RESEARCH ARTICLE

The maternal-zygotic transition and zygotic activation of the *Mnemiopsis leidyi* genome occurs within the first three cleavage cycles[†]

Short title: MZT and ZGA in the ctenophore *M. leidyi*

Phillip L Davidson[1†], Bernard J. Koch[2], Christine E. Schnitzler[2,3], Jonathan Q. Henry[4], Mark Q. Martindale[3], Andreas D. Baxevanis[2], William E. Browne[1]*

1. Department of Biology, University of Miami, Coral Gables, Florida, USA
2. Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland USA
3. Whitney Laboratory for Marine Biosciences, University of Florida, St. Augustine, Florida, USA and Department of Biology, University of Florida, Gainesville, FL 32611, USA
4. Department of Cell & Developmental Biology, University of Illinois, Urbana, Illinois, USA

†. Current affiliation: Department of Biology, Duke University, Durham, North Carolina, USA

*Corresponding author:
William E. Browne
Department of Biology, University of Miami
1301 Memorial Drive, Coral Gables, FL 33146
Phone: 305-284-3319
Email: wbrowne@bio.miami.edu

[†]This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/mrd.22926]

Additional Supporting Information may be found in the online version of this article.

Received 2 August 2017; Revised 1 October 2017; Accepted 3 October 2017
Molecular Reproduction & Development

DOI 10.1002/mrd.22926

Abstract

The maternal-zygotic transition (MZT) describes the developmental reprogramming of gene expression marked by the degradation of maternally supplied gene products and activation of the zygotic genome. While the timing and duration of the MZT vary among taxa, little is known about early stage transcriptional dynamics in the non-bilaterian phylum Ctenophora. We sought to better understand the extent of maternal mRNA loading and subsequent differential transcript abundance during the earliest stages of development by performing comprehensive RNA-sequencing-based analyses of mRNA abundance in single- and eight-cell-stage embryos in the lobate ctenophore *Mnemiopsis leidyi*. We found 1,908 contigs with significant differential abundance between single- and eight-cell stages, of which 1,208 contigs were more abundant at the single-cell stage and 700 contigs were more abundant at the eight-cell stage. Of the differentially abundant contigs, 267 were exclusively present in the eight-cell samples, providing strong evidence that both the MZT and zygotic genome activation (ZGA) have commenced by the eight-cell stage. Many highly abundant transcripts encode genes involved in molecular mechanisms critical to the MZT, such as maternal transcript degradation, serine/threonine kinase activity, and chromatin remodeling. Our results suggest that chromosomal restructuring, which is critical to ZGA and the initiation of transcriptional regulation necessary for normal development, begins by the third cleavage within 1.5 hours post-fertilization in *Mnemiopsis leidyi*.

Keywords: Ctenophora, RNAseq, gene expression, maternal transcript degradation, chromatin remodeling

1 Introduction

The maternal-zygotic transition (MZT) describes the developmental reprogramming of gene expression marked by the degradation of maternally supplied gene products and zygotic genome activation (ZGA). A number of molecular mechanisms drive the variable duration and timing of the MZT among plants and animals, shaping maternal and zygotic transcriptional dynamics and thus regulation of organismal development (Lee et al., 2014). Across metazoans (animals) there is minimal coupling between the MZT and cleavage-cycle stages. For example, among bilaterian species such as the urchin *Strongylocentrotus purpuratus* and the mouse *Mus musculus*, ZGA occurs at the single-cell stage: onset of this event begins almost immediately after fertilization in the sea urchin (Poccia et al., 1985), but does not take place until approximately 10 hours post-fertilization (hpf) in the mouse (Bouniol et al., 1995). In the zebrafish *Danio rerio*, ZGA begins around the seventh cleavage cycle, about 2 hpf (Heyn et al., 2014), while zygotic expression in the frog *Xenopus laevis* begins midway through blastula stages, about 6 hpf (Newport and Kirschner, 1982a,b). In the ecdysozoan nematode *Caenorhabditis elegans*, ZGA begins by the second cleavage cycle, less than 2 hpf (Edgar et al., 1994; Baugh et al., 2003), while in the crustacean *Parhyale hawaiensis*, MZT occurs at the 32-cell stage after germ layer specification (Nestorov et al., 2013). In the non-bilaterian cnidarian *Nematostella vectensis*, the MZT begins during formation of the embryonic blastula, between 2-7 hpf (Helm et al., 2013). Presumably the lack of temporal uniformity in zygotic transcription and chromosomal restructuring marking the onset of ZGA during early development among animals is attributable to diverse life histories and to the differential deployment of mechanisms responsible for the MZT.

The molecular drivers of maternal transcript degradation and ZGA have been extensively studied for over 30 years (reviewed by Schier, 2007; Tadros and Lipshitz, 2009; and Lee et al., 2014.). Deadenylation of mRNA transcripts (Semotok et al., 2005) and microRNA (miRNA) activity (Giraldez et al., 2006; Bushati et al., 2008) are critical for maternal mRNA destabilization and degradation, and both can be linked to maternally supplied and zygotically transcribed factors (Bashirullah et al., 1999). Transcriptional repression of the zygotic genome preceding ZGA is often a result of chromatin-mediated silencing (Newport and Kirschner, 1982a,b; Ruzov et al., 2004), lack of transcriptional machinery (Prioleau et al., 1994; Kikyo and Wolffe, 2000), or mitosis-induced transcript abortion in rapidly dividing embryos (Shermoen and O'Farrell, 1991).

Mechanisms responsible for lifting transcriptional repression and subsequent ZGA include titration of transcriptional repressors by increasing amounts of chromatin, via the nucleocytoplasmic ratio model (Newport and Kirschner, 1982a,b); transcriptional activation following initiation of gene regulatory cascades at fertilization, via the maternal-clock model (Howe et al., 1995); and large-scale chromatin remodeling that can alter transcription rates in different chromosomal regions (Prioleau et al., 1994; Kikyo and Wolffe, 2000; Stancheva and Meehan, 2000; Iwafuchi-Doi and Zaret, 2014). These mechanisms play a critical role in shifting the balance of chromatin accessibility and transcriptional complex assembly as the developing embryo transitions from dependence on maternal transcripts to activation of gene expression from the zygotic genome. Given the temporal and mechanistic variation associated with the MZT, investigating this critical early developmental period in non-bilaterians can reveal evolutionarily conserved aspects of the MZT as well as highlight potentially novel, lineage-specific aspects of MZT regulation (Helm et al., 2013).

Ctenophora is a non-bilaterian phylum in which the MZT and early developmental transcriptional dynamics remain largely unexplored. Ctenophores, or comb jellies, are gelatinous marine animals characterized by the presence of eight ciliated comb (*ctene*) rows used for locomotion (Pang and Martindale, 2008a; Dunn et al., 2015). Ctenophora represent one of the earliest branching extant metazoan lineages (Dunn et al., 2008; Hejnol et al., 2009; Ryan et al., 2013; Moroz et al., 2014; Whelan et al., 2015: Shen et al., 2017), evolving independently from other animal lineages for well over 600 million years (Wang et al., 1999; Hedges et al., 2004). A number of recent studies have focused on some of the unique and derived features of ctenophore biology, such as a highly derived mitochondrial genome (Pett et al., 2011; Kohn et al., 2012), a lack of genes required for mesoderm specification, and a nervous system missing key neurotransmitters (Ryan et al., 2013; Moroz et al., 2014).

