ASSESSING COMPARATIVE FUNCTIONS FOR THE BZIP TRANSCRIPTION FACTOR NFE2 IN *NEMATOSTELLA VECTENSIS*

by

Devin J. Clegg

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Biological Sciences

Charlotte

2020

Approved by:

_____
Dr. Adam Reitzel

_____
Dr. Richard Chi

_____
Dr. Andrew Truman

ABSTRACT

DEVIN J. CLEGG. Assessing comparative functions for the bZIP transcription factor NFE2 in *Nematostella vectensis*. (Under the direction of Dr. ADAM REITZEL)

The bZIP gene superfamily is a family of transcription factors that have gone through multiple duplication events within bilaterians, specifically in vertebrates. This has led to multiple paralogs in each bZIP family while early diverging phyla have a single ortholog. These duplication events have motivated questions if the paralogs evolved to have new functions (neofunctionalization) or evolved to split the functions of the ancestor ortholog (subfunctionalization). The transcription factor, Nuclear Factor Erythroid 2 (NFE2) duplicated within vertebrates into at least four paralogs of Nrf1, Nrf2, Nrf3, and Nf-e2. Of these paralogs Nrf2 is well studied and has been characterized to upregulate antioxidant genes to alleviate stress from oxidizing molecules such as reactive oxygen species in the cell. The other three paralogs are involved proteasomal activity, redox homeostasis, and developmental processes. Knowing these paralogs have differing roles in vertebrates, NFE2 needs to be studied in an organism that has a single ortholog to decipher between neofunctionalization vs. subfunctionalization. Cnidarians are an insightful group to study NFE2 due to being sister group to bilaterians and possess a single ortholog. Using the sea anemone *Nematostella vectensis* as a model my research was focused to provide information on the ancestral function of NFE2 to compare with vertebrate paralogs and orthologs present in other invertebrates. In this study, database, phylogenetic, and molecular approaches were used to identify domains within and establish an interactome to begin characterization of function of *Nematostella* NFE2. Domain comparisons in NFE2 throughout animals showed that cnidarians have core domains important for DNA binding but lack some that elicit protein-protein interactions. In studying what proteins interact with

Nematostella NFE2 I characterized an antibody specific to NFE2 to perform proteomic analysis. Of the resulting hundreds of characterized interactors, few were commonly characterized in vertebrates. From this proteomic analysis, cnidarian NFE2 may have a protein interactome unique in comparison to vertebrates, regulation mechanisms. Future research to characterize these proteins have been summarized to better understand the mechanisms of NFE2 regulation in animal evolution.

ACKNOWLEDGEMENTS

First I would I like to thank my committee members Dr. Adam Reitzel, Dr. Richard Chi, and Dr. Andrew Truman. Dr. Reitzel was helpful in guiding me throughout my thesis research. Dr. Reitzel taught me how to perform phylogenetic analyses and how to identify conservation within DNA and amino acid sequences. He also encouraged me to apply for grants and attend conferences to further myself as a scientist. Dr. Chi provided my first research opportunity with his lab that initiated my interest in research. Dr. Chi allowed me to intern in his lab with no experience in biology research. Dr. Chi taught me the basics of molecular biology and biochemistry. Dr. Truman was helpful for the mass-spectrometry analysis. Dr. Truman provided insights on preparing samples for mass-spec along with his connections with University of Chicago wade mass-spec analysis possible. Additionally, I would like to thank the NSF for awarding me with the Graduate Research Fellowship. This award provided opportunities to attend conferences and visit collaborating labs.

TABLE OF CONTENTS

LIST OF TABLES VIII
LIST OF FIGURES IX
LIST OF ABBREVIATIONS X
CHAPTER 1: INTRODUCTION 1
1.1. NFE2 Gene Duplication and Function in Vertebrates 2
1.2. NFE2 in Invertebrates 7
1.3. Studying NFE2 in a Cnidarian (*Nematostella vectensis)* 8
CHAPTER 2: MATERIALS AND METHODS 10
2.1. Identification of NFE2 Proteins in Diverse Cnidarians and other Animals 10
2.2. MEME Analysis 10
2.3. NFE2 and MAF Phylogenetic Analysis 10
2.4. Search for Keap1 in *Nematostella* 11
2.5. Identification of a Specific NFE2 Antibody for *Nematostella* 11
2.6. Preparation of *Nematostella* homogenates 13
2.7. Immunoprecipitation of NFE2 from *Nematostella* Homogenates 13
2.8. Mass-Spectrometry Analysis 14
2.9. NFE2 Expression under Oxidative Stress 15
2.10. Developmental Timing and Localization of NFE2 Expression 16
CHAPTER 3: RESULTS 17
3.1. Identification of NFE2 Proteins in Diverse Cnidarians and Other Species 17
3.2. NFE2 and MAF Phylogeny 18
3.3. Identification of Keap1 in *Nematostella* 20
3.4. Motif Analysis using MEME Suite 21
3.5. *Nematostella* NFE2 Mass-spectrometry 22
3.6. NFE2 expression under Oxidative Stress 23
3.7. Developmental Timing and Localization of NFE2 Expression 24
CHAPTER 4: DISCUSSION 26
4.1. Phylogenetic Analysis 26
4.2. MEME Suite Domain Analysis 27
4.3. Nematostella NFE2 29

4.4. NFE2 Mass-spec Analysis 31
4.5. Future Directions 34
REFERENCES 40
APPENDIX: SUPPLEMENTARY FILES 46

LIST OF TABLES

TABLE 1: NFE2 Proteins Identified in Diverse Cnidarians and Other Species

TABLE 2: Motifs Identified Using MEME Suite

TABLE 3: Domain Composition of NFE2 Proteins

TABLE 4: Binary Motif Comparison Within NEH1 domain

TABLE 5: Top 35 Identified Peptides by Mass-spectrometry Analysis

TABLE 6: Identified Peptides with >20 counts by Mass-spectrometry Analysis

TABLE 7: Identified Peptides Based on Literature

LIST OF FIGURES

FIGURE 1: Screening of NFE2 Antibodies

FIGURE 2: Immunoprecipitation of NFE2 from Nematostella Homogenates

FIGURE 3: NFE2 & Maf Proteins Phylogenic Tree

FIGURE 4: Keap1 & Kelch Domain Proteins Phylogenetic Tree

FIGURE 5: Domain Comparison of NFE2 Proteins

FIGURE 6: NFE2 Expression in Response to Oxidative Stress

FIGURE 7: NFE2 Expression During Regeneration Embryogenesis

LIST OF ABBREVIATIONS

PCR	polymerase chain reaction
bZIP	basic leucine zipper
NFE2	Nuclear Factor Erythroid 2
CNC	cap ’n’ collar
β-TrCP	beta-transducin repeat-containing protein
ERAD	Endoplasmic-Reticulum-Associated protein Degradation
ARE	Antioxidant Response Element
MEME	Multiple Em for Motif Elicitation
tBOOH	tert-Butyl hydroperoxide
DNA	deoxyribonucleic acid
Keap1	Kelch-like ECH-associated protein 1
SCF	Skp1-Cullin-F-box-protein
ER	endoplasmic reticulum

CHAPTER 1: INTRODUCTION

Gene families diversify through duplication and divergence over evolutionary time [1]. How these duplicated proteins acquire new functions and to what extent protein function is maintained despite duplication events is a central question in molecular evolution [2]. The basic leucine zipper family (bZIP) superfamily is a large group of transcription factors that originated before the last common ancestor of eukaryotes and future diversified during animal evolution due to frequent duplication of particular family members in distinct lineages/phyla. Recently, Jindrich and Degan [3] have shown that a rapid wave of bZIP duplication occurred in Bilateria, particularly in vertebrates following whole genome duplication events, such that these species have multiple copies of a bZIP family whereas early diverging lineages have only a single ortholog. Bilaterian bZIPs possess specific roles in tissues or gene networks, where paralogs can have differential expression as well as dimerization partners relative to one another [4]. A fundamental open question is to what extent these specific roles represent subfunctionalization of an ancestral protein with multiple functions vs. the acquisition of new functions (neofunctionalization) over evolutionary time following duplication events. Nuclear Factor Erythroid 2 (NFE2) is one bZIP that has emerged with a critical role in both cell physiology, disease, and development through studies in vertebrates and insects. This transcription factor is a central transcription factor in the anti-oxidant stress response and involved in aerobic respiration, embryogenesis, inflammation and carcinogenesis [5]. Vertebrate models have been shown to possess four homologs of NFE2, due to multiple rounds of whole genome duplication, whereas cnidarians and other invertebrates have only one copy of the gene [6]. Studying the sequence

diversity and function of NFE2 within cnidarians and comparing these results to bilaterian NFE2 will allow me to discern the question of subfunctionalization vs. acquisition of new functions.

1.1. NFE2 Gene Duplication and Function in Vertebrates

Like most other bZIP families, the ancestral NFE2 gene in the stem chordate was a part of the bZIP duplication event within vertebrates. NFE2 has been duplicated into at least four genes present in extant vertebrates; NRF1, NRF2, NRF3, and NF-E2. NFE2, as a member of the bZIP family of transcription factors, primarily functions to regulate transcription of genes responsible for antioxidant response, proteasome subunits, and globin expression [6,21,26,28,29]. Despite the duplication of these more than 500 million years ago, the paralogs retain conserved protein domains that are referred to as NEH domains. The domains NEH1, NEH3, and NEH6 are conserved across the four vertebrate paralogs. The NEH1 domain contains three regions within it, CNC, Basic, and ZIP. The CNC and basic regions are required for DNA binding. The sequence of the basic region is highly conserved and is mainly composed of arginine and lysine residues. The ZIP region is needed for dimerization with Mafβ proteins. This dimerization is required for NFE2 proteins to bind to DNA and for transactivation of genes. Since the NEH1 region is essential and highly conserved within several species, it is indicative of an ancestral gene and is very helpful for phylogenetic analysis. The NEH3 and NEH6 domains are important for transactivation. The NEH3 domain contains a VFLVPK domain that interacts with chromodomain helicase DNA binding protein 6 [6]. The NEH6 domain has the motif DSGIS. This motif is relatively conserved between the genes and is a phosphorylation site for glycogen synthase kinase [6]. Once phosphorylated, β-TrCP binds and signals for ubiquitination resulting in degradation [6,10]. So, the NEH6 domain serves as avenue for lowering the NFE2 proteins levels and thus their activity. There are other important domains that affect the roles,

target genes, and stability levels of these NFE2 genes. The domains are an ER binding domain, NEH2, NEH4, and NEH5 domain [6]. The ER binding domain sequesters the protein to the ER keeping it in an inactive state [24,27]. This leads to the protein ability to be degraded in ERAD. The NEH2 domain is responsible for binding to Keap1 which leads to ubiquitination and subsequent proteasome degradation. The NEH4 and NEH5 domains are transcriptional activation domains [6]. The combination of these domains in different NFE2 paralogs as well as differential expression in cells or tissue types, are what makes the four genes functionally distinct from one another.

Our understanding of the expression and function of the NFE2 (also called cap 'n' collar in insects) subfamily has largely derived from research in vertebrates, particularly mammals. Although there are multiple members of the NFE2 subfamily, most research to date in vertebrates has focused on just one: nuclear factor erythroid 2-related factor 2, or (NRF2) [7]. NRF2 possess a NEH2, NEH4, and NEH5 domain in addition of the three common domains. The most well-known function of Nrf2 is its ability to upregulate transcription of genes that build a defense to oxidative stressors. Nrf2 forms a heterodimer with small maf proteins and binds to antioxidant response element (ARE) within DNA [6]. This sequence is typically found upstream of antioxidant genes and phase II enzymes, which together function to remove toxic chemicals and reactive metabolites from the cell [6,7]. This cellular defense mechanism is helpful for organisms in changing environments, but mis regulation or aberrant functions of Nrf2 can lead to cancerous cells that have increased resistance to chemotherapies and increase proliferation [7]. Thus, tight regulation of Nrf2 is important for the balance of oxidative stress resistance and proliferation of cancerous cells.

Nrf2 protein levels are regulated in two different ways. Both molecular mechanisms lead to ubiquitin mediated proteasome degradation. The main regulation mechanism of Nrf2 is the Cul3-Rbx1 E3 ubiquitin complex with Keap1 being the substrate recognition unit of the complex. Keap1 binds to the DLG and ETGE motif within the NEH2 domain of Nrf2 [6]. This Keap1/Nrf2 interaction holds Nrf2 in the cytoplasm and inhibits transactivation activity. Cysteine residues within Keap1 can be either oxidized or reduced, leading to a conformation change disrupting Nrf2 binding [8]. Once dissociated from Keap1, Nrf2 can translocate to the nucleus and increase transcription of target genes. The kinase glycogen synthase kinase 3 (GSK3) is another avenue of Nrf2 regulation. GSK3 phosphorylation stimulates nuclear export of Nrf2 and ubiquitin mediated degradation by the βTrcP E3 ubiquitin ligase [9,10]. These pathways of regulation have been studied and characterized mainly in mammalian models. Studying Nrf2 regulation in non-mammalian or invertebrate models will assist in identifying the likely ancestral function and regulation Nrf2.

In zebrafish, Nrf2 functions in the same capacity as Nrf2 in mammals. Zebrafish Nrf2 dimerizes with small Maf proteins and binds to the ARE sequence, equivalent to mammals [6]. It has been shown the Nrf2 target genes are conserved in zebrafish as well [6]. This conservation suggests Nrf2 functions to relieve oxidative stress in zebrafish. Also, the Keap1 regulation system of Nrf2 is present within zebrafish [6].

The protein Nrf1, or nuclear factor erythroid 2-related factor 1, has NEH1, NEH3, and NEH6 as well as an ER binding, NEH2, NEH4, and NEH5 domain. Thus, Nrf1 is similar in sequence to Nrf2 with the addition of the ER binding domain being the exception. It has been shown that Nrf1 elevates the expression of proteasome subunits in response to the inhibition of proteasomes, called ‘bounce-back’ [21,22]. Nrf1 activates transcription of proteasome subunits

through antioxidant response elements [21]. Radhakrisknan et al. have shown that Nrf1 is tethered to the membrane of the ER and then has to processed to its active form. The n-terminal of Nrf1 is embedded in the membrane of the ER with the c-terminal of the protein in the lumen [22]. The c-terminal of Nrf1 is then exposed to the cytosol of the cell via p97 to be either degraded or processed to its active form [22]. When proteasome activity is inhibited, Nrf1 that isn't degraded is processed by proteases to generate active Nrf1 that translocates to the nucleus to upregulate expression of proteasome subunits to increase proteasome activity [22]. Although the NEH2 domain is present in Nrf1, it has been demonstrated that Keap1 does not take part in Nrf1 degradation but is subjected to degradation in two different manners [23]. Nrf1 is either degraded in the nucleus or when the Nrf1 is exposed to the cytosol. Once Nrf1 has fulfilled its role in proteasome 'bounce-back', Nrf1 needs to be removed from the nucleus avoid proteasome hyperactivity. Inside the nucleus, β-TrCP interacts with DSGLS within the NEH6 domain in Nrf1 [23]. Resulting in ubiquitination and degradation of the active forms of Nrf1. Because Nrf1 is ER resident protein, Nrf1 is signaled for degradation by ERAD specific E3 ligases. HRD1, a part of ERAD E3 ligase, has been noted to be responsible for the degradation of Nrf1 [23,24]. So, Nrf1 maintains proteasome activity levels within cells to mitigate proteotoxic levels. Nrf1 initially resides in and being degraded in the ER until proteasome activity decreases. Then Nrf1 is able to be processed to its active form and elevate expression of proteasome subunits. Once proteasome activity has been reestablished, active forms of Nrf1 are degraded in the nucleus by β-TrCP E3 ligases.

