UFC500396). The flow through was freeze-dried, reconstituted by addition of 21 µl nuclease free water, incubated at 55°C, and spun down at 6000 x *g* for 2 minutes. I took 2 µl of this solution to be assessed by Nanodrop, Qubit dsDNA HS, and Qubit RNA HS assay kit. The quality of digested DNA was determined by the similar amount of DNA detected by Nanodrop before and after digestion alongside with the undetected level of DNA and RNA by the respective Qubit HS assay kit after digestion (but DNA was detected by Qubit dsDNA BR before digestion).

2.2.5 HPLC-MS/MS

To quantify the level of 5mC and 6mA in early diverging animals, I established a high performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS) system and in this case (due to availability in NUI Galway), the HPLC was tandem with a triple

[2] Kindly provided by Franziska Weichmann and Prof. Gunter Meister, University of Regensburg

quadrupole (QQQ-6490, an MS/MS system) and jet stream electrospray ionization source (Agilent, Santa Clara, CA). Hence, hereafter I use the abbreviation HPLC-QQQ. The column used in the HPLC is ZORBAX SB-C18 2.1 mm width x 50 mm length 1.8 µm particles. HPLC was performed with flow rate 250 µl/min using mobile phase A (0.1% formic acid solutions in water) and mobile phase B (0.1% formic acid in acetonitrile). The injection volume was 2 µl and a gradient elution for HPLC separation was used: 0-3 min, 5.0% B; 3-7min, 15.0% B; 7-10 min, 100% B. The QQQ set to multiple reaction monitoring (MRM, seeTable 2.2) in positive electrospray ionization mode in common 10eV CID and 80.0 V Fragmentation. Nucleosides were identified using the nucleoside precursor (parent) ion to product (daughter) ion mass transitions (Table 2.2).

Table 2.2 List of Nucleoside Ion Mass for Multi Reaction Monitoring (MRM)

Nucleoside	**Precursor Ion Mass**	**Product Ion Mass**
dC	228.1	112.1
5mC	242.1	126.1
dA	252.1	136.1
6mA	266.1	150.1
dG	268.1	152.1
dT	243.1	127.1

The output data from triple quadrupole (QQQ) machine were recorded and visualised using Agilent MassHunter B.07.00. Chromatograms were extracted for visualization and calculation was based on the multi reaction monitoring assigned (Table 2.2), then imported and compiled into Microsoft Excel for illustration purposes. The raw data in are provided as Appendix (see page. 113). Modified oligo 1 (Table 2.3) was used as external standard, which was digested and injected to the HPLC-QQQ system as the samples were treated. This was done in series of dilutions and served as standard curve for calculation of the DNA samples from animals (Figure 2.1). The 6mA/dA% were calculated as the mol of 6mA (inter- or extrapolated from standard curve,) per total mol of deoxyadenosine (dA + 6mA, Figure 2.1).

Table 2.3 List of oligonucleotides for HPLC-QQQ

Oligonucleotide	Abbrev.	Sequence (5’ → 3’)
Unmodified oligo	UO	GGGCAGTACACAGACTATGTTG
Modified oligo1	MO1	C^{5m} GC A^{6m} TA
Modified oligo2	MO2	CGC A^{1m} TA

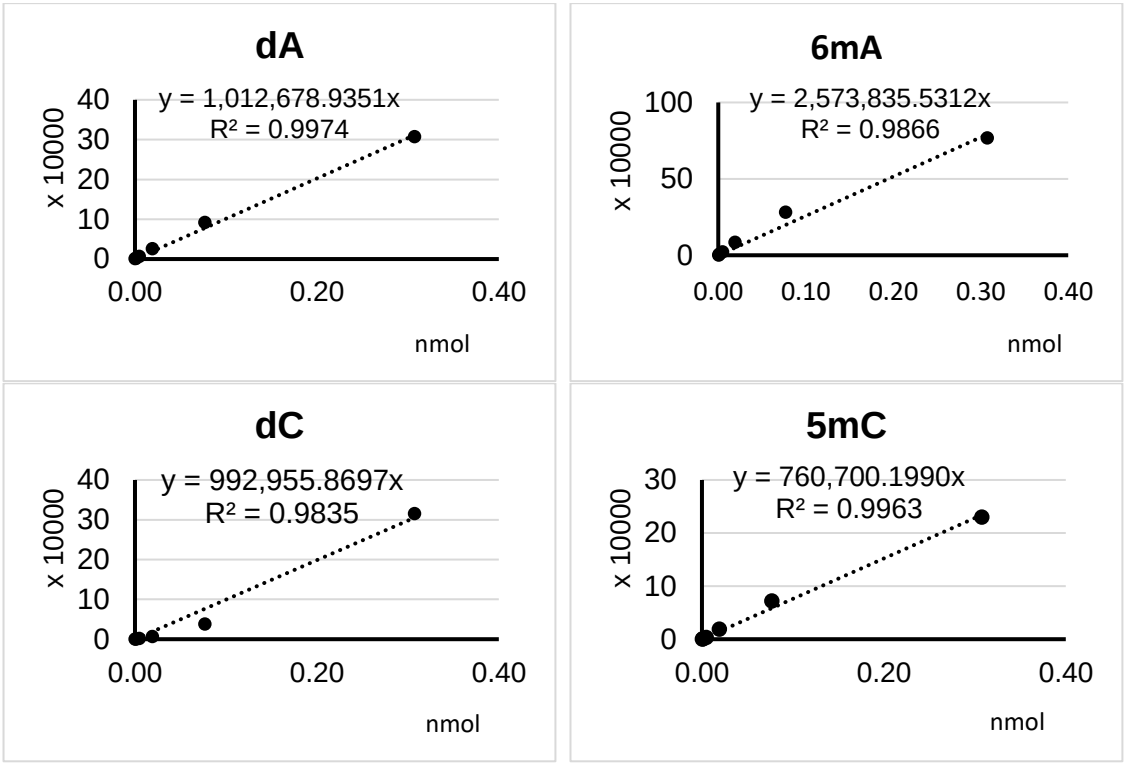


Figure 2.1 External standard curves of respective nucleotides from samples.

Y-axes are the intensity of the peak of the chromatogram in arbitrary unit. R^2 > 95% indicate the linearity relationship between X-axes (nmol of respective nucleotides) and Y axes. There are 15 run of HPLC-MS/MS and each run have a set of external standard curve and samples mol were intra-extrapolated to the standard curve from the same run, and this figure is representing one of the run (Run14, see Supplementary Document 1)

Additionally, unmodified oligo (Table 2.3) was used as negative control, which is also digested, ultra-filtered, and injected to the HPLC-QQQ system as the protocol applied to the DNA samples. Furthermore, I mixed unmodified and modified oligo 1 to create a series of concocted standard as described in Table 2.4. These concoction of known standard solutions then used as an extra control to confirm the quality of the protocol and system I performed to detect DNA methylation on early diverging animals.

Table 2.4 List of known concocted standard solutions

Known Standard Solutions	Unmodified oligo (nmol)	Modified oligo1 (nmol)	dA (nmol)	6mA (nmol)	dC (nmol)	5mC (nmol)	6mA/dA (% mol/mol)	5mC/dC (% mol/mol)
Solution A	0.33	0.01	nc	nc	1.670	0.010	nc	0.5988%
Solution B	0.33	0.02	nc	nc	1.690	0.020	nc	1.1834%
Solution C	0.33	0.04	nc	nc	1.730	0.040	nc	2.3121%
Solution D	0.33	0.0003	2.3109	0.0003	nc	nc	0.0130%	nc
Solution E	0.33	0.0006	2.3118	0.0006	nc	nc	0.0260%	nc
Solution F	0.33	0.0012	2.3136	0.0012	nc	nc	0.0519%	nc
Solution G	0.33	0.0023	2.3169	0.0023	nc	nc	0.0993%	nc

nc : not calculated

2.2.6 Searching for and retrieving DNA methylation associated genes from the *Hydractinia symbiolongicarpus* transcriptome

There are 10 genes that act (or have been proposed to act) as "writer" and "eraser" of 5mC and 6mA (see Table 2.5). I retrieved the orthologous sequences from available omics data of the animal model where the functional study of the genes had been reported (see Table 2.5). I translated these cDNA sequences to amino acid sequences if they were not yet in amino acid (aa) sequence form. Then, I used these aa sequences as query to search for the orthologous transcripts from *Hydractinia symbiolongicarpus* by performing tblastn search at https://blast.ncbi.nlm.nih.gov/. The tblastn search was limited, which set by TSA project of 'GAWH: TSA: *Hydractinia symbiolongicarpus*, transcriptiome shotgun assembly'. From the tblastn results page, the top hit *Hydractinia*'s transcript sequences were retrieved. From this nucleotide (nt) sequence, the longest in frame start-stop codon was treated as the open reading frame and most likely the coding sequence for the respective gene from *Hydractinia*.

Table 2.5 List of DNA methylation associated genes

Gene	Animal	Uniprot ID	Functional Confirmation
DNMT1	Human	P26358	(Yen et al. 1992; Rhee et al. 2000; Schermelleh et al. 2007)
Dnmt2	Fruit fly	Q9VKB3	(Goll et al. 2006; Tuorto et al. 2012)
DNMT3a	Human	Q9Y6K1	(Okano et al. 1999; Gowher and Jeltsch 2001; Heyn et al. 2019)
Dnmt3b	Mouse	O88509	(Okano et al. 1999)
mettl4	*C. elegans*	Q09956	(Greer et al. 2015)
Mettl3	Mouse	Q8C3P7	(Liu et al. 2014; Wang et al. 2016; Xu et al. 2017)
Mettl14	Mouse	Q3UIK4	(Liu et al. 2014; Wang et al. 2016)
N6AMT1	Human	Q9Y5N5	(Ratel et al. 2006; Xiao et al. 2018; Woodcock et al. 2019)
N6AMT2	Human	Q8WVE0	(Hamey et al. 2016)
Tet	Fruit fly	M9NEY8	(Zhang et al. 2015; Wakisaka et al. 2019)
Alkbh1	Mouse	P0CB42	(Xie et al. 2018; Tian et al. 2020; Zhang et al. 2020)
alkbh4	*C. elegans*	Q8MNT9	(Greer et al. 2015)
Alkbh5	Mouse	Q3TSG4	(Zheng et al. 2013; Tang et al. 2018)

2.2.7 Multiple Sequence Alignment (MSA)

To infer phylogenetic tree, multiple sequence alignment (MSA) is required, thus I built a dataset of amino acid sequences for each homolog from various species. The list of species I used to build datasets is presented on Table 2.6. I retrieved the sequences of the respective homologs from each species mainly from the uniprot database (www.uniprot.org) and ensembl omics database (https://metazoa.ensembl.org/), which were imported into Geneious Prime 2019.0.4 software (Geneious). Specifically, for *Mnemiopsis leidyi*, *Hydra vulgaris*, *Hydractinia echinata, Salpingoeca rosetta*, *Nematostella vectensis*, *Saccoglossus kowalevskii*, and *Acropora digitifera*, these organisms' transcriptomes were downloaded from their specific respective database (link provided in Table 2.6). Sequences were aligned in Geneious using MAFFT with E-INS-i

algorithm, JTT 100 PAM scoring matrix, and gap penalty of 1.53 (Katoh and Standley 2013).

Table 2.6 List of species and their abbreviation used for MSA

Represented Phyla	Species	Abbrev.	Main Database
Choanoflagellata	*Salpingoeca rosetta*	SROS	uniprot, ensembl
Choanoflagellata	*Capsaspora owczarzaki*	COWC	uniprot, ensembl
Placozoa	*Trichoplax adhaerens*	TADH	uniprot, ensembl
Porifera	*Amphimedon queenslandica*	AQUE	uniprot, ensembl
Cnidaria: Anthozoa	*Nematostella vectensis*	NVEC	NvERTx-ircan
Cnidaria: Anthozoa	*Acropora digitifera*	ADIG	OIST
Cnidaria: Hydrozoa	*Hydra vulgaris*	HVUL	nhgri.nih
Cnidaria: Hydrozoa	*Hydractinia echinata*	HECH	nhgri.nih
Cnidaria: Hydrozoa	*Hydractinia symbiolongicarpus*	HSYM	This study, nhgri.nih
Ctenophora	*Mnemiopsis leidyi*	MLEI	nhgri.nih
Xenacoelomorpha	*Hofstenia miamia*	HMIA	uniprot, ensembl
Ecdysozoa: Arthropoda	*Drosophila melanogaster*	DMEL	uniprot, ensembl
Ecdysozoa: Nematoda	*Caenorhabditis elegans*	CELE	uniprot, ensembl
Lophotrochozoa	*Capitella teleta*	CTEL	uniprot, ensembl
Lophotrochozoa	*Crassostrea gigas*	CGIG	uniprot, ensembl
Hemichordata	*Saccoglossus kowalevskii*	SKOW	OIST
Chordata: … : Tunicata	*Ciona intestinalis*	CINT	uniprot, ensembl
Chordata: … : Teleostei	*Danio rerio*	DRER	uniprot, ensembl
Chordata: … : Amphibia	*Xenopus laevis*	XLAE	uniprot, ensemble
Chordata: … : Amphibia	*Xenopus tropicalis*	XTRO	uniprot, ensemble
Chordata: … : Mammalia	*Mus musculus*	MMUS	uniprot, ensemble
Chordata: … : Mammalia	*Homo sapiens*	HSAP	uniprot, ensembl

2.2.8 Phylogenetic Tree Inference

The phylogenetic trees are built from gap-free (manually edited) multiple sequence alignments in three ways. Firstly, phylogenetic trees are inferred by neighbour-joining method (Saitou and Nei 1987) using Jukes-Cantor distance model, then a consensus tree is built by 10,000 bootstrap resampling with 60% support values thresholds. Secondly, a phylogenetic tree is built by RAxML 8.2.11 (Stamatakis 2014) using GAMMA LG protein model (default), rapid bootstrapping (10,000 replicates) and search for best-scoring maximum likelihood tree algorithm. Thirdly, a Bayesian phylogenetic tree was produced using MrBayes v.3.2.2. The program was run using a fixed WAG substitution model with gamma distributed rate variation across sites ("lset rates=gamma") with four chains for 4 million generations. The run was sampled every 500th generation and analysed with a 20% burnin. These three methods of phylogenetic tree inference are available in Geneious. The consensus tree from neighbor-joining analysis then exported and manually edited in Microsoft Publisher to display the support values from the two different methods of phylogenetic inference on each strongly supported node (>85% support values) which presented as figures in this chapter. The unedited trees are available in Appendix (see page 114-117).

2.2.9 Localization Signal, Protein Domain and Catalytic Site Analysis

Homologs from *Hydractinia symbiolongicarpus* were analysed for signal peptide by SignalP 5.0 (Nielsen et al. 2019), target peptide by TargetP 2.0 (Almagro Armenteros et al. 2019), protein sorting in general by Wolf Psort (Horton et al. 2007) and for nuclear localisation signals by cNLS Mapper (Kosugi et al. 2009). The results retrieved and imported to Microsoft Excel for data visualization. Protein domain structures were searched and predicted using InterProScan, embedded in Geneious, (Mitchell et al. 2018). Annotated domains then compared in the context of multiple sequence alignment, and an animated representation was created as figures in this chapter. The raw unedited but InterProScan annotated MSA is presented in Appendix (see page 122&127).

2.3 The presence of methylated DNA in *Hydractinia* and early diverging animals

To detect 6mA in the DNA of *Hydractinia*, I validated four available anti-m6A/6mA antibodies. These antibodies were designed to recognize m6A in RNA but they have been used to detect 6mA of DNA as well previously (Fu et al. 2015; Greer et al. 2015; Zhang et al. 2015; Wu et al. 2016; Mondo et al. 2017; Xiao et al. 2018). Nonetheless, I validate the specificity of these antibodies against negative control (unmodified oligo), positive 6mA control (modified oligo 1) and 1mA control (modified oligo 2) by dot blot alongside a total DNA extracted from adults specimen of *Hydractinia symbiolongicarpus* and *Nematostella vectensis* (Figure 2.2). As expected, rabbit polyclonal anti-m6A/6mA (Synaptic Systems #202003) exhibited specificity towards 6mA, while mouse monoclonal anti-m6A/6mA (Synaptic Systems #202111) displayed reactivity against all DNA fragments, regardless of the modification (Figure 2.2b).

During late 2017, I attended the 1st Nucleic Acid Modification Symposium in Mainz, Germany. In that symposium, I listened to the presentation of Franziska Weichmann from Universität Regensburg, displaying the superiority of the anti-m6A that she produced. Thus, I asked for an aliquot of her anti-m6A to test it by dot blot. However, the two anti-m6A she has provided did not prove to be any better than the rabbit-polyclonal anti-m6A (Synaptic Systems #202003, Figure 2.2a). Hence, I decided to keep using the latter antibody for the rest of my project. Moreover, this anti m6A/6mA (SS#202003) remains the main antibody used in multiple publications on m6A/6mA studies across different kingdoms.

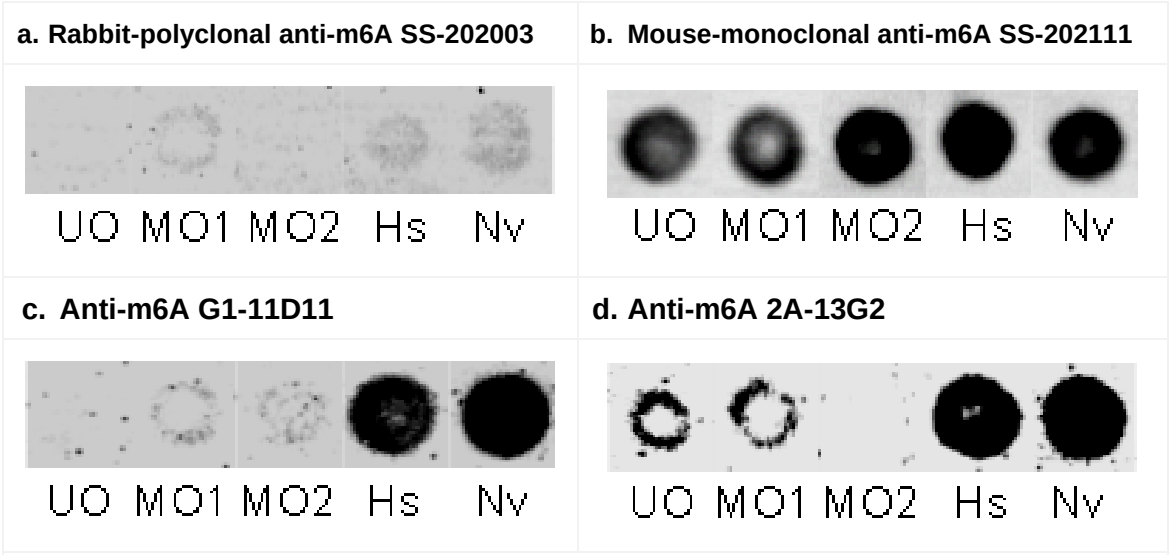


Figure 2.2 Anti-m6A/6mA Dot-Blot.

Spots are each 200 ng DNA. UO: Unmodified oligo, MO1: concocted solution from unmodified oligo and modified oligo1 (6mA/dA (0.1%)), MO2: concocted solution from unmodified oligo and modified oligo2 (1mA/dA (0.1%)), Hs: *Hydractinia symbiolongicarpus*, Nv: *Nematostella vectensis*.

The dot blot with the SS-202003 antibody showed the presence of 6mA in the DNA originated from adult specimens of *Hydractinia* and *Nematostella* (Figure 2.2a). Finally, I also validated the method by detecting 6mA from bacterial plasmid in a concentration dependent manner, as well as detecting 5mC from *Hydractinia* (Figure 2.3). These results from dot blot experiments provide evidence for the presence of DNA methylation (5mC and 6mA) in *Hydractinia symbiolongicarpus* and *Nematostella vectensis*. However, there are still possibilities of modified RNA or bacterial DNA contamination, despite of extensive RNase treatments and washes by filter-sterilized sea water.

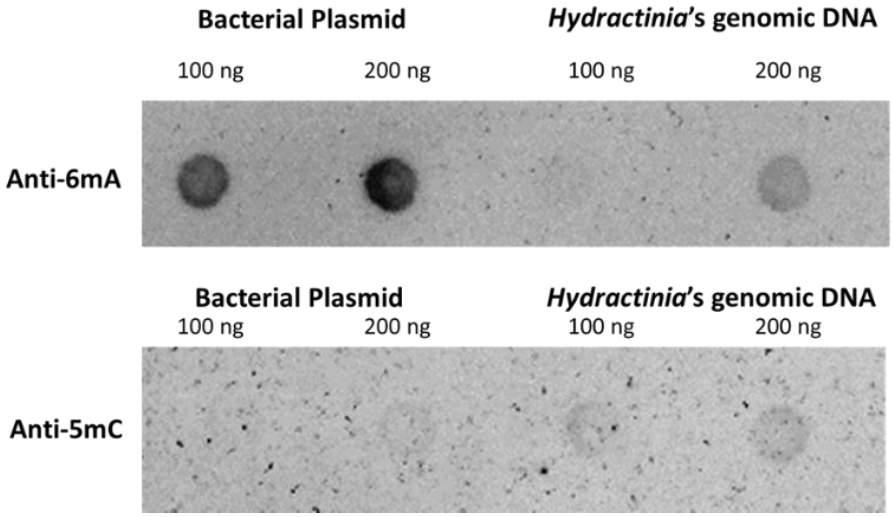
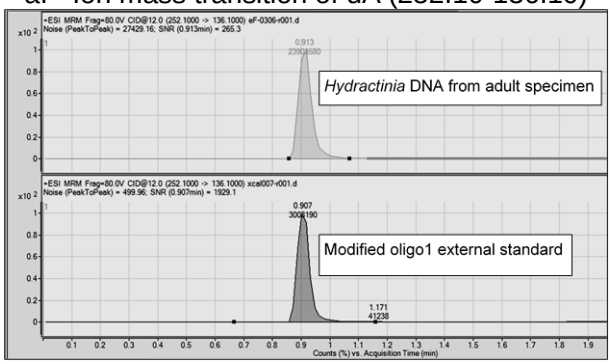


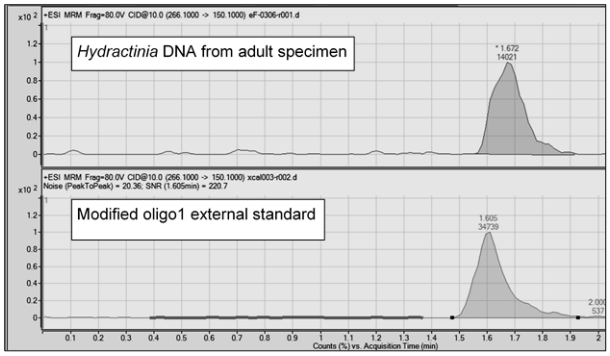
Figure 2.3 anti-m6A/6mA and anti-5mC dot blot of bacterial plasmid and *Hydractinia* genomic DNA.

To quantify 5mC and 6mA levels from DNA, I have established an HPLC-QQQ protocol and validated it in two ways. First, as commonly used, I validated the retention time of individual ion mass transition between the external DNA standards and DNA from biological samples (Figure 2.4).

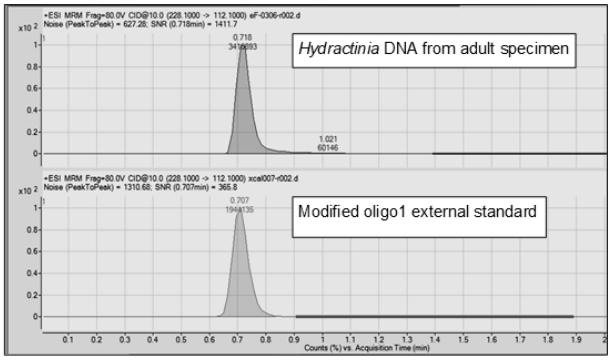
a. Ion mass transition of dA (252.10-136.10)



b. Ion mass transition of 6mA (266.10-150.10)



c. Ion mass transition of dC (228.10-112.10)



d. Ion mass transition of 5mC (242.10-126.10)

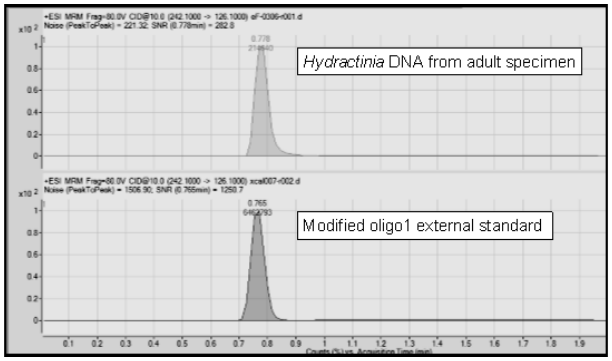


Figure 2.4 Chromatogram of dA, 6mA, dC and 5mC from HPLC-QQQ.

Second, I validated the methods by concocting a standard solution with known concentrations of 6mA/dA and 5mC/dC (Table 2.7). I found that HPLC-QQQ could detect the levels of 6mA/dA and 5mC/dC in reasonable precision; however, the method is apparently unable to detect 6mA/dA below ~0.02%. Unmodified oligo concocted from known standard solutions and external standards were treated like the biological DNA samples.

Table 2.7 HPLC-QQQ validation by known standard solutions

Known Standard Solutions	Concocted 6mA/dA (% mol/mol)	Concocted 5mC/dC (% mol/mol)	Detected 6mA/dA (% mol/mol)	Detected 5mC/dC (% mol/mol)
Solution A	nc	0.5988%	nc	0.6892%
Solution B	nc	1.1834%	nc	1.4463%
Solution C	nc	2.3121%	nc	2.9031%
Solution D	0.0130%	nc	undetected	nc
Solution E	0.0260%	nc	0.0224%	nc
Solution F	0.0519%	nc	0.0534%	nc
Solution G	0.0993%	nc	0.1030%	nc

nc : not relevant for calculation

I, then, used this established HPLC-QQQ system to detect DNA methylation in four species of early diverging animals: *Hydractinia symbiolongicarpus, Hydractinia echinata, Nematostella vectensis* and *Mnemiopsis leidyi*. I found similar levels of 5mC/dC in *Nematostella* (~1.8%) as previously reported (Figure 2.5b, (Zemach et al. 2010; de Mendoza et al. 2019)) indicating the robustness of the method used to detect 5mC/dC in this thesis. I also found 6mA/dA in early diverging animals (Figure 2.5a), but the levels are close to background (6mA/dA levels of unmodified oligo, Figure 2.5a). This indicates that 6mA are extremely scarce in the genome of adult specimens of *Hydractinia*.

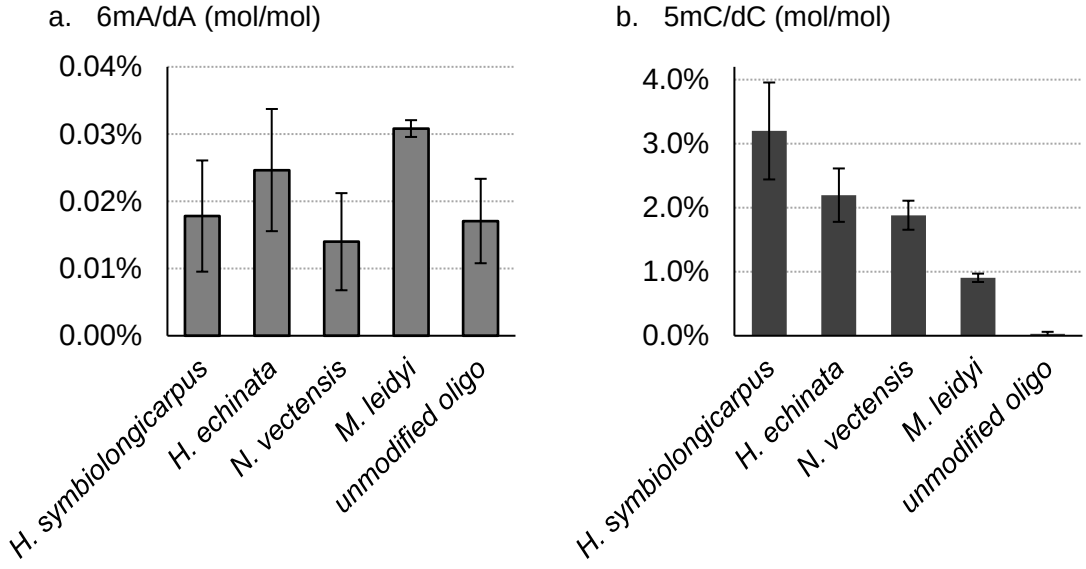


Figure 2.5. Quantification of DNA methylation from adult individuals of early diverging animals by HPLC-QQQ.

2.4 The writers and erasers of DNA methylation in early diverging animals.

Apart from detecting the actual methylated DNA from the genome of early diverging animals, discovering the methyltransferases (writers) and oxidoreductases that initiates demethylation (erasers) of DNA in their genetic repertoire would provide an indication of their functionalities. Thus, I searched for the writers and erasers of 5mC and 6mA among early diverging animals through publicly available transcriptomes, built multiple sequence alignments, inferred their phylogenetic relationship with their potential homologs from more complex animals, and whenever possible, I aligned the essential protein domain and catalytic sites as comparison to the known functional homologs. Lastly, I assessed the possibility of these enzyme to localize to the nucleus or some other cellular compartments.

Most bilaterian animal models have three DNMTs. DNMT1 methylates deoxycytidine on the hemi methylated CpG at replication fork. DNMT3 is more associated with *de novo,* replication-independent 5mC methylation. DNMT2, however, is confirmed to work on tRNA 5mC methylation in mammals (Goll et al. 2006). I utilized mainly the conserved DNA methylation protein domain for multiple sequence alignment (Figure 2.8 and Supplementary_Document_2), to build phylogenetic trees of DNMTs. DNMT phylogeny showed that DNMT1 and 3 of early diverging animals cluster with their respective mammalian orthologs with strong support using two different algorithms (neighbour-joining and maximum likelihood, Figure 2.6). However, no DNMT1 and 3 homologs can be found from unicellular cousins of animals nor from placozoans. This is consistent with previous reports (de Mendoza et al. 2019; de Mendoza et al. 2020).

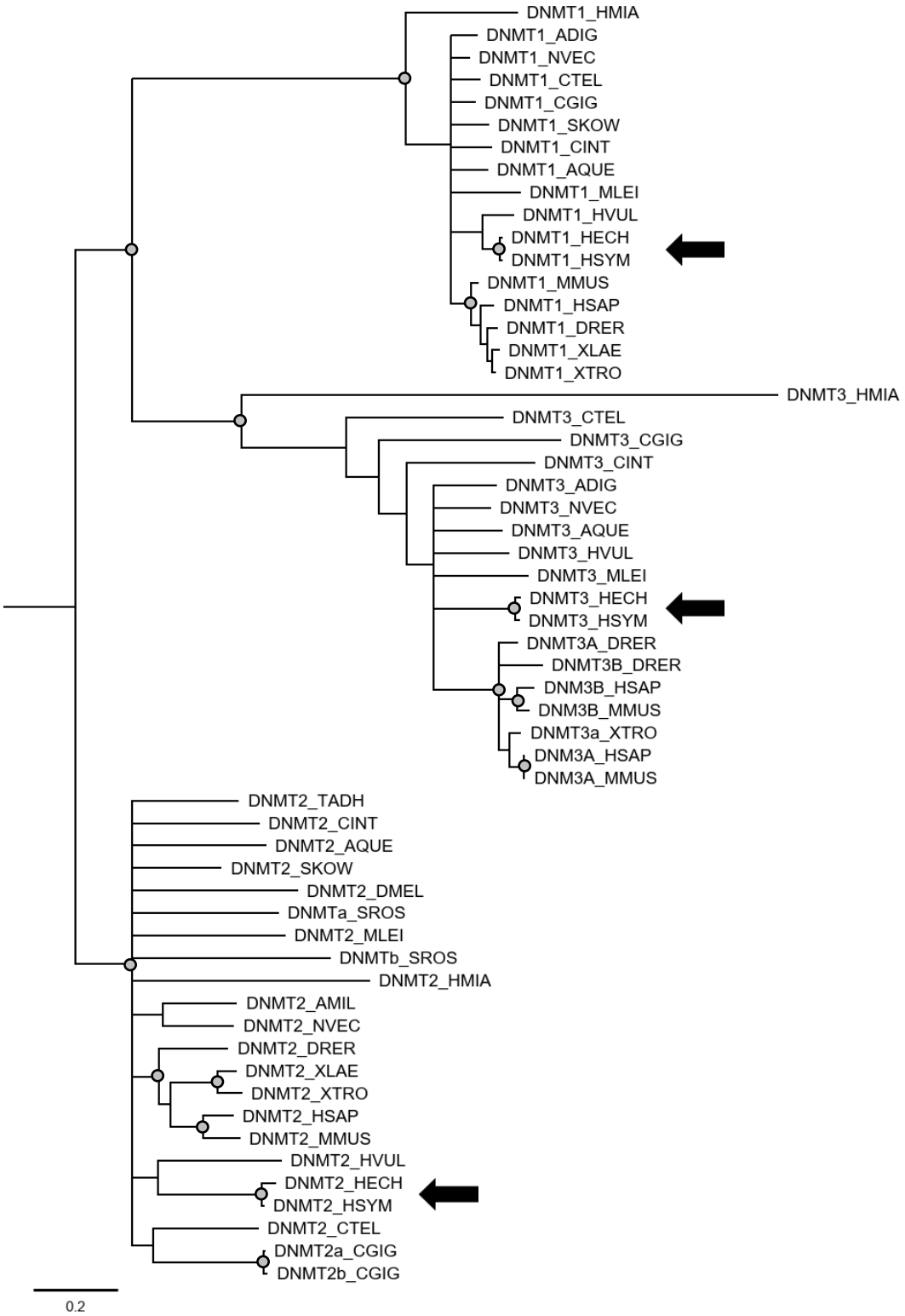


Figure 2.6. Molecular phylogeny of DNMTs

The tree is a neighbour-joining consensus phylogenetic tree with 65% support threshold, thus nodes with below 65% support values were collapsed. The nodes where strong supports from neighbour-joining (>85%), maximum likelihood (>85%), and Bayesian inference posterior probability (>0.95) are marked with grey circle. The raw unedited trees from both methods are presented in Appendix (see page 110-111). The MSA of DNMTs is provided in Supplementary_Document_2 *Hydractinia* Dnmt homologs are pointed by black arrows. The abbreviation of the species were described in Table 2.6

However, hydrozoans' Dnmt1 has lost the DMAP1-binding domain, while anthozoans' Dnmt1 possess the DMAP1-binding domain as the mammalian DNMT1 (Figure 2.8). The loss of DMAP1 binding domain has been reported as well from sponges and ctenophores (de Mendoza et al. 2019). This DMAP1-Binding domain allows DNMT1 to interact to DMAP1 (a transcriptional corepressor by interacting with histone deacetylase 2, (Rountree et al. 2000)) that plays important roles in DNA double strand break repair (Negishi et al. 2009; Lee et al. 2010). *Hydractinia* and anthozoan genomes and transcriptomes contain *Dmap1* with no indication of domain loss compared to their vertebrate's homologs. Mouse oocytes, however, contain two type of Dnmt1 derived from two mRNA isoforms. One of the two is Dnmt1o is truncated without the DMAP1-binding domain like hydrozoans' Dnmt1. However, Dnmt1o can still interact with Dmap1 and exhibits essential roles in embryonic development (Mohan et al. 2011). Therefore, it would be interesting to investigate the consequences of this DMAP1-binding domain loss in hydrozoans.

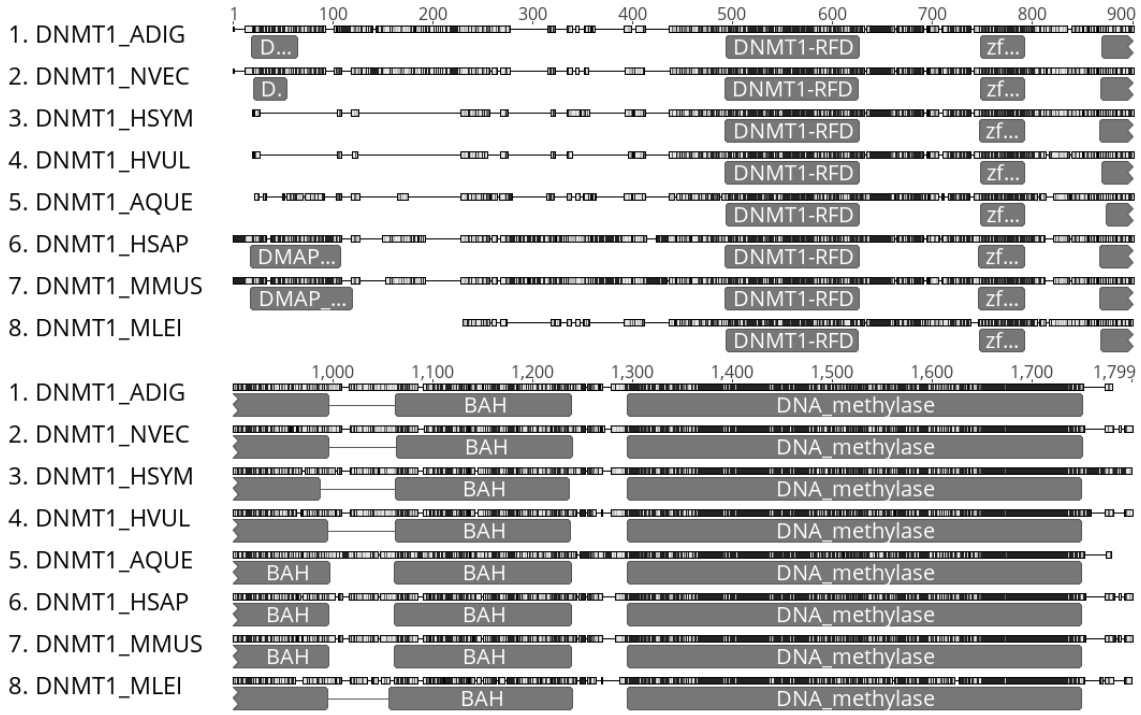


Figure 2.7. Eight sequences of DNMT1 MAFFT alignment with annotation from InterProScan.

Sequence name are described in Table 2.6. DMAP1-Binding domain were undetected from DNMT1_HVUL, _HSYM, _AQUE, and _MLEI. Detail alignment and domain Prosite/Pfam ID are provided in Appendix (see page 122). The abbreviation of the species were described in Table 2.6.

In addition, hydrozoans' Dnmt3 contain an N-terminus extension with an extra PWWP domain, different from the mammalian and anthozoan DNMT3 (Figure 2.8). This is possibly a hydrozoan-specific extension, which has never been reported in other animals

and I found it to be conserved between *Hydractinia* and *Hydra*. PWWP domain in DNMT3a and b in mammals are required for their localization onto satellite repeats at pericentric heterochromatin (Chen et al. 2004). However, A ChIP-Seq experiment where the satellite repeats were masked and removed indicate that DNMT3b selectively bound to gene bodies of active genes, and this binding requires the PWWP domain interaction with histone H3 lysine 36 trimethylation (H3K36me3) (Baubec et al. 2015). Interestingly, PWWP domain of DNMT3a interaction with H3K36me3 remains disputable. A study in *Exaiptasia,* an anthozoan, indicates strong association between H3K36me3 ChIP with gene body DNA (5mC) methylation as well as low level of intragenic transcripts (Li et al. 2018b). Nonetheless, how the extra PWWP domain in hydrozoans affects this interaction is still unknown. Thus, these distinct architectures of Dnmt1 and 3 in *Hydractinia* are, likely, a hydrozoan trait but its physiological function remains unknown.

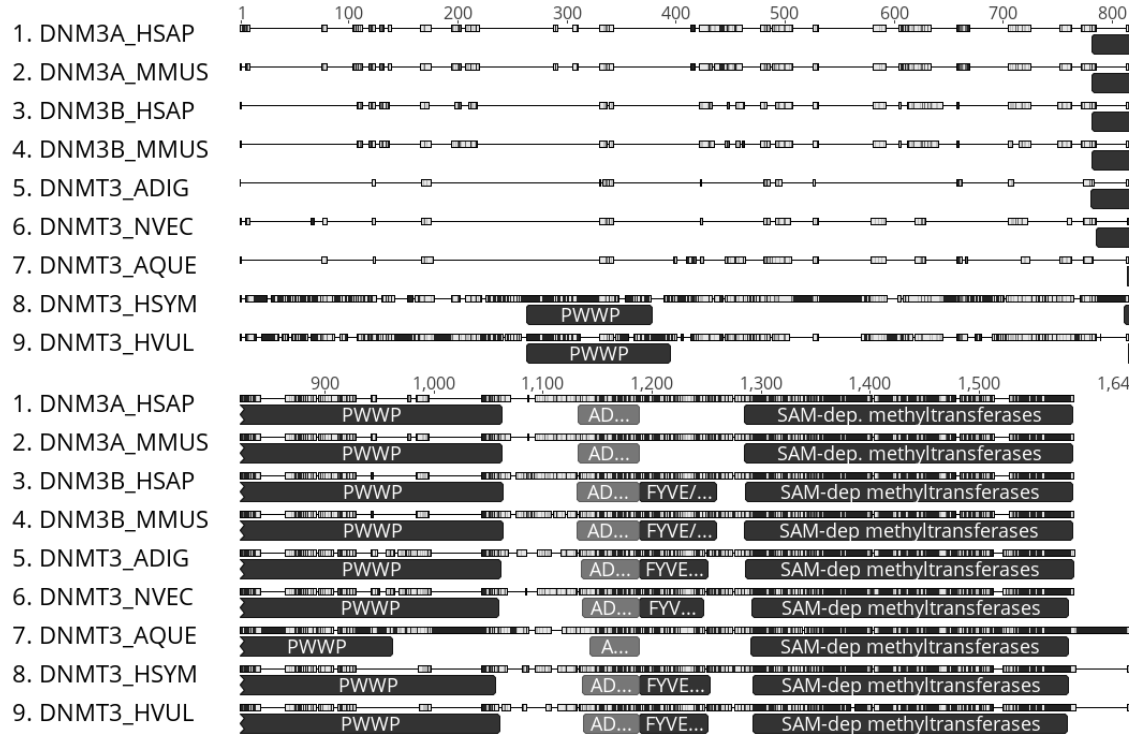


Figure 2.8. Nine sequence of DNMT3 MAFFT alignment with annotation from InterProScan.

Sequence name are described in Table 2.6. Extra PWWP domains were detected from DNMT3_HVUL and DNMT3_HSYM. Detail alignment and domain Prosite/Pfam ID are provided in Appendix (see page 127). The abbreviation of the species were described in Table 2.6.

The demethylation of 5mC in mammals is initiated by TET1, TET2 and TET3. Invertebrates possess only one homolog of TET (Figure 2.9). However, TET_DMEL has been described as the 6mA (instead of 5mC) demethylation initiator (Zhang et al. 2015). Despite the undetected TET homologs from choanoflagellates, all early diverging animals possess TET homologs with strong support in the node clustering all

invertebrates' TET homologs together, separated from the vertebrates homologs (Figure 2.9).

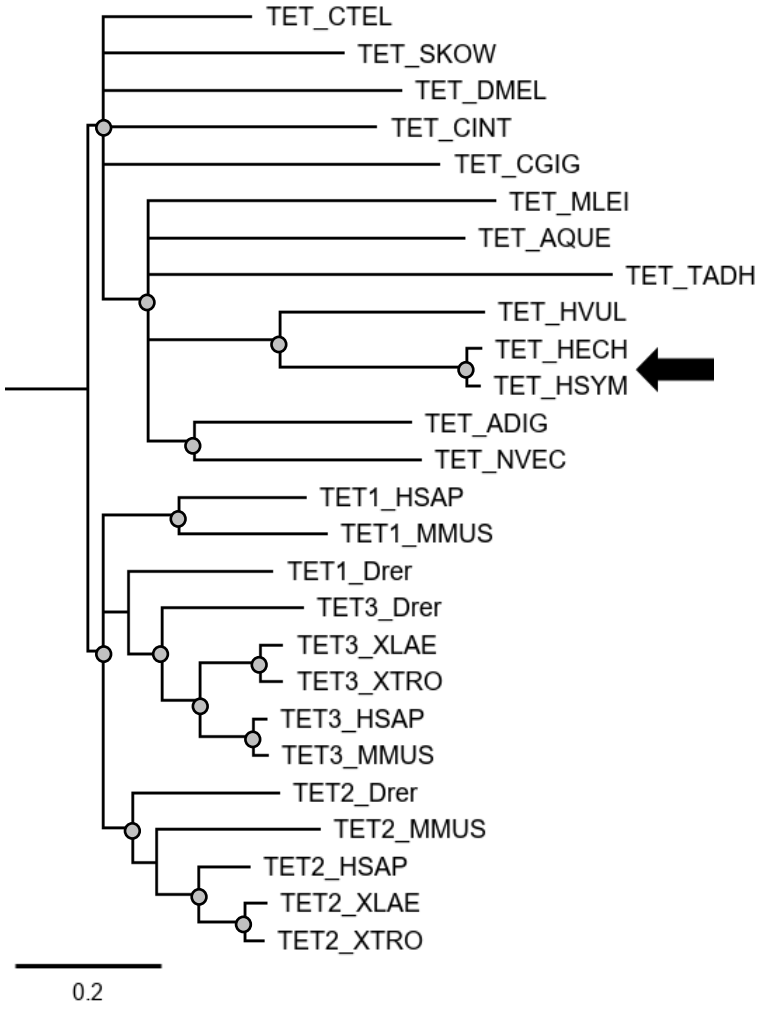


Figure 2.9. Molecular Phylogeny of TETs

The tree represents a neighbour-joining consensus with 65% support threshold, thus clades supported below 65% were collapsed. The nodes where strong supports from neighbour-joining (>85%), maximum likelihood (>85%), and Bayesian inference posterior probability (>0.95) are marked with grey circle. The raw unedited trees from both methods are presented in Appendix (see page 114-118) The MSA of TET is provided in Supplementary_Document_3. *Hydractinia* Tet homologs are pointed by black arrows. The abbreviation of the species were described in Table 2.6.

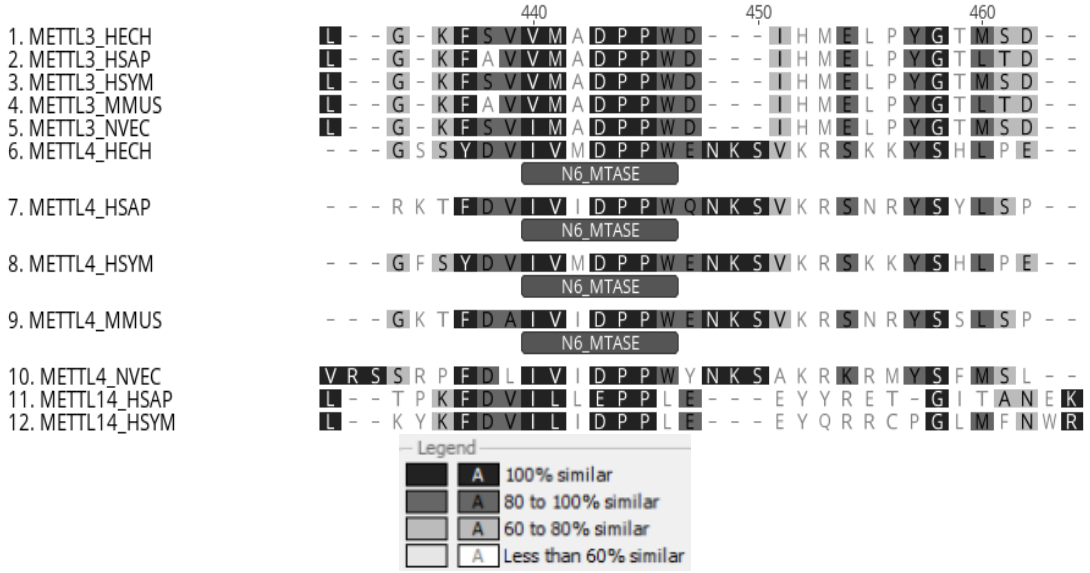


Figure 2.10. Conserved N6-MTase Catalytic Site (DPPW) of METTL3 and 4.

Writers of 6mA have been controversial topics. METTL4 has been reported as the methyltransferase for 6mA in DNA in *C. elegans,* mammalian cell culture, mouse and silkworm (Greer et al. 2015; Wang et al. 2018; Kweon et al. 2019; Hao et al. 2020). Recently, however, the 6mA methylator role of METTL4 has been disputed by studies on fruit fly and mammalian cultured cells, shown to be the RNA methyltransferase for U2 snRNA (Chen et al. 2020; Goh et al. 2020; Gu et al. 2020). METTL3 and METTL14 are the paralogs of METTL4 that methylate mRNA, depositing m6A (Bokar et al. 1997; Liu et al. 2014). Removal of the DPPW catalytic sites abolishes METTL3's and 4's ability to perform methylation on RNA and DNA, respectively (Greer et al. 2015; Wang et al. 2016). I confirmed that early diverging animals, including *Hydractinia*, have METTL4 as well as METTL3 and METTL14 (the m6A methyltransferases) (Figure 2.11). Nevertheless, the catalytic site (DPPW) of all m6A/6mA methyltransferases is conserved across animals (Figure 2.10).

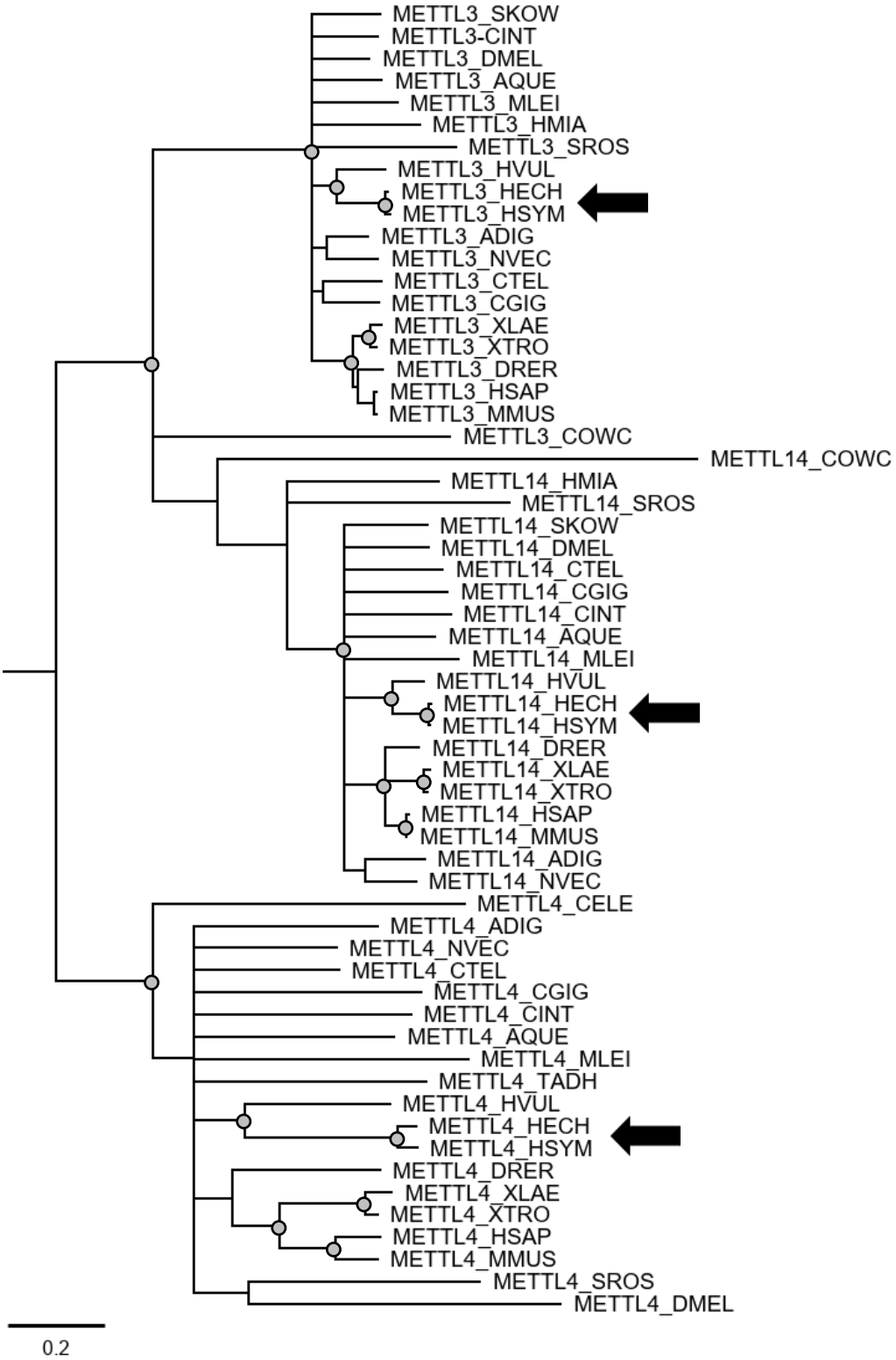


Figure 2.11. Molecular Phylogeny of METTLs

The tree represents a neighbour-joining consensus with 65% support threshold, thus clades supported below 65% were collapsed. The nodes where strong supports from neighbour-joining (>85%), maximum likelihood (>85%), and Bayesian inference posterior probability (>0.95) are marked with grey circle. The raw unedited trees from both methods are presented in Appendix (see page 114-118) The MSA of METTLs is provided in Supplementary_Document_4. *Hydractinia* Mettl homologs are pointed by black arrows. The abbreviation of the species were described in Table 2.6.

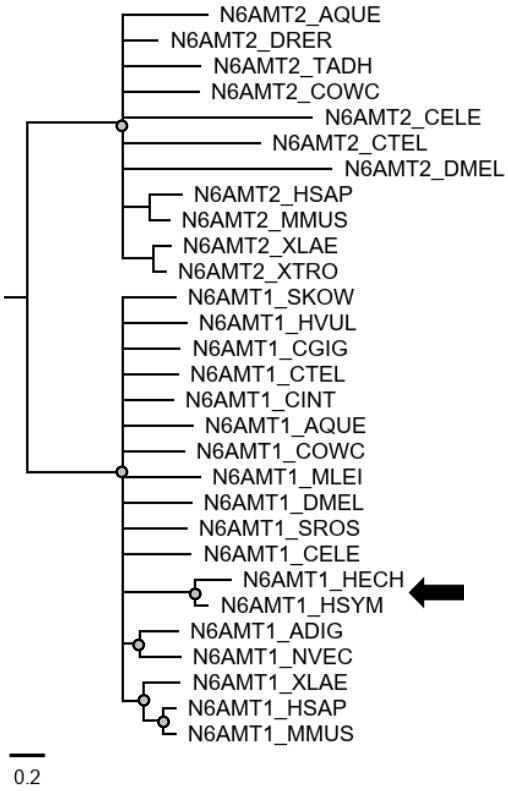


Figure 2.12. Molecular Phylogeny of N6AMT

The tree represents a neighbour-joining consensus with 65% support threshold, thus clades supported below 65% were collapsed. The nodes where strong supports from neighbour-joining (>85%), maximum likelihood (>85%), and Bayesian inference posterior probability (>0.95) are marked with grey circle. The raw unedited trees from both methods are presented in Appendix (see page 114-118) The MSA of N6AMT is provided in Supplementary_Document_5. *Hydractinia* N6amt1s homologs are pointed by black arrows. The abbreviation of the species were described in Table 2.6.

Additionally, N6AMT1 has been proposed as another candidate 6mA methyltransferase (Xiao et al. 2018). *Hydractinia* and other early-diverging animals possess N6AMT1 (Figure 2.12), and the catalytic site (NPPY) responsible for the 6mA activity is conserved in these animals (Figure 2.13). However, recently, the role of N6AMT1 on 6mA has been strongly disputed (Xie et al. 2018; Woodcock et al. 2019).