Mnemiopsis leidyi is one of the most well-studied ctenophore species, and has emerged as a model system for ctenophore biology (e.g., Martindale and Henry, 1999; Pang and Martindale, 2008b; Yamada et al., 2010; Schnitzler et al., 2012; Presnell et al., 2016; Vandepas et al., 2017). *M. leidyi* is a pelagic lobate ctenophore and, like most ctenophore species, is simultaneously hermaphroditic. *M. leidyi* embryonic development is rapid, reaching a juvenile cydippid stage less than a day after fertilization. Identification and analyses of the MZT in ctenophores should offer additional insight into the evolution of early embryonic developmental mechanisms related to the MZT, including maternal gene product degradation and ZGA in animals. For example, miRNAs are known to play a role in maternal transcript degradation (Giraldez et al., 2006; Bushati et al., 2008); however, the *M. leidyi* genome appears to lack both miRNAs and the necessary processing machinery for miRNA biogenesis (Maxwell et al., 2012). Subsequent genomic and transcriptomic analyses of the ctenophore *Pleurobrachia bachei* have corroborated this absence of the canonical miRNA pathway in ctenophores, while also highlighting the expression of many ctenophore-specific genes during early embryogenesis (Moroz et al., 2014).

Embryonic lineage tracing in *M. leidyi* suggested differential cell fate potential as early as the four-cell stage (Martindale and Henry, 1995), and revealed a clear segregation of developmental potential at the eight-cell stage (Martindale and Henry, 1999). Subsequent cell lineage analyses paired with pharmacological reagents inhibiting transcription, protein synthesis, and DNA synthesis have confirmed that the specification of some cell fates in *M. leidyi* are governed by a cleavage clock in which determinants are spatially segregated at the eight-cell stage into the E- and M-blastomeres (Fischer et al., 2014). An additional conclusion drawn from this pharmacological study is that ZGA in *M. leidyi* initiates just prior to gastrulation, at the 60-cell stage.

Here, we present the results of RNA-sequencing based analyses of single- and eight-cell samples from *M. leidyi* embryos spanning the first 1.5 h of development. These data were used to explore differential transcript abundance during the initial segregation of developmental potential during *M. leidyi* embryogenesis (Figure 1a). In contrast to previous embryological and pharmacological studies, our transcriptomic approach provides a new perspective of the MZT and the initiation of ZGA in ctenophores. Specifically, our analyses identified gene activity associated with the MZT and the initiation of new gene expression at the eight-cell stage during early *M. leidyi* development.

2 Results

2.1 Assembly of Reference Transcriptome and Read Counts

Twelve paired-end RNA-seq samples were used to assemble a reference transcriptome, including 4 single-cell, 3 complete eight-cell, 3 eight-cell M-blastomere, and 2 eight-cell E-blastomere samples. After trimming adapters from the sequences and evaluating the sequences using appropriate quality control measures (see section 3), a *de novo* reference transcriptome was assembled using Trinity (Grabherr et al., 2011), as implemented within Agalma (Dunn et al., 2013) (Table 1). The *de novo* transcriptome assembly approach was chosen to maximize contig assembly, read mapping, and subsequent read counting, thereby providing a comprehensive analysis of mRNA abundance and transcript variation. On average, 94.6% of the reads mapped to the *de novo* transcriptome, whereas only 53.0% of the reads mapped to the reference genome (Table S1). The depressed mapping rates to the reference genome are partially a reflection of sequence heterogeneity between the *M. leidyi* population used for this study and the reference genome sequence, as well as incomplete annotation of the reference genome gene models. Our *de novo* transcriptome includes 682 contigs containing 562 unique genes with significant BLAST (Basic Local Alignment Search Tool) hits (Altschul et al., 1990) (E-value $\leq 1 \times 10^{-5}$) to the Swiss-Prot protein database (UniProt, 2015) that do not align to any *M. leidyi* gene models (File S1). We additionally scanned the *de novo* transcriptome for contamination using an alien index (Gladyshev et al., 2008), as implemented in *alien_index* (Ryan, 2014); this process recovered minimal expression of non-animal transcripts across all samples (File S2).

Alignment of reads from the 12 individual samples to the *de novo* reference transcriptome identified a total of 28,122 expressed contigs encompassing 6,216 unique *M. leidyi* gene models (Figure 1b-c; File S3). Downstream analyses were carried out at the contig-level to highlight transcript diversity and to maximize read mapping rates not attainable by directly referencing the *M. leidyi* genome assembly (see above). Read counts were further normalized for between-sample variation using RUVseq (Risso et al., 2014) (Figure 2a-b). Principal component analyses (PCA) distinguished single-cell expression profiles from eight-cell expression profiles, with the first two principal components accounting for 47% of the variation, indicating substantial variation between single-cell and eight-cell stages (Figure 2c-d).

Despite capturing significant variation, PCA analysis does not discriminate between the occurrence of maternal mRNA degradation and/or zygotic transcription, as either process could produce different expression profiles at the single- and eight-cell stage. One way to gain insight into the roles of these developmental mechanisms was to closely examine the eight-cell samples in the PCA (Figure 2c). Complete eight-cell embryos, isolated eight-cell M-blastomere, and isolated eight-cell E-blastomere samples did not group as distinct sample types in the analysis, as reflected in less-robust pairwise differential abundance between these three sample groups. Importantly, the eight-cell samples did group at the individual specimen level, regardless of sample type (i.e., complete, M-, or E-blastomere specific; Figure 2c). One possible explanation for this result is that by the eight-cell stage, sequencing identified the differential timing of maternal transcript degradation and/or ZGA cascades at the individual-sample level. Thus, gene expression patterns associated with a specimen slightly further along or behind in development are reflected in read-frequency data as being slightly out of synchrony between individuals, but highly correlated within samples from a given individual. As a result, the differential timing of large-scale transcript degradation pathways or the initiation of zygotic transcription can magnify between-sample transcriptional dynamics within the 1.5-hpf period examined. Two plausible scenarios still remain: either maternal transcript degradation alone or a combination of transcript degradation and zygotic transcription could be occurring at this stage of development. Differential abundance analyses were carried out in an attempt to distinguish between these two possibilities.

2.2 Differential Abundance Analysis

Noise in the differential abundance analysis was reduced by requiring a minimum mapping threshold of at least five reads in two or more samples to identify reference contigs for inclusion in our analysis using edgeR (Robinson et al., 2010). We generated a final set of 23,326 reference contigs for differential abundance analysis. We defined differentially abundant contigs as having an absolute log2 fold change ≥ 2 between sample groups supported by a false-discovery rate ≤ 0.05 (Figure 3; File S4). We identified 1,908 contigs as significantly differentially abundant between single-cell (0 hpf) and complete eight-cell samples (Figure 3a-b; Figure S1). Of these contigs, 1,208 had a higher representation in the maternally loaded single-cell stage, whereas 700 had a higher representation at the eight-cell stage (Figure 3b), which is consistent with initiation of the MZT. Critically, 217 contigs are exclusively present in the single-cell stage and 267 contigs are exclusively present in the eight-cell stage (Figure 3b-c). The presence of 267 contigs in the eight-cell samples that are completely absent in single-cell samples rules out the possibility that all 700 contigs with higher representation at the eight-cell stage are due solely to the differential processing of maternally supplied mRNAs. Taken together, these results confirm both the MZT and ZGA have initiated by the third cell cycle in *M. leidyi* at ~1.5 hpf.