Nuclear factor erythroid 2-related factor 3, or Nrf3, has an ER binding site in addition to the NEH6, NEH1, and NEH3 domains [6]. The ER binding domain is present suggesting that Nrf3 levels are regulated in the similar manner as Nrf1. The ER binding domain also suggests

that Nrf3 has to be proteolytically processed to become active. In Chowdhury et al., it was demonstrated that Nrf3 is degraded in the ER through ERAD via HRD1 and Nrf3 is degraded in the nucleus via β-TrCP [27]. Also, it was shown that Nrf3 needed processing by the same protease as Nrf1 [27]. In contrast though Nrf3 wouldn't be processed and accumulate into the nucleus in response to the same stimulus that triggers Nrf1 processing. So, even though the ER binding sequesters both Nrf1 and Nrf3 to the ER, Nrf3 targets a different set of genes. Some literature states that Nrf3 negatively antioxidant response [25,26]. It was shown that Nrf3 binds to the ARE of NQO1 and represses the expression of NQO1 in a concentration dependent manner [26]. NQO1, NAD(P)H:quinone oxidoreductase 1, is one of several antioxidant genes upregulated in response to oxidative stress. Since Nrf3 can repress gene expression through AREs and Nrf2 increase expression of genes though AREs, Nrf3 competes with Nrf2 to maintain a balance of oxidation and reduction.

NF-E2 only has the NEH5 domain in addition to the common domains. This paralog does not have the ER binding or NEH2 domains, so there are no known mechanisms of sequestering and degradation like the other three NFE2 proteins. NF-E2 does have NEH6 domain suggesting the possibility that NF-E2 is regulated in the nucleus in a similar manner. NF-E2 does have another phosphorylation site at n-terminus of the protein [28]. In the NEH5 transactivation domain, NF-E2 has been shown to be phosphorylated by protein kinase A but no direct effect has been observed [28]. NF-E2 is hypothesized to be important for erythroid, megakaryocytic gene expression, and platelet function genes [28,29].

Thus, the four paralogs of NFE2 are distinct in their overall domain composition and their primary target genes. Nrf2 and Nrf3 differ in they are sequestered and how activity is regulated, but they work in antagonist manner to each other to maintain a redox homeostasis

needed within cells. Nrf1 functions to alleviate proteotoxic stress within cells. NF-E2 plays a role in platelet function and erythroid expression. These proteins do have the highly conserved NEH1 domain that is also generally in invertebrate and non-bilaterian NFE2 proteins as well. Generally, the NFE2 paralogs are only seen in bilaterians, specifically vertebrates, while other species through the animal kingdom have one NFE2 gene. Using the highly conserved NEH1 will be useful to identify and compare NFE2 genes from multiple species to infer the ancestral proteins and its function.

1.2. NFE2 in Invertebrates

As discussed earlier, there were numerous duplication events within the bZIP protein family over family evolution [3]. Vertebrates have four or more paralogs of NFE2 while earlier diverging organisms possess only one gene [6]. Previous research has characterized function of NFE2 within *C. elegans* and *D. melanogaster*. The NFE2 ortholog in *C. elegans* is called Skn-1 and it functions in the same manner as vertebrate Nrf2 in response to oxidative stress. However, the molecular mechanisms for Skn-1's regulation and how it regulates transcription of downstream genes are very different. First, Skn-1 regulated transcription of other genes as a monomer because it has lost the leucine zipper domain that is responsible for maf heterodimerization [6]. Also, SKN-1 is missing the DLG and ETGE motifs for cytoplasmic degradation SKN-1 constitutively accumulates in the nucleus but is subjected to ubiquitin degradation by the Cul4-ddb1 E3 complex with Wdr23 recognizing Skn-1 [6,11] Skn-1 has been shown to be important for embryonic development [14].

In fruit flies and other studied arthropods, the NFE2 homolog is known as cap ‘n’ collar or Cnc [6]. There are multiple isoforms of Cnc in fruit flies due to splice variants. Cnc was initially characterized for its importance for development of the mouth portions of fruit flies, but

the isoform CncC was shown to function in the same manner in adult fruit flies as in mammals in response to oxidative stress [6]. CncC target genes were similar and the protein possesses the motif necessary for Keap1 interaction ubiquitin proteasome degradation [6].

To my knowledge, NFE2 genes have not been studied in other invertebrates. While Jindrich and Degnan have shown that NFE2 genes are present throughout the animal kingdom, with the possible exception of some ctenophores or sponges, the function and regulation of this important transcription factor remains uncharacterized. NFE2 has implications in development as well, because cnidarians having only one version of NFE2, studying NFE2 in cnidarians will give understanding to the evolution of function of NFE2.

1.3. Studying NFE2 in a Cnidarian (*Nematostella vectensis)*

Dunn et. al. [12] defined the term 'hidden biology' in animal evolution as the comparatively limited understanding of complex traits in non-bilaterian organisms due to a majority of work being focused on bilaterians. This biased view limits our ability to understand the evolution and function of particular proteins because when we use findings within bilateria to understand non bilaterian biology, we are biased to identify only similar characteristics between bilaterians and non-bilaterians, but miss characteristics, functions, and traits that are specific to non-bilaterians. This results in the inference that non bilaterians are simpler. This is misleading and results in gaps in understanding of early animal evolution and how protein function may have changed over deep evolutionary time.

Studying NFE2 (or any other protein) function within a phylum outside Bilateria will give insight into deeply conserved and novel features of during animal evolution. *Nematostella vectensis* is a useful model for studying the evolution of function of NFE2. *Nematostella*

possesses only one copy of NFE2. Also, *Nematostella* is a member of the phylum Cnidaria. This is an informative group for comparative research because cnidarians are a sister group to bilaterians and have a relatively conserved genome [13]. Using a cnidarian model to study the evolution of this protein as well as its function will give insight to the ancestral function of NFE2 in comparison to the highly divergent orthologs in other invertebrates (e.g. *C. elegans*) and the paralogs within vertebrates.

CHAPTER 2: MATERIALS AND METHODS

2.1. Identification of NFE2 Proteins in Diverse Cnidarians and other Animals

NFE2 proteins were identified in 14 cnidarian species through tBLASTn searches of publicly available transcriptomes (Table 1). The *Nematostella* NFE2 protein was used as the query sequence and the subject sequence with the lowest e-score was downloaded. Transcripts from each species with translated using NCBI's Open Reading Frame Finder and correct translation was confirmed through a BLASTp search at NCBI. This search was to verify that the protein was a bZIP protein. The orthology of each protein was determined through phylogenetic analyses (see 2.3. below).

2.2. MEME Analysis

MEME Suite was used to compare the conserved domains of *Nematostella* NFE2 to NFE2 in other species [20]. The conserved domains identified from this analysis can be used to determine the domain composition of the cnidarian NFE2 genes in general and the *Nematostella* NFE2 gene specifically. 55 NFE2 protein sequences from species of chordate (n=21), arthropods (n=4), annelides (n=2), nematodes (n=1), echinoderms (n=1), acoelomorphs (n=1), placozoa (n=1), brachiopods (n=1) molluscs (n=4), cnidarians (n=15), ctenophores (n=1), and sponges (n=3) were used for MEME analysis. The domains identified were matched with previously annotated functional domains. The presence of each domain was recorded with the corresponding e-values (Table 2). The domain composition for each NFE2 gene was recorded with the corresponding p-values (Table 3).

2.3. NFE2 and MAF Phylogenetic Analysis

For phylogenetic analysis, the 55 NFE2 sequences along with sequences from the MAF family (the sister bZIP family to NFE2) from Jindrich and Degnan [3] were aligned in MEGA v10.0.5 using MUSCLE. The alignment was then trimmed to the DNA binding and leucine zipper region. RAxML [19] was used for phylogenetic analyses using the LG model of protein evolution. Support for nodes in the phylogeny was determined with 1000 bootstraps.

2.4. Search for Keap1 in *Nematostella*

Doonan et al. (2019) recently published a phylogenetic analysis of the Kelch-domain proteins in a few selected cnidarians to determine if any of these species contained an ortholog to Keap1. They identified a Keap1 ortholog in three species from different taxonomic classes in the Cnidaria:: the anthozoan *Acropora digitifera*, the hydrozoan *Hydra vulgaris,* and the staurozoan *Clavadosia cruxmelitensis* but failed to identify one in cnidarians belonging to the Myxozoa (e.g. *Polypodium*), I searched for a Keap1 ortholog from *Nematostella* using BLASTp searches with the *Acropora* Keap1 to identify kelch-domain containing proteins at the JGI database. These proteins (n=62) were then aligned to Keap1 and closely related kelch-domain protein families (KLHL-27, 28, 29) from cnidarians and bilaterians (human, fly). Phylogenetic trees were constructed as described above using RAxML.

2.5. Identification of a Specific NFE2 Antibody for *Nematostella*

Antibodies specific to particular proteins remain uncommon for most non-model organisms, which limits the ability to study molecular mechanisms beyond transcription. Three peptide antigen polyclonal antibodies were designed to unique portions of the *Nematostella* NFE2 protein using GenScript's peptide design software. These antibodies were distributed across different parts of the protein at the following amino acid positions: Nrf_67 (67-82),

Nrf_111 (111-126), and Nrf_243 (243-268). I refer to the antibody names below based on the first position of the targeted sequence. Purified antibodies were delivered to UNC Charlotte and screened by Western blot to determine specificity.

The three antibodies were tested for effectiveness in extracts from e. coli overexpressing GST tagged NFE2 (~75 kDa) and endogenous *Nematostella* NFE2 (~50 kDa). The antibodies were tested at 1:100, 1:500, and 1:1000 dilutions. Out of the three antibodies two had good specificity and signal at a 1:1000 dilution (Nrf_111, Nrf_243). The Nrf_243 antibody was used for immunoprecipitation of NFE2 (see 2.7. below).

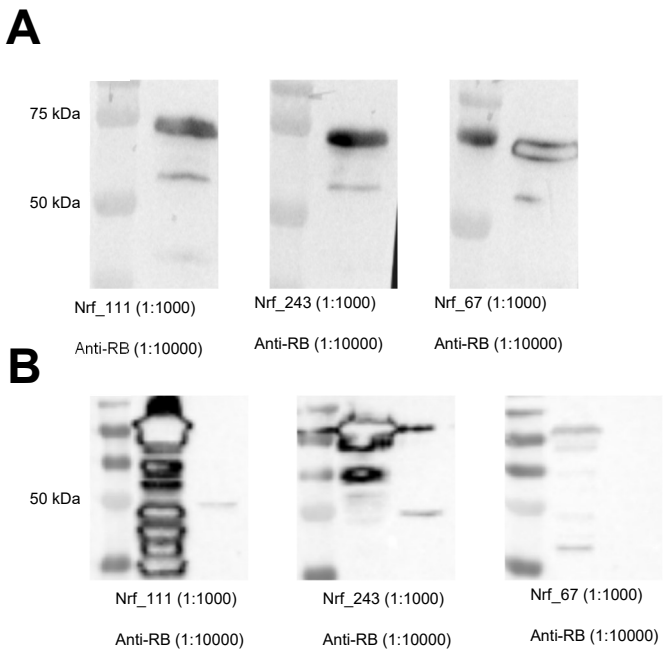

Figure 1. Antibody Screening of NFE2. **A.** Three antibodies tested on recombinantly expressed GST-NFE2 from *e. coli* **B.** Antibodies tested in *Nematostella* homogenates. Lane 1: pellet Lane 2: clarified lysate.

2.6. Preparation of *Nematostella* homogenates

Four *Nematostella* of each treatment group was collected into a tube. The two treatment groups were control (without antibody incubation) to remove non-specific interactors and the experimental group (with antibody incubation). *Nematostella* were then washed three times with artificial sea water and once with pure water. Then the anemones were transferred into 0.5 mL of AT lysis buffer (20mM Hepes, 150mM NaCl, 20% glycerol) supplemented with protease inhibitors (Pierce) [34]. The *Nematostella* were then pulverized using mortar and pestle and sonicated afterwards at 90% amplitude 3 times for 10 seconds each with 30 seconds rest on ice in between. The extracts were then centrifuged at 16,000 x g for 20 minutes at 4°C. The supernatant was transferred to a clean microcentrifuge tube and saved for immunoprecipitation of NFE2 from *Nematostella* homogenates and Western blot analysis.

2.7. Immunoprecipitation of NFE2 from *Nematostella* Homogenates

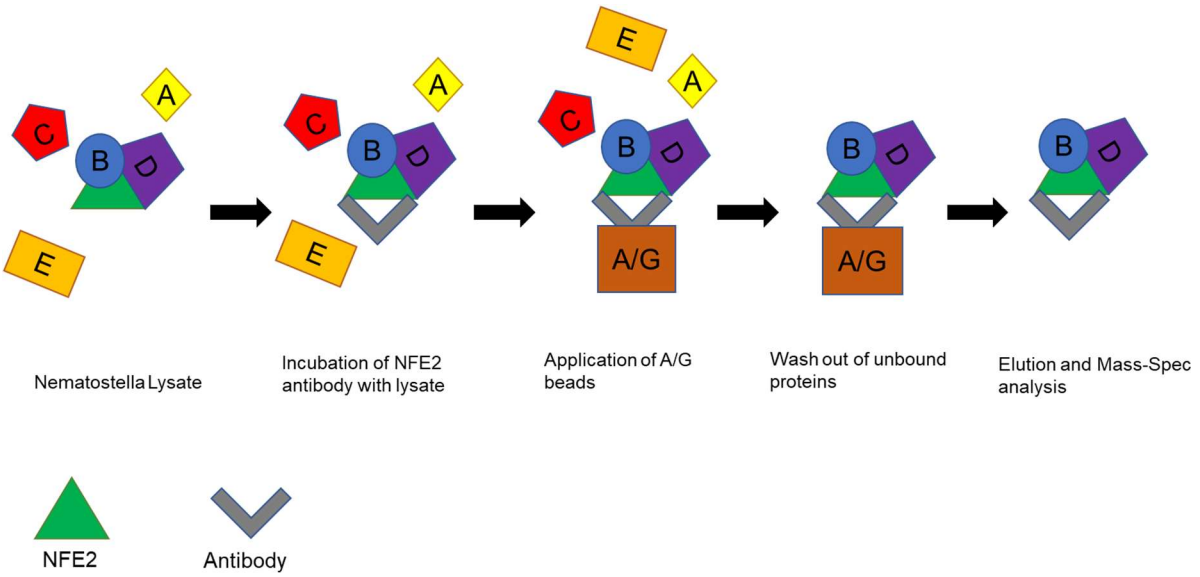

Figure 2. Schematic showing immunoprecipitation of NFE2 from *Nematostella* homogenates. This animation summarizes the strategy for using an antibody for immunoprecipitation of NFE2 from *Nematostella* homogenates. Non-interacting proteins (A, C, and E) will be washed away leaving interacting proteins (B and D) for mass-spectrometry.

