Gene silencing in the nucleus: mechanisms and new phenomena

By

VAISHNAVI NAGARAJAN

Submitted to the graduate degree program in the Department of Molecular Biosciences and the Graduate Faculty of University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Chair Person: Dr. Lisa Timmons

Dr. Berl Oakley

Dr. David Davido

Dr. Liang Xu

Dr. Kirsten Jensen

Date Defended: 06/07/2019

The Dissertation Committee for Vaishnavi Nagarajan

certifies that this is the approved version of the following dissertation

Gene silencing in the nucleus: mechanisms and new phenomena

Chairperson/Mentor: Dr. Lisa Timmons

Date approved: 06/10/2019

ABSTRACT

Silencing of homologous genes by exogenously introduced dsRNA was first observed in *C. elegans*. Endogenous small RNAs (siRNAs, piRNAs, miRNAs) mediate regulation of expression of genes post-transcriptionally or at the level of transcription, when argonaute proteins complex with small RNAs to target genomic loci for chromatin modifications in a sequence-specific manner, in the nucleus. There have been previous reports of regulation of expression by targeting mRNA for silencing in the cytoplasm. We identified a region in the genome, *flp-17* locus that is amenable to nuclear silencing mechanisms in a wildtype animal.

C. elegans exhibits strong anti-foreign genome silencing in their germline as defense against invading viral or transposon DNA. This activity is extended to transgene DNA, resulting in its silencing in the germline. *Mos1* is a foreign element, a transposon from *Drosophila* that is heterologously inserted in *C. elegans* genome. Trangenes are integrated in the chromosomal DNA by *Mos1* based Single Copy insertion technology where homologous regions flanking the gene of interest promotes recombination and thus its integration at a specific *mos* site on the worm genome. In our experiments we observed robust silencing in the somatic cells of transgenes that were intended to be integrated in ttTi5605 *mos* site on *C. elegans* genome by homologous recombination. The silencing phenomenon involves epigenetic mechanisms. We hypothesize that over several generations, the worm has "learned" that mos is a foreign element and when transgene is integrated at that site, it is amenable to silencing by the epigenetic machinery. Furthermore, we identified a previously undescribed mutation (*yy14*) in *eri-6* gene and show evidences pointing to a role in silencing of transgenes in somatic tissues. We observed an increase in trans-spliced mRNA from *eri-6/7* genes in *yy14* mutants. Our model reasons out that increase in trans-spliced mRNA

results will lead to upregulation of small RNAs (26G RNAs) that efficiently function in the silencing of our transgene, which needs further verification.

In addition, ABC transporters in *C. elegans* have been previously shown to be required for efficient RNAi. *haf-2, haf-6 and haf-9* mutants also exhibit defects in transposon mobilization. Chromatin modification by epigenetic machinery prevents mobilization of transposable elements. This includes RNAi effectors like siRNAs, Dicer and argonaute proteins. Thus, a previous study in the lab, highlights a link between ABC transporters and RNAi mechanisms in *C. elegans*. It is important to characterize the dimerization pattern of half ABC transporter proteins to gain insights on their functions and precise roles in RNAi.

ACKNOWLEDGEMENT

First, I would like to thank my mentor Dr. Lisa Timmons for her sustained mentoring, encouragement and valuable guidance. Her inquisitive nature towards scientific questions and diligence have been constant sources of inspiration to me. I am thankful to her, for her incessant support which was instrumental in helping me cross tough times during my PhD. For their time and valuable suggestions, I am thankful to the professors in my committee: Dr. Mark Richter, Dr. Berl Oakley, Dr. David Davido, Dr. Liang Xu and Dr. Kirsten Jensen. I thank my past committee members Dr. Peter Gegenheimer and Dr. Kathy Suprenant for stimulating a deep interest in science. I am thankful to all friends from Timmons lab for creating a fun and inspiring environment to perform science. My special thanks to Dr. Nadeem Asad who had been a great companion in the lab making the time here enjoyable. He had been truly inspirational and immensely helpful in the lab. I thank him and Emily Smith for their contributions to my projects. I thank Dr. Deborah Taylor, the teaching in- charge, Mammalian Physiology.

My father, Nagarajan and my mother Kamakshi, have been the primary source of encouragement in pursuing PhD and I take this chance to express my gratitude to them. I acknowledge their sacrifices towards making this happen. I am ever grateful to them for their love, time, effort and sincere prayers without which my dream of PhD would have been impossible. I thank them for supporting all my life decisions. I would like to thank my caring brother Srinivasan, for helping me think through and for keeping me from sulking, during my tough times, and also for being there for my parents, when I was away from home.

I would like to thank my husband Srivatsan, my intellectual companion, my best friend, who was pivotal in seeding in me the desire to pursue science. I cannot thank him

enough for believing in my abilities and for being my inner strength. I am indebted to him for the sleepless nights, for his endless patience, for escorting me to my lab at odd hours and what not. I would like to thank my baby girl Shakthi Srivatsan for her best behavior during the times I was away from her. She, I believe is the harbinger of luck and hope.

I thank my grandmother Gowri for the valuable morals and life lessons she instilled in me. I thank her for the wealth of cherished memories she left behind for me. I am grateful to my aunt Anuradha and my uncle Narayana Raja for their never-ending love and kindness. I thank my in-laws for their support and motivation.

DEDICATION

This work is dedicated to my husband Srivatsan Parthasarathy, my daughter Shakthi and my parents V. Nagarajan and J. Kamakshi.

INDEX

CHAPTER 1

Determining the functional dimerization partners of half molecule ABC transporters

ABSTRACT

The existence of a link between a conserved phenomenon, RNA interference and a ubiquitous protein superfamily of ABC transporters, was discerned when it was shown that some ABC transporters have endogenous roles in gene silencing mechanism (Sundaram, Echalier et al. 2006). Such ABC transporters are named ABC_{RNAi} transporters. ABC_{RNAi} transporter genes, when mutated, lead to RNAi defects that normally respond in the cytoplasm to environmentally delivered dsRNAs as well as defects in the nucleus that normally facilitate silencing of transposable elements. We hypothesized that these ABC_{RNAi} proteins may participate in RNAi mechanisms through the transport of specific substrates essential for RNAi; alternatively, trafficking of substrates may produce a subcellular environment that is conducive to RNAi. Some ABC_{RNAi} transporter genes are configured as half molecules that dimerize to constitute a functional transporter protein. Our goal was to determine the homodimer or heterodimer configurations of half ABC transporters in *C. elegans*. We approached this question with Bimolecular Fluorescence Complementation (BiFC) where we utilize parallel orientation of the interacting proteins, to bring our fluorescent reporter fragments in close proximity. The regaining of fluorescence of the two fragments will indicate protein:protein interaction. The knowledge of organized arrangement of domains of half ABC transporters acquired from this study will be integral for 1) designing substrate trafficking assays in order to verify potential substrates and 2) for a better understanding of their RNAi functions in *C. elegans*.

INTRODUCTION

ABC TRANSPORTERS

Biochemistry of ABC transporters

ATP Binding Cassette (ABC) transporters are a superfamily of integral membrane proteins that are involved in the translocation of many substrates across membranes. The transport of small molecules against a chemical gradient from one side of the membrane to the other is powered by the binding and hydrolysis of ATP. ABC transporters minimally contain four functional domains. Two transmembrane-spanning domains (TMD) form the pore through which molecules cross the membranes. The other two domains are nucleotide-binding domains (NBD) to which ATP molecules bind. ABC transporters move substrate molecules by a two-step cycle. An ATP molecule first binds to each NBD (Sharom 2008). Binding of a substrate then enables NBD dimerization, ATP hydrolysis and movement of substrate through the pore. ATP is hydrolyzed using an unusual hydrolysis mechanism in which the ATPase activity of one NBD hydrolyzes the ATP bound to the second NBD domain, releasing inorganic phosphate (Rees, 2009). The ATP binding domain of ABC transporters is unique from other ATP binding domains in that it harbors a phosphate-binding P-loop with a LSGGQ (ABC motif) sequence that is involved in nucleotide binding (Wilkens 2015, Ford and Beis 2019).

Full and half ABC transporter proteins

Full transporters consist of two TMDs and two NBDs, while half transporters contain one of each domain (Figure 1). In mammalian cells, full transporters such as P-glycoprotein, Multi-Drug Resistance Protein (MRP) and the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) can be found in plasma membranes. By contrast, all human half transporters are found in membranes of various subcellular organelles (Zhang, Zhang et al. 2000). Half transporters are required to form homodimers or heterodimers to functionally transport substrates. For example, in humans, TAP1 (ABCB2) and TAP2 (ABCB3) form a heterodimer in the endoplasmic reticulum where they perform essential immune functions by binding peptides derived from the cytoplasmic proteasome and traffick those into the endoplasmic reticulum where they are loaded on to the Major Histocompatibility Complex I (MHCI).

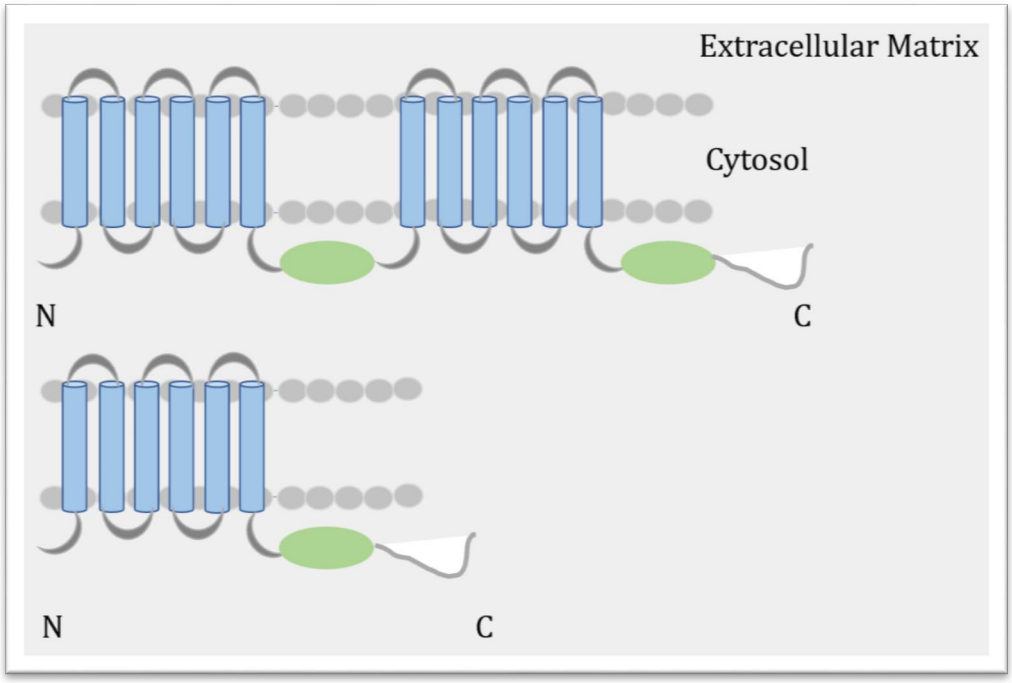


Figure 1. Full and half ABC transporters. Top panel: A full ABC transporter consists of two transmembrane domains (TMD, blue cylinders) and two nucleotide binding domains (NBD, green ovals). Bottom panel: Genes encoding half ABC transporters encode only one TMD and one NBD; half transporter molecules function as homodimers or heterodimers. In both cases, the TMD is shown to be towards the N-terminus; although, some transporter genes harbor different structures with the NBD domain appearing at the N-terminus, for example. ABCG subfamily half transporters are unique that they have their NBD at their N-terminus.

ABC transporters are conserved across prokaryotic and eukaryotic cells. ABC transporters are classified as importers and exporters, depending on the directionality of transport across the membrane. While bacteria have both types, eukaryotic ABC transporters are predominantly exporters. In prokaryotic organisms, transport by importers

is facilitated by the association of substrates to specific binding proteins (Heppel 1969).

Human ABC transporters

4% of human genes encode transporter proteins of which ABC transporters constitute the largest family (Dermauw and Van Leeuwen 2014). There are 49 ABC transporters in the human genome (Vasiliou, Vasiliou et al. 2009) comprising seven subfamilies, A-G. (Stefkova, Poledne et al. 2004). Mutations in 14 of these genes can cause 13 genetic disorders. ABC transporters are instrumental for human health and development and the study of human transporters is a relatively highly developed field in comparison to the study of ABC transporters from other organisms. One can acquire an appreciation of the breadth of functions that ABC transporters provide to multicellular organisms from the nature of human genetic disorders cause by mutations in ABC transporter genes as well as the identity of substrates trafficked, where known.

Twelve of the 49 human ABC transporter genes belong to the ABCA subfamily. Mutations in ABCA subfamily genes are associated with various human disorders. Tangier disease (caused by defects in the cholesterol efflux of ABCA1) is characterized by reduced high-density lipoprotein HDL in the blood leading to premature cardiovascular disease. Stargardt disease is a type of hereditary macular degeneration caused by buildup of retinoids in the eye due to lack of ABCA4 which is found in the disc cell membranes of the eye where it flips N-retinylidene-phosphatidylethanolamine from the lumen-facing leaflet to the cytoplasmic leaflet (Garces, Jiang et al. 2018). (Stefkova, Poledne et al. 2004).

The ABCC subfamily includes the ABCC7 gene which encodes CFTR protein, a chloride ion channel. Cystic fibrosis is a disorder affecting lungs and other tissue due to defects in transmembrane conductance caused by mutations in ABCC7.

The D subfamily consists of half transporters which are all expressed in the peroxisomal membranes where they traffick mostly hydrophobic lipid substrates. Mutations in two of these half transporters are linked to adrenoleukodystrophies.

ABC E and F subfamilies have proteins that do not have transmembrane domains. The F subfamily ABC proteins are predicted to be involved in regulation of mRNA translation (Tyzack, 2000).

The ABCG subfamily consists of full and half transporters. The five half ABC G transporter genes are also referred to as the "white" transporters due to their similarity to *Drosophila white* genes. Unlike *white*, which transports nucleotide substrates that are used in the biochemical manufacture of eye pigments, the human ABC G transporters are involved in cholesterol and sterol transport (although the list of substrates for this gene family in humans are incomplete). ABCG5 and ABCG8 cause Sitosterolemia – a lipid disorder characterized by the inability to efflux dietary sterols. (Stefkova, Poledne et al. 2004). ABCG5 and ABCG8 proteins form heterodimers at the plasma membrane to export dietary, including plant derived sterols. Lipid driven three-dimensional crystallization has shown that the dimer assumes inward facing conformation (Lee, Kinch et al. 2016).

The ABCB subfamily (MDR) includes 11 genes that are configured as full- or half-transporter proteins. ABCB1 is a full transporter protein that localizes in liver and

constitutes the blood brain barrier and is implicated in multidrug resistance in cancer and virus infected cells. ABCB half transporters are found in the inner membrane of the mitochondria where they facilitate iron transport for the purpose of producing functional heme (Vasiliou, Vasiliou et al. 2009).

ABC transporters in C. elegans

There are 60 ABC transporter genes in *C. elegans* which are categorized into 8 subfamilies (A-H) (Figure 2) based on amino acid sequence and domain organization from phylogenetic analyses with transporters from *Homo sapiens* and *Saccharomyces cerevisiae* (Sheps, Ralph, Zhao, Baillie, & Ling, 2004). Most ABC transporters have the domains (two TMDs and two NBDs) in tandem while some transporters have only one of each domain. Half transporter genes in *C. elegans* are distributed among B, D, G and H subfamilies.

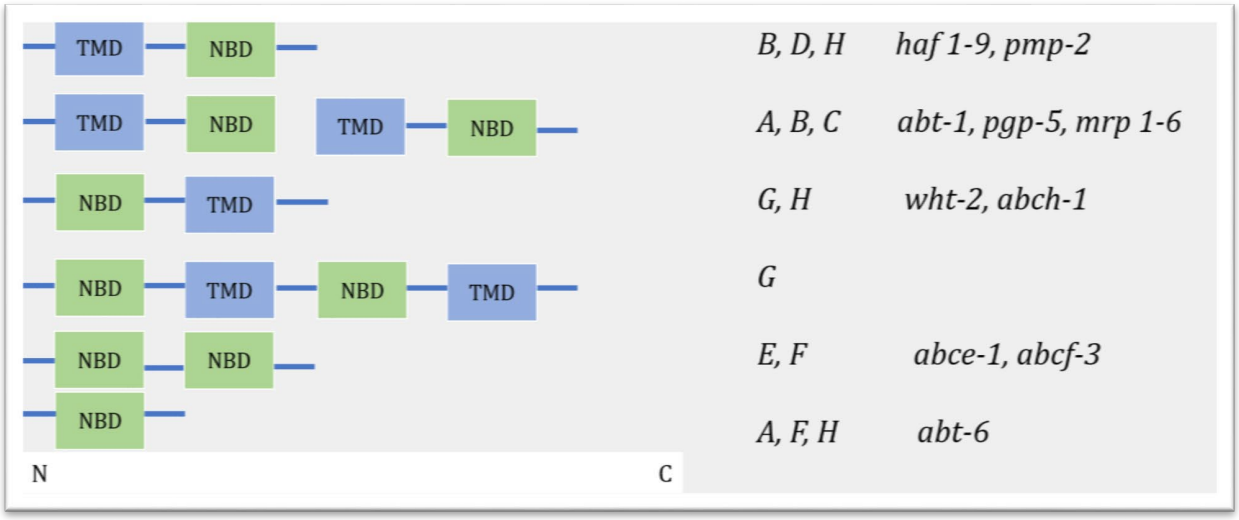


Figure 2 : Categorization of *C. elegans* ABC proteins based on homology and domain arrangements. TM - transmembrane domain and ABC - ATP-binding cassette domain. The right column is a list of *C. elegans* genes in from each category. Figure modified from (Sheps, Ralph et al. 2004).

The *Pgp* family is a set of genes that is configured as full ABC transporters from the B subfamily. P-glycoproteins in *C. elegans* are found in intestine and excretory cells. Pgp-1, 3 and MRP-1 (from C subfamily) in *C. elegans* respond to heavy metals such as cadmium and arsenite (Broeks, Gerrard et al. 1996). *C. elegans* P-gp transporter has been solved for its structure at a 3.4 Å resolution in the absence of any bound substrates. The crystal structure reveals an inward-open conformation. Substrates can bind and exit the cells from the cytoplasm as well as the membrane because the transport pathway is shown to be continuous with the inner leaflet of the membrane (Jin, Oldham et al. 2012).

