THE CATHOLIC UNIVERSITY OF AMERICA

Identification and characterization of novel homodimer targets of CeTwist

A DISSERTATION

Submitted to the Faculty of the

Department of Biology

School of Arts and Sciences

Of The Catholic University of America

In partial fulfillments of the Requirements

For the Degree

Doctor of Philosophy

By

Nirupama Janardansingh Singh

Washington D.C.

2012

Identification and characterization of novel homodimer targets of CeTwist

Nirupama Janardansingh Singh, Ph. D.

Director: Ann K. Corsi, Ph. D.

Twist, a basic-helix-loop-helix (bHLH) transcription factor, functions in the mesoderm in many organisms. bHLH transcription factors dimerize with their helix-loop-helix region forming a heterodimer or a homodimer and the basic regions of the dimers bind to a consensus sequence known as an E box on their target genes. Previous evidence suggests a distinct role for CeTwist homodimers and heterodimers (with CeE/DA) in different mesoderm cells in the model organism *C. elegans*. We have already identified a number of CeTwist heterodimer targets. However, no CeTwist homodimer targets have been identified to date. The aim of this project is to discover and study such genes to understand bHLH dimer-specific transcriptional regulation. To find CeTwist homodimer targets, we overexpressed CeTwist by using an inducible heat shock promoter and found the potential target genes that are overexpressed using a global gene expression measurement technique called Affymetrix oligonucleotide microarrays. Nineteen genes from a list of overexpressed genes were prioritized based on higher microarray value. Using transcriptional GFP reporters, we examined the expression pattern of these genes.

We were interested in a pattern that may partially overlap with the CeTwist expression pattern, and we found five genes in this category. Then GFP expression was studied in mutant animals of the *hlh-2* (CeE/DA) and *hlh-8* (CeTwist) genes for further validation. The results suggest that we found one CeTwist homodimer target, two targets of CeTwist and CeE/DA heterodimers and two more CeTwist targets that need further characterization. Next, we studied the promoter region of the new CeTwist targets by using 5'deletion reporter constructs. In one of the heterodimer targets, we found a previously identified E box playing a role in target gene expression whereas for the homodimer target, we identified a palindromic sequence CACGTG that plays an important role in expression. This interesting result is consistent with bHLH homodimers preferentially binding to palindromic sequences with identical half-sites for regulation. Moreover, we also found that two of the new CeTwist target genes identified in this study, have human homologues which are mutated in human diseases. Therefore, understanding the transcriptional regulation of these genes is important.

This dissertation by Nirupama Janardansingh Singh fulfills the dissertation requirement for the doctoral degree in Cell and Microbial Biology approved by Ann K. Corsi, Ph.D., as Director, and by John E. Golin, Ph.D., and Pamela L. Tuma Ph.D. as Readers.

_____

Ann K. Corsi, Ph.D., Director

_____

John E. Golin, Ph.D., Reader

_____

Pamela L. Tuma, Ph.D., Reader

This work is dedicated to

My husband,

For his unconditional help and support that encouraged me to move forward in difficult times and whose dedicated love and confidence in me, gave me the strength to achieve heights I never anticipated

And

My parents,

Their emotional support at each step of my life was important

Without their love, support and prayers this work could have never been finished

And,

My loving and adorable kids Astha and Parth,

Their smile at the end of the day used to be the most beautiful gift

TABLE OF CONTENTS

Preliminaries:

Text:

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

bHLH: Basic helix-loop-helix

PCR: polymerase chain reaction

bp: base pair

gfp: green fluorescent protein

L1: the first larval stage

L2: the second larval stage

L3: the third larval stage

Egl: egg-laying defective

Con: constipated

dNTPs: deoxynucleotides

SOE: sequence overlap extension pcr

µL: microlitre

mM: millimolar

NGM: nematode growth medium

ACKNOWLEDGEMENTS

I would like to acknowledge the following:

Dr. Ann Corsi: For providing this interesting project and also to be available all the times during this last 6 yrs especially in the beginning when nothing was making sense to me. I am thankful to her for her continuous guidance that I could finish this research work. I would also like to thank her for drafting of the thesis through long distance communications.

Dr. Golin and Dr. Tuma: For providing support and encouragement in yearly committee meetings. I would also like to thank Dr. Tuma for her insightful ideas on my thesis. I would like to thank Dr. Golin for providing reagents for site-directed mutagenesis.

Dr. Lori Estes, Dr. Golin and Dr. Corsi: For freedom and their trust in my work as a T.A. for their classes.

Dr. Peng Wang: Dr Wang started this project and helped me to understand the project in the beginning along with helping me to train in microinjection technique.

Yanhan Huang: Her enthusiasm used to keep me motivated to move forward in my research. She was a great help in every day lab work as a sincere friend. Her support and help as a friend and lab mate meant a lot to me.

Robert Gereige: For matings in *hlh-2* deletion mutants for few of the reporter constructs.

UMBC worm club and NIH data club: I would like to thank them for their helpful advice and suggestions.

INTRODUCTION

bHLH transcription factors

Transcription factors are proteins required for gene regulation where the factors can bind to the promoter of a gene to activate or repress its transcription. Most of the transcription factors have three domains. The first domain is a DNA binding domain which binds to the target gene. The second is a transactivating domain which activates or represses the transcription of the target gene. The third domain is a protein-protein interaction domain which binds to other proteins e.g. another transcription factor. Transcription factors are required for the development of multicellular organisms (Lobe et al., 1992). These proteins can be grouped into families based on structural similarity. An example of a family of transcription factors is the basic helix-loop-helix (bHLH) transcription factors. The bHLH transcription factors are required for regulating developmental processes such as myogenesis, neurogenesis, and hematopoesis in multicellular organisms. The bHLH motif is composed of a basic region and two helices with an interhelical loop. The bHLH proteins dimerize by binding to other HLH proteins through their helix-loop-helix regions. After dimerization, the two basic regions bind to target genes and activate or repress the transcription of the gene depending on the specific dimer partners (Fig 1). The dimer binds to an E box, a hexanucleotide sequence, CANNTG where each monomer binds to the half-site CAN or NTG. Different dimers

prefer a unique composition of N bases which is important in determining the binding affinity of dimers and their role in regulating specific genes. The bHLH proteins can be classified further, into different groups based upon their partner choice for dimerization, tissue distribution, DNA binding properties and structural features (Massari et al., 2000).

Twist, a bHLH transcription factor

bHLH transcription factors are classified into several groups. The class I proteins, also known as E proteins, are ubiquitous in expression and capable of forming homodimers and heterodimers (Murre et al., 1989). Class II bHLH transcription factors include proteins such as MyoD, myogenin, Twist and others. Class II bHLH proteins can form homodimers or heterodimers with class I E proteins and have a tissue-restricted expression pattern (Maleki et al., 1997, Castanon et al., 2001, Connerney et al., 2006, Philogene et al., 2012). Class III HLH proteins include the Myc family of transcription factors, SREBP-1 and others. A Leucine-Zipper is the characteristic of proteins of this class. Class IV group of proteins include Mad, Max and others which are capable of forming dimers with Myc proteins or with one another. Class V members include Id proteins which lack a basic region and thus, cannot bind DNA. Id proteins are negative regulators of class I and class II proteins. The other classes of HLH transcription factors are grouped based on their structure and functional properties. Some of these proteins

may form homodimers and/or heterodimers with different proteins of the same family of transcription factors or with members of different families of transcription factors.

Twist, a class II transcription factor, was first identified in *Drosophila melanogaster*, and it is required for establishment of dorsal-ventral patterning, gastrulation and differentiation of the mesoderm (Thisse et al., 1987). Subsequently, its homologs were identified in mice (Wolf et al., 1991), humans (Wang et al., 1997) and nematodes (Harfe et al., 1991). Twist is an essential gene required for differentiation of mesoderm at different stages of development in *Drosophila* and mouse. For example, in *Drosophila* it is required in early stages of development and a loss-of-function mutant is embryonic lethal (Simpson et al., 1983, Thisse et al., 1987). In mouse, a Twist null mutant dies at embryonic day 11.5 with defects in head mesenchyme, somites and limb buds where it was shown that head mesenchyme was required for cranial neural tube closure (Chen et al., 1995). The null mutant of CeTwist in *C. elegans* is viable and fertile (Corsi et al., 2000) giving an added advantage to study the Twist protein in *C. elegans*.

Role of Twist in Metazoans

In humans, class II bHLH transcription factors include the subfamily of Twist proteins. The Twist family includes Paraxis, Scleraxis, Hand1, Hand2, Twist1 and Twist2 (Cserjesi et al., 1995, Li et al., 1995, Burgess et al., 1995). Twist1 and Twist2 share 66% identity across the entire protein length whereas in the bHLH region of the proteins, the

identity is 98%. Regulation of Twist protein level is tightly controlled in humans where loss-of-function mutations are involved in developmental disorders and gain-of-function mutations are involved in cancer metastasis. In humans, cranial-sutures (joints in skull-bones) are open at birth and fuse later in life. Twist1 haploinsufficiency causes the autosomal dominant Saethre-Chotzen Syndrome (SCS) (Fig. 2), characterized by premature fusion of cranial sutures (craniosynostosis) leading to facial asymmetry (craniofacial phenotype) along with, syndactyly (fused digits), polydactyly (supernumerary digits) and, rarely, associated with congenital heart malformations (Howard et al., 1997). Twist2 loss-of-function mutations cause the autosomal recessive Setleis syndrome, characterized by a typical facial appearance (leonine face) and bitemporal scar-like lesions that resemble forceps marks. This disease is also classified as ectodermal dysplasia (Tukel et al., 2010).

The function of the Twist protein in mammals is shown to be both activation and inhibition of mesenchymal specification and differentiation (Baylies et al., 2002, Tam et al., 2002), but it is still unclear how Twist is performing these two opposite functions. It was proposed by scientists working in different model organisms that the different functions by a single bHLH protein might be dependent on its dimer choice (Castanon et al., 2001, Connerney et al., 2006). MyoD, another class II bHLH transcription factor, regulates establishment and differentiation of myogenic lineages with other proteins e.g., myogenin (Weintraub et al., 1991). Previous evidence suggests that E proteins are

required for heterodimerization with myogenic proteins MyoD and myogenin in vivo for activation of several downstream muscle regulators (Lassar et al., 1991). Initial work suggested that Twist functions as a repressor when it binds to E proteins to form dimers and, thus, not allowing dimerization of MyoD and E protein complexes (Kophengnavong et al., 2000). Also, Twist and MyoD can dimerize in a unique way with their basic and helix1 regions therefore, inhibiting MyoD regulated activation of myogenesis (Hamamori et al., 1997).

Twist mutants in mice have phenotypes similar to human diseases where Twist1 null-heterozygous mice develop polydactyly and craniofacial phenotypes similar to human Saethre-Chotzen Syndrome (Bourgeois et al., 1998). Another group showed that mice heterozygous for Twist null mutations exhibit fusion of cranial sutures (craniosynostosis), the landmark phenotype for SCS (Carver et al., 2002). Similarly, Twist2 knockout mutants resemble a phenotype similar to Setleis syndrome patients where they share the typical facial appearance (leonine face), bitemporal lesions and other facial and eye abnormalities (pointed chin, sparse eyelashes) (Tukel et al., 2010). This set of data suggested the evolutionarily conserved role of Twist in mice and humans where mice can be a molecular model to study the Twist protein.

In vertebrates, Id proteins represent another class of HLH proteins which have no basic domain and therefore cannot bind DNA. Id proteins bind to class I-E proteins and, thus, do not allow heterodimer formation of class II and class I. Spicer and his group,

working on the mouse model suggested Twist1 can form functional homodimers along with the heterodimers with the E proteins in the sutures of the mouse skull (Connerney et al., 2006). In an open suture, the bones of the skull comprise two opposing osteogenic fronts and in the middle is the mesenchyme (mid-suture). The evidence suggested that dimer choice of Twist depends on the availability of E and Id proteins. In osteogenic fronts, Id proteins titrate away the E proteins and, therefore, Twist homodimers form. In the mid-suture mesenchyme, there is no expression of Id protein and, thus, Twist and E proteins are available to heterodimerize in this region. The different Twist dimers function at specific areas in the suture to regulate different sets of genes. Twist homodimers (T/T) and genes regulated by them, such as FGFR2 and periostin, are expressed in osteogenic fronts, whereas Twist heterodimers (T/E) and their target thrombospondin-1 are expressed in the mid-suture. The data also suggested that the function of the Twist protein depends on its dimer choice where Twist homodimers activate bone differentiation whereas heterodimers function by inhibiting bone differentiation in cranial suture mesenchyme by regulating different sets of genes (Connerney et al., 2006). Further, evidence suggests that the FGFR2 is differentially regulated by T/T and T/E dimers where Twist homodimers activate gene expression whereas heterodimers of Twist and E proteins inhibit FGFR2 expression. Increased expression of FGR2 activates FGF signaling that promotes differentiation of the mid-suture cells and further leads to craniosynostosis. Thus, in Twist haploinsufficiency,

Twist prefers homodimerization and activation of FGFR2 leading to craniosynostosis (Connerney et al., 2008).

In *Drosophila,* the Twist protein is known as D-Twist and the E protein homologue is Daughterless (Da). It has been shown previously, that D-Twist forms homodimers in vitro (Lee et al., 1997; Halfon et al., 2000) but the first evidence of a functional Twist homodimer in vivo was shown by the Baylies group where the data suggested that Twist homodimers function as activators, required for mesoderm induction and somatic muscle development whereas D-Twist/Da heterodimers repress genes required for somatic myogenesis (Castanon et al., 2001). In order to test the role of D-Twist homodimers, tethered D-Twist proteins were expressed, in the *Twist* null mutant. In this experiment, two D-Twist monomers were connected via a flexible polylinker and they were predicted to function as solely D-Twist homodimers in vivo. *Twist* null mutants are devoid of the mesoderm layer and tethered D-Twist homodimers can rescue the mesoderm induction. Also, it was shown that misexpression of D-Twist homodimers in ectoderm or endoderm cells induces ectopic somatic muscle formation (Castanon et al., 2001). Thus, the cumulated evidence suggests a distinct role of Twist homodimers and heterodimers in both vertebrates and invertebrates.

C. elegans mesoderm and CeTwist

The germ layers are formed during embryogenesis in multicellular organisms making three layers, the ectoderm, mesoderm and endoderm. The outermost layer the ectoderm, forms the nervous system and epidermis whereas the innermost layer, the endoderm, gives rise to many organs, e.g. lungs, thyroid, and liver. The middle germ layer, the mesoderm, forms a variety of tissues, e.g. blood, muscles, connective tissue and some organs like heart. *C. elegans* is a eukaryotic organism, where its mesoderm differentiates into somatic gonad and non-gonadal tissues. CeTwist is expressed only in a subset of mesoderm tissues. CeTwist does not have blood and organs as higher eukaryotic organisms do, thus the simpler mesoderm along with limited CeTwist expression in a subset of mesoderm cells is advantageous to use *C. elegans* as a model organism to study this transcription factor. CeTwist plays an important role in the development of non-gonadal mesodermal derived tissues in *C. elegans. C. elegans* non-gonadal mesoderm tissue is composed of striated muscles, non-striated muscles and non-muscle cells. Striated muscles include 95 body wall muscles (bwm) and non-striated muscles include 20 pharyngeal muscles (pm), 16 sex muscles (sm) and 4 enteric muscles (em). There are a few non-muscle cells of mesoderm origin: one head mesodermal cell (hmc) and 6 coelomocytes (cc). Body wall muscles are arranged in four quadrants along the longitudinal axis of the animal and the contraction and relaxation of these muscles allows worms to move in a sinusoidal pattern. The pharyngeal muscles are present in the

anterior region of worm which is required for food uptake. Out of 16 sex muscles, 8 are vulval and 8 are uterine muscles; together they function in egg laying. The vulval muscles appear as cross-like projections around the vulval opening when seen from a ventral view and the uterine muscles can be easily identified as finger-like projections around the uterus in the region of the vulval muscles. There are four muscles grouped in the enteric muscles: 2 intestinal muscles, 1 anal depressor muscle and 1 anal sphincter. All four of these muscles are required for efficient defecation in worms. In the non-muscle cell category, the function of the head mesodermal cell is not identified yet and coelomocytes have a scavenging function like macrophages. The mesoderm cells where CeTwist plays a role arise from separate cell lineages during various times of worm development, which can be divided into embryonic and post-embryonic mesoderm development. The four enteric muscles and the head mesodermal cell arise embryonically from three different cell lineages whereas CeTwist dependent post-embryonic mesoderm arises from a single M cell which gives rise to 16 sex muscles, 14 body wall muscles and 2 coelomocytes (Fig 3). *C. elegans* presumptive CeTwist null mutants have defects in enteric muscles and sex muscles causing constipated (Con) and egg-laying defective (Egl) phenotypes.