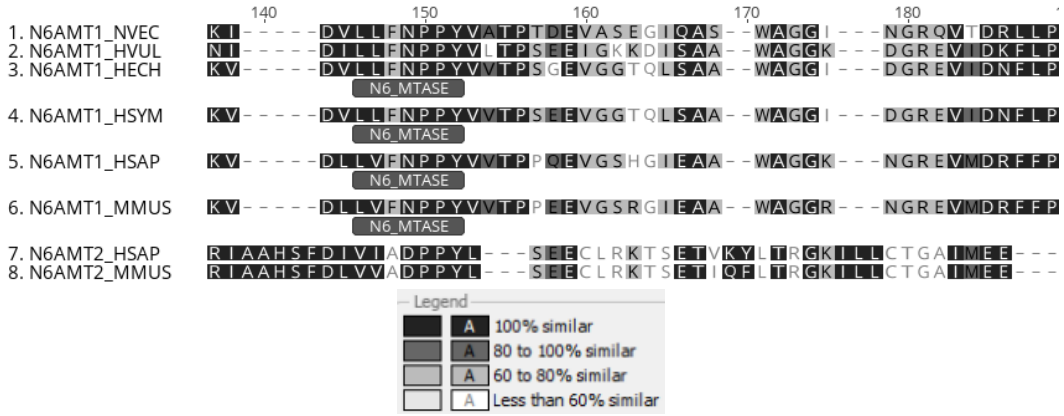


Figure 2.13. Conserved N6-MTase Catalytic Site of N6AMT1

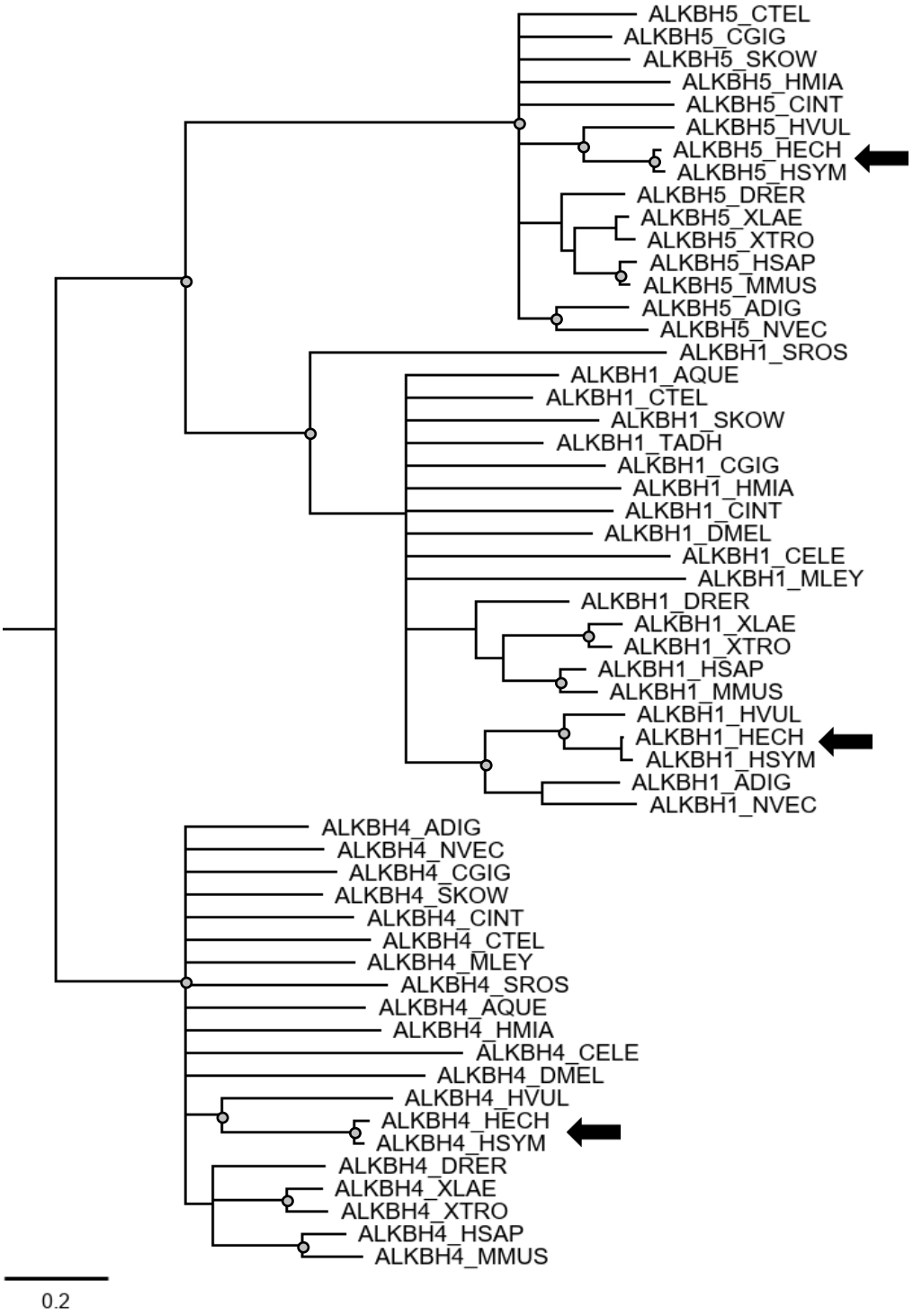


Figure 2.14 Molecular Phylogeny of ALKBHs

The tree represents a neighbour-joining consensus with 65% support threshold, thus clades supported below 65% were collapsed. The nodes where strong supports from neighbour-joining (>85%), maximum likelihood (>85%), and Bayesian inference posterior probability (>0.95) are marked with grey circle. The raw unedited trees from both methods are presented in Appendix (see page 114-118). The MSA of ALKBH is provided in Supplementary_Document_6. *Hydractinia* Alkbh homologs are pointed with black arrows. The abbreviation of the species were described in Table 2.6.

Lastly, ALKBH4 has been reported as the 6mA eraser in *C. elegans* (Greer et al. 2015).

ALKBH1 has been reported as 6mA eraser in mouse and human cells (Wu et al. 2016;

Xiao et al. 2018; Zhang et al. 2020). However, these reports have been disputed by others (Schiffers et al. 2017). Nonetheless, *Hydractinia* and early diverging animals have these two enzymes that belong to the ALKBH family (Figure 2.14). Moreover, recent studies display a strong emphasis on ALKBH1 as the enzyme that removes 6mA from gDNA in mammals (Xie et al. 2018; Tian et al. 2020; Zhang et al. 2020).

2.5 Localization signals of the writers and erasers of DNA methylation

The writers and erasers of 5mC and 6mA must be localized to the nucleus to act on genomic DNA. Enzymes localization is determined by a short peptide within the sequence of the enzymes. Thus, I have analysed all putative DNA methylation writers and erasers sequences of *Hydractinia* for the availability of short peptides that direct them to several subcellular structures. Bioinformatically, there are four established methods to predict these “tag” peptide. The main software was cNLS mapper that map the nuclear localization sequence (NLS), one or a pair short peptides that can be anywhere in the primary sequence of a protein but exposed on the protein surface and recognized by the nuclear transport machineries in the cell and imports the protein into the nucleus (Kosugi et al. 2009). Next, SignalP is a software to detect signal peptides, a short peptide of 10-35 amino acids in the N-terminus of a protein that will direct it to endoplasmic reticulum (Almagro Armenteros et al. 2019). Next, I used TargetP, a software that predicts target peptides, a short peptide (15-40 amino acids) following a signal peptide, thus a bipartite signal, that leads the protein to organelles such as mitochondria or chloroplast after the signal peptide is cleaved inside the reticulum endoplasm (Almagro Armenteros et al. 2019). Lastly, Wolf-PSORT is a software that generally maps similarity of an enquired protein with a list of proteins with known localization (Horton et al. 2007). I used signalP and targetP to exclude the possibility that any of the six enzymes associated with DNA methylation in this thesis is targeted to the secretory pathways, mitochondria, or any other organelles apart from nucleus.

All enzymes displayed strong signalP and targetP results towards “other”. This means that the proteins would be targeted into organelles other than the endoplasmic reticulum, mitochondria, or peroxisomes. From cNLS mapper, all but N6amt1 displayed a high score of nuclear localisation signal (>6.8). N6amt1 is also predicted by Wolf-PSORT to be a protein that is targeted towards cytoplasm and cytoskeleton. Wolf-PSORT pointed *Hydractinia* Dnmt1 and Alkbh1 as being cytoplasmic but they also displayed cytoplasmic-nuclear features as the second-best prediction. This was also the case for Alkbh4. Thus, apart from N6amt1, all other enzymes are predicted to be targeted to the nucleus through the cytoplasm (Table 2.8).

Table 2.8 Signal Peptide and Protein Targeting Prediction

Enzyme	cNLS score	SignalP/TargetP Prediction score for "other"	1st PSORT	1st PSORT score	2nd PSORT	2nd PSORT score
DNMT1	8.3	0.999/1.000	Cytoplasmic	17.5	Cytoplasmic-Nuclear	14.67
DNMT3	7.5	0.999/0.999	Nuclear	16.5	Cytoplasmic-Nuclear	15.5
TET	6.8	0.999/0.956	Nuclear	18	Plasma membrane	8
METTL4	7.3	0.998/0.999	Nuclear	19	Cytoplasmic-Nuclear	16.5
N6AMT1	0	0.998/1.000	Cytoskeleton	15	Cytoplasmic	7
ALKBH1	8.5	0.999/0.999	Cytoplasmic	23	Cytoplasmic-Nuclear	17
ALKBH4	9.5	0.995/1.000	Extracellular	15	Cytoplasmic-Nuclear	8

The cNLS mapper score is mark from 0 to 10 for the enquired protein sequence contain NLS with 10 indicating the most likelihood to have NLS. SignalP/TargetP prediction of "others" is a score from 0.0 to 1.0 for the enquired protein sequence containing no signal or target peptide, thus predicted to be directed to "other" than mitochondria, chloroplast, thylakoid luminal, or peroxisome within the cell, with 1.0 indicating the most confident prediction. 1st PSORT is the first (most likely) localization prediction made by Wolf-PSORT. The 1st PSORT score indicates the number of proteins (with known localization) considered similar to the enquired protein within the 1st prediction. 2nd PSORT is the second-best localization prediction by Wolf-PSORT.

2.6 Adult specimens of early diverging animals have 5mC but their 6mA level is close to background level

The presence of 5mC writers and erasers are aligned with detected 5mC in the total DNA of early diverging animals, consistent with previous reports (de Mendoza et al. 2019). Interestingly, I have discovered distinct domains in DNMT1 and 3 of hydrozoans but their function is yet unknown. The presence of 6mA writers and erasers candidates - especially *Mettl4*, *Alkbh1*, and *Alkbh4* - is consistent with the presence of 6mA in adult early diverging animals; however, the results of the HPLC-QQQ analysis revealed only very low levels of 6mA in these animals (Figure 2.5). 6mA was detected in dot blot experiment (Figure 2.2-Figure 2.3), demonstrating the sensitivity of this method and, most likely, the reliability of the anti-6mA I used. Low levels of 6mA/dA have also been reported in other animals (Greer et al. 2015; Zhang et al. 2015; Liu et al. 2016b; Wu et al. 2016; Schiffers et al. 2017; Xiao et al. 2018; Hao et al. 2020). Higher levels of 6mA/dA in animals - ~0.15% inzebrafish (Liu et al. 2016b) and ~0.08% in fruit fly (Zhang et al. 2015) - have been reported previously. These, however, were detected during early embryogenesis while during adulthood, most studied animals display an extremely low level of 6mA. My data suggests that 6mA level from adult specimens of *Nematostella vectensis, Mnemiopsis leidyi, Hydractinia echinata* and *Hydractinia symbiolongicarpus* are close to the background level. Thus, measuring the 6mA level of *Hydractinia* during

early embryogenesis are necessary to move forward in this investigation, as will be described in the next chapter.

2.7 References

Achwal, Chhaya W, Chitra A Iyer, and H Sharat Chandra. 1983. 'Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, *Drosophila* and mealybug DNA', *FEBS letters*, 158: 353-358.

Almagro Armenteros, Jose Juan, Marco Salvatore, Olof Emanuelsson, Ole Winther, Gunnar von Heijne, Arne Elofsson, and Henrik Nielsen. 2019. 'Detecting sequence signals in targeting peptides using deep learning', *Life Science Alliance*, 2: e201900429.

Arber, Werner, and Daisy Dussoix. 1962. 'Host specificity of DNA produced by *Escherichia coli*: I. Host controlled modification of bacteriophage λ', *Journal of Molecular Biology*, 5: 18-36.

Bartolomei, Marisa S, Andrea L Webber, Mary E Brunkow, and Shirley Marie Tilghman. 1993. 'Epigenetic mechanisms underlying the imprinting of the mouse H19 gene', *Genes & development*, 7: 1663-1673.

Baubec, Tuncay, Daniele F. Colombo, Christiane Wirbelauer, Juliane Schmidt, Lukas Burger, Arnaud R. Krebs, . . . Dirk Schübeler. 2015. 'Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation', *Nature*, 520: 243-247.

Bird, Adrian P, and Edwin M Southern. 1978. 'Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from *Xenopus laevis*', *Journal of Molecular Biology*, 118: 27-47.

Bokar, JA, ME Shambaugh, D Polayes, AG Matera, and FM Rottman. 1997. 'Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase', *Rna*, 3: 1233-1247.

Capuano, Floriana, Michael Mülleder, Robert Kok, Henk J. Blom, and Markus Ralser. 2014. 'Cytosine DNA Methylation Is Found in *Drosophila melanogaster* but Absent in *Saccharomyces cerevisiae, Schizosaccharomyces pombe,* and Other Yeast Species', *Analytical Chemistry*, 86: 3697-3702.

Chen, Hao, Lei Gu, Esteban A. Orellana, Yuanyuan Wang, Jiaojiao Guo, Qi Liu, . . . Yang Shi. 2020. 'METTL4 is an snRNA m6Am methyltransferase that regulates RNA splicing', *Cell Research*.

Chen, Taiping, Naomi Tsujimoto, and En Li. 2004. 'The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin', *Molecular and Cellular Biology*, 24: 9048-9058.

Dabe, Emily C, Rachel S Sanford, Andrea B Kohn, Yelena Bobkova, and Leonid L Moroz. 2015. 'DNA methylation in basal metazoans: Insights from ctenophores', *Integrative and comparative biology*, 55: 1096-1110.

de Mendoza, Alex, William L. Hatleberg, Kevin Pang, Sven Leininger, Ozren Bogdanovic, Jahnvi Pflueger, . . . Ryan Lister. 2019. 'Convergent evolution of a vertebrate-like methylome in a marine sponge', *Nature Ecology & Evolution*, 3: 1464-1473.

de Mendoza, Alex, Ryan Lister, and Ozren Bogdanovic. 2020. 'Evolution of DNA methylome diversity in eukaryotes', *Journal of Molecular Biology*, 432: 1687-1705.

Delatte, Benjamin, Fei Wang, Long Vo Ngoc, Evelyne Collignon, Elise Bonvin, Rachel Deplus, . . . François Fuks. 2016. 'Transcriptome-wide distribution and function of RNA hydroxymethylcytosine', *Science*, 351: 282-285.

Dunn, D. B., and J. D. Smith. 1958. 'The occurrence of 6-methylaminopurine in deoxyribonucleic acids', *The Biochemical journal*, 68: 627-636.

Frommer, M, L E McDonald, D S Millar, C M Collis, F Watt, G W Grigg, . . . C L Paul. 1992. 'A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands', *Proceedings of the National Academy of Sciences*, 89: 1827-1831.

Fu, Lijuan, Candace R Guerrero, Na Zhong, Nicholas J Amato, Yunhua Liu, Shuo Liu, . . . Laura J Niedernhofer. 2014. 'Tet-mediated formation of 5-hydroxymethylcytosine in RNA', *Journal of the American Chemical Society*, 136: 11582-11585.

Fu, Ye, Guan-Zheng Luo, Kai Chen, Xin Deng, Miao Yu, Dali Han, . . . Louis C Doré. 2015. 'N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*', *Cell*, 161: 879-892.

Geier, Gail E, and Paul Modrich. 1979. 'Recognition sequence of the dam methylase of *Escherichia coli* K12 and mode of cleavage of Dpn I endonuclease', *Journal of Biological Chemistry*, 254: 1408-1413.

Goh, Yeek Teck, Casslynn W. Q. Koh, Donald Yuhui Sim, Xavier Roca, and W. S. Sho Goh. 2020. 'METTL4 catalyzes m6Am methylation in *U2 snRNA* to regulate pre-mRNA splicing', *bioRxiv*: 2020.2001.2024.917575.

Goll, Mary Grace, Finn Kirpekar, Keith A. Maggert, Jeffrey A. Yoder, Chih-Lin Hsieh, Xiaoyu Zhang, . . . Timothy H. Bestor. 2006. 'Methylation of $tRNA^{Asp}$ by the DNA Methyltransferase Homolog Dnmt2', *Science*, 311: 395-398.

Gowher, Humaira, and Albert Jeltsch. 2001. 'Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites', *Journal of Molecular Biology*, 309: 1201-1208.

Greer, Eric Lieberman, Mario Andres Blanco, Lei Gu, Erdem Sendinc, Jianzhao Liu, David Aristizábal-Corrales, . . . Yang Shi. 2015. 'DNA methylation on N6-adenine in *C. elegans*', *Cell*, 161: 868-878.

Grippo, P., M. Iaccarino, E. Parisi, and E. Scarano. 1968. 'Methylation of DNA in developing sea urchin embryos', *Journal of Molecular Biology*, 36: 195-208.

Gu, Lei, Longfei Wang, Hao Chen, Jiaxu Hong, Zhangfei Shen, Abhinav Dhall, . . . Yang Shi. 2020. 'CG14906 (*mettl4*) mediates m6A methylation of U2 snRNA in *Drosophila*', *bioRxiv*: 2020.2001.2009.890764.

Hamey, Joshua J, Daniel L Winter, Daniel Yagoub, Christopher M Overall, Gene Hart-Smith, and Marc R Wilkins. 2016. 'Novel N-terminal and lysine methyltransferases that target translation elongation factor 1A in yeast and human', *Molecular & Cellular Proteomics*, 15: 164-176.

Hao, Ziyang, Tong Wu, Xiaolong Cui, Pingping Zhu, Caiping Tan, Xiaoyang Dou, . . . Chuan He. 2020. 'N6-deoxyadenosine methylation in mammalian mitochondrial DNA', *Molecular cell*.

Heyn, Patricia, Clare V. Logan, Adeline Fluteau, Rachel C. Challis, Tatsiana Auchynnikava, Carol-Anne Martin, . . . Andrew P. Jackson. 2019. 'Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions', *Nature genetics*, 51: 96-105.

Horton, Paul, Keun-Joon Park, Takeshi Obayashi, Naoya Fujita, Hajime Harada, C.J. Adams-Collier, and Kenta Nakai. 2007. 'WoLF PSORT: protein localization predictor', *Nucleic Acids Research*, 35: W585-W587.

Hu, Chiung-Wen, Jian-Lian Chen, Yu-Wen Hsu, Cheng-Chieh Yen, and Mu-Rong Chao. 2014. 'Trace analysis of methylated and hydroxymethylated cytosines in DNA by isotope-dilution LC–MS/MS: first evidence of DNA methylation in *Caenorhabditis elegans*', *Biochemical Journal*, 465: 39-47.

Katoh, Kazutaka, and Daron M. Standley. 2013. 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772-780.

Kosugi, Shunichi, Masako Hasebe, Masaru Tomita, and Hiroshi Yanagawa. 2009. 'Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs', *Proceedings of the National Academy of Sciences*, 106: 10171-10176.

Koziol, Magdalena J., Charles R. Bradshaw, George E. Allen, Ana S. H. Costa, Christian Frezza, and John B. Gurdon. 2016. 'Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications', *Nature Structural & Molecular Biology*, 23: 24-30.

Kweon, Soo-Mi, Yibu Chen, Eugene Moon, Kotryna Kvederaviciutė, Saulius Klimasauskas, and Douglas E. Feldman. 2019. 'An Adversarial DNA N6-Methyladenine-Sensor Network Preserves Polycomb Silencing', *Molecular cell*, 74: 1138-1147.

Lee, Gun E, Joo Hee Kim, Michael Taylor, and Mark T Muller. 2010. 'DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair', *Journal of Biological Chemistry*, 285: 37630-37640.

Li, Yong, Yi Jin Liew, Guoxin Cui, Maha J. Cziesielski, Noura Zahran, Craig T. Michell, . . . Manuel Aranda. 2018b. 'DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model Aiptasia', *Science Advances*, 4: eaat2142.

Liu, Jianzhao, Yanan Yue, Dali Han, Xiao Wang, Ye Fu, Liang Zhang, . . . Chuan He. 2014. 'A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation', *Nature Chemical Biology*, 10: 93-95.

Liu, Jianzhao, Yuanxiang Zhu, Guan-Zheng Luo, Xinxia Wang, Yanan Yue, Xiaona Wang, . . . Ye Fu. 2016b. 'Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig', *Nature communications*, 7: 1-7.

Mitchell, Alex L, Teresa K Attwood, Patricia C Babbitt, Matthias Blum, Peer Bork, Alan Bridge, . . . Robert D Finn. 2018. 'InterPro in 2019: improving coverage, classification and access to protein sequence annotations', *Nucleic Acids Research*, 47: D351-D360.

Mohan, K. Naga, Feng Ding, and J. Richard Chaillet. 2011. 'Distinct Roles of DMAP1 in Mouse Development', *Molecular and Cellular Biology*, 31: 1861-1869.

Mondo, Stephen J, Richard O Dannebaum, Rita C Kuo, Katherine B Louie, Adam J Bewick, Kurt LaButti, . . . Steven R Ahrendt. 2017. 'Widespread adenine N6-methylation of active genes in fungi', *Nature genetics*, 49: 964-968.

Negishi, Masamitsu, Tetsuhiro Chiba, Atsunori Saraya, Satoru Miyagi, and Atsushi Iwama. 2009. 'Dmap1 plays an essential role in the maintenance of genome integrity through the DNA repair process', *Genes to Cells*, 14: 1347-1357.

Nielsen, Henrik, Konstantinos D. Tsirigos, Søren Brunak, and Gunnar von Heijne. 2019. 'A Brief History of Protein Sorting Prediction', *The Protein Journal*, 38: 200-216.

Okano, Masaki, Daphne W. Bell, Daniel A. Haber, and En Li. 1999. 'DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development', *Cell*, 99: 247-257.

Ratel, David, Jean-Luc Ravanat, Marie-Pierre Charles, Nadine Platet, Lionel Breuillaud, Joël Lunardi, . . . Didier Wion. 2006. 'Undetectable levels of N6-methyl adenine in mouse DNA: cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase', *FEBS letters*, 580: 3179-3184.

Rhee, Ina, Kam-Wing Jair, Ray-Whay Chiu Yen, Christoph Lengauer, James G. Herman, Kenneth W. Kinzler, . . . Kornel E. Schuebel. 2000. 'CpG methylation is maintained in human cancer cells lacking DNMT1', *Nature*, 404: 1003-1007.

Rountree, Michael R, Kurtis E Bachman, and Stephen B Baylin. 2000. 'DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci', *Nature genetics*, 25: 269-277.

Saitou, N, and M Nei. 1987. 'The neighbor-joining method: a new method for reconstructing phylogenetic trees', *Molecular Biology and Evolution*, 4: 406-425.

Schermelleh, Lothar, Andrea Haemmer, Fabio Spada, Nicole Rösing, Daniela Meilinger, Ulrich Rothbauer, . . . Heinrich Leonhardt. 2007. 'Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation', *Nucleic Acids Research*, 35: 4301-4312.

Schiffers, Sarah, Charlotte Ebert, René Rahimoff, Olesea Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, . . . Thomas Carell. 2017. 'Quantitative LC-MS provides no evidence for m6dA or m4dC in the genome of mouse embryonic stem cells and tissues', *Angewandte Chemie International Edition*, 56: 11268-11271.

Shied, Bertrum, Parithychery R Srinivasan, and Ernest Borek. 1968. 'Deoxyribonucleic acid methylase of mammalian tissues', *Biochemistry*, 7: 280-285.

Simpson, Victoria J., Thomas E. Johnson, and Richard F. Hammen. 1986. '*Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging', *Nucleic Acids Research*, 14: 6711-6719.

Stamatakis, Alexandros. 2014. 'RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics*, 30: 1312-1313.

Stein, R, Y Gruenbaum, Y Pollack, A Razin, and H Cedar. 1982a. 'Clonal inheritance of the pattern of DNA methylation in mouse cells', *Proceedings of the National Academy of Sciences*, 79: 61-65.

Stöger, R, P Kubička, C-G Liu, T Kafri, A Razin, H Cedar, and DP Barlow. 1993. 'Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal', *Cell*, 73: 61-71.

Susan, J.CIark, Janet Harrison, Cheryl L. Paul, and Marianne Frommer. 1994. 'High sensitivity mapping of methylated cytosines', *Nucleic Acids Research*, 22: 2990-2997.

Tahiliani, Mamta, Kian Peng Koh, Yinghua Shen, William A Pastor, Hozefa Bandukwala, Yevgeny Brudno, . . . L Aravind. 2009. 'Conversion of 5-methylcytosine to 5-

hydroxymethylcytosine in mammalian DNA by MLL partner TET1', *Science*, 324: 930-935.

Tang, Chong, Rachel Klukovich, Hongying Peng, Zhuqing Wang, Tian Yu, Ying Zhang, . . . Wei Yan. 2018. 'ALKBH5-dependent m6A demethylation controls splicing and stability of long 3′-UTR mRNAs in male germ cells', *Proceedings of the National Academy of Sciences*, 115: E325-E333.

Tian, Li-Fei, Yan-Ping Liu, Lianqi Chen, Qun Tang, Wei Wu, Wei Sun, . . . Xiao-Xue Yan. 2020. 'Structural basis of nucleic acid recognition and 6mA demethylation by human ALKBH1', *Cell Research*, 30: 272-275.

Tuorto, Francesca, Reinhard Liebers, Tanja Musch, Matthias Schaefer, Sarah Hofmann, Stefanie Kellner, . . . Frank Lyko. 2012. 'RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis', *Nature Structural & Molecular Biology*, 19: 900-905.

Urieli-Shoval, Simcha, Yosef Gruenbaum, John Sedat, and Aharon Razin. 1982. 'The absence of detectable methylated bases in *Drosophila melanogaster* DNA', *FEBS letters*, 146: 148-152.

Waalwijk, Câ, and RA Flavell. 1978. 'MspI, an isoschizomer of hpaII which cleaves both unmethylated and methylated hpaII sites', *Nucleic Acids Research*, 5: 3231-3236.

Wakisaka, Keiko Tsuji, Yuuka Muraoka, Jo Shimizu, Mizuki Yamaguchi, Ibuki Ueoka, Ikuko Mizuta, . . . Masamitsu Yamaguchi. 2019. '*Drosophila* Alpha-ketoglutarate-dependent dioxygenase AlkB is involved in repair from neuronal disorders induced by ultraviolet damage', *NeuroReport*, 30: 1039-1047.

Wang, Xiang, Jing Feng, Yuan Xue, Zeyuan Guan, Delin Zhang, Zhu Liu, . . . Ping Yin. 2016. 'Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex', *Nature*, 534: 575-578.

Wang, Xiaoyan, Zhiqing Li, Quan Zhang, Bingqian Li, Chenchen Lu, Wanshun Li, . . . Ping Zhao. 2018. 'DNA methylation on N6-adenine in lepidopteran *Bombyx mori*', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1861: 815-825.

Woodcock, Clayton B., Dan Yu, Xing Zhang, and Xiaodong Cheng. 2019. 'Human HemK2/KMT9/N6AMT1 is an active protein methyltransferase, but does not act on DNA in vitro, in the presence of Trm112', *Cell Discovery*, 5: 50.

Wu, Tao P, Tao Wang, Matthew G Seetin, Yongquan Lai, Shijia Zhu, Kaixuan Lin, . . . Mei Zhong. 2016. 'DNA methylation on N 6-adenine in mammalian embryonic stem cells', *Nature*, 532: 329-333.

Xiao, Chuan-Le, Song Zhu, Minghui He, De Chen, Qian Zhang, Ying Chen, . . . Feng Luo. 2018. 'N6-methyladenine DNA modification in the human genome', *Molecular cell*, 71: 306-318.

Xie, Qi, Tao P. Wu, Ryan C. Gimple, Zheng Li, Briana C. Prager, Qiulian Wu, . . . Jeremy N. Rich. 2018. 'N6-methyladenine DNA Modification in Glioblastoma', *Cell*, 175: 1228-1243.

Xu, Kai, Ying Yang, Gui-Hai Feng, Bao-Fa Sun, Jun-Qing Chen, Yu-Fei Li, . . . Wei Li. 2017. 'Mettl3-mediated m6A regulates spermatogonial differentiation and meiosis initiation', *Cell Research*, 27: 1100-1114.

Yen, R. W., P. M. Vertino, B. D. Nelkin, J. J. Yu, W. el-Deiry, A. Cumaraswamy, . . . S. B. Baylin. 1992. 'Isolation and characterization of the cDNA encoding human DNA methyltransferase', *Nucleic Acids Research*, 20: 2287-2291.

Zemach, Assaf, Ivy E. McDaniel, Pedro Silva, and Daniel Zilberman. 2010. 'Genome-wide evolutionary analysis of eukaryotic DNA methylation', *Science*, 328: 916.

Zhang, Guoqiang, Hua Huang, Di Liu, Ying Cheng, Xiaoling Liu, Wenxin Zhang, . . . Jianzhao Liu. 2015. 'N6-methyladenine DNA modification in *Drosophila*', *Cell*, 161: 893-906.

Zhang, Min, Shumin Yang, Raman Nelakanti, Wentao Zhao, Gaochao Liu, Zheng Li, . . . Haitao Li. 2020. 'Mammalian ALKBH1 serves as an N6-mA demethylase of unpairing DNA', *Cell Research*.

Zheng, Guanqun, John Arne Dahl, Yamei Niu, Peter Fedorcsak, Chun-Min Huang, Charles J Li, . . . Chuan He. 2013. 'ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility', *Molecular cell*, 49: 18-29.

3 ACCUMULATION OF 6mA INVERSELY COINCIDES WITH ZYGOTIC GENOME ACTIVATION

From all model animals in which 6mA was investigated, only *Drosophila* and zebrafish provided evidence of above background 6mA levels during embryogenesis (Zhang et al. 2015; Liu et al. 2016b). Furthermore, only from these two reports 6mA levels are significantly and rapidly increases and decreases, hence dynamic ((Zhang et al. 2015; Liu et al. 2016b), Figure 3.2a). Accumulation and removal of 6mA during early embryogenesis may reflect a functional epigenetic role. Meanwhile, reports from adult tissues display only very subtle changes of 6mA levels (Koziol et al. 2016; Wu et al. 2016; Schiffers et al. 2017; Xie et al. 2018; Hao et al. 2020; Zhang et al. 2020). As 6mA levels of adult specimens of *Hydractinia* are extremely low (Figure 2.5), I investigated 6mA levels during early embryogenesis.

3.1 Embryonic Development of *Hydractinia symbiolongicarpus*.

An effort to investigate DNA methylation during embryogenesis of *Hydractinia* at very high temporal resolution would be too exhaustive and costly. Thus, it was necessary for me to limit my investigation to few embryonic stages where major biological events occur.

Embryonic development of *Hydractinia* begins with holoblastic and equal cleavage from 1 to 16 blastomeres ((Kraus et al. 2014), Figure 3.1a-e). As invagination never clearly occurs, the inward cell divisions from 16, 32, and to 64 cells are described as the gastrulation stages in *Hydractinia* embryogenesis ((Kraus et al. 2014), Figure 3.1c-f). The definitive segregation between endoderm and ectoderm, however, is established only between 7 and 11 hpf embryos (Figure 3.1i and j). During these time frame (9-12 hpf), *Nematostella* embryos are at the prawn chip stages of oscillating blastula marked distinctively by pseudo-invagination generating pits on the embryo's surface ((Fritzenwanker et al. 2007), Figure 3.1h[3]).

[3] Permission to reproduce this figure 3.1.h has been granted by Elsevier

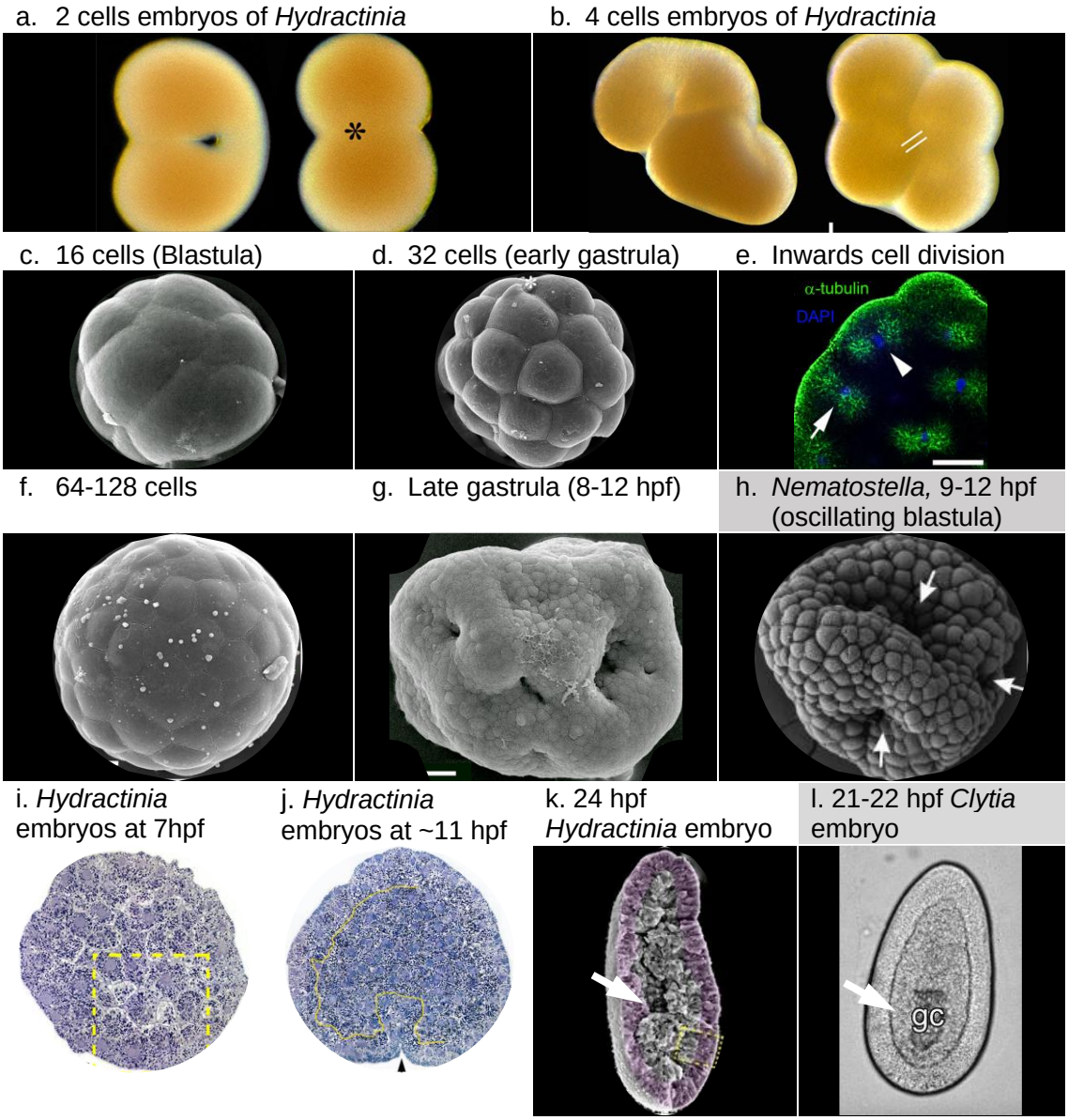


Figure 3.1. *Hydractinia* early embryogenesis description and the comparison with other cnidarians

(a) Two cells embryos. The furrow initiation site is marked with the asterisk. (b) Four cells embryos. (c) scanning electron microscopy (SEM) image embryos with 16 blastomeres. (d) SEM image of embryos at the 16-32 cell stage. (e) Immunofluorescence image shows orientation of the mitotic spindles at 32-cell stage. Arrowhead points to a spindle that is oriented parallel to the embryo surface, arrow points to a spindle oriented perpendicular to the embryo surface. Mitotic spindles were visualized using an anti a-tubulin antibody. DNA was stained with DAPI. (f) SEM image of 64-128 cell stage embryo. (g) Late gastrula ten hpf embryo, SEM data. Pits on the embryo surface are visible. (h) Oscillating blastula of 8-12 hpf embryos of *Nematostella*. Arrowhead points to pits on the embryo surface. (i) Histological semithin section of the 7 hpf embryo of *Hydractinia*. The region framed on highlights the unclear distinction between ectodermal and endodermal cells. (j) Histological semithin section of the 10-12 hpf embryo of *Hydractinia*. The clear demarcation line between ecto- and endodermal cells is accented by a yellow line. (k) SEM image of parenchymula stage of 24 hpf *Hydractinia* embryo, gastric cavity indicated by white arrow. (l) Parenchymula stage of 22 hpf *Clytia* embryo, gastrocoel indicated by white arrow. Figure (h) was taken from (Fritzenwanker et al. 2007), figure (l) from (Kraus et al. 2019) and all the other figures (*Hydractinia*) were taken from (Kraus et al. 2014).

Pits and concave region can also be observed at 10-15 hours post fertilization (hpf) embryos of *Hydractinia*. This stage of *Hydractinia* embryos, however, considered as late gastrula stage instead of blastula despite the striking morphological similarity with the oscillating blastula stage of *Nematostella* embryos ((Kraus et al. 2014), Figure 3.1j and g).

In *Nematostella,* definitive gastrulation-invagination occurs around 20-24 hours post fertilization embryos (Fritzenwanker et al. 2007); likewise, gastrulation-ingression in *Clytia* occurs from 20 to 24 hpf embryos and gastrocoel formed once the gastrulation completed as a distinctive trait of parenchymula (preplanula) stage. This gastrocoel of *Clytia* is morphologically resembling the gastric cavity of parenchymula stage of 24 hpf *Hydractinia* embryo (Kraus et al. 2014; Kraus et al. 2019) Figure 3.1k and l[4]). Therefore, through these comparative efforts, I narrowed my focus of investigation to embryos at early cleavage 2-4 cell stage; to the stage that it is formally described as the gastrulation period 16-32 cell and 64-128 cell stage; to the stage where a clear separation between endoderm and ectoderm by a basal lamina occurs at around 8 hpf, and to stage of parenchymula at 24 hpf where gastric cavity had formed.

3.2 Maternal to Zygotic Transition (MZT) and Zygotic Genome Activation (ZGA)

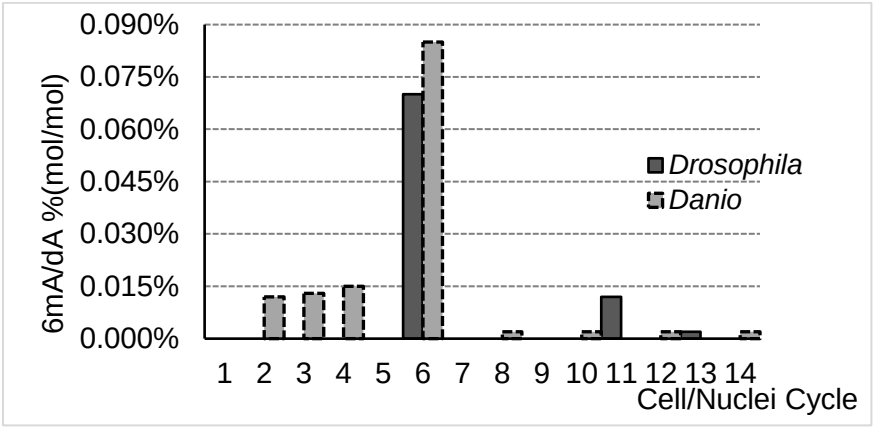
Comparing morphologically *Hydractinia* embryonic development to other animals outside Cnidaria like *Drosophila* and zebra fish is even more difficult then comparing it within cnidarians (Figure 3.2). Common among animals at early embryogenesis, however, is an event known as maternal to zygotic transition (MZT), which comprises rapid degradation of maternally deposited RNA (Maternal degradation) and the commencement of transcription from the zygotic genome (Zygotic genome activation, ZGA).

Maternal mRNA degradation begins immediately after fertilization, while the zygotic genome is activated at various stages across animals. Mammalian embryos tend to activate their genome early at 2-8 cell stages. As cleavage time in mammalian embryos is relatively slow, this early cleavage, thus zygotic genome activation, could last for 24 hours after fertilization (Jukam et al. 2017; Chen et al. 2019). In *Drosophila*, zebrafish, and frog, however, early cleavages are rapid and it takes only few hours to activate their genomes at the middle of the blastula stage before gastrulation occurs (Jukam et al. 2017; Schulz and Harrison 2019). Nonetheless, zygotic genome activation is a feature

[4] Permission to reproduce this figure 3.1.l has been granted by Elsevier

of embryogenesis that occurs in all animal embryos. Thus, using ZGA as a timepoint of comparison between early embryonic developmental stages of *Hydractinia* to other more evolutionarily distant animals would be an attractive approach.

a.



b.

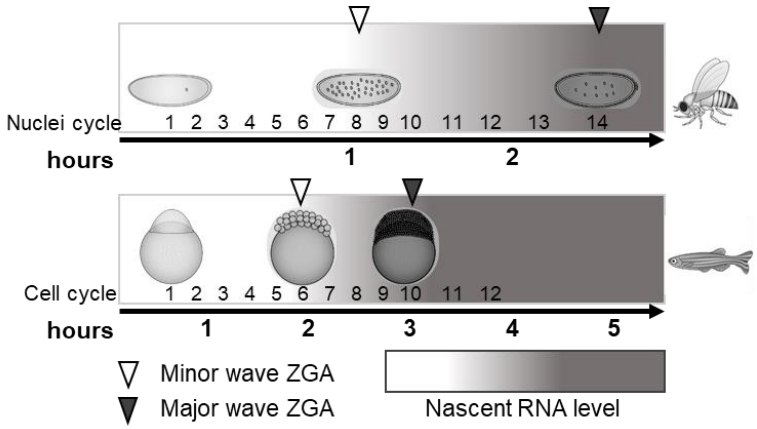


Figure 3.2. Dynamics of 6mA and ZGA at early embryogenesis of fruit fly and zebra fish.

(a). Levels of 6mA/dA at early embryogenesis of *Drosophila* and zebrafish (adapted from (Zhang et al. 2015; Liu et al. 2016b; O'Brown et al. 2019)). (b). Waves of zygotic genome activation in *Drosophila* and zebrafish (adapted from (Jukam et al. 2017; Chen et al. 2019; Schulz and Harrison 2019)).

3.3 ZGA is associated with 6mA in different animals

ZGA is the first transcription of an individual. Transcription can be regulated by DNA methylation. In yeast, 6mA pauses the transcriptional machinery (Wang et al. 2017a). Consistent with this, (Wu et al. 2016) found 6mA to mark and silence transposons in mouse embryonic stem cells. However, in *Drosophila*, 6mA is associated with active transposons (Zhang et al. 2015) while in ciliates 6mA is associated with H2A.Z and RNA polymerase II, hence active transcription (Wang et al. 2017b), and 6mA marks active transcription start sites in the genome of *Chlamydomonas reinhardtii* too (Fu et al. 2015). Thus, I decided to investigate the first transcription (ZGA) in *Hydractinia* and its association, if any, with 6mA levels.

Recently, the Fox-family protein Jumu was described to bind 6mA marked *zelda*, thereby negatively regulating Zelda that controls zygotic genome activation during early embryonic development of *Drosophila* (He et al. 2019), suggesting a mechanistic link between 6mA and ZGA. However, Zelda, the pioneer of *Drosophila* ZGA, is a derived character of arthropods (Ribeiro et al. 2017), absent in other taxa. Meanwhile, zebrafish that contains higher level of 6mA during early embryogenesis, controls ZGA in different ways (Giraldez et al. 2006; Zhao et al. 2017). Therefore, a link between 6mA and transcription during early embryogenesis could have evolved independently in different animals. In this chapter, I report a striking coincidence between the dynamic level of 6mA and ZGA during early embryogenesis of *Hydractinia*. I also discuss the comparison of 6mA-ZGA relationships in *Hydractinia, Drosophila*, and zebrafish embryos.

3.4 Methods

3.4.1 Collecting embryos of *Hydractinia symbiolongicarpus*.

Male and female colonies of *Hydractinia* were exposed to light to induce spawning in a sea water container outside of the main culture tanks. One and a half hours after light onset, *Hydractinia* colonies release eggs and sperm into the water column (Ballard 1942). Male and female colonies were then transferred back to the main culture tank, leaving behind the fertilized eggs which tend to settle to the bottom following fertilization. Then, embryos were swirled to the centre of the container. The embryos were collected and transferred to ice-cold seawater in a glass Petri dish. Once the embryos are collected, they can be stored at 4°C for up to two hours to halt their development providing a wider timeframe for manipulations such as injections. In 18°C incubator, *Hydractinia* embryos are dividing every ~45 minutes until reaching 512 cell stage. Afterwards, the divisions become non-synchronous (Kraus et al. 2014).

3.4.2 Quantification of DNA methylation by HPLC-QQQ during embryogenesis

Embryos were washed in filtered-sterile sea water for 15 times and observed under a stereomicroscope every 30-40 minutes to capture the specific developmental stages. These embryos were immediately transferred to 2 ml tube for DNA extraction and processed following the procedures as described in the previous chapter (see 2.2.4-5).

3.4.3 Wholemount immunofluorescence (IF)

3.4.3.1 Before fixation (EU/EdU soaking)

After embryos were collected and washed with sterile-filtered seawater, at the designated time (45 minutes before fixation), the embryos were incubated in 0.5-1.0 mM EU (Jena Bioscience #CLK-N002) or 10 µM EdU (Jena Bioscience # CLK-N001) solution in sterile-filtered sea water.

3.4.3.2 Fixation

Collected embryos were fixed by four different fixatives; PFA+GA (4% paraformalydehyde (Alfa Aesar #043368.9M), 0.4% Glutaraldehyde (Alfa Aesar #11448900) in filtered-sterile seawater), PFA+Ac (4% Paraformaldehyde, 0.27 M acetic acid (Fisher # 10384970) in filtered-sterile seawater, (Fernández and Fuentes 2013)), PAGA-T (20% PEG 6000 (Sigma #81260), 4% Glycerol (Sigma #G5516), 2.5% Acetic Acid, 56% of Ethanol in 100 mM Tris-HCl pH 6.0 (Invitrogen # 15568025), (Zanini et al. 2012)), and Methacarn (60% Methanol (Acros Organics #10649492), 30% Chloroform (Alfa Aesar # 15498679), 10% Acetic Acid, (Puchtler et al. 1970)). All fixatives were freshly prepared on ice and fixations were performed on a rocker for ~1.5 hour at 4 °C.

3.4.3.3 Permeabilization and pre-antibody preparation

Fixed embryos were rehydrated once by PBS-Triton (PTx, 0.5% Triton-X (MP Biomedicals #11471632) in 1x PBS). For embryos that were fixed by PAGA-T and Methacarn, a transition step was done with 1:3 mixture of the respective fixatives in PTx for 15 minutes. For embryos that were fixed by PFA+GA or PFA+Ac, a transition step with 200 mM Glycine (in PTx) for 15 minutes was performed to stop the cross-linking reaction by PFA.

The fixed embryos were then permeabilized with 1 ml PTx solution for 3 x 15 minutes at room temperature. This was followed by one wash in 1 ml 1x PBS solution for 15 minutes. Samples were then treated with 1:50 RNase solution[5] and/or DNase (2 U/µl, NEB #M0303) at 37°C overnight. Then, after one PBS wash, the embryos were rinsed in 1 ml of HCl 2 M for 45 minutes to denature the DNA as antigen retrieval step. The HCl was replaced and embryos were neutralized with 1 ml 100 mM Tris-HCl pH 8.0 for 2 x 15 minutes. The embryos were then rinsed in 1 ml block-i1 solution (3% BSA (MP Biomedicals #11444296) in PTx) for overnight at 4°C (or 1.5 hours at room temperature) on a rocker.

3.4.3.4 The CuAAC reaction

The Cu(I)-catalyzed alkyne-azide chemistry (CuAAC) reaction is a selective and rapid reaction between alkyne tagged molecules and azide tagged molecules. Ethynyl groups in EU/EdU act as the alkyne, which can react with fluorophore tagged azide through CuAAC reaction (Presolski et al. 2011). Na-Ascorbate solution was prepared by dissolving 200 mg of sodium ascorbate (Acros Organics # 10113731) in 1010 ml nuclease-free water[6]. Afterwards, the CuAAC solutions (Jena Bioscience #CLK-074)

[5] Mixture of RNaseA, T1 and H. (20 mg/ml, 1000 U/µl, and 10 U/µl respectively)

[6] Only freshly prepared and any excess solution were discarded

were prepared freshly as described in Table 3.1. The block-i1 solution was replaced with 500 µl CuAAC solutions then incubated on the rocker for at least 45 minutes in the dark and room temperature followed by two PTx washes.

Table 3.1 CuAAC solution recipe

	Volume for 1 reaction [µl]	*Final Concentration [mM]*
Na-Phosphate	Add up to 500	Irrelevant
Alexaflour488-[picolyl]azides	1	2×10^{-3}
$CuSO_4$/THPTA	15	1 / 5 [Cu/THPTA]
Na-Ascorbate	50	100

Note: The click-it solution must be used within 15 minutes after preparation.

3.4.3.5 Antibody reaction

The fixed embryos were rinsed in 1 ml block-i1 solution (3% BSA in PTx) overnight at 4°C before replaced with 200 µl of the primary antibody (diluted in block-i1, see Table 3.2) for one hour at room temperature (or 4°C overnight). Then, the fixed embryos were washed in 1x PBS for 2x15 minutes then rinsed in 400 µl block-i2 solution (5% goat serum (ThermoFisher #16210064) and 3% BSA in PTx) for overnight at 4°C. Then, embryos soaked in the secondary antibody (diluted in block-i2 solution, all antibodies used in this protocol is listed in Table 3.2) at 4°C overnight followed by PTx wash once for 15 minutes.

Table 3.2. List of antibodies used for immunofluorescence

Antibody	**Sources (code)**	**Host**	**Antigen**	**Dilution**
Polyclonal anti-m6A	SynapticSystem (202003)	Rabbit	m6A/6mA	1:1000
Monoclonal anti-5mC	Abcam (ab10805)	Mouse	5mC	1:200
Anti-rabbit Alexafluor488	Abcam (ab150077)	Goat	Rabbit IgG	1:2000
Anti-rabbit Alexafluor594	Abcam (ab150088)	Goat	Rabbit IgG	1:2000
Anti-rabbit Alexofluor647	Abcam (ab150083)	Goat	Rabbit IgG	1:2000
Anti-mouse Alexafluor488	Abcam (ab150117)	Goat	Mouse IgG	1:2000
Anti-mouse Alexafluor594	Abcam (ab150116)	Goat	Mouse IgG	1:2000

3.4.3.6 Nuclear staining, mounting and imaging

Embryos were then incubated in DAPI (Merck # 10236276001) solution [dissolved in 1x PBS at 1:2000 dilution] for 15 minutes followed by 4x washes with 1x PBS. Then, the embryos were mounted onto glass-slides in mounting medium (0.05 gram of N-Propyl gallate (Sigma # 02370) dissolved in 1 ml of Tris-HCl pH 8.0 and topped up with 9 ml glycerol, heated up to 60 °C for 10 minutes and stored at -20°C before use). The mounted embryos were observed, and the images were taken by, a confocal laser scanning microscope (Olympus FV1000). Known positive control samples were used to calibrate the confocal setup against the negative control samples (replacing primary antibody solution with block-i1, replacing EU/EdU soaking steps with seawater only), once balance between the two controls achieved at particular setup, this setup then

consistently performed when images taken from samples slides on the same day of image acquisition.

3.4.3.7 Image Preparation and Quantification

Images were imported to ImageJ software (Schneider et al. 2012). Stacked images were z-projected in “Max Intensity” and “Sum Slices” types. DAPI (or EdU) channel images from “Max Intensity” projections were used to determine nuclear area from the images by setting the threshold on MaxEntrophy mode and particle analysis at 10-300 micron^2 of size, 0.25-1.00 of circularity. The nuclear region of interest (nucROI) resulted were then manually sorted and redirected to the second (EU-Clik-iT and anti-6mA) channel images of “Sum Slices” projection for intensity measurement (set to measure Area, Mean Grey Values, and Integral Density).

Background areas were determined by setting the threshold value 300 in reverse from “MaxEntrophy” minimum. The ROI were then applied to the second channel, manually selected only for those that surrounded the nuclear ROI and the intensity measured. The measurement data from nuclear ROI and background ROI then transferred to Microsoft Excel for calculation. The corrected total cell fluorescence (CTCF) was calculated by correcting the nuclear ROI by the background ROI through the following equation:

$$CTCF = IntDen\ of\ nucROI - (Area\ of\ nucROI \times avg.background\ ROI)$$

$$avgbackground\ ROI = \frac{Total\ Integral\ Density\ of\ background\ ROI}{Total\ Area\ of\ background\ ROI}$$

$$IntDen\ of\ nucROI = Integral\ Density\ of\ nuclear\ ROI = \ Area\ \times Mean\ Grey\ of\ nucROI$$

3.4.4 RNA sequencing and time course analysis in embryonic development

3.4.4.1 RNA extraction

Total RNAs were extracted from embryos of 2-4 cell, 16-32 cell, 64-128 cell stages and 24 hours post fertilization. The extraction begun by lyses of the embryos in 500 µl TRIzol solutions at room temperature with rigorous mixing for 5 minutes[7]. Afterwards, 100 µl chloroform were added into the suspension and strongly mixed. The suspensions were then centrifuged at maximum speed for 15 minutes at 4°C. I transferred the upper-aqueous phase [±500 µl] into a new clean 2 ml tube. Then, 1.5 volume of chilled ethanol absolute [±750 µl] were added. After short inversion, the whole suspension was transferred into an RNA column (EpochLifeScience #1940) in 2 ml flowthrough-collection tube. The RNA bound to the column-membrane by centrifugation at 6,000 x g for 2 minutes, then the flowthrough was discarded. Afterwards, 400 µl of wash-r1 buffer (1 M Gu-HCl, 10 mM Tris-HCl pH 7.0, in nuclease-free water) were added to the columns, which were then centrifuged at 8,000 x *g* for 1 minute. After the flowthrough was

[7] This suspension can be stored at -80 °C for 2 months.

discarded, 80 µl DNase solution (5 µl DNaseI + 75 µl digestion buffer, Qiagen #79254) was added carefully into the centre of the column then incubated at room temperature for two hours. Thereafter, the columns were washed with 500 µl of wash-r1 buffer, then with 500 µl of wash-r2 buffer (10mM Tris-HCl pH 7.5, 60 mM potassium acetate, 80% ethanol in nuclease-free water) twice. These washes were done by centrifugation at 11,000 x *g* for 1 minute. Then, the columns were dried by one more centrifugation at maximum speed for 2 minutes. The columns were then transferred into new 2 ml collection tubes. Finally, the bond RNAs were eluted from the column with 20-40 µl nuclease-free water and 8,000 x g centrifugation for 1.5 minutes.

3.4.4.2 RNA assessment

The eluted RNA solution was measured and assessed by Nanodrop and Qubit fluorometer (Qubit RNA HS assay). Then, RNA solutions were mixed with RNA loading dyes (1:1, ThermoScientific #R0641) heated up to 70°C for 10 minutes and loaded into a denaturing formaldehyde agarose gel (1.5% agarose, 0.06% formaldehyde, in 1x MEN buffer (20 mM MOPS, 5 mM Na-Acetate, 1 mM EDTA, pH = 7.0)) for electrophoresis.

3.4.4.3 RNA-seq[8]

After RNA assessed and high quality ensured, the RNA samples were shipped on dry ice to the NIH Intramural Sequencing Center (NISC) for further processing and sequencing. Here, the RNAs were further assessed using Bioanalyzer System (Agilent #G2939BA). RNA-Seq libraries were constructed from 1 µg RNA (RIN ≥ 9.5) using the Illumina TruSeq RNA Sample Prep Kit, version 2. The resulting cDNA was fragmented using a Covaris E210 focused ultrasonicator. Library amplification was performed using 10 cycles to minimize the risk of over-amplification. Unique barcode adapters were applied to each library. Libraries were pooled in an equimolar ratio and sequenced together on three lanes of an Illumina HiSeq 2500 sequencing system using version 4 flow cells and sequencing reagents. At least 65 million reads of 75-base pairs were generated for each individual library. These raw reads were then transferred to National University of Ireland Galway High Performance Computing resources, Syd.

3.4.4.4 Quality control and trimming

Read quality control for both sets of samples was performed using FastQC v0.11.5. Overrepresented sequences and low-quality (<32) bases were trimmed using Trimmomatic v0.30. After trimming, unpaired reads and reads shorter than 25 bp were discarded.

[8] This section was done by collaborators at NIH, USA.

3.4.4.5 Reference transcriptome assembly[9]

Reads from all samples were combined to generate two transcriptome assemblies. The first transcriptome was assembled using Trinity (v2.8.4) with default parameters. For the second assembly, reads were mapped to a draft assembly of the *H. symbiolongicarpus* genome using HISAT2 (v2.1.0). The resulting alignments were sorted using SAMtools and used to assemble transcripts with Stringtie (v1.3.4). A third transcriptome was downloaded from previous assemblies done on reads originated from adult specimens of *Hydractinia symbiolongicarpus* (DuBuc et al. 2020).

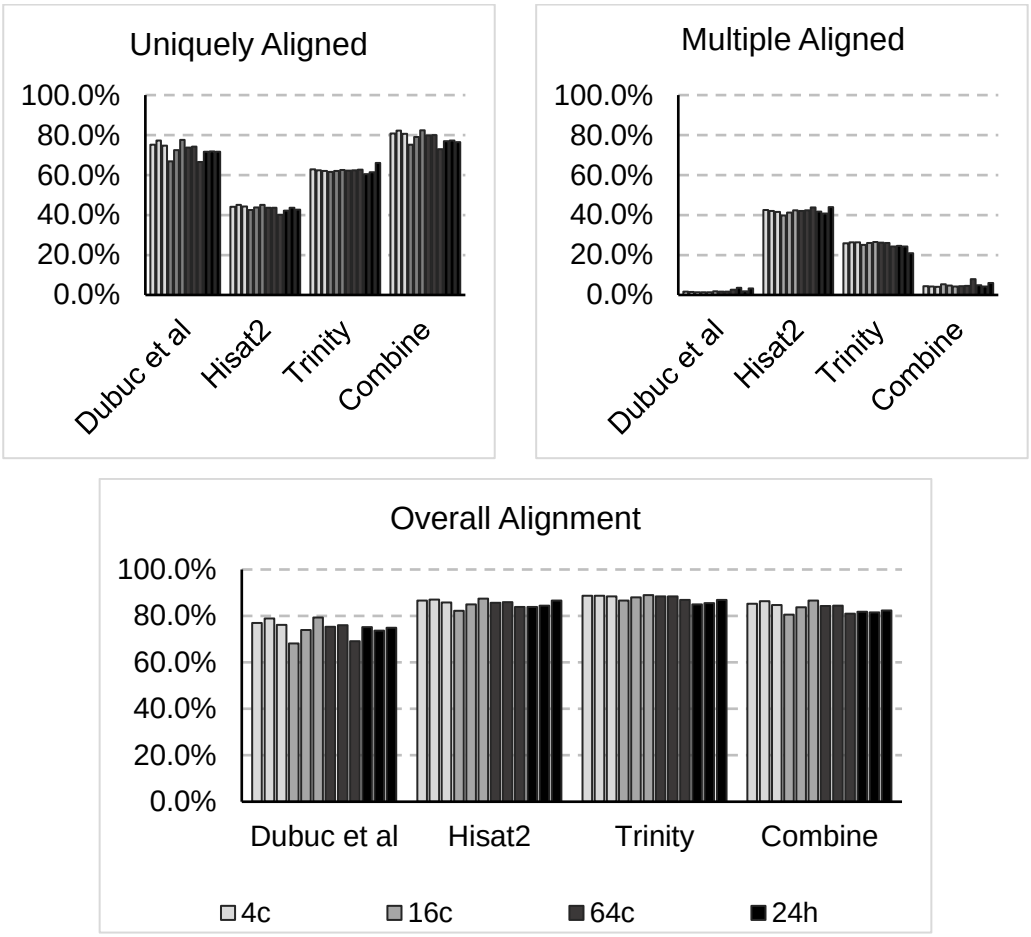


Figure 3.3. Mappability of transcriptomes of *Hydractinia symbiolongicarpus.*

Pair reads from four embryonic stages were aligned to four transcriptomes by RSEM-bowtie2. Dubuc et al. is the publicly available transcriptome of *Hydractinia symbiolongicarpus* (DuBuc et al. 2020). “Combine” is a transcriptome that was built by combining Dubuc et al, Hisat2 and Trinity transcriptomes by tr2aacds.pl. Upper-Left are the percentage of reads that aligned one time to the reference. Upper-Right are the percentage of read that aligned to the reference multiple time, thus indicative of redundancies of the reference. Lower graph shows the overall alignment of the reads to the references.