For the immunoprecipitation of NFE2 we used the antibody NRF_243 due to its specificity. 50µg antibody was incubated in the prepared homogenates for one hour at 4°C. The prepared homogenate and antibody mixture were then incubated with magnetic Protein A/G beads for one hour at 4°C. After incubation, the beads were collected using a magnetic stand and the supernatant was discarded. The beads were the washed twice with a wash buffer (20mM Tris, 0.15M NaCl pH7.5) and then washed once with pure water. For elution, low-pH elution buffer (0.1M glycine pH 2.0) was added to the beads and was incubated at room temperature for ten minutes. The beads were separated from the supernatant using a magnetic stand. The supernatant was transferred to a clean microcentrifuge tube and neutralization buffer (1M Tris pH 8) was used to neutralize the pH. The supernatant was saved for mass-spectrometry to identify interactors of *Nematostella* NFE2.

2.8. Mass-Spectrometry Analysis

The super-natant from the triplicate samples from the two conditions were sent to the University of Chicago for mass spectrometry. The methods performed by the Mayo Clinic Proteomics Core facility for HPLC and LC-MS/MS data acquisition were essentially the same as reported in [35]. Briefly, the proteins in each homogenate were digested with trypsin, purified with HPLC using anOptiPak trapping cartridge custom packed with Michrom Magic, and then analyzed by electrospray tandem mass spectrometry (LC–MS/MS on a Thermo Q-Exactive Orbitrap mass spectrometer. All LC-MS/MS *.raw data files were analyzed with MaxQuant version 1.5.2.8, searching against the *Nematostella* proteome derived from the "Vienna Models" from *Nematostella* with a 1% FDR. The abundance data from each biological replicate were normalized median of the quantitative values and not a particular protein. Data were initially filtered at >= 2 experiments to be included and then missing data were imputed with the Perseus

algorithm (Supp. Figure 1, imputed data in red, original data in blue). Proteins were categorized as differentially abundant based on Log2 ratios for up (>= Log2 0.26) and down (<= Log2 -0.32) in experimental vs. control, equivalent to anything up or down 20% or more.

I used three approaches to identify potential proteins of interests from the mass-spec data. First, I selected the top 35 proteins based on the ratio of occurrence in the experimental (i.e. antibody incubated with anemone homogenates) vs. control (i.e., anemone homogenate without antibody incubation). Second, I selected protein based on the presence of more than 20 peptide fragments matched to the *Nematostella* reference proteins. The filtering step resulted in 37 proteins. Last, I surveyed the entire list of proteins to identify proteins that have previously been identified to be important for altering activity of NFE2 based on studies in other species. The proteins identified from each of my approaches were then cross referenced with a human Nrf2 interactome database, NRF-ome database [30,31].

2.9. NFE2 Expression under Oxidative Stress

Previous research has shown a diversity of responses of NFE2 abundance in response to oxidative stress. Here, I used Western blots to determine if NFE2 increased in abundance in adult Nematostella when exposed to a model pro-oxidant, tBOOH. There were experimental groups of anemones were control, 15µm tBOOH for 1hr (acute), and 15µm tBOOH for 24 hours (chronic). After exposure the anemones were homogenized as specified above in section 2.6. The anemone homogenates were used for Western blot analysis to compare NFE2 expression between the experimental groups. For analysis, the band intensity for each experimental group was measured in ImageJ. The measured band intensity for each group was tested in JMP Pro 15. A Dunnett's with control test was used to compare the mean band intensity of control group between the acute stress and chronic stress groups.

2.10. Developmental Timing and Localization of NFE2 Expression

In order to understand when during development and in what cells or tissues NFE2 is expressed, I used a combination of database approaches to query data from previous gene expression studies. To determine expression during embryogenesis and regeneration. I located and downloaded expression data from NvERTx [15]. This database has combined data from former transcriptomics studies on expression of all genes during development from fertilization (0 hours post-fertilization) to an adult polyp stage (192 hours post-fertilization). I also downloaded gene expression data available at NvERTx on regeneration. These data were collected from adult Nematostella that were bisected and sampled for gene expression for 144 hours post-amputation. To determine in what cell types NFE2 is expressed, I located and downloaded expression data of NFE2 from a former study by Sebe-Pedros et al. [32]. These data were generated by isolating cell types of *Nematostella* and then conducting transcriptome sequencing to qualify gene expression in each cell type. These data are summarized based on expression at two different developmental stages: larvae and adults. The cell type data are classified based on “meta-cells”, which are groups of single cell types based on correlated gene expression and separation on a cell sorter.

CHAPTER 3: RESULTS

3.1. Identification of NFE2 Proteins in Diverse Cnidarians and Other Species

Table 1. Table of NFE2 genes identified in cnidarians, mollusca, an acoelomorpha, an annelid and a brachiopod

Phylum	Species	Database	Gene ID
Annelida	Platynereis dumerilii	http://pdumbase.gdcb.iastate.edu	comp225037_c1_seq8
Brachiopoda	Lingula anatina	GenBank	XP_013384291.1
Cnidaria	Aurelia aurita	NCBI TSA	GBRG01161313.1
Cnidaria	Clytia hemispherica	http://marimba.obs-vlfr.fr/blast	TCONS_00021876-protein
Cnidaria	Hydractinia symbiolongicarpus	NCBI, TSA	GAWH01056891.1
Cnidaria	Alatina alata	NCBI: PUGI00000000 TSA	comp75450_c0_seq2
Cnidaria	Tripedalia cystophora	NCBI accession numbers SRR7791343-SRR7791345 and GGWE01000000, TSA	GHAQ01131580.1
Cnidaria	Exaptasia pallida	GenBank	XP_020903769.1
Cnidaria	Podocoryna carnea	NCBI, TSA	GBEH01051662.1
Cnidaria	Anthopleura elegantissima	NCBI, TSA	GBXJ01138408.1
Cnidaria	Polypodium hydriform	TSA	c21979_g1_i1
Cnidaria	Amenomia viridis	NCBI, TSA	GHCD01000975.1
Cnidaria	Actinia tenebrosa	GenBank	XP_031555350.1
Cnidaria	Orbicella faveolata	NCBI, TSA	XP_020623055.1
Cnidaria	Acropora millepora	NCBI, TSA	XP_029182828.1
Cnidaria	Pocillopora damicornis	NCBI, TSA	XP_027060548.1
Acoelomorpha	Hofstenia miamia	NCBI, TSA	GHHH01000876.1
Mollusca	Cristaria plicata	Genbank	AZM32563.1
Mollusca	Crassostrea virginica	GenBank	XP_022333013.1

NFE2 protein sequences were identified in 19 species. Of the 19 species, two were molluscs, one was an acoelomorpha, one brachiopod, one annelid, and 14 were cnidarians. 35 NFE2 sequences identified in Jindrich and Degnan [3]. and Fuse and Kobayashi [6] as well as *Nematostella* NFE2 from JGI was used in addition to the 19 newly identified sequences for analysis in the MEME suite and phylogenetic analysis.

3.2. NFE2 and MAF Phylogeny

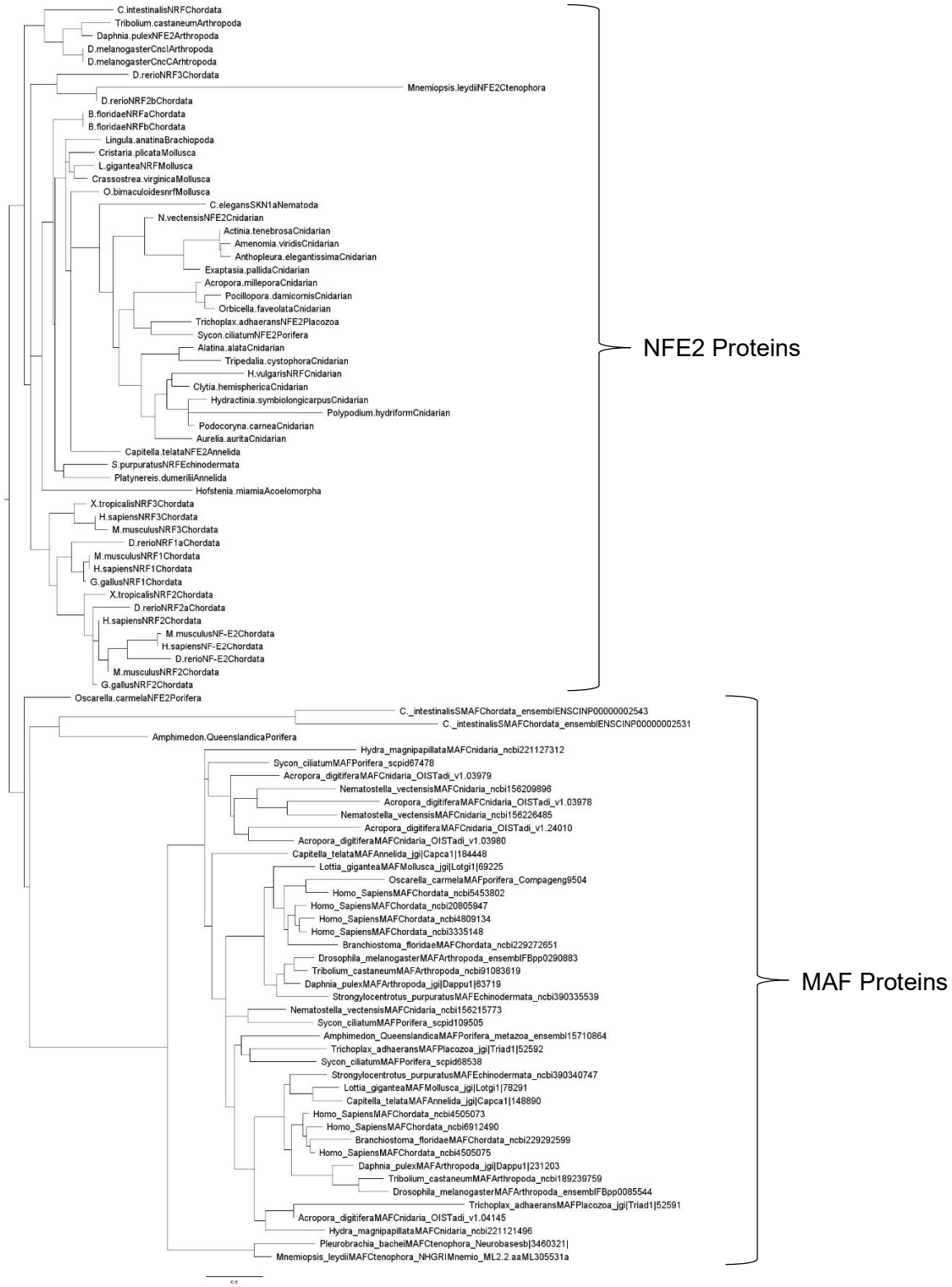

Figure 3. Maximum-likelihood phylogenetic tree of the NFE2 and MAF protein sequences.

The phylogenetic analysis was used in conjunction with the MEME suite analysis. Maf protein sequences pulled from Jindrich and Degnan et al. were used as an outgroup to decipher if the NFE2 sequences are using are actually NFE2 or better supported as a protein outside of the NFE2 family. Of the NFE2 sequences, all cnidarian protein grouped with NFE2 proteins from bilaterians and are supported as NFE2 proteins. The proteins of two sponges, *A. queenslandica* and *O. Carmella* did not group with NFE2 proteins thus seem to be possible Maf proteins. The phylogeny shows the genes used in the MEME analysis are true NFE2 genes. This supports the MEME analysis that domains are being compared between conserved ancestral NFE2 genes (see section 3.4.).

3.3. Identification of Keap1 in *Nematostella*

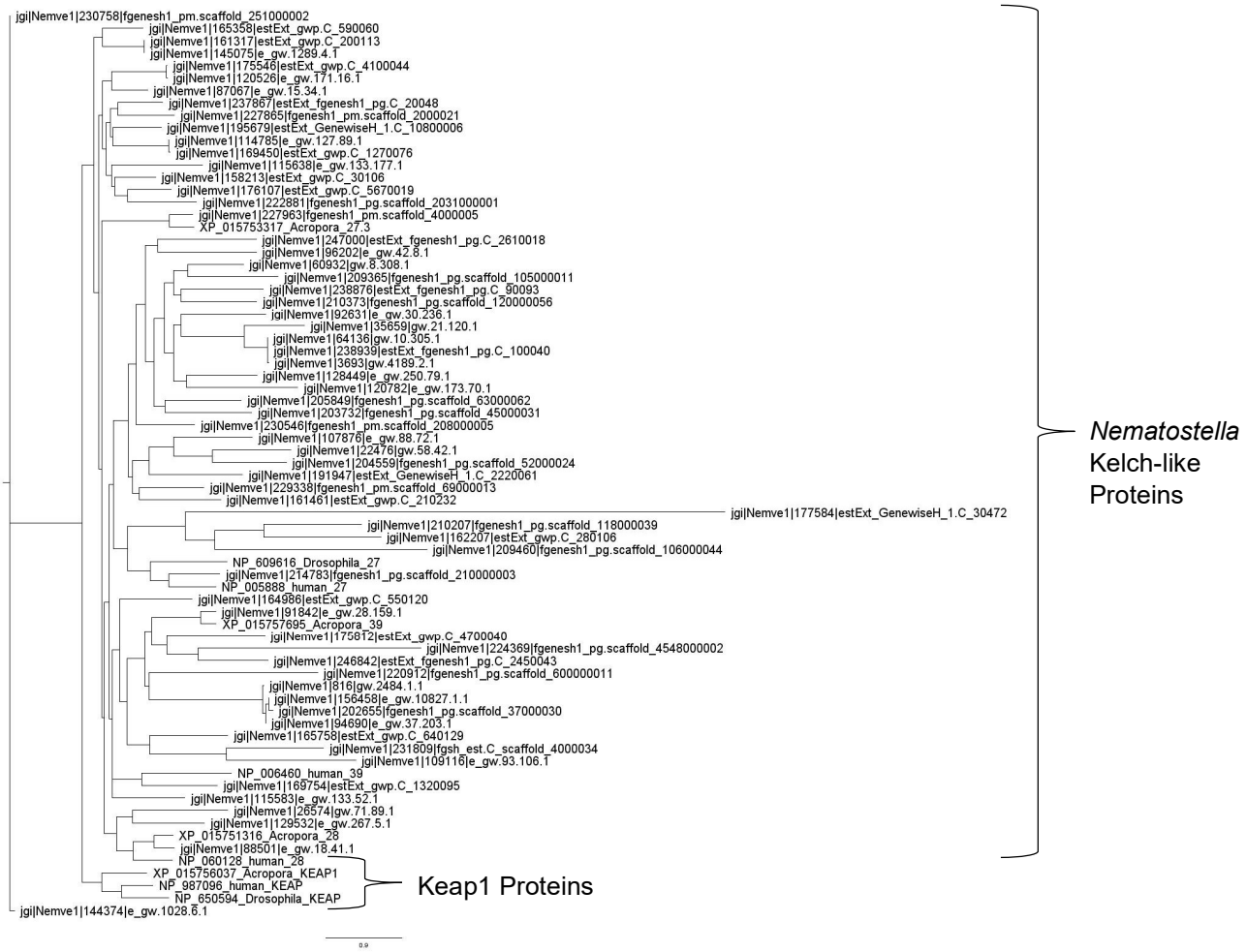

Figure 4. Maximum-likelihood phylogenetic tree of Kelch domain containing *Nematostella* proteins and human, Acropora (coral), and Drosophila Keap1 proteins.

