TRANSDIFFERENTATION IN *Turritopsis dohrnii* (IMMORTAL JELLYFISH): MODEL SYSTEM FOR REGENERATION, CELLULAR PLASTICITY AND AGING

A Thesis

by

YUI MATSUMOTO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Maria Pia Miglietta
Committee Members,	Jaime Alvarado-Bremer
	Anja Schulze
	Noushin Ghaffari
Intercollegiate Faculty Chair,	Anna Armitage

December 2017

Major Subject: Marine Biology

Copyright 2017 Yui Matsumoto

ABSTRACT

Turritopsis dohrnii (Cnidaria, Hydrozoa) undergoes life cycle reversal to avoid death caused by physical damage, adverse environmental conditions, or aging. This unique ability has granted the species the name, the "Immortal Jellyfish". *T. dohrnii* exhibits an additional developmental stage to the typical hydrozoan life cycle which provides a new paradigm to further understand regeneration, cellular plasticity and aging. Weakened jellyfish will undergo a whole-body transformation into a cluster of uncharacterized tissue (cyst stage) and then metamorphoses back into an earlier life cycle stage, the polyp. The underlying cellular processes that permit its reverse development is called transdifferentiation, a mechanism in which a fully mature and differentiated cell can switch into a new cell type. It was hypothesized that the unique characteristics of the cyst would be mirrored by differential gene expression patterns when compared to the jellyfish and polyp stages. Specifically, it was predicted that the gene categories exhibiting significant differential expression may play a large role in the reverse development and transdifferentiation in *T. dohrnii*.

The polyp, jellyfish and cyst stage of *T. dohrnii* were sequenced through RNA-sequencing, and the transcriptomes were assembled *de novo*, and then annotated to create the gene expression profile of each stage. Comparative functional gene enrichment analyses with the cyst as the central stage of comparison reported significant GO categories that were over-expressed, such as telomere maintenance and DNA repair, in the cyst as compared to other stages. The enrichment analyses also reported

significantly under-expressed categories, such as mitotic cell division and cellular differentiation, in the cyst as compared to the other stages. Additionally, candidate genes, such as the Yamanaka (Oct4, Sox2, Klf4, c-Myc) and Thompson Factors (Lin28, Nanog) that exhibit potential association with the transdifferentiation processes were found among the three stages for downstream differential gene expression analyses. Ultimately, our work produced a foundation to develop an alternative model system to further investigate and comprehend regeneration, cellular plasticity and aging in metazoans.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my M.S. graduate adviser and committee chair, Dr. Maria Pia Miglietta, for consistently providing invaluable guidance throughout the duration of my thesis project. Her patience, enthusiasm and knowledge have been fundamental to my deep understanding of genomics and the advancement of my project. Furthermore, Dr. Miglietta's investment in the success of my research and M.S. career has facilitated my further development as a scientist, and has inspired my future as a researcher in the academic system.

Additionally, I would like to thank the rest of my thesis committee, Dr. Jaime Alvarado-Bremer, Dr. Noushin Ghaffari and Dr. Anja Schulze, and essential collaborator, Dr. Stefano Piraino, for the encouragement and support they provided throughout the progression of my research. Without their provision, my project would have failed to incorporate different perspectives which drastically expanded the scope of my project.

Sample collection would not have been possible without the research facilities of the Smithsonian Tropical Research Institute and support from the members of the Systematics and Biology of Hydrozoa Summer Course (2015) who invested long hours helping me collect specimens in Bocas del Toro, Panama. This project would also not have been possible without the troubleshooting and bioinformatic support from the Texas A&M Genomics and Bioinformatics Service team. Funding for this project was

generously provided by Texas A&M University's Agrilife Research Center and Texas Sea Grant. In addition, graciously being awarded the Graduate Student Boost Funding scholarship significantly assisted the efficiency and intricacy of my research during my final year as a graduate student as it enabled the complete devotion to the progression of my project.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Professor Maria Pia Miglietta [adviser], Jaime Alvarado-Bremer, and Anja Schulze of the Department of Marine Biology, and Dr. Noushin Ghaffari of the Department of Electrical and Computer Engineering.

All work for this thesis was completed by the student, under the advisement of Dr. Maria Pia Miglietta of the Department of Marine Biology.

Funding Sources

Graduate study was supported by a Graduate Student 'Boost' Funding Scholarship (2016-2017) from Texas A&M University, Galveston.

This work was made possible in part by Texas Sea Grant College Program's Grants-In-Aid of Graduate Research Program. Additionally, this work was made possible in part by TAMUG's Marine Biology Mini Travel Grants, Galveston Graduate Student Association Travel Grant, and Erma Lee and Luke Mooney Graduate Student Travel Grant, TAMU's Agrilife Research Division and TAMU's PESCA Research Division. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Texas Sea Grant College Program and TAMUG.

NOMENCLATURE

BUSCOs	Benchmarking Universal Single-Copy Orthologs
ESc	Embryonic Stem cell
FDR	False Discovery Rate
GO	Gene Ontology
iPSc	Induced Pluripotent Stem cell
KEGG	Kyoto Encyclopedia of Genes and Genomes
NCBI	National Center for Biotechnology Information
PE	Paired Ends
RNA-seq	RNA-sequencing
TAMUG	Texas A&M University Galveston

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iv

CONTRIBUTORS AND FUNDING SOURCES vi

NOMENCLATURE vii

TABLE OF CONTENTS viii

LIST OF TABLES x

LIST OF FIGURES xii

1. INTRODUCTION 1

  1.1 Background 1
  1.2 Study system: Transdifferentiation in *Turritopsis dohrnii* 9
  1.3 Research approach: RNA-Sequencing and *de novo* assembly 14
  1.4 Research goals 17

2. METHODOLOGY 19

  2.1 Specimen collection/Species identification 19
  2.2 RNA extraction/Library construction/RNA-sequencing 21
  2.3 *De novo* transcriptome assembly 22
  2.4 Completeness of transcriptomes 23
  2.5 Functional annotation of transcriptomes 24
  2.6 Functional enrichment analyses 25
  2.7 Screening of candidate genes 26

3. RESULTS AND DISCUSSION 27

  3.1 Transcriptome sequencing and assembly 27
  3.2 Completeness of transcriptomes 35
  3.3 Functional annotation and categorization of transcriptomes 38
  3.4 Functional enrichment analyses 53
  3.5 Candidate genes of interest 68

Page

4. CONCLUSION 83

REFERENCES 85

APPENDIX A 104

APPENDIX B 105

LIST OF TABLES

Page

Table 1: Functions and significance of Yamanaka factors 5

Table 2: Number of raw, trimmed and normalized reads 28

Table 3: Trinity *de novo* assembly statistics 29

Table 4: Transcript length distribution of transcriptomes... 30

Table 5: Cyst RNA-seq alignment 33

Table 6: Read mapping evaluation of transcriptomes.. 36

Table 7: BUSCO evaluations of transcriptomes 37

Table 8: Summary of transcriptomes annotations 39

Table 9: Number of gene enrichment categories over- and under-expressed................. 54

Table 10: Top 10 over-expressed enrichment categories in biological processes 56

Table 11: GO enrichment categories involved in telomere maintenance 57

Table 12: GO enrichment categories potentially involved in the chitinous cyst exterior 59

Table 13: Top 10 under-expressed enrichment categories in biological processes 61

Table 14: GO enrichment categories associated with lifespan and aging 62

Table 15: GO enrichment categories associated with mitotic cell division.................... 64

Table 16: GO enrichment categories associated with DNA repair................................. 64

Table 17: GO enrichment categories associated with the response to stimuli................ 65

Table 18: GO enrichment categories associated with cell differentiation/organization .. 67

Table 19: GO enrichment categories associated with development 67

Page

Table 20: Candidate genes for the induction of pluripotency 69

Table 21: tBlastx screening of Yamanaka factors 71

Table 22: tBlastx screening of additional Thompson factors 72

Table 23: tBlastx screening of additional factors tested by Yamanaka group 73

Table 24: Candidate genes for longevity and aging 75

Table 25: tBlastx screening of telomerase subunits 77

Table 26: tBlastx screening of human shelterin subunits 78

Table 27: tBlastx screening of fission yeast shelterin subunits 79

Table 28: Accession name of sirtuin hits with highest bit scores 81

Table 30: tBlastx screening of other genes associated with aging and longevity 82

LIST OF FIGURES

Page

Figure 1: Life cycle of *T. dohrnii* (reprinted from Miglietta and Lessios, 2009) 10

Figure 2: Reverse development of *T. dohrnii* (reprinted from Schmich et al., 2007) 11

Figure 3: Workflow of research methods 19

Figure 4: Transcript length distribution of transcriptomes 30

Figure 5: Pre-filtered top-hit species distribution of the polyp transcriptome................. 40

Figure 6: Top-hit species distribution of transcriptomes 41

Figure 7: E-value distribution of transcriptomes 46

Figure 8: Hit species distribution of transcriptomes 48

1. INTRODUCTION

1.1 Background

Gene Expression

Gene expression occurs universally in all forms of life, including viruses, and allows the vital assembly of molecular machinery required for cellular function. Recent comparative studies of developmental processes of different species have continued to reaffirm the concept that many genetic programs for development are preserved throughout many taxa, and evolutionary selection exhibit a tendency to act upon and alter conserved gene networks that may induce new functional proteins and cellular processes (Wagner, 1994; Levine and Tjian, 2003; Butland et al., 2005; Wildwater et al., 2005; Costa and Shaw, 2006). The presence of homologous genetic networks among different taxa have expanded and reshaped the biomedical genetics sector by providing rationalization for the use of a variety of model systems, including mice (Jun-Li et al., 2001; Orsulic et al., 2003; Takahashi and Yamanaka, 2006), amphibians (Brockes, 1997; Brockes and Kumar, 2002; Nye et al., 2003), Zebrafish (Poss et al., 2002; Lien et al., 2006; Kikuchi et al., 2010), flatworms (Pellettieri and Sánchez-Alvarado, 2007; Gentile et al., 2011; Elliott and Sánchez-Alvarado, 2013), plants (Birnbaum and Sánchez-Alvarado, 2008; Cary et al., 2002) and even non-metazoans such as E. coli (Lee et al., 2003; Augusto-Pinto et al., 2003; Gogoi et al., 2006) and yeast (Botstein et al., 1997;

Longo and Kennedy, 2006; Fontana et al., 2010) to further our understanding and exploration of human health and disease.

Regeneration

Virtually all animals are capable of wound healing. However, after a substantial loss or damage of tissue, such as limbs and organs, only a few taxonomic groups appear capable of completely redeveloping complements of lost structures. Despite the limited number of taxa that have high regenerative capacities, a diverse array of approaches to overcome damage and injuries has evolved. Regenerative potential is often compromised as complexity in body organization increases. However, even though a larger number of early metazoan groups exhibit higher regenerative capacities, close evolutionary relationships with highly regenerating taxa or the complexity of body organization cannot be used to predict regenerative ability. *C. elegans* is a simple roundworm that cannot undergo cellular proliferation after development and exhibits no regenerative ability (Sánchez-Alvarado and Tsonis, 2006; Bely and Nyberg, 2010; Sánchez-Alvarado and Yamanaka 2014). On the other hand, some vertebrates such as urodele amphibians and zebrafish, are able to regenerate their tail, jaw, limbs and strikingly, parts of their eye (Tsonis and Rio-Tsonis, 2004; Brockes and Kumar, 2005; Jopling et al., 2011). A crucial question is whether genetic networks used for regeneration are highly conserved throughout metazoans, or are the outcomes of multiple independent evolutionary events. Numerous genes and cellular processes involved in differentiation during embryonic development have been shown to be key components of regeneration (Birnbaum and

Sánchez-Alvarado, 2008; Sánchez-Alvarado and Yamanaka, 2014). Despite the drastic differences in regenerative mechanisms among metazoan species, these components are often highly conserved across diverse taxa (Brockes and Kumar, 2008; Bely and Nyberg, 2010).

Comparative studies of regeneration in metazoan groups with high regenerative capacities and those that lack such capabilities may permit the discovery of cellular components that are indispensable to regeneration. Although traditional model systems have yielded much progress in exploring human health, there are countless understudied taxa that exhibit interesting and pertinent traits that can facilitate better understanding of mechanisms of regeneration. It is hard not to deliberate whether modern research sectors are fully utilizing all of biological resources available, and therefore missing a large part of the picture to be uncovered if there were to use both traditional and non-traditional model systems. Few cnidarians, such as *Hydra* (Class Hydrozoa) and *Nematostella* (Class Anthozoa), have been used as models to understand animal evolution. Cnidarians are sister taxa to bilaterians and thus represents a critical phylogenetic link to protostome and deuterostome divergence in metazoans (Putnam et al., 2007; Watanabe et al., 2009; DuBuc et al., 2014). Additionally, the phylum exhibits a remarkably high level of conservation in genomic content and organization in comparison to vertebrate taxa (Spring et al., 2000; Gauchat et al., 2000; Miller et al., 2000). Combined with their highly regenerative capabilities, Cnidarians are the optimal candidates to identify constituents and mechanisms that are relevant to regenerative processes of early metazoans.

Genetic Network of Regeneration

The ultimate goal of regenerative medicine is to restore function in cells and replace lost tissue. Irreversible damage to the body, particularly in species with long life spans, can be caused by environmental stresses, aging, trauma and disease, resulting in a decline or cessation of physiological and physical functions. The drive to eliminate many of the threatening challenges in human health has uncovered the ever-expanding field of cellular transplantation therapy in hopes of engineering cells with "stem cell-like" plasticity (Yu et al., 2007; Yamanaka, et al., 2009; Kim and De Vellis, 2009; Sánchez-Alvarado and Yamanaka, 2014). These cells are characterized to have capabilities of self-renewal and the ability to differentiate into any cell lineages that ultimately allow them to regenerate lost or damaged parts. Shinya Yamanaka, recipient to the Nobel Prize in Physiology/Medicine in 2012, demonstrated that the overexpression of four transcription factors, Oct4, Sox2, Klf4 and c-Myc (functions described in Table 1) can generate an induced pluripotent stem cell (iPSC) from fibroblasts taken from an adult mouse tail (Takahashi and Yamanaka, 2006).

Transcription Factor	Function
Oct4 (octamer-binding transcription factor 4)	Mammalian transcription factor (encoded by POU5F1 gene) that is associated with the maintenance and renewal of undifferentiated embryonic stem cells that remain active during the embryonic stages (Niwa et al., 2000; Takahashi and Yamanaka, 2006; Yamanaka et al., 2007; Lengner et al., 2008; Shi and Jin, 2010). Among the four transcription factors that generated an iPSc, Oct4 is indispensable and is a fundamental factor to inducing a pluripotent state. This transcript alone was capable of inducing a pluripotent state in neural cells (Kim et al., 2009). This transcription factor is highly expressed and is a repressor for cell specialization in undifferentiated states but becomes silenced when the cell undergoes differentiation (Pesce and Schöler, 2001; Zaehres et al., 2005). Oct4 has been reported to promote tumorigenesis by inhibiting cell death (Tai et al., 2005; Chiou et al., 2008; Saigusa et al., 2009).
Sox2 (sex determining region Y-box 2)	Similar to Oct4, Sox2 also is associated with the maintenance of pluripotency in undifferentiated embryonic stem cells and neural stem cells, and plays a fundamental role in the embryonic development of mammals (Avilion et al., 2003; Kiernan et al., 2005). The genes in the Sox family are known to be highly conserved amongst eukaryotic taxa, attracting much interest from evolutionary and clinical geneticists (Sasai, 2003; Uchikawa et al., 2003). Sox2 and Oct4 interact and bind DNA cooperatively (Chambers and Tomlinson, 2009). This transcript has been reported to be also be involved in formation of cancers and tumors (Liu et al., 2013; Boumahdi et al., 2014)
c-Myc	Proto-oncogene (mutation contributes to cancer) encodes a transcript that regulated gene expression associated with maintaining the cell cycle, regulating apoptosis and metabolism (Kerr et al., 1994; Dang, 1999). This regulator is alleged to universally control gene expression of 15% of all genes in humans (Gearhart et al., 2007) and is up-regulated in many cancerous cells (Little et al., 1983; Dubik et al., 1987; Chen et al., 2001)
Klf4 (kruppel-like factor 4)	Transcript that cellular proliferation, apoptosis, differentiation and reprogramming, thus playing a large role in modulating reparative responses to DNA damage and maintaining its stability (Segre et al., 1999; El-Karim et al., 2013). Klf4 is known to promote the formation of tumorigenesis but has also been shown to be an oncogenic suppressor (Rowland et al., 2005; Guan et al., 2010).

Table 1: Function and significance of Yamanaka factors. Oct4, Sox2, Klf4 and c-Myc are the core transcription factors that induced pluripotent stem cells in mammalian somatic cells.

In addition to the Yamanaka factors, the reprogramming of human cells into IPSC were demonstrated using the combination of transcription factors Oct4, Sox2, Lin28 and Nanog by the Thompson Lab of University of Wisconsin, Madison (Yu et al., 2007). Lin28 and Nanog can replace c-Myc and Klf4, while Oct4 and Sox2 are indispensable. Transcription factors became recognized inducers of pluripotency and regenerative

mechanisms. The Yamanaka group later demonstrated that the same four transcription factors that induced pluripotency in mice could reprogram fibroblast in iPSC in humans (Takahashi et al., 2007). These discoveries opened up the potential of creating disease- and patient-specific stem cells to optimize health screenings and therapy, as well as overcoming the constant battle with ethics in using human embryos and oocytes for stem cell research. Because most of what we know about regenerative capabilities in cells and induction of pluripotency comes from *in vitro* experimentation, there is still much to be discovered about the underlying mechanisms that prompt transcription factor induced nuclear reprogramming to a pluripotent state *in vivo.*

Aging

Biological aging is the physiological deterioration and decline of body functionality as a result of the reduced capabilities to regenerate and restore cellular machinery (Hayflick, 1985). In general, cellular regenerative potential decreases with aging (Taub and Longo, 2005; Sousounis et al., 2014). The decrease in regenerative potential results in the accumulation of cellular damage that may increase the development and progression of many late-onset diseases (Mutation-Accumulation Theory of Aging), especially in higher metazoans with long lifespans (Conboy and Rando 2005; Moskalev et al., 2012; Sousounis et al., 2014). Aging has been hypothesized to be the result of an evolutionary trade-off often associated with higher reproductive potential in earlier stages of sexual maturity (Schatten et al., 1999; Williams and Day, 2003; Blagosklonny, 2010). This compensation is known as an antagonistic pleiotropy

effect, in which beneficial genes used in early development become detrimental later in ontogeny. Two factors that have been shown to be correlated with aging are stem cell exhaustion and telomerase attrition (Silva and Conboy, 2008; Seifert and Voss, 2013). Self-renewing stem cells and molecular signaling pathways that regularly repair cellular machinery become fatigued and malfunction as we age (Terman and Brunk, 2004; López-Otín et al., 2013). The length of telomeres and telomerase activity are believed to directly influence the productivity of stem cells and has been recognized as a method to measure aging (Djojosubroto et al., 2003; Blasco, 2007). Telomeres, the ends of linear chromosomes, are long nucleotide repeats that are maintained by the reverse transcriptase enzyme telomerase. Germline cells continually produce this enzyme and therefore are considered "immortal", but its production is limited in somatic cells of most metazoans. As a result, telomeres shorten at every cell division. The cumulative and detrimental loss of genetic material has been linked to cell dysfunction and aging (Kaul et al., 2012; Campisi, 2013). Cancerous cells on the contrary exhibit large amounts of telomerase at all times (Shay, 1997; Shay and Gazdar, 1997; Bechter et al., 1998; Neumann et al., 2013; Hoxha et al., 2014).

Much effort has been allocated to research that may provide further knowledge of the range of senescence and cellular processes of aging in Metazoa. There is evidence that some of the eukaryotic genetic networks and biochemical processes involved with aging and longevity are conserved within Metazoans (Pitt and Kaeberlein, 2015; Faisal et al., 2014; Murthy and Ram, 2015).

Traditional invertebrate model systems, particularly round worms and fruit flies, have provided much insight on the consequences of senescence and genetic pathways associated with lifespan. There are, however, clear limitations in fully understanding aging when utilizing only traditional and familiar model systems. Roundworms and fruit flies are in phyla that have undergone broad divergence from the most recent ancestor with mammals (Kortschak et al., 2003; Austad, 2009; Bosch, 2012; Murthy and Ram, 2015; Sell, 2015). Although cnidarians have diverged from mammals much earlier than roundworms and fruit flies, it has been reported that many inherited features of the genome have been conserved from hydrozoans to mammals that are lost, replaced, or modified in *C. elegans* and *D. melanogaster* (Bignell et al., 1997; Lo et al., 2003; Kusserow et al., 2005; Sullivan and Finnerty, 2007; Martinez and Bridge, 2012). In addition, the phyla exhibit high regenerative potential in comparison to roundworms and fruit flies.

Cnidarians such as *Hydra* (Martinez, 1998; Vaupel et al., 2004) and some anthozoans (Prouty et al., 2011; Reiter et al., 2012) do not follow common trajectories of aging and are considered to display negligible senescence. The common symptoms of aging, such as the increase in mortality rate and the decline in reproductive rates, are not observed in these organisms (Finch, 1994). Various coral species (Class Anthozoa) have been utilized for aging research as they have the ability to sexually and asexually reproduce continuously throughout their lifetime with no evidence of senescence (Vaupel et al., 2004; Roark et al., 2009; Sherwood and Edinger, 2009; Prouty et al., 2011). Some species of *Hydra* (Class Hydrozoa) are well-studied solitary cnidarians that have

demonstrated negligible senescence for multiple years when kept in laboratory settings (Martinez, 1998; Watanabe et al., 2009; Martinez and Bridge, 2012; Reiter et al., 2012). Similar to corals, *Hydra* can undergo asexual reproduction to bud clonal polyps from adults in addition to sexual reproduction and have unlimited capacities to self-renew tissue. The entire body is composed of three different cell lineages that all possess their own stem cell population. Progenitor interstitial stem cells (I-cells) give rise to all cells in the body as *Hydra* do not pertain a separate soma and germline throughout ontogeny (Bosch and David, 1984; Galliot, 2012; Bosch, 2012). A new promising system to study both regeneration and aging is the hydrozoan *Turritopsis dohrnii.*

1.2 Study System: Transdifferentiation in *Turritopsis dohrnii*

Hydrozoa are a class of the phylum Cnidaria that exhibit complex life cycles. They remain comparatively unexplored due to a limited number of experts that can collect Hydrozoa in the field or identify to the species level. The typical hydrozoan life cycle starts with a planula larval stage that settles on a substrate and develops into the juvenile polyp stage (Figure 1A). The polyps asexually form new polyps, and expand into a sessile colony. During optimal environmental conditions and seasons, polyps bud the free-swimming medusae (jellyfish). Adult medusae release gametes into the water column and planulae larvae are produced through external fertilization. The planula settles on the appropriate substrate and metamorphoses into the polyp, thus closing the life cycle. The benthic polyp and planktonic medusa dramatically differ in morphology as

each stage exhibits stage specific structures. Within the class Hydrozoa, the species *Turritopsis dohrnii* (Filifera, Family Oceaniidae) additionally exhibits a unique developmental trait that can provide a new paradigm for studies in a vast array of developmental biology sectors.

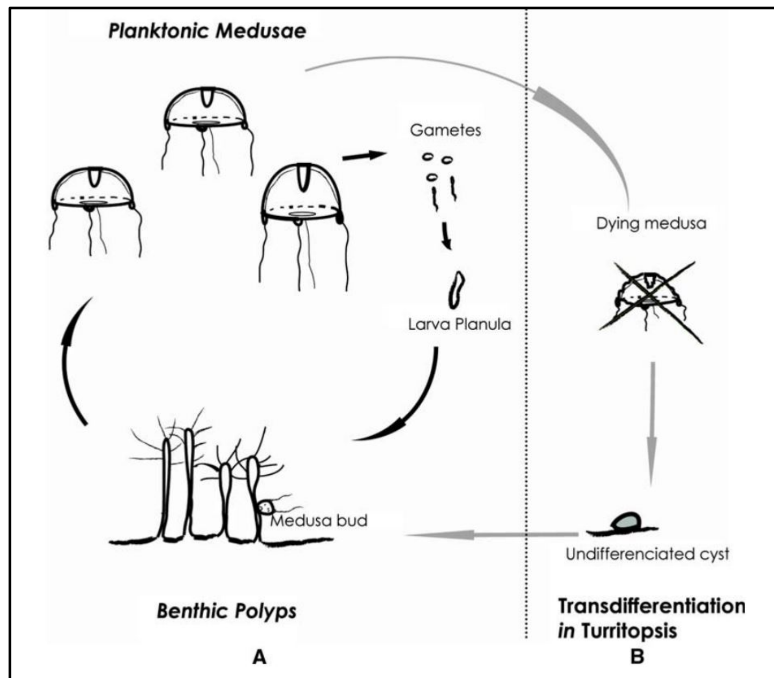

Figure 1: Life cycle of *T. dohrnii* (reprinted from Miglietta and Lessios, 2009). A) Typical hydrozoan lifecycle. B) Dying medusa transforms into a cyst, in which cellular transdifferentiation occurs.

Turritopsis dohrnii can avoid death caused by physical damage, sub-optimal environmental conditions, and senescence through reverse development. When exposed to stressful conditions, medusae of *T. dohrnii* are able to revert back into a polyp (Figure 1B). During the metamorphosis, the mesoglea (gelatinous layer between ecto-and endoderm) of the jellyfish is reduced and tentacles retract (Figure 2A-C), and the medusa

becomes an uncharacterized cyst-like structure with chitinous exterior that settles and attaches to substrate (Figure 2D-E). This cyst stage has not been adequately characterized in terms of tissue and cell component and thus, is referred to as uncharacterized. The cyst metamorphoses back into a polyp within three days post-induction (Figure 2F-J), and then undergoes the typical hydrozoan lifecycle (Schmich et al., 2007). Due its small size (Figure 2), the species is very inconspicuous and not much regarding their ecological nature is known (e.g. food source, predators, number of gametes spawn).

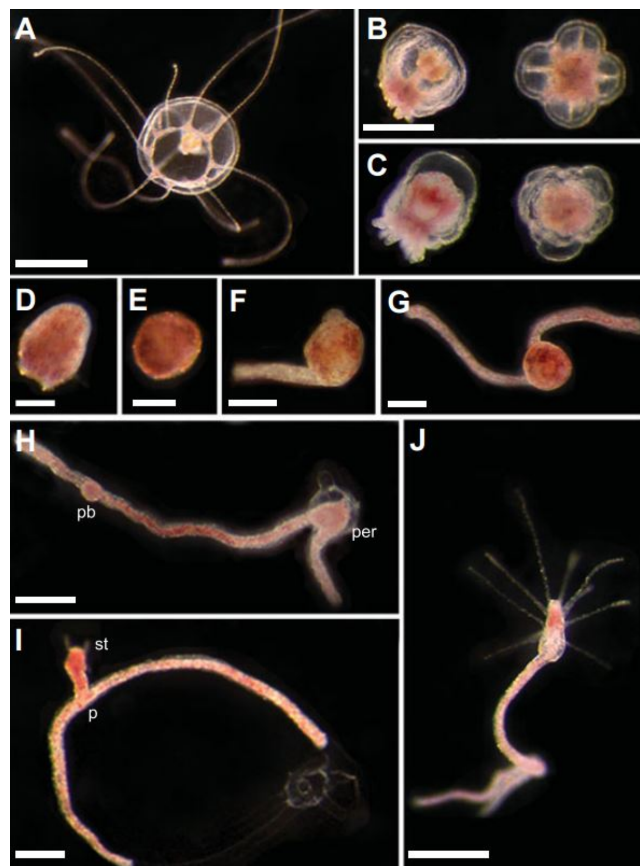

Figure 2: Reverse development of *T. dohrnii* (reprinted from Schmich et al., 2007). Scale bars: A) 500µm; B) 300µm; C) 300µm; D) 150µm; E) 150µm; F) 400µm; G: 400µm; H) 300µm; I) 300µm; J) 400µm.