[9] This section was done by collaborators at NIH, USA.

A consensus, non-redundant transcriptome was made by combining the Trinity, Stringtie and assemblies from the previous study with the tr2aacds.pl (a Perl script from the 10 EvidentialGene package)[10]. This combined transcriptome was used as the reference transcriptome hereafter.

3.4.4.6 Read alignment and gene-level read counts

Reads from the 4 stages (each in triplicates) of embryogenesis samples were aligned to the *H. symbiolongicarpus* combined transcriptome using bowtie2 (v2.2.6) and gene-level read counts were generated with RSEM (Li and Dewey 2011).

```
rsem-prepare-reference --bowtie2 -p 12 Hsymbio_combine.fasta RefHsym
rsem-calculate-expression --bowtie2 -p 12 --paired-end --calc-ci --ci-memory 2048
    input_R1.fq input_R2.fq RefHsym output/
rsem-generate-data-matrix  4c_1.genes.results 4c_2.genes.results ... > SymbDevMat
```

Script 3.1. RSEM main steps for alignments and counting the expression level.

Overall alignment rates from bowtie2 ranged from 78-86% for all samples (Figure 3.3). This alignment rates to the combined transcriptome perform much better compare to alignment rates of all other prior transcriptomes (Figure 3.3). Afterward, RSEM-generate-data-matrix were performed to generate count.matrix ("SymbDevMat"). The resulting counts.matrix file was used for downstream analyses of differential expression.

3.4.4.7 Time course and differential expression analysis

Pair-wise differential expression (DE) analysis among all four embryonic stages was performed by collaborators at NIH with R packages HTseq, DEseq2 and ImpulseDE2. In addition to the pair-wise comparisons, I was interested in understanding how clusters of genes are differentially expressed over the time course of development. Thus, I performed time course analyses in EBSeqHMM (R package), which utilizes Bayesian approach with a hidden Markov model to identify DE between ordered conditions. The count.matrix resulted from RSEM were unnormalized and uploaded into the EBseq-HMM environment in R and name it "DevData". The count then normalized using upper-quartile normalization (Bullard et al. 2010) generating normalized expression level for each time point. Then, genes were grouped into expression paths (i.e. "Up-Down", "Down-Down", "Equally Expressed"), in which DE occurs when expression paths change between at least one adjacent condition using "GetConfidentCalls". As I had four time points, three step paths were generated (i.e., "Up-Up-Up", "Down-Down-Down", "EE-Up-Down").

[10] http://arthropods.eugenes.org/EvidentialGene/trassembly.html

```
> Sizes = QuartileNorm(DevData,0.75)
> levels (Conditions)
      [1] "4c" "16c" "64c" "24h"
>GeneNormData = GetNormalizedMat(DevData, Sizes)
> EBSeqHMMGeneOut = EBSeqHMMTest(Data=GeneNormData, sizeFactors=Sizes,
      Conditions=Conditions, UpdateRd=1000)
>GeneConfCalls = GetConfidentCalls(EBSeqHMMGeneOut, FDR=0.01,cutoff=0.8,
      OnlyDynamic=FALSE)
```

Script 3.2. EBseq-HMM main steps performed to get gene expression paths.

I extracted the gene list of all expression paths, then plotted the expression level in the course of 4 developmental stages into a boxplot. I then visually inspected the expression profiles of genes with a known expression path from previous in situ hybridization studies. This allowed me to validate the expression path resulted from the analysis and assigned the transcripts into expression groups. For instances, transcripts of *RFamide* and *Tfap2* were never expressed (extremely rare) along developmental course, consistent with previous studies (Figure 3.9a, (Plickert et al. 2004; Gahan et al. 2017; DuBuc et al. 2020)). *Piwi1* and *Wnt3* were grouped in the maternal degradation and late developmental group, which corresponds well with previous results (Figure 3.9b&c, (Plickert et al. 2006; Duffy et al. 2010; McMahon 2018; DuBuc et al. 2020)). Thus, these data indicate the reliability of the expression profile and path assigned by EBseqHMM.

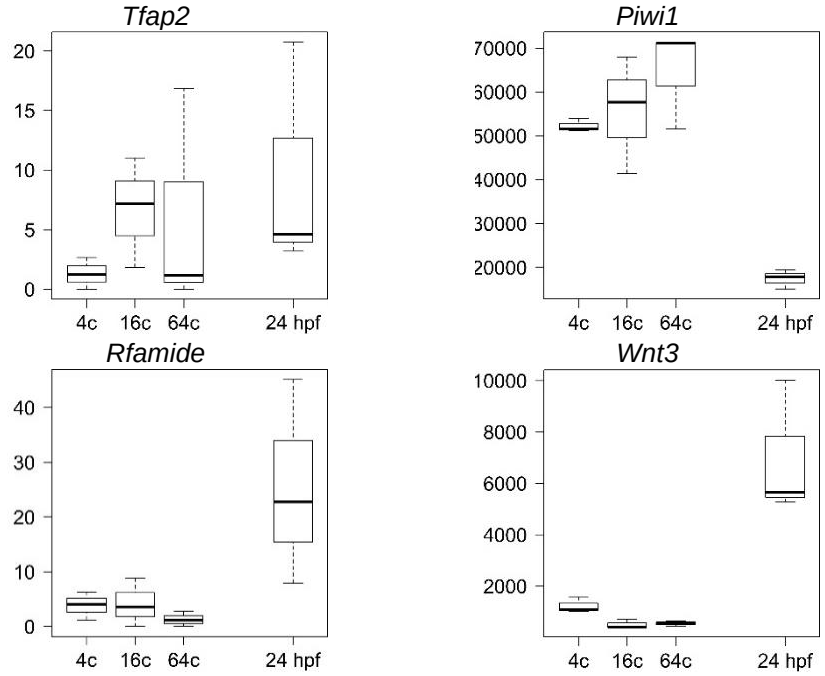


Figure 3.4. EBseqHMM validation from genes with known expression profile from previous studies in *Hydractinia.*

Y-axis represent upper-quartile normalized expression levels. Plots and extensive list of the transcripts presented at Appendix (see page 133-138) and supplementary documents.

3.5 The dynamics of DNA methylation during embryogenesis of *Hydractinia*.

With the established HPLC-QQQ procedures (see section 2.2.5), DNA from embryonic stages of *Hydractinia* were collected and the DNA methylation levels were measured. I discovered that 5mC/dC levels are incrementally rising from 2 cell stage (~2%) to 24 hours post fertilization embryos, then remain steady towards adulthood at ~3% (Figure 3.5a). I also found that 6mA levels are accumulated from fertilization to 16-32 cell stage at ~0.06% of 6mA/dA, but rapidly decline to background level at 64-128 cell stage and maintained at this level until adulthood (Figure 3.5b).

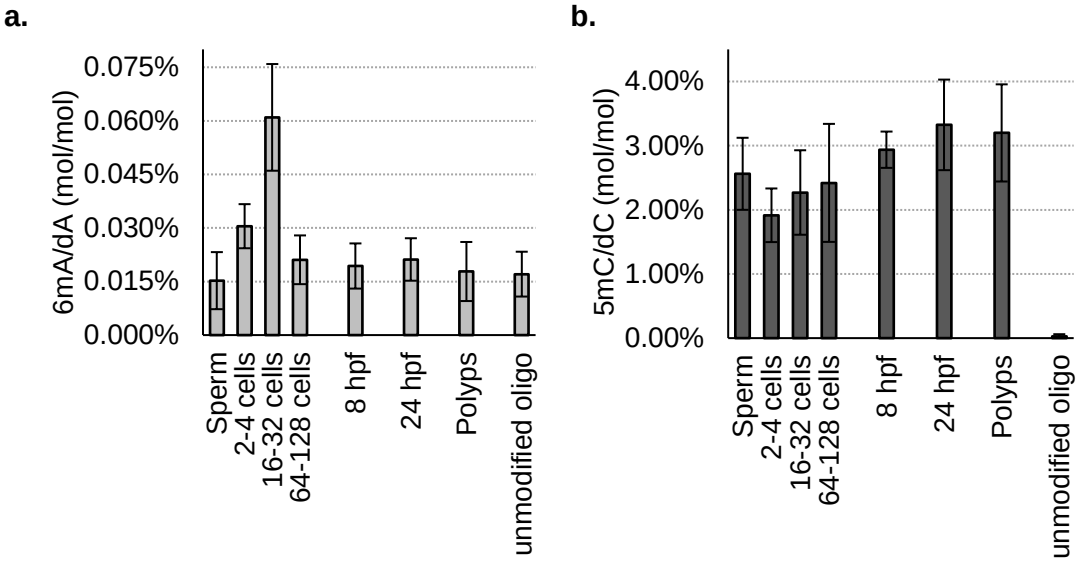


Figure 3.5. Dynamic levels of DNA methylation during embryogenesis of *Hydractinia.*

(a) 6mA/dA (%, mol/mol) measured by HPLC-QQQ. (b) 5mC/dC (% mol/mol). Unmodified oligo was used as negative control, hence indicating background level of any DNA methylation signals from samples.

These dynamics of 5mC and 6mA levels are strikingly similar to DNA methylation dynamics during embryogenesis of zebrafish with respect to ZGA (Figure 3.2a , (O'Brown et al. 2019)). Further, despite the lack of significant 5mC in *Drosophila*, their 6mA/dA dynamics during embryogenesis (Figure 3.2a, (Zhang et al. 2015)) also resembles 6mA dynamics in *Hydractinia* embryogenesis. Thus, rapid changes of 6mA/dA levels during early embryogenesis with a drop before ZGA are, perhaps, a feature shared across animals.

3.6 Immunofluorescence confirms 5mC and 6mA in early embryos of *Hydractinia.*

Positive results by HPLC-QQQ system may result from bacterial contaminations instead of the actual 6mA detection from *Hydractinia* genomic DNA. This issue has been discussed in a study on zebrafish embryos (O'Brown et al. 2019). Thus, I had to control the HPLC-QQQ data with whole mount immunofluorescence using anti-6mA that was

validated previously by dot-blot in Figure 2.3. Indeed, 5mC and 6mA signals colocalised with the DAPI staining (Figure 3.6a and b). This excluded the bacterial sources of 6mA signals.

However, anti-6mA may recognize both m6A (RNA) and 6mA (DNA), thus RNase and DNase treatment were necessary as additional control to ensure that the signals came out from 6mA (DNA), not from m6A (RNA). RNase treatment did not eradicate the anti-6mA or anti 5mC signals. However, additional DNase treatment also failed to deplete the signals (Figure 3.6c). I suspected that this failure is due to the crosslinking fixation by paraformaldehyde and glutaraldehyde, which fixed and strongly preserved the DNA structure.

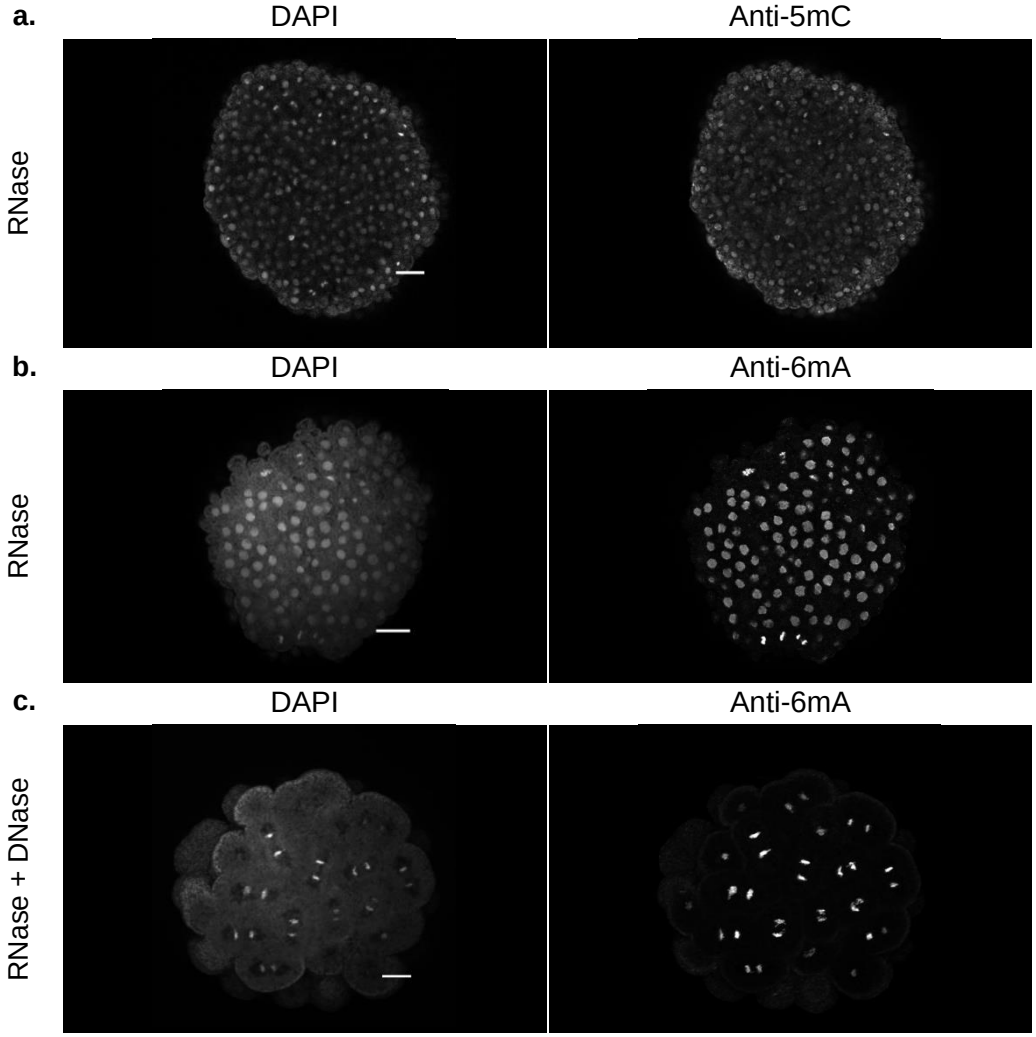


Figure 3.6. Wholemount Immunofluorescence of anti-5mC and anti-6mA on embryos of *Hydractinia.*

a. Anti-5mC with RNase treatment only. b. Anti-6mA with RNase treatment only. and c. Anti-6mA with RNase and DNase treatment. All embryos were fixed by paraformaldehyde and glutaraldehyde (PFA+GA). Left panels of all rows are DAPI channels for the respective experiments. Scale bars at DAPI channels indicate 20 µm in length.

Moreover, the very strong signals of 6mA at embryonic stages later than 64-128 cells seems inconsistent with the HPLC-QQQ data (Figure 3.5 and Figure 3.6b). This, perhaps, due to oversaturation of the antibody used. Thus, I reassessed the fixatives and the anti-6mA antibody dilution in my whole mount immunofluorescence protocols to be able to differentiates 6mA levels on embryos of *Hydractinia*.

Three different fixatives were tested: PFA+GA, PAGA-T (Zanini et al. 2012), and Methacarn (Puchtler et al. 1970). PAGA-T proved to be the best fixative because DNase treatment post fixation removed the 6mA signal (Figure 3.7b). Moreover, 6mA was still detected despites the RNase treatment (Figure 3.7a and b). Thus, these results together excluded an 6mA bacterial sources or m6A (RNA) contamination.

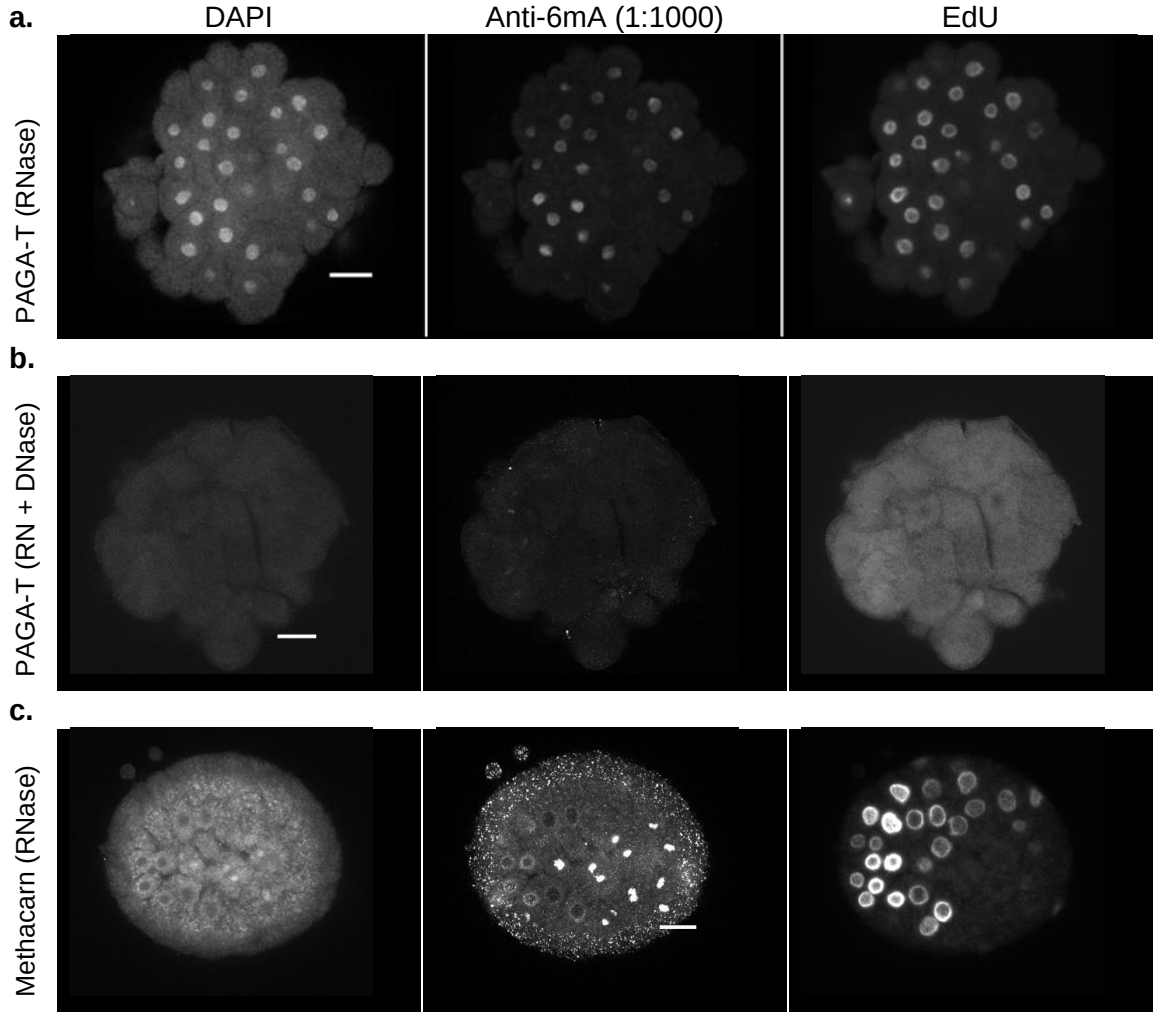


Figure 3.7. PAGA-T is the best fixative for anti-6mA

All scale bars are indicating length of 20 µm. a. PAGA-T fixation can indicate anti-6mA at nuclei of 64-128 cells embryo of *Hydractinia*, despite of the RNase treatment. b. RNase plus DNase treatment on PAGA-T fixation protocols shows no signals from DAPI, EdU or anti-6mA. c. Methacarn fixation was inconsistent with nuclear staining and anti-6mA in RNase treatment, thus the DNase treatment experiment was not pursued further.

I also found that anti-6mA dilution to 1:8000 is necessary to display the differential signals of 6mA between 16-32 and 64-128 stages (Figure 3.8). Therefore, these results provided strong evidence for 6mA dynamic during early embryogenesis and suggest that the HPLC-QQQ results are not biased by bacterial sources (Figure 3.5).

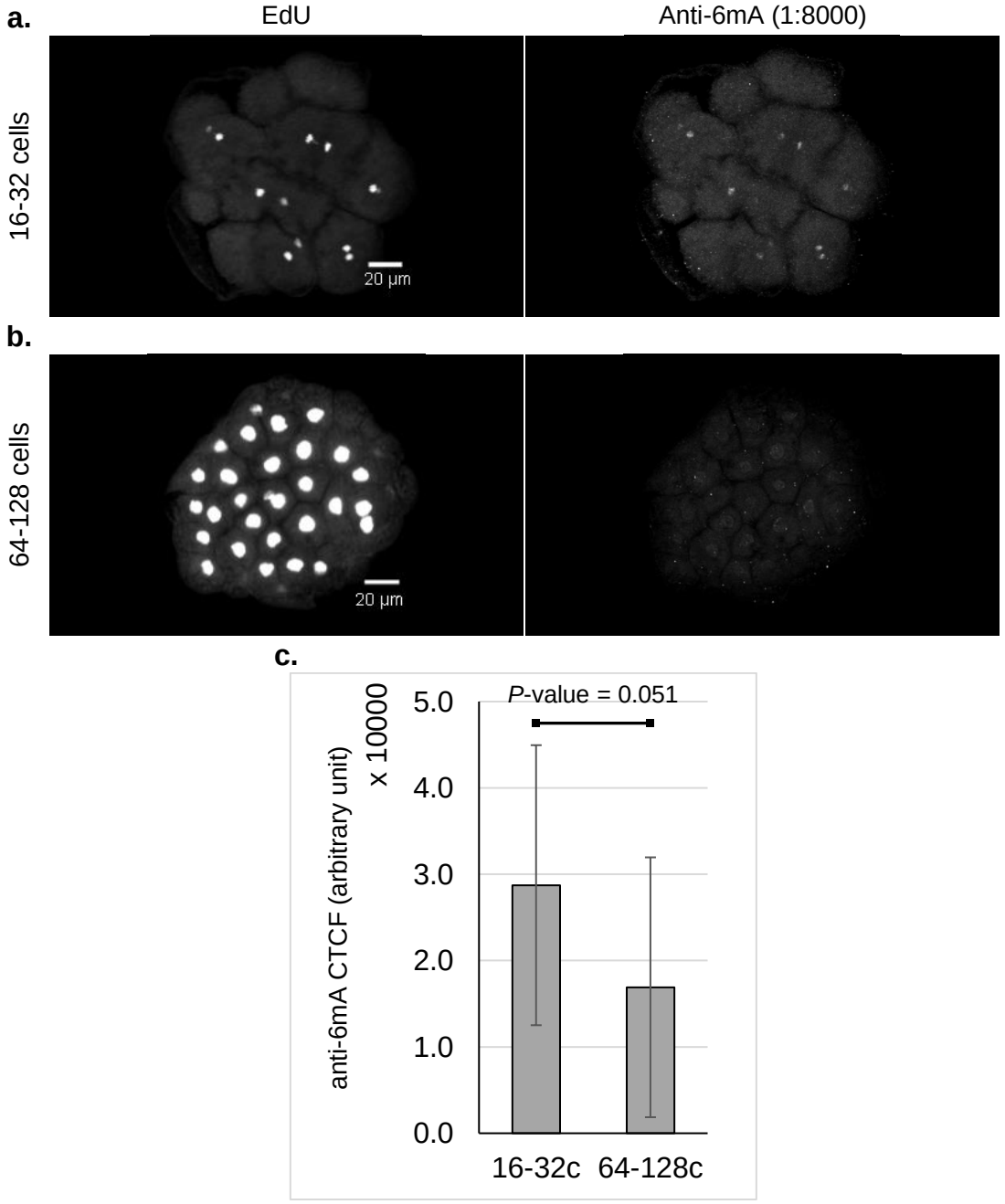


Figure 3.8 Differential 6mA signals from 16-32 cells and 64-128 cell stage in *Hydractinia* embryos.

a. Wholemount-IF of 16-32 cells embryo. b. Wholemount-IF of 64-128 cells embryo. Left pannels of a and b are EdU channels indicating DNA signals, right pannels of a and b are anti-6mA channels (with 1:8000 dilution). Scale bars at EdU channels are indicating.length of 20 µm. c. image quantification of anti-6mA channels from 16-32 cells and 64-128 cells embryos. CTCF (corrected total cell fluorescence, calculated as described in section 3.4.3.7) Raw data available in Supplementary Document 7.

3.7 Expression of genes during early embryogenesis.

Next, I questioned the functional role of 6mA in the transcriptional profile of 16-32 and 64-128 cells embryos of *Hydractinia*. For this, gene expression profile from these embryonic stages was necessary. Hence, I (and collaborators[11]) performed RNA-Seq on 2-4 cells, 16-32 cells, 64-128 cells, and 24 hours post fertilization embryos. For my thesis project, I performed EBseqHMM to understand the path of gene expression in the course of embryonic development, which generated thirteen most likely expression paths of genes in the course of 4 developmental stages of *Hydractinia*. From the thirteen generated expression paths, I visually inspected the boxplot and concluded that there are 8 major expression groups that share similar expression path (Appendix (133-138)[12] and Figure 3.9).

First and most obvious group is maternal degradation, which comprises early rapid degradation (Figure 3.9a) and late degradation (Figure 3.9b, Appendix 6.6). These are genes that are consistently downregulated across the four stages (D-D-D), or equally expressed between the first three stages but strongly down regulated at the transition from 64-128 cell stage to 24 hpf embryos (D/E-D/E-D) Figure 3.9a and b).

During my visual inspection, I found ~120 transcripts differentially expressed in the time course analysis, thus assigned to an expression path termed equal-down-up (E-D-U) or other type of expression path, but the expression levels of these transcripts were extremely low or nearly undetectable in all four embryonic stages (Appendix 6.10). I regroup the transcripts from these various expression path and call the group as lineage specific expression group. For instance, transcripts of *RFamide receptor* (Figure 3.9c) were never expressed as the *RFamide* precursor transcripts (Figure 3.4) along the developmental course, consistent with previous studies (Plickert et al. 2004; Gahan et al. 2017; DuBuc et al. 2020).

Next expression group is the late developmental transcripts which are highly up regulated at 24 hpf preplanula stage. I further divided the transcripts within this “high-preplanula” expression group into three: (A) those transcripts that are extremely low at all previous stages (Figure 3.9f), (B) the transcripts that are substantially and dynamically expressed at earlier stage (Figure 3.9g) and (C) the transcripts that follows rapid early maternal degradation path but highly expressed at preplanula (24 hpf) stage (Figure 3.9h). Lastly, I also uncovered two noteworthy expression groups: the transiently

[11] These RNA-Seq efforts were part of the *Hydractinia* genome project of our collaborators at NIH.

[12] I provided boxplot examples of 9 transcripts foreach of the 6 on the expression paths, the complete raw data can be accessed in Supplementary Documents.

upregulated genes at 16-32 cell stage (Figure 3.9d) and the significantly upregulated once at 64-128 cell stage but then either up or down regulated at the next (preplanula) stage (Figure 3.9e).

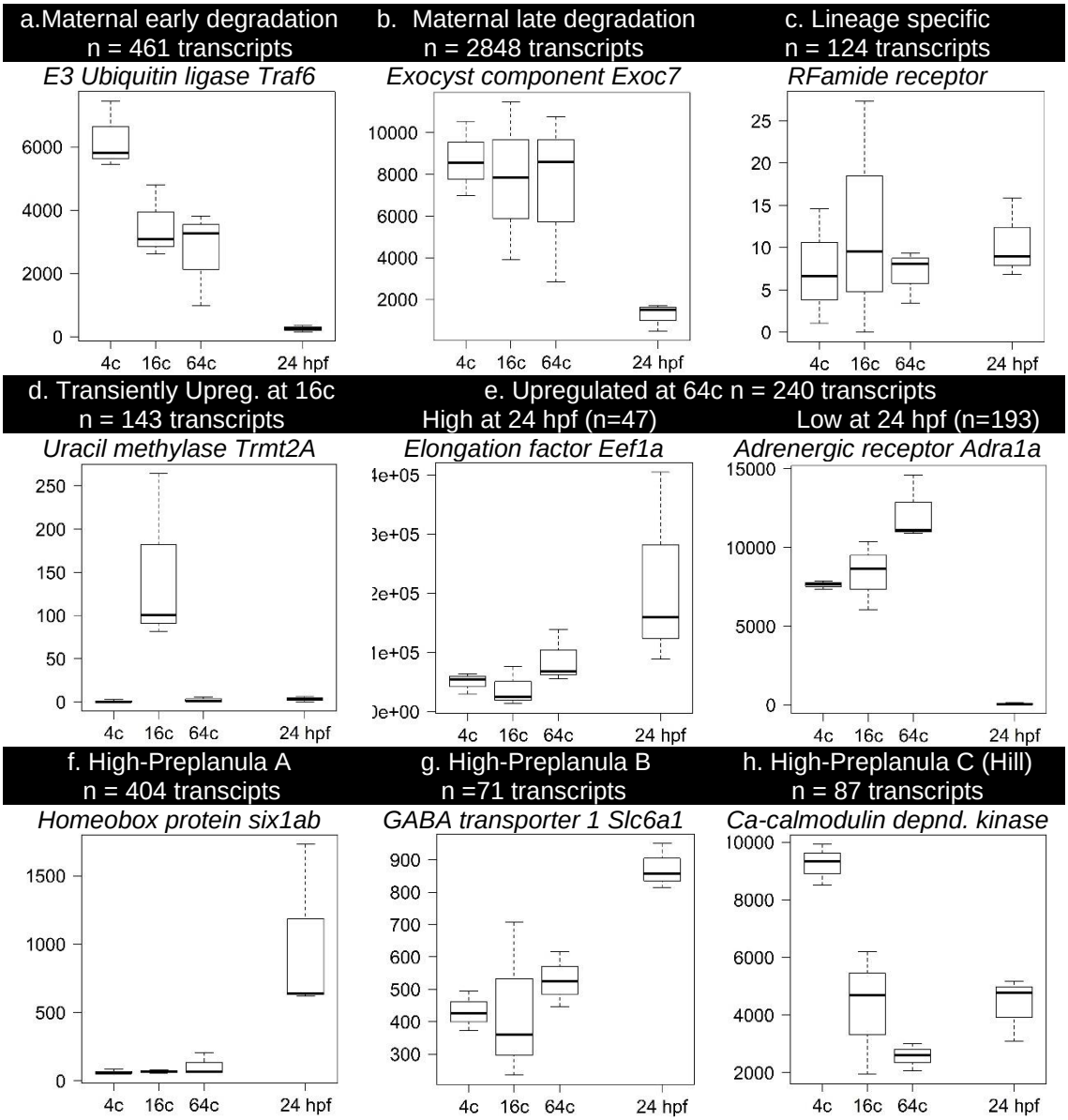


Figure 3.9. Expression group of genes during embryogenesis of *Hydractinia.*
a. Maternal early degradation expression group (D-D/E-D/E). b. Maternal late degradation expression group (E-D/E-D). c. Lineage specific expression group were grouped based on extremely low level of count (<50) regardless the expression path. d. Transiently upregulated at 16-32 cell stage (U-D-E). e. Upregulated at 64-128 cell stage and either upregulated at 24 hpf (High, D/E-U-U) or downregulated at 24 hpf (Low, D/E-U-D). f. High preplanula A expression group exhibit extremely low count at early stage but significantly upregulated at 24 hpf (E-E-U). g. High preplanula B expression group exhibit dynamic count level at early stages but significantly upregulated at 24 hpf (D/E-D/E-U). h. High preplanula C expression group exhibit rapid decline of expression at early stage but significantly upregulated at 24 hpf (D-D-U). E (equally expressed), D (down regulated) and U (up regulated). Y-axes represent upper quartile normalized read counts. Plots and extensive list of the transcripts presented at Appendix (see page 133-138) and supplementary folder.

Important to note, the normalized count level of the transiently upregulated genes at 16-32 cell stage expression group are mostly below 250 (Figure 3.9d and Appendix 6.11), which indicates a minor wave of zygotic genome activation. Meanwhile, those transcripts that are upregulated at 64-128 cell stage are expressed at the range of thousands to 10^5 of normalized count level (Figure 3.9e and Appendix 6.9), which indicates the major wave of zygotic genome activation. This led me to question the maternal late degradation expression group (Figure 3.9b). There is a possibility that the transcripts that are maternally degraded, are concomitantly zygotically activated, thus manifested as equally expressed at early stages.

Furthermore, principal component analysis (PCA) by DEseq2 found no significant differences between the three early embryonic stages (Figure 3.10). This is not surprising as differential expression from RNAseq with few time points without additional experimental treatment have failed to distinguish maternal from zygotic transcripts in a previous study, thus being less effective to identify zygotic genome activation timepoint (Owens et al. 2016). To address this problem, I developed another method to pinpoint ZGA during early embryogenesis of *Hydractinia* as explained in the next section.

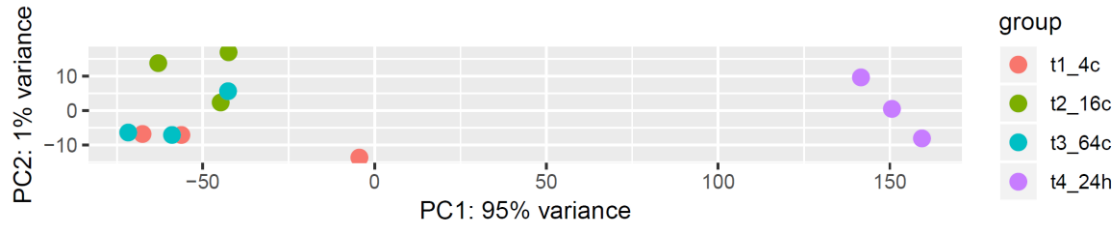


Figure 3.10. PCA from DEseq2 analysis.
The PCA analysis displays insignificant variance between samples from the three early embryonic stages [13].

3.8 Zygotic genome activation at 64 cells embryos of *Hydractinia*

To detect the first nascent RNA at the ZGA time point, *Hydractinia* embryos were soaked in EU for 30 minutes then fixed and stained using the CuAAC reaction. EU is a nucleoside analogue, which if taken up by cells will be incorporated into RNA by RNA polymerases, hence marking nascent RNA and indicates zygotic genome activation during early embryogenesis (Chen et al. 2019). I have found that *Hydractinia* strongly incorporates EU only after the 64-cell stage (Figure 3.11), which I determined as being the major wave of ZGA. It is interesting to note that a weak incorporation of EU is detectable at 32-64 cell stage consistent with the weak and transiently upregulated at 16-32 cell stage expression group (Figure 3.9d).

[13] Thanks to Sofia Barreira and Paul Gonzalez, my collaborator at NIH for this figure.

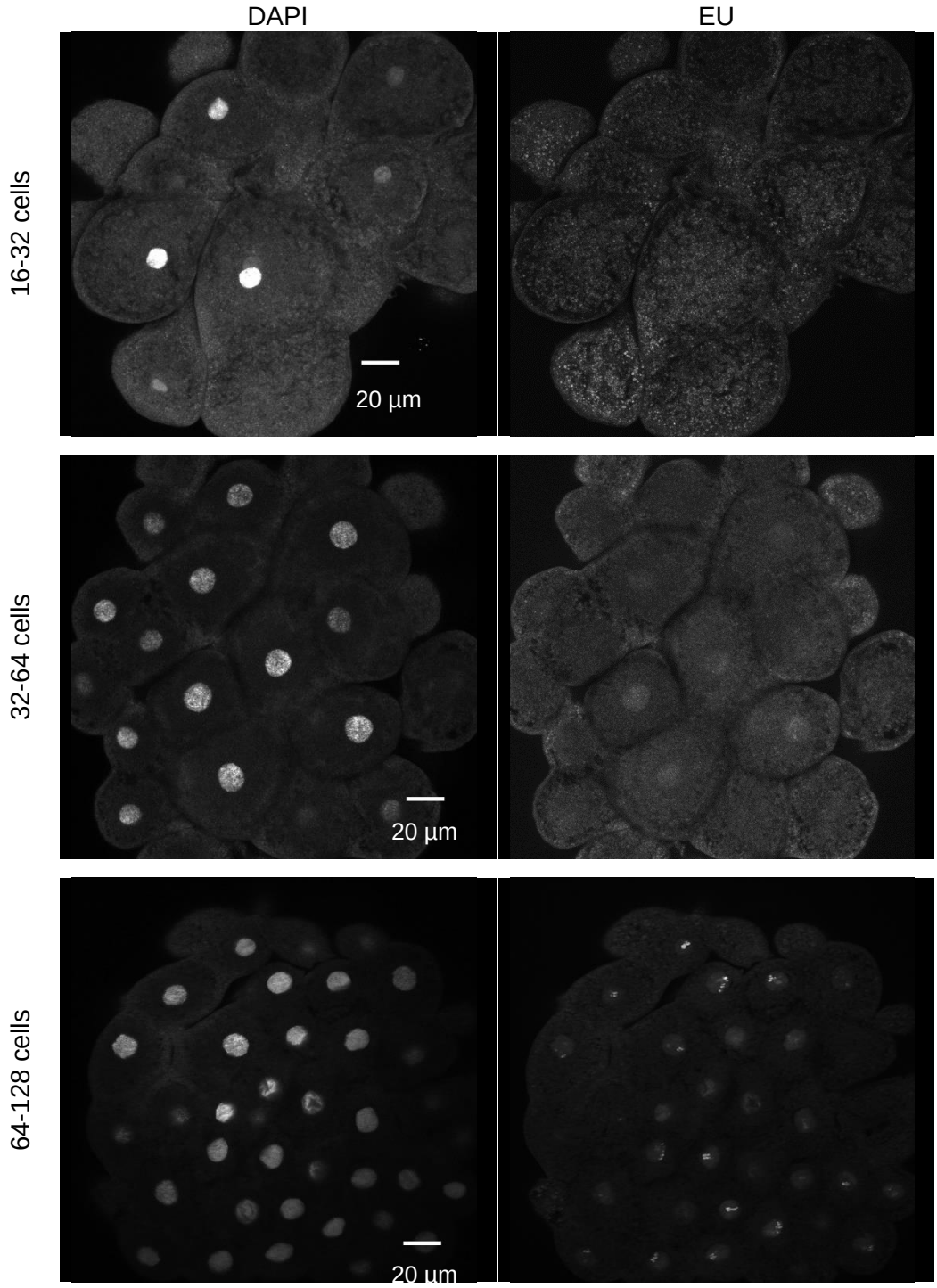


Figure 3.11. EU incorporation at 64 cells is indicative of major wave ZGA.

However, there is a possibility that 16 cells (or earlier) embryos have difficulties to take up the EU. Hence, I extent the soaking period to two hours, and used EdU as a control for penetration because EdU (marking nascent DNA) is structurally and chemically similar to EU. This experiment showed that EdU incorporation, but not EU, stained the 16 cell embryos (Figure 3.11). Furthermore, RNase treatment removed the EU signals at late embryonic stages but not the EdU. Therefore, EU incorporation at 64 cells

embryos marks nascent RNA, hence zygotic genome activation (Figure 3.11 & Figure 3.12).

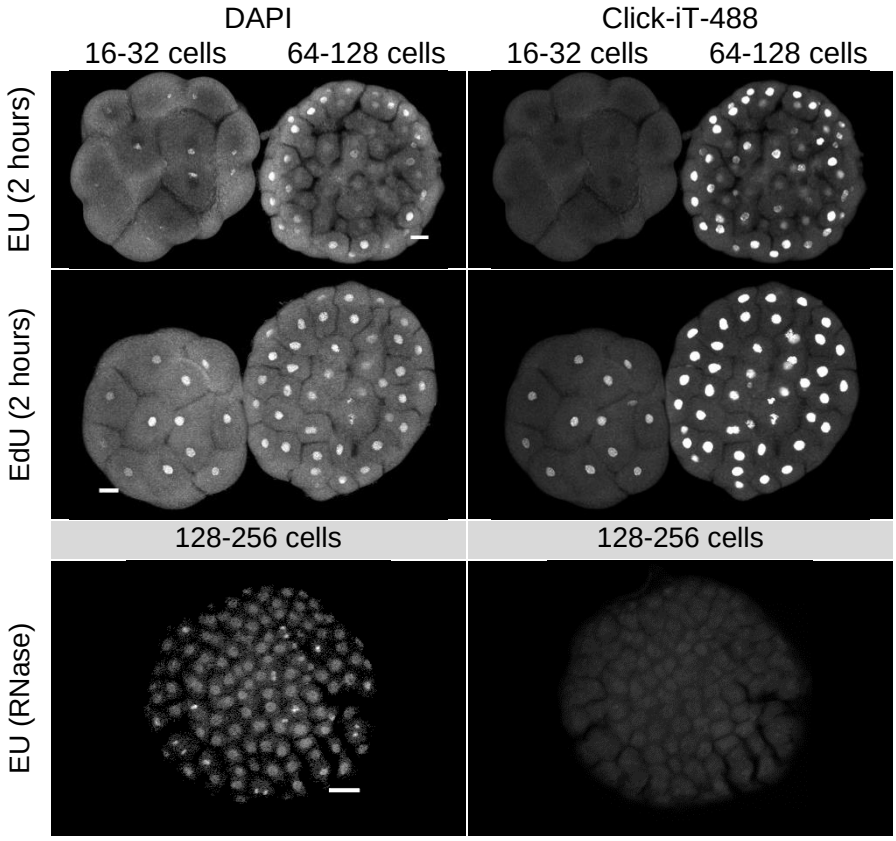


Figure 3.12 EU incorporation displays nascent RNA only from 64-128 cell stage.

3.9 Accumulation of 6mA inversely coincides with zygotic genome activation

Altogether, it is evidentially established that 6mA accumulation inversely coincides with EU incorporation and zygotic genome activation (Figure 3.13). Such a relationship can also be drawn from *Drosophila* and zebrafish published work (Figure 3.2). This indicates that ZGA occurs only after a rapid decline of 6mA levels in the zygotic genome of early embryos across animals. This result guided me to hypothesize that high levels of 6mA prevent transcription and their rapid decline allows transcription to commence at 64-128 cell embryos of *Hydractinia*. To test this hypothesis, I decided to miss-express genes that are expected to act as methylators and removers of 6mA in *Hydractinia* embryogenesis. For this, I searched the EBseqHMM data for genes associated with DNA methylation/demethylation during early embryogenesis.

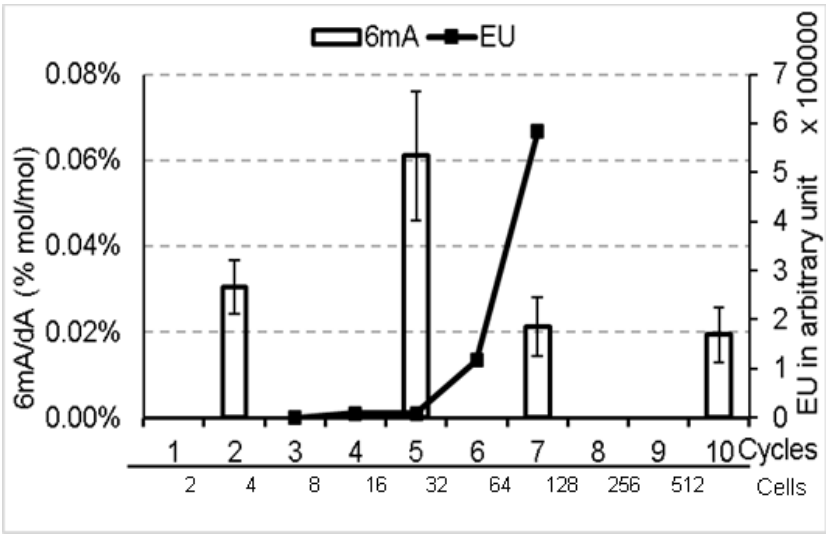


Figure 3.13. Inverse correlation between 6mA/dA and nascent RNA level at early embryogenesis of *Hydractinia*.

3.10 Candidate genes of DNA methylation at early embryonic stages

Revisiting the EBseqHMM, I analyzed the expression path of genes associated with DNA methylation. Logically, based on the dynamic of 5mC DNA methylation levels that steadily increase during early development (Figure 3.5), I expected that *Dnmt1* - the maintenance methyltransferase - would generally follow the maternal degradation expression group, but *Dnmt3* as the *de novo* methyltransferase would be expressed highly later, either at 64-128 cell following ZGA pattern or at 24 hpf following high-preplanula group.

The time course analysis, however, showed that *Dnmt1* and *3* are constantly downregulated, following the maternal degradation expression group decreasing to a near zero level (Figure 3.14). This raises a question on how the 5mC DNA methylation levels are maintained beyond 24 hpf embryos (Figure 3.5). Nonetheless, the high levels of both Dnmts at the 2-4 cell stages are consistent with the 5mC dynamics at early embryogenesis.

The 6mA methyltransferase candidate[14] *Mettl4*, however, displayed a puzzling expression pattern following the high-preplanula group with strong upregulation at 24 hpf (Figure 3.14), a time point when 6mA levels drops to background (Figure 3.5). The alternative candidate, *N6amt1*, by contrast, displayed a consistent expression path with 6mA dynamic during embryogenesis following the maternal degradation group (Figure 3.14). Thus, *N6amt1* (not *Mettl4*) is the better candidate to act as methyltransferase for

[14] Based on the conclusion from phylogenetic, domain, catalytic and localization signal analysis on Chapter 2.

6mA at early embryo of *Hydractinia* based on its expression path. These results, however, are inconsistent with the results from section 2.4.

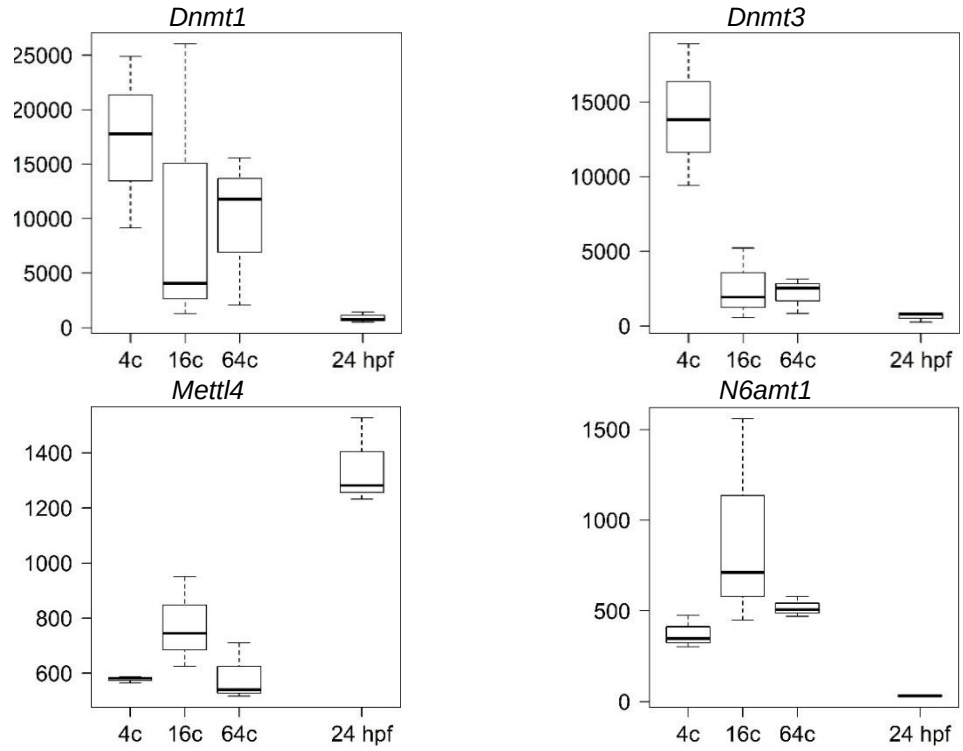


Figure 3.14. Expression path of methyltransferases transcripts

As the demethylation initiator, there were three candidate transcripts: *Tet* for 5mC, and *Alkbh1* and *Alkbh4* for 6mA. According to the dynamic of 5mC expression during embryogenesis, I expected *Tet* transcripts to remain steady along the embryogenesis, which met consistently with the expression group I found despite of a slight fluctuation (Figure 3.15). The expression group of *Alkbh1* and *4* were precisely as expected to be, increasing towards 64-128 cells and dropping down at 24 hpf (Figure 3.15).

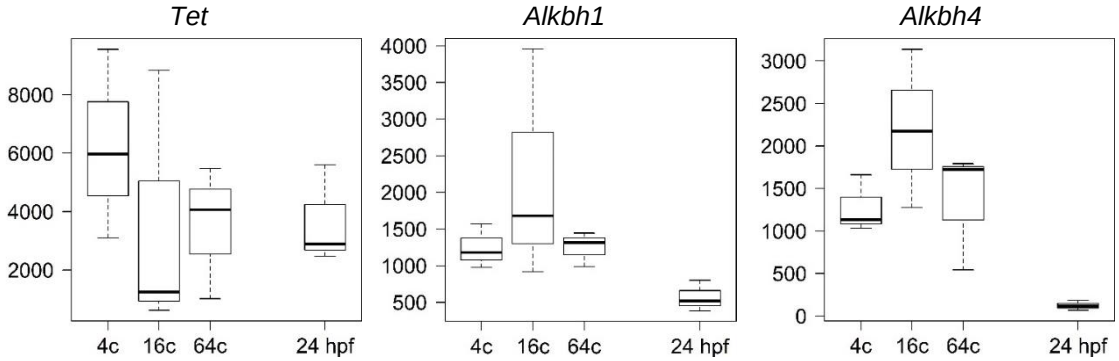


Figure 3.15. Expression path of demethylation transcripts.

Focusing on 6mA, I analysed the two candidate methyltransferases for 6mA. N6amt1 displayed no nuclear localization signal and is therefore more likely to localize to the cytoplasm or cytoskeleton (subchapter 2.4). The other candidate, Mettl4, despite predicted to have nuclear localization signals (Table 2.8), its transcript followed the high-

preplanula group, inconsistent with high levels of 6mA at the 16-32 cell stage (Figure 3.5). However, the 6mA demethylation initiator candidates, Alkbh1 and Alkbh4, displayed a fair number of signals for nuclear localization and their transcripts followed the expected expression group consistent with the dynamic level of 6mA at early embryogenesis of *Hydractinia*.

3.11 ZGA-6mA inverse correlation and promising ALKBHs

In summary, the results of this chapter emphasize the inverse correlation between the level of 6mA and nascent RNA (Figure 3.5,Figure 3.8, Figure 3.11 &Figure 3.13). Meanwhile, although *N6amt1*, *Alkbh1*, and *Alkbh4* display expression patterns consistent with 6mA dynamic in the course of *Hydractinia* embryogenesis (Figure 3.14 Figure 3.15), N6amt1 was predicted to localize to the cytoplasm or cytoskeleton instead of being nuclear (Table 2.8). Thus, both demethylating enzyme candidates, Alkbh1 and 4, are the best candidates to carry on functional investigations relating to 6mA. Therefore, I decided to knockdown their genes and confirm their expected functions in 6mA and ZGA during early embryogenesis, as will be investigated in the next chapter.

3.12 References

Ballard, W. 1942. 'The mechanism for synchronous spawning in *Hydractinia* and *Pennaria*. ', *The Biological Bulletin*, 82: 329-339.

Bullard, James H, Elizabeth Purdom, Kasper D Hansen, and Sandrine Dudoit. 2010. 'Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments', *BMC bioinformatics*, 11: 1-13.

Chen, Hui, Lily C. Einstein, Shawn C. Little, and Matthew C. Good. 2019. 'Spatiotemporal Patterning of Zygotic Genome Activation in a Model Vertebrate Embryo', *Developmental Cell*, 49: 852-866.e857.

DuBuc, Timothy Q, Christine E Schnitzler, Eleni Chrysostomou, Emma T McMahon, Febrimarsa, James M Gahan, . . . Uri Frank. 2020. 'Transcription factor AP2 controls cnidarian germ cell induction', *Science*, 367: 757-762.

Duffy, David J., Günter Plickert, Timo Kuenzel, Wido Tilmann, and Uri Frank. 2010. 'Wnt signaling promotes oral but suppresses aboral structures in *Hydractinia* metamorphosis and regeneration', *Development*, 137: 3057-3066.

Fernández, Juan, and Ricardo Fuentes. 2013. 'Fixation/permeabilization: new alternative procedure for immunofluorescence and mRNA in situ hybridization of vertebrate and invertebrate embryos', *Developmental Dynamics*, 242: 503-517.

Fritzenwanker, Jens H., Grigory Genikhovich, Yulia Kraus, and Ulrich Technau. 2007. 'Early development and axis specification in the sea anemone *Nematostella vectensis*', *Developmental Biology*, 310: 264-279.

Fu, Ye, Guan-Zheng Luo, Kai Chen, Xin Deng, Miao Yu, Dali Han, . . . Louis C Doré. 2015. 'N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*', *Cell*, 161: 879-892.

Gahan, James M., Christine E. Schnitzler, Timothy Q. DuBuc, Liam B. Doonan, Justyna Kanska, Sebastian G. Gornik, . . . Uri Frank. 2017. 'Functional studies on the role of Notch signaling in *Hydractinia* development', *Developmental Biology*, 428: 224-231.

Giraldez, Antonio J, Yuichiro Mishima, Jason Rihel, Russell J Grocock, Stijn Van Dongen, Kunio Inoue, . . . Alexander F Schier. 2006. 'Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs', *Science*, 312: 75-79.

Hao, Ziyang, Tong Wu, Xiaolong Cui, Pingping Zhu, Caiping Tan, Xiaoyang Dou, . . . Chuan He. 2020. 'N6-deoxyadenosine methylation in mammalian mitochondrial DNA', *Molecular cell*.

He, Shunmin, Guoqiang Zhang, Jiajia Wang, Yajie Gao, Ruidi Sun, Zhijie Cao, . . . Dahua Chen. 2019. '6mA-DNA-binding factor Jumu controls maternal-to-zygotic transition upstream of Zelda', *Nature communications*, 10: 2219.

Jukam, David, S Ali M Shariati, and Jan M Skotheim. 2017. 'Zygotic genome activation in vertebrates', *Developmental Cell*, 42: 316-332.

Koziol, Magdalena J., Charles R. Bradshaw, George E. Allen, Ana S. H. Costa, Christian Frezza, and John B. Gurdon. 2016. 'Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications', *Nature Structural & Molecular Biology*, 23: 24-30.

Kraus, Yulia, Sandra Chevalier, and Evelyn Houliston. 2019. 'Cell shape changes during larval body plan development in *Clytia hemisphaerica*', *bioRxiv*: 864223.

Kraus, Yulia, Hakima Flici, Katrin Hensel, Günter Plickert, Thomas Leitz, and Uri Frank. 2014. 'The embryonic development of the cnidarian *Hydractinia echinata*', *Evolution & development*, 16: 323-338.

Li, Bo, and Colin N Dewey. 2011. 'RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome', *BMC bioinformatics*, 12: 323.

Liu, Jianzhao, Yuanxiang Zhu, Guan-Zheng Luo, Xinxia Wang, Yanan Yue, Xiaona Wang, . . . Ye Fu. 2016b. 'Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig', *Nature communications*, 7: 1-7.

McMahon, Emma. 2018. 'Characterization of PIWI+ stem cells in *Hydractinia*', National University of Ireland Galway.

O'Brown, Zach K, Konstantinos Boulias, Jie Wang, Simon Yuan Wang, Natasha M O'Brown, Ziyang Hao, . . . Eric Lieberman Greer. 2019. 'Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA', *BMC genomics*, 20: 445.

Owens, Nick DL, Ira L Blitz, Maura A Lane, Ilya Patrushev, John D Overton, Michael J Gilchrist, . . . Mustafa K Khokha. 2016. 'Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development', *Cell reports*, 14: 632-647.

Plickert, Günter, Vered Jacoby, Uri Frank, Werner A. Müller, and Ofer Mokady. 2006. 'Wnt signaling in hydroid development: Formation of the primary body axis in embryogenesis and its subsequent patterning', *Developmental Biology*, 298: 368-378.

Plickert, Gunter, Eva Schetter, Nicole Verhey-Van-Wijk, Jorg Schlossherr, Marlis Steinbuchel, and Martin Gajewski. 2004. 'The role of alpha-amidated neuropeptides in hydroid development--LWamides and metamorphosis in *Hydractinia echinata*', *International Journal of Developmental Biology*, 47: 439-450.

Presolski, Stanislav I, Vu Phong Hong, and MG Finn. 2011. 'Copper-catalyzed azide–alkyne click chemistry for bioconjugation', *Current protocols in chemical biology*, 3: 153-162.

Puchtler, Holde, Faye Sweat Waldrop, Susan N Meloan, Mary S Terry, and HM Conner. 1970. 'Methacarn (methanol-Carnoy) fixation', *Histochemie*, 21: 97-116.

Ribeiro, Lupis, Vitoria Tobias-Santos, Daniele Santos, Felipe Antunes, Georgia Feltran, Jackson de Souza Menezes, . . . Rodrigo Nunes da Fonseca. 2017. 'Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects', *PLoS genetics*, 13: e1006868.

Schiffers, Sarah, Charlotte Ebert, René Rahimoff, Olesea Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, . . . Thomas Carell. 2017. 'Quantitative LC-MS provides no evidence for m6dA or m4dC in the genome of mouse embryonic stem cells and tissues', *Angewandte Chemie International Edition*, 56: 11268-11271.

Schneider, Caroline A, Wayne S Rasband, and Kevin W Eliceiri. 2012. 'NIH Image to ImageJ: 25 years of image analysis', *Nature methods*, 9: 671-675.

Schulz, Katharine N., and Melissa M. Harrison. 2019. 'Mechanisms regulating zygotic genome activation', *Nature Reviews Genetics*, 20: 221-234.

Wang, Wei, Liang Xu, Lulu Hu, Jenny Chong, Chuan He, and Dong Wang. 2017a. 'Epigenetic DNA modification N 6-methyladenine causes site-specific RNA

polymerase II transcriptional pausing', *Journal of the American Chemical Society*, 139: 14436-14442.

Wang, Yuanyuan, Xiao Chen, Yalan Sheng, Yifan Liu, and Shan Gao. 2017b. 'N6-adenine DNA methylation is associated with the linker DNA of H2A. Z-containing well-positioned nucleosomes in Pol II-transcribed genes in *Tetrahymena*', *Nucleic Acids Research*, 45: 11594-11606.

Wu, Tao P, Tao Wang, Matthew G Seetin, Yongquan Lai, Shijia Zhu, Kaixuan Lin, . . . Mei Zhong. 2016. 'DNA methylation on N 6-adenine in mammalian embryonic stem cells', *Nature*, 532: 329-333.