Turnover of transcript abundance can be visualized as a heat map of log counts per million (Figure 3c; Figure S1). Many contig clusters decrease in relative abundance from single-cell to eight-cell stages (Figure S1), reflecting either the degradation of maternally loaded mRNAs, the dilution of maternal transcripts by nascent zygotic transcription, or possibly both processes. Conversely, a number of contig clusters also increase between single- and eight-cell stages, supporting zygotic transcription (Figure S1). Expression dynamics at the individual-level uncovered some variable expression changes associated with contig clusters between samples, lending additional support to the idea that the differential abundance analyses are sensitive to slight differences in timing of gene expression changes between individuals (Figure 3c; Figure S1). Despite some noise detected in expression counts, contig annotation revealed common MZT processes at work.

We also carried out three pairwise comparisons between complete eight-cell, isolated eight-cell E-blastomeres, and isolated eight-cell M-blastomeres in an attempt to identify transcriptional differences associated with the segregation of developmental potential between E- and M-blastomeres at the eight-cell stage. We found 157 differentially abundant contigs between complete eight and isolated E-blastomere samples; 77 differentially abundant contigs between complete eight and isolated M-blastomere samples; and 7 differentially abundant contigs between isolated E- and isolated M-blastomere samples (File S4). The seven differentially abundant genes between E- and M-blastomere samples included a $Na^{+}/K^{+}/Ca^{2+}$ exchanger (ML261714a), Exportin2 (ML29681a), Actin-binding LIM protein 3 (ML435828a), three lineage-specific genes (ML12703a, ML065736a, ML15402a), and one contig that did not map to the reference genome (contig29091). Two possible explanations for the limited differential abundance observed between the E- and M-blastomeres are the masking of differential abundance signal due to stress-induced gene expression from mechanical separation and isolation of E- and M-blastomeres and/or variability in gene expression between individuals. Alternatively, maternally deposited proteins rather than maternal or zygotic mRNAs may be responsible for the fate determination of these cell populations. This analysis, along with the findings of Fischer et al. (2014), supports the idea that maternal proteins and their differential partitioning during the first three cleavage cycles, rather than nascent translation of mRNA, drives early cell fate specification of blastomeres during *M. leidyi* embryogenesis.

2.3 Annotation of Assembled Reference Contigs

Using the reference *de novo* transcriptome contigs as the query, BLAST identified 22,648 statistically significant hits against *M. leidyi* protein models and 11,718 statistically significant hits against the Swiss-Prot protein database (*E*-value ≤ 1 x 10^{-5}) (File S5). In addition, 14,106 contigs could be assigned one or more gene ontology terms (Ashburner et al., 2000) (File S6). Contig sequence counts of some of the most highly represented Biological Process and Molecular Function terms across the entire transcriptome, as well as broad terms that are specifically related to MZT mechanisms, are listed in Figure 4. These include regulating maternal transcripts for degradation (serine/threonine/tyrosine kinase activity); silencing and activating the zygotic genome; large-scale chromatin structural modifications; and regulation of DNA-templated transcription.

We recovered a number of contigs containing genes thought to play important roles during the MZT. Each is among the top 20% most-abundant contigs across both single- and eight-cell stage samples. The serine threonine kinase SMAUG (SMG), for example, recruits the deadenylase complex CCR4/POP2/NOT to mediate maternal transcript degradation (Semotok et al., 2005; Tadros et al., 2007; Benoit et al., 2009); we recovered expressed contigs for both the kinase *Smg1* (contig12703) and the deadenylase complex subunit *Cnot4* (contig2682). Chromatin-mediated silencing represses zygotic transcription early in development, and depletion of the methyltransferase DNMT1 results in early zygotic transcription in *Xenopus* through decreased CpG methylation (Stancheva and Meehan, 2000); we recovered a contig for *Dnmt1* (contig9358) during the first three cell cycles, in addition to the related methyltransferase *Dnmt3a* (contig2787, contig2788, contig7017). Chromatin remodelers are crucial for both zygotic transcriptional repression as well as activation. BRG1 (Brahma-related gene 1; also called SMARCA4 [Switch/sucrose non-fermentable-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A, member 4]) is a catalytic subunit of the chromatin remodeling SWI/SNF (Switch/sucrose non-fermentable)-related complexes that function in gene expression reprogramming during *M. musculus* ZGA (Bultman et al., 2006). *Smca4* (contig6831), a gene highly similar to *Brg1*, was also expressed in single- and eight-cell samples.

The OSKM transcription factor genes, *Oct4* (also called *Pou5f1*), *Sox2*, *Klf4* (*Krüppel-like factor 4*), and *c-Myc*, participate in chromatin remodeling during pluripotency-related cell reprogramming and are associated with ZGA in vertebrates (Zeng and Schultz 2005; Lee et al., 2013; Leichsenring et al., 2013; Soufi et al., 2015). During the first ~1.5 h of *M. leidyi* development, we recovered expressed contigs for homologs to the OSKM cellular reprogramming circuit. The POU class homeodomain containing gene family includes *Oct4* (Ryan and Rosenfeld 1997). We detected significant expression of *MlePou26a* (contig16781, contig16782), which contains a POU-specific domain upstream of a POU class homeodomain (Ryan et al., 2010). The Sox gene family is characterized by the presence of a high mobility group (HMG) DNA-binding domain, and the *Sox2* transcription factor, along with other members of the *Sox* gene family, play key roles in regulating cell fate during development in metazoans (Sarkar and Hochedlinger 2013). During the first three cell cycles, we detected expression of a single Sox gene, *MleSox2* (contig22607, contig22608), a member of the SoxC group (Schnitzler et al., 2014). Members of the Krüppel-like factor and specificity protein (KLF/SP) gene family are characterized by the presence of a highly conserved triple C2H2 zinc finger DNA-binding domain and play key roles in essential biological processes, including balancing stem cell proliferation and differentiation (Presnell et al., 2015; Bialkowska et al., 2017). We detected significant expression for both *MleKlf5a* (contig2761, contig16159, contig16164, contig26148) and *MleKlf5b* (contig4663, contig27551) (Presnell et al., 2015). The MYC transcription factor contains a highly conserved basic-helix-loop-helix leucine zipper (bHLH-LZ) DNA-binding domain, and plays a central role in cell growth and proliferation (Grandori et al., 2000; Young et al., 2011; Kress et al., 2015). We detected significant expression of a *Myc* ortholog (contig20351) during the first three cell cycles. Expression of these and other chromatin remodelers suggests that large-scale chromatin restructuring involved in the repression and activation of zygotic gene expression likely occurs within the first 1.5 h of *M. leidyi* development.