Medusae and polyps possess different cell types. For example, specialized sensory structures called ocelli and statocysts are only present in the free swimming medusae, while only sessile polyps can produce a chitinous sheath (perisarc) that protect the delicate body from peripheral damage (Piraino et al., 1996). Hydrozoans contain migratory, self-renewing stem cell reserves between epithelial tissues called interstitial cells (I-cells), most heavily explored in Hydra (Bode, 1996; Martinez, 1998; Martinez and Bridge, 2012). Although I-cells play a role in tissue and organ regeneration in both polyps and medusae of Hydrozoans, *T. dohrnii* life cycle reversal can occur even from fragments of medusae that do not contain I-cells (Piraino et al., 1996). Selective tissue excisions of *T. dohrnii* have shown that the manubria of the medusa cannot revert into perisarc-secreting tissue despite the abundant population of I-cells. On the other hand, the exumbrellar epidermis that does not contain any populations of I-cells was able to transform back into the polyp. This implies that during the reverse development of *T. dohrnii*, cells undergo transdifferentiation (Piraino et al., 1996). Cell transdifferentiation, also known as cellular reprogramming, allows mature, differentiated cells to reprogram themselves into a new (needed) cell type (Okada, 1991). It occurs in two stages: first the cells de-differentiate, and then a new developmental program is activated, allowing the cells to develop into a different cell type. It is also believed that the regulation of transcription factors represents a transition bridge between initial and target cells (MacQuarrie et al., 2011; Iwafuchi-Doi and Zaret, 2014). The cellular transdifferentiation occurs at the cyst stage in *T. dohrnii*, when the dying medusa is about to metamorphose back into the polyp (Figure 1B). Reverse development and cell transdifferentiation in *T.*

dohrnii can be induced through starvation, mechanical damage, heat shock or chemical exposure to Cesium Chloride (CsCl) (Piraino et al., 1996; Piraino et al., 2004; Schmich et al., 2007).

T. dohrnii contradicts the paradigm of aging and ontogenesis of most metazoans. For example, the onset of sexual reproduction has long been accepted as the point of no return in the ontogenetic sequence of an organism, and is normally followed by senescence and death (Stearns, 1976; Piraino et al., 1996; Mahmoudi and Brunet, 2012). However, the reverse metamorphosis can be induced by senescence in *T. dohrnii*, as reproductively senescent jellyfish can still revert back into a juvenile polyp stage (Piraino et al., 1996; Schmich et al., 2007). In fact, reverse development can occur at any time during the medusa stage, from newly released newborns to senescent. *T. dohrnii*'s unique life cycle makes it an excellent system to study cell transdifferentiation, life cycle reversal and aging in basal metazoans.

Understanding the genetic basis of transdifferentiation is especially important to study the flexibility of cellular phenotypes and regeneration. Much of the effort in regenerative research is often conducted *in vitro*. This approach, although extremely useful to identify key underlying genetic components, neglects to address the complexity of *in vivo* transdifferentiation. Removing a cell from its natural environment has commonly induced changes in phenotype, cellular stability and behavior (Holtfreter, 1947; Goldberg et al., 2007; MacArthur and Lemischka, 2013). *T. dohrnii*, thus represent a relatively simple model system to assess natural transdifferentiation *in vivo* and address

the mechanisms by which cells spontaneously (*in vivo*) leave a differentiated state to become a new lineage.

Finally, *T. dohrnii*'s sister species (*T. nutricula*) does not have the capability to reverse its life cycle (Miglietta et al., 2007). This allows for interesting comparative analyses that can shed light on how cell transdifferentiation and life cycle reversal can occur in one species (*T. dohrnii*), but not the other (*T. nutricula*).

1.3. Research Approach: RNA-Sequencing and *de novo* assembly

Transcriptomics, the study of gene expression profiles, can be used to investigate how gene activity is influenced by development, different environmental conditions, or the progression of a disease. Gene expression landscapes can be explored using a broad-ranged, high-throughput technique called RNA-sequencing (RNA-seq), that utilizes complementary DNA (cDNA) synthesized from messenger RNA (mRNA) by an RNA-dependent, retroviral DNA polymerase (Metzker, 2010). RNA-seq allows the quantification of RNA transcripts, including non-coding RNA, to create a snapshot of the genes expressed at the specific stage. Transcriptomes (gene expression profile) offer distinct advantages over genome (entire genetic profile) sequencing because it focuses on gene expression and activity rather than gene structure.

RNA-seq has been utilized on a broad range of living organisms, from prokaryotes to humans, to reveal the genetic constituents involved in a cell, organ or entire organism (Core et al., 2008; Nagalakshmi et al., 2008; Oliver et al., 2009; Ozsolak

and Milos, 2011; Iacobucci et al., 2012; Shpirer et al., 2014). Prior to the development of RNAseq, gene activity was studied primarily using DNA hybridization microarray technologies. In comparison, microarrays are unable to completely catalog all levels of gene expression as it requires a prior genomic reference alignment, have significantly lower sequence throughput, low range of detection, and is additionally cost and labor intensive.

RNAseq offers the ability to explore non-traditional systems lacking a corresponding reference genome by enabling the construction of *de novo* transcriptome alignments through the reassembly of fragmented sequences (Metzker, 2010). In addition, *de novo* approach assemblies can uncover novel genes or sequence variation that is masked or left out during reference-guided alignment. Thus, researchers are even conducting *de novo* assemblies of the human genome to fully capture and further complete the genomic profile of our species (Chaisson et al., 2015; Li et al., 2010; Mostovoy et al., 2016).

A *de novo* assembler serves to arrange short reads into contiguous sequences (contigs) that will represent part or an entire mRNA transcript (Zerbino, 2010; Luo et al., 2012; Haas et al., 2013). A popular algorithm approach to *de novo* assembly, used in assemblers such as Trinity, is to partition the sequencing data into multiple individual de Bruijn graphs, directed graphs that represent the overlapping region between sequencing reads, which are then utilize these graphs to identify all transcript variants, including alternatively spliced isoforms to capture the complexity of the transcriptome (Compeau et al., 2011; Haas et al., 2013).

RNA-seq projects with goals to assemble a transcriptome using the *de novo* approach follows four fundamental steps (Baker, 2012). First, the mRNA is extracted, fragmented and then sequenced. Second, a *de novo* assembler will find overlaps between sequencing reads. Third, the overlaps in the reads will be assembled into contigs, or contiguous sequences that best represents the mRNA transcript. Fourth, these contigs can be assembled into scaffolds, which are the contigs accumulated in the order represented in its genome with gaps that represent non-coding regions.

1.4. Research Goals

The long-term goal is to establish *Turritopsis dohrnii* as an alternate model system to further understand regeneration, cellular plasticity and aging. The objective is to characterize the transcriptome profiles that underlie *T. dohrnii's* life cycle with emphasis on its cell transdifferentiation and reverse development. Specifically, this research aims to: A) characterize the transcriptome landscapes through *T. dohrnii*'s life cycle with focus on the polyp, the newborn jellyfish and the cyst, and B) to identify the genes involved in its transdifferentiation. Because transdifferentiation occurs at the cyst stage, it is hypothesized that the transcriptome profile of the cyst would showcase over-expressed gene categories that are involved in longevity and the undifferentiated cell state, and under-expressed gene categories would reflect the seemingly dormant state of the cyst, such as cellular characterization and development.

RNA-sequencing reads from **the fertile polyps, newborn jellyfish, and cyst** were generated and then assembled into separate transcriptomes using a *de novo* approach, followed by bioinformatic analyses. The stage-specific transcriptomes were assigned functional annotation and comparative analyses to explore the different gene networks and pathways that may potentially be involved *T. dohrnii*'s ability to undergo cellular transdifferentiation. The cyst is the central stage of our comparison as it directly involved in the process of transdifferentiation. The generated transcriptomes were additionally screened for the presence candidate genes that exhibit potential association

with transdifferentiation in *T. dohrnii*. The candidate genes of interest were categorized into its association with the induction of pluripotency or longevity and aging.

2. METHODOLOGY

An overall workflow of the research methods described in this section can be found below in Figure 3.

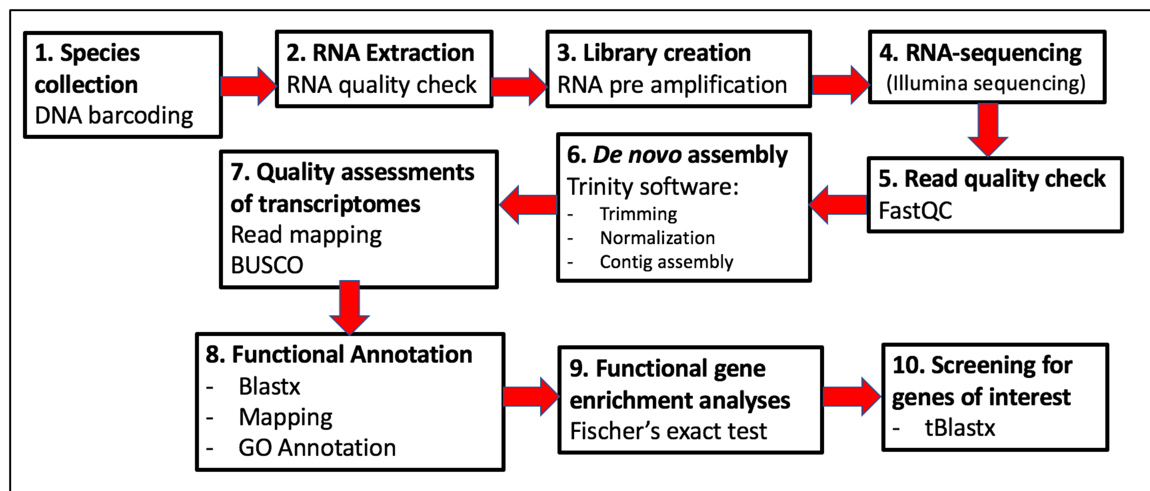

Figure 3: Workflow of research methods.

2.1 Specimen collection/Species identification

Colonies bearing medusa buds of *Turritopsis dohrnii* were collected in Otranto, Italy in July 2013 by M.P. Miglietta at 5mt depth, and in a rocky and exposed environment. The individual colonies were kept isolated in glass bowls in room temperature near a light to collect newly released jellyfish. RNAlater RNA stabilization solution (ThermoFischer Scientific) was used as the specimen preserving reagent to

stabilize and prevent the degradation of RNA. The fertile colony and newborn jellyfish were preserved individually in RNAlater and stored in -80° until further processing.

Fertile colonies of *T. dohrnii* were collected in Bocas del Toro, Panama in 2015 at 1-2mt depth range. Individual colonies were once again kept isolated in glassware in room temperature near light to encourage the release of jellyfish. The newborn jellyfish were immediately transferred following release to a separate glass bowl with new filtered sea water to best eliminate possible biological contaminants. The jellyfish were starved in order to avoid genetic contamination from undigested non-target organisms, and induced to transform into cysts (24-48 hours). The cysts were preserved in RNAlater when the jellyfish retracted all of its tentacles to form a ball-like structure and has settled on substrate (Piraino et al., 1996). In addition, the colony in which the cyst originated from was preserved to confirm the species identification. All specimen samples were stored in -80° until further processing.

Prior to RNA extraction, total DNA was extracted from a piece of polyp tissue from both collections using a protocol provided by M.P. Miglietta (Miglietta Lab, TAMUG). The purity and concentration of DNA was confirmed using the Thermo Scientific NanoDrop2000 Spectrophotometer. A fragment of the mitochondrial 16S gene was amplified using forward SHA and reverse SHB primers [Forward (SHA): 5'-TCGACTGTTTACCAAAAACATAGC-3', Reverse (SHB): 5'-ACGGAATGAACTCAAATCATGTAAG-3'] for DNA barcoding to identify the species. PCR products were digested through agarose gel electrophoresis and visualized

for the presence of a DNA band, treated with ExoSAP, and sequenced to confirm the species identification.

2.2 RNA extraction/Library construction/RNA-sequencing

Total RNA was extracted from the fertile polyps and a pooled sample of jellyfish (approximately 25-50 individuals) using Qiagen's RNAeasy Mini Kit (Cat. #: 74104) following manufacturer's instructions. Epicentre's MasterPureTM RNA Purification Kit (Cat.#: MCR85102), specifically made for samples with low number of cells or little tissue, was used to extract total RNA from individual cysts. Multiple replicates of cysts were extracted. The protocol was carried out with the following minor modifications recommended by the manufacturer: 1) Pre-rinsing and removing RNAlater from tissue using the T&C Buffer, 2) 1/5th of reagents were used due to the small amounts of tissue, 3) increased number of ethanol washes for the additional purification of RNA. The concentration of all RNA extracts was determined using the Thermo Scientific NanoDrop2000 Spectrophotometer, and their integrity was validated via Agilent Technologies Bioanalyzer prior to initiating library construction for sequencing. Three replicates of the cyst stage with the highest quantity of RNA and quality were selected for further processing.

The sequencing library for the polyp stage was constructed by the University of Nortre Dame's Genomic and Bioinformatic Core facility using the TruSeq RNA Sample Prep Kit v2 from Illumina (Cat. #: RS-122-2001) following the manufacturer's protocols.

For the jellyfish and cyst RNA extractions, the amount of output RNA was less than 100ng and thus, a pre-prep amplification step was performed prior to library preparation at TAMU's Agrilife Fenomics and Bioinformatics sequencing facility. A Poly A-based SMARTer Ultra Low Input RNA kit for Sequencing v4 kit from Clontech Laboratories (Cat. #: 634888) was used for the amplification according to manufacturer's protocol. 100uL of the resulting jellyfish and cyst cDNA generated in the amplification was constructed as separate libraries using the TruSeq Nano DNA Library Preparation Kit from Illumina (Cat. #: FC-121-4001) following manufacturer's instructions.

The cDNA from the fertile polyp library was sequenced using the Illumina MiSeq platform (Illumina Sequencing Technologies) generating 83 bp paired-end (PE) reads. The cDNA from the jellyfish library was sequenced using the Illumina HiSeq2500v4 platform generating 125 bp PE reads. The cDNA from the three cyst libraries were individually sequenced using the Illumina HiSeq4000 platform generating 150bp PE reads for each library.

2.3 *De novo* transcriptome assembly

Polyp and jellyfish stage

The quality of the PE reads generated from fertile polyps and jellyfish library was examined using the FastQC tool (Andrews, 2010). Reads from polyps and jellyfish were trimmed prior to assembly using the default setting in the trimmomatic tool (MacManes, 2014) within Trinity version 2.2.0. Trimmed reads were normalized using *in silico*

normalization within Trinity's toolkit using default parameters. The normalized RNA-seq reads of the fertile polyp and jellyfish stage were individually assembled *de novo* via Trinity using a k-mer size of 25. Assembly statistics were generated using the trinitystats.pl script within Trinity's toolkit.

Cyst Stage

The quality of the PE reads of the three cyst libraries was examined using the FastQC tool. In addition, each dataset was aligned to the fertile polyp and jellyfish assembly to assess the quality of the reads. Two of the datasets that had high alignment percentages and low percentages of broken reads were chosen and pooled together. The same trimmomatic and *de novo* assembly via Trinity workflow (see subsection above for parameters) followed, and assembly statistics were generated.

2.4 Completeness of transcriptomes

Input reads resulting from Trinity's *in silico normalization* were mapped back to assembled contigs via TopHat v2.1 and CLC Genomic Workbench v8 alignment software using default parameters to evaluate the quality of the assemblies. The BUSCO v2.0 software (Simão et al., 2015) was used to assess the completeness of gene content in each assembled transcriptomes. The contents of the Eukaryota and Metazoa database (files: metazoan_odb9 and eukaryota_odb9 downloaded from the software website (http://busco.ezlab.org)) was used for the assessment with default settings (expectation value of e^{-3}).

2.5 Functional annotation of transcriptomes

The annotation pipeline from Blast2GO PRO program (Conesa et al., 2005; Götz et al., 2008) was used to assign functional annotation to the three stage-specific transcriptomes. Blastx through CloudBlast (Matsunaga et al., 2008) was performed on the assemblies against NCBI's Non-Redundant (NR) database using the following parameters: expectation value of e^{-3}, word size of 6 and HSP length cutoff of 33, and the top 20 hits were saved. Sequences that exceeded the upper limit for NR database searches of 18,000 bps (3 sequences from Jellyfish; 12 from Cyst) were blasted on CloudBlast against the Metazoa database (as recommended by Blast2GO support team). Gene Ontology (GO) functional annotation was conducted with the following parameters: annotation cutoff of 55, GO weight of 5, expectation value of e^{-6} and a HSP-Hit coverage cutoff of 0. InterProScan (Zdobnov and Apweiler, 2001) was also used to perform protein domain-based identifications in InterPro databases to build upon and confirm existing GO annotations. The following databases were searched in InterProScan: BlastProDom, FPrintScan, HMMPIR, HMMPfam, HMMSmart, HMMTigr, ProfileScan, HAMAP, SuperFamily, HMMPanther and Gene3Ds. Sequences with open reading frames that do not translate to the minimum of 11 amino acids were unable to be utilized in InterProScan (15 sequences from fertile polyps; 9 from Jellyfish; 6 from cyst). ANNEX Augmentation was performed after merging the Interpro annotations, and 1st level annotations were manually removed as recommended in a Blast2GO methodology report (Conesa and Götz, 2008). Lastly, the Enzyme Comission (EC) codes mapping was

conducted in order to generate Kyoto Encyclopedia of Genes and Genomes (KEGG) maps downstream.

Due to predicted ecological contamination in the polyp transcriptome of taxa closely related to the Protista species *Reticulomyxa filosa*, the sequences that had at least one top hit to *R. filosa* were re-blasted against the metazoan database and the Blast2GO annotation pipeline was re-performed.

2.6 Functional enrichment analyses

Functional enrichment analyses were conducted using the FatiGO (Al-Shahrour et al., 2004) package within the data mining tools of the Blast2GO PRO program. Prior to the analyses, all three transcriptomic datasets of the stages were combined into one dataset and a sequence ID list that corresponds to each stage's dataset was generated as the working dataset. The Fischer's Exact Test (two-tailed) was performed using the False Discovery Rate (FDR) p-value of 0.01 to compare the datasets of the following: 1) Cyst vs. polyps; 2) Cyst vs. jellyfish; 3) Cyst vs. polyps and jellyfish combined. The results were generated as graphs showing the over- and under-expression of functional subgroups broadly categorized into the following: biological function, molecular function and cellular components.

2.7 Screening of candidate genes

Genes of interest were searched NCBI's GenBank database (https://www.ncbi.nlm.nih.gov/) under the 'Gene' category, and the orthologs of specific taxa were chosen. All mRNA variant sequences for the genes of interest were downloaded as fasta files from NCBI's Reference Sequences (RefSeq) database. The sequences were uploaded onto CLC Genomics Workbench Version 8, and tBlastx was performed on the downloaded fasta files against all three of the *T. dohrnii* transcriptomes. An e-value cutoff of e^{-8} was used for tBlastx to set a stringent cutoff and to ensure truly significant results. The outputs of the blast search and the presence or absence of genes were recorded all three stages.

3. RESULTS AND DISCUSSION

3.1 Transcriptome sequencing and assembly

Innovative deep sequencing techniques have allowed the assembly of transcriptomes in non-model organisms that lack a reference genome. A *de novo* approach using the Trinity software was performed to generate the three stage-specific *T. dohrnii* transcriptomes. The Trinity program utilizes the Inchworm, Chrysalis and Butterfly software modules to construct *de novo* assemblies. Briefly, Inchworm assembles the reads into unique contiguous sequences (contigs) and reports distinctive regions of alternatively spliced isoforms. Chrysalis then clusters the generated contigs to build de Bruijn graphs that represent transcriptional complexity within or among several genes. Butterfly then processes the resulting de Bruijn graphs and reports full-length isoforms and separates paralogous transcripts.

Prior to the *de novo* assembly, quality trimming was conducted using the Trimmomatic tool within Trinity version 2.2.0. Default parameters based off the optimized parameters of a Trimmomatic methodology article by MacManes (2014) was utilized. The trimmed reads were then normalized through Trinity's *in silico normalization* to condense reads based on sequencing depth (maximum coverage of 30x) and improve the run time of the assembly. *In silico normalization* reduced the redundancy of highly expressed reads while conserving low coverage reads (Haas et al.,

2013). Thus, the normalization process will not reduce the overall coverage of *de novo* transcriptome assembly, as long as the input reads undergone sequencing inclusive of all transcripts and genes. The resulting normalized reads were then fed to the *de novo* assembly pipeline to build the transcriptomes.

The trinitystats.pl script from the Trinity toolkit utilities was used to generate assembly statistics based on overall transcript length for all three assemblies. These statistics can be used to assess the quality of the assembly, as longer reads are more desirable as they have higher confidence than the shorter reads. Shorter reads may be composed of repetitive sequences that can be aligned to other sequences within its genome creating ambiguities during downstream analyses.

Polyp Transcriptome

The Illumina MiSeq generated 47,008,570 reads (23,504,285 paired) from polyp library. 46,947,184 reads (23,473,592 paired) remained after quality trimming and 17,552,220 reads (8,776,110 paired) were left after normalization (Table 2).

	Polyp	**Jellyfish**	**Cyst**
Raw Reads	47,008,570 (23,504,285 paired)	42,190,870 (21,095,435 paired)	90,019,818 (45,009,909 paired)
Trimmed Reads	46,947,184 (23,473,592 paired)	42,002,298 (21,001,149 paired)	89,811,296 (44,905,648 paired)
Normalized Reads	17,552,220 (8,776,110 paired)	6,220,660 (3,110,330 paired)	11,289,162 (5,644,581 paired)

Table 2: Number of raw, trimmed and normalized reads.

The normalized reads were assembled into a transcriptome with a total length of approximately 68.32 Mb and produced 92,659 transcript isoforms and 76,456 unique transcripts with a GC content of 37.33%. 10% (N10) of the 92,659 contigs are larger than 3,576 bps, 30% (N30) are larger than 2,116 bps and half (N50) are larger than 1,332 bps. The assembly ranged from 201bps to 11,079bps with the median length of 376bps and mean length of 737.35bps (Table 3). The summary of transcript length distribution of the transcriptome can be found in Table 4 and visualized in Figure 4.

	Polyps	**Jellyfish**	**Cyst**
# of unique transcripts	76,456	58,312	62,666
# of transcripts	92,659	74,641	86,373
% GC	37.33	37.17	38.25
Contig N10	3,576	3,766	3,935
Contig N30	2,116	2,196	2,420
Contig N50	1,332	1,434	1,634
Median contig length	376	493	645
Mean contig length	737.35	864.82	1,010.57
Minimum length	201	201	201
Maximum length	11,079	20,570	20,736
Total assembled bases	68,322,250	64,550,812	87,289,185

Table 3: Trinity *de novo* assembly statistics.

	Polyps		Jellyfish		Cyst	
Length (bps)	Transcripts	%	Transcripts	%	Transcripts	%
200-500	56,902	61.41%	37,717	50.53%	35,498	41.10%
501-1000	15,558	16.79%	15,958	21.38%	20,986	24.30%
1001-1500	7,726	8.34%	8,633	11.57%	11,420	13.22%
1501-2000	4,999	5.40%	5,115	6.85%	7,142	8.27%
1001-2500	3,031	3.27%	2,860	3.83%	4,377	5.07%
2501-3000	1,765	1.90%	1,704	2.28%	2,627	3.04%
3001-3500	1,077	1.16%	1,061	1.42%	1,693	1.96%
3501-4000	627	0.68%	579	0.78%	1,021	1.18%
4001-4500	386	0.42%	331	0.44%	704	0.82%
4501-5000	236	0.25%	238	0.32%	340	0.39%
>5000	352	0.38%	443	0.59%	565	0.65%

Table 4: Transcript length distribution of transcriptomes.

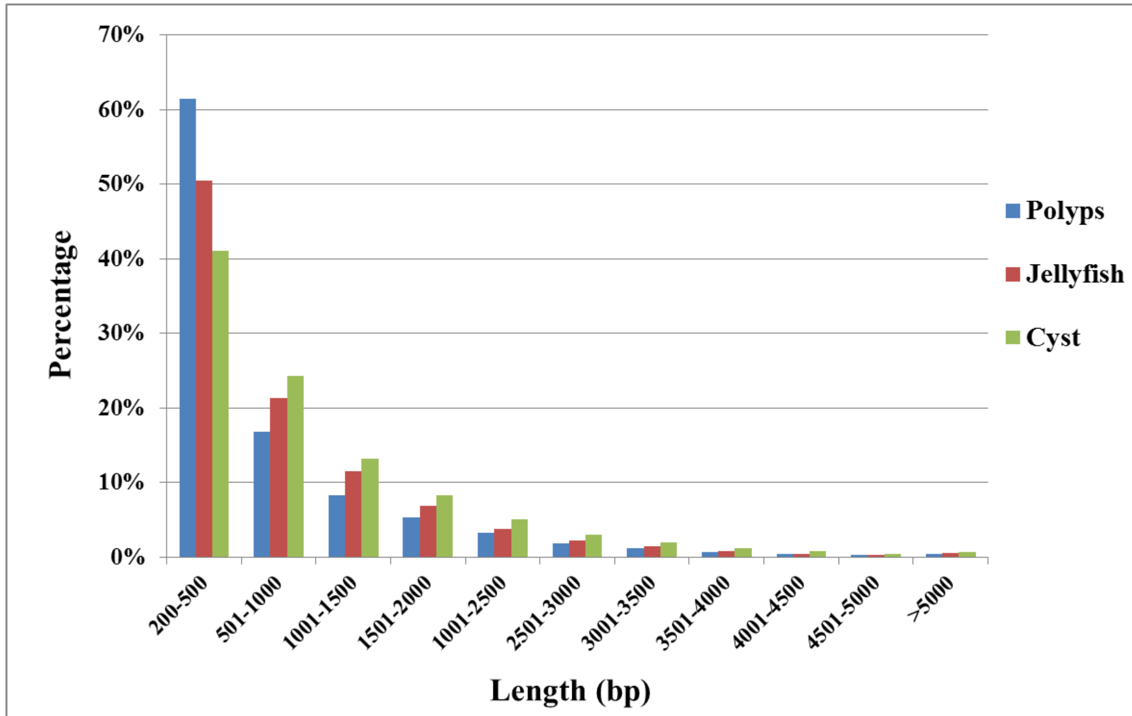

Figure 4: Transcript length distribution of transcriptomes.

Jellyfish Transcriptome

The Illumina HiSeq2500v4 generated 42,190,870 reads (21,095,435 paired) from the jellyfish library. 42,002,298 reads (21,001,149 paired) remained after quality trimming and 6,220,660 reads (3,110,330 paired) were left after normalization (Table 2). The normalized reads were assembled into a transcriptome with a total length of 64.55 Mb and produced 74,639 transcript isoforms and 58,312 unique transcripts with a GC content of 37.17%. 10% of the 74,639 contigs are larger than 3,766bps, 30% are larger than 2,196bps and half are larger than 1434 bps. The assembly ranged from 201bps to 20,570bps with the median length of 493 bps and mean length of 864.82 bps (Table 3). The summary and figure of transcript length distribution of the transcriptome can be found in Table 4 and Figure 4.

Cyst Stage

The Illumina HiSeq4000 platform generated the following number of reads for three cyst replicates: Cyst1) 51,524,006 reads (25,762,003 paired); Cyst2) 54,006,982 reads (27,003,491 paired); Cyst3) 38,495,812 reads (19,247,906 paired). The FastQC report was generated and an abnormally high and unparalled GC% content for the Cyst2 PE reads (48% and 50%) in comparison to Cyst1 (both 39%) and Cyst3 (both 41%), and even more so when compared to the polyps (37.22%) and jellyfish (37.17%) assemblies (Table 5). Additionally, the quality of the reads in the Cyst2 dataset had significantly lower quality than the other two cyst replicates.

Due to the cyst stage being the central stage of interest and the importance of high quality input reads for systems without a reference genome, RNA-seq alignments were

performed using the fertile polyp and jellyfish stage as the reference to further assess the quality of the cyst datasets prior to *de novo* assembly. Using the CLC Genomics Workbench Version 9 software, the reads from Cyst1 and Cyst3 were trimmed using the limit score of 0.1, equivalent to a Phred score cutoff of 10. The reads from Cyst2 were trimmed using the limit score of 0.316, equivalent to a Phred score cutoff of 15, due to its lower overall quality of the dataset. The trimmed reads for all three datasets were then aligned to the polyp and jellyfish transcriptomes separately as well as combined length fraction value of 0.5 and similarity fraction of 0.8, and the paired-distances were estimated. Cyst1 had an estimated paired distance range of 128-594 bp and had high alignment percentages in the range of approximately 93%-97% against the polyp and jellyfish transcriptomes (Table 5). The broken reads percentage for Cyst1 was relatively low in the range of approximately 11%-17%. Similarly, Cyst3 had an estimated paired distance range of 128-596 bp, had a high alignment percentage in the range of approximately 89%-95%, and a slightly higher percentage range of broken reads, approximately 15%-20%. Cyst2 in comparison had an estimated paired distance range of 0-3,830 bp with few pairs mapped, had an unusually large range in alignment percentages of approximately 10%-92%, and broken read pairs of approximately 8%-61%. The very low range in the estimated paired distance for the Cyst2 replicate shows that a large portion of the dataset consist of pairs that have short reads or do not align correctly, also portrayed in the differences of %GC content between the PE datasets. The very low alignment rate (9.68%) against the polyp assembly as compared to the other two replicates, but a high alignment rate against the jellyfish reference most likely signifies

that only a small variety of mRNA transcripts were incorporated into the library construction and has mapped multiple times to alignments exclusively found in the jellyfish. In addition to the short reads and low coverage, the high percentages of broken reads are particularly not suitable for quality *de novo* approach assemblies. Thus, Cyst2 was excluded from our dataset.

