Regulation of early cell fate specification in the crustacean embryo, *Parhyale hawaiensis*



Tzu-Chieh Wen

Wolfson College
University of Oxford

A thesis submitted for the degree of
Master of Science by Research

Trinity Term 2022

Declaration

I hereby declare that the thesis entitled "Regulation of early cell lineage specification in the crustacean embryo, *Parhyale hawaiensis*" has been originally carried out by me under the supervision of Professor Aziz Aboobaker. This work has not formed the basis for award of any degree or diploma previously. The particulars given in the thesis are true to the best of my knowledge.

Tzu-Chieh Wen

Department of Zoology
Wolfson College
University of Oxford

Acknowledgements

I would like to thank my supervisor Professor Aziz Aboobaker for his guidance and support over the past year that I was working in his lab. I have learned a tremendous amount and have had a productive year of science and discovery. I would also like to extend my gratitude to the Elizabeth Hannah Jenkinson Fund for a research grant to fund part of this work.

I would also like to thank Dr. Manuel Jara Espejo for helping me with the bioinformatics and analysis of the transcriptomic data as well as troubleshooting protocols in the lab together. It was a pleasure to work alongside you and learn from you.

Thank you to Salha Rawas for teaching me all the necessary lab skills for working with *Parhyale*. You figured out a lot of the challenges to working with *Parhyale* in our lab, so I was able to hit the ground running without having to worry too much about establishing protocols and maintaining animal cultures.

Many thanks to Simon Kershenbaum for being a supportive colleague and so keen to discuss my work with me and throw around ideas regarding future step to take.

Finally, many thanks to my parents, Wen-Ching Wen and Chia-Hui Yu, and sister, Julia Wen, for supporting me throughout the year and encouraging to pursue an educational year in the United Kingdom before my PhD.

Abstract

Early cell fate specification is important in the development of animal embryos. This specification is reliant on maternally deposited mRNAs and proteins which help activate lineage-specific gene regulatory networks and requires mechanisms that can establish asymmetries. Previous studies have shown that there is germ layer specification at the 8-cell stage and that some mRNAs are asymmetrically distributed amongst the early progenitors of the germ layers of the amphipod crustacean, *Parhyale hawaiensis*. These early asymmetries must translate into differential regulation of maternal to zygotic transition (MZT) with different germ layers to ensure correct specification and patterning during development. This makes *Parhyale* an attractive model to study how early cell fate specification occurs and how early lineage restriction subsequently regulates MZT. In this thesis, we establish single blastomere lineage transcriptomic datasets and test various CRISPR/Cas9 strategies to aid in our understanding of early cell fate specification and the role of maternally loaded mRNAs.

Firstly, I describe a method to generate single blastomere lineage transcriptomic datasets. I sequenced mRNAs of each blastomere of the 4- and 8-cell *Parhyale* embryo to describe genes implicated by asymmetric distribution to be potentially involved in early specification. I show that many early patterning genes, germline specific genes, and translational repression RNA-binding proteins that are asymmetrically distributed in *Drosophila* and *C. elegans* are ubiquitously expressed in the *Parhyale* embryo by *in situ* hybridization, verifying that transcriptomic data correctly reflects expression of these genes. The establishment of these transcriptomic datasets will now allow us to pick candidate genes for functional studies of early cell fate specifiers as well as analysis of lineage specification as cells divide. Using principal component analysis, we show that macromeres of the 8-cell embryo (Er, Ep, El, Mav) are more similar to their parent cells (Emr,

Epen, Eml, Mavg) than their sister cells (mr, en, ml, g) are to their parents. Further analysis also shows that many genes while ubiquitously expressed throughout the embryo have higher expression in specific blastomeres. In addition, I tested various CRISPR/Cas9 strategies for knock-out and epitope tag knock-in experiments to improve current CRISPR mutagenesis methods on *Parhyale* and to combine other imaging and -omic technologies with the CRISPR/Cas9 system. I used the pioneer transcription factor *Zelda* as the study gene because it is expressed both before and after MZT, and may be a conserved regulator of MZT in *Parhyale*. I designed guide RNAs (sgRNAs) to target *Zelda* for CRISPR knock-outs and designed plasmid constructs to tag *Zelda* with various epitope tags. I tested non-homologous end-joining and microhomology-mediated end-joining methods for efficiency of CRISPR knock-ins and found that both methods fail to increase penetrance. In the future, we hope to further analyze and test candidate genes from the blastomere-specific transcriptomic datasets as well as improve CRISPR/Cas9 mutagenesis strategies so we can better study maternally expressed genes in the *Parhyale* embryo.

Key words: Cell fate specification, Maternal-to-zygotic transition, Parhyale hawaiensis, CRISPR/Cas9, Embryogenesis, Transcriptomics, Zelda

Table of Contents

Declaration……..…i
Acknowledgements……..…..ii
Abstract…….iii

Chapter 1. Introduction……..……..1
1.1. Early cell fate specification 2
1.2. Maternal-to-zygotic transition 5
1.3. *Parhyale hawaienses*: a model system to study arthropod evolution and development 7
1.4. *Parhyale* as a model to study cell fate specification and MZT 10
1.5. Thesis aim and outline 11

Chapter 2. Materials and Methods……..……...12
2.1. Sequences 13
2.2. Animals 13
2.3. Cloning 13
2.4. Embryo deshelling and fixation 14
2.5. Blastomere isolation and fixation 14
2.6. RNA extraction of fixed cell suspensions 14
2.7. mRNA-seq data analysis 15
2.8. RNA probe synthesis 16
2.9. *In situ* hybridization 17
2.10. CRISPR/Cas9 construct design and synthesis 18
2.11. CRISPR/Cas9 genome editing 19

Chapter 3. Identification of genes important in early cell fate specification……..21
3.1. Introduction 22
3.1.1. Early cell fate specification in the *Parhyale* system 22
3.1.2. RNA sequencing and its applications for developmental biology research 22
3.1.3. RNA sequencing in the *Parhyale* model 23
3.2. Aim of this chapter 25
3.3. Results and discussion 26
3.3.1. Separation of individual blastomeres at the 4- and 8-cell stage 26
3.3.2. Analysis of RNA sequencing of 4- and 8-cell stage blastomeres 28
3.3.3. Validation of transcriptomic datasets 30
3.4. Conclusion 36

Chapter 4. Development of CRISPR/Cas9 strategies in *Parhyale hawaiensis* to study maternally expressed genes……...….37
4.1. Introduction 38
4.1.1. CRISPR/Cas9 strategies 39
4.1.2. CRISPR/Cas9 knock-in applications 40
4.2. Aim of this chapter 42
4.3. Results and discussion 43
4.3.1. Design and testing of guide RNAs for CRISPR/Cas9 mutagenesis 43
4.3.2. *Zelda* loss-of-function knockout in 1-cell embryos using CRISPR/Cas9 shows significant decrease in survival after 5 days 46
4.3.3. Design and assembly of plasmid constructs for CRISPR/Cas9 knock-in 47
4.3.4. NHEJ and MMEJ methods with linearized DNA do not successfully tag *Zelda* with epitope tags 52
4.4. Conclusion 53

Chapter 5. Thesis discussion and future directions……..54
5.1. Identifying early cell fate specifiers during early crustacean development 55
5.2. Establishing genetic and genomic tools and resources to study *Parhyale* embryonic development 56

References……..…58

Appendix……...69
Appendix A: Nucleotide coding sequences of candidate genes in this study 70
Appendix B: Details of primer sequences for CRISPR/Cas9 knock-in constructs 76
Appendix C: Details of primer sequences for *in situ* hybridization 77

Chapter 1

Introduction

1. Introduction

1.1. Early cell fate specification

The accurate specification of cell fates is crucial in the normal development of an embryo. Although all organisms with multiple tissue types require specification of distinct germ lines, cell fate specification widely varies across the animal kingdom. For example, mammals, echinoderms, nematodes, and ascidians display holoblastic cleavage which means that they generate individual separate cells during early embryonic development (Fig 1.1A; Gilbert, 2000). These individual cells known as blastomeres have distinct lineages and will contribute to different parts of the organism. Another alternative cleavage pattern is meroblastic cleavage, common in fish, reptiles, and birds. In meroblastic cleavage, the embryo does not divide fully but rather the cell divides partially while keeping the pole of the cell containing the yolk intact. This results in a blastodisc that spreads over the yolk with a syncytial layer adjacent to the yolk (Fig 1.1B; Hasley *et al.*, 2017). Although the type of cleavage pattern is different, the blastomeres with different mRNA and protein profiles will result in specification of the germ layers. In contrast, the fruit fly *Drosophila melanogaster* displays superficial cleavage, meaning that they do not create individual blastomeres during early development but rather a multinucleate syncytium where the nuclei can migrate throughout the cytoplasm (Fig 1.1C; Gilbert, 2000). Although there are no distinct blastomeres in the *Drosophila* embryo, the gradient of factors in the syncytium results in distinct cell fate decisions during early specification.

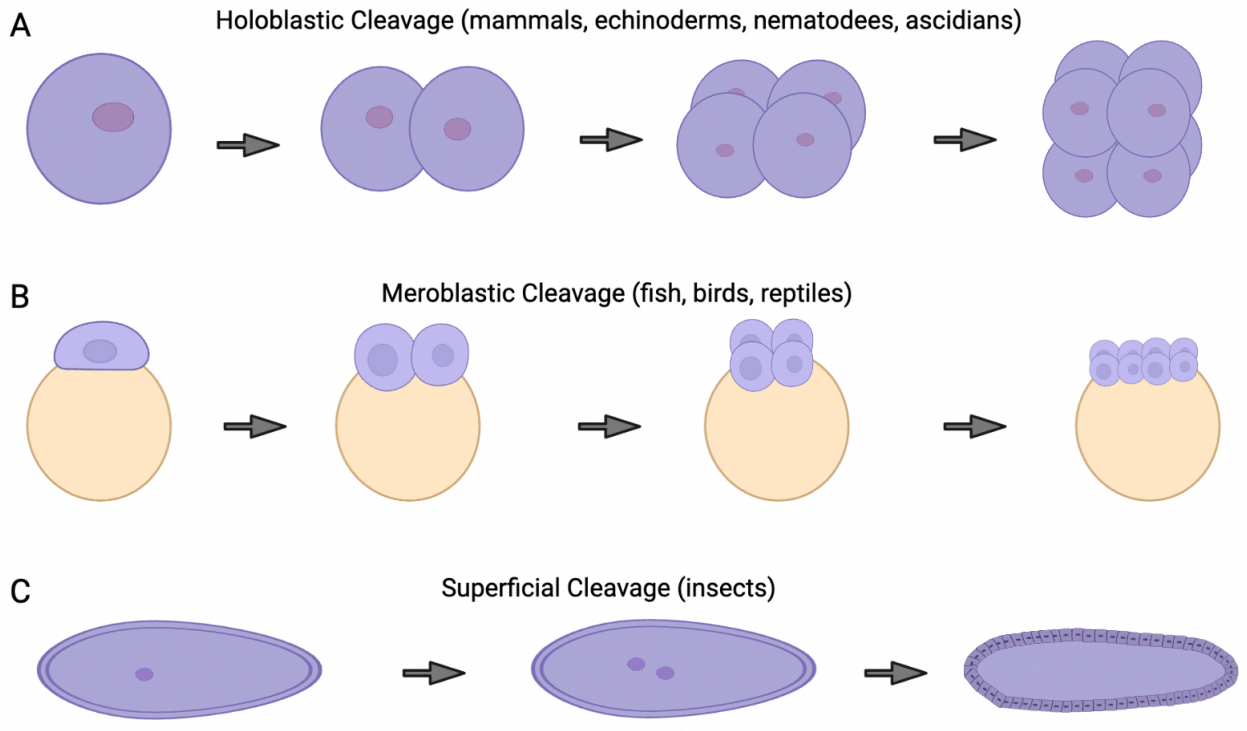


Fig 1.1: Overview of general early cleavage patterns across animals including A) holoblastic cleavage, B) meroblastic cleavage, and C) superficial cleavage.

Regardless of the type of cleavage pattern and specification that occurs during early embryogenesis, there are key processes that result in the development of different germ layers. The establishment of polarity and the asymmetric distribution of RNAs and proteins are crucial in the formation of distinct cell lineages. In organisms that go through holoblastic cleavage such as tunicates, morphogenetic determinants such as proteins and mRNAs are localized to different parts of the embryo before cleavage. Once cleavage occurs, each individual blastomere will contain different factors that are associated with a specific cell lineage (Whittaker, 1973). Similarly, the nematode *C. elegans* requires multiple factors to result in anterior-posterior (A-P) axis determination at the first division. MEX3 and GLP1 proteins are expressed anteriorly while SKN1 and PAL1 are expressed posteriorly (Bowerman *et al.*, 1997). These A-P determinants are shown to work independently of one another, and loss-of-function experiments of individual factors does

not result in a full disruption of A-P asymmetry (Bowerman *et al.*, 1997). In addition, interactions with neighboring cells also result in the formation of different germ layers as seen in most vertebrate embryos and a few invertebrate species (Gilbert, 2000). Sea urchin embryos provide a phenomenal example of inhibitory interactions between different cell types. In the sea urchin, removal of skeletal mesoderm cells results in neighboring cell types to change their cell fate to that of skeletal mesoderm (Cheng *et al.*, 2014). In mouse embryos, cell-cell signaling also plays an important role in cell fate specification. Fgf signaling is essential for the specification of the primitive endoderm as the embryonic epiblast precursor cells upregulate *fgf4* which inhibits the repression of *Gata6* by *Nanog* (Bedzhov *et al*., 2014). Therefore, cell fate specification may also depend on cell-cell interactions and may not be cell autonomous. Another way in which cell fate is specified is through morphogen gradients in which soluble compounds diffuse through the embryo and the varying levels of the morphogen result in specification of certain cell types. In *Drosophila* embryos and other fly species, the morphogen *Bicoid*, is present in the anterior end and results in a gradient causing the specification of head and tail (St. Johnston *et al.*, 1989; Yoon *et al.*, 2019). Varying concentrations of morphogen activate and repress downstream genes, resulting in differential cell fate specification along the A-P axis embryo.

Taken together, cell fate specification is important in the formation of different tissue types to form organisms with a complex body plan. At its core, specification requires asymmetric distribution of mRNAs, proteins, and morphogens as well as establishment of polarity so distinct blastomeres and different parts of the zygote will commit to specific cell lineages.

1.2. Maternal-to-zygotic transition

The development of a multi-cellular organism with multiple tissue types is initially controlled by maternal gene products but later directed by the zygotic genome. In many species, maternally provided factors including RNAs and proteins are important in this transition because they help activate the early embryonic genome. This maternal-to-zygotic transition (MZT) results in zygotic genome activation (ZGA), degradation of many/most maternally loaded factors, and replacement of maternal transcripts with zygotic transcripts so that embryonic development is controlled by the zygote rather than by the mother (Fig 1.2). For MZT to occur, there needs to be coordination of multiple processes including changes in cell cycle regulation (Yuan *et al.*, 2016; Blythe and Wieschaus, 2015; Tadros *et al.*, 2007), transcriptional activation of the zygotic genome (Anderson and Lengyel, 1979; Lamb and Laird, 1976), regulation and degradation of maternal gene products (De Renzis et al., 2007; Thomsen et al., 2010; Tadros et al., 2007).

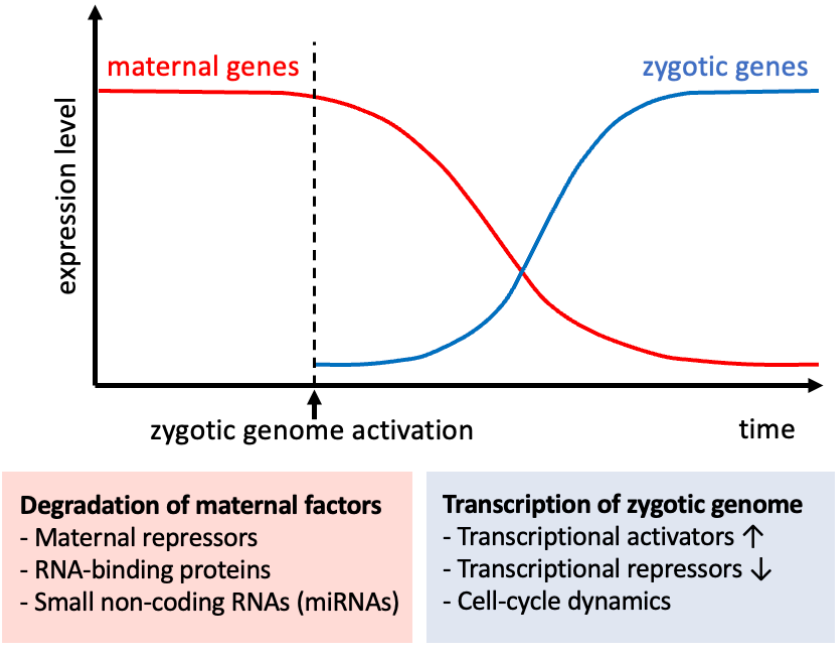


Fig 1.2: Overview of maternal-to-zygotic transition and mechanisms for the degradation of maternal factors and activation of the zygotic genome.

Although MZT occurs in all higher plants and animals, the time and processes differ between organisms. Before MZT occurs, a fraction of the protein-coding transcripts is expressed maternally from one-third in mouse (Wang et al., 2004) and *C. elegans* (Stoeckius et al., 2014), to three-quarters in the zebrafish (Harvey et al., 2013) and Drosophila (Tadros et al., 2007; Thomsen et al., 2010; Lécuyer *et al.*, 2007). During MZT, a range of maternal transcripts are eliminated from the embryo from one-quarter in zebrafish (Aanes *et al.*, 2011), one-third in mouse (Hamatani *et al.*, 2004) and *C. elegans* (Baugh *et al.*, 2003), and two-thirds in *Drosophila* (Thomsen *et al.*, 2010).

There are multiple ways in which maternally loaded factors are degraded. RNA-binding proteins regulate destabilization of maternal RNA during MZT. For example, in *Drosophila*, the RNA-binding proteins SMAUG and BRAT regulate mRNA decay by binding to hairpin structures on its targets and recruiting degradation machinery (Aviv *et al.*, 2003; Semotok *et al.*, 2008; Chen *et al.*, 2014; Laver *et al.,* 2015). Small non-coding RNAs such as microRNAs (miRNAs) also play a role in maternal transcript decay. In zebrafish, the *miR-430* family of miRNAs is required for the clearance of approximately 40% of maternal transcripts (Bazzini *et al.*, 2012).

During ZGA, cell cycle length, transcriptional activators and transcriptional repressors play a role in the transition from a repressed to an activated zygotic genome. Previous work has shown that cell cycle length affects the production of transcripts by their number and length in *Drosophila* and zebrafish (Edgar and Schubiger, 1986; Dalle Nogare *et al.*, 2009). The interplay between transcriptional repressors and activators is important in the activation of the zygotic genome. Histones are considered general transcriptional repressors and are found in large excess during the early embryo (Anderson and Lengyel, 1980). They block off DNA so transcriptional activators are unable to bind (Workman and Kingston, 1998). During ZGA of zebrafish embryos, the nuclear histone concentration in the embryo drops because of the increase in DNA content as well as the

changes in cytoplasm to nuclei ratio (Joseph *et al.*, 2017). This drop in the concentration of histones allows for transcriptional activators to then bind to DNA (Joseph *et al.*, 2017). Such transcription factors include *Zelda* and *Clamp* in *Drosophila* (Liang *et al.*, 2008; Harrison *et al.*, 2011; Duan *et al.*, 2021), *Pou5f3, Sox19b,* and *Nanog* in zebrafish (Lee *et al.*, 2013), as well as *Oct4* and *Dux4* in humans (Gao *et al.*, 2018). RNAs of these transcriptional activators are maternally loaded and then translated in the early embryo to result in activation of the zygotic genome (Blythe and Wieschaus, 2016).

Together, cell fate specification and maternal-to-zygotic transition occur during early embryonic development and often alongside one another. Whether the cell fate is specified before, during, or after MZT depends on the organism. However, the development of any multicellular organism ultimately depends on the transition from maternal to zygotic control of the genome as well as the specification of germ layers to form various tissue types.

1.3. *Parhyale hawaienses*: a model system to study arthropod evolution and development

Since the twentieth century, *Drosophila melanogaster* has been the predominant model system in arthropod development because of the ease to maintain laboratory cultures, its capacity to be manipulated genetically, and its convenience to test evolutionary hypotheses (Arias, 2008). This model system has been central in our understanding of Homeotic genes and segmentation (Lewis, 1978), segmental patterning (Nüsslein-Volhard and Wieschaus, 1980), as well as other developmental processes. *Drosophila* has been chosen to represent arthropods out of the traditional model organisms because of its strength as a genetic system and has played a tremendous role in the field of genetics and developmental biology. However, it represents only a small subset of arthropods which not only includes all orders of insects but also crustaceans, myriapods, and chelicerates.

In recent years, the amphipod crustacean *Parhyale hawaiensis* has emerged as a model to study arthropod development, regeneration, and immunity (Gerberding *et al.*, 2002; Extavour, 2005; Browne *et al.*, 2006; Pavlopoulos *et al.*, 2009; Vargas-Vila *et al.*, 2010; Price *et al.*, 2010; Blythe *et al.*, 2012; Serano *et al.*, 2016). *Parhyale* being a crustacean represents a clade of non-insect arthropods and can thus serve as a phylogenetic outgroup for insects (Fig 1.3A; Schwentner *et al.*, 2017). In addition to its phylogenetic advantage, it has many characteristics which make it an attractive model to work with. *Parhyale* are found in shallow water along tropical coastlines globally and have a high tolerance for environmental changes (Barnard, 1965; Shoemaker, 1956). They are small and easy to rear, making them ideal for laboratory culture. *Parhyale* females can lay eggs every two weeks all year round and the embryos develop into hatchlings in around 10 days (Fig 1.3B; Browne *et al.*, 2005). Their embryos are large and can be easily collected from the female's ventral brood pouch (Rehm *et al.*, 2009). In addition, they can be imaged since the chorion and the embryo itself are transparent. Finally, *Parhyale* will reach reproductive age two months after hatching, thus allowing researchers to establish transgenic and mutant lines.

Due to the ease of studying *Parhyale* embryos, many embryological techniques and genetic tools have been applied to this model. Easily accessible protocols have been published and described in detail for embryo microinjections (Rehm *et al.*, 2009a), dissections of the eggshell and fixation (Rehm *et al.*, 2009b), antibody staining (Rehm *et al.*, 2009c), and in situ hybridization of labeled RNA probes (Rehm *et al.*, 2009d). Other embryological techniques that have been performed include blastomere isolations (Extavour, 2005), cell ablation (Price *et al.*, 2010; Alwes *et al.*, 2011; Nast and Extavour, 2014), and live imaging (Alwes *et al.*, 2011; Hannibal *et al.*, 2012; Chaw and Patel, 2012; Alwes *et al.*, 2016). Genetic tools have also been developed for the *Parhyale* system. Loss-of-function techniques such as CRISPR/Cas9 knock-out (Martin *et al.*, 2016; Kao *et al.*,

2016), RNAi knock-down (Liubicich *et al.*, 2009) and morpholino knock-down (Ozhan-Kizil *et al.*, 2009) have also been used to understand gene functions. Finally, the genome and transcriptome of *Parhyale* were sequenced to guide functional studies (Zeng *et al.*, 2011; Blythe *et al.*, 2012; Kao *et al.*, 2016; Calvo *et al.*, 2022).