Subfamily	*Gene name*	Chromosome	Full/Half	Domain organization	No of helices per TMD
ABCA/Abt	*Abt-1*	IV	Full	2 *(TMD-NBD)	6
	Abt-2	I	Full	2 *(TMD-NBD)	6
	Abt-3	IV	Full	2 *(TMD-NBD)	4 or 8
	Abt-4	V	Full	2 *(TMD-NBD)	6 or 8
	Abt-5	I	Full	2 *(TMD-NBD)	6
	ced-7	III	Full	2 *(TMD-NBD)	8
	F56F4.6	I		NBD	
ABCB/Pgp	*Pgp-2*	I	Full	2 *(TMD-NBD)	6
	Pgp-3	X	Full	2 *(TMD-NBD)	6
	Pgp-10	X	Full	2 *(TMD-NBD)	4
	Pgp-11	II	Full	2 *(TMD-NBD)	6
	Pgp-12	X	Full	2 *(TMD-NBD)	6
	Pgp-13	X	Full	2 *(TMD-NBD)	6
	Pgp-14	X	Full	2 *(TMD-NBD)	6
	Pgp-15	X	Full	2 *(TMD-NBD)	6
	Pgp-4	X	Full	2 *(TMD-NBD)	6
	Pgp-1	IV	Full	2 *(TMD-NBD)	6
	Pgp-6	X	Full	2 *(TMD-NBD)	6
	Pgp-7	X	Full	2 *(TMD-NBD)	6
	Pgp-8	X	Full	2 *(TMD-NBD)	6
	Pgp-5	X	Full	2 *(TMD-NBD)	6
	Haf-1	IV	Half	TMD-NBD	4
	Haf-2	II	Half	TMD-NBD	8
	Haf-3	V	Half	TMD-NBD	6
	Haf-4	I	Half	TMD-NBD	8
	Haf-5	III	Half	TMD-NBD	8
	Haf-6	I	Half	TMD-NBD	4
	Haf-7	V	Half	TMD-NBD	6
	Haf-8	IV	Half	TMD-NBD	4
	Haf-9	I	Half	TMD-NBD	8
ABCC/Mrp	*Mrp-3*	X	Full	2 *(TMD-NBD)	6
	Mrp-5	X	Full	2 *(TMD-NBD)	6
	Mrp-6	X	Full	2 *(TMD-NBD)	6
	Mrp-4	X	Full	2 *(TMD-NBD)	6 or 10
	Mrp-2	X	Full	2 *(TMD-NBD)	6 or 10
	Mrp-1	X	Full	2 *(TMD-NBD)	6 or 12
	Mrp-7	V	Full	2 *(TMD-NBD)	2 or 12
	Mrp-8	X	Full	2 *(TMD-NBD)	4 or 6
	Cft-1	V	Full	2 *(TMD-NBD)	5 or 6
ABCD	*C44B7.9*	II	Half	TMD-NBD	4
	C54G10.3	V	Half	TMD-NBD	6
	T02D1.5	IV	Half	TMD-NBD	6
	T10H9.5	V	Half	TMD-NBD	6
ABCD	*C44B7.8*	II	Half	TMD-NBD	4
ABCE	*Y39E4B.1*	III		NBD-NBD	
ABCF	*F18E2.2*	V		NBD-NBD	
	F42A10.1	III		NBD-NBD	
	T27E9.7/ GCN20-2	III		NBD-NBD	
ABCG	*C05D10.3*	III		NBD-TMD	4
ABCH	*C56E6.1*	II	Half	NBD-TMD	12
	C56E6.5	II	Half	NBD-TMD	6
	C10C6.5	IV	Half	NBD-TMD	6
	C16C10.12	III	Half	NBD-TMD	4
	F02E11.1	II	Half	NBD-TMD	6
	F19B6.4	IV	Half	NBD-TMD	6
	T26A5.1	III	Half	NBD-TMD	4
	Y42G9A.6	III	Half	NBD-TMD	6
	Y47D3A.11	III	Half	NBD-TMD	6
	Y49E10.9	III	Half	NBD-TMD	4

Figure 3 : *C. elegans* ABC transporter subfamilies. *C. elegans* ABC transporters full and half ABC transporters are categorized into eight subfamilies A-H.

Half ABC Transporters

ABCB subfamily in Homo sapiens

The ABCB subfamily in humans consists of seven half transporters – ABCB2, BACB3, ABCB6, ABCB7, ABCB8, ABCB9 and ABCB10.

The half transporter ABCB2 (TAP1 – Transporter associated with antigen processing) and ABCB3 (TAP2) form heterodimers, localize at the endoplasmic reticulum (ER) membrane and function in delivering antigenic peptides to the lumen of ER. The peptides are then presented at the cell surface as antigens by class I MHC molecules (Abele, 1999). X ray crystallography was used to solve the structure of TAP homolog TmrAB (*Thermus thermophilus* multidrug resistance proteins A and B). The TMD sequences between human TAP1/TAP2 and TmrAB share 22% identity. The 2.7 Å structure shows that the substrate cavity of the TmrAB transport is towards the cytoplasmic side transporting lipids. This protein has also been shown to rescue antigen processing defects in human immuno-compromised cells, showing that these proteins from two organisms share functional homology (Noll, Thomas et al. 2017).

ABCB6, B7, B8 and B10 localize to the mitochondria (Paterson, Shukla et al. 2007), but recent studies have shown that ABCB6 localizes to additional organelles including the endolysosomal compartment (Koszarska, Kucsma et al. 2014). This underscores the fact that knowledge of subcellular localization and thus, of functions of any ABC transporter is currently incomplete. ABCB7 is associated with X-linked sideroblastic anemia and ataxia

(XLSA/A), suggesting it has functions related to iron homeostasis (Allikmets, Raskind et al. 1999). ABCB10 is a homodimer half transporter which functions in erythropoiesis which suggests that it could export heme molecule (Shintre, Pike et al. 2013). ABCB9 localization to the lysosome has been demonstrated using subcellular fractionation and immunofluorescence and is expressed in brain, spinal cord and testis (Zhang, Zhang et al. 2000).

ABCB subfamily in Caenorhabditis elegans

There are nine half transporter proteins in *C. elegans* (*haf-1 to 9*) and the knowledge of their functions and substrates are far from complete. The half transporters that belong to the B subfamily have an N-terminal TMDs and a C-terminal ABC domain (Figure 2).

HAF-1 in *C. elegans* harbors an N-terminal mitochondrial targeting sequence conferring localization to the mitochondria where it is believed to function in exporting peptides that come from the matrix proteins**,** although evidence for this function is indirect, having come from isolation of intact mitochondria from wild type and mutant *C. elegans* and not from in vitro assessment of biochemical activity. Also, there is no direct evidence for dimerization partners for this protein (Haynes, Yang et al. 2010). Peptide transport by HAF-1 transporters may be involved in unfolded protein responses, a function that is particularly important during stress (Nargund, Pellegrino et al. 2012). As Mdl1p, the yeast homologue of HAF-1 has broad substrate specificity, it may be the case that HAF-1 might also transport substrates other than peptides.

C. elegans HAF-4 and HAF-9 transporter proteins are shown to be homologous to human TAPL (TAP-like) based on amino acid identity (38%). TAPL is shown to be localized to lysosomes based on immunofluorescence and subcellular fractionation methods. TAPL possesses two localization signals – YXXΨ and dileucine motifs – in its C terminal tail. The presence of these motifs in non-lysosomal transporters also suggests TAPL might localize in locations other than lysosomes. (Zhang, Zhang et al. 2000). The *C. elegans* HAF-4 and HAF-9 proteins colocalize specifically in non-acidic gut granules, which may represent a lysosomal like organelle. Because of the rather limited homology between human and *C. elegans* proteins (homology between HAF-4 and HAF-9 with TAPL is based on the alignment of their ABC domains), the lack of conserved localization signals and the question of the precise function of the subcellular compartment, it is difficult to make definitive statements regarding functional conservation in the absence of biochemical data. Clearly, this kind of data does not indicate the possibility of shared substrates between (HAF-4,HAF-9) and TAPL. Co-immunoprecipitation studies have shown that HAF-4 and HAF-9 form heterodimers (Tanji, Nishikori et al. 2013), yet the few studies that have been performed on these proteins, coupled with the high probability for antibody cross reactivity between related family members requires independent assessments of dimerization data.

HMT-1(HAF-5) is required for Cd2+ tolerance in *C. elegans* based on phenotypic analysis of RNAi knockdown of *hmt-1.* HMT-1 has most homology to HAF-7 and HAF-9 within *C. elegans* half transporters. But, as stated earlier, HAF-9 localizes to lysosomes, whereas, heterologously expressed *C. elegans* HMT-1 is localized in vacuolar membranes (Vatamaniuk, Bucher et al. 2005). These distinct subcellular localizations hint that these half

transporters have divergent roles, likely through the transport of distinct substrates. HMT-1 was demonstrated to form homodimers by mating based on data obtained using the Split-Ubiquitin Yeast 2 Hybrid method (Kim, Selote et al. 2010). However, this method did not eliminate the possibility of HMT-1 heterodimerization with other half transporters in *C. elegans*. Based on phylogenetic analyses, *C. elegans* HAF-5 is orthologous to ABCB6 (human) (Sheps, Ralph et al. 2004). Nevertheless, their organellar localization and transport substrates are not similar between the two organisms, as human ABCB6 localizes to mitochondria where it functions in the trafficking of porphyrins.

In an effort to help elucidate dimerization partners as well as the identification of potential substrates, we compared the homology between *C. elegans* half transporters and the better-studied human transporters. We focused on HAF-6 because mutants defective for *haf-6* display the strongest RNAi defects amongthe three half transporters. Reasoning that important functional differences might lie the in transmembrane region, which forms the pore through which the substrate trafficks, we compared the homology between HAF-6 protein and the human half transporters. We obtained the TM domain sequence from hydropathy analysis and aligned the domains in pair-wise fashion using MUSCLE (Supplementary information 1). By this criterion HAF-6 is more similar to human ABCB8 (isoform d), with 36.5% identity and 49% similarity to HAF-6A and 45.9% identity and 61.5% similarity to HAF-6B. The human ABCB8 is a half ABC transporter localized in the inner membrane of mitochondria. ABCB8 is thought to transport heme (Elliott and Al-Hajj 2009). This leaves us to speculate that ABCB8 might have roles in oxidative stress. By overexpressing ABCB8 in HEK293 cells, it was shown that the first 55 amino acids are required for their localization in the mitochondria (Hogue et al, 1999). Like many

mitochondrial proteins, ABCB8 is coded in the nuclear genome and is transported by the TOM (Translocase of the Outer Membrane) complex (Schaedler, Faust et al. 2015). Overexpression of ABCB8 also showed increased export of iron from mitochondria (Ichikawa, Bayeva et al. 2012). Along with these functions, ABCB8 has been shown to confer resistance against doxorubicin (Elliott and Al-Hajj 2009). Mouse studies that deleted ABCB8 showed decreased levels of mitochondrial succinate dehydrogenase (Ardehali, Xue et al. 2005). All these observations point out to mitochondrial localization of ABCB8. However, in contrast to the human mitochondrial ABCB8, HAF-6 localization is confined to ER membrane as revealed from immunofluorescence staining and studies using GFP fusion proteins (Sundaram, Echalier et al. 2006). These examples highlight why it is not meaningful to make predictions on subcellular localization or substrates transported based on homology between two proteins.

RNA INTERFERENCE (RNAi)

RNA interference (RNAi) describes a gene silencing phenomenon that was discovered in *C. elegans* by Andrew Fire and Craig Mello in 1998 (Fire, Xu et al. 1998). RNAi is the ability of exogenous dsRNA to interfere with the expression of mRNAs, homologous in sequence to the dsRNA trigger. RNAi is conserved in eukaryotes and is thought to have evolved as a molecular immune function against viruses and transposons (Plasterk 2002). RNAi has natural functions that also include regulation of gene expression during development (Agrawal, Dasaradhi et al. 2003).

In *C. elegans*, dsRNA that is introduced experimentally, is cleaved by a double-stranded RNA-specific endonuclease protein called Dicer, resulting in short interfering RNAs (siRNAs). siRNAs are discrete, double-stranded RNAs 21-28 nucleotides long that are important in silencing complementary nucleic acids. Argonaute proteins are important effector molecules for RNA-based silencing mechanisms. Argonaute proteins are highly conserved, being found in prokaryotes, eukaryotes and archaea. They can be found in association with cytoplasmic or nuclear Dicer proteins where they are loaded with the Dicer RNA product and are instrumental in converting the product to single-stranded RNA form. In general, Argonaute proteins bind different classes of small, noncoding RNAs, recognizing features such as length, 5' end modifications (the nature of the phosphate) and 3' end modifications (such as 2'-O-methylation). Acting in concert with other proteins and using RNA as a guide to direct protein activities, Argonaute proteins are instrumental in mediating RNA-directed silencing mechanisms. Some Argonaute proteins have Slicer activity (cleaving mRNAs with homology to the siRNA sequence), some Argonautes shuttle siRNAs into the

nucleus; other Argonautes are found in the nucleus in association with chromatin modifying enzymes and facilitate transcription-level silencing (Huang, 2014).

RNAi mechanisms can affect gene expression at the transcriptional and post-transcriptional levels. In the nucleus, RNAi mechanisms regulate endogenous gene expression and are critically important for chromatin processes such as transposon silencing, maintenance of heterochromatin, and establishment and maintenance of centromeres. In worms and plants, silencing of genes by experimental delivery of dsRNA can result in histone modifications—methylation patterns that can be inherited through multiple generations, even in the absence of the trigger dsRNA. This provides further evidence of RNAi mechanisms in the nucleus.

A clear understanding of different RNAi mechanisms, their pathways and the numerous conserved proteins involved have been elucidated by using *C. elegans* as the genetics tool. RNAi mutants are viable and fertile. Unlike mammalian cells, *C. elegans* does not harbor interferon genes, which makes it an easier system to study RNAi. *C. elegans* has a transparent body making the visualization of reporters direct and simple. In *C. elegans*, RNAi can be induced by more than one method – feeding or injecting dsRNA or soaking worms in dsRNA.

In our studies of RNAi mechanisms, we have observed that some ABC transporters are required for RNAi-related mechanisms in the nucleus. In these ABC transporter mutants, transposon mobilization is activated. Half transporter genes that display this interesting RNAi-related defect include *haf-6* and *haf-9.* The precise molecular role of these transporters in transposon silencing mechanisms will require a thorough analysis of their biochemical

function. Here, we set to exploit Split-GFP approach towards determining dimerization partners of half ABC transporters. Knowing how the protein is put together is a critical first step in order for us to design biochemical assays for protein function.

Significance

ABC transporters are conserved proteins and are critical for various physiological functions such as maintaining homeostasis of biological compounds or preventing intracellular buildup of toxic compounds. Mutations in ABC transporters are the cause of several genetic disorders in humans (Figure 4) and those in the multidrug transport class, are especially important in drug design and therapy—in particular, in chemoresistance in cancer and virus-infected cells.

SID-1 (Systemic RNAi Deficient) channel protein is a good example to highlight the role of transporter proteins in RNAi. SID-1 is an integral membrane protein with 11 transmembrane helices, and an extracellular N-terminal glycosylated domain. *sid-1* mutants are deficient for systemic RNAi, presumably due to impairment of this protein to act as a pore for the trafficking of dsRNA molecules to the cytoplasm in different cells and tissues (Whangbo, Weisman et al. 2017).

Our lab has previously described roles for the half ABC transporters *haf-2, haf-6 and haf-9* in RNA interference in *C. elegans* (Sundaram, Echalier et al. 2006), although their precise biochemical roles are unknown. In order to determine the roles of half transporters in RNAi mechanism, it is important to elucidate their specific substrates. This is contingent upon knowing the functional partners of these half ABC transporters.

Gene symbol (Common names)	Major tissue distribu	substrate/function	Disease (loss of function)
ABCA1	Ubiquitous	Phospholipid, cholesterol	Tangier disease
ABCA2	Brain, kidney, lung, hea	Cholesterol /Drug resistance	Alzheimer's disease
ABCA3	Lung	Surfactant lipids	Surfactant metabolism Dysfunction 3
ABCA4	Retina	N-retinylphosphatidylethanolamin	Stargardt disease
ABCA5	Muscle, heart, testes		
ABCA6	Liver		
ABCA7	Spleen, thymus		
ABCA8	Ovary		
ABCA9	Heart		
ABCA10	Muscle, heart		
ABCA12	Lung, skin, stomach	Lipids	Harlequin ichthyosis
ABCA13	Low in all tissues		
ABCB1 (MDR1)	epithelia, blood-brain barrier, Adrenal, kidney, brain	Glucosylceramides /Multidrug resistance	Inflammatory bowel disease
ABCB2 (TAP1)	All cells	Peptide transport	Immune deficiency
ABCB3 (TAP2)	All cells	Peptide transport	Immune deficiency
ABCB4 (MDR3)	Hepatocytes	Long chain phosphatidylcholines	Progressive familial intrahepatic cholestasis 3 (PFIC3)
ABCB5	Ubiquitous		
ABCB6 (MTABC	Mitochondria	Iron transport	
ABCB7	Mitochondria	Fe/S cluster transport	X-linked sideroblastosis and anemia (XLSA/A)
ABCB8 (MABC1	Mitochondria	Fe/S cluster transport	
ABCB9	Heart, brain		
ABCB10 (MTAB	Mitochondria		
ABCB11 (BSEP)	Hepatocytes	Bile salts	PFIC2

Gene symbol (Common names)	Major tissue distribution	substrate/function	Disease (loss of function)
ABCC1 (MRP1)	Lung, testes, PBMC	Drug resistance	
ABCC2 (MRP2)	Liver, kidney, intestine	Bilirubin/bile salts	Dubin-Johnson syndrome
ABCC3 (MRP3)	Lung, intestine, liver	Drug resistance	
ABCC4 (MRP4)	Prostate	Nucleoside transport	
ABCC5 (MRP5)	Ubiquitous	Nucleoside transport	
ABCC6 (MRP6)	Kidney, liver		Pseudoxanthoma elasticum
ABCC7 (CFTR)	Exocrine tissues	Chloride ion channel	Cystic fibrosis, CBAVD
ABCC8 (SUR)	Pancreas	Sulfonylurea receptor	familial persistent hyperinsulinemic hypoglycemia of infancy (FPHHI)
ABCC9 (SUR2)	Heart, muscle		dilated cardiomyopathy with ventricular tachycardia (DCVT)
ABCC10	Low in all tissues		
ABCC11	Low in all tissues		
ABCC12	Low in all tissues		
ABCD1 (ALD)	Many (Peroxisomes)	Very long chain fatty acids (VLCFA	Adrenoleukodystrophy
ABCD2 (ALDR)	Peroxisomes		
ABCD3	Peroxisomes		
ABCD4	Peroxisomes		
ABCE1	Ovary, testes, spleen		
ABCF1	Ubiquitous		
ABCF2	Ubiquitous		
ABCF3	Ubiquitous		
ABCG1 (White)	Ubiquitous	Cholesterol transport?	implicated in cardiovascular disease, obesity and diabetes
ABCG2 (BCRP)	Placenta, intestine	Toxin efflux, drug resistance	
ABCG4 (White 2	Liver		
ABCG5 (White 3	Enterocytes, hepatocyt	Cholesterol, plant sterols	Sitosterolemia
ABCG8	Enterocytes, hepatocyt	Cholesterol, plant sterols	Sitosterolemia, gallstones

Figure 4: ABC superfamily of transporters in humans. This table lists human ABC transporters with their substrates, functions and their associated diseases.

BIMOLECULAR FLUORESCENCE COMPLEMENTATION (BiFC)

BiFC analysis employs a fragmented reporter protein. In our case, we utilized green fluorescent protein that is split into two non-fluorescent fragments. The split GFP reconstitution method is based on reobtaining the fluorescence when the two fragments of GFP reassociate. This reassociation can be promoted by fusing the fragments to two proteins that normally interact with each other, bringing the attached GFP fragments into close proximity where they can associate and gain fluorescence (Figure 5).