		Polyps	**Jellyfish**	**Combined**
Cyst1	% Reads mapped	92.98%	94.90%	96.95%
	% Broken pairs	10.61%	14.36%	16.64%
	Est. paired distance	128-594 bp	130-594 bp	128-594 bp
Cyst 2	% Reads mapped	9.68%	91.91%	91.90%
	% Broken pairs	8.73%	60.54%	60.58%
	Est. paired distance	0-3830 bp *	0-240 bp	0-240 bp
Cyst3	% Reads mapped	90.97%	88.71%	95.02%
	% Broken pairs	14.65%	15.02%	19.54%
	Est. paired distance	172-596 bp	128-596 bp	172-596 bp
	*= Few pairs mapped			

Table 5: Cyst RNA-seq alignment. The percentage of reads mapped, percentage of broken reads, and estimated paired distance of three cyst replicates aligned to the polyp and jellyfish transcriptomes.

The raw RNA-seq reads of Cyst1 and Cyst 3 were pooled into one dataset and the same protocol for base trimming, normalization and *de novo* assembly of the polyp and jellyfish assembly was followed. In total, 90,019,818 reads (45,009,909 paired) raw reads were used as input. 89,811,296 reads (44,905,648 paired) remained after quality trimming and 11,298,162 reads (5,644,581 paired) were left after normalization (Table 2). The normalized reads were assembled into a transcriptome with a total length of approximately 87.29 Mb and produced 86,373 transcript isoforms and 62,666 unique

transcripts with a GC content of 38.25%. 10% (N10) of the 86,373 transcripts are larger than 3,935 bp, 30% (N30) are larger than 2,420 bp, and half (N50) are larger than 1,332 bp. The assembly ranged from 201-20,736 bp with the median length of 645 bp and the mean length of 1010.57 bp. The summary and transcript length distribution can be found alongside the other stages in Table 4 and in Figure 4.

The cyst stage had a higher quality assembly based on the higher overall transcript lengths when compared to the polyp and jellyfish transcriptome (Table 3). Although the cyst stage did not have the highest number of assembled transcripts, the stage exhibited the highest number of assembled bases, N50 statistics, median and mean contig length, and incorporated the largest assembled contig among the transcriptomes. The number of assembled and unique transcripts may be dependent on the nullified physiological nature of the cyst stage, in which most phenotypic characters are reduced. 72.55% of the total number of assembled transcripts were unique in the cyst stage, while the polyp and jellyfish stage had 82.51% and 78.12%, respectively, the lower percentage of unique transcripts potentially reflecting its reduced physiological state. The higher overall quality of the cyst transcriptome is most likely attributed to the longer lengths of PE sequencing reads, the larger number of input reads, and the upgraded Illumina sequencing platform, as well as the more stringent quality analyses of the reads prior to the assembly. Overall, a higher quality assembly provides a better baseline to investigating genes that are involved in the cyst stage, the stage directly involved in the process of transdifferentiation.

3.2 Completeness of transcriptomes

Read mapping evaluation of assembly

Normalized reads resulting from Trinity's *in silico* normalization were mapped back to the assembled contigs for each transcriptome to evaluate the completeness of its assembly. Two different alignment software based on different algorithms, TopHat v2.1 and CLC Genomic Workbench v8 read mapper tool, were utilized for the assessment. A high mapping percentage indicates that the *de novo* assembly incorporated the majority of sequenced reads and represents high completeness of the transcriptome in terms of its input content (Zhao et al., 2011; Ghaffari et al., 2015).

The TopHat read mapping evaluation exhibited an overall lower trend in mapping percentages than the alignments in CLC Genomic Workbench (Table 6). In both assessments, the polyp assembly exhibited more than 10% lower mapping percentages in both TopHat and CLC Genomic Workbench assessments (69.5%, 84.2%) than the jellyfish (81.2%, 95.1%) and cyst (80.0%, 98.8%). A closer look at the completeness of the transcriptome in terms of gene content will provide insight on whether a significant amount of genetic information was lost in the polyp stage as a lower percentage of reads were incorporated into the assembly.

	TopHat			CLC Genomic Workbench		
	Polyp	**Jellyfish**	**Cyst**	**Polyp**	**Jellyfish**	**Cyst**
Input Reads (normalized)	17,552,220	6,220,660	11,289,162	17,552,220	6,220,660	11,289,162
Mapped Reads	12,205,483	5,050,091	9,029,379	14,770,198	5,915,526	11,168,291
Read Mapping %	69.5%	81.2%	80.0%	84.2%	95.1%	98.9%

Table 6: Read mapping evaluation of transcriptomes.

BUSCO evaluation of gene content

Completeness of the three assembled transcriptomes in terms of gene content was assessed using BUSCO v2.0 software (Simão et al., 2015). The software utilizes taxa-specific datasets consisting of Benchmarking Universal Single-Copy Orthologs (BUSCOs) from OrthoDB v9 (Zdobnov et al., 2016; PMID: 27899580) found to be present in more than 90% of species in all categories. This approach utilizes the BLAST+ v2.2.28 program and HMMER v3.1b2 hidden Markov models to screen for sets of genes that represent the completeness of a transcriptome while accounting for rare gene duplications and/or loss events that often occur when species diverge. A large portion of BUSCOs missing from an assembly is an indication poor coverage in capturing the entire gene expression profile of the sequenced specimen. BUSCOs will be evaluated as complete, fragmented, or missing. Complete BUSCOs have match scores and length alignments within the expected range, while fragmented BUSCOs have match scores within the expected range but do not meet the expected range of length alignments and are considered to be incomplete transcripts. Missing BUSCOs have no significant matches or match score is below the threshold to be considered fragmented.

Two assessments, one with the Eukaryota database and other with the Metazoa database, were conducted on BUSCO v2.0 on each stage-specific assembly. Two

different databases of differing phylogenetic distances were utilized to comparatively confirm the results of the assessments. All assemblies in both assessments exhibited a high coverage and completeness of the transcriptome, as all percentages of complete BUSCOs were above 90% (Table 7). Both assessments showed similar trends in which the polyp stage had the highest percentage of complete BUSCOs (97.0%, 95.4%), the jellyfish had the 2nd highest (95.7%, 93.0%), and the cyst stage had the lowest percentage (95.4%, 91.3%). This implies that despite exhibiting the lowest percentage of reads mapped back to the contigs (see section above), no significant amount of genetic information was lost in the unmapped reads. It should be noted that the cyst stage may not incorporate a higher number of BUSCOs due to its unusual physiological state as well as the small sample size used to prepare the sequencing library (N=2). Overall, the BUSCO results indicate the assemblies provide a quality foundation to a diverse range of downstream analyses such as quantitative gene enrichment, screening for candidate genes, and differential gene expression.

	Eukaryota database (Total BUSCOS: 303)						**Metazoa database (Toal BUSCOS: 978)**					
	Polyp		**Jellyfish**		**Cyst**		**Polyp**		**Jellyfish**		**Cyst**	
	#	%	#	%	#	%	#	%	#	%	#	%
Complete BUSCOs	294	97.0%	290	95.7%	289	95.4%	933	95.4%	909	93.0%	893	91.3%
Fragmented BUSCOs	9	3.0%	8	2.6%	9	3.0%	22	2.2%	33	3.4%	35	3.6%
Missing BUSCOs	0	0.0%	5	1.7%	5	1.6%	23	2.4%	36	3.6%	50	5.1%

Table 7: BUSCO evaluation of transcriptomes.

3.3 Functional Annotation and Characterization of Transcriptomes

Assigning functions to transcripts is a fundamental step to better understand and investigate genetic networks and activity. Functional annotation using Gene Ontology (GO) terms categorizes the data into broader physiological subgroups based on molecular function, biological processes or cellular components. The Blast2GO annotation pipeline consisting of Blastx, mapping and GO annotation steps were used to assign functional terms to the polyp, jellyfish, and cyst assemblies.

Blastx analyses

For the Blastx step, an e-value cutoff of $1e^{-3}$ was utilized. Though not set it stone, the e-value cutoff of $1e^{-3}$ is considered a standard for an acceptable but liberal threshold (Wit et al., 2012). This threshold was chosen at the annotation stage because such stage is considered more exploratory, as opposed to a stricter comparison to only closely related taxa (e.g. Hydrozoans, Cnidarians). In future projects, it will be valuable to perform the Blastx search using more stringent e-value cutoffs (such as $1e^{-6}$ or $1e^{-9}$) to determine whether our resulting annotation content changes in any significant way as different e-value cutoffs are imposed.

45,516 out of 92,659 contigs in the fertile polyp assembly, 34,171 out of 74,639 contigs in the jellyfish assembly, and 38,565 out of 86,373 contigs showed significant similarity (e-value $\leq 1e^{-3}$) to proteins in the GenBank Non-Redundant (NR) database

(Table 8). A maximum of top 20 blast hits were saved for each contig, and each contig was assigned names based on the top-blast hit.

	Fertile Polyps	Jellyfish	Cyst
Total # contigs	92,659	74,639	86,373
Successful contigs after Blast	45,904	34,171	38,565
Sucessful contigs after Mapping	38,686	27,443	29,634
Successful contigs after B2G Annotation	31,719	21,898	25,440

Table 8: Summary of transcriptome annotation.

Investigation of ecological contamination

For all stages, the majority of sequences had top hits that matched *Hydra vulgaris* sequences (Figure 5 for the polyp, Figure 6B for the jellyfish, and Figure 6C for the cyst). This was expected as *Hydra* is the most prominently researched hydrozoan that has a sequenced genome. However, the 2nd highest top-hit species in the polyp assembly (but not of the jellyfish and the cyst) was the foraminifera *Reticulomyxa filosa* (Protista) with 4,585 top-hits (Figure 5 for the polyp, Figure 6B for the jellyfish, and Figure 6C for the cyst).

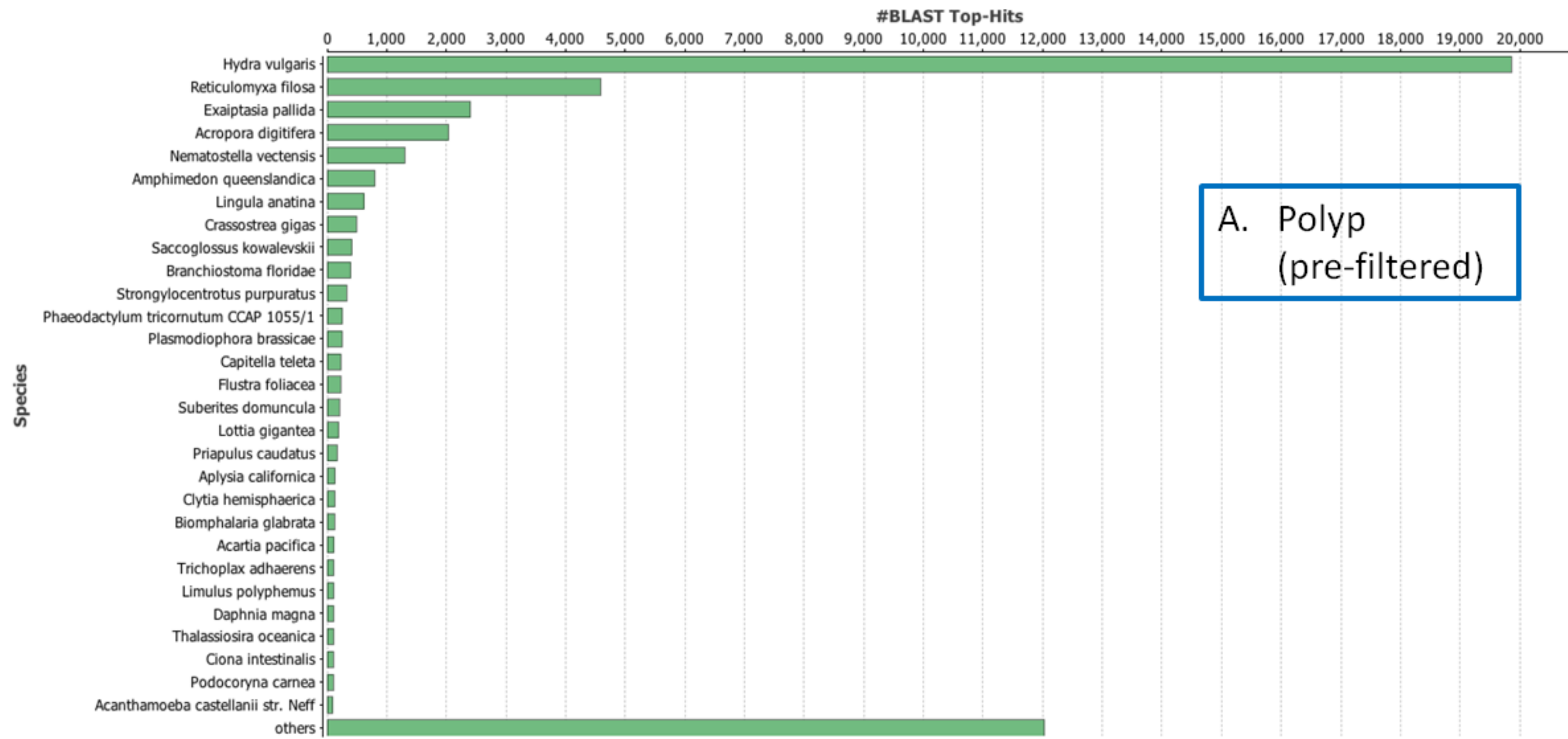

Figure 5: Pre-filter top-hit species distribution of the polyp transcriptome.

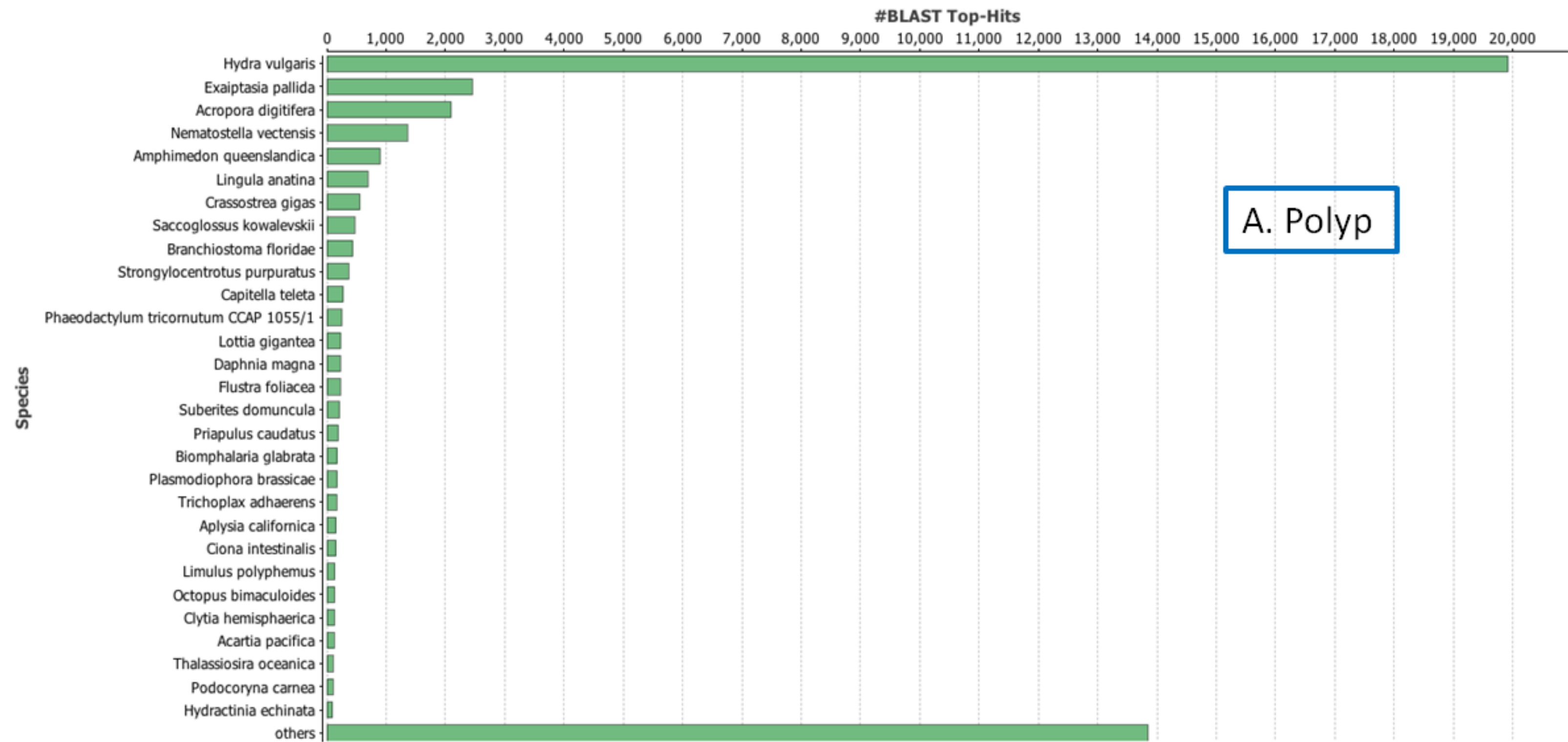

Figure 6: Top-hit species distribution of transcriptomes. A) Polyps, B) Jellyfish, C) Cyst stage

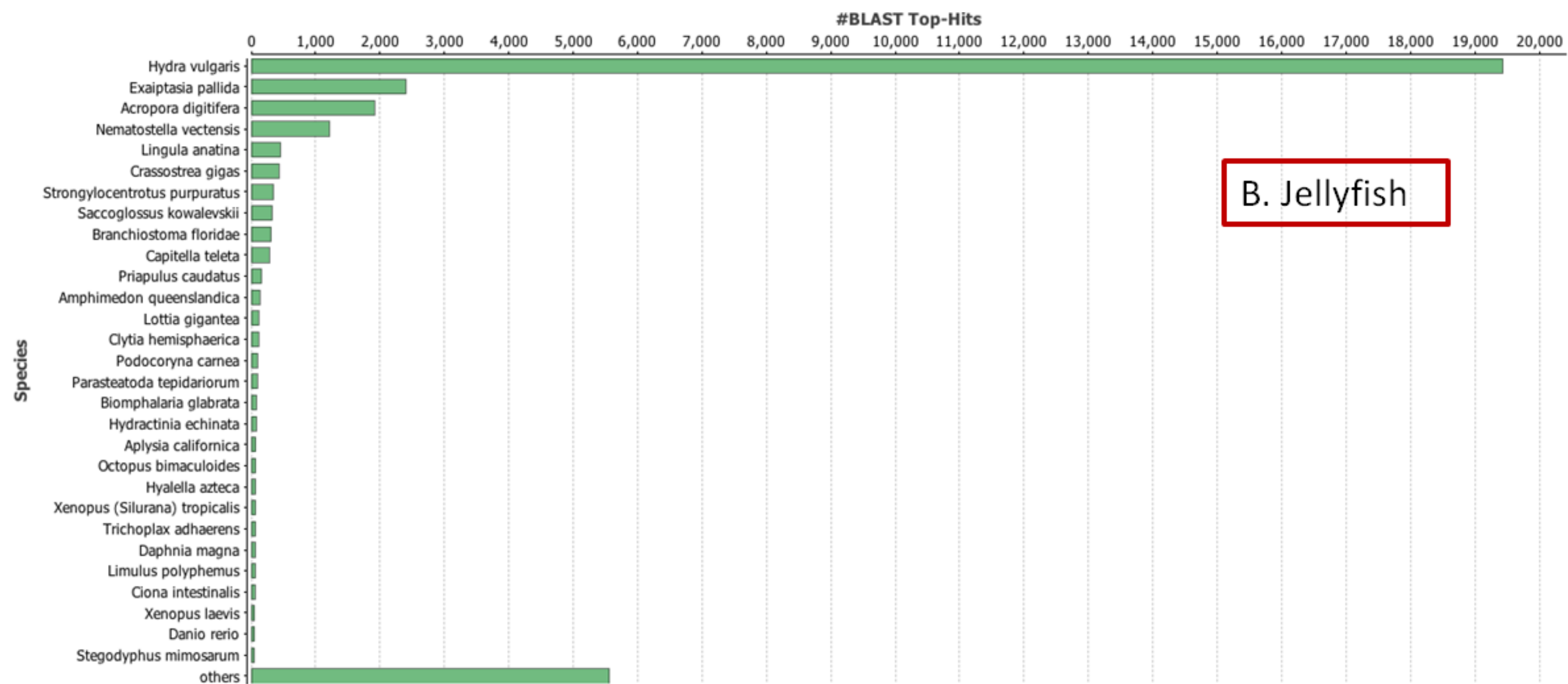

Figure 6: Continued.

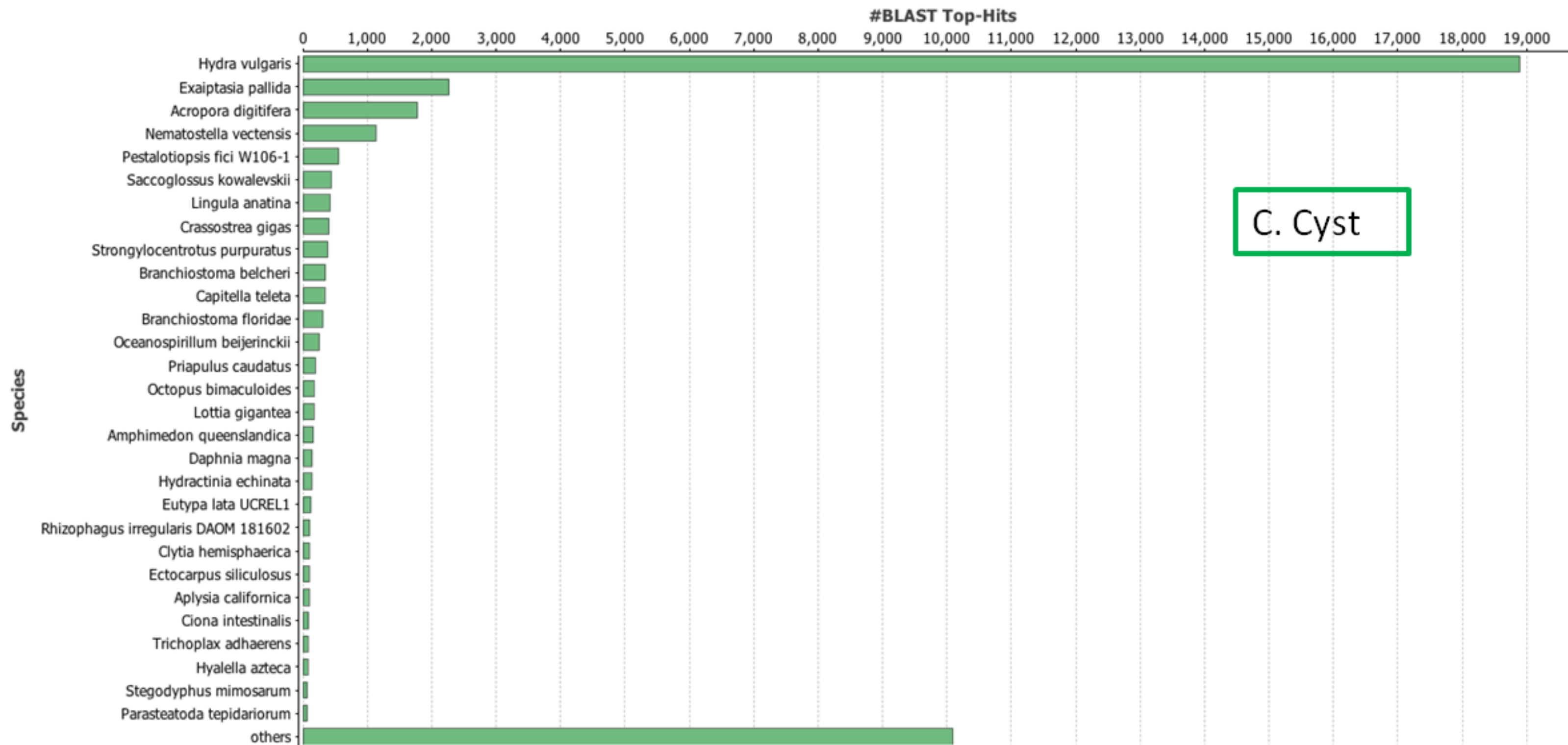

Figure 6: Continued.

This could be due to contamination of foraminiferan RNA within the polyp transcriptome. It has been reported that protists are epibiontic (living associated with organisms of various taxa) on many hydrozoan colonies (Bavestrello et al., 2008). Jellyfish on the contrary do not host epibionts because they are solitary and planktonic, and the cyst stage was obtained under laboratory conditions and contamination is unlikely. When further investigated, the polyp transcriptome had 5,540 sequences with at least one hit against *R. filosa.* Among the 5,540 sequences (82.76%) resulted as top hits. As a comparison, the jellyfish stage only had 65 sequences with at least one hit against *R. filosa,* and the cyst stage had 98 sequences. Among the *R. filosa* hits in the jellyfish and cyst transcriptomes, a majority of the sequences had matched with higher similarity with cnidarian species than *R. filosa*, reaffirming that the fertile polyp transcriptome most likely had ecological contamination within its assembly. In order to address this ecological contaminant issue, the hydrozoan contigs were filtered from the set of 5440 contigs that had at least one top-hit with *R. filosa* by re-blasting the sequences against the metazoan database from NCBI. 3769 sequences out the 5540 sequences had hits against the metazoan database indicating a small portion of most likely contaminate sequences within the overall transcriptome (1.91% of total contigs, 3.86% of total contigs with blast hits), and these blast results were replaced the former result. Post-filtering, the polyp transcriptome had 45,904 out of 92,659 total contigs showed significant similarity (e-value $\leq 1e^{-3}$) to proteins in the GenBank Non-Redundant (NR) database (Table 8). All statistics and graphs were generated using the updated polyp transcriptome.

Blastx output and visualization

The e-value distribution graph of all stages (Figure 7A-C) portray that a very large portion of the blast hits correlate to matches that were below the e-value of $\leq 1e^{-180}$, indicating a very high similarity of matches against the NR database. Shorter contigs are less likely to have matches against the NR database. This is due to the nature of the blast search, where significant sequence similarity is highly dependent on the length of the query sequence. This indeed is seen in both of our transcriptomes as 36,133 of the 46,755 (77.28%) sequences that had no hits were 500 bps or less in the fertile polyp assembly, 29,903 out of 40,469 (73.89%) sequences in the jellyfish, and 29,753 out of 47,808 (62.23%) sequences in the cyst. Longer contigs without blast hits may be novel genes that may have no record in the NR database. 281 contigs with no blast hits were greater than or equal to 2000 bp for the polyp, 345 contigs for the jellyfish, and 1,030 contigs for the cyst stage. The cyst stage had a surprisingly large number of long contigs that had no blast hits as compared to the other two stages. This may partly be due to its unique and unusual physiological state and expression of genes that are specific to the cyst. A follow-up downstream differential gene expression (DGE) analysis of the cyst stage will help determine whether these genes have significant roles in transdifferentiation.