It has been shown previously that the Twist pathway in humans to *C.elegans* is conserved where Twist and its targets involved in human craniosynostotic diseases are conserved (Harfe et al., 1998, Wang et al., 2006, Fig. 2). Therefore, we predicted if Twist

homodimers function in *Drosophila* and mouse, they might be functioning in *C. elegans* also, along with CeTwist and CeE/DA heterodimers that have been studied previously. Recently new evidence in *C. elegans* suggests the existence of functional homodimers along with the heterodimers (Philogene et al., 2012). The evidence suggests that different CeTwist dimers might be functioning at different stages of the worm life cycle. CeTwist tethered homodimers can rescue the abnormal M cell division and constipation phenotype caused by defective enteric muscle but it cannot rescue the egg-laying defective phenotype caused by abnormal vulval muscle in null-mutants of CeTwist. Thus, it was predicted that CeTwist homodimers (T/T) function in the early mesoderm and enteric muscle development whereas CeTwist and CeE/DA (T/E) heterodimers are predicted to function in the development of vulval muscle. The data also suggested that *arg-1,* a target of CeTwist might be regulated by different dimers of CeTwist similar to the regulation of the FGFR2 gene by different Twist dimer partners in mouse (Philogene et al., 2012, Connerney et al., 2008). Thus, we predict if CeTwist functions as a homodimer and heterodimer in *C. elegans,* it might have targets regulated exclusively by each dimer. So far, we have identified only CeTwist and CeE/DA heterodimer targets, and thus, our aim in this project is to identify targets that are exclusively regulated by CeTwist homodimers. Therefore, having several CeTwist homodimer targets, combined with already known heterodimer targets, we might gain a better understanding of the dimer-specific transcriptional regulation by CeTwist using a model organism, *Caenorhabditis*

elegans. And due to the conserved nature of the Twist pathways, what we learn in *C. elegans* might shed light on the vertebrate pathway.

The model organism, *C. elegans*

C. elegans was studied as a model organism by Brenner, Sulston, and Horvitz and appreciated by many more scientists because of its many advantages listed below. The nematode, *C. elegans*, is a non-parasitic, soil-dwelling organism of 1 mm in length when fully developed. It can be easily cultivated in the Nematode Growth Medium petri-dishes or liquid culture, if large numbers are required. Its food is the non virulent strain of *Escherichia coli*, OP50. The handling of animals is very easy, where they can be used as a scientific tool by a high school student or an undergraduate, making it a simple model organism for young scientists.

The life cycle of the worm starts in the hermaphrodite gonad and embryos at the 28-cell stage are laid into the environment. The embryo hatches within the next 24 hours, depending on the incubation temperature, because the animal's life cycle can be sped up or slowed down by manipulating the incubator temperature, which is another interesting feature of the worm. Embryos grow from the first larval stage to the fourth in the next few days by molting to become reproductive adults. Under stressful conditions, the worm goes to an alternative developmental stage called the dauer stage, which can survive

harsh conditions. When the stress is relieved, the dauer stage animals molt to the fourth larval stage (L4) and further complete the life cycle. The nematodes can be frozen also, in liquid nitrogen tanks at -80^{o}C and animals can be recovered after years in the freezer.

Another interesting feature of *C. elegans* is its transparency where each cell can be easily visualized under the microscope. The transparency of the worm allowed mapping of the whole cell lineage during embryogenesis and post-embryogenesis (Sulston and Horvitz 1977) from a single cell through all of the divisions that give rise to 959 somatic cells in hermaphrodites and 1031 cells in males. Each cell division and differentiation is tightly controlled, and therefore, factors playing a role in division and differentiation of a cell can be easily investigated. The transparency was advantageous to this work since GFP was used to observe gene expression in different cells at different stages of worm development. GFP can be used with different approaches. Here, a defined promoter region of the gene of interest was fused upstream to *gfp* sequences to study gene expression and regulation of the predicted promoter. The *gfp* reporter construct was studied in live animals at different stages of worm development from embryo to larvae to adult animal.

The worm has two sexes: self fertilizing hermaphrodites and males, which make the organism a powerful tool for genetics. Homozygous animals can be easily propagated through self fertilizing hermaphrodites with no mating necessary. Additionally mutations can be brought together by using crosses between males and hermaphrodites. This tool

was advantageous to this work because mutant strains of CeTwist and CeE/DA were crossed with the wild-type animals containing the *gfp* reporter constructs to study the expression in mutant animals while comparing with the wild-type nematodes. *C. elegans* short generation time of 3-4 days and an average brood size of 300-1000 offspring was beneficial to us since expression of *gfp* reporter constructs was studied in hundreds of animals throughout development.

Another advantage of *C. elegans* is its small genome size of 97 Mb (compared to human 3000 Mb) which was completely sequenced in 1998 (*C. elegans* sequencing consortium, 1998). It was discovered that 40% of genes associated with human diseases have homologues in *C. elegans* (Mango et al., 2000). In *C. elegans* CeTwist and its target genes have human homologues, mutations of which cause different craniosynostotic diseases (Fig. 2). Our goal is to identify novel targets of CeTwist that might have human homologues mutated in other craniosynostotic diseases. There is only one Twist homolog in *C. elegans*, which is encoded by the *hlh-8* gene in contrast to vertebrates where there are six Twist family members. Having only one homolog of the *C. elegans* gene reduces the redundancy in its function and makes it easy to study the protein, in this case, CeTwist. Moreover, Twist is required for some early essential developmental events in *Drosophila* and mouse generating nonviable loss-of-function mutants as compared to viable null mutants in *C. elegans*. Therefore, there are many reasons that make *C. elegans* a favorable model organism to study Twist. Altogether, the nematode *C. elegans* is an

excellent model for this work because of its completely sequenced genome, transparency, one homolog for Twist, and its many unique features.

CeTwist targets previously identified and characterized

In *C.elegans*, many CeTwist and CeE/DA heterodimer targets have been identified, e.g. *ceh-24*, *arg-1*, and *egl-15* (Harfe et al., 1998, Wang et al., 2006). The *ceh-24* has CATATG E boxes in its promoter and mutation of this E box or its flanking sequences, abolishes its activity. Similarly the E box CATATG was important for *egl-15* activity (Harfe et al, 1998). In a detailed study of the *arg-1* promoter, it was found that CeTwist homodimers and heterodimers might have preferential binding to specific E boxes and also, that different E boxes might be regulating the tissue-specific expression pattern. The *arg-1::gfp* is expressed in the head mesodermal cell, vulval, uterine and enteric muscles. The data suggested that the E box, CATCTG is predicted to be involved in all of the tissue-specific expression of *arg-1::gfp*. The CATATG E box was predicted to be required for head mesodermal cell and vulval muscle expression and the CACATG E box for enteric muscle expression (Zhao et al., 2007). New evidence suggests that *arg-1* might be a target of CeTwist homodimers (T/T) along with the CeTwist and CeE/DA heterodimer (T/E) where T/T might be binding to an E box for enteric muscle expression

and the vulval muscle expression might be controlled by T/E binding to another E box in the promoter of *arg-1* (Philogene et al., 2012).

Focus of this study, to identify and characterize novel homodimer targets

The recent evidence suggests that CeTwist homodimers are required for M patterning and enteric muscle development whereas CeTwist heterodimers might be required for vulval muscle development (Philogene et al., 2012). CeTwist null mutants have abnormal M cell division along with defective enteric and vulval muscles causing a constipated and egg-laying defective phenotype (Corsi et al., 2000). Tethered Twist homodimers can rescue the abnormal M cell patterning and constipated phenotype of CeTwist null mutants but not the egg-laying defective phenotype. Moreover, *hlh-2* mutants have normal M patterning, and do not have defects in enteric muscles but partially formed vulval muscles. Together, these results suggest the role of CeTwist homodimers in early M lineage patterning and enteric muscle development whereas the role of CeTwist and CeE/DA heterodimers are in the development of vulval muscles. Also, CeTwist null mutants have no properly formed vulval muscles (Corsi et al., 2000) as compared to *hlh-2* deletion mutants. This difference might be because of residual CeE/DA activity in the mutant animals or it could be because CeE/DA is not an obligate partner of CeTwist in vulval muscles such that CeTwist alone or another unidentified partner might be functioning in vulval muscle development. The latter result is also

supported by the evidence that *egl-15::gfp*, a CeTwist target, does not lose expression in reduced *hlh-2* function and also, *hlh-8* tethered homodimers cannot rescue the *gfp* expression of *hlh-8* mutants containing *egl-15::gfp* which is not expressed in the mutant animals suggesting a role of another CeTwist containing dimer.

Thus, the previous evidence suggested CeTwist homodimers were playing an important role in *C. elegans*, and we predicted we would identify novel targets of CeTwist homodimers by overexpressing CeTwist. We overexpressed CeTwist with the help of an inducible heat shock promoter and found the potential target genes that are overexpressed using a global gene expression measurement technique called Affymetrix oligonucleotide microarrays. We identified 107 overexpressed genes by this technique and genes were then studied on the wormbase website (www.wormbase.org) to obtain information regarding the description of the gene, protein domains, RNAi information, homologues in other organisms and known expression pattern. Genes which have expression that is coincident with CeTwist (data identified from wormbase) or the genes which have a higher value in microarrays were chosen further to make *gfp* reporter constructs. Out of 19 genes which were chosen to make *gfp* reporter constructs, we identified five genes which have an expression pattern partially coincident with CeTwist. When these genes were further validated in CeTwist and CeE/DA mutants and *hlh-2* RNAi, we found one CeTwist homodimer target, two heterodimer targets and two CeTwist targets that need further characterization. In the next step, to identify the region

regulating the expression that is coincident with CeTwist expression pattern, we made 5'deletion reporter constructs. Our results suggest that CeTwist homodimer targets might be regulated by a palindromic E box similar to other bHLH homodimer targets. In addition, out of the five genes identified two genes have human homologues whose mutation may causes disease similar to disorders where Twist may play a role.

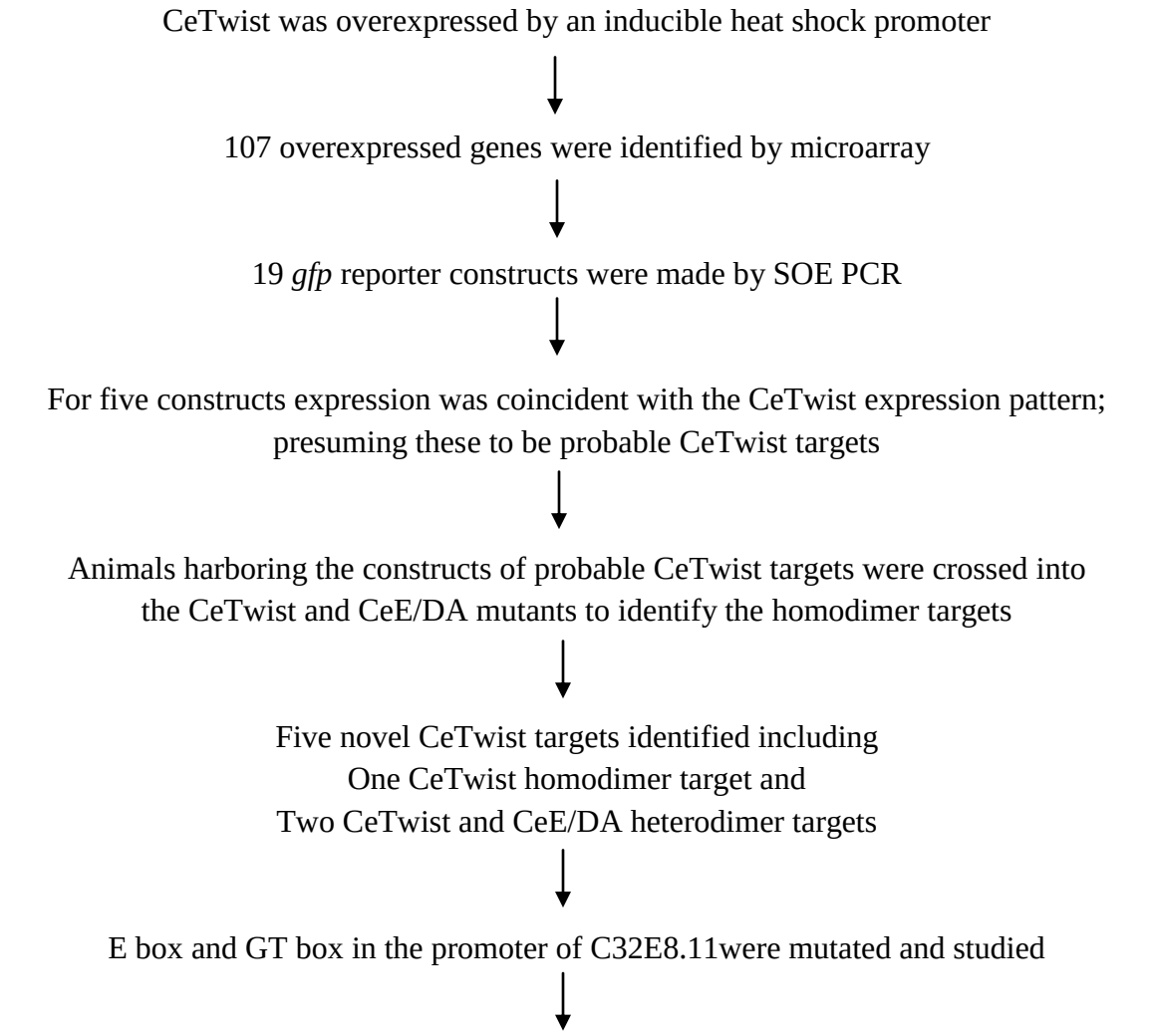
Experimental Design

The goal of this project was to identify novel homodimer targets of CeTwist which is encoded by the *hlh-8* gene in nematodes. The approach we used was to overexpress *hlh-8* with the help of an inducible heat shock promoter to find the potential target genes that are overexpressed using a global gene expression measurement technique known as Affymetrix Oligonucleotide Microarrays. We made two strains: experimental strain (exp) with the heat shock promoter driving *hlh-8* and a marker of transgenesis and the control strain with the transgenic marker (con). Both the strains were treated with (+) or without (-) heat shock, and thus, we have data in four sets exp(+), exp(-), con(+), con(-).We expected that when *hlh-8* is overexpressed in the experimental animals with the inducible heat shock promoter exp(+), CeTwist will bind and overexpress its targets. These targets will be upregulated when comparing mRNA expression levels in exp(+) to con(-) and con(+). We identified 107 overexpressed genes while comparing the results obtained from exp(+) and con(+). We predicted that the overexpressed genes which have expression patterns partially overlapping with the CeTwist expression might be the probable targets of CeTwist. Out of the list of 107 overexpressed genes identified, 19 genes were prioritized to study further based on their higher microarray value, multiple E boxes and homologues in other species. Of the 19 genes, *gfp* reporter constructs were generated and the expression pattern was studied. We found five candidate genes that have expression partially overlapping with CeTwist. To

further validate these genes to be true targets of CeTwist, the animals harboring the reporter constructs were crossed into the CeTwist and CeE/DA mutants. CeE/DA is the only known binding partner identified of CeTwist (Grove et al., 2009, deMasi et al., 2011). We predicted that the true CeTwist homodimer targets would lose expression only in the CeTwist mutant whereas a heterodimer target would lose expression in both the mutants and a target of another transcription factor should not lose any expression in both the mutants.

After identifying the probable targets of CeTwist, we made a series of 5' deletion reporter constructs to study the predicted promoter region. In each new reporter construct, a region containing an E box was deleted. For most of the targets, we identified the smallest promoter region required for the vulval muscle expression. For homodimer target C32E8.11, the smallest region showing expression in the vulval muscles has an E box and a 22 bp conserved region. The conserved region has a GT box which is an E box in the opposite orientation. A GT box was previously shown to be important for expression patterns of an *arg-1::gfp* reporter, another CeTwist target. To identify the role of the E box and the GT box in the promoter of C32E8.11, we mutated the E box and GT box individually, and the reporter constructs were then injected into the wild-type animals to study the expression.

A summary flowchart for identification of targets of CeTwist homodimers



MATERIALS AND METHODS

Nematode strains and maintenance

Nematode strains were maintained and studied at 20^{o}C (Brenner 1974), unless noted otherwise. Strains used in this study are: N2 wild type, PD4444 *ccls4444 [arg-1::gfp; dpy-20(+)*] , *hlh-8 (nr2061)* referred to as *hlh-8(-),* and *hlh-2(tm1768)* referred to as *hlh-2(*Δ*)* in this work. The *arg-1::gfp* reporter is a known heterodimer target of CeTwist and CeE/DA and is used as a control in the *hlh-2* RNA interference assay (RNAi) (Zhao et al., 2007, Philogene et al., 2012). The reporter is expressed in the vulval muscles (vm) along with the head mesodermal cell (hmc) and enteric muscles (em). The *hlh-8(-)* mutant is a presumptive null mutant for CeTwist with a deletion of 1267 bp that removes its HLH domain (Corsi et al., 2000). The *hlh-2(*Δ*)* is characterized as a partial loss of function mutant for CeE/DA with the deletion of 633 bp and an insertion of 18 bp (Chesney et al., 2007, Philogene et al., 2012) (Table 1.1, 1.2)

GFP reporter constructs

GFP reporter constructs were made of candidate CeTwist targets, by Sequence Overlap Extension (SOE) PCR (Hobert et al., 2002), to study the expression pattern of these genes. In general for the *gfp* reporters, the 5' upstream region of 3-4 kb from the start site ATG was used to make the construct whenever possible. If there is a gene

immediately upstream (<1 kb) the entire intergenic region was used. In cases where genes have a first intron of >300 bp in size, the intron was included in the reporter construct along with part of Exon 1.