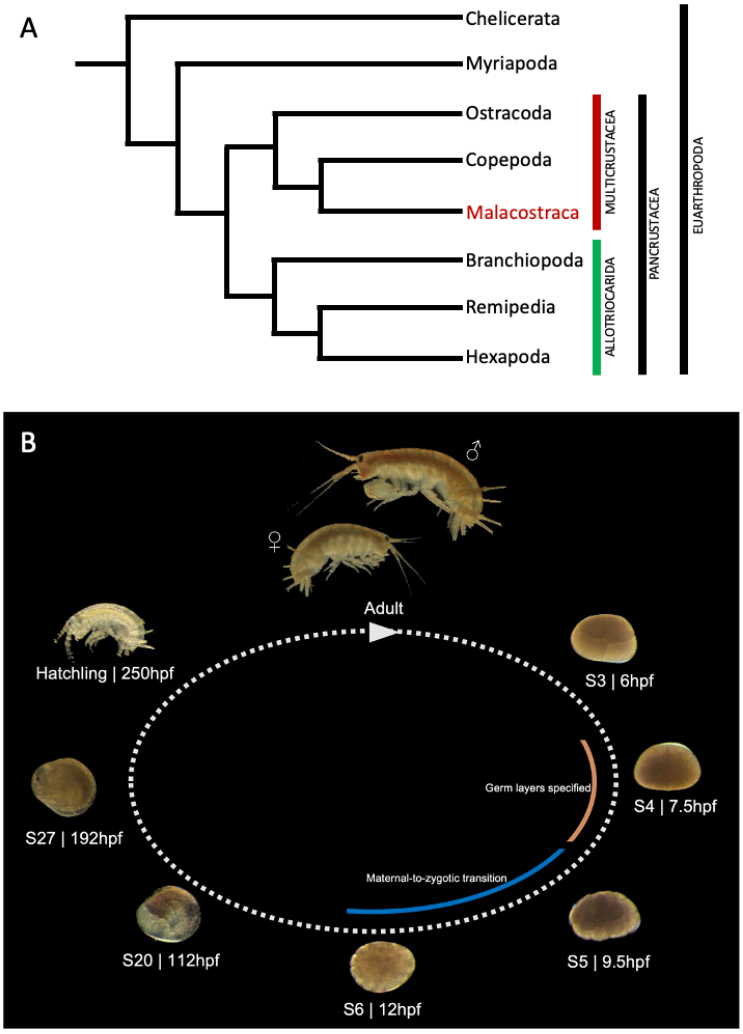


Fig 1.3: Introduction to *Parhyale* life history (A) Phylogeny of arthropods showing *Parhyale,* a malacostraca crustacean, as an outgroup to hexapods which includes all insects (modified from Sun and Patel, 2019). (B) *Parhyale* are direct developers and development occurs over approximately 10 days. Germ layers are already specified at S4 (8-cell stage) and maternal-to-zygotic transition occurs around S5-6 (Images from Kao *et al.*, 2016).

1.4. *Parhyale* as a model to study cell fate specification and MZT

In addition to their life history and the technologies and protocols developed to study them, *Parhyale* is an exceptional model to study cell fate specification and MZT because of its early specification and well documented cell lineages. After the third cleavage, *Parhyale* embryos contain four macromeres and four micromeres which can be distinguished based on their position and shape (Browne *et al.*, 2005, Gerberding *et al.*, 2002). The *in vivo* dye labeling of each blastomere shows that each cell results in distinct cell lineages and cell fates and each cell fate is restricted to only one germ layer (Gerberding *et al.*, 2002). Three of the macromeres are restricted to the ectoderm germ layer and the final macromere results in the anterior and visceral mesoderm. Two of the micromeres results in the trunk mesoderm, one micromere results in the endoderm, and the final micromere results in the germ line (Gerberding *et al.*, 2002). When one or two of the mesoderm precursors are ablated, the remaining mesoderm precursor will replace the lost cell and result in a normal hatchling (Price *et al.*, 2010). However, ectoderm, endoderm, and germ line precursors are not able to replace lost mesoderm. This reveals that there is an early restriction of cell fates and limited plasticity amongst the different germ layers. Transcriptomic studies have shown that MZT occurs at around the 32-cell stage in two waves: a major wave and a minor wave (Nestorov *et al.*, 2013; Calvo *et al.*, 2022). Custom gene expression microarray data on various groups of dissected blastomeres show that 129 RNAs included in the microarray are asymmetrically distributed at the eight-cell stage in *Parhyale* embryos (Nestorov *et al.*, 2013). Since there is an asymmetric distribution of RNAs at the 8-cell stage before MZT occurs, previous work suggests that maternally loaded RNAs result in the specification of early cell fate.

1.5. Thesis aim and outline

The primary aim of my thesis project is to understand the regulation of early cell fate specification before MZT in the *Parhyale* embryo. From a transcriptomic dataset, our lab has uncovered *Zelda* as a maternally expressed gene present before MZT. Since *Zelda* is a pioneer transcription factor and plays an important role in MZT in *Drosophila,* I hypothesize that it is an essential gene in early *Parhyale* embryonic development as it binds to and activates genes related to cell fate specification, ultimately resulting in distinct germ layers.

Chapter 1 has given a brief overview of early cell fate specification and maternal-to-zygotic transition along with an introduction to *Parhyale hawaiensis* as an emerging model system to study early embryonic development. **Chapter 2** describes the methods and materials used to study *Parhyale* development as presented in the results chapters of this thesis. In **Chapter 3**, I have produced transcriptomic datasets for each blastomere of the 4- and 8-cell embryo and identified asymmetrically distributed mRNAs at early embryonic stages. In addition, I show that these transcriptomic datasets reflect what we see in the actual embryo by *in situ* hybridization.

In **Chapter 4**, I discuss the current state of CRISPR/Cas9 technology and its applications, as well as the use of CRISPR/Cas9 technology in the *Parhyale* model. I have developed and tested various strategies for CRISPR/Cas9 knock-outs and knock-in of epitope tags. In order to do so, I designed guide RNAs for knock-ins and knock-outs, tested them *in vitro* and *in vivo*, and, for the knock-ins, delivered them with donor DNA constructs containing the desired tag.

Chapter 5 provides a summary of this thesis. Here, I discuss the main findings of all the results mentioned in the previous chapters, the significance of the results presented in this thesis, and future directions to understand early cell fate specification and the application of CRISPR/Cas9 technologies to understand maternal genes in a non-traditional model organism.

Chapter 2

Materials and Methods

Chapter 2. Materials and Methods

2.1. Sequences

Coding sequences of Zelda and cell lineage specification candidate genes come from a reference-guided transcriptome assembly that used polyA RNA-seq data from multiple *Parhyale* developmental stages (Espejo, unpublished; Appendix A).

2.2. Animals

Parhyale hawaiensis cultures, derived from the 'German' line, were used throughout this study. *Parhyale* were kept in plastic boxes or glass tanks with artificial seawater at 25°C and salinity of 30 PPT. Animals were fed with carrots and fish flakes once a week and tanks were cleaned once a week. Animal cultures were aerated using aquarium pumps and airstones. For embryo collections, mating pairs were collected 1 day prior to collection, gravid females were anesthetized with clove oil (diluted 1:10000 in artificial seawater), and embryos were collected from the ventral brood pouch using forceps and a blunted glass pasteur pipette.

2.3. Cloning

Total RNA was extracted from approximately 40 mg of embryos using a TRIzol based protocol (Thermo-Fischer). 20 μL of cDNA was synthesized using the SuperScript Reverse Transcriptase III Kit (Invitrogen). Primers were designed against a *Parhyale* transcriptome dataset for all full length and partial length coding sequences used in this study (Appendix B; Appendix C). Targets were amplified from Polymerase Chain Reaction (PCR) of *Parhyale* early embryo cDNA and cloned into pGEM T-Easy Vector (Promega). Plasmids containing the amplicons were prepared using the Wizard plus SV Minipreps DNA Purification System (Promega) and sent off for Sanger sequencing.

2.4. Embryo deshelling and fixation

Embryo deshelling and fixation was performed according to published protocols (Rehm *et al.*, 2009). In short, eggshells were dissected off using No. 5 Forceps (World Precision Instruments) and Tungsten Needles (Fine Science Tools) in 3.7% paraformaldehyde (PFA) solution in filtered artificial seawater. Embryos were left in PFA for 2-3 hours before dissection for easier removal of the eggshell. From the initial puncturing of the eggshell, embryos were left in PFA for at least 15 minutes. After fixation, embryos were washed three times with PT (1X PBS with 0.1% Triton X-100) and dehydrated with sequential washes of 50%, 70%, 90%, and 100% methanol in PT.

2.5. Blastomere isolation and fixation

To isolate individual blastomeres, 4-cell and 8-cell embryos were treated in 3.7% PFA in 10X PBS for at least 20 minutes. By applying pressure using No. 5 forceps (World Precision Instruments), blastomeres became loose from each other. Embryos were oriented in a way as to determine the identity of the blastomeres according to the position from the g cell and the Mavg cell for the 8-cell and 4-cell embryo respectively. Using forceps and Tungsten Needles (Fine Science Tools), individual blastomeres were able to be fully separated and the eggshell was able to be removed. Blastomeres of the same cell type were pooled together and transferred to 1.5 mL tubes containing 3.7% PFA in 10X PBS and kept on ice until all dissections were completed. Once all dissections for the day were completed, as much PFA as possible was removed without drying out the embryos, and embryos were washed with PT containing 1% BSA. Embryos were then kept at -20°C for up to two weeks until RNA extraction was performed.

2.6. RNA extraction of fixed cell suspensions

Total RNA was isolated from PFA-fixed blastomeres using a combination of the RNeasy FFPE kit (Qiagen) and the RNeasy Plus Micro Kit (Qiagen). Blastomeres were first washed with 1mL

of PT containing 1% BSA and 2 µL of RNase inhibitor (Thermo Fisher Scientific). After washes, RNA was extracted using the kits as described previously (Russell *et al.*, 2013). Cells were suspended in Buffer PKD (Qiagen) and 10 µL of proteinase K (Qiagen) solution were added to each sample. Samples were incubated at 56°C for 15 minutes and vortexed every 3-5 minutes. Samples were then incubated at 80°C for 15 minutes and shaken at 400 rpm. 500 µL of Buffer RBC (Qiagen) was added to each sample and the samples was passed through a gDNA Eliminator column. The column was discarded, and flowthrough kept. 1200 µL of 100% ethanol was added to the flowthrough, then passed through a RNeasy MiniElute spin column. Samples were washed with 800 µL of Buffer RW1 (Qiagen), 500 µL of Buffer RPE (Qiagen), then 500 µL of 80% ethanol. Finally, samples were eluted with 15 µL of RNase-free water and stored at -80°C. Total RNA was sent to Novogene (Cambridge) for sequencing. Novogene handles the mRNA library preparation including mRNA enrichment by poly A capture, RNA fragmentation, and cDNA reverse transcription. Sequencing was performed using the Illumina platform and the PE150 strategy for paired-end, non-directional sequencing of 20 million short reads (6 G raw data) per sample.

2.7. mRNA-seq data analysis

RNA-seq paired-end reads were trimmed using Trimmomatic V0.39 to result in clean reads (Bolger *et al.*, 2014). The SLIDINGWINDOW was set to 4 bp with a quality threshold of 20 where it scans from the 5' end of the read and will clip the read once the average quality within the window of 4 bp falls below the quality threshold of 20. The LEADING and TRAILING parameters were set to 10 to cut off bases at the start and end of the read below a quality threshold of 10. MINLEN was set to 36 which drops all reads that are shorter than 36 bp in length. The clean reads were mapped using HISAT2 onto the *Parhyale* genome assembly described in (Kao *et al*., 2016).

Using an updated transcriptome annotation, we quantified mRNA expression using the featureCounts function in the Rsubread package (Liao *et al.*, 2019). Expression values were normalized as TPM (transcripts per million) using the TPMCalculator software (Vera Alvarez *et al.*, 2019). Principle component analysis (PCA) was performed using the plotPCA function of the DESeq2 package (Love *et al.*, 2014).

We then inferred blastomere-specific and lineage-specific genes based on candidates with the highest fold change between the TPM of each blastomere/lineage and that of the other blastomeres. This includes comparing each blastomere to the other blastomeres, comparing the average TPM of ectodermal blastomeres (El, Ep, Er) to the TPMs of the other blastomeres, and comparing the average TPM of mesodermal blastomeres (ml, mr) to the TPMs of the other blastomeres. In addition, we inferred micromere-specific genes based on top candidates with the highest fold change between the average TPM of micromeres and the average TPM of macromeres. From here, we selected genes for experimental validation using *in situ* hybridization including genes that are specific to a lineage and genes that might play an important role in cell fate specification based on the literature.

To better visualize the gene expression in each blastomere at the 4- and 8-cell stages, we created coverage plots using the SuperIntronic package (Lee *et al.*, 2020).

2.8. RNA probe synthesis

Full length and partial length coding sequences were amplified from cDNA and a subsequent PCR was run with a reverse primer containing a T7-adaptor appended to the 5’ end, allowing direct transcription from the PCR products instead of cloning each candidate gene. The T7-adaptor also allowed for Sanger sequencing of PCR products to confirm correct amplification of candidate genes. PCR product was purified using a phenol/chloroform extraction method.

RNA probes were synthesized by *in vitro* transcription using T7 RNA polymerase with its buffer, RNAse inhibitor, and were labeled with Digoxigenin-11-UTP (Roche) for 3 hours at 37°C. The reaction mix was treated with DNAse I for 15 minutes at 37°C. The RNA probe was precipitated out using 3M sodium acetate and ethanol at -20°C overnight. The sample was centrifuged at 4°C for 30 minutes, followed by 3 washes using 70% ethanol. The pellet was left to air dry then resuspended in 20 µL of RNAse-free water. RNA probes were stored at -20°C.

2.9. *In situ* hybridization

In situ hybridization was performed according to a published protocol but in a 24-well plate (Rehm *et al.*, 2009). PFA-fixed embryos were rehydrated with 70%, 50%, and 30% MeOH/PT and washed three times with PT. Embryos were fixed again with 3.7% PFA in PT for 30 minutes then washed twice with PT. Embryos were then pre-hybridized in hybridization buffer for 60 minutes or more at 65°C then hybridized with 1 ng/µL of RNA probe in hybridization buffer overnight at 65°C. Samples were washed with hybridization buffer, then four times with 2X SSC buffer for 30 minutes at 65°C. After two 10-minute washes with 2X SSC buffer at room temperature, embryos were washed four times with PT for 20 minutes with PT at room temperature. Samples were blocked in PT + 1% BSA for at least 30 minutes then incubated with AP-conjugated anti-DIG antibody (1:3000, Roche) overnight at 4°C. Samples were washed four times with PT for 30 minutes at room temperature then three times with AP reaction buffer for 5 minutes. RNA probes were visualized using NBT/BCIP (Roche) in the dark at room temperature. Once color has developed, reactions were stopped by washing embryos with PT three times for 5 minutes. Embryos were stored in 70% glycerol/PT until imaging. The specificity of the *armadillo* probe being only expressed in the g cell acts as a control for the ISH assays since the other blastomeres fail to produce the dark blue color (Fig 3.4).

2.10 CRISPR/Cas9 construct design and synthesis

For the knock-in experiments, multiple tagging plasmids were designed and constructed for CRISPR/Cas9 mutagenesis using Non-Homologous End Joining (NHEJ) and microhomology-mediated end joining (MMEJ) strategies. pCRISPR-Zld-T2A-H2B-mRuby2 contained the partial Zld coding sequence fused in-frame with the T2A self-cleaving peptide, the histone H2B and mRuby2 fluorescent protein, followed by the pGEM T-easy vector backbone (Promega). pCRISPR-Zld-3XFLAG-T2A-H2B-mRuby2 contained an additional 3XFLAG sequence in between the coding sequence and T2A self-cleaving peptide. Finally, pCRISPR-Zld-FLAG-4C contained the partial Zld coding sequence fused in frame with FLAG tag sequence and tetracysteine provided by Patrik Verstreken (Addgene plasmid #19175). Different parts of the constructs were isolated using PCR amplification and restriction enzyme digestion followed by gel isolation (Qiagen). Primers were designed to allow overlapping of sections to be used in Gibson Assembly (Appendix B). Using the Gibson Assembly kit (NEB) or T4 DNA Ligase (Promega), various sections of the plasmid construct were assembled. Plasmid constructs were transformed into NEB 5-alpha competent cells, grown overnight, and isolated using the Wizard plus SV Minipreps DNA Purification System (Promega).

Donor plasmids were linearized using a restriction enzyme on the 3' end and used for microinjections to avoid incorporating the pGEM T-easy Vector to the genome. Linearized plasmid was purified using a phenol-chloroform extraction method and used as donor DNA.

Separate NHEJ and MMEJ strategies were also used for the knock-in microinjections. For NHEJ, PCR of plasmid constructs were performed with forward primers starting at the cut site with a mutated PAM site or mutated guide sequence and reverse primers at the end of the construct (mRuby2-Rev, FLAG-4C-Rev). PCR product was purified with phenol-chloroform extraction and

used as donor DNA. For MMEJ, PCR of plasmid constructs were performed with primers used for NHEJ but with additions of a 48-bp homology arm on each end.

2.11 CRISPR/Cas9 genome editing

sgRNA synthesis was prepared by oligo annealing and PCR amplification of the sgRNA-specific forward primer, a universal reverse primer, and the Phusion DNA Polymerase (Thermo Fisher Scientific). The PCR products were checked on a gel and then purified using PCR Column Purification Kit (Qiagen). 150 ng of DNA was used as a template in an in vitro transcription reaction with the Megashortscript T7 kit (Thermo Fisher Scientific). The reactions were incubated for 5 hours at 37°C, digested with TURBO DNase (Invitrogen), extracted with phenol/chloroform, and precipitated with ammonium acetate and ethanol. The pellet was resuspended in nuclease-free water and the concentration was quantified by Qubit. The Cas9 was provided as recombinant protein (PNA Bio Inc) resuspended in nuclease-free water at a concentration of 1 μg/μL and stored in 2 μL aliquots in -80°C.

sgRNAs were tested using an in vitro digestion assay. sgRNAs were pre-incubated at 25°C with NEBuffer r3.1, nuclease-free water, and Cas9 Nuclease before the addition of plasmid DNA containing the target sequence. The reaction was incubated at 37°C for 15 minutes followed by addition of Proteinase K and incubation at room temperature for 10 minutes. Samples were run on a gel alongside intact plasmid DNA and analyzed to see if the Cas9 was able to cut the substrate DNA.

For microinjections, a mixture of 500 ng/μL Cas9 protein and 250 ng/μL sgRNA was incubated at 37°C for 7 minutes (Kao *et al.*, 2016). The mixture was then transferred on ice and the inert dye phenol red (5X Sigma Aldrich) was added. For knock-in experiments, purified donor DNA was also added at a concentration of 10 ng/μL.

For knock-out experiments, embryos were imaged and scored from 1 to 7 days post fertilization (dpf) under a stereomicroscope.

Chapter 3

Identification of genes important in early cell fate specification

3.1. Introduction

3.1.1. Early cell fate specification in the *Parhyale* system

Cell fate specification is an important developmental process that requires asymmetric distribution of mRNAs, proteins, and morphogens to establish the different germ layers. Previous work to understand how cell lineages specify utilized lineage tracing, perturbations (such as knock-downs and knock-outs), as well as -omic technologies. Lineage tracing allows us to use dyes to follow the clones of injected cells in order to map the fate of blastomeres (Kretzchmar and Watt, 2012). Initial work using lineage tracing allowed for the entire mapping of the *C. elegans* embryo to understand when cells divide, what they become, and when they go through apoptosis (Sulston *et al.*, 1983). By combining lineage tracing with other methods such as pharmacologic perturbations, transgenic or mutant animals, and micromanipulations, we can begin to understand the mechanisms that drive lineage specification (Wolf *et al.*, 2021; Shah *et al.*, 2019). Recently, lineage tracing has been coupled with -omic technologies such as single-cell RNA sequencing, allowing for a more comprehensive understanding of how different gene expression signatures emerge in different lineages (Wagner and Klein, 2020; VanHorn and Morris, 2021; Kester and van Oudenaarden, 2018). Many of these methods separately and together have aided in our understanding of how germ layers specify overtime.

3.1.2. RNA sequencing and its applications for developmental biology research

For more than a decade, RNA sequencing has become a technology used in many labs and in many different fields of study. RNA-seq is used to understand the transcriptomic profile of an organism and allows for comparisons of transcriptomes using differential gene expression analysis (Rapaport *et al.*, 2013). In order to sequence the transcriptome, RNA needs to be extracted from

live or fixed tissue. Next, polyadenylation enrichment is required if looking at mature mRNAs. cDNA is then synthesized and an adaptor is added to create the mRNA library. Finally, the library is sequenced using a high-throughput platform to result in 20-50 million raw reads. After receiving the raw data, pre-processing such as trimming is required to clean up the data. The sequence reads are then aligned to a reference genome or aligned *de novo* without a reference genome. The number of transcripts is then counted and normalized to the total number of reads. Finally, this data can then be used for further analysis.

In addition to traditional bulk RNA sequencing, single-cell RNA sequencing (scRNA-seq) has become more and more widely used as it provides a way to understand gene expression at a resolution of a single cell. This is crucial to decipher gene expression differences in cell populations, and has allowed us to understand heterogeneity within tissue samples. However, there is less coverage than bulk RNA-seq methods because of the failure to detect specific transcripts and the potential of variable transcriptional activity within single cells (Haque *et al.*, 2017). In addition, scRNA-seq data is more prone to technical variation, resulting in noisier data (Haque *et al.*, 2017). When comparing cell samples in a specific population to determine small changes resulting in heterogeneity, the variation, noise, and lower coverage of scRNA-seq methods are challenges that make analyzing scRNA-seq data difficult. When working with scRNA-seq technology in development, the variation in cell size also causes difficulties in analyzing the data (Haque *et al.*, 2017).

3.1.3. RNA sequencing in the *Parhyale* model

Previous work in *Parhyale* has been done to understand the transcriptome as the embryo develops from zygote to hatchling stages (Zeng *et al.*, 2011; Blythe *et al.*, 2012; Nestorov *et al.*, 2013; Kao *et al.*, 2016; Calvo *et al.*, 2022). Initial RNA sequencing experiments used a *de novo* assembly

approach to provide a foundation for further analysis and direction for transcriptional regulation during development (Zeng *et al.*, 2011; Blythe *et al.*, 2012). However, since the assembly and annotation of the *Parhyale* genome, more studies have been done to decipher the importance of maternal mRNAs and micro-RNAs (miRNAs) during maternal-to-zygotic transition (Kao *et al*., 2016; Calvo *et al.*, 2022). Although these studies have provided a resource for the study of certain processes such as maternal-to-zygotic transition, they lack the spatial information which can help us understand how different lineages are formed. Nestorov *et al.* (2013) provides a method to begin understanding how cell fates are specified early on in the 8-cell embryo through the use of microarray analysis coupled with microdissections of different parts of the embryo. They compare the maternal transcriptome of ectodermal blastomeres at the 8-cell stage to the rest of the embryo and find that there are asymmetrically distributed mRNAs present already after the third cleavage. This is particularly interesting because lineage tracing experiments have shown that there a restriction to cell fates as early as the 8-cell stage (Gerberding *et al.*, 2002). Although they find that the eight clones have distinct cell lineages, it is unclear which germ layers certain blastomeres (like the Mav and en blastomeres) commit towards.

In order to further our understanding of how cell fates are specified early, single blastomere transcriptomes are necessary to shed light on the important molecular players in this process. Although the microarray study provides an initial look into the asymmetric distribution of mRNAs, it does not study cell fate specification at the single blastomere level and microarray data comes with limitations due to the bias in choosing which sequences to design probes for. Therefore, there is a need in the *Parhyale* system for transcriptomic datasets of good quality and coverage that includes spatial information at the single blastomere level as well.