To determine if *Nematostella* have a Keap1 ortholog, kelch domain containing proteins were used with Keap1 proteins to build a phylogenetic tree. The phylogenetic tree shows the Keap1 orthologs and kelch domain containing proteins are grouping separately from each other. With no protein showing orthology to Keap1, there is no support that *Nematostella* has a Keap1 ortholog. However, dozens of kelch domain containing proteins was detected and some grouped with high bootstrap support to other families (27, 28, 39). This result suggests *Nematostella* has a

diverse set of kelch domain containing proteins, but Keap1 was lost in the lineage leading to *Nematostella* because it is present in corals and other cnidarians.

3.4. Motif Analysis using MEME Suite

Table 2. Table of consensus motifs from MEME Suite analysis with corresponding E-values. The consensus motifs were match with functional domains known from literature.

MEME#	Motif (Consensus)	Domain	E-Value
MEME1	KLSEAQLQLIRDIRRRGKNKVAAQNCRKRKJDVIVTLEDEV	NEH1 (DNA binding)	7.8E-1413
MEME2	VFSSLRDENGRPYSPSEYSLQQT		7.0E-483
MEME3	EERLSRDEKRAKALKIPFSVDKIINLPVDEFNEMLSKY	NEH1 (CNC)	3.9E-596
MEME4	MDLIDILWRQDIDLGVGREVFDYSLRQKE	NEH2 (DLG)	4.20E-286
MEME5	SDGNVFLVPRNKK	NEH3	8.50E-133
MEME6	EHVGHNHTYPLPPGA		1.50E-113
MEME7	LKQEKEKLLKERGEJ	NEH1 (Leucine Zipper)	6.40E-116
MEME8	LEQTWQDLMSILELQ	NEH5	8.50E-94
MEME9	EGLLQLAJLLSLLRP	ER Binding	4.00E-77
MEME10	FPYSEDEJIEMPVEEFNEFJE	Majority Cnidarian genes	6.90E-89
MEME11	NYTLDGETGEYIPLQ	NEH2 (ETGE)	4.20E-81
MEME12	QFHNLRNTLDGYGIHPKSVDLDNYFTARRLLSQVRALD		7.30E-77
MEME13	KELREMKQKLSELYQ		2.20E-69
MEME14	NDSDSGLSLBSSHSPSSPSSS	NEH6	3.60E-62
MEME15	LPSPLSGLLDEALLDEISLMDLALEEGFN		1.90E-43
MEME16	VQVPATEVSAWLVHSDPDGAVSG		8.40E-40
MEME17	EKEREVZLQKEREK		1.50E-38
MEME18	YSQLPPLQEIILGQSSAYTQT		7.80E-41
MEME19	TLLYLPDFKTSPDGSDFQDSLEMELENKFFDPFTIDFGNTITNSSYQF		6.50E-38
MEME20	DALSFDECMQLLAETFPFGEDNE	NEH4	8.10E-38

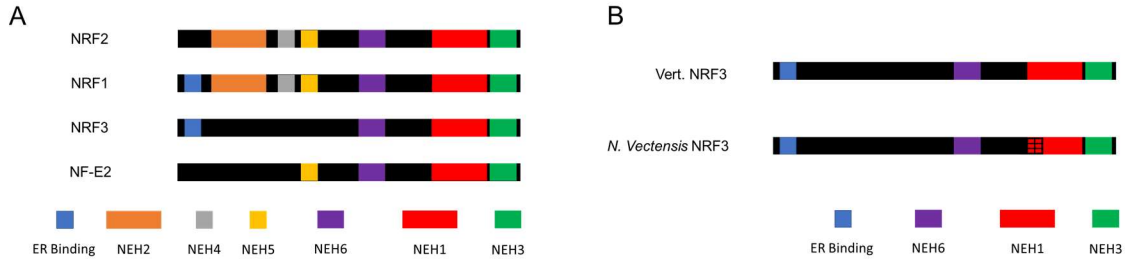

Figure 5. A. Domain composition of vertebrate NFE2 paralogs. B. Comparison of vertebrate Nrf3 to *Nematostella* NFE2 with slightly different NEH1 due to having MEME 10 instead of CNC region.

After analysis using MEME suite, 20 statistically significant motifs were identified from the 55 NFE2 sequences. Within the twenty motifs identified they were matched with characterized functional domains of NFE2 proteins. One motif (MEME 1) was present in all 55 sequences with the consensus sequence being

(KLSEAQLQLIRDIRRRGKNKVAAQNCRKRKJDVIVTLEDEV). This motif was identified as the highly conserved DNA biding domain, NEH1. The second most prevalent motif, MEME 7, was present in 53 of the 55 sequences (LKQEKEKLLKERGEJ) was the leucine zipper region that is responsible for heterodimerization with small Maf proteins. MEMEs 4 and 11 make up the NEH2 domain that is responsible for Keap1 binding. The NEH6 domain was represented as MEME14 with the important DSGLS sequence. Also MEME9 was recognized as the ER binding domain present in Nrf1 and Nrf3 proteins. MEME10, FPYSEDEJIEMPVEEFNEFJE, was present in early diverging lineages and if the species NFE2 gene possessed this MEME, it lacked MEME3, the CnC NEH1. MEME suite identified eight of the twenty motifs within the *Nematostella* NFE2 sequence with a combined p-value of 6.63E-74. The identified in *Nematostella* NFE2 were MEMEs 1, 2, 5, 7, 9, 10, 13, and 14.

3.5. *Nematostella* NFE2 Mass-spectrometry

From the mass-spec analysis of the immunoprecipitated NFE2 from *Nematostella* extracts there were 888 interacting proteins identified. Overall, only the samples where the antibody was used for the isolation had consistent hits of the mass-spec data to the reference proteome (Supp. Figure 2). Of the almost 900 proteins, a large majority (n=788) were only present in the homogenates where the antibody was used and 27 were found in both the experimental and control samples (Supp. Figure 3). Proteins sequences for all proteins identified were then used in BLASTp at NCBI to determine the interactors identity and function. Of the top 35 hits one of these proteins suggested a potential functional interaction, A-kinase anchor protein 9. The presence of this interacting protein with *Nematostella* NFE2 may suggest NFE2 is phosphorylated at its n-terminus like the vertebrate paralog NF-E2. A few other interesting interactors identified through mass-spec were F-box protein 9 and Spry domain containing SOCS

box protein 3 (SSB3). These two proteins are a part of ubiquitin ligase complexes that signal substrates for proteasomal degradation. Additional potential protein interactors of NFE2 identified were SSR1 and the chaperone BiP. These proteins are ER resident proteins.

I also selected proteins identified in the mass-spec data that had at least 20 matches in an effort to identify if any proteins have These 38 proteins were then given a best match identify to human proteins using BLASTp (Table 6). The human protein was then used a query sequence to the NRF-ome database, which is a catalog of NRF interactions taken from curated data from human NFE2-proteins (website: http://nrf2.elte.hu/). While a majority of the *Nematostella* proteins had high similarity to one, or in the case of myosin, tubulin, and collagen, many matches, few of these proteins have annotated direct or indirect interactions with NFE2 paralogs in human. In fact, the only protein with any confirmed direction interaction was HSP90B1 that inhibits Keap1, but Keap1 is not present in *Nematostella* (see above). Together, this approach did not suggest any clear interaction candidate with NFE2 likely to be conserved between the anemone and human.

3.6. NFE2 expression under Oxidative Stress

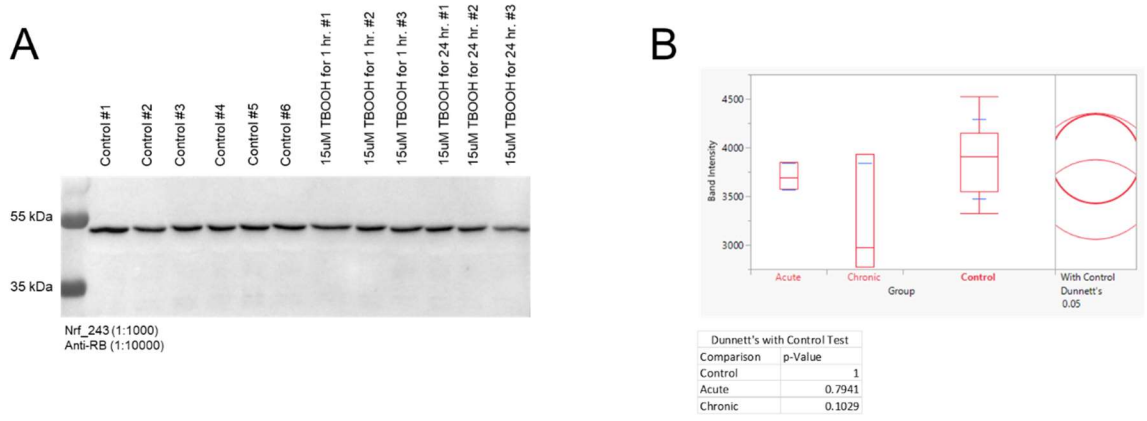

Figure 6. A. Western blot image of *Nematostella* NFE2 from anemone homogenates at different stress levels; Control, Acute (15µm TBOOH for 1 hr.), Chronic (15µm TBOOH for 24 hr.) B. Box plot and Dunnett's with control of band intensities

ImageJ was used to measure the band intensity of NFE2 from each experimental group. The values were recorded in JMP Pro 15. The average band intensity for each experimental group was compared to the control using a Dunnett's with control test. For comparison of control group mean with the acute stress group mean the p value was 0.7941. For comparison of control group mean with the chronic stress group mean the p value was 0.1029. There was no significant change in NFE2 protein abundance when anemones were exposed to either acute or chronic stress.

3.7. Developmental Timing and Localization of NFE2 Expression

I used two different databases, NvERTX for developmental time courses data and the single-cell data from two stages (larva, adult) to compare the transcription of NFE2 from *Nematostella*. From the NvERTX data, NFE2 was differentially expressed during embryogenesis and regeneration. In both developmental contexts, NFE2 had higher expression early in development. For embryogenesis, the highest expression peaked during hours 2-9 post-fertilization, which coincides with the early cell divisions to form the blastula. For the expression separated by cell type, NFE2 was expressed in all cell types (Supp. Table 2, e.g., neurons, gland cells, cnidocytes). Expression data from the study that generated these data used 'metacells' to classify cell types that differed in morphology and gene expression. In the larval stage, NFE2 had highest relative expression in two populations of neurons (metacells 28 and 31). Similarly, in the adult stage NFE2 showed highest relative expression in a subset of neuron cells (metacell 55, 36 to a lesser extent) but also had relatively high expression in a few subsets of gland secretory cells (metacells 86 and 97).

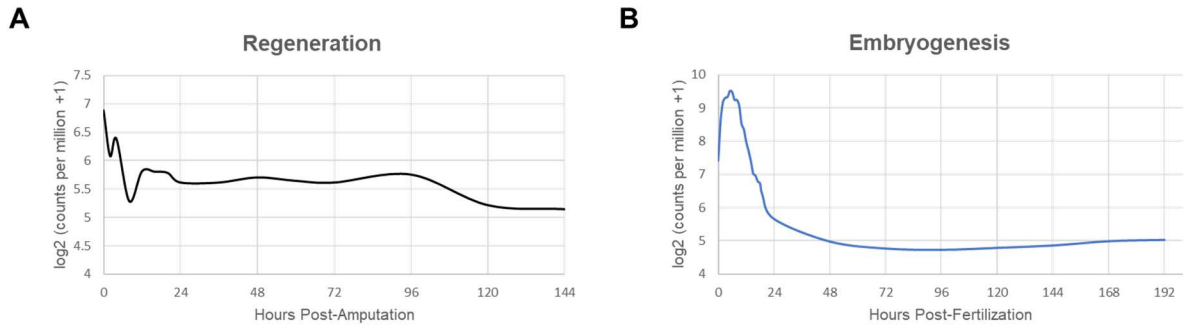

Figure 7. NFE2 Expression in *Nematostella* A. NFE2 expression profile up to 144 hours after amputation. B. NFE2 expression profile up to 192 hours after egg fertilization.

CHAPTER 4: DISCUSSION

Previous research in other animals has shown that Nrf2 and other NFE2-related transcription factors function in oxidative stress response and development. In vertebrates, NFE2 has duplicated into at least four copies with evidence of subfunctionalization. Using *Nematostella* as a model is beneficial because they have one NFE2 gene. Also, *Nematostella* is a member of the cnidarians, thus their phylogenetic position can give insight into the evolution of protein function. The combination of phylogenetic and molecular approaches has been helpful in understanding the function of NFE2 in *Nematostella*. Phylogenetic analysis along with MEME suite have determined the domain composition of NFE2 as well as shown that *Nematostella* do not have a Keap1 ortholog. Also, the mass-spec analysis has established an interactome for NFE2 in unstressed *Nematostella* that provides a valuable initiation point for later oxidative stress experiments.

4.1. Phylogenetic Analysis

After NFE2 proteins were initially identified through tBLASTn, a phylogenetic tree was produced to determine if the protein identified were a part of the NFE2 family. Maf amino acid sequences, from Degnan and Jindrich, were used as an outgroup to confirm NFE2 identity. Maf proteins are a part of the bZIP family, so using as an outgroup to differentiate proteins as NFE2. All proteins identified was indeed NFE2 and was used in MEME suite to observe domain conservation throughout al of the sequences.

Knowing Keap1 is important for control of Nrf2 activity, a Keap1 ortholog within *Nematostella* would be informative on NFE2 function. The kelch domain containing proteins that were identified from the *Nematostella* JGI database were used to build a phylogeny with

human, fly, and *Acropora* Keap1 sequences to identify a possible ortholog in *Nematostella*. None of the kelch domain containing proteins identified showed orthology to the known Keap1 proteins. This indicates that *Nematostella* does not have a Keap1-like protein and that NFE2 possibly does not function to alleviate oxidative stress in *Nematostella* through a similar protein-protein interaction known from bilaterians.