In the first round of the PCR, a defined region (described above) of the predicted promoter was amplified from *C. elegans* genomic DNA and in the second round of PCR, the *gfp* vector pPD95.67 was amplified. In the third set, a nesting PCR was performed that fused the amplified promoter with the *gfp* vector product through the nucleotides complementary to *gfp* that were added to the amplified promoter in the first set of PCR. All three sets of PCR were performed by using the Roche Expand Long Template PCR System (catalog no#11681834001) (Hobert et. al, 2002). A typical PCR reaction of 50 µL contained 5µL of 10X PCR Buffer 2, 2 µL of 10 mM dNTPs, 2.5 µL of 6 µM forward and reverse primers, 0.75µL of Expand Long Template Enzyme Mix, 2.5 µL of N2 worm lysate (in the fusion PCR, 1µL of the primary product) and autoclaved milliQ water. PCR conditions used were: one cycle of 95^{o}C for 2 min, followed by 10 cycles at 95^{o}C for 10 sec, 56^{o}C for 30 sec and 68^{o}C for 4 min (for PCR products of 3-5 kb), followed by 20 cycles of 95^{o}C for 10 sec, 56^{o}C for 30 sec and 68^{o}C for 4 min 30 sec with addition of 30 sec at 68^{o}C in each step and finally 68^{o}C for 7 min. The extension time varied based on the length of the product (Fig. 4)

For C32E8.11, cloning was also used to generate the *gfp* reporters. Primers were designed with the restriction sites *Pst*I and *BamH*I to amplify the 636 bp upstream

of the ATG. The multiple cloning site of the *gfp* vector pPD95.67 was used to clone the 636 bp PCR fragment of C32E8.11.

The GFP reporter constructs, either the plasmid (100 µg/ml) or SOE PCR product (50-80 µg/ml) plus the roller marker pRF4 (50 µg/ml), which contains the dominant *rol-6 (su1006)* allele, were injected into the N2 wild-type animals to obtain the transgenic lines. Multiple lines have been obtained, except for two of the constructs, to study the expression pattern. For all the lines, expression was studied at different stages of the worm life cycle under the fluorescent compound microscope (Table 2).

The line generated by cloning for the C32E8.11 predicted promoter has almost a similar expression pattern as the lines generated by the SOE PCR product except that the percentage of animals showing the expression was lower with the plasmid.

The GFP reporters of five probable CeTwist homodimer targets in the *hlh-8(-)* mutants

GFP reporters from probable homodimer targets were crossed into the *hlh-8(-)* mutants by standard mating techniques. The homozygous strain of *hlh-8(-)* animals can be recognized easily by its obvious phenotype of constipation (Con) and egg laying defect (Egl) (Corsi et al., 2000). The matings and the expression studies were performed at 20^{o}C and the *gfp* reporter constructs in the wild-type background were used as control, which was maintained in the similar conditions. Expression of the reporter constructs was

recorded as present (+) or absent (-) when observed under the fluorescent compound microscope.

The GFP reporters in the *hlh-2(Δ)* mutants

We predicted that the expression of probable homodimer targets of CeTwist will persist in the loss of CeE/DA, the only known binding partner of the CeTwist. For this experiment, we used *hlh-2(Δ)* animals which are temperature-sensitive loss-of-function mutants at 25^{o}C. The embryos, L1, L2 or L3 *hlh-2(Δ)* larvae when moved to the non-permissive temperature, become sterile and many of them had a protruded vulva (Pvl). Whereas, when late L4 or young adult *hlh-2(Δ)* worms are moved to 25^{o}C, these animals lay a few eggs before becoming sterile. The eggs can hatch and grow to the adult stage with sterile and Pvl phenotypes. Reporter constructs from probable CeTwist targets were crossed into the *hlh-2(Δ)* animals at the permissive temperature of 20^{o}C and expression of the reporter constructs for the five genes were studied at the non-permissive temperature of 25^{o}C under the fluorescent compound microscope where the reporter in the wild-type background was used as the control. To study the vulval muscle expression in the loss of CeE/DA for the five candidate genes, late L4 animals were moved to 25^{o}C and lay a few eggs before becoming sterile. When the eggs hatch and grow to become late L4 larvae or adults, the vulval muscle expression was studied in these animals which were 100% sterile suggestive of the loss of CeE/DA (Philogene et al., 2012).

hlh-2 RNA interference

To examine the role of CeE/DA in the expression of the five probable targets of CeTwist, an RNA interference assay (RNAi) was done for *hlh-2*. Briefly, HT115 (DE3) bacteria containing a ds *hlh-2* vector or a control L4440 empty vector were grown overnight in culture without induction followed by seeding NGM plates containing IPTG (0.35 mM) and ampicillin (100 µg/ml) with the culture and growing at room temperature overnight. Feeding at the L4 stage of the *C. elegans* does not generate viable embryos suggesting the role of *hlh-2* in embryogenesis is critical (Krause et al., 1997, Kamath et al., 2001). To bypass this critical role, L1 worms were used for the RNAi. They were generated by hypochlorite treatment of gravid adults grown at 20^{o}C, and the embryos obtained from this treatment were allowed to hatch on NGM plates with no food overnight. Synchronized L1 animals were then washed off and transferred to the RNAi plate. Animals with reporter constructs were moved every 24 hr by picking and expression was studied only after 48 hr of exposure to RNAi (Kamath et al., 2001) (Fig. 8).

Homologous alignment of promoter region for two genes in distantly related *C. elegans* species

Sequences were obtained from www.wormbase.org for *C. elegans*, *C. remanei* and *C. briggsae*, nematode species. ClustalW alignment to identify homologous regions was generated from http://www.ebi.ac.uk/Tools/clustalw2/index.html.

Sequential deletion of the 5' promoter region for CeTwist targets

A set of 5' deletion reporter constructs were made for four out of the five novel CeTwist targets. SOE was used to generate more *gfp* reporter constructs with the deletion of regions containing an E box for genes, C32E8.11, D1046.4, F08D12.7 and W06H8.8. For C32E8.11 gene, a total of 4 constructs were made with sequential deletion of regions containing 1-2 E boxes where the largest 3kb SOE has 4 E boxes and 2 GT boxes and the smallest 2 kb SOE has none of the E boxes and GT boxes. For D1046.4, a total of 2 constructs were made where the larger 4 kb construct has 4 E boxes and the smaller 3.5 kb construct has 3 E boxes. Similarly for F08D12.7 and W06H8.8, two constructs were made for each gene where the larger 3 kb construct have 4 and 2 E boxes, respectively, and the smallest 2 kb SOE construct has no E box.

Site-directed mutagenesis of the E box and the GT box in the C32E8.11 promoter

A Quick Change Lightning Site Directed Mutagenesis Kit (catalog#210518) was used to mutate the E box and the GT box in pNS51 containing the 636 bp from the start site ATG of the gene C32E8.11 that is required for its expression in the vulval, body wall and pharyngeal muscles. The sequence of the E box, CACGTG, was mutated to AACGAG and the sequence of the GT box, GTCCAC, to GACCAA. Three constructs were made, one wild type with no mutation and two with each element mutated individually, each mutated construct was sequenced to confirm that only the desired mutation was obtained. The mutated constructs were then injected with the roller marker (pRF4) into N2 wild-type animals to obtain the transgenic lines. The expression of the transgenic animals containing the wild-type and the mutated constructs was studied under the fluorescent compound microscope.

RESULTS

The approach to identify novel CeTwist homodimer targets

To identify homodimer targets, the approach we took was to overexpress the CeTwist protein encoded by the *hlh-8* gene in the nematode, *C.elegans*, by using an inducible heat shock promoter and finding the potential target genes that are overexpressed using a global gene expression measurement technique called Affymetrix Oligonucleotide Microarrays. Two strains were made: the experimental strain containing heat shock promoter driven *hlh-8* plus the roller marker (exp) whereas the control strain (con) has only the roller marker. One group of synchronized L2 animals from both the strains received heat shock treatment for 33^{o}C for 20 min followed by recovery at 20^{o}C for 40 minutes; these animals are referred to as exp(+) and con(+) and the second group of untreated animals are referred to as exp(-) and con(-). Using this approach, we predicted we would find genes which were upregulated due to CeTwist overexpression as compared to control *C. elegans* where there is a normal amount of CeTwist when comparing exp(+) with exp(-), con(-) or con(+). The animals containing the heat shock plasmid can express transcription factors from the heat shock promoter even when kept at low temperature (Wang et al., 2006). Therefore, to avoid the leaky nature of the heat shock promoter and to be consistent, we compared the results obtained from the exp(+)

group with the con(+) group of animals. The microarray experiment was done in triplicate where three sets of two groups of animals (exp and con) were treated with or without heat shock. We identified 107 overexpressed genes when comparing the results of exp(+) to con(+) in three sets of experiments.

In the first step, we studied all 107 genes on the wormbase website (wormbase.org release WS 170-200) and collected the information regarding gene descriptions, protein domains, homologues in the other organisms, expression patterns, RNAi results and any other information available (Table 2). CeTwist and its targets participate in signal transduction, transcriptional regulation and cell fate determination in the mesoderm, and therefore, we prioritized those genes that participate in similar processes and are expressed in mesoderm. Based on the above information, we could prioritize only 3 genes to study further because for approximately 80 genes, not much information was available, and for the rest of the genes information present on the wormbase website was suggestive of no relation to CeTwist or its targets. Therefore, in the second step to prioritize genes, we set a high priority for genes which have microarray values >1.7 overexpressed when comparing the results of exp(+) with con(+) while comparing three set of experiments. We added 12 more genes based on the above criteria to our list. In the next step, we compared the results of the microarray values of two sets of experiments leaving the value of the third set of experiment which is offset to the values obtained in the other two experiments. This allowed us to add 4 more genes with

microarray values >1.7 to our list. Thus in total, we prioritized 19 genes out of the 107 to study further based on known expression patterns and the microarray values.

Expression pattern of the *gfp* reporter constructs

We predicted that if a gene has an expression pattern partially overlapping with CeTwist expression from the list of overexpressed genes then, it might be a probable CeTwist target. Therefore, we constructed the *gfp* reporters for each prioritized gene to study the expression patterns.

We used up to 3-4 kb upstream from the initiator start site ATG of each gene, if available, to make the constructs. In genes where there is a gene immediately upstream (~1 kb) the entire intergenic region was used to make the reporter construct. Since the 5' upstream regulatory sequence of each gene was fused with the *gfp* cDNA, we expected the expression pattern of the GFP will reflect the expression of each gene. In addition, evidence suggests that the larger introns (>300 bp) might contain transcriptional control elements (Jantsch-Plunger et al., 1994, Meyers et al., 2010). Out of the list of 19 genes, 2 genes have >300 bp in the first intron, and therefore, introns with part of the first exon for these two genes were included to make the GFP reporter construct.

Out of 19 genes, we got multiple transgenic *C. elegans* lines of reporter constructs for 17 and only one line for two of the genes. The GFP expression was studied in all the lines and at different stages of the life cycle from embryos to late adult.

CeTwist, like other bHLH transcription factors requires only 6 bp CANNTG (E box) to bind DNA (Harfe et al., 1998, Zhao et al., 2007, Meyers et al., 2010) which might occur every 256 bp in a gene. Therefore, in an artificial heat-shock system with overexpressed of CeTwist protein, it might also bind to an E box of non-target genes. Therefore, we knew by overexpressing CeTwist to identify its targets, we might also overexpress genes that are non-CeTwist targets. Moreover, we used only 1-4 kb upstream of the ATG to make the reporter constructs, thus we might have missed enhancer elements of these genes if they are present further upstream or downstream or in the intronic region of the gene, giving us no expression or not a complete expression pattern for the genes. Thus, based on the above information and similar experiments done before (Wang et al., 2006), we grouped the *gfp* expression pattern of 19 genes in three different categories; the expression pattern coincident with the CeTwist expression making them potential CeTwist target genes, expression not co-incident with the CeTwist or no expression (Table 3).

CeTwist is expressed in the 4 enteric muscles (2 intestinal muscles, anal depressor and anal sphincter), all undifferentiated cells of the M lineage and the mature vulval muscles (Harfe et al., 1998, Wang et al., 2006). Out of the 19 genes, we found five candidate genes which have expression similar to the CeTwist expression pattern, C32E8.11, D1046.4, F54D7.4, F08D12.7 and W06H8.8. For all the five genes, expression was present in the mature vulval muscles. In addition to these cells, for

C32E8.11::gfp expression was also present occasionally in the anal muscles (15/100 animals), *D1046.4::gfp* was expressed in the uterine muscles along with the intestinal and anal muscles, and F08D12.7 was also expressed in the intestinal and anal muscles (Fig. 5, Table 3). There were 4 genes where no expression was identified and for the remaining 10 genes expression was not coincident with the CeTwist targets e.g. C06G8.2 was expressed in the vulval epithelial cells along with head and tail neurons. Therefore, we identified five probable CeTwist targets based on their expression pattern partially overlapping with the CeTwist expression pattern.

Studying expression of probable CeTwist targets in the *hlh-8* null mutant for further validation

To validate the five candidate genes to be true CeTwist targets, animals having the constructs were crossed into the *hlh-8(-)* mutants to study the *gfp* expression in the absence of CeTwist. In the wild-type nematodes, vulval muscles differentiate from sex myoblasts (sm) precursors present in the center of the animal at the L4 stage and look like a V-shaped structure in a lateral view (Sulston et al., 1977) whereas in the *hlh-8(-)* mutants, the vulval muscles are non-functional and only partially differentiated from SM descendants seen as small cells [also known as vm-like cells (Corsi et al., 2000)] in the vulval region. Previous evidence has shown that the reporters of known CeTwist targets are not expressed in the *hlh-8(-)* mutants whereas *gfp* reporters of non-targets normally

found in the vulval muscles e.g., *myo-3* & T12D8.9 were expressed in vm-like cells (Corsi et al., 2000, Wang et al., 2006). We predicted that the novel targets of CeTwist identified in this project would lose expression in *hlh-8(-)* animals (Wang et al., 2006). Out of the five genes, three genes (C32E8.11, D1046.4, and F54D7.4) lost expression in 100% of the animals (n = 45, 90, 45) whereas the other two genes F08D12.7 and W06H8.8 lost expression in approximately 75% of the animals (n = 50, 41) (Fig. 6). Based on the above experiment, we identified five genes which lose vulval muscle expression partially or completely in the loss of CeTwist, suggesting they are probable targets of CeTwist.

hlh-2 is required for the expression of some reporters of the CeTwist candidate genes

To test if the above identified novel CeTwist targets are CeTwist homodimer targets or CeTwist and CeE/DA heterodimer targets, the *gfp* reporters of the five genes were crossed into the partial loss-of-function mutant of CeE/DA *hlh-2(Δ*) to study the vulval muscle expression in the decreased amount of CeE/DA. CeE/DA is the only known binding partner of CeTwist in *C. elegans*. It was shown previously that *arg-1::gfp*, a known heterodimer target of CeTwist and CeE/DA, lost expression in the vulval muscles in the *hlh-2(Δ)* mutants whereas another target *egl-15::gfp* retained vulval muscle expression in the adults. We predicted that the reporter gene of CeTwist homodimers would not lose expression in the *hlh-2(Δ)* mutants. We found that in the

C32E8.11::gfp, vulval muscle expression was retained in 90% of the animals (n=50). F08D12.7 and W06H8.8 retained expression in 84% and 78% of the animals (n=50 and 41) respectively whereas D1046.4 and F54D7.4 lost expression in 90% and 50% of the animals (n=80) (Fig. 7). These results suggested that CeE/DA is not required for the expression of vulval muscles for C32E8.11, F08D12.7 and W06H8.8, three of which retained expression in a significant number of the *hlh-2(Δ)* animals whereas the role of CeE/DA is significant for the vulval muscle expression in F54D7.4 and D1046.4 *gfp* reporter constructs.

RNA interference of the *hlh-2* gene reinforces the results of *hlh-2(Δ)* mutants

To confirm our results with the *hlh-2* deletion mutant, we used RNAi to decrease *hlh-2* expression in animals containing the five reporter constructs. In this experiment, L1 animals were fed *E. coli* HT115 (DE3) harboring ds RNA from *hlh-2* and then the treated late L4 and young adults were observed for *gfp* expression in the vulval muscles. We used *arg-1::gfp* as a control which significantly loses its expression in RNAi for *hlh*-2 (Zhao et al., 2007, Philogene et al., 2012). It was shown previously that *hlh-2* RNAi treatment of *egl-15::gfp* reporter causes vulval muscle defects where *gfp* expressing cells near the vulval opening are deformed (Philogene et al., 2012). Results of *hlh-2* RNAi for the five CeTwist candidate constructs were similar to the results obtained from the *hlh-2(Δ)* mutants where *gfp* reporters for genes, C32E8.11, F08D12.7 and W06H8.8 retained

expression in 78%, 80% and 72% of the animals (n=35,80 and 80), respectively, whereas D1046.4 and F54D7.4 lost expression in 90% and 40% of the animals (n=80). Also, *gfp* reporter constructs for genes, C32E8.11, F08D12.7 and W06H8.8 showed deformed vulval muscle cells near the vulval opening similar to *egl-15::gfp* reporter (Fig. 8). The *hlh-2* RNAi results were similar to the results obtained from the *hlh-2(Δ)* mutant indicating that, CeE/DA is not required for the vulval muscle expression of C32E8.11, F08D12.7 and W06H8.8 whereas *gfp* reporter constructs of D1046.4 and F54D7.4 require CeE/DA for the vulval muscle expression.