Xie, Qi, Tao P. Wu, Ryan C. Gimple, Zheng Li, Briana C. Prager, Qiulian Wu, . . . Jeremy N. Rich. 2018. 'N6-methyladenine DNA Modification in Glioblastoma', *Cell*, 175: 1228-1243.

Zanini, Cristina, Elisa Gerbaudo, Elisabetta Ercole, Anna Vendramin, and Marco Forni. 2012. 'Evaluation of two commercial and three home-made fixatives for the substitution of formalin: a formaldehyde–free laboratory is possible', *Environmental Health*, 11: 59.

Zhang, Guoqiang, Hua Huang, Di Liu, Ying Cheng, Xiaoling Liu, Wenxin Zhang, . . . Jianzhao Liu. 2015. 'N6-methyladenine DNA modification in *Drosophila*', *Cell*, 161: 893-906.

Zhang, Min, Shumin Yang, Raman Nelakanti, Wentao Zhao, Gaochao Liu, Zheng Li, . . . Haitao Li. 2020. 'Mammalian ALKBH1 serves as an N6-mA demethylase of unpairing DNA', *Cell Research*.

Zhao, Boxuan Simen, Xiao Wang, Alana V Beadell, Zhike Lu, Hailing Shi, Adam Kuuspalu, . . . Chuan He. 2017. 'm6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition', *Nature*, 542: 475-478.

4 ALKBH1 REMOVES GENOMIC CONTAMINATION OF 6MA, PROMOTING ZGA

I established that 6mA levels inversely coincide with nascent RNA levels during early embryogenesis (Chapter 3). The genes associated with 6mA, especially on the demethylation side, indicate promising characteristics in terms of nuclear localization signals (Chapter 2) and their high expression levels during early embryogenesis (Chapter 3). Thus, in this chapter the main focus is to alter the expression level of these genes (*Mettl4, N6amt1, Alkbh1* and *Alkbh4*), then study the change in the EU incorporation phenotype of *Hydractinia* embryos at the respective stages (Figure 4.1). Thus, I had to choose and select the better strategy to miss-express these genes from the plethora of tools available in *Hydractinia* as an emerging model organism (Ballard 1942; Sanders et al. 2018; Frank et al. 2020).

4.1 Strategy of the functional investigation of 6mA associated genes

The Frank Lab have developed molecular tools to knockdown, knockout, ectopically express and overexpress genes in *Hydractinia*. Knockout and knockin by CRISPR-Cas9 system have been established (Gahan et al. 2017; Sanders et al. 2018) but the lengthy experimental time is discouraging as only at second generation one can observe the phenotype at the early embryo. Ectopic expression of genes in specific cell types (Millane et al. 2011; Kanska and Frank 2013) that are associated with DNA methylation is not necessary for functional investigation at the early embryo as the expression and the efficacies will be very likely broad and not in a cell type specific manner.

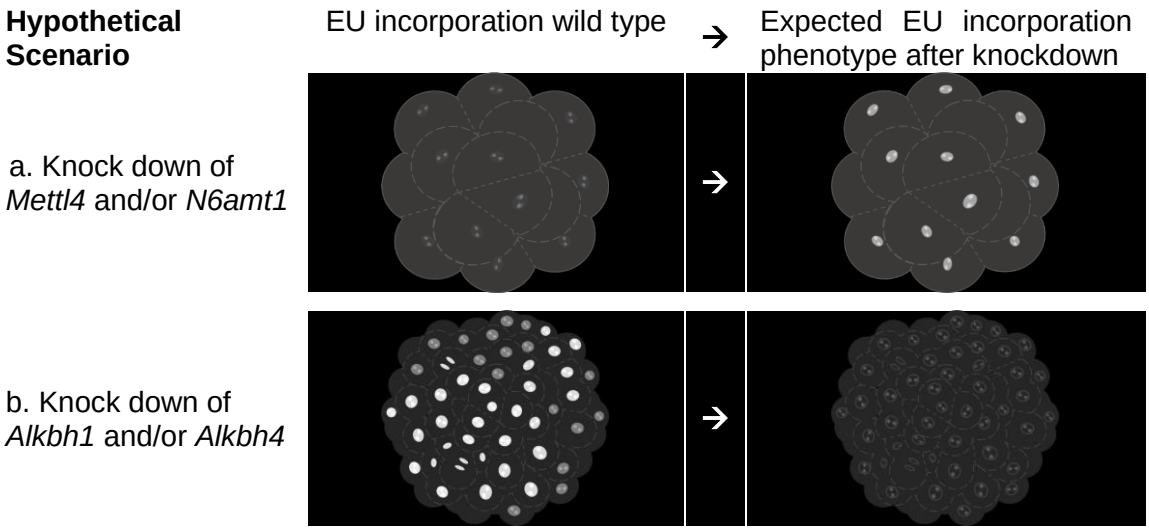


Figure 4.1. Expected EU incorporation phenotype after knocking down experiment. a. Expected EU incorporation after knocking down of *Mettl4* or *N6amt1* at 16-32 cell stage of embryos b. Expected EU incorporation after knocking down of *Alkbh1* or *Alkbh4* at 64-128 cell stage of *Hydractinia* embryos.

Overexpression by mRNA microinjection in *Hydractinia* have only short period of strong expression beyond 12 hpf embryos (McMahon 2018). However, a knockdown system would give a definitive answer to the research question I asked. Knocking down methyltransferases of 6mA, for instances, is hypothesized to decrease the level of 6mA at 16-32 cell embryos. If the relationship between 6mA and ZGA is functional, then this knockdown would also accelerate ZGA to commence at 16-32 cell stage (Figure 4.1a). Conversely, knockdown of methylases *Alkbh1* or *Alkbh4* would let the 6mA level at 16-32 cell stage be carried on to 64-128 cell (and, perhaps, beyond), thus delaying (or stopping altogether) the ZGA (Figure 4.1b).

Knocking down a gene in early embryogenesis of *Hydractinia* can be done by morpholino antisense oligomer (Kanska and Frank 2013). However, in 2018, a novel, less expensive and perhaps more effective short hairpin RNA (shRNA) mediated knock down of genes have been established for *Nematostella vectensis* (He et al. 2018). Knockdown mediated by shRNA is fundamentally based on the regulation of mRNA by miRNA, which in cnidarians work in high (19-21 nucleotides) complementarity as oppose to the 8-9 nucleotide complementary miRNA in bilaterians (Moran et al. 2014). The conservation of the miRNA machineries among cnidarians are very likely (Moran et al. 2017; Tripathi et al. 2020). Thus, I decided that establishing this shRNA mediated knockdown method for *Hydractinia* would be an advantage for my investigation during early embryogenesis.

4.2 Methods

4.2.1 Retrieving sequence information of gene of interest (*goi*) from genomic and transcriptomic of *Hydractinia*.

Our collaborators at NIH were the ones who sequenced the genome of *Hydractinia*. Our lab members have access to the yet unpublished assembled genome of *Hydractinia symbiolongicarpus* from the NHGRI webpage[15]. This assembly was also used as reference genome to perform SMRTseq, IPD, and Motif maker analysis to identify 6mA in the genome of *Hydractinia* extracted from adult polyps, 16-32 cell and 64-128 cell embryos. This reference genome and the transcriptome assembled from Chapter 3 were set as searchable database in Geneious. Genome information and transcripts sequence from genes of interest (*goi*) were retrieved from these databases for downstream work, i.e. designing template cassettes for in vitro mRNA synthesis.

[15] https://research.nhgri.nih.gov/hydractinia/download/index.cgi?dl=s_assembly.

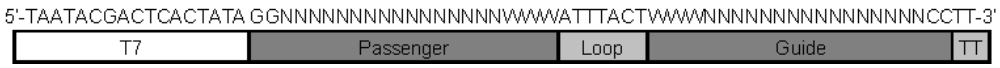
4.2.2 Designing shRNA for transient knock down of gene expression

Transcripts of *goi* were retrieved from the transcriptomes and used to find siRNA motif at http://www.invivogen.com/sirnawizard/design.php. The default 21 nucleotides were chosen for desired motif size. I left blank the option of mRNA database and miRNA SEED database. From the output motif, few of them were picked based on three characteristics: higher GC content, begin the sequence with 'GG', and having the last 3 nucleotides AT rich. At the 5' of the passenger sequence, GG dinucleotides are required for high efficacy T7 in vitro transcription reaction. GGG trinucleotide would be better but GG or GC are also effective. At the 3' of the passenger sequence, AT richness is necessary to enable the cell distinguishing passenger from guide sequence.

Few selected motif output sequences were BLASTed against the *Hydractinia symbiolongicarpus* transcriptomes with loose parameter (e.g. expected value 1). To confirm specificity, the first top hit with expected value equal or lower than 1×10^{-1} had to be the *goi* that was used to find siRNA motif at the beginning, while the rest of the hit list had to show expected values significantly higher than $1x10^{-1}$. All other motif outputs that did not pass this BLAST search criteria were discarded. The best motif output was then BLASTed against the miRNA database of *Hydractinia symbiolongicarpus*[16]*.* The motif that did not get any hit from the miRNA database was then chosen for next step.

The selected 21-nucleotide motif served as the passenger sequence of the shRNA after replacement of all Ts with Us. The guide sequence of the shRNA was generated by reverse complementing the passenger sequence and pasted at the 3' of the passenger sequence. The sequence that would generate loop secondary structure [5'-AUUUACU-3'] was then added between passenger and guide sequences. The full sequence of an shRNA was completed after addition of 'UU' dinucleotide at the 3' of the sequence to create overhang.

a.



b.

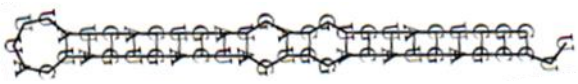


Figure 4.2. Design of shRNA
(a) Final DNA primary structure of the shRNA cassette. (b) Example of expected output from secondary structure assessment.

The full shRNA sequence was then assessed for the secondary structure folding in http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi using default parameters.

[16] Provided by Dr. Christine Schnitzler, University of Florida. USA.

One or two mismatches were added in the middle of the passenger sequence to generate a small bulb in the middle of the hairpin structure (Figure 4.2b). The sequence was finalized by reverting it back to DNA sequence (by replacing U to T) and T7 promoter [5'-TAATACGACTCACTATA-3'] was added at the 5' of the sequence to form the shRNA cassette (Figure 4.2). The complete list of shRNA sequences used in this chapter is available at Table 4.1.

Table 4.1. List of shRNA

Name	Endogenous Target	DNA cassette for shRNA production
sh*Gfp*	GATGACGGGAACTACAAGACA	TAATACGACTCACTATAGGATGACG**C**GA**T**CT **G**CAAGACAATTTACTTGTCTTGTAGTTCCCG TCATC*TT*
sh*Mock*	GATGACGGGA**G**CTACAAGACA	TAATACGACTCACTATAGGATGACGGGAGCT ACAAGACAATTTACTTGTCTTGTAGCTCCCG TCATC*TT*
sh*Mettl4*	GAGAACTCTGTTACGTACTCA	TAATACGACTCACTATAGAGAACTCTG**C**TA**G** GTACTCAATTTACTTGAGTACGTAACAGAGT TCTC*TT*
sh*N6amt1*	GCTTCATATGCCACTCTTCAA	TAATACGACTCACTATAGCTTCATATG**G**CA**G** T**G**TTCAAATTTACTTTGAAGAGTGGCATATG AAGC*TT*
sh*Alkbh1*	GGCTCATGTCCACTAGTCACT	TAATACGACTCACTATAGGCTCATGTGCAGT AGTCACTATTTACTAGTGACTAGTGGACATG AGCC*TT*
sh*Alkbh4*	GGACCTTTCTCACGTTGTCTT	TAATACGACTCACTATAGGACCTTTCTC**T**CG **A**TGTCTTATTTACTAAGACAACGTGAGAAAG GTCC*TT*

T7sequence, **mismatches**, loop, *UU dinucleotide tail*.

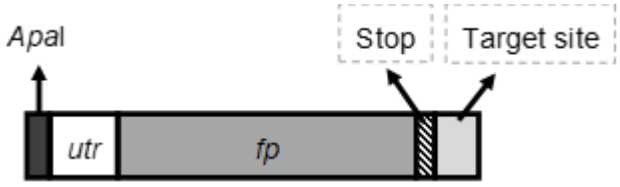
4.2.3 Designing plasmid for mRNA synthesis

There are two purposes of constructing mRNA in this project. First, mRNA encoding fluorescence protein followed by sequence targeted by the shRNA so when co-injected with the respective shRNA displayed the efficacy of shRNA in knocking down the gene. Secondly, mRNA of *goi* with silent mutation at the specific sequence targeted by the shRNA, thus evading the knockdown effect as a rescue strategy.

I used the transcripts and genome information of *goi* from *Hydractinia* to design an mRNA expression cassette that will be inserted into a plasmid vector. For this project, pGEM-T-easy (Promega #A137A) was used as the backbone to construct templates for mRNA synthesis (Figure 4.3). The map of pGEM-T-easy (Appendix 6.12, see page 139) indicates that the nearest restriction site to the 3' end of T7 promotor sequence is *Apa*I, thus I designed the expression cassettes beginning with an *Apa*I site (5'G^GGCC^C'3) so the cassette placed as near as possible to the T7 promotor. The rest of the cassettes were constructed as described in Figure 4.3. Briefly, shRNA confirmation cassette mainly comprises of gene encoding a fluorescent protein (*fp*) and passenger sequence of the shRNA without the mismatches after the stop codon (Figure 4.3a). The rescue cassette comprised of a gene encoding the methyltransferases or the oxidoreductases (*goi*) with

two silent mutations at the sequence corresponding to the shRNA to render it resistant to the shRNA that was targeted against the endogenous gene (Figure 4.3b). The complete sequence for the cassettes that was actually used in this study are presented in Appendix 6.13 (see page 143).

a. mRNA expression cassette for shRNA confirmation



b. mRNA expression cassette for rescue strategy

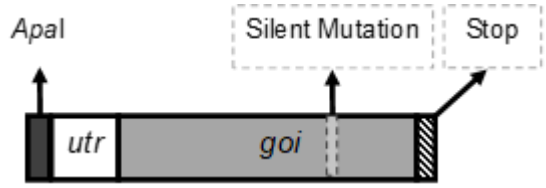


Figure 4.3. Design of mRNA expression cassettes.
utr: short 5' untranslated region (5'TGCAGCCCCGGTAGAAAAA'3). *Apa*I: restriction site (5'G$_{\wedge}$GGCC$^{\vee}$C'3). *fp*: gene encoding fluorescent protein. *goi*: gene of interest.

4.2.4 DNA cassette template preparation for shRNA synthesis

Forward and reverse oligonucleotides corresponding to the shRNA cassette sequence were synthesized using commercial services. Upon arrival, both (forward and reserve) oligonucleotides were dissolved separately in ultrapure nuclease free water to a final concentration of 100 μM. To create a template DNA cassette solution, those 100 μM oligonucleotides were diluted 1:10 in nuclease-free water while combining forward and reserved oligos. This combined oligonucleotide solution was then denatured at 98°C for 5 minutes and cooled down at room temperature for 10 minutes. This solution was then immediately used as a template for T7 based in vitro transcription (IVT) synthesizing the RNA.

4.2.5 DNA cassette template preparation for mRNA synthesise

To synthesis the shRNA confirmation cassettes, I added the specific target site (Figure 4.3a) to reverse primers amplifying the *fp* sequence (in this project specifically I used *mScarlett*). Sequence of *Apa*I and *utr* were also added to forward primers amplifying the *fp* (Table 4.2, no.1&2). These primer pairs were then used to amplify *fp* using available in-house plasmid as template (Appendix 6.13.1, see page 143). Normal *fp* primer pairs (Table 4.2, no.3&4) were used to amplify the plasmid to act as control for size of the fragments selected downstream during agarose gel electrophoresis.

To synthesize the rescue mRNA expression cassettes, a double stranded DNA gBlock fragment was ordered from IDT (Integrated DNA Technology, https://www.idtdna.com)

with sequence corresponding to the design (Figure 4.3b and Appendix 6.13.2, see page 148). At the same time, I also ordered a pair of oligonucleotide primers (Table 4.2, no 5&6) corresponding exactly to the gBlock fragment. Upon arrival, the gBlock fragment was immediately dissolved, amplified, and loaded into agarose gel for electrophoresis.

The selected fragments (for shRNA confirmation cassette) and the amplified gBlock fragment (for rescue cassette) were then extracted from the agarose gel, A-tailed[17], rapidly ligated into pGEM-T-easy, and inserted into XL1-Blue *E. coli* using standard heat shock transformation protocol. The *E. coli* colonies were then selected by Blue/White screening of X-Gal/IPTG layer on the LB-plate (all LBs either plate or broth contained 100 µg/ml carbenicillin). Several white colonies were then transferred to another LB-plate, picked for PCR using T7 and SP6 primers, and transferred to 4 ml LB-broth culture, which was then incubated at 37°C and 250 rpm for 18 hours. During this incubation period, colony PCRs were performed, and the amplified fragments size after agarose gel electrophoresis used as basis for culture selection. The selected LB broth cultures were then used for plasmid extraction. These plasmids were sent to Eurofins for Sanger sequencing from both (T7 and SP6) direction to ensure sequence and directionality of the cloned fragment.

Table 4.2. Oligonucleotides used as primers for PCR.

No.	Name	Sequence (5' → '3)
1	Scarlett-alkbh1-F	TTGGGCCCGGATCTGAAAAAATGGTATCTAAAGGTG
2	Scarlett-alkbh1-R	AGTGACTAGTGGACATGAGCCCTTAGTATAGTTCATCC
3	Scarlett-F	GATCTGAAAAAATGGTATCTAAAGGTG
4	Scarlett-R	CTTAGTATAGTTCATCCATTCCTCCAGT
5	gB-Alkbh1-rescue-F	TAGGGCCCTGAGTTTAAGATCA
6	gB-Alkbh1-rescue-R	GGCCTGCAGGTCAAAACAG
7	Hydractinia-Splice Leader	ACTATTTCTAGGTCCCTGAGTTTAAG
8	T7-F	TAATACGACTCACTATAGGG
9	SP6-R	ATTTAGGTGACACTATAGAAT

The correctly orientated plasmids were then digested with ApaI. The linearized, and lighter, plasmid was then extracted from an agarose gel, re-ligated, cloned and sequenced. The plasmid with shortest distance between T7 promotor sequence and *utr* was then renamed and linearized using SdaI (an *Sda*I restriction site is available at the 3' region of the insertion site in the pGEM-T-easy map). This linearized plasmid was then used as the template for T7 IVT reaction.

[17] A-tailing is a step necessary only if the fragments produced by a PCR using high-fidelity polymerase.

4.2.6 T7/SP6 In Vitro Transcription (RNA synthesis)

For shRNA synthesis, T7/SP6 IVTs were performed at triple to quadruple of the recommended volume to get higher yield of RNA. For mRNA synthesis, the T7/SP6 IVT was performed at the recommended volume, but in duplicate. Table 4.3 below describes the recipe for the T7/SP6 IVT solution at the recommended volume as per the manufacturer's instruction (NEB #E2040S & #E2060S).

Table 4.3. Recipe of T7/SP6 In vitro transcription (IVT) solution for 1x reaction.

Component	**Volume (µl)**	**Final concentration**
Nuclease-free H_2O	8-x	NA
dsDNA template	x	5-10 µM or 1 µg
NTP (for mRNA synthesis use Arca/NTP instead)	8	10 mM each
T7/SP6 Buffer	2	1 x
T7/SP6 Polymerase mix	2	Unknown
Total:	20	

For shRNA synthesis, the resulted IVT solutions were then immediately used for RNA extraction step. For mRNA synthesis, the IVT solutions were subjected to DNase treatment and Poly A tailing before the total RNA was extracted.

4.2.7 RNA extraction

The IVT suspension turned very viscous after the reaction. Thus, the IVT suspension was diluted with nuclease-free H_2O to made up 200 µl solutions. One and a half volume of ethanol absolute was then added into the solutions and mixed. Afterwards, the whole solution was transferred into an RNA column in 2 ml tube and the RNA bound into the column membrane by centrifugation at 8,000 x *g* for 2 minutes then the flow-through discarded. Afterwards, 400 µl of wash-r1 buffer (1 M Gu-HCl; 10 mM Tris-HCl pH 7.0; in nuclease-free water) was added into the column then centrifuged at 11,000 x *g* for 1 minute. Then, 80 µl DNaseI solution was added (5 µl DNaseI + 75 µl digestion buffer) into the centre of the column and incubated at room temperature for at least 30 minutes. Thereafter, the column was washed with 500 µl of wash-r1 buffer then twice with 500 µl of wash-r2 buffer (10mM Tris-HCl pH 7.5; 60 mM potassium acetate; 80% Ethanol; in nuclease-free water). These washes were done by centrifugation at 11,000 x *g* for 1 minute. Then, the column was dried by one more centrifugation at maximum speed for 2 minutes. Finally, the RNA eluted from the column with 40 µl nuclease-free water by 11,000 x *g* centrifugation for 1.5 minutes.

4.2.8 RNA assessment

The eluted RNA solutions were assessed by Nanodrop (on RNA mode), Qubit microRNA assay kit (for shRNA), Qubit RNA high sensitivity assay kit (for mRNA) and RNA-gel electrophoresis (for shRNA using low molecular weight RNA marker). The RNAs were

then immediately aliquoted in several nuclease-free 1.5 ml tube and stored in -70°C. I took one tube at a time for preparing injection solution to avoid multiple freeze/thaw cycles.

4.2.9 Preparation of microinjection solution

Microinjection solutions were prepared freshly on ice with the following recipes (Table 4.4).

Table 4.4. Injection Solution Recipe

	Volume [µl]	**Final Concentration**
Nuclease-free water	9-x	NA
Alexafluor555-Dextran	0.5	10 ng/µl
KCl (2 mM)	0.5	0.1 mM
shRNA (and mRNA)	x	0.5-1.0 µg/µl (and 0.8-1.2 µg/µl)

4.2.10 Microinjector Preparation

The microinjector (Narishige IM 300) was filled with pressurized air at 60 psi. The prepared injection capillary (the needle) was filled with injection solution using a micro loader pipette tip. The needle was then inserted into the capillary holder, which is connected to the microinjector, then the capillary holder (with the needle) was secured into a micromanipulator. The microinjector was switched on and the pressure in the capillary holder was balanced, then the microinjector was altered into the injection mode. I pushed the foot pedal and adjusted the injection pressure to ~40 psi.

4.2.11 Microinjection

Injection dish (35 mm petri dish with a 180 µm plankton net attached) was filled with pre-chilled artificial sea water (ASW). Fertilized eggs were pipetted into the dish using a glass pipet. This injection dish then was set on the stage of a stereomicroscope and the position of the injection dish was adjusted so that the embryos got into the field of view. Most embryos were positioned onto the pockets of a preglued 200 µm plankton net by gently wiggling the dish. The micromanipulator (Narishige MN-151) was adjusted to lower the capillary holder until the tip of the needle was visibly submerged in ASW through the stereo microscope, while positioning the tip into the center of the field of view inside a pocket of the net. Then, by adjusting the micromanipulator, the tip of the needle lifted off the sea water and cut with curved micro scissors (Ted Pella 1341, 80 mm) then the microinjector foot pedal was pushed once to let a droplet form at the tip of the needle. The formed droplet was re-sucked up if the cutting was too low or increased in size if the cutting was too far away from the tip. Thus, I incrementally cut the tip from low until one whole droplet became visible (~2-10 µm) from one push of the injector foot pedal. If the

droplet uncontrollably increased in size, the needle (with the solution) was replaced. At the microinjector, injection pressure/time can also be used to control the droplet size.

By the (up-down) adjuster of the micromanipulator, the needle was lowered penetrating the embryo and the droplet of injection solution was inserted into the embryos by one push of the injector foot pedal. Then, the needle was moved up to be released from the injected embryos. Afterwards, the injection dish was slightly moved to bring the next embryo into the field of injection and the needles were lowered again for the next injection. Once most of the embryos were injected, the injection dish was viewed under an epifluorescence stereo microscope to sort out the uninjected embryos from the injected, based on tracer dye fluorescence.

4.2.12 Metamorphosis and main culture tank of husbandry

Embryos with a clear visible fluorescent dye (through epifluorescence stereo microscope) were transferred from the injection dish to an ASW filled Petri dish, which was then moved into a 16-18 °C incubator. Stock solution of CsCl (580 mM) was prepared in deionized water. This stock CsCl solution was freshly diluted with artificial sea water in a 1 to 5 dilution (aiming for ~100 mM CsCl in artificial sea water). Three days old planula larva (Figure 4.4a) were transferred into freshly prepared 100 mM CsCl solution in a glass Petri dish, which was then incubated in room temperature for 3 hours. After three hours incubation, the contracting planula larvae were observed every 15 minutes through stereo microscope until they completely contracted (Figure 4.4b). They were then transferred onto a prelabelled microscope glass slide in a fresh artificial sea water in a larger container. The whole container with the metamorphosing larva was left untouched for 24 hours.

a. Planula larva (3 dpf) b. Contracted planula c. Primary polyps

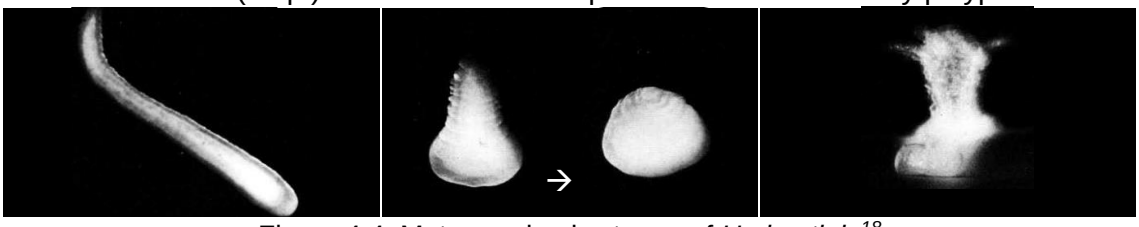


Figure 4.4. Metamorphosis stages of *Hydractinia*[18]
Metamorphosis in *Hydractinia.* (a) normal planula larva. (b) contracted planula, 2 to 3 hours after initiation of CsCl treatment. (C) Metamorphosed polyps at 1-2 days after metamorphosis induction. Adapted from (Müller 1984).

After 24 hours, primary polyps developed on the glass slide (Figure 4.4c), which was then transferred into a glass rack in a bowl of sea water. The primary polyps were fed with filtered-suspended oyster puree for ~15 minutes in sea water while shaking or

[18] Permission to reproduce this figure 4.4 has been granted by The Company Biologists Ltd.

aerating. Afterwards, the whole glass rack was transferred into another container filled with fresh and clean sea water. After few hours, the whole glass rack was transferred into the main culture system of *Hydractinia* and followed routine feeding regime.

4.2.13 Main culture system of *Hydractinia* colonies maintenance

Clones of wild type/transgenic *Hydractinia* are grown on glass microscope slides in glass racks (Hauenschild 1956; Toth 1965) inside 500-liter sea water culture tank system. The culture tank system is in a room with the light being controlled by an electric timer set to be on at 8.30 am every morning and off at 20.30 pm every evening, allowing a 12:12 light:dark regime. The food was prepared by hatching *Artemia* cysts in sea water in 2 litre brine shrimp hatchers with aeration incubated in room temperature for 48 hours. The hatched, actively moving nauplii were harvested, avoiding any remaining cyst and chicken egg powder was added to the freshly hatched *Artemia*. These egg coated nauplii were fed to the *Hydractinia* colonies four times a week in the morning after spawning (Ballard 1942; Hauenschild 1956; Toth 1965). Once a week, however, *Hydractinia* were fed with finely blend oyster filtered through 100 µm plankton net[19]. The colonies were observed under epifluorescence stereo microscope and cleaned from algae and any other contamination once fortnightly. Sexual polyps were developing in the colonies 2-3 months after metamorphosing. The colonies that reached sexual maturity were used for cross breeding between various lines of animals. *Hydractinia* colonies were crossbred by spawning the male and female colonies in one isolated sea water column.

4.3 Establishment of shRNA mediated knock down

To prove the efficacy of shRNA mediated knockdown in *Hydractinia*, I designed an experiment in which *Gfp* was knocked down in embryos that were offspring of a heterozygous male transgenic animal expressing GFP in i-cells under the *Piwi1* promoter, crossed with a wild type female, which generated an embryo population of which 50% expressed GFP, as expected from Mendelian genetics (Figure 4.5a). I injected these embryos with sh*Gfp*. The results showed indeed that 50% of the non-injected embryos displayed a GFP+ phenotype (Figure 4.5b), while all those that were injected with sh*Gfp* displayed a wild type phenotype without any apparent GFP+ cells (Figure 4.5c). Half of this population, however, displayed a GFP+ phenotype 7-15 days after injection, indicating the timeframe where shRNA remains effective (Figure 4.5d).

[19] This oyster puree can be stored in -20°C for 6 months.

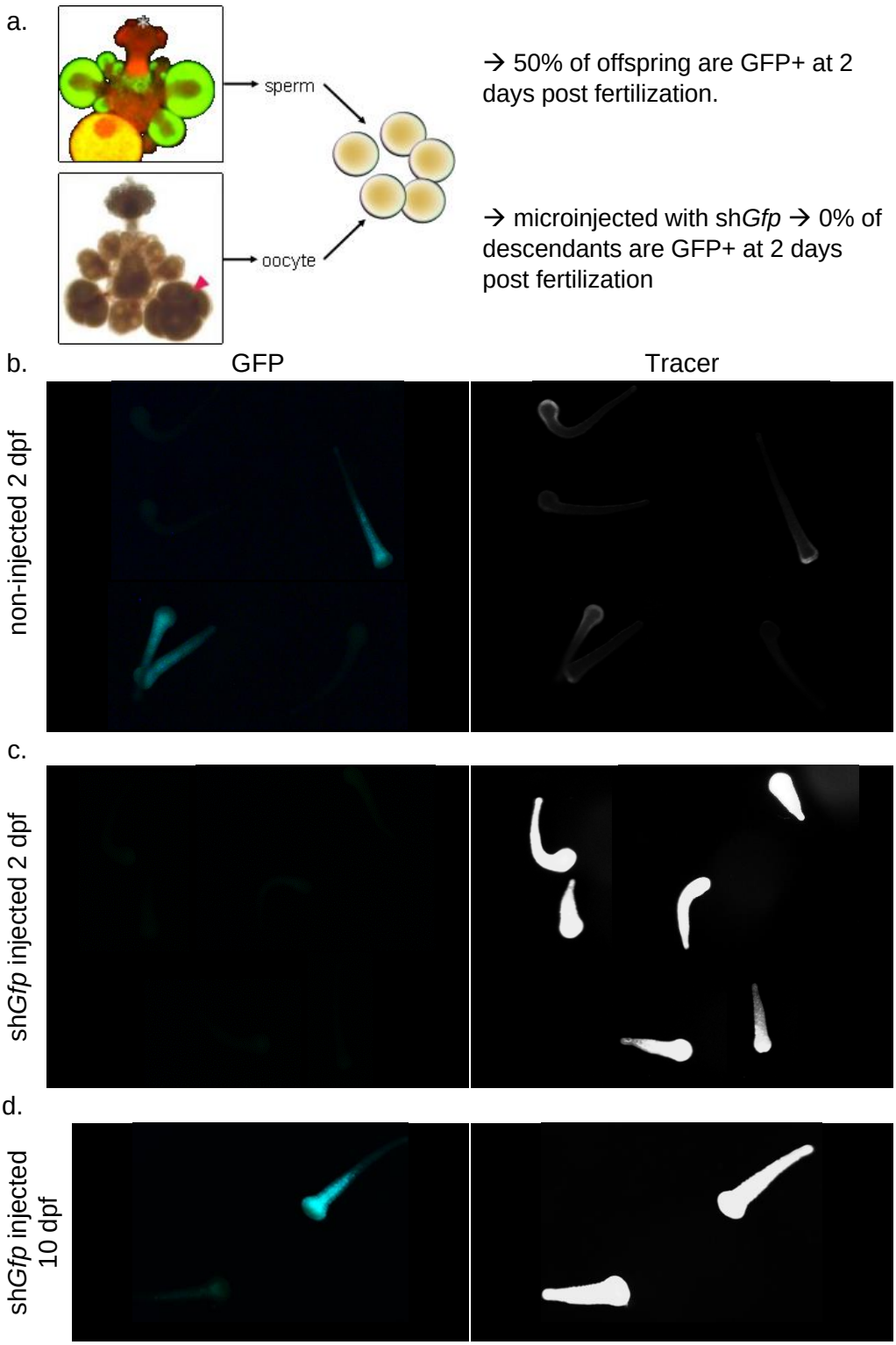


Figure 4.5. Confirmation of shRNA efficacies

a. Experimental set up. Crossing a *Piwi1*::*Gfp* male with a wild type female generated a population of embryos where half displayed GFP+ phenotypes 2 days post fertilization (2 dpf). If sh*Gfp* effectively knocked down *piwi*::*gfp* then the whole population descending from this cross is expected to be GFP-. b. Non-injected embryos where half of the population is GFP+. c. none of the sh*Gfp* injected embryos was GFP+ 2 days post fertilization. d. Half of the population of the sh*Gfp* injected embryos was GFP+ 7-15 days post fertilization.

4.3.1 Knockdown of Alkbh1 delayed the EU incorporation for two cell divisions.

Next, I designed shRNA targeting putative methyltransferases and oxidoreductases of 6mA, then followed the injection of these shRNA with the EU incorporation protocol (Chapter 3) to see the effect of knockdown of 6mA associated genes on ZGA. The results show that sh*Mettl4*, sh*N6amt1*, and sh*Alkbh4* do not affect ZGA (Figure 4.6), rejecting the hypothesis shown in Figure 4.1. It is important to note that no further experiments to confirm that these shRNA were indeed targeting the respective mRNAs were preformed, thus ineffective design of shRNA might be an issue, and no efforts were done to trace the injected embryos to adulthood to reveal any phenotypic changes later in development.

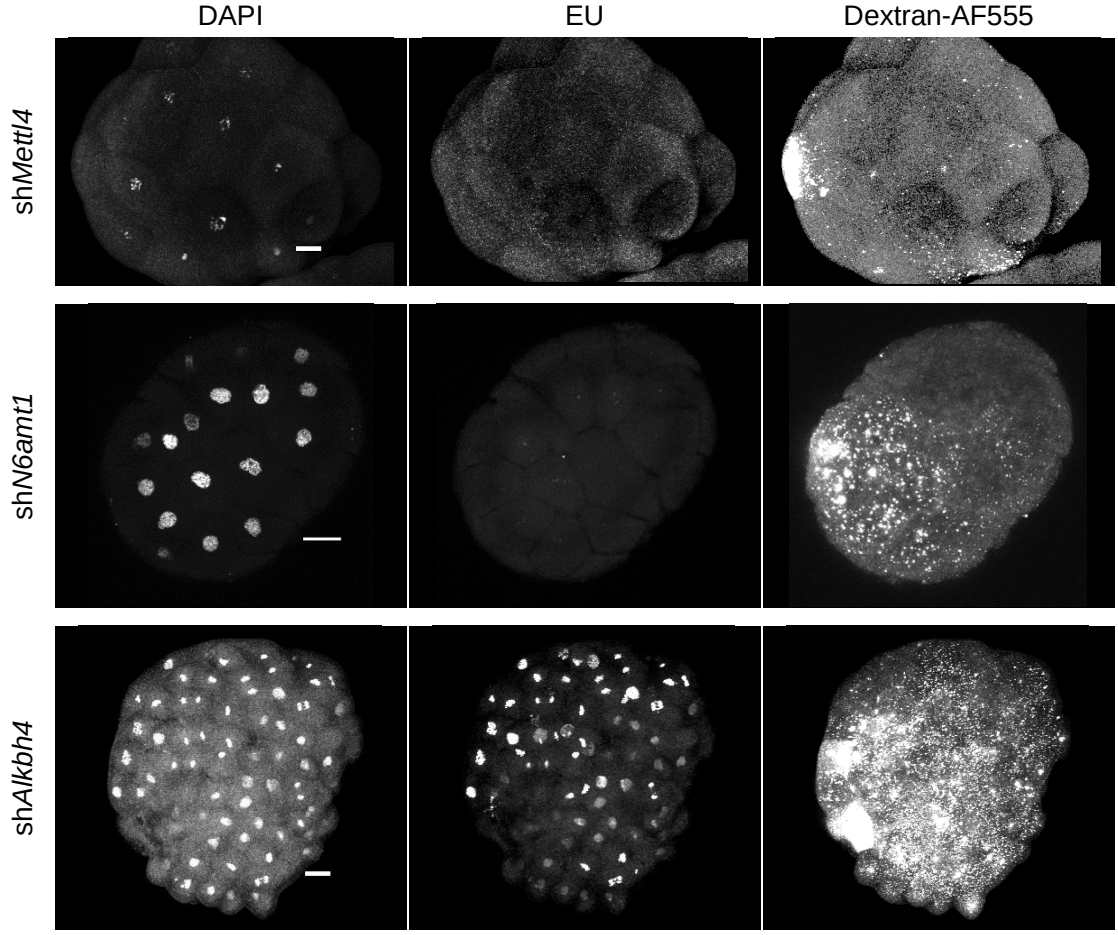


Figure 4.6. EU incorporation is not affected by shRNA knockdown of *Mettl4, N6amt1* and *Alkbh4.*
Scale bar = 20 µm.

The sh*Alkbh1*, however, strongly affected EU incorporation, hence zygotic genome activation, at 64-128 cells (Figure 4.9a), while sh*Gfp* injected embryos displayed wild type EU incorporation at 64-128 cells (Figure 4.9b).

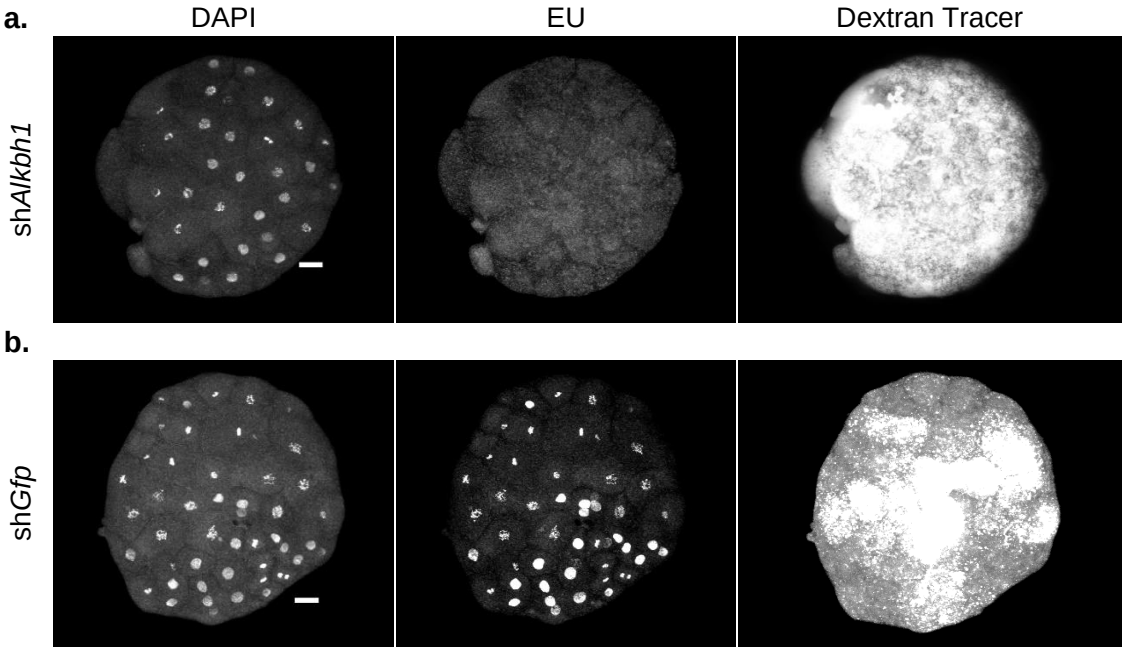


Figure 4.7. *Hydratinia* embryos injected by sh*Alkbh1* do not incorporate EU at 64-128 cell.

a. representative image of the sh*Alkbh1* injected embryo. b. representative image of the sh*Gfp* injected embryo. Scale bars = 20 µm.

To demonstrates the specificity of the sh*Alkbh1*, I co-injected sh*Alkbh1* with mRNA of *Scarlett-Alkbh1*. Embryos injected with the sh*Alkbh1* remained fluorescence free, while sh*Gfp* injection were fluorescent (Figure 4.8). Moreover, synthetic mRNA of *Alkbh1*, which was designed with silent mutation at the sh*Alkbh1* targeted sequence (Figure 4.3b), co-injected together with sh*Alkbh1* could rescue the EU incorporation phenotype (Figure 4.9).

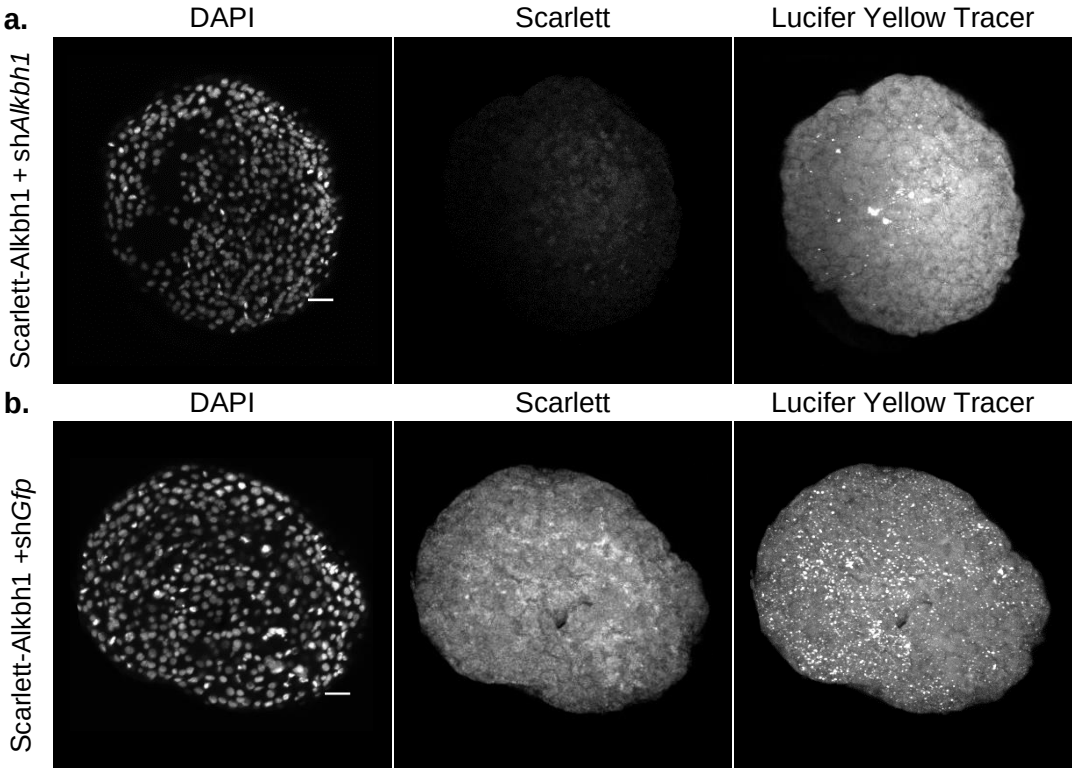


Figure 4.8. Confirmation of sh*Alkbh1* efficacy.

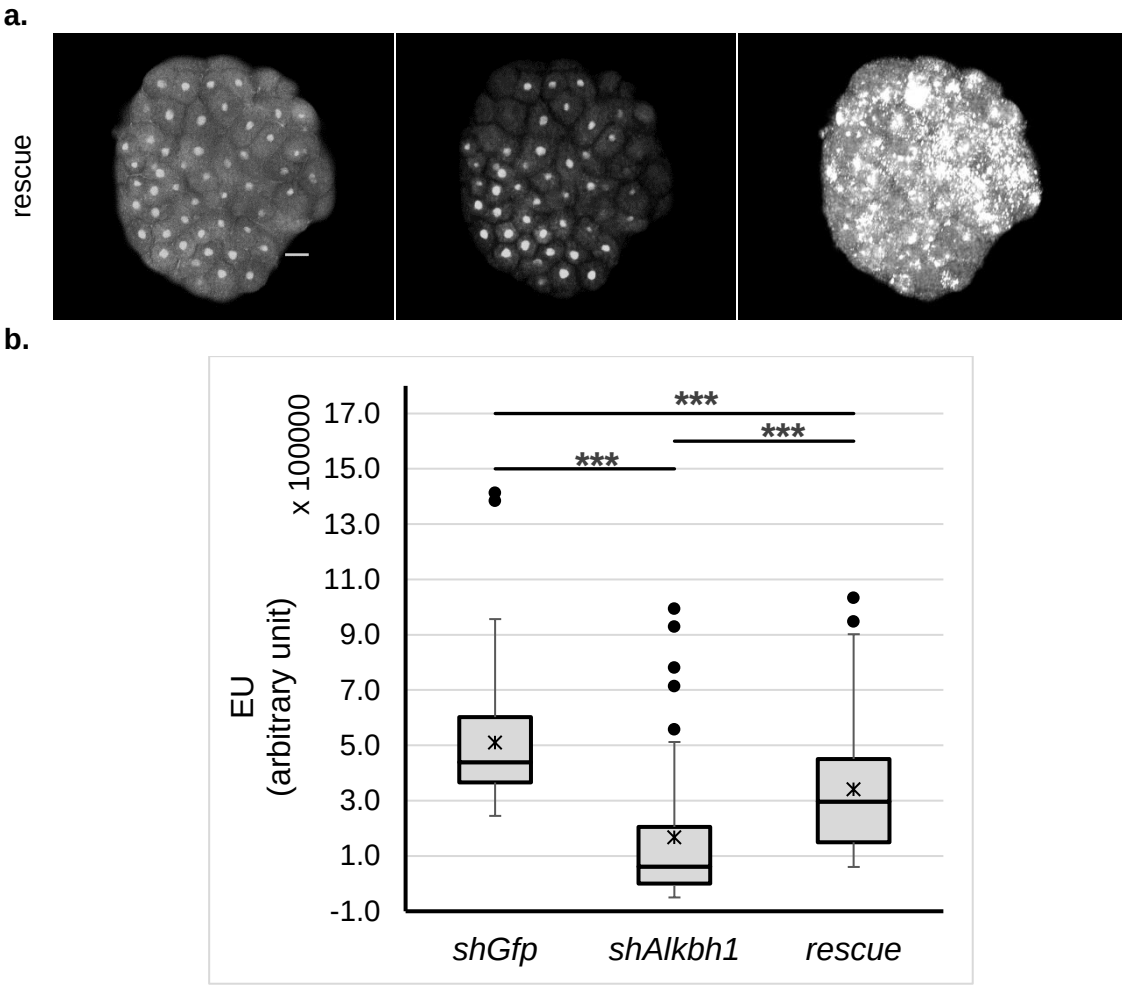


Figure 4.9. *Hydratinia* embryos injected by sh*Alkbh1* do not incorporate EU at 64-128 cell.

a. representative image of embryos from the rescue experiment where silently mutated mRNA encoding *Alkbh1* co-injected with sh*Alkbh1*. b. image quantification by ImageJ of the EU channels from three independent replicates for each experiment. Scale bars = 20 µm. *** indicates *P*-value < 0.01. Data and calculation in Supplementary Document 8.

However, EU incorporation recommence at 256-512 cells (Figure 4.10) in non-rescued embryos (i.e. 1-2 cell cycles later than in control embryos), and no phenotypic changes was observed post-metamorphosis in *Alkbh1* knockdown embryos (Figure 4.11).

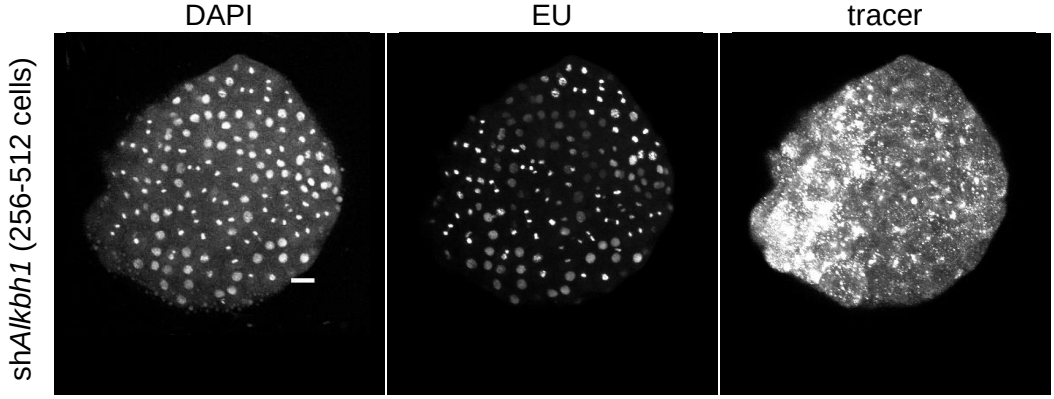


Figure 4.10. Recommencement of EU incorporation at 256-512 cells embryo after *Alkbh1* knockdown.

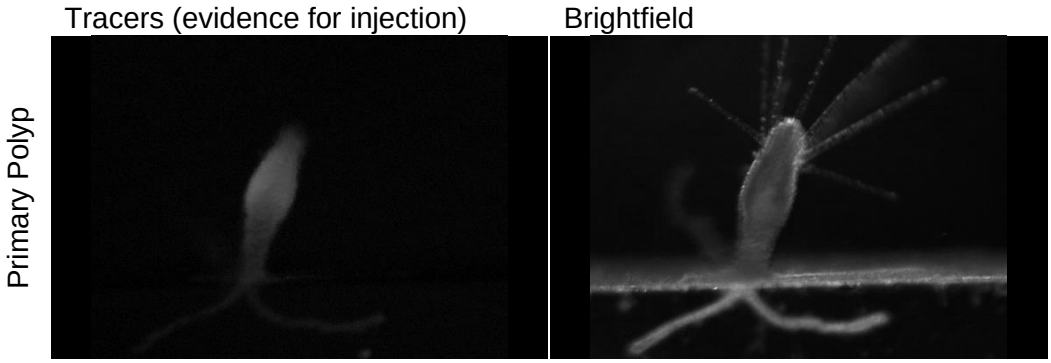


Figure 4.11. Wild type phenotype of sh*Alkbh1* knockdown animal after metamorphosis.

4.4 *Alkbh1* knocked-down embryos maintain high level of 6mA, thus delaying ZGA.

Anti-6mA immunofluorescence images taken from sh*Alkbh1* injected embryos (Figure 4.12) displayed a relatively high level of 6mA at 64-128 cell embryos compare to the sh*Gfp* injected embryos (Figure 4.12). Furthermore, two nucleotide silent mutation of *Alkbh1* mRNA avoided the sh*Alkbh1* knockdown effect, rescuing the level of 6mA to background level (Figure 4.12), consistent with timely commencement of ZGA (Figure 4.9). Collectively, these results suggest that *Alkbh1* acts as the oxidoreductase that initiates demethylation of 6mA between the 16 and 128 cells stages of *Hydractinia* embryos. Together with the EU incorporation data, they also indicate that failure to remove 6mA from the genome of 64-128 cells delays zygotic genome activation for at least two cell cycles. In summary, I have been able to show that *Alkbh1* plays a role in removing 6mA from the genome of *Hydractinia* between 16-128 cells stage, thus facilitating the timely activation of the zygotic genome. The non-palindromic context of 6mA distribution in the *Hydractinia* embryonic genome (see section 4.6.2, page 94 below) probably resulted in dilution of 6mA levels two cell cycles later, allowing ZGA to commence in *Alkbh1* knockdown embryos.

A relationship between 6mA and zygotic genome activation during early embryogenesis has not been suggested before. However, in *Drosophila melanogaster*, zygotic genome activation is pioneered by transcription factor Zelda. A Fox family protein, Jumu, binds and regulates 6mA methylated *Zelda* at the embryonic stage where zygotic genome activation commences (He et al. 2019). Zygotic genome activation pioneered by Zelda is, however, an evolutionarily derived and specific characteristic of arthropods (Liang et al. 2008; Pires et al. 2016). Knockdown of single or a combination of *Alkbh* genes in *Drosophila* does not induce any phenotypes in RNA/DNA methylation (Lence et al. 2017); instead, *Drosophila*'s Alkbh1 is required to repair UV-damaged DNA in the central nervous system (Wakisaka et al. 2019). Moreover, *Drosophila* 6mA demethylation is initiated by a Tet protein rather than by an Alkbh (Zhang et al. 2015; He et al. 2019).

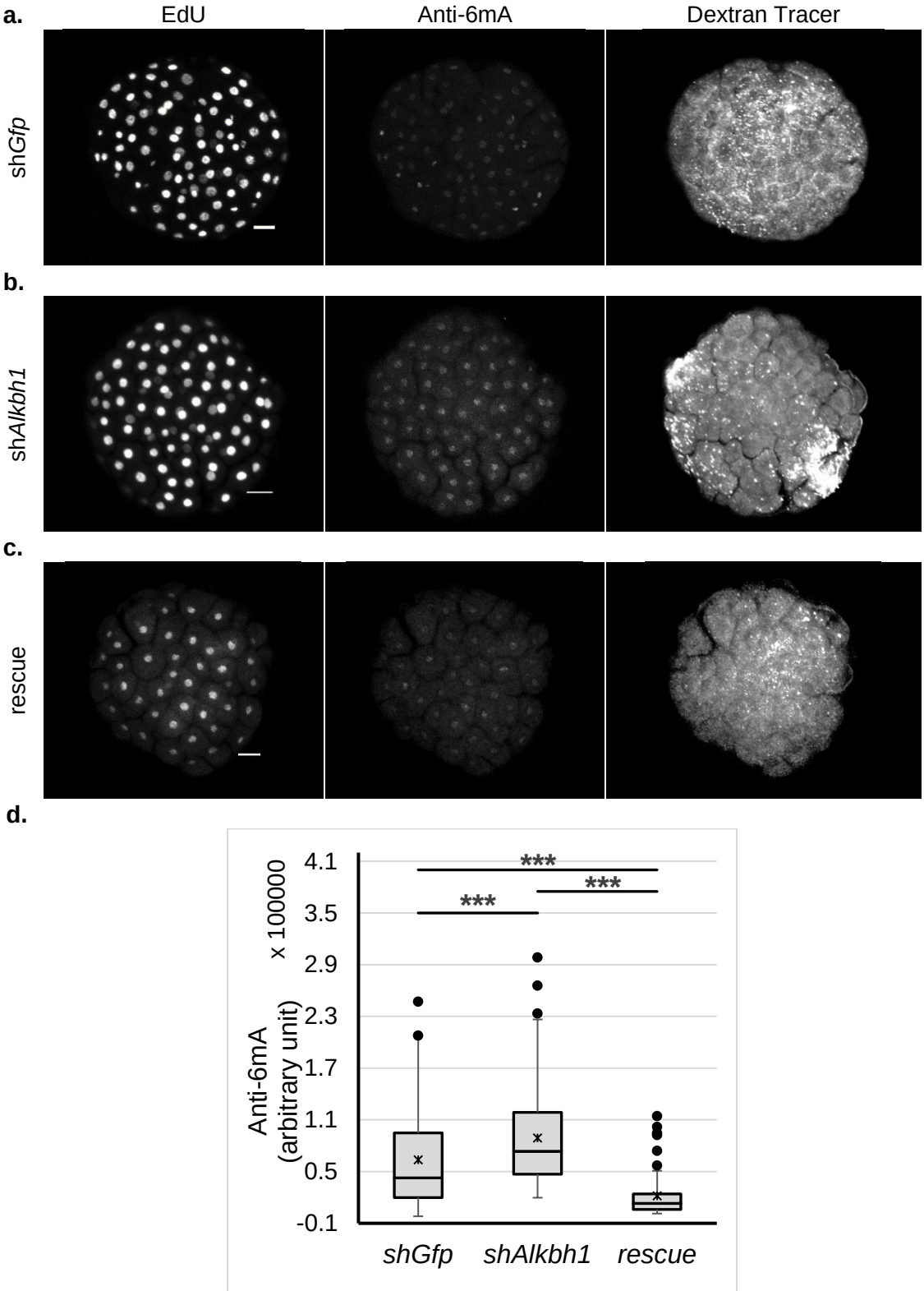


Figure 4.12. *Alkbh1* knock down maintain high level of 6mA at 64-128 cell embryo. a. representative image of the sh*Alkbh1* injected embryo. b. representative image of the sh*Gfp* injected embryo. c. representative image of embryos from the rescue experiment where silently mutated mRNA encoding *Alkbh1* co-injected with sh*Alkbh1*. d. image quantification by ImageJ of the anti-6mA channels from three independent replicates for each experiment. Scale bars = 20 µm. *** indicates *P*-value < 0.01. Data and calculation in Supplementary Document 8.

As established in Chapter 3, 6mA and ZGA cooccurs in zebrafish similar to *Hydractinia* and *Drosophila*. Attempts to investigate ALKBHs in zebrafish, however, were limited to Alkbh4 and Alkbh5, which were found to play a role in actomyosin ring formation (Sun et al. 2017) and RNA demethylation, respectively (Chen et al. 2014). Finally, ZGA in zebrafish, frogs, and mammals have different regulatory mechanisms from flies (Ribeiro et al. 2017; Schulz and Harrison 2019) and probably from *Hydractinia*. Therefore, a link between *Alkbh1,* 6mA, and zygotic genome activation has never been established before; my discoveries on how Alkbh1 removes 6mA to allow the ZGA to commence in a timely manner during early embryonic development is the first describes in animals.

4.5 A glaring Question

One question, however, remains open. How is 6mA incorporated into the genome of *Hydractinia* embryos between the 4 and 32 cell stages? Investigation of 6mA in both *Drosophila* and zebrafish did not provide a clear answer to the question on how the 6mA accumulated in early embryogenesis before rapidly decreased. In the present study, *Mettl4* and *N6amt1* knock down failed to display any role in ZGA of *Hydractinia*. Although these shRNA experiments lack rigorous efficacy controls, the absence of localization signals in *Mettl4* and *N6amt1* methyltransferases (Chapter 2) and of their transcript's distribution in early embryos (Chapter 3) discouraged further attempts to clarify this issue. Moreover, *Mettl4* has been reported convincingly multiple times to functions in methylation of snRNA (Chen et al. 2020; Goh et al. 2020; Gu et al. 2020) and no evidence in the literature points to a role of *N6amt1* in 6mA (Ratel et al. 2006; Kweon et al. 2019; Woodcock et al. 2019). Therefore, I conclude that these two genes most likely have no direct role in *Hydractinia* 6mA as well.