3.2. Aim of this chapter

Transcriptomics have played an immense role in our understanding of gene regulation at the whole organism, tissue-specific, and single cell level. Although there are pipelines and technologies that have been established to perform RNA-sequencing experiments on various organisms, obtaining a cell-specific transcriptome in a non-traditional model remains challenging as it requires different methods to create single-cell suspensions, might not be compatible with commercial options such as 10X genomics, and might not have a large enough number of cells during early embryonic development for droplet-based methods. In this chapter, I present a method to dissect, fix, and extract the RNA of distinct blastomeres of early stage *Parhyale* embryos in order to perform bulk RNA sequencing on specific cell types. This method departs from single-cell transcriptomics as it does not sequence individual cells but rather a pool of cells with the same cell identity. By using bulk RNA-sequencing, we are able to obtain transcriptomes with better depth and coverage compared to single-cell RNA-seq while still obtaining the cell-specific information. The aim of the RNA-sequencing of blastomeres is to provide transcriptomic datasets of each blastomere at the 4- and 8-cell embryo for future analysis on cell fate specification, maternal mRNA degradation, and to determine candidates for further study. In addition to making the transcriptomic datasets, I validated them by *in situ* hybridization of various mRNAs involved in cell fate specification based on the literature as well as *armadillo*, a top g-cell specific gene candidate.

3.3. Results and Discussion

3.3.1. Separation of individual blastomeres at the 4- and 8-cell stage

Dissection of 4- and 8- cell *Parhyale* embryos occurred manually with the use of No. 5 forceps and tungsten wires. The embryos were placed in 3.7% PFA for at least 20 minutes before dissection occurred. This moment fixation occurs is important as creating single cell suspensions before fixation may result in a stress response, affecting the transcriptome of the sequenced sample (Denisenko *et al.*, 2020). Therefore, by dissecting in 3.7% PFA, I was able to fix the transcriptome early before a stress response could occur. As described in Fig. 3.1A, the eggshell was removed with forceps to reveal the embryo. The embryo was oriented in a way where the Mavg or Mav and g cells were facing the bottom left corner (Fig 3.1 B,C,E,F). This is important to determine the blastomere identity of the other cells after separation as the blastomeres look similar after dissected. Using forceps, slight pressure was the applied to the embryo which resulted in the blastomeres coming loose. The tungsten wires were then used to aid in the full separation of the blastomeres from one another (Fig 3.1 D,G). Each blastomere was then transferred to a 1.5 mL tube for that cell identity and kept on ice until the end of the dissections. These samples were then washed and stored at -20 °C until RNA extraction using a combination of the Qiagen RNeasy Micro kit and the Qiagen RNeasy FFPE kit. I dissected a total of 300 4-cell embryos and 1150 8-cell embryos before RNA extraction.

To test whether the nuclei in the single cell suspensions were intact, I used DAPI to stain for the nuclei of single blastomeres and observe that the nuclei are still intact (Fig 3.1 I). Since the nuclei of the blastomeres are intact, the integrity of the cell membrane remained intact as well, and the global RNA amount of the pooled samples corresponded with the relative size of the blastomeres

(Table 3.1), I was confident in sending the samples for RNA sequencing by Novogene using the Illumina platform.

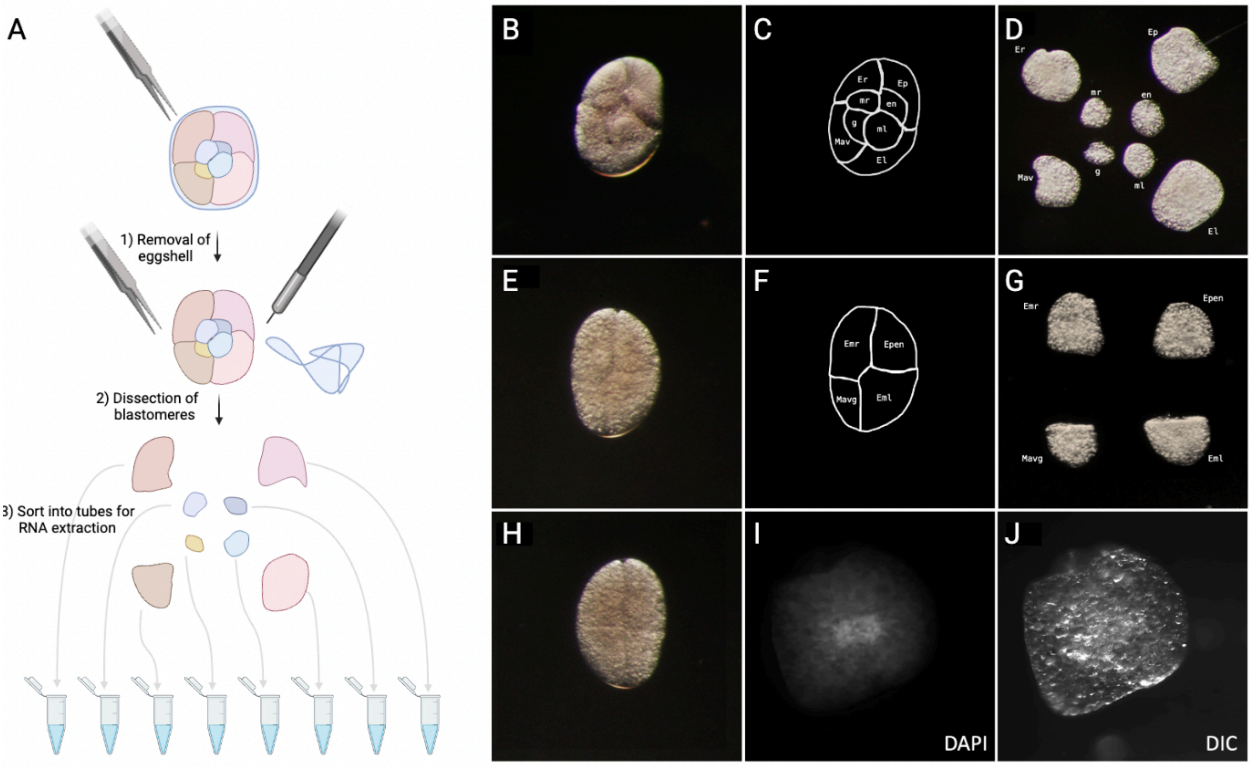


Fig 3.1: Overview of blastomere separation of 8- and 4-cell embryos. (A) Schematic of blastomere separation of 8-cell embryos (B) Vegetal view of 8-cell embryo fixed in paraformaldehyde. (C) Nomenclature of blastomeres projected on embryo of (B). (D) Separated blastomeres of fixed 8-cell embryo. (E) Vegetal view of 4-cell embryo fixed in paraformaldehyde. (F) Nomenclature of blastomeres project on embryo of (E). (G) Separated blastomeres of fixed 4-cell embryo. (H) Animal view of 4-cell embryo fixed in paraformaldehyde. (I) DAPI staining of blastomere showing intact nucleus of (J). (J) DIC image of individual blastomere.

Table 3.1: RNA quantity of each pooled blastomere sample sent to Novogene

Blastomere	**Total RNA amount(ng)**
Mavg	207.95
Emr	340.78
Eml	243.43
Epen	149.98
Mav	729.0
Er	146.0
El	145.8
Ep	141.3
g	126.0
mr	208.0
ml	240.0
en	171.0

3.3.2. Analysis of RNA sequencing of 4- and 8-cell stage blastomeres

After receiving the raw data of the 4- and 8- cell blastomere samples, I trimmed the data using Trimmomatic V0.39 to result in clean FASTQ reads by filtering out poor quality reads from the data. Next, I mapped the clean reads to the *Parhyale* reference genome. The lowest number of clean reads after trimming the data was from the en blastomere sample with 19464771 clean reads (Fig 3.2A). The mapping rates for all the samples to the *Parhyale* genome were approximately 80% (Fig 3.2B).

I performed principal component analysis (PCA) to show the similarities and differences in the gene expression signatures of the blastomeres of the 4- and 8- cell stages. Principal component analysis is a statistical technique used to display a pattern of similarity between different sample observations (Abdi and Williams, 2010). When looking at the 8-cell blastomeres, the PCA plot shows clustering of the ectodermal blastomeres (El, Er, Ep) at the bottom left of the plot (Fig 3.2 D). Interestingly, the three blastomeres at the 4-cell stage with ectodermal lineage also cluster near the 8-cell ectodermal blastomeres but form a distinct group. This suggests that the ectoderm is specified early on and that it is quite distinct from the rest of the lineages. When the 4-cell embryo

divides to become the 8-cell embryo, there is high asymmetry in the blastomeres which results in daughter cells with very distinct gene expression signatures. In addition, the Mavg blastomere of the 4-cell stage is distinct from the rest of the embryo, possible due to the fact it is the only blastomere at the 4-cell stage without any ectodermal lineage (Fig 3.2D). When it divides, it creates daughter cells g and Mav (Fig 3.2C). According to the PCA plot, Mav is more similar to its parent cell Mavg then g is to its parent cell. This is similar to how the ectodermal macromeres of the 8-cell embryo are more similar to their parent cell than their sister micromeres are to the parent. This suggests that when the *Parhyale* embryo divides and creates cells of varying sized, the larger macromeres tend to retain the transcriptional signature of its parent while the micromeres begin to deviate to result in new germ layer identities.

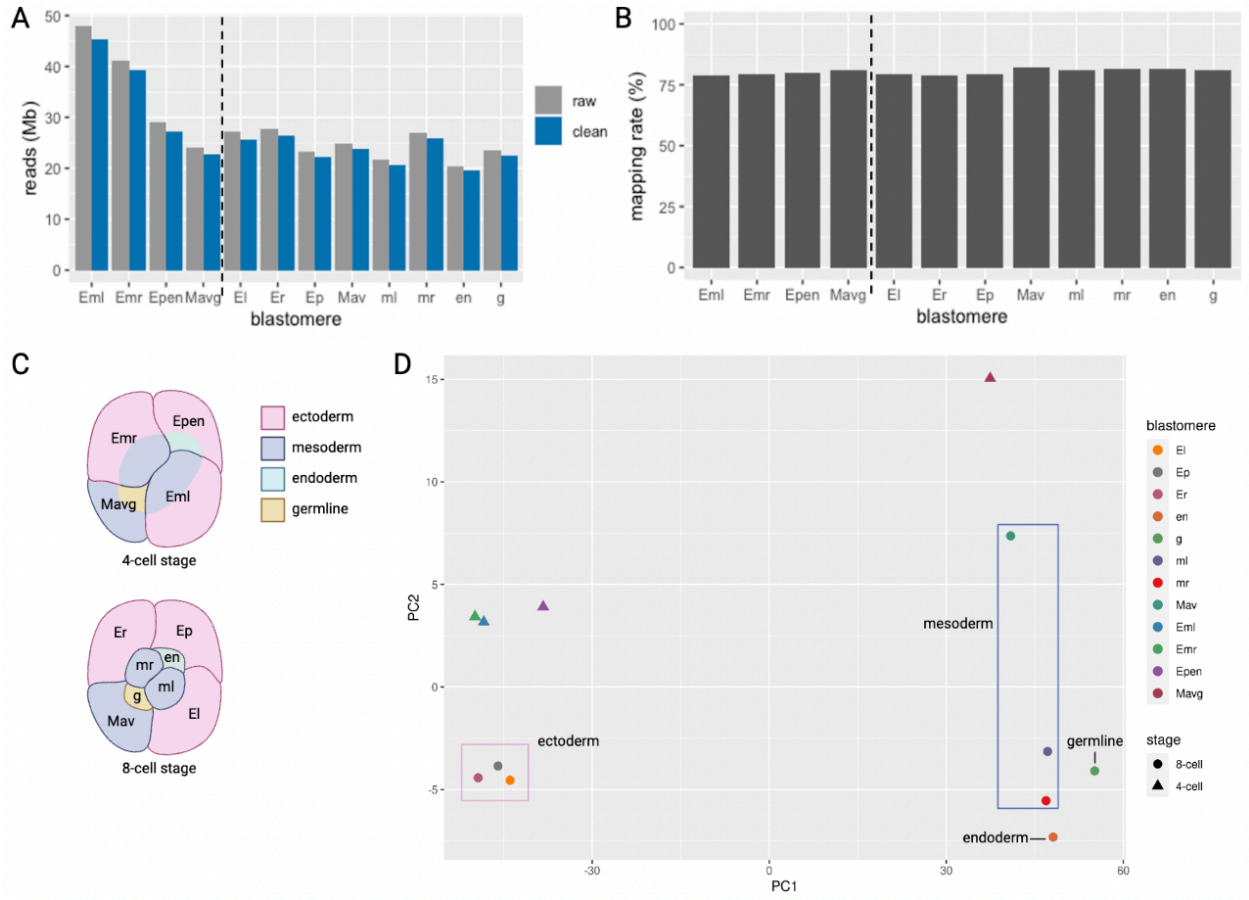


Fig 3.2: Quality check and analysis of RNA-seq data. (A) Abundance of sequence reads per blastomere, total sequence reads (grey), clean reads after trimming (blue). (B) Mapping rate of sequence reads to annotated genome per blastomere. (C) Schematic of 4- and 8-cell embryos colored by germ layer. (D) Principal component analysis of 12 RNA-seq datasets showing ectodermal blastomeres clustering together. PC1 explains 93% of the variance between the datasets. PC2 explains 2% of the variance between the datasets.

3.3.3. Validation of transcriptomic datasets

In order to validate the transcriptomic datasets, I performed in situ hybridization on 8-cell embryos with probes for genes that were thought to be important in cell fate specification based on literature in other model systems such as *Drosophila* and *C. elegans* (Fig. 3.3). In *Drosophila*, *vasa* is a gene encoding an RNA-binding protein important in the development of the germ cell (Raz, 2000). Therefore, it is a potential candidate for g cell specific localization. Özhan-Kizil et al., (2008) show that *vasa* is expressed ubiquitously in the 8-cell embryo but is later degraded in non-germ line lineages. Using *vasa* as a positive control for the *in situ* hybridizations performed, I designed a probe targeting *vasa* and show similar results in that *vasa* is ubiquitously expressed in the 8-cell embryo (Fig 3.3A). In *C. elegans*, *mex3* is a maternal protein that is localized to anterior fated AB cell after the first cell divisin, and as an RNA binding protein represses translation of key posterior determinant pal-1/caudal (Bowerman et al., 1997). Although *mex3* is asymmetrically distributed through the action of PAR proteins in the *C. elegans* zygotes, it is expressed strongly in all blastomeres at the 8-cell stage in *Parhyale* (Fig. 3.3B). *Argonaute2* is a PIWI-family protein involved in forming RNA-induced silencing complexes with small interfering RNAs and has a potential role in germ line development (Okamura et al., 2004). Although *argonaute2* is expected to be expressed in the g cell, ISH shows that *argonaute2* is expressed ubiquitously at the 8-cell stage (Fig. 3.3C). *Snail* is a transcription factor involved in the development of the mesoderm and the process of epithelial-mesenchymal transition (Rembold et al., 2014). Based on the literature, we expect to see *snail* expressed in Mav, mr, and/or ml blastomeres. However, ISH shows that

snail expression is weak throughout the 8-cell embryo (Fig. 3.3D). *Hunchback* is a transcription factor involved in the establishment of anterior-posterior polarity and segmentation in the *Drosophila* embryo (Lehmann and Nüsslein-Volhard, 1987). As it is a maternally-expressed gene involved in polarity, we expected it to be asymmetrically distributed in the *Parhyale* embryo as well. However, ISH shows that *hunchback* is expressed in all blastomeres with slightly stronger expression in the macromeres (Fig. 3.3E). *Pumilio* is a gene coding an RNA-binding protein and acts as a translational regulator of mRNAs by binding to their 3'UTR. I show that *pumilio* is expressed ubiquitously as well with elevated expression in the macromeres (Fig. 3.3F). Although the candidate genes picked were thought to be potentially asymmetrically distributed in the *Parhyale* embryo but were all ubiquitously expressed, these *in situ* hybridization experiments still help validate the transcriptomic datasets that were produced. All these genes except *snail* had transcripts present in every blastomere at varying levels. There were no transcripts present for *snail* in any of the eight transcriptomic datasets for the 8-cell stage. This could be attributed to the very low expression of *snail* compared to other genes. Regardless, none of these candidate genes had much higher expression in a single cell or subset of cells (Table 3.2). This reflects the ISH experiments as all these candidate genes are ubiquitously expressed.

One reason why these genes that are thought to be asymmetrically distributed are actually ubiquitously expressed may be because of later specification past the 8-cell stage. Özhan-Kizil et al. (2008) describe that *vasa* is asymmetrically distributed from the 16-cell onwards and is important in the maintenance rather than the formation of the germ cell. In addition, *vasa* mRNA was shown to be degraded overtime in non-g cell derived blastomeres (Özhan-Kizil *et al.*, 2008). A similar mechanism of initial ubiquitous expression then followed by targeted degradation may occur for these genes as well.

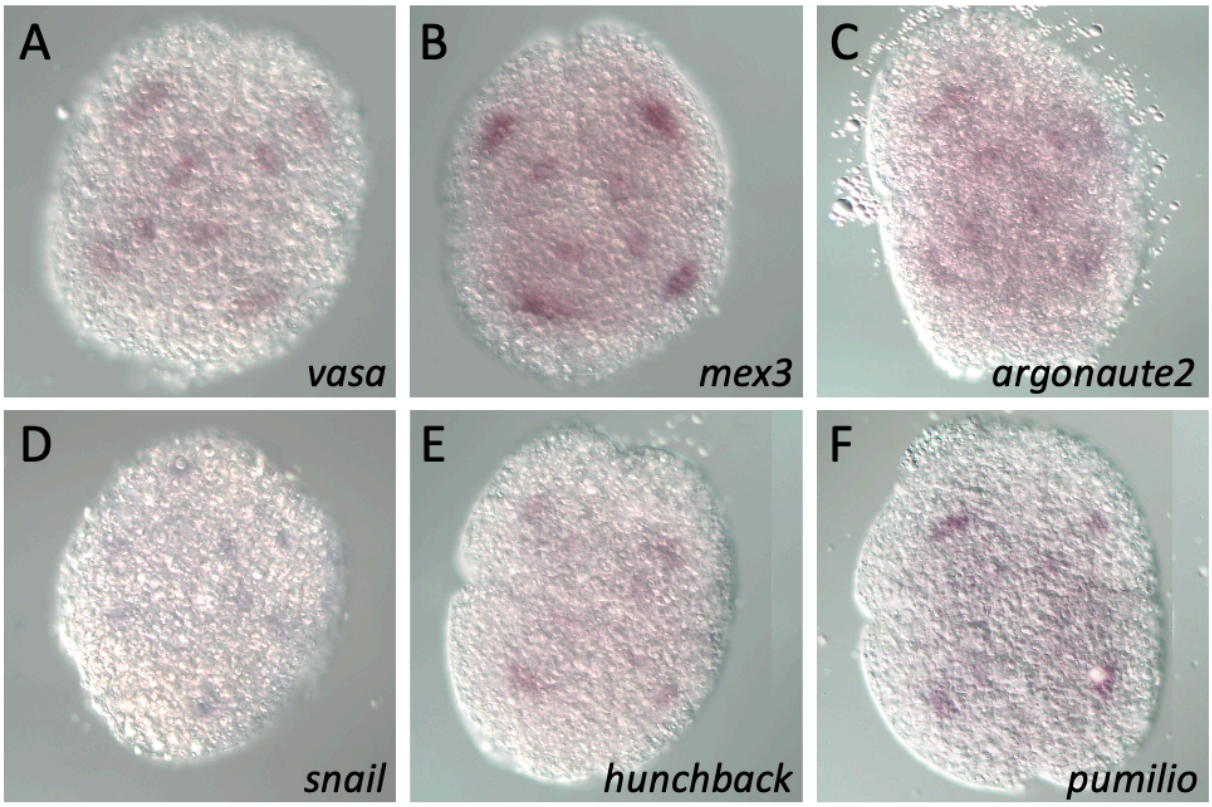


Fig 3.3: Whole mount *in situ* hybridization analysis of mRNAs with potential asymmetric distribution based on literature on cell fate specification and RNA-binding proteins. (A) *Vasa,* (B) *mex3,* and (C) *argonaute2* show strong signal in all blastomeres. (D) *Snail* shows weak expression in all blastomeres. (E) *Hunchback* and (F) *pumilio* show signal in all blastomeres with slightly stronger expression in the macromeres.

Table 3.2: Transcripts per million (TPM) of ubiquitously-expressed genes by blastomere

Gene	Gene ID	El	Er	Ep	Mav	ml	mr	en	g
Mex3	MSTRG.64035	279.45	261.33	307.63	356.57	224.83	221.24	154.23	213.83
Pumilio	MSTRG.22416	22.63	24.58	25.85	10.68	8.23	10.11	8.46	10.26
Vasa	MSTRG.26077	27.59	29.13	29.63	82.58	113.93	91.42	105.85	34.41
Argonaute2	MSTRG.32662	18.49	19.95	19.11	8.48	8.15	6.52	7.23	6.83
Hunchback	MSTRG.22414	25.98	28.56	18.79	24.84	44.76	46.55	35.18	37.47

Using the transcriptomic datasets, I compared the transcripts per million (TPMs) of each blastomere to the others to determine the fold change between blastomeres. In doing so, I was able to generate lists top candidate genes that are expressed in certain blastomeres or certain germ layers. One of the top candidates when looking at the TPM fold change between the g cell and the

other blastomeres is the *wnt*-signaling molecule, Armadillo, which is the *Drosophila* homolog of beta-catenin. A*rmadillo* plays roles in Wingless signaling as well as cell adhesion, particularly in the development of the central nervous system (Loureiro and Peifer, 1998). This is interesting as *armadillo* has never been associated with the development of the germ line before. When performing *in situ* hybridization using a probe designed for *armadillo*, we see that there is expression of *armadillo* in the g cell (Fig. 3.4A). This reflects what we see from the transcriptomic datasets as the TPM fold change between the g cell and the rest of the blastomeres range from 3.49 to 27.47 (Fig 3.4B). Although there is some *armadillo* expression throughout the 8-cell embryo as shown by the transcriptomic datasets, the expression of *armadillo* based on the number of raw reads is much higher in the g cell (Fig 3.4C). The discrepancies between the TPM fold change and the coverage plots is due to the normalization of the counts to calculate the TPM fold change while the coverage plots were made using raw read counts (Fig 3.4B,C). It is particularly interesting that the *armadillo* TPM fold change between the g cell and the ectodermal blastomeres is extremely high (17.36-27.47) since the g cell is shown to be dissimilar from El, Er, and Ep based on the PCA plot (Fig 3.2D). All in all, the *in situ* hybridization experiment of *armadillo* provides evidence to support that the transcriptomic datasets reflect actual mRNA expression and shows that the candidates we pick based on the RNA-seq data are asymmetrically expressed.