4.2. MEME Suite Domain Analysis

Comparing domain conservation is one way to characterize function of NFE2 or any protein. To identify conserved domains in *Nematostella*, NFE2 proteins from vertebrates along with confirmed orthologs were used in MEME suite. Through MEME suite 20 motifs were identified to have conservation. Of the 20 motifs identified by MEME suite, nine can be recognized as part of characterized domains found in all NFE2 proteins. MEMEs 1, 3, and 7 make up the NEH1 domain. This domain is present in all of these transcription factors because NEH1 is responsible for DNA binding. MEME 1 and MEME 3 are the basic and CNC regions respectively. These are essential for DNA binding, with the basic region, MEME 1, being highly conserved across species [6], which explains MEME1 prevalence and highly significant e-value. MEME 3 representing the CNC region of NEH1 was present Chordata, Arthropoda, Echinodermata, Acoelomorpha and Mollusca NFE2 proteins, but was missing in Cnidarian, Porifera, Placozoa, and Ctenophora NFE2 proteins. These non-bilaterian NFE2 genes missing MEME3 had MEME10 instead. MEME 7 is the leucine zipper or ZIP region of the NEH1 domain. This region is responsible for dimerization with small MAF proteins. MEME 7 was identified in 53 of the 55 NFE2 sequences except for *C. elegans* homolog SKN-1 and the parasitic cnidarian *P. hydriform*. In previous studies, SKN-1 is known to have lost this region and has lost the ability to dimerize with MAF proteins [6,11]. So, SKN-1 is known to bind to DNA as a monomer. The NEH2

domain has the sites responsible for KEAP1 binding, DLG and ETGE, and lysine residues for ubiquitination. This domain is represented by MEMEs 4 and 11. MEME 4 has the DLG motif and MEME 11 has the ETGE motif. These two MEMEs were found in all of the vertebrate NRF2 and NRF1 proteins along with Arthropoda, Brachiopoda, Mollusca, and Echinodermata NFE2 proteins. This domain distribution suggests the protein levels of these genes may be controlled in the similar manner. MEME 9 represents the ER binding region seen in NRF1 and NRF3 proteins. The ER binding region isn't found within NRF2 or NF-E2 proteins. MEME 9 was present in majority of cnidarian species tested including *Nematostella*, *Drosophila,* and Mollusca. This region has importance for sequestering NRF1 and NRF3 to the ER until activated. Proteins that have this region are localized to the ER in an inactive state being degraded until stressors activate them. Once certain conditions are met the ER membrane bound NFE2 protein are then cleaved at the n-terminus freeing the transcription factor and activating it. For invertebrate species whose NFE2 protein has this domain NFE2 is likely inactivated on the ER membrane under normal conditions and then when active a smaller NFE2 is translocated to the nucleus. The NEH3 is located at the c-terminus of NFE2 genes. This domain is important for the transactivation of through the ARE response element for NRF2. This domain is represented by MEME 5. This MEME was found in 39 of the NFE2 protein sequences and there is not a pattern of which protein has it or not. Although not all proteins possess this domain, this does not insist that the rest are non-functional proteins. SKN-1 is missing this domain along with the ZIP region but functions properly for transactivation. NEH4 and NEH5 are transcriptional activation domains that are only present in deuterostomes such as chordates for example [6]. Through this MEME analysis, MEMEs 8 and 20 represent these domains and as expected protostomes such as cnidarians, ctenophores, and sponges are lacking these domains. Lacking these transcriptional

activation domains does not indicate that protostome NFE2 genes cannot function as transcription factors. NRF3 and NF-E2 genes lack NEH4 domains but still function properly as transcription factors. The NEH6 domain is represented by MEME 14 in the MEME Suite analysis. MEME 14 was found in the vertebrate NFE2 proteins and in three cnidarian species; *N. vectensis, P. hydriform, and A. tenebrosa*. The NEH6 domain is responsible for the phosphorylation needed to signal nuclear NFE2 to be degraded to reduce activity. Being that this domain was identified in three cnidarian species outside of the bilaterian NFE2 proteins,

The NEH1 domain is the most important domain within this protein family. The three regions CNC, Basic, and Zip were identified as three separate motifs MEME's 1, 3, and 7 respectively. The Basic and Zip regions were used to identify possible NFE2 genes in other organisms because of the high conservation of these regions. The Basic region is the part of NFE2 that binds DNA and the Zip region is where NFE2 dimerizes with Maf proteins to be fully active. There's not much known on the importance of the CNC region of the NEH1 domain. Through MEME analysis the CNC region was not identified in cnidarians, ctenophores, and some porifera. MEME 10 was identified (FPYSEDEJIEMPVEEFNEFJE) instead of the CNC region. This implies the CNC region was added within and the MEME 10 region was lost before the bZIP duplication event in bilaterians. Since CNC regions importance for NFE2 function is not known and NFE2 is not well studied in early diverging species, we cannot deduce function for how NFE2 is potentially regulated with either region.

4.3. Nematostella NFE2

Domain Composition of *Nematostella* NFE2

MEME suite analysis was helpful in determining domain composition of NFE2. Through MEME, an ER binding, NEH1, NEH6, and NEH3 domains were identified. The NEH1 domain is different in comparison to vertebrate NFE2 proteins. The NEH1 domain was identified as three separate MEMEs. MEMEs 1, 3, and 7 correspond to the DNA binding, CNC, and leucine zipper region of the NEH1 domain. These memes were identified in vertebrate, chordate, arthropod, molluscs, annelid, and echinoderm NFE2 proteins. In cnidarian NFE2 proteins including Nematostella have MEMEs 1 and 7 but have MEME 10 (FPYSEDEJIEMPVEEFNEFJE) instead of MEME 3. All the organism NFE2 proteins besides *C. elegans* and *S. ciliatum* have either one or the other. This does not suggest the proteins that possess MEME 10 are not NFE2. The highly conserved DNA binding region, MEME1, is still there, so these are within the NFE2 protein family. This is reiterated in the phylogenetic tree showing the proteins are indeed NFE2 protein and that there is a shared ancestor. Since the NEH2 domain was not identified in Nv NFE2, it is doubtful that Keap1 can sequester NFE2 from the nucleus. Also, as summarized above. *Nematostella* lacks a Keap1 orthology, although it is conceivable one of the other kelch-domain containing proteins could perform a similar function. However, no kelch-domain proteins were detected in the proteomics data (see below). The ER binding domain is conserved in *Nematostella* NFE2. This provides one mechanism how NFE2 is sequestered and activity levels are controlled within this anemone. This model would then suggest that Nv NFE2 is sequestered in the ER and subjected to degradation through ERAD and there is a second proteolytically processed NFE2 that is active inside the nucleus. The NEH6 domain was also identified in *Nematostella* NFE2. It is known that this domain is responsible for lowering activity levels of Nrf1, Nrf2, Nrf3, and NFE2 within the nucleus. Within the domain is the conserved DSGLS sequence. This sequence is a phosphorylation site for glycogen synthase

kinase. Once phosphorylated, β -TrCP are able to bind to NFE2 proteins leading ubiquitination by the SCF ligase complex signaling for degradation of NFE2 proteins. This domain being identified in *Nematostella* NFE2 informs that nuclear NFE2 activity levels are controlled in a similar manner.

The combination of phylogenetic analysis and MEME suite have been informative of which vertebrate NFE2 protein *Nematostella* NFE2 is most similar to. With the ER binding, NEH1, NEH3, and NEH6 domain being identified, *Nematostella* NFE2 is most similar to Nrf3. *Nematostella* NFE2 possesses the same domain composition as Nrf3 with slightly different NEH1 domain. This gives direction into how NFE2 may be functioning in *Nematostella*.

To identify the function of NFE2 in *Nematostella* we need to understand how the expression levels are controlled. Even though it was shown that NFE2 does not have the NEH2 domain in *Nematostella* and that *Nematostella* does not have Keap1 ortholog, NFE2 can still have oxidative stress response. *A. millepora* has Keap1 even though NEH2 domain was not identified though MEME suite. Anemones were exposed to a chemical oxidant and NFE2 protein expression was measured. Since there was no significant change in NFE2 protein levels in stressed anemones compared to unstressed anemones, *Nematostella* NFE2 activity is not increased in response to oxidative stress. This suggests that the oxidative stress response studied in Nrf2 may be a secondarily evolved function of the NFE2 family.

4.4. NFE2 Mass-spec Analysis

Because transcription factors affect expression of a multitude of genes, tight regulation is needed prevent overactivity and underactivity. This mass-spec approach was performed with unstressed anemones to give a base for the interactome of NFE2 during normal conditions. Of

the peptides identified from the mass-spec three approaches were used to identify potential proteins of interest (section 2.8.). The peptides from each approach were then searched through the interactome of human Nrf2, “Nrf-ome”. The first approach to look at the top 35 peptides identified based on ratio of occurrence in the experimental group vs the control. Within these 35 peptides none were characterized to have a direct interaction with Nrf2. One peptide worth noting was A kinase anchoring protein 9. This is a scaffolding protein for protein kinase A. This is interesting because it has been noted the vertebrate paralog NF-E2 can be phosphorylated at its n-terminus by protein kinase A. Protein kinase A is not present in the top 35 peptides, but the anchoring protein suggests a conserved phosphorylation site. No peptides from the second approach was found to interact with Nrf2 or interesting due to literature on the four paralogs. This is due to this approach looking at the peptides with the >20 counts and not considering if peptides had counts within the control experiment. Through the third approach I identified multiple peptides that may be of interest to study for interaction: ubiquitin, f-box protein 9, SplA ryanodine receptor domain and SOCS box containing 3 (SPSB3), the ER chaperone BiP, and SSR1. Ubiquitin’s presence in the data suggests that NEF2 is being signaled for degradation in *Nematostella*. F-box protein 9 and SPSB3 are interesting because both are the substrate recognition protein of ubiquitin ligase complexes just like Keap1 and β -TrCP. F-box protein 9 is a part of a SCF ubiquitin ligase complexes just like β -TrCP. Substrate recognition motifs from SCF ligase complexes bind to phosphorylated substrates. So, even though β -TrCP was not identified through the mass-spec data, a protein with a similar function was identified. The protein BiP and SSR1 prove to be very interesting and reassuring with the finding of the ER binding domain within *Nematostella* NFE2. The ER binding domain is a signal sequence that signals for proteins to be cotranslated within the ER. Having the ER resident chaperone BiP and

SSR1 that functions to keep proteins in the ER coincides with MEME suite identifying the ER binding domain within NFE2 for *Nematostella*.

Comparisons of transcription dynamics of NFE2 during development and in particular cell types suggests that NFE2 is differentially regulated in both contexts. NFE2 is expressed at relatively higher levels during embryogenesis during the first 24 hours post fertilization. At this time frame the embryo is undergoing cell division and differentiation to form the blastula. The analysis of cell type data suggests that NFE2 has highest relative expression in subsets of neuronal cells in both the larval and adults stages of the anemone. Previous research in vertebrates has shown that Nrf2 and potentially other NFE2 genes are important for development, particularly of the neurons. This synexpression of NFE2 and the duplicated NFE2/Nrf genes in vertebrates may suggest an important and conserved role for NFE2 in neural patterning and function.

The combination of database and molecular approaches have direction on NFE2 function within *Nematostella* and determining the evolution of NFE2. Phylogenetic analysis has shown that *Nematostella* do not have Keap1 or a Keap1-like protein. MEME suite showed that *Nematostella* NFE2 is most similar to Nrf3 due to its domain composition. Along with the NFE2 protein levels not being affected by oxidative stress suggests that NFE2 does not alleviate oxidative stress in *Nematostella* and the oxidative stress response in Nrf2 is possibly a result of neofunctionalization. Also, interactors such BiP and SSR1 were identified through mass-spec affirming the identification of the ER binding domain in NFE2.

4.5. Future Directions

The next steps for studying NFE2 in *Nematostella* will be to study how NFE2 is activated. Knowing that NFE2 is most similar to Nrf3 and that oxidative stress did not change protein levels, NFE2 most likely does not function to upregulate antioxidant genes. As of now the conditions needed to activate Nrf3 are unknown but it is known Nrf1 is activated when proteasomal activity levels are low. So, the next step would be to test if NFE2 is responsible for 'bounce back' as seen by Nrf1 in response to the inhibition of proteasome activity. Because NFE2 has an ER binding domain detection for NFE2 activation can be assayed because NFE2 will need to be processed by a protease to become unbound to the ER membrane and be active. Also, performing invitro pulldown assays to confirm interaction with some of the interesting protein identified through the mass-spec. F-box protein 9 and SPSB3 are a part of ubiquitin ligase complexes and thus testing these potential proteins would identify how NFE2 protein levels are regulated in *Nematostella.*

Another future direction steps in studying NFE2 in *Nematostella* will be to determine the expression and function of this transcription factor during development. It was discussed earlier that some NFE2 genes had roles in development. For example, CncC is important for development of the cranial portion of *D. melanogaster* [6]. Also, loss of function of Skn-1 results in embryonic lethality in *C. elegans* [14]. Through the NVERTx database, it has been shown that NFE2 levels are elevated after fertilization, peaking around 6 hours post fertilization [15]. This suggests that NFE2 could be important in early development, particularly during formation of the germ layers and body axes. The cell type data suggest part of this differentiation may be related to subsets of neurons. Using the approach proposed here through the life stages of *Nematostella* will assist in understanding NFE2's role in development. Also, knocking down NFE2 expression

can be useful in understanding NFE2 role in development. There is a recent paper showing that gene expression in *Nematostella* can be knocked down using shRNA [16]. Using this technique, phenotypic changes of development could be observed after effective shRNA knockdown of NFE2. Another avenue of studying NFE2 in *Nematostella* is immunohistochemistry (IHC). We already have an antibody against NFE2 available. Using this antibody with IHC, we will be able to see if NFE2 is tissue specific and the location of NFE2 within the cell during normal and stressed conditions. These approaches along with the approaches proposed here will be useful in uncovering the evolution of function of NFE2 in animal evolution.