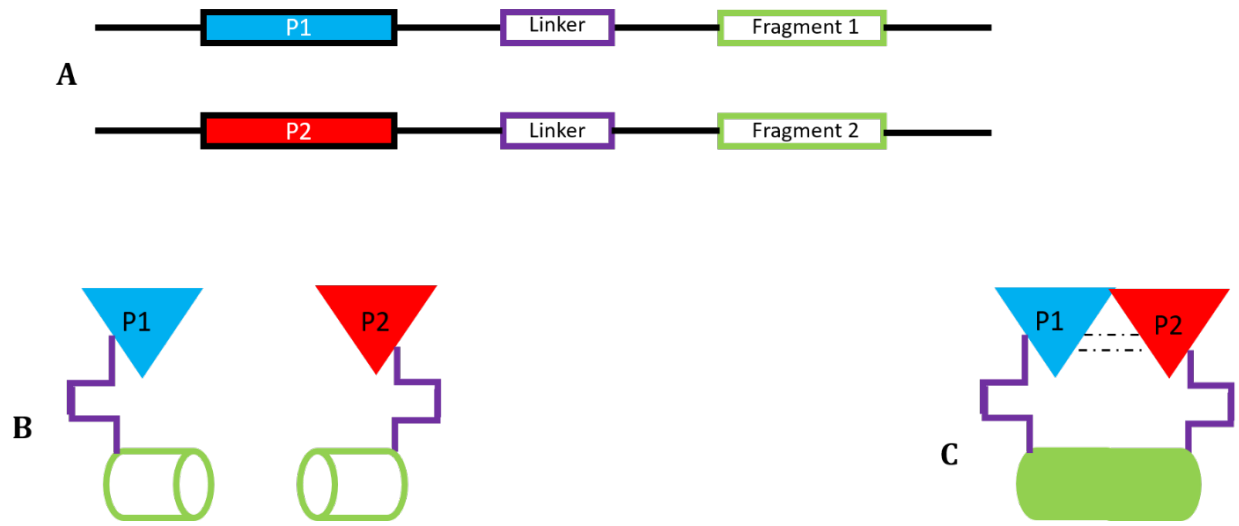


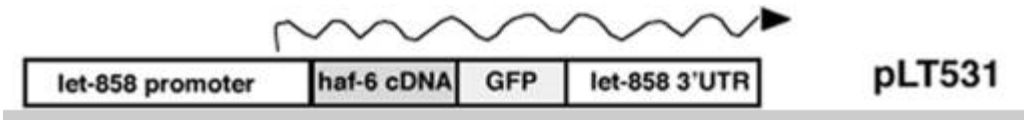
Figure 5 : Schematics of BiFC. A- Classical BiFC vectors require a protein coding gene (P1 &or P2) connected by a short peptide (Linker) to a fragment (Fragment 1 or 2) of a reporter (a fluorescent molecule, in this figure). B- Expression of fusion proteins P1:: Fragment1 and P2::Fragment2 where the reporter fragments are attached to the C-termini of P1 and P2 through a linker. C- Protein interactions between P1 and P2 bring the reporter fragments together, which results in re association of the GFP molecule and formation offluorescence.

Protein-protein interactions have been studied using several methods. BiFC is our method of choice because it has several advantages over other methods. For example, the classical yeast two hybrid (Y2H) method often faces a major issue of improper protein folding, preventing interaction of the proteins being tested. Split ubiquitin is another BiFC type assay that has been in use to detect membrane protein interactions. Split Ubiquitin has been shown to successfully determine the protein partner of a *C. elegans* transporter (Kim, Selote et al. 2010). Although Y2H and mating based split-ubiquitin assay in yeast have been used in studying *C. elegans* protein interactions, endogenous expression pattern and localization information is missing when *C. elegans* proteins are expressed in a yeast system (Boxem, Maliga et al. 2008). In contrast, our study expresses proteins in the native organism. Also, the obligatory expression of the protein of interest in the nucleus in Y2H disfavors it from being the method of our study since our proteins are membrane transporters. Co-immunoprecipitation could also be used to investigate dimerization partners; however, there is a drawback of non-specific binding or cross reactivity, especially when working with membrane proteins that have a high degree of amino acid similarity. False negatives are common with co-immunoprecipitation (co-IP) as low affinity interactions are not detected in this method. Methods such as co-IP are more suited toward verifying suspected dimerization partners, and not for a study that is exploratory in nature like ours. More practically, the antibodies we generated against the HAF proteins are anti-peptide antibodies (due to the high degree of amino acid similarity between proteins in this family), and most are not suitable for co-IP.

Split-GFP based detection of protein interactions averts many of these challenges by the in-vivo nature of the technique. In addition to traditional protein interaction methods, in

recent years, there has been a rise in BiFC methods--strategies that generally employ reassociation of various protein fragments including ubiquitin, variants of GFP and β-galactosidase, β-lactamase and others (Luker and Piwnica-Worms 2004). Such BiFC of split protein-based methods side steps the disadvantages associated with the purification of membrane proteins. In our study, the split-GFP system was favored due to its potential for rapid detection of interaction by simple instruments such as inverted fluorescence microscopes. Bimolecular fluorescent complex formation suffers from disadvantages that the fluorescence from interaction is irreversible and that transient interactions may not be detectable ((Shyu, Hiatt et al. 2008), (Luker and Piwnica-Worms 2004)). Nevertheless, in our experiments, these properties should work to our advantage, as we do not anticipate transient interactions between functional dimers. Furthermore, HAF-6::GFP fusion proteins are proven to successfully express in *C. elegans* in our previous studies (Sundaram, Echalier et al. 2006) (Figure 6), an indication that C-terminal tagging of our proteins is functionally tolerated.

A



B

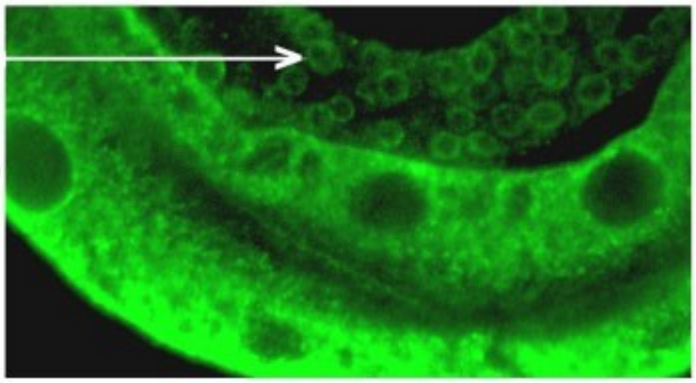


Figure 6: HAF-6::GFP is functional in *C. elegans* tissues. A GFP Fused to *haf-6* at the C-terminus driven by a ubiquitous promoter was membrane localized and overlapped with patterns obtained using anti-HAF-6 antibodies. The transgene rescued the RNAi defects in *haf-6* mutants (Sundaram, 2006).

Previous Results

It was previously reported by our lab that *haf-6* transporter in *C. elegans, is* required for efficient RNAi (Sundaram, Echalier et al. 2006). In addition to *haf-6*, eight other *C. elegans* ABC transporter genes were shown to be implicated in RNAi. (A list of all the nine ABC transporters shown to be required for efficient RNAi is given in Table 1.) HAF-6 is a half ABC protein that belongs to the B subfamily of ABC transporters; the *haf-2* and *haf-9* genes, also B subfamily half transporter genes, were also implicated in RNAi.

Gene name	ABC subfamily
abt-1	A
pgp-4	B
pgp-11	B
haf-2	B
haf-6	B
mrp-1	C
pmp-1	D
wht-3	G
wht-1	G

Table 1 : ABC transporter strains required for RNAi. The corresponding mutants displayed RNAi defects in the germ line (Sundaram, Echalier et al. 2006).

The RNAi defects observed in *haf-6* mutants were reminiscent of RNAi defects observed in *mut*-7 and *rde*-2 mutants. Like these mutants, *haf*-6 mutants also exhibit defects in transposon silencing (Tops, Tabara et al. 2005). Similarly, mutants of *haf-9* also exhibit increased transposon mobilization in a temperature-dependent manner. Additionally, *haf-9, haf-2* and *haf-6* interact genetically with *mut-7* and with *rde-2,* displaying second-site non-complementation interactions (Sundaram, Han et al. 2008). These findings, along with RNAi defects in *haf-2* and *haf-6* mutants, suggest that half ABC_{RNAi} transporters might form homo/ heterodimers to traffic substrates required for RNAi in a MUT-7/RDE-2 dependent manner. It is, therefore, essential to elucidate the dimerization configuration and the specific substrates transported by half ABC_{RNAi} transporters to understand their role in RNAi processes.

MATERIALS

STRATEGY 1 – PREDICTED *C. elegans* ZIPPER SEQUENCES – PLASMIDS pLT642 / pLT643

Primers

LT1237 -TATATCCGGAGCGCGCTTGTTTGTCTGCCATGATGT

LT1240 - TATATCCGGAGCGCGCTTTGTATAGTTCATCCATGCC

LT1369 -

ACCGCGGTCAACGCCTTGACGCTGAACTCCTCGAGCTCAACAGAGCTCTTGAACATTTCAGAGCC

G GCGCTGCTGCTAAC

LT1370 -

CCACCAAATGCCGTCAAAAGAAAATGGATCGCATCAAGGAGCTGGAAGAACAGGTTCTCCACGA
GAAGCACCGCGGTCAACGC

LT1372 -

AGCAACGACAAAAAAATGGCGGAATGCAACAACATCCGAAATAAGCTCAACAGTCTCGCCGGCG
CT GCTGCTAAC

LT1373 -

ATTGATTTGATGAAGGAATTGCAAGATCAAGTAAATGACTTCAAAAATAGCAACGACAAAAAAA
TGGCG

LT1375 -

TATAGCTAGCATGCAAGAAAAGAAGAAGCTTGAGAGAAAGAGAGCTCGCAATAGGCAAGCCGC
CAC CAAATGCCGTCAAA

LT1376 -

TATAGCTAGCATGCAAAGGAATAAAGAAGCTGCTGCAAGATGTCGGCAAAGGAGAATTGATTTG
AT GAAGGAATT

Plasmids

pLT 642 → *let858* :: Jun zipper-CMV linker-N-GFP – *fog-2* 3' UTR with *unc-119* rescue sequence flanked by ttTi 5605 mos left and right homologous regions.

pLT 643 →*let858* :: Fos zipper-CMV linker-C-GFP – *fog-2* 3' UTR with *unc-119* rescue sequence flanked by ttTi 5605 mos left and right homologous regions.

Fos & Jun zipper sequence – Predicted *C. elegans* zipper sequences

Linker sequence – pCMV-FLAG from Sigma – AAANSSIDLISVPVDSR

Transgenic Strains

XX2090 – XX2098- Single Copy Integrants Plasmid LT642. yyIs322 [*unc-119*+ transgene in *unc-119* mutant] [*let-858* promoter::Jun LeuZip::CMV linker (Sigma)::N-GFP::*fog-2* 3' UTR in ttTi5605 mosSci insertion vector injected into XX1327]

XX2208, XX2211 – XX2213 - Single Copy Integrants Plasmid LT643. yyIs322 [*unc-119*+ transgene in *unc-119* mutant] [*let-858* promoter::Fos LeuZip::CMV linker (Sigma)::C-GFP::*fog-2* 3' UTR in ttTi5605 mosSci insertion vector injected into XX1327]

STARTEGY 2 – C. elegans OPTIMIZED HUMAN ZIPPRER SEQUENCES – PLASMIDS pLT694/ pLT695

Primers

LT1442 –

TATATAGCTAGCATGTCCATCGCTCGTCTCGAGGAGAAGGTCAAGACCCTCAAGG

CTCAG

LT1443 - TATATAGCCGGCGTTCATGACCTTCTGCTTGAGCTGAGCGACCTGCTCACGGAGCATGTT

LT1444 - CAAGACCCTCAAGGCTCAGAACTCCGAGCTCGCTTCCACCGCTAACATGCTCCGTGAGCA

LT1445 –

TATATAGCTAGCATGATGCTCACCGACACCCTCCAGGCTGAGACCGACCAGCTCG

AGGAC

LT1446 –

TATATAGCCGGCGTGAGCAGCGAGGATGAACTCGAGCTTCTCCTTCTCCTTGAGGA

GGTT

LT1447 -

GACCAGCTCGAGGACGAGAAGTCCGCTCTCCAGACCGAGATCGCTAACCTCCTCAAGGAG

Plasmids

pLT694 → *let858* :: *C. eleganised* human *jun* zipper-CMV linker-N-GFP – *fog-2* 3' UTR with *C. briggsae unc-119* rescue sequence flanked by ttTi 5605 mos left and right homologous regions.

pLT695 → *let858* :: *C. eleganised* human *fos* zipper-CMV linker-C-GFP – *fog-2* 3' UTR with *C. briggsae unc-119* rescue sequence flanked by ttTi 5605 mos left and right homologous regions.

Fos & Jun zipper sequences – *C. elegans* optimized human zipper sequences

Linker sequence – pCMV-FLAG from Sigma - AAANSSIDLISVPVDSR

Transgenic Strains

XX2217 - Single copy integrant for pLT 694 injected in XX1327 (mutant in *unc-119*) with ttTi5605 MosSCI insertion vector.

XX2218 - Single copy integrant for pLT 695 injected in XX1327 (mutant in *unc-119*) with ttTi5605 MosSCI insertion vector.

STRATEGY 3 – *C. elegans* OPTIMIZED HUMAN ZIPPERS DRIVEN BY eft-3 PROMOTER – PLASMIDS pLT729/ pLT730

Primers

LT1563 - TATAGCTAGCTACGGAGTGAGCAAAGTGTTTCCCAACTGAAAA

LT1562 - TATACTTAAGGCACCTTTGGTCTTTTATTGTCAACTTCCA

Plasmids

pLT729 → *eft-3* :: *C. eleganised* human *jun* zipper-CMV linker-N-GFP – *fog-2* 3' UTR with *C. briggsae unc-119* rescue sequence flanked by ttTi 5605 mos left and right homologous regions.

pLT730 → *eft-3* :: *C. eleganised* human *fos* zipper-CMV linker-C-GFP – *fog-2* 3' UTR with *C. briggsae unc-119* rescue sequence flanked by ttTi 5605 mos left and right homologous regions.

Fos & Jun zipper sequences – *C. elegans* optimized human zipper sequences

Linker sequence – pCMV-FLAG from Sigma - AAANSSIDLISVPVDSR

Transgenic Strains

XX2219 – pLT 729/pLT730 co-injection array line

STRATEGY 4 – *C. elegans* OPTIMIZED HUMAN ZIPPERS DRIVEN BY eft-3 PROMOTER in DIFFERENT Mos VECTOR BACKGROUNDS – pLT729/ pLT753

Primers

1534 –TATA CCTAGG GCACCTTTGGTCTTTTATTGTCAA

1546 – TATATAAGATCTAATGCCGAGTACGCAGTTGTGT

Plasmids

pLT729 → *eft-3* :: *C. eleganised* human jun zipper-CMV linker-N-GFP – fog-2 3' UTR with *C. briggsae* unc-119 rescue sequence flanked by ttTi 5605 mos left and right homologous regions.

pLT 753 → *eft-3* :: *C. eleganised* human fos zipper-CMV linker-C-GFP – fog-2 3' UTR with *C. briggsae* unc-119 rescue sequence flanked by cxTi 10816 mos left and right homologous regions.

Fos & Jun zipper sequences – *C. elegans* optimized human zipper sequences

Linker sequence – pCMV-FLAG from Sigma - AAANSSIDLISVPVDSR

METHODS

PLASMID CLONING

1. Amplification of *haf* genes

haf 1-9 genes were amplified by PCR based methods. Optimization was carried out for each *haf* gene with different polymerase enzymes, annealing temperatures and primer sequences (For conditions, see Table 2).

	template	primer 1	primer 2	DNA Pol	size(bp)	cycle						
haf-1	Gen DNA	537	547	Phusion	2839	95C 1m	55C 2m	72C 3m	36 X			
haf-2	Gen DNA	539	1548	Phusion	2645	95C 1m	49C 45s	72C 6m	30 X	Touchdown	56-46C 45s	10X
haf-3	Gen DNA	541	1549	Phusion	5537	95C 1m	49C 45s	72C 6m	30 X	Touchdown	56-46C 45s	10X
haf-4	Gen DNA	1550	1556	Goldstar Takara	6038	98C 10s	55C 30s	68C 8m	34 X			
haf-4A	Gen DNA	1609	1556	Goldstar Takara	4856	98C 10s	55C 30s	68C 8m	34 X			
haf-5	Gen DNA	545	1551	Goldstar Takara	5537	98C 10s	55C 30s	68C 8m	30 X			
haf-6	p662	1524	1552	Q5 HF	2934	95C 1m	53.8C 2m	72C 4m	39 X			
haf-6B	p662	1610	1524	Q5 HF	2421	95C 1m	53.8C 2m	72C 4m	39 X			
haf-7	Gen DNA	547	1553	Taq Pol	3966	95C 2m	62C 2m	68C 6m	35 X	Touchdown	70-60C 2m	10x
haf-8	Gen DNA	1554	549	Goldstar Takara	5934	98C 10s	55C 30s	68C 8m	34 X			
haf-9	Gen DNA	1555	1557	Phusion	4038	95C 1m	56C 2m	72C 4m	34 X			
haf-9B	Gen DNA	1555	1531	Q5 HF	3231	95C 1m	57C 2m	68C 4m	34 X			

Table 2 : Optimized PCR conditions for generating ABC transporters coding regions.

The DNA sequence for several of the *haf* genes is surrounded by difficult-to-PCR repetitive DNA (and some introns also harbored repetitive sequences). PCR optimization was required for most genes. Genomic DNA was used as introns are apparently required in *C. elegans* for efficient nuclear export of mRNAs.

2. Final plasmid cloning

Due to the size of the inserts and plasmid vectors, cloning of *haf* transgenes was carried out using classical restriction enzyme – based strategies as indicated in the chart below. Top10 or SURE *E. coli* cells were used in $CaCl_2$ -mediated transformations or electroporation. Ampicillin resistance selected for transformants. Colonies were screened based on size changes, and/or PCR-based screening for the presence of inserts. Candidates were extensively analyzed for correct restriction-enzyme fragment length. All plasmids were subjected to DNA sequencing of the entire region of interest.

Plasmid Vector with N-GFP	Fragments	Restriction Enzymes		Final	Size (bp)
pLT 729 (Nhe1/NgoM4)	*haf-1*	NheI	NgoM4	pLT 741	12147
	haf-2	NheI	NgoM4	pLT 742	11953
	haf-3	NheI	BspE1	pLT 743	14846
	haf-4A	NheI	Xma1	pLT 744	15354
	haf-4B	NheI	Xma1	pLT 764	14162
	haf-5	NheI	NgoM4	pLT 745	17577
	haf-6A	NheI	Age1	pLT 746	17884
	haf-6B	NheI	Age1	pLT 766	17431
	haf-7	NheI	NgoM4	pLT 747	13278
	haf-8	NheI	NgoM4	pLT 748	15246
	haf-9A	NheI	Xma1	pLT 749	13354
	haf-9B	NheI	Age1	pLT 769	12545

TABLE 3A : Cloning of BiFC vectors by classical restriction enzyme based strategy. This table includes transgenic plasmids pLT741-pLT749 (and pLT764, pLT769) that are constructed from proof-of-principle vector pLT729.