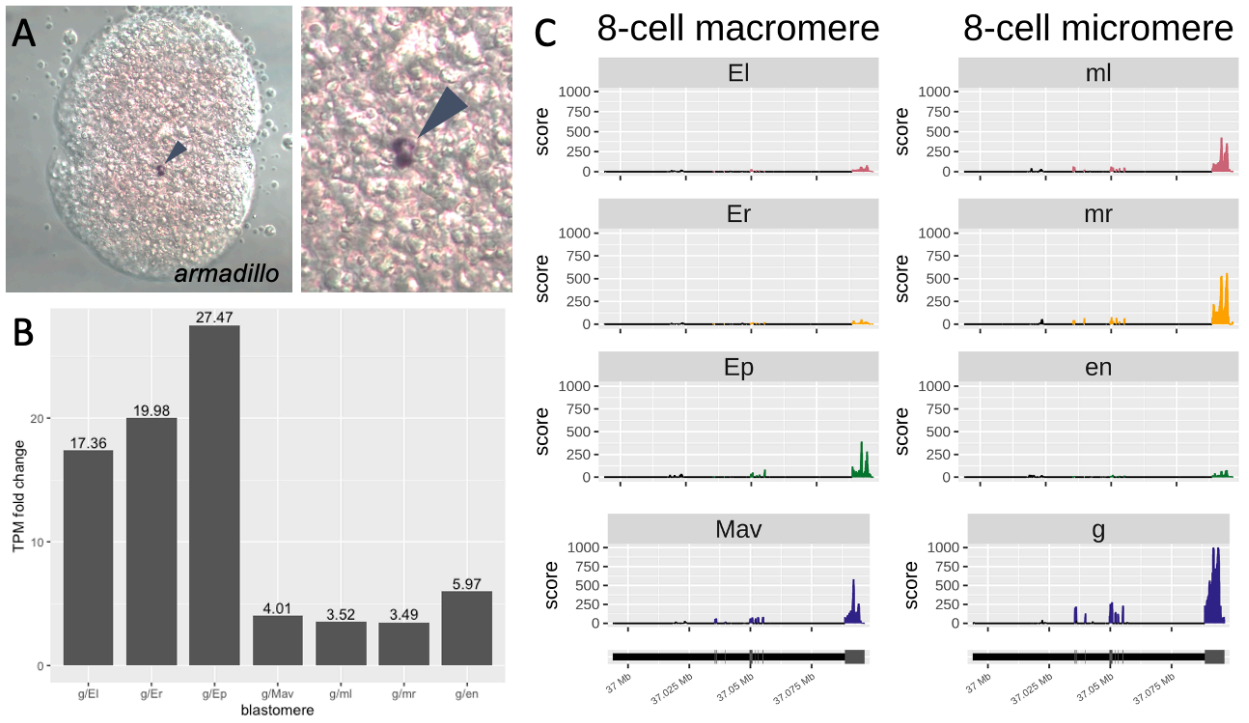


Fig 3.4: Gene expression of *armadillo* in the 8-cell embryo. (A) Whole mount *in situ* hybridization showing that *armadillo* is strongly expressed in the g cell perinuclear cytoplasm. (B) Fold change between g cell and other blastomeres based on normalized transcripts per million. (C) Coverage plots of *armadillo* in the different blastomeres based on raw read counts.

3.4. Conclusion

In this study, I present a method to effectively fix, dissect, and separate blastomeres of early Parhyale embryos. I show that the membranes and nuclei of blastomeres remain intact and that we are able to pool blastomeres of the same type together for RNA extraction and sequencing. Based on the number of clean reads and the mapping rate to the Parhyale genome, I confirm that the RNA was of good quality and the transcriptomic datasets have good coverage. Initial analysis of the transcriptomic datasets by principal component analysis found that the ectoderm is specified early on in the embryo and that the macromeres resemble their parent cell more so than the micromeres as the micromeres have very different gene expression signatures. I validated the transcriptomic datasets by showing that many ubiquitously expressed genes based on the transcriptomic datasets are also ubiquitously expressed when I performed in situ hybridization of those genes on 8-cell embryos. In addition, candidate gene lists based on the pairwise comparisons of TPM fold change were produced and I performed in situ hybridization on armadillo, a gene with asymmetric distribution to the g cell. I confirm that armadillo is expressed highly in the g cell, thus validating the transcriptomic datasets again. Ultimately, these transcriptomic datasets provide a resource for further analysis and selection of candidate genes involved in early cell fate specification and maternal mRNA degradation.

Chapter 4

Development of CRISPR/Cas9 strategies in *Parhyale hawaiensis* to study maternally expressed genes

4.1. Introduction

Since the establishment of *Parhyale* as a model system for developmental biology and genomics, many genetic manipulation techniques have been applied to this system. Pavlopoulos *et al*. (2014) generated a heat-inducible transgenic line to misexpress *Ultrabithorax*. In addition, morpholinos and RNA interference have provided methods to knockdown genes of interest, enabling us to understand gene function (Liubcich *et al.*, 2009; Martin *et al.*, 2015; Ozhan-Kizil *et al.*, 2017). Many techniques have also been used to allow us to understand gene expression at the mRNA and protein levels. *In situ* hybridization and hybridization chain reaction (HCR) have been applied to *Parhyale* to visualize mRNA expression within the embryo (Rehm *et al.*, 2009; Bruce *et al.*, 2021). Immunostaining has been used to visualize protein localization in the embryo (Rehm *et al.*, 2009). Although these knockdown and gene expression methods have allowed us to increase our understanding on homeobox genes like *Ubx*, RNA-binding proteins like *Vasa*, and other genes of interest, these techniques have downsides. Morpholino and RNAi knockdown methods only result in moderate decreases of target mRNAs and may have off-target effects. In addition, protein visualization methods like immunostaining can be time intensive and costly as many commercial antibodies are not cross-reactive with models like *Parhyale* that are less worked with.

As an alternative to previous knockdown and gene expression methods, CRISPR/Cas9 provides a way to knockout a gene as well as to visualize protein localization in live embryos. *Parhyale* has already been studied using both CRISPR/Cas9 loss-of-function knockout methods (Martin *et al.*, 2016; Kao *et al.*, 2016; Sun *et al.*, 2022; Alberstat *et al.*, 2022) and CRISPR/Cas9 protein-tagging knock-in methods (Kao *et al.*, 2016; Serano *et al.*, 2016). Although CRISPR methods have been established in *Parhyale*, the penetrance for knock-out phenotypes (9 – 76%) and the expression of

tagged fluorescent proteins (7%) is quite variable and low. Therefore, development of more efficient CRISPR/Cas9 methods will aid in the study of gene function and localization.

4.1.1. CRISPR/Cas9 strategies

CRISPR/Cas9 can generate non-functional proteins and/or insert epitope tags by leveraging two main DNA repair pathways: non-homologous end-joining (NHEJ) and homology-directed repair (HDR). When Cas9 is delivered with sgRNAs into a cell, the sgRNAs allow for precise targeting to the correct site of the DNA which generally consists of 20 nucleotides followed by a PAM sequence (NGG). This is because the sgRNA consists of two components: one to create a scaffold for Cas9 binding, and the other which is a complementary sequence to the target DNA sequence. The Cas9 protein will cause a double-stranded break (DSB) 3-4 nucleotides upstream of the PAM site. The DSBs will activate DNA repair mechanisms to fix the break in the DNA. NHEJ is an error-prone mechanism that may result in indels when reattaching the two sections of DNA. HDR is a more precise mechanism that requires donor DNA flanked by DNA segments that are homologous to the ends adjacent to the cut site.

Microhomology-mediated end-joining (MMEJ) is another DNA repair mechanism that acts as an intermediate between NHEJ and HDR. It is similar to NHEJ in that it is an error-prone mechanism and may still result in indels when reattached DNA. However, it relies on a microhomologous sequence flanking the cut site to repair the break. In short, the BRCA1-MRN-CtIP complex has been shown to help resect the ends of the cut DNA (Chen *et al.*, 2008). This allows the microhomologous sections to be exposed, allowing for annealing of the two strands. The flaps are removed, and DNA polymerase fills in the gaps until the DNA is ligated back together (Wang and Xu, 2017). As DNA can be repaired using microhomologous strands of DNA, researchers have

utilized this method with CRISPR knock-ins as an alternative to HDR as HDR requires a much longer donor DNA template (Sakuma *et al.*, 2015; Vu *et al.*, 2021).

4.1.2. CRISPR/Cas9 knock-in applications

CRISPR/Cas9 technology has been adapted to allow for epitope tagging of endogenous DNA for multiple applications. One basic approach for CRISPR/Cas9 knock-in systems is to tag an endogenous protein with a fluorescent protein to allow us to visualize protein expression (Matsuda and Oinuma, 2019; Sharma *et al.*, 2018; Zhong *et al.*, 2021). This is often attached to a nuclear localization signal sequence so the fluorescent protein will aggregate in the nucleus to allow for easier identification of fluorescent-positive cells (Kitamura *et al.*, 2015). There are multiple ways to design donor DNA for fluorescent tagging of endogenous genes. First, a fusion protein can be made using CRISPR/Cas9 by attaching the fluorescent tag in the N-terminal or C-terminal end of the endogenous gene (Geny *et al.*, 2020). The fluorescent tag must be in frame with the protein coding sequence and can be separated by a linker sequence. Another option is to knock-out the gene via CRISPR by disrupting an allele while inserting a coding sequence attached to a fluorescent tag (Kao *et al.*, 2016). This will result in a rescue phenotype which should be similar to the wild-type phenotype but with the tagged protein. Finally, one way in which to minimize the disruption to protein folding of the endogenous gene is to use a 2A self-cleaving peptide such as P2A or T2A to attach the fluorescent protein (Oikawa *et al.*, 2022; Botman *et al.*, 2019). The 2A self-cleaving peptide can attach the fluorescent protein sequence to the gene of interest while allowing both the fluorescent protein and the endogenous protein to be functional as there is no attached peptide that may affect the folding of the protein.

Another technology that can be implemented using CRISPR/Cas9 is the FlAsH/ReAsH labeling system (Irtegun *et al.*, 2011). This system involves a short tetracysteine motif (CCXXCC) inserted

into a target protein. The tetracysteine binds permanently to biarsenical dyes, FlAsH (green fluorescent) and ReAsH (red fluorescent) *in vivo* allowing for visualization of protein expression. The ability to use two dyes also allows for temporal understanding of protein synthesis by treating cells with one dye first then treating later with the other dye to visualize newly synthesized protein. Finally, this technology also allows for photoinactivation of proteins tagged with the tetracysteine motif (Venken *et al.*, 2008).

Another application of CRISPR/Cas9 is its implementation with chromatin immunoprecipitation followed by next-generation DNA sequencing (ChIP-seq). ChIP-seq is a technology that allows us to determine what certain transcription factors bind to throughout an entire genome. Although ChIP-seq has been widely used in many traditional model organisms, a limitation with applying this technology to a non-traditional model organism is that there are few commercial antibodies that are cross-reactive with these species. It is also very difficult to raise antibodies due to the time and money cost. Savic *et al.* (2015) have provided a method utilizing CRISPR to easily perform ChIP-seq on cell lines. This technology, CRISPR epitope-tagging ChIP-seq (CETCh-seq), tags DNA-binding proteins with a 3XFLAG epitope tag so commercial antibodies for 3XFLAG can be used for ChIP-seq of these transcription factors. CETCh-seq was shown to provide a robust way to perform ChIP-seq as there were strong correlations between using transcription factor antibodies and using antibodies for 3XFLAG (Savic *et al.*, 2015). Therefore, CRISPR/Cas9 systems have been implemented in the development of genomic technologies to better understand DNA-binding interactions.

4.2. Aim of this chapter

Previous work using the CRISPR/Cas9 system in *Parhyale* has been to knock-out genes as well as tagging proteins with fluorescent proteins. However, efficiency for doing so is variable and low. In addition, there have been many other applications and technologies that have been integrated with the CRISPR/Cas9 system. Therefore, there is room to implement these other technologies using CRISPR/Cas9 into the *Parhyale* model in a more efficient and precise way.

In this chapter, I want to use the flexibility of the CRISPR/Cas9 system and its many applications to tag the pioneer transcription factor *Zelda* with various epitope tags *in vivo*. I use *Zelda* as the study gene because transcriptomic analysis has shown that it is expressed both before and after zygotic genome activation (Jara Espejo, unpublished). In addition, *Zelda* plays an important role in *Drosophila* as a regulator of MZT and may have a conserved role in *Parhyale*. The aim of this chapter is to understand the role of *Zelda* through CRISPR/Cas9 knock-outs as well as develop and test alternative CRISPR/Cas9 knock-in methods in the *Parhyale* model to increase penetrance of epitope tags for use in protein visualization, integration of -omic technologies, and photoinactivation of tagged proteins.

4.3. Results and discussion

4.3.1. Design and testing of guide RNAs for CRISPR/Cas9 mutagenesis

Cloning and sequencing of full-length *Zelda* coding sequence (CDS) was performed to confirm the DNA sequence and to identify single-nucleotide polymorphisms (SNPs) in the gene. This is important when designing guide RNAs (sgRNAs) for CRISPR/Cas9 because areas with many SNPs should be avoided as the CRISPR/Cas9 system might not cut the endogenous gene if it contains a guide site with a SNP. Fig 4.1 shows the gene structure of *Zelda* based on transcriptomic data as well as the identified polymorphisms in the coding sequence. *Zelda* contains two exons of 803 bp and 16932 bp length. The 3078 bp CDS is located at the 5' end of the second exon. Based on transcriptome data and sequencing of four samples of *Zelda* CDS, there is a total of 43 SNPs in the coding sequence.

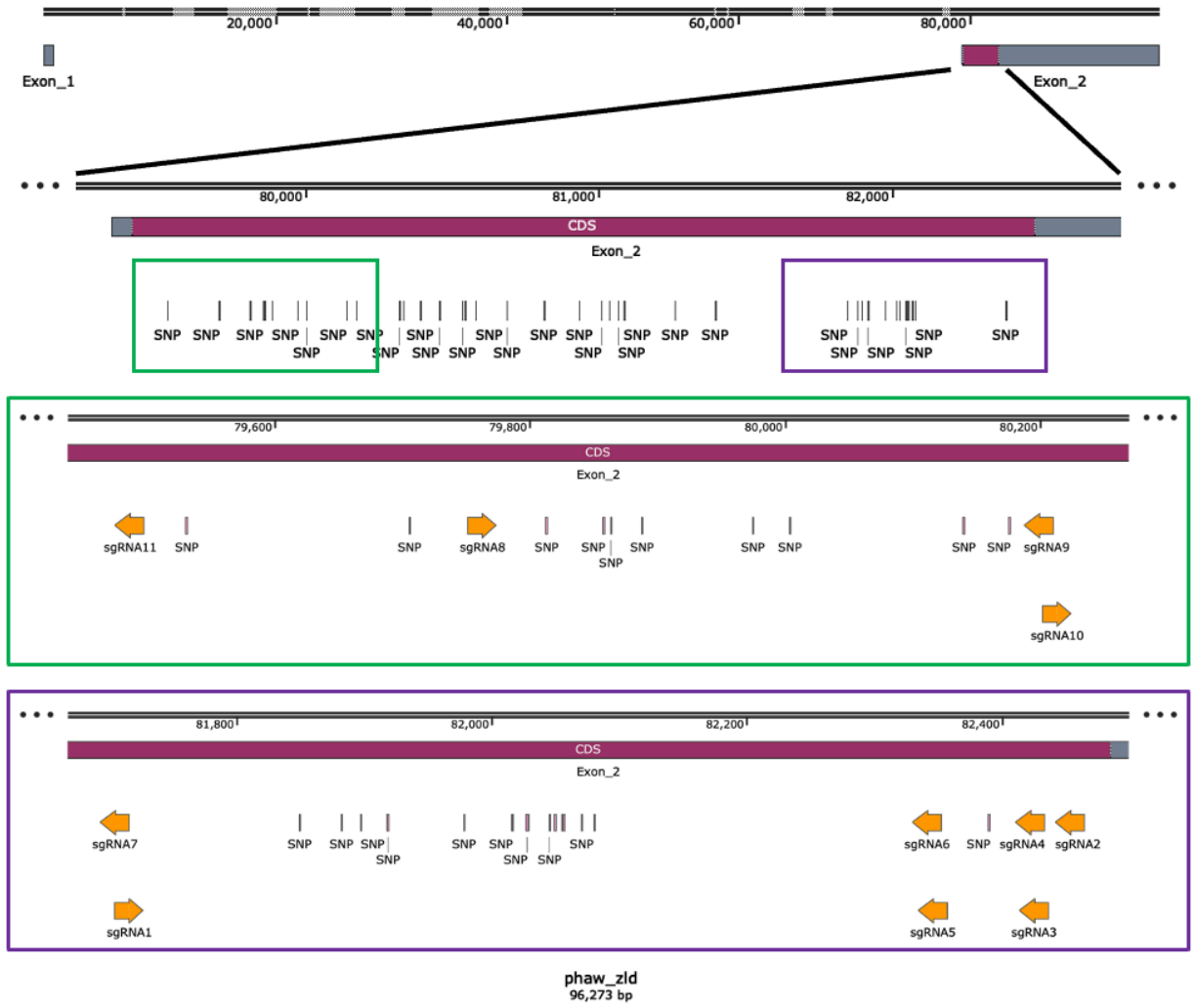


Fig 4.1: Gene structure of *Zelda* and location of 43 SNPs and 11 sgRNAs in the coding sequence

As displayed in Table 4.1, I designed 11 sgRNAs for CRISPR/Cas9 that avoided the regions containing SNPs. Seven of the sgRNAs (sgRNAs 1-7) were designed within 800 bp of the 3' end of the coding sequence for use in knock-in experiments to decrease the chance of creating nonfunctional protein. Four of the sgRNAs (sgRNAs 8-11) were designed within 850 bp of the 5' end of the coding sequence for use in knock-out experiments to increase the likelihood of frameshift mutations affecting the function of the protein.

Table 4.1: *Zelda* sgRNA sequences designed for CRISPR/Cas9 Knock-In and Knock-Out experiments

sgRNA	Direction	Sequence (Guide Site + PAM)
1	Forward	GGGCTCTATGCCAAAGGAACTGG
2	Reverse	GGAAGGGTGCTGAGACAAAGTGG
3	Reverse	GGAGGACCATTGTGATGAGGCGG
4	Reverse	GGACCATTGTGATGAGGCGGTGG
5	Reverse	GGATAGGGTTAGGCCCCATGTGG
6	Reverse	GGGTTAGGCCCCATGTGGGGAGG
7	Reverse	GGCATAGAGCCCGAGCCTAACGG
8	Forward	GGATATACACAACCAACCATTGG
9	Reverse	GGTTCTTCCTTTACAACACTGGG
10	Forward	GGAAGAACCGGTAATTAAGACGG
11	Reverse	GGCAGGTGAAGCTGATGATGAGG

The sgRNAs were synthesized by in-vitro transcription and tested using an in-vitro digestion assay with a DNA plasmid containing the full-length *Zelda* coding sequence (Fig 4.2). sgRNAs 1, 2, 3, 5, 7, 8, and 9 were shown to efficiently cut the *Zelda* CDS while the remaining sgRNAs did not cut the plasmid as efficiently, causing a smear in the gel (Fig 4.2). From this, I chose to use sgRNA8 for CRISPR/Cas9 knock-out experiments and sgRNA2 for CRISPR/Cas9 knock-in experiments.

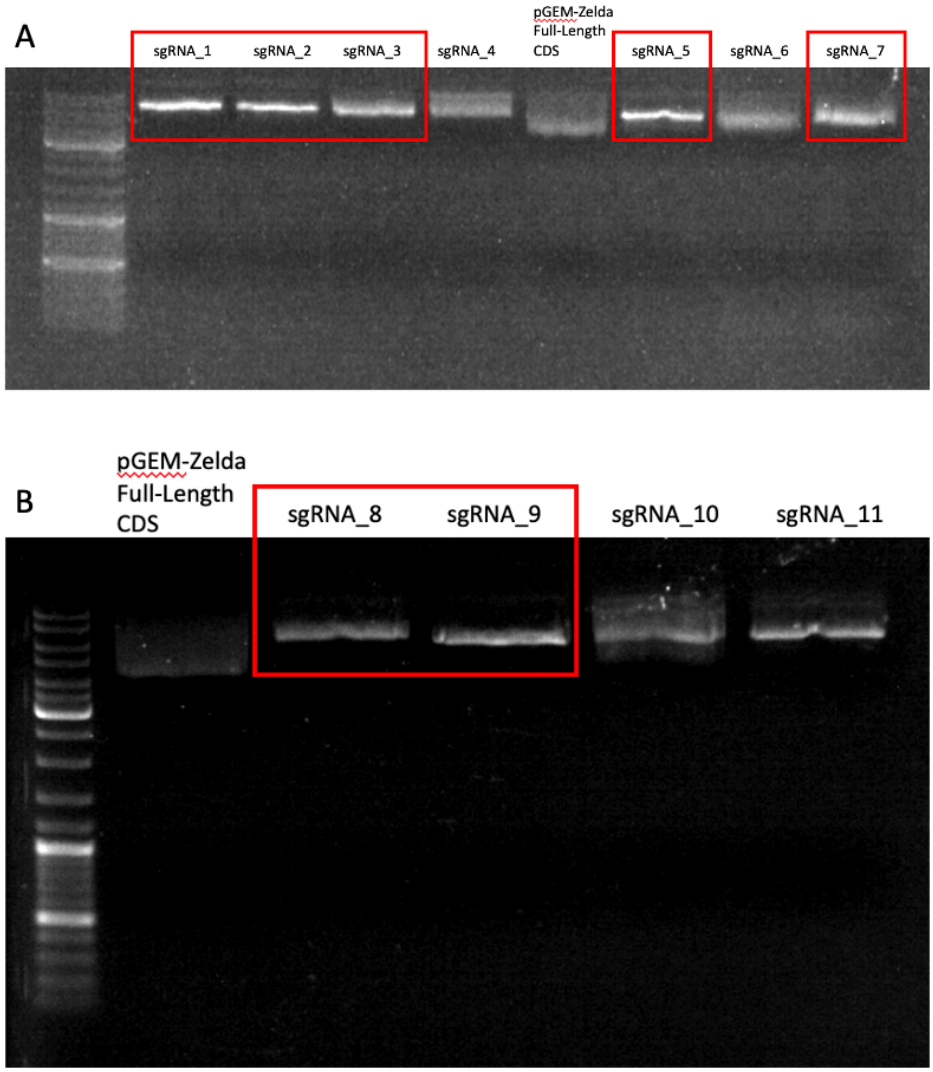


Fig 4.2: In Vitro Digestion Assay of Full-Length *Zelda* CDS in pGEM with sgRNAs designed for A) CRISPR/Cas9 Knock-Ins close to the 3' end of *Zelda* CDS and B) CRISPR/Cas9 Knock-Outs close to 5' end of *Zelda* CDS

4.3.2. Zelda loss-of-function knockout in 1-cell embryos using CRISPR/Cas9 shows significant decrease in survival after 5 days

To knock-out *Zelda*, I injected 1-cell *Parhyale* embryos with Cas9 protein (500 ng/uL) and sgRNA8 (250 ng/uL) to create indels at the cut site due to NHEJ machinery. As a control, I omitted Cas9 protein from the injection mix. I performed three batches of injections with at least 70 control

embryos and at least 200 experimental embryos each. Fig 4.3A shows representative embryo phenotypes of live and dead embryos for each day. Fig 4.3B shows the survival of embryos over time. At 5 dpf, there is a significant decrease in survival for the knock-out embryos. Since *Zelda* is expressed both maternally and after zygotic genome activation, and CRISPR/Cas9 system will only be able to knock-out zygotic *Zelda*, there may have been a delay in the death of embryos as maternal *Zelda* might need to be depleted for a phenotype to show up. From this data, we show that zygotically transcribed *Zelda* is required for embryo survival past 5 dpf.

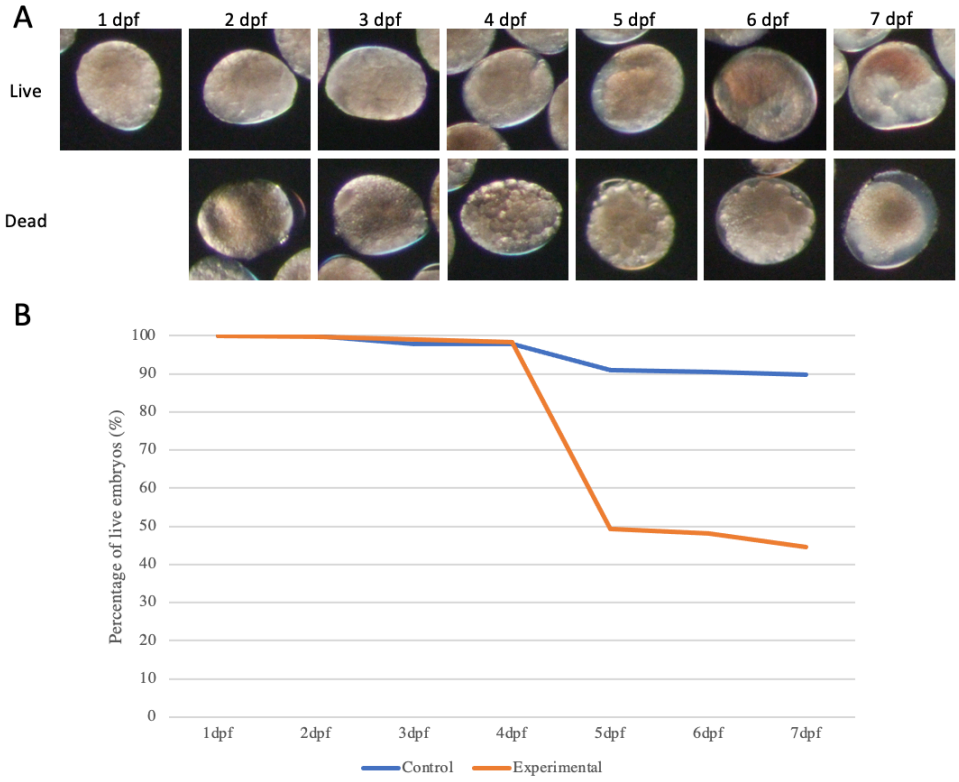


Fig 4.3: A) Phenotypes of live and dead *Parhyale* embryos over a seven-day period, B) average percentage of embryo survival over time for knock-out injections.