Homology in the promoter regions of two genes in distantly related *C. elegans* species

Important promoter elements are expected to be conserved throughout the evolution. The promoter regions present in the *gfp* reporter construct of the five genes were examined for homology in distantly related *C. elegans* species, *C.elegans, C. briggsae* and *C. remanei.* We found 2 conserved regions in two of the genes. In D1046.4, 21 bp are conserved with an E box CATCTG at 300 bp from the start site ATG (Fig. 11). In the C32E8.11, we found a conserved region of 22 bp with one of its GT boxes (Fig. 10). The conservation in C32E8.11 region and the GT box was not a perfect match. We could not find any other region of homology when comparing the promoter region of the three other CeTwist candidate target genes

Identifying the region responsible for vulval muscle expression by the sequential 5' deletions in the promoters of the CeTwist targets

To identify the region responsible for the vulval muscle expression in four out of the five CeTwist candidate target genes, a set of 5' deletion reporter constructs were made (Fig. 12). For C32E8.11, the original SOE construct has four E boxes: CATTTG and CAAATG at around 1200 bp, CATGTG at 1050 bp, CACGTG at 450 bp and a GT box at 200 bp upstream from the ATG start site. When deleting the region containing the CATTTG and CAAATG in one construct and then further deleting a region containing the CATGTG E box, both constructs do not lose expression in any cell type. Further deleting the region containing the CACGTG E box, along with the GT box, the expression was lost in all the cell types (Fig. 12) The CACGTG E box and the GT box were then further evaluated by site-directed mutagenesis in the next section.

In the D1046.4 SOE reporter, there are four E boxes CATCTG at 1.6 kb and 300 bp where the E box near to the ATG start site is conserved, CAAATG at 800 bp and a palindromic sequence CAGCTG in the first intron. When deleting the region containing the E box, CATCTG, at 1.6 kb, the construct loses expression in the vulval muscles along with the enteric and anal muscles, but is retained in vulval-uterine precursors, excretory cell soma, anterior and posterior canal cells. It has been shown previously while studying the promoter of *arg-1* that CeTwist heterodimers can bind to a CATCTG E box in an Electrophoretic Mobility Shift Assay (EMSA) and also, this E box is crucial for

expression since mutating it causes loss of expression in all cells where *arg-1* is expressed (Zhao et al., 2007). We expect that the E boxes CATCTG near to the start site ATG and CAGCTG may play some role in the expression of *D1046.4::gfp* that can be further evaluated by mutational analysis in future studies. For F08D12.7 and W06H8.8 genes, even with losing all the E boxes present in their 5' upstream region, there is no change in the *gfp* expression in the vulval muscles. Therefore, we can conclude that for three different groups of CeTwist targets, there must be different E boxes or sequences for regulation.

The palindromic sequence, CACGTG, in the promoter of C32E8.11 is responsible for the vulval muscle expression

The gene C32E8.11 was overexpressed in the microarray experiment while overexpressing the *hlh-8* and the *C32E8.11::gfp* is expressed in the vulval muscles and anal muscles, partially overlapping the CeTwist expression pattern. These two results suggested that the gene C32E8.11 might be a CeTwist target. Further, the vulval muscle expression of *C32E8.11::gfp* is completely lost in the absence of CeTwist (Fig. 6) whereas the vulval muscles expression is retained in the loss of CeE/DA (Fig. 7, 8) where CeE/DA is the only known binding partner of CeTwist in *C. elegans*. Based on these validation experiments, we predicted that C32E8.11 is the best candidate for a CeTwist homodimer target. In the next step, we were interested to identify the elements, in the

promoter of C32E8.11, controlling the vulval muscle expression. Previously, it was shown that the alteration of the sequence of the E box can eliminate its function (Harfe et al., 1998). Moreover, evidence suggests that the homodimers prefer a palindromic sequence to bind DNA for the regulation (Grove, de Masi et al., 2009). In a minimal promoter region for the gene C32E8.11 that is expressed in the vulval, body wall and pharyngeal muscles, there is a palindromic E box, CACGTG, and also a conserved region of 22 bp with a GT box in it. A GT box is an E box in an opposite orientation that may have some role in the expression of CeTwist targets (Zhao et al., 2007). Thus, we were interested in studying the GT box along with the palindromic E box in the promoter of C32E8.11. To investigate the role of the E box and GT box, site-directed mutagenesis was used. The E box, CACGTG, was mutated to AACGAG and the GT box, GTCCAC, was mutated to GACCAA. The mutated constructs were then injected into the wild-type animals and expression was studied. Interestingly, we found that mutating the E box, CACGTG, leads to loss of expression only in the vulval muscles whereas the expression is retained in body wall and pharyngeal muscles (Fig. 13). For the construct with a mutation only in the GT box expression is lost in the pharyngeal muscles but not in the vulval muscles and body wall muscles (Fig 13). The results from the expression of the mutated constructs suggested that the palindromic E box might be the required E box for the CeTwist homodimer binding for its role in the vulval muscle expression whereas the GT box which lies in between the stretch of a conserved sequence might be the binding

site of another transcription factor, which is required for the expression in pharyngeal muscles.

DISCUSSION

CeTwist, a bHLH transcription factor, forms homodimers and heterodimers with CeE/DA. After dimerization, the dimers bind to an E box, CANNTG, in target genes to regulate transcription. Several targets of CeTwist in *C.elegans* have human homologs which are mutated in different craniosynostotic diseases. New target genes of CeTwist might help us to identify novel genes that are mutated in other human diseases. In addition, because we do not fully understand the mechanism by which CeTwist functions, we predict that novel targets of CeTwist homodimers along with the known heterodimer targets will help us to understand how CeTwist regulates these different targets in mesoderm development. In this study, we identified one homodimer target, two heterodimer targets and two other CeTwist targets that need further characterization. Also, we predict that CeTwist homodimers regulate target genes through a palindromic E box, CACGTG, similar to other bHLH homodimers.

Identification of novel targets of CeTwist

Affymetrix oligonucleotide microarray, a global gene expression measurement technique, was used to identify novel targets of CeTwist. Upregulated genes, in the experimental animals that have more CeTwist as compared to control *C. elegans* where

there is a normal amount of CeTwist, were identified as potential targets of CeTwist. Out of 107 overexpressed genes, we prioritized 19 genes to study further based on the information available on the wormbase website and higher values in microarray experiments. We found five novel CeTwist targets which met the criteria designed to identify novel targets of CeTwist. First, they were overexpressed in the oligonucleotide microarrays while overexpressing CeTwist, and secondly the *gfp* reporters of these genes were expressed in a subset of mesodermal cells where CeTwist is expressed such as vulval and enteric muscles (Fig. 5, Table 4). GFP reporter constructs were made for the 19 genes and 5 *gfp* reporters had an expression pattern coincident with CeTwist, 10 *gfp* reporters were not coincident with CeTwist and for the rest no expression was seen. We predict that for the 14 genes where expression was not coincident with CeTwist, it might be because while making *gfp* reporter constructs we are using only 1-4 kb of 5' upstream region from the ATG start site thus potentially missing important regulatory elements located further upstream, downstream or in the intronic region. In addition, in the microarray experiment, we might have genes that are false positives which were overexpressed because they contain multiple E boxes where CeTwist does not normally bind but in the presence of excess CeTwist are now activated. An example of a gene that might be in this category is C56E6.2, which has 11 E boxes in its 5' promoter region. Out of 11 E boxes one is CATATG and three are CATCTG which are potential CeTwist binding sites. Thus, in normal conditions CeTwist might not bind to these E boxes

whereas when a large amount of CeTwist was present it might have bound to these E boxes leading to overexpression in the microarray. Therefore, the expression of its *gfp* reporter does not overlap with the CeTwist expression since C56E6.2 is not a true target of CeTwist.

Categorizing the CeTwist targets into regulated CeTwist homodimers, CeTwist and CeE/DA heterodimers or undetermined CeTwist-containing dimers

The only known binding partner of CeTwist in *C. elegans* is CeE/DA. Thus, the expression of a bonafide CeTwist homodimer target should be decreased or lost in the absence of CeTwist but it should be retained when CeE/DA levels are decreased. We tested our five probable CeTwist targets in *hlh-8* null mutants to observe whether the expression was found in the vm-like cells in *hlh-8*. For three out of five reporter constructs, CeTwist was found to be necessary for the expression, C32E8.11, D1046.4 and F54D7.4. The other two genes, F08D12.7 and W06H8.8, also lost expression in 70% of the animals in contrast to three non-targets of CeTwist: *myo-3* which loses expression only in 30% of animals, *cpn-3* in approximately 20%, and T12D8.9 in almost no animals (Corsi et al., 2000, Wang et al., 2006, Philogene et al., 2012). These results suggested that F08D12.7 and W06H8.8 might also be targets of CeTwist but may not absolutely require

CeTwist for full expression and therefore, fall in a category that needs to be further characterized.

To test for the role of CeE/DA in the five CeTwist targets, we studied the expression of five *gfp* reporters in the background of *hlh-2*(Δ) mutant and RNAi for *hlh-2*. We found that C32E8.11 loses its expression in less than 10% of the animals in the vulval muscles in the *hlh-2*(Δ) mutants and less than 20% in the *hlh-2* RNAi background. Similarly, F08D12.7 and W06H8.8 loses expression only in approximately 20% of the animals in both the experiments (p value = 0.07, 0.06 for the *hlh-2*(Δ) and the *hlh-2*RNAi experiment, respectively). Therefore, the results suggest that for three genes C32E8.11, F08D12.7 and W06H8.8, CeE/DA is not required for the vulval muscle expression. F54D7.4 and D1046.4 lose expression in approximately 50% and 90% of the animals respectively, in both the validation experiments (p value = 0.045, 0.001 for the *hlh-2*(Δ) and the *hlh-2* RNAi experiment, respectively) (Fig. 7, 9, Table 4) suggesting that CeE/DA is required for the vulval muscle expression in F54D7.4 and D1046.4.

Thus, we can categorize our five probable targets of CeTwist into three groups; in the first category of CeTwist homodimer targets, the reporter construct completely lost expression in the *hlh-8* null background but retained the expression in the *hlh-2(Δ)* background and *hlh-2* RNAi in a significant number of the animals where CeE/DA encoded by the *hlh-2* gene, is the only binding partner of CeTwist identified so far in *C. elegans*. In the second category of CeTwist and CeE/DA heterodimer targets, reporters

lost expression in the *hlh-8* null background along with the loss of expression in the *hlh-2(Δ)* and *hlh-2* RNAi experiments. And in the third category of CeTwist targets, which need further characterization, are *gfp* reporters of genes which lost some but not all expression in the *hlh-8* null background and also did not lose expression with the loss of CeE/DA in the *hlh-2(Δ)* or *hlh-2* RNAi. In the first category of homodimer targets, is gene C32E8.11 and in the second category of heterodimer targets, are D1046.4 and F54D7.4, and in the third category of targets of CeTwist are F08D12.7 and W06H8.8, which do not require CeE/DA for the expression similar to homodimer targets but are also not dependent completely on CeTwist.

These two targets of CeTwist F08D12.7 and W06H8.8 might fall in a category with *egl-15,* another CeTwist target. The *egl-15* gene is predicted to be a CeTwist heterodimer target with an unidentified bHLH partner of CeTwist (Philogene et al., 2012).

Identifying the region responsible for the vulval muscle expression in four out of five targets

Next, we studied the promoter region of novel CeTwist targets identified in this project. First, we made a set of 5' deletion reporter constructs to identify the region responsible for the vulval expression. In the promoter region of D1046.4, we predict an E

box CATCTG at 1.6 kb is responsible for the vulval muscle expression; loss of the region in which this E box resides also results in some loss of expression in anal muscles and intestinal muscles. We predict that the other CATCTG E box at 300 bp which is conserved in different nematode species might be playing some role in expression of enteric muscles. While studying an *arg-1::gfp* reporter, the authors have identified that CeTwist and CeE/DA heterodimers bind to a CATCTG E box and is required for the expression in the vulval, intestinal and anal muscles similar to our *D1046.4::gfp* construct (Zhao et al., 2007)

In F08D12.7, deleting a region that contains the E box, CACATG, results in partial loss of expression in intestinal muscles suggesting that this E box might be responsible for expression in intestinal muscles. But even the smallest reporter construct for F08D12.7 which has no E boxes in it, has expression in the vulval muscle, anal muscle and some expression in the intestinal muscles. Similarly for the W06H8.8, the smallest *gfp* construct with no E boxes has expression in the vulval muscle. We predict that for these two CeTwist targets the regulation might be different than the other CeTwist targets where CeTwist binds to an E box for the regulation, but first we need to determine if these smallest constructs actually depend on CeTwist for expression. ADD1/SREBP1, a bHLH transcription factor binds to an E box CACGTG and a non-E box ATCACCCCAC sequences. This binding to E box and non-E box depends on an atypical tyrosine amino acid in the basic region of ADD1/SREBP1. In the basic region of

ADD1/SREBP1, substitution of tyrosine to arginine found in most bHLH proteins abolished the non-E box binding but retained binding to E box (Kim et al., 1995). Therefore, we predict if we study the structure of CeTwist and the smallest promoter region of F08D12.7 and W06H8.8 in detail, we might be able to find the non-E sequences where CeTwist might bind to regulate these targets.

For C32E8.11, we identified a region containing an E box and a GT box that was required for vulval, body wall and pharyngeal muscle expression. We further mutated the E box and GT box to identify the region responsible for the vulval muscle expression.

Characterization of CeTwist homodimer target, C32E8.11

Many bHLH proteins form homodimers e.g. HLH-1, HLH-11, HLH-30, MXL-3, and others. The basic region of each dimer binds to the half-site CAN of the E box CANNTG. Therefore, it is predicted that homodimers might bind to palindromic sequences. It was inferred by using invitro DNA binding specificities, bHLH-DNA co-crystal structures and computational binding that the most preferential binding site for bHLH-homodimers is a palindromic sequence CACGTG followed by CAGCTG and very rarely other E boxes (de Masi et al., 2011). Also, other bHLH transcription factors, MyoD and Stra13 homodimers, preferentially bind to CAGCTG and CACGTG E boxes, respectively (Kophengnavong et al., 2000, St-Pierre et al., 2002). Similarly, the HEN1

protein that contains a bHLH motif binds to an extended E box that contains a CAGCTG sequence (Brown et al., 1994). These results lead to the hypothesis that the most preferential binding site for bHLH homodimers is a palindrome.

The *C32E8.11::gfp* reporter construct has a palindromic sequence CACGTG and a GT box that shares partial homology in three different nematode species (Fig. 11). Thus, we predicted that one of these elements might be responsible for the vulval muscle expression in the *C32E8.11::gfp* reporter. To identify the element in the promoter of C32E8.11 responsible for vulval muscle expression, we used site-directed mutagenesis. When the palindrome E box CACGTG was mutated to AACGAG, the *C32E8.11::gfp* reporter construct loses expression in the vulval muscles but retains it in body wall and pharyngeal muscles whereas mutating GTCCTC to GACCTA in the *C32E8.11::gfp* construct loses expression in pharyngeal muscles but retains the expression in vulval and body wall muscles leading to hypothesis that the palindromic E box, CACGTG, is responsible for the vulval muscle expression. After identifying a homodimer target, two heterodimer targets and two more CeTwist targets along with characterizing the promoter of homodimer target, we were interested to identify human homologues for the five novel CeTwist targets.

Human homologues of candidate genes

We examined five novel CeTwist targets on wormbase website to identify human homologues. We found human homologues for two genes C32E8.11 and W06H8.8. C32E8.11's human homologue is UBR1 & 2, mutation of which in humans causes Johanson-Blizzard syndrome which is a disease that is categorized into ectodermal dysplasia similar to Setleis syndrome which is caused by mutation of Twist2. The human homologue of W06H8.8 is titin. Mutation of titin in humans causes hypertrophy of heart and muscular dystrophies. The Twist family of bHLH transcription factors, including Hand1 and Hand2 along with Twist1 and Twist2 play role in heart and muscle development, respectively. Titin has protein domains of immunoglobin and protein kinases similar to *egl-15* which is another target of CeTwist containing dimers.

Future direction and significance of this work

We predicted that CeTwist homodimers might be functioning in M lineage patterning and enteric muscle development whereas heterodimers might be functioning in the vulval muscles. We have already identified heterodimer targets that are expressed in the vulval muscles. Therefore, we predicted we would identify CeTwist homodimer targets that are expressed in early M lineage and enteric muscles. But, we identified a novel CeTwist homodimer target that is also expressed in the vulval muscles where the

role of CeTwist and CeE/DA has already been identified. The results from this work suggest that in vulval muscle each CeTwist dimer might be regulating a different set of genes by binding to specific E boxes. This dimer-specific regulation may be similar to what is happening in the sutures in the mouse skull. In the different cells of mouse skull suture (mid-suture vs. osteogenic front cells), Twist homodimers and heterodimers are functioning by regulating unique sets of genes (Connerney et al., 2006). However in *C. elegans* in contrast to mice, we predict both the CeTwist-containing dimers are functioning in the same cells at the same time of development. We have already characterized the potential binding sites of CeTwist and CeE/DA heterodimers i.e., CATCTG and CATATG for vulval muscle expression. Here, in this work, we identified a CACGTG E box which is most likely binding CeTwist homodimers in vivo for regulation of vulval muscle expression. In addition, we have also identified two genes D1046.4 and F08D12.7, which are expressed in enteric muscles where CeTwist homodimers are predicted to function. We characterized D1046.4 as a CeTwist and CeE/DA heterodimer target and F08D12.7 as target of CeTwist containing dimer with an unidentified partner. We predict that, D1046.4 and F08D12.7 might be targets of CeTwist homodimers for the enteric muscle expression whereas other dimers might be regulating their vulval muscle expression similar to *arg-1* (Philogene et al., 2012). However, this prediction will require further study.