4.6 Alternative possibilities of 6mA deposition at early embryos

With the unlikeliness of either *Mettl4* or *N6amt1* playing a role in 6mA deposition at early embryogenesis, it is imperative to find alternative possibilities for how 6mA incorporates into the genome. Most plausible would be other methyltransferases for 6mA, as proposed in ciliates (Beh et al. 2019) and the oomycete *Phytophthora* (Chen et al. 2018). However, I failed to discover orthologs of these two methyltransferases in *Hydractinia*, which is phylogenetically remote from ciliates and oomycetes. Thus, with 6mA methyltransferases being most likely out of the picture, I would like to end this chapter with another hypothesis for 6mA deposition during early embryogenesis.

Upon fertilization, the zygote immediately enters replication requiring dNTP pools, which are presumably maternally provided (Song et al. 2017). However, the maternal dNTP pool only suffices for several cell cycles, thus ribonucleotide reductase (RNR) is activated

to convert NTPs to dNTPs (Song et al. 2017; Djabrayan et al. 2019; Liu et al. 2019). The source of NTPs is yet unknown. One possibility is that NTPs are provided by the degradation of maternal RNA, which has been described to occur in embryogenesis immediately after sperm entry, before the two pronuclei fuse (Paynton et al. 1988; Giraldez et al. 2006; Ma et al. 2013; Ivanova et al. 2017; Zhao et al. 2017). *De novo* synthesis and the salvage pathway are the conventional pathway to build NTPs. However, in early embryos of mouse, *de novo* synthesis and salvage pathways are only significantly functional after zygotic genome activation (Epstein 1970; Alexiou and Leese 1992). Nevertheless, mammalian early cleavages are slow in comparison to *Drosophila*, zebrafish, and *Hydractinia*, while investigation on *de novo* synthesis and salvage pathway in rapid cleaving early embryos are lacking. Moreover, the rate of NTP-dNTP consumption in rapid cleaving embryos demands a faster provision of NTPs than could be supplied by the slow-multiple steps of *de novo* synthesis. Hence, we[20] argue that RNA nucleotide recycling is probably the major source of nucleosides for dNTP synthesis, supporting the rate of replication during the early, rapid embryonic cleavage.

In zebrafish, a third of maternal RNA are m6A-marked and degraded upon fertilization (Ivanova et al. 2017; Zhao et al. 2017). This degradation generates NTP pools that contain m6A. A recent study has shown that mouse Rnr can convert m6A to 6mA (Musheev et al. 2020), providing a mechanism for random 6mA incorporation into the genome.

Therefore, I hypothesize that maternally deposited but embryonically degraded m6A-marked RNAs are converted by Rnr to 6mATP in *Hydractinia* embryos. Methylated deoxynucleotides are randomly incorporated into DNA during replication, resulting in elevated 6mA at the 16-cell stage. This “RNA recycling” hypothesis may also apply to *Drosophila*, zebrafish, and other animals and plants.

4.6.1 Identifying the source of dNTP at 16 cells embryo of *Hydractinia*.

Testing the above RNA recycling hypothesis in *Hydractinia* early embryo would require too much time to be completed in my thesis work. However, the data I have generated, together with recently published work by others, provide a theoretical framework that makes RNA recycling a plausible if not directly proven source for dNTPs. First, I provide evidence that Rnr inhibition by hydroxyurea results in stalling the replication at 8-16 nuclei stage of embryos in *Hydractinia* (Figure 4.13). This indicated that the maternally

[20] We instead of I, as the idea of this hypothesis formulated during poster session discussion (and follow up personal emails correspondences) with Dr. Markus Mueller at 1st Nucleic Acid Symposium, Institute of Molecular Biology, Mainz, Germany in September 2017.

provided dNTP pool suffices for three-four cell cycles; hence, 6mA incorporation at the 16-cell stage can be associated with the commencement of RNR activities around this stage.

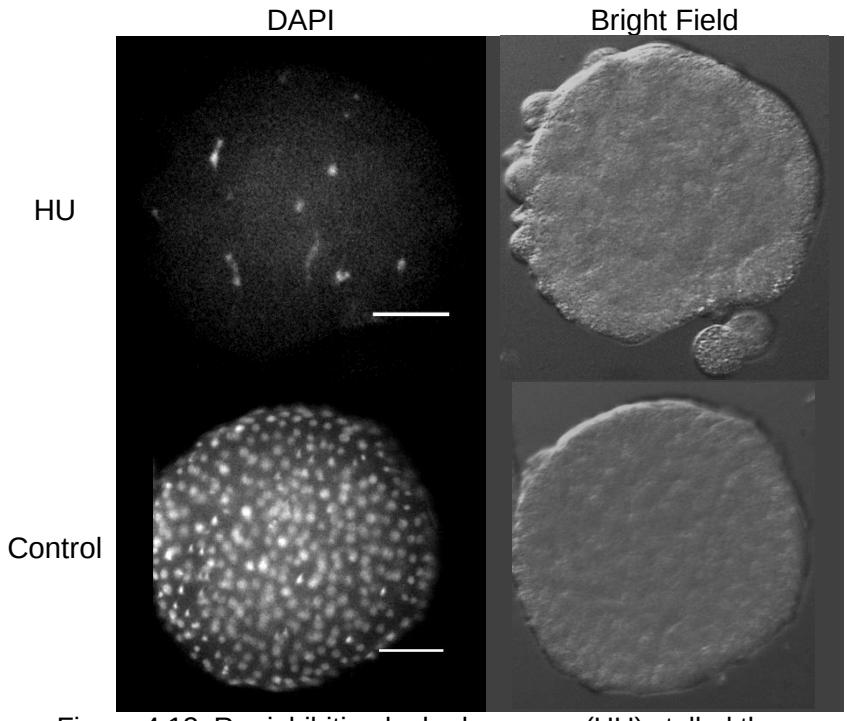


Figure 4.13. Rnr inhibition by hydroxyurea (HU) stalled the embryonic development at 8-16 nuclei.

4.6.2 SMRTseq indicates the non-palindromic motif of 6mA sites in *Hydractinia* genome.

The other, indirect evidence for RNA recycling comes from sequencing data generated in collaboration with NIH. Here, we have sequenced the genome of two *Hydractinia* embryo pools, one was at the 16-32-cell stage and the other at the 64-128 cell stage. We used the single molecule real time sequencing (SMRTseq) system (Pacific Biosciences, PacBio). This new generation of direct long read sequencing records the kinetics of the read at single nucleotide resolution. The kinetics of a modified base (like 6mA) is significantly distinct from a canonical base (like dA). Thus, extracting these records and performing interpulse duration (IPD) analysis plus motif maker analysis allowed our collaborators to map the 6mA sites in the genome of *Hydractinia* embryos of these two stages (and from adult polyps) at single nucleotide resolution. The gff files (documents of *Hydractinia* genome with the 6mA mark information within) were used by me to perform downstream analysis of the results.

To ensure that the pipeline used was performed at high standard, our collaborators performed the analysis on the published *E. coli* genome alongside with *Hydractinia*. An overall read number to obtain genome coverage above 80x is required for genomes with less than 0.1% level of 6mA/dA (Zhu et al. 2018). Thus, 66x coverage for *E. coli* genome

which has ~2.8% 6mA/dA level is more than enough to get ~5 of IPD ratio (Table 4.5, 1st row). The IPD ratio indicates how the IPD of a single base is distinct from the background level. Thus, the higher the IPD ratio the more confident one can assign the modification mark to one base (Table 4.5, 7th column). The coverage of reads for adult polyps and 64-128 cells were ~120x, good enough to avoid false discoveries (Zhu et al. 2018). However, the coverage of the 16-32 cell genome was 72.8x (Table 4.5, 9th column). Therefore, I expect that the reads from the 64-128 cell embryos and adult polyp provide more reliable data compared to the data extracted from the 16-32 cell embryos.

However, the IPD ratios were difficult to interpret (Table 4.5, 7th column). The mean of I PD ratio from all bases considered to be 6mA from 64-128 cells and adult polyps reads were 2.5 and 3.0, respectively, while the mean IPD ratio from 16-32 cell reads was 3.4. These results indicate that 6mA discovered by SMRTseq from adult polyps and the 64-128-cell embryos are less reliable compared to the 6mA found in 16-32-cell embryos. This interpretation is also supported by the level of 6mA/dA calculated from the SMRTseq of 16-32 cell genome (Table 4.5, 4th column), which is consistent with the HPLC results (Table 4.5, 5th column). By contrast, the 6mA/dA level of 64-128 cell-embryos and adult polyps were inconsistent between SMRTseq and HPLC data.

Table 4.5. Overall Data Performance of SMRT Sequencing

Data	dA	6mA [motif]	6mA/dA [motif]	6mA [all]	6mA/dA [all]	6mA/dA [HPLC]	IPD	Cover [motif]	Cover [all]
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
E. coli	1376521	38967	2.831	38967	2.831	?	4.94	66.2	66.2
Hydractinia adult polyp	80713888	2122	0.0026	2278	0.0028	0.018	3.02	81.1	120.5
Hydractinia 16-32 cell	80713888	34633	0.0429	51495	0.0638	0.061	3.38	66.0	72.8
Hydractinia 64-128 cell	80713888	61913	0.0767	112939	0.1399	0.021	2.52	87.9	116.5

(1) Total count of dA in the reference genome. (2) Amount of 6mA within an assigned motif. (3) Percentage of 6mA/dA where the 6mA within an assigned motif. (4) Total of all 6mA identified by SMRTseq. (5) Percentage of 6mA/dA from SMRTseq. (6) Percentage of 6mA/dA measured by HPLC (as in Chapter 2&3) for comparison. (7) Mean of IPD ratio of 6mA within an assigned motif. (8) Mean coverage of the base which marked with 6mA within an assigned motif. (9) Mean coverage of the overall reads of the data.

The motif maker analysis indicated a strong correlation between the palindromic motif of GATC with 6mA methylation in *E. coli*. This is consistent with the decades long established knowledge of 6mA marked GATC motif in *E. coli* (Table 4.6), thus indicating the reliability of SMRTseq for detecting 6mA from a genome with high level of 6mA/dA. Regardless the issues of the IPD ratio, motif maker analysis revealed that motif where 6mA were found in *Hydractinia* embryonic genomes are the non-palindromic GAVBB (Table 4.6). These 6mA only accounts for a third (~0.02% 6mA/dA) of the total 6mA

(~0.06% 6mA/dA) found in 16-32 cell genome. Thus, more than half of the 6mA found in the genome of 16-32 cell embryos of *Hydractinia* are without motif.

Furthermore, out of all GAVBB motif found in the *Hydractinia* genome, only 0.27% of them were methylated, a stark contrast to the 99% methylation of GATC motif in *E. coli*. These results from SMRTseq, IPD and motif maker analysis provide a strong indication that the 6mA incorporation into the genome is sequence independent, thus randomly deposited, which is consistent with the RNA recycle hypothesis.

The preservation of a methyl mark following replication is possible for 5mC due to the replication dependent DNMT1, which works on the hemi methylated palindromic CG motif. Therefore, the random, sequence independent incorporation of 6mA into the genome is also consistent with the lack of functional activities of either *Mettl4* or *N6amt1*. Moreover, it also explains the return of 6mA level to the background level after two cell divisions despites the *Alkbh1* downregulation with specific shRNA injection. Without any replication dependent methyltransferases in play, any methyl mark will be diluted exponentially after every DNA replication in each cell divisions. Therefore, two cell divisions are enough to dilute 6mA level as low as an eight (1/8x) of the original level, allowing ZGA to commence in *Alkbh1* knockdown embryos.

Table 4.6. Motif Maker Analysis Results

Data	Motif	Palind-Partner	nMotif	nMeth Motif	Fract. (%)	mIPD	mCover	nMethMotif /dA (%)
E. coli	GATC	GATC	38150	38369	99.63	5.45	66.84	2.8310
	GCACNNNNNNGTT	AACNNNNNNGTGC	600	598	99.67	4.94	67.49	
Hydractinia adult polyp	ANCGRNCG	–	45496	2122	4.66	3.02	81.13	0.0026
Hydractinia 16-32 cell	GAVBB	–	7202095	19461	0.27	3.29	74.34	0.0241
	ANYGVNBR	–	854873	6825	0.80	3.41	60.52	0.0084
Hydractinia 64-128 cell	GAVBB	–	7227383	33639	0.47	2.52	90.04	0.0417
	ANYGVNYR	–	1847269	16984	0.92	2.60	84.63	0.0210

Motif: Motif where 6mA found, the methylated A is underlined. Palind-Partner: The partner motif from the reverse-complementary strand. nMotif: the number of instances such motif found in the reference genome. nMethMotif: the number of occurrences where such motif is modified. Fract: The fraction of the motif to be modified in percentage. mIPD: Mean IPD ratio of instances of this motif when the motif is modified. mCover: Mean coverages of instances of this motif when the motif is modified. nMethMotif/dA: nMethMotif per total dA in the reference genome in percentage. V: A/C/G, B: C/G/T, N: A/C/G/T, Y:C/T, R:A/G.

4.6.3 Maternal m6A-marked RNAs are degraded after the 2 cell-stage

Apart from the DNA, I also extracted RNA from various embryonic stages of *Hydractinia*. These RNAs were used for RNAseq experiment and for HPLC-QQQ analysis. Although lacking the appropriate external or internal standard, the mass transition and retention time of m6A and A are unique, thus reliable to identify m6A and A from a very pure RNA

samples of very high quality assessed by multiple assessment methods (section 3.4.4.2). A simple calculation using arbitrary unit can indicate m6A/A level (Figure 4.14).

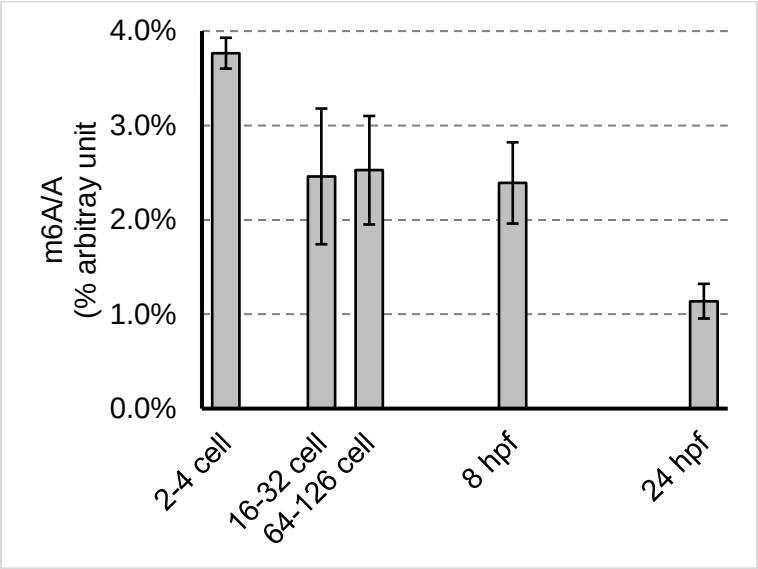


Figure 4.14. Levels of m6A/A indicative of RNA degradation rate during embryonic development of *Hydractinia*.

The levels of m6A/A are arbitrary. Therefore, although it could not calculate the actual level of m6A, the results are, nevertheless, good enough to indicate that m6A marked RNA are degraded at faster rates between 2-4 cell stage to 16-32 cell stage of embryos compare to the total RNA. Furthermore, the m6A marked RNA are degraded at the same rate as the degradation of total RNA between 16-32 cell, 64-128 cell and 8 hpf embryos. This is consistent with the RNA recycling hypothesis, which required evidence that m6A marked RNAs are degraded prior to the Rnr activities at 8-16 cell stage.

4.7 Contamination of 6mA in the embryonic genome of *Hydractinia* are from degraded m6A marked RNA and their clearance by ALKBH1 is necessary for ZGA commencement.

Although I have no direct evidence to support the RNA recycle hypothesis, four indirect pieces of evidences support it: first, RNR inhibition stalls replication at 8-16 nuclei; second, random motif of 6mA distribution in the genome; third, m6A marked RNA degradation rate; four, the lack of any functional methyltransferases for 6mA. Maternal m6A marked RNA degradation immediately after fertilization has been described in zebrafish (Ivanova et al. 2017; Zhao et al. 2017). Thus, the possibility of contribution of the m6A marked RNA degradation to the high level of 6mA/dA level at 128 cells embryo of zebrafish is possible. The replication stalled due to Rnr inhibition by hydroxyurea has also been described in *Drosophila* (Song et al. 2017; Djabrayan et al. 2019; Liu et al. 2019). Interestingly, replication stalls at the stage close to the one where 6mA/dA level at the highest (Zhang et al. 2015). Finally, further support of this RNA recycle hypothesis

comes from the very recent discoveries of m6A-6mA conversion that leads to 6mA incorporation into the genome of mammalian cell cultures (Musheev et al. 2020). Therefore, I believe the model I propose for *Hydractinia* is plausible.

Taken together, I conclude that 6mA is a genomic contamination that prevents transcription but does not function as an epigenetic mark. My third hypothesis (see section 1.7, page 9), however, that miss-expressing a demethylation initiator will alter the 6mA level and affecting transcription is accepted. My functional data show that Alkbh1 is a demethylation agent, necessary for the clearance of 6mA from the genome, enabling timely ZGA.

4.8 References

Alexiou, M., and H.J. Leese. 1992. 'Purine utilisation, de novo synthesis and degradation in mouse preimplantation embryos', *Development*, 114: 185-192.

Ballard, W. 1942. 'The mechanism for synchronous spawning in *Hydractinia* and *Pennaria*. ', *The Biological Bulletin*, 82: 329-339.

Beh, Leslie Y, Galia T Debelouchina, Derek M Clay, Robert E Thompson, Kelsi A Lindblad, Elizabeth R Hutton, . . . Laura F Landweber. 2019. 'Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization', *Cell*, 177: 1781-1796. e1725.

Chen, Han, Haidong Shu, Liyuan Wang, Fan Zhang, Xi Li, Sylvans Ochieng Ochola, . . . Tingting Gu. 2018. '*Phytophthora* methylomes are modulated by 6mA methyltransferases and associated with adaptive genome regions', *Genome biology*, 19: 1-16.

Chen, Hao, Lei Gu, Esteban A. Orellana, Yuanyuan Wang, Jiaojiao Guo, Qi Liu, . . . Yang Shi. 2020. 'METTL4 is an snRNA m6Am methyltransferase that regulates RNA splicing', *Cell Research*.

Chen, Weizhong, Liang Zhang, Guanqun Zheng, Ye Fu, Quanjiang Ji, Fange Liu, . . . Chuan He. 2014. 'Crystal structure of the RNA demethylase ALKBH5 from zebrafish', *FEBS letters*, 588: 892-898.

Djabrayan, Nareg J. V., Celia M. Smits, Matej Krajnc, Tomer Stern, Shigehiro Yamada, William C. Lemon, . . . Stanislav Y. Shvartsman. 2019. 'Metabolic Regulation of Developmental Cell Cycles and Zygotic Transcription', *Current Biology*, 29: 1193-1198.e1195.

Epstein, Charles J. 1970. 'Phosphoribosyltransferase activity during early mammalian development', *Journal of Biological Chemistry*, 245: 3289-3294.

Frank, Uri, Matthew L Nicotra, and Christine E Schnitzler. 2020. 'The colonial cnidarian *Hydractinia*', *EvoDevo*, 11: 1-6.

Gahan, James M., Christine E. Schnitzler, Timothy Q. DuBuc, Liam B. Doonan, Justyna Kanska, Sebastian G. Gornik, . . . Uri Frank. 2017. 'Functional studies on the role of Notch signaling in *Hydractinia* development', *Developmental Biology*, 428: 224-231.

Giraldez, Antonio J, Yuichiro Mishima, Jason Rihel, Russell J Grocock, Stijn Van Dongen, Kunio Inoue, . . . Alexander F Schier. 2006. 'Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs', *Science*, 312: 75-79.

Goh, Yeek Teck, Casslynn W. Q. Koh, Donald Yuhui Sim, Xavier Roca, and W. S. Sho Goh. 2020. 'METTL4 catalyzes m6Am methylation in *U2 snRNA* to regulate pre-mRNA splicing', *bioRxiv*: 2020.2001.2024.917575.

Gu, Lei, Longfei Wang, Hao Chen, Jiaxu Hong, Zhangfei Shen, Abhinav Dhall, . . . Yang Shi. 2020. 'CG14906 (*mettl4*) mediates m6A methylation of U2 snRNA in *Drosophila*', *bioRxiv*: 2020.2001.2009.890764.

Hauenschild, C. 1956. 'Über die Vererbung einer Gewebeverträglichkeits-Eigenschaft bei dem Hydroidpolypen *Hydractinia echinata*', 11: 132.

He, Shunmin, Guoqiang Zhang, Jiajia Wang, Yajie Gao, Ruidi Sun, Zhijie Cao, . . . Dahua Chen. 2019. '6mA-DNA-binding factor Jumu controls maternal-to-zygotic transition upstream of Zelda', *Nature communications*, 10: 2219.

He, Shuonan, Florencia del Viso, Cheng-Yi Chen, Aissam Ikmi, Amanda E. Kroesen, and Matthew C. Gibson. 2018. 'An axial Hox code controls tissue segmentation and body patterning in *Nematostella vectensis*', *Science*, 361: 1377-1380.

Ivanova, Ivayla, Christian Much, Monica Di Giacomo, Chiara Azzi, Marcos Morgan, Pedro N. Moreira, . . . Dónal O'Carroll. 2017. 'The RNA m6A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence', *Molecular cell*, 67: 1059-1067.e1054.

Kanska, Justyna, and Uri Frank. 2013. 'New roles for Nanos in neural cell fate determination revealed by studies in a cnidarian', *Journal of Cell Science*, 126: 3192-3203.

Kweon, Soo-Mi, Yibu Chen, Eugene Moon, Kotryna Kvederaviciutė, Saulius Klimasauskas, and Douglas E. Feldman. 2019. 'An Adversarial DNA N6-Methyladenine-Sensor Network Preserves Polycomb Silencing', *Molecular cell*, 74: 1138-1147.

Lence, Tina, Matthias Soller, and Jean-Yves Roignant. 2017. 'A fly view on the roles and mechanisms of the m6A mRNA modification and its players', *RNA Biology*, 14: 1232-1240.

Liang, Hsiao-Lan, Chung-Yi Nien, Hsiao-Yun Liu, Mark M Metzstein, Nikolai Kirov, and Christine Rushlow. 2008. 'The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*', *Nature*, 456: 400-403.

Liu, Boyang, Franziska Winkler, Marco Herde, Claus-Peter Witte, and Jörg Großhans. 2019. 'A Link between Deoxyribonucleotide Metabolites and Embryonic Cell-Cycle Control', *Current Biology*, 29: 1187-1192.e1183.

Ma, Jun, Matyas Flemr, Hynek Strnad, Petr Svoboda, and Richard M Schultz. 2013. 'Maternally recruited DCP1A and DCP2 contribute to messenger RNA degradation during oocyte maturation and genome activation in mouse', *Biology of reproduction*, 88: 11, 11-12.

McMahon, Emma. 2018. 'Characterization of PIWI+ stem cells in *Hydractinia*', National University of Ireland Galway.

Millane, R. Cathriona, Justyna Kanska, David J. Duffy, Cathal Seoighe, Stephen Cunningham, Günter Plickert, and Uri Frank. 2011. 'Induced stem cell neoplasia in a cnidarian by ectopic expression of a POU domain transcription factor', *Development*, 138: 2429-2439.

Moran, Yehu, Maayan Agron, Daniela Praher, and Ulrich Technau. 2017. 'The evolutionary origin of plant and animal microRNAs', *Nature Ecology & Evolution*, 1: 0027.

Moran, Yehu, David Fredman, Daniela Praher, Xin Z Li, Liang Meng Wee, Fabian Rentzsch, . . . Hervé Seitz. 2014. 'Cnidarian microRNAs frequently regulate targets by cleavage', *Genome research*, 24: 651-663.

Müller, Werner. A. 1984. 'Retinoids and pattern formation in a hydroid', *Journal of Embryology and Experimental Morphology*, 81: 253-271.

Musheev, Michael U., Anne Baumgärtner, Laura Krebs, and Christof Niehrs. 2020. 'The origin of genomic N6-methyl-deoxyadenosine in mammalian cells', *Nature Chemical Biology*, 16: 630-634.

Paynton, Barbara V, Rachel Rempel, and Rosemary Bachvarova. 1988. 'Changes in state of adenylation and time course of degradation of maternal mRNAs during oocyte maturation and early embryonic development in the mouse', *Developmental Biology*, 129: 304-314.

Pires, Camilla Valente, Flavia Cristina de Paula Freitas, Alexandre S Cristino, Peter K Dearden, and Zila Luz Paulino Simoes. 2016. 'Transcriptome analysis of honeybee (*Apis mellifera*) haploid and diploid embryos reveals early zygotic transcription during cleavage', *PLoS One*, 11.

Ratel, David, Jean-Luc Ravanat, Marie-Pierre Charles, Nadine Platet, Lionel Breuillaud, Joël Lunardi, . . . Didier Wion. 2006. 'Undetectable levels of N6-methyl adenine in mouse DNA: cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase', *FEBS letters*, 580: 3179-3184.

Ribeiro, Lupis, Vitoria Tobias-Santos, Daniele Santos, Felipe Antunes, Georgia Feltran, Jackson de Souza Menezes, . . . Rodrigo Nunes da Fonseca. 2017. 'Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects', *PLoS genetics*, 13: e1006868.

Sanders, Steven M, Zhiwei Ma, Julia M Hughes, Brooke M Riscoe, Gregory A Gibson, Alan M Watson, . . . Andreas D Baxevanis. 2018. 'CRISPR/Cas9-mediated gene knockin in the hydroid *Hydractinia symbiolongicarpus*', *BMC genomics*, 19: 649.

Schulz, Katharine N., and Melissa M. Harrison. 2019. 'Mechanisms regulating zygotic genome activation', *Nature Reviews Genetics*, 20: 221-234.

Song, Yonghyun, Robert A Marmion, Junyoung O Park, Debopriyo Biswas, Joshua D Rabinowitz, and Stanislav Y Shvartsman. 2017. 'Dynamic control of dNTP synthesis in early embryos', *Developmental Cell*, 42: 301-308. e303.

Sun, Qingrui, Xingfeng Liu, Bo Gong, Di Wu, Anming Meng, and Shunji Jia. 2017. 'Alkbh4 and Atrn Act Maternally to Regulate Zebrafish Epiboly', *International journal of biological sciences*, 13: 1051-1066.

Toth, Steven E. 1965. 'Cultivation of Marine Hydroids', *Bios*, 36: 63-65.

Tripathi, Abhinandan Mani, Arie Fridrich, Magda Lewandowska, and Yehu Moran. 2020. 'Functional characterization of a “plant-like” HYL1 homolog in the cnidarian *Nematostella vectensis* indicates a conserved involvement in microRNA biogenesis', *bioRxiv*: 2020.2005.2031.126003.

Wakisaka, Keiko Tsuji, Yuuka Muraoka, Jo Shimizu, Mizuki Yamaguchi, Ibuki Ueoka, Ikuko Mizuta, . . . Masamitsu Yamaguchi. 2019. '*Drosophila* Alpha-ketoglutarate-dependent dioxygenase AlkB is involved in repair from neuronal disorders induced by ultraviolet damage', *NeuroReport*, 30: 1039-1047.

Woodcock, Clayton B., Dan Yu, Xing Zhang, and Xiaodong Cheng. 2019. 'Human HemK2/KMT9/N6AMT1 is an active protein methyltransferase, but does not act on DNA in vitro, in the presence of Trm112', *Cell Discovery*, 5: 50.

Zhang, Guoqiang, Hua Huang, Di Liu, Ying Cheng, Xiaoling Liu, Wenxin Zhang, . . . Jianzhao Liu. 2015. 'N6-methyladenine DNA modification in *Drosophila*', *Cell*, 161: 893-906.

Zhao, Boxuan Simen, Xiao Wang, Alana V Beadell, Zhike Lu, Hailing Shi, Adam Kuuspalu, . . . Chuan He. 2017. 'm6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition', *Nature*, 542: 475-478.

Zhu, Shijia, John Beaulaurier, Gintaras Deikus, Tao P Wu, Maya Strahl, Ziyang Hao, . . . Chuan He. 2018. 'Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing', *Genome research*, 28: 1067-1078.

5 DNA METHYLATION AND ZYGOTIC GENOME ACTIVATION DURING EARLY EMBRYOGENESIS OF ANIMALS

5.1 The evolution of DNA methylation in animals

In vertebrates, most cytosines in CpG dinucleotides are methylated, distributed at most contexts such as exons, introns, intergenic, and DNA repeats but glaringly devoid at the promoter, transcription start site and enhancer sites. This distribution is not shared with most accessible and studied invertebrates, with sponge as the only exception so far (Zemach et al. 2010; de Mendoza et al. 2019; de Mendoza et al. 2020). Most invertebrates only methylate their CpGs at gene bodies and not anywhere else not even at repeats or transposable elements. Gene bodies methylation is an enigma in several ways. DNA methylation is prone to mutation (Wang et al. 1982; Holliday and Grigg 1993). Thus, the conservation of DNA methylation at gene bodies in evolution seems very costly and indicates an essential function. Moreover, distinct from the major silencing role in genomic imprinting, germline specific genes, and X chromosome inactivation, gene bodies methylation is associated with active transcription. There are several hypotheses discussed. One is that DNA methylation has a role in splicing and transcriptional elongation. Another hypothesis is that gene bodies methylation is associated with H3K36me3, a histone modification that follow RNA pol II during elongation of transcription. From this understanding, it is proposed that DNA methylation at gene bodies are deposited to avoid intragenic cryptic transcriptional start sites (Greenberg and Bourc'his 2019). An investigation on a cnidarian, *Exaiptasia pallida*, reinforces the "control of spurious transcription" hypothesis of gene bodies methylation (Li et al. 2018b). These studies from sponge and *E. pallida*, reminded us the importance of early diverging animals' insight to understand the biology of DNA methylation in animals.

The evolution of the methyltransferases and oxidoreductases that methylates and initiate demethylation, respectively, are also noteworthy. The methyltransferase domains (MTase) among DNMT, METTL and N6AMT indicate methylation but not necessarily of DNA. DNMT2 and METTL3, for instances, methylate RNA (Goll et al. 2006; Liu et al. 2014) rather than DNA, and N6AMT2 methylates translation elongation factor 1A (Hamey et al. 2016), although *in vitro* these methyltransferases are weakly active on DNA. Thus, in vivo evidence is necessary to confirm their methylation role. DNMT1 and 3's roles on 5mC methylation have been confirmed in mammals. The conservation of major domain and catalytic units suggests that DNMT would have similar role in cnidarians and other early diverging animals, but in vivo functional investigation of DNMTs has not yet been replicated in early diverging animals.

In *Hydractinia*, I have shown that 5mC/dC levels are increased from below 2% at 2-4 cells to above 3% at 24 hpf, and remain so through adulthood (Figure 3.5). Mammals contain comparable levels of 5mC (~3%, (Holliday and Grigg 1993)). However, the mRNA expression path of *Dnmt1, Dnmt3,* and *Tet* are puzzling in *Hydractinia*. The transcripts for 5mC methyltransferases are rapidly degraded reached to near zero at 24 hpf, while *Tet* remain around 2000 upper-quartile normalized counts (Figure 3.14 & Figure 3.15), a comparable level to *Wnt3*, *Ca-almodulin Kinase*, *Alkbh4,* and *Alkbh1* (Figure 3.4, Figure 3.9, and Figure 3.15). These expression patterns indicate that 5mC level should be decreased at 24 hpf and more so at adulthood, which contradicts the actual 5mC/dC levels displayed by HPLC-QQQ data (Figure 3.5). The conservation of the RFD domain of *Hydractinia* Dnmt1 (Figure 2.7) should indicate its function as replication-dependent (mediated by interaction with UHRF1 (Li et al. 2018a)) 5mC methylation, which maintains methylated CpG through replication (Garvilles et al. 2015). The conservation of the PWWP and ADD domains in Dnmt3 (Figure 2.8) indicate a functional Dnmt3 that can interact with H3K36me3 and H3K4 then *de novo* methylating dC. Therefore, the low level of *Dnmt1* and *Dnmt3* transcripts at the 24 hpf stage of embryos where 5mC level are highest is a paradox. Dnmt1's half-life is ~3 hours, thus Dnmt1 posttranslational modification (which stabilizes and degrades the enzymes) may play a role to extend the life of of the protein (Estève et al. 2009; Wang et al. 2009; Estève et al. 2011). The high expression level of *Tet* in *Hydractinia* embryos (Figure 3.15), however, indicates a tightly intricate regulation of 5mC methylation.

The N-terminal extension of Dnmt3 in hydrozoans, with an extra PWWP domain (Figure 2.8), may complicate even more the regulation of methylation and demethylation in *Hydractinia*. The loss of DMAP1-binding domain of Dnmt1 indicates the distinct interactome between Dnmt1 with Double-Strand Break (DSB) repair machineries (Negishi et al. 2009; Lee et al. 2010; Mohan et al. 2011; Garvilles et al. 2015). A variant of Dnmt1 (Dnmt1o) that lack of DMAP1-binding domain has been reported in mouse oocytes (Ratnam et al. 2002). The role of this Dnmt1 variants, however, has been disputed due to the complexity and redundancy between Dnmt1o with the canonical Dnmt1, Dnmt3a and Dnmt3b (Mohan et al. 2011; Garvilles et al. 2015). Regardless, Dnmt1o alone is apparently dispensable but essential in the Dmap1-/- background. and play important role in embryonic development of mouse (Mohan et al. 2011). The loss of such domain from sponge, ctenophore and hydrozoan DNMTs, indicates a more ancestral state of this DNMT1 variant (Figure 2.7,(de Mendoza et al. 2019)).

5.2 Artifact, biological bystander, or epigenetic mark: the 6mA case.

The extremely low level of 6mA at adulthood in most animal models rises questions (Figure 1.3 & Figure 2.5). Furthermore, the reports of 6mA among metazoans are conflicting to each other. The research group who discovered it for the first time in 2015 (Greer et al. 2015; Zhang et al. 2015), for instance, rethink their position and discussed the possibility of 6mA as an artifact resulting from the method they used or from the bacterial contamination in the biological samples (O'Brown et al. 2019). From the method perspective, research groups that specialize in HPLC-MS/MS dismissed 6mA in mouse embryonic stem cells (Schiffers et al. 2017). From the bacterial contamination stand point, the efforts to eradicate bacterial contamination never completely removed high 6mA level in zebrafish and *Drosophila* (He et al. 2019; O'Brown et al. 2019). The 6mA/dA level from *Hydractinia* that I reported here using immunofluorescence and HPLC-QQQ with the respective controls are arguably devoid of bacterial contamination (Figure 2.2-5 & Table 2.4). Moreover, the dynamic change of 6mA levels after methyltransferases and demethylases manipulation *in vivo* is consistent with a potential epigenetic function.

I failed to associate *Mettl4* and *Alkbh4* with zygotic genome activation, thus perhaps also to 6mA deposition and removal, respectively (Figure 4.6). METTL4 as methyltransferase and the oxidoreductase ALKBH4, a putative initiator of 6mA demethylation were first reported in *C. elegans* (Greer et al. 2015). However, follow up studies provided no further evidence for a role of 6mA methyltransferase for METTL4 in *C. elegans*, while follow up work on ALKBH4 indicated other functions through interaction with topoisomerase-2 (O'Brown et al. 2019; Wang et al. 2019a). Another independent report of METTL4 in *C. elegans* indicated a strong association with mitochondrial stress response inheritance, but its association with 6mA requires further investigation (Ma et al. 2019). An effort to replicate METTL4 role as 6mA DNA methyltransferase in other animals are failed to gain acceptance in general (Chen et al. 2020; Goh et al. 2020; Gu et al. 2020). However, knock out of *Mettl4* and *Alkbh4* in mouse cause sub lethality in the progeny and increase the level of Asxl1 and Mpnd. These proteins have a conserved 6mA-recognition domain (RAMA) and have a role in deubiquitinylation of H2AK119ub1, a repressive mark deposited by PRC1. Mettl4 overexpression initiates degradation of Asxl1 and Mpnd, thus maintaining Prc1 mediated repressed genes, but such degradation was not observed when bacterial methyltransferase was overexpressed in mouse despite the relatively similar 6mA accumulation observed (Kweon et al. 2019). Therefore, the unfitting expression path of *Mettl4* is not consistent with the association between *Hydractinia* Mettl4 and 6mA deposition (Figure 3.14). This, together with the discrepancies in the literature of METTL4 and ALKBH4 role on 6mA, discouraged me to pursue the exact

effect of sh*Mettl4* and sh*Alkbh4* mediated knock down on 6mA level of *Hydractinia* embryos.

Another couple of methylator-demethylator of 6mA are N6AMT1 and ALKBH1 which were reported first to be functional on 6mA in human cells (Xiao et al. 2018). Two earlier efforts were done separately; Alkbh1 functionally initiated demethylation 6mA in mouse embryonic stem cells (Wu et al. 2016). However, earlier efforts to functionally investigate N6amt1's role as 6mA methyltransferase in mouse have failed (Ratel et al. 2006). Multiple efforts to reproduce N6AMT1's role in 6mA in mouse and human models have failed too (Schiffers et al. 2017; Xie et al. 2018; Woodcock et al. 2019). In *Hydractinia,* N6amt1 has no nuclear localization signal (Table 2.8) and sh*N6amt1* knockdown embryos exhibited no changes in timing of ZGA (Figure 4.6). Thus, the role of N6amt1 in 6mA deposition in *Hydractinia* is, like in other animal models, unlikely.

The role of ALKBH1 as oxidoreductase that initiates 6mA removal also cannot escape scepticism. While the first functional report from human was still in preprint, three independent groups of researchers reported the role of ALKBH1 in mitochondrial tRNA (Haag et al. 2016; Liu et al. 2016a; Kawarada et al. 2017). Subsequently, however, the study that failed to replicate N6AMT1's function in glioblastoma cells, has successfully shown ALKBH1's function in regulating 6mA in glioblastoma cells (Xie et al. 2018). Apparently, ALKBH1 has several other demethylation functions in e.g. H2A (Ougland et al. 2012), 1mA on tRNA (Liu et al. 2016a), and 3-methylcytosine in DNA and RNA (Westbye et al. 2008). However, very recently, multiple groups of researchers provided multiple lines of evidences that strongly emphasize the 6mA DNA demethylation function of ALKBH1 in animals (Xiong et al. 2018; Wu et al. 2019; Tian et al. 2020; Zhang et al. 2020). Thus, the changes of 6mA level in sh*Alkbh1*-KD embryos of *Hydractinia* (Figure 4.12) provide strong evidence of not only the presence of 6mA in early embryos of *Hydractinia* but also the functional association with Alkbh1 as 6mA demethylator.

5.3 RNA recycling hypothesis, DNA methylation, and zygotic genome activation

My results on *Alkbh1* knockdown associated with 6mA in *Hydractinia* are the first confirmation of Alkbh1's role in 6mA demethylation in early diverging animals (Section 4.4). The potential other functions of Alkbh1 require further investigation. Hence, I cannot exclude the possibility of other Alkbh1 functions to convolutedly affecting zygotic genome activation along with the 6mA demethylation function. However, the *Hydractinia* Alkbh1 is not predicted to be targeted to mitochondria (Table 2.8), thus its role on mitochondrial tRNA is less likely. Nonetheless, RNA polymerase II was kinetically paused by 6mA on DNA (Wang et al. 2017a). Moreover, if the association of Asxl1 and Mpnd with 6mA is

true, the removal of 6mA by ALKBH1 would also promote deubiquitination of H2AK119ub1 allowing active transcription. In *Hydractinia*, zygotic genome activation was resumed in *Alkbh1*-KD embryos in 256-512 cell stage embryos (Figure 4.10). The dilution effect due to non-palindromic motif of the 6mA mark in the genome of *Hydractinia* (Table 4.6) is a plausible and simple explanation for recommencement of ZGA after a two cell-division delay (Figure 1.2).

The facts that no methyltransferases can be firmly associated with the increasing level of 6mA are puzzling. In *Hydractinia*, *Drosophila,* and zebrafish genomes, 6mA was accumulated upon fertilization until few cell-cycles before ZGA (Figure 3.2 & Figure 3.13). This indicates that the mechanism for 6mA accumulation in these three evolutionarily distant animals is possibly conserved. Therefore, the RNA recycling hypothesis as the mechanism for 6mA accumulation during early embryogenesis is appealing. This hypothesis is supported in *Hydractinia* by several indirect lines of evidence: (1) Rnr inhibition by hydroxyurea halted nuclear division at stage where 6mA level reaches the highest level (Figure 4.13), which necessitates a flood of NTP to dNTP conversion just before it; (2) more than 60% of the 6mA site discovered in SMRTseq analysis were found without sequence motif (while those found with motif are non-palindromic), thus 6mA are randomly distributed in early embryos of *Hydractinia* (Table 4.6); and (3) m6A in RNA are rapidly degraded from 2-4 cell stage to 16-32 cell stage of *Hydractinia* embryo, way before the accumulation of 6mA in the genome occurred (Figure 4.14). Finally, the RNA recycling hypothesis is consistent with very recent reports that isotopic-labelled m6A can be converted into 6mA and incorporated into DNA in mammalian culture cells by DNA polymerase (Liu et al. 2020; Musheev et al. 2020). One of this reports also attempted to test METTL4 and ALKBH1 function, and found unequivocally that *Mettl4* knockout did not alter 6mA level, while in embryonic stem cell lines from five strains of *Alkbh1* knockout mice, three exhibited increase of 6mA levels (Liu et al. 2020). These data reiterate that 6mA is not an artifact and Alkbh1 contributes to its demethylation, although its rises doubt about 6mA being an epigenetic mark and leverages the arguments that 6mA in the genome of animals is a biological fortuitous bystander.

In single cell eukaryotes, however, 6mA found in palindromic ApT context (Fu et al. 2015; Mondo et al. 2017; Wang et al. 2017b; Chen et al. 2018). Methyltransferases for 6mA found in ciliates and oomycete are evolutionary distinct from animals' 6mA methyltransferases (Chen et al. 2018; Beh et al. 2019; Wang et al. 2019b). Furthermore, 6mA are strongly associated with transcription start site and active genes in these single cell eukaryotes, hence indicates a role of 6mA as epigenetic mark. However, this

probably a specific trait of single cell eukaryotes that are not shared with animals, while no studies, so far, have been done in choanoflagellates, the closest single cell eukaryotes to animals evolutionarily.

While I was writing this very last section of my thesis, Alkbh1 was reported to bind preferably to stress-induced DNA double helix destabilization (SIDD), which maintains the hetero/eu-chromatin boundaries (Li et al. 2020; Zhang et al. 2020). Furthermore, 6mA deposition at SIDD sites repels Satb1-SIDD interaction, hence repressing euchromatin expansion in trophoblast stem cells (Li et al. 2020). Satb1 is a chromatin regulator that bind to SIDD sites allowing for euchromatin expansion to heterochromatin area. Thus, 6mA in mouse trophoblast stem cells maintains a repressed state of the neighbouring heterochromatin by inhibiting the interaction of Satb1 and SIDD. Nevertheless, these studies failed to provide evidence for functional 6mA methyltransferases and to describe how 6mA were deposited, indicating the probable fortuitous incorporation due to high level of m6A degradation in early embryonic development. These new developments plus my discoveries rejuvenate the importance of 6mA and Alkbh1 in affecting the epigenetic landscape of the genome, despite incorporated as a biological by-product.

I searched for Satb1 homologs in early diverging animals but failed. I concluded instead that SATB1 is a vertebrate innovation. SIDD, however, is conserved between plants and animals. It is more likely to happen in AT-rich area in the genome (Bode et al. 2006). Thus, the AT richness of the *Hydractinia* genome might serve a fertile ground for SIDD to occur, thus 6mA misincorporation into these sites have to be regulated. This would serve as a very fascinating future direction to follow up research from this thesis, as no studies ever reported on SIDD and the proteins that interact with it in early diverging animals. Furthermore, it will contribute to the discussion of 6mA's role in the biology of animals.

5.4 References

Beh, Leslie Y, Galia T Debelouchina, Derek M Clay, Robert E Thompson, Kelsi A Lindblad, Elizabeth R Hutton, . . . Laura F Landweber. 2019. 'Identification of a DNA N6-adenine methyltransferase complex and its impact on chromatin organization', *Cell*, 177: 1781-1796. e1725.

Bode, Jürgen, Silke Winkelmann, Sandra Götze, Steven Spiker, Ken Tsutsui, Chengpeng Bi, . . . Craig Benham. 2006. 'Correlations between scaffold/matrix attachment region (S/MAR) binding activity and DNA duplex destabilization energy', *Journal of Molecular Biology*, 358: 597-613.

Chen, Han, Haidong Shu, Liyuan Wang, Fan Zhang, Xi Li, Sylvans Ochieng Ochola, . . . Tingting Gu. 2018. '*Phytophthora* methylomes are modulated by 6mA methyltransferases and associated with adaptive genome regions', *Genome biology*, 19: 1-16.

Chen, Hao, Lei Gu, Esteban A. Orellana, Yuanyuan Wang, Jiaojiao Guo, Qi Liu, . . . Yang Shi. 2020. 'METTL4 is an snRNA m6Am methyltransferase that regulates RNA splicing', *Cell Research*.

de Mendoza, Alex, William L. Hatleberg, Kevin Pang, Sven Leininger, Ozren Bogdanovic, Jahnvi Pflueger, . . . Ryan Lister. 2019. 'Convergent evolution of a vertebrate-like methylome in a marine sponge', *Nature Ecology & Evolution*, 3: 1464-1473.

de Mendoza, Alex, Ryan Lister, and Ozren Bogdanovic. 2020. 'Evolution of DNA methylome diversity in eukaryotes', *Journal of Molecular Biology*, 432: 1687-1705.

Estève, Pierre-Olivier, Yanqi Chang, Mala Samaranayake, Anup K Upadhyay, John R Horton, George R Feehery, . . . Sriharsa Pradhan. 2011. 'A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability', *Nature Structural & Molecular Biology*, 18: 42.

Estève, Pierre-Olivier, Hang Gyeong Chin, Jack Benner, George R Feehery, Mala Samaranayake, Gregory A Horwitz, . . . Sriharsa Pradhan. 2009. 'Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells', *Proceedings of the National Academy of Sciences*, 106: 5076-5081.

Fu, Ye, Guan-Zheng Luo, Kai Chen, Xin Deng, Miao Yu, Dali Han, . . . Louis C Doré. 2015. 'N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*', *Cell*, 161: 879-892.

Garvilles, Ronald Garingalao, Takashi Hasegawa, Hironobu Kimura, Jafar Sharif, Masahiro Muto, Haruhiko Koseki, . . . Shoji Tajima. 2015. 'Dual functions of the RFTS domain of Dnmt1 in replication-coupled DNA methylation and in protection of the genome from aberrant methylation', *PLoS One*, 10: e0137509.

Goh, Yeek Teck, Casslynn W. Q. Koh, Donald Yuhui Sim, Xavier Roca, and W. S. Sho Goh. 2020. 'METTL4 catalyzes m6Am methylation in *U2 snRNA* to regulate pre-mRNA splicing', *bioRxiv*: 2020.2001.2024.917575.

Goll, Mary Grace, Finn Kirpekar, Keith A. Maggert, Jeffrey A. Yoder, Chih-Lin Hsieh, Xiaoyu Zhang, . . . Timothy H. Bestor. 2006. 'Methylation of $tRNA^{Asp}$ by the DNA Methyltransferase Homolog Dnmt2', *Science*, 311: 395-398.

Greenberg, Maxim VC, and Deborah Bourc'his. 2019. 'The diverse roles of DNA methylation in mammalian development and disease', *Nature reviews Molecular cell biology*: 1-18.

Greer, Eric Lieberman, Mario Andres Blanco, Lei Gu, Erdem Sendinc, Jianzhao Liu, David Aristizábal-Corrales, . . . Yang Shi. 2015. 'DNA methylation on N6-adenine in *C. elegans*', *Cell*, 161: 868-878.

Gu, Lei, Longfei Wang, Hao Chen, Jiaxu Hong, Zhangfei Shen, Abhinav Dhall, . . . Yang Shi. 2020. 'CG14906 (*mettl4*) mediates m6A methylation of U2 snRNA in *Drosophila*', *bioRxiv*: 2020.2001.2009.890764.

Haag, Sara, Katherine E Sloan, Namit Ranjan, Ahmed S Warda, Jens Kretschmer, Charlotte Blessing, . . . Peter Rehling. 2016. 'NSUN 3 and ABH 1 modify the wobble position of mt-t RNA Met to expand codon recognition in mitochondrial translation', *The EMBO journal*, 35: 2104-2119.

Hamey, Joshua J, Daniel L Winter, Daniel Yagoub, Christopher M Overall, Gene Hart-Smith, and Marc R Wilkins. 2016. 'Novel N-terminal and lysine methyltransferases that target translation elongation factor 1A in yeast and human', *Molecular & Cellular Proteomics*, 15: 164-176.

He, Shunmin, Guoqiang Zhang, Jiajia Wang, Yajie Gao, Ruidi Sun, Zhijie Cao, . . . Dahua Chen. 2019. '6mA-DNA-binding factor Jumu controls maternal-to-zygotic transition upstream of Zelda', *Nature communications*, 10: 2219.

Holliday, R., and G. W. Grigg. 1993. 'DNA methylation and mutation', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 285: 61-67.

Kawarada, Layla, Takeo Suzuki, Takayuki Ohira, Shoji Hirata, Kenjyo Miyauchi, and Tsutomu Suzuki. 2017. 'ALKBH1 is an RNA dioxygenase responsible for cytoplasmic and mitochondrial tRNA modifications', *Nucleic Acids Research*, 45: 7401-7415.

Kweon, Soo-Mi, Yibu Chen, Eugene Moon, Kotryna Kvederaviciutė, Saulius Klimasauskas, and Douglas E. Feldman. 2019. 'An Adversarial DNA N6-Methyladenine-Sensor Network Preserves Polycomb Silencing', *Molecular cell*, 74: 1138-1147.

Lee, Gun E, Joo Hee Kim, Michael Taylor, and Mark T Muller. 2010. 'DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair', *Journal of Biological Chemistry*, 285: 37630-37640.

Li, Tao, Linsheng Wang, Yongming Du, Si Xie, Xi Yang, Fuming Lian, . . . Chengmin Qian. 2018a. 'Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation', *Nucleic Acids Research*, 46: 3218-3231.

Li, Yong, Yi Jin Liew, Guoxin Cui, Maha J. Cziesielski, Noura Zahran, Craig T. Michell, . . . Manuel Aranda. 2018b. 'DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model Aiptasia', *Science Advances*, 4: eaat2142.

Li, Zheng, Shuai Zhao, Raman V. Nelakanti, Kaixuan Lin, Tao P. Wu, Myles H. Alderman, . . . Andrew Z. Xiao. 2020. 'N6-methyladenine in DNA antagonizes SATB1 in early development', *Nature*.

Liu, Fange, Wesley Clark, Guanzheng Luo, Xiaoyun Wang, Ye Fu, Jiangbo Wei, . . . Guanqun Zheng. 2016a. 'ALKBH1-mediated tRNA demethylation regulates translation', *Cell*, 167: 816-828. e816.

Liu, Jianzhao, Yanan Yue, Dali Han, Xiao Wang, Ye Fu, Liang Zhang, . . . Chuan He. 2014. 'A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation', *Nature Chemical Biology*, 10: 93-95.

Liu, Xiaoling, Weiyi Lai, Yao Li, Shaokun Chen, Baodong Liu, Ning Zhang, . . . Hailin Wang. 2020. 'N6-methyladenine is incorporated into mammalian genome by DNA polymerase', *Cell Research*.

Ma, Chengchuan, Rong Niu, Tianxiao Huang, Li-Wa Shao, Yong Peng, Wanqiu Ding, . . . Ying Liu. 2019. 'N6-methyldeoxyadenine is a transgenerational epigenetic signal for mitochondrial stress adaptation', *Nature Cell Biology*, 21: 319-327.

Mohan, K. Naga, Feng Ding, and J. Richard Chaillet. 2011. 'Distinct Roles of DMAP1 in Mouse Development', *Molecular and Cellular Biology*, 31: 1861-1869.

Mondo, Stephen J, Richard O Dannebaum, Rita C Kuo, Katherine B Louie, Adam J Bewick, Kurt LaButti, . . . Steven R Ahrendt. 2017. 'Widespread adenine N6-methylation of active genes in fungi', *Nature genetics*, 49: 964-968.

Musheev, Michael U., Anne Baumgärtner, Laura Krebs, and Christof Niehrs. 2020. 'The origin of genomic N6-methyl-deoxyadenosine in mammalian cells', *Nature Chemical Biology*, 16: 630-634.

Negishi, Masamitsu, Tetsuhiro Chiba, Atsunori Saraya, Satoru Miyagi, and Atsushi Iwama. 2009. 'Dmap1 plays an essential role in the maintenance of genome integrity through the DNA repair process', *Genes to Cells*, 14: 1347-1357.

O'Brown, Zach K, Konstantinos Boulias, Jie Wang, Simon Yuan Wang, Natasha M O'Brown, Ziyang Hao, . . . Eric Lieberman Greer. 2019. 'Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA', *BMC genomics*, 20: 445.

Ougland, Rune, David Lando, Ida Jonson, John A Dahl, Marivi Nabong Moen, Line M Nordstrand, . . . Tony Kouzarides. 2012. 'ALKBH1 is a histone H2A dioxygenase involved in neural differentiation', *Stem cells*, 30: 2672-2682.

Ratel, David, Jean-Luc Ravanat, Marie-Pierre Charles, Nadine Platet, Lionel Breuillaud, Joël Lunardi, . . . Didier Wion. 2006. 'Undetectable levels of N6-methyl adenine in mouse DNA: cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase', *FEBS letters*, 580: 3179-3184.

Ratnam, Sarayu, Carmen Mertineit, Feng Ding, Carina Y. Howell, Hugh J. Clarke, Timothy H. Bestor, . . . Jacquetta M. Trasler. 2002. 'Dynamics of Dnmt1 Methyltransferase Expression and Intracellular Localization during Oogenesis and Preimplantation Development', *Developmental Biology*, 245: 304-314.

Schiffers, Sarah, Charlotte Ebert, René Rahimoff, Olesea Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, . . . Thomas Carell. 2017. 'Quantitative LC-MS provides no evidence for m6dA or m4dC in the genome of mouse embryonic stem cells and tissues', *Angewandte Chemie International Edition*, 56: 11268-11271.

Tian, Li-Fei, Yan-Ping Liu, Lianqi Chen, Qun Tang, Wei Wu, Wei Sun, . . . Xiao-Xue Yan. 2020. 'Structural basis of nucleic acid recognition and 6mA demethylation by human ALKBH1', *Cell Research*, 30: 272-275.

Wang, Jing, Sarah Hevi, Julia K Kurash, Hong Lei, Frédérique Gay, Jeffrey Bajko, . . . Guoliang Xu. 2009. 'The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation', *Nature genetics*, 41: 125-129.

Wang, Richard Y-H, Kenneth C Kuo, Charles W Gehrke, Lan-Hsiang Huang, and Melanie Ehrlich. 1982. 'Heat-and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA', *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 697: 371-377.

Wang, Simon Yuan, Hui Mao, Hiroki Shibuya, Satoru Uzawa, Zach Klapholz O'Brown, Sage Wesenberg, . . . Eric Lieberman Greer. 2019a. 'The demethylase NMAD-1 regulates DNA replication and repair in the *Caenorhabditis elegans* germline', *PLoS genetics*, 15: e1008252.

Wang, Wei, Liang Xu, Lulu Hu, Jenny Chong, Chuan He, and Dong Wang. 2017a. 'Epigenetic DNA modification N 6-methyladenine causes site-specific RNA polymerase II transcriptional pausing', *Journal of the American Chemical Society*, 139: 14436-14442.

Wang, Yuanyuan, Xiao Chen, Yalan Sheng, Yifan Liu, and Shan Gao. 2017b. 'N6-adenine DNA methylation is associated with the linker DNA of H2A. Z-containing well-positioned nucleosomes in Pol II-transcribed genes in *Tetrahymena*', *Nucleic Acids Research*, 45: 11594-11606.

Wang, Yuanyuan, Yalan Sheng, Yongqiang Liu, Wenxin Zhang, Ting Cheng, Lili Duan, . . . Shan Gao. 2019b. 'A distinct class of eukaryotic MT-A70 methyltransferases maintain symmetric DNA N6-adenine methylation at the ApT dinucleotides as an epigenetic mark associated with transcription', *Nucleic Acids Research*, 47: 11771-11789.

Westbye, Marianne Pedersen, Emadoldin Feyzi, Per Arne Aas, Cathrine Broberg Vågbø, Vivi Anita Talstad, Bodil Kavli, . . . Nina-Beate Liabakk. 2008. 'Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA', *Journal of Biological Chemistry*, 283: 25046-25056.

Woodcock, Clayton B., Dan Yu, Xing Zhang, and Xiaodong Cheng. 2019. 'Human HemK2/KMT9/N6AMT1 is an active protein methyltransferase, but does not act on DNA in vitro, in the presence of Trm112', *Cell Discovery*, 5: 50.

Wu, Lianpin, Yuqing Pei, Yinhuan Zhu, Minghua Jiang, Cheng Wang, Wei Cui, and Donghong Zhang. 2019. 'Association of N6-methyladenine DNA with plaque progression in atherosclerosis via myocardial infarction-associated transcripts', *Cell Death & Disease*, 10: 909.

Wu, Tao P, Tao Wang, Matthew G Seetin, Yongquan Lai, Shijia Zhu, Kaixuan Lin, . . . Mei Zhong. 2016. 'DNA methylation on N 6-adenine in mammalian embryonic stem cells', *Nature*, 532: 329-333.

Xiao, Chuan-Le, Song Zhu, Minghui He, De Chen, Qian Zhang, Ying Chen, . . . Feng Luo. 2018. 'N6-methyladenine DNA modification in the human genome', *Molecular cell*, 71: 306-318.

Xie, Qi, Tao P. Wu, Ryan C. Gimple, Zheng Li, Briana C. Prager, Qiulian Wu, . . . Jeremy N. Rich. 2018. 'N6-methyladenine DNA Modification in Glioblastoma', *Cell*, 175: 1228-1243.

Xiong, Jun, Tian-Tian Ye, Cheng-Jie Ma, Qing-Yun Cheng, Bi-Feng Yuan, and Yu-Qi Feng. 2018. 'N6-Hydroxymethyladenine: a hydroxylation derivative of N6-methyladenine in genomic DNA of mammals', *Nucleic Acids Research*, 47: 1268-1277.