4.3.3. Design and assembly of plasmid constructs for CRISPR/Cas9 knock-in

In addition to knocking out *Zelda*, I attempt to use the CRISPR/Cas9 system to knock-in combinations of a 3XFLAG tag, a fluorescent tag, and a tetracysteine by non-homologous end-

joining and microhomology-mediated end-joining mechanisms. By attaching fluorescent tags to the *Zelda* coding sequence, I will be able to image live embryos and show Zelda protein localization during early development. By attaching a 3XFLAG epitope tag to the coding sequence, I will be able to perform immunostaining of fixed embryos to show localization of Zelda protein and validate that the 3XFLAG accurately tags Zelda like the fluorescent tag. In addition, the 3XFLAG tag will allow for future integration of -omic technologies such as CETCH-seq to understand which genes Zelda binds to (Savic *et al.*, 2015). This is particularly interesting as *Zelda* is a pioneer transcription factor and binds to genes to activate the zygotic genome and initiate MZT in *Drosophila* and may play a conserved role in *Parhyale*. Finally, the addition of a tetracysteine to the *Zelda* coding sequence will allow for ReAsH/FlAsH dye labeling of the protein (Irtegun *et al.*, 2011). I will be able to track Zelda protein by treating live embryos with the ReAsH/FlAsH system to turn cells containing Zelda protein fluorescent red or fluorescent green. By using both ReAsH and FlAsH subsequently, we might also be able to distinguish between maternally expressed *Zelda* and zygotically expressed *Zelda.* Finally, we will also be able to perform photoinactivation of Zelda protein tagged with the tetracysteine, giving temporal control over when *Zelda* is knocked down (Marek and Davis, 2002).

Fig 4.4 shows the design of the CRISPR/Cas9 knock-in constructs which will be the basis of the injected donor DNA.

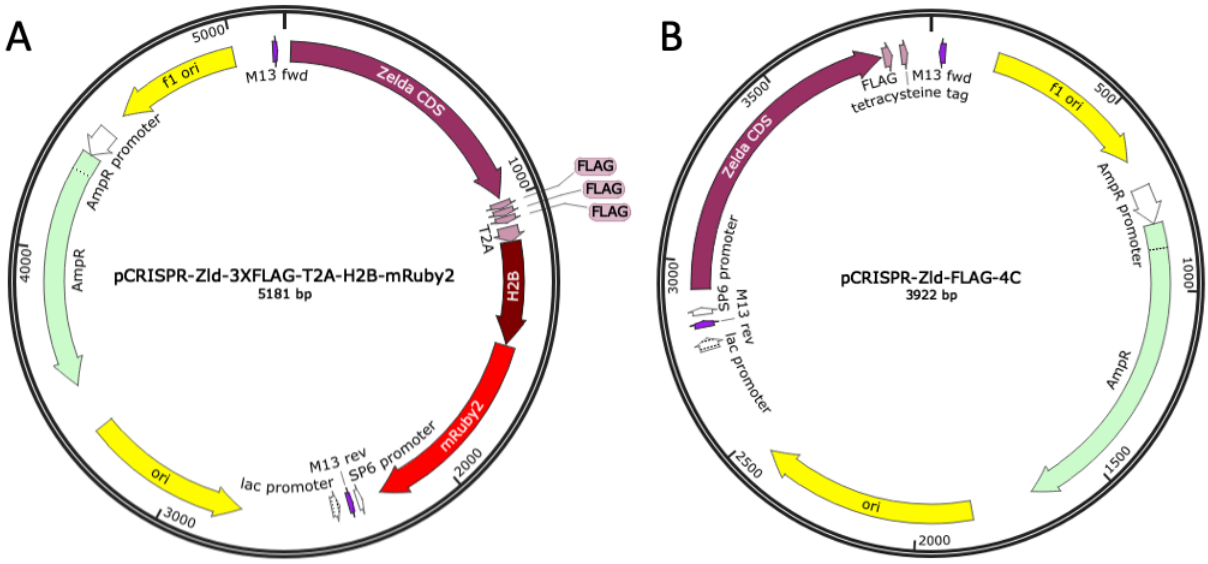


Fig 4.4: Construct design of A) pCRISPR-Zld-3XFLAG-T2A-H2B-mRuby2 and B) pCRISPR-Zld-FLAG-4C

To create transgenic *Parhyale* lines with epitope and fluorescent tags attached to *Zelda*, I used traditional restriction enzyme ligation methods and Gibson Assembly to create the donor DNA constructs described above. I successfully assembled the pCRISPR-Zld-3XFLAG-T2A-H2B-mRuby2 construct using Gibson Assembly and sequenced it to ensure that the epitope and fluorescent tags are in-frame with the coding sequence. I attempted to assemble the pCRISPR-Zld-FLAG-4C construct using Gibson Assembly as well, but the short length of the epitope tag makes it difficult as there is possibility of the exonuclease digesting the entire length of the epitope tag before it can be ligated to the other parts of the construct. Therefore, I used traditional restriction enzyme cut and ligation methods to create this construct.

To increase the chances of successfully knocking-in the desired tags, I used three different methods for CRISPR/Cas9 knock-ins. The first is to linearize the plasmid at the 3' end of the donor DNA so the pGEM T-Easy vector will not be inserted into the genome (Fig 4.5). The donor DNA is then delivered with Cas9 recombinant protein and sgRNA into the 1-cell embryo for NHEJ.

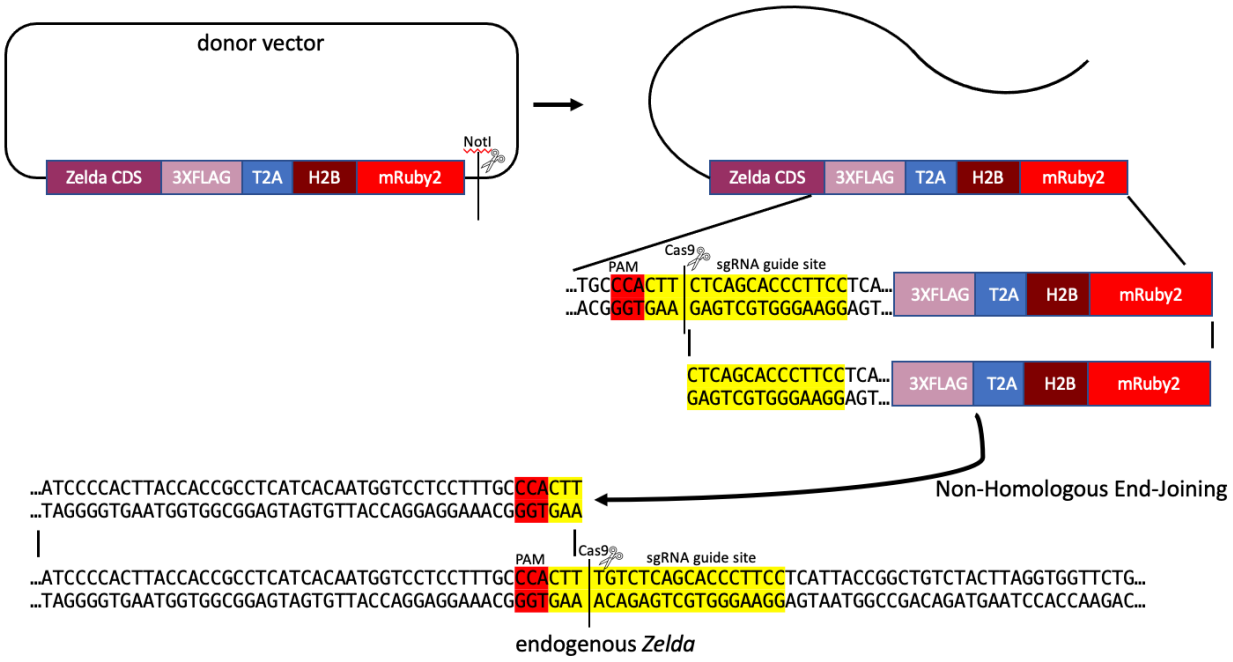


Fig 4.5: CRISPR/Cas9 Knock-In Strategy using non-homologous end joining of endogenous *Zelda* and linearized donor vector containing desired epitope tags

The second method is to amplify by PCR the insert that I wish to knock into the genome while also mutating the guide site or PAM site so subsequent cutting of the genome does not occur. The PCR product will then be used as the donor DNA in the injection mix for NHEJ (Fig 4.6).

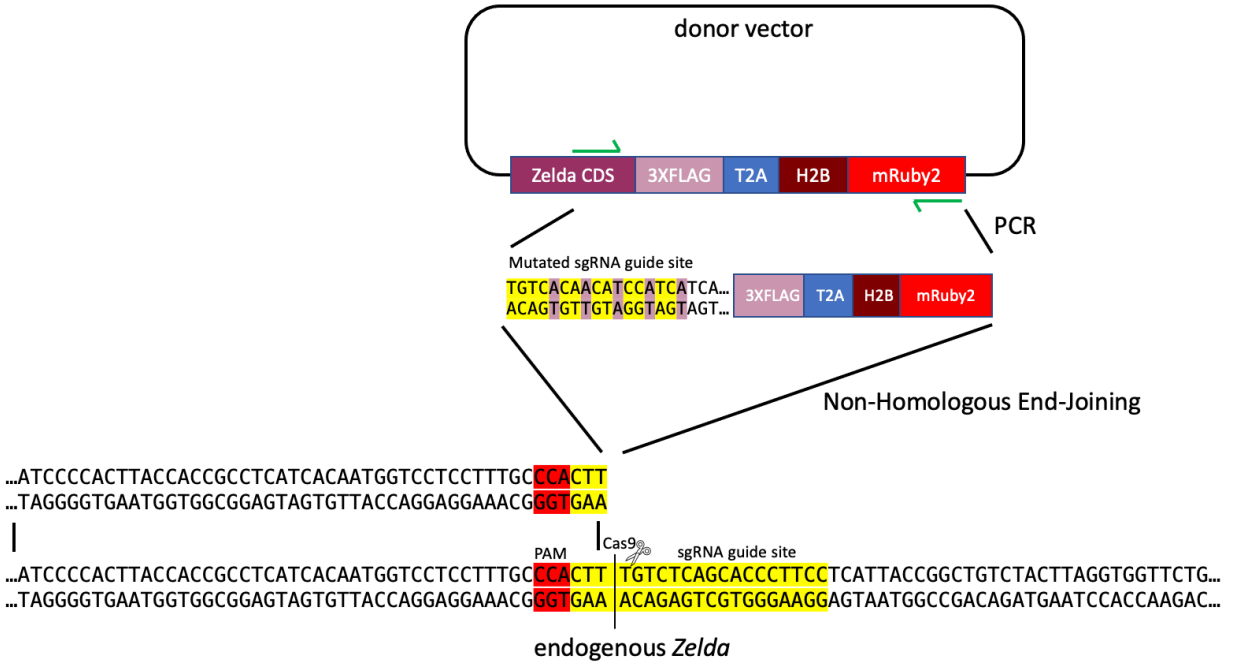


Fig 4.6: CRISPR/Cas9 Knock-In strategy using non-homologous end joining of endogenous *Zelda* with PCR product containing desired epitope tags and mutated guide site

Finally, MMEJ has been shown to increase efficiency in CRISPR/Cas9 knock-ins both *in vitro* and *in vivo* (Hisano *et al.*, 2015; Sfeir and Symington, 2016). Therefore, I have attached 48 bp homology arms to the insert with a mutated PAM site in order to replicate the precise in-frame knock-in strategy described before (Fig 4.7).

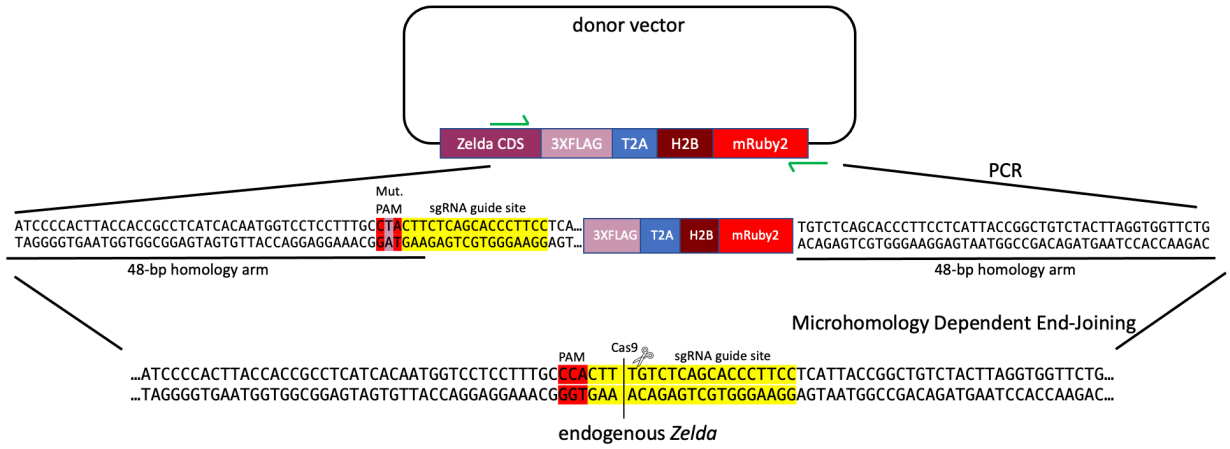


Fig 4.7: CRISPR/Cas9 Knock-In strategy by microhomology-mediated end-joining of endogenous *Zelda* with PCR product containing desired epitope tags, mutated PAM site, and 48-bp homology arms

4.3.4. NHEJ and MMEJ methods with linearized DNA do not successfully tag *Zelda* with epitope tags

Using the pCRISPR-Zld-3XFLAG-T2A-H2B-mRuby2 construct digested with NotI, I injected 371 embryos with Cas9 recombinant protein (500 ng/uL), sgRNA2 (250 ng/uL), phenol red, and donor DNA (10 ng/uL) and 59 embryos where I omitted the Cas9 protein as the control. I checked for expression of mRuby2 protein each day from 1-7 dpf. Although embryos were glowing red at 4 dpf onwards and there was variation in the amount of fluorescence being emitted between embryos from the same batch, further injections and experience leads me to believe that the red fluorescence is due to autofluorescence of later-stage embryos rather than proper knock-in of fluorescent tags.

As an alternative method, I amplified parts of the construct to mutate the guide site and PAM site in order to perform NHEJ knock-ins and ensure that the genomic DNA will not be cut again after the initial knock-in of the epitope tags. I injected 27 control embryos and 146 experimental embryos using this method, but no embryos expressed red fluorescent protein.

Finally, I attached 48-bp homology arms to the desired insert with mutated PAM to perform MMEJ on CRISPR/Cas9 injected embryos. I injected a total of 50 control embryos and 96 experimental embryos using this method, but no embryos expressed red fluorescent protein either.

4.4. Conclusion

In this study, we test out various CRISPR/Cas9 knock-out and knock-in strategies to understand the role of *Zelda* in early *Parhyale* development as well as to increase efficiency and integrate other technologies to the CRISPR/Cas9 system. We show that knock-out of zygotically transcribed *Zelda* results in a significant decrease in survival at around 5 dpf as maternal *Zelda* has been depleted, thus not providing a protective effect. In addition, we developed and tested CRISPR/Cas9 knock-in strategies with donor constructs containing various epitope tags. Using donor DNA linearized with a restriction enzyme, donor DNA amplified by PCR to mutate the guide and PAM sites, and donor DNA amplified to attached 48 bp homology arms, we injected embryos to tag *Zelda* by NHEJ and MMEJ methods. However, no method resulted in *Zelda* being tagged correctly. This negative result can possibly be attributed to low concentration of donor DNA or a need for alternative methods such as conventional homology-directed repair. In addition, selection of a gene with a more obvious phenotype might help in testing these different CRISPR/Cas9 strategies.

Chapter 5

Thesis discussion and future directions

5.1. Identifying early cell fate specifiers during early crustacean development

In this thesis, I have produced and validated transcriptomic datasets to be used as a resource for future analysis and avenues of research in regard to early cell fate specification, maternal-to-zygotic transition, and maternal transcript degradation. By generating transcriptomic datasets for each blastomere at the 4- and 8-cell stage and confirming that the method in which to produce these transcriptomic datasets create good quality data, candidate genes of importance to cell fate specification can easily be selected for future avenues of research.

In the *Parhyale* model, there are studies that look into the germline (Özhan-Kizil *et al.*, 2009; Gerberding *et al.*, 2002), ectodermal (Price *et al.*, 2010; Nestorov *et al.*, 2013; Gerberding *et al.*, 2002), and mesodermal (Gerberding *et al.*, 2002; Price *et al.,* 2010) lineages of the *Parhyale* embryo. Although these cell lineages have been studied in some capacity, there are few studies that show mechanistically how these cell lineages are formed and specified. In addition, the Mav and en blastomeres which are known to restrict to the visceral mesoderm and endoderm lineages respectively are not well studied (Gerberding *et al.*, 2002). Therefore, the blastomere-specific transcriptomic datasets at the 4- and 8-cell stage are starting places to understand the mechanisms that drive early formation of these germ layers. To test which genes drive the formation of the germ layers, knockdown experiments using RNAi can be performed on the top candidates of mRNAs asymmetrically distributed to the different lineages. By knocking down these genes, we can begin to understand the function and role they play in formation of the cell lineages. More specific knockdown of certain blastomeres by injecting at the 8-cell stage can provide a more targeted way to study the function of the gene or lack of function in blastomeres that have minimal to no expression of that gene. In addition, cell ablation methods as described previously (Price *et al.*, 2010; Alwes *et al.*, 2011; Nast and Extavour, 2014) can stop formation of certain germ layers.

By performing ISH on the embryos with ablated cells using probes designed from the candidate lists generated by the transcriptomic data, we will be able to see if other blastomeres may compensate for the loss of cell lineages.

Preliminary analysis of the datasets shows that the gene expression signatures of the macromeres of the 8-cell stage show a higher resemblance to the transcriptome of their parent cell than the micromeres do with their parent cell. Having the single blastomere data provides a way for us to understand cell fate decision events from the 4-cell to 8-cell embryo. Why do the micromeres differ much more from their parents than their macromere sister cells? Why are the micromere transcriptomes more similar to one another than to their sister cells, especially considering they come from the same parent cell as the macromeres? Previous work in the sea urchin and sea star embryo has shown that cell fate decisions are often associated with asymmetries in the cell size (Poon *et al.*, 2019; Barone *et al.*, 2022). Although cell size asymmetries may predict how cell fates are specified, they are not necessary in the formation of different axes. Future analysis on which mRNAs localize to the vegetal pole of the embryo at the 3rd cleavage to create differences in the sister cells may begin to answer these questions.

5.2. Establishing genetic and genomic tools and resources to study *Parhyale* embryonic development

In addition to actually generating the transcriptomic datasets for use in future analysis of early cell fate specification and cell fate decisions, I tested out and established methods to generate single blastomere suspensions for RNA-sequencing as well as for CRISPR/Cas9 technology applications. The method to fix and separate to collect blastomere suspensions for RNA-sequencing can be performed with other stages of interest. However, by the 16-cell stage, it is difficult to determine the blastomere identities as the cells are much smaller (Browne *et al.*, 2005). In addition, it provides

a challenge when dissecting as it is much easier to mistaken blastomeres since cell sizes are more similar. Therefore, scRNA-seq can be used to generate transcriptomic data of each blastomere at the 16-cell stage onwards. SPLiT-seq, which was developed initially using PFA-fixed cells, provides a way to perform scRNA-seq on blastomere suspensions for *Parhyale* embryos after the 8-cell stage (Rosenberg *et al.*, 2018).

In addition to potential scRNA-seq datasets that can be generated using the described method of fixation and separation of blastomeres, other -omic technologies can be used such as ATAC-seq and Hi-C to understand the chromatin structure of the *Parhyale* genome in different blastomere identities. ATAC-seq and Hi-C have been used on fixed samples, making it plausible to use these technologies on the samples generated from the blastomere separations (Yadav *et al.*, 2022; Zhang *et al.*, 2022; Ing-Simmons *et al.*, 2021).

Finally, I also tested out various CRISPR/Cas9 knock-in methods in order to improve the efficiency of tagging endogenous proteins with epitope tags. The methods attempted including NHEJ and MMEJ did not result in correct tagging of endogenous genes. Although unsuccessful, the attempts at improving CRISPR/Cas9 efficiency for knock-ins provides information for other strategies or steps to take to increase the efficiency. For example, the use of higher amounts of donor DNA than described previously in Kao *et al.*, (2016) may be necessary as other model systems uses a higher concentration of donor DNA template (Ata *et al.*, 2018). Another way to increase efficiency might be to use HDR as a method for DNA repair (Liu *et al.*, 2019).

In addition to increasing efficiency for CRISPR/Cas9 knock-ins, I have also synthesized plasmid constructs that include combinations of 3xFLAG, fluorescent mRuby2, and a tetracysteine sequence. These plasmid constructs can be used in the future for imaging experiments once the CRISPR/Cas9 knock-in system works.

References

References

Aanes, H., Winata, C. L., Lin, C. H., Chen, J. P., Srinivasan, K. G., Lee, S. G. P., Lim, A. Y. M., Hajan, H. S., Collas, P., Bourque, G., et al. (2011). Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res.* **21**, 1328–1338.

Abdi, H. and Williams, L. J. (2010). Principal component analysis: Principal component analysis. *WIREs Comp Stat* **2**, 433–459.

Alberstat, E. J., Chung, K., Sun, D. A., Ray, S. and Patel, N. H. (2022). *Combinatorial interactions of* Hox *genes establish appendage diversity of the amphipod crustacean* Parhyale hawaiensis. Developmental Biology.

Alwes, F., Hinchen, B. and Extavour, C. G. (2011). Patterns of cell lineage, movement, and migration from germ layer specification to gastrulation in the amphipod crustacean Parhyale hawaiensis. *Developmental Biology* **359**, 110–123.

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

Anderson, K. V. and Lengyel, J. A. (1979). Rates of synthesis of major classes of RNA in Drosophila embryos. *Developmental Biology* **70**, 217–231.

Arias, A. M. (2008). Drosophila melanogaster and the Development of Biology in the 20th Century. In *Drosophila* (ed. Dahmann, C.), pp. 1–25. Totowa, NJ: Humana Press.

Ata, H., Ekstrom, T. L., Martínez-Gálvez, G., Mann, C. M., Dvornikov, A. V., Schaefbauer, K. J., Ma, A. C., Dobbs, D., Clark, K. J. and Ekker, S. C. (2018). Robust activation of microhomology-mediated end joining for precision gene editing applications. *PLoS Genet* **14**, e1007652.

Aviv, T., Lin, Z., Lau, S., Rendl, L. M., Sicheri, F. and Smibert, C. A. (2003). The RNA-binding SAM domain of Smaug defines a new family of post-transcriptional regulators. *Nat Struct Mol Biol* **10**, 614–621.