In this work, we have identified target genes which we predict are regulated by three different types of CeTwist-containing dimer i.e., CeTwist homodimer, CeTwist and CeE/DA heterodimer and CeTwist-dimer containing an unidentified partner. The results of 5' deletion reporter constructs suggest that the three different set of genes might be regulated by different sequences where a palindromic E box CACGTG might bind CeTwist homodimer in vivo for regulation of its targets and the already known E box CATCTG might be bound to CeTwist and CeE/DA heterodimers for regulation. We predict identifying more CeTwist homodimer targets along with studying in detail the promoter region of the three categories of five novel CeTwist targets might help us to elucidate the role of different CeTwist-containing dimers in *C. elegans* mesoderm development. More specifically, site directed mutagenesis of the E boxes and sequences present in these targets will shed more light on regulation by CeTwist. Moreover, we can also add more genes from the microarray list to make *gfp* reporter constructs to study expression. This time, we can also add another criterion for choosing these genes by studying E boxes in the promoter region and identifying genes which have the palindromic E box CACGTG. We have identified human homologues for two novel CeTwist targets which are mutated in human diseases, suggesting that *C. elegans* is an excellent model organism and study of CeTwist transcription might help us to understand about human diseases related to Twist function.

Conclusions:

We identified a CeTwist homodimer target C32E8.11 and identified a palindromic sequence that the CeTwist homodimer might be binding to in the promoter of C32E8.11 for regulation. Moreover, we have also identified two more novel CeTwist and CeE/DA heterodimer targets, D1046.4 and F54D7.4 and two targets of CeTwist, F08D12.7 and W06H8.8, which need to be characterized further. The biggest challenge will be to identify bHLH partners of CeTwist other than CeE/DA since we predict that unidentified partners function with CeTwist in *C. elegans* mesoderm development.

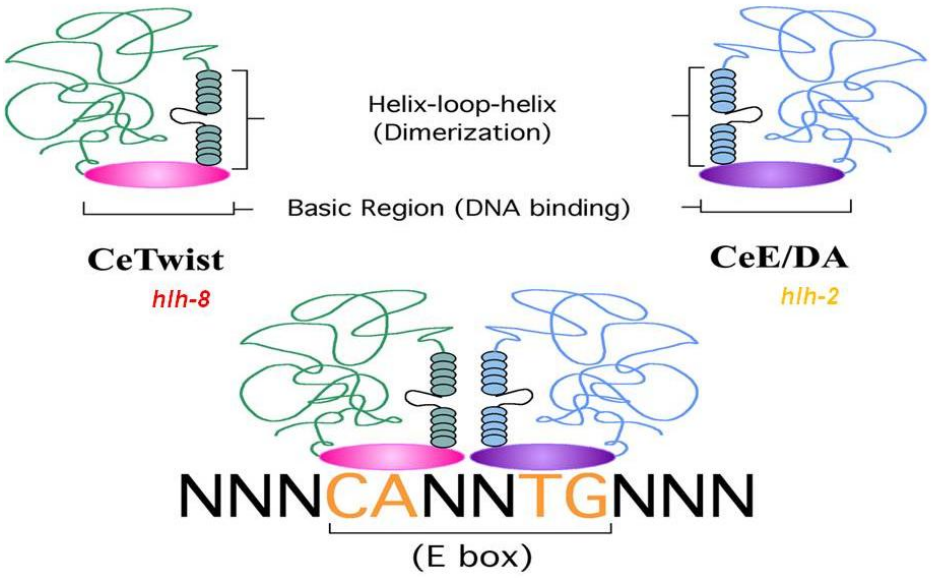


Figure 1. CeTwist as a homodimer and heterodimer. CeTwist is a bHLH protein which can form homodimers and heterodimers with its helix loop helix (HLH) region. The only known heterodimeric binding partner of CeTwist is CeE/DA. After dimerization, the basic region of the protein dimer can bind to a consensus sequence CANNTG, also known as an E box, on its target gene. It has been shown before that the 6 bases of this sequence are highly specific for different bHLH dimers binding and also important for differential tissue expression.

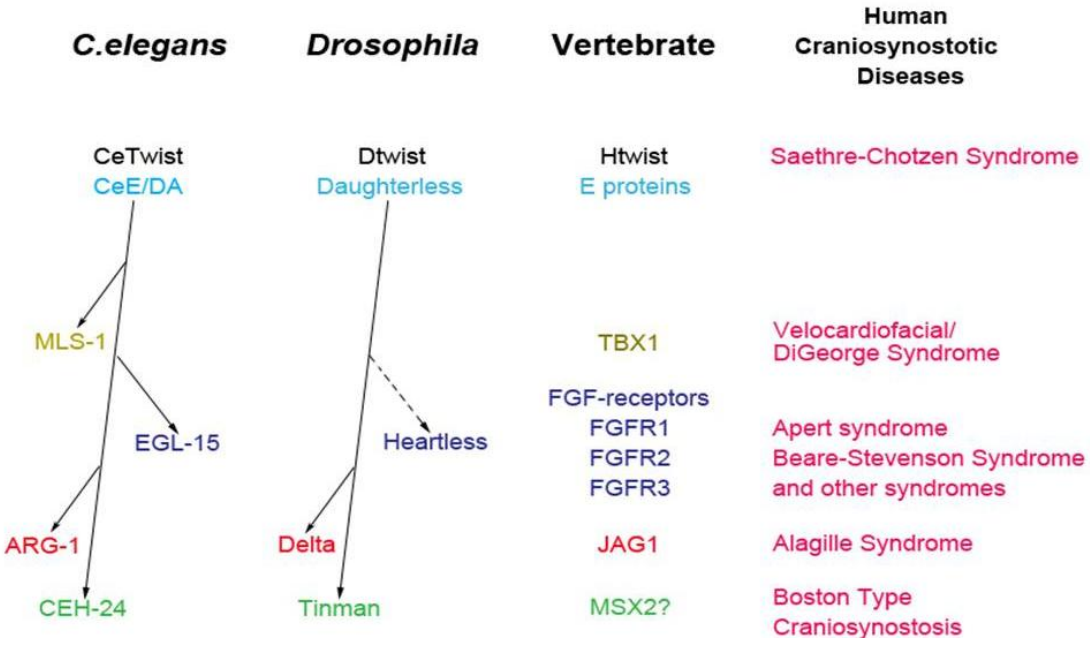


Figure 2. Role of CeTwist in human diseases. In humans, Twist mutations cause Saethre-Chotzen syndrome, which is a type of craniosynostotic disease where sutures of newborns fuse early in life. There are more than 100 diseases in this group of disorders and for approximately 30, single gene defects have been predicted but only for very few the concerned genes have been identified. In *C.elegans*, only the CeTwist heterodimer targets have been identified, several are shown here on the left which also have the human homologues (shown on the right side). So far, no CeTwist homodimer targets have been identified, and we predict that if we can identify novel homodimer targets, it will help us in understanding how the bHLH transcription factor CeTwist regulates mesoderm development as both a homodimer and heterodimer. Also, if these novel homodimer targets have human homologues, it may help us in understanding the craniosynostotic diseases.

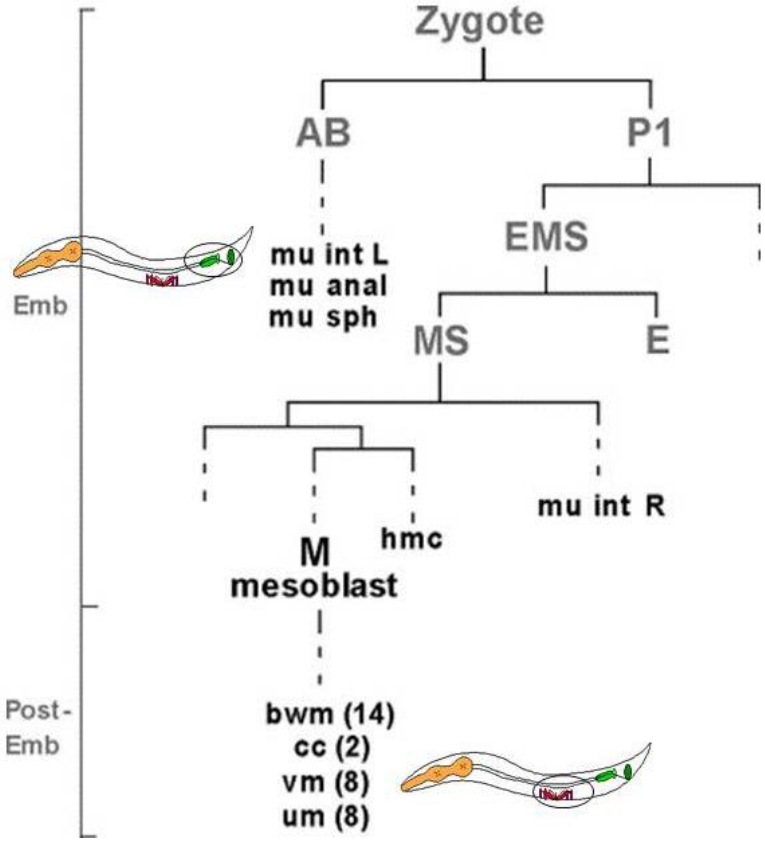


Figure 3. Mesoderm development and *hlh-8* expression pattern in the mesoderm of *C. elegans.* CeTwist is expressed in a group of enteric muscles, 2 intestinal muscles (mu int L & R), the anal depressor muscle (mu anal), and the anal sphincter muscle (mu sph) (these circled cells are depicted in the top worm diagram) and the head mesodermal cell (hmc) which arise embryonically (Emb) from different cell lineages. CeTwist is also expressed in the M mesoblast which gives rise to all postembryonic (Post-emb) nongonadal mesoderm: 14 body wall muscles (bwm), 2 coelomocytes (cc), 8 vulval muscles (vm) and 8 uterine muscles (um) (vm & um are shown in a circle in the bottom worm diagram). This single transcription factor is required in multiple tissues which develop from different cell lineages at different time intervals of development. Our goal is to identify novel homodimer targets that, together with already known heterodimer targets, will help us in understanding how bHLH transcription factors function in the mesoderm development of *C. elegans.*

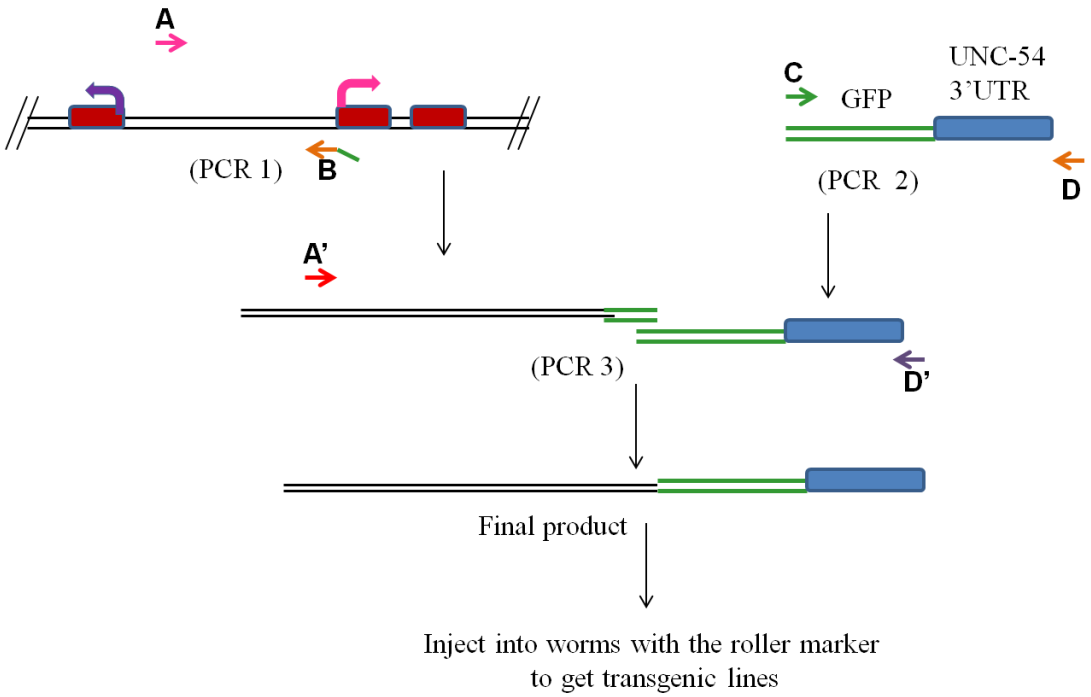


Figure 4. PCR fusion approach to generate *gfp* reporters. Sequence overlap extension (SOE) PCR was used to generate the *gfp* reporter constructs where three sets of PCR were performed. In the first set of PCR (PCR 1), a predicted promoter region was amplified (primers A and B) and in the second round of PCR (PCR 2) the *gfp* vector pPD95.67 was amplified (primers C and D). In the third set (PCR 3), a nesting PCR was performed (primers A' and D') that fused the amplified promoter with the *gfp* vector through the nucleotides complementary to the *gfp* that were added to the amplified promoter in the first set of PCR (shown with green line in primer B).
Black lines represent the DNA where red boxes are the exons of genes. Block arrows are the start site ATG where the purple arrow represents the gene of interest and the blue arrow represents the upstream gene. The horizontal arrows labelled, A, B, C, D, A' and D' are the primers used in the PCR.

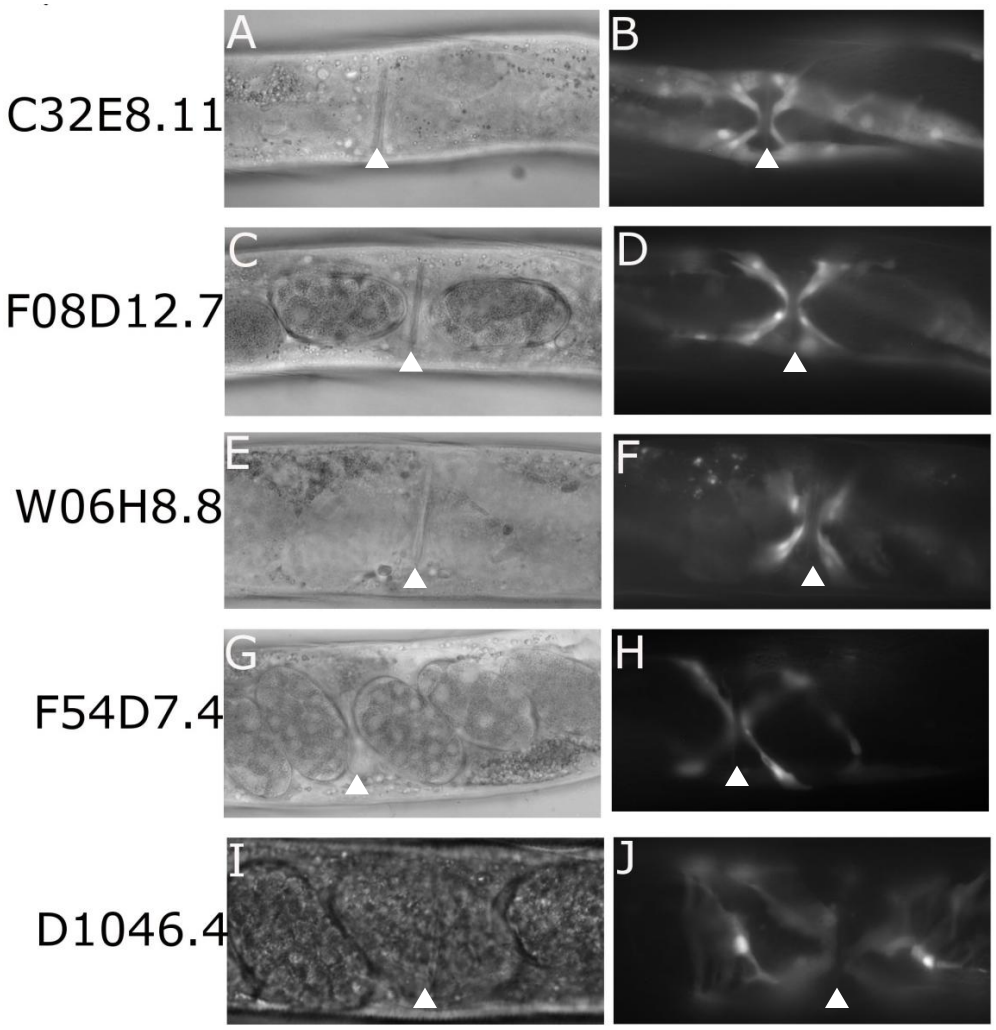


Figure 5. Expression pattern of probable CeTwist homodimer targets. Five genes show a *gfp* expression pattern partially overlapping the CeTwist expression pattern. In the micrographs, on the left are the Nomarski images (A, C, E, G, I) and on the right are the GFP images of the same adult animals (B, D, F, H, J). All the animals are oriented with the ventral side to the top and anterior to the left. The white arrowheads point to the vulval opening. In all the GFP images, the vulval muscles are seen as an "X" shape and the uterine muscles (in J) are seen as finger-like projections near the vulval muscles.