Predicted protein-coding sequences from the *M. leidyi* genome assembly include a large fraction of expressed lineage-specific genes (7,798/16,548) (Ryan et al., 2013), defined as having an identity derived from the *Mnemiopsis* reference genome (Moreland et al., 2014) but no orthology to genes in other organisms. Transcript profiling in the ctenophore *Pleurobrachia bachei* also identified a significant number of ctenophore-specific genes that are differentially expressed during early development (Moroz et al., 2014). Our results in *M. leidyi* similarly identified differential abundance of a number of lineage-specific genes within the first 1.5 h of development. For example, ~13% of contigs exclusively expressed in both single- and eight-cell stage embryos were lineage-specific (29 and 35 contigs respectively; File S4). Most contigs exclusively detected at the single-cell stage were characterized by highly variable expression between samples, and may reflect non-uniform loading of maternal transcripts (Figure 3c). In contrast, a number of uniformly expressed contigs exclusively detected at the eight-cell stage encode homeodomain-containing transcription factors. Among these are three contigs (contig8798, contig31508, contig10728) that respectively correspond to the ctenophore-specific SINE (*Sine oculis*) class genes, *Mlesix59a*, *Mlesix13a*, and *Mlesix13f* (Ryan et al., 2010). We also detected differential activation at the eight-cell stage of three divergent, unclassified homeodomain genes – *Mlehd07c* (contig14607, contig14733), *Mlehd11a* (contig25571), *Mlehd11b* (contig23590) – and a single ANTP (*Antennapedia*) class HOXL-related gene – *Mleantp03c* (contig23663) (Ryan et al., 2010).

3 Discussion

The results of this study demonstrate that molecular components essential for carrying out the MZT are present within the first three cleavage cycles of the ctenophore *M. leidyi*. ZGA has commenced by the eight-cell stage, ~1.5 hpf. Expression profiles of single- and eight-cell samples reflect a large turnover between maternally supplied and zygotic transcripts, and our differential abundance analyses definitively rule out the possibility that maternal gene products alone account for this difference between samples. Furthermore, annotation of the expressed transcripts is consistent with maternal transcript degradation and large-scale chromatin remodeling activity involved in both zygotic genome silencing and activation.

The early onset of ZGA in *M. leidyi* is consistent with a rapid, highly deterministic cleavage program that promotes development into a juvenile cydippid within 24 hpf. Recent RNA-sequencing analysis of development in another non-bilaterian lineage, the cnidarian *Nematostella vectensis*, also uncovered molecular players involved in the MZT and identified initiation of the MZT between 2-7 hpf, followed by a major wave of zygotic transcription between 7-12 hpf (Helm et al. 2013). Our study in *M. leidyi* provides an additional example of the variable timing of the MZT and ZGA among metazoans. Additionally, our results highlight the deep conservation of molecular mechanisms underlying MZT and ZGA, including factors associated with the OSKM cellular pluripotency reprogramming circuit.

In contrast to a recent study using actinomycin D drug treatment to inhibit transcription (Fischer et al., 2014), RNA-sequencing afforded a fine-scale granular analysis of early developmental transcriptional dynamics in *M. leidyi*. We found that ZGA initiates much earlier in *M. leidyi* than previously reported and, as in other animals, a temporal overlap between maternal transcript degradation and zygotic transcription exists – i.e. MZT checkpoints are not occurring in a sequential on/off fashion. Thus, our results highlight that the potential of maternally loaded transcripts to support the early cleavage program through gastrulation does not necessarily imply that zygotic transcription has not commenced prior to this stage of development, explaining why actinomycin D treatment alone was unable to accurately define the onset of ZGA. The temporal dynamics of these developmental mechanisms result, in part, from both mRNA transcript destabilization pathways and chromatin restructuring.

Additional fine-scale RNA-sequencing analyses in *M. leidyi* paired with assessments of chromatin states (open/closed) will offer more biological context to early genome-wide remodeling associated with ZGA. Importantly, proteomic analyses of ctenophore development may provide much needed insight into the identification and partitioning of maternally loaded proteins responsible for the remarkably stereotyped early-cleavage fate-specification program (Freeman, 1976; Martindale and Henry, 1999) and the results of our differential abundance analysis. Here, we provided a new and revised understanding of the MZT in ctenophores reflecting the deep conservation of this critical early developmental event in metazoans.

4 Materials and Methods

4.1 Collection, RNA extraction, and sequencing

Adult *M. leidyi* individuals were collected in Biscayne Bay, Miami, FL (25°28’05”N 80°12’45”W). Individual adult animals were induced to spawn following a light-dark cycle.

Single- and eight-cell embryos were collected after fertilization. For eight-cell embryo collections, synchronized spawns were monitored until the third cleavage. The egg envelopes from a subset of eight-cell embryos were removed using fine-tipped forceps and sharpened tungsten wire needles. For each individual spawn, the E- and M-blastomeres were then separated by hand with glass needles and pooled.

RNA was extracted and isolated from all 12 pooled samples using TRIzol. RNA quantity was assessed via Qubit flourometer (Invitrogen, Carlsbad, CA), and RNA quality was assessed via Bioanalyzer (Agilent Technologies, Santa Clara, CA). Paired-end sequencing for each sample was carried out on an Illumina HiSeq 2000 platform (Husman Institute of Human Genomics, Miami, FL). Sequenced RNA reads included 4 single-cell samples (0 hpf), 3 complete eight-cell samples (~1.5 hpf), and 5 mechanically separated E- and M-blastomere-specific samples from matched eight-cell samples (~1.5 hpf). Raw sequencing files can be obtained from the National Center for Biotechnology Information (NCBI) SRA repository (NCBI BioProject ID PRJNA396768).

4.2 Raw read preparation

Before transcriptome assembly, a series of trimming and quality control measures were carried out on the 12 RNA-sequencing samples. Reads were subject to Trimmomatic processing, including cropping the hexamer primers and sequencing adaptors from each read; removing low quality reads (reads with four consecutive bases averaging a sequencing score less than 15); clipping leading or trailing undefined (N) bases; and discarding reads less than 36 basepairs in length (Bolger et al., 2014). A total of 374,560,885 reads resulted from this filtering step. Next, forward and reverse reads from all 12 samples were concatenated into two separate files, and each set of forward and reverse reads were digitally normalized (kmer length: 20) using the khmer-2.2 software package to decrease sample variation and discard redundant data (Brown et al., 2012; Crusoe et al., 2015). Resultant reads with quality scores less than 33 (Andrews, 2010) were removed, and ribosomal RNAs were identified and removed. A master reference transcriptome was then assembled with Trinity-r20140413p1 (Grabherr et al., 2011) (File S7). Transcriptome preparation and assembly was executed with the Agalma-0.5.0 workflow (Dunn et al., 2013) and Biolite (Howison, 2012), and summary statistics were generated with Transrate v1.0.3 (Smith-Unna et al., 2016). Identification of putative non-metazoan gene expression contamination was carried out with the *alien_index* program (Ryan, 2014) (File S2).

4.3 Read counts, count normalization, differential abundance

Sample reads were aligned to the assembled transcriptome and the reference genome using Bowtie2-2.2.6 (Langmead and Salzberg, 2012), and RSEM-1.2.9 (Li and Dewey, 2011) was implemented to generate counts from each sample's reads against the reference transcriptome (File S3 displays a table of raw read counts). Bioconductor (Gentleman et al., 2004) RUVseq package (Risso et al., 2014) was then used to normalize between-sample count variation in an effort to mitigate batch effects and to remove lowly expressed contigs (>5 reads in 2 or more samples were kept).