Barnard, J. L. (1965). Marine Amphipoda of atolls in Micronesia. *Proceedings of the United States National Museum* **117**, 459–552, 11 tables, 35 figures.

Barone, V., Byrne, M. and Lyons, D. C. (2022). Lineage tracing shows that cell size asymmetries predict the dorsoventral axis in the sea star embryo. *BMC Biol* **20**, 179.

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

Bazzini, A. A., Lee, M. T. and Giraldez, A. J. (2012). Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science* **336**, 233–237.

Bedzhov, I., Graham, S. J. L., Leung, C. Y. and Zernicka-Goetz, M. (2014). Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. *Phil. Trans. R. Soc. B* **369**, 20130538.

Blythe, S. A. and Wieschaus, E. F. (2015). Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition. *Cell* **160**, 1169–1181.

Blythe, S. A. and Wieschaus, E. F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. *eLife* **5**, e20148.

Blythe, M. J., Malla, S., Everall, R., Shih, Y., Lemay, V., Moreton, J., Wilson, R. and Aboobaker, A. A. (2012). High Through-Put Sequencing of the Parhyale hawaiensis mRNAs and microRNAs to Aid Comparative Developmental Studies. *PLoS ONE* **7**, e33784.

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

Botman, D., de Groot, D. H., Schmidt, P., Goedhart, J. and Teusink, B. (2019). In vivo characterisation of fluorescent proteins in budding yeast. *Sci Rep* **9**, 2234.

Bowerman, B., Ingram, M. K. and Hunter, C. P. (1997). The maternal par genes and the segregation of cell fate specification activities in early Caenorhabditis elegans embryos. *Development* **124**, 3815–3826.

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

Browne, W. E., Schmid, B. G. M., Wimmer, E. A. and Martindale, M. Q. (2005b). Expression of otd orthologs in the amphipod crustacean, Parhyale hawaiensis. *Dev Genes Evol* **216**, 581–595.

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

Chaw, R. C. and Patel, N. H. (2012). Independent migration of cell populations in the early gastrulation of the amphipod crustacean Parhyale hawaiensis. *Developmental Biology* **371**, 94–109.

Chen, L., Nievera, C. J., Lee, A. Y.-L. and Wu, X. (2008). Cell Cycle-dependent Complex Formation of BRCA1·CtIP·MRN Is Important for DNA Double-strand Break Repair. *Journal of Biological Chemistry* **283**, 7713–7720.

Chen, L., Dumelie, J. G., Li, X., Cheng, M. H., Yang, Z., Laver, J. D., Siddiqui, N. U., Westwood, J. T., Morris, Q., Lipshitz, H. D., et al. (2014). Global regulation of mRNA translation and stability in the early Drosophila embryo by the Smaug RNA-binding protein. *Genome Biol* **15**, R4.

Cheng, X., Lyons, D. C., Socolar, J. E. S. and McClay, D. R. (2014). Delayed transition to new cell fates during cellular reprogramming. *Developmental Biology* **391**, 147–157.

Dalle Nogare, D. E., Pauerstein, P. T. and Lane, M. E. (2009). G2 acquisition by transcription-independent mechanism at the zebrafish midblastula transition. *Dev Biol* **326**, 131–142.

De Renzis, S., Elemento, O., Tavazoie, S. and Wieschaus, E. F. (2007). Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. *PLoS Biol* **5**, e117.

Denisenko, E., Guo, B. B., Jones, M., Hou, R., de Kock, L., Lassmann, T., Poppe, D., Clément, O., Simmons, R. K., Lister, R., et al. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol* **21**, 130.

Duan, J., Rieder, L., Colonnetta, M. M., Huang, A., Mckenney, M., Watters, S., Deshpande, G., Jordan, W., Fawzi, N. and Larschan, E. (2021). CLAMP and Zelda function together to promote Drosophila zygotic genome activation. *eLife* **10**, e69937.

Edgar, B. and Schubiger, B A (1986). Parameters controlling transcriptional activation during early drosophila development. *Cell* **44**, 871–877.

Extavour, C. G. (2005). The fate of isolated blastomeres with respect to germ cell formation in the amphipod crustacean Parhyale hawaiensis. *Developmental Biology* **277**, 387–402.

Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., Yi, L., Yu, G., Hou, Z., Fan, D., et al. (2018). Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell* **173**, 248-259.e15.

Geny, S., Pichard, S., Brion, A., Renaud, J.-B., Jacquemin, S., Concordet, J.-P. and Poterszman, A. (2021). Tagging Proteins with Fluorescent Reporters Using the CRISPR/Cas9 System and Double-Stranded DNA Donors. In *Multiprotein Complexes* (ed. Poterszman, A.), pp. 39–57. New York, NY: Springer US.

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

Gilbert, S. F. (2000). *Developmental biology*. 6th ed. Sunderland, Mass: Sinauer Associates.

Hamatani, T., Carter, M. G., Sharov, A. A. and Ko, M. S. H. (2004). Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. *Developmental Cell* **6**, 117–131.

Hannibal, R. L., Price, A. L., Parchem, R. J. and Patel, N. H. (2012). Analysis of snail genes in the crustacean Parhyale hawaiensis: insight into snail gene family evolution. *Dev Genes Evol* **222**, 139–151.

Haque, A., Engel, J., Teichmann, S. A. and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* **9**, 75.

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. and Eisen, M. B. (2011). Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. *PLoS Genet* **7**, e1002266.

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

Hasley, A., Chavez, S., Danilchik, M., Wühr, M. and Pelegri, F. (2017). Vertebrate Embryonic Cleavage Pattern Determination. In *Vertebrate Development* (ed. Pelegri, F.), Danilchik, M.), and Sutherland, A.), pp. 117–171. Cham: Springer International Publishing.

Hisano, Y., Sakuma, T., Nakade, S., Ohga, R., Ota, S., Okamoto, H., Yamamoto, T. and Kawahara, A. (2015). Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci Rep* **5**, 8841.

Ing-Simmons, E., Vaid, R., Bing, X. Y., Levine, M., Mannervik, M. and Vaquerizas, J. M. (2021). Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning. *Nat Genet* **53**, 487–499.

Irtegun, S., Ramdzan, Y. M., Mulhern, T. D. and Hatters, D. M. (2011). ReAsH/FlAsH Labeling and Image Analysis of Tetracysteine Sensor Proteins in Cells. *JoVE* 2857.

Joseph, S. R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A. and Vastenhouw, N. L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *eLife* **6**, e23326.

Kao, D., Lai, A. G., Stamataki, E., Rosic, S., Konstantinides, N., Jarvis, E., Di Donfrancesco, A., Pouchkina-Stancheva, N., Sémon, M., Grillo, M., et al. (2016). The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife* **5**, e20062.

Kester, L. and van Oudenaarden, A. (2018). Single-Cell Transcriptomics Meets Lineage Tracing. *Cell Stem Cell* **23**, 166–179.

Kitamura, A., Nakayama, Y. and Kinjo, M. (2015). Efficient and dynamic nuclear localization of green fluorescent protein via RNA binding. *Biochemical and Biophysical Research Communications* **463**, 401–406.

Kretzschmar, K. and Watt, F. M. (2012). Lineage Tracing. *Cell* **148**, 33–45.

Lamb, M. M. and Laird, C. D. (1976). Increase in nuclear poly(A)-containing RNA at syncytial blastoderm in Drosophila melanogaster embryos. *Developmental Biology* **52**, 31–42.

Laver, J. D., Li, X., Ray, D., Cook, K. B., Hahn, N. A., Nabeel-Shah, S., Kekis, M., Luo, H., Marsolais, A. J., Fung, K. Y., et al. (2015). Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. *Genome Biol* **16**, 94.

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P. and Krause, H. M. (2007). Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function. *Cell* **131**, 174–187.

Lee, M. T., Bonneau, A. R., Takacs, C. M., Bazzini, A. A., DiVito, K. R., Fleming, E. S. and Giraldez, A. J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364.

Lee, S., Zhang, A. Y., Su, S., Ng, A. P., Holik, A. Z., Asselin-Labat, M.-L., Ritchie, M. E. and Law, C. W. (2020). Covering all your bases: incorporating intron signal from RNA-seq data. *NAR Genomics and Bioinformatics* **2**, lqaa073.

Lehmann, R. and Nüsslein-Volhard, C. (1987). hunchback, a gene required for segmentation of an anterior and posterior region of the Drosophila embryo. *Developmental Biology* **119**, 402–417.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570.

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M. M., Kirov, N. and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**, 400–403.

Liao, Y., Smyth, G. K. and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research* **47**, e47–e47.

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D. and Ma, W. (2019). Methodologies for Improving HDR Efficiency. *Front. Genet.* **9**, 691.

Liubicich, D. M., Serano, J. M., Pavlopoulos, A., Kontarakis, Z., Protas, M. E., Kwan, E., Chatterjee, S., Tran, K. D., Averof, M. and Patel, N. H. (2009). Knockdown of

Parhyale Ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13892–13896.

Loureiro, J. and Peifer, M. (1998). Roles of Armadillo, a Drosophila catenin, during central nervous system development. *Current Biology* **8**, 622–633.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550.

Marek, K. W. and Davis, G. W. (2002). Transgenically Encoded Protein Photoinactivation (FlAsH-FALI). *Neuron* **36**, 805–813.

Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C. and Patel, N. H. (2016). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. *Current Biology* **26**, 14–26.

Matsuda, T. and Oinuma, I. (2019). Optimized CRISPR/Cas9-mediated in vivo genome engineering applicable to monitoring dynamics of endogenous proteins in the mouse neural tissues. *Sci Rep* **9**, 11309.

Nast, A. R. and Extavour, C. G. (2014). Ablation of a Single Cell From Eight-cell Embryos of the Amphipod Crustacean Parhyale hawaiensis. *JoVE* 51073.

Nestorov, P., Battke, F., Levesque, M. P. and Gerberding, M. (2013). The Maternal Transcriptome of the Crustacean Parhyale hawaiensis Is Inherited Asymmetrically to Invariant Cell Lineages of the Ectoderm and Mesoderm. *PLoS ONE* **8**, e56049.

Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795–801.

Oikawa, M., Nagae, M., Mizuno, N., Iwatsuki, K., Yoshida, F., Inoue, N., Uenoyama, Y., Tsukamura, H., Nakauchi, H., Hirabayashi, M., et al. (2022). Generation of *Tfap2c-T2A-tdTomato* knock-in reporter rats via adeno-associated virus-mediated efficient gene targeting. *Molecular Reproduction Devel* **89**, 129–132.

Okamura, K., Ishizuka, A., Siomi, H. and Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666.

Özhan-Kizil, G., Havemann, J. and Gerberding, M. (2009). Germ cells in the crustacean Parhyale hawaiensis depend on Vasa protein for their maintenance but not for their formation. *Developmental Biology* **327**, 230–239.

Pavlopoulos, A., Kontarakis, Z., Liubicich, D. M., Serano, J. M., Akam, M., Patel, N. H. and Averof, M. (2009). Probing the evolution of appendage specialization by Hox gene

misexpression in an emerging model crustacean. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13897–13902.

Poon, J., Fries, A., Wessel, G. M. and Yajima, M. (2019). Evolutionary modification of AGS protein contributes to formation of micromeres in sea urchins. *Nat Commun* **10**, 3779.

Price, A. L., Modrell, M. S., Hannibal, R. L. and Patel, N. H. (2010). Mesoderm and ectoderm lineages in the crustacean Parhyale hawaiensis display intra-germ layer compensation. *Developmental Biology* **341**, 256–266.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C. E., Socci, N. D. and Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* **14**, R95.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A. and Patel, N. H. (2009a). Antibody Staining of *Parhyale hawaiensis* Embryos. *Cold Spring Harb Protoc* **2009**, pdb.prot5129.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A. and Patel, N. H. (2009b). Fixation and Dissection of *Parhyale hawaiensis* Embryos. *Cold Spring Harb Protoc* **2009**, pdb.prot5127.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A. and Patel, N. H. (2009c). In Situ Hybridization of Labeled RNA Probes to Fixed *Parhyale hawaiensis* Embryos. *Cold Spring Harb Protoc* **2009**, pdb.prot5130.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A. and Patel, N. H. (2009d). Injection of *Parhyale hawaiensis* Blastomeres with Fluorescently Labeled Tracers. *Cold Spring Harb Protoc* **2009**, pdb.prot5128.

Rembold, M., Ciglar, L., Yáñez-Cuna, J. O., Zinzen, R. P., Girardot, C., Jain, A., Welte, M. A., Stark, A., Leptin, M. and Furlong, E. E. M. (2014). A conserved role for Snail as a potentiator of active transcription. *Genes Dev.* **28**, 167–181.

Rosenberg, A. B., Roco, C. M., Muscat, R. A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L. T., Peeler, D. J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182.

Russell, J. N., Clements, J. E. and Gama, L. (2013). Quantitation of Gene Expression in Formaldehyde-Fixed and Fluorescence-Activated Sorted Cells. *PLoS ONE* **8**, e73849.

S Bruce, H., Jerz, G., R Kelly, S., McCarthy, J., Pomerantz, A., Senevirathne, G., Sherrard, A., A Sun, D., Wolff, C. and H Patel, N. (2021). *Hybridization Chain Reaction (HCR) In Situ Protocol v1*.

Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.-I. T. and Yamamoto, T. (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat Protoc* **11**, 118–133.

Savic, D., Partridge, E. C., Newberry, K. M., Smith, S. B., Meadows, S. K., Roberts, B. S., Mackiewicz, M., Mendenhall, E. M. and Myers, R. M. (2015). CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Res* **25**, 1581–1589.

Schwentner, M., Combosch, D. J., Pakes Nelson, J. and Giribet, G. (2017). A Phylogenomic Solution to the Origin of Insects by Resolving Crustacean-Hexapod Relationships. *Current Biology* **27**, 1818-1824.e5.

Semotok, J. L., Luo, H., Cooperstock, R. L., Karaiskakis, A., Vari, H. K., Smibert, C. A. and Lipshitz, H. D. (2008). Drosophila Maternal Hsp83 mRNA Destabilization Is Directed by Multiple SMAUG Recognition Elements in the Open Reading Frame. *Mol Cell Biol* **28**, 6757–6772.

Serano, J. M., Martin, A., Liubicich, D. M., Jarvis, E., Bruce, H. S., La, K., Browne, W. E., Grimwood, J. and Patel, N. H. (2016). Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiensis. *Developmental Biology* **409**, 297–309.

Sfeir, A. and Symington, L. S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences* **40**, 701–714.

Shah, G., Thierbach, K., Schmid, B., Waschke, J., Reade, A., Hlawitschka, M., Roeder, I., Scherf, N. and Huisken, J. (2019). Multi-scale imaging and analysis identify pan-embryo cell dynamics of germlayer formation in zebrafish. *Nat Commun* **10**, 5753.

Sharma, A., Toepfer, C. N., Ward, T., Wasson, L., Agarwal, R., Conner, D. A., Hu, J. H. and Seidman, C. E. (2018). CRISPR/Cas9-Mediated Fluorescent Tagging of Endogenous Proteins in Human Pluripotent Stem Cells. *Current Protocols in Human Genetics* **96**,.

Shoemaker, C. R. (1956). Observations on the Amphipod Genus Parhyale. *Proceedings of the United States National Museum* **106**, 345–358.

St. Johnston, D., Driever, W., Berleth, T., Richstein, S. and Nüsslein-Volhard, C. (1989). Multiple steps in the localization of *bicoid* RNA to the anterior pole of the *Drosophila* oocyte. *Development* **107**, 13–19.

Stoeckius, M., Grün, D., Kirchner, M., Ayoub, S., Torti, F., Piano, F., Herzog, M., Selbach, M. and Rajewsky, N. (2014). Global characterization of the oocyte-to-embryo transition in *C aenorhabditis elegans* uncovers a novel m RNA clearance mechanism. *EMBO J* **33**, 1751–1766.

Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. *Developmental Biology* **100**, 64–119.

Sun, D. A. and Patel, N. H. (2019). The amphipod crustacean *Parhyale hawaiensis* : An emerging comparative model of arthropod development, evolution, and regeneration. *WIREs Dev Biol* **8**,.

Sun, D. A., Bredeson, J. V., Bruce, H. S. and Patel, N. H. (2022). Identification and classification of cis-regulatory elements in the amphipod crustacean *Parhyale hawaiensis*. *Development* **149**, dev200793.

Tadros, W., Goldman, A. L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T. R., Westwood, J. T., Smibert, C. A. and Lipshitz, H. D. (2007). SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase. *Developmental Cell* **12**, 143–155.

Thomsen, S., Anders, S., Janga, S. C., Huber, W. and Alonso, C. R. (2010). Genome-wide analysis of mRNA decay patterns during early Drosophila development. *Genome Biol* **11**, R93.

VanHorn, S. and Morris, S. A. (2021). Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development. *Developmental Cell* **56**, 7–21.

Vargas-Vila, M. A., Hannibal, R. L., Parchem, R. J., Liu, P. Z. and Patel, N. H. (2010). A prominent requirement for *single-minded* and the ventral midline in patterning the dorsoventral axis of the crustacean *Parhyale hawaiensis*. *Development* **137**, 3469–3476.

Venken, K. J. T., Kasprowicz, J., Kuenen, S., Yan, J., Hassan, B. A. and Verstreken, P. (2008). Recombineering-mediated tagging of Drosophila genomic constructs for in vivo localization and acute protein inactivation. *Nucleic Acids Research* **36**, e114–e114.

Vera Alvarez, R., Pongor, L. S., Mariño-Ramírez, L. and Landsman, D. (2019). TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35**, 1960–1962.

Vu, T., Doan, D., Kim, J., Sung, Y. W., Tran, M., Song, Y. J., Das, S. and Kim, J. (2021). CRISPR/Cas-based precision genome editing via microhomology-mediated end joining. *Plant Biotechnology Journal* **19**, 230–239.

Wagner, D. E. and Klein, A. M. (2020). Lineage tracing meets single-cell omics: opportunities and challenges. *Nat Rev Genet* **21**, 410–427.

Wang, H. and Xu, X. (2017). Microhomology-mediated end joining: new players join the team. *Cell Biosci* **7**, 6.

Wang, Q. T., Piotrowska, K., Ciemerych, M. A., Milenkovic, L., Scott, M. P., Davis, R. W. and Zernicka-Goetz, M. (2004). A Genome-Wide Study of Gene Activity Reveals

Developmental Signaling Pathways in the Preimplantation Mouse Embryo. *Developmental Cell* **6**, 133–144.

Whittaker, J. R. (1973). Segregation During Ascidian Embryogenesis of Egg Cytoplasmic Information for Tissue-Specific Enzyme Development. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2096–2100.

Wolf, S., Wan, Y. and McDole, K. (2021). Current approaches to fate mapping and lineage tracing using image data. *Development* **148**, dev198994.

Workman, J. L. and Kingston, R. E. (1998). Alterature of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**, 545–579.

Yadav, R. P., Polavarapu, V. K., Xing, P. and Chen, X. (2022). FFPE-ATAC: A Highly Sensitive Method for Profiling Chromatin Accessibility in Formalin-Fixed Paraffin-Embedded Samples. *Current Protocols* **2**,.

Yoon, Y., Klomp, J., Martin-Martin, I., Criscione, F., Calvo, E., Ribeiro, J. and Schmidt-Ott, U. (2019). Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. *eLife* **8**, e46711.

Yuan, K., Seller, C. A., Shermoen, A. W. and O'Farrell, P. H. (2016). Timing the Drosophila Mid-Blastula Transition: A Cell Cycle-Centered View. *Trends in Genetics* **32**, 496–507.

Zeng, V., Villanueva, K. E., Ewen-Campen, B. S., Alwes, F., Browne, W. E. and Extavour, C. G. (2011). De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis. *BMC Genomics* **12**, 581.

Zhang, H., Polavarapu, V. K., Xing, P., Zhao, M., Mathot, L., Zhao, L., Rosen, G., Swartling, F. J., Sjöblom, T. and Chen, X. (2022). Profiling chromatin accessibility in formalin-fixed paraffin-embedded samples. *Genome Res.* **32**, 150–161.

Zhong, H., Ceballos, C. C., Massengill, C. I., Muniak, M. A., Ma, L., Qin, M., Petrie, S. K. and Mao, T. (2021). High-fidelity, efficient, and reversible labeling of endogenous proteins using CRISPR-based designer exon insertion. *eLife* **10**, e64911.

Appendix

Appendix A: Nucleotide coding sequences of candidate genes in this study

phaw Zelda
>MSTRG.30332

ATGGAAGAAACCACCCTTCCAACCGATTCTACGTCCGTCAGTTCCTCATCTCCTACTACTCTAACCGCCTCATCATC
AGCTTCACCTGCCATCGATACGACAGTAAACGATCGCTCCACCCCCACTGTCTACAGCTGCGAACCTTGCAAAAAAC
GATTTGTAGACGACCAGTCCTTGCATATCCACCTTGCCTACGATCACGGCAGGGGCAACCTGATGAGCAGCATGTGG
ATGAGCTCAAACGACAACCCAGACGGCACACAGTCGTCCTCCACCATGGATTCTCAAAATGCTGATAAATCAGTTAG
TCATATTCCCAAAGTCGAACCCTCTGTCATAACAGGGGATATACACAACCAACCATTGGACAACATGGATTTCGCCA
GCATCCTCACAAGTGTCGATTCGATGCAGATGCAGAACAGCCCTCACCAAAAACTCGCGACTTCGTCAATGTCGCCC
GTACAGTTTGATTTTGCCTCAACCACCGGCATGATGCCTTCCCCTCAAACCCCTGTTGTGTCTCAAGGTCCGTCATC
AACTCAAAGCAACGTGAGCGGATCTTTTCAGGTTTCGGCAGGTCAGTATCACAATGATTATGGGCAGTTACAGAGTA
CCGGAGACTATGGAGTTCCTCAACCTTACACATCGCCCCTCCCGTCTTACGGAAGAATCAACGGATATTCCTCTGAA
GATACAGACTACCAGTCAGCTTACGATAGACAGTTCTCTACATCTGCGGGGAGATACCATCCCTATGGCAGATCACC
ACTTCCTTCACCCAGTGTTGTAAAGGAAGAACCGGTAATTAAGACGGAACCTCAAGAAATTCTAGATCTCGACTCTC
ATCGATTTCTACAGAACAGCACGCATACGACTGCTGTACCTCAACTCTCGGGGGGAACTCTATCGCAGATATGGAGG
CCTGTAGATCAGCAGCTTAGCAATTCGCCTACTCTTCCGCCGATGCCGTTAACCAGTGGACCTCTGCAATCAATGCC
CTCGAGCTCCATGTTTGGGGGTTCAGCTGTACATCCTCAGCAGAGCATGTCGCAAAACGTAATGAACGGTGTACCTC
ACGGATACCCATCTCCCGGCCCTTCACACAGCTACCACCCTCAAGGGTTTTCACAAGCGGGAGTCCTAGCTCCTCTT
AATCTACCCAAGCCTCCTCCAAGTTTCATAACGACTGATACTCCGCAAAGTTCGCCTGTTATGCAGTCTCAGCCGCC
CGTGCCCCCACCGCGGCCTTCTAGTGATGGACCACCGACTGAGTTTCCAACGAACGTGAAAAGGCCGAAGAATTTCA
AATGCGAAATTTGCCTCAAGTGGTTCACCAGTCAAGGTCACTTGAAGAGACATTACAACACTACTCTTCACAAGAAT
ATGGTGAAACAGAAAGAACCTAGGGAGGCTGATGTGACTGAGCCCGCTCCTATGCAAAATACGCTTCAAAATTCTCA
TTTACAGCCGGCTATGTCTCCGGCGATATCGCCTGCCATGGCTTCAGCCATGTCACCGGCTATGACTTCTTCCCTCC
CAACTCCTGATACACCTTTGGTCTCATCTAGTCCGAATCTCTCTAATGGTCCAACTGGCCTTCCATGTCCTGATAAC
TCAAATCAACTGCCATCTGAGCAAGAAGATAAAACTGCTCTGGTAAATCCCAGTCCTTTATCCTATCAAATGTCCAA
TGATATTCAATACTCACAGCCAGCCGAGCAGGGCATACAATATCATTCAACTTACGGTTCCCTGCCGACTTACGATG
CTACCTACAGTAGTGACTTTCACAATAGACCCCAGGGTATGGATGGGTTTCCCGTAGTTTCTTCGAACAGCCAATCT
ATTCATCCCAATGTTCCTTCAACTATGTATCATGGCCATTTTAGTAGCAGTTCTAATGCATCTCAGGCTCTGGAGAC
CAGTACTAGTGGGGGATATTCGTCTCCGTTTTTTCAGGCAGGATCCACCCCAATGGTAGGGGGTGGGACTACCTTTC
TCCGACCCAATGAATCCAACGTCCGTCCTTCGATTTCACCTCAACAGCCGTCCTCCCAGGCTCACAATGGTGGATAT
AACCCAAGTTTAGGTCTAGTGACGAACCCTACTTACCCTTCTGCCACGTGTATCACTACAAGTACAAATGAATATCT
TAGTCAGAGTCATTGTGACACTTCTGGGCATATCAACGCACCCAGTCGCTTTTCTCCTCTGCAGAGCAACCATGCAG
ATTTACAGAATGTAGATGCAATTATAGACAATTTTAGCTTCACGAATTTGCATCCGTTAGGCTCGGGCTCTATGCCA
AAGGAACTGGAATCATCGCAGCATTTTCCTGAGCATGACGGGAGATCGAATTCGATATCTCCTGAAGGGAGACTGAG
GTCGGCATTGGATAGTCCATCTGACGGGTCCAGAGACTCAGAGGCTACCAGTGCCAGCGGGGGCAGCGCGCCGGCCA
CGACGCCTACCCCTAAAACCGTCAAGGATAGCAAAGAGGGGACCTCAATTAAGTGTCCCGACTGTACTAAAGTATGT
GCCAGCCCTTGTTATCTTACACAACATAGGAAACGCCACCATGCGGGAGCGTTAGATCACAAGTGCGAAAAATGTGG
CAAAAAGTTTTCTGATCGCGTTACTTATGAGGAACACAAATTGAAACACGCCGGTGAAAAACCTTACAAATGTGATG
AATGTCCCAAGCAGTTTAACCATAAAACAGATTTGAGGCGACACAGTTGTCTGCATACTGGAGAAAAACCCTTTCAC
TGTCCCACTTGCCAGAAAGGGTTCATAAGGAAGGACCATATGATGAAACACCAATTGACTCACGAAAAGAAAAGACA
TCAAGCTGCTGCCGCTTCTGTCAACAGCCTCGCTGCTTCTCAACTCCCTGTTACCATGAATGGTTTCGCCCCTCCTC
CCCACATGGGGCCTAACCCTATCCTTCCTGATCATTTTGTCCCAACCCAGTATGCTACAATGGGTCATCCCCACTTA
CCACCGCCTCATCACAATGGTCCTCCTTTGCCCACTTTGTCTCAGCACCCTTCCTCATTACCGGCTGTCTACTGA