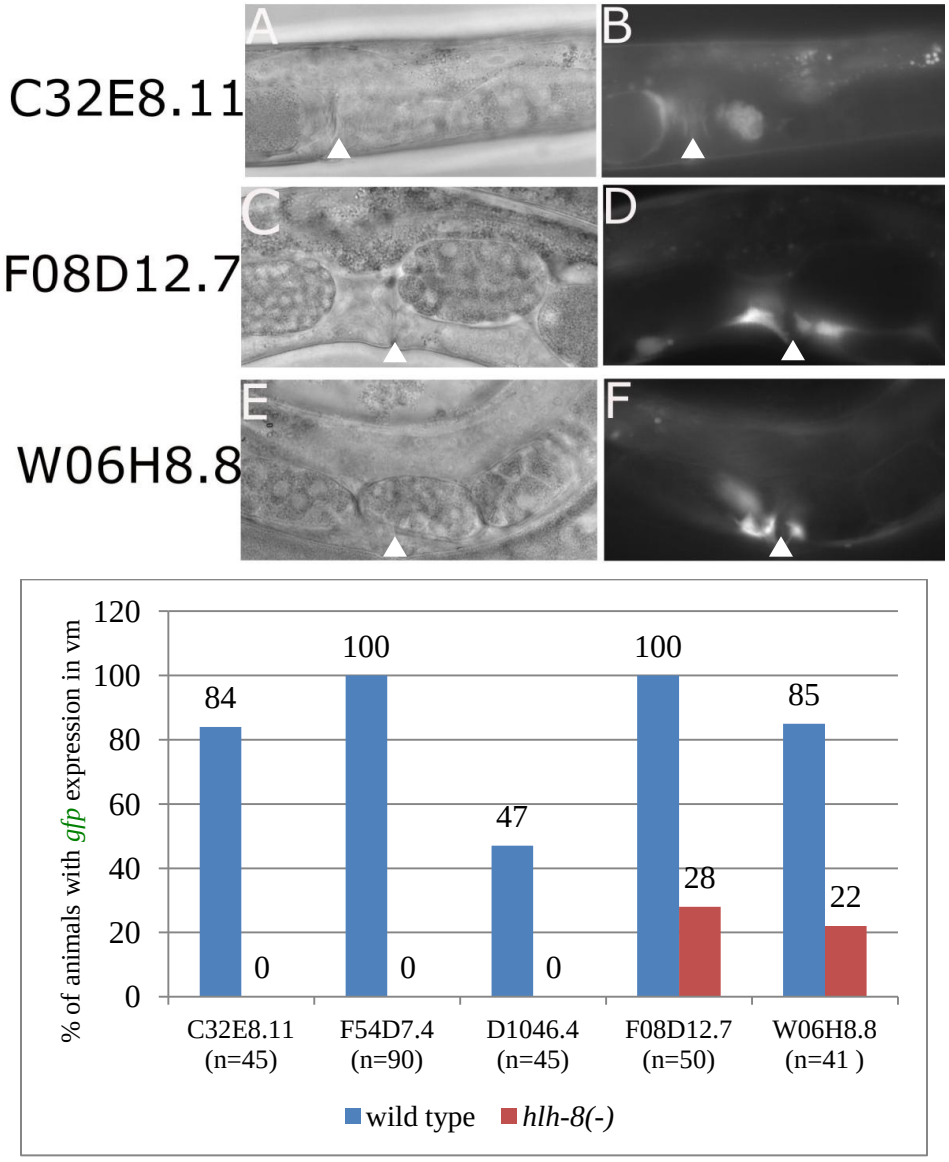


Figure 6. Expression of the reporter constructs in the *hlh-8* null mutant animals. Some GFP reporters lose expression significantly in the vm-like cells of the *hlh-8(-)* mutant. Representative micrographs of the vulval region is shown for three genes, C32E8.11, F08D12.7 and W06H8.8 in the *hlh-8(-)* mutants. Nomarski images (A, C, E) are shown on the left and GFP images (B, D, F) are on the right. In all the animals, the vulval opening is on the ventral side of the lateral views shown by the white arrowheads.

C32E8.11::gfp in the mutant does not show any expression in vm-like cells as compared to *F08D12.7::gfp* and *W06H8.8::gfp*. In the graph, the Y axis is the percentage of animals expressing *gfp* in the vulval muscle and on the X axis are different genes. In blue are the reporter constructs in the wild-type background and in red are the reporter constructs in the *hlh-8* null-mutants. For genes, C32E8.11, F54D7.4 and D1046.4 there is complete loss of expression in the null-mutants whereas for F08D12.7 expression is lost in the 72% of the animals and for gene W06H8.8 loss of expression is in 78% of the animals.

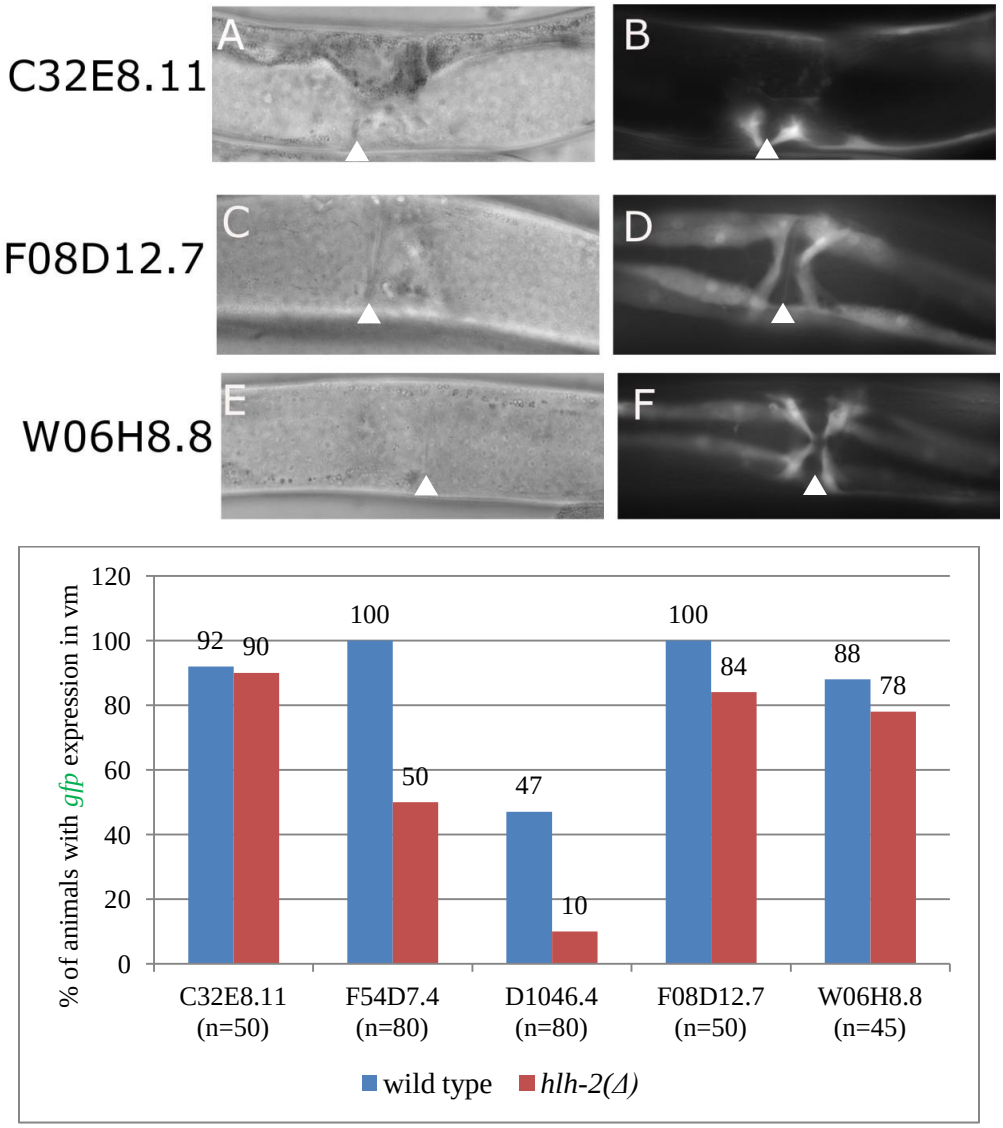


Figure 7. Reporter constructs in the *hlh-2* deletion background. In the micrographs Nomarski images (A, C, E) are on the left and the GFP images (B, D, F) are on the right where white arrowheads show the vulval opening. In the micrographs A & B the vulval opening is on the ventral side of the lateral view and in all the other images the vulval opening is on the ventral view. In the *hlh-2(Δ)* mutant, three of the reporters do not lose vulval muscle (vm) expression. In the graph, the Y axis is the percentage of animals expressing *gfp* in the vulval muscle and on the X axis are different genes. In blue are the

reporter constructs in the wild-type background and in red are the reporter constructs in the *hlh-2* deletion mutants. For C32E8.11, the *gfp* reporter does not lose expression in 90% of the *hlh-2* mutant animals, for genes, F08D12.7 and W06H8.8 in ~80% whereas *gfp* reporter constructs for F54D7.4 and D1046.4, lost expression in 50% & 90% of the animals, respectively.

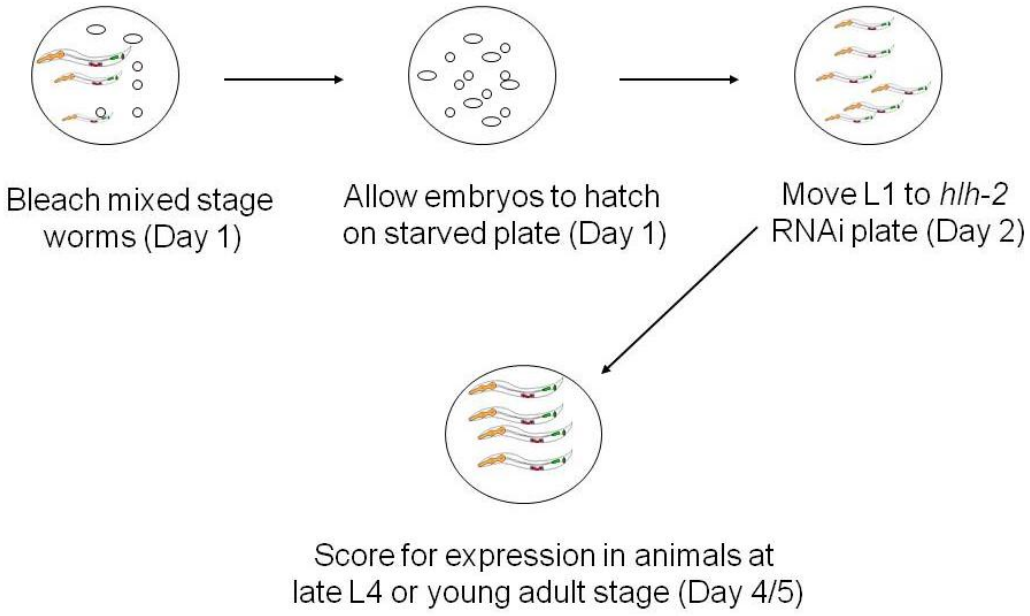


Figure 8. Experimental methodology for RNA interference (RNAi) . RNAi plates were made by growing overnight cultures of HT115 (DE3) bacteria containing a ds *hlh-2* vector or a control L4440 empty vector without induction followed by seeding NGM plates containing IPTG (0.35 mM) and ampicillin (100 µg/ml) with the culture and growing at room temperature overnight. L1 animals were generated by bleaching mixed stages of worms grown at 20^{o}C, and the embryos obtained from this treatment were allowed to hatch on NGM plates with no food. Synchronized L1 animals were then washed off and transferred to the RNAi plate. Animals with reporter constructs were moved every 24 hr by picking and expression was studied only after 48 hrs of exposure to RNAi (Kamath et al., 2000).

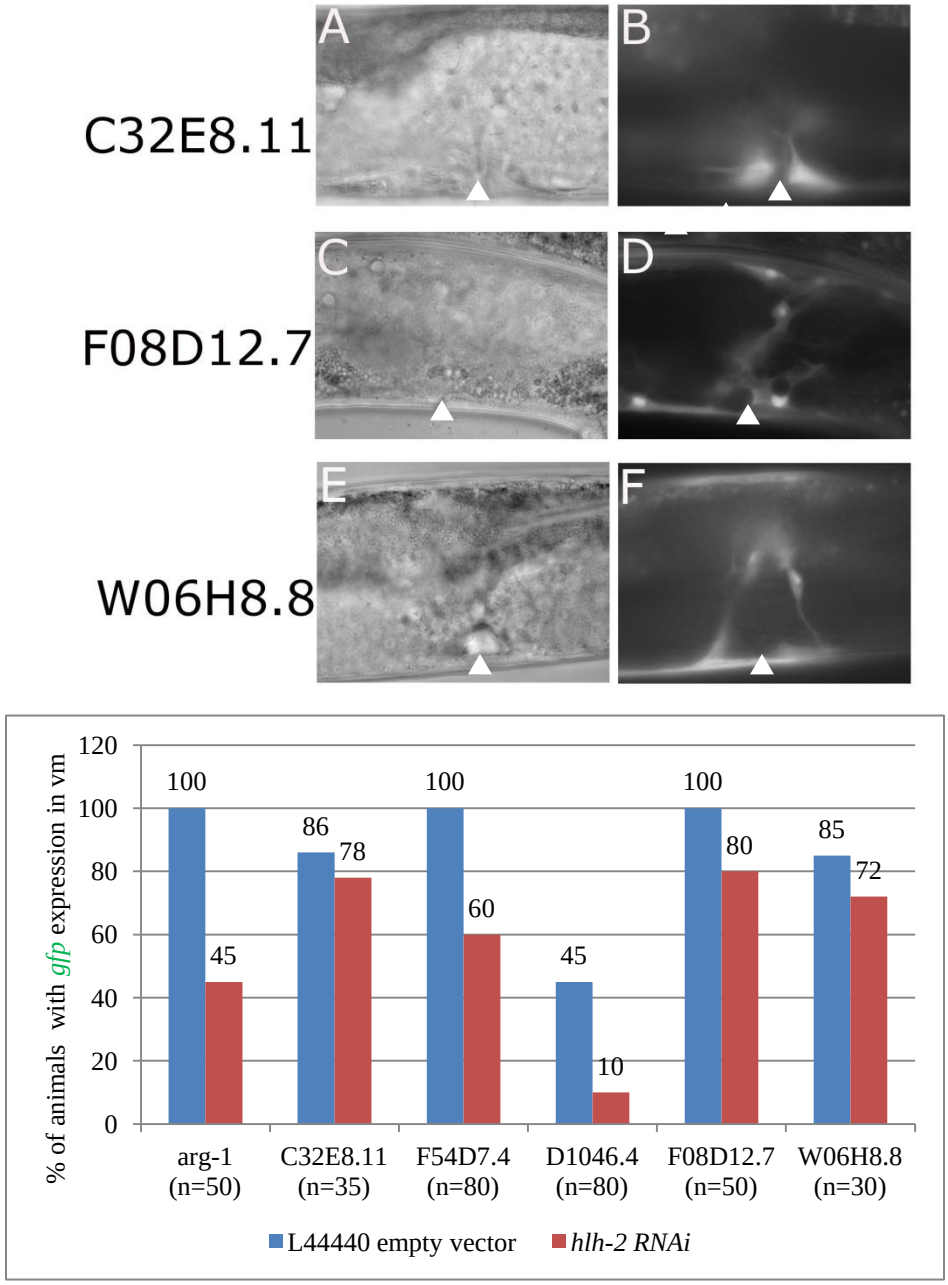


Figure 9. *hlh-2* RNAi for probable CeTwist targets. An *hlh-2* RNAi assay was performed on all five reporter constructs where *arg-1::gfp* was used as a control. Nomarski images are on the left side of the micrograph (A, C, E) and GFP images are on the right side (B, D, F) where white arrowheads show the vulval opening. In the

micrographs A & B the vulval opening is on the ventral side of the lateral view and in all the other images the vulval opening is on the ventral view.
Similar to the previous graphs, on the Y axis is the percentage of animals expressing *gfp* in the vulval muscle and on the X axis are different genes. In blue are the reporter constructs in the wild-type background and in red are animals treated with *hlh-2* RNAi. *arg-1::gfp* is used as a control which is a known heterodimer target and RNAi-treated animals lose expression in 55% of the animals. The C32E8.11::*gfp* reporter construct does not lose any expression in the RNAi background (A, B) whereas reporter constructs of F08D12.7 and W06H8.8 lose expression in approximately 20% of the RNAi-treated animals (C, D, E, F), as compared to F54D7.4 which loses expression in 40% of the animals and D1046.4 loses expression in 90% of the RNAi-treated animals.