Zemach, Assaf, Ivy E. McDaniel, Pedro Silva, and Daniel Zilberman. 2010. 'Genome-wide evolutionary analysis of eukaryotic DNA methylation', *Science*, 328: 916.

Zhang, Guoqiang, Hua Huang, Di Liu, Ying Cheng, Xiaoling Liu, Wenxin Zhang, . . . Jianzhao Liu. 2015. 'N6-methyladenine DNA modification in *Drosophila*', *Cell*, 161: 893-906.

Zhang, Min, Shumin Yang, Raman Nelakanti, Wentao Zhao, Gaochao Liu, Zheng Li, . . . Haitao Li. 2020. 'Mammalian ALKBH1 serves as an N6-mA demethylase of unpairing DNA', *Cell Research*.

6 Appendices

6.1 Raw data of HPLC calculation

See Supplementary Document 1 for more details from other HPLC-QQQ run.

Suppl. Table 1. External Standard Curve Raw Data from Run 14

Mass Transition	252.10-136.10	266.10-150.10	228.10-112.10	242.10-126.10
Retention Time	0.91	1.60	0.70	0.76
mol	**dA**	**6m-dA**	**dC**	**5m-dC**
0.307500	307380.50	767930.50	315531.00	230237.00
0.076875	92228.50	282430.00	38701.00	72016.50
0.019219	25701.50	85164.00	7180.50	19295.00
0.004805	6823.00	21118.50	2034.00	3687.00
0.001201	1862.00	6614.50	601.50	1192.50
0.000300	839.00	1474.00	352.00	328.00
R	0.9963	0.9799	0.9789	0.9946

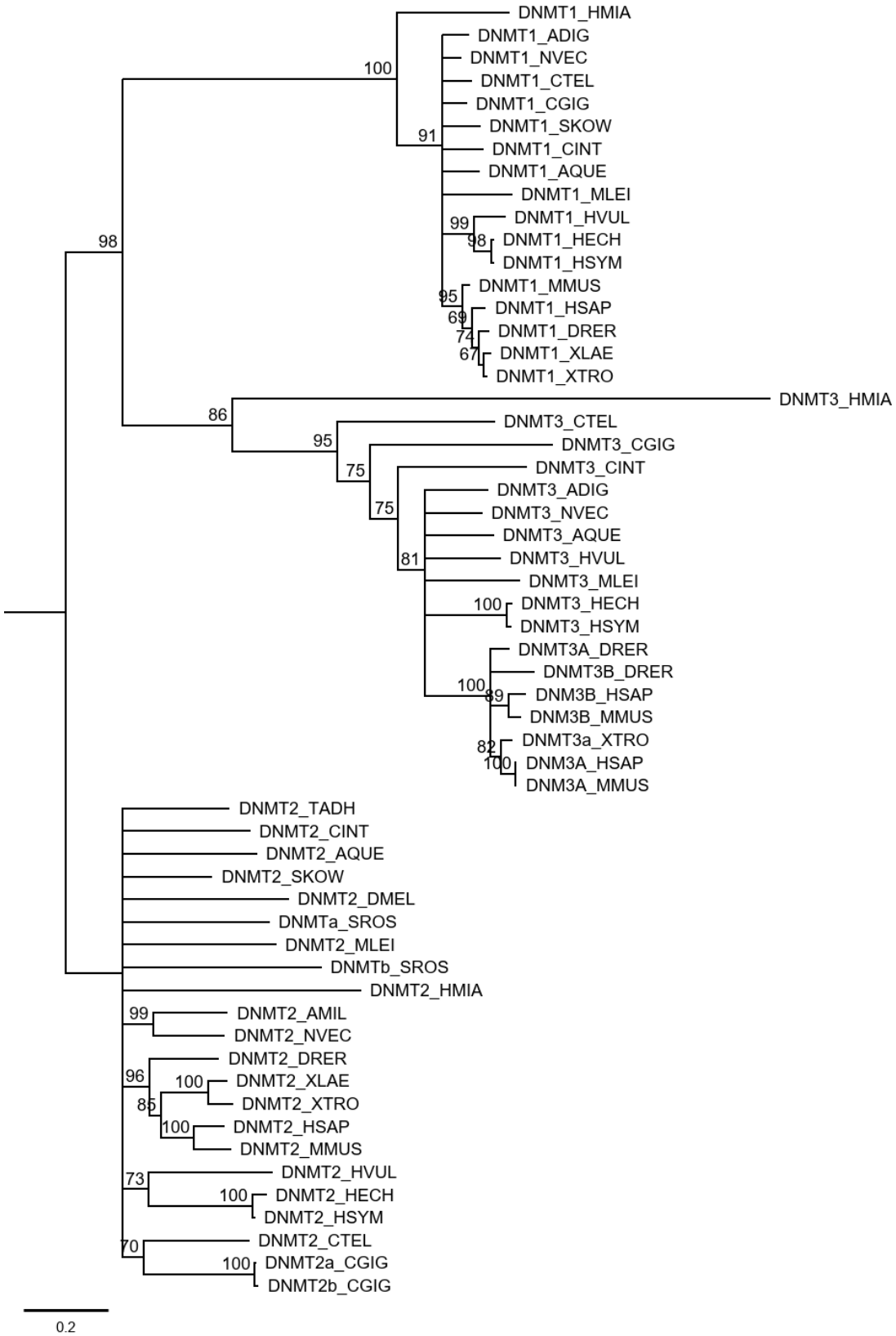
Suppl. Table 2. HPLC-QQQ data from Adult specimens of early diverging animals

Species	6mA/A	5mC/C	stdev 6mA	stdev 5mC	replicates
H. symbiolongicarpus	0.018%	3.198%	0.008%	0.757%	4
H. echinata	0.025%	2.196%	0.009%	0.417%	3
N. vectensis	0.014%	1.882%	0.007%	0.227%	3
M. leidyi	0.031%	0.905%	0.001%	0.065%	4
unmodified oligo	0.017%	0.028%	0.006%	0.033%	3

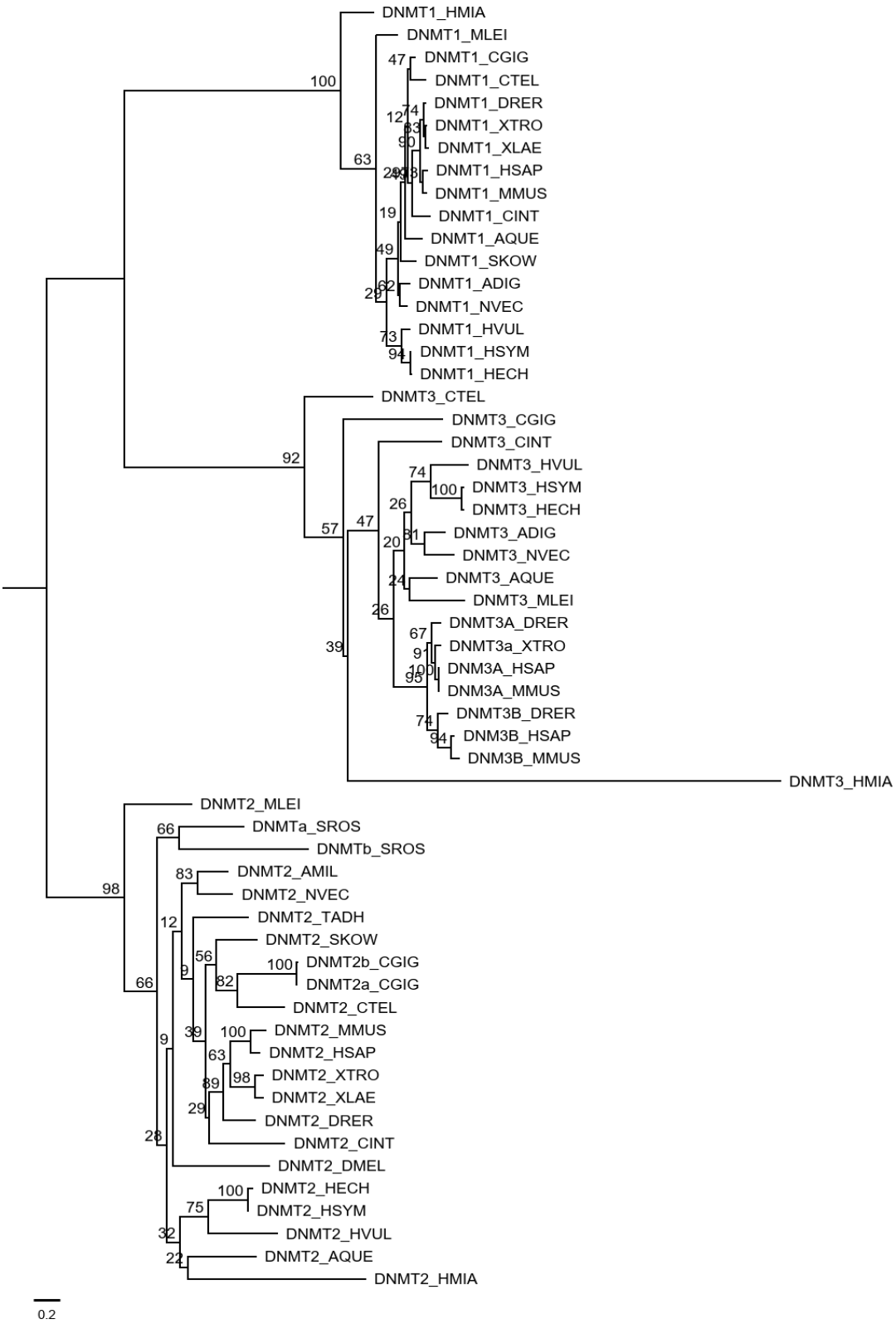
Suppl. Table 3. HPLC-QQQ data of embryonic stages of *Hydractinia*.

Samples	6mA/A	5mC/C	stdev 6mA	stdev 5mC	From Run*	Rep.
Sperm	0.0152%	2.5611%	0.0080%	0.5614%	9_1	2
2-4 cells	0.0305%	1.9146%	0.0062%	0.4178%	7_3 & 9_3	4
16-32 cells	0.0610%	2.2695%	0.0149%	0.6585%	5_1 & 7_2	4
64-128 cells	0.0211%	2.4194%	0.0068%	0.9198%	5, 7,11 &12	6
8 hpf	0.0193%	2.9366%	0.0063%	0.2828%	9_2 & 12_3	6
24 hpf	0.0212%	3.3237%	0.0059%	0.7058%	5_2 & 7_1	6
Polyps	0.0178%	3.1983%	0.0083%	0.7571%	12	4
unmodified oligo	0.0171%	0.0277%	0.0063%	0.0335%	9_1 & 14_2	3

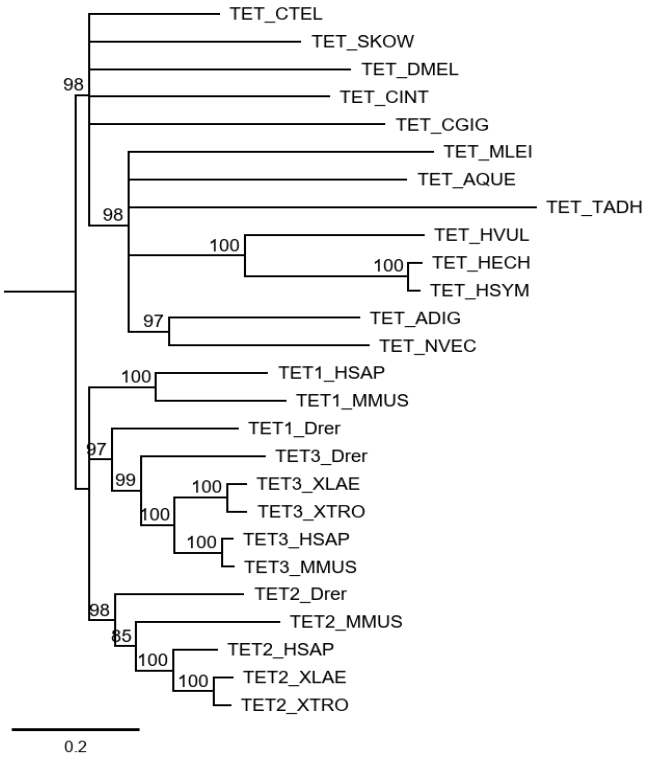
6.2 Unedited phylogenetic trees



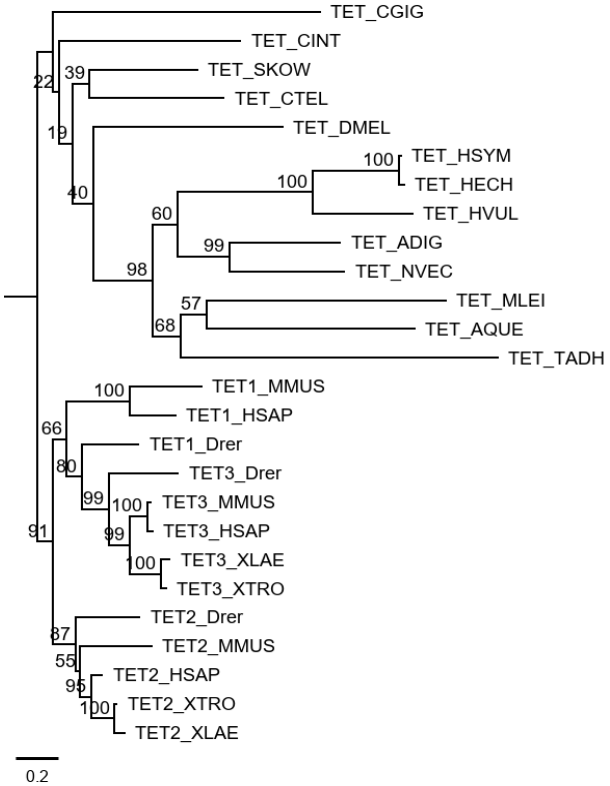
Suppl. Figure 1. Neighbor-Joining consensus phylogenetic tree of DNMT multiple sequence alignment.



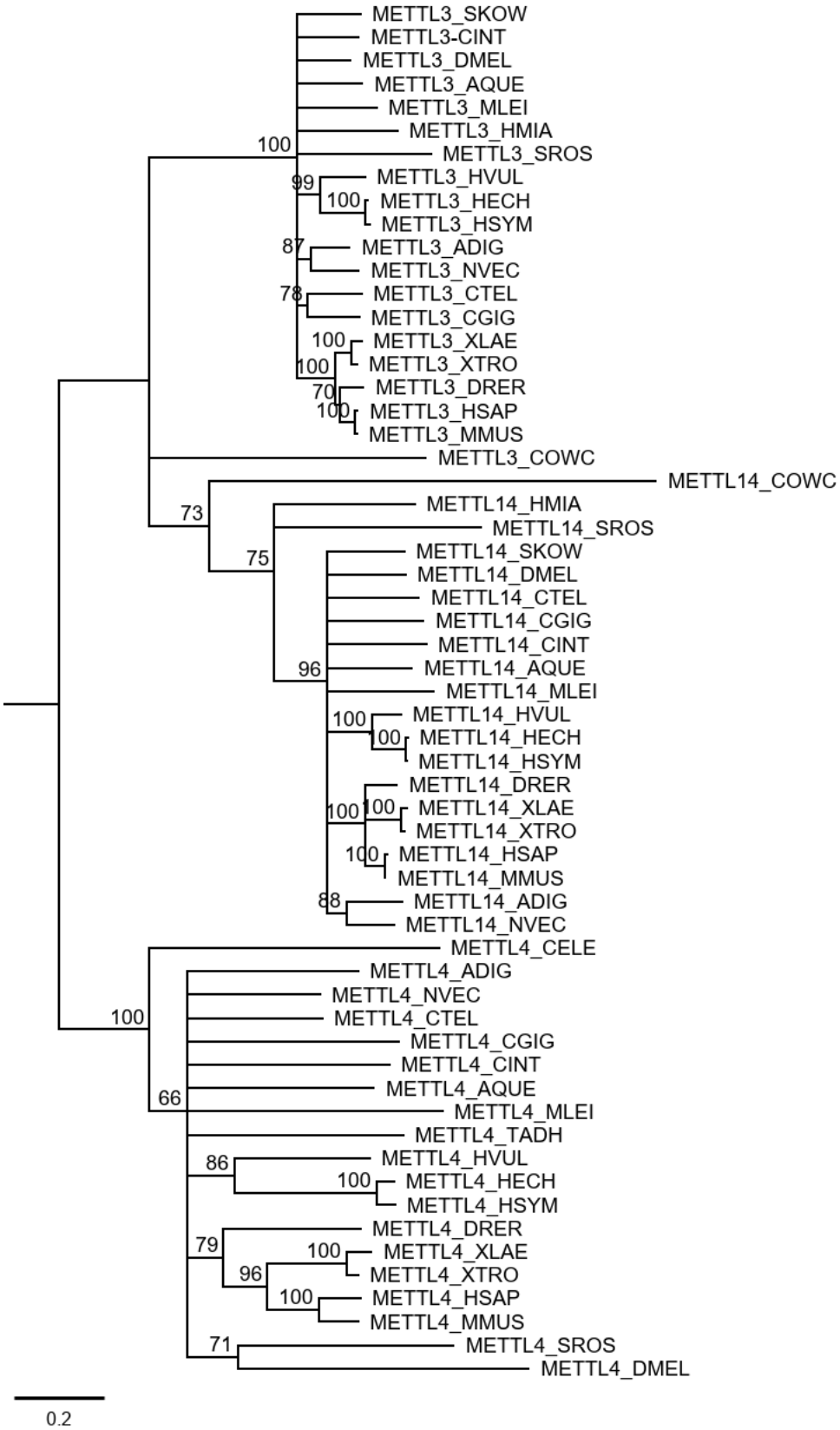
Suppl. Figure 2. Best-scoring maximum likelihood tree (RAxML) of DNMT MSA. Bootstrap support values from 1000 bootstrapping were executed by RAxML.script : /raxmlHPC-SSE3.exe -s input.phy -n output -m PROTGAMMALG -f a -x 1 -N 1000 -p 1 -d. DNMT multiple sequence alignment used were provided in Supplementary_Document_2.



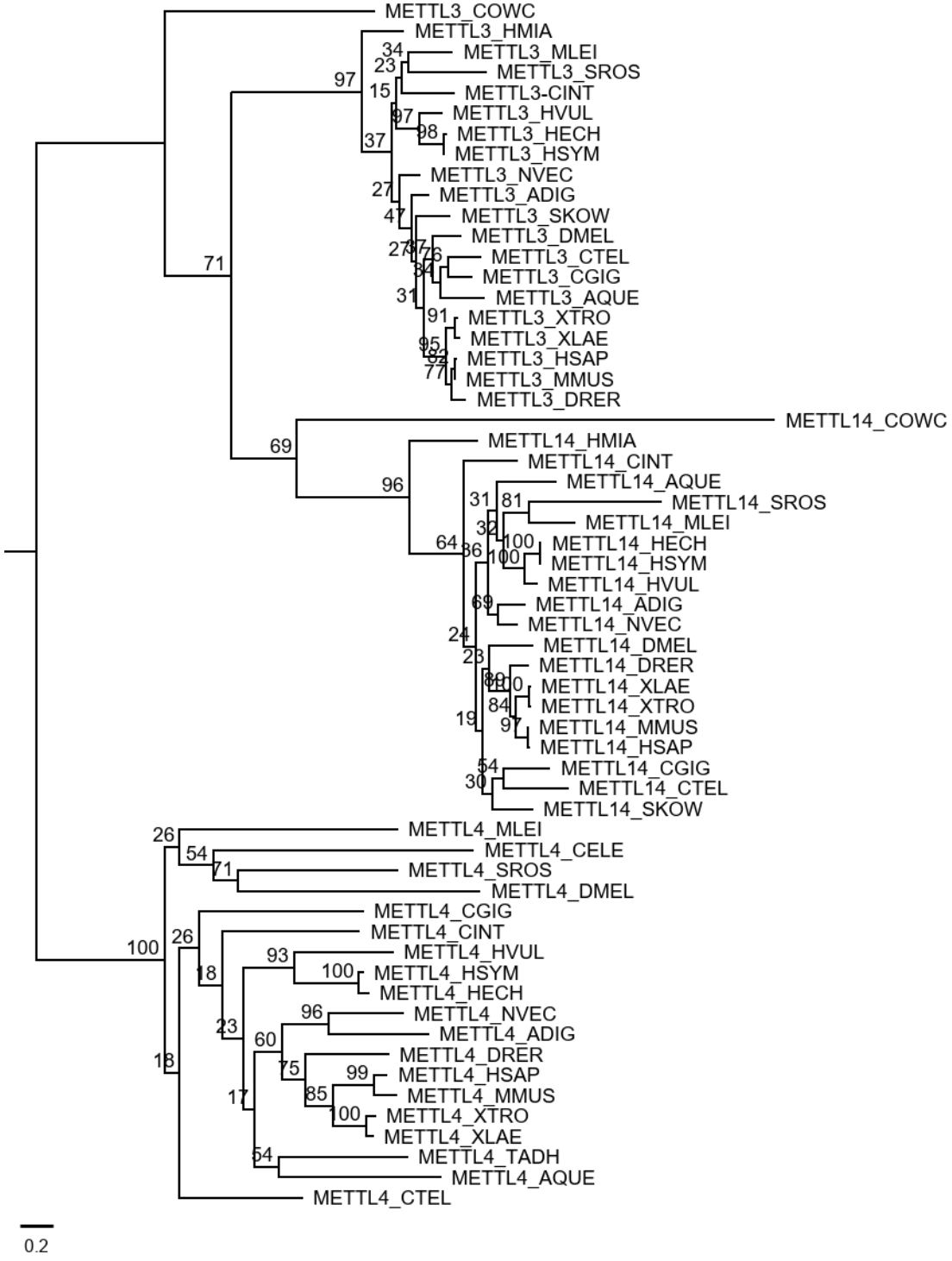
Suppl. Figure 3. Neighbor-Joining consensus phylogenetic tree of TET multiple sequence alignment.



Suppl. Figure 4. Best-scoring maximum likelihood tree (RAxML) of TET MSA. Bootstrap support values from 1000 bootstrapping were executed by RAxML.script : /raxmlHPC-SSE3.exe -s input.phy -n output -m PROTGAMMALG -f a -x 1 -N 1000 -p 1 -d. TET multiple sequence alignment used were provided in Supplementary_Document_3.

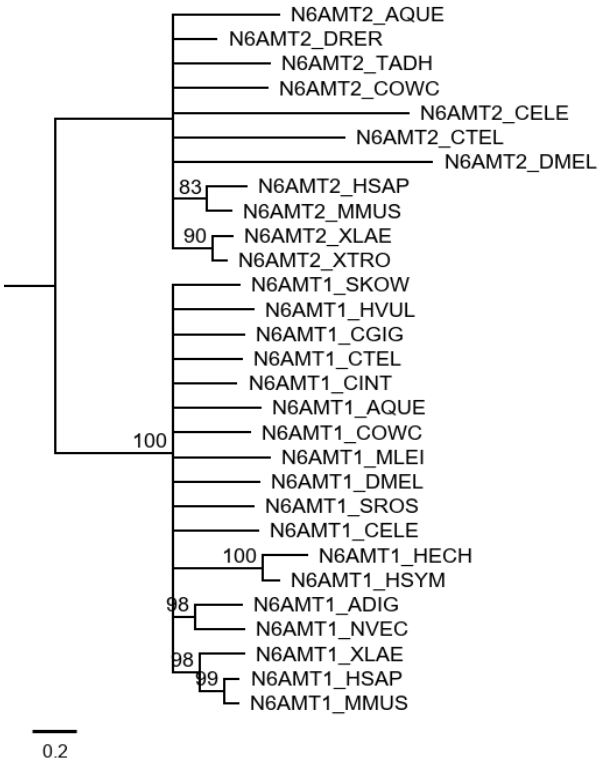


Suppl. Figure 5.Neighbor-Joining consensus phylogenetic tree of METTL multiple sequence alignment.

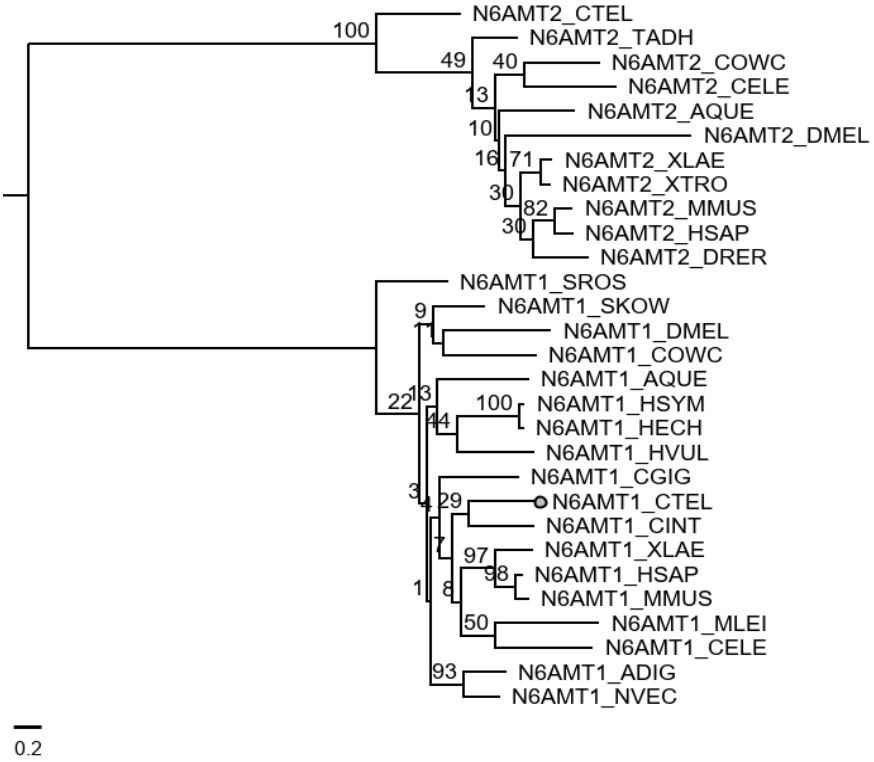


Suppl. Figure 6. Best-scoring maximum likelihood tree (RAxML) of METTL MSA.

Bootstrap support values from 1000 bootstrapping were executed by RAxML.script : /raxmlHPC-SSE3.exe -s input.phy -n output -m PROTGAMMALG -f a -x 1 -N 1000 -p 1 -d. METTL multiple sequence alignment used were provided in Supplementary_Document_4.

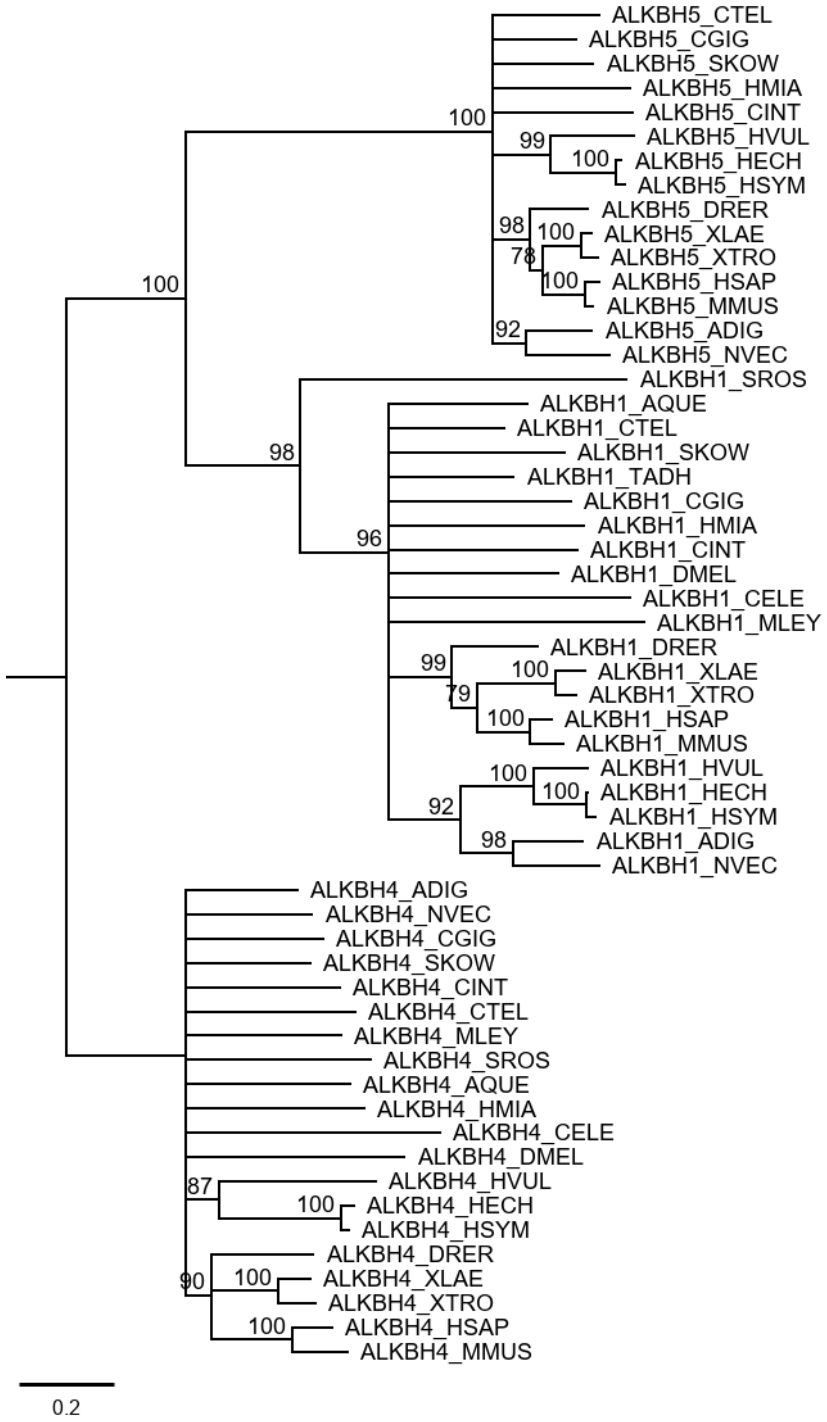


Suppl. Figure 7. Neighbor-Joining consensus phylogenetic tree of N6AMT MSA

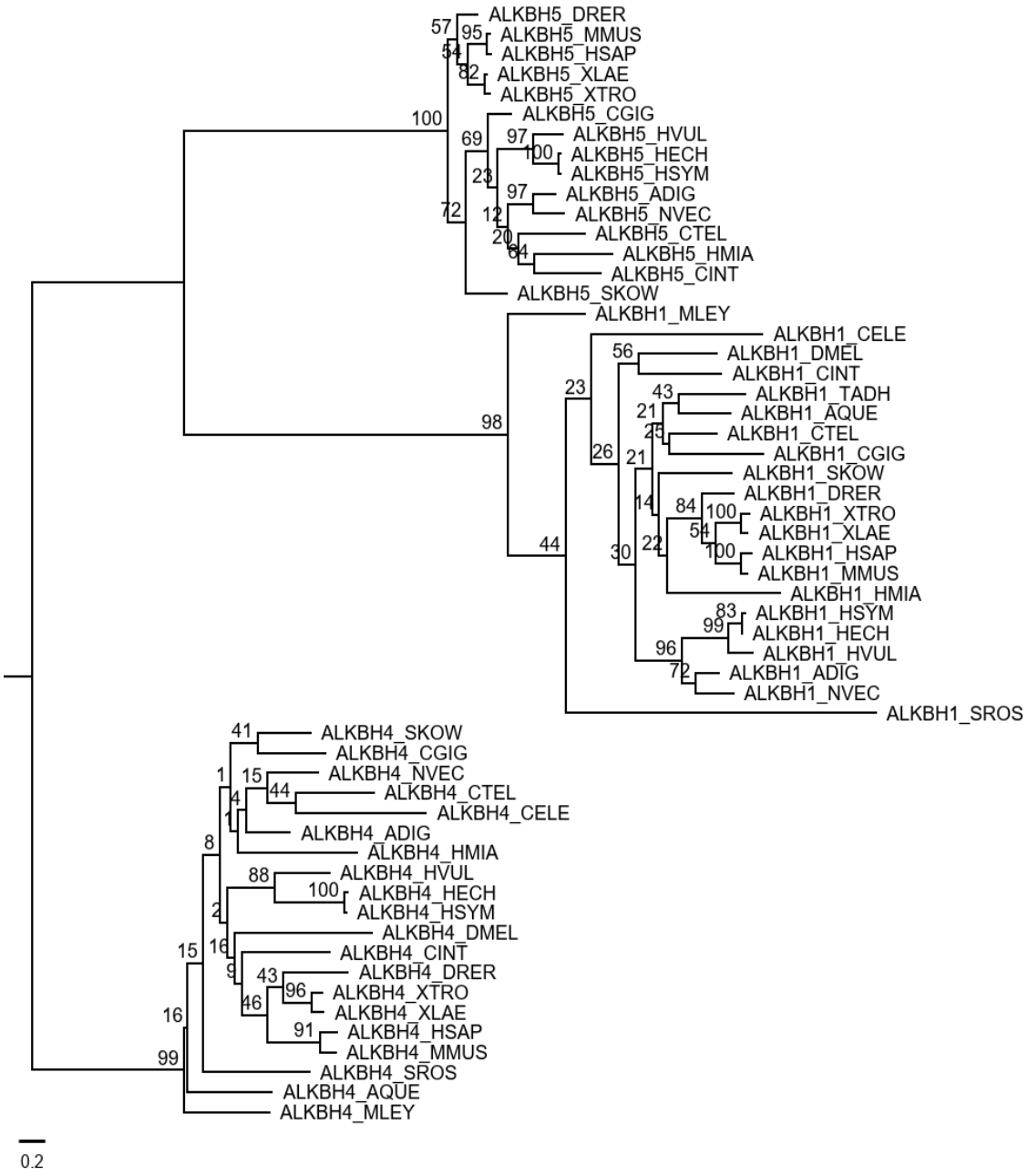


Suppl. Figure 8. Best-scoring maximum likelihood tree (RAxML) of N6AMT MSA.

Bootstrap support values from 1000 bootstrapping were executed by RAxML.script : /raxmlHPC-SSE3.exe -s input.phy -n output -m PROTGAMMALG -f a -x 1 -N 1000 -p 1 -d. N6AMT multiple sequence alignment used were provided in Supplementary_Document_5.



Suppl. Figure 9. Neighbor-Joining consensus phylogenetic tree of ALKBH multiple sequence alignment.



Suppl. Figure 10. Best-scoring maximum likelihood tree (RAxML) of ALKBH MSA.

Bootstrap support values from 1000 bootstrapping were executed by RAxML.script : /raxmlHPC-SSE3.exe -s input.phy -n output -m PROTGAMMALG -f a -x 1 -N 1000 -p 1 -d. ALKBH multiple sequence alignment used were provided in Supplementary_Document_6.

6.3 DNMT1 nine sequence alignment with annotated domain

```
#NEXUS
begin taxa; dimensions ntax=8;
taxlabels
'DNMT1_ADIG'
'DNMT1_NVEC'
'DNMT1_HSYM'
'DNMT1_HVUL'
'DNMT1_AQUE'
'DNMT1_HSAP'
'DNMT1_MMUS'
'DNMT1_MLEI'
;
end;

begin characters; dimensions nchar=1799;
format datatype=protein missing=? gap=- interleave=yes;
matrix
'DNMT1_ADIG' M----------VQLKDDFPP-AISERLSELEDEYNDGDITEKGYVRKKCKLMKPLLANF
'DNMT1_NVEC' M----------VACEEILAPEVISQRLQDLEAEFLDGDLTEKGYIKKKCKLLKSSLPSD
'DNMT1_HSYM' ------------------MP-VPAEE---------------------------------
'DNMT1_HVUL' ------------------MP-VLTDG---------------------------------
'DNMT1_AQUE' ---------------------MLGGE----EEE---------------ELL--FQTSS
'DNMT1_HSAP' MPARTAPARVPTLAVPAISLPDDVRRRLKDLERD----SLTEKECVKEKLNLLHEFLQTE
'DNMT1_MMUS' MPARTAPARVPALASPAGSLPDHVRRRLKDLERD----GLTEKECVREKLNLLHEFLQTE
'DNMT1_MLEI' ------------------------------------------------------------
                                          DMAP1-Binding Domain (PF06464)

'DNMT1_ADIG' QQERIQEIEDDLKAGKFSEEQFISHLKELLTEI--------GH-RSSNGCGAKPLPPVCN
'DNMT1_NVEC' VKTRLTQLEQRFAEKRITEDKYLSLMKQLLNEK--------TQTQESNG----------H
'DNMT1_HSYM' ----------------------------------------KAASG----------CS
'DNMT1_HVUL' ----------------------------------------REATD----------H
'DNMT1_AQUE' KKKRVPRIIDD--DDDEMEVSRVTTAKRSNNK------------RKSSG----------T
'DNMT1_HSAP' IKNQLCDLETKLRKEELSEEGYLAKVKSLLNKDLSLENGAHAYNREVNG---------RL
'DNMT1_MMUS' IKSQLCDLETKLHKEELSEEGYLAKVKSLLNKDLSLENGTHTLTQKANG---------CP
'DNMT1_MLEI' ------------------------------------------------------------
                          DMAP1-Binding Domain (PF06464)

'DNMT1_ADIG' KDSNNVTLNAEPMETGMEIN---ASDIPYTQASDCHEMEEANDSEGEPSSQSEPGLRQSS
'DNMT1_NVEC' TSSSHVNQNGVAMPTDQGVSSEVSTEPALSSTLSSSGMADRVDNKGGNSESEVEGVGGSD
'DNMT1_HSYM' KDGPSV------------------------------------------------------
'DNMT1_HVUL' KACGS-------------------------------------------------------
'DNMT1_AQUE' GASSNLS-----------------------------------GGNSTNKSSKI----
'DNMT1_HSAP' ENGNQAR----------------------SEARRVGMADANSPPKPLSKPRTPRRSKSD
'DNMT1_MMUS' ANGSRPT-------------------------WRAEMADSNRSPRSRPKPRGPRRSKSD
'DNMT1_MLEI' ------------------------------------------------------------

'DNMT1_ADIG' CN-NSNEPTSSNGAERSQCLLDIKVSLTDLKKSTSPTNKCSQEGNKIKVNKKESKKQPGI
'DNMT1_NVEC' MDVDSGVASDAMTSDFSQPSASSKSSDEETETPAKKQQKAKHSSQKKKGKKSSPMRQAGI
'DNMT1_HSYM' ------------------------------------------------DKEQKTSKQATI
'DNMT1_HVUL' -----------------------------------------------SKTNRQSKQSTI
'DNMT1_AQUE' ---------------------------------------------QSSSTSGRQLGI
'DNMT1_HSAP' GEAKP-EPSPSP-----------------------------------RITRKSTRQTTI
'DNMT1_MMUS' SDTLSVETSPSS-----------------------------------VATRRTTRQTTI
'DNMT1_MLEI' ------------------------------------------------MVRNVKQTDL

'DNMT1_ADIG' KEMFAKNALKRKKTEDSSTDDSAGQAELSSENSEASN----------------------
'DNMT1_NVEC' TEMFSKVSTKRKETEKD--EGVTG----CSSVDEEAK----------------------
'DNMT1_HSYM' ASLFQRSSKRKKCKEES----------LSVVGNEN-------------------------
'DNMT1_HVUL' NALFKKSVKKTKRS-------------ISPSKEEN-------------------------
'DNMT1_AQUE' GSFFNKDPQKRSKVEEE-KEEVSSA--IAPTGEEEREEEE--------------------
'DNMT1_HSAP' TSHFAKGPAKRKPQEES--ERAKS---DESIKEEDKDQDEKRRRVTSRERVARPLPAEEP
'DNMT1_MMUS' TAHFTKGPTKRKPKEES--EEGNS----AESAAEERDQDKKRRVVDTESGAAAAV--EKL
'DNMT1_MLEI' RQMFSAKVEKIQKSNEK--TNIRT----SAVIAD-------------------------
```

```
'DNMT1_ADIG'  ---------------YSQDGKR------------QKTSG----DDC-----DEEESSG-D
'DNMT1_NVEC'  ---------------SAQEEKK------------QKLVD----DEH-----EQKDK----
'DNMT1_HSYM'  ------------------EKKK------------QKTEDVVPPDSNG-IKGEENEK----
'DNMT1_HVUL'  ------------------ENKK------------KRTDG---------------------
'DNMT1_AQUE'  --------------GKEKEAKK------------AKTSE----TGSS----SNGTSAA-T
'DNMT1_HSAP'  ERAKSGTRTEKEEERDEKEEKR-LRSQTKEPTPKQKLKEEPDREARAGVQADEDEDGD-E
'DNMT1_MMUS'  EEVTAGTQLGPEEPCEQEDDNRSLRRHTRELSLRRKSKEDPDREARPETHLDEDEDGKKD
'DNMT1_MLEI'  ------------------EGQSSLTDIA------SKNVE----DNCL----AEGEPA---

'DNMT1_ADIG'  SG----------------------------IERQSLRE----STVGIRSEE--------
'DNMT1_NVEC'  --------------------------------ESYGLRENNSQNTNTVKNE---------
'DNMT1_HSYM'  ---------------------------------AKMDVDE-KTDIKQGVSLEK--------
'DNMT1_HVUL'  -----------------------------------DIKE--KNTSFLTNSEK--------
'DNMT1_AQUE'  NGN---------------------------NKSHDLRP-HDDIPANKTSEAP--------
'DNMT1_HSAP'  KDEKKHRSQPKDLAAKRRPEEKEPEKVNPQISDEKDEDE-KEEKRRKTTPKEP-------
'DNMT1_MMUS'  KRSSRPRSQPRDPAAKRRPKEAEPEQVAPETPEDRDEDE-REEKRRKTTRKKLESHTVPV
'DNMT1_MLEI'  --------------------------------SKRQKTEKYNREINESTKRKS--------

'DNMT1_ADIG'  ----------------KRAQPPAKCKECKQLLNSPDLRLFPGDSNDAVEEFVALTDPRLS
'DNMT1_NVEC'  -----------------KPQPPARCKECRQLIDSPDLCLFAGDPADAVEEFVALVDPRLS
'DNMT1_HSYM'  ----------------SSKKVIERCHQCRQIIHKDEIKMFGGDSPDAVDEFVMLCDPRLS
'DNMT1_HVUL'  ----------------PGVKTVEKCKECRQLLNSNDIKLFQGDHSDALEEFAMLVDPRLS
'DNMT1_AQUE'  ----------------QSSKPP--CTYCKRSSDDPRLKIFIGDPPNANDEFITLADPSLS
'DNMT1_HSAP'  ---TEKKMARAKTVMNSKTHPP-KCIQCGQYLDDPDLK-YGQHPPDAVDEPQMLTNEKLS
'DNMT1_MMUS'  QSRSERKAAQSKSVI-PKINSP-KCPECGQHLDDPNLK-YQQHPEDAVDEPQMLTSEKLS
'DNMT1_MLEI'  ----------------PALKPV-KCDVCKQLLNSEALKVFKGDESGAVEEFIALASDTLS

'DNMT1_ADIG'  LFSGEEEQCDS YSDTPQHKITNFSIYDKNTHLCPFDSGLIEKNVELFFSGYLKPIYDENP
'DNMT1_NVEC'  LFSGDEEQFDS YEDKPQHKITNFSVYDKNTHLCAFDTGLIEKNVELFISGYVKPIYDENP
'DNMT1_HSYM'  LLTGNEQEYDN YDDRPQHKLTDFSVYDKCDHLCAFDTGLIDKNIELFFSGYVKPVYDENP
'DNMT1_HVUL'  LLSGNEQDFDA YEDRPQHKVTEFSVYDKCGHLCAFDTGLIEKNVELFFSGYVKPIFDENP
'DNMT1_AQUE'  VLSADQE--SA LDDVPQHKITGFSVYDKNHHLCHFDTGLVEKNVELFFSGWVKPIYDENP
'DNMT1_HSAP'  IFDANESGFES YEALPQHKLTCFSVYCKHGHLCPIDTGLIEKNIELFFSGSAKPIYDDDP
'DNMT1_MMUS'  IYDSTSTWFD TYEDSPMHRFTSFSVYCSRGHLCPVDTGLIEKNVELYFSGCAKAIHDENP
'DNMT1_MLEI'  LFDADSNGESE WTEKPQHRITNFTVYDEYGHMCPFDSGLIEKNVPLYFSGYIKPIYDDSP
```

DNMT1-RFD (PF12047)

```
'DNMT1_ADIG'  SPEGGVPTKNIGPINEWWVAGFDGGENALIGFTTAFAEYILMQASEDYMPFMNIMREKIA
'DNMT1_NVEC'  SPEGGVPTKILGPINEWWTAGFDGGENALIGLTTAYGEYFLMNPSKEYAPFIHTMREKIQ
'DNMT1_HSYM'  DKEGGVCTKKMGPINEWWISGFDGGENALIGFSTAFAEYILMRPSDEYTSFMNAVSEKIC
'DNMT1_HVUL'  DIEGGISTKAMGPINEWWIAGFDGGENALIGFGTAFAEYILMRPSDDYASFMNAVTEKIY
'DNMT1_AQUE'  DPSDGIPTRQLGPINSWWIAGFDGGEKALVGFSTAYAEYILMDASDDYAEIMASVQEKIY
'DNMT1_HSAP'  SLEGGVNGKNLGPINEWWITGFDGGEKALIGFSTSFAEYILMDPSPEYAPIFGLMQEKIY
'DNMT1_MMUS'  SMEGGINGKNLGPINQWWLSGFDGGEKVLIGFSTAFAEYILMEPSKEYEPIFGLMQEKIY
'DNMT1_MLEI'  SLEGGVAGSKLGPIDSWYIAGFDGGEKELIGFTTAYADYLLSQPSDQYSPIFDMLREKTY
```

DNMT1-RFD (PF12047)

```
'DNMT1_ADIG'  MSKIVVEFMQSNPEARYEDLLNKVETSVPPA--NCSTFTEDTLLRHAQFLVEQVESYDQA
'DNMT1_NVEC'  LSKIVIEFLLNNPEARYEDLLNKVQVSVPPE--GCPSFTEDSLLRHAQFLVEQIENYDSA
'DNMT1_HSYM'  MSKVVIEFIISNPEARYEELLNKIETTVPPE--NCSQLTEDSLLRHAQFLVDQVESFDSA
'DNMT1_HVUL'  MSKIVIEFLLNNPDSEYEELLNKLETTVPPE--NCAKFTEDTLLQHAQFLVEQVESYDSA
'DNMT1_AQUE'  LSKILIEYLEEFPAATYEDLLNKIQTSVPPQSIGCTSFTEDSLLRHAQFIVEQVESYDQY
'DNMT1_HSAP'  ISKIVVEFLQSNSDSTYEDLINKIETTVPPSGLNLNRFTEDSLLRHAQFVVEQVESYDEA
'DNMT1_MMUS'  ISKIVVEFLQNNPDAVYEDLINKIETTVPPSTINVNRFTEDSLLRHAQFVVSQVESYDEA
'DNMT1_MLEI'  LSKTVIEILVEDSGMTMEDLLNCIQLVVPPE--NCSKFTEETLLRHAQFLVEQIESYDHA
```

DNMT1-RFD (PF12047)

```
'DNMT1_ADIG'  A--DDDELPLLISPCMRDLIKLAGVTLGKR---RA-ARGVKVRTE----KKQARPSKATT
'DNMT1_NVEC'  G--DTDELPLIVTPCIRDFIKLAGVTLGKR---RT-ARKIKVKSRA-ELKKKQAPTKATT
'DNMT1_HSYM'  AAENEEEPPLIATNCMRDLIKLAGVTLGKR---RQ-LRGVKVKAE----KKALGPTLATT
'DNMT1_HVUL'  ALDEEDSPRLITSCCMRDLIKLAGVTLGKR---RQ-MRGLKVKEE----KKTLGPTLATT
'DNMT1_AQUE'  C--DEDEDLLLVSPCMRALIKLAGVTLGKRLSSRR-PHQPKVKKT----KQ----TKATT
'DNMT1_HSAP'  G--DSDEQPIFLTPCMRDLIKLAGVTLGQR---RAQARRQTIRHST--REKDRGPTKATT
'DNMT1_MMUS'  K--DDDETPIFLSPCMRALIHLAGVSLGQR---RA-TRRVMGATK----EKDKAPTKATT
'DNMT1_MLEI'  A--DEDEDLLITIPAVRGLIELAGVTLGGNQQSRARPERPVVKSKAAGKTKSTDISHATV
```

```
'DNMT1_ADIG'  TPLVRQIFDIFFKDQID----GKEIGAIRRKRCGVCEVCQLPDCGKCKSCKDMVKFGGTG
'DNMT1_NVEC'  TELVRHIFDIFFQNQID----GKNGNAPRRKRCGVCETCQQPNCGKCSACRDMVKFGGTG
'DNMT1_HSYM'  TPLVRYVFDTFFKDQID----LKATSAQRKRRCGVCEICQQPDCGKCKACKDMVKFGGTG
'DNMT1_HVUL'  TPLVRHVFDTFFKNQID----LKGVTTQRRKRCGVCEICQQPDCGVCRSCKDMVKFGGSG
'DNMT1_AQUE'  TPLVRDIFETFFKNQID----NKSSSAPRRRRCGVCETCQQPDCGKCNACADMIKFGGTG
'DNMT1_HSAP'  TKLVYQIFDTFFAEQIEKDDREDKENAFKRRRCGVCEVCQQPECGKCKACKDMVKFGGSG
'DNMT1_MMUS'  TKLVYQIFDTFFSEQIEKYDKEDKENAMKRRRCGVCEVCQQPECGKCKACKDMVKFGGTG
'DNMT1_MLEI'  TPLVQSSFEAIFVDQMQ--------ATSRKTRCNECEVCNLPDCGECAACKDMFKFGGSG
```

Zf-CXXC (PF02008)

```
'DNMT1_ADIG'  KKKQCCEERRCPNMAVKEADEDDDIGEDIEDAENVTLNKRITKKKSPTPKKTKS----KS
'DNMT1_NVEC'  KSKQCCINRRCPNMMVREADDDEGLEDEEKDDN-GDKENKAESPHKPHHKTIKGQKTIKT
'DNMT1_HSYM'  KAKQCCINRRCPNLAVQQADENAESQDEDAEPLKSPLKSPLKEDRSPKLSRKGSEKIVKT
'DNMT1_HVUL'  RSKQCCINRRCPNLAVQEAEEDALCSQDEDEPM---LKSPSKTDLSPRHHRKGQEK--KS
'DNMT1_AQUE'  RSKQACVNRRCPYMAVQTAEEEED--NDAADPD-------LKNVKDLKSPSKKIKK-IKT
'DNMT1_HSAP'  RSKQACQERRCPNMAMKEADDDEEVDDNIPE---------MPSPKKMHQGKKKKQN--KN
'DNMT1_MMUS'  RSKQACLKRRCPNLAVKEADDDEEADDDVSE---------MPSPKKLHQGKKKKQN--KD
'DNMT1_MLEI'  KKRQACVKRKCPKIAVIRAEENAK--DETDEPN---LAQSNGQKKKLPEVKTEKKVSSNI
```

Zf-CXXC (PF02008)

```
'DNMT1_ADIG'  KVQWIGDPE-VHGK-QNYYTSVLI-DKEEICVGDFVMFRPDENSHLALYIACVRYMWEEP
'DNMT1_NVEC'  KVEWLGEPAWIEGK-KKYYSSVLI-NKKKVSLGDFVKVCPDD-PHIPLYIGCVQYMWESS
'DNMT1_HSYM'  KCKWACEPH-IEKKSKKMYKSVYI-NNEMITVGDYVKVCPDN-PTDPLYICRVTYMWEDF
'DNMT1_HVUL'  FVKWACEDF-VERKGKKLYKSVQI-NNELINIGEFVQVYPTD-PSDPLYICRVMYMWEDL
'DNMT1_AQUE'  KVEWIGEPDFVEGG-KSYYTEVLINNKEKVCLYDVVSVCPEV-PDDPLYLTRIMSMYEDS
'DNMT1_HSAP'  RISWVGEAVKTDGK-KSYYKKVCI-DAETLEVGDCVSVIPDD-SSKPLYLARVTALWEDS
'DNMT1_MMUS'  RISWLGQPMKIEEN-RTYYQKVSI-DEEMLEVGDCVSVIPDD-SSKPLYLARVTALWEDK
'DNMT1_MLEI'  DGKLTGDPVSKSGK-TTYYKTAEV-DGNSLTLGDYVTIAPDT-PDIAPYVGRIISLFRK-
```

BAH (PF01426)

```
'DNMT1_ADIG'  NGDKMFHCRWFSRGSETILGETSDPREVFLLDKCDDNLLGCIKQKCTVTYNKPDADWFMK
'DNMT1_NVEC'  SGEKLFHARWMTRGAETVLGETADSAELFLCDDCDDNPLGSVLETCEVTYKDPSVDW-LQ
'DNMT1_HSYM'  NGKKMFHAQWLYRSAETILGETGDPAELFLCDDCDNNPLGSIMCKCKVQFNLPTKDWFNE
'DNMT1_HVUL'  NGDKKFHAQWLYRSSETVLGEVGDPSEVFLSDDCDDIKLGAIMSKCNVSSKFASENWFME
'DNMT1_AQUE'  NGKKMFHGWWFHRSTDTVLGETGDPRELFLIDDCEDNPLGAIMDKVEVEYKPPVSNWFMC
'DNMT1_HSAP'  SNGQMFHAHWFCAGTDTVLGATSDPLELFLVDECEDMQLSYIHSKVKVIYKAPSENWAME
'DNMT1_MMUS'  NGQMMFHAHWFCAGTDTVLGATSDPLELFLVGECENMQLSYIHSKVKVIYKAPSENWAME
'DNMT1_MLEI'  GGDNNLHVHWLNRASDTVLGEAHDEHELFLSDLCDDAPLGSVLGKVTVSYTPPSSDWSMS
```

BAH (PF01426)

```
'DNMT1_ADIG'  GGLDEPEGDVSMEENDGNTFFFQKWYDPDHGRFTDPPAEYLHVVNSEDN-------FRYC
'DNMT1_NVEC'  GSAG-MEIETSDDEHNDNTFFYHKWYHPEMGRFVDPPPEYQTKVDRTQQ--------KFC
'DNMT1_HSYM'  GGSNVIEHD---ADDDKEILTLQKWYDVEEGAFIDPPVEFLY--NCFDE--------RYC
'DNMT1_HVUL'  GGKE----D-CIISEENNELFYQKWYDYEDGLFTDPPTEFLL--SKFDE--------KYC
'DNMT1_AQUE'  GGEEVSDDE-KEIEEDGKTFFVQKYYDQSLARFEDIPSEYIRYLISDDAHPPTDGFIPQC
'DNMT1_HSAP'  GGMD-PES--LLEGDDGKTYFYQLWYDQDYARFESPPKTQPT----EDNK------FKFC
'DNMT1_MMUS'  GGTD-PETT-LPGAEDGKTYFFQLWYNQEYARFESPPKTQPT----EDNK------HKFC
'DNMT1_MLEI'  GGM-------PVPEKQNEELFFKFWYDSDRARFEAPREEYLD--WNVDN--------GCC
```

BAH (PF01426)

```
'DNMT1_ADIG'  ESCVRQSAEENLETLSLGEPLEVDSKPSSKSYYKSCSKGGNRYKSGDCVFLEPDAFSFNV
'DNMT1_NVEC'  ESCIRISAEEKLREPTVQERIDDNMDKTSKEFYRSFTLNGSEFTVSDCVYLDPEAFSFNV
'DNMT1_HSYM'  PCCDRESRKKKVNTPQLMEELEDEN-TDSTVYYRKFLWMEQEFAIGDCVYINPEAYKFKK
'DNMT1_HVUL'  PSCERTKTKLKYKQPTLGDELETENSTKSEHCFKTVTWQGVCYSLGDCVYLDPDAFTFKI
'DNMT1_AQUE'  ASCERRNHMKTVHSAVPINKIESEK--PKITNYSSFQLNEESYSIGDCVYLSPHTYSFPN
'DNMT1_HSAP'  VSCARLAEMRQKEIPRVLEQLEDLD---SRVLYYSATKNGILYRVGDGVYLPPEAFTFNI
'DNMT1_MMUS'  LSCIRLAELRQKEMPKVLEQIEEVD---GRVYCSSITKNGVVYRLGDSVYLPPEAFTFNI
'DNMT1_MLEI'  PACIRKSEAERADIP----HLDTDT---NVLMYKNVQ-----YSIGDYVYITPDAFSMPN
```

BAH (PF01426)

```
'DNMT1_ADIG'  KPKES-----KKGFKKDDMVDEEKYPEYYRKPLEYVKGSNYDVPESFKIGRIINIF-TKS
'DNMT1_NVEC'  RPKES------KPHKREGSVDEDLYPEYYRKSSEYVKGSNQNVPEPLRIGRIKSIY-VKS
'DNMT1_HSYM'  RVKND-----QIKSKKEETFDEETYPELYRKTKDYIKGSNVDIADPYRIGCILNIM-QKK
'DNMT1_HVUL'  KQKSE-----LPKNKSV-VKDEDEYPELYRKSNDYIKGSNINIVEPFRIGKIISII-KKT
'DNMT1_AQUE'  VKKTSGTAANKKQKKEEEEFDESEYPEKYRKVSDYVKGSNIDSPSPFQIGQVLEIF-SKS
'DNMT1_HSAP'  KLSSP-----VKRPRKE-PVDEDLYPEHYRKYSDYIKGSNLDAPEPYRIGRIKEIFCPKK
'DNMT1_MMUS'  KVASP-----VKRPKKD-PVNETLYPEHYRKYSDYIKGSNLDAPEPYRIGRIKEIHCGKK
'DNMT1_MLEI'  RAKIS-----STKIRKD-NVDDEMFPEYYRK-TDYIKGNNDLVPELYRIGRIVSFS-TKQ
```