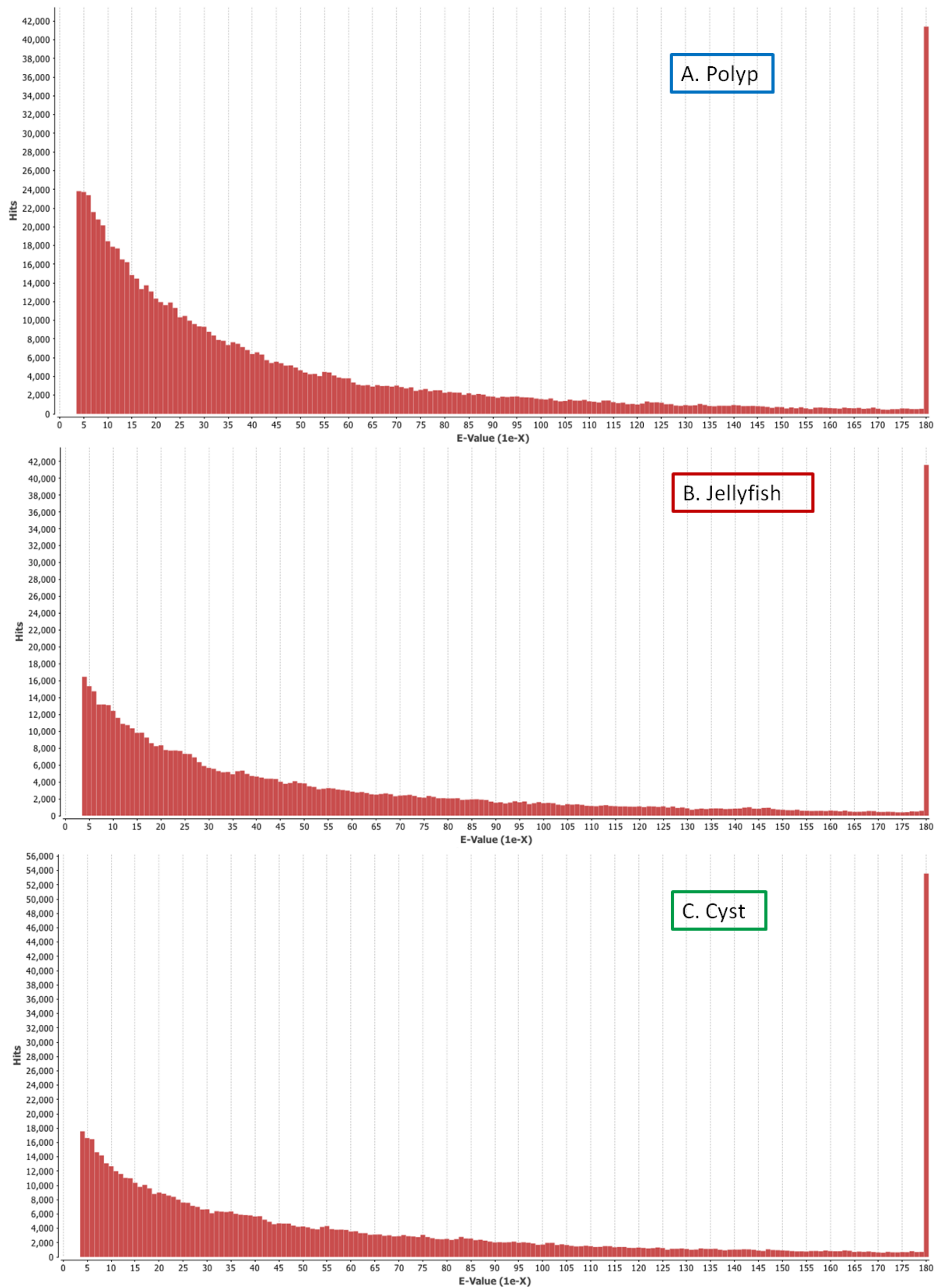

Figure 7: E-value distribution of transcriptomes. (A) Polyp, B) Jellyfish and C) Cyst stage

As expected, the majority of the sequences with top-blast hits belonged to the top four cnidarian species in all stages (Figure 6A-C). *Hydra vulgaris* had the highest number top-blast hits for all assemblies with 19,922 hits in the polyp, 19,435 from the jellyfish, and 18,895 from the cyst assembly. As previously stated, having the highest similarity with *H. vulgaris* is expected as it is the main traditional model hydrozoan system with a sequenced genome. The *H. vulgaris* top-hits accounted for a significant portion in all assemblies, 43.40% of all sequences with blast hits in the polyps, 56.87% in the jellyfish, and 49.00% in the cyst. The 2nd – 4th top-hit species for all stages were all of class Anthozoa and totaled 5,905 top-hits in the polyp, 5,527 in the jellyfish, and 5,190 in the cyst transcriptome. In total, the top four cnidarian species with the most top-hits accounted for 56.26% in the polyp, 73.05% in the jellyfish, and 62.45% in the cyst assembly. The consistency of the order and composition of highest top-hit species across the stages and similarity to closely related taxa belonging to the same phylum indicates high quality assemblies.

Finally, 3,300 sequences (6,358 hits) matched with human transcripts were found in the polyp assembly, 2,137 sequences (4,296 hits) in the jellyfish, and 2,277 sequences (4,632 hits) in the cyst (Figure 8A-C). Matched with the house mouse (*Mus musculus*), the polyp had 2,686 sequences (5,121 hits), jellyfish had 1,882 matches (3,727 hits), and the cyst had 1,786 matches (3,442 hits) (Figure 8A-C). This implies the conservational nature of genome evolution as a portion the contigs that have alignments that showed significant similarity with mammalian sequences.

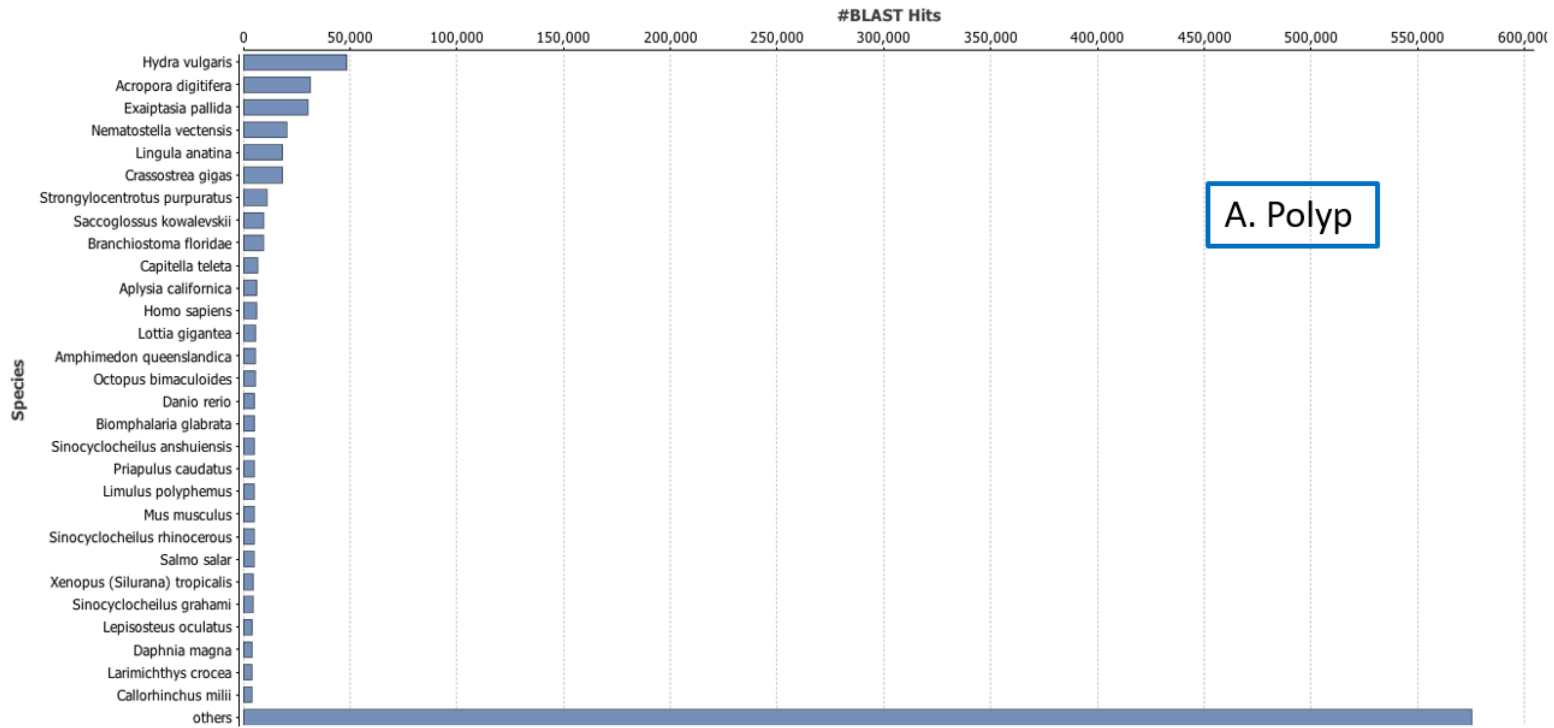

Figure 8: Hit species distribution of transcriptomes. A) Polyp, B) Jellyfish and C) Cyst stage

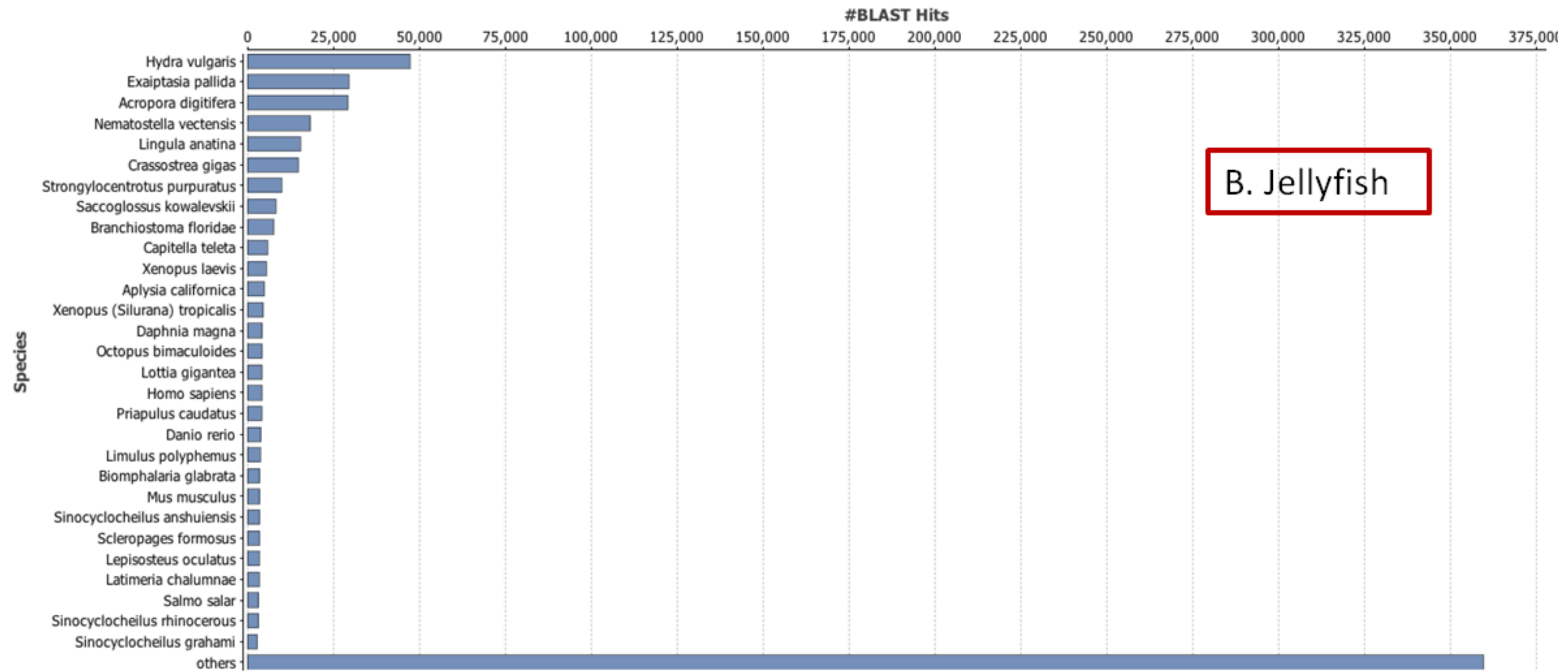

Figure 8: Continued.

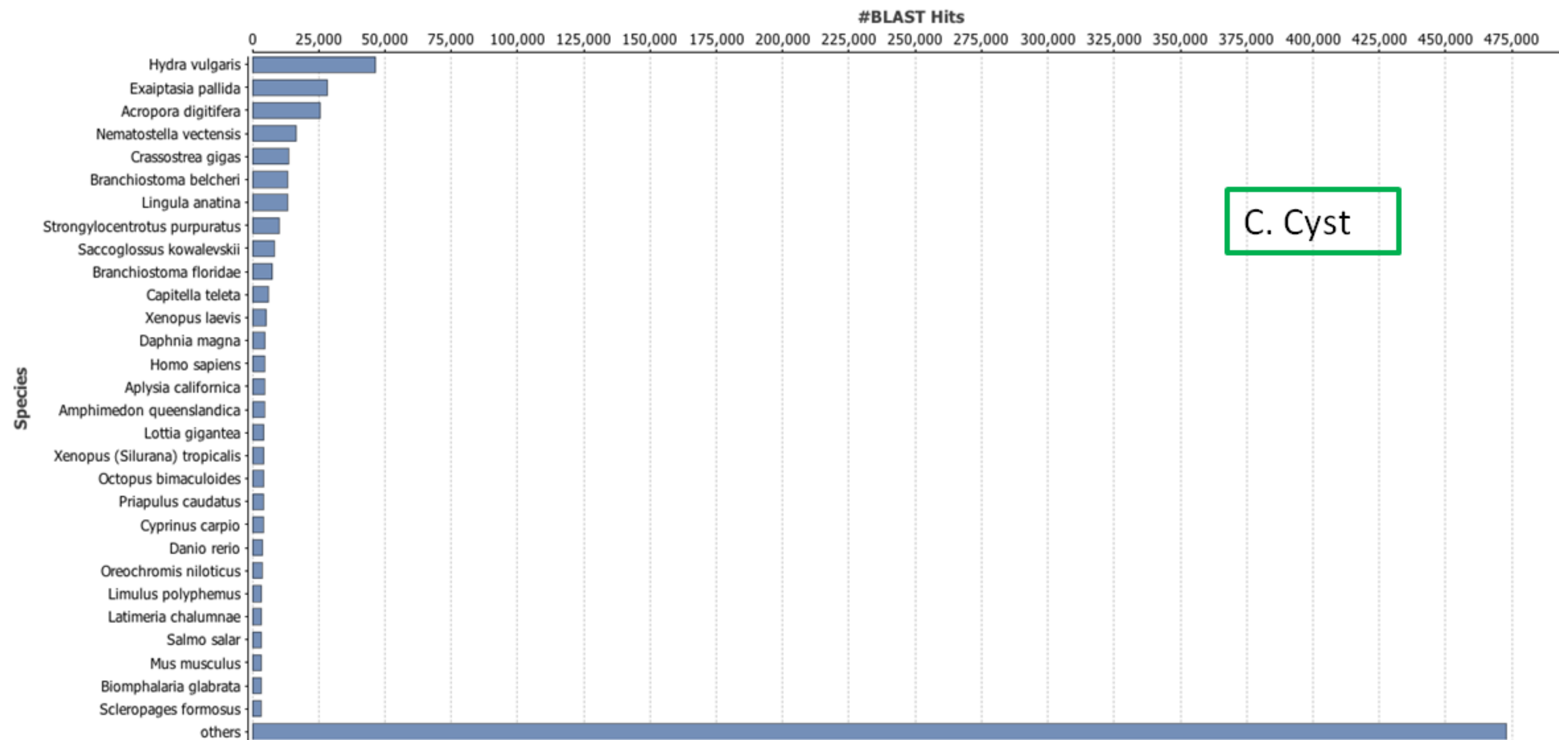

Figure 8: Continued.

Mapping

Mapping is the process where GO terms associated with the top-blast hits are retrieved for each contig. The Blast2GO program performs the mapping process against three databases: gene product table of the GO database, Non-redundant Reference Protein database (Uniprot, Swiss-prot, TrEMBL, etc.), and Database Cross Reference (DBXRef) Table of the GO database. 38,686 out of 45,904 contigs with blast hits were successfully mapped in polyp, 24,443 out of 34,171 contigs in the jellyfish, and 29,634 out of 38,565 contigs in the cyst assembly (Table 8).

GO ontology assignment

The final step consists of assigning functional annotations to the mapped sequences based on their Gene Ontology (GO) terms. GO terms distinguish and assign transcripts to one of three categories: biological process, molecular function and cellular component. InterProScan was additionally used to conduct domain-based searches against the BlastProDom, FPrintScan, HMMPIR, HMMPfam, HMMSmart, HMMTigr, ProfileScan, HAMAP, SuperFamily, HMMPanther and Gene3D databases to assign new GO terms and confirm existing annotations. In total, the polyp transcriptome resulted with 31,719 annotations, jellyfish with 21,898 annotations, and the cyst stage with 25,440 annotations (Table 8).

Enzymatic pathway analyses

Contigs with enzyme commission (EC) codes were mapped to pathways found in the Kyoto Encyclopedia of Genes and Genomes (KEGG) via the Blast2GO software (Kanehisa et al., 2015). In the polyp transcriptome, 13,539 sequences mapped to 1,221

enzymes within 131 different pathways. In the jellyfish transcriptome, 7,581 sequences mapped to 1075 enzymes within 130 different pathways. Lastly, 10,219 sequences mapped to 1,240 enzymes within 130 different pathways in the cyst transcriptome. The top 50 pathways with the most sequences in all three stages were compiled and can be found in APPENDIX A.

Though the ranking order based on number of sequences differed, the total number of mapped enzyme pathways was similar among the stages, with the majority of mapped pathways found consistently throughout all of the stages. Among all stages, the pathway with the highest number of sequences was the purine metabolism pathway, and the pathway with the highest number of enzymes was the biosynthesis of antibiotics pathway. Downstream differential gene expression analyses will be interesting to determine whether any of the enzymes reported in these pathways are substantially over- or under-expressed in the cyst as compared to the polyp and jellyfish despite exhibiting similar pathways.

3.4 Functional enrichment analyses

Categorizing the data based on broader GO terms creates the opportunity to further understand the specific physiological components found in *Turritopsis dohrnii* and assess the functional differences between the different stages. Blast2GO Pro incorporates the FatiGO (Al-Shahrour et al., 2004) tool for enrichment analyses to statistically assess the functional annotation differences between two datasets. FatiGO utilizes the Fischer's Exact Test, commonly used in genomics to compare two RNA-seq datasets that have been constructed with no biological replication, often seen in *de novo* transcriptome assemblies. The output is based on the three GO categories, biological processes, molecular function and cellular components, and reports the over- or under-expression of subgroups, providing an extremely useful tool to understand the differences in gene pathways and networks between the polyp, jellyfish and cyst stage of *T. dohrnii*. The biological processes domain includes recognized series of molecular events and networks pertinent to the functionality of the cell, organs or organism as a whole. The molecular function domain comprises of the specific function of a gene, such as enzyme catalysis and molecular binding. Lastly, the cellular component consists of subcellular and macromolecular structures in which the gene products are located.

A False Discovery Rate (FDR) adjusted p-value, also known as a q-value, of 0.01 was utilized in the gene enrichment analysis. The q-value is a corrected p-value that accounts for and addresses multiple testing by controlling the number of false discoveries within the significant results. Thus, a q-value of 0.01 implies that 1% of significant tests,

rather than all tests, will result in false positives. Since our results from this enrichment analysis will further be used downstream as a foundation to further investigate networks and genes that are associated with *T. dohrnii*'s transdifferentiation, a stringent q-value of 0.01 was used for the enrichment analyses to ensure most results are truly significant.

The following analyses were conducted with the cyst as the central stage in our comparison as the stage directly involved in the process of transdifferentiation: 1) Cyst vs. Polyp; 2) Cyst vs. Jellyfish; and 3) Cyst vs. Polyp/Jellyfish combined. Due to the large number of categories reported as significant, an overall summary of the number of gene enrichment categories that were over- or under- expressed in each of the comparative tests can be found in Table 9.

		Cyst vs. Polyp	**Cyst vs. Jellyfish**	**Cyst vs. Both**
Biological processes	Over	44 (3.27%)	147 (64.19%)	66 (8.99%)
	Under	1302 (96.73%)	82 (35.81%)	734 (91.01%)
	Total	1346	229	780
Molecular function	Over	17 (7.00%)	38 (39.18%)	32 (16.16%)
	Under	243 (93.00%)	59 (60.82%)	166 (83.84%)
	Total	260	97	198
Cellular component	Over	12 (4.71%)	36 (49.31%)	19 (10.92%)
	Under	243 (95.29%)	37 (50.69%)	155 (89.08%)
	Total	255	73	174

Table 9: Number of gene enrichment categories over-or under-expressed. FDR adjusted p-value (q-value) of 0.01 was utilized in all comparative analyses.

Overall, a vastly larger proportion of under-expressed gene categories were found with the highest proportion in the comparison between the cyst and the polyp among all categories (Table 9). The suppression of a large number of genetic pathways is likely due

to the uncharacteristic state of the cyst stage in which morphological structures are reduced. However, the cyst exhibited more over-expressed functional categories in biological processes than suppressed categories as compared to the jellyfish, while the polyp stage reported a substantially large number of under-expressed categories. The biological processes domain is of highest interest to our investigation as it will shed light on the processes of the transcriptional products that is elevated or suppressed in the cyst, and thus, further analyzed. Due to the larger number of total annotated sequences in the polyp transcriptome, the Cyst vs. Both analyses will show a slight bias towards the polyp stage in the reference dataset. This test however, will help uncover the over- and under-expression of gene categories that are highly specific to the cyst.

Biological processes

Over-expression in the cyst

The top 10 over-expressed gene enrichment categories for each analysis are cumulatively reported in Table 10. In the cyst vs. polyp comparison, most of the highest expressed categories are directly involved in DNA synthesis and metabolic processes. The comparison with the jellyfish indicates that there are a heightened number of gene products that are associated with the synthesis of cellular compounds and metabolic processes, particularly related to nitrogen. In addition, the ‘Nucleobase-containing compound metabolic process (GO:006139)’ category was heightened in the cyst in comparison to the jellyfish, indicating heightened pathways involved in the maintenance of genetic material.

Exp.	Analysis	GO ID	GO Name	q-value	p-value	#GO Test	#GO Ref
Over	Cyst vs. Polyp	GO:0006259	DNA metabolic process	4.92e^-39	1.35e^-41	2643	1904
		GO:0006278	RNA-dependent DNA biosynthetic process	9.88e^-31	4.74e^-33	704	355
		GO:0090304	Nucleic acid metabolic process	1.56e^-23	1.32e^-25	6004	5324
		GO:0071897	DNA biosynthetic process	7.87e^-21	8.58e^-23	821	511
		GO:0015074	DNA integration	5.73e^-17	8.31e^-19	636	390
		GO:0000160	Phosphorelay signal transduction system	1.14e^-08	3.74e^-10	43	4
		GO:0006139	Nucleobase-containing compound metabolic process	1.62e^-07	6.38e^-09	7059	6890
		GO:0090305	Nucleic acid phosphodiester bond hydrolysis	7.56e^-07	3.22e^-08	495	363
		GO:0006725	Cellular aromatic compound metabolic process	9.51e^-07	4.13e^-08	7395	7273
		GO:0006399	tRNA metabolic process	9.65e^-07	4.21e^-08	399	280
	Cyst vs. Jellyfish	GO:0034641	Cellular nitrogen compound metabolic process	1.57e^-32	1.45e^-36	8466	5973
		GO:0006807	Nitrogen compound metabolic process	2.07e^-27	3.81e^-31	9497	6900
		GO:1901566	Organonitrogen compound biosynthetic process	1.64e^-23	4.53e^-27	2124	1262
		GO:0044237	Cellular metabolic process	1.83e^-23	6.76e^E-27	12530	9454
		GO:0009058	Biosynthetic process	1.85e^-23	8.54e^-27	6341	4481
		GO:1901576	Organic substance biosynthetic process	2.04e^-23	1.12e^-26	6213	4383
		GO:0044249	Cellular biosynthetic process	1.56e^-22	1.00e^-25	6122	4329
		GO:0046483	Heterocycle metabolic process	3.59e^-22	2.64e^-25	7420	5366
		GO:0006139	Nucleobase-containing compound metabolic process	4.58e^-22	3.79e^-25	7059	5082
		GO:1901360	Organic cyclic compound metabolic process	8.41e^-22	8.51e^-25	7585	5508
	Cyst vs. Both	GO:0006259	DNA metabolic process	1.21e^-34	7.03e^-38	2643	3690
		GO:0090304	Nucleic acid metabolic process	7.82e^-28	1.16e^-30	6004	9669
		GO:0006278	RNA-dependent DNA biosynthetic process	4.11e^-20	1.32e^-22	704	821
		GO:0006139	Nucleobase-containing compound metabolic process	1.42e^-17	5.76e^-20	7059	11972
		GO:0015074	DNA integration	3.78e^-17	1.56e^-19	636	754
		GO:0046483	Heterocycle metabolic process	1.02e^-16	4.50e^-19	7420	12672
		GO:0006725	Cellular aromatic compound metabolic process	1.47e^-16	6.58e^-19	7395	12635
		GO:1901360	Organic cyclic compound metabolic process	3.49e^-16	1.64e^-18	7585	13001
		GO:0071897	DNA biosynthetic process	8.17e^-16	4.12e^18	821	1059
		GO:0000160	Phosphorelay signal transduction system	4.66e^-14	2.67e^-16	43	4

Table 10: Top 10 over-expressed enrichment categories in biological processes.

Telomere maintenance and organization

Telomeres are composed of repetitive sequences at the end of chromosomes that protects the genetic material from deterioration or fusion with adjacent chromosomes. Telomeres are shortened each time the cell undergoes DNA replication, and this loss of genetic material is attributed to cellular senescence and the aging phenotype. In immortal cell lineages, such as germ cells and cancerous cells, telomeres length is replenished and maintained by the telomerase enzyme.

The GO term ‘Telomere maintenance (GO:0000723)’ and ‘Telomere organization (GO:0032200)’ was over-expressed in the cyst as compared to the polyp and both stages combined (Table 11). Telomere maintenance includes processes that are involved with the maintenance of telomere length and structure through replication and repair (Gene Ontology Consortium, 2015). Telomere organization consists of processes that are involved in the assembly and arrangement of telomeres at the cellular level (Gene Ontology Consortium, 2015). The cyst had 234 sequences out of 25,440 total annotated sequences in the telomere maintenance category, and 236 sequences in the telomere organization category. On the other hand, the polyp stage exhibited a considerably less 154 out of 31,719 total annotated sequences and 160 sequences, respectively. The cyst vs. both comparison exhibited similar patterns with a significantly less proportion of sequences involved in telomerase maintenance and organization. Though the comparison with the jellyfish stage did not report these processes as significant, perhaps due to our very stringent q-value of 0.01, the jellyfish contained 156 sequences associated with telomere maintenance and 163 sequences with telomere organization. It was noted that the polyp and jellyfish stage exhibited a similar number of transcripts associated with telomerase maintenance and organization.

Exp.	GO ID	GO Name	Analysis	q-value	p-value	#GO Test	#GO Ref
Over	GO:0000723	Telomere maintenance	Cyst vs. Polyp	3.50e^-05	2.04e^-06	234	154
			Cyst vs. Both	4.12e^-04	1.60e^-05	234	310
	GO:0032200	Telomere organization	Cyst vs. Polyp	1.06e^-04	6.86e^-06	236	160
			Cyst vs. Both	1.49e^-03	6.25e^-05	236	323

Table 11: GO enrichment categories involved in telomere maintenance.

T. dohrnii's ability to undergo reverse development and rejuvenate its life cycle may involve a mechanism in which DNA is continuously repaired to enable the organism to avoid aging and repeat its lifecycle. Heightened activity of processes that are involved in the regulation of telomere length may be a key part of the overall cellular network that enables the ability for life cycle reversal. To further look into whether telomere maintenance is fundamental to *T. dohrnii*'s life cycle reversal, quantifying the expression of these transcripts found in the both categories will be very insightful. Quantification of expression levels (e.g. differential gene expression analysis) will help interpret the degree of expression change in telomere maintenance and organization in terms of both content and level of activity.

Chitinous cyst exterior

Perhaps the only characteristic that is prominent in the cyst stage is the presence of a hard chitinous sheath exterior that develops after the jellyfish reduces all of its morphological features and into uncharacterized tissue (Schmich et al., 2007). This external structure, called the perisarc, is composed of chitin, the same cell wall component of fungi. Polyps also produce a thin chitinous sheath that protects their soft bodies, but jellyfish do not exhibit this character (Piraino et al., 1996).

Although a large number of processes involved in differentiation, biogenesis and development of bodily structures were significantly suppressed in the cyst, there was an upregulation in pathways that associated with the formation of a cell wall in fungi and the process of sporulation (Table 12). Consistent to the phenotypic features the three *T. dohrnii* stages, the cyst stage exhibited the most proportion of sequences in chitin-related

pathways and the polyp exhibited a few sequences, while the jellyfish exhibited none. In addition, a more general category of ‘External encapsulating structure organization (GO:0045229) was also found to be upregulated in the cyst in comparison to the jellyfish and when the polyp and jellyfish stages were combined. Thus, there is a potential in which the transcripts involved in the formation of fungus-type cell walls and sporulation processes attribute to the perisarc of the cyst. However, further analyses need to be carefully implemented as single celled fungal organism such as yeast can easily be contaminants of the polyp and cyst stage accidentally considered in the analysis, particularly due to the small number of overall sequences in each category among the cyst and polyp stage.