Plasmid Vector with C-GFP	Fragments	Restriction Enzymes		Final	Size (bp)
pLT 753 (Nhe1/NgoM4)	*haf-1*	NheI	NgoM4	pLT 731	12272
	haf-2	NheI	NgoM4	pLT 732	12078
	haf-3	NheI	BspE1	pLT 733	14971
	haf-4A	NheI	Xma1	pLT 734	15479
	haf-4B	NheI	Xma1	pLT 754	14287
	haf-5	NheI	NgoM4	pLT 735	17702
	haf-6A	NheI	Age1	pLT 736	18009
	haf-6B	NheI	Age1	pLT 756	17556
	haf-7	NheI	NgoM4	pLT 737	13403
	haf-8	NheI	NgoM4	pLT 738	15371
	haf-9A	NheI	Xma1	pLT 739	13479
	haf-9B	NheI	Age1	pLT 759	12670

TABLE 3B : Cloning of second set of BiFC vectors by classical restriction enzyme based strategy. This table includes transgenic plasmids pLT731-pLT739 (and pLT754, pLT759) that are constructed from proof-of-principle vector pLT753.

3. Worm strains and maintenance

All *C. elegans* strains were maintained on NGM plates that contained OP-50, an auxotrophic strain of *E. coli* (Brenner 1974). N2 wildtype and *unc-119 C. elegans* were maintained at room temperature (22°C). We used *pha-1(e2123),* a temperature-sensitive mutant which grows normally in 15°C. *pha-1* embryonic lethal mutants do not grow in 25°C (Granato, Schnabel et al. 1994). Microinjecting transgenic DNA with *pha-1* wildtype copy rescues the mutant from embryonic lethality phenotype. Thus, it is easy to select for transformants by growing F1 larvae from injected hermaphrodites at 25°C.

4. Microinjection

Three different strategies that will produce transgenic lines are possible, using the plasmids we generated: a simple injection strategy that will produce extrachromosomal arrays with multiple copies of our plasmid and two strategies that can produce single-copy

integration for each plasmid (MosSCI- and CRISPR-mediated integration). The plasmid vectors in each construct we generated harbor a rescuing segment of the *unc-119* gene derived from *C. briggsae*; therefore, our injections utilize *C. elegans unc-119* mutants. Successful uptake of DNA and generation of stably inherited lines is indicated by wild-type movement (*unc*-119 mutants are paralyzed). For MosSCI, co-injection of a plasmid capable of expressing Mos transposase into *unc-119* mutants that also harbor a single copy of a defective Mos transposon will allow for single copy integration into a particular locus (the plasmid vectors harbor recombinogenic sequences that flank the Mos integration site.) Co-injection of GFP- or RFP-expressing plasmids allow extrachromosomal arrays to be distinguished from integrants. For CRISPR, injection mixes are similarly made, but instead of using Mos transposase to make dsDNA breaks, Cas9-expressing plasmid is used, along with guide RNA sequences that match the location of the Mos insertion. A co-CRISPR injection strategy was used to generate CRISPR integrants (Arribere, Bell et al. 2014). By including *pha*-1 guide RNA and recombination template DNA in the injection mix and injecting into temperature sensitive *pha-1* mutants, animals that have both taken up DNA and have obtained a *pha-1* reversion can be screened for half transporter plasmid integration, All candidate integrants will be extensively analyzed genetically, for appropriate Mendelian segregation patterns, and molecularly, by PCR analysis of integration sites for validation.

GENETIC CROSSES TO REVEAL HALF TRANSPORTER DIMERIZATION PARTNERS

Genetic crosses require males but the frequency of *C. elegans* males naturally arising from self-fertilizing hermaphrodites is low in wildtype. Therefore, RNAi reagents that

produce increased frequency of males were developed (discussed in Chapter 2). Male stocks will be generated for all strains that express each of *haf::N-GFP* transgenes (shown along y axis in black letters in the table).

haf : *C-gfp* transgene (pLT753 backbone)

haf : *N-gfp* transgenes (pLT729 backbone)	*haf-1*	*haf-2*	*haf-3*	*haf-4*	*haf-5*	*haf-6*	*haf-7*	*haf-8*	*haf-9*
haf-1									
haf-2									
haf-3									
haf-4									
haf-5									
haf-6									
haf-7									
haf-8									
haf-9									

Figure 7 : Genetic crosses between transgenic worms carrying BiFC vectors. This figure shows how genetic crosses will be set-up between worms.

These males will be allowed to mate with hermaphrodites each of which harbors one of *haf::C-GFP* transgenes (shown along x axis in red letters in the Figure 7). The presence of of male progeny in the F1 generation indicates successful mating between two parents, and GFP fluorescence in the progeny will indicate that the two half transporters being

investigated can act as dimers . In addition, the F1 animals will be verified for inheritance of transgenes from both parents by PCR amplification of regions unique to each transgene. For crosses that fail to produce GFP fluorescence, PCR analysis will help confirm the presence of both halves of the GFP and help validate that the lack of GFP fluorescence is due to a non-interaction. In this nine-by-nine set of crosses will be reciprocal crosses—each interaction will be effectively tested twice, helping to confirm our interactions. Crosses between parents having *haf-5::N-GFP* and *haf-5::C-GFP* will serve as positive control as HAF-5 protein has previously been shown to homodimerize (Kim, Selote et al. 2010).

In our experiments, we will have the opportunity to investigate if half transporters can form distinct patterns in different cells. An ability of half transporters to dimerize with different partners in different cells has not been fully investigated in any organism, and C. elegans, being a simple multicellular system with transparent cells, is an ideal system to investigate such a possibility. We also predict that we will observe different subcellular localization patterns in different tissues, as this has been observed for human transporters. Interestingly, this will highlight the potential for underlying differences in specificities towards different substrates in these membranes.

RESULTS AND DISCUSSION

DESIGN AND CONSTRUCTION OF PROOF-OF-PRINCIPLE VECTORS

Choice of reporter protein and the split site

Bimolecular Fluorescence Complementation is a technology that allows for the determination or validation of protein:protein interactions in vivo. It is based on the ability of a "split" fluorescent protein to regain fluorescence when the two parts of the protein come into close proximity. When two proteins that normally interact are fused to each half of a split fluorescent molecule, the associating proteins bring the non-fluorescent fragments together, allowing them to associate and regain fluorescence.

GFP is the first fluorescent reporter used in bimolecular fluorescent complementation studies by Regan and colleagues. Green fluorescent protein is made of 238 amino acids constituting 11 β strands forming a cylindrical protein. An alpha helix runs through the center of the barrel. Each strand is connected by loops in between. Three of the four conserved residues, part of chromophore of the fluorescent molecule, Y66, G67, R96 are in the first 10 strands whereas the 11^{th} strand contains E222 residue. All the four residues are required for complete maturation of the chromophore which then results in fluorescence of GFP (Craggs 2009). A number of ways of severing the fluorescent molecule has been adapted in BiFC type studies. The most prevalent one that has been shown to function in different systems is the splitting of GFP into 157 residue long N-GFP (1-157 amino acids) fragment and an 81 residue long C-GFP (158-238 amino acids) fragment (Wilson, Magliery et al. 2004). Remy et al, 2004 in their research, aiming at identifying interacting partners of the protein kinase PKB/Akt, employ GFP based BiFC. Here, they split the GFP such that the N-terminal comprises the first

158 amino acids with the rest of residues in a C-terminal fragment. We chose to split the GFP between residues 157 and 158 because this split has previously been shown to be optimal for reconstitution through leucine zippers in *C. elegans* by Chalfie et al (Zhang, Ma et al. 2004).

In addition to GFP, many research groups, in their BiFC assays, adopted the use of other fluorescent reporters such as YFP, CFP, RFP and other variants of GFP attached to interacting proteins at their N and C termini (Hu and Kerppola 2003) ((Nagai, Sawano et al. 2001) (Hu, Chinenov et al. 2002)) (Jach, Pesch et al. 2006) (Hiatt, Shyu et al. 2008). For our experiments, we chose green fluorescent protein as the reporter for detecting interactions due to its photostability and its ability to fluoresce in all intracellular organelles. Also, a number of experiments using *C. elegans* have demonstrated that the GFP reporter, when fragmented, does not spontaneously re-associate in the absence of putative interactor proteins/peptides (Lindman, Johansson et al. 2009) . Furthermore, since we are analyzing protein:protein interactions between membrane proteins, the fluorescence will be concentrated into a two-dimensional lipid bilayer; thus, we anticipate that we will obtain a sufficiently bright signal for detection. From our preliminary data of ABC transporter :: GFP fusion proteins, we observed that GFP tagged HAF-6 ABC transporter protein colocalized with anti-HAF-6 antibodies and did not disrupt the ability of the transgene to rescue phenotypes, suggesting that C-terminal tagging of ABC transporters is functionally tolerated (Sundaram, Echalier et al. 2006).

Choice of linkers

Peptide linkers are short sequences of amino acids that separate consecutive domains of a protein. Linkers can facilitate proper folding of fusion proteins, the mobility of their domains and in improving their expression (Chen, Zaro et al. 2013). Especially, in our experiments which involve two non-functional reporter fragments that associate and fluoresce only when the protein partners interact, it is important for the linker to prevent steric constraints imposed by the interacting proteins and provide flexibility to C- and N-GFP (Kerppola 2006). Ghosh et al (2004) constructed their fusion protein using 6 and 4 amino acids long peptides as linkers. A similar short peptide linker was used to connect the helical zippers with GFP fragments in research by another group (Regan, 2000). (Gly_n-Ser_x) glycine spacer linkers are also commonly used to impart flexibility in such BiFC fusion vectors ((Hiatt, Shyu et al. 2008) and (Remy and Michnick 2004)), although these sequences can confer protein instability. A linker peptide derived from cytomegalovirus (CMV linker) with amino-acid sequence AAANSSIDLISVPVDSR has been successfully used in a number of BiFC experiments expressing interacting proteins in a variety of different subcellular organelles (Hu and Kerppola 2003). This is the linker peptide that we used in most of our experiments.

Choice of interacting proteins – Zipper sequences

Zippers, made of leucine rich hydrophobic core sequences, have been extensively used in protein ::protein interaction studies. Leucine zipper helices are capable of interacting in parallel or anti-parallel orientations. Such zippers, fused to truncated reporter polypeptides, have been regularly used in BiFC experiments (Kerppola 2008). The first

demonstration of the BiFC technique performed in *C. elegans*, involved the use of leucine zippers heterodimerizing in antiparallel orientation (Zhang, Ma et al. 2004). As our experimental design depends upon re-association of the GFP fragments as fused in a parallel fashion to ABC transporter proteins, we sought out to determine if this parallel configuration for fusion proteins was compatible with the BiFC method. One hint that parallel fusions might be the case compatible was a study that made use of the C. *elegans* FOS-1 and JUN-1 proteins to demonstrate BiFC in live worms (Hiatt, Shyu et al. 2008). While this group reported success with these, putative parallel, zippers, their study used whole *C. elegans* protein and not isolated zipper motifs. Furthermore, they did not observe reconstituted FOS/JUN-dependent fluorescence in a wild type background—they utilized a sensitized *smg-1(cc546ts)* background to observe fluorescence. Isolated leucine zippers from mammalian FOS and JUN form parallel heterodimers, yet the nature of the dimerization in *C. elegans* FOS and JUN proteins has not been analyzed biochemically. It may be the case that *C. elegans* FOS and JUN proteins utilize different dimerization motifs, or the entire proteins might be prevented from forming dimers as fusion proteins. Therefore, during the course of building our plasmid vectors, we decided to incorporate the putative leucine zipper motifs from C. *elegans* FOS and JUN proteins as a test of how the BiFC method might work when GFP fragments are brought together as parallel fusion proteins.

The first pair of BiFC vectors in strategy 1 (Figure 8A), pLT642/pLT643 utilized such *C. elegans fos-1* and *jun-1* putative zippers (leucine zipper transcription factor genes in *C. elegans* orthologous to mammalian *fos* and *jun* genes respectively). We did not observe GFP fluorescence in worms that harbored both these control plasmids. We considered that our plasmid design may have failed to include important FOS and JUN dimerization residues;

Alternatively, the *C. elegans* FOS and JUN proteins may dimerize in a manner different from the mammalian FOS/JUN (the biochemistry of FOS and JUN dimerization in *C. elegans* is not well-studied.) We therefore replaced these *C. elegans* with human *fos* and *jun* zipper motifs, codon optimizing the sequences for expression in *C. elegans*. We refer to these dimerization motifs as '*C. eleganised* human zipper sequences'. These zippers were efficient in interacting with each other as predicted, which was evident from fluorescence obtained from GFP reassociation in array worms that expressed transgenes from vectors pLT694 and pLT695 (Figure 8B).

Choice of promoters

While it is relatively straightforward to generate transgenic strains of *C. elegans*, expression of proteins from the introduced DNA has been demonstrated to be problematic (Kelly, Xu et al. 1997). In somatic cells, the transgenic DNA is often well-expressed; however, the DNA is frequently silenced in germline tissue. This is likely due to anti-foreign genome responses that are particularly robust in germline tissue (Kelly, Xu et al. 1997). The *let-858* promoter is expressed well in somatic cells and has also been demonstrated to drive expression in germline tissue (Kelly, Xu et al. 1997). *let-858* promoter has been used largely to drive the expression of transgenes in *C. elegans* due to its ubiquitous nature.

We first tested our *C. elegans fos-1* and *jun-1* are cloned in plasmids pLT643/pLT695 and pLT642/pLT694 respectively under the transcriptional control of *let-858* promoter. A report by Jorgensen et al, 2012 demonstrated that *eft-3*(known as *eef-1A.1*) is the most effective promoter in transgene insertion by MosSCI method. Here, the author shows that

transposase driven by *eft-3* has an improved Mos1 transposition (Frokjaer-Jensen, Davis et al. 2012). This led us to hypothesize that transgenes driven from *eft-3* p

will result in a ubiquitous expression. Moreover, evidences showing high levels of transgene expression from *eft-3* encouraged us to choose this ubiquitous promoter to regulate expression of our BiFC vectors pLT729, pLT730 and pLT753 in cloning strategies 3 and 4 (Mitrovich and Anderson 2000).

Choice of Mos1 insertion sites

Many insertion sites are available for recombination of transgenes by *Mos1* mediated SCI(Single Copy Insertion) in *C. elegans* (Frokjaer-Jensen, Davis et al. 2012). Generally, transgenes are expressed in two copies from a single *Mos1* site after homologous recombination in *C. elegans* genome. But in our BiFC assay experiments, on crossing single copy integrant worms, the progeny harbored one copy of each BiFC vector. In our experiments from strategies 1 & 2, that used vectors pLT 642(or pLT 694) and pLT643 (or pLT 695), Mos based Single Copy Insertion was carried out at the TTti5605 locus on chromosome II. In an effort to strengthen the reporter signals from our BiFC expression vectors, we proposed the insertion of each of a pair of BiFC vector in one of the two Mos1 sites. This will result in two copies of each transgene construct being recombined in *C. elegans* genome. To achieve this, in pLT753, one of the constructs in the set of BiFC vectors from strategy 3, we cloned our transgene *Peft-3*::*C.eleganised* human *fos* zipper:CMV linker:C-GFP::*fog-2* 3' UTR such that recombination occurs at cxTi10816 *Mos1* locus on chromosome IV (Figure 8D). Thus, F1 progeny obtained from mating single copy integrants of pLT729 and pLT753 vectors will possess two copies of each expression plasmid.

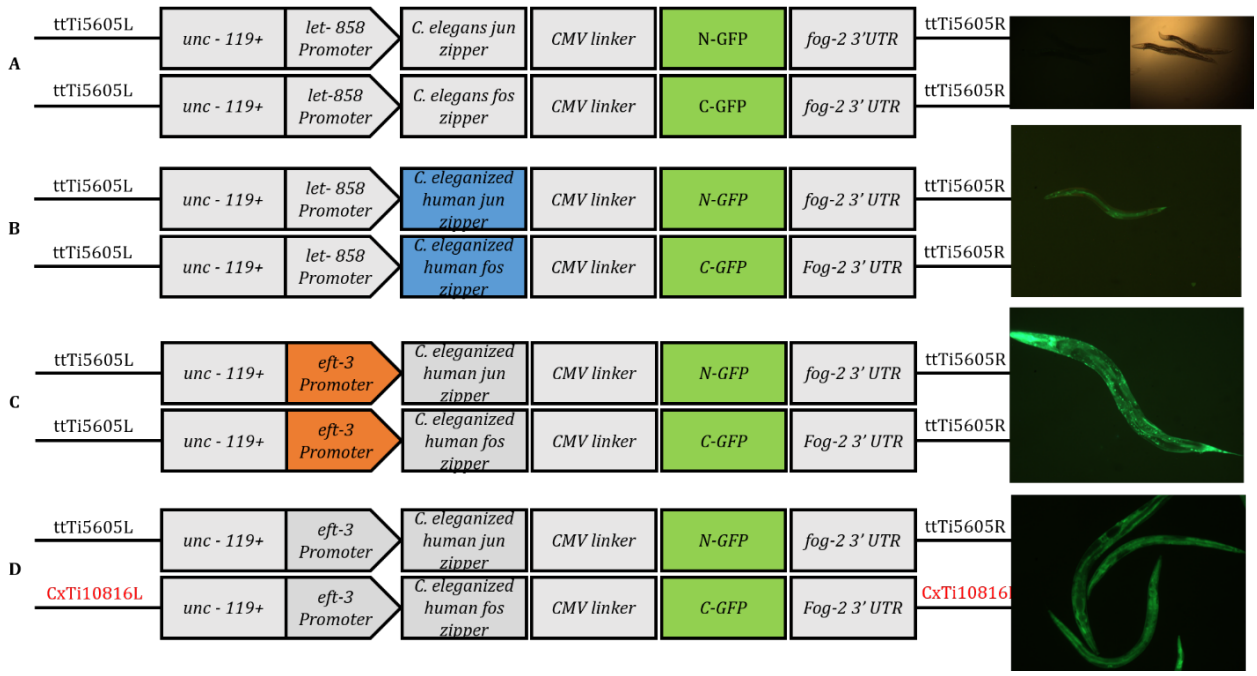


Figure 8: BiFC vectors and GFP complementation in *C. elegans*. A) Transgenic worms did not express GFP when *C. elegans jun* and *fos* zippers were used. B) Human FOS and JUN zipper optimized for *C. elegans* successfully resulted in GFP reassociation resulting in fluorescence. Shown in B) is an array worm having the two vectors shown in left. C) An array worm expressing GFP after co-injection of the two vectors shown on its left. D) GFP fluorescence in worms after co-injection of the two vectors shown on its left. It is expected that when these transgenes in D) are microinjected into separate worms, it will result in two copies of each transgene being expressed.

These experiments demonstrate the utility of our expression plasmids in determining protein-protein interactions. This study provides guidance in designing and constructing optimal GFP based BiFC vectors interacting in parallel orientation (Figure 9). Additionally,

this study also provides vector tools that can be of further use in protein interaction studies in *C. elegans*.