BAH (PF01426)

```
'DNMT1_ADIG'  SPGKLSKDL-DIMLTVGKFYRPENTHKGSSFAYQADLNLLYWSK-EEATVPFDAVLGSCT
'DNMT1_NVEC'  SSGKLKGDGDDVMLTVTKFYRPENTHKGANASHQADLNLLYWTD-EEAKVSALSVKGKCM
'DNMT1_HSYM'  GS---YGDNTDIVLHVRKFYRPENTHKGFTGSTNSDLNLLYWCD-EEATVSFDYVEGKCY
'DNMT1_HVUL'  DN---YGGMKSIMLKVRKFYRPENTHKGMSGTMNCDLNMVYWSN-EEATVDFQMVEGKCY
'DNMT1_AQUE'  LGGKLIDNNKIVHIKLRMYYRPQDTHKGDEATAQYDLNLLYWSDLVTTVVAGDVVCGKCF
'DNMT1_HSAP'  SNGRPNET--DIKIRVNKFYRPENTHKSTPASYHADINLLYWSD-EEAVVDFKAVQGRCT
'DNMT1_MMUS'  -KGKVNEA--DIKLRLYKFYRPENTHRSYNGSYHTDINMLYWSD-EEAVVNFSDVQGRCT
'DNMT1_MLEI'  EG--LDGS--VVKVTVKKMYRPENI-ENCRVES-ADINLLFYSL-DQVKIRADKLSGKCT
```

BAH (PF01426)

```
'DNMT1_ADIG'  VTCGEDLNCSIAEYTAKAINNFYFLEAYNSETKDFEEPPLEARNAALKGKGKG--KGKGK
'DNMT1_NVEC'  VACGEDIHSSIQEYSSKG-DCFYYLEAYNSQTKEFEELPPQARDCSSKGKGKG--KGKGK
'DNMT1_HSYM'  VVYIDDPDIDVNEYTNDGPDRFYFTESYDPNTEEFDIPPSKARNELSGGKSSG--KGKGK
'DNMT1_HVUL'  VLFADDADMDMFKFTEAGPDRFYFREAYDADKKEFDVPPREAWN---SNKGKG--KGKSK
'DNMT1_AQUE'  VKFKEDITEDIDSYFSNKPNHFYFVEAYCADTKEFEDPPVHAMN---KGKGKGVSKGKGK
'DNMT1_HSAP'  VEYGEDLPECVQVYSMGGPNRFYFLEAYNAKSKSFEDPPNHARS--PGNKGKG--KGKGK
'DNMT1_MMUS'  VEYGEDLLESIQDYSQGGPDRFYFLEAYNSKTKNFEDPPNHARS--PGNKGKG--KGKGK
'DNMT1_MLEI'  VKNSEDI-EDLSRYTTLS-DHFYFTESWNSASKCLEDVPSEGRS---SNKGK---KKRAG
```

BAH (PF01426)

```
'DNMT1_ADIG'  ARSACQT-EE--------QNLRQDSKPSETVPKLRCLDVFAGCGGLSEGLHQAGTAESLW
'DNMT1_NVEC'  GKGKSSA-DSGP------ESCMHDQQKEAHVRKLRSLDVFAGCGGLSEGLHQAGVAESLW
'DNMT1_HSYM'  ARKKGSIFQQ--------SALLTNHPVYEKVDKLRTLDIFAGCGGLSEGLHQVGVADSRW
'DNMT1_HVUL'  STK-----QQ--------NGLESNFPAYEKVEKLRTLDIFAGCGGLSEGLDQVGVVNSCW
'DNMT1_AQUE'  GKGPAKGSKESIATTSTDDDKKDEITPDSSFKKLRMLDVFAGCGGLSEGFHQAGVADSCW
'DNMT1_HSAP'  GKPKSQACEP----------SEPEIEIKLPKLRTLDVFSGCGGLSEGFHQAGISDTLW
'DNMT1_MMUS'  GKGKHQVSEP----------KEPEAAIKLPKLRTLDVFSGCGGLSEGFHQAGISETLW
'DNMT1_MLEI'  IKS-----EK----------------ETQFRKLRTLDVFAGCGGLSAGFHQAGIAESCW
```

DNA_methylase (PF00145)

```
'DNMT1_ADIG'  AVEKEEPAAHAFSLNNPGCTVFTDDCNLLLKLVMDGEKKNSRGQTLPQKGEVELLCGGPP
'DNMT1_NVEC'  AIEKEEPAAQAYRLNNPGCTVFTDDCNTLLKLAMEGEATNSTGQKIPQRGEVELLCGGPP
'DNMT1_HSYM'  AIEFEPSAAQAYRLNNPGTIVFNADCNHILKLIMEGKETNESGQRLPRKGEVDLLCGGPP
'DNMT1_HVUL'  AIEFEPSAAQAYRLNNPSAIVFNQDCNNVLKQIMEGKEKDDLGQRLPRRGEVDLLCGGPP
'DNMT1_AQUE'  AVEIDEPAAQAFRLNNSQTTVFTDDCNILLSLVMEGAKTNSRGQLLPQKGDVELLCGGPP
'DNMT1_HSAP'  AIEMWDPAAQAFRLNNPGSTVFTEDCNILLKLVMAGETTNSRGQRLPQKGDVEMLCGGPP
'DNMT1_MMUS'  AIEMWDPAAQAFRLNNPGTTVFTEDCNVLLKLVMAGEVTNSLGQRLPQKGDVEMLCGGPP
'DNMT1_MLEI'  AIEFFSEAAQAYKLNNPQAEVFNEDCNAVLRMAMEGVLTNKLGQKIPQKGEVELLCGGPP
```

DNA_methylase (PF00145)

```
'DNMT1_ADIG'  CQGFSGMNRFSSRDYSQFKNSLVVSYLSYCDYYRPRFFILENVRNFVSFKRSMVLKLTLR
'DNMT1_NVEC'  CQGFSGMNRFNTREYSLFKNSLVVSYLSYCDFYRPRFFILENVRNFVSFKKSMVLKLTLR
'DNMT1_HSYM'  CQGFSGMNRFNHREYSMFKNSLVTSYLSYCDYFRPKFFILENVRNFVSFKKGMVLKLIMK
'DNMT1_HVUL'  CQGFSGMNRFNQREYSMFKNSLVTSYLSYCDYFRPKFFILENVRNFVSFKKSMVLKLTLS
'DNMT1_AQUE'  CQGFSGMNRFNSREYSQFKNSLVISYLSFCEYYRPRFFLLENVRNFVSFKKSMVLKLTLR
'DNMT1_HSAP'  CQGFSGMNRFNSRTYSKFKNSLVVSFLSYCDYYRPRFFLLENVRNFVSFKRSMVLKLTLR
'DNMT1_MMUS'  CQGFSGMNRFNSRTYSKFKNSLVVSFLSYCDYYRPRFFLLENVRNFVSYRRSMVLKLTLR
'DNMT1_MLEI'  CQGFSGMNRFNAREYSMFKNSLISSYLSYCEFYRPRYFLLENVRNFVSYKKNMVLKLCLS
```

DNA_methylase (PF00145)

```
'DNMT1_ADIG'  CLIKMGYQCTFGVLQAGCYGVPQTRRRAIIMAAAPGEVLPLYPEPTHCFSPRAIQLTVMV
'DNMT1_NVEC'  CLLRMGYQCTFGVLQAGCYGVPQTRRRAIIMAAAPGEELPLYPEPTHCFSPRTCQLTVMV
'DNMT1_HSYM'  CLLKMGYQCEFGVLQAGSYGVPQTRRRAILIAAAPGEILPKYPECQHVFAPKGLALSVTI
'DNMT1_HVUL'  CLVKMGYQCEFGVLQAGSYGVPQTRRRAIIIAAAPGEILPKFPEPQHVFASKALSLSVTI
'DNMT1_AQUE'  CLVKMGYQCTFGVLQAGQYGVPQTRRRAIILAAAPGEKLPHFPNPTHVFSPRACQLTVVV
'DNMT1_HSAP'  CLVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPLFPEPLHVFAPRACQLSVVV
'DNMT1_MMUS'  CLVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPLFPEPLHVFAPRACQLSVVV
'DNMT1_MLEI'  SLVNMGYQCTFGVLQAGCYGVAQTRRRAFLLAAAPGEVLPQFPEPRHVFSPKAMSLSVTI
                                DNA_methylase (PF00145)

'DNMT1_ADIG'  DDKKFESNITRLSSAPFRTITVRDSMSDLPEIRNGASNAEISYSGDSISHFQRQVRGSQY
'DNMT1_NVEC'  DEKKFTSNITRTASAPYRTITVRDAMSDLPEIRNGASAAESSYEGEAISHFQRQIRGNQY
'DNMT1_HSYM'  DDKKYQA-VKRLHTAPYRTITVRDAMSDLPEIKNGAKKPEIGYDSEPISHFQKMMRGKHV
'DNMT1_HVUL'  NSHKYQA-LKRLHSAPYRTTTVYDAMSDLPEIKNGANKMEIGYDTEPLTHFQKKIRGKHM
'DNMT1_AQUE'  NDIKYEGSI-RMDSAPYRTITVRDSMSDLPHIKNGSAVRSMNYNGEPHCHYQRLMRGNQH
'DNMT1_HSAP'  DDKKFVSNITRLSSGPFRTITVRDTMSDLPEVRNGASALEISYNGEPQSWFQRQLRGAQY
'DNMT1_MMUS'  DDKKFVSNITRLSSGPFRTITVRDTMSDLPEIQNGASNSEIPYNGEPLSWFQRQLRGSHY
'DNMT1_MLEI'  DNKQYSQHITRFDSAPLRNITVRDIMSDLPAITNGATKRELAYDCEPDSWFQRQIRGNS-
                                DNA_methylase (PF00145)

'DNMT1_ADIG'  QPVLRDHICKEMNPLVAVRMRYIPLAPGSDWRDLPNIEVQLPDGTKTKKLAYTHHDKKNG
'DNMT1_NVEC'  QPLLRDHICKEMSALVEARMRHIPLAPGSDWRDLPNKEIRLSDGTYSKKLQYTHHDKKNE
'DNMT1_HSYM'  QQVLRDHVCKDMSPLVEARMQYIPCKPGSDWRDLPNTVVKLRDGNYTKKLLYLHHDKKQG
'DNMT1_HVUL'  Q-VLRDHICKDMSPLVEARMSYIPCIPGSDWRDLPNKVVKLRDGNTTKLLLYLHKDKKQG
'DNMT1_AQUE'  QPVLYDHICKEMNPLVAARMRFIPIGPGSDWRDLPNKCIRLSDGTTAPKLQYTHHDKKNG
'DNMT1_HSAP'  QPILRDHICKDMSALVAARMRHIPLAPGSDWRDLPNIEVRLSDGTMARKLRYTHHDRKNG
'DNMT1_MMUS'  QPILRDHICKDMSPLVAARMRHIPLFPGSDWRDLPNIQVRLGDGVIAHKLQYTFHDVKNG
'DNMT1_MLEI'  -EVLTDHICKEMNPLVAIRMRHIPLTPGSDWRDLPNISVTLPDGKKTPKLVYTHNDLKNG
                                DNA_methylase (PF00145)

'DNMT1_ADIG'  RSSEGQLRGVCSCAE-DKPCDPADRQFNTLVPWCLPHTGNRHNHWAGLYGRLEWDGYFST
'DNMT1_NVEC'  KGSSQRLRGVCSCAE-GRPCEPADRQFNTLIPWCLPHTGNRHNHWAGLYGRLEWDGYFST
'DNMT1_HSYM'  PANNGDMRGVCACAT-GASCDSADRQFNTLIPWCLPHTSNRHNHWAGLYGRLEWDGFFST
'DNMT1_HVUL'  PGKNGENRGVCACAN-GKPCDPSDRQFNTLIPWCLPHTGNRHNNWAGLYGRLEWDGFFST
'DNMT1_AQUE'  RAKNKSLRGVCPCAT-GQPCDSSYRQYGTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFST
'DNMT1_HSAP'  RSSSGALRGVCSCVEAGKACDPAARQFNTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFST
'DNMT1_MMUS'  YSSTGALRGVCSCAE-GKACDPESRQFSTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFST
'DNMT1_MLEI'  P-----MKGVCSCAE-GKKCDPMDKQNMTLIPWCLPHTGNRHNNWAGLYGRVFWDGYFST
                                DNA_methylase (PF00145)

'DNMT1_ADIG'  TITNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPDTYRFYGSILDKHRQVGNAVPPPLA
'DNMT1_NVEC'  TITNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPDTYRFYGSILDKHRQVGNAVPPPLA
'DNMT1_HSYM'  TVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPDTFRFFGNILDKHRQIGNAVAPPMA
'DNMT1_HVUL'  TVTNPEPMGKQGRVLHPEQHRVVSVRECSRSQGFPDSFRFYGNILDKHRQIGNAVAPPMS
'DNMT1_AQUE'  TVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPDTFRFFGTILDKHRQVGNAVPPPLA
'DNMT1_HSAP'  TVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPDTYRLFGNILDKHRQVGNAVPPPLA
'DNMT1_MMUS'  TVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPDSYRFFGNILDRHRQVGNAVPPPLA
'DNMT1_MLEI'  TITNPEPMGKQGRVLHPEQHRLVSVRECARSQGFPDTYKFYGNLLDKHRQIGNAVPPPLA
                                DNA_methylase (PF00145)

'DNMT1_ADIG'  AAIGREIKKSL----------------------QATQQNA--------------------
'DNMT1_NVEC'  AAIGREIKKGL----------------------ELTQGVRK------KD---ESKMDITS
'DNMT1_HSYM'  AAIGQEIRKSIVQKKMKNNENETNRSDVIEESEKKREVKETVKQENKDGDIEDDGNSS-
'DNMT1_HVUL'  AAIGQEIRKSIIAKKLRN--------------EQLQSIKE-----------ENNHIATF
'DNMT1_AQUE'  KAIGLEIKRSV----------------------EKKDK----------------------
'DNMT1_HSAP'  KAIGLEIKLCML--------------------AKARESASA-----KI---KEEEAAKD
'DNMT1_MMUS'  KAIGLEIKLCLL--------------------SSARESASAA---VKA---KEEAATKD
'DNMT1_MLEI'  RAIGDEIAKCL---------------RVV---SQRKELKV-----------ESKLDVN-
              DNA_methylase (PF00145)
;
```

end;

6.4 DNMT3 nine sequence alignment with annotated domain

```
#NEXUS
begin taxa; dimensions ntax=9; taxlabels
'DNM3A_HSAP'
'DNM3A_MMUS'
'DNM3B_HSAP'
'DNM3B_MMUS'
'DNMT3_ADIG'
'DNMT3_NVEC'
'DNMT3_AQUE'
'DNMT3_HSYM'
'DNMT3_HVUL'
;end;

begin characters; dimensions nchar=1643;
format datatype=protein missing=? gap=- interleave=yes;
matrix
'DNM3A_HSAP'  MPAMPSSGPG----------------------------------------------------
'DNM3A_MMUS'  MP---SSGPG----------------------------------------------------
'DNM3B_HSAP'  MK------------------------------------------------------------
'DNM3B_MMUS'  MK------------------------------------------------------------
'DNMT3_ADIG'  M-------------------------------------------------------------
'DNMT3_NVEC'  LS---ATMAE----------------------------------------------------
'DNMT3_AQUE'  MA------------------------------------------------------------
'DNMT3_HSYM'  ML---SSAVDDCEQEKETYIQKTHLQ---QNIPSFYQTGDMTTDLNVTKTDVRKNEQEGI
'DNMT3_HVUL'  MN---KTSVKLVAKRK----RKSEVDSLFDNLSFIY--SEKSCSL--PKIDLINAKNDGI

'DNM3A_HSAP'  ---------------DTSSSA----------------------AEREEDRKDG-----EE
'DNM3A_MMUS'  ---------------DTSSSS----------------------LEREDDRKEG-----EE
'DNM3B_HSAP'  ------------------------------------------------GDTRHL-----NG
'DNM3B_MMUS'  ------------------------------------------------GDSRHL-----NE
'DNMT3_ADIG'  --------------------------------------------------------------
'DNMT3_NVEC'  -----VTHA-------SPPLC-----------------------------------------
'DNMT3_AQUE'  ---------------EGPANC-----------------------------------------
'DNMT3_HSYM'  RHFAEEVHSGNQREFKTAPKCTSDTEPSHDISPKAPIVNLPQMMEIQQEIRNVTEVLTND
'DNMT3_HVUL'  ESL--KTHGGRRWCLDSTLYCDRVIDHN---SQKFSMAM--------EDHSQVNQQSNKD

'DNM3A_HSAP'  QEEPR---GKEER---QEPS--------------------------TTARKVGRPGR----
'DNM3A_MMUS'  QEENR---GKEER---QEPS--------------------------ATARKVGRPGR----
'DNM3B_HSAP'  EEDAG---GREDSILVNG----------------------------ACSDQSSDS------
'DNM3B_MMUS'  EEGAS---GYEECIIVNG----------------------------NFSDQSSDTKD----
'DNMT3_ADIG'  -EATP-------------------------------------------TSEAIQAPKK----
'DNMT3_NVEC'  -QDTP-------------------------------------------FLENANEHDK----
'DNMT3_AQUE'  --SIP-------------------------------------------LYRKRKAPDHLA--
'DNMT3_HSYM'  EDTTPY----EDLDFSQDSKLDS----------NHISFS-NMLKEALNEQIVDPEKKAD-
'DNMT3_HVUL'  KEDMLFKCVKADKNILQNKNKKNKRKSEIDALFDHLSFIYSDKKSCLPKEDKQARKLPNK

'DNM3A_HSAP'  --------------KRKHPPVESG--DTPKDPAVISKSPS--------------------
'DNM3A_MMUS'  --------------KRKHPPVESS--DTPKDPAVTTKSQP--------------------
'DNM3B_HSAP'  -----------------PPILEAI-----RTPEIRGRR----------------------
'DNM3B_MMUS'  --------------APSPPVLEAICTEPVCTPETRGRR----------------------
'DNMT3_ADIG'  --------------------------------------------------------------
'DNMT3_NVEC'  --------------------------------------------------------------
'DNMT3_AQUE'  --------------------------------------------------------------
'DNMT3_HSYM'  --------------IKTDMIIQT---YMCDHPQNDGQLKNEK---HRAKRITMYKRKRTK
'DNMT3_HVUL'  ELVNDENDKEWSLYNKIAHSSEIIDHDTPRTPEIAEQLQINKIFNNDANNNILFEGNKVK

'DNM3A_HSAP'  ------------------------------------------------MAQDS--------
'DNM3A_MMUS'  ------------------------------------------------MAQDS--------
'DNM3B_HSAP'  --------------------------------------------------------------
'DNM3B_MMUS'  --------------------------------------------------------------
'DNMT3_ADIG'  --------------------------------------------------------------
'DNMT3_NVEC'  --------------------------------------------------------------
'DNMT3_AQUE'  --------------------------------------------------------------
'DNMT3_HSYM'  PNVNIDNYKFIIEKGCRKVISVK**RKKTFSTGDIVVGKFKHMGWWFGEIIEHTEKPLKGHH**
'DNMT3_HVUL'  SK-NKETLQTQLIGGSKLV----**KKKNFKVGDVVICYFKYFGWWFAKIIQHNKRAAEGNC**
                                                **PWWP (SSF63748)**
```

```
'DNM3A_HSAP' -----GASELL-----------------PNGDLEKRSEPQPE-----------------
'DNM3A_MMUS' -----GPSDLL-----------------PNGDLEKRSEPQPE-----------------
'DNM3B_HSAP' ----------------------------SSSRLSKR-EVSSL-----------------
'DNM3B_MMUS' ----------------------------SSSRLSKR-EVSSL-----------------
'DNMT3_ADIG' ----------------------------KP-KFKRRLEFSAS-----------------
'DNMT3_NVEC' ----------------------------KPLKAKRRLNFGSI-----------------
'DNMT3_AQUE' ----------------------------SPYTAAKHLELSPS-----------------
'DNMT3_HSYM' FVFWYGDHHILQIPTKYLDVYTNFSKVFSKPKLAKRKYEVATMEFLEE-----IYKKLNI
'DNMT3_HVUL' WVSSFGDHKILMV---------------PLHSLRHHYDFSSK-FNQTKMKKPLYKK---
                                          PWWP (SSF63748)

'DNM3A_HSAP' ------------------------------------------------------EGSPA--
'DNM3A_MMUS' ------------------------------------------------------EGSPA--
'DNM3B_HSAP' ------------------------------------------------------------
'DNM3B_MMUS' ------------------------------------------------------------
'DNMT3_ADIG' ------------------------------------------------------------
'DNMT3_NVEC' ------------------------------------------------------------
'DNMT3_AQUE' ----------------------------------PAPD--------STLSPNQPVE-
'DNMT3_HSYM' IDIHKVNEDEINHSRFHI----------LLAWGLASFPVVEKTSLKNNLCVMFQNNPTEK
'DNMT3_HVUL' -AVKEFLSELAEQSNFKIMFGQKGYLNELLEWGLNGFAAHV---IKD------KNKPCQG
             PWWP (SSF63748)

'DNM3A_HSAP' -GGQKGGAPAEGEGAAETLPEASRAVENGCCTPKEGRGAPA----------------EAG
'DNM3A_MMUS' -AGQKGGAPAEGEG-TETPPEASRAVENGCCVTKEGRGASA----------------GEG
'DNM3B_HSAP' -LSYTQDLTGDGDG------------EDG-----DGSDTPVMP--------------KLF
'DNM3B_MMUS' -LNYTQDMTGDGDRDD--------EVDDG-----NGSDI-LMP--------------KLT
'DNMT3_ADIG' --DY------------------------------------------------------
'DNMT3_NVEC' --DYV-----------------------------------------------------
'DNMT3_AQUE' --NYVD------------------PIEELKSLEEDMSNFRRFSG----------------
'DNMT3_HSYM' KGNYTHMESNDKIE-ERKKTETD-NPEKERVDTNEISNMKVLGVPSNSVHPTRNVHDKEL
'DNMT3_HVUL' NLDFTNTSLNQSIDFVIESSDTS-NTESSKLTSKDNSESIDFLLKNYSLSPNLEDSDNYF

'DNM3A_HSAP' KEQKETN-IESMKMEGSRGRLRGGLGW------------------ESSLRQ---------
'DNM3A_MMUS' KEQKQTN-IESMKMEGSRGRLRGGLGW------------------ESSLRQ---------
'DNM3B_HSAP' RETR-------TRSESPAVRTRNNNSV------------------SSRERH---------
'DNM3B_MMUS' RETKDTR----TRSESPAVRTRHSNGT------------------SSLERQ---------
'DNMT3_ADIG' KSKKKIR----VKSDVFI--------------------------SVT------------
'DNMT3_NVEC' KRKKKRK----ISKQNLVGKKSKAS--------------------DVTAEP---------
'DNMT3_AQUE' RSKKQTNFFCPIAPASFHRRRSNSAT------------------NAAVRA---------
'DNMT3_HSYM' SKKKQEKGTPEIKQDTLIKTSSLIGSTQEKTIVQTTFKVTREEKENISGKADFGHSSEIN
'DNMT3_HVUL' SSDQHSS-INSFKNNEVDSKKSLC---------------------KVKNRS---------

'DNM3A_HSAP' ----------------------------------------RPMPRLTFQAGDP-------
'DNM3A_MMUS' ----------------------------------------RPMPRLTFQAGDP-------
'DNM3B_HSAP' ----------------------------------------RPSPRSTRGRQGR-------
'DNM3B_MMUS' ----------------------------------------RASPRITRGRQGR-------
'DNMT3_ADIG' ------------------------------------------------------------
'DNMT3_NVEC' -------------------------------------TVCTDSDSGFDTQDN-------
'DNMT3_AQUE' -------------------------------------LKFPSSTSPVDSGGP-------
'DNMT3_HSYM' QNLTLKQNFFSDPPTRRSFAKAIDHFANRKKPVLEYTILNRTDGTNVVECKENI------
'DNMT3_HVUL' ----------------------------KINVLKSSIINHTAQSTNLECQNTVIKNSEC

'DNM3A_HSAP' ----YYISKRKRDEWLARWKREAEKKAKVIAGMN----------------------AVE
'DNM3A_MMUS' ----YYISKRKRDEWLARWKREAEKKAKVIAVMN----------------------AVE
'DNM3B_HSAP' ----NHV-----DESPVEFPATRSLRRRATASAGTPWPSPPSSYL------------TID
'DNM3B_MMUS' ----HHV-----QEYPVEFPATRSRRRRASSSASTPWSSPA----------------SVD
'DNMT3_ADIG' ------------------------------------------------------SVE
'DNMT3_NVEC' ------------------QTPEKGKEQRE-------------------------------
'DNMT3_AQUE' ------------------AGIAASEEKKE---------------------------SID
'DNMT3_HSYM' ---LSNL-NSNEKNLSSDNSVNNPKDPKEITVTSYADVVTPRIITNASVDRASMSDKSIE
'DNMT3_HVUL' ENIKKNIENYNAQDWFL-NKVDHVRKDCLEKMTHTIQRSPDKILS------------NIE

'DNM3A_HSAP' ENQGPGESQK----------------------------------VEEASPPAVQQPTDPA
'DNM3A_MMUS' ENQASGESQK----------------------------------VEEASPPAVQQPTDPA
'DNM3B_HSAP' ----------------------------------------LTDDTEDTHGTPQSSS
'DNM3B_MMUS' ----------------------------------------FMEEV-----TPKSVS
'DNMT3_ADIG' DNH-------------------------------------LATSSD----------
'DNMT3_NVEC' ----------------------------------------LDKNNSHTPKKSKNVD
```

```
'DNMT3_AQUE'  NGE--GED-------------------------------------------------KVVA
'DNMT3_HSYM'  HEDSCSDDSGIIDISLSKSDLSYESSSLSHSGSNADVSELY-SSVDESISSPKRKTKDGI
'DNMT3_HVUL'  HNDVTCKD------FLCNSK---------HPSFHSNSTPLQTSCLKENSNQTMRLYKDGI

'DNM3A_HSAP'  SPTVAT------------------------TPEPVGSDAGDK--------NATKAGDDE
'DNM3A_MMUS'  SPTVAT------------------------TPEPVGGDAGDK--------NATKAADDE
'DNM3B_HSAP'  TPYARL----------------------AQDSQQGGMESPQV--------EADSGDGDS
'DNM3B_MMUS'  TPSVDL----------------------SQDGDQEGMDTTQV--------DAESRDGDS
'DNMT3_ADIG'  ---------------------------------------------------DHLSTSD
'DNMT3_NVEC'  QMA---------------------------------KRKHK----------GSAVSVD
'DNMT3_AQUE'  AMEIA-------------------------TGEEVEINKPQK----------PQQQTVP
'DNMT3_HSYM'  TLDFDLLFTHASKRQCFISAEKKFTSPKRKQSESDGKLKRKQRVK-----VSAHSSPSLD
'DNMT3_HVUL'  KLDVTLLNEKTPKRACSIKNQTLSSVSPNNVSCASITIGKKQKCVIENKYSSKKNDKTCD

'DNM3A_HSAP'  PEYEDG-------------------------RGFGIGELVWGKLRGFSWWPGRIVSWWM
'DNM3A_MMUS'  PEYEDG-------------------------RGFGIGELVWGKLRGFSWWPGRIVSWWM
'DNM3B_HSAP'  SEYQDG-------------------------KEFGIGDLVWGKIKGFSWWPAMVVSWKA
'DNM3B_MMUS'  TEYQDD-------------------------KEFGIGDLVWGKIKGFSWWPAMVVSWKA
'DNMT3_ADIG'  DENG---------------------------ATFVHGQLIWGRLKGYDWWPGLIVSHLE
'DNMT3_NVEC'  KEVGEG-------------------------TFFPGVLVWSKLKGYDWWPGRVVTYME
'DNMT3_AQUE'  QEI----------------------------KKLPIGSLIWGKLPGYEWWPGCIISFDK
'DNMT3_HSYM'  VEIKERMKAFGSWTKSIVTSKVLEHSSDDSSKRKIRVGDLVLGKLKGFDWWFGKVITFRK
'DNMT3_HVUL'  REIGSD---F--------------------FQIKIGDLVLGKLKGYDWWFGMVVSHRV
                                          PWWP (SSF63748)

'DNM3A_HSAP'  TG---------------------RSRAAEGTRWVMWFGDGKFSVVCVEKLMPL-SSFCSA
'DNM3A_MMUS'  TG---------------------RSRAAEGTRWVMWFGDGKFSVVCVEKLMPL-SSFCSA
'DNM3B_HSAP'  TS---------------------KRQAMSGMRWVQWFGDGKFSEVSADKLVAL-GLFSQH
'DNM3B_MMUS'  TS---------------------KRQAMPGMRWVQWFGDGKFSEISADKLVAL-GLFSQH
'DNMT3_ADIG'  AQ---------------------KAPPAPSNHWIKWFGDNKLSLLPFQCLRPF-SKFKES
'DNMT3_NVEC'  AG---------------------RPPPGPGNHWVKWFGDNKFSQVPQDTVLPF-AEFKSN
'DNMT3_AQUE'  KKLIEVVKDEDDDDEDTEEDEEEKEEDGGESVWVKWYGDNQLSQISFKKIFPFGSNFLEF
'DNMT3_HSYM'  AK---------------------ESPPSEGCSWVYWYGDHKKSEMLTDRLESL-SLFPLR
'DNMT3_HVUL'  IR---------------------QRPAANDCHWIRWYGDHKVSEVHLQNIELL-TSFSNR
                                          PWWP (SSF63748)

'DNM3A_HSAP'  FHQATYNKQPMYRKAIYEVLQVASSRAGK-------------LFPVCH------------
'DNM3A_MMUS'  FHQATYNKQPMYRKAIYEVLQVASSRAGK-------------LFPACH------------
'DNM3B_HSAP'  FNLATFNKLVSYRKAMYHALEKARVRAGK-------------TFP---------------
'DNM3B_MMUS'  FNLATFNKLVSYRKAMYHTLEKARVRAGK-------------TFS---------------
'DNMT3_ADIG'  LIPSKMRGI--YKRAVFDSLEVAVKRSGK-------------VFAKCPI-------KQTT
'DNMT3_NVEC'  FLVTKMKGL--YKKAVLDALELAASRSNQ-------------SFTRADLHL-----KLSP
'DNMT3_AQUE'  FSPNKLRGL--YKKAVVNSLKEAARRCSKDIGCTLDERKPSWIFHSSQFQVFLDERKALN
'DNMT3_HSYM'  FAPKKMKGL--YRKAVEEMLTEAAFRCSK--------------------------------
'DNMT3_HVUL'  YLPSKMQGL--YLRAIKELLEEAARRCCK--------------------------------
                                          PWWP (SSF63748)

'DNM3A_HSAP'  ---------------DSDE----SDTAKAVEVQNKP------------------------
'DNM3A_MMUS'  ---------------DSDE----SDSGKAVEVQNKQ------------------------
'DNM3B_HSAP'  ----------------------SSPGDSLEDQLKP------------------------
'DNM3B_MMUS'  ----------------------SSPGESLEDQLKP------------------------
'DNMT3_ADIG'  QTKKK--TATRRAKVGAEQIAGLHEDQEPLGERERQEM----------------------
'DNMT3_NVEC'  KVSKK---QRRKKKLLSKDTETPAEPDEPLSEEEKCDA----------------------
'DNMT3_AQUE'  NLTQEEAIEATPEKIEQGEKSPVYSPDSSFPNLQQEESSIFTQIESDQESDSSPANGVKE
'DNMT3_HSYM'  -------------------------EDLPLDDDKRLQC----------------------
'DNMT3_HVUL'  --------------------------TDLPDDEKRLGV----------------------
                                          PWWP (SSF63748)

'DNM3A_HSAP'  ----------------------MIEWALGGFQPSGPKGLEPPEEEKNPYK---------
'DNM3A_MMUS'  ----------------------MIEWALGGFQPSGPKGLEPPEEEKNPYK---------
'DNM3B_HSAP'  ----------------------MLEWAHGGFKPTGIEGLKPNNTQPVVNK----SKVRR
'DNM3B_MMUS'  ----------------------MLEWAHGGFKPTGIEGLKPNKKQPVVNK----SKVRR
'DNMT3_ADIG'  ----------------------MVQWAMNGFKPNGPKGFAPPEDDMIS-------DTSV
'DNMT3_NVEC'  ----------------------MVNWALTGFQPQGASGFAPTLEETI------------
'DNMT3_AQUE'  KGGGRGAKKTKTGQFLLSLMEKDLLQWALTGFLPTGPDGFRPRPEHLIDIKMFPFGESPA
'DNMT3_HSYM'  ----------------------LVDWALTGFKPMGVNALYSVVEDTE------------
'DNMT3_HVUL'  ----------------------LVDWALNGFQPYGFDKL--SIEETLTLE---------
```

PWWP (SSF63748)

```
'DNM3A_HSAP'  ------E-----VYTDMWVEPEAAAYAPPPPAKKPRKSTAEKPKVKEIIDERTRERLVYE
'DNM3A_MMUS'  ------E-----VYTDMWVEPEAAAYAPPPPAKKPRKSTTEKPKVKEIIDERTRERLVYE
'DNM3B_HSAP'  AGSRKLESRKYENKTRRRTADDSATSDYCPAPKRLKTNCYNNGKDRGDEDQS-REQMASD
'DNM3B_MMUS'  SDSRNLEPRRRENKSRRRTTNDSAASES-PPPKRLKTNSYG-GKDRGEDEES-RERMASE
'DNMT3_ADIG'  SVTSFM--------DVISSDDSESTADH--------KSRFGKPVNVPNLAEN-IKALFEE
'DNMT3_NVEC'  ----FL--------PPVSSDSDEEEADQTVNGSESTKGRFGKPVKEPNLVQD-IKALFDE
'DNMT3_AQUE'  PGNQFII-----KRNELPMERGRAVIDH---------DYDKKPENKLVLLQQRNKAQFEL
'DNMT3_HSYM'  -GSHSLE-----QKSNIESSRELESGEG-----ALLKDRVGQILQSPILSDG-IKALFDR
'DNMT3_HVUL'  -DDAFI----------------QEGSET-----DSLTHNEGEPLPSPILSED-VKALFDK
```

(PF17980)

```
'DNM3A_HSAP'  VRQKCRNIEDICISCGSLNVTLEHPLFVGGMCQNCKNCFLECAYQYDDDGYQSYCTICCG
'DNM3A_MMUS'  VRQKCRNIEDICISCGSLNVTLEHPLFIGGMCQNCKNCFLECAYQYDDDGYQSYCTICCG
'DNM3B_HSAP'  VANNKSSLEDGCLSCGRKNPVSFHPLFEGGLCQTCRDRFLELFYMYDDDGYQSYCTVCCE
'DNM3B_MMUS'  VTNNKGNLEDRCLSCGKKNPVSFHPLFEGGLCQSCRDRFLELFYMYDEDGYQSYCTVCCE
'DNMT3_ADIG'  VSKGKRDLECICLACGDLRVCAKHPLFKGGLCKECKTSFLGNIYLYDEDGSQMYCTICGD
'DNMT3_NVEC'  VVDGKKGIEDICLACGDCKVYAQHPLFEGGLCKECKQSFLECTYLFDEDGYQMYCTICGD
'DNMT3_AQUE'  IKLGQLAIENICIACGSVNISAQHPLFYGGLCKHCKETFMECAYMFDEDGSQMYCTICSG
'DNMT3_HSYM'  VHKGNYDIENLCLGCGDTKVITQHPLFEGGLCKECKDSFIEHAYLYDDDGSQMYCTICSD
'DNMT3_HVUL'  AANGNMDLNKICLGCGDLKCIMEHPLFIGGLCKECKESFMETAYLYDEDGSQMYCSICSD
```

ADD (PF17980) overlap **PHD-zf**

```
'DNM3A_HSAP'  GREVLMCGNNNCCRCFCVECVDLLVGPGAAQAAIKEDPWNCYMCGHKGTYGLLRRREDWP
'DNM3A_MMUS'  GREVLMCGNNNCCRCFCVECVDLLVGPGAAQAAIKEDPWNCYMCGHKGTYGLLRRREDWP
'DNM3B_HSAP'  GRELLLCSNTSCCRCFCVECLEVLVGTGTAAEAKLQEPWSCYMCLPQRCHGVLRRRKDWN
'DNM3B_MMUS'  GRELLLCSNTSCCRCFCVECLEVLVGAGTAEDAKLQEPWSCYMCLPQRCHGVLRRRKDWN
'DNMT3_ADIG'  GKEVFMCDNDGCFRSYCNLCIGMLCGVNAVRKIAACETWVCFMCSSKTQ-GLISPRQDWS
'DNMT3_NVEC'  GQEVFMCDNEGCFRSYCGPCLEMLAGRGTVREIASREKWICYMCSGKGD-RLIHRRKDWQ
'DNMT3_AQUE'  GSQVFMCDTPNCSKVYCNPCIEKLCGPEERMKVQDATEWQCYMCSGEMN-GLLRKREDWD
'DNMT3_HSYM'  GEQVILCDKVGCCRSYCPVCIDMLCGAGYAEKLSKEDSWECFMCSGKPM-RILRRRDDWQ
'DNMT3_HVUL'  GREIILCDVPGCYRSYCSVCLDMFCGVGYSRTVSAGLDWHCFMCTGEKV-RLIQRRDDWQ
```

PHD-zf (SSF57903)

```
'DNM3A_HSAP'  SRLQMFFANNHDQE-FDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYI
'DNM3A_MMUS'  SRLQMFFANNHDQE-FDPPKVYPPVPAEKRKPIRVLSLFDGIATGLLVLKDLGIQVDRYI
'DNM3B_HSAP'  VRLQAFFTSDTGLE-YEAPKLYPAIPAARRRPIRVLSLFDGIATGYLVLKELGIKVGKYV
'DNM3B_MMUS'  MRLQDFFTTDPDLEEFEPPKLYPAIPAAKRRPIRVLSLFDGIATGYLVLKELGIKVEKYI
'DNMT3_ADIG'  ARLQEHFMNDKEQE-FKPPHTFSAIPPEERRPLRVLSLFDGIASGLQALKELGIEIEILYC
'DNMT3_NVEC'  SKLHELFLSDREKE-YDTPIVYPVVAPEDRKPIRVLALFDGIATGLQALNELGIVSDKYY
'DNMT3_AQUE'  KRLHDLFVTDIETE-FEPPKFFPPIPAEMRKPMRVLGLFDGIGTGLLVLKELGINVEVYI
'DNMT3_HSYM'  KKLKDLFQDEMGSEDYGTMTFYDPIPIEERKAIRVLALFDGIGTGFHVLKELDFDVELYI
'DNMT3_HVUL'  QRLKNVFSVNNEG--YPLPFFYDPVPFEDRPSLRVLSLFDGLSTGYLALSELGLDILSYQ
```

DNA_methyltransferase (SSF53335)

```
'DNM3A_HSAP'  ASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARK
'DNM3A_MMUS'  ASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPCNDLSIVNPARK
'DNM3B_HSAP'  ASEVCEESIAVGTVKHEGNIKYVNDVRNITKKNIEEWGPFDLVIGGSPCNDLSNVNPARK
'DNM3B_MMUS'  ASEVCAESIAVGTVKHEGQIKYVNDVRKITKKNIEEWGPFDLVIGGSPCNDLSNVNPARK
'DNMT3_ADIG'  ASEIDENAVQVAKVQHGRKIHHIGDIQKISRDQIQDLGPFDFVFGGSPCNDLSIANPIRR
'DNMT3_NVEC'  SAEIDEQAIQVTKVNHGDRITHLGDIKDLTESQIRELGPFDLVIGGSPCQDLSIANPARR
'DNMT3_AQUE'  ASEIDPDAIKVSRIKHPE-IIHVGAIEQVTEKEVRSWGPFDLVFGGSPCNDLSIVNPARK
'DNMT3_HSYM'  ASEIDENAKTVTLVHYGDKIRHVGDVCKIREEDIKEWGPFDLVIGGSPCNDLSIANPLRK
'DNMT3_HVUL'  ASEIDPLAIKVSKVHHSMRVEQIGDVQKITKQDIENWGPFDLVIGGSPCDELSIANPFRK
```

DNA_methyltransferase (SSF53335)

```
'DNM3A_HSAP'  GLYEGTGRLFFEFYRLLHDARP-KEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMI
'DNM3A_MMUS'  GLYEGTGRLFFEFYRLLHDARP-KEGDDRPFFWLFENVVAMGVSDKRDISRFLESNPVMI
'DNM3B_HSAP'  GLYEGTGRLFFEFYHLLNYSRP-KEGDDRPFFWMFENVVAMKVGDKRDISRFLECNPVMI
'DNM3B_MMUS'  GLYEGTGRLFFEFYHLLNYTRP-KEGDNRPFFWMFENVVAMKVNDKKDISRFLACNPVMI
'DNMT3_ADIG'  GIYEGTGRLFFDFFRLLEYARP-KSELERPFFWLFENVVGMRAEDKAVISRFLECNPVVV
'DNMT3_NVEC'  GIFEGSGRLFFEFFRLLMHAKPSRTCPSRPFFWLFENVVGMRAEDKKTISRFLQSNPVVV
'DNMT3_AQUE'  GIYDGTGKLFFEFFRILSYAKP-QPQDERPFFWLYENVVSMRAPDKKIISRFLQCNPVVV
'DNMT3_HSYM'  GIYDGTGRLFFEFYRILQYAKP-HPIESRPFFWLFENVVGMQHVDREVINRFLKCHPVVV
'DNMT3_HVUL'  GIY-GTGQLFFEFYRILEYCKP-QPLTARPFFWLFENVVGMRYTDRAVISRFLQCHPVII
```

DNA_methyltransferase (SSF53335)

```
'DNM3A_HSAP'  DAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLE--HGRIAKFSKVRTITTRSN
'DNM3A_MMUS'  DAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECLE--HGRIAKFSKVRTITTRSN
'DNM3B_HSAP'  DAIKVSAAHRARYFWGNLPGMNRPVIASKNDKLELQDCLE--YNRIAKLKKVQTITTKSN
'DNM3B_MMUS'  DAIKVSAAHRARYFWGNLPGMNRPVMASKNDKLELQDCLE--FSRTAKLKKVQTITTKSN
'DNMT3_ADIG'  DAKEISPAHRARYFWGNLPGMNRPTIPLPGDKLCLQDCLEPNCGRQAKFTKIQTLTTNAN
'DNMT3_NVEC'  DAKEVSPAHRPRYFWGNLPGMNRPAIPLPGDRLTLQECLEPNCGRKARFTKVQTITTNAN
'DNMT3_AQUE'  DARDISAAHRARFFWGNLPGMNRPAVPLPGDRLTLQDCLEPNCFRHAQFTKLRTITTKMN
'DNMT3_HSYM'  SAKDVSAQHRTRFFWGNLPGMNRSLRPIPTDKVLLQDCLEPGCGRVAKVDKVRCITTMRH
'DNMT3_HVUL'  NAKEISAQQRTRYFWGNLPGMNRAMCPLPSDRLKLQQCLEKDCGRVAKFDKVRCITTSQN
```

DNA_methyltransferase (SSF53335)

```
'DNM3A_HSAP'  SIKQGKDQHFPVFM------------NEKEDILWCTEMERVFGFPVHYTDVSNMSRLAR
'DNM3A_MMUS'  SIKQGKDQHFPVFM------------NEKEDILWCTEMERVFGFPVHYTDVSNMSRLAR
'DNM3B_HSAP'  SIKQGKNQLFPVVM------------NGKEDVLWCTELERIFGFPVHYTDVSNMGRGAR
'DNM3B_MMUS'  SIRQGKNQLFPVVM------------NGKDDVLWCTELERIFGFPAHYTDVSNMGRGAR
'DNMT3_ADIG'  SMLQTKKALLPVQ--------YVASDGHEKEDILWITEMERLFGLPSHYTDVSNMGRSQR
'DNMT3_NVEC'  SLTQTKKNILPVA--------VNDDGGQEREDILWCTEMERLFGFPSHYTDVNNMGRTQR
'DNMT3_AQUE'  SIKQTKRAIFPVRVFSGGMFDITGEGEEEEGDVLWCTEMERLFGFPDHYTDVANLGRSGR
'DNMT3_HSYM'  SLRQTKKSILPVK--------LTRSVNDEIEDGLWLTEIERIFGFPDHYTDVCNMGQRQR
'DNMT3_HVUL'  SLKQTKAAMLPVK--------WTISRNSQLDDGLWLTEIERIFGFPDHYTDVANMGQRDR
```

DNA_methyltransferase (SSF53335)

```
'DNM3A_HSAP'  QRLLGRSWSVPVIRHLFAPLKEYFACV---------------------------------
'DNM3A_MMUS'  QRLLGRSWSVPVIRHLFAPLKEYFACV---------------------------------
'DNM3B_HSAP'  QKLLGRSWSVPVIRHLFAPLKDYFACE---------------------------------
'DNM3B_MMUS'  QKLLGRSWSVPVIRHLFAPLKDYFACE---------------------------------
'DNMT3_ADIG'  QRLLGKSWSVPVIRHLLSPLKDYYKCET---------------------------------
'DNMT3_NVEC'  QRLLGNAWSVPVVRHLLSPLKDYFKCT----------------------------------
'DNMT3_AQUE'  QKLLGKAWSVPVIRHLLSPLKDYFKSDTTGVPSSLKQTHIPITPVSITTAPPAPVLPVIA
'DNMT3_HSYM'  QKLLGKAWSVPVIRHLFTPLKDYFRTNSS---------------------------------
'DNMT3_HVUL'  QKLLGKSWSVPVLRHLFAPLRNYFRSKVY--------------------------------
```

DNA_methyltransferase (SSF53335)

```
'DNM3A_HSAP'  ------------------------
'DNM3A_MMUS'  ------------------------
'DNM3B_HSAP'  ------------------------
'DNM3B_MMUS'  ------------------------
'DNMT3_ADIG'  ------------------------
'DNMT3_NVEC'  ------------------------
'DNMT3_AQUE'  PPPKYLSPEPETVIVIMKTDEMD
'DNMT3_HSYM'  ---------------PEKTADV-
'DNMT3_HVUL'  ---------------IDNYNSML

;end;
```

6.5 Protein Sorting and Localization Analysis

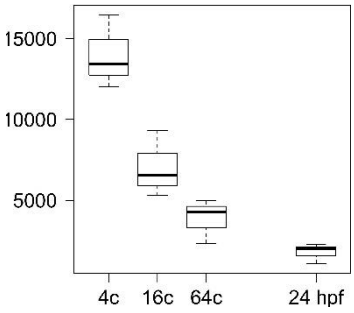
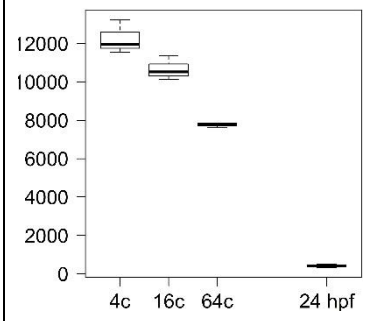
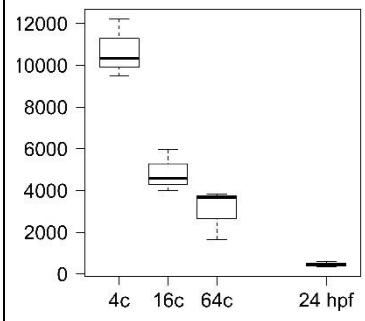
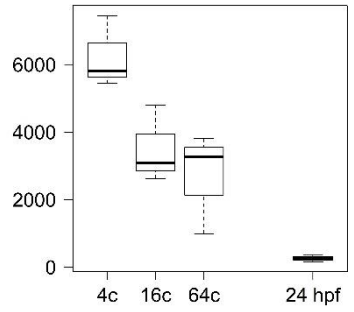
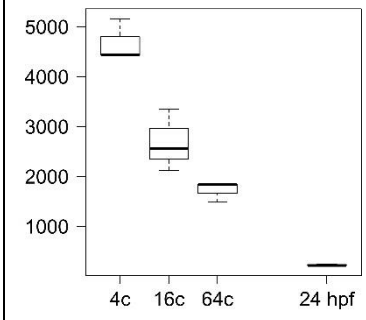
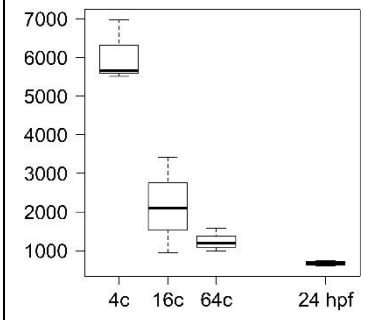
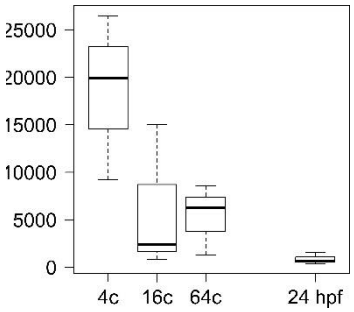
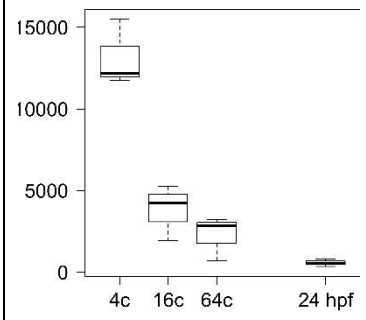
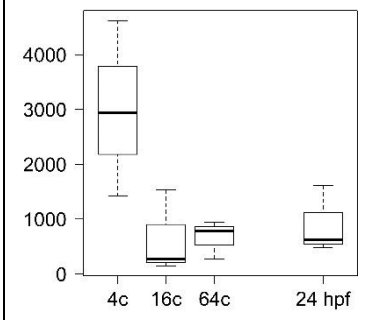
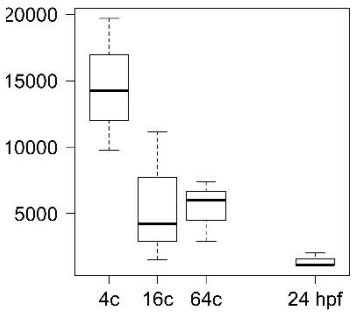
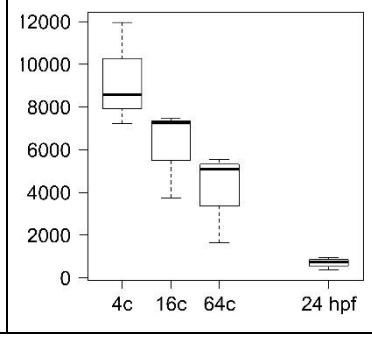
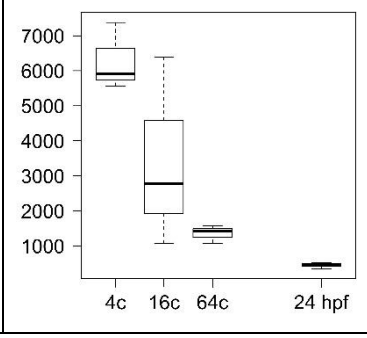
Enzyme	cNLS_Type	cNLS_Pos	cNLS_score	SignalP (Sec)	SignalP_Other	TargetP_SP	TargetP_mTP	TargetP_Other
DNMT1	Bi	38	8.3	0.001	0.999	0.000	0.000	1.000
DNMT3	Bi	34	7.5	0.001	0.999	0.000	0.001	0.999
TET	Bi	148	6.8	0.001	0.999	0.000	0.044	0.956
METTL4	Bi	14	7.3	0.002	0.998	0.001	0.000	0.999
N6AMT1	None	-	0	0.002	0.998	0.000	0.000	1.000
ALKBH1	Mono	21	8.5	0.001	0.999	0.000	0.001	0.999
ALKBH4	Mono	118	9.5	0.005	0.995	0.000	0.000	1.000

Enzyme	1st PSORT	1st PSORT_score	2nd PSORT	2nd PSORT_score
DNMT1	Cytoplasmic	17.5	Cytoplasmic-Nuclear	14.67
DNMT3	Nuclear	16.5	Cytoplasmic-Nuclear	15.5
TET	Nuclear	18	Plasma membrane	8
METTL4	Nuclear	19	Cytoplasmic-Nuclear	16.5
N6AMT1	Cytoskeleton	15	Cytoplasmic	7
ALKBH1	Cytoplasmic	23	Cytoplasmic-Nuclear	17
ALKBH4	Extracellular	15	Cytoplasmic-Nuclear	8

Enzyme	Nuclear	Cytoplasmic	Cytoplasmic-Nuclear	Cytoplasmic-Mitochondria	Mitochondria	Peroxisome	Plasma Membrane	Cytoskeleton	Extracellular
DNMT1	10.5	17.5	14.67	10.33		3			
DNMT3	16.5	13.5	15.5				1	1	
TET	18		1		5		8		
METTL4	19	12	16.5				1		
N6AMT1	3	7				2	5	15	
ALKBH1	9	23	17						
ALKBH4	6.5	6.5	8		3		1		15

6.6 EBseq-HMM plot of degradation expression path

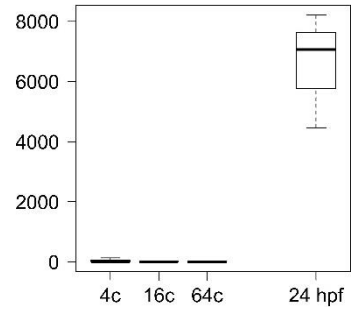
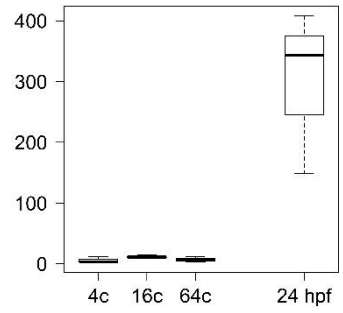
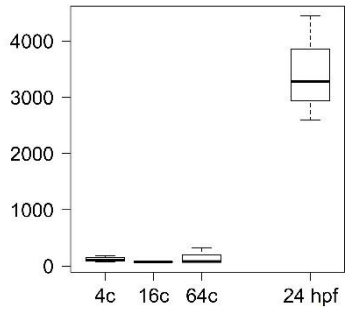
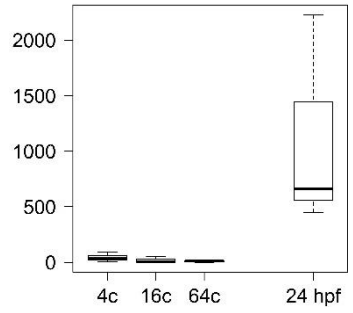
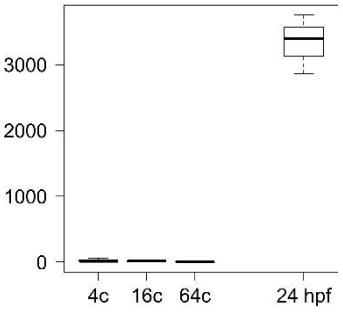
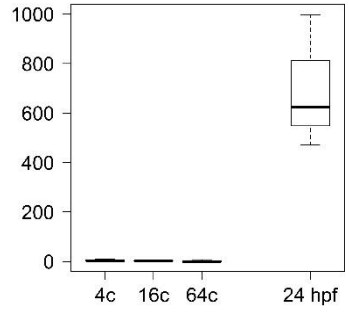
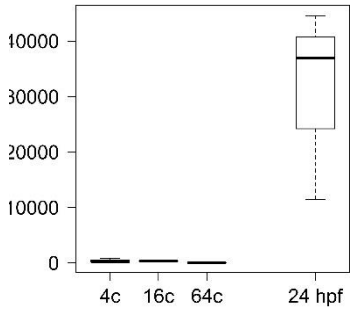
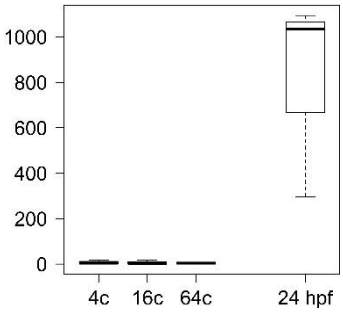
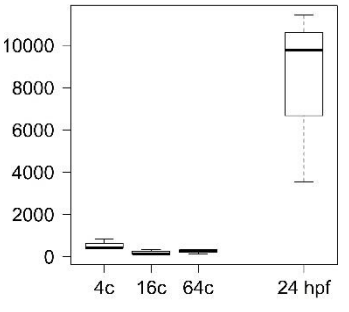
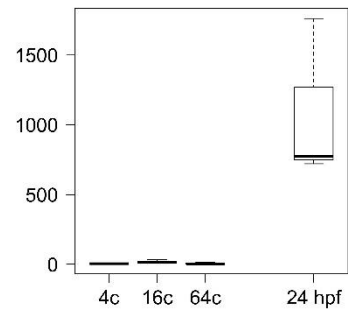
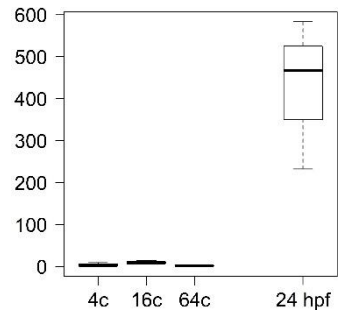
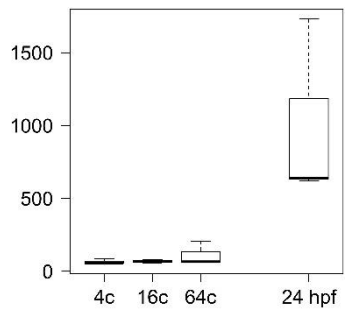
a. Maternal early degradation expression path (461 transcripts)
b. Maternal late degradation expression path (2848 transcripts)

Tyrosine kinase Tec Lyk	***Selenocysteine synthase***	***Denn6db***
		
E3 Ubiquitin ligase Traf6	***Apoptosis induc. factor 3***	***Choline Kinase Chckb***
		
S phase ass. prot. *scaper*	***Glutamate dehydrogenase***	***Cell death protein ced6***
		
Tgf-beta signalling Vps39	***Uncharac. Ubx domain Prot***	***Protein phosphatase 1k***
		