Table 3. Domain composition of the NFE2 protein sequences from MEME Suite analysis

				Motifs Present (Y or N), P-value																	
	NFE2 Protein Name	Phylum	Gene	MEME1	MEME2	MEME3	MEME4	MEME5	MEME6	MEME7	MEME8	MEME9	MEME10	MEME11	MEME12	MEME13	MEME14	MEME15	MEME16	MEME17	MEME18	MEME19	MEME20	Combined P-value
1	H.sapiensNRF2Chordata (ENSG00000116044)	Chordata	NRF2	Y, 1.21E-42	Y, 2.84E-24	Y, 9.03E-36	Y, 5.99E-34	Y, 2.91E-13	N	Y, 4.76E-13	Y, 1.31E-14	N	N	Y, 6.8E-14	N	Y, 2.65E-9	Y, 1.13E-18	Y, 4E-19	N	Y, 5.05E-15	N	N	Y, 1.28E-24	1.03E-205
2	M.musculusNRF2Chordata (ENSMUSP00000099733)	Chordata	NRF2	Y, 1.21E-42	Y, 2.84E-24	Y, 2.78E-37	Y, 3.21E-33	Y, 2.91E-13	N	Y, 7.49E-13	Y, 9.93E-17	N	N	Y, 1.67E-14	N	Y, 1.69E-8	Y, 1.34E-18	Y, 1.28E-16	N	Y, 5.05E-15	N	N	Y, 2.23E-24	2.39E-205
3	G.gallusNRF2Chordata (ENSGALP00000032649)	Chordata	NRF2	Y, 6.87E-43	Y, 2.54E-22	Y, 2.78E-37	Y, 1.13E-31	Y, 1.38E-11	N	Y, 5.54E-13	Y, 7.23E-14	N	N	Y, 9.42E-14	N	Y, 1.69E-13	Y, 2.91E-17	Y, 5.91E-16	N	Y, 8.1E-16	N	N	Y, 1.23E-23	7.50E-200
4	X.tropicalisNRF2Chordata (ENSXETP00000039783)	Chordata	NRF2	Y, 1.73E-43	Y, 7.46E-21	Y, 1.08E-36	Y, 3.21E-33	Y, 6.84E-11	N	Y, 4.54E-12	Y, 4.15E-14	N	N	Y, 1.15E-14	N	Y, 1.06E-9	Y, 1.79E-14	Y, 2.57E-18	N	Y, 8.1E-16	N	N	Y, 4.99E-18	4.30E-191
5	D.rerioNRF2Chordata (ENSDARP00000062853)	Chordata	NRF2	Y, 1.76E-42	Y, 6.40E-19	Y, 3.32E-39	Y, 1.01E-33	Y, 8.85E-13	N	Y, 6.16E-11	Y, 1.24E-13	N	N	Y, 2.37E-12	N	Y, 1.99E-10	Y, 2.51E-15	N	N	N	N	N	Y, 1.49E-22	1.99E-171
6	D.rerioNRF2Chordata (ENSDARP00000106581)	Chordata	NRF2	Y, 9.42E-37	Y, 6.26e-16	Y, 1.08E-25	Y, 1.12E-25	Y, 5.59E-11	N	Y, 1.09E-8	Y, 1.12E-17	N	N	Y, 1.07E-10	N	Y, 4.94E-10	N	N	N	Y, 2.44E-9	N	N	N	1.14E-107
7	H.sapiensNRF1Chordata (ENSG00000182641)	Chordata	NRF1	Y, 1.32E-44	Y, 7.13E-23	Y, 8.57E-40	Y, 4.98E-35	Y, 1.49E-10	Y, 1.46E-15	Y, 3.96E-11	Y, 1.12E-17	Y, 7.14E-15	N	Y, 2.16E-13	Y, 2.83E-47	Y, 1.18E-13	Y, 1.06E-19	Y, 7.44E-31	Y, 3.44E-26	Y, 2.12E-12	Y, 3.1E-23	N	Y, 2.53E-20	0.00E+00
8	M.musculusNRF1Chordata (ENSMUSP00000080467)	Chordata	NRF1	Y, 1.32E-44	Y, 5.31E-23	Y, 8.57E-40	Y,4.98E-35	Y, 1.49E-10	Y, 1.46E-15	Y, 3.96E-11	Y, 1.12E-17	Y, 7.14E-15	N	Y, 8.25E-13	Y, 2.83E-47	Y, 1.18E-13	Y, 1.06E-19	Y, 7.44E-31	Y, 3.44E-26	Y, 2.12E-12	Y, 3.1E-23	N	N	0.00E+00
9	G.gallusNRF1Chordata (ENSGALP00000035179)	Chordata	NRF1	Y, 1.32E-44	Y, 3.33E-22	Y, 8.57E-40	Y, 8.47E-34	Y, 9.19E-11	Y, 1.81E-16	Y, 2.14E-10	Y, 2.35E-14	Y, 1.01E-14	N	Y, 2.16E-13	Y, 2.27E-48	Y, 9.99E-14	Y, 3.04E-18	Y, 7.44E-31	Y, 4.53E-26	Y, 2.12E-12	Y, 4.25E-23	N	Y, 1.22E-20	0.00E+00
10	D.rerioNRF1Chordata (ENSDARP00000094757)	Chordata	NRF1	Y, 1.62E-39	Y, 1.1E-13	Y, 1.52E-29	N	N	Y, 3.04E-13	Y, 2.36E-10	N	Y, 6.79E-14	N	N	Y, 3.41E-38	Y, 3.76E-9	Y, 2.2E-18	Y, 1.12E-20	Y, 1.23E-17	N	Y, 2.72E-19	N	N	4.18E-188
11	H.sapiensNRF3Chordata (ENSG00000050344)	Chordata	NRF3	Y, 4.36E-42	Y, 4.96E-20	Y, 8.21E-34	N	Y, 2.51E-9	Y, 8.85E-16	Y, 2.32E-9	N	Y, 2.59E-11	N	N	N	Y, 4.18E-12	Y, 7.22E-16	Y, 3.57E-16	Y, 2.19E-20	Y, 6.44E-9	N	N	N	6.69E-152
12	M.musculusNRF3Chordata (ENSMUSP00000005103)	Chordata	NRF3	Y, 1.47E-41	Y, 3.20E-19	Y, 7.29E-34	N	Y, 1.22E-9	Y, 6.57E-13	Y, 5.27E-9	N	Y, 9.46E-13	N	N	N	Y, 5.31E-9	Y, 1.05E-12	N	Y, 6.11E-21	N	N	N	N	8.69E-133
13	X.tropicalisNRF3Chordata (ENSXETP00000026569)	Chordata	NRF3	Y, 1.76E-42	Y, 2.75E-17	Y, 1.96E-31	N	Y, 6.77E-10	Y, 8.73E-13	Y, 7.57E-12	N	Y, 3.18E-14	N	N	Y, 3.19E-34	Y, 2.15E-11	Y, 5.66E-19	Y, 3.41E-18	Y, 1.07E-19	N	Y, 3.15E-18	N	N	2.16E-197
14	D.rerioNRF3Chordata (ENSDARP00000019027)	Chordata	NRF3	Y, 2.19E-33	Y, 5.02E-17	Y, 9.05E-25	N	Y, 3.99E-8	Y, 3.23E-11	Y, 2.05E-7	N	Y, 1.01E-14	N	N	Y, 1.29E-33	Y, 1.85E-9	Y, 5.34E-12	Y, 6.98E-16	Y, 3.9E-16	N	N	N	N	6.17E-138
15	H.sapiensNF-E2Chordata (ENSG00000123401)	Chordata	NF-E2	Y, 7.40E-42	Y, 4.67E-21	Y, 1.53E-35	N	Y, 3.41E-13	N	Y, 3.52E-9	Y, 3.07E-15	N	N	N	N	Y, 1.85E-9	Y, 2.7E-12	Y, 3.57E-16	N	N	N	N	N	1.10E-124
16	M.musculusNF-E2Chordata (ENSMUSP00000122476)	Chordata	NF-E2	Y, 7.40E-42	Y, 2.07E-20	Y, 3.78E-35	N	Y, 1.82E-12	N	Y, 4.49E-9	Y, 3.07E-15	N	N	N	N	Y, 1.20E-7	Y, 2.7E-12	Y, 6.36E-17	N	N	N	N	N	2.98E-123
17	D.rerioNF-E2Chordata (ENSDARP00000022745)	Chordata	NF-E2	Y, 5.57E-41	Y, 1.14E-18	Y, 5.27E-33	N	Y, 5.59E-11	N	Y, 1.96E-9	Y, 5.8E-14	N	N	N	N	N	Y, 1.14E-12	N	N	N	N	N	N	2.36E-102
18	D.melanogasterCncCArthropoda (NP_732833.1)	Arthropoda	CrC	Y, 1.73E-43	Y, 9.38E-20	Y, 1.51E-34	N	Y, 3.32E-11	Y, 1.28E-11	Y, 3.24E-9	N	Y, 1.35E-10	N	Y, 1.16E-13	N	N	N	N	Y, 1.44E-15	N	N	N	N	4.69E-118
19	D.melanogasterCncIArthropoda (NP_001247258.1)	Arthropoda	CrC	Y, 1.73E-43	Y, 9.38E-20	Y, 1.51E-34	N	Y, 3.32E-11	Y, 1.28E-11	Y, 3.24E-9	N	N	N	Y, 1.16E-13	N	N	N	N	Y, 1.44E-15	N	N	N	N	4.15E-113
20	C.elegansSKN1aNematoda (NP_741404.1)	Nematoda	SKN1	Y, 6.60E-28	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	6.73E-24
21	N.vectensisNFE2Cnidarian	Cnidarian	NFE2	Y, 5.10E-38	Y, 6.69E-18	N	N	Y, 1.56E-7	N	Y, 2.05E-7	N	Y, 4.08E-10	Y, 9.87E-16	N	N	Y, 2.89E-9	Y, 2.8E-11	N	N	N	N	N	N	6.63E-74
22	C.intestinalisNRFChordata (ENSCINP00000024999)	Chordata	NFE2	Y, 4.96E-40	Y, 1.52E-18	Y, 1.22E-27	N	N	Y, 7.88E-11	Y, 7.20E-7	N	Y, 3.48E-8	N	Y, 5.19E-13	N	Y, 8.37E-8	N	N	N	N	N	N	N	1.59E-94
23	H.vulgarisNRFCnidarian (XP_002160548.1)	Cnidarian	NFE2	Y, 7.55E-30	Y, 1.49E-17	N	N	Y, 8.31E-11	N	Y, 5.71E-9	N	N	Y, 7.28E-19	N	N	N	N	N	N	N	N	N	N	1.28E-58
24	S.purpuratusNRFEchinodermata (XP_011683763)	Echinodermata	NFE2	Y, 8.33E-37	Y, 1.10E-14	Y, 2.71E-29	Y, 1.16E-22	Y, 5.05E-11	Y, 2.73E-11	Y, 1.75E-10	Y, 2.64E-14	Y, 1.07E-9	N	Y, 2.64E-13	N	Y, 2.50E-8	N	N	N	N	N	N	N	3.40E-127
25	B.floridaeNRFaChordata (gi: 131476)	Chordata	NFE2	Y, 3.22E-44	Y, 1.27E-22	Y, 3.13E-30	Y, 2.36E-29	Y, 2.09E-12	Y, 4.7E-18	Y, 2.33E-12	Y, 3.73E-12	Y, 5.05E-14	N	Y, 6.09E-14	N	Y, 5.44E-12	Y, 2.15E-13	Y, 1.81E-16	N	Y, 3.76E-11	Y, 1.06E-16	N	Y, 6.12E-19	1.51E-214
26	B.floridaeNRFbChordata (gi: 127500)	Chordata	NFE2	Y, 3.22E-44	Y, 1.27E-22	Y, 3.13E-30	Y, 2.36E-29	Y, 2.09E-12	Y, 4.7E-18	Y, 2.33E-12	Y, 3.73E-12	Y, 5.05E-14	N	Y, 6.09E-14	N	Y, 5.44E-12	Y, 2.15E-13	Y, 1.81E-16	N	Y, 3.76E-11	Y, 1.06E-16	N	Y, 6.12E-19	1.55E-214
27	O.bimaculoidesnrfMollusca (Ocbimv22001599m)	Mollusca	NFE2	Y, 7.69E-41	Y, 7.46E-21	Y, 3.16E-25	Y, 3.36E-29	Y, 8.33E-11	Y, 6.02E-14	Y, 1.16E-9	N	Y, 2.59E-11	N	Y, 8.83E-15	N	Y, 7.59E-11	N	N	N	Y, 7.72E-14	Y, 9.78E-17	N	N	1.75E-145
28	L.giganteaNRFMollusca (LgGnFWreduced.7631)	Mollusca	NFE2	Y, 4.27E-40	Y, 2.22E-22	Y, 2.08E-27	Y, 1.86E-29	Y, 3.14E-12	Y, 4.07E-16	Y, 7.39E-10	Y, 1.14E-11	N	N	Y, 1.68E-17	N	Y, 1.17E-10	N	N	N	Y, 1.95E-12	N	N	N	1.13E-151
29	Capitella. telataNFE2Annelida (gi: 19335)	Annelida	NFE2	Y, 4.27E-40	Y, 1.26E-18	Y, 2.71E-27	Y, 3.4E-30	Y, 2.32E-9	Y, 8.83E-18	Y, 4.49E-9	N	N	N	Y, 1.97E-10	N	Y, 1.05E-10	N	N	N	Y, 3.77E-13	N	N	N	2.69E-133
30	Oscarella.carmelaNFE2Porifera (Compagen: g5281)	Porifera	NFE2	Y, 7.37E-37	N	N	N	N	N	Y, 1.92E-7	N	N	Y, 1.6E-13	N	N	Y, 1.83E-8	N	N	N	N	N	N	N	1.14E-41
31	Mnemiopsis.leydiiNFE2Ctenophora (NHGRI: Mnemio_ML2.2.aa: ML016353a)	Ctenophora	NFE2	Y, 1.48E-20	N	N	N	N	N	Y, 6.41E-7	N	N	Y, 3.26E-13	N	N	N	N	N	N	N	N	N	N	3.50E-22
32	Trichoplax.adhaerensNFE2Placozoa (gi: 60816)	Placozoa	NFE2	Y, 9.48E-34	N	N	N	N	N	Y, 1.03E-8	N	N	Y, 5.59E-15	N	N	N	N	N	N	N	N	N	N	6.76E-35
33	Sycon.ciliatumNFE2Porifera (Compagen: scpid35695)	Porifera	NFE2	Y, 3.50E-37	N	N	N	N	N	Y, 6.38E-8	N	N	N	N	N	N	N	N	N	N	N	N	N	6.29E-27
34	Daphnia.pulexNFE2Arthropoda (gi: 307821)	Arthropoda	NFE2	Y, 8.32E-43	Y, 1.57E-21	Y, 3.14E-34	N	Y, 1.73E-11	Y, 1.37E-13	Y, 1.67E-8	Y, 1.07E-12	N	N	N	N	N	N	Y, 5.79E-13	N	N	N	N	N	1.65E-114
35	Platynereis.dumeriiAnnelida (jgtumbase.gb3.iastate.edu: comp225037_c1_seq8)	Annelida	NFE2	Y, 2.88E-39	Y, 3.13E-22	Y, 5.43E-32	N	Y, 5.95E-12	Y, 2.36E-15	Y, 2.61E-10	N	N	N	N	N	Y, 8.01E-10	N	N	N	N	N	N	N	N	6.46E-105
36	Lingula.anatinaBrachiopoda (GenBank: XP_013384291.1)	Brachiopoda	NFE2	Y, 6.64E-45	Y, 1.07E-25	Y, 1.3E-32	Y, 9.77E-28	Y, 4.06E-13	Y, 2.49E-18	Y, 4.23E-10	Y, 3.4E-12	N	N	Y, 9.42E-14	N	Y, 3.76E-9	N	N	N	Y, 1.51E-13	N	N	N	2.74E-161
37	Aurelia.auritaCnidarian (NCBI TSA: GBNG01163133.1)	Cnidarian	NFE2	Y, 1.89E-34	Y, 6.11E-18	N	N	Y, 2.83E-10	N	Y, 6.38E-8	N	N	Y, 8.19E-13	N	N	Y, 1.50E-6	N	N	N	N	N	N	N	1.88E-56
38	Clytia.hemisphericaCnidarian (marimba.obs-vlfr.fr/blast: TCONS_00021876-protein)	Cnidarian	NFE2	Y, 1.78E-33	Y, 1.00E-15	N	N	Y, 3.38E-10	N	Y, 1.58E-11	N	Y, 5.22E-10	Y, 7.93E-18	N	N	Y, 2.14E-8	N	N	N	N	N	N	N	1.84E-64
39	Hydractinia.symbiolongicarpusCnidarian (NCBI: GAWH03056891.1)	Cnidarian	NFE2	Y, 2.70E-33	Y, 1.93E-19	N	N	Y, 4.11E-11	N	Y, 6.75E-10	N	Y, 2.14E-11	Y, 1.06E-18	N	N	N	N	N	N	N	N	N	N	1.46E-70
40	Alatina.alataCnidarian (NCBI: PUGI00000000.1TSA: comp75450_c0_seq2)	Cnidarian	NFE2	Y, 2.03E-35	Y, 1.85E-20	N	N	Y, 1.73E-11	N	Y, 4.86E-8	N	Y, 7.82E-12	Y, 3.22E-20	N	N	Y, 1.31E-10	N	N	N	N	N	N	N	1.85E-83
41	Tripedalia.cystophoraCnidarian (NCBI: GHAQ01131580.1)	Cnidarian	NFE2	Y, 5.58E-32	Y, 1.75E-19	N	N	Y, 1.54E-11	N	Y, 1.34E-6	N	Y, 2.39E-12	Y, 3.3E-17	N	N	Y, 9.55E-9	N	N	N	N	N	N	N	3.07E-70
42	Exaptasia.pallidaCnidarian (GenBank: XP_020903769.1)	Cnidarian	NFE2	Y, 5.05E-36	Y, 1.00E-16	N	N	N	N	Y, 1.75E-10	N	Y, 6.63E-11	Y, 5.11E-15	N	N	Y, 1.44E-8	N	N	N	N	N	N	Y, 1.44E-34	1.82E-90
43	Aethozpleura.elegantissimaCnidarian (NCBI: GBXI01138408.1)	Cnidarian	NFE2	Y, 6.20E-34	Y, 6.69E-18	N	N	N	N	Y, 1.75E-10	N	Y, 3.8E-11	Y, 2.88E-21	N	N	Y, 1.69E-8	N	N	N	N	N	N	Y, 2.24E-59	1.43E-127
44	Podocoryna.carneaCnidarian (NCBI: GBEH01051662.1)	Cnidarian	NFE2	Y, 1.60E-33	Y, 3.22E-20	N	N	Y, 6.74E-12	N	Y, 7.39E-10	N	N	Y, 3.83E-19	N	N	N	N	N	N	N	N	N	N	3.92E-69
45	Polypodium.hydriformCnidarian (TSA: c21979_g2_i1)	Cnidarian	NFE2	Y, 1.68E-28	N	N	N	N	N	N	N	Y, 1.28E-8	Y, 1.28E-16	N	N	N	Y, 1.23E-12	N	N	N	N	N	N	2.38E-34
46	Amanemia.viridisCnidarian (NCBI: GHCD01020975.1)	Cnidarian	NFE2	Y, 4.04E-34	Y, 3.20E-19	N	N	N	N	Y, 3.49E-10	N	N	N	Y, 8.31E-22	N	N	Y, 1.83E-8	N	N	N	N	N	Y, 4.96E-52	4.10E-109
47	Actinia.tenebrosaCnidarian (GenBank: XP_031555350.1)	Cnidarian	NFE2	Y, 1.14E-35	Y, 1.26E-18	N	N	N	N	Y, 1.43E-10	N	Y, 3.8E-11	Y, 2.88E-21	N	N	Y, 6.29E-9	Y, 2.31E-12	N	N	N	N	Y, 2.08E-54	N	1.86E-124
48	Orbicella.faveolataCnidarian (NCBI: XP_020623055.1)	Cnidarian	NFE2	Y, 6.20E-34	Y, 7.09E-19	N	N	N	N	Y, 8.10E-10	N	Y, 5.03E-11	Y, 2.57E-16	N	N	N	N	N	N	N	N	N	N	1.07E-63
49	Acropora.milleporaCnidarian (NCBI: XP_029182828.1)	Cnidarian	NFE2	Y, 1.81E-35	Y, 3.53E-18	N	N	N	N	Y, 1.58E-11	N	Y, 6.63E-11	Y, 5.59E-16	N	N	N	N	N	N	N	N	N	N	1.26E-64
50	Pocillopora.damicornisCnidarian (NCBI: XP_027060548.1)	Cnidarian	NFE2	Y, 6.39E-36	Y, 8.76E-18	N	N	N	Y, 9.78E-9	Y, 7.82E-9	N	Y, 7.26E-11	Y, 1.72E-16	N	N	Y, 2.11E-7	N	N	N	N	N	N	N	6.85E-69
51	Hofstenia.miamiaAcoelomorpha (NCBI: GHPH01000876.1)	Acoelomorpha	NFE2	Y, 1.83E-28	N	Y, 4.28E-21	N	N	N	Y, 9.55E-6	N	N	N	N	N	Y, 3.12E-13	N	N	N	N	N	N	N	4.93E-39
52	Cristaria.plicataMollusca (GenBank: ADM32563.1)	Mollusca	NFE2	Y, 6.70E-40	Y, 2.32E-20	Y, 6.05E-32	Y, 8.42E-31	Y, 2.75E-12	Y, 1.2E-16	Y, 4.23E-10	Y, 9.2E-15	Y, 2.32E-14	N	Y, 3.86E-15	N	Y, 6.71E-13	N	N	N	Y, 2.56E-14	Y, 3.52E-20	N	N	8.45E-182
53	Crassostrea.virginicaMollusca (GenBank: XP_022333013.1)	Mollusca	NFE2	Y, 1.07E-46	Y, 7.46E-21	Y, 9.16E-26	Y, 1.09E-30	Y, 1.75E-13	Y, 7.8E-16	Y, 3.54E-11	Y, 3.92E-13	Y, 2.32E-14	N	Y, 3.87E-14	N	Y, 2.21E-10	N	N	N	Y, 1.26E-16	Y, 2.46E-19	N	N	3.47E-185
54	Tribolium.castaneumArthropoda (NCBI: 91093885)	Arthropoda	NFE2	Y, 3.41E-41	Y, 4.59E-17	Y, 1.17E-36	Y, 3.67E-20	Y, 4.40E-10	Y, 4.88E-14	Y, 1.23E-7	Y, 4.64E-14	Y, 5.51E-11	N	Y, 2.18E-12	N	N	N	N	N	N	N	N	N	3.22E-134
55	AmphimedonQueenslandlicaPorifera (metazoa ensembl: 15723397)	Porifera	NFE2	Y, 3.96E-28	N	Y, 5.28E-16	N	Y, 6.35E-8	N	Y, 1.49E-6	N	N	N	N	N	N	N	N	N	N	N	N	N	9.52E-39