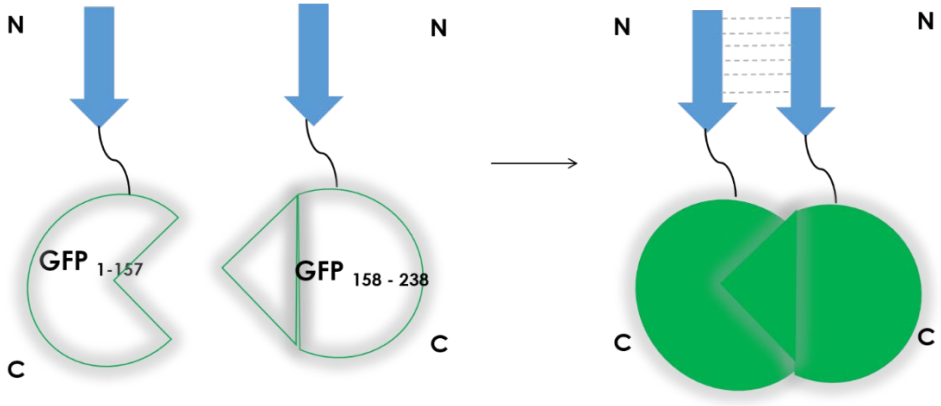


Figure 9 : BiFC model using GFP fluorescence to report protein-protein interaction. Blue arrows indicate the interacting proteins – *C. elegans jun* zipper and *fos* zipper in case of pLT642 and pLT643 respectively, *C. eleganised* human *jun* zipper in pLT729 and *C. eleganised* human *fos* zipper in pLT730 and pLT753 vectors. A CMV linker is seen attached to the C terminal ends of zipper peptides. It is to be noted that the zipper sequences are oriented in the same direction to enable parallel dimerization. GFP reporter fragments (N-GFP1-157 and C-GFP158-238) are attached through the linker to the zipper peptides at their C termini. Fragmented GFP is non-fluorescent while interaction between the zippers cause the GFP molecule to regain fluorescence.

	Promoter	Zipper	Linker	3' UTR	Fluorescence
642/643	*let-858*	*C. elegans*	CMV	*fog-2*	No
694/695	*let-858*	C.eleganised human zippers	CMV	*fog-2*	**Yes**
729/730 1 copy of transgene	*eft-3*	C.eleganised human zippers	CMV	*fog-2*	**Yes**
729/753 2 copies of transgene	*eft-3*	C.eleganised human zippers	CMV	*fog-2*	**Yes**

Table 4 : Our expression plasmids are capable of demonstrating protein interactions.

CONSTRUCTION OF HALF ABC VECTORS WITH SPLIT-GFP TAGS

Two plasmids from above, pLT729 and pLT753 were further used as backbone vectors for constructing *haf-(1-9)* expression plasmids. In our design, *C. eleganised* human zippers in the vectors were replaced with each of *haf-1* to *haf-9* including isoforms *haf-4b*, *haf-6B* and *haf-9b*. Thus, a total of 24 plasmids were designed to be expressed in *C. elegans* in somatic and germline tissues.

CONCLUSION

Twenty two, out of the above mentioned twenty four transgenes are constructed and verified for accuracy by sequencing of *haf::GFP* inserts. Cloning of *haf-3* insert in the pLT729 backbone and *haf-7* insert in the pLT753 backbone requires further optimization. Vectors pLT729 and pLT753, when injected will recombine in two different *Mos1* sites, *ttTi5605* and *CxTi10816* respectively, in *C. elegans* genome. Therefore, two copies of *each* transgene, coming from each allele in one *Mos1* locus will be expressed in an animal from reassociation of GFP fragments in prospective partners. This is expected to aid in increased intensity of fluorescence in germline cells. Transformation of *C. elegans* worms with split-GFP transgenes will be carried out by micro-injection, followed by genetic crosses of transformants, to determine the dimerization pattern between half transporters (elaborated in Methods section).

This project will help in gaining insights about how homo/ heterodimerization of half transporters will influence their function. More specifically, knowledge of binding of these transporters to their functional partners will lead to developing assays aimed at determining substrates that are required for RNAi processes. Although many of their functions are identified, a link between ABC transporters and RNAi opens the possibility of unprecedented roles. RNAi related silencing is known to have evolved as a basic protective mechanism against viral genomes, foreign genomes and transposon mobilization. Characterizing the nature of substrates related to RNAi will lead to the emergence of new therapies to disorders linked to ABC transporter functions.

REFERENCES

1. Agrawal, N., P. V. Dasaradhi, A. Mohmmed, P. Malhotra, R. K. Bhatnagar and S. K. Mukherjee (2003). "RNA interference: biology, mechanism, and applications." Microbiol Mol Biol Rev **67**(4): 657-685.
2. Allikmets, R., W. H. Raskind, A. Hutchinson, N. D. Schueck, M. Dean and D. M. Koeller (1999). "Mutation of a putative mitochondrial iron transporter gene (ABC7) in X-linked sideroblastic anemia and ataxia (XLSA/A)." Hum Mol Genet **8**(5): 743-749.
3. Ardehali, H., T. Xue, P. Dong and C. Machamer (2005). "Targeting of the mitochondrial membrane proteins to the cell surface for functional studies." Biochem Biophys Res Commun **338**(2): 1143-1151.
4. Arribere, J. A., R. T. Bell, B. X. Fu, K. L. Artiles, P. S. Hartman and A. Z. Fire (2014). "Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans." Genetics **198**(3): 837-846.
5. Boxem, M., Z. Maliga, N. Klitgord, N. Li, I. Lemmens, M. Mana, L. de Lichtervelde, J. D. Mul, D. van de Peut, M. Devos, N. Simonis, M. A. Yildirim, M. Cokol, H. L. Kao, A. S. de Smet, H. Wang, A. L. Schlaitz, T. Hao, S. Milstein, C. Fan, M. Tipsword, K. Drew, M. Galli, K. Rhrissorrakrai, D. Drechsel, D. Koller, F. P. Roth, L. M. Iakoucheva, A. K. Dunker, R. Bonneau, K. C. Gunsalus, D. E. Hill, F. Piano, J. Tavernier, S. van den Heuvel, A. A. Hyman and M. Vidal (2008). "A protein domain-based interactome network for C. elegans early embryogenesis." Cell **134**(3): 534-545.
6. Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.

7. Chen, X., J. L. Zaro and W. C. Shen (2013). "Fusion protein linkers: property, design and functionality." Adv Drug Deliv Rev **65**(10): 1357-1369.
8. Craggs, T. D. (2009). "Green fluorescent protein: structure, folding and chromophore maturation." Chem Soc Rev **38**(10): 2865-2875.
9. Dermauw, W. and T. Van Leeuwen (2014). "The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance." Insect Biochem Mol Biol **45**: 89-110.
10. Edgar, R. C. (2004). "MUSCLE: a multiple sequence alignment method with reduced time and space complexity." BMC Bioinformatics **5**: 113.
11. Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Res **32**(5): 1792-1797.
12. Elliott, A. M. and M. A. Al-Hajj (2009). "ABCB8 mediates doxorubicin resistance in melanoma cells by protecting the mitochondrial genome." Mol Cancer Res **7**(1): 79-87.
13. Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.
14. Ford, R. C. and K. Beis (2019). "Learning the ABCs one at a time: structure and mechanism of ABC transporters." Biochem Soc Trans **47**(1): 23-36.
15. Frokjaer-Jensen, C., M. W. Davis, M. Ailion and E. M. Jorgensen (2012). "Improved Mos1-mediated transgenesis in C. elegans." Nat Methods **9**(2): 117-118.
16. Garces, F., K. Jiang, L. L. Molday, H. Stohr, B. H. Weber, C. J. Lyons, D. Maberley and R. S. Molday (2018). "Correlating the Expression and Functional Activity of ABCA4

Disease Variants With the Phenotype of Patients With Stargardt Disease." Invest Ophthalmol Vis Sci **59**(6): 2305-2315.

17. Granato, M., H. Schnabel and R. Schnabel (1994). "pha-1, a selectable marker for gene transfer in C. elegans." Nucleic Acids Res **22**(9): 1762-1763.
18. Haynes, C. M., Y. Yang, S. P. Blais, T. A. Neubert and D. Ron (2010). "The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in C. elegans." Mol Cell **37**(4): 529-540.
19. Heppel, L. A. (1969). "The effect of osmotic shock on release of bacterial proteins and on active transport." J Gen Physiol **54**(1): 95-113.
20. Hiatt, S. M., Y. J. Shyu, H. M. Duren and C. D. Hu (2008). "Bimolecular fluorescence complementation (BiFC) analysis of protein interactions in Caenorhabditis elegans." Methods **45**(3): 185-191.
21. Hu, C. D., Y. Chinenov and T. K. Kerppola (2002). "Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation." Mol Cell **9**(4): 789-798.
22. Hu, C. D. and T. K. Kerppola (2003). "Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis." Nat Biotechnol **21**(5): 539-545.
23. Ichikawa, Y., M. Bayeva, M. Ghanefar, V. Potini, L. Sun, R. K. Mutharasan, R. Wu, A. Khechaduri, T. Jairaj Naik and H. Ardehali (2012). "Disruption of ATP-binding cassette B8 in mice leads to cardiomyopathy through a decrease in mitochondrial iron export." Proc Natl Acad Sci U S A **109**(11): 4152-4157.

24. Jach, G., M. Pesch, K. Richter, S. Frings and J. F. Uhrig (2006). "An improved mRFP1 adds red to bimolecular fluorescence complementation." Nat Methods **3**(8): 597-600.

25. Jin, M. S., M. L. Oldham, Q. Zhang and J. Chen (2012). "Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans." Nature **490**(7421): 566-569.

26. Kawai, H., T. Tanji, H. Shiraishi, M. Yamada, R. Iijima, T. Inoue, Y. Kezuka, K. Ohashi, Y. Yoshida, K. Tohyama, K. Gengyo-Ando, S. Mitani, H. Arai, A. Ohashi-Kobayashi and M. Maeda (2009). "Normal formation of a subset of intestinal granules in Caenorhabditis elegans requires ATP-binding cassette transporters HAF-4 and HAF-9, which are highly homologous to human lysosomal peptide transporter TAP-like." Mol Biol Cell **20**(12): 2979-2990.

27. Kelly, W. G., S. Xu, M. K. Montgomery and A. Fire (1997). "Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene." Genetics **146**(1): 227-238.

28. Kerppola, T. K. (2006). "Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells." Nat Protoc **1**(3): 1278-1286.

29. Kerppola, T. K. (2008). "Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells." Annu Rev Biophys **37**: 465-487.

30. Kim, S., D. S. Selote and O. K. Vatamaniuk (2010). "The N-terminal extension domain of the C. elegans half-molecule ABC transporter, HMT-1, is required for protein-protein interactions and function." PLoS One **5**(9): e12938.

31. Koszarska, M., N. Kucsma, K. Kiss, G. Varady, M. Gera, G. Antalffy, H. Andrikovics, A. Tordai, M. Studzian, D. Strapagiel, L. Pulaski, Y. Tani, B. Sarkadi and G. Szakacs (2014). "Screening the expression of ABCB6 in erythrocytes reveals an unexpectedly high frequency of Lan mutations in healthy individuals." PLoS One **9**(10): e111590.
32. Krogh, A., B. Larsson, G. von Heijne and E. L. Sonnhammer (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." J Mol Biol **305**(3): 567-580.
33. Lee, J. Y., L. N. Kinch, D. M. Borek, J. Wang, J. Wang, I. L. Urbatsch, X. S. Xie, N. V. Grishin, J. C. Cohen, Z. Otwinowski, H. H. Hobbs and D. M. Rosenbaum (2016). "Crystal structure of the human sterol transporter ABCG5/ABCG8." Nature **533**(7604): 561-564.
34. Lindman, S., I. Johansson, E. Thulin and S. Linse (2009). "Green fluorescence induced by EF-hand assembly in a split GFP system." Protein Sci **18**(6): 1221-1229.
35. Luker, K. E. and D. Piwnica-Worms (2004). "Optimizing luciferase protein fragment complementation for bioluminescent imaging of protein-protein interactions in live cells and animals." Methods Enzymol **385**: 349-360.
36. Mitrovich, Q. M. and P. Anderson (2000). "Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in C. elegans." Genes Dev **14**(17): 2173-2184.
37. Nagai, T., A. Sawano, E. S. Park and A. Miyawaki (2001). "Circularly permuted green fluorescent proteins engineered to sense Ca2+." Proc Natl Acad Sci U S A **98**(6): 3197-3202.

38. Nargund, A. M., M. W. Pellegrino, C. J. Fiorese, B. M. Baker and C. M. Haynes (2012). "Mitochondrial import efficiency of ATFS-1 regulates mitochondrial UPR activation." Science **337**(6094): 587-590.

39. Noll, A., C. Thomas, V. Herbring, T. Zollmann, K. Barth, A. R. Mehdipour, T. M. Tomasiak, S. Bruchert, B. Joseph, R. Abele, V. Olieric, M. Wang, K. Diederichs, G. Hummer, R. M. Stroud, K. M. Pos and R. Tampe (2017). "Crystal structure and mechanistic basis of a functional homolog of the antigen transporter TAP." Proc Natl Acad Sci U S A **114**(4): E438-E447.

40. Paterson, J. K., S. Shukla, C. M. Black, T. Tachiwada, S. Garfield, S. Wincovitch, D. N. Ernst, A. Agadir, X. Li, S. V. Ambudkar, G. Szakacs, S. Akiyama and M. M. Gottesman (2007). "Human ABCB6 localizes to both the outer mitochondrial membrane and the plasma membrane." Biochemistry **46**(33): 9443-9452.

41. Plasterk, R. H. (2002). "RNA silencing: the genome's immune system." Science **296**(5571): 1263-1265.

42. Remy, I. and S. W. Michnick (2004). "Regulation of apoptosis by the Ft1 protein, a new modulator of protein kinase B/Akt." Mol Cell Biol **24**(4): 1493-1504.

43. Schaedler, T. A., B. Faust, C. A. Shintre, E. P. Carpenter, V. Srinivasan, H. W. van Veen and J. Balk (2015). "Structures and functions of mitochondrial ABC transporters." Biochem Soc Trans **43**(5): 943-951.

44. Sharom, F. J. (2008). "ABC multidrug transporters: structure, function and role in chemoresistance." Pharmacogenomics **9**(1): 105-127.

45. Sheps, J. A., S. Ralph, Z. Zhao, D. L. Baillie and V. Ling (2004). "The ABC transporter gene family of Caenorhabditis elegans has implications for the evolutionary dynamics of multidrug resistance in eukaryotes." Genome Biol **5**(3): R15.

46. Shintre, C. A., A. C. Pike, Q. Li, J. I. Kim, A. J. Barr, S. Goubin, L. Shrestha, J. Yang, G. Berridge, J. Ross, P. J. Stansfeld, M. S. Sansom, A. M. Edwards, C. Bountra, B. D. Marsden, F. von Delft, A. N. Bullock, O. Gileadi, N. A. Burgess-Brown and E. P. Carpenter (2013). "Structures of ABCB10, a human ATP-binding cassette transporter in apo- andnucleotide-bound states." Proc Natl Acad Sci U S A **110**(24): 9710-9715.

47. Shyu, Y. J., S. M. Hiatt, H. M. Duren, R. E. Ellis, T. K. Kerppola and C. D. Hu (2008). "Visualization of protein interactions in living Caenorhabditis elegans using bimolecular fluorescence complementation analysis." Nat Protoc **3**(4): 588-596.

48. Stefkova, J., R. Poledne and J. A. Hubacek (2004). "ATP-binding cassette (ABC) transporters in human metabolism and diseases." Physiol Res **53**(3): 235-243.

49. Sundaram, P., B. Echalier, W. Han, D. Hull and L. Timmons (2006). "ATP-binding cassette transporters are required for efficient RNA interference in Caenorhabditis elegans." Mol Biol Cell **17**(8): 3678-3688.

50. Sundaram, P., W. Han, N. Cohen, B. Echalier, J. Albin and L. Timmons (2008). "Caenorhabditis elegans ABCRNAi transporters interact genetically with rde-2 and mut-7." Genetics **178**(2): 801-814.

51. Tanji, T., K. Nishikori, H. Shiraishi, M. Maeda and A. Ohashi-Kobayashi (2013). "Co-operative function and mutual stabilization of the half ATP-binding cassette

transporters HAF-4 and HAF-9 in Caenorhabditis elegans." Biochem J **452**(3): 467-475.

52. Tops, B. B., H. Tabara, T. Sijen, F. Simmer, C. C. Mello, R. H. Plasterk and R. F. Ketting (2005). "RDE-2 interacts with MUT-7 to mediate RNA interference in Caenorhabditis elegans." Nucleic Acids Res **33**(1): 347-355.
53. Vasiliou, V., K. Vasiliou and D. W. Nebert (2009). "Human ATP-binding cassette (ABC) transporter family." Hum Genomics **3**(3): 281-290.
54. Vatamaniuk, O. K., E. A. Bucher, M. V. Sundaram and P. A. Rea (2005). "CeHMT-1, a putative phytochelatin transporter, is required for cadmium tolerance in Caenorhabditis elegans." J Biol Chem **280**(25): 23684-23690.
55. Whangbo, J. S., A. S. Weisman, J. Chae and C. P. Hunter (2017). "SID-1 Domains Important for dsRNA Import in Caenorhabditis elegans." G3 (Bethesda) **7**(12): 3887-3899.
56. Wilkens, S. (2015). "Structure and mechanism of ABC transporters." F1000Prime Rep **7**: 14.
57. Wilson, C. G., T. J. Magliery and L. Regan (2004). "Detecting protein-protein interactions with GFP-fragment reassembly." Nat Methods **1**(3): 255-262.
58. Zhang, F., W. Zhang, L. Liu, C. L. Fisher, D. Hui, S. Childs, K. Dorovini-Zis and V. Ling (2000). "Characterization of ABCB9, an ATP binding cassette protein associated with lysosomes." J Biol Chem **275**(30): 23287-23294.
59. Zhang, S., C. Ma and M. Chalfie (2004). "Combinatorial marking of cells and organelles with reconstituted fluorescent proteins." Cell **119**(1): 137-144.

CHAPTER 2

RNAi-induced non-disjunction of X chromosomes

ABSTRACT

In *C. elegans*, there are two genders – males and hermaphrodites. Gender in *Caenorhabditis elegans* is established by the number of X chromosomes in the developing embryo, with hermaphrodites having genotype XX and males, XO. Males typically arise by non-disjunction during meiosis, and the frequency of an X-chromosome non-disjunction event is relatively low. About 1 in 500 to 1 in 2000 progeny from a self-fertilizing wildtype hermaphrodite are males, depending on the incubation temperature and strain. Males are required for genetic crosses and phenotypic analysis, yet current methods to generate large numbers of males can be cumbersome. In our lab, we identify RNAi reagents (dsRNA-expressing bacteria) with improved effectiveness for eliciting males. Specifically, we used RNAi to systematically reduce the expression of over two hundred genes with meiotic chromosome segregation functions, and we identified a set of RNAi reagents that robustly and reproducibly elicited male progeny.