phaw Mex3
>MSTRG.64035

ATGCGCAGAGCGATGCCCAGTTACCAGGGCGAGTATTGTTTGAACTCGGGCTTTGGTATAAGTTACTTTGCTTTCAG
GACATGCGTAGGTGACGACTTGAACGATCTTTGCTCTAGGAAGCACGTTCATGAAATTAAGTTCATGCCTGCCAACA
GTTCTTACGATTTGGGCACCATGGTTACAAACTCTATTCTTTCACAATCTGTCACACCAGGAGTCACCATGGAGGAG
ACTAGAGCTTTTAATCTTGCTCTCGAATTAAATCAGTACATACAACATCAGCAGCAACAACATCAACAGCAGCAACA
TCAGCAGCCGCAGCAACAAAATATCGAATCTATTGCGTACATGAACGCCTTATCCCAGATGGAAGATGCCTCTGTTC
TGGCAAATCTGTCAGAAGCTGAACGAAGACGTTCCCAGAACATGACGGAATGTGTTGCTGTTCCCTCATCGGAGCAC
GTTGCTGAAATCGTTGGCAGGCAGGGTTGCAAAATAAAGGCGCTGCGAGCAAAGACAAACACGTACATCAAGACCCC

AGTCCGTGGAGAGGAACCAGTCTTCGTTGTGACAGGTCGCAAGGAAGACGTGGCGAGCGCCAAGCGTGAGATCCTGT
CAGCGGCTGACCACTTCAGCCAGATCCGTGCCTGCCGGAAAAATAACATCAATGGATCCATGGCGCCGGGACCGCCA
GCAAACGTGCCTGGTCAGAAGACCATAAACGTCACCGTGCCCTACAGAGTAGTGGGTCTTGTTGTAGGACCCAAAGG
GGCAACCATCAAGCGCATCCAACAAGCCACCCAGACCTACATTGTAACGCCATCGAGAGACAAGGAGCCGATCTTCG
AAGTCACCGGTCTGCCGGAATCCGTGGAAGCTGCTAAGAGAGAAATTGAGAGCCACATCGCCATGCGCACCGGAAAC
ATGTTCGAGGAAGACCTTGCAATGCCGGGCATGGCTCCGTATGATGCGCTCAACCTCACCGCGCAAGAACAGTACAA
CATGTACAGCCATCCTAAGCCAATCACTACTGCTGCTACATTGAGTGCCTTACTGGCCAATGGTTACTCTCAACAAA
GCTCAATGCTCAGCAACGTGAATTCTCAGTATCCCGGTATGTACAGTGGTGCATCTCAGAGCTCCAACAGTGTCCAG
TCGAGCCTACCATCCCTCTCAACCAACGGGCTCAATGGACTTGGTGGTGGTTCCTCTTACAACTCCACGCTCAACAA
CAGCACAATGAATTCCCTCAGCGCGCTCAACCACGCACTGTCCTCCAACAATGTGCTCCCCTCCAACTCGTCTGGTA
ACCTAGTGGATTTCTCCAGCCTCAATGGGCTCCACACTTCCAACAACTCCAGTAGTATGCTCAACGGCGGAAGTTCT
TCCTTCGGTAATGGAGGCATCAACGGTGCTGCTCAATCTGCTTTTGGATCGGCTTCAAATATCTTCAATGACTTCCC
TGGATTCAACTTGCTTGGCGAACAAGCCAATAACCTCATCAACTCGAAGAGCTCCAGCCTTGATCAGAATTCCTCCA
ATGCCAGCTTGCTTTCGTCCCTGAACAATTGCTCGACCATACCATCTGGTAGCTCGGCTTTGCCGCCCTACGGCGGA
CTTGACGAGGGCCTTGGATCACTCGGGTCTTCAATGGGGTCGTCTCCCGCACTTGATCAGAGCCAGGCCACTCAGAG
CATCTGGAGTGATGCCAAATATCAGACTCCACTGTCCGGATACAACTTCAGCAGCCGCCCATTTGGTGGATCGAGCG
GCAATCCGTCACGTCACCTGTCGCCTACCCGCGCGTTCAGCAGTCACGGTTCTGTTCAGCGCTCCAACTCTGATCCC
ACTCCCAAGATTGACGTGCTGTCTTCTGCTCTTTCACAGCTAAGCACCACGGGAATCCGTCGCAGCAACACTTCCTC
TCCTGAGATTTTGCCAGACTTCGATGTAACGGCGCTGGGTAACTCCATCTCGGGCTTCTCTAGGAGCAGTCCTGTCG
GCTGTTCCCTTGTGGAAGGCGCCTCCGCCGCTCTCCTTGCTGAAATGGGCGTCAAAACTTCTGCCGTCAGTTCAGGC
TCGTCCCTCTTGAGCAGCTTGCCTTCATCCAGCGCCGTTGCCGGCATGACTGCTGCTTCTGCCATTCCCATTGGCAC
GCCTGCTTCAGTTTCCGCCGCTGTTTCAGAAGCGGTTGCTGCCGGCGTCTCAGCGTGCGGCAGCCGCCGCAGCAGTG
TCGATGGCATTGGTGGAGAGAAGATCTGCCCTTCATGCAAGGGGGAAAGCTTCGAGGTTGCTTTCGTGCCCTGCGGT
CATCGCCGCTTCTGCTTGCCTTGCGCCCAAAAAATCCAAGACACCACCGGCTGCTGCCCCTCATGCCATGCTGAATC
CAAGGAGGTCATCCGCATCAGAGATGCCCACGAATGA

phaw Pumilio
>MSTRG.22416

ATGGCCACGAGTCACATTCCCTATCCTCCAGTCTCATCTTTCCAGGGTATAATGGTGGGGCCTCACATGCAGCAACA
GCAGCACCAACAACAGCAACAGCTGGCTGCCTTTTCCAATGCCTCCTTACTTCCAGTAACTCAGGGGGCATCTGCCT
TCCCTCCTCCTCCTGGCCTGGGCAATGTTGTCGGAGGTAATGAGGACGCCAGCACTGCGCACTTTCTTCCCCCTTTG
CCTCCGCATATTCTCCAACAGCAACAGCAGCAGCAACAACAACAGCAACAGTTGGATCCATCACAAGAGAGCATGCA
ACACACTATGCAAGGCATGGGCTTGGACATCAGCAGTGTCAGTGGTGGCATGTTGGTTGGACCTCCAATGCCTGGCC
AGTACTACACCACTGCCTCGGGAGCACTAGTATCTCATCCTTTGCATCAGCCGGATCAGAACGATCCGAGCTCGAGT
GTACAGCAGCAACAACAGGACCAGCAGCAGCAGCAACACCAGCAACAGCATCAACAGCAGCAGCAACAACAGCAGCA
ACAGGGTGATCAAGCTTTTCCTTCCGAAGCCGATCAGAAATGGAACGCTATCGGTCAAAGCGACAACCAGGGTCTCT
CCATTGCTCCCATGAAGAATCTTTGGTGCGCGGGCGGCGAAGAGACGAGCAAGGCAGACGGTGGCTCCTCCAGCAAG
GCGGCCACAGCCAACCACATTTTCAACTCTGCTGCCGAGCAGACCTGGAGAAATGCTGCCTGGAGCACGGGCAGCAG
CAGCGGACACCCTGGGCAGCAGCACCAGCCACAGCACCCTTACGAACACGGCGTTGGCGGTGGGCTGGTGATGGGCC
GTGCAGGCGGCGCGACGGCAGGTGGCAGCGGCAGCGGAGGAGCTCCCCACTCTCAGTTCGTTGGGGGAGGTGGTGGC
CCCGGACCTGAGGGCGTGCACTCGACGCGTAATGTTGTTACGGACGGCACTGGTGTGGTTAAAATGGTGGAACAGCT
ACTGGGCTCCTCACCGACTGGCACTAAGGATATCGATAGAAGCATGGCCAATCTCTCGCTCAGAAACGGCGGCACCG
AACCCGGTAAAGAGAAGAAAGATAAACCTCCCTCTCCTTACGACCCAAGCAAGAAGGATGTGGAGAATGGCAACCTC
ACGGCCCAAGCTAATGGCATCGTACAGAATGGTTTAGATGACGACAAAAGTTATAACAGGCAGGGCTCTCCCTCGGA
GGAAGATCTCAACAAGAACTCCGGAGTGGCAGGCGCTGGCGGCGGCGGGGGTGGCGGAGGTGGAGCTGTAGGCCTCA
TGCCGGGTGTCGTGGTAAAGGACGAACTGGTGGCGGGTGTAGGCGGGGGCGTTGTTATGGGCGGCGGTGCTGGCGGT
CCCACTGCTGCCCAGCAGATGCTGGGACACCTCGATCATGCCGCATTCGACGGCGTCACCGGCATGGGCAGCCACCA
CCACACCGGCATGCTCGACCCCACCGCCATGAGCGCGTCTATGGGCGGGGTGAGCATGGCCAACATGGGCGGCGTGA
GCATGGGCATGGCTGGCGCTATGGGCGGCATCAATATGGGCACGAGTATGGGCCAGGCGGGAGGCATGAACGCCGCG
GCGGTGGCGGCAGCCATGGGCGGCGTGGGGGGTCTCAACGCTGCTGGAGTAGCCGGCATGGCCGGGGCTGCCGGCTT
CAACAATCCCGGCAACACCGGAGCTGCCGGCACCGCCACCGCGGACTACACCAGCATGGCTAACAACATCTCCATGG
ACTCGCCCTCTCTTATGCAGCAACAGCAGCAGCAACAACAGCAACAGCAGGCTGCGGCTGCTGCCGTCCAGAACTTT
GCTGCTGTTGCTGCCGCCGCCCAGCAACAACAACAGCAGCAGCAACAACAACAACAGCAACAACAACAGCAGCAGCA
ACAGCAACAACAGCAGCAGCAACAGCAAAACGACAATCAGTTGTTCCGGAGCCCACAGCAACAGCAACAACAGGCTG
CAGCAGCAGCGGCTGCTGGACCTCAGCAGTTGCAGCTCCTAGCTGCTCAACAACAGTACTTGGCCCAGCAGGGTTTG

GCCCCGGCGGCGGCTGGTTTCCCTAACGCTCCCTATGTCATTAATGCTGCACAGGAGCCGTACACGCTTATTGCTGC
TCCCGGAGCTGCTACAGTCATCCCGCAGTACTATGGCGTCCACTGGGGTGTATACCCAGCCAATATCATACAACAAC
AGCAGCAGCAGCAACAACAGCAACAGCAGCAGCAACAGCAACAACAAGGAGCTGCAGCGCAGCTTGCGGCGGCCGGA
CAGCAGCAACAGCAGAGACGTCCTCTGTCTCCCACACAGAGTGACGTCACGAACCAAAACAATGGAGGAATTGCTCA
GAACATGCAGCAGTATCAGGTGATTCCTACGTATTATGATCAGAACGGCACTATGCTGATGGCTGGGAACTTGCGTG
GTGTCGGAAATGGCGGTGCTGTTCGCCTTGTGTCACCGGCACCTGTCATCATCAACTCCGCCAACGGTCCCCAGCCG
CAGCCTCAACCCACCGCCCTTAGGCTCATGCCTTCTCAGGCGACTCAAATCGGCACGCCTCCAGCCTCCCTGTACAG
CAGCACCCAAAGTGGCAACGGCCTCACCACACCATCGTTTACTGCCAACAACTCCATCGCTGATTCCATCTCGCAAG
GTTTGCAGACTTCCAACTCTCTAACAGGCAACAACTACTCTCTCAACCTTCAAAGTAGTAGTCTAGGTTTTGGAAAT
AACACTCTGGGTACCCTGGGATCCCCTTCCATCGGCAGCCTTGGAGGTATGGGCAGCGGATCCAGTATGGCCCGTCG
TGACTCCCTCGAACGCGCCTCCGTTCTGTCTCCTCTGTCTCAAGATTTCCGAGCTCCTAAATGGACGAGCAGTTACG
GTGCTGTAGGCAACGTCGCTGCATCTCCTGGCCCAATGCTAGGGGGAAGTCTAACGCCACCGCCTCTGGGCGTTACC
TCAGGACTTGGATTAGGAAATATTGTATCAGGGAACAGGATGATCTCGGCAGCTCCTGGTGCTGAAGCCAAATTTAG
GAACGGTGCTAGCAATGGTGGCTACTTCAGTTCATCCAACACGTCGGCTCTCTTCCCCACTCTCAAGAAGAACATCA
ACAATGACAAGTGCGCTGGCCGGTCTCGGCTTTTGGAAGATTTTCGAAATAACCGTTATCCAAACCTTCAGCTTCGT
GATCTTGTCAACCACATTGTGGAATTCAGCCAGGATCAACATGGATCGAGGTTTATTCAGCAGAAACTCGAACGTGC
CACTCCCGCTGACAAGCAGATGGTCTTTGCGGAGATACTAAACGCCTCATACTCCCTCATGACTGATGTTTTCGGGA
ATTACGTTATCCAGAAATTCTTCGAGTTTGGTTCTCCCGAACAAAAAAGCATCCTGGCCAGCCGCATTAGAGGTCAC
GTGTTGCCGCTGGCGCTGCAGATGTACGGCTGTCGTGTTATCCAGAAAGCTCTGGAATGCATCACCCAGGAGCAGCA
GAAGGAAATTGTGCGCGAACTGGATGGACATGTTCTCAAATGTGTCAAGGACCAAAATGGAAACCACGTAGTGCAGA
AGTGCATCGAGTGTGTGGAGCCTGGCGCCCTACAGTTCATCATCAACGCGTTCCAAGGTCAAGTGTTCAGTCTGTCA
ACTCACCCATACGGTTGTCGCGTGATCCAGCGGATCCTAGAACACTGCACTGCTGAGCAGACGGGGCCGGTGCTGGA
GGAGTTGCATCAGCATACCGATCAACTGCTCCAAGATCAGTATGGCAATTACGTCATTCAGCATGTGCTCGAACATG
GGAAGACCGAAGAGAAGAGCAAGATCGTGAACTGTGTCCGTGGCGCGGTACTGAAGCTGTCGCAGCATAAATTTGCG
AGCAACGTTGTCGAGAAGTGTGTCACTCATGCTACAAGATCTGAACGTGCAATGCTCATTGAAGAGGTGTGCGGCGG
CTACAACGACACAAGTCACAGCACTCTCTTAGTGCTTATGAAAGATCAGTATGCCAATTACGTGGTGCAGAAGATGA
TAGATGTTGCTGAGCCTGCTCAGCGCAAGATACTGATGCATAAGATCAGGCCGCACATTGCTACTCTGCGCAAGTAC
ACTTACGGCAAGCACATCTTGGCTAAATTGGAGAAATACTTCTTGAAGAATAACGACTCCTTGGCTGGCAGCATGGG
CAACTCGATGTAA

phaw Vasa
>MSTRG.26077

ATGAGTGATTGGGAAGAAGAGGAAGCAGCAGCAGGAAGCTGTGAAGTTGTGGAATGGAACGGAGGAGCCGAAACCAA
TGGTTACGGCTCCGGGAGTCGAGGGCGTGGCGGTTTCGGTGGTGGTGGCAGAGGCCGAGGAGGTGGTCGAGGAGGTT
CGACCGGTTGCAGGAAGTGTGGCGAGGAAGGCCATAGAGCGTTCGAGTGCACTTCTGGAGGTGGCGGCGGCGGCGGC
AACAGAGCTTGTTTCAAGTGTGGGGAAGAAGGGCACATGTCGCGCGAGTGTCCCCAGGGAGGAGGACAATCGTTCGG
CGGTGGTGGCGGCGGTAACAGAGGCTGCTTCAAGTGTGGAGAAGAAGGGCACATGTCCCGTGACTGTCCCAACAGCG
GTGGTGGTGGTGGTTCTGGCGGGAAAGGCTGCTTCAAGTGTGGAGAAGAAGGACACATGTCTCGTGACTGCCCTAGC
AGCGGTAATGGCGGCGGCANAGGTTGCTTCAAGTGCGGCGAGGACGGCCATATGGCTAGAGACTGTCCTCAAGGTGG
AGATGGTGGTGGCGGTGGCGGCGGCAACAGGGGTTGCTTCAACTGTGGAGAGCAAGGGCACAACAAAGCTGATTGCC
CCAACCCTCCCAAGGACAACACGGGCGAGCTGGGTCCTGACGGCAAGCCCAGACCGCCGCTCTACGTACCTGAGCAC
ATTGCTGATGAGCAGCTCTTTTCTGAGGGTGTTAATCCCGGCATCAACTTCGACGCCTACCACAACATCCCAGTCTC
GGTCTCCGGTGAAGGAGAGATTCCGGATCCCATCGATACTTTTGGCGCTAGCGGTCTCCGCGATCTGCTCATTTCCA
ACATTGAAAGGGCTGGGTATAAAACGCCCACACCCATCCAGCGTGTGTGCATCCCTACCATTATGGCGGGCAGGGAC
ATCATGGGCTGTGCTCAGACTGGATCCGGCAAAACGGCTGCTTTCCTTCTGCCCATTTTGCACGGGATCCTGGCAAG
CGGTGGCGGAAACAGTGGCTCCATGTCGAGCACTGCGGAACCCTCGGCTGTAGTTGTGGCCCCCACTCGCGAGCTCG
CCATTCAAATACACAACGAGGCGAGAAAATTCGCTCTCAATAGCATCGTGCGCACCGTAGTTTGTTACGGCGGTGCT
TCCATGAACAGCCAGTACCGCCAGCTCCAGAACGGATGTGCTGTGTTGGTGGCTACCCCGGGTCGCCTCAATGACTT
CGTGACCCGCGGACGCGTGAGCTTCTCCAGCGTCAGGTACCTGGTGCTGGACGAGGCGGACCGCATGCTCGACATGG
GCTTCATTGGTGATATTGAGAAGATCGTTAACCATCAGACGATGCCTGCCGTGGGACAGCGACAGACATTGCTCTTC
AGCGCAACCTTCCCTGAAGAGATCCAAACATTGGCGTGCAACCACCTGCAGAATTACGTCTTCTACGCTGTGGGAAC
TGTGGGCGCGGCCAACACGGACGTGTGCCAGGAGGTGCTCAATGTGCCCCGCCAGCAGAAGAGAGAAGTTCTCATGT
CTAAGATAGAAGAGTTCATGGCTAACGGCGACAATAAGGTGCTCATCTTCGTGGAGACGAAACGCACTGCTGACTTT
CTGGCCACGCTTCTGAGCTCCCAGCAGCTCAACAGCACCAGTATTCACGGCGACAGATTCCAGTCGCAGCGAGAGGA
GGCGCTCGCGCAGTTCAAAGCAGGCATCAGGTCGATCCTTGTTGCTACTGCAGTAGCTGCTCGCGGGTTGGATATCC