Table 4. Comparing MEME 3 vs MEME 10 presence in NFE2 sequences. MEME 10 was identified in cnidarian, ctenophore, sponges, and placozoa. No protein possesses both MEMEs.

NFE2 Protein Name	Phylum	Gene	MEME3	MEME10
H.sapiensNRF2Chordata (ENSG00000116044)	Chordata	NRF2	Y, 9.03E-36	N
M.musculusNRF2Chordata (ENSMUSP00000099733)	Chordata	NRF2	Y, 2.78E-37	N
G.gallusNRF2Chordata (ENSGALP00000032649)	Chordata	NRF2	Y, 2.78E-37	N
X.tropicalisNRF2Chordata (ENSXETP00000003783)	Chordata	NRF2	Y, 3.08E-36	N
D.rerioNRF2aChordata (ENSDARP00000062853)	Chordata	NRF2	Y, 9.32E-39	N
D.rerioNRF2bChordata (ENSDARP00000106581)	Chordata	NRF2	Y, 1.08E-25	N
H.sapiensNRF1Chordata (ENSG00000082641)	Chordata	NRF1	Y, 8.57E-40	N
M.musculusNRF1Chordata (ENSMUSP00000080467)	Chordata	NRF1	Y, 8.57E-40	N
G.gallusNRF1Chordata (ENSGALP00000035379)	Chordata	NRF1	Y, 8.57E-40	N
D.rerioNRF1aChordata (ENSDARP00000094757)	Chordata	NRF1	Y, 1.52E-29	N
H.sapiensNRF3Chordata (ENSG00000050344)	Chordata	NRF3	Y, 8.21E-34	N
M.musculusNRF3Chordata (ENSMUSP00000005103)	Chordata	NRF3	Y. 7.29E-34	N
X.tropicalisNRF3Chordata (ENSXETP00000026569)	Chordata	NRF3	Y, 1.96E-31	N
D.rerioNRF3Chordata (ENSDARP00000015027)	Chordata	NRF3	Y, 9.05E-25	N
H.sapiensNF-E2Chordata (ENSG00000123405)	Chordata	NF-E2	Y, 1.53E-35	N
M.musculusNF-E2Chordata (ENSMUSP00000122476)	Chordata	NF-E2	Y, 3.78E-35	N
D.rerioNF-E2Chordata (ENSDARP00000002745)	Chordata	NF-E2	Y, 5.27E-33	N
D.melanogasterCncCArhtropoda (NP_732833.1)	Arthropoda	CnC	Y, 1.51E-34	N
D.melanogasterCncIArthropoda (NP_001247258.1)	Arthropoda	CnC	Y, 1.51E-34	N
C.elegansSKN1aNematoda (NP_741404.1)	Nematoda	SKN1	N	N
N.vectensisNFE2Cnidarian	Cnidarian	NFE2	N	Y, 9.87E-16
C.intestinalisNRFChordata (ENSCINP00000024999)	Chordata	NFE2	Y, 1.22E-27	N
H.vulgarisNRFCnidarian (XP_002160548.1)	Cnidarian	NFE2	N	Y, 7.28E-19
S.purpuratusNRFEchinodermata (XP_011683763)	Echinodermata	NFE2	Y, 2.71E-29	N
B.floridaeNRFaChordata (jgi: 131476)	Chordata	NFE2	Y, 3.13E-30	N
B.floridaeNRFbChordata (jgi: 127500)	Chordata	NFE2	Y, 3.13E-30	N
O.bimaculoidesnrfMollusca (Ocbimv22001599m)	Mollusca	NFE2	Y, 3.16E-25	N
L.giganteaNRFMollusca (LgGsHFWreduced.7631)	Mollusca	NFE2	Y, 2.08E-27	N
Capitella.telataNFE2Annelida (jgi: 19335)	Annelida	NFE2	Y, 2.71E-27	N
Oscarella.carmelaNFE2Porifera (Compagen: g5281)	Porifera	NFE2	N	Y, 1.6E-13
Mnemiopsis.leydiiNFE2Ctenophora (NHGRI: Mnemio_ML2.2.aa: ML016353a)	Ctenophora	NFE2	N	Y, 3.26E-13
Trichoplax.adhaeransNFE2Placozoa (jgi: 60616)	Placozoa	NFE2	N	Y, 5.59E-15
Sycon.ciliatumNFE2Porifera (Compagen: scpid35693)	Porifera	NFE2	N	N
Daphnia.pulexNFE2Arthropoda (jgi: 307821)	Arthropoda	NFE2	Y, 3.14E-34	N
Platynereis.dumeriliiAnnelida (pdumbase.gdcb.iastate.edu: comp225037_c1_seq8)	Annelida	NFE2	Y, 5.43E-32	N
Lingula.anatinaBrachiopoda (GenBank: XP_013384291.1)	Brachiopoda	NFE2	Y, 1.3E-32	N
Aurelia.auritaCnidarian (NCBI TSA: GBRG01161313.1)	Cnidarian	NFE2	N	Y, 8.19E-13
Clytia.hemisphericaCnidarian (marimba.obs-vlfr.fr/blast: TCONS_00021876-protein)	Cnidarian	NFE2	N	Y, 7.93E-18
Hydractinia.symbiolongicarpusCnidarian (NCBI: GAWH01056891.1)	Cnidarian	NFE2	N	Y, 1.06E-18
Alatina.alataCnidarian (NCBI: PUGI00000000 TSA: comp75450_c0_seq2)	Cnidarian	NFE2	N	Y, 2.32E-20
Tripedalia.cystophoraCnidarian (NCBI: GHAQ01131580.1)	Cnidarian	NFE2	N	Y, 3.3E-17
Exaptasia.pallidaCnidarian (GenBank: XP_020903769.1)	Cnidarian	NFE2	N	Y, 5.11E-15
Anthopleura.elegantissimaCnidarian (NCBI: GBXJ01138408.1)	Cnidarian	NFE2	N	Y, 2.88E-21
Podocoryna.carneaCnidarian (NCBI: GBEH01051662.1)	Cnidarian	NFE2	N	Y, 3.83E-19
Polypodium.hydriformCnidarian (TSA: c21979_g1_i1)	Cnidarian	NFE2	N	Y, 1.28E-16
Amenomia.viridisCnidarian (NCBI: GHCD01000975.1)	Cnidarian	NFE2	N	Y, 8.31E-22
Actinia.tenebrosaCnidarian (GenBank: XP_031555350.1)	Cnidarian	NFE2	N	Y, 2.88E-21
Orbicella.faveolataCnidarian (NCBI: XP_020623055.1)	Cnidarian	NFE2	N	Y, 2.57E-16
Acropora.milleporaCnidarian (NCBI: XP_029182828.1)	Cnidarian	NFE2	N	Y, 5.59E-16
Pocillopora.damicornisCnidarian (NCBI: XP_027060548.1)	Cnidarian	NFE2	N	Y, 1.72E-16
Hofstenia.miamiaAcoelomorpha (NCBI: GHHH01000876.1)	Acoelomorpha	NFE2	Y, 4.28E-21	N
Cristaria.plicataMollusca (GenBank: AZM32563.1)	Mollusca	NFE2	Y, 6.05E-32	N
Crassostrea.virginicaMollusca (GenBank: XP_022333013.1)	Mollusca	NFE2	Y, 9.16E-26	N
Tribolium.castaneumArthropoda (NCBI: 91093885)	Arthropoda	NFE2	Y, 1.17E-36	N
Amphimedon.QueenslandicaPorifera (metazoa ensembl: 15723397)	Porifera	NFE2	Y, 5.28E-16	N

Table 5. Top 35 identified interactors of *Nematostella* NFE2 based on peptide counts from experimental (antibody incubated with anemone homogenates) vs. control (anemone homogenate without antibody incubation) search in "Nrf-ome."