INTRODUCTION

Gender in *Caenorhabditis elegans* is established by non-disjunction of X chromosomes. *C. elegans* has six chromosomes, including the sex chromosomes. Gender classification is based on the number of X chromosomes. Hermaphrodites have 5 diploid autosomes and one diploid set of X chromosomes whereas males are diploid for the 5 autosomes and have a haploid X chromosome. The haplo-X males arise by non-disjunction of X chromosome during meiosis in a hermaphrodite worm (Hodgkin, Horvitz et al. 1979). Thus, the transparent, simple nematode is a fitting organism for meiotic chromosome segregation studies, as the presence of males is an indicator of an X-chromosome non-disjunction event during gametogenesis.

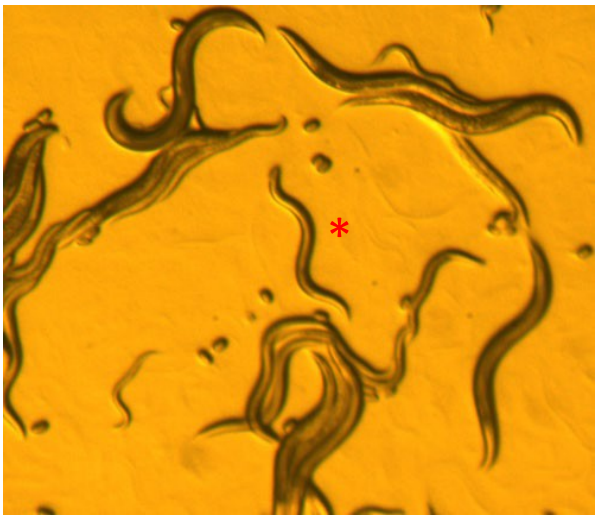


Figure 10: *C. elegans* male from self-fertilized hermaphrodites. The red star mark shows a male. Adult males, in a population of hermaphrodites are identified by their distinguishable morphology that bears a distinctive tail with a copulatory apparatus (pointed tip at the tail),

slimmer body and a clear ventral gonad. Males are also smaller and move faster than hermaphrodites.

This rare event in the hermaphrodite germline makes for infrequent male progeny as low as 1 in 500 worms. During its lifetime, a single wildtype hermaphrodite worm produces about 300 progeny. As a result, it is typically the case that a single hermaphrodite will produce no male offspring. Nevertheless, the need for males is realized in genetic studies that require mating between two genders.

Several strategies have been devised in order to obtain male animals, though each has its limitations. A simple method to increase the number of males in the progeny is to perform mating crosses between hermaphrodite worms and male worms. A successful cross results in half of the progeny being males (Ward and Carrel 1979). But when the parents of the cross have different genotypes, for example with one parent harboring a transgene and other multiple mutations, the strategy usually requires genotypic or phenotypic selection amongst the progeny, and again, using crosses to maintain the population of males with the desired genotype. Depending on the genotype of the strain, obtaining males with the desired set of mutations and transgenes can be quite laborious. Other methods shown to increase the occurrence of male progeny involve exposure to ethanol or heat shock of L4 hermaphrodites (Lyons and Hecht, 1997 and Sulston and Hodgkin, 1988). Unfortunately, these methods yield few males; A third method involves introducing *him* mutations into the strain. "*him*" refers to "high incidence of males" phenotype, and mutations in *him-5* and *him-8* are commonly introduced into the background of strains to increase the frequency of males to about 20-40% males (David S Fay, Wormbook and (Hodgkin, Horvitz et al. 1979). However, the

strategy of introducing *him* mutations into strains can be problematic, as discussed above. Furthermore, introduction of the *him* mutations using CRISPR requires a screening step, and many strains may not perform CRISPR efficiently. Finally, RNAi can be used to increase the number of males in a population by targeting the *him-14* gene. dsRNA corresponding in sequence to *him-14* gene can induce non-disjunction of X chromosomes (RNAi feeding to produce males, Worm Breeder's Gazette p. 32). The population of males with the desired genotype can be maintained by simply encouraging mating. This approach has the advantage of producing males in homozygotes with the required genotype, and as RNAi effects are reversible, there should be no long-lasting deleterious effects on the offspring. Of all the male-producing strategies mentioned above, RNAi using *him-14* is the least effective.

We hypothesized that it might be possible to produce improved RNAi tools for the purposes of producing males. We anticipated a need to generate males from a large number of different transgenic strains produced in the half transporter dimerization study. These strains have an *unc-119* mutation in the background, which renders the animals paralyzed, unless rescued by the *unc-119+* gene in our injected DNA. Performing crosses to generate males would require selection and examination of the *unc-119* background in the progeny of the cross, which would be challenging due to the presence of a wild type *unc-119* allele from the male (*unc-119* homozygous males are paralyzed and do not mate). Given the number of strains for which we require males, producing them by genetic crosses would be quite laborious. Instead, we anticipate using an RNAi strategy that will elicit males with the correct genotype in one generation. Our goal was to develop a better RNAi tool for the efficient production of males.

MATERIALS

Worm and Bacterial Strains

NL3531 [*rde-2(pk1657)*], PD8186 [*rde-2(ne221)*], NL1820 [*mut-7(pk720)*], and NL917 [*mut-7(pk204)*]. EG4322 [*ttTi5605* II; *unc-119(ed3)* III]. Wild type strains used: XX935 [N2 received from *C. elegans* Genetics Center Feb. 2006. DR sub clone of CB original, Tc1 pattern I]. HT115(DE3): W3110, rnc14::DTn10(Dasgupta et al., 1998; Takiff et al., 1989).

Feeding plasmids

Feeding plasmids were obtained from Source Biosciences LifeSciences (Fraser, Kamath et al. 2000); the genomic sequences residing in these plasmids can be found in (Timmons, Luna et al. 2014), Supplementary Table S2. Additional plasmids constructed include: pLT651 (BA) *klp-16* (BW) + *him-8* (FW) in double T7 RNAi feeding vector L4440 (L4440 was a gift from Andrew Fire); pLT651 (BB) *klp-16* (BW) + *him-8* (BW) in double T7 RNAi feeding vector L4440 ; pLT652 (BA) *dhc-3* (FW) + *him-8* (BW) in double T7 RNAi feeding vector L4440 ; pLT652 (BB) *dhc-3* (BW) + *him-8* (BW) in double T7 RNAi feeding vector L4440 ; pLT653 (BA) *klp-16* (BW) + *dhc-3* (FW) in double T7 RNAi feeding vector L4440—as described in Methods, below.

METHODS

Identifying candidate genes to be targeted

We compiled a list of genes that are likely to induce chromosomal non-disjunction as targets for RNAi in our assays for male production by mining through publications and through data sets deposited on Wormbase using WormMart and WormMine data mining tools. Plasmid templates that produce dsRNA against these genes were obtained from Source BioScience LifeSciences (Fraser, Kamath et al. 2000). The clones were verified using PCR and also by using CelRNAi (http://biocompute.bmi.ac.cn/CelRNAi/) (Qu, Ren et al. 2011).

RNAi by feeding

C. elegans is maintained in the laboratory on Nematode Growth medium (NGM) plates with *E. coli* bacterial cells. Normally, worms feed on these bacteria. In RNAi by feeding, HT115 (DE3) cells are engineered to express dsRNA. HT115 bacterial cells have RNAse III deletion and an IPTG inducible T7 polymerase that comes from DE3 lysogen. HT115 cells are Tetracycline resistant. Feeding plasmids are designed to harbor the specific gene fragments between two opposable T7 promoters. HT115 (DE3) bacterial cells used in feeding experiments were transformed using standard $CaCl_2$ transformation protocol with the feeding plasmid of interest using LB-Ampicillin plates. A fresh colony from these plates containing the plasmid was grown in LB-Amp (100 ug/ml) +Tet (12.5 ug/ml) liquid culture overnight at 37°C. The overnight culture, on the next day was diluted more than 100 times

and was allowed to grow until it reached an OD600 of 0.4 - 0.8 before inducing with Isopropyl-b-d-thiogalactopyranoside (IPTG) at a final concentration of 0.4mM at 37℃ for 2-4 hrs. The cells were then directly added to NGM (Nematode Growth Medium) plates. These NGM plates were made by following standard procedure (Brenner 1974) butsupplemented with 100 mg/ml ampicillin, 12.5 mg/ml tetracycline (TAI), and 0.4 mM IPTG. Thus in a feeding protocol, large numbers of worms on a plate are fed with HT115 (DE3) cells expressing dsRNA against target regions of the worm genome. Wildtype worms (4 worms/plate) at L1/L2 stage were added onto each NGM-TAI plate and F1 progeny were assayed for RNAi phenotype (Him phenotype, in our case) from the next 1-5 days depending on the gene targeted. Plates that were freshly seeded were used for all experiments and it was ensured that no plates were depleted of the bacterial food during the course of the experiment (Timmons, Court et al. 2001).

Him phenotype

RNAi feeding assays were performed at 22℃ or 25℃. In some experiments performed at 25℃, a higher percentage of males was observed, due to a general increase in lethality of progeny by increased non-disjunction of autosomes at this higher temperature. Serial transfer of treated animals to RNAi plates also led to sterility of treated animals for some targeted genes which indirectly increased the percentage of males in the F1 progeny. We utilized two procedures previously demonstrated to improve the consistency of the feeding method: i) we used recently transformed HT115 (DE3) cells + dsRNA-expressing plasmid—transformation. Transformation plates older than one week were not used in

inoculations; ii) we used freshly seeded RNAi feeding plates – plates were seeded with induced cells the day before adding worms.

The presence of males may not indicate the efficiency of the knock-down experiment because the phenotype we are scoring (Him) is a readout for reduction in expression of genes involved in chromosome disjunction, and non-disjunction events are rare and unpredictable. The percentage of viable progeny could be skewed as non-disjunction of autosomes leads to sterility of treated animals. This is evident because we observed reduced brood sizes (0-100 progeny from three parental worms) in at least one experiment for most genes. Therefore, we do not report our observations as 'percentage of males in total progeny' as this may reflect a bias in attainment of autosomal aneuploidy and not X-chromosome non-disjunction or male production. For example, although the percentage of males from animals with reduced brood sizes was higher than that in animals with higher brood sizes, the total number of males was not proportionally increased in animals with higher brood sizes. We were looking for feeding strains that consistently produced greatest number of males. To circumvent the problem of experimental inconsistencies within and between experiments, we compared median and average results to help identify consistent and reliable male-producing bacterial food.

Wormbase ID	Description	Gene Seq Name	Library Position	Sample number	
WBGene00020466	(28.5 kDa basic protein)	T12F5.1	I-1F23	1	
WBGene00022386	(protein coding)	Y95B8A.6 &	I-1I19	2	
WBGene00022390	him-19 (DDX6 DEAD/H-box RNA helicase family)	Y95B8A.11	I-1I19	2	plasmid overlap
WBGene00002225	klp-15 (kinesin-like protein)	M01E11.6	I-2F15	3	
WBGene00002226	klp-16 (kinesin-like protein)	C41G7.2	I-2F15	3	BLAST hit
WBGene00015691	viln-1 (VILliN related)	C10H11.1	I-2G15	4	
WBGene00019767	rpa-2 (replication protein A)	M04F3.1	I-2I11	5	
WBGene00002034	htp-3 (him-3 paralog)	F57C9.5	I-2K13	6	
WBGene00003878	opt-3/pept-3 (H+-coupled oligopeptide transporter)	F56F4.5	I-2L08	7	
WBGene00199766	(ncRNA)	F56F4.10	I-2L08	7	
WBGene00023072	(tRNA)	F56F4.T1	I-2L08	7	plasmid overlap
				8	
WBGene00001856	hil-5 (Histone H1-Like)	B0414.3	I-2N01	9	
WBGene00001853	hil-2 (Histone H1-Like)	Y73B6BL.9	I-2N01	9	BLAST hit
WBGene00003601	nhr-2 (Nuclear Hormone Receptor family)	C32F10.6	I-2N21	10	
WBGene00001860	him-1 (SMC1 homolog)	F28B3.7	I-2O07	11	
WBGene00018000	(Zn finger, LIM-domain protein)	F33D11.1	I-2P03	12	
WBGene00016389	(Zn finger, LIM-domain protein)	C34B2.4	I-2P03	12	BLAST hit
WBGene00017816	RNA binding protein (K homology domain, type 1 & 2)	F26B1.2	I-3C09	13	
WBGene00006377	syp-3 (SYnaPsis in meiosis abnormal)	F39H2.4	I-4C22	14	
WBGene00045237	mrt-1(MoRTal germline)	F39H2.5	I-4C22	14	plasmid overlap
WBGene00003183	mei-1 (defective MEiosis)	T01G9.5	I-4G05	15	
WBGene00000467	cep-1 (C. elegans P-53-like protein)	F52B5.5	I-4G17	16	
WBGene00004808	skr-2 (SKp1 Related (ubiquitin ligase complex component)	F46A9.4	I-4N09	17	
WBGene00004807	skr-1 (SKp1 Related (ubiquitin ligase complex component)	F46A9.5	I-4N09	17	BLAST hit
WBGene00002226	klp-16 (kinesin-like protein)	C41G7.2	I-4P13	18	
WBGene00002225	klp-15 (kinesin-like protein)	M01E11.6	I-4P13	18	BLAST hit
WBGene00008061	(RNA binding protein, K Homology domain, type 1)	C41G7.3	I-4P15	19	
WBGene00007710	rsa-1 (Regulator of Spindle Assembly)	C25A1.9	I-5I05	20	
WBGene00016389	LIM-type Zinc finger	C34B2.4	I-5K14	21	
WBGene00018000	LIM-type Zinc finger	F33D11.1	I-5K14	21	BLAST hit
WBGene00011549	(conserved protein, pseudogene?)	T06G6.8	I-6B03	22	
WBGene00006595	top-1 (TOPoisomerase)	M01E5.5	I-6B20	23	
WBGene00012209	hmg-20	W02D9.3	I-6I04	24	
WBGene00012370	(Williams-Beuren syndrome chromosoal region 16 protein, RCC1 domain (Regulator of chromosome condensation)	W09G3.7	I-7A13	25	
WBGene00012367	(dead gene)	W09G3.3	I-7A13	25	BLAST hit
WBGene00011612	(chk-1 paralog (Forkhead-associated, protein kinase domains))	T08D2.7 (in i	I-7A13	25	BLAST hit
WBGene00013510	(predicted pseudogene)	Y71A12B.13	I-7A13	25	BLAST hit
WBGene00007691	(carboxyesterase)	C23H4.2 (in	I-7A13	25	BLAST hit
WBGene00195377	ncRNA (in intron of pag-3)	F45B8.6	I-7A13	25	BLAST hit
WBGene00044134	(similar to Huntingtin-assoc protein 1, and a yeast nucleolar protein required for preRNA processing)	D1025.10 (in	I-7A13	25	BLAST hit
WBGene00009369	(conserved protein, regulator of microtubule dynamics)	F33H2.6	I-7M20	26	
WBGene00001049	dog-1(Deletions Of G-rich DNA)	F33H2.1	I-7M20	26	plasmid overlap
WBGene00009284	(nucleotidyl transferase)	F31C3.2	I-7O04	27	
WBGene00004775	sep-1 (SEParase)	Y47G6A.12	I-7P07	28	
WBGene00021846	ztf-23 (Zn finger)	Y54E10BR.8	I-7P07	28	
WBGene00022124	protein coding (in intron)	Y71F9AR.3	I-7P07	28	BLAST hit
WBGene00022042	icd-2 (Inhibitor of Cell Death)	Y65B4BR.5	I-8A07	29	

Wormbase ID	Description	Gene Seq Name	Library Position	Sample number	
WBGene00021406	frm-5.2 (FERM domain, protein 4.1, exrin, radixin, moesin, family)	Y38C1AB.4	II-1M02	30	
WBGene00001492	frm-5.1 (FERM domain, protein 4.1, exrin, radixin, moesin, family)	Y38C1AB.8	II-1M02	30	BLAST hit
WBGene00015915	(Oxoglutarate/iron-dependent dioxygenase, prolyl 4-hydroxylase, short-chain dehydrogease/reductase)	C17G10.1	II-4D09	31	
WBGene00195182	C17G10.10	C17G10.10	II-4D09	31	plasmid overlap
WBGene00002089	ins-6 (INSulin relates)	ZK84.6	II-4D10	32	
WBGene00200396	ncRNA	ZK84.14	II-4D10	32	plasmid overlap
WBGene00000252	bli-2	F59E12.12	II-4F24	33	
WBGene00206373	C27A2.12 histidine acid phosphatase)	C27A2.12	II-4O07	34	
WBGene00016152		C27A2.T1	II-4O07	34	
WBGene00196833	ncRNA	C27A2.10	II-4O07	34	plasmid overlap
WBGene00016277		C30G12.6	II-5F07	35	
WBGene00002216	klp-3 (kinesin-like protein)	T09A5.2	II-5H18	36	
WBGene00007434	arp-6 (Actin-Related Proteins)	C08B11.6	II-5N16	37	
WBGene00008431	piga-1 (Phosphatidylinositol-Glycan biosunthesis class A protein)	D2085.6	II-6G02	38	
WBGene00013143	Y53C12B.1 (G-protein beta WD-40 repeat, small-subunit processome, Utp13)	Y53C12B.1	II-6P02	39	
WBGene00007332	nucleolar protein?	C05C10.5	II-7E23	40	
WBGene00009946		F52H3.4	II-7I15	41	
WBGene00009364	WW domain binding protein 11, splicing	F33H1.3	II-7M15	42	
WBGene00008740		F13D12.5	II-8A01	43	
WBGene00009395	clec-64 (C-type LECtin)	F35C5.7	II-8B19	44	
WBGene00010267	lips-9 (LIPaSe related)	F58G1.5	II-8D15	45	
WBGene00006381	tac-1 (TACC, transforming acid coiled coil, protein family)	Y54E2A.3	II-9C20	46	
WBGene00000264	brc-1	C36A4.8	III-1N14	47	
WBGene00021936	ZincViron permease?	Y55F3BL.2 (ir	III-1N14	47	BLAST hit
WBGene00014810		M142.7	III-1N14	47	BLAST hit
WBGene00006735	ula-1	C26E6.8	III-2H07	48	
WBGene00004782	set- 2(SET, trithorax/polycomb domain containing)	C26E6.9	III-2H09	49	
WBGene00011415	him-18 (High Incidence of Males)	T04A8.15	III-2K10	50	
WBGene00001514	xnd-1 (X chromosome NonDisjunction factor)	C05D2.5	III-2P10	51	
WBGene00001869	him-10 (High Incidence of Males), kinetochore protein Nuf2 domain	R12B2.4	III-3E15	52	
WBGene00200903	ncRNA	K04C2.10	III-3J17	53/54	
WBGene00000265	brd-1 (BaRD homolog (tumor suppressor gene Bard1	K04C2.4	III-3J17	53/54	
WBGene00023099	tRNA	K04C2.t3	III-3J17	53/54	plasmid overlap
WBGene00015813	thoc-2 (THO Complex (transcription factor/nuclear export) subunit)	C16A3.8	III-3K12	55	
WBGene00022694	(conserved protein, SprT domain)	ZK328.4	III-3M11	56	
WBGene00018953	(conserved protein, Breast carcinoma amplified sequence 3 domain; nucleolar Ser-rich)	F56C9.10	III-3N08	57	
WBGene00000939	dcr-1	K12H4.8	III-4C08	58	
WBGene00018371	ess-2 (ES2, conserved nuclear protein)	F42H10.7	III-4F01	59	
WBGene00011142	(Bud13 domain)	R08D7.1	III-4F14	60	
WBGene00004342	him-15/rfs-1 (rad-51-like)	C30A5.2	III-4G22	61	
WBGene00016238	(conserved protein, Mob1/phocein domain)	C30A5.3	III-4G24	62	
WBGene00002219	klp-7 (kinesin-like protein)	K11D9.1	III-5B24	63	