PCA of the normalized reads used for differential abundance analysis was carried out in R through the RUVseq package. Differential abundance analysis of the remaining 23,326 contigs was accomplished at the contig level using the generalized linear model (GLM) approach in edgeR (Robinson et al., 2010): dispersion estimated with *estimateGLMCommonDisp* and *estimateGLMTagwiseDisp* functions; *glmFit* and *glmLRT* were used to fit the negative binomial GLM for each tag and to carry out a likelihood-ratio test, respectively. Significantly differentially abundant contigs were defined in this analysis as having a false-discovery rate ≤

0.05 and an absolute log2 FC > 2 (File S4). The heat map of log-counts per million (logCPM) of contigs was created using the mixOmics package (Le Cao et al., 2009) in R. The Venn diagram summarizing results of pairwise differential abundance analyses was generated with the VIB/UGent Venn diagram web tool (http://bioinformatics.psb.ugent.be/webtools/Venn/).

4.4 Annotation and Enrichment Analyses

Contigs within the assembled transcriptome were annotated for significant hits (E-value ≤ 1 x 10^{-5}) against the *Mnemiopsis* genome protein models (https://research.nhgri.nih.gov/mnemiopsis) and Swiss-Prot protein database (UniProt, 2015) (File S5). Blast2GO (Conesa et al., 2005) was implemented to assess and compile these annotations and to assign gene ontology terms (Ashburner et al., 2000) (File S6). Manual screening of the transcriptome was conducted for contigs of interest, including ctenophore-specific genes and genes potentially critical to the MZT, as identified from the literature.

Acknowledgements

This work was supported by startup funds from the University of Miami College of Arts and Sciences and Provost Research Award to W.E.B. This work was also supported, in part, by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health. J.Q.H. (J.J.H.) was supported by National Sciences Foundation grant IOS-1558061. We thank the Center for Computational Science at the University of Miami for the fellowship provided to P.L.D. We thank the University of Miami's John P. Hussman Institute for Human Genomics sequencing core for data generation.

Author contributions:

Conceptualization, W.E.B.; Methodology, P.L.D., B.J.K., C.E.S., A.D.B., and W.E.B.; Formal Analysis, P.L.D., B.J.K., C.E.S., A.D.B., and W.E.B.; Investigation, P.L.D., B.J.K., C.E.S., J.Q.H., M.Q.M., A.D.B., and W.E.B.; Writing – Original Draft, P.L.D., and W.E.B.; Writing – Review & Editing, P.L.D., B.J.K., C.E.S., J.Q.H., M.Q.M., A.D.B., and W.E.B.; Visualization, P.L.D. and W.E.B.; Resources, A.D.B. and W.E.B.

Conflicts of Interest

The authors declare no competing financial interests.

Abbreviations

hpf, hours post-fertilization; MZT, maternal-zygotic transition; OSKM, Oct4, Sox2, Klf4, and c-Myc*;* PCA, principal component analysis; ZGA, zygotic genome activation.

Funding

National Sciences Foundation grant IOS-1558061.
National Human Genome Research Institute, National Institutes of Health.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic Local Alignment Search Tool. J Mol Biol 215:403-410.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. Nat Genet 25:25-29.

Bashirullah A, Halsell SR, Cooperstock RL, Kloc M, Karaiskakis A, Fisher WW, Fu W, Hamilton JK, Etkin LD, Lipshitz HD. 1999. Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in *Drosophila melanogaster*. EMBO J 18:2610-2620.

Baugh LR, Hill AA, Slonim DK, Brown EL, Hunter CP. 2003. Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. Development 130:889-900.

Benoit B, He CH, Zhang F, Votruba SM, Tadros W, Westwood JT, Smibert CA, Lipshitz HD, Theurkauf WE. 2009. An essential role for the RNA-binding protein Smaug during the *Drosophila* maternal-to-zygotic transition. Development 136:923-932.

Bialkowska AB, Yang VW, Mallipattu SK. 2017. Krüppel-like factors in mammalian stem cells and development. Development 144:737-754.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Bouniol C, Nguyen E, Debey P. 1995. Endogenous transcription occurs at the 1-cell stage in the mouse embryo. Exp Cell Res 218:57-62.

Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH. 2012 A reference-free algorithm for computational normalization of shotgun sequencing data. arXiv 1203.4802.

Bultman SJ, Gebuhr TC, Pan H, Svoboda P, Schultz RM, Magnuson T. 2006. Maternal BRG1 regulates zygotic genome activation in the mouse. Genes Dev 20:1744-1754.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edvenson G, Fay S, Fenton J, Fenzl T, Fish J, Garcia-Gutierrez L, Garland P, Gluck J, Gonzalez I, Guermond S, Guo J, Gupta A, Herr JR, Howe A, Hyer A, Harpfer A, Irber L, Kidd R, Lin D, Lippi J, Mansour T, McA'Nulty P, McDonald E, Mizzi J, Murray KD, Nahum JR, Nanlohy K, Nederbragt AJ, Ortiz-Zuazaga H, Ory J, Pell J, Pepe-Ranney C, Russ ZN, Schwarz E, Scott C, Seaman J, Sievert S, Simpson J, Skennerton CT, Spencer J, Srinivasan R, Standage D, Stapleton JA, Steinman SR, Stein J, Taylor B, Trimble W, Wiencko HL, Wright M, Wyss B, Zhang Q, Zyme E, Brown CT. 2015. The khmer software package: enabling efficient nucleotide sequence analysis. F1000Res 4:900.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745-749.

Dunn CW, Howison M, Zapata F. 2013. Agalma: an automated phylogenomics workflow. BMC Bioinformatics 14:330.

Dunn CW, Leys SP, Haddock SHD. 2015. The hidden biology of sponges and ctenophores. Trends Ecol Evol 30:282-291.

Edgar LG, Wolf N, Wood WB. 1994. Early transcription in *Caenorhabditis elegans* embryos. Development 120:443-451.

Fischer AH, Pang K, Henry JQ, Martindale MQ. 2014. A cleavage clock regulates features of lineage-specific differentiation in the development of a basal branching metazoan, the ctenophore *Mnemiopsis leidyi*. Evodevo 5:4.

Freeman G. 1976. The role of cleavage in the localization of developmental potential in the ctenophore *Mnemiopsis leidyi*. Developmental Biology 49:143-177.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80.

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science 312:75-79.

Gladyshev EA, Meselson M, Arkhipova IR. 2008. Massive horizontal gene transfer in bdelloid rotifers. Science 320:1210-1213.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644-652.

Grandori C, Cowley SM, James LP, Eisenman RN. 2000. The Myc/Max/Mad network and the transcriptional control of cell behavior. Annu. Rev. Cell Dev. Biol. 16:653-699.

Hedges, SB, Blair, JE, Venturi, ML, Shoe, JL. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evolutionary Biology 4:2.

Helm RR, Siebert S, Tulin S, Smith J, Dunn CW. 2013. Characterization of differential transcript abundance through time during *Nematostella vectensis* development. BMC Genomics 14:266.

Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, Martinez P, Baguna J, Bailly X, Jondelius U, Wiens M, Muller WE, Seaver E, Wheeler WC, Martindale MQ, Giribet G, Dunn CW. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc Biol Sci 276:4261-4270.

Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, Neugebauer KM. 2014. The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep 6:285-292.

Howe JA, Howell M, Hunt T, Newport JW. 1995. Identification of a developmental timer regulating the stability of embryonic cyclin A and a new somatic A-type cyclin at gastrulation. Genes Dev 9:1164-1176.

Howison MS, Sinnott-Armstrong NA, Dunn CW. 2012. Biolite, a lightweight bioinformatics framework with automated tracking of diagnostics and provenance. Proceedings of the 4th USENIX Workshop on the Theory and Practice of Provenace (TaPP '12).