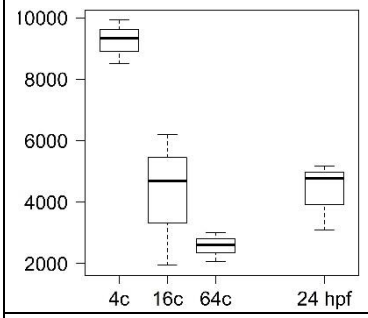
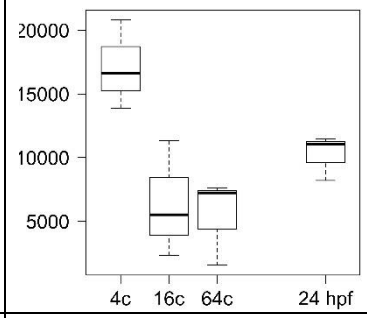
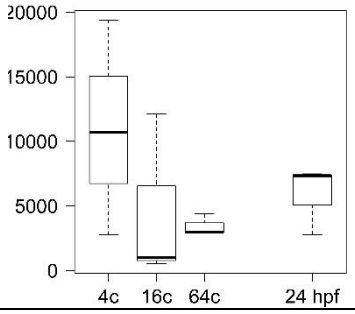
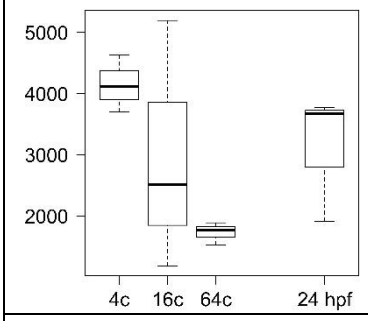
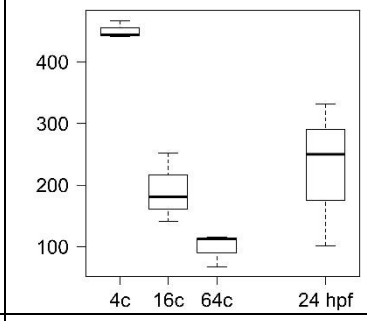
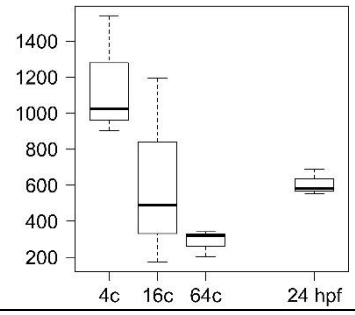
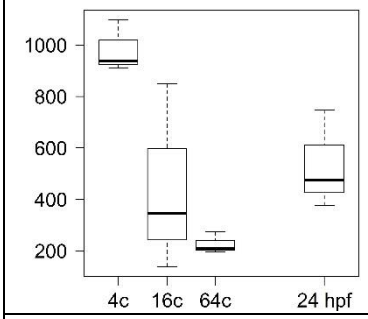
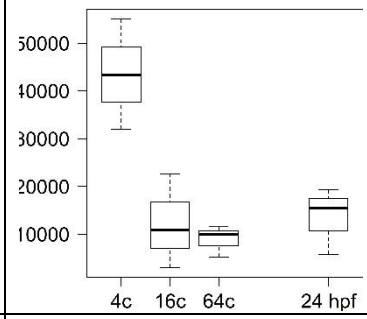
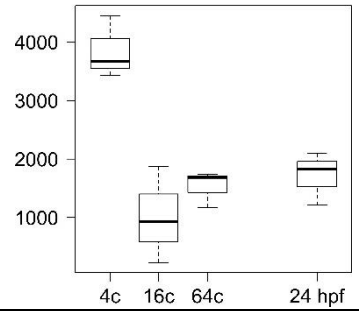
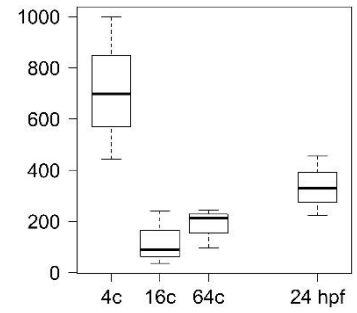
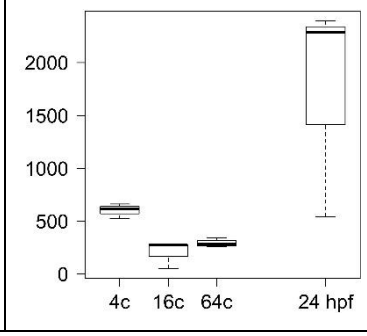
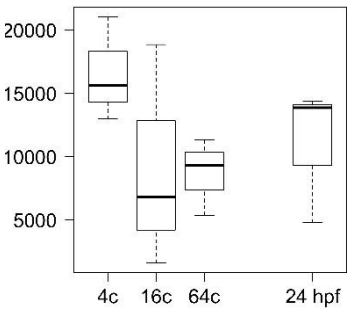
6.7 EBseq-HMM plot of high-preplanula expression path

a. High preplanula A (near zero to high at preplanula, 404 transcripts)

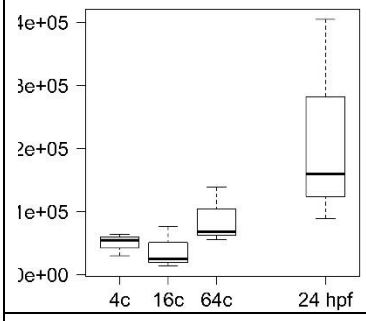
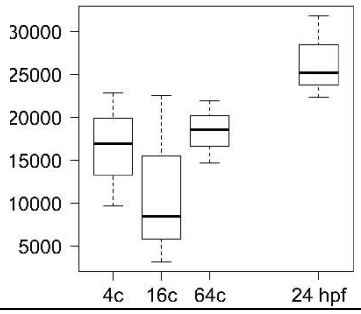
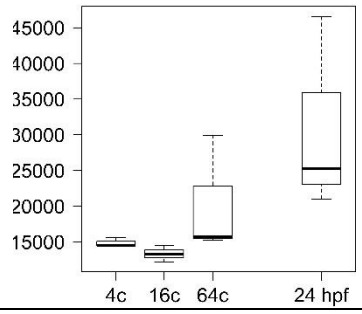
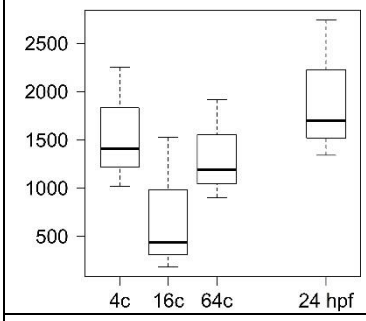
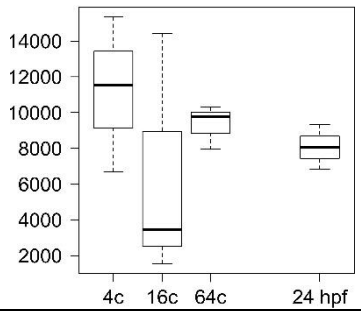
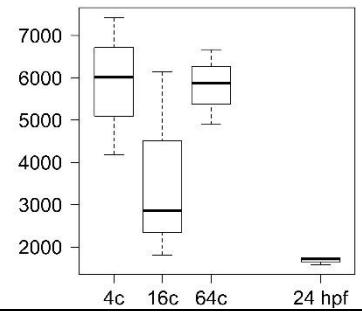
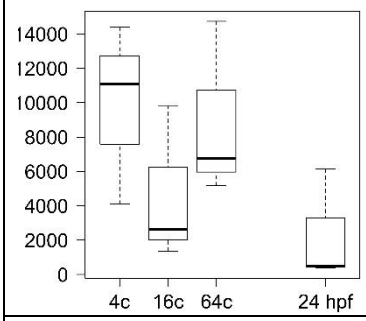
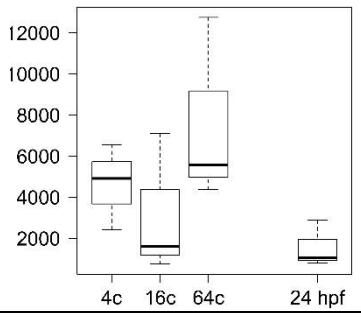
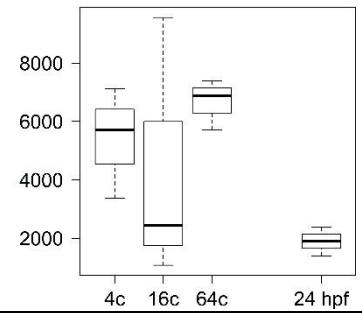
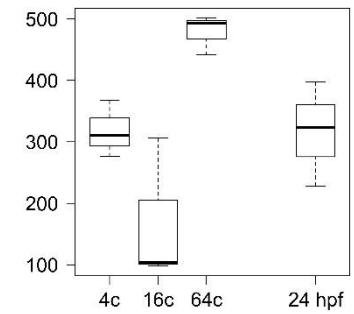
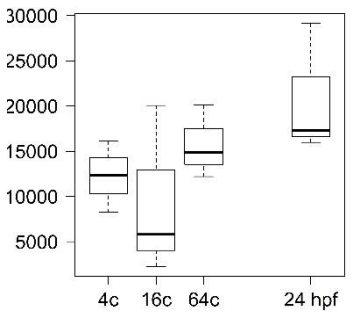
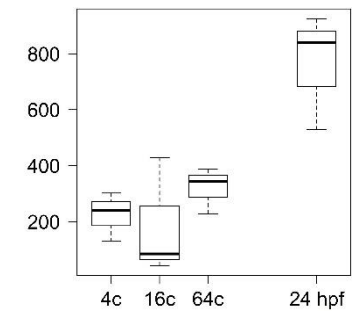
b. High preplanula B (substantial early even higher at preplanula, 71 transcripts)

Ectonucleotide pyrophosphatase ***Enpp5***	***Reverse transcriptase transposon***	***Axin interactor dorsalization Aida***
		
Tetratricopeptide ***Ttc28***	*Arachidonate 5-lipoxygenase* ***Alox5***	*bHLH transcription factor p48*
		
Collagen alpha-1(IV) Col4a1	***Uncharacterized Aminotransferase***	*F-box only protein 11* ***Fbxo11***
		
Cysteine rich protein 1	***Amyloid-like protein Aplp2***	*Homeobox protein six1a****b***
		

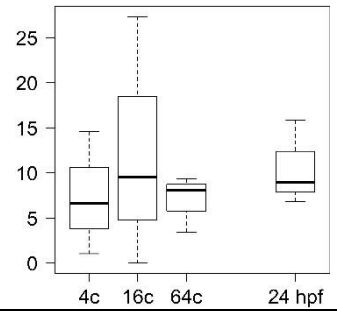
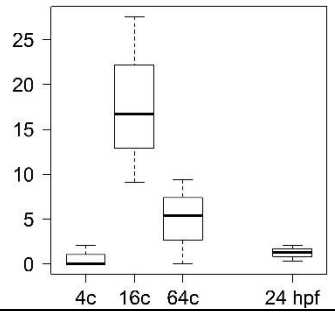
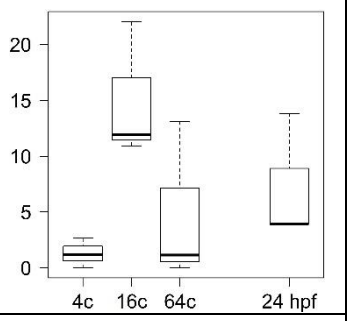
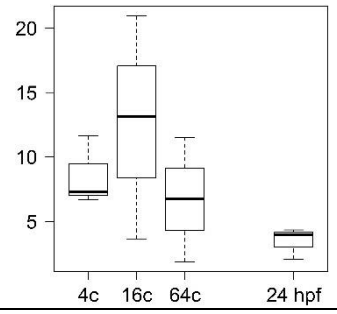
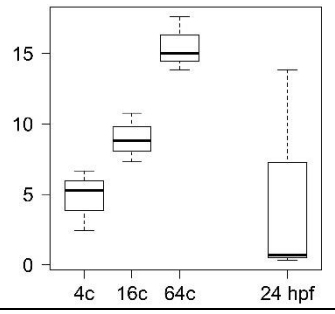
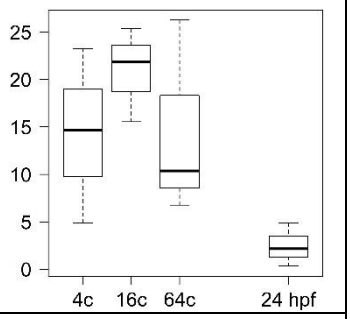
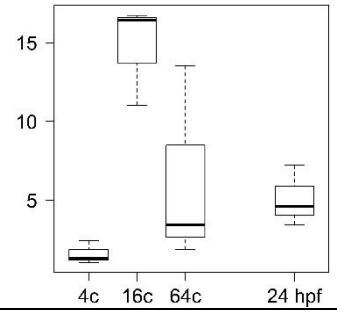
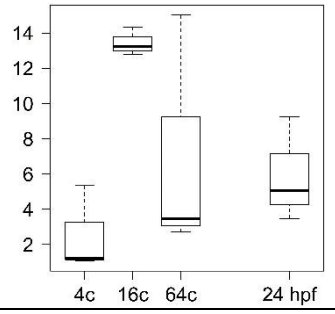
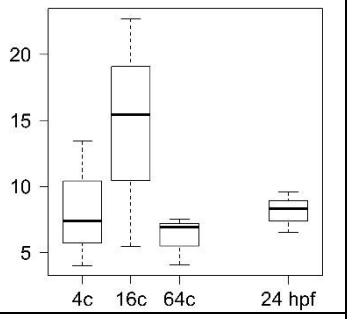
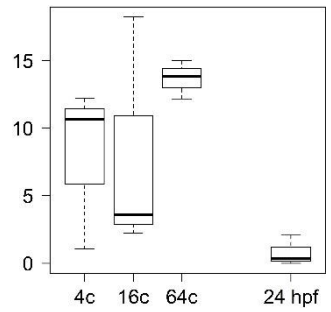
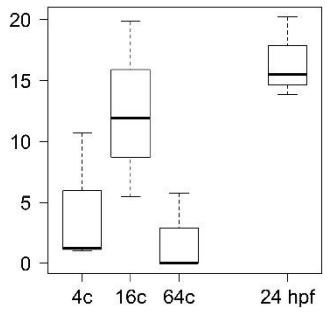
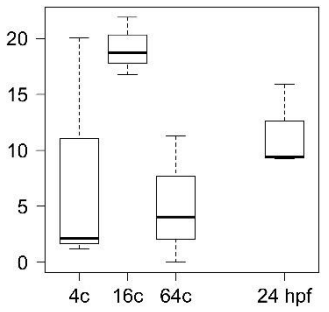
6.8 EBseq-HMM boxplot of D-DE-U (Early degraded but late upregulated) expression path. (87 transcripts)

Calcium-calmodulin depnd. kinase Camk2g	***Hemicentin-Coadhesin 1***	***Aconitase***
		
Amino acid transporter 1	***Meiosis inhibitory protein 1***	***Ras related protein Rab36***
		
*E3 ubiquitin-protein ligase M**archf**1*	*Alpha-actinin **Actn***	*BICD family-like cargo adapter 1 **Bicdl1***
		
Gastrula zinc finger protein XlCGF57	*Golgin subfamily A 1 **Golga1***	*Coatomer subunit beta **Copb1***
		

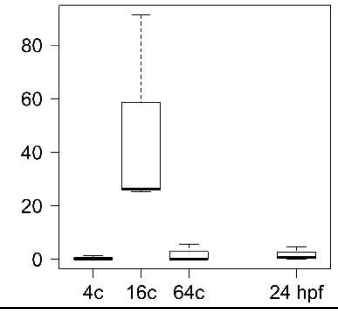
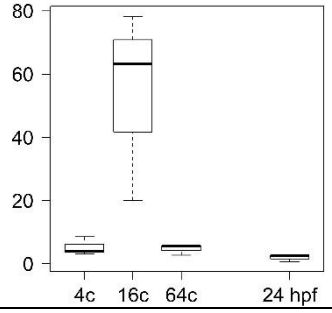
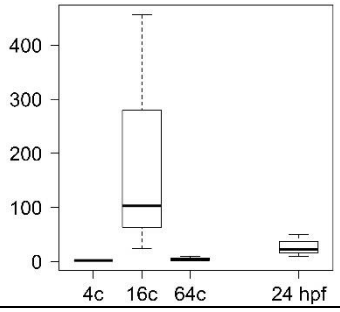
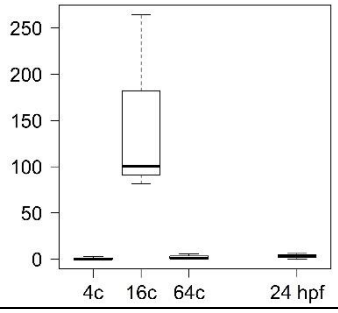
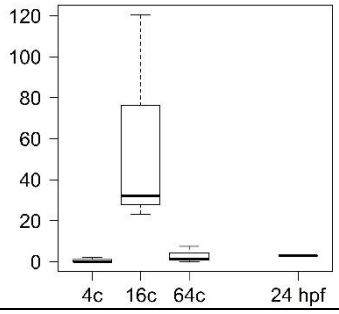
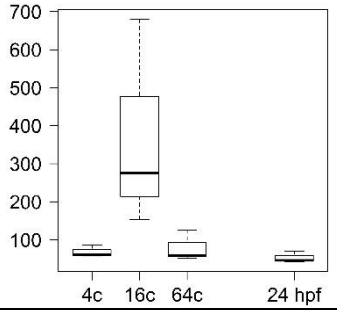
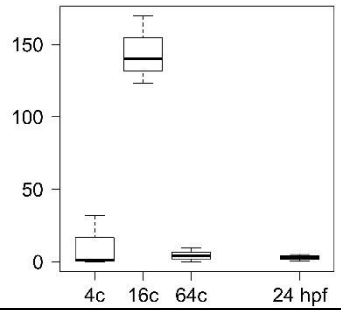
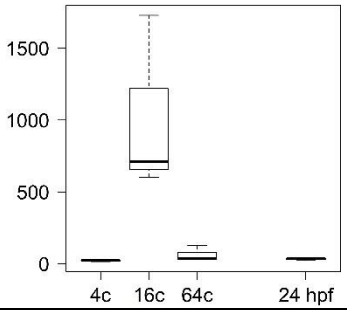
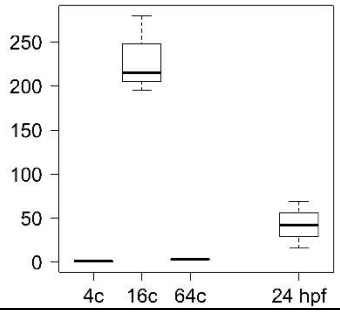
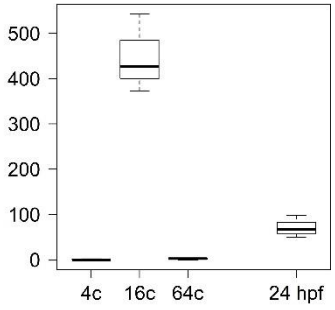
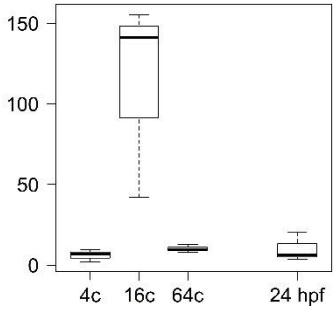
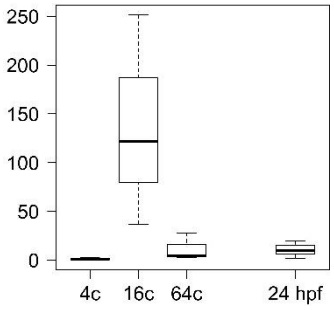
6.9 EBseq-HMM plots of upregulated at 64c expression path (EUD-U-EUD) (240 transcripts)

Elongation factor 1-alpha 1 Eef1a	*Dolichyl glycosyltransferase Rpn1*	*MFS transporter ceroid-lipofuscinosis neuronal 7*
		
Unchar. TNF receptor-associated factor 3 Traf3	*Islet cell autoantigen 1 Ica1*	*Forkhead box protein J3 FoxJ3*
		
Keratin-associated protein 5-1-like	*Rest corepressor 2 CoREST*	*Neural precursor cell protein Nedd1*
		
WD40 repeat-containing protein	*Band 4.1-like protein 3 EPB41L3*	*Protein SON*
		

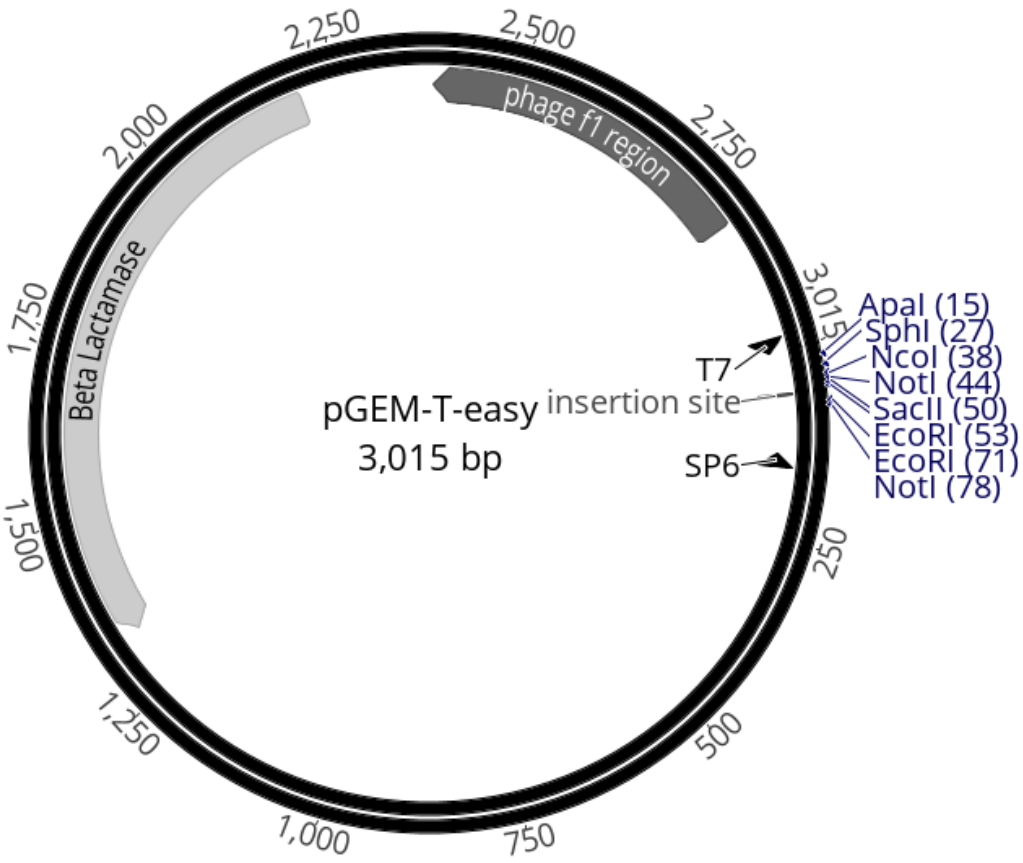
6.10 EBseq-HMM boxplot of various expression paths but upon visual inspection considered as lineage and context dependent expression. (124 transcripts)

Pyroglutamylated RFamide peptide receptor	*Zinc metalloproteinase nas-13/15*	*Uncharacterized DUF4243 domain*
		
Uncharacterized contain YfaL domain	*Endosomal potassium channel Tmem175*	*Pancreatic triacylglycerol lipase Pnlip*
		
Macrophage MHC class I receptor 2	*Uncharacterized Roundabout homolog 1*	*Unchar. Cysteine-rich domain of tyrosine kinase-like orphan receptors*
		
Dual specificity mitogen-activated protein kinase kinase 1 Mapkk1	*4-beta-N-acetyl-glucosaminyl transferase A Mgat4a*	*Protein mab-21-like 3*
		

6.11 EBseqHMM boxplot for transient expression path (143 transcripts)

Carboxypeptidase B-like	*Unchar. Coagulation factor 5/8*	*Leucine-rich repeat G-protein coupled receptor 4 Lgr4*
		
tRNA uracil-5-methyltransferase homolog A-like Trmt2A	*Uncharacterized contain domain Thyroglobulin type I repeats*	*Uncharacterized Calcium binding protein, EF-hand superfamily*
		
Lutropinchoriogonadotropic hormone receptor Lhcgr	*Parkin coregulated gene protein homolog Pacrg*	*Potassium-sodium activated cyclic nucleotide-gated channel 1-like Hcn1*
		
Ankyrin-1 Ank1	*Uncharacterized Apolipoprotein A1 A4 E*	*Chitinase-1 Chitotriosidase1 Chit1*
		

6.12 pGEM-T-easy



pGEM-T-easy annotated sequence

```
1    TAATACGACTCA CTATAGGGCGAA TTGGGCCCGACG TCGCATGCTCCC GGCCGCCATGGC
          T7                    ApaI          SphI          NcoI
61   GGCCGCGGGAAT TCGATATCACTA GTGAATTC GCGGCCGCCTGCAGGT CGACCATATGGG
           EcoRI  Insertion site  EcoRI    NotI
121  AGAGCTCCCAAC GCGTTGGATGCA TAGCTTGAGTAT TCTATAGTGTCA CCTAAATAGCTT
                                          SP6
181  GGCGTAATCATG GTCATAGCTGTT TCCTGTGTGAAA TTGTTATCCGCT CACAATTCCACA

241  CAACATACGAGC CGGAAGCATAAA GTGTAAAGCCTG GGGTGCCTAATG AGTGAGCTAACT

301  CACATTAATTGC GTTGCGCTCACT GCCCGCTTTCCA GTCGGGAAACCT GTCGTGCCAGCT

361  GCATTAATGAAT CGGCCAACGCGC GGGGAGAGGCGG TTTGCGTATTGG GCGCTCTTCCGC

421  TTCCTCGCTCAC TGACTCGCTGCG CTCGGTCGTTCG GCTGCGGCGAGC GGTATCAGCTCA

481  CTCAAAGGCGGT AATACGGTTATC CACAGAATCAGG GGATAACGCAGG AAAGAACATGTG

541  AGCAAAAGGCCA GCAAAAGGCCAG GAACCGTAAAAA GGCCGCGTTGCT GGCGTTTTTCCA

601  TAGGCTCCGCCC CCCTGACGAGCA TCACAAAAATCG ACGCTCAAGTCA GAGGTGGCGAAA

661  CCCGACAGGACT ATAAAGATACCA GGCGTTTCCCCC TGGAAGCTCCCT CGTGCGCTCTCC

721  TGTTCCGACCCT GCCGCTTACCGG ATACCTGTCCGC CTTTCTCCCTTC GGGAAGCGTGGC

781  GCTTTCTCATAG CTCACGCTGTAG GTATCTCAGTTC GGTGTAGGTCGT TCGCTCCAAGCT

841  GGGCTGTGTGCA CGAACCCCCCGT TCAGCCCGACCG CTGCGCCTTATC CGGTAACTATCG

901  TCTTGAGTCCAA CCCGGTAAGACA CGACTTATCGCC ACTGGCAGCAGC CACTGGTAACAG

961  GATTAGCAGAGC GAGGTATGTAGG CGGTGCTACAGA GTTCTTGAAGTG GTGGCCTAACTA

1021 CGGCTACACTAG AAGAACAGTATT TGGTATCTGCGC TCTGCTGAAGCC AGTTACCTTCGG

1081 AAAAAGAGTTGG TAGCTCTTGATC CGGCAAACAAAC CACCGCTGGTAG CGGTGGTTTTTT
```

1141 TGTTTGCAAGCA GCAGATTACGCG CAGAAAAAAAGG ATCTCAAGAAGA TCCTTTGATCTT

1201 TTCTACGGGGTC TGACGCTCAGTG GAACGAAAACTC ACGTTAAGGGAT TTTGGTCATGAG

1261 ATTATCAAAAAG GATCTTCACCTA GATCCTTTTAAA TTAAAAATGAAG TTTTAAATCAAT

1321 CTAAAGTATATA TGAGTAAACTTG GTCTGACAGTTA CCAATGCTTAAT CAGTGAGGCACC

1381 TATCTCAGCGAT CTGTCTATTTCG TTCATCCATAGT TGCCTGACTCCC CGTCGTGTAGAT

1441 AACTACGATACG GGAGGGCTTACC ATCTGGCCCCAG TGCTGCAATGAT ACCGCGAGACCC

1501 ACGCTCACCGGC TCCAGATTTATC AGCAATAAACCA GCCAGCCGGAAG GGCCGAGCGCAG

1561 AAGTGGTCCTGC AACTTTATCCGC CTCCATCCAGTC TATTAATTGTTG CCGGGAAGCTAG

1621 AGTAAGTAGTTC GCCAGTTAATAG TTTGCGCAACGT TGTTGCCATTGC TACAGGCATCGT

1681 GGTGTCACGCTC GTCGTTTGGTAT GGCTTCATTCAG CTCCGGTTCCCA ACGATCAAGGCG

1741 AGTTACATGATC CCCCATGTTGTG CAAAAAAGCGGT TAGCTCCTTCGG TCCTCCGATCGT

1801 TGTCAGAAGTAA GTTGGCCGCAGT GTTATCACTCAT GGTTATGGCAGC ACTGCATAATTC

1861 TCTTACTGTCAT GCCATCCGTAAG ATGCTTTTCTGT GACTGGTGAGTA CTCAACCAAGTC

1921 ATTCTGAGAATA GTGTATGCGGCG ACCGAGTTGCTC TTGCCCGGCGTC AATACGGGATAA

1981 TACCGCGCCACA TAGCAGAACTTT AAAAGTGCTCAT CATTGGAAAACG TTCTTCGGGGCG

2041 AAAACTCTCAAG GATCTTACCGCT GTTGAGATCCAG TTCGATGTAACC CACTCGTGCACC

2101 CAACTGATCTTC AGCATCTTTTAC TTTCACCAGCGT TTCTGGGTGAGC AAAAACAGGAAG

2161 GCAAAATGCCGC AAAAAAGGGAAT AAGGGCGACACG GAAATGTTGAAT ACTCATACTCTT

2221 CCTTTTTCAATA TTATTGAAGCAT TTATCAGGGTTA TTGTCTCATGAG CGGATACATATT

2281 TGAATGTATTTA GAAAAATAAACA AATAGGGGTTCC GCGCACATTTCC CCGAAAAGTGCC

2341 ACCTGATGCGGT GTGAAATACCGC ACAGATGCGTAA GGAGAAAATACC GCATCAGGAAAT

2401 TGTAAGCGTTAA TATTTTGTTAAA ATTCGCGTTAAA TTTTTGTTAAAT CAGCTCATTTTT

2461 TAACCAATAGGC CGAAATCGGCAA AATCCCTTATAA ATCAAAAGAATA GACCGAGATAGG

2521 GTTGAGTGTTGT TCCAGTTTGGAA CAAGAGTCCACT ATTAAAGAACGT GGACTCCAACGT

2581 CAAAGGGCGAAA AACCGTCTATCA GGGCGATGGCCC ACTACGTGAACC ATCACCCTAATC

2641 AAGTTTTTTGGG GTCGAGGTGCCG TAAAGCACTAAA TCGGAACCCTAA AGGGAGCCCCCG

2701 ATTTAGAGCTTG ACGGGGAAAGCC GGCGAACGTGGC GAGAAAGGAAGG GAAGAAAGCGAA

2761 AGGAGCGGGCGC TAGGGCGCTGGC AAGTGTAGCGGT CACGCTGCGCGT AACCACCACACC

2821 CGCCGCGCTTAA TGCGCCGCTACA GGGCGCGTCCAT TCGCCATTCAGG CTGCGCAACTGT

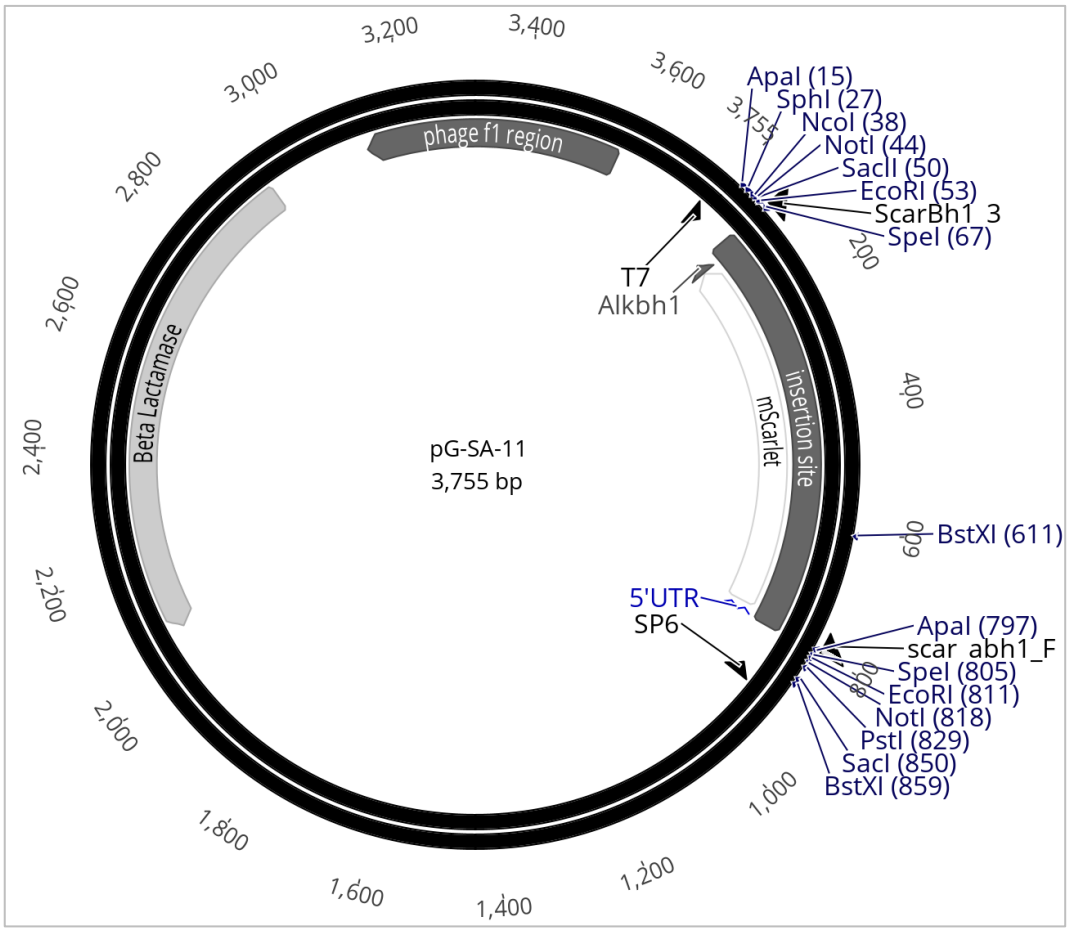
2881 TGGGAAGGGCGA TCGGTGCGGGCC TCTTCGCTATTA CGCCAGCTGGCG AAAGGGGGATGT

2941 GCTGCAAGGCGA TTAAGTTGGGTA ACGCCAGGGTTT TCCCAGTCACGA CGTTGTAAAACG

3001 ACGGCCAGTGAA TTG

6.13 Sequence of the cassettes inserted into pGEM-T-easy.

6.13.1 Map and sequence of shRNA confirmation plasmid



pG-SA-11 (pGEM-T-easy, mScarlett-Alkbh1-target-clone11, the cassette is inverted)

1 **TAATACGACTCA CTATAGGG**CGAA TTGGGCCCGACG TCGCATGCTCCC GGCCGCCATGGC

T7 *Apa*I *Sph*I *Nco*I

61 GGCCGCGGGAAT TCGATT***AGTGAC TAGTGGACATGA GCC***C**TTAGTATA GTTCATCCATTC**

*Eco*RI ***Alkbh1-target*** *mScarlette*

121 **CTCCAGTACTGT GCCTGCCTTCAC TTCGCTCATATT GTTCGACAACTG TATAATCTTCGT**

mScarlette

181 **TGTGTGATGTGA TATCTAATTTCC GATCAACATTAT AAGCACCAGGCA TTTGAACGGGTT**

mScarlette

241 **TCTTTGCCTTAT ATGTCGTTTTAA AATCAGCTAGAT AACGGCCACCGT CCTTTAATCGTA**

mScarlette

301 **AAGCCATTTTTA TATCTCCCTTCA AGACGCCGTCTT CGGGGTATAAAC GCTCAGTCGATG**

mScarlette

361 **CCTCCCAACCCA TTGTCTTCTTCT GCATTACAGGGC CATCAGGCGGAA AATTTGTTCCTC**

mScarlette

421 **GTAATTTAACTT TGTAAATCAAAG TACCATCTTCCA GACTGGTATCCT GGGTTACTGTAA**

mScarlette

481 **CAGCTCCTCCGT CTTCAAAGTTCA TAACTCTTTCCC ATTTGAATCCCT CCGGAAAACTTT**

mScarlette

541 **GTTTATAATAGT CCGGTATATCAG CTGGGTGCTTGA TGAAAGCTCTGG AACCATACATAA**

mScarlette

601 **ATTGAGGAGATA AAATATCCCACG AGAATGGCAGTG GACCTCCTTTTG TAACTTTTAACT**

mScarlette

661 **TAGCTGTTTGTG TTCCTTCGTATG GTCTACCTTCAC CTTCTCCTTCAA TTTCAAATTCAT**

mScarlette

721 **GTCCATTCATGG AACCTTCCATAT GAACTTTAAATC TCATAAACTCTT TTATAACTGCTT**

mScarlette

781 **CACCTTTAGATA CCATT*TTTTCAG ATCC***GGGCCCAA AATCACTAGT GAATTC GCGGCCGC

mScarlette ***short-5'UTR*** *Apa*I *Spe*I *Eco*RI *Not*I

841 CTGCAGGTCGAC CATATGGGAGAG CTCCCAACGCGT TGGATGCATAGC TTGAGT**ATTCTA**

SP6

901 **TAGTGTCACCTA AAT**AGCTTGGCG TAATCATGGTCA TAGCTGTTTCCT GTGTGAAATTGT

SP6

961 TATCCGCTCACA ATTCCACACAAC ATACGAGCCGGA AGCATAAAGTGT AAAGCCTGGGGT

1021 GCCTAATGAGTG AGCTAACTCACA TTAATTGCGTTG CGCTCACTGCCC GCTTTCCAGTCG

1081 GGAAACCTGTCG TGCCAGCTGCAT TAATGAATCGGC CAACGCGCGGGG AGAGGCGGTTTG

1141 CGTATTGGGCGC TCTTCCGCTTCC TCGCTCACTGAC TCGCTGCGCTCG GTCGTTCGGCTG

1201 CGGCGAGCGGTA TCAGCTCACTCA AAGGCGGTAATA CGGTTATCCACA GAATCAGGGGAT

1261 AACGCAGGAAAG AACATGTGAGCA AAAGGCCAGCAA AAGGCCAGGAAC CGTAAAAAGGCC

1321 GCGTTGCTGGCG TTTTTCCATAGG CTCCGCCCCCCT GACGAGCATCAC AAAAATCGACGC

1381 TCAAGTCAGAGG TGGCGAAACCCG ACAGGACTATAA AGATACCAGGCG TTTCCCCCTGGA

1441 AGCTCCCTCGTG CGCTCTCCTGTT CCGACCCTGCCG CTTACCGGATAC CTGTCCGCCTTT

1501 CTCCCTTCGGGA AGCGTGGCGCTT TCTCATAGCTCA CGCTGTAGGTAT CTCAGTTCGGTG

1561 TAGGTCGTTCGC TCCAAGCTGGGC TGTGTGCACGAA CCCCCCGTTCAG CCCGACCGCTGC

1621 GCCTTATCCGGT AACTATCGTCTT GAGTCCAACCCG GTAAGACACGAC TTATCGCCACTG

1681 GCAGCAGCCACT GGTAACAGGATT AGCAGAGCGAGG TATGTAGGCGGT GCTACAGAGTTC

1741 TTGAAGTGGTGG CCTAACTACGGC TACACTAGAAGA ACAGTATTTGGT ATCTGCGCTCTG

1801 CTGAAGCCAGTT ACCTTCGGAAAA AGAGTTGGTAGC TCTTGATCCGGC AAACAAACCACC

1861 GCTGGTAGCGGT GGTTTTTTTGTT TGCAAGCAGCAG ATTACGCGCAGA AAAAAAGGATCT

1921 CAAGAAGATCCT TTGATCTTTTCT ACGGGGTCTGAC GCTCAGTGGAAC GAAAACTCACGT

1981 TAAGGGATTTTG GTCATGAGATTA TCAAAAAGGATC TTCACCTAGATC CTTTTAAATTAA

2041 AAATGAAGTTTT AAATCAATCTAA AGTATATATGAG TAAACTTGGTCT GACAGTTACCAA

2101 TGCTTAATCAGT GAGGCACCTATC TCAGCGATCTGT CTATTTCGTTCA TCCATAGTTGCC

2161 TGACTCCCCGTC GTGTAGATAACT ACGATACGGGAG GGCTTACCATCT GGCCCCAGTGCT

2221 GCAATGATACCG CGAGACCCACGC TCACCGGCTCCA GATTTATCAGCA ATAAACCAGCCA

2281 GCCGGAAGGGCC GAGCGCAGAAGT GGTCCTGCAACT TTATCCGCCTCC ATCCAGTCTATT

2341 AATTGTTGCCGG GAAGCTAGAGTA AGTAGTTCGCCA GTTAATAGTTTG CGCAACGTTGTT

2401 GCCATTGCTACA GGCATCGTGGTG TCACGCTCGTCG TTTGGTATGGCT TCATTCAGCTCC

2461 GGTTCCCAACGA TCAAGGCGAGTT ACATGATCCCCC ATGTTGTGCAAA AAAGCGGTTAGC

2521 TCCTTCGGTCCT CCGATCGTTGTC AGAAGTAAGTTG GCCGCAGTGTTA TCACTCATGGTT

2581 ATGGCAGCACTG CATAATTCTCTT ACTGTCATGCCA TCCGTAAGATGC TTTTCTGTGACT

2641 GGTGAGTACTCA ACCAAGTCATTC TGAGAATAGTGT ATGCGGCGACCG AGTTGCTCTTGC

2701 CCGGCGTCAATA CGGGATAATACC GCGCCACATAGC AGAACTTTAAAA GTGCTCATCATT

2761 GGAAAACGTTCT TCGGGGCGAAAA CTCTCAAGGATC TTACCGCTGTTG AGATCCAGTTCG

2821 ATGTAACCCACT CGTGCACCCAAC TGATCTTCAGCA TCTTTTACTTTC ACCAGCGTTTCT

2881 GGGTGAGCAAAA ACAGGAAGGCAA AATGCCGCAAAA AAGGGAATAAGG GCGACACGGAAA

2941 TGTTGAATACTC ATACTCTTCCTT TTTCAATATTAT TGAAGCATTTAT CAGGGTTATTGT

3001 CTCATGAGCGGA TACATATTTGAA TGTATTTAGAAA AATAAACAAATA GGGGTTCCGCGC

3061 ACATTTCCCCGA AAAGTGCCACCT GATGCGGTGTGA AATACCGCACAG ATGCGTAAGGAG

3121 AAAATACCGCAT CAGGAAATTGTA AGCGTTAATATT TTGTTAAAATTC GCGTTAAATTTT

3181 TGTTAAATCAGC TCATTTTTTAAC CAATAGGCCGAA ATCGGCAAAATC CCTTATAAATCA

3241 AAAGAATAGACC GAGATAGGGTTG AGTGTTGTTCCA GTTTGGAACAAG AGTCCACTATTA

3301 AAGAACGTGGAC TCCAACGTCAAA GGGCGAAAAACC GTCTATCAGGGC GATGGCCCACTA

3361 CGTGAACCATCA CCCTAATCAAGT TTTTTGGGGTCG AGGTGCCGTAAA GCACTAAATCGG

3421 AACCCTAAAGGG AGCCCCCGATTT AGAGCTTGACGG GGAAAGCCGGCG AACGTGGCGAGA

3481 AAGGAAGGGAAG AAAGCGAAAGGA GCGGGCGCTAGG GCGCTGGCAAGT GTAGCGGTCACG

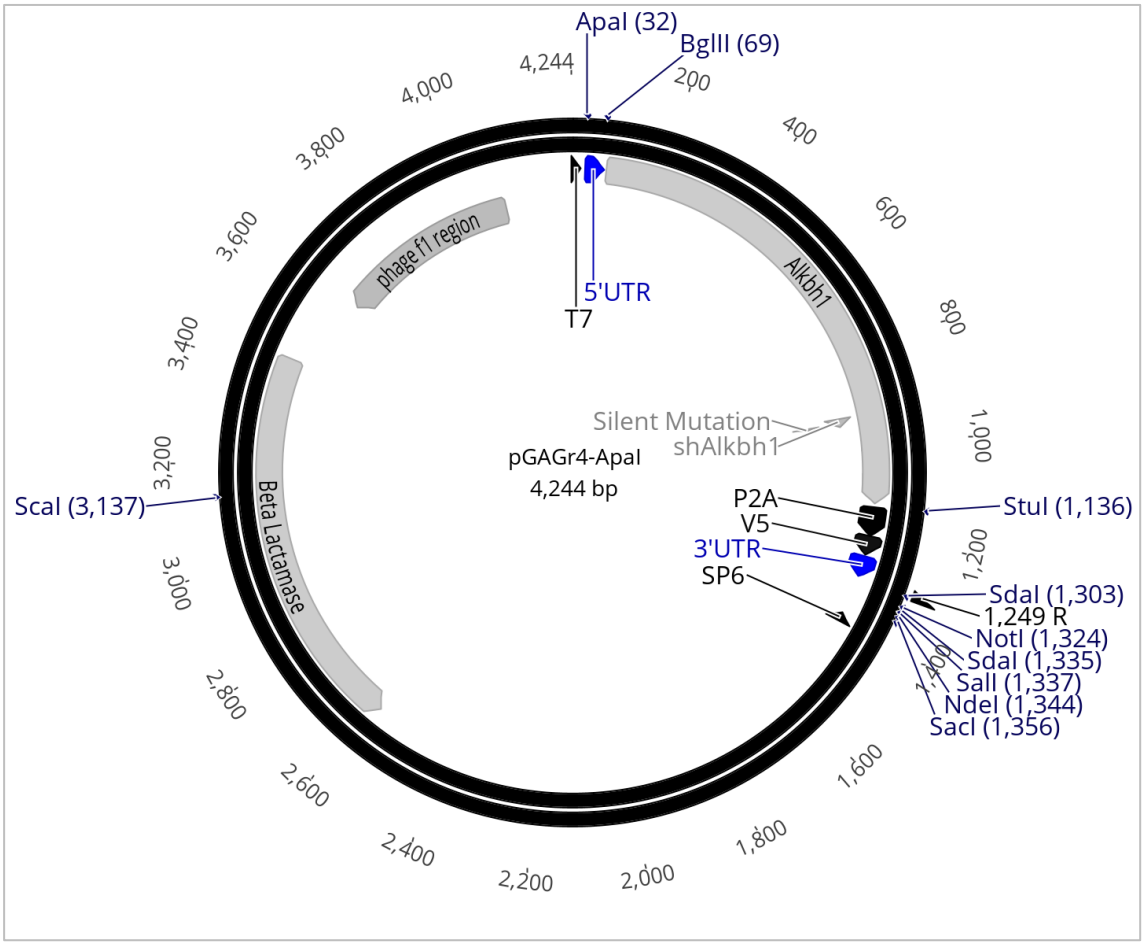
3541 CTGCGCGTAACC ACCACACCCGCC GCGCTTAATGCG CCGCTACAGGGC GCGTCCATTCGC

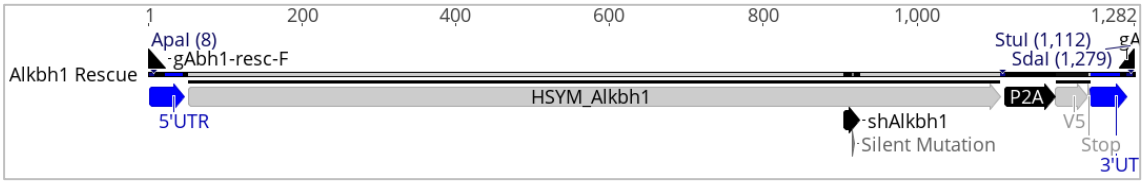
3601 CATTCAGGCTGC GCAACTGTTGGG AAGGGCGATCGG TGCGGGCCTCTT CGCTATTACGCC

3661 AGCTGGCGAAAG GGGGATGTGCTG CAAGGCGATTAA GTTGGGTAACGC CAGGGTTTTCCC

3721 AGTCACGACGTT GTAAAACGACGG CCAGTGAATTG

6.13.2 Map and sequence of Alkbh1-rescue plasmid





pGAGr4-ApaI (pGEMTeasy-Alkbh1GblockRescue clone 4-ApaI digested religated)

1 TAATACGACTCA CTATAGGGCGAA TTGGGCCCTGAG TTTAAGATCAAT TCATCAAAAATT

T7 ***Apa*I** **short-5'UTR**

61 GAGGATGAGATC TATGATGTACAG GGAAAACAAGTT GTGCAATATAGA CCATACAAATGA

Bgl*II** ***Alkbh1

121 TCTTGTTCGAAA GGAATATAAACG TTACAAAAAGCG AAAACCACCACC GGATTTTAGCAA

Alkbh1

181 TGTTATAGATTT TAACGAACCAGC TAGATTTCAGGA TTACATAGAAAG AGTACATTTGGA

Alkbh1

241 TTTAGTTGATTT CAACAATGAACA GTTTTGTTCAAT TGGATTAAAGCA TCCAAAGGAATG

Alkbh1

301 GAATATTTACAA ACTAAAAACGTG TCCCGGTTTTAT AGTTGTAACAAA TCCATTTGTTTC

Alkbh1

361 ATATGGTGCTCA AAGTTACTGGGT GAAACAAACTTT GACGAACTATAC AAAAGGACCATA

Alkbh1

421 TGCAAACAACTT AGACATGTTAAT GAAATTGGATAA AGAAAAGACAAT GTGGCAGATTAG

Alkbh1

481 TCATGAGAAATC AGTGGACAATAC ATTCATTAACCA ATTACGTTGGGT GCATATGGGCTA

Alkbh1

541 CCATTTTGACTA TAATACAGTAGA TTATAAAGCAAA ACAATATTTCGG ATTTCCGGATGA

Alkbh1

601 TCTTGCAAAACT TACTGAAACAAT AGCAAAAGTGTT TAACTTTTCTGA TTATTTTGCAGA

Alkbh1

661 GACTGGTATTGT AAATTACTATCC AGAGGGTAGTAG CATGGGAGGCCA TACTGACCACTA

Alkbh1

721 TGAGGAGGAGCT AAGGCAACCCCT CATCTCTTATAG TTTTGGTCAGTC TGCAGTGTACTT

Alkbh1

781 AATTGGAGGAAA CACTCGAGATAT TAAACCAGAAGC TATATGGGTTCG ATCTGGTGACTT

Alkbh1

841 TATGCTCATGAC AGGCCCATCTCG AACTGCATTCCA TGCTGTACCTTG TATCCTTACCAA

Alkbh1

901 GCCTGTTAAAAC AGGACCTGAGGA TATT **GGCTCATGCCC**GT**TGGTCACT**—CATGGCAACAA

Alkbh1 **shRNA** rescue mutation***Alkbh1***

961 AACTAACGAGCA AGACAATTTTTT GCATTGTACATC AACACATGAACA CAGCCATATGTG

Alkbh1

1021 GAAAAATGAAAT GGATGGTGATTG GGAAGAATTCTC ACAATATTTATC TAATACAAGAAT

Alkbh1

1081 AAATATAAACAT ACGTCAAGTTCA TAAACATGAAGA TACAAAAGAGAATTCT AGGCCT GG

Alkbh1 ***Stu*I**

1141 TTCAGGTGCTAC AAATTTTTCATT ATTAAAACAAGC TGGTGATGTTGA AGAAAATCCAGG

P2A

1201 TCCA GGTAAGCC TATCCCTAACCC TCTCCTCGGTCT CGATTCTACG TAG*TGTGGATATTT*

V5 *****

1261 *GCAAATAAATAT TTACTCTTTACT TTACTGTTTTGA* CCTGCAGGCCAT CACTAGTGAATT

3'UTR ***Sda*I**

1321 CGCGGCCGCCTG CAGGTCGACCAT ATGGGAGAGCTC CCAACGCGTTGG ATGCATAGCTTG

1381 AGTATTCTATAG TGTCACCTAAAT AGCTTGGCGTAA TCATGGTCATAG CTGTTTCCTGTG

1441 TGAAATTGTTAT CCGCTCACAATT CCACACAACATA CGAGCCGGAAGC ATAAAGTGTAAA

1501 GCCTGGGGTGCC TAATGAGTGAGC TAACTCACATTA ATTGCGTTGCGC TCACTGCCCGCT

1561 TTCCAGTCGGGA AACCTGTCGTGC CAGCTGCATTAA TGAATCGGCCAA CGCGCGGGGAGA

1621 GGCGGTTTGCGT ATTGGGCGCTCT TCCGCTTCCTCG CTCACTGACTCG CTGCGCTCGGTC

1681 GTTCGGCTGCGG CGAGCGGTATCA GCTCACTCAAAG GCGGTAATACGG TTATCCACAGAA

1741 TCAGGGGATAAC GCAGGAAAGAAC ATGTGAGCAAAA GGCCAGCAAAAG GCCAGGAACCGT

1801 AAAAAGGCCGCG TTGCTGGCGTTT TTCCATAGGCTC CGCCCCCCTGAC GAGCATCACAAA

1861 AATCGACGCTCA AGTCAGAGGTGG CGAAACCCGACA GGACTATAAAGA TACCAGGCGTTT

1921 CCCCCTGGAAGC TCCCTCGTGCGC TCTCCTGTTCCG ACCCTGCCGCTT ACCGGATACCTG

1981 TCCGCCTTTCTC CCTTCGGGAAGC GTGGCGCTTTCT CATAGCTCACGC TGTAGGTATCTC

2041 AGTTCGGTGTAG GTCGTTCGCTCC AAGCTGGGCTGT GTGCACGAACCC CCCGTTCAGCCC

2101 GACCGCTGCGCC TTATCCGGTAAC TATCGTCTTGAG TCCAACCCGGTA AGACACGACTTA

2161 TCGCCACTGGCA GCAGCCACTGGT AACAGGATTAGC AGAGCGAGGTAT GTAGGCGGTGCT

2221 ACAGAGTTCTTG AAGTGGTGGCCT AACTACGGCTAC ACTAGAAGAACA GTATTTGGTATC

2281 TGCGCTCTGCTG AAGCCAGTTACC TTCGGAAAAAGA GTTGGTAGCTCT TGATCCGGCAAA

2341 CAAACCACCGCT GGTAGCGGTGGT TTTTTTGTTTGC AAGCAGCAGATT ACGCGCAGAAAA

2401 AAAGGATCTCAA GAAGATCCTTTG ATCTTTTCTACG GGGTCTGACGCT CAGTGGAACGAA

2461 AACTCACGTTAA GGGATTTTGGTC ATGAGATTATCA AAAAGGATCTTC ACCTAGATCCTT

2521 TTAAATTAAAAA TGAAGTTTTAAA TCAATCTAAAGT ATATATGAGTAA ACTTGGTCTGAC

2581 AGTTACCAATGC TTAATCAGTGAG GCACCTATCTCA GCGATCTGTCTA TTTCGTTCATCC

2641 ATAGTTGCCTGA CTCCCCGTCGTG TAGATAACTACG ATACGGGAGGGC TTACCATCTGGC

2701 CCCAGTGCTGCA ATGATACCGCGA GACCCACGCTCA CCGGCTCCAGAT TTATCAGCAATA

2761 AACCAGCCAGCC GGAAGGGCCGAG CGCAGAAGTGGT CCTGCAACTTTA TCCGCCTCCATC

2821 CAGTCTATTAAT TGTTGCCGGGAA GCTAGAGTAAGT AGTTCGCCAGTT AATAGTTTGCGC

2881 AACGTTGTTGCC ATTGCTACAGGC ATCGTGGTGTCA CGCTCGTCGTTT GGTATGGCTTCA

2941 TTCAGCTCCGGT TCCCAACGATCA AGGCGAGTTACA TGATCCCCCATG TTGTGCAAAAAA

3001 GCGGTTAGCTCC TTCGGTCCTCCG ATCGTTGTCAGA AGTAAGTTGGCC GCAGTGTTATCA

3061 CTCATGGTTATG GCAGCACTGCAT AATTCTCTTACT GTCATGCCATCC GTAAGATGCTTT

3121 TCTGTGACTGGT GAGTACTCAACC AAGTCATTCTGA GAATAGTGTATG CGGCGACCGAGT

3181 TGCTCTTGCCCG GCGTCAATACGG GATAATACCGCG CCACATAGCAGA ACTTTAAAAGTG

3241 CTCATCATTGGA AAACGTTCTTCG GGGCGAAAACTC TCAAGGATCTTA CCGCTGTTGAGA

3301 TCCAGTTCGATG TAACCCACTCGT GCACCCAACTGA TCTTCAGCATCT TTTACTTTCACC

3361 AGCGTTTCTGGG TGAGCAAAAACA GGAAGGCAAAAT GCCGCAAAAAAG GGAATAAGGGCG

3421 ACACGGAAATGT TGAATACTCATA CTCTTCCTTTTT CAATATTATTGA AGCATTTATCAG

3481 GGTTATTGTCTC ATGAGCGGATAC ATATTTGAATGT ATTTAGAAAAAT AAACAAATAGGG

3541 GTTCCGCGCACA TTTCCCCGAAAA GTGCCACCTGAT GCGGTGTGAAAT ACCGCACAGATG

3601 CGTAAGGAGAAA ATACCGCATCAG GAAATTGTAAGC GTTAATATTTTG TTAAAATTCGCG

3661 TTAAATTTTTGT TAAATCAGCTCA TTTTTTAACCAA TAGGCCGAAATC GGCAAAATCCCT

3721 TATAAATCAAAA GAATAGACCGAG ATAGGGTTGAGT GTTGTTCCAGTT TGGAACAAGAGT

3781 CCACTATTAAAG AACGTGGACTCC AACGTCAAAGGG CGAAAAACCGTC TATCAGGGCGAT

3841 GGCCCACTACGT GAACCATCACCC TAATCAAGTTTT TTGGGGTCGAGG TGCCGTAAAGCA

3901 CTAAATCGGAAC CCTAAAGGGAGC CCCCGATTTAGA GCTTGACGGGGA AAGCCGGCGAAC

3961 GTGGCGAGAAAG GAAGGGAAGAAA GCGAAAGGAGCG GGCGCTAGGGCG CTGGCAAGTGTA

4021 GCGGTCACGCTG CGCGTAACCACC ACACCCGCCGCG CTTAATGCGCCG CTACAGGGCGCG

4081 TCCATTCGCCAT TCAGGCTGCGCA ACTGTTGGGAAG GGCGATCGGTGC GGGCCTCTTCGC

4141 TATTACGCCAGC TGGCGAAAGGGG GATGTGCTGCAA GGCGATTAAGTT GGGTAACGCCAG

4201 GGTTTTCCCAGT CACGACGTTGTA AAACGACGGCCA GTGAATTG

6.14 Link to Supplementary Document

https://aran.library.nuigalway.ie/