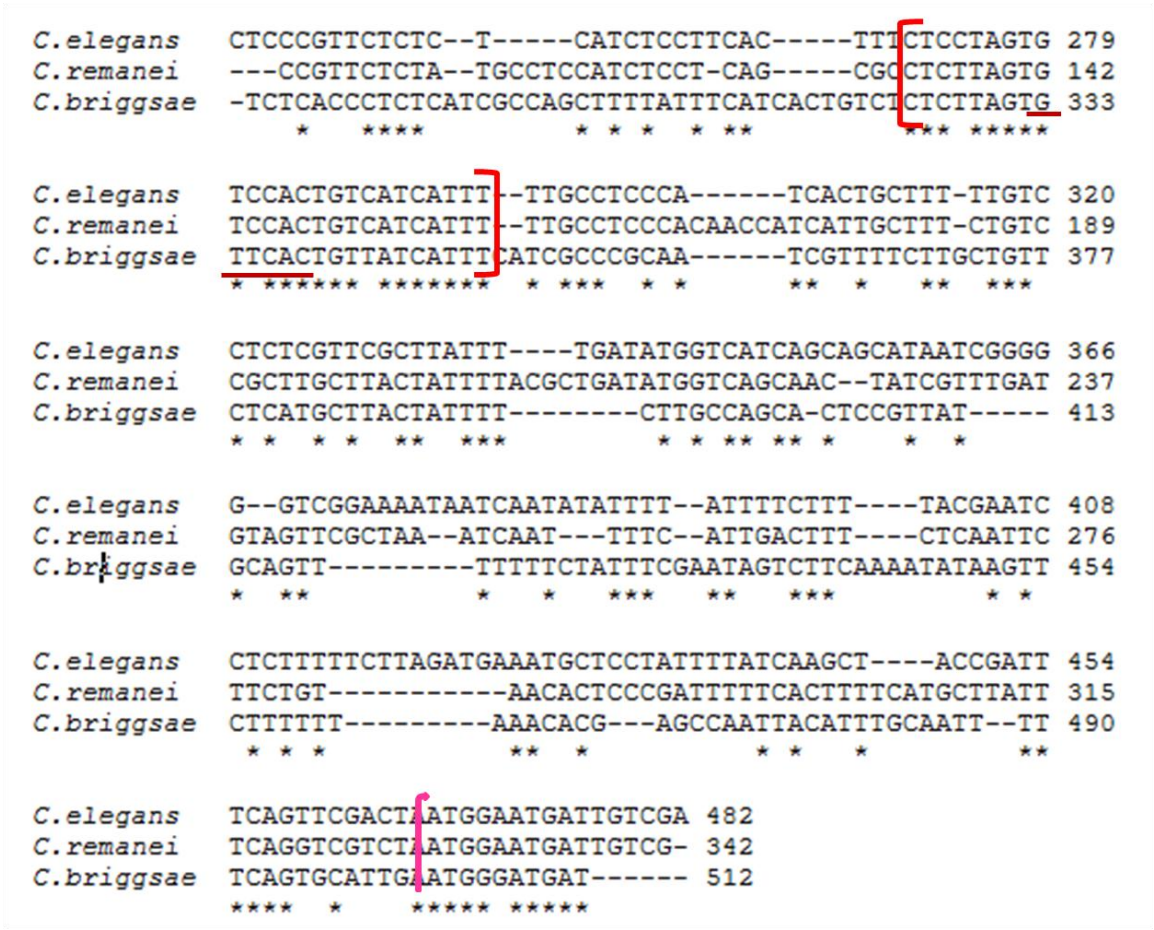


Figure 10. Conserved region in the promoter of C32E8.11 which contains a GT box. While comparing the promoter region of the C32E8.11 gene in the *C. elegans, C. remanei* and *C. briggsae* nematode species, we found a conserved region of 22 bp shown in the red bracket. In this conserved region lies the GT box GTCCAC which is an E box in an opposite orientation (underlined in red). The pink arrow head shows the start site ATG of the C32E8.11 in the three nematode species.

Figure 11. Conserved E box, CATCTG in the promoter of D1046.4. While comparing a part of the promoter regions of D1046.4 in *C. elegans, C. remanei* and *C. briggsae*, we found a conserved region (shown in blue bracket) containing the E box CATCTG (underlined in red).

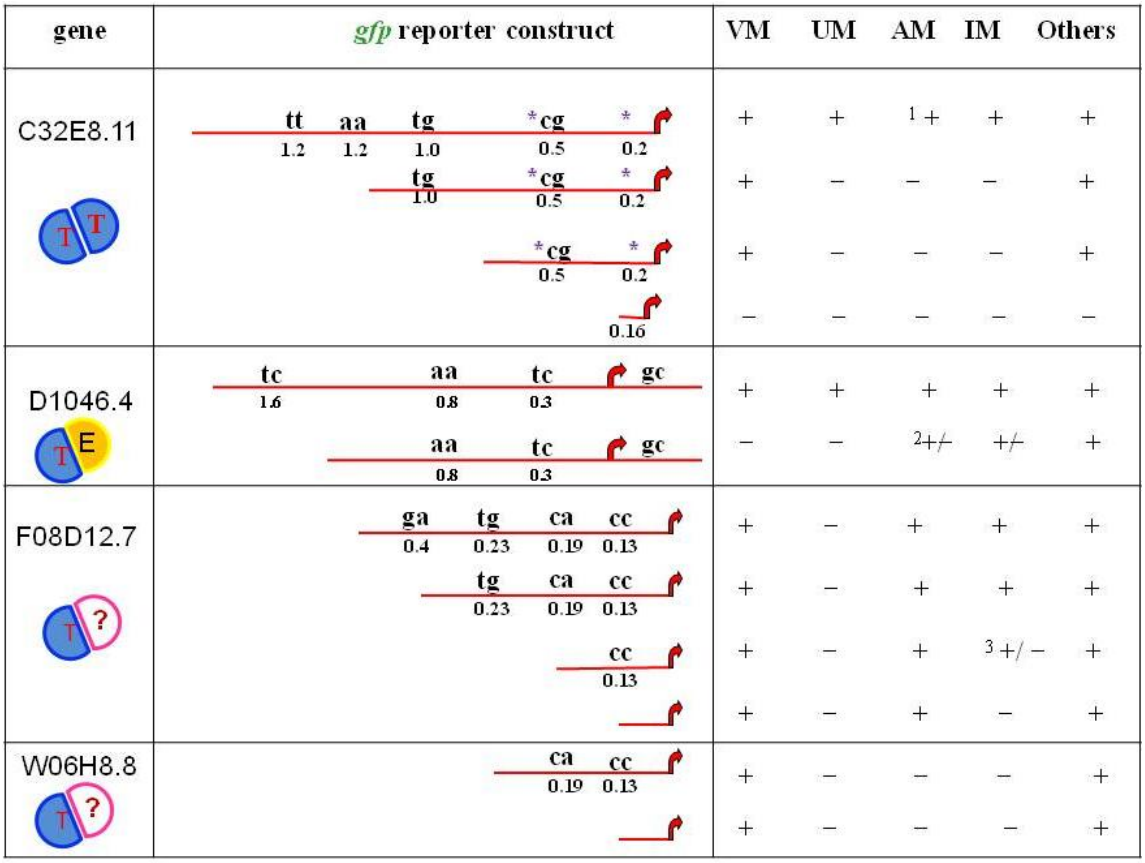


Figure 12. **Identifying the region responsible for the vulval muscle expression in CeTwist targets by deleting the 5' promoter region.** For C32E8.1, we identified a region of 532 bp from the ATG that shows expression in the vulval muscles along with expression in other cells. For D1046.4, we identified a region of 1700 bp which is required for the vulval, uterine, intestinal and anal muscle expression and with the loss of E box CATCTG at 1.6 kb expression was lost in the vulval and uterine muscles and decreased in the intestinal muscles and anal muscles whereas expression was retained in the other cells. For the F08D12.7 construct, we identified a region of 338 bp from the start site ATG which is responsible for the vulval, intestinal and anal muscle expression and losing of 188 bp from 338 bp to 150 bp that includes two E boxes CATGTG and CACATG decreases the expression only in the intestinal muscles but retains the expression in the vulval and anal muscles. In the W06H8.8 construct, 123 bp upstream of the ATG start site retains the expression in the vulval muscle and other cells.

* GT boxes, which are E boxes in the opposite orientation, are present at around 500 bp and 200 bp
[1] Anal muscle expression is present only in 15% of the animals
[2] Expression in the intestinal and anal muscles is present only in <50% of the animals as compared to ~80% with the larger construct which has the E box CATCTG at 1.6 kb
[3] Expression was present occasionally (<10% of the animals) as compared to the expression when a region containing the E boxes CATGTG and CACATG was present (90% of the animals)
Red arrowhead: The start site ATG of the gene of interest, red line: the predicted promoter
Two letters represent the middle two bases of the E box CANNTG.
Numbers: the distance in kb from the start site ATG
+: presence of expression, −: absence of expression
Others: expression pattern in tissues not co-incident with CeTwist expression
T: CeTwist, E: CeE/DA

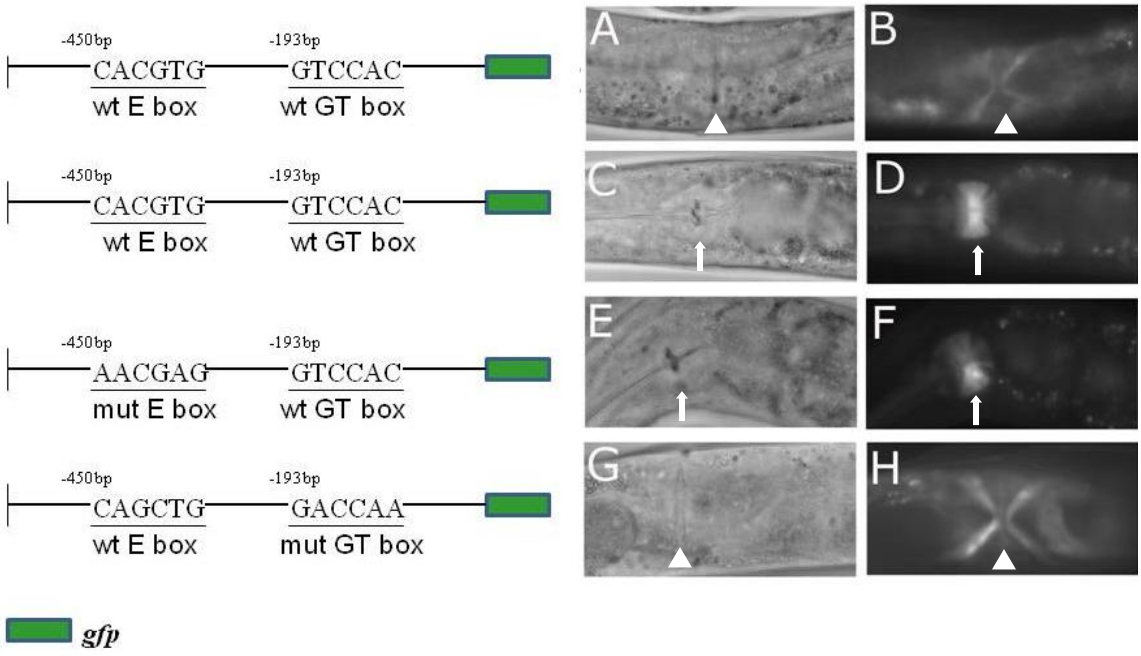


Figure 13. Site-directed mutagenesis of the E box and the GT box from C32E8.11. In the micrograph, *gfp* reporter constructs are shown on the left, Nomarski images in the center and GFP images on the right. White arrowheads represent the vulval opening and the arrows represent the pharyngeal muscles. All worms are shown with anterior to the left and the vulval opening in A, B, G, H is on the ventral view. In the presence of the wild-type E box and the GT box in the *C32E8.11::gfp* construct, GFP is expressed in the vulval and the pharyngeal muscles (A, B, C, D).When mutating the E box CACGTG expression is present only in the pharyngeal muscles (E, F) but there is loss of expression in the vulval muscles (not shown) whereas mutating the GT box leads to loss of expression in the pharyngeal muscles but expression is still present in the vulval muscles (G, H).

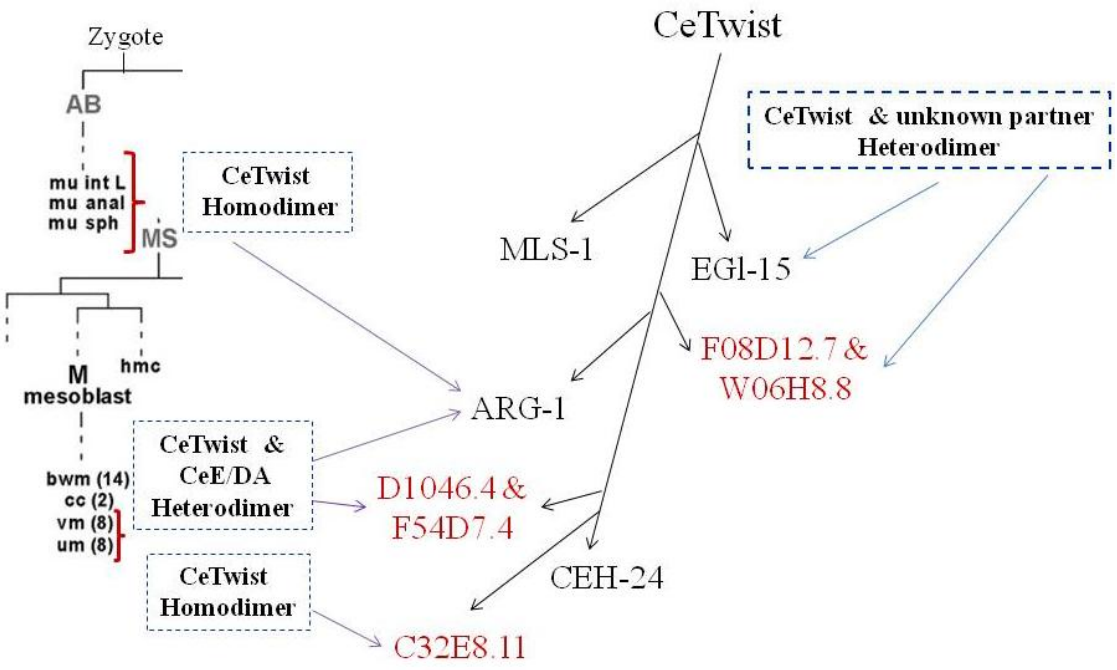


Figure 14. Model for the CeTwist regulation of target genes. On the left is shown the partial cell lineage of *C. elegans* with bold lettering depicting the cells where CeTwist is expressed. On the right in black letters, are previously identified CeTwist targets and in red are novel targets identified in this work. In dashed boxes are shown CeTwist homodimers and CeTwist and CeE/DA heterodimers that are predicted to regulate the individual genes.

Table 1.1 Plasmids used in this study

Plasmid	Description	From
pPD95.67	GFP cDNA with no promoter insert	Andrew Fire
pRF4	rol-6 *(su1006)*	Mello et al., 1991
hlh-2 RNAi	ds *hlh-2* RNAi vector	Zhao et al., 2007
L4440	L4440 RNAi empty vector	Zhao et al., 2007

Table 1.2 *C. elegans* strains used in this study

Strain	Genotype	G expression[a]/ Phenotype[b]
N2	Wild type	NA
PD4444	*ccIs4444*[*arg-1::gfp;dpy-20(+)*]*II*	the head mesodermal cell (hmc), vulval muscles (vm1) and 4 enteric muscles (em)
hlh-8(nr2061)	Null mutant of CeTwist	constipation (Con) and egg laying defective (Egl)
hlh-2(tm1768)	Partial loss-of-function mutant of CeE/DA	sterile, protruded vulva (Pvl) @ 25^{o}C

[a] GFP expression of the strain PD4444 *(arg-1::gfp)*

[b] Phenotype of *hlh-8(nr2061)* and *hlh-2(tm1768)* mutant worms

Table 2. Information available from the wormbase website for the 19 genes [a]

Gene	Gene Description	Protein domains	RNAi [b]	Known anatomic expression	Homo-logues
C06G8.2	proton-coupled oligopeptide transporter	TGF-beta receptor, type I/II extracellular region	NA	NA	NA
F28C6.2	F28C6.2 encodes an AP-2-like transcription factor	Transcription factor AP-2 , C-terminal	unc,emb	NA	H, M
D1046.4	Na(+)-dependent inorganic phosphate cotransporter	NA	NA	NA	NA
F54C9.4	encodes a member of the collagen superfamily containing collagen triple helix repeats required for normal body morphology	Collagen triple helix repeat	NA	NA	NA
K09C4.5	NA	Major facilitator superfamily	fat content increased	NA	H,D,S
ZK792.7	NA	Metallophosphoesterase ,Calcineurin-like phosphoesterase	NA	NA	NA
C32E8.5	NA	SMAD/FHA ,FHA domain	emb,,unc ,ste	NA	NA
T07G12.3	NA	Protein of unknown function DUF229	NA	NA	NA
Y76A2A.1	NA	PapD-like	NA	NA	NA
C49G7.5	NA	NA	NA	NA	NA
T24E12.5	NA	NA	NA	NA	NA
C09B8.4	NA	Proteinof unknown function DUF829	NA	NA	NA
C56E6.2	NA	Ras GTPase	NA	NA	NA

F22F7.6	NA	NA	NA	NA	NA
C32E8.11	NA	Zn finger	NA	NA	H,M,D, S
Y38C9A.2	NA	Phosphorylation site	NA	NA	NA
F54D7.4	NA	Immunoglobulin I-set domain	aldicarb resistant	Adult Expression: bwm, vm	NA
F08D12.7	F08D12.7 transcripts are classified as strictly maternal	NA	NA	Adult Expression: bwm, vm,am, int m,vnc, others	NA
W06H8.8	mutated in cardiomyopathy or tibial muscular dystrophy	Immunoglobulin subtype 2	NA	Adult Expression: bwm, vm	H, D

a Some of the information available from the www.wormbase.org is shown here for the 19 prioritized genes

b RNA interference (RNAi) information for the gene

NA: information not available on wormbase website

H*: Homo sapiens*, M: *Mus musculus* D: *Drosophila melanogaster*, S: *Saccharomyces cerevisiae*

emb: embryonic lethal, unc: uncoordinated, ste: sterile

bwm: body wall muscles, vm: vulval muscles, am: anal muscles, int m: intestinal muscles, vnc: ventral nerve cord others: expression was present in other cells

Table 3. Expression Pattern of the 19 candidate genes studied by transcriptional GFP Reporter constructs

Gene	Exp+/ Con+[a]	#of lines[b]	GFP Pattern[c]
C06G8.2	1.9	6	vulval epithelial cells, head and tail neurons
F28C6.2	1.7	2	head neurons, few intestinal cells
D1046.4	2	4	vulval and uterine m., 2 intestinal and anal m.
F54C9.4	2.2	3	body wall muscles, head and tail neurons
K09C4.5	1.5	1	no expression in hermaphrodites, an unidentified cell in tail of males
ZK792.7	1.7	4	head and tail neurons
C32E8.5	1.9	2	no expression
T07G12.3	2	2	no expression
Y76A2A.1	2.5	3	no expression
C49G7.5	2	3	few neurons
T24E12.5	1.9	2	few hypodermal cells and intestinal cells
C09B8.4	4.5	4	pharyngeal & hypodermal m., head neurons
C56E6.2	1.4	2	intestinal cells, vulval epithelial cells, head and body n.
F22F7.6	1.5	4	no expression
C32E8.11	1.9	3	vulval, body wall and pharyngeal m.
Y38C9A.2	1.9	1	vulval epithelial cells, hypodermal and pharyngeal m
F54D7.4	1.7	2	vulval and body wall m.
F08D12.7	1.4	3	vulval, 2 intestinal & anal m., body wall m., head n.
W06H8.8	1.9	2	vulval & body wall m.

m.: muscles

n.: neurons

[a] values of microarray while comparing group of experimental (exp+) and control animals (con+) under the heat shock treatment

[b] number of lines generated for each gene

[c] GFP pattern was recorded for the 19 genes in all the stages of worm's life cycle. Expression shown in red, partially overlaps with the CeTwist expression and thus, the gene might be the probable CeTwist target.