Iwafuchi-Doi M, Zaret KS. 2014. Pioneer transcription factors in cell reprograming. Genes Dev 28:2679-2692.

Kikyo N, Wolffe AP. 2000. Reprogramming nuclei: insights from cloning, nuclear transfer and heterokaryons. J Cell Sci 113(Pt 1):11-20.

Kress T, Sabò A, Amati B. 2015. MYC: connecting selective transcriptional control to global RNA production. Nat. Rev. Cancer 15:593-607.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357-359.

Le Cao KA, Gonzalez I, Dejean S. 2009. integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics 25:2855-2856.

Lee MT, Bonneau AR, Giraldez AJ. 2014. Zygotic genome activation during the maternal-to-zygotic transition. Annu Rev Cell Dev Biol 30:581-613.

Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, Giraldez AJ. 2013. Nanog, Pou5f1, and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature 503:360-364.

Leichsenring M, Maes J, Mössner R, Driever W, Onichtchouk D. 2013. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science 341:1005-1009.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Martindale MQ, Henry JQ. 1995. Diagonal development: Establishment of the anal axis in the ctenophore *Mnemiopsis leidyi*. Biological Bulletin 189, 190-192.

Martindale MQ, Henry JQ. 1999. Intracellular fate mapping in a basal metazoan, the ctenophore *Mnemiopsis leidyi*, reveals the origins of mesoderm and the existence of indeterminate cell lineages. Developmental Biology 214, 243-257.

Maxwell EK, Ryan JF, Schnitzler CE, Browne WE, Baxevanis AD. 2012. MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore *Mnemiopsis leidyi*. BMC Genomics 13:714.

Moreland RT, Nguyen A-D, Ryan JF, Schnitzler CE, Koch BJ, Siewert K, Wolfsberg TG, Baxevanis AD. 2014. A customized web portal for the genome of the ctenophore *Mnemiopsis leidyi*. BMC Genomics 15:316.

Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM, Ptitsyn A, Reshetov D, Mukherjee K, Moroz TP, Bobkova Y, Yu F, Kapitonov VV, Jurka J, Bobkov YV, Swore JJ, Girardo DO, Fodor A, Gusev F, Sanford R, Bruders R, Kittler E, Mills CE, Rast JP, Derelle R, Solovyev VV, Kondrashov FA, Swalla BJ, Sweedler JV, Rogaev EI, Halanych KM, Kohn AB. 2014. The ctenophore genome and the evolutionary origins of neural systems. Nature 510: 109-114.

Nestorov P, Battke F, Levesque MP, Gerberding M. 2013. The maternal transcriptome of the crustacean *Parhyale hawaiensis* is inherited asymmetrically to invariant cell lineages of the ectoderm and mesoderm. PLoS ONE 8(2): e56049.

Newport J, Kirschner M. 1982a. A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell 30:675-686.

Newport J, Kirschner M. 1982b. A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. Cell 30:687-696.

Pang K, Martindale MQ. 2008a. Ctenophores. Curr Biol 18: 1119-1120.

Pang K, Martindale MQ. 2008b. Comb jellies (Ctenophora): A model for basal metazoan evolution and development. Cold Spring Harb Protoc, doi:10.1101/pdb.emo106.

Pett W, Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxevanis AD, Lavrov DV. 2011. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: Insight from mtDNA and the nuclear genome. Mitochondrial DNA 22, 130-142.

Poccia D, Wolff R, Kragh S, Williamson P, 1985. RNA synthesis in male pronuclei of the sea urchin. Biochim Biophys Acta 824, 349-356.

Presnell JS, Schnitzler CE, Browne WE. 2015. KLF/SP transcription factor family evolution: expansion, diversification, and innovation in eukaryotes. Genome Biol. Evol. 7:2289-2309.

Presnell JS, Vandepas LE, Warren KJ, Swalla BJ, Amemiya CT, Browne WE. 2016. The presence of a functionally tripartite through-gut in Ctenophora has implications for metazoan character trait evolution. Curr Biol 26, 2814-2820.

Prioleau MN, Huet J, Sentenac A, Mechali M. 1994. Competition between chromatin and transcription complex assembly regulates gene expression during early development. Cell 77:439-449.

Risso D, Ngai J, Speed TP, Dudoit S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 32:896-902.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.

Ruzov A, Dunican DS, Prokhortchouk A, Pennings S, Stancheva I, Prokhortchouk E, Meehan RR. 2004. Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. Development 131:6185-6194.

Ryan JF. 2014. Alien Index: identify potential non-animal transcripts or horizontally transferred genes in animal transcriptomes. DOI:http://dx.doi.org/10.5281/zenodo.21029.

Ryan JF, Pang K, Program NICS, Mullikin JC, Martindale MQ, Baxevanis AD. 2010. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. EvoDevo 1:9.

Ryan JF, Pang K, Schnitzler CE, Nguyen AD, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, Program NICS, Smith SA, Putnam NH, Haddock SHD, Dunn CW, Wolfsberg TG, Mullikin JC, Martindale MQ, Baxevanis AD. 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science 342:1242592.

Ryan AK, Rosenfeld MG. 1997. POU domain family values: flexibility, partnerships, and developmental codes. Genes Dev. 11:1207-1225.

Sarkar A, Hochedlinger K. 2013. The Sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell 12:15-30.

Schier AF. 2007. The maternal-zygotic transition: death and birth of RNAs. Science 316:406-407.

Schnitzler CE, Pang K, Powers ML, Reitzel AM, Ryan JF, Simmons D, Tada T, Park M, Gupta J, Brooks SY, Blakesley RW, Yokoyama S, Haddock SHD, Martindale MQ, Baxevanis AD. 2012. Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidyi*: a new view of ctenophore photocytes. BMC Biology 10:107.

Schnitzler CE, Simmons D, Pang K, Martindale MQ, Baxevanis AD. 2014. Expression of multiple *Sox* genes through embryonic development in the ctenophore *Mnemiopsis leidyi* is spatially restricted to zones of cell proliferation. EvoDevo 5:15.

Semotok JL, Cooperstock RL, Pinder BD, Vari HK, Lipshitz HD, Smibert CA. 2005. Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early *Drosophila* embryo. Curr Biol 15: 284-294.

Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat Eco Evo 1: 0126.

Shermoen AW, O'Farrell PH. 1991. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell 67:303-310.

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, and Kelly S. 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res 26:1134-1144.

Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 161:555-568.

Stancheva I, Meehan RR. 2000. Transient depletion of xDnmt1 leads to premature gene activation in *Xenopus* embryos. Genes Dev 14:313-327.

Tadros W, Goldman AL, Babak T, Menzies F, Vardy L, Orr-Weaver T, Hughes TR, Westwood JT, Smibert CA, Lipshitz HD. 2007. SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. Dev Cell 12:143-155.

Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. Development 136:3033-3042.

UniProt Consortium. 2015. UniProt: a hub for protein information. Nucleic Acids Res 43:D204-212.

Vandepas LE, Warren KJ, Amemiya CT, Browne WE. 2017. Establishing and maintaining primary cell cultures from the ctenophore *Mnemiopsis leidyi*. Journal of Experimental Biology 220, 1197-1201.

Wang, DYC, Kumar, S, Hedges, SB. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc. R. Soc. Lond. B 266 163-171.