Wormbase ID	Description	Gene Seq Name	Library Position	Sample number	
WBGene00001689	gpr-2 (G-protein regulator)	C38C10.4	III-5C03	64	
WBGene00001688	gpr-1 (G-protein regulator)	F22B7.13	III-5C03	64	BLAST hit
WBGene00003417	mrt-2 (MoRTal germ line)	Y41C4A.14	III-6G01	65	
WBGene00003804	npp-18 (Nuclear Pore complex Protein)	Y43F4B.4	III-6H19	66	
WBGene00000203	arx-5 (ARp2/3 compleX component)	Y37D8A.1	III-6M08	67	
WBGene00006804	unc-71 (ADAM disintegrin and metalloproteinase)	Y37D8A.13	III-6O08	68	
WBGene00019947	htz-1 (HisTone variant H2AZ homolog)	R08C7.3	IV-2C22	69	
WBGene00002032	htp-1 (Him-Three Paralog)	F41H10.10	IV-2H08	70	
WBGene00002033	htp-2(Him-Three Paralog)	Y73B6BL.2	IV-2H08	70	BLAST hit
WBGene00019441	(basic protein)	K06B9.4	IV-2K13	71	
WBGene00019439	(UDP-N-acetylglucosamine pyrophosphorylase)	K06B9.2 (in i	IV-2K13	71	BLAST hit
WBGene00015989	(protein coding)	C18H2.2	IV-2K13	71	BLAST hit
WBGene00017993	cec-5 (C. Elegans Chromodomain protein)	F32E10.6	IV-3H08	72	
WBGene00002221	klp-10 (kinesin-like protein)	C33H5.4	IV-3N16	73	
WBGene0000222	klp-18 (Kinesin-Like Protein)	C06G3.2	IV-3N16	73	BLAST hit
WBGene00002228	klp-18 (Kinesin-Like Protein)	C06G3.2	IV-3O14	74	
WBGene00002221	klp-10 (kinesin-like protein)	C33H5.4	IV-3O14	74	BLAST hit
WBGene00016381	sgo-1 (ShuGOshin, yeast chromosome segregation protein)	C33H5.15	IV-3P12	75	
WBGene00016178	mesp-1 (Meiotic Spindle)	C28C12.2	IV-4G08	76	
WBGene00004051	pme-3 (Poly(ADP-ribose) Metabolism Enzyme	F20C5.1	IV-4O12	77	
WBGene00012795	(BTB/POZ-like domain)	Y43E12A.3	IV-5H06	78	
WBGene00019254	(BTB/POZ-like domain)	H31G24.3	IV-5H06	78	BLAST hit
WBGene00001867	him-8 (C2H2 Zn fingers, zim-1/3& C02F5.12 homologs)	T07G12.12	IV-5H19	79	
WBGene00004297	rad-51 (RADiation sensitivity abnormal)	Y43C5A.6	IV-5O08	80	
WBGene00006406	srgp-1 (Slit-Robo GAP homolog)	F12F6.5	IV-6K23	81	
WBGene00008682	lex-1 (Lin-48 Expression abnormal, ATPase and bromodomains)	F11A10.1	IV-6M12	82	
WBGene00003150	mbk-2 (MiniBrain Kinase, DYRK dual specificity Yak1-related kinase)	F49E11.1	IV-6N06	83	
WBGene00004818	skr-12 (SKp1 Related, ubiquitin ligase complex component)	C52D10.6	IV-8K23	84	
WBGene00023259	C52D10.10 (pseudogene?)	C52D10.10	IV-8K23	84	plasmid overlap
WBGene00004819	skr-13 (SKp1-Related, ubiquitin ligase complex component)	C52D10.8	IV-8K23	84	BLAST hit
WBGene00004820	skr-14 (SKp-1-Related, ubiquitin ligase complex component)	Y47D7A.8	IV-8K23	84	BLAST hit
WBGene00004819	skr-13 (SKp1-Related, ubiquitin ligase complex component)	C52D10.8	IV-8M03	85&86	
WBGene00172905	21ur-14953 (21U-RNA)	C52D10.111	IV-8M03	85&86	plasmid overlap
WBGene00004818	skr-12 (SKp1 Related, ubiquitin ligase complex component)	C52D10.6	IV-8M03	85&86	BLAST hit
WBGene00004820	skr-14 (SKp-1-Related, ubiquitin ligase complex component)	Y47D7A.8	IV-8M03	85&86	BLAST hit
WBGene00001460	flp-17 (FMRFamide-related neuropeptide)	C52D10.11	IV-8M07	87	
WBGene00046720	piRNA 21ur-3217	C52D10.26	IV-8M07	87	plasmid overlap
WBGene00047812	piRNA 21ur-2713	C52D10.31	IV-8M07	87	plasmid overlap
WBGene00170510	piRNA 21ur-12283	C52D10.95	IV-8M07	87	plasmid overlap
WBGene00171842	piRNA 21ur-7573	C52D10.105	IV-8M07	87	plasmid overlap
WBGene00006375	syp-1 (SYnaPsis in meiosis abnormal, coiled-coil)	F26D2.2	V-10P20	88	
WBGene00010293	fbxa-108 (F-box A protein)	F59A1.7	V-11F18	89	
WBGene00012564	fbxa-107 (F-boxA protein)	Y37H2A.4	V-11F18	89	BLAST hit
WBGene00011612	(Forkhead-associated domain, kinase domain)	T08D2.7	V-13E12	90	
WBGene00000499	chk-2 (Cds1Chk2 checkpoint kinase)	Y60A3A.12	V-13E12	90	BLAST hit
WBGene00020016	lipl-3 (LIPase-like)	R11G11.14	V-1A16	91	
WBGene00022642	lipl-5 (LIPase-Like)	ZK6.7	V-1A16	91	BLAST hit

Wormbase ID	Description	Gene Seq Name	Library Position	Sample number	
WBGene00022642	lipl-5 (LIPase-Like)	ZK6.7	V-1I21	92	
WBGene00020016	lipl-3 (LIPase-like)	R11G11.14	V-1I21	92	BLAST hit
WBGene00004820	skr-14 (SKp-1-Related, ubiquitin ligase complex component)	Y47D7A.8	V-3P05	93	
WBGene00006376	syp-2 (SYnaPsis in meiosis abnormal)	C24G6.1	V-4C12	94	
WBGene00015463	hpo-1 (Hypersensitive to Pore-forming toxin, BTB/POZ-like domain)	C05C8.6	V-5O15	95	
WBGene00000226	atl-1 (ATM, ataxia telangectasia mutated-Like)	T06E4.3	V-6B02	96	
WBGene00017800	(protein coding)	F25G6.9	V-6I23	97	
WBGene00018285	smk-1 (SMEK (Dictyostelium Suppressor of MEK null) homolog)	F41E6.4	V-6K13	98	
WBGene00001830	hcp-2 (HoloCentric chromosome binding Protein)	T06E4.1	V-6P23	99	
WBGene00008921	SepB domain (conserved protein with WD repeat)	F17C11.10	V-7D23	100	
WBGene00001110	duo-1 (Deubiquitylating with USP/UBP and OTU domains)	F38B7.5	V-7J06	101	
WBGene00003405	mre-11 (yeast MRE recombination/repair homolog)	ZC302.1	V-7K02	102	
WBGene00008380	prp-38 (yeast PRP splicing factor-related)	D1054.14	V-7O04	103	
WBGene00001874	him-17 (High Incidence of Males, THAP domains)	T09E8.2	V-8B04	104	
WBGene00004296	rad-50 (RADiation sensitivity abnormal/yeast RAD-related)	T04H1.4	V-8O13	105	
WBGene00007154	dhc-3 (dynein heavy chain)	B0365.7	V-8P19	106	
WBGene00001864	him-5 (High Incidence of Males)	D1086.4	V-9C22	107	
WBGene00000592	coh-3 (COHesin family)	F08H9.1	V-9D01	108	
WBGene00021560	coh-4 (COHesin family)	Y45G5AM.8	V-9D01	108	BLAST hit
WBGene00007938	(conserved protein, cysteine-rich repeat)	C34F6.1	X-5K04	109	
WBGene00006831	unc-104 (klp-1, kinesin-like protein)	C52E12.2	II-5G20	110	
WBGene00003884	osm-3 (klp-2, kinesin-like protein)	M02B7.3	IV-1N20	111	
WBGene00002217	klp-4	F56E3.3	X-2K05	112	
WBGene00006874	vab-8 (klp-5, kinesin-like protein)	K12F2.2	V-8O05	113	
WBGene00002218	klp-6 (Kinesin-like protein)	R144.1	III-2J07	114	
WBGene00002220	klp-8 (Kinesin-Like Protein)	C15C7.2	X-2I17	115	
WBGene00201126	ncRNA	C15C7.12	X-2I17	115	plasmid overlap
WBGene00006974	zen-4 (klp-9, Kinesin-Like Protein)	M03D4.1	IV-3I07	116	
WBGene00002222	klp-11 (Kinesin-Like Protein)	F20C5.2	IV-4O14	117	
WBGene00002223	klp-12 (Kinesin-Like Protein)	T01G1.1	IV-6G03	118	
WBGene00002224	klp-13 (Kinesin-Like Protein)	F22F4.3	X-3C14	119	
WBGene00000257	bmk-1 (klp-14, Kinesin-Like Protein)	F23B12.8	V-9B23	120	
WBGene00002227	klp-17 (Kinesin-Like Protein)	W02B12.7	II-7F24	121	
WBGene00012203	rga-1 (Rho GTPase Activating protein)	W02B12.8 (ir	II-7F24	121	plasmid overlap
WBGene00002229	klp-19 (Kinesin-Like Protein)	Y43F4B.6	III-6H23	122	
WBGene00004985	spo-11	T05E11.4	IV-5L20	123	
WBGene00004333	rec-8	W02A2.6	IV-7G13	124	
WBGene00000591	coh-1 (COHesin family)	K08A8.3	X-3L08	125	
WBGene00001862	him-3 (High Incidence of Males, DNA binding HORMA)	ZK381.1	IV-3M14	126	
WBGene00003784	nos-2 (An finger, nanos-type)	ZK1127.1	II-5I04	127	
WBGene00001872	him-14 (MutS family of DNA mismatch repair proteins)	ZK1127.11	II-5I04	127	plasmid overlap
WBGene00001863	him-4 (hemicentin)	F15G9.4	X-4F24	128	
WBGene00001865	him-6 (RecQ-like DNAhelicase)	T04A11.6	IV-6L21	129	
WBGene00004737	scc-1 (yeast SCCmitotic condensin subunit homolog)/coh-2/rad-21.1	F10G7.4	II-3P06	130	
WBGene00001085	dpy-26 (acidic protein involved in dosage compensation and meiotic chromosome segregation)	C25G4.5	IV-6L01	131	

Wormbase ID	Description	Gene Seq Name	Library Position	Sample number	
WBGene00001872	him-14 (MutS family of DNA mismatch repair proteins)	ZK1127.11	II-5I24	132	
WBGene00003209	mel-26 (Maternal Effect Lethal, MATH and BTB/POZ domains)	ZK858.4	I-4D17	133	
WBGene00004187	prp-8 (similar to U5 snRNA-specific protein)	C50C3.6	III-4G10	134	
WBGene00006414	raga-1 (ras-related GTPase RagA)	T24F1.1	II-7B10	135	
WBGene00006818	unc-86 (POU homeodomain)	C30A5.7	III-4I08	136	
WBGene00015347	cids-1 (RNA Pol II CTD-interacting domain)	C02F5.4	III-4I18	137	
WBGene00018357	moc-3 (yeast urmylation and tRNA thiolation homology)	F42G8.6	IV-4I05	138	
WBGene00011597	zim-1 (C2H2 Zn fingers, paralog of him-8, zim-2/3 and C02F5.12)	T07G12.6	IV-5H07	139	
WBGene00011601	zim-3 (C2H2 Zn fingers, paralog of him-8, zim-1/2 and C02F5.12)	T07G12.11	IV-5H07	139	
WBGene00011600	zim-2 (C2H2 Zn fingers, paralog of him-8, zim-1/3 and C02F5.12)	T07G12.10	IV-5H07	139	BLAST hit
WBGene00011601	zim-3 (C2H2 Zn fingers, paralog of him-8, zim-1/2 and C02F5.12)	T07G12.11	IV-5H15	140	
WBGene00011600	zim-2 (C2H2 Zn fingers, paralog of him-8, zim-1/3 and C02F5.12)	T07G12.10	IV-5H15	140	BLAST hit
WBGene00011597	zim-1 (C2H2 Zn fingers, paralog of him-8, zim-2/3 and C02F5.12)	T07G12.6	IV-5H15	140	BLAST hit
WBGene00011601	zim-3 (C2H2 Zn fingers, paralog of him-8, zim-1/2 and C02F5.12)	T07G12.11	IV-5H17	141	
WBGene00011597	zim-1 (C2H2 Zn fingers, paralog of him-8, zim-2/3 and C02F5.12)	T07G12.6	IV-5H17	141	BLAST hit
WBGene00009451	cids-2 (RNA Pol II CTD interacting domain)	F36A2.1	I-4I12	142	
WBGene00000517	cki-2 (CDK inhibitor of CIP/KIP family Beckwith-Wiedemann syndrome))	T05A6.2	II-5H10	143	
WBGene00017641	csr-1 (Argonaute)	F20D12.1	IV-4E01	144	
WBGene00001332	eri-1 (conserved Rnase)	T07A9.5	IV-1I21	145	
WBGene00004684	rsd-6	F16D3.2	I-4I05	146	
WBGene00001402	fbf-2	F21H12.5	II-4F23	147	
WBGene00001401	fbf-1	H12I13.4	II-4F23	147	BLAST hit
WBGene00001688	gpr-1 (G Protein Regulator)	F22B7.13	not in library		
WBGene00006976	zhp-3 (Zip (yeast meiotic zipper) Homologous Protein)	K02B12.8	not in library		
WBGene00197950	ncRNA (in intron of coh-1)	K02B12.11	not in library		
WBGene00198376	ncRNA (in intron of coh-1)	K08A8.15	not in library		
WBGene00202160	ncRNA (in intron of coh-1)	K08A8.30	not in library		
WBGene00201525	ncRNA (in intron of coh-1)	K08A8.27	not in library		
WBGene00044276	(pseudogene) nspa-11 (Nematode Specific Peptide family, group A)	T06E4.13	not in library		
WBGene00001280	him-12 (emb-26)		not in library		
WBGene00001861	him-2		not in library		
WBGene00001866	him-7		not in library		
WBGene00001868	him-9		not in library		
WBGene00001870	him-11		not in library		
WBGene00001871	him-13		not in library		
WBGene00004043	plk-2(POLO kinase)	Y71F9B.7	not in library		
WBGene00004510	rrf-3 (RNA-dependent RNA Polymerase)	F10B5.7	not in library		
WBGene00004681	rsd-2	F52G2.2	not in library		
WBGene00001748	gsp-2 yeast Glc7-like phosphatase	F56C9.1	not in library		

Table 5. *C. elegans* genes associated with a Him phenotype. The selection of genes to target was made as described in Materials and Methods. Some dsRNA-expressing plasmids harbor sequences from adjacent genes and will produce hybrid dsRNAs that could simultaneously knock down multiple genes (the adjacent genes are marked as "plasmid overlap"). Other dsRNA-expressing plasmids harbor sequences that are members of conserved gene families (the similar sequences, with at least 80% identity over a 200bp region, are marked as "BLAST hit").

RESULTS AND DISCUSSION

RNAi assays to identify the most effective male foods

We first identified approximately 150 genes with roles in disjunction mechanisms in *C. elegans* (Table 5). This information is derived from data obtained from other labs that observed Him phenotypes in *C. elegans* mutants and also in worms undergoing RNAi. The Him phenotype observed in these experiments was an indicator of improper disjunction of chromosomes, specifically the X chromosome. We obtained plasmids capable of expressing dsRNA corresponding to each of the genes in bacteria for our RNAi experiments from Source Bioscience Lifesciences (Fraser, Kamath et al. 2000). Multiplex RNAi was possible because some plasmid clones had sequences that could target more than one gene simultaneously (for eg., genes within a homologous family). (Table 5). This led to an increase in the number of predicted gene targets to more than two hundred.

RNAi experiments were carried out as described in Methods. The relative effectiveness of each knockdown in eliciting a Him phenotype was measured by counting the number of male progeny from worms that ingested bacteria harboring dsRNA. The percentage of male progeny is not an ideal score for measuring non-disjunction because most of the gene knockdowns affect chromosomal segregation for autosomes, which is a lethal condition. For experiments that resulted in fewer progeny, we did not obtain larger numbers of males, though the proportion was larger in comparison to experiments that resulted in larger brood sizes. Therefore, we took the absolute number of males as well as the median as a score for non-disjunction, as this better reflects our goal, to obtain large numbers of males. To illustrate with an example, upon treating animals with dsRNA against

ula-1, fewer than ten progeny (including one male) was observed on each of the two experimental plates. Table S3 from (Timmons, Luna et al. 2014)). From our analysis, we found that the plasmids targeting *klp-16*, and *klp-15* ranked the highest in producing males.

Target gene	Off-target sequences	median # males (experiment)	average # males (experiment)
klp-16	*klp-15*	10 (n=4)	**19.8** (SD=27)
klp-15	*klp-16*	5 (n=5)	**15.4** (SD=22)
klp-7		0 (n=5)	**7.0** (SD=18)
zim-1	*zim-2, zim-3*	6.5 (n=2)	**6.5** (SD=3.5)
him-4		2 (n=6)	**5.2** (SD=6.6)
htp-1	*htp-2*	0 (n=4)	**4.8** (SD=9.5)
syp-2		2 (n=4)	**4.5** (SD=6.5)
syp-1		2 (n=3)	**4.0** (SD=5.3)
rad-50		1 (n=5)	**4.0** (SD=7.9)
rec-8		1 (n=6)	**3.8** (SD=7.0)
ZK328.4		0 (n=5)	**3.8** (SD=8.0)
him-17		1 (n=5)	**3.6** (SD=4.0)
hcp-2		2 (n=4)	**3.5** (SD=4.7)
him-19	*Y95B8A.6*	0.5 (n=4)	**3.3** (SD=5.9)
dhc-3		1 (n=5)	**3.2** (SD=3.5)
him-8		5 (n=5)	**3.0** (SD=9.6)
coh-1		0 (n=6)	**3.0** (SD=6.4)
klp-18	*klp-10*	1 (n=5)	**2.8** (SD=3.0)

Table 6: The most effective "male foods". The effectiveness of each dsRNA-expressing bacterial strain is expressed by means of the median and average number of males per experiment. As the bacterial feeding method is inherently inconsistent. Each experiment includes three to four replica plates. Assessments were made for individual plates as well. Strains that are bolded and with grey background are the RNAi reagents with best production of males (as indicated by the experimental median, the average per experiment, and average per plate). Strains that are shown with grey background indicate those with high rankings based on the experimental median and average per plate. Strains in bold ranked highly in average male production per experiment and per plate. (Original data sets can be found as Supplementary Table S3 in (Timmons, Luna et al. 2014).