Exp.	GO ID	GO Name	Analysis	q-value	p-value	#GO Test	#GO Ref
Over	GO:0071852	Fungal-type cell wall organization or biogenesis	Cyst vs. Polyp	6.35e^-03	7.04e^-04	22	5
			Cyst vs. Jellyfish	1.24e^-04	1.16e^-06		0
			Cyst vs. Both	2.40e^-05	1.04e^-06		5
	GO:0071940	Fungal-type cell wall assembly	Cyst vs. Both	3.47e^-03	1.81e^-04	8	0
	GO:0043934	Sporulation	Cyst vs. Jellyfish	6.30e^-04	7.70e^-06	19	0
			Cyst vs. Both	2.80e^-03	1.41e^-04		8
	GO:0030435	Sporulation resulting in formation of a cellular spore	Cyst vs. Jellyfish	1.85e^-03	2.73e^-05	17	0
	GO:0034293	Sexual sporulation	Cyst vs. Jellyfish	5.26e^-05	9.74e^-05	15	0
	GO:0043935	Sexual sporulation resulting in formation of a cellular spore	Cyst vs. Jellyfish	8.99e^-03	1.84e^-04	14	0
	GO: 0030437	Ascopore formation	Cyst vs. Jellyfish	8.99e^-03	1.84e^-04	14	0
	GO:0070591	Ascospore wall biogenesis	Cyst vs. Both	8.56e^-03	5.30e^-04	7	0
	GO:0030476	Ascospore wall assembly	Cyst vs. Both	8.56e^-03	5.30e^-04	7	0
	GO:0045229	External encapsulating structure organization	Cyst vs. Jellyfish	6.74e^-03	1.30e^-04	19	1
			Cyst vs. Both	9.76e^-03	2.77e^-04	19	10

Table 12: GO enrichment categories potentially involved in the chitinous cyst exterior.

Under-expression in the cyst

The top 10 under-expressed gene enrichment categories for each analysis are cumulatively reported in Table 13. In the cyst vs. polyp analysis, the prominent processes suppressed in the cyst are directly related to the maintenance, specifically the synthesis, of proteins. This includes pathways such as ‘Protein metabolic process GO:0019538’, ‘Translation (GO:0006412)’, ‘Peptide biosynthetic process (GO:0043043)’, and ‘Ribosome biogenesis (GO:0042254)’. In addition, processes involved in the creation of components in cells were highly suppressed, which can be easily justified when comparing the highly-specialized phenotypes of the polyp versus the uncharacteristic cyst. The cyst in comparison to the jellyfish exhibited a down-regulation in biological processes involved in cell signaling and communication, such as ‘G-protein coupled receptor signaling pathway (GO:0007186)’, ‘Signal transduction (GO:0007165)’, ‘Single organism signaling (GO:0044700)’, and ‘Cell communication (GO:0007154)’. The general categories of biological processes necessary for functionality of a multicellular organism (in reference to GO:0032501 and GO:0044707) were also downregulated along with cell division (in reference to GO:0098763). Overall, the top-10 under-expressed gene categories among the three comparative analyses indicate that cellular communication, characterization and growth seem to be suppressed in the cyst stages in comparison to the polyp and jellyfish.

Exp.	Analysis	GO ID	GO Name	q-value	p-value	#GO Test	#GO Ref
Under	Cyst vs. Polyp	GO:0019538	Protein metabolic process	5.07e^-126	4.35e^-130	5699	9022
		GO:0008152	Metabolic process	2.04e^-122	4.13e^-126	14989	20234
		GO:0006412	Translation	2.04e^-122	5.27e^-126	1352	3048
		GO:0043043	Peptide biosynthetic process	1.10e^-119	3.80e^-123	1385	3076
		GO:0044267	Cellular protein metabolic process	3.07e^-119	1.32e^-122	4629	7551
		GO:0043604	Amide biosynthetic process	2.41e^-116	1.24e^-119	1465	3169
		GO:0006518	Peptide metabolic process	1.81e^-113	1.08e^-116	1534	3249
		GO:0042254	Ribosome biogenesis	4.70e^-113	3.23e^-116	982	2397
		GO:0044085	Cellular component biogenesis	1.07e^-109	8.31e^-113	2367	4411
		GO:0043603	Cellular amide metabolic process	1.07e^-107	9.20e^-111	1729	3485
	Cyst vs. Jellyfish	GO:0007186	G-protein coupled receptor signaling pathway	6.31e^-18	9.29e^-21	439	668
		GO:0032501	Multicellular organismal process	6.41e^-13	1.76e^-15	2484	2670
		GO:0007165	Signal transduction	8.41e^-12	2.47e^-14	3116	3248
		GO:0023052	Signaling	2.75e^-11	8.37e^-14	3293	3403
		GO:0044700	Single organism signaling	4.72e^-11	1.49e^-13	3285	3389
		GO:0007154	Cell communication	4.72e^-11	1.52e^-13	3349	3449
		GO:0044707	Single-multicellular organism process	1.14e^-09	4.12e^-12	2269	2396
		GO:0030001	Metal ion transport	1.14e^-07	4.52e^-10	456	581
		GO:0098763	Mitotic cell cycle phase	2.37e^-07	1.02e^-09	2	34
		GO:0006813	Potassium ion transport	4.77e^-07	2.19e^-09	111	193
	Cyst vs. Both	GO:0019538	Protein metabolic process	3.35e^-46	2.77e^-50	5699	13791
		GO:0044699	Single-organism process	2.87e^39	4.74e^-43	11375	25406
		GO:0044267	Cellular protein metabolic process	9.47e^-38	2.34e^-41	4629	11226
		GO:0051179	Localization	2.19e^-35	7.26e^-39	4593	11073
		GO:0044085	Cellular component biogenesis	2.73e^-35	1.13e^-38	2367	6206
		GO:0032501	Multicellular organismal process	9.66e^-35	4.79e^-38	2484	6452
		GO:0042254	Ribosome biogenesis	8.78e^-32	6.26e^-35	982	2943
		GO:0051234	Establishment of localization	8.78e^-32	6.53e^-35	4156	10012
		GO:0006810	Transport	1.92e^-31	1.59e^-34	4095	9870
		GO:0006412	Translation	1.88e^-30	1.71e^-33	1352	3786

Table 13: Top 10 under-expressed enrichment categories in biological processes.

Lifespan and aging

Physical deterioration and the decline in cellular functionality and regenerative potential are characteristics associated with the process of aging. Post-reproductive *T. dohrnii* jellyfish have the ability reverse its ontogenetic sequence to avoid the ultimate consequence of aging, death. To avoid the detrimental effects of replicative senescence, it is hypothesized that *T. dohrnii* has developed mechanisms that prevent the loss of cellular functionality and regenerative potential is halted.

Groups associated the cellular senescence and biological aging of organisms were found to be under-expressed in the cyst, particularly in comparison to the polyp stage (Table 14). The category 'Determination of adult lifespan (GO:0008340)' was under-expressed in the cyst vs. polyp and cyst vs. both analyses, and is defined as processes that control the viability and duration of the adult life cycle phase in organisms (Gene Ontology Consortium, 2015), constituting an important aspect of aging. This GO term is a part of 'Multicellular organism aging (GO:0010259)', and the broader category of 'Aging (GO:007568)', which were also under-expressed in the cyst in comparison to the polyp. Though the comparison with the jellyfish did not report as significant, 17 sequences were found to be associated with process involved in the 'Determination of adult lifespan' category. Though not nearly as over-expressed as the polyp with 77 sequences, the number is considerably higher than the 7 sequences found in the cyst associated with this category.

Exp.	GO ID	GO Name	Analysis	q-value	p-value	#GO Test	#GO Ref
Under	GO:0008340	Determination of adult lifespan	Cyst vs. Polyp	2.46e^-13	4.81e^-15	7	77
			Cyst vs. Both	1.27e^-08	1.44e^-10	7	94
	GO:0010259	Multicellular organism aging	Cyst vs. Polyp	4.60e^-15	7.50e^-17	7	84
	GO:007568	Aging	Cyst vs. Polyp	7.90e^-07	3.39e^-08	70	161

Table 14: GO enrichment categories associated with lifespan and aging.

Biological processes associated with telomere length maintenance are over-expressed in the cyst (Table 11), fundamental processes that diminish replicative senescence and in turn, and potentially may have suppressed pathways involved in

cellular senescence and aging. A more detailed investigation of gene content and expression quantification, in which the reversion process is divided into more numerous and specific stages, will shed further insight on how these processes involved in the lifespan determination and aging change throughout its ontogeny reversal.

Cell division

Mitotic or asexual cell division is a fundamental process that occurs during cellular proliferation events and overall organismal growth and development. Enrichment categories associated with mitotic cell division were under-expressed in the cyst stage as compared to both the polyp and jellyfish (Table 15). Table 15 is not an exhaustive list due to the large number of highly specific categories related to mitotic cell division reported to be under-expressed in the cyst across the three analyses. Although more numerous and specific categories were reported to be suppressed when the cyst was compared to the polyp, 'Mitotic cell cycle phase' (GO:0098763) was 9th most under-expressed enrichment category when compared to the jellyfish (Table 13).

Exp.	GO ID	GO Name	Analysis	q-value	p-value	#GO Test	#GO Ref
Under	GO:0098763	Mitotic cell cycle phase	Cyst vs. Polyp	3.01e^-11	7.19e^-13	2	52
			Cyst vs. Jellyfish	2.37e^-07	1.03e^-09	2	34
			Cyst vs. Both	3.68e^-11	2.80e^-13	2	86
	GO:0000087	Mitotic M phase	Cyst vs. Polyp	4.17e^-10	1.14e^-11	1	43
			Cyst vs. Jellyfish	9.75e^-06	6.46e^-08	1	25
			Cyst vs. Both	1.65e^-09	1.50e^-11	1	68
	GO:0000090	Mitotic anaphase	Cyst vs. Polyp	1.82e^-09	5.42e^-11	0	36
			Cyst vs. Jellyfish	1.06e^-04	9.76e^-07	0	18
			Cyst vs. Both	1.95e^-08	2.32e^-10	0	54
	GO:0051329	Mitotic interphase	Cyst vs. Jellyfish	6.92e^-03	1.36e^-06	1	15
			Cyst vs. Both	3.69e^-03	1.94e^-04	1	27
	GO:0007067	Mitotic nuclear division	Cyst vs. Polyp	1.28e^-07	5.02e^-09	200	357
			Cyst vs. Both	3.19e^-04	1.49e^-05	200	550
	GO:0007052	Mitotic spindle organization	Cyst vs. Polyp	5.88e^-06	2.98e^-07	49	121
			Cyst vs. Both	1.61e^-04	7.39e^-05	49	176
	GO:2000045	Regulation of G1/S transition of the mitotic cell cycle	Cyst vs. Polyp	7.98e^-05	5.07e^-06	44	105
			Cyst vs. Both	9.80e^-03	6.30e^-04	44	151
	GO:0000080	Mitotic G1 phase	Cyst vs. Both	7.65e^-03	4.61e^-04	0	20

Table 15: GO enrichment categories associated with mitotic cell division.

Prior to cell division, the cell undergoes growth and DNA replication. Although processes involved in DNA biosynthesis and maintenance are highly over-expressed in the cyst stage as compared to the polyp and jellyfish stage (Table 10), mitotic cell division processes are highly under-expressed (Table 15). In addition, there is a heightened number of sequences involved in the process of DNA repair in the cyst in comparison to the other stages (Table 16). These results potentially indicate that the cells in the cyst are synthesizing and maintaining DNA, but cell division, proliferation and growth are inhibited.

Exp.	GO ID	GO Name	Analysis	q-value	p-value	#GO Test	#GO Ref
Over	GO:0006281	DNA repair	Cyst vs. Polyp	2.76e^-03	2.62e^-04	831	742
			Cyst vs. Jellyfish	2.27e^-03	3.45e^-05	831	574
			Cyst vs. Both	1.74e^-04	5.88e^-06	831	1316

Table 16: GO enrichment categories associated with DNA repair.

Response to stimuli

The majority of processes associated with the response to external and internal stimuli were a prominently reported to be under-expressed in the cyst stage as compared to both the polyp and jellyfish. Due to the large number of categories under-expressed categories, the results were condensed to categories representing the least specific under-represented categories along with the small number of over-expressed processes associated with response to stimuli (Table 17). ‘Cellular response to stress (GO:0033554)’ was the only over-expressed process on the cyst in comparison to the jellyfish is likely attributed to the nature of the cyst, a stage induced when jellyfish are damaged or weakened and is an exaggerated response to physiological stress. ‘Cellular response to DNA damage stimulus (GO:0006974)’ is over-expressed in the cyst in comparison to the combination of the polyp and jellyfish, consistent with the over-expression of processes involved in DNA repair (Table 16).

Exp.	GO ID	GO Name	Analysis	q-value	p-value	#GO Test	#GO Ref
Over	GO:0033554	Cellular response to stress	Cyst vs. Jellyfish	9.64^-03	2.00e^-04	1434	1067
Over	GO:0006974	Cellular response to DNA damage stimulus	Cyst vs. Both	8.60e^-03	5.36e^-04	999	1684
Under	GO:0050896	Response to stimulus	Cyst vs. Polyp	4.47e^-47	8.05e^-50	5005	6992
			Cyst vs. Jellyfish	2.34e^-04	2.42e^-06	5005	4745
			Cyst vs. Both	1.95e^-29	2.25e^-32	5005	11737
	GO:0009628	Response to abiotic stimulus	Cyst vs. Polyp	8.71e^-24	7.10e^-26	315	676
			Cyst vs. Both	7.05e^-13	4.55e^-15	315	994
	GO:0009607	Response to biotic stimulus	Cyst vs. Polyp	1.12e^-15	1.75e^-17	167	383
			Cyst vs. Both	4.91e^10	4.30e^-12	167	578

Table 17: Go enrichment categories associated with response to stimuli.

Processes in ‘Response to stimulus (GO:0050896)’ were found to be under-represented in the cyst as compared to the other stages (Table 17). The GO term is

broadly defined as processes triggered by a stimulus that result in the change in state or activity within a cell/organism (Gene Ontology Consortium, 2015). The inhibition of response pathways can be physiologically justified by the inactive state of the cyst, in which bodily structures that can detect and respond to stimuli is presumably absent. A large number of more specific categories in part of the broad ‘Response to stimuli’ category, such as the ‘Response to abiotic stimulus (GO:0009628)’ and ‘Response to biotic stimulus (GO:0009607), were particularly reported to be under-expressed in the cyst in comparison to the polyp stage. 145 GO enrichment categories including the key word ‘response’ were reported against the polyp, while 4 were reported against the jellyfish (data not shown).

Cellular differentiation and development

A large number of that involved cell differentiation and organization were consistently down-regulated in the cyst as compared to the polyp and jellyfish (Table 18, list not exhaustive). In addition, 54 processes with the term ‘development’ were reported to be consistently under-expressed in the cyst in comparison to the polyp, 6 compared to the jellyfish, and 29 compared to the polyp and jellyfish combined (data not shown). A few of the broader categories associated with development reported to be under-expressed can be found in Table 19. These results are expected due to the uncharacterized and seemingly inactive physiological state of the cyst stage. A quantification of the gene activity (e.g. differential gene expression) involved in processes associated with cell differentiation and development will provide more insight on how much the transcripts assigned to each GO category are expressed and play a role in the physiology of the cyst.

Exp.	GO ID	GO Name	Analysis	q-value	p-value	#GO Test	#GO Ref
Under	GO:0030154	Cell differentiation	Cyst vs. Polyp	1.46e^-18	1.97e^20	1177	1778
			Cyst vs. Jellyfish	9.98e^-03	1.98e^-04	1177	1185
			Cyst vs. Both	9.73e^-13	6.35e^-15	1177	2963
	GO:0045595	Regulation of cell differentiation	Cyst vs. Polyp	5.91e^-05	3.61e^-06	346	511
			Cyst vs. Both	8.11e^-05	2.43e^-06	346	898
	GO:0000904	Cell morphogenesis involved in differentation	Cyst vs. Polyp	4.86e^-03	4.94e^-04	193	286
			Cyst vs. Both	9.75e^-03	6.25e^-04	193	498
	GO:0051179	Localization	Cyst vs. Polyp	2.13e^-63	3.11e^-66	4593	6757
			Cyst vs. Jellyfish	3.07e^-03	5.01e^-05	4593	4316
			Cyst vs. Both	2.20e^-35	7.26e^-39	4593	11073
	GO:0001709	Cell fate determination	Cyst vs. Polyp	3.35e^-06	1.63e^-07	1	28
			Cyst vs. Both	9.89e^-04	4.28e^-05	1	31
	GO:0045165	Cell fate commitment	Cyst vs. Polyp	4.96e^-03	5.06e^-04	50	98
			Cyst vs. Both	6.02e^-03	3.49e^-04	50	170
	GO:0030030	Cell projection organization	Cyst vs. Polyp	4.14e^-06	2.05e^-07	514	741
			Cyst vs. Jellyfish	2.24e^-05	1.63e^-07	514	607
			Cyst vs. Both	1.70e^-07	2.33e^-09	514	1348
	GO: 0016043	Cellular component organization	Cyst vs. Polyp	4.45e^-23	2.33e^-25	3662	4895
			Cyst vs. Both	1.89e^-09	1.78e^-11	3662	8078
	GO:0048872	Homeostasis of number of cells	Cyst vs. Polyp	1.10e^-04	6.52e^-06	80	158
			Cyst vs. Both	8.03e^-03	4.88e^-04	80	241
	GO:0035567	Non-canonical Wnt signaling pathway	Cyst vs. Polyp	9.45e^-06	4.98e^-07	52	124
			Cyst vs. Both	8.47e^-03	5.20e^-04	52	172
	GO:0009798	Axis specification	Cyst vs Polyp	8.32e^-03	9.31d^-04	32	69
	GO:0007389	Pattern specification process	Cyst vs. Jellyfish	3.75e^-03	6.42e^-05	230	283

Table 18: GO enrichment categories associated with cell differentiation/organization.

Exp.	GO ID	GO Name	Analysis	q-value	p-value	#GO Test	#GO Ref
Under	GO:0032502	Developmental process	Cyst vs. Polyp	4.61e^-37	1.35e^-39	2429	3647
			Cyst vs. Jellyfish	2.29e^-06	1.24e^-08	2429	2465
			Cyst vs. Both	9.08e^-27	1.65e^-29	2429	6112
	GO:0032502	Anatomical structure development	Cyst vs. Polyp	4.66e^-35	1.60e^-37	2242	3380
			Cyst vs. Jellyfish	2.55e^-06	1.43e^-08	2242	2288
			Cyst vs. Both	1.22e^-25	2.72e^-28	2242	5668
	GO:0048731	System development	Cyst vs. Polyp	9.81e^-19	1.30e^-20	1698	2433
			Cyst vs. Jellyfish	6.89e^-06	4.23e^-08	1698	1765
			Cyst vs. Both	6.16e^-16	2.95e^-18	1698	4198
	GO:0048513	Animal organ development	Cyst vs. Polyp	4.77e^-18	6.63e^-20	1175	1768
			Cyst vs. Jellyfish	3.43e^-03	5.68e^-05	1175	1768
			Cyst vs. Both	5.95e^-13	3.74e^-15	1175	1197
	GO:0048468	Cell development	Cyst vs. Polyp	1.15e^-08	3.77e^-10	666	976
			Cyst vs. Both	1.49e^-05	3.43e^-07	666	1625
	GO:0009888	Tissue development	Cyst vs. Polyp	7.73e^-07	3.31e^-08	652	926
			Cyst vs. Both	5.91e^-05	1.67e^-06	652	1573
	GO:0050793	Regulation of developmental process	Cyst vs. Polyp	1.04e^-12	2.19e^-14	557	897
			Cyst vs. Both	2.24e^-09	2.34e^-11	557	1492

Table 19: GO enrichment categories associated with development.

3.5 Candidate genes of interest

The assembled transcriptomes were screened for the presence or absence of candidate genes associated with transdifferentiation, induction of pluripotency, and aging/longevity. The functions and significance of the candidate genes of interest are compiled in APPENDIX B. tBlastx was used to screen for potential orthologs of the candidate genes of interest within each of the polyp, jellyfish and cyst transcriptomes. tBlastx compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database, and its main purpose is to find distant relationships between the compared sequences. An e-value of $\leq 1e^{-8}$ was utilized for the screening.

The lowest e-value, highest bit-score, and highest aligning variant, if applicable, were recorded for all candidate genes with matches. If no variant type was listed for a particular gene, outputs were the same for all variants or the gene encompasses only one variant. Although a low e-value is a key indicator of a significant alignment, the parameter is dependent on the number of sequences in the reference database. On the other hand, the highest bit score represents the best alignment and is not dependent on the database size, providing a constant statistical measure of alignment.

Induction of Pluripotency

The induction of a pluripotent state is a fundamental step in the process of cellular transdifferentiation, permitting the reprogramming of a specialized cell into a new cell

type. Candidate genes associated with the induction of pluripotency were screened in the *T. dohrnii* transcriptomes. The candidate gene's GenBank ID, full gene name, species, and number of variants are listed in Table 20. This list is centered on genes reported to associate with the induction of iPSC and/or play a critical role in maintaining the self-renewing, pluripotent state of embryonic stem (ES) cells.

#	Gene ID	mRNA Symbol	Gene Name	Species	Variants
1	5460	Oct3/4 (POU5F1)	POU class 5 homeobox 1	Homo sapiens	5
2	6657	Sox-2	SRY-box 2	Homo sapiens	1
3	4609	c-Myc	v-myc avian myelocytomatosis viral oncogene homolog	Homo sapiens	1
4	9314	Klf4	Kruppel like factor 4	Homo sapiens	2
5	79923	Nanog	Nanog homeobox	Homo sapiens	2
6	79727	Lin28	lin-28 homolog A	Homo sapiens	1
7	154288	Ecat1 (KHFC3L)	KH domain containing 3 like, subcortical maternal complex	Homo sapiens	1
8	340168	Dppa5	developmental pluripotency associated 5	Homo sapiens	1
9	201456	Fbxo15	F-box protein 15	Homo sapiens	2
10	3266	ERas	ES cell expressed Ras	Homo sapiens	1
11	29947	Dnmt3l	DNA methyltransferase 3 like	Homo sapiens	2
12	91646	Ecat8 (TDRD12)	tudor domain containing 12	Homo sapiens	1
13	9573	Gdf3	growth differentiation factor 3	Homo sapiens	1
14	6665	Sox15	SRY-box 15	Homo sapiens	1
15	55211	Dppa4	developmental pluripotency associated 4	Homo sapiens	1
16	151871	Dppa2	developmental pluripotency associated 2	Homo sapiens	1
17	83457	Fthl17	ferritin, heavy polypeptide-like 17, member E	Mus musculus	1
18	57167	Sall4	spalt like transcription factor 4	Homo sapiens	2
19	132625	Rex1 (Zfp42)	ZFP42 zinc finger protein	Homo sapiens	2
20	8433	Utf1	undifferentiated embryonic cell transcription factor 1	Homo sapiens	1
21	21432	Tcl1	T cell lymphoma breakpoint 1	Mus musculus	4
22	359787	Dppa3 (Stella)	developmental pluripotency associated 3	Homo sapiens	1
23	1499	β-catenin	catenin beta 1	Homo sapiens	4
24	6774	Stat3	signal transducer and activator of transcription 3	Homo sapiens	3
25	2885	Grb2	growth factor receptor bound protein 2	Homo sapiens	2

Table 20: Candidate genes for the induction of pluripotency.

Yamanaka factors: Oct4, Sox2, c-Myc and Klf4

All human variants of the Yamanaka factors (Oct4, Sox2, c-Myc, Klf4) were downloaded from NCBI and used in the screening. Details regarding the functions of each gene can be found in the introduction (Table 1 or APPENDIX B). All four transcription factors had significant matches against all three stages, indicating sequences with significant similarity in its amino acid sequence to the queried transcripts (Table 21). This is surprising because it differs from what was reported for the *Hydra* genome (Chapman et al., 2010). In *Hydra*, 4 Myc homologues and the Sox B group, which includes Sox2, were present, while Oct4 and Klf4 were absent (Chapman et al., 2010). This led to the conclusion that the genetic network of stem cells in hydrozoans has an independent evolutionary origin compared to mammals, governed by the Yamanaka factors (Chapman et al., 2010; Steele et al., 2012). Our results suggest the contrary, in which the genetic network of stem cells and pluripotency induction in mammals is shared with early metazoans such as Hydrozoa. This has a profound evolutionary implication. It can also identify *T. dohrnii* as a potential laboratory model system to understand the expression patterns of the Yamanaka factors *in vivo* rather than *in vitro*.

Downstream analyses to confirm that these matches contain similar domains and are true orthologs will be conducted to verify whether these genes truly are present in *T. dohrnii*. In addition, a differential gene expression analysis of the Yamanaka factors among the three stages will provide further insight on whether these transcripts play a role in *T. dohrnii* transdifferentiation.

				e<=1e^-8		
#	**Genbank Accession #**	**mRNA Symbol**	**Species**	**Polyps**	**Jellyfish**	**Cyst**
1	NM_002701.5, NM_203289.5, NM_001173531.2, NM_001285986.1, NM_001285987.1	Oct4	H. sapiens	e-value: **1.07e^-47** Bit score: **127.99** (Variant 1)	e-value: **3.11e^-32** Bit score: **127.99** (Variant 2/5)	e-value: **4.78e^-39** Bit score: **149.53** (Variant 2/5)
2	NM_003106.3	Sox2	H. sapiens	e-value: **4.85e^-42** Bit score: **171.98**	e-value: **9.72e^-43** Bit score: **174.27**	e-value: **6.63e^-42** Bit score: **171.98**
3	NM_002467.4	c-Myc	H. sapiens	e-value: **1.64e^-21** Bit score: **103.71**	e-value: **1.61e^-21** Bit score: **103.71**	e-value: **2.23e^-21** Bit score: **103.71**
4	NM_001314052.1, NM_004235.5	Klf4	H. sapiens	e-value: **1.64e^-21** Bit score: **187.10** (Variant 2)	e-value: **1.55e^-46** Bit score: **187.10** (Variant 2)	e-value: **2.15e^-46** Bit score: **187.10** (Variant 2)

Table 21: tBlastx screening of Yamanaka factors.

Thompson factors: Oct4, Sox2, Lin28 and Nanog

In addition to the Yamanaka factors, Lin28 and Nanog were screened for in all three transcriptomes. Both transcription factors have significant hits in all three stages (Table 22). Nanog is reported to be absent in the genomes of *H. vulgaris* (Chapman et al., 2010), as well as *Nematostella vectensis* (Steele et al., 2012), another prominent hydrozoan model system. Similar to the Yamanaka factor screening, our results imply the contrary to current belief that hydrozoans possess an independent origin of gene networks that underlie pluripotency, and instead share similar origins of networks. Further analyses to confirm our results will be conducted.

					e<=1e^-8		
#	Genbank Accession #	mRNA Symbol	Species		Polyps	Jellyfish	Cyst
5	NM_024865.3, NM_001297698.1	Nanog	H. sapiens		e-value: **1.93e^-13** Bit score: **76.67** (Variant 2)	e-value: **1.99e^-18** Bit score: **76.67** (Variant 2)	e-value: **3.20e^-29** Bit score: **122.49** (Variant 2)
6	NM_024674.5	Lin28	H. sapiens		e-value: **6.18e^-15** Bit score: **68.42**	e-value: **4.58e^-19** Bit score: **104.62**	e-value: **5.15e^-33** Bit score: **105.54**

Table 22: tBlastx screening of Thompson factors.

Additional factors tested by the Yamanaka Group

In order to isolate the four core transcription factors that induce pluripotency, 24 transcription factors involved in the maintenance of the ES cell state were initially selected by the Yamanka Group for the assay to detect the pluripotency (Takahashi and Yamanaka, 2006). The 24 genes include the Yamanaka factors, Nanog, and 19 other candidate transcripts. The human orthologs of these candidate genes were used as the query for screening. If unavailable, the mouse ortholog was utilized.

11 out of 19 transcripts had significant hits in the polyp transcriptome, 12 out of 19 transcripts in the jellyfish, and 11 out of 19 in the cyst (Table 23). Among the screened transcripts, β-catenin stood out to have an e-value of 0 and a very high bit score of 823.09 among all stages, highly implying its presence in *T. dohrnii*. In addition, Rex1 had a very low e-value and bit score around 250 in all stages, and Grb2 exhibited a low e-score around 250 in the polyp and jellyfish stage. In future research, it will be interesting to further investigate whether these genes are true orthologs and quantify their expression among the stages to explore whether they may or may not be involved in *T. dohrnii*'s ability to undergo reverse development.