Whelan NV, Kocot KM, Moroz LL, Halanych KM. 2015. Error, signal, and the placement of Ctenophora sister to all other animals. Proc Natl Acad Sci U S A 112:5773-5778.

Yamada A, Martindale MQ, Fukui A, Tochinai S. 2010. Highly conserved functions of the *Brachyury* gene on morphogenetic movements: Insight from the early-diverging phylum Ctenophora. Developmental Biology 339, 212-222.

Young SL, Diolaiti D, Conacci-Sorrell M, Ruiz-Trillo I, Eisenman RN, King N. 2011. Premetazoan ancestry of the Myc-Max network. Mol. Biol. Evol. 28:2961-2971.

Zeng F, Schultz RM. 2005. RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. Dev. Biol. 283:40-57.

Legends

Figure 1: Experimental design, read count abundances, and contig length distribution. **(a)** RNA-sequencing was performed on pooled single-cell (0 hpf), eight-cell (~1.5 hpf), and isolated eight-cell E- and isolated M-blastomere-specific *M. leidyi* embryo samples for *de novo* transcriptome assembly and differential abundance analysis. **(b)** Read count abundance prior to normalization. A total of 374,560,403 reads distributed across the 12 samples were mapped to a reference *de novo* transcriptome. **(c)** Reference transcriptome contig length distribution. A total of 33,887 assembled contigs ranged in length from 301 to 21,725 basepairs (see Table 1), with a mean length of 1,319 basepairs.

Figure 2: Read count normalization and PCA. **(a)** Relative log expression prior to normalization, showing the ratio of expression to the median expression across all samples. **(b)** Relative log expression post-normalization. Normalized read counts across all samples were used for subsequent analyses. **(c-d)** The first principal component (PC1) distinguishes single-cell samples from each of the eight-cell samples, and explains 25.64% of the variation. The second principal component (PC2) explains 21.59% of the variation, and separates samples by individual parent. Similarly numbered labels of the eight-cell samples indicate embryos coming from the same self-fertilizing ctenophore parent. The single-cell samples all came from different self-fertilizing parents. This second principal component may also be distinguishing embryos from an individual fertilization (for example, embryos from individual '2' and '5') that are at slightly different developmental stages when sampled within the first 1.5 hpf.

Figure 3: Differential abundance of contigs. Significantly differentially abundant contigs were defined as having a log2 fold change (log2 FC) > 2 and a false-discovery rate ≤ 0.05. **(a)** Venn

diagram displaying the distribution of differentially abundant, unique contigs among the samples: 1,908 contigs between single- and complete eight-cell samples; 157 contigs between complete eight-cell and eight-cell M-blastomeres; 77 contigs between complete eight-cell and eight-cell E-blastomeres; and 7 contigs between eight-cell E- and M-blastomeres. **(b)** Of the 1,908 differentially abundant contigs (red) between single- and eight-cell samples, 1,208 are more abundant at the single-cell stage and 700 are more abundant at the eight-cell stage. Of these, 217 and 267 contigs (blue) are exclusively expressed at the single- and eight-cell stage, respectively. **(c)** Heat map of log counts per million (logCPM) of contigs exclusively expressed in either single- or eight-cell samples, highlighting clusters of gene expression associated with maternal mRNA degradation and initial activation of the zygotic genome. SC, single-cell samples; Eight, eight-cell samples.

Figure 4: Gene ontology (GO) identifiers, description, and sequence counts of the most highly represented GO terms as well as terms that are broadly related to MZT/ZGA processes. **(a-b)** Biological Process and **(c-d)** Molecular Function. GO identifiers from each respective table correspond to the *x*-axis labels.

Table 1

Title: Transcriptome Summary Statistics

Raw Reads	374,560,403	**# Contigs > 1kbp**	15,910
Contigs Assembled	33,887	**N10**	4601bp
Smallest Contig	301bp	**N50**	2819bp
Largest Contig	21,725bp	**N70**	1283bp
Mean Contig Length	1319bp	**N90**	565bp
GC Content	41.6%	**# Contigs with ORF**	17,706

Legend: Transcriptome summary statistics generated by Transrate (Smith-Unna et al., 2016). The term "contig" is used rather than "gene" to distinguish between the method of *de novo* transcriptome assembly and referencing the *Mnemiopsis* genome during RNA-seq read alignment. Open reading frame (ORF).

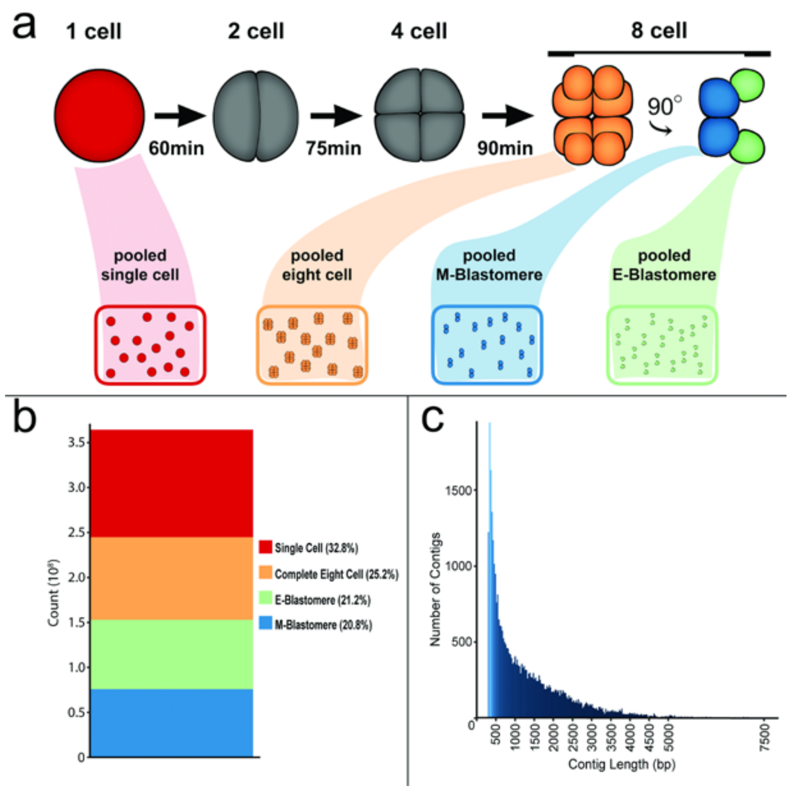

a
1 cell
2 cell
4 cell
8 cell
60min
75min
90min
90°
pooled
single cell
pooled
eight cell
pooled
M-Blastomere
pooled
E-Blastomere
b
Count (10⁸)
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Single Cell (32.8%)
Complete Eight Cell (25.2%)
E-Blastomere (21.2%)
M-Blastomere (20.8%)
c
Number of Contigs
0
500
1000
1500
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
7500
Contig Length (bp)