GTGGAGTTTCGCATGTCATCAACTACGACTTGCCCAAAGAAGTTGACGAGTACGTCCATCGCATCGGACGGACGGGG
CGCGTTGGCAACAAGGGCCACGCTGTCAGTTTCTACGATGAAGAACAGGACGGAGCTTTGGCTAAGAATTTGGTCAA
AATCTTGACTGACGCGAGTCAGGAAGTTCCAGAGTGGCTGAAGAACGCTGCTGCGCACAGCGGTCACTCCCAGACGT
ACCACGGAGTTGGAGACTTCGCTTCTCACGATATTAGAGGAGAGGGCTTGCAGCGAGAAGGGTCCTCTCAAGCATTC
GGAGGTCCAGCACCAGTGGATGAGGAAGAAGAGTGGTAA

phaw Argonaute2
>MSTRG.32662

ATGTACCCCGTAGGGCAGCCTCCTGGTCCCCCCGGCGCCCCTCCTCCTACAGGTGGTGGGGGAGCTGGCCCGGGAGG
TCCAGGGGGGCCAGCTCGTGCCCTTGCTCCGTCGCTGCCTCCTACTGCTGCTCCACTTGCTGCCCAGCCACCTGGCC
CTGCCATCACCACCATTGTCCCTGTCACCCCACACCAACCACCAGGTGCTACACAGACAGCTTTGGCATCCGGTGAC
CAAACCCAGAGCAATGTTGCCGCTGCTGCAGCTGCCGCCCTCCTCCCTACAGACCTGCCCACTTTTGTGGCTCCCCG
GCGCCCTAACCTGGGCAGAGAGGGACGCCCCATTACCCTCAGGGCCAACCACTTCCAGATCACCATGCCCCGCGGCT
ACGTACAGCACTACGACGTCAGCATCCAGCCAGACAAGTGCCCCCGTAAGGTCAATAGGGAGATCATTGAGACCATG
GTACATGCCTTCCCCCGGATATTTGGTACGCTTAAGCCTGTGTTTGATGGCCGCAGCAACTTGTACACACGCGATCC
ACTGCCTTTCAACGATAAAATCGAACTGGAGGTGACACTCCCCGGAGAAGGTCGTGACAGAGTGTTCAAAGTGGCCA
TCAAGTGGGTTGCACAAGTCAACCTCTACACTTTGGAAGAAGCTCTGGAAGGGAGAACTCGAACTATTCCTTTTGAT
GCGATCCAGGCATTGGACGTAGTGATGAGGCACCTTCCCTCCATGACATACACGCCTGTGGGGCGTTCGTTCTTCTC
GGCACCTGACGGCTATTACCATCCATTGGGAGGCGGCAGGGAGGTCTGGTTTGGCTTCCATCAGAGTGTCAGGCCCT
CCCAGTGGAAGATGATGCTCAATATTGATGTGTCGGCGACGGCATTCTACAAAGCACAGGCCGTCATCGAGTTTATG
TGCGAGGTGCTGGACATTCGAGAGATCAACGAGCAGCGCAAGGTGCTCACCGACTCTCAGCGTGTCAAGTTCACCAA
GGAAATCAAAGGCCTCAAGATCGAAATCACACATTGTGGGGCCATGCGACGCAAGTATCGAGTGTGCAACGTGACCA
GACGACCTGCCCAGATGCAGTCATTCCCACTGCAGCTGGAGAACGGCCAGACTGTGGAGTGCACCGTAGCGAAATAT
TTCTTGGACAAGTACCGTATGCGCCTTCGGTTCCCTCACTTGCCCTGCCTCCAGGTCGGTCAAGAGCACAAGCACAC
CTACCTGCCCCTGGAAGTGTGTAACATTGTTGCAGGCCAGCGCTGCATCAAGAAGCTAACTGACATGCAGACGAGCA
CCATGATCAAAGCCACGGCCCGCTCTGCCCCTGACAGGGAGAAGGAAATCAACACGCTTGTGAGGAAGGCAGACTTT
AACAATGATCCCTACATGAGGGAGTTTGGGCTTACCATCAGCAATGCCATGATGGAAGTCCGTGGCCGAGTGCTGCC
GCCTCCCCGCCTTCAGTATGGCGGGAGGTCTAGGCAGCAGGCCATGCCTAACCAAGGTGTGTGGGACATGCGAGGTA
AGCAGTTCTTTACTGGTGTAGAAATTAGAGTCTGGGCCATCGCTTGCTTTGCTCCCTCTCGTACCGTGAAGGAAGAT
TCTCTTCGTAACTTCACGACGCAGCTGCAAAAAATCAGTAATGATGCCGGCATGCCTATCATGGGACAGCCCTGCTT
CTGTAAATACGCCACTGGACCTGACCAAGTCGAGCCCATGTTCCGCTACCTCAAGTCCCAGTTCACAGGTCTGCAGC
TTGTTTGTGTTGTTCTGCCTGGCAAAACTCCCGTGTATGCTGAAGTTAAGCGTGTCGGTGACACGGTGCTGGGTATG
GCTACCCAGTGTGTCCAAGCGAAGAATGTCAACAAAACATCCCCTCAGACCCTCTCTAATCTTTGTCTAAAAATCAA
CGTCAAACTCGGCGGTATCAACTCCATCCTAGTGCCTACTATCAGGCCTAAAGTTTTCAATGAACCTGTCATATTCC
TCGGTGCAGATGTAACGCATCCTCCTGCTGGTGACAATAAGAAGCCATCCATCGCTGCTGTGGTCGGCAGCATGGAT
GCCCATCCGTCCCGCTACGCTGCCACCGTCAGGGTTCAGCAACATCGTCAGGTCGGTGGCATGAGGGAAGGTAAGGA
AGAGGTCATCACCGAATTCTCCTCAATGGTGAAGGAACTGCTCATTCAGTTCTACAAGTGCACTCGTTTCAAACCCA
ACCGAATTATCCTGTACCGTGATGGTGTGAGTGAGGGGCAGTTCCAGACGGTCCTGCAGCACGAGCTCACTGCCATG
AGGGAAGCCTGCATCAGGCTAGAGAGCGAGTACAAGCCCGGCATCACGTATATTGCCGTCCAGAAGCGCCACCACAC
CAGGCTCTTCTGCTCCGACAAGAAGGAACAAAGTGGCAAGTCTGGCAATATCCCAGCAGGTACCACAGTTGACGTCG
GTATTACCCATCCAACTGAGTTTGATTTCTACCTATGTTCTCACCAAGGTATTCAGGGCACCAGCCGACCGTCGCAT
TACCACGTACTTTGGGACGATAATCATTTCGACAGCGACGAGCTGCAGTGCTTGACGTACCAGTTGTGCCACACGTA
CGTACGGTGTACGCGGTCCGTGTCTATACCTGCTCCTGCGTACTATGCCCATCTCGTGGCCTTCCGAGCCCGGTACC
ACTTGGTTGAGAAGGAGCACGACAGTGGAGAAGGTTCTCACCAGAGCGGGAATAGCGAGGACCGAACGCCCACTGCT
ATGGCTCGTGCCGTCACCGTGCACACCGATACCAACAAAGTCATGTACTTCGCTTAA

phaw Hunchback
>MSTRG.22414

ATGGCTGATTTTGGAAGCGTGGATTCAGGGCCTGTAGCTACTCAGGCTAGAGCTACTGCAGGGTCCAGCACAGCGAC
AGCGCAACTTGACTTAGGATCCCTAACGGTGGACGGGCAGCAACACCTGTACACCGAACAGCAAAAGTTCATTGCTG
CTTTGTCACGTGAGTCGTTTCTTCGCCGTCAGCAGCAGCTTCTTTCAACTTCTTCCATTGTGCCCCAGCGAGCAAAC
TTGCTTCACGCCATGAGTAGTGGGGAATACTACTCTCACAACGAGGTACCACGCGTAGTCAATATGCACTCCCTCCC
ACAGTGGCCCATGCATGCACCAAAGAGAATTTTGCCAAATACCACTGAGAGGCCCCGTGAGCCGCCACCTCTACCAC
CAGAAGAACCGATAAGAGTGCCTTCTGCCACAAGTTCTTTGAAGAGAAAATCTCCATCGCCTGACGGTTTTACTCCA

AAGCGCACTCGATACTCGCCCGAATTTATGTTGAATCCACTGCTGTATTATTCACTTCAAACTGCGCAAAAACCTTA
TGGCGGCAGTCCTCAAGGAATATCCGACTCTAGTCCTTTGCCAATTCAGGAGCCGCAAAGGCATTTTCTGGAAGCGC
CATCGTGTCAAAATAGTCCTTCAGAACCATCTCCCGTCCAAAGATCAACAGCGTCTTCTTCGACGTCAGAGTTACAT
GCCATGAAGACTGAACCACCTGAATACAACTCTCTGCCTACTACAACACTTTTATCTGTTTTAAATTCTTTGAAAAA
AGAAAATACTGAAAACAACTATAGCTTACCAAAAACATCAATCATTAATGCATCCAAAGCCGACAATTCACATGATG
CTCGAGGAGCAGCAGCTATACAAGTTAAGAAGGAACAAGATTTATGTAATCCGAAAGAGATTCAGAGCAACTTGGAA
GCTGCTGAAAAAGAAATTTCTGAGAATATCAGAACTGACGACTTTGACGAGTCCAGCAATTCGTCACAAAAAGATTC
TGACTCGAATGATCCAGTAAATGGCAGTGCGGATCATAACAGAACGGAGGCTTCGCAACGCGAGAACCACCTCAGCT
GTATGTTGTGCGAGTTTGCCACGGACGATGTTGTCATCTATAACGATCACGTGCAGTCTGTTCACCGTAATGCGTGT
TTTGAATGTAGGTATGTCGCTGAAAATGCCACCCAGTTGTTAGAGCATCGCAAGGTCTGCGGGGGCAGCTGTTCTCG
CGGTCCTGCAGCTCCGTCTTCTTCTTCTTCTTCAGCAAAACCTAAAATTAAATCCAAAAAGCTCTTGTGCAAGA
CCTGTAAGGAAATTTCAGAGAGCGAAGAAGAATTTTATCAGCACAGACGGCAGCATATTAGCCCTGACAAATTACTG
TGTTGTTCATTCTGTCCTTTCGTGACCCAATACAAACACCATCTCGACTACCACGTGAAAAATCACGCAGGCAACAA
ACCATTTAAATGCCAAAAATGTGAATACTCTTGCGTGAATAAATCAATGTTAAACTCGCACATGAAGTCACACAGTA
ATTTTTATTCGTACAGGTGCAAGGATTGTAATTACGAAGCAAAATACATGCATGCCCTGAAGTGTCATTGTCGTAAG
TACAATCACGGGGCATTGCCTGTTCTTAACCCAGACGGTACCCCTAATCCGTACCCCATTATTGATATATACGGAAC
TAGGCGTGGGCCCAAAGTAAAGAGAGATGAAGAAGGCAACATTATTTATCCATCTAATTCTCAATATACCGGCAAAT
TCGCCATGCAGCTAGCCGATAACTCGGCGACAACTTCTGTTTCTCCTGTCACTGATACCCCCAGTACTTCCACGGTT
TCTCCCGTGCCATCGTTAGTATTTCCAACTAATCCTCTATATCAGCATGCTGCCACAAGTTTGCCTACGTTCTCTAG
TCTTTCGCCGCTCGATAACAGATTATTTGGCAGCCACGCTCCTCTTTTCAACAGTGCTTGTTCAACTTTAGCATATC
AGCAGCCTTCGGCTTTGAATTCTCCCCAATCCCCAGCTACGTCAACTCCAAGAAGTATTTTCGAGTCTCTTGGCTTT
TCCGCCGCGGTAAGAAAGGATTCAACTGAAGCAGAGACTGCCCCAAGCACTCTAAGGTGTGAGCCTTGTGGGATTCT
CTTTTCATCAAAAGAACAAGCGGGCAACCATTTCATAGTCTATCATATTTTACCGAACAGGTCCTTTGCTCATGGCT
CGGGTACCCCTCTTCGAGGCGCTTCTAGATCAGACAGCGTAGCTTCTAACGAGCCCATTTGGTCAAAACTCTTCAGA
CCGTCGGGACCCGCGTTTCCTCCGATTCCGAAAGCACCAGATTTCTTGCAACACCAGCCTCCAGTTGAGGAGGTTTC
AAACTTCGACTGTGTCGAGGAAGCGGAGGGCCGAGAGGATGCAGGCTCTCCTTCTCCACTGGACCTGACCACAGACC
ACACTCCTCCTCAACAGATCATGCGGAGGGGATTGTCAGAAACTCCACCAGATGTCGCAGCTACATCCCTGACACCT
CCGGACACCTGCTCTCCTCCCAAAAAACCACCGAGGCGCAGTGACCACCTCACGCTCACAAACAGGCGCTTCGAAAA
TATGGAAGAGGAAGGAACTGCAGTCACCTGCCAGGATCAGAGAGGCCAAGAAGCTTCGGCAGCTGCTCTTGATAAAA
ATGAACCTGAAGAACTATTGGCCGAAAAAGAAACAACCAGAGCTACGTCTGAAAAACCGCTCTCCGAAGATGCATTG
GAGTGTTCCATTTGTAAAATCACTTTCCGCGTGAGGGAAATGTACAATACGCACATCGCCTTCCACGACGACGAAGA
ACCTCTGAAATGTTCCTTCTGCAAGGAAAAGTTCGAAGATTCATCTGTATTCTTTGAGCATTGCTTCAACTGCAAAC
ATCGTCATTGTGTTTCAGCGTCGAGAAAATGTTCCACAACAGCCAATGAAACAGAAATGTCGAGGGCACGGTAA

phaw Snail
>MSTRG.54290

ATGCCAAGTTCATTCATGGCGAAACGAAATTACGCACACTGCCCCCTCAAGAAGAGGCCCCTTGCTGCCCTCTACGA
GCAAGACGAGCCCGAAGATTTAAGCGTGAAGAAACCCAAAATTGAAGATGAAGAGCGAGTGCCAAAAGTGGAAGCAG
TGACTCCAGTTCCGAGTTTTTCTTCATCAGCGGCACTCACGCCTCCGCCTGCCAGCCCGCCACCAGCCAACGTTGCT
TCTCTGCCAACTTCTTCCGCGCGGGAGTCAACACAAGACAGTTTCTTGGGCAAGAGCCCTCTACCAAATTACTATCC
AAGTGTTCTGCCATCGTACAGTGTGTCTGTTAAGCCATGTGTAAGCCAGCCGTACTCTCCTGTGCCTTTTACTCGTC
ACGTACTGCCGGTTTCATCGTACGCGTTCAGTGCTCCAGTTAGACAGCGGACTTTTGTGCCACCGGTGGTCACGTCC
ATCAACAATTCCCGTGTGCCCACCACAAGTCCTCCAGTATCATCACCTACGGTGGACACCACCAGCAGGCCATCAAG
AGATTCGCAAATGCCAAACTTGGCCGAGTATTACTTGCAGCTATTCAGGGCCACAAGTGAGGCCGAGTCCAGGGCCT
CATCGTGGATGCCTACCTCGCACATGCCACCTTCTCCTCAAGCCTCACCCCGGTCGGGGGATCCGTGTCTGTTATCA
CCTCCTGAGATCCACTACCCAGAGACGTCTGCACTTATTTCTCCAGCTTCTTCCATATCGACGATATCAGAATCTGA
AGAAGAACCATTATCGCCCAGTGGCAGCGTCAGCAGCCAGACATCATCTACCTCCAAGGCTCGCTACAACTGCAGTG
AGTGCGGTAAGGCTTACGCCACCAGACCTGGTCTCGCTAAACATAAGGAACTACACTGCACAGCTACGTTAGGCGTC
CGTGCTTACTCGTGCCCTACATGTAACAAATCCTACACCACGTGTGGGGCTCTCAAGATGCACATTCGCACCCATAC
CTTGCCTTGCAAGTGTAACATCTGCGGCAAGTCTTTCTCGCGACCGTGGCTGCTGCAGGGCCACATCCGCACCCACA
CCGGAGAGAAGCCCTTCCAGTGCACAGAATGTTGTCGTAGTTTTGCTGACCGCAGCAATTTAAGAGCCCATCAGCAG
ACCCACGCAACTGTCAAAAAATATGCCTGCGAGATCTGCACTAAGACTTTCTCCAGAATGTCGCTGCTCAACAAACA
TAAAGAAGGCACCTGCACTCCAGGGTGCCCCAGCAACGAGCTTGAGGAAGACGAAGAAGTTGACGTTGTGGGTGAAG
ACAACTGGTCTTAG

phaw Armadillo
>MSTRG.18049

ATGAGCTACCAGATGTCACAACAGGCTCGACCATTGTCACAGAGTGGAAGTTACCAGCACCCTGGTGATATGCCACT
TGCCAACCCCAAGGATCAAGTTGCCATGTGGCAACAGTCTCCCTACATGGCTGACTCGGGCATCCACTCAGCCACCA
CAACACAAGCTCCGTCCTTGAGTGGCAAGGAGGAGGACATGGACACCAGTGGCGTGCCCACTCCTGGCACGCAGGAG
TGGATTGAAGGCCAGCAGCAGCCTGGTGGCGGCTTTGCTCAGGCTGGCTTTACTTCTGATCAGGTGGATGAGATGAA
CCAGCAGCTGAATCAAACCAGGTCTCAGCGTGTGAGGGCAGCCATGTTCCCAGAGACGCTGGATGAGGGCCTTGAGA
TTCCTTCCACTCAATACGATCCTTCCCAGCCCACCGCAGTGCAGCGGCTCGCTGAGCCTTCTCAGATGCTGAAACAT
GCAGTTGTGAATCTCATCAATTACCAGGACGATGCAGATTTGGCTACACGCGCCGTGCCTGAGCTCATCAAGCTGCT
CAATGATGAGGACCAGGTGGTGGTGTCCCAGGCCGCTATGATGGTGCACAGGCTGAGCAAGAAGGAAGCGTCTCGCC
ACGCCATCATGAACTCCCAGCAGATGGTTGCTGCTCTGGTGCGTGCTCTCAACAACTCCTCCGACCTGGAGACTACG
CGTGTCACTGTGGGAGCTCTCCACAACCTGTCCCACCACCGCCAGGGGCTGCTTGCTATCTTCAAAAGTGGCGGTAT
TCCCGCGCTCGTGAAGTTGCTCAGTTCGCCGGTGGAGAGCGTCCTGTTCTATGCCATCACTACCCTGCACAGTCTCC
TGCTGCATCAGGAGGGCAGTAAGACTGCTGTCCGGATGGCTGGCGGTCTGCAGAAGATGGTGGCTCTCCTTCAGAGG
AATAATGTTAAATTCCTCGCTATTGTCACTGACTGCCTTCAGATTTTGGCTTATGGCAACCAAGAGAGCAAACTCAT
CATCCTTGCCTCTCAAGGCCCTCAGGAGCTGGTGCGCATCATGCGCTCATACACGTACGAGAAGCTCCTGTGGACGA
CATCCAGAGTGCTGAAGGTGTTGTCGGTGTGCTCCAGCAACAAGCCTGCCCTTGTGGAGGCTGGTGGTATGCAGGCC
TTGGCTATGCACCTCAACCACCAGAGCCACCGACTCCTGCAGAACTGCCTGTGGACCCTGCGAAATCTCTCTGATGC
AGGCACCAAGGTTGAGAACCTGGACAATTTGCTGCAGAGCCTGGTGCAGTTGCTGCGTTCCAACGATATCAATGTGG
TCACATGCGCTGCCGGCATCCTCTCCAATCTCACGTGCAACAACCAAAGGAACAAAGTCACTGTGTGCCAGGTGGGT
GGCATCGAGGCACTCGTATCAACAATATACAATGCAGGAGAGCGAGAGGAGATCACAGAACCGGCAGTGTGTGCCCT
GCGGCACCTCACCGCCCGACACCCCGAGTCTGATGTGGGTCAGAACACGGTGCGCCTCAAGAATGGTATCCCGGTCA
TTGTGAACCTCCTCCATCCACCCTCACGGTGGCCCCTTGTCAAAGCTGTTATTGGCCTCATCAGGAACCTTGCCCTG
TGTCCTGCCAACCACGCTCCCCTCAGGGAAGCTGGGGCCATCCCTCGTCTTGTGCAGCTTCTCATCAGGGCTTTCCA
AGACACACAGCGGCAACGAGGCTCTGTGGCCAGTGGAGGCAGTGGCAGCGCTGGTGGGGCTTATGCTGATGGTGTGC
GTATGGAGGAGATCGTGGAGGGCACTGTGGGGGCTCTCCACATCCTGGCCCGAGAGTCTCACAACAGGGCTATTATC
CGTGGCCTCAACGTCATCCCAATCTTCGTACAGCTCTTGTACAACGAGATAGAGAATATCGAGCGCGTGGCGGCGGG
AGTGCTCTGTGAACTCGCCGCTGATAAGGAAGGTGCGGAACTTATTGAACAGGACGGCGCCACGGCACCACTCACTG
AACTACTGCACTCTCGCAATGAGGGCGTCGCAACATACGCGGCCGCTGTGCTGTTCCGTATGTCAGAAGACAAGCCG
CAGGACTATAAGAAACGCCTCAGTCTGGAGCTCACAAACTCCTTACTACGCGAGGACGGCGGCGTGTGGGGAGGAGA
TCTGTCACAAATGCCGCCCGATATGCAAGAGTTTATGGCTCCTCCTGAGGGAGGTTATGACCCTATGTACGGACAAG
GCCCTCCCAGTGTCCACAGTGCTGGTCATCCTCCACCGCAGAATTATCCCGGCGGCTATGACCCTCACCTGCCTGTG
GAATCTCTACAAGGACTTGACATCGCAGCAGCACAACAGCATCGTTCAGGTGGCAGTGGCAGTGTAGTTGGTGGAGG
AGGCGGTGGTGGCAGTAACTATGGTACCCTGGACCCTATGGAACTGCCACCTGGTCACCCCGCTGCTGCCCCGCCTC
CTCATCCCGCCGATATCCCTCCGTCTGGACAGCCCGGTTACGATCAGCAAGTTCCACCGCAGCCTCCCCGCGACAAC
AATCAGATGGCGGCCTGGTATGATACGGACCTCTAA

Appendix B: Details of primer sequences for CRISPR/Cas9 knock-in constructs

Primer	Primer Sequence (5' → 3')
Zelda_For	ATGGAAGAAACCACCCTTC
Zelda_Rev	TCAGTAGACAGCCGGTAATG
GA pGEM-CDS For	GCTCTCCCATATGGTCGACCCTACTTACCCTTCTGCCACG
GA CDS-3XFLAG Rev	TAATCTCCAGAACCACCTAAGTAGACAGCCGGTAATGAGG
GA CDS-3XFLAG For	CCTCATTACCGGCTGTCTACTTAGGTGGTTCTGGAGATTA
GA 3XFLAG-4C Rev	CAGCAATTCAAAAACGATCCGCTTCCTGAAACCTTATC
GA 3XFLAG-4C For	GATAAGGTTTCAGGAAGCGGATCGTTTTTGAATTGCTG
GA 4C-pGEM Rev	CATGGCGGCCGCGGGGCTGCAAGGCGATTAAGTT
GA mRuby-CDS For	AATTGGGCCCGACGTCGCTACTTACCCTTCTG
GA 3XFLAG-pGEM Rev	GAGTGAGCCGCGGCCCTCCGAGCTTCCTGAAACCTT
Flag-4C-BamHI-F	CGCGGATCCATGGATTACAAGGATGACGAC
Flag-4C-SacII-R	TCCCCGCGGTTAGGGCTCCATGCAGCAG

Appendix C: Details of primer sequences for *in situ* hybridization

Gene Name	Gene ID	Primer	Primer Sequence (5' → 3')
Mex3	MSTRG.64035	mex3_For	TACAGCCATCCTAAGCCAAT
		mex3_Rev	ACATCGAAGTCTGGCAAAAT
		mex3_Rev+T7	GCGTAATACGACTCACTATAGGGAGAACATCGAAGTCTGGCAAAAT
Pumilio	MSTRG.22416	pumilio_For	ATTCCCTATCCTCCAGTCTCAT
		pumilio_Rev	TATCCTTAGTGCCAGTCGGT
		pumilio_Rev+T7	GCGTAATACGACTCACTATAGGGAGATATCCTTAGTGCCAGTCGGT
Vasa	MSTRG.26077	vasa_For	GAGTGATTGGGAAGAAGAGGA
		vasa_Rev	GCCAAAAGTATCGATGGGAT
		vasa_Rev+T7	GCGTAATACGACTCACTATAGGGAGAGCCAAAAGTATCGATGGGAT
Argonaute2	MSTRG.32662	argonaute2_For	CTACAGACCTGCCCACTTTT
		argonaute2_Rev	GTCTGCATGTCAGTTAGCTTCT
		argonaute2_Rev+T7	GCGTAATACGACTCACTATAGGGAGAGTCTGCATGTCAGTTAGCTTCT
Hunchback	MSTRG.22414	hunchback_For	AGCGCAACTTGACTTAGGAT
		hunchback_Rev	GCACAACATACAGCTGAGGT
		hunchback_Rev+T7	GCGTAATACGACTCACTATAGGGAGAGCACAACATACAGCTGAGGT
Snail	MSTRG.54290	snail_For	GAAACGAAATTACGCACACT
		snail_Rev	GCTGTGCAGTGTAGTTCCTTA
		snail_Rev+T7	GCGTAATACGACTCACTATAGGGAGAGCTGTGCAGTGTAGTTCCTTA
Armadillo	MSTRG.18049	armadillo_For	TCTCCCTACATGGCTGACTC
		armadillo_Rev	GCCATAAGCCAAAATCTGAA
		armadillo_Rev+T7	GCGTAATACGACTCACTATAGGGAGAGCCATAAGCCAAAATCTGAA