#	Genbank Accession #	mRNA Symbol	Species	Polyps (e<=1e^-8)	Jellyfish (e<=1e^-8)	Cyst (e<=1e^-8)
7	NM_001017361.2	Ecat1	H. sapiens	Absent	Absent	Absent
8	NM_001025290.2	Dppa5 (Esg1)	H. sapiens	Absent	Absent	Absent
9	NM_001142958.1, NM_152676.2	Fbxo15	H. sapiens	e-value: **4.80e^-15** Bit score: **42.76** (Variant 1)	e-value: **5.72^-19** Bit score: **42.76**	e-value: **1.47^-11** Bit score: **35.43** (Variant 1)
10	NM_181532.3	ERas	H. sapiens	e-value: **1.74e^-24** Bit score: **87.67**	e-value: **4.33e^-24** Bit score: **80.8**	e-value: **5.93e^-24** Bit score: **80.8**
11	NM_013369.3, NM_175867.2	Dnmt3l	H. sapiens	e-value: **4.24e^-36** Bit score: **132.12**	e-value: **7.03e^-20** Bit score: **97.75** (Variant 2)	Absent
12	NM_001110822.1	Ecat8	H. sapiens	Absent	Absent	Absent
13	NM_020634.2	Gdf3	H. sapiens	e-value: **3.13e^-28** Bit score: **89.96**	e-value: **3.11e^-28** Bit score: **89.96**	e-value: **4.29e^-28** Bit score: **89.96**
14	NM_006942.1	Sox15	H. sapiens	e-value: **1.08.e^-27** Bit score: **123.41**	e-value: **1.06e^-27** Bit score: **123.41**	e-value: **1.46e^-27** Bit score: **123.41**
15	NM_018189.3	Dppa4	H. sapiens	Absent	e-value: **7.55e^-18** Bit score: **53.3**	e-value: **2.32e^-29** Bit score: **117.91**
16	NM_138815.3	Dppa2	H. sapiens	Absent	Absent	Absent
17	NM_031261.2	Fthl17	M. musculus	e-value: **3.62e^-41** Bit score: **167.4**	e-value: **6.68e^-37** Bit score: **153.19**	e-value: **2.82e^-30** Bit score: **131.66**
18	NM_020436.4, NM_001318031.1	Sall4	H. sapiens	e-value: **1.40e^-36** Bit score: **80.34** (Variant 1)	e-value: **3.30e^-39** Bit score: **80.34** (Variant 1)	e-value: **1.05e^-41** Bit score: **125.7** (Variant 1)
19	NM_174900.4, NM_001304358.1	Rex1 (Zfp42)	H. sapiens	e-value: **1.58e^-65** Bit score: **246.67** (Variant 2)	e-value: **1.55e^-64** Bit score: **246.67** (Variant 2)	e-value: **2.15e^-64** Bit score: **246.67** (Variant 2)
20	NM_003577.2	Utf1	H. sapiens	Absent	Absent	Absent
21	NM_009337.3, NM_001289468.1, NM_001309485.1, NM_001309484.1	Tcl1	M. musculus	Absent	Absent	Absent
22	NM_199286.3	Dppa3 (Stella)	H. sapiens	Absent	Absent	Absent
23	NM_001904.3, NM_001098209.1, NM_001098210.1, NM_001330729.1	β-catenin	H. sapiens	e-value: **0** Bit score: **823.09**	e-value: **0** bit score: **823.09**	e-value: **0** bit score: **823.09**
24	NM_139276.2, NM_003150.3, NM_213662.1	Stat3	H. sapiens	e-value: **6.92e^-18** Bit score: **41.39** (Variant 3)	e-value: **6.05e^-24** Bit score: **60.18** (Variant 3)	e-value: **1.84e^-31** Bit score: **104.62** (Variant 2)
25	NM_002086.4, NM_203506.2	Grb2	H. sapiens	e-value: **3.12e^-83** Bit score: **254.46** (Variant 1)	e-value: **3.06e^-83** Bit score: **254.46** (Variant 1)	e-value: **5.52e^-36** Bit score: **106.91** (Variant 2)

Table 23: tBlastx screening of additional factors tested by the Yamanaka group.

Aging/Longevity

Transgenic organisms and RNA interference experiments have uncovered genes that directly influence the progression of the aging process and phenotype. Contributing genes and networks that influence the progression of aging includes telomerase activity (Whitaker et al., 1995; Nowak et al., 2006; Blasco, 2007), sirtuin proteins (Dali-Youcef et al., 2007; Someya et al., 2010; Park et al., 2013), heat shock proteins (Gong et al. 2009; Benoit et al., 2010), and the reactivation of oncogenes (Newbold and Overell, 1983; Garbe et al., 1999; Bringold and Serrano, 2000). The candidate genes for aging and longevity, GenBank ID, full gene name, species, and number of variants are listed in Table 24.

#	Genbank Gene ID	mRNA Symbol	Gene Name	Species	Transcript Variants
1	7015	TERT	Telomerase reverse transcriptase	*H. sapiens*	2
2	7012	TERC	Telomerase RNA component	*H. sapiens*	1
3	1736	DKC1	Dyskerin pseudouridine synthase 1	*H. sapiens*	3
4	7011	TEP1	Telomerase associated protein 1	*H. sapiens*	2
5	7013	TRF1	Telomeric repeat binding factor 1	*H. sapiens*	2
6	7014	TRF2	Telomeric repeat binding factor 2	*H. sapiens*	1
7	26277	TINF2	TERF1 interacting nuclear factor 2	*H. sapiens*	2
8	25913	POT2	Protection of telomeres 1	*H. sapiens*	2
9	5906	RAP1	Repressor/activator protein 1	*H. sapiens*	3
10	65057	PTOP	POT1-TIN2 organizer protein	*H. sapiens*	3
11	2542313	Taz1	Telomere length regulator	*S. pombe*	1
12	2540115	Rap1	Telomere binding protein	*S. pombe*	1
13	2542244	Poz1	Pot1 associated protein	*S. pombe*	1
14	2542414	Tpz1	Telomere-protecting protein	*S. pombe*	1
15	2542703	Pot1	Telomere end-binding protein	*S. pombe*	1
16	2539352	Ccq1	Telomere maintenance protein	*S. pombe*	1
17	851520	SIR2	NAD-dependent histone deacetylase SIR2	*Saccharomyces cerevisiae*	1
18	177924	Sir-2.1	NAD-dependent protein deacetylase sir-2.1	*C. elegans*	2
19	34708	dSir2	NAD-dependent protein deacetylase dSir2	*D. melanogaster*	1
20	23411	SIRT1	Sirtuin 1	*H. sapiens*	3
21	22933	SIRT2	Sirtuin 2	*H. sapiens*	3
22	23410	SIRT3	Sirtuin 3	*H. sapiens*	2
23	23409	SIRT4	Sirtuin 4	*H. sapiens*	1
24	23408	SIRT5	Sirtuin 5	*H. sapiens*	4
25	51548	SIRT6	Sirtuin 6	*H. sapiens*	9
26	51547	SIRT7	Sirtuin 7	*H. sapiens*	1
27	173078	HSF-1	Heat shock factor 1	*C. elegans*	1
28	3297	HSF-1	Heat shock factor 1	*H. sapiens*	1
29	175410	Daf-2	Insulin-like receptor subunit beta	*C. elegans*	4
30	172981	Daf-16	Forkhead box protein O	*C. elegans*	8
31	176608	MRT-2	MoRTal germline	*C. elegans*	1

Table 24: Candidate genes for longevity and aging.

Telomerase Activity

Telomerase subunits

Telomerase is a specialized DNA polymerase that elongates the protective ends of DNA molecules, called telomeres, to inhibit the loss of genetic information during DNA replication. Telomerase maintains the telomere length across a large range of both invertebrate and vertebrate species, such as sea stars, crustaceans, fish and mammals, and is involved in aging and/or regenerative processes (Elmore et al., 2008). The telomere motif $(TTAGGN)_n$ is present in all animal lineages, from basal metazoans to mammals,

except for phylum Arthropoda and Nematoda (Traut et al., 2007; Gomes et al., 2010). The majority of metazoan somatic cells do not express telomerase, but the enzyme is continuously active in immortal germline cells and cancerous cells. Telomerase has been detected via telomerase repeat amplification protocol (TRAP) assays in cnidarian systems such as the moon jellyfish and ctenophores, but not in *Hydra* or *Nematostella* (Traut et al., 2007).

The human orthologs of the telomerase subunits, TERT, TERC, DKC1 and TEP1, were screened in all transcriptomes. TERT is the catalytic subunit of telomerase and works with TERC to maintain the length of telomeres, forming the most fundamental parts of the enzyme. All subunits have significant matches against the queried factors (Table 25). DKC1, the Dsykerin component, in particular exhibited an e-value of 0 and a very high bit score (714.5) in all stages, thus have very high potential of being true orthologs. Further analyses that detect telomerase, such as a TRAP detection assay, on the different stages of *T. dohrnii* would provide insight on whether telomerase is active among the stages and whether it may play a role in its reverse development.

#	Genbank Accession #	mRNA Symbol	Species	e<=1e^-8 Polyps	Jellyfish	Cyst
1	NM_198253.2, NM_001193376.1	TERT	*H. sapiens*	e-value: **2.50e^-31** Bit score: **69.34** (Variant 1)	e-value: **7.63e^-46** Bit score: **69.34** (Variant 1)	e-value: **1.48e^-43** Bit score: **69.8** (Variant 1)
2	NR_001566.1	TERC	*H. sapiens*	e-value: **1.42e^-113** Bit score: **221.92**	e-value: **1.75e^-113** Bit score: **216.88**	e-value: **1.94e^-113** Bit score: **216.88**
3	NM_001363.4, NM_001142463.2, NM_001288747.1	DKC1	*H. sapiens*	e-value: **0** Bit score: **714.5**	e-value: **0** Bit score: **714.5**	e-value: **0** Bit score: **714.5**
4	NM_007110.4, NM_001319035.1	TEP1	*H. sapiens*	e-value: **1.19e^-65** Bit score: **193.97** (Variant 2)	e-value: **4.20e^-70** Bit score: **135.32** (Variant 2)	e-value: **5.18e^-36** Bit score: **132.12** (Variant 2)

Table 25: tBlastx screening of telomerase subunits.

Shelterin

The shelterin complex's main function is to protect telomeres from DNA repair and maintain telomerase activity (Xin et al., 2008; Martínez and Blasco, 2010). By binding to the telomere motif (TTAGGN)$_n$ and forming a loop-like cap structure, the complex prevents DNA repairing machinery to access telomeres. Without the shelterin complex, cells will often undergo senescence or apoptosis. The shelterin complex has been found surprisingly found in yeast (Martinez et al., 2016), but absent in fruit flies, which also lacks telomerase (Raffa et al., 2013). Whether basal metazoans possess similar shelterin complex as vertebrates remains to be unknown. The human shelterin subunits, TRF1, TRF2, TINF2, POT2, RAP1, and PTOP were screened for in each of the transcriptomes. Many subunits were missing from some or all of the stages (Table 26), implying that the mammalian shelterin complex is absent in *T. dohrnii*.

#	Genbank Accession #	mRNA Symbol	Species	e<=1e^-8		
				Polyps	**Jellyfish**	**Cyst**
5	NM_017489.2, NM_003218.3	TRF1	*H. sapiens*	Absent	e-value: **3.92e^-22** Bit score: **89.5** (Variant 2)	e-value: **1.57e^-34** Bit score: **147.69** (Variant 2)
6	NM_005652.4	TRF2	*H. sapiens*	Absent	e-value: **7.99e^-25** Bit score: **58.34**	e-value: **7.95e^-34** Bit score: **134.41**
7	NM_001099274.1, NM_012461.2	TINF2	*H. sapiens*	Absent	Absent	Absent
8	NM_015450.2. NM_001042594.1	POT2	*H. sapiens*	e-value: **4.25e^-12** Bit score: **37.72** (Variant 1)	Absent	Absent
9	NM_001010935.2, NM_002884.3, NM_001291896.1	RAP1	*H. sapiens*	e-value: **1.87e^-89** Bit score: **330.52** (Variant 2)	e-value: **1.69e^-88** Bit score: **327.31** (Variant 2)	e-value: **2.35e^-88** Bit score: **327.31** (Variant 2)
10	NM_001082486.1, NM_022914.2, NM_001082487.1	PTOP	*H. sapiens*	Absent	Absent	Absent

Table 26: tBlastx screening of human shelterin subunits.

Interestingly, the only subunit with potential orthologs consistently among the three stages was RAP1, a member of an oncogene family that plays a role in cellular proliferation. RAP1 is an essential part of the shelterin complex that is evolutionarily conserved from the yeast to humans (Martinez et al., 2016). To investigate whether *T. dohrnii* may potentially possess a shelterin complex more similar to fission yeast (*Schizosaccharomyces pombe*), the six yeast shelterin subunits Taz1, Rap1, Poz1, Tpz1, Pot1, and Ccq1 were screened for in all three stages. It should been noted that the mRNA sequences of the yeast shelterin subunits are considered provisional in the NCBI Reference Sequences (RefSeq) database.

The only yeast shelterin subunit with matches was Tpz1among all three stages, implying that a similar multiprotein complex is absent in *T. dohrnii* (Table 27). The yeast Rap1 does not seem to be present in any of the stages. To check whether any of the yeast

subunits may have been the same tBlastx hits as humans (Table 26), the accession numbers of hits with the highest bit score were compared in each stage (data not shown). None of the accession values overlapped between the human and yeast subunits, implying no potential orthologs between the two groups were found. Overall, more potential orthologs of subunits were found in the human shelterin as compared to the yeast. Further investigation on whether these matches are true orthologs would potentially contribute to the characterization of the shelterin or shelterin-like complex, in *T. dohrnii*.

				e<=1e^-8		
#	Genbank Accession #	mRNA Symbol	Species	Polyps	Jellyfish	Cyst
11	NM_001019472.2	Taz1	*S. pombe*	Absent	Absent	Absent
12	NM_001022207.2	Rap1	*S. pombe*	Absent	Absent	Absent
13	NM_001019857.2	Poz1	*S. pombe*	Absent	Absent	Absent
14	NM_001019338.3	Tpz1	*S. pombe*	e-value: **6.34e^-32** Bit score: **71.63**	e-value: **6.31e^-32** Bit score: **71.63**	e-value: **8.74e^-32** Bit score: **71.63**
15	NM_001019882.2	Pot1	*S. pombe*	Absent	Absent	Absent
16	NM_001023200.2	Ccq1	*S. pombe*	Absent	Absent	Absent

Table 27: tBlastx screening of fission yeast shelterin subunits.

Sirtuins

The overexpression of the Sir2 (silent information regulator 2) gene has been found to extend life span in brewer's yeast (Kaeberlein et al., 1999), nematodes (Tissenbaum and Guarente, 2001) and fruit flies (Rogina and Helfand, 2004). This discovery has led to a series of subsequent studies on Sir2 related genes in murine and human system, in which seven mammalian SIRT proteins have been identified. Among

the SIRT proteins, the mammalian Sir2 ortholog SIRT1, has been extensively investigated on whether the protein have similar functions in promoting longevity as Sir2. Though SIRT1 did not increase lifespan like the invertebrate counterparts, it plays a role in stem cell functionality and took part in improving cellular function through enhancements such as stabilizing genomic DNA and increasing metabolic efficiency (Herranz et al., 2010; Park et al., 2013). SIRT6 however, has been found to extend the lifespan in mammals (Park et al., 2013) and its deletion has been found to induce premature defects among test subjects (Dalli-Youcef et al., 2007).

The Sir2 gene from yeast, nematodes (Sir-2.1) and fruit flies (dSir2), and the seven mammalian SIRTs were screened for in the transcriptomes. Due to the high potential of the SIRT genes exhibiting similar sequence composition and structure, the accession name of the top hit sequence of the highest bit score was also used to evaluate the results.

All four Sir2 orthologs, Sir2. Sir-2.1, dSir2, SIRT1, matched the same accession number, which was expected (Table 28). Interestingly, among the vertebrate and invertebrate Sir2 orthologs, the mammalian SIRT1 gene had the highest alignment score throughout all of the *T. dohrnii* stages (Table 29). Each of the stages have 3-4 other potential SIRT orthologs (Table 28). When the SIRT genes that matched the same accession name were compared, SIRT6 in the polyp and cyst stage in particular exhibited alignment scores (Table 29). Downstream analyses to further investigate whether Sirtuin proteins orthologs are truly expressed and whether these genes are differential expression among the *T. dohrnii* stages will provide more insight on their potential involvement in its ability to undergo life cycle reversal.

mRNA Symbol	Polyp	Jellyfish	Cyst
Sir2	37701_g1_i1	17367_c0_g1_i1	9731_c0_g1_i1
Sir-2.1			
dSir2			
SIRT1			
SIRT2	35468_g1_i1	22923_c0_g1_i1	15891_c0_g1_i1
SIRT3			
SIRT4	41091_g1_i1	10771_c1_g1_i1	14866_c1_g1_i2
SIRT5		1036_c0_g1_i1	11979_c1_g1_i1
SIRT6	33775_g1_i1	13219_c2_g1_i1	26844_c0_g1_i1
SIRT7			

Table 28: Accession name of sirtuin hits with highest bit scores.

				e<=1e^-8		
#	Genbank Accession #	mRNA Symbol	Species	Polyps	Jellyfish	Cyst
17	NM_001180101.1	Sir2	*S. cerevisiae*	e-value: **6.13e^-59** Bit score: **115.16**	e-value: **6.05e^-59** Bit score: **115.16**	e-value: **8.45e^-59** Bit score: **115.16**
18	NM_001268555.1, NM_001268556.1	Sir-2.1	*C. elegans*	e-value: **7.87e^-101** Bit score: **251.25** (Variant 2)	e-value: **7.71e^-101** Bit score: **251.25** (Variant 2)	e-value: **1.08e^-100** Bit score: **251.25** (Variant 2)
19	NM_058003.4	dSir2	*D. melanogaster*	e-value: **8.11e^-110** Bit score: **295.69**	e-value: **7.95e^-110** Bit score: **295.69**	e-value: **1.53e^-109** Bit score: **295.69**
20	NM_012238.4, NM_001142498.1, NM_001314049.1	SIRT1	*H. sapiens*	e-value: **1.46e^-133** Bit score: **347.01** (Variant 1)	e-value: **1.15e^-133** Bit score: **347.01** (Variant 1)	e-value: **3.14e^-131** Bit score: **345.18** (Variant 1)
21	NM_012237.3, NM_030593.2, NM_001193286.1	SIRT2	*H. sapiens*	e-value: **1.55e^-77** Bit score: **195.35** (Variant 1)	e-value: **1.55e^-77** Bit score: **195.35** (Variant 1)	e-value: **4.52e^-65** Bit score: **191.22** (Variant 2)
22	NM_012239.5, NM_001017524.2	SIRT3	*H. sapiens*	e-value: **3.16e^-78** Bit score: **208.18** (Variant 2)	e-value: **8.83e^-79** Bit score: **259.04** (Variant 2)	e-value: **2.84e^-77** Bit score: **205.43** (Variant 2)
23	NM_012240.2	SIRT4	*H. sapiens*	e-value: **8.84e^-92** Bit score: **205.89**	e-value: **1.99e^-45** Bit score: **182.06**	e-value: 8.55**e^-11** Bit score: **42.76**
24	NM_012241.4, NM_031244.3, NM_001193267.2, NM_001242827.1	SIRT5	*H. sapiens*	e-value: **1.70e^-17** Bit score: **55.59** (Variant 1)	e-value: **1.94e^-26** Bit score: **84.92** (Variant 4)	e-value: **1.17e^-33** Bit score: **128.91** (Variant 4)
25	NM_016539.3, NM_001193285.2, NM_001321058.1, NM_001321059.1, NM_001321060.1, NM_001321061.1, NM_001321062.1, NM_001321063.1, NM_001321064.1	SIRT6	*H. sapiens*	e-value: **9.34e^-107** Bit score: **382.3** (Variant 1)	e-value: **1.76e^-36** Bit score: **136.24** (Variant 1)	e-value: **6.82e^-84** Bit score: **310.82** (Variant 1)
26	NM_016538.2	SIRT7	*H. sapiens*	e-value: **1.32e^-40** Bit score: **92.25**	e-value: **9.12e^-15** Bit score: **73.46**	e-value: 7**.12e^-31** Bit score: **91.33**

Table 29: tBlastx screening of sirtuin genes.

Other aging/longevity genes

Other genes that have been reported to influence longevity and aging, HSF-1, Daf-2, Daf-16 and MRT2 were screened in *T. dohrnii*. Both the nematode and mammalian HSF-1 had top matches with the same accession name (data not shown), and similar to the Sir2 orthologs, the mammalian query had a higher alignment score (Table 30). Among the queried genes, Daf-2 in all stages and Daf-16 in the polyp and jellyfish stage are among high interest for downstream analyses to determine whether these matches are true orthologs.

				e<=1e^-8		
#	**Genbank Accession #**	**mRNA Symbol**	**Species**	**Polyps**	**Jellyfish**	**Cyst**
27	NM_060630.3	HSF-1	*C. elegans*	e-value: **8.48e^-26** Bit score: **103.9**	e-value: **8.45e^-26** Bit score: **103.9**	e-value: **8.45e^-25** Bit score: **103.9**
28	NM_005526.3	HSF-1	*H. sapiens*	e-Value: **3.88e^-42** Bit score: **115.1**	e-value: **5.47e^-25** Bit score: **115.1**	e-value: **3.21e^-37** Bit score: **115.1**
29	NM_065249.4, NM_001312987.1, NM_001312988.1, NM_001312989.1	Daf-2	*C. elegans*	e-value: **6.75e^-77** Bit score: **182.52** (Variant 4)	e-value: **1.85e^-68** Bit score: **182.52** (Variant 4)	e-value: **3.04e^-42** Bit score: **144.94** (Variant 4)
30	NM_001264561.1, NM_001026427.4, NM_001026425.3, NM_001026424.4, NM_001026423.4, NM_001026422.4, NM_001264563.1, NM_001026426.2	Daf-16	*C. elegans*	e-value: **1.18e^-51** Bit score: **203.24** (Variant 1)	e-value: **1.16e^-51** Bit score: **203.24** (Variant 1)	e-value: **1.20e^-18** Bit score: **94.03** (Variant 1)
31	NM_067120.4	MRT-2	*C. elegans*	e-value: **1.91e^-12** Bit score: **54.68**	e-value: **1.86e^-12** Bit score: **54.68**	e-value: **2.55e^-12** Bit score: **54.68**

Table 30: tBlastx screening of other aging and longevity genes.

4. CONCLUSION

The polyp, jellyfish and cyst stage of *Turritopsis dohrnii* were sequenced through RNA-sequencing, assembled via a *de novo* approach, and assigned GO functional annotation to create a gene expression profile of each stage. Comparative functional gene enrichment analyses were conducted using the cyst as the central stage of our comparison to identify genetic networks and pathways that are potentially involved in the reverse development and transdifferentiation in *T. dohrnii*. Our results suggest that DNA synthesis and metabolic processes, telomerase activity, and chitin related enrichment categories are over-expressed in the cyst in comparison to the polyp and jellyfish stage. On the other hand, gene enrichment categories involving lifespan and aging, cellular characterization, division, differentiation and development, and response to stimuli are under-expressed in the cyst in comparison to the polyp and jellyfish stage. Our results confirm our starting hypothesis that biological processes associated with longevity, such as telomerase activity, were over-expressed, while there was a variety of suppressed processes associated to cellular differentiation and development when compared to the other two stages. The data is consistent with the hypothesis where the cyst is actively preparing to re-differentiate and revert back into the juvenile polyp stage while repressing a variety of biological processes that associate to the characterization of an organism. In addition, potential orthologs of candidate genes involved in the induction of pluripotency and aging were discovered among all stages. Further studies are necessary to determine whether these transcripts within the over- and under-expressed categories and potential

orthologs of candidate genes are directly involved in the life-cycle reversal and transdifferentiation in *T. dohrnii.*

REFERENCES

Ahmed, S., & Hodgkin, J. (2000). MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. Nature, 403(6766), 159-164.

Al-Shahrour, F., Díaz-Uriarte, R., & Dopazo, J. (2004). FatiGO: a web tool for finding significant associations of gene ontology terms with groups of genes. Bioinformatics, 20(4), 578-580.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Augusto-Pinto, L., Da Silva, C. G., Lopes, D. D. O., Machado-Silva, A., & Machado, C. R. (2003). *Escherichia coli* as a model system to study DNA repair genes of eukaryotic organisms. Genetics and Molecular Research, 2(1), 77-91.

Austad S. N. (2009). Is there a role for new invertebrate models for aging research? Journal of Gerontology, 64A(2), 192-194.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & Development, 17(1), 126-140.

Baker, M. (2012). De novo genome assembly: what every biologist should know. Nature Methods, 9(4), 333-337.

Bavestrello, G., Cerrano, C., Di Camillo, C., Puce, S., Romagnoli, T., Tazioli, S., & Totti, C. (2008). The ecology of protists epibiontic on marine hydroids. Journal of the Marine Biological Association of the UK, 88(08), 1611-1617.

Bechter, O. E., Eisterer, W., Pall, G., Hilbe, W., Kühr, T., & Thaler, J. (1998). Telomere length and telomerase activity predict survival in patients with B cell chronic lymphocytic leukemia. Cancer Research, 58(21), 4918-4922.

Bely, A. E., & Nyberg, K. G. (2010). Evolution of animal regeneration: re-emergence of a field. Trends in Ecology & Evolution, 25(3), 161-170.

Benoit, J. B., Lopez-Martinez, G., Phillips, Z. P., Patrick, K. R., & Denlinger, D. L. (2010). Heat shock proteins contribute to mosquito dehydration tolerance. Journal of Insect Physiology, 56(2), 151-156.

Bignell, G., Micklem, G., Stratton, M. R., Ashworth, A., & Wooster, R. (1997). The BRC repeats are conserved in mammalian BRCA2 proteins. Human Molecular Genetics, 6(1), 53-58.

Birnbaum, K. D., & Sánchez-Alvarado, A. (2008). Slicing across kingdoms: regeneration in plants and animals. Cell, 132(4), 697-710.

Blagosklonny, M. V. (2010). Revisiting the antagonistic pleiotropy theory of aging: TOR-driven program and quasi-program. Cell Cycle, 9(16), 3171-3176.

Blasco, M. A. (2005). Telomeres and human disease: ageing, cancer and beyond. Nature Reviews Genetics, 6(8), 611-622.

Blasco, M. A. (2007). Telomere length, stem cells and aging. Nature Chemical Biology, 3(10), 640-649.

Bode, H. R. (1996). The interstitial cell lineage of *Hydra*: a stem cell system that arose early in evolution. Journal of Cell Science, 109(6), 1155-1164.

Bosch, T. C. (2012). What *Hydra* has to say about the role and origin of symbiotic interactions. The Biological Bulletin, 223(1), 78-84.

Bosch, T. C., & David, C. N. (1984). Growth regulation in *Hydra*: relationship between epithelial cell cycle length and growth rate. Developmental Biology, 104(1), 161-171.

Botstein, D., Chervitz, S. A., & Cherry, J. M. (1997). Yeast as a model organism. Science, 277(5330), 1259.

Boumahdi, S., Driessens, G., Lapouge, G., Rorive, S., Nassar, D., Le Mercier, M., Delatte B., Caauwe, A., Lenglez, S., Nkusi, E. & Brohée, S. (2014). SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. Nature, 511(7508), 246-250.

Bowles, J., Teasdale, R. P., James, K., & Koopman, P. (2003). Dppa3 is a marker of pluripotency and has a human homologue that is expressed in germ cell tumours. Cytogenetic and Genome Research, 101(3-4), 261-265.

Bringold, F., & Serrano, M. (2000). Tumor suppressors and oncogenes in cellular senescence☆. Experimental Gerontology, 35(3), 317-329.

Brockes, J. P. (1997). Amphibian limb regeneration: rebuilding a complex structure. Science, 276(5309), 81-87.

Brockes, J. P., & Kumar, A. (2002). Plasticity and reprogramming of differentiated cells in amphibian regeneration. Nature Reviews Molecular Cell Biology, 3(8), 566-574.

Brockes, J. P., & Kumar, A. (2005). Appendage regeneration in adult vertebrates and implications for regenerative medicine. Science, 310(5756), 1919-1923.

Brockes, J. P., & Kumar, A. (2008). Comparative aspects of animal regeneration. Annual Review of Cell and Developmental Biology, 24, 525-549.

Bromberg, J. F., Wrzeszczynska, M. H., Devgan, G., Zhao, Y., Pestell, R. G., Albanese, C., & Darnell, J. E. (1999). Stat3 as an oncogene. Cell, 98(3), 295-303.

Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N. & Davey, M. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature, 433(7025), 531-537.

Calvanese, V., Lara, E., Suárez-Álvarez, B., Dawud, R. A., Vázquez-Chantada, M., Martínez-Chantar, M. L., Embade, N., López-Nieva, P., Horrillo, A., Hmadcha, A. & Soria, B. (2010). Sirtuin 1 regulation of developmental genes during differentiation of stem cells. Proceedings of the National Academy of Sciences, 107(31), 13736-13741.

Campisi, J. (2013). Aging, cellular senescence, and cancer. Annual Review of Physiology, 75, 685.

Cary, A. J., Che, P., & Howell, S. H. (2002). Developmental events and shoot apical meristem gene expression patterns during shoot development in *Arabidopsis thaliana*. The Plant Journal, 32(6), 867-877.

Chaisson, M. J., Wilson, R. K., & Eichler, E. E. (2015). Genetic variation and the de novo assembly of human genomes. Nature Reviews Genetics, 16(11), 627-640.