Gene ID	Gene Search Name	NRF2-ome hit	Interactions with NRF2	Interactions with KEAP1
NVE22591	adhesion G-protein coupled receptor	None	None	None
NVE6160	filamin A, alpha	33 Interactions	None	None
NVE7520	calponin family repeat	None	None	None
NVE18733	basement membrane-specific heparan sulfate proteoglycan core protein-like	12 interactions	None (Indirect Transcriptional regulation)	None
NVE4716	flotillin	None	None	None
NVE25480	cytochrome b-c1 complex subunit 2, mitochondrial	1 interaction	None (Indirect Transcriptional regulation)	None
NVE9857	ATP synthase subunit beta	None	None	None
NVE24413	sodium/potassium-transporting ATPase subunit alpha	None	None	None
NVE16136	rab GDP dissociation inhibitor alpha	None	None	None
NVE1879	succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	None	None	None
NVE24872	Collagen alpha-4(VI) chain	11 Interactons	None (Indirect Transcriptional regulation)	None
NVE5875	uncharacterized protein	None	None	None
NVE14118	sodium/potassium-transporting ATPase subunit beta-1	4 Interactions	None (Indirect Transcriptional regulation)	None
NVE3482	flotillin	None	None	None
NVE11712	Phosphoglycerate kinase	None	None	None
NVE15323	phosphate carrier protein	None	None	None
NVE13578	ATP-dependent RNA helicase eIF4A	None	None	None
NVE12751	peroxisomal multifunctional enzyme type 2	None	None	None
NVE2876	actin, cytoplasmic 1	None	None	None
NVE9487	adenylyl cyclase-associated protein 1	None	None	None
NVE23692	uncharacterized protein	None	None	None
NVE25966	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	43 Interactions	None (Indirect Transcriptional regulation)	None
NVE7749	propionyl-CoA carboxylase alpha chain, mitochondrial	None	None	None
NVE160	uncharacterized protein	None	None	None
NVE25082	calpain 2, (m II) large subunit	None	None	None
NVE15469	clathrin, heavy	25 Interactiona	None (Indirect Transcriptional regulation)	None
NVE21257	14-3-3 protein beta/alpha-B	60 Interactions	None	None
NVE20499	vitellogenin-1	None	None	None
NVE14700	Sorting nexin-2	None	None	None
NVE3191	NAD(P) transhydrogenase	None	None	None
NVE14971	calcium uniporter protein	2 Interactions	None (Indirect Transcriptional regulation)	None
NVE16306	gelsolin-like protein 2	14 Interactions	None (Indirect Transcriptional regulation)	None
NVE14874	A-kinase anchor protein 9	None	None	None
NVE6411	Major vault protein	None	None	None
NVE21815	Uncharacterized protein	None	None	None

Table 6. Identified interactors of *Nematostella* NFE2 based on peptide counts >20 search within the "Nrf-ome."

Gene ID	Gene Search Name	NRF2-ome hit	Interactions with NRF2	Interactions with KEAP1
NVE22591	Fibropellin-1	None	None	None
NVE6160	Filamin-1 (FLNA)	33 direct interactions, mostly with other filamins and smads	None	None
NVE9857	ATP5G3	1 direct interaction, ABHD16A	None	None
NVE24413	ATP1A1	None	None	None
NVE24872	MATN2	4 direct interactions, ATXN1, COL4A1, COL4A4, FNA	None (indirect trancsriptional)	None
NVE5875	No hits to use as reference	None	None	None
NVE25082	CAPN14 and 7 (other calpains had no hits)	None	None (indrect transcriptional)	None
NVE15469	CLTC	25 direct interactions	None	None
NVE20499	VIT and Vitelloginin	None	None	None
NVE14422	No clear search term to match	None	None	None
NVE13575	ERN1	7 direct interactions, BAX, MAP3K5, PSEN1, TAOK3,TNFRSF1A	None	None
NVE22525	HSP90AB1	50 direct interactions,	None	Direct inhibition of KEAP
NVE7035	SPTA1	Yes, 6 direct with ABI1, ABL1, ENAH, ERCC4, YWHAQ	None (indirect transcriptional)	None
NVE20824	No hits to use as reference	None	None	None
NVE4604	No hits to use as reference	None	None	None
NVE15896	Moesin	5 direct, ELAVL1, H2AFX, MAPK8, RHDA, TNFRSF1A	None (indirect transcriptional)	None
NVE22317	No hits to use as reference	None	None	None
NVE20604	SPTA1 (SPTBN1, no hits)	Yes, 6 direct with ABI1, ABL1, ENAH, ERCC4, YWHAQ	None (indirect transcriptional)	None
NVE1270	No hits to use as reference	None	None	None
NVE6605	Many tubulins	None	None	None
NVE24810	Many collagens	None	None	None
NVE7034	SPTA1	Yes, 6 direct with ABI1, ABL1, ENAH, ERCC4, YWHAQ	None (indirect transcriptional)	None
NVE9071	No hits to use as reference	None	None	None
NVE20813	No hits to use as reference	None	None	None
NVE14885	Many tubulins	None	None	None
NVE25565	Many tubulins	None	None	None
NVE2016	Agrin (AGRN)	2 direct interactions, GRI1B, PAK1IP1	None	None
NVE4823	Many myosins	None	None	None
NVE2702	Many tubulins	None	None	None
NVE16356	No hits to use as reference	None	None	None
NVE8262	No hits to use as reference	None	None	None
NVE15133	Many myosins	None	None	None
NVE14552	Many myosins	None	None	None
NVE2701	Many tubulins	None	None	None
NVE12470	No hits to use as reference	None	None	None
NVE8266	No hits to use as reference	None	None	None
NVE3113	No hits to use as reference	None	None	None

Table 7. Identified interactors of *Nematostella* NFE2 based on studies of Vertebrate NFE2 paralogs search within the "Nrf-ome."

Gene ID	Gene Search Name	NRF2-ome hit	Interactions with NRF2	Interactions with KEAP1
NVE12138	Ubiquitin	21 total interactions	1 direct interaction	n/a
NVE11974	ubiquitin-like protein FUBI	n/a	n/a	n/a
NVE25649	f-box protein 9	n/a	n/a	n/a
NVE24130	SplA ryanodine receptor domain and SOCS box containing 3	n/a	n/a	n/a
NVE11918	Endoplasmic reticulum chaperone BiP	n/a	n/a	n/a
NVE10888	SSR1	n/a	n/a	n/a

REFERENCES

1. Lynch, M. (2002). Gene duplication and evolution. *Science*, *297*(5583), 945-947.

2. Prince, V. E., & Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics*, *3*(11).

3. Jindrich, K., & Degnan, B. M. (2016). The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity. *BMC Evolutionary Biology*, *16*(1), 28.

4. Amoutzias, G. D., Veron, A. S., Weiner III, J., Robinson-Rechavi, M., Bornberg-Bauer, E., Oliver, S. G., & Robertson, D. L. (2006). One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Molecular Biology and Evolution*, *24*(3), 827-835.

5. J Biol Chem. 2012 Feb 10;287(7):4609-27. doi: 10.1074/jbc.M111.260125. Epub 2011 Dec 15.

6. Fuse, Y.; Kobayashi, M. Conservation of the Keap1-Nrf2 System: An Evolutionary Journey through Stressful Space and Time. *Molecules* **2017**, *22*, 436.

7. Ma, Qiang. “Role of nrf2 in oxidative stress and toxicity.” *Annual review of pharmacology and toxicology* vol. 53 (2013): 401-26. doi:10.1146/annurev-pharmtox-011112-140320

8. Kobayashi, Makoto et al. “The antioxidant defense system Keap1-Nrf2 comprises a multiple sensing mechanism for responding to a wide range of chemical compounds.” *Molecular and cellular biology* 29 2 (2009): 493-502 .

9. Culbreth, Megan, and Michael Aschner. “GSK-3β, a double-edged sword in Nrf2 regulation: Implications for neurological dysfunction and disease.” *F1000Research* vol. 7 1043. 10 Jul. 2018, doi:10.12688/f1000research.15239.1

10. Matzinger, Manuel et al. “Activation of Nrf2 signaling by natural products-can it alleviate diabetes?.” Biotechnology advances vol. 36,6 (2018): 1738-1767. doi:10.1016/j.biotechadv.2017.12.015

11. Choe, Keith P et al. “The WD40 repeat protein WDR-23 functions with the CUL4/DDB1 ubiquitin ligase to regulate nuclear abundance and activity of SKN-1 in Caenorhabditis elegans.” Molecular and cellular biology vol. 29,10 (2009): 2704-15. doi:10.1128/MCB.01811-08

12. Dunn et al. “The hidden biology of sponges and ctenophores.” Trends in Ecology and Evolution vol. 30, 5 (2015): 282-291.

13. Technau, Ulrich and Steele, Robert E. “Evolutionary crossroads in developmental biology: Cnidaria.” Development vol 138 (2011) 1447-1458; doi: 10.1242/dev8959

14. Tang, Lanlan et al. “Isolation of a Hypomorphic skn-1 Allele That Does Not Require a Balancer for Maintenance.” G3 (Bethesda, Md.) vol. 6,3 551-8. 29 Dec. 2015, doi:10.1534/g3.115.023010

15. Warner et al. “NvERTx: a gene expression database to compare embryogenesis and regeneration in the sea anemone Nematostella vectensis.” Development 2018 145: dev162867 doi: 10.1242/dev.162867 Published 17 May 2018

16. Karabulut et al. “Electroporation of short hairpin RNAs for rapid and efficient gene knockdown in the starlet sea anemone, Nematostella vectensis.” Developmental Biology vol 448, 1 (2019) 7-15 doi: 10.1016/j.ydbio.2019.01.005

17. Ikmi, A., McKinney, S., Delventhal, K. et al. TALEN and CRISPR/Cas9-mediated genome editing in the early-branching metazoan Nematostella vectensis. Nat Commun 5, 5486 (2014). https://doi.org/10.1038/ncomms6486

18. Asis Datta, Ayushi Kamthan, Mohan Kamthan et al. A simple protocol to detect interacting proteins by GST pull down assay coupled with MALDI or LC-MS/MS analysis., 12 October 2015, PROTOCOL (Version 1) available at Protocol Exchange [+https://doi.org/10.1038/protex.2015.093+]

19. A. Stamatakis: "RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies". In *Bioinformatics*, 2014, open access link: http://bioinformatics.oxfordjournals.org/content/early/2014/01/21/bioinformatics.btu033.abstract?keytype=ref&ijkey=VTEqgUJYCDcf0kP

20. Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Park, California, 1994. [pdf]

21. Radhakrishnan SK, Lee CS, Young P, Beskow A, Chan JY, Deshaies RJ. Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. *Mol Cell*. 2010;38(1):17-28. doi:10.1016/j.molcel.2010.02.029

22. eLife 2014;3:e01856 DOI: 10.7554/eLife.01856

23. Yoshiki Tsuchiya, Tomoko Morita, Mehee Kim, Shun-ichiro Iemura, Tohru Natsume, Masayuki Yamamoto, Akira Kobayashi "Dual Regulation of the Transcriptional Activity of Nrf1 by β-TrCP- and Hrd1-Dependent Degradation Mechanisms" Molecular and Cellular Biology Oct 2011, 31 (22) 4500-4512; DOI: 10.1128/MCB.05663-11

24. Steffen J, Seeger M, Koch A, Krüger E. Proteasomal degradation is transcriptionally controlled by TCF11 via an ERAD-dependent feedback loop. Mol Cell. 2010;40(1):147-158. doi:10.1016/j.molcel.2010.09.012

25. Anna Elena Pepe , Qingzhong Xiao , Anna Zampetaki , Zhongyi Zhang , Akira Kobayashi , Yanhua Hu , and Qingbo Xu "Crucial Role of Nrf3 in Smooth Muscle Cell Differentiation From Stem Cells" Jan 2010https://doi.org/10.1161/CIRCRESAHA.109.211417Circulation Research. 2010;106:870–879

26. Sankaranarayanan K, Jaiswal AK. Nrf3 negatively regulates antioxidant-response element-mediated expression and antioxidant induction of NAD(P)H:quinone oxidoreductase1 gene. J Biol Chem. 2004;279(49):50810-50817. doi:10.1074/jbc.M404984200

27. Chowdhury, A.M.M.A., Katoh, H., Hatanaka, A. et al. Multiple regulatory mechanisms of the biological function of NRF3 (NFE2L3) control cancer cell proliferation. Sci Rep 7, 12494 (2017). https://doi.org/10.1038/s41598-017-12675-y

28. Casteel D, Suhasini M, Gudi T, Naima R, Pilz RB. Regulation of the erythroid transcription factor NF-E2 by cyclic adenosine monophosphate-dependent protein kinase. *Blood*. 1998;91(9):3193-3201.

29. Fujita, Rie et al. “NF-E2 p45 is important for establishing normal function of platelets.” Molecular and cellular biology vol. 33,14 (2013): 2659-70. doi:10.1128/MCB.01274-12

30. NRF2-ome, an integrated web resource to discover protein interaction and regulatory networks of NRF2. Türei D, Papp D, Fazekas D, Földvári-Nagy L, Módos D, Lenti K, Csermely P, Korcsmáros T. (2013) — Oxidative Medicine and Cellular Longevity, 2013:737591

31. The NRF2-related interactome and regulome contain multifunctional proteins and fine-tuned autoregulatory loops. (2012) Papp D, Lenti K, Módos D, Fazekas D, Dúl Z, Türei D, Földvári-Nagy L, Nussinov R, Csermely P, Korcsmáros T. FEBS Letters 586, 13, 1795–1802.

32. Sebé-Pedrós A, Saudemont B, Chomsky E, et al. Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell*. 2018;173(6):1520-1534.e20. doi:10.1016/j.cell.2018.05.019

33. “NvERTx: a gene expression database to compare embryogenesis and regeneration in the sea anemone Nematostella vectensis” Jacob F. Warner, Vincent Guerlais, Aldine R. Amiel, Hereroa Johnston, Karine Nedoncelle, Eric Röttinger Development 2018 145: dev162867 doi: 10.1242/dev.162867 Published 17 May 2018

34. Wolenski FS, Chandani S, Stefanik DJ, et al. Two polymorphic residues account for the differences in DNA binding and transcriptional activation by NF-κB proteins encoded by naturally occurring alleles in Nematostella vectensis. J Mol Evol. 2011;73(5-6):325-336. doi:10.1007/s00239-011-9479-7

35. Knighton LE, Nitika, Waller SJ, et al. Dynamic remodeling of the interactomes of Nematostella vectensis Hsp70 isoforms under heat shock. *J Proteomics*. 2019;206:103416. doi:10.1016/j.jprot.2019.103416

APPENDIX: SUPPLEMENTARY FILES

Supplementary Table 1: Excel file of NFE2 expression profile in *Nematostella* 144 hours post-amputation (regeneration) and 192 hours post-fertilization (embryogenesis).

Supplementary Table 2: NFE2 expression within cell types of larval and adult *Nematostella*

Supplementary Figure 1: Histogram of LFQ Intensity of Mass Spectrometry

Supplementary Figure 2: Volcano Plot of Mass Spectrometry

Supplementary Figure 3: Venn-Diagram of Peptides Identified Between Control and NFE2