Table 4. Summary of the five probable CeTwist targets

Gene	Animals with no vulval muscle (vm) *gfp* reporter expression		
	hlh-8(-) [a]	***hlh-2(Δ)*** [b]	***hlh-2* RNAi** [c]
C32E8.11	100% (n=45)	10% (n=50)	22% (n=35)
F08D12.7	72% (n=50)*	16% (n=50)	20% (n=50)
W06H8.8	78% (n=41)*	22% (n=45)	18% (n=30)
F54D7. 4	100% (n=45)	50% (n=80)	40% (n=80)
D1046.4	100% (n=90)	90% (n=80)	90% (n=80)

[a] loss of expression (exp) in the vm for the *gfp* reporter constructs in the *hlh-8*(-) mutant

[b] loss of expression (exp) in the vm for the *gfp* reporter constructs in the *hlh-2(Δ)* mutant

[c] loss of expression (exp) in the vm for the *gfp* reporter constructs in the *hlh-2* RNAi

* In F08D12.7 and W06H8.8, there was a significant loss of expression in *hlh-8(-)* mutants (~70%) as compared to other non-targets of CeTwist e.g., *myo-3* (30%, n=20) [Corsi et al., 2000, Wang et al., 2007] but not 100% as found in the other CeTwist targets identified in this work or previously (Wang et al., 2007)

Table 5. List of primer sequences for the *gfp* reporter constructs for 19 genes

Gene	SOE (length bp)[a]	5'oligo A	5'oligo A2	3'oligo B
C06G8.2	5145	AC-582-O ttc caa gtg cat cta gag cg	AC-583-I gct acc gct acc tgg act tgg	AC584cat ata cga tcg cct ata cag aac g
Y38C9A.2	4489	AC-585-O ccc tac ggt acc agg tct ca	AC-586-I gcg cct tta agg agg act cg	AC-665 tcc gag aat gtc gtt tcc cac tg
D1046.4	4157	AC-588-O ccc tgt tac tct agg att gcc c	AC-589-I gtc cat ctg gat tga gtc ac	AC590ggc ggg aag atg atc att gg
F54C9.4	2680	AC-591-O gac ctt gac gag acc cat gt	AC- 592-I cgg gaa gtt cta gcg ttc gcc	AC593cat ttt tag acg atg tct gag tgt tcc
C32E8.5	2484	AC-609-O cgg cga tgg agc aga cga cg	AC- 610-I gag gga gca cgg cac cct cc	AC611cat cgt ttt cgt gtg ttt tat tct g
T07G12.3	3017	AC-612-O ctt cca att gtt ccg tct ga	AC-613-I gtc ctt caa ttg ccc att cgg	AC614cat tgt tat ttc aaa aaa aaa ttg cgc
Y76A2A.1	4934	AC-615-O cat cgt gac gag acc cac tca g	AC-616-I ggt cgt ttg ggg tca ctt tcg g	AC617cat ttg cga gcc gga cgt ctc c
C49G7.5	2740	AC-621-O ccc aaa atg gaa cca aaa tga c	AC-622-I gct aat ttt tgg aag tcg tcg atc	AC623cat agc tgt cgt att ttg cga a
T24E12.5	3517	AC- 624-O cga tgc aac cat ggc aga ag	AC-625-I gaa cta gca gga cac ctg tg	AC626cat tcg tta ttc agg ctc tcc
C09B8.4	2918	AC-627-O gtg aat gtc tat acc agt ctg ag	AC-628-I caa ttt tgt tac tac gtt tcc tg	AC-629cat att ctg atc gta aat tgg aa
ZK792.7	5253	AC-630-O gga gaa gga ttt ctg tgg gag c	AC-631-I gag cag ctc tgc cta tcg gc	AC632cat ttc agt cag tag aat aca aaa
F28C6.2	3572	AC-638-O cag aaa ggg gca tcc ttg cc	AC-639-I gtg gaa tgg cgg atg aga tc	AC640 cat tct gaa gac gtt gag tag aag
K09C4.5	4375	AC-641-O ctt tgt ccg ccg acg acc tga g	AC-642-I cca gtt ggc gga cat ggt tta cc	AC643 cat gat tgc taa tgg tgt gat ag
C56E6.2	5897	AC-656-O gtc aag acc cgg tac cgt att c	AC-657-I cag gcg tag gtc att ctt cca g	AC658 cat tat ctt gtc tgc aaa ttg gtt c
F22F7.6	5326	AC-659-O cca cta cgg gta ata agg tgg c	AC-660-I ggt gag cgc act ata cgt caa ag	AC661 cat ttc agt ttg aat tga ata att
C32E8.11	3132	AC-662-O ggg cag caa tat tat ttg ac	AC-663 -I cga cag aga ggc aga aaa tag	AC664 cat tcc att agt cga act gaa atc gg

W06H8.8	4250	AC-709 gat gat gag ctg ttc cag gga gg	AC710-gga tct aat acc gct gca	AC711 cat gtt tgt gag agg gcc tc
F08D2.7	2248	AC-260-o ggt tgg cgatgc aga cgt act tat ttg	AC707- gtg-acc-tat-cac-ttt-tga aac g	AC259cat ggc tgg aat ata aat tcg att a
F54D7.4	2158	AC704- cag cac cca atg agc acc tta tg	AC705- cag cca tcc gaa tca caa gac c	AC687catatcttcagtgaagaaagatat

a length of the *gfp* reporter construct

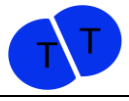
b All 3' oligo B primers contain GTCCTTTGGCCAATCCCGGGGATC in front of the sequences shown that is overlap to the *gfp* vector

Table 6. Primers used for the mutation of the E box and the GT box from C32E8.11

Construct[a]	Forward Primer	Reverse Primer
Wild type	AC721 GCGC CTGCAG ctt cag ggt ctc gcc acg atc	AC722 CGCG GGATCC cat tcc att agt cga act gaa atc gg
E box mutation	AC723 gac gtt ata tga gtg aaa cat tca aag tc**a** **acg** **ag**c tct aaa gtc tg	AC 724 ca gac ttt aga g**c**t **cgt** tga ctt tga atg ttt cac tca tat aac gtc
GT box mutation	AC725 ctc ctt cac ttt ctc cta gt**g** **acc** **a**at gtc atc att ttt gcc tcc	AC726 gga ggc aaa aat gat gac **att** **gg**t **c**ac tag gag aaa gtg aag gag

[a] Type of the construct
In red are the mutation of the nucleotide
In bold are the sequences of the E box and the GT box

Table 7. Summary Table

Proposed Twist dimer	E box responsible for the expression [a]	Gene	Human homologue and the predicted function
	CACGTG	C32E8.11	UBR 1&2 (Johanson-Blizzard syndrome)[b]
	No E box	F08D12.7	None
	No E box	W06H8.8	Titin (cancer metastasis, craniosynostosis)[c]
	ND	F54D7. 4	None
	CATATG	D1046.4	None

ND: not determined

[a] The loss or mutation of the E box CACGTG in the reporter construct of C32E8.11, loses expression in the vulval muscle whereas retaining expression in all other cells. Similarly losing the E box CATATG in *D1046.4::gfp* loses expression in the vulval muscle but retains all the other expression . In *gfp* constructs made for F08D12.7 and W06H8.8 with no E box, the animals still retain the expression in the vulval muscle.

[b] Mutations in UBR1 and UBR2 causes Johanson-Blizzard syndrome, an autosomal recessive multisystem congenital disorder, which is also described as a form of ectodermal dysplasia. In humans, Twist 2 mutation causes Setleis syndrome which is also an autosomal recessive congenital disorder and is a form of ectodermal dysplasia (www.rarediseases.org) (Tukel et al., 2010, Fallahi et al., 2010)

[c] The human homologue for W06H8.8 is titin, which is already shown to be involved in cardiomyopathy, where the role of Twist is also shown (www.wormbase.org)

REFERENCES

Bourgeois P, Bolcato-Bellemin AL, Danse JM, Bloch-Zupan A, Yoshiba K, Stoetzel C, Perrin-Schmitt F.(1998) The variable expressivity and incomplete penetrance of the twist-null heterozygous mouse phenotype resemble those of human Saethre-Chotzen syndrome. Hum Mol Genet. 7, 945-57.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. Genetics 77, 71-94.

Brown L, **Baer R** (1994). HEN1 encodes a 20-kilodalton phosphoprotein that binds an extended E-box motif as a homodimer. Mol Cell Biol. 14, 1245-55.

Burgess R, Cserjesi P, Ligon KL, Olson EN (1995).Paraxis: a basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. Dev Biol., 68, 296-306.

Carver EA, Oram KF, Gridley T.(2002). Craniosynostosis in Twist heterozygous mice: a model for Saethre-Chotzen syndrome. Anat Rec. 268, 90-2.

Castanon I, Von Stetina S, Kass J, Baylies MK (2001). Dimerization partners determine the activity of the Twist bHLH protein during Drosophila mesoderm development. Development. 2001 16, 3145-59.

***C. elegans* Sequencing Consortium** (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012-2018.

Chen, Z. F. and Behringer, R. R. (1995). Ttwist is required in head mesenchyme for cranial neural tube morphogenesis. Genes Dev, 9, 686-99.

Chesney MA, Lam N, Morgan DE, Phillips BT, Kimble J (2007). C. elegans HLH-2/E/Daughterless controls key regulatory cells during gonadogenesis. Dev Biol. 2009 Jul 1;331(1):14-25. Epub 2009 Apr 17.

Connerney J, Andreeva V, Leshem Y, Muentener C, Mercado MA, Spicer DB (2006). Twist1 dimer selection regulates cranial suture patterning and fusion. Dev Dyn. 235(5):1345-57.

Connerney J, Andreeva V, Leshem Y, Mercado MA, Dowell K, Yang X, Lindner V, Friesel RE, Spicer DB (2008). Twist1 homodimers enhance FGF responsiveness of the cranial sutures and promote suture closure.Dev Biol. 318, 323-34.

Corsi, A. K., Brodigan, T. M., Jorgensen, E. M. and Krause, M. (2002) Characterization of a dominant negative *C. elegans* Twist mutant protein with implications for human Saethre-Chotzen syndrome. Development 129, 2761-72.

Corsi, A.K., Kostas, S.A, Fire, A. and Krause, M. (2002). *Caenorhabditis elegans* twist plays an essential role in non-striated muscles development. Development 127, 2041-51.

Cserjesi P, Brown D, Ligon KL, Lyons GE, Copeland NG, Gilbert DJ, Jenkins NA, Olson EN (1995). Scleraxis: a basic helix-loop-helix protein that prefigures skeletal formation during mouse embryogenesis. Development. 121, 1099-110

Danciu TE, Whitman M. Oxidative stress drives disulfide bond formation between basic helix-loop-helix transcription factors. J Cell Biochem. 109, 417-24.

De Masi F, Grove CA, Vedenko A, Alibés A, Gisselbrecht SS, Serrano L, Bulyk ML, Walhout AJ (2011). Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. Nucleic Acids Res. 2011 39, 4553-63.

Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJ. A multiparameter network reveals extensive divergence between *C. elegans* bHLH transcription factors. Cell. 138, 314-27.

Halfon MS, Carmena A, Gisselbrecht S, Sackerson CM, Jiménez F, Baylies MK, Michelson AM (2000) Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. Cell. 103, 63-74.

Hamamori Y, Wu HY, Sartorelli V, Kedes L. (1997). The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, Twist. Mol Cell Biol. Nov 11, 6563-73.

Harfe, B. D., Vaz Gomes, A., Kenyon C., Liu J., Krause, M. and Fire, A. (1998). Analysis of a *Caenorhabditis elegans* twist homolog identifies conserved and divergent aspects of mesodermal patterning. Genes Dev 12, 2623-35.

Hobert, O. (2002). PCR fusion-based approach to create reporter constructs for expression analysis in transgenic *Caenorhabditis elegans*. Biotechniques 32, 728-30.

Howard TD, Paznekas WA, Green ED, Chiang LC, Ma N, Ortiz de Luna RI, Garcia Delgado C, Gonzalez-Ramos M, Kline AD, Jabs EW (1997). Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. Nat Genet. 15, 3-4.

Lobe CG. (1992). Transcription factors and mammalian development. Curr Top Dev Biol. 27, 351-83.

Jantsch-Plunger, V. and Fire, A. (1994). Ombinatorial structure of a body muscle-specific transcriptional enhancer in *Caenorhabditis elegans*. J Biol Chem 269, 27021-8.

Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J. (2001) Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. Genome Biol. 2, 1-10.

Kim JB, Spotts GD, Halvorsen YD, Shih HM, Ellenberger T, Towle HC, Spiegelman BM (1995). Dual DNA binding specificity of ADD1/SREBP1 controlled by a single amino acid in the basic helix-loop-helix domain. Mol Cell Biol.15, 2582-8.

Kophengnavong T, Michnowicz JE, Blackwell TK (2000). Establishment of distinct MyoD, E2A, and twist DNA binding specificities by different basic region-DNA conformations. Mol Cell Biol. 20, 261-72.

Lassar AB, Davis RL, Wright WE, Kadesch T, Murre C, Voronova A, Baltimore D, Weintraub H.(1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell. 66, 305-15.

Lee YM, Park T, Schulz RA, Kim Y.(1997). Twist-mediated activation of the NK-4 homeobox gene in the visceral mesoderm of *Drosophila* requires two distinct clusters of E-box regulatory elements. J Biol Chem. 272, 17531-41.

Li L, Cserjesi P, Olson EN (1995). Dermo-1: a novel twist-related bHLH protein expressed in the developing dermis. Dev Biol. 172, 280-92.

Massari ME, Murre C (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol. 20,429-40.

Maleki SJ, Royer CA, Hurlburt BK (1997). MyoD-E12 heterodimers and MyoD-MyoD homodimers are equally stable. Biochemistry, 36, 6762-7.

Meyers SG, Corsi AK (2010). *C. elegans* twist gene expression in differentiated cell types is controlled by autoregulation through intron elements. Dev Biol. 346, 224-36.

Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 58, 537-44.

Philogene MC, Meyers Small SG, Wang P, Corsi AK (2012). Distinct *Caenorhabditis elegans* HLH-8/twist-containing dimers function in the mesoderm. Dev Dyn. 241, 481-92.

Simpson P. (1983) Maternal-Zygotic Gene Interactions during Formation of the Dorsoventral Pattern in *Drosophila* Embryos. Genetics. 105, 615-32.

St-Pierre B, Flock G, Zacksenhaus E, Egan SE (2002).Stra13 homodimers repress transcription through class B E-box elements. J Biol Chem. 277, 46544-51.

Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic lineages of the nematode. *Caenorhabditis elegans.* Developmental Biology, 56, 110-156.

Thisse, B., Messal, M. E. and Perrin-Schmitt, F. (1987). The twist gene: isolation of a *Drosophila* zygotic gene necessary for the establishment of dorsoventral pattern. *Nucl.Acids Res.* **15**, 3439-3453.

Tukel T, Šošić D, Al-Gazali LI, Erazo M, Casasnovas J, Franco HL, Richardson JA, Olson EN, Cadilla CL, Desnick RJ.(2010). Homozygous nonsense mutations in TWIST2 cause Setleis syndrome. Am J Hum Genet. Aug 87, 289-96.

Wang P, Zhao J, Corsi AK (2006). Identification of novel target genes of CeTwist and CeE/DA. Dev Biol. 293, 486-9.

Wang SM, Coljee VW, Pignolo RJ, Rotenberg MO, Cristofalo VJ, Sierra F. (1997) Cloning of the human twist gene: its expression is retained in adult mesodermally-derived tissues. Gene. 187, 83-92.

Wolf C, Thisse C, Stoetzel C, Thisse B, Gerlinger P, Perrin-Schmitt F (1991). The M-twist gene of Mus is expressed in subsets of mesodermal cells and is closely related to the *Xenopus* X-twi and the *Drosophila* twist genes. Dev Biol. 143, 363-73.

Wong MC, Castanon I, Baylies MK (2008). Daughterless dictates Twist activity in a context-dependent manner during somatic myogenesis. Dev Biol. 317, 417-29.

Weintraub H, Davis R, Tapscott S, Thayer M, Krause M, Benezra R, Blackwell TK, Turner D, Rupp R, Hollenberg S. (1991). The myoD gene family: nodal point during specification of the muscle cell lineage. Science 251, 761-6.

Zhao J, Wang P, Corsi AK (2007). The *C. elegans* Twist target gene, *arg-1*, is regulated by distinct E box promoter elements. Mech Dev 124, 377-89