Chambers, I., & Tomlinson, S. R. (2009). The transcriptional foundation of pluripotency. Development, 136(14), 2311-2322.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., & Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell, 113(5), 643-655.

Chang, H. M., Triboulet, R., Thornton, J. E., & Gregory, R. I. (2013). A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature, 497(7448), 244-248.

Chang, J. T. C., Chen, Y. L., Yang, H. T., Chen, C. Y., & Cheng, A. J. (2002). Differential regulation of telomerase activity by six telomerase subunits. The FEBS Journal, 269(14), 3442-3450.

Chapman, J. A., Kirkness, E. F., Simakov, O., Hampson, S. E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D. & Disbennett, K. (2010). The dynamic genome of *Hydra*. Nature, 464(7288), 592-596.

Chen, C. R., Kang, Y., & Massagué, J. (2001). Defective repression of c-myc in breast cancer cells: a loss at the core of the transforming growth factor β growth arrest program. Proceedings of the National Academy of Sciences, 98(3), 992-999.

Chiang, W. C., Ching, T. T., Lee, H. C., Mousigian, C., & Hsu, A. L. (2012). HSF-1 regulators DDL-1/2 link insulin-like signaling to heat-shock responses and modulation of longevity. Cell, 148(1), 322-334.

Chiou, S. H., Yu, C. C., Huang, C. Y., Lin, S. C., Liu, C. J., Tsai, T. H., Chou, S.H., Chien, C.S., Ku, H.H. & Lo, J. F. (2008). Positive correlations of Oct-4 and Nanog in oral cancer stem-like cells and high-grade oral squamous cell carcinoma. Clinical Cancer Research, 14(13), 4085-4095.

Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N. & Lovell-Badge, R. (1996). A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. Development, 122(2), 509-520.

Compeau, P. E., Pevzner, P. A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. Nature Biotechnology, 29(11), 987-991.

Conboy, I. M., & Rando, T. A. (2005). Aging, stem cells and tissue regeneration: lessons from muscle. Cell Cycle, 4(3), 407-410.

Conesa, A., & Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics, 2008.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21(18), 3674-3676.

Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science, 322(5909), 1845-1848.

Costa, S., & Shaw, P. (2006). Chromatin organization and cell fate switch respond to positional information in *Arabidopsis*. Nature, 439(7075), 493-496.

Dali-Youcef, N., Lagouge, M., Froelich, S., Koehl, C., Schoonjans, K., & Auwerx, J. (2007). Sirtuins: the 'magnificent seven', function, metabolism and longevity. Annals of Medicine, 39(5), 335-345.

Dang, C. V. (1999). c-Myc target genes involved in cell growth, apoptosis, and metabolism. Molecular and Cellular Biology, 19(1), 1-11.

Deng, Y., Chan, S. S., & Chang, S. (2008). Telomere dysfunction and tumour suppression: the senescence connection. Nature Reviews Cancer, 8(6), 450-458.

Djojosubroto, M. W., Choi, Y. S., Lee, H. W., & Rudolph, K. L. (2003). Telomeres and telomerase in aging, regeneration and cancer. Molecules and Cells, 15(2), 164-175.

Dome, J. S., Chung, S., Bergemann, T., Umbricht, C. B., Saji, M., Carey, L. A., Grundy, P.E., Perlman, E.J., Breslow, N.E. & Sukumar, S. (1999). High telomerase reverse transcriptase (hTERT) messenger RNA level correlates with tumor recurrence in patients with favorable histology Wilms' tumor. Cancer Research, 59(17), 4301-4307.

Du, J., Chen, T., Zou, X., Xiong, B., & Lu, G. (2010). Dppa2 knockdown-induced differentiation and repressed proliferation of mouse embryonic stem cells. Journal of Biochemistry, 147(2), 265-271.

Dubik, D., Dembinski, T. C., & Shiu, R. P. (1987). Stimulation of c-myc oncogene expression associated with estrogen-induced proliferation of human breast cancer cells. Cancer Research, 47(24), 6517-6521.

DuBuc, T. Q., Traylor-Knowles, N., & Martindale, M. Q. (2014). Initiating a regenerative response; cellular and molecular features of wound healing in the cnidarian *Nematostella vectensis*. BMC Biology, 12(1), 24.

El-Karim, E. A., Hagos, E. G., Ghaleb, A. M., Yu, B., & Yang, V. W. (2013). Krüppel-like factor 4 regulates genetic stability in mouse embryonic fibroblasts. Molecular Cancer, 12, 89.

Elliott, S. A., & Sánchez-Alvarado, A. (2013). The history and enduring contributions of planarians to the study of animal regeneration. Wiley Interdisciplinary Reviews: Developmental Biology, 2(3), 301-326.

Elmore, L. W., Norris, M. W., Sircar, S., Bright, A. T., McChesney, P. A., Winn, R. N., & Holt, S. E. (2008). Upregulation of telomerase function during tissue regeneration. Experimental Biology and Medicine, 233(8), 958-967.

Ezeh, U. I., Turek, P. J., Reijo, R. A., & Clark, A. T. (2005). Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma. Cancer, 104(10), 2255-2265.

Faisal, F. E., Zhao, H., & Milenković, T. (2014). Global network alignment in the context of aging. In Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 580-580). ACM.

Feng, B., Ng, J. H., Heng, J. C. D., & Ng, H. H. (2009). Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. Cell Stem Cell, 4(4), 301-312.

Feng, J., Funk, W. D., Wang, S. S., Weinrich, S. L., Avilion, A. A., Chiu, C. P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J. & Le, S. (1995). The RNA component of human telomerase. Science, 269(5228), 1236-1241.

Feng, Q., Lu, S. J., Klimanskaya, I., Gomes, I., Kim, D., Chung, Y., ... & Lanza, R. (2010). Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. Stem Cells, 28(4), 704-712.

Finch, C. E. (1994). Longevity, senescence, and the genome. University of Chicago Press.

Fontana, L., Partridge, L., & Longo, V. D. (2010). Extending healthy life span—from yeast to humans. Science, 328(5976), 321-326.

Galliot, B. (2012). *Hydra*, a fruitful model system for 270 years. International Journal of Developmental Biology, 56, 411-423.

Garbe, J., Wong, M., Wigington, D., Yaswen, P., & Stampfer, M. R. (1999). Viral oncogenes accelerate conversion to immortality of cultured conditionally immortal human mammary epithelial cells. Oncogene, 18(13).

Gauchat, D., Mazet, F., Berney, C., Schummer, M., Kreger, S., Pawlowski, J., & Galliot, B. (2000). Evolution of Antp-class genes and differential expression of *Hydra* Hox/paraHox genes in anterior patterning. Proceedings of the National Academy of Sciences, 97(9), 4493-4498.

Gearhart, J., Pashos, E. E., & Prasad, M. K. (2007). Pluripotency redux—advances in stem-cell research. New England Journal of Medicine, 357(15), 1469-1472.

Gene Ontology Consortium. (2015). Gene ontology consortium: going forward. Nucleic Acids Research, 43(D1), D1049-D1056.

Gentile, L., Cebrià, F., & Bartscherer, K. (2011). The planarian flatworm: an in vivo model for stem cell biology and nervous system regeneration. Disease Models & Mechanisms, 4(1), 12-19.

Ghaffari, N., Arshad, O., Jeong, H., Thiltges, J., Criscitiello, M., Yoon, B. J., Ghaffari, N., Arshad, O., Jeong, H., Thiltges, J., Criscitiello, M., Yoon, B.J., Datta, A. & Johnson, C. (2015). Examining De Novo transcriptome assemblies via a quality assessment pipeline. IEEE/ACM Transactions on Computational Biology and Bioinformatics.

Gogoi, S. K., Gopinath, P., Paul, A., Ramesh, A., Ghosh, S. S., & Chattopadhyay, A. (2006). Green fluorescent protein-expressing *Escherichia coli* as a model system for investigating the antimicrobial activities of silver nanoparticles. Langmuir, 22(22), 9322-9328.

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

Gomes, N., Shay, J. W., & Wright, W. E. (2010). Telomere biology in Metazoa. FEBS letters, 584(17), 3741-3751.

Gong, J., Zhu, B., Murshid, A., Adachi, H., Song, B., Lee, A., Liu C. & Calderwood, S. K. (2009). T cell activation by heat shock protein 70 vaccine requires TLR signaling and scavenger receptor expressed by endothelial cells-1. The Journal of Immunology, 183(5), 3092-3098.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research, 36(10), 3420-3435.

Greenberg, R. A., O'Hagan, R. C., Deng, H., Xiao, Q., Hann, S. R., Adams, R. R., Lichtsteiner, S., Chin, L., Morin, G.B. & DePinho, R. A. (1999). Telomerase reverse transcriptase gene is a direct target of c-Myc but is not functionally equivalent in cellular transformation. Oncogene, 18(5).

Guan, H., Xie, L., Leithäuser, F., Flossbach, L., Möller, P., Wirth, T., & Ushmorov, A. (2010). KLF4 is a tumor suppressor in B-cell non-Hodgkin lymphoma and in classic Hodgkin lymphoma. Blood, 116(9), 1469-1478.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. & MacManes, M. D. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature Protocols, 8(8).

Hamazaki, T., Kehoe, S. M., Nakano, T., & Terada, N. (2006). The Grb2/Mek pathway represses Nanog in murine embryonic stem cells. Molecular and Cellular Biology, 26(20), 7539-7549.

Hayflick, L. (1989). Antecedents of cell aging research. Experimental gerontology, 24(5), 355-365.

Herranz, D., Muñoz-Martin, M., Cañamero, M., Mulero, F., Martinez-Pastor, B., Fernandez-Capetillo, O., & Serrano, M. (2010). Sirt1 improves healthy ageing and protects from metabolic syndrome-associated cancer. Nature Communications, 1, 3.

Hiratsuka, M., Inoue, T., Toda, T., Kimura, N., Shirayoshi, Y., Kamitani, H., Watanabe, T., Ohama, E., Tahimic, C.G., Kurimasa, A. & Oshimura, M. (2003). Proteomics-based identification of differentially expressed genes in human gliomas: down-regulation of SIRT2 gene. Biochemical and biophysical research communications, 309(3), 558-566.

Holtfreter, J. (1947). Observations on the migration, aggregation and phagocytosis of embryonic cells. Journal of Morphology, 80(1), 25-55.

Hoxha, M., Fabris, S., Agnelli, L., Bollati, V., Cutrona, G., Matis, S., Recchia, A.G., Gentile, M., Cortelezzi, A., Morabito, F. & Bertazzi, P. A. (2014). Relevance of telomere/telomerase system impairment in early stage chronic lymphocytic leukemia. Genes, Chromosomes and Cancer, 53(7), 612-621.

Iacobucci, I., Ferrarini, A., Sazzini, M., Giacomelli, E., Lonetti, A., Xumerle, L., Ferrari, A., Papayannidis, C., Malerba, G., Luiselli, D. & Boattini, A. (2012). Application of the whole-transcriptome shotgun sequencing approach to the study of Philadelphia-positive acute lymphoblastic leukemia. Blood Cancer Journal, 2(3), 61.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y. & Lemischka, I. R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature, 442(7102), 533-538.

Iwafuchi-Doi, M., & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. Genes & Development, 28(24), 2679-2692.

Jopling, C., Boue, S., & Belmonte, J. C. I. (2011). Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. Nature reviews molecular cell biology, 12(2), 79-89.

Jun-Li, L., Yang, Q., Wei-Min, T., Hergenhahn, M., Zhao-Qi, W., & Hollstein, M. (2001). Knock-in mice with a chimeric human/murine p53 gene develop normally and show wild-type p53 responses to DNA damaging agents: a new biomedical research tool. Oncogene, 20(3), 320.

Kaeberlein, M., McVey, M., & Guarente, L. (1999). The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes & Development, 13(19), 2570-2580.

Kaul, Z., Cesare, A. J., Huschtscha, L. I., Neumann, A. A., & Reddel, R. R. (2012). Five dysfunctional telomeres predict onset of senescence in human cells. EMBO Reports, 13(1), 52-59.

Kerr, J. F., Winterford, C. M., & Harmon, B. V. (1994). Apoptosis. Its significance in cancer and cancer therapy. Cancer, 73(8), 2013-2026.

Kielman, M. F., Rindapää, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., Taketo, M.M., Roberts, S., Smits, R. & Fodde, R. (2002). Apc modulates embryonic stem-cell differentiation by controlling the dosage of β-catenin signaling. Nature genetics, 32(4), 594-605.

Kiernan, A. E., Pelling, A. L., Leung, K. K., Tang, A. S., Bell, D. M., Tease, C., Lovell-Badge, R., Steel, K.P. & Cheah, K. S. (2005). Sox2 is required for sensory organ development in the mammalian inner ear. Nature, 434(7036), 1031-1035.

Kikuchi, K., Holdway, J. E., Werdich, A. A., Anderson, R. M., Fang, Y., Egnaczyk, G. F., Evans, T., MacRae, C.A., Stainier, D.Y. & Poss, K. D. (2010). Primary contribution to zebrafish heart regeneration by gata4+ cardiomyocytes. Nature, 464(7288), 601-605.

Kim, S. K., Suh, M. R., Yoon, H. S., Lee, J. B., Oh, S. K., Moon, S. Y., Lee, J.Y., Hwang, J.H., Cho, W.J. & Kim, K. S. (2005). Identification of developmental pluripotency associated 5 expression in human pluripotent stem cells. Stem Cells, 23(4), 458-462.

Kim, S. U., & De Vellis, J. (2009). Stem cell-based cell therapy in neurological diseases. Journal of Neuroscience Research, 87(10), 2183-2200.

Kimura, K. D., Tissenbaum, H. A., Liu, Y., & Ruvkun, G. (1997). daf-2, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. Science, 277(5328), 942-946.

Kortschak, R. D., Samuel, G., Saint, R., Miller, D.J. (2003). EST analysis of the Cnidarian Acropora millepora reveals extensive gene loss and rapid sequence divergence in the model invertebrates. Current Biology. 13, 2190–2195.

Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H. A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M.Q. & Holstein, T. W. (2005). Unexpected complexity of the Wnt gene family in a sea anemone. Nature, 433(7022), 156-160.

Lee, H. J., Göring, W., Ochs, M., Mühlfeld, C., Steding, G., Paprotta, I., Lee, H.J., Göring, W., Ochs, M., Mühlfeld, C., Steding, G., Paprotta, I., Engel, W. & Adham, I. M. (2004). Sox15 is required for skeletal muscle regeneration. Molecular and Cellular Biology, 24(19), 8428-8436.

Lee, K., Zhan, X., Gao, J., Qiu, J., Feng, Y., Meganathan, R., Cohen, S.N. & Georgiou, G. (2003). RraA: a protein inhibitor of RNase E activity that globally modulates RNA abundance in *E. coli*. Cell, 114(5), 623-634.

Lengner, C. J., Welstead, G. G., & Jaenisch, R. (2008). The pluripotency regulator Oct4: a role in somatic stem cells?. Cell Cycle, 7(6), 725-728.

Levine, A. J., & Brivanlou, A. H. (2006). GDF3 at the crossroads of TGF-beta signaling. Cell Cycle, 5(10), 1069-1073.

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. Nature, 424(6945), 147-151.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K.& Li, S. (2010). De novo assembly of human genomes with massively parallel short read sequencing. Genome Research, 20(2), 265-272.

Lien, C. L., Schebesta, M., Makino, S., Weber, G. J., & Keating, M. T. (2006). Gene expression analysis of zebrafish heart regeneration. PLoS Biology, 4(8), 260.

Little, C. D., Nau, M. M., Carney, D. N., Gazdar, A. F., & Minna, J. D. (1983). Amplification and expression of the c-myc oncogene in human lung cancer cell lines. Nature, 306(5939), 194-196.

Liu, K., Lin, B., Zhao, M., Yang, X., Chen, M., Gao, A., Liu, F., Que, J. & Lan, X. (2013). The multiple roles for Sox2 in stem cell maintenance and tumorigenesis. Cellular Signaling, 25(5), 1264-1271.

Lo, H. S., Wang, Z., Hu, Y., Yang, H. H., Gere, S., Buetow, K. H., & Lee, M. P. (2003). Allelic variation in gene expression is common in the human genome. Genome Research, 13(8), 1855-1862.

Longo, V. D., & Kennedy, B. K. (2006). Sirtuins in aging and age-related disease. Cell, 126(2), 257-268.

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. Cell, 153(6), 1194-1217.

Loriot, A., Boon, T., & De Smet, C. (2003). Five new human cancer-germline genes identified among 12 genes expressed in spermatogonia. International Journal of Cancer, 105(3), 371-376.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y. & Tang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience, 1(1), 18.

MacArthur, B. D., & Lemischka, I. R. (2013). Statistical mechanics of pluripotency. Cell, 154(3), 484-489.

MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. Frontiers in Genetics, 5, 13.

MacQuarrie, K. L., Fong, A. P., Morse, R. H., & Tapscott, S. J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. Trends in Genetics, 27(4), 141-148.

Mahmoudi, S., & Brunet, A. (2012). Aging and reprogramming: a two-way street. Current Opinion in Cell Biology, 24(6), 744-756.

Martinez, D. E. (1998). Mortality patterns suggest lack of senescence in *hydra*. Experimental Gerontology, 33(3), 217-225.

Martinez, D. E., & Bridge, D. (2012). *Hydra*, the everlasting embryo, confronts aging. International Journal of Developmental Biology, 56(6-7-8), 479-487.

Martínez, P., & Blasco, M. A. (2010). Role of shelterin in cancer and aging. Aging Cell, 9(5), 653-666.

Martínez, P., Gómez‐López, G., Pisano, D. G., Flores, J. M., & Blasco, M. A. (2016). A genetic interaction between RAP1 and telomerase reveals an unanticipated role for RAP1 in telomere maintenance. Aging cell, 15(6), 1113-1125.

Maruyama, M., Ichisaka, T., Nakagawa, M., & Yamanaka, S. (2005). Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. Journal of Biological Chemistry, 280(26), 24371-24379.

Masaki, H., Nishida, T., Kitajima, S., Asahina, K., & Teraoka, H. (2007). Developmental pluripotency-associated 4 (DPPA4) localized in active chromatin inhibits mouse embryonic stem cell differentiation into a primitive ectoderm lineage. Journal of Biological Chemistry, 282(45), 33034-33042.

Matsunaga, A., Tsugawa, M., & Fortes, J. (2008). Cloudblast: combining mapreduce and virtualization on distributed resources for bioinformatics applications. eScience 2008. IEEE Fourth International Conference, 222-229.

Metzker, M. L. (2010). Sequencing technologies—the next generation. Nature Reviews Genetics, 11(1), 31-46.

Miglietta, M. P., & Lessios, H. A. (2009). A silent invasion. Biological Invasions, 11(4), 825-834.

Miglietta, M. P., Piraino, S., Kubota, S., & Schuchert, P. (2007). Species in the genus *Turritopsis* (Cnidaria, Hydrozoa): a molecular evaluation. Journal of Zoological Systematics and Evolutionary Research, 45(1), 11-19.

Miller, D. J., Hayward, D. C., Reece-Hoyes, J. S., Scholten, I., Catmull, J., Gehring, W. J., Callaerts, P., Larsen, J.E. & Ball, E. E. (2000). Pax gene diversity in the basal

cnidarian *Acropora millepora* (Cnidaria, Anthozoa): implications for the evolution of the Pax gene family. Proceedings of the National Academy of Sciences, 97(9), 4475-4480.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. & Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell, 113(5), 631-642.

Mochizuki, Y., He, J., Kulkarni, S., Bessler, M., & Mason, P. J. (2004). Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. Proceedings of the National Academy of Sciences of the United States of America, 101(29), 10756-10761.

Moskalev, A. A., Smit-McBride, Z., Shaposhnikov, M. V., Plyusnina, E. N., Zhavoronkov, A., Budovsky, A., Tacutu, R. & Fraifeld, V. E. (2012). Gadd45 proteins: relevance to aging, longevity and age-related pathologies. Ageing research reviews, 11(1), 51-66.

Mostovoy, Y., Levy-Sakin, M., Lam, J., Lam, E. T., Hastie, A. R., Marks, P., Lee, J., Chu, C., Lin, C., Džakula, Ž. & Cao, H. (2016). A hybrid approach for de novo human genome sequence assembly and phasing. Nature Methods, 13(7), 587-590.

Muñoz-Sanjuán, I., & Brivanlou, A. H. (2002). Neural induction, the default model and embryonic stem cells. Nature Reviews Neuroscience, 3(4), 271-280.

Murthy, M., & Ram, J. L. (2015). Invertebrates as model organisms for research on aging biology. Invertebrate Reproduction & Development, 59(1), 1-4.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, 320(5881), 1344-1349.

Neumann, A. A., Watson, C. M., Noble, J. R., Pickett, H. A., Tam, P. P., & Reddel, R. R. (2013). Alternative lengthening of telomeres in normal mammalian somatic cells. Genes & Development, 27(1), 18-23.

Newbold, R. F., & Overell, R. W. (1983). Fibroblast immortality is a prerequisite for transformation by EJ c-Ha-ras oncogene. Nature, 304(5927), 648-651.

Niwa, H., Miyazaki, J. I., & Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of embryonic stem cells. Nature Genetics, 24(4), 372-376.

Nowak, T., Januszkiewicz, D., Zawada, M., Pernak, M., Lewandowski, K., Rembowska, J., Nowicka, K., Mankowski, P. & Nowak, J. (2006). Amplification of hTERT and

hTERC genes in leukemic cells with high expression and activity of telomerase. Oncology reports, 16(2), 301-305.

Nye, H. L., Cameron, J. A., Chernoff, E. A., & Stocum, D. L. (2003). Regeneration of the urodele limb: a review. Developmental Dynamics, 226(2), 280-294.

Okada, T. S. (1991). Transdifferentiation: flexibility in cell differentiation. Oxford University Press, USA.

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell, 99(3), 247-257.

Okuda, A., Fukushima, A., Nishimoto, M., Orimo, A., Yamagishi, T., Nabeshima, Y., Kuro-o, M., Nabeshima, Y.I., Boon, K., Keaveney, M. & Stunnenberg, H. G. (1998). UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. The EMBO Journal, 17(7), 2019-2032.

Oliver, H. F., Orsi, R. H., Ponnala, L., Keich, U., Wang, W., Sun, Q., Cartinhour, S.W., Filiatrault, M.J., Wiedmann, M. & Boor, K. J. (2009). Deep RNA sequencing of L. monocytogenes reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs. BMC Genomics, 10(1), 641.

Orsulic, S., Li, Y., Soslow, R. A., Vitale-Cross, L. A., Gutkind, J. S., & Varmus, H. E. (2002). Induction of ovarian cancer by defined multiple genetic changes in a mouse model system. Cancer Cell, 1(1), 53-62.

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. Nature Reviews Genetics, 12(2), 87-98.

Park, S., Mori, R., & Shimokawa, I. (2013). Do sirtuins promote mammalian longevity?: A Critical review on its relevance to the longevity effect induced by calorie restriction. Molecules and Cells, 35(6), 474-480.

Pellettieri, J., & Sánchez-Alvarado, A. (2007). Cell turnover and adult tissue homeostasis: from humans to planarians. Annual Review of Genetics., 41, 83-105.

Pesce, M., & Schöler, H. R. (2001). Oct-4: gatekeeper in the beginnings of mammalian development. Stem Cells, 19(4), 271-278.

Piraino, S., Boero, F., Aeschbach, B., & Schmid, V. (1996). Reversing the life cycle: medusae transforming into polyps and cell transdifferentiation in *Turritopsis nutricula* (Cnidaria, Hydrozoa). Biological Bulletin, 302-312.

Piraino, S., De Vito, D., Schmich, J., Bouillon, J., & Boero, F. (2004). Reverse development in Cnidaria. Canadian Journal of Zoology, 82(11), 1748-1754.

Pitt, J. N., & Kaeberlein, M. (2015). Why is aging conserved and what can we do about it?. PLoS Biology, 13(4), 1002131.

Poss, K. D., Wilson, L. G., & Keating, M. T. (2002). Heart regeneration in zebrafish. Science, 298(5601), 2188-2190.

Prouty, N. G., Roark, E. B., Buster, N. A., & Ross, S. W. (2011). Growth rate and age distribution of deep-sea black corals in the Gulf of Mexico. Marine Ecology Progress Series, 423, 101-115.

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V. & Jurka, J. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science, 317(5834), 86-94.

Qian, X., Kim, J. K., Tong, W., Villa-Diaz, L. G., & Krebsbach, P. H. (2016). DPPA5 supports pluripotency and reprogramming by regulating NANOG turnover. Stem Cells, 34(3), 588-600.

Raffa, G. D., Cenci, G., Ciapponi, L., & Gatti, M. (2013). Organization and evolution of Drosophila terminin: similarities and differences between Drosophila and human telomeres. Frontiers in oncology, 3.

Reiter, S., Crescenzi, M., Galliot, B., & Buzgariu, W. C. (2012). *Hydra*, a versatile model to study the homeostatic and developmental functions of cell death. International Journal of Developmental Biology, 56(6-7-8), 593-604.

Roark, E. B., Guilderson, T. P., Dunbar, R. B., Fallon, S. J., & Mucciarone, D. A. (2009). Extreme longevity in proteinaceous deep-sea corals. Proceedings of the National Academy of Sciences, 106(13), 5204-5208.

Rogina, B., & Helfand, S. L. (2004). Sir2 mediates longevity in the fly through a pathway related to calorie restriction. Proceedings of the National Academy of Sciences of the United States of America, 101(45), 15998-16003.

Rowland, B. D., Bernards, R., & Peeper, D. S. (2005). The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene. Nature Cell Biology, 7(11), 1074-1082.

Saigusa, S., Tanaka, K., Toiyama, Y., Yokoe, T., Okugawa, Y., Ioue, Y., Miki, C. & Kusunoki, M. (2009). Correlation of CD133, OCT4, and SOX2 in rectal cancer and their

association with distant recurrence after chemoradiotherapy. Annals of Surgical Oncology, 16(12), 3488-3498.

Sánchez-Alvarado, A., & Tsonis, P. A. (2006). Bridging the regeneration gap: genetic insights from diverse animal models. Nature Reviews Genetics, 7(11), 873-884.

Sánchez-Alvarado, A., & Yamanaka, S. (2014). Rethinking differentiation: stem cells, regeneration, and plasticity. Cell, 157(1), 110-119.

Sasai, Y. (2003). Roles of Sox factors in neural determination: conserved signaling in evolution?. International Journal of Developmental Biology, 45(1), 321-326.

Schatten, H., Chakrabarti, A., & Hedrick, J. (1999). Centrosome and microtubule instability in aging *Drosophila* cells. Journal of Cellular Biochemistry, 74(2), 229-241.

Schmich, J., Kraus, Y., De Vito, D., Graziussi, D., Boero, F., & Piraino, S. (2007). Induction of reverse development in two marine hydrozoans. International Journal of Developmental Biology, 51(1), 45.

Scotland, K. B., Chen, S., Sylvester, R., & Gudas, L. J. (2009). Analysis of Rex1 (zfp42) function in embryonic stem cell differentiation. Developmental Dynamics, 238(8), 1863-1877.

Segre, J. A., Bauer, C., & Fuchs, E. (1999). Klf4 is a transcription factor required for establishing the barrier function of the skin. Nature Genetics, 22(4), 356-360.

Seifert, A. W., & Voss, S. R. (2013). Revisiting the relationship between regenerative ability and aging. BMC Biology, 11(1), 2.

Sell, C. (2015). Minireview: The complexities of IGF/Insulin signaling in aging: why flies and worms are not humans. Molecular Endocrinology, 29(8), 1107-1113.

Shay, J. W. (1997). Telomerase in human development and cancer. Journal of Cellular Physiology, 173(2), 266-270.

Shay, J. W., & Gazdar, A. F. (1997). Telomerase in the early detection of cancer. Journal of Clinical Pathology, 50(2), 106-109.

Sherwood, O. A., & Edinger, E. N. (2009). Ages and growth rates of some deep-sea gorgonian and antipatharian corals of Newfoundland and Labrador. Canadian Journal of Fisheries and Aquatic Sciences, 66(1), 142-152.

Shi, G., & Jin, Y. (2010). Role of Oct4 in maintaining and regaining stem cell pluripotency. Stem Cell Research and Therapy, 1(5), 39.

Shpirer, E., Chang, E. S., Diamant, A., Rubinstein, N., Cartwright, P., & Huchon, D. (2014). Diversity and evolution of myxozoan minicollagens and nematogalectins. BMC Evolutionary Biology, 14(1), 205.

Silva, H., & Conboy, I. M. (2008). Aging and stem cell renewal. StemBook, the Stem Cell Research Community.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 351.

Smogorzewska, A., & de Lange, T. (2004). Regulation of telomerase by telomeric proteins. Annual Review of Biochemistry, 73(1), 177-208.