Figure 1

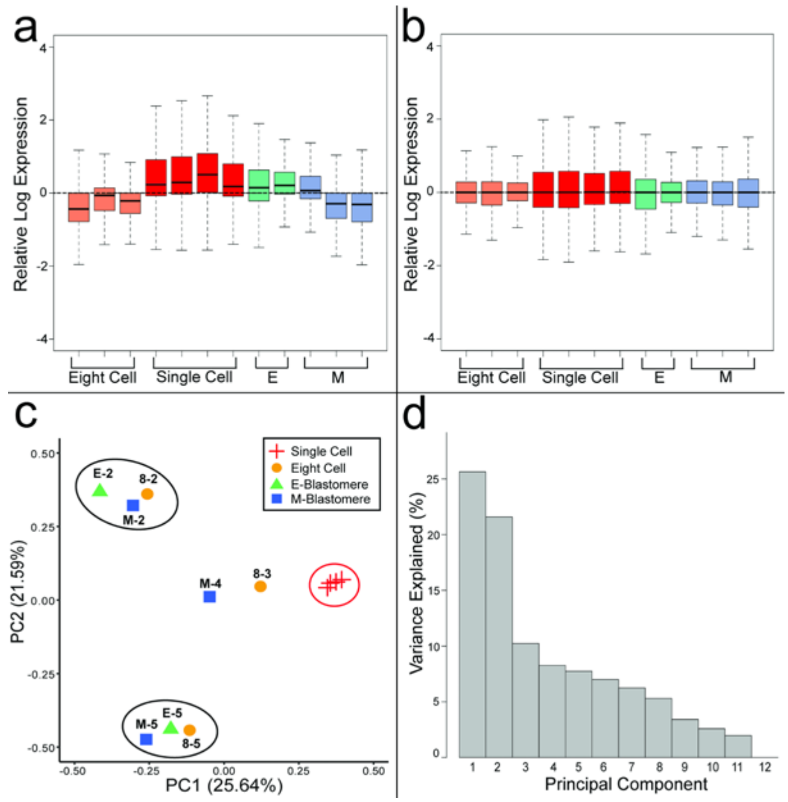

a
Relative Log Expression
Eight Cell
Single Cell
E
M
b
Relative Log Expression
Eight Cell
Single Cell
E
M
c
Single Cell
Eight Cell
E-Blastomere
M-Blastomere
E-2
8-2
M-2
M-4
8-3
M-5
E-5
8-5
PC2 (21.59%)
PC1 (25.64%)
d
Variance Explained (%)
Principal Component

Figure 2

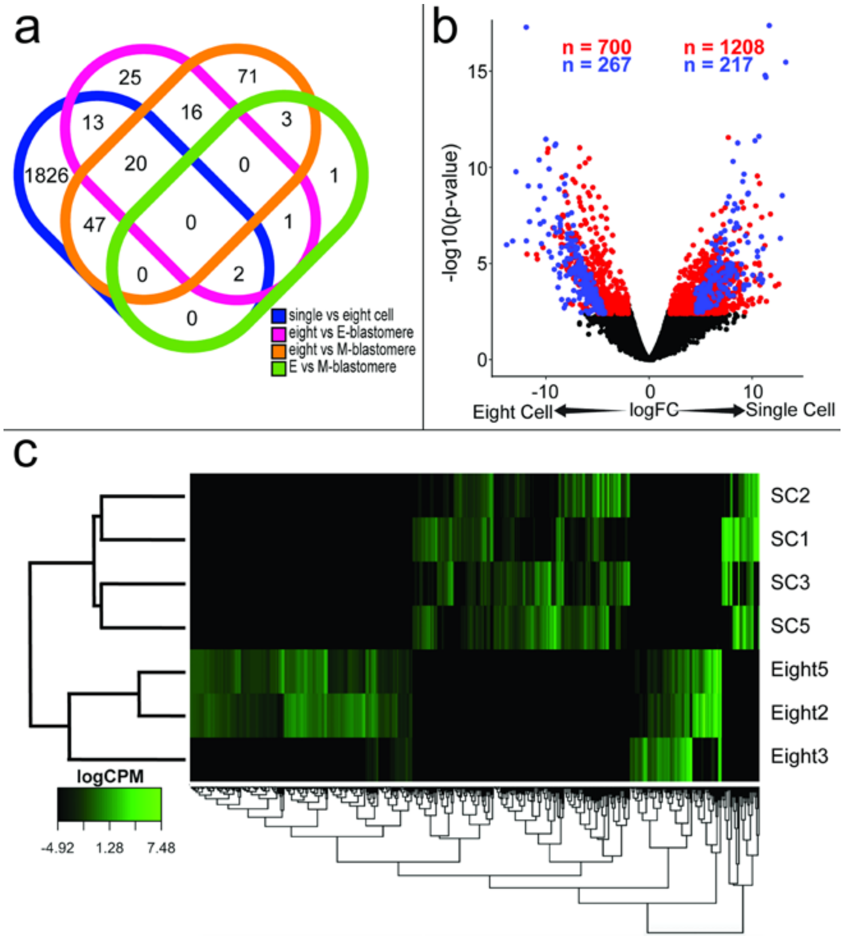
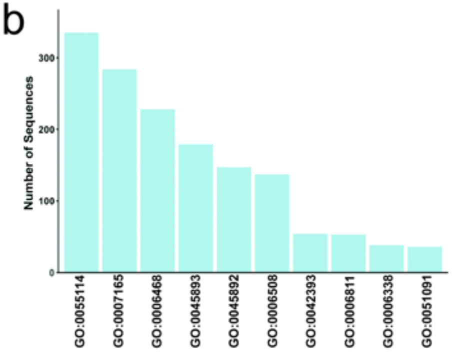

a
25
71
13
16
3
1826
20
0
1
47
0
1
0
2
0
single vs eight cell
eight vs E-blastomere
eight vs M-blastomere
E vs M-blastomere
b
n = 700
n = 267
n = 1208
n = 217
-log10(p-value)
15
10
5
0
-10
0
10
Eight Cell
logFC
Single Cell
c
SC2
SC1
SC3
SC5
Eight5
Eight2
Eight3
logCPM
-4.92
1.28
7.48

Figure 3

a

GO ID	Description	Number of Sequences
GO:0055114	oxidation-reduction process	335
GO:0007165	signal transduction	284
GO:0006468	protein phosphorylation	228
GO:0045893	positive regulation of transcription, DNA-templated	179
GO:0045892	negative regulation of transcription, DNA-templated	147
GO:0006508	proteolysis	137
GO:0042393	histone binding	54
GO:0006811	ion transport	53
GO:0006338	chromatin remodeling	38
GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity	36

b



c

GO ID	Description	Number of Sequences
GO:0005524	ATP binding	1088
GO:0008270	zinc ion binding	396
GO:0004674	protein serine/threonine kinase activity	201
GO:0003723	RNA binding	201
GO:0008134	transcription factor binding	105
GO:0008168	methyltransferase activity	33
GO:0004713	protein tyrosine kinase activity	33
GO:0004402	histone acetyltransferase activity	27
GO:0003964	RNA-directed DNA polymerase activity	18
GO:0042054	histone methyltransferase activity	3

d

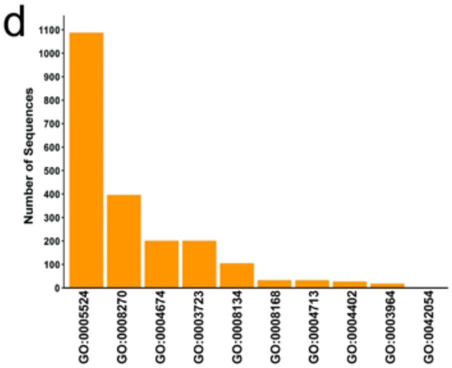


Figure 4