Soder, A. I., Hoare, S. F., Muire, S., Balmain, A., Parkinson, E. K., & Keith, W. N. (1997). Mapping of the Gene for the Mouse Telomerase RNA Component, Terc, to Chromosome 3 by Fluorescencein SituHybridization and Mouse Chromosome Painting. Genomics, 41(2), 293-294.

Someya, S., Yu, W., Hallows, W. C., Xu, J., Vann, J. M., Leeuwenburgh, C., Tanokura, M., Denu, J.M. & Prolla, T. A. (2010). Sirt3 mediates reduction of oxidative damage and prevention of age-related hearing loss under caloric restriction. Cell, 143(5), 802-812.

Sousounis, K., Baddour, J. A., & Tsonis, P. A. (2014). Aging and regeneration in vertebrates. Current Topics of Developmental Biology, 108, 217-246.

Spring, J., Yanze, N., Middel, A. M., Stierwald, M., Gröger, H., & Schmid, V. (2000). The mesoderm specification factor twist in the life cycle of jellyfish. Developmental Biology, 228(2), 363-375.

Stearns, S. C. (1976). Life-history tactics: a review of the ideas. Quarterly Review of Biology, 3-47.

Steele, R. E. (2012). The *Hydra* genome: insights, puzzles and opportunities for developmental biologists. International Journal of Developmental Biology, 56(6-7-8), 535-542.

Sullivan, J. C., & Finnerty, J. R. (2007). A surprising abundance of human disease genes in a simple “basal” animal, the starlet sea anemone (*Nematostella vectensis*). Genome, 50(7), 689-692.

Tai, M. H., Chang, C. C., Olson, L. K., & Trosko, J. E. (2005). Oct4 expression in adult human stem cells: evidence in support of the stem cell theory of carcinogenesis. Carcinogenesis, 26(2), 495-502.

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell, 126(4), 663-676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell, 131(5), 861-872.

Taub, D. D., & Longo, D. L. (2005). Insights into thymic aging and regeneration. Immunological Reviews, 205(1), 72-93.

Tejera, A. M., d'Alcontres, M. S., Thanasoula, M., Marion, R. M., Martinez, P., Liao, C., Flores, J.M., Tarsounas, M. & Blasco, M. A. (2010). TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. Developmental Cell, 18(5), 775-789.

Terman, A., & Brunk, U. T. (2004). Aging as a catabolic malfunction. The International Journal of Biochemistry & Cell Biology, 36(12), 2365-2375.

Tissenbaum, H. A., & Guarente, L. (2001). Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature, 410(6825), 227-230.

Traut, W., Szczepanowski, M., Vítková, M., Opitz, C., Marec, F., & Zrzavý, J. (2007). The telomere repeat motif of basal Metazoa. Chromosome Research, 15(3), 371-382.

Tsonis, P. A., & Del Rio-Tsonis, K. (2004). Lens and retina regeneration: transdifferentiation, stem cells and clinical applications. Experimental Eye Research, 78(2), 161-172.

Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y., & Kondoh, H. (2003). Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. Developmental Cell, 4(4), 509-519.

Vaupel, J. W., Baudisch, A., Dölling, M., Roach, D. A., & Gampe, J. (2004). The case for negative senescence. Theoretical Population Biology, 65(4), 339-351.

Vilchez, D., Morantte, I., Liu, Z., Douglas, P. M., Merkwirth, C., Rodrigues, A. P., Manning, G. & Dillin, A. (2012). RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions. Nature, 489(7415), 263-268.

Wagner, A. (1994). Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. Proceedings of the National Academy of Sciences, 91(10), 4387-4391.

Watanabe, H., Mättner, R., & Holstein, T. W. (2009). Immortality and the base of multicellular life: lessons from cnidarian stem cells. Seminars in Cell & Developmental Biology, 20(9), 1114-1125.

Westerheide, S. D., Anckar, J., Stevens, S. M., Sistonen, L., & Morimoto, R. I. (2009). Stress-inducible regulation of heat shock factor 1 by the deacetylase SIRT1. Science, 323(5917), 1063-1066.

Whitaker, N. J., Bryan, T. M., Bonnefin, P., Chang, A. C., Musgrove, E. A., Braithwaite, A. W., & Reddel, R. R. (1995). Involvement of RB-1, p53, p16INK4 and telomerase in immortalisation of human cells. Oncogene, 11(5), 971-976.

Wildwater, M., Campilho, A., Perez-Perez, J. M., Heidstra, R., Blilou, I., Korthout, H., Chatterjee, J., Mariconti, L., Gruissem, W.& Scheres, B. (2005). The retinoblastoma-related gene regulates stem cell maintenance in *Arabidopsis* roots. Cell, 123(7), 1337-1349.

Williams, P. D., & Day, T. (2003). Antagonistic pleiotropy, mortality source interactions, and the evolutionary theory of senescence. Evolution, 57(7), 1478-1488.

Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., Therkildsen, N.O., Morikawa, M. & Palumbi, S. R. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Molecular Ecology Resources, 12(6), 1058-1067.

Xin, H., Liu, D., & Songyang, Z. (2008). The telosome/shelterin complex and its functions. Genome Biology, 9(9), 232.

Yang, J., Chai, L., Fowles, T. C., Alipio, Z., Xu, D., Fink, L. M., Ward, D.C. & Ma, Y. (2008). Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. Proceedings of the National Academy of Sciences, 105(50), 19756-19761.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R. & Slukvin, I. I. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science, 318(5858), 1917-1920.

Zaehres, H., Lensch, M. W., Daheron, L., Stewart, S. A., Itskovitz-Eldor, J., & Daley, G. Q. (2005). High-efficiency RNA interference in human embryonic stem cells. Stem Cells, 23(3), 299-305.

Zahmatkesh, A., Ansari Mahyari, S., Daliri Joupari, M., Shirazi, A., & Rahmani, H. (2015). Expression of bovine Ecat1 gene in immature and in vitro matured oocytes as

well as during early embryonic development. Reproduction in Domestic Animals, 50(1), 34-40.

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics, 17(9), 847-848.

Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simão, F. A., Ioannidis, P., Seppey, M., Loetscher, A.& Kriventseva, E. V. (2016). OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Research, 1119.

Zerbino, D. R. (2010). Using the velvet de novo assembler for short-read sequencing technologies. Current Protocols in Bioinformatics, 11-5.

Zhang, J., Tam, W. L., Tong, G. Q., Wu, Q., Chan, H. Y., Soh, B. S., Lou, Y., Yang, J., Ma, Y., Chai, L. & Ng, H. H. (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nature Cell Biology, 8(10), 1114-1123.

Zhao, Q. Y., Wang, Y., Kong, Y. M., Luo, D., Li, X., & Hao, P. (2011). Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics, 12(14), S2.

APPENDIX A

Polyp				Jellyfish				Cyst			
#	Pathways	# Sequences	# Enzymes	#	Pathways	# Sequences	# Enzymes	#	Pathways	# Sequences	# Enzymes
1	Purine Metabolism	2712	47	1	Purine Metabolism	1550	45	1	Purine metabolism	2045	50
2	Thiamine metabolism	2350	4	2	Thiamine metabolism	1272	4	2	Thiamine metabolism	1665	3
3	Biosynthesis of antibiotics	802	127	3	Biosynthesis of antibiotics	309	106	3	Biosynthesis of antibiotics	599	144
4	Aminobenzoate degredation	376	7	4	Aminobenzoate degredation	270	6	4	Aminobenzoate degradation	318	6
5	Oxidative phosphorylation	351	7	5	Pyrimidine metabolism	209	31	5	Pyrimidine metabolism	300	32
6	T cell receptor signaling pathway	289	2	6	T cell receptor signaling pathway	196	2	6	T cell receptor signaling pathway	229	2
7	Pyrimidine metabolism	279	32	7	Th1 and Th2 cell differentiation	175	1	7	Th1 and Th2 cell differentiation	204	1
8	Glycolysis/Gluconeogenesis	262	25	8	Phosphatidylinositol signaling pathway	125	19	8	Oxidative phosphorylation	176	8
9	Th1 and Th2 cell differentiation	256	1	9	Drug metabolism - other enzymes	124	13	9	Citrate cycle (TCA cycle)	148	19
10	Cysteine and methionine metabolism	223	29	10	Glutathione metabolism	116	15	10	Cysteine and methionine metabolism	138	35
11	Carbon fixation in photosynthetic organisms	176	18	11	Lysine degradation	110	9	11	Glycolysis / Gluconeogenesis	132	25
12	Glutathione metabolism	173	15	12	Oxidative phosphorylation	101	9	12	Glutathione metabolism	131	15
13	Citrate cycle (TCA cycle)	162	18	13	Inositol phosphate metabolism	93	7	13	Aminoacyl-tRNA biosynthesis	129	20
14	Pyruvate metabolism	161	23	14	Cysteine and methionine metabolism	92	27	14	Tryptophan metabolism	125	13
15	Glyoxylate and dicarboxylate metabolism	157	23	15	Glycerophospholipid metabolism	85	19	15	Carbon fixation pathways in prokaryotes	125	20
16	Carbon fixation pathway in prokaryotes	153	22	16	Amino sugar and nucleotide sugar metabolism	79	28	16	Amino sugar and nucleotide sugar metabolism	124	33
17	Arginine and proline metabolism	141	23	17	Arginine and proline metabolism	78	28	17	Pyruvate metabolism	123	26
18	Methane metabolism	136	17	18	Pyruvate metabolism	71	21	18	Phosphatidylinositol signaling system	114	17
19	Drug metabolism - other enzymes	135	13	19	Glycolysis/Gluconeogenesis	69	22	19	Valine, leucine and isoleucine degradation	110	18
20	Glycine, serine and threonine metabolism	132	23	20	Valine, leucine and isoleucine degradation	68	17	20	Arginine and proline metabolism	104	20
21	Alanine, aspartate and glutamate metabolism	130	23	21	Alanine, aspartate and glutamate metabolism	66	23	21	Alanine, aspartate and glutamate metabolism	100	22
22	Inositol phosphate metabolism	123	22	22	Tryptophan metabolism	65	13	22	Glycine, serine and threonine metabolism	98	35
23	Phenylpropanoid biosynthesis	121	3	23	Aminoacyl-tRNA biosynthesis	65	20	23	Glycerophospholipid metabolism	91	20
24	Valine, leucine and isoleucine degredation	118	21	24	Glycine, serine and threonine metabolism	61	27	24	Glyoxylate and dicarboxylate metabolism	84	15
25	Phosphatidylinositol signaling system	118	21	25	Carbon fixation pathways in prokaryotes	61	18	25	Inositol phosphate metabolism	83	18
26	Pentose phosphate pathway	118	16	26	Phenylpropanoid biosynthesis	58	2	26	Fatty acid degradation	82	9
27	Amino sugar and nucleotide sugar metabolism	117	30	27	beta-Alanine metabolism	57	12	27	Pentose phosphate pathway	82	15
28	Lysine degredation	115	12	28	Glycerolipid metabolism	57	13	28	Glycerolipid metabolism	81	14
29	Tryptophan metabolism	108	16	29	Glyoxylate and dicarboxylate metabolism	55	17	29	Phenylpropanoid biosynthesis	74	2
30	Aminoacyl-tRNA biosynthesis	101	20	30	Citrate cycle (TCA cycle)	55	16	30	Butanoate metabolism	73	15
31	Arginine biosynthesis	94	13	31	Other glycan degradation	49	7	31	Nicotinate and nicotinamide metabolism	71	15
32	Glycerophospholipid metabolism	94	17	32	Fatty acid degradation	47	8	32	beta-Alanine metabolism	65	11
33	Fatty acid degredation	92	11	33	Pentose phosphate pathway	47	15	33	Propanoate metabolism	65	19
34	Fructose and mannose metabolism	90	19	34	Histidine metabolism	46	7	34	Fructose and mannose metabolism	62	16
35	Porphyrin and chlorophyll metabolism	86	14	35	Sphingolipid metabolism	43	10	35	Arginine biosynthesis	62	12
36	Glycerolipid metablism	83	14	36	Ether lipid metabolism	43	5	36	Other glycan degradation	58	7
37	Ascorbate abd aldarate metabolism	74	6	37	Cyanoamino acid metabolism	42	3	37	Histidine metabolism	57	10
38	beta-Alanine metabolism	72	16	38	Arachidonic acid metabolism	42	8	38	Starch and sucrose metabolism	56	17
39	Nitgoren metabolism	68	9	39	Porphyrin and chlorophyll metabolism	40	14	39	Methane metabolism	54	13
40	Ether lipid metabolism	68	7	40	Nicotinate and nicotinamide metabolism	40	14	40	Ascorbate and aldarate metabolism	54	5
41	Propanoate metabolism	68	17	41	Taurine and hypotaurine metabolism	39	4	41	Nitrogen metabolism	52	10
42	Histidine metabolism	64	7	42	Methane metabolism	39	14	42	alpha-Linolenic acid metabolism	50	6
43	Sphingolipid metabolism	63	13	43	Butanoate metabolism	39	12	43	Carbon fixation in photosynthetic organisms	50	15
44	One carbon pool by folate	60	14	44	Ascorbate and aldarate metabolism	39	5	44	One carbon pool by folate	49	13
45	Arachidonic acid metabolism	59	10	45	Stach and sucrose metabolism	38	13	45	Arachidonic acid metabolism	49	8
46	Starch and sucrose metabolism	58	17	46	N-Glycan biosynthesis	36	10	46	Limonene and pinene degradation	46	2
47	Other glycan degradation	57	8	47	Arginine biosynthesis	35	10	47	N-Glycan biosynthesis	45	12
48	alpha-Linolenic acid metabolism	56	5	48	alpha-Linolenic acid metabolism	35	5	48	Phenylalanine metabolism	45	13
49	Drug metabolism - cytochrome p450	56	5	49	mTOR signaling pathway	34	3	49	Chloroalkane and chloroalkene degradation	44	4
50	Tyrosine metabolism	53	13	50	One carbon pool by folate	34	14	50	Porphyrin and chlorophyll metabolism	43	15

APPENDIX B

Induction of pluripotency

Yamanaka factors

#	mRNA Symbol	Function/Significance
1	Oct4 (POU5F1)	Mammalian transcription factor (encoded by POU5F1 gene) that is associated with the maintenance and renewal of undifferentiated embryonic stem cells that remain active during the embryonic stages (Niwa et al., 2000; Takahashi and Yamanaka, 2006; Yamanaka et al., 2007; Lengner et al., 2008; Shi and Jin, 2010).Among the four transcription factors that generated an iPSc, Oct4 is indispensable and is a fundamental factor to inducing a pluripotent state, no factor has been found to replace it (Feng et al., 2009). This transcript alone was capable of inducing a pluripotent state in neural cells (Kim et al., 2009).This transcription factor is highly expressed and is a repressor for cell specialization in undifferentiated states but becomes silenced when the cell undergoes differentiation (Pesce and Scholer, 2001; Zaehres et al., 2005). Oct4 has been reported to promote tumorigenesis by inhibiting cell death (Tai et al., 2005; Chiou et al., 2008; Saigusa et al., 2009).
2	Sox-2	Sox2 is associated with the maintenance of pluripotency in undifferentiated embryonic stem cells and neural stem cells, and plays a fundamental role in the embryonic development of mammals (Avilion et al., 2003; Kiernan et al., 2005). The genes in the Sox family are known to be highly conserved amongst eukaryotic taxa, attracting much interest from evolutionary and clinical geneticists (Collignon et al., 1996; Uchikawa et al., 2003). Sox2 and Oct4 interact and bind DNA cooperatively (Chambers and Tomlinson, 2009). This transcript has been reported to be also be involved in formation of cancers and tumors (Liu et al., 2013; Boumahdi et al., 2014).
3	c-Myc	Proto-oncogene (mutation contributes to cancer) encodes a transcript that regulated gene expression associated with maintains the cell cycle, regulating apoptosis and metabolism (Kerr et al., 1994; Dang, 1999). This regulator is alleged to universally control gene expression of 15% of all genes in humans (Gearhart et al., 2007) and is up-regulated in many cancerous cells (Little et al., 1983; Dubik et al., 1987; Chen et al., 2001).
4	Klf4	Transcript that cellular proliferation, apoptosis, differentiation and reprogramming, thus playing a large role in modulating reparative responses to DNA damage and maintaining its stability (Segre et al., 1999; El-Karim et al., 2013). Klf4 is known to promote the formation of tumorigenesis but has also been shown to be an oncogenic suppressor (Rowland et al., 2005; Guan et al., 2009).

Thompson factors

#	mRNA Symbol	Function/Significance
5	Nanog	Sustains the pluripotent and self-renewing state in mammalian embryonic stem cells and development (Chambers et al., 2003; Mitsui et al., 2003). Human embryonic stem cell gene expressed in cancerous cells (Ezeh et al., 2005). Knockout experiments have resulted in early embryonic death (Jopling et al., 2011).
6	Lin28	Encodes a highly conserved miRNA-binding protein that promotes human embryonic stem cell differentiation and development; directly enhances the production of insulin-like growth factors; found to be involved in tissue regeneration/repair though cellular reprogramming (Chang et al., 2013). When combined with the factors Oct4, Sox3 and Nanog, human somatic cells are able to be reprogrammed into a induced pluripotent stem cell (Yu et al., 2007).

Additional factors tested by Yamanaka group

7	Ecat1 (KHFC3L)	Maternal gene exclusively expressed in female gametes and embryonic stem cells in bovine; through to be required for its sequential development (Zahmatkesh et al., 2015).
8	Dppa5	Specifically and differentially expressed in human cells that have pluripotency; expression is downregulated during induced differentiation of human embryonic stem cells which suggests a critical role in inducing and maintaining a pluripotent state, and can be used as a marker of stem cells (Kim et al., 2005). Directly stablizes and enhances the function of Nanog in human pluropotent stem cells (Qian et al., 2016).
9	Fbxo15	Often used as a marker transcript of pluripotent embryonic stem cells as it is expressed in murine embryonic stem cells and during embryonic development (Shi and Jin, 2010).
10	ERas	Found to be upregulated in cancers/tumors, contributes to the maintenance of morphological changes in embryonic stem cell changes and rapid divison of *in vitro* embryonic stem cells (Takahashi and Yamanaka, 2006).
11	Dnmt3l	Inactive regultory factor of DNA methyltransferases that is essential for the activation and function of Dnmt3A and Dnmt3B, which contibute to the formation of methylation patterns during mammalian embryogenesis (Okano et al., 1999).
12	Ecat8 (TDRD12)	Associated in enbyronic cell transcripts and is prominently associated with cellular reprogramming Feng et al., 2010).
13	Gdf3	Part of a superfamily of a superfamily of ligands that is essential to human embryonic development and specifically active in stem cells (Levine and Brivanlou, 2006). Have not been found many of the tradtional vertebrate model systems, but found in the *Xenopus* embryo (Munoz-Sanjuán and Brivanlou, 2002).
14	Sox15	Influences cell fate determination during development and is associated with the regulation of skeletal muscle regeneration in mice (Lee et al., 2004). Highest activity during pluripotent state and represses during cellular differentiation; demonstrated to have distinct differential roles and functions when compared to Sox2 (Maruyama et al., 2005).
15	Dppa4	Highly active in early embryonic development and embryonic stem cells; overexpression in murine cells caused large scale cellular death and halt in proliferation; regulator of differentiation into ectoderm cell types (Masaki et al., 2007).
16	Dppa2	Specifically active in embryonic stem cells; knockdown of the gene resulted in the differentiation of murine embryonic stem cells and downregulated the activity of Oct4 and Nanog, while increasing the genes that mark differentiation, such as Fst and Psk1; additionally found to decrease the process of cellular proliferation when repressed (Du et al., 2009).
17	Fthl17	Associated with human cancer germlines and giving rise to cancerous cells with stem cell-like functions which often results in tumors (Loriot et al., 2003).
18	Sall4	Regulates the pluripotent state of murine embryonic stem cells through controlling transcription of the gene encoding Oct4; binds to a highly conserved region that regulates the activation of Oct4 *in vivo* and *in vitro*; interference of the transcript has resulted in differentiation of cells (Yang et al., 2008; Zhang et al., 2006).
19	Rex1 (Zfp42)	Expressed in undifferentiated embryonic and adult stem cells in mammals; disruption enhances differential expression of germ layers; believed to influence differentiation, cell cycle maintenance and the progression of malignant tumors (Scotland et al., 2009). Low and high levels of Oct4 represses as intermediate levels activate this gene; thought to maintain the self-renewing and undifferentiated state in stem cells (Shi and Jin, 2010).
20	Utf1	Transcriptional cofactor expressed in murine embryonic stem cells but is rapidly downregulated during differentiation; activated by Oct4 and Sox2 (Okuda et al., 1998; Shi and Jin, 2010). Approximately a 100 fold increase in reprogramming efficiency was seen with the introduction of this factor in human fibroblasts at the same time as a p53 knockout; a known target of the Oct4-Sox2 heterodimer and believed to activate other important reprogramming factors (Feng et al., 2009)
21	Tcl1	Directly associated with human T-cell leukemia and serves as a stem cell marker; found to increase the efficiency of embryonic stem cells in culture and its downregulation induced differentiation (Ivanova et al., 2006).
22	Dppa3 (Stella)	Regulates the pluripotent state in mammalian embryonic stem and embryonic germ cells; associated with the development of matured germ cell tumors (Bowles et al., 2003).
23	β-catenin	Induced by Wnt signal transduction pathways and is fundamental to cell-fate determination; uncontrolled accumulation results in tumors and other developmental defects (Keilman et al., 2002). Contributes to the maintenance of morphological embryonic stem cell changes and rapid divison of embryonic stem cells *in vitro* (Takahashi and Yamanaka, 2006).
24	Stat3	Oncogene that is persistently active in tumors and causes cell immortalization; normal function is to mediate cytokine and growth factor related processes (Bromberg et al., 1999).
25	Grb2	Heavily influences the MAPK/ERK signaling pathway by associating with MEK resulting in the regulation of the expression of Nanog in mammalian embryonic stem cells (Hamazaki et al., 2006).

Telomerase Activity

#	mRNA Symbol	Function/Significance
1	TERT	Contains the reverse transcriptase motif and is the catalytic subunit of of telomerase; One of the major components (other major component TERC) of the telomerase enzyme; The human component produced from the gene, hTERT, adds new segments of DNA to the chromosome ends using the component produced from the TERC gene, hTR, as a template for the elongatio of telomeres by the telomerase enzyme (Feng et al., 1995). Is the only telomerase associated gene that changed expression proportionally to telomerase production, as others either remain high or unchanged (Chang et al., 2002). Not expressed in somatic cells, upregulated in immortal germ cells and tumors/cancerous cells; Direct target of one of the Yamanaka factors, c-Myc, which mediated the induction of this gene (Greenberg et al., 1999). Increased the reprogramming efficiency of human fibroblasts in addition to the core transcription factors (Feng et al., 2009).
2	TERC	One of the major components of the telomerase enzyme (other major component TERT); Produced hTR component in humans and provides a template for the elongation of telomeres by telomerase (Feng et al., 2009). Introduction of this gene in malignant cell lines resulted in the loss of telomeres and promoted senescence, the implication of playing a indispensible role in the maintenance of telomeres (Deng et al., 2008).
3	DKC1	Provides instruction for the production of a protein called dyskerin involved in the maintenance of telomeres; Mutations in this gene lead to a significant destablization of hTR; Plays a primary function in the processing or assembly of the hTR gene into the telomerase complex (Chang et al., 2002; Mochizuki et al., 2004). Mutations that affect this gene result in premature aging syndromes due to decreased telomerase activity and shorter telomeres (Blasco, 2005).
4	TEP1	Suspected to be associated with RNA and protein binding; Perhaps involved in the progression of cancer; Inhibitory experiments with the gene reported the inhibition of telomerase activity in the cells, despite a indirect correlation to telomerase production (Chang et al., 2002). Overexpression of the TERT gene in combination of the underexpression of this gene was correlated to the heightened telomerase levels in cancer cell lines (Dome et al., 1999).
5	TRF1	Part of the shelterin complex that is responsible for telomerase production and activity by creating a loop to stablize the telomeres (Martinez and Blasco, 2010). Directly binds to the telomere repeats as well as other factors to form large complexes to regulate the telomere length and functionality; Function extends outside of telomere regulation as the deletion of the gene in murine models result in embryonic lethality; Expression is significantly altered in maliginant cells (Blasco, 2005).
6	TRF2	Part of the shelterin complex that is responsible for telomerase production and activity by creating a loop to stablize the telomeres (Martinez and Blasco, 2010). Directly binds to the telomere repeats as well as other factors to form large complexes to regulate the telomere length and functionality; Expression is signifciantly altered in malignant cells; Responsible for many diseases that cause chromosomal instability (Blasco, 2005). Specific role of protecting the single-stranded telomeric overhang from degradation and DNA repairation in order to prevent the fusion of the ends of telomeres (Smogorzewska and de Lange, 2004).
7	TIN2	Part of the shelterin complex that is responsible for telomerase production and activity by creating a loop to stablize the telomeres; Deletion led to early embryonic lethality (Martinez and Blasco, 2010). Believed to be a regulator for telomere length by controlling the access of telomerases to the repetitive ends through the interaction with TRF1; Mice deficient in this gene exhbit embryo lethality; Overexpressed in murine tumors (Blasco, 2005).
8	POT1	Part of the shelterin complex that is responsible for telomerase production and activity by creating a loop to stablize the telomeres (Martinez and Blasco, 2010). Interacts with TRF1 and PTOP to regulate the access of telomerase to telomeres and associated with its capping; Overexpressed in murine tumors (Blasco, 2005).
9	RAP1	Part of the shelterin complex that is responsible for telomerase production and activity by creating a loop to stablize the telomeres; Plays a role in protecting from telomere recombination and combats fragility of the complex (Martinez and Blasco, 2010).
10	PTOP	Part of the shelterin complex that is responsible for telomerase production and activity by creating a loop to stablize the telomeres; Deletion directly affected the presence of POT1 at telomeres and results in excess of single-stranded telomere DNA sequences (Martinez and Blasco, 2010). Deletion of this gene resulted in the failed elognation of telomeres when used in pluripotency experiment involving the core reprogramming factors that induce pluripotency (Tejera et al., 2010).

Sirtuins

11	Sir2	Protein found in yeast [Sir2], nematodes [Sir-2.1], and fruit flies [dSir2]; Overexpression leads to the decrease in histone acetylation activity and an increase in life span in all three taxa; Compounds that activiate this gene was shown to promote longevity in taxa ranging from yeast to the mouse through its involvement in caloric restriction (Dali-Youcef et al., 2007).
12	Sir2.1	
13	dSir2	
14	SIRT1	Closest homolog to invertegrate Sir2 gene in mammals; overexpression improves aspect of bodily function though enhacements suchs as genomic stability and increased metabolic efficiency but did not increase lifespan (Herranz et al., 2010). Plays a role in stem cell functionality through controlling cell fate pathways and/or regulating p53 independent expression of Nanog; downregulated during human embryonic stem cell differentiation (Calvanese et al., 2010). Studies of mice deficient and overexpressive of this gene portrayed that it the gene regulates the shortening of telomeres as animals age (Park et al., 2013).
15	SIRT2	Downregulated/deleted in one of the most frequent and malignant brain tumors in humans called gliomas (Hiratsuka et al., 2003). Thought to be a potential tumor surpressor as it plays a critical role in cell cycle progression; Specifically it plays a role as the mitotic checkpoint and prevents chromosome instability (Dali-Youcef et al., 2007).
16	SIRT3	Mitochondrial mammalian sirtuin that increases the deacetylation of proteins and is linked to longevity in murine models; Reduction of reactive oxygen species in response to stress (Dali-Youcef et al., 2007). Have been found highly expressed in long living individuals (Park et al., 2013).
17	SIRT6	Suggested to regulate the stability of genomic DNA and its repair; Believed to play an essential role in maintaining organ functionality as animals age; Deletion in murine models exhbit premature defects and anomolies in test subjects (Dali-Youcef et al., 2007). Only sirtuin isoform that was able to extend the lifespan in mammals, and is beleived to play a similar role to SIRT1 (Park et al., 2013).

Other aging/longevity genes

27, 28	HSF-1	Master regulator of the heat-shock response that increases the lifespan and tolerance against thermal stress in nematodes (Chiang et al., 2012). In mammals, this gene activates other heat shock proteins in conjunction with SIRT1 (Westerheide et al., 2009).
29	Daf-2	One of the key components that modulates aging/longevity in roundworms, aditionally a regulator of reproduction, lipid metabolism and entry into developmental diapause that arrests metabolism to ensure long term survival (Kimura et al., 1997).
30	Daf-16	Homolog of the mammalian FOXO gene; Increased expression increases the resistance to proteotoxic stress an extends lifepsan in roundworms (Vilchez et al., 2012).
31	MRT-2	Checkpoint protein required for germline immortality in nematodes; facilitator of telomerase and telomere replication as mutants will exhibit progressive shortening of telomeres (Ahmed and Hodgkin, 2000).