Simultaneous knock-down of two genes - Supermale plasmids

We next hypothesized that the methodology might be further improved by knocking down two genes simultaneously. We engineered three plasmids (in the L4440 background (Timmons and Fire 1998)) with double T7 promoters flanking a pair-wise combination of two genes. We constructed the two-gene hybrid plasmid by selecting genes that were likely involved in different chromosome disjunction. The gene sequences for the plasmids were obtained from the Source BioScience LifeSciences *C. elegans* RNAi library (Fraser, Kamath et al. 2000). These plasmid constructs were named "Supermale foods" (Figure 11). Plasmids harbored fragments of exons 2-4 (1169bp) from *klp-16*, exons 24-28 (1180bp) of *dhc-3* and that of exons 3-6 (1019bp) of *him-8*. *klp-16* is important for chromosome segregation by forming proper meiotic spindles (Robin, DeBonis et al. 2005). HIM-8 binds to the pairing center of X chromosome and is shown to be required for meiotic segregation of X chromosome (Phillips, Wong et al. 2005). RNAi assays using supermale foods were performed by the protocol described in "RNAi by feeding" in the METHODS section. The assay recorded the number of male progeny as described previously.

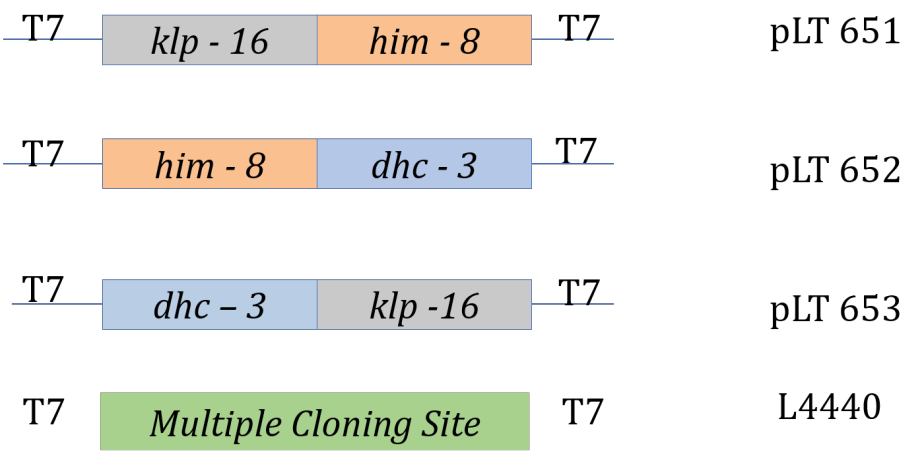


Figure 11: Production of a hybrid dsRNA molecule in order to simultaneously target two genes. RNAi feeding plasmids (constructed from plasmid L4440) containing two-hybrid gene sequence flanked by opposable bacteriophage T7 promoters (see Methods).

We found that all the three Supermale foods were consistently effective in eliciting male progeny at 25 °C (8-20%) (Figure 12). All males obtained from 'super male food' experiments were fertile.

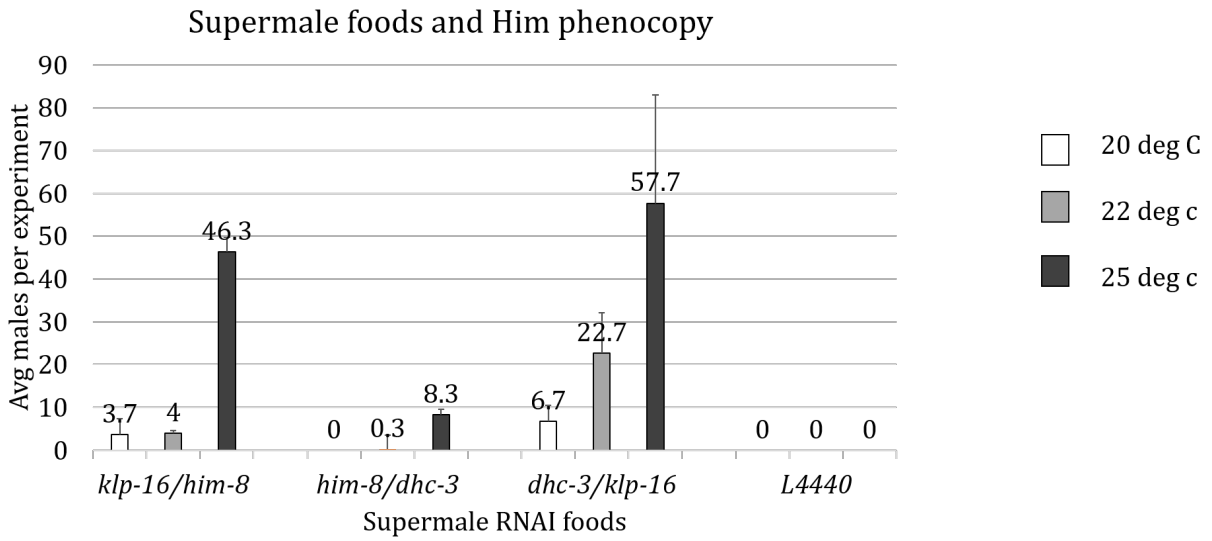


Figure 12: Effectiveness of Super Male food in generating male progeny. Y axis represents the average number of males per experiment. Total experiments in each case = 3. X axis represents three experiments using Supermale food, with L4440 as negative control (plasmid only). Experiments were conducted at three different temperatures in each case. (Original data sets can be found as Supplementary Table S3 in (Timmons, Luna et al.2014))

An application of RNAi induced non-disjunction

One successful application of RNAi-induced non-disjunction is the production of males in *mut-7* and *rde-2* mutant strains (Sundaram, Han et al. 2008). *rde-2* and *mut-7* animals are Him, but the mutant males are not fertile. On the other hand, *rde-2* and *mut-7* males obtained from male food can procreate. Having the same genotype with different male fertility phenotypes suggests a role for epigenetic mechanisms. The *mut-7* and *rde-2* mutants not only possess germline non-disjunction defects but also defects that contribute to infertility in adults. We investigated the nature of the fertility defects in *mut-7* and *rde-2*

mutant males. These males were able to produce cross progeny when mated with *unc* (paralyzed worms exhibiting Uncoordinated phenotype) hermaphrodites. This suggests the mutant males are incapable of mating with non *unc* hermaphrodite due to movement or coupling defects (Table 3). In such a case, non-disjunction induced by male foods offers with *mut-7* and *rde-2* males that are fertile and are capable of mating with hermaphrodites to give mutant progeny.

Strain	Source of Males	Hermaphrodite	Successful Matings
rde-2(pk1657)	mutation-induced males	wild type	0/6
rde-2(ne221)	mutation-induced males	wild type	0/3
rde-2(pk1657)	mutation-induced males	*unc-119(ed3)*	7/7
rde-2(ne221)	mutation-induced males	*unc-119(ed3)*	3/3
rde-2(pk1657)	RNAi-induced males	wild type	10/10
rde-2(ne221)	RNAi-induced males	wild type	9/10
mut-7(pk720)	mutation-induced males	wild type	0/4
mut-7(pk204)	mutation-induced males	wild type	0/2
mut-7(pk720)	mutation-induced males	*unc-119(ed3)*	0/5
mut-7(pk204)	mutation-induced males	*unc-119(ed3)*	2/2
mut-7(pk720)	RNAi-induced males	wild type	8/10
mut-7(pk204)	RNAi-induced males	wild type	9/10

Table 7. Infertility in *mut-7* and *rde-2* Him males is not observed in males obtained using "male food". *mut-7* and *rde-2* males were readily obtained due to the inherent Him phenotype of the mutants. *mut-7* and *rde-2* males were also obtained using *klp-16*"male food". (As *mut-7* and *rde-2* mutants are not completely RNAi defective, males can be isolated using bacterial feeding.) Each cross plate contained one hermaphrodite and seven males; male progeny, as well as non-Uncs (where appropriate), were indicative of a successful mating. These results, which reflect more extensive observations, implicate a link between mutant-induced defects that promote non-disjunction in early germline development and movement defects that affect mating ability later in adult males.

CONCLUSION

We identified genes that are important for proper chromosome disjunction in *Caenorhabditis elegans*. We constructed RNAi-based tools that were able to knock down these genes and showed that these RNAi reagents (Supermale foods) can be employed to produce large numbers of fertile males from different strains easily. We also demonstrated that these male foods can be used to produce fertile males from strains that had mutations which interfered with their ability to mate. Thus, having a number of effective "male foods" identified improves flexibility in experimental design.

REFERENCES

1. Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.
2. Fraser, A. G., R. S. Kamath, P. Zipperlen, M. Martinez-Campos, M. Sohrmann and J. Ahringer (2000). "Functional genomic analysis of C. elegans chromosome I by systematic RNA interference." Nature **408**(6810): 325-330.
3. Hodgkin, J., H. R. Horvitz and S. Brenner (1979). "Nondisjunction Mutants of the Nematode CAENORHABDITIS ELEGANS." Genetics **91**(1): 67-94.
4. Phillips, C. M., C. Wong, N. Bhalla, P. M. Carlton, P. Weiser, P. M. Meneely and A. F. Dernburg (2005). "HIM-8 binds to the X chromosome pairing center and mediates chromosome-specific meiotic synapsis." Cell **123**(6): 1051-1063.
5. Qu, W., C. Ren, Y. Li, J. Shi, J. Zhang, X. Wang, X. Hang, Y. Lu, D. Zhao and C. Zhang (2011). "Reliability analysis of the Ahringer Caenorhabditis elegans RNAi feeding library: a guide for genome-wide screens." BMC Genomics **12**: 170.
6. Robin, G., S. DeBonis, A. Dornier, G. Cappello, C. Ebel, R. H. Wade, D. Thierry-Mieg and F. Kozielski (2005). "Essential kinesins: characterization of Caenorhabditis elegans KLP-15." Biochemistry **44**(17): 6526-6536.
7. Sundaram, P., W. Han, N. Cohen, B. Echalier, J. Albin and L. Timmons (2008). "Caenorhabditis elegans ABCRNAi transporters interact genetically with rde-2 and mut-7." Genetics **178**(2): 801-814.
8. Timmons, L., D. L. Court and A. Fire (2001). "Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans." Gene **263**(1-2): 103-112.

9. Timmons, L. and A. Fire (1998). "Specific interference by ingested dsRNA." Nature **395**(6705): 854-854.
10. Timmons, L., H. Luna, J. Martinez, Z. Moore, V. Nagarajan, J. M. Kemege and N. Asad (2014). "Systematic comparison of bacterial feeding strains for increased yield of Caenorhabditis elegans males by RNA interference-induced non-disjunction." FEBS Lett **588**(18): 3347-3351.
11. Ward, S. and J. S. Carrel (1979). "Fertilization and sperm competition in the nematode Caenorhabditis elegans." Dev Biol **73**(2): 304-321.

CHAPTER 3

dsRNA-induced nuclear gene silencing and epigenetic transgene silencing in *C. elegans*

ABSTRACT

We previously performed an RNAi-based screen for Him phenotype in order to identify RNAi reagents (bacterial strains expressing dsRNA against gene targets) that are most effective in the efficient and reproducible production of males. Surprisingly, in the screening for genes associated with Him phenotype, was a sequence encoding a secreted neuropeptide. Since the role of such a protein in chromosome disjunction mechanisms is difficult to resolve, we further examined this locus and noticed the presence of non-coding RNAs included in the intron of the secreted neuropeptide gene. Our RNAi experiments led to the realization that nuclear RNAi silencing mechanisms are active at this locus, in a NRDE-3-dependent manner. In contrast to previous reports describing phenotypes that depend on nuclear RNAi mechanisms in an *eri*-1 mutant background, our results were obtained in a wild-type background. Thus, we have serendipitously identified a region in the genome that is naturally amenable to nuclear silencing mechanisms.

During the course of our investigation of the above locus, and its ability to be targeted by nuclear RNAi mechanisms, we observed a second, unusually robust silencing phenomenon associated with a transgene that was introduced into a wild type strain. It has long been observed that transgenes can express well in somatic tissues of *C. elegans*, yet often fail to express in the germ line. This provides evidence for a more robust set of anti-foreign genome responses in the germ line, in comparison to those in the somatic cells. In our experiments, we repeatedly observe silencing of single-copy sequences in somatic tissue, a phenomenon that is not typically observed. As we and others have successfully utilized the

same *mos* strain in the past for transgene expression, we hypothesize that the worm has "learned" that the *mos* element is foreign, and any DNA introduced in place of the *mos* element is now engulfed by epigenetic silencing machinery. We describe our efforts to understand the nature of these genomic loci that are unusually amenable to silencing in the nucleus.

Further investigation led to identifying a previously undescribed mutation in *eri-6/7* locus. We show here that the mutation *yy14* is associated with increase in trans-splicing of *eri-6/7*, which might play a role in silencing of transgenes in somatic cells.

INTRODUCTION

This project is an extension of the previous chapter that identified RNAi reagents (dsRNA-expressing bacterial clones) that increased the frequency of *C. elegans* males. As previously shown, the RNAi reagents effectively and reproducibly produced a Him (High Incidence of Males) phenotype. The function of most of these genes can be reconciled with an involvement in chromosome disjunction mechanisms, but surprisingly in the list was a gene, *flp-17* that encodes a FMRF amide-related protein. We further investigated this locus, because it was intriguing how a secreted neuropeptide might be required for meiotic chromosome segregation.

TRANS-SPLICING OF mRNAs IN *C. elegans*

One unique feature of *C. elegans* RNAs is that the pre mRNAs are trans-spliced with either of the two Splice Leader (SL1 or SL2) sequences. The 22 nt SL sequence, which is part of a longer (~100 nt) SL snRNP (small nuclear ribonuclear protein) forms the new 5' end of mRNA. About 70% of all genes in *C. elegans* are trans-spliced with SL sequences (Allen, Hillier, Waterston, & Blumenthal, 2011). A majority of these are trans-spliced with SL1 sequence. mRNAs that receive an SL2 trans-splicing leader sequence are mRNAs that typically reside downstream in a co-transcriptionally expressed gene operon. The polyadenylation signal AAUAAA of the upstream gene is critical for trans splicing by SL2 (Kuersten, Lea, MacMorris, Spieth, & Blumenthal, 1997). This form of trans-splicing was first observed in trypanosomes (Sutton & Boothroyd, 1986) and has been described for some chordates (Vandenberghe, Meedel, & Hastings, 2001).

Another form of trans-splicing, not unique to nematodes, involves intergenic splicing of exons. This type of splicing has been observed in various organisms, including Drosophila, mouse, and humans, and is relatively rare and typically observed in "complex" gene loci that include alternative cis-splicing and that harbor repetitive DNA sequences (Flouriot, Brand, Seraphin, & Gannon, 2002; McManus, Duff, Eipper-Mains, & Graveley, 2010; Zaphiropoulos, 2011). One example of such exon-exon trans-splicing is seen in *eri6/ eri-7* locus of *C. elegans*. The *eri-6* and *eri-7* genes are oppositely oriented, flank a bidirectional promoter, and are independently transcribed using different DNA strands as template. The individual pre mRNAs are joined together by a trans-splicing event. Direct repeats flanking the *eri-6* gene are thought to facilitate the trans-splicing process by forming complementary sequences between the mRNAs.

TRANSGENICS IN *C. elegans* AND RESPONSES TO FOREIGN DNA

Transgenic *C. elegans* are generated by microinjection of DNA of interest (plasmids or PCR products) along with a transformational marker into the distal arm of the gonad. Some commonly used co-transformational markers are pharyngeal GFP (*myo-2::GFP*), mcherry fluorescence (*myo-2::mcherry, myo-3::mcherry*), dominant roller allele (*rol-6*), *unc-119* rescue, *pha-1* rescue (Mello, Kramer, Stinchcomb, & Ambros, 1991) (Frokjaer-Jensen et al., 2008). Extrachromosomal arrays contain multiple copies (upto 300 copies) of different genes included in the injection. These arrays are mitotically unstable but are heritable through several generations. When animals inherit arrays, the genes present in them are usually overexpressed. Animals having the arrays can be mosaic. The arrays are silenced in

the germline as a part of anti-foreign genome response (Kelly, Xu, Montgomery, & Fire, 1997). To avoid this problem and have expression of transgene in the germline, formation of complex arrays is employed, where the number of repeats is reduced by adding fragmented genomic DNA along with the DNA of interest. But in this method, transgenes get silenced in these tissues after a few generations. To overcome this issue, extrachromosomal arrays can be integrated using gamma or UV irradiation or using single strand oligos along with transgenic DNA (Mello et al., 1991). But, in all these cases, the repetitive nature of array DNA is still retained to some extent. Therefore, this does not always allow germline expression. Also, the irradiation method introduces other mutations. Moreover, when integration of arrays happens, it becomes difficult to distinguish animals with extrachromosomal arrays from those that have the arrays integrated in the genome. Therefore, due to the disadvantages associated with these older methods of introducing transgenes, a next-generation technology of Mos1-mediated Single Copy Insertion (MosSCI) was developed. Mos1 is a transposon from *Drosophila*. A transgenic plasmid is constructed such that the gene of interest is flanked by homologous regions that will aid in recombination, following the excision of Mos1 (Frokjaer-Jensen et al., 2008). Many of our transgenes are inserted at the *ttTi5605* Mos1 site. In this method, transgenes are injected along with a plasmid that codes for transposase under a germline promoter (*Pglh-2::transposase*) and an array marker, (usually *Pmyo-3::mCherry*) . The *ttTi5605* locus allows germline expression from single copy transgenes.

There are several gene silencing phenomena associated with transgenes injected by the above methods and these are discussed below.

SILENCING IN *C. elegans*

C. elegans, like all other organisms, has evolved multiple strategies to protect against the invasion of foreign genomes in the germ line, such as viruses or transposons, and transgenes which are commonly used in genetic engineering experiments. Foreign genetic elements are introduced into *C. elegans* as transgenes that integrate in the genome or exist as independently heritable extra chromosomes. Transgenes can integrate as single copy elements or be tandemly repeated arrays. The repetitive arrays, but no single copy integrants are silenced in the germline. The transgenic arrays are silenced in somatic cells in mutants that have enhanced RNAi phenotype (*eri* mutants)(Simmer et al., 2002) and can be desilenced in RNAi defective mutants. While that is the case, other RNAi factors involved in exogenous RNAi pathway like RDE-4, DCR-1, DRH-1, DCR-1 are required for silencing of transgenes in *eri-6/7* mutants (Fischer, Butler, Pan, & Ruvkun, 2008). This shows transgene silencing and RNAi mechanisms have common effectors. The silencing of transgenes can utilize nuclear RNAi factors like NRDE-3 argonaute. In this case, the silencing occurs by the deposition of H3K9 methylation marks in the chromatin of somatic cells by SET-25 and LIN-61 proteins in an RNAi-compromised strain that lacks ERI-6/7 (Fischer et al., 2013). On the other hand, CSR-1 is an anti-silencing argonaute in *C. elegans*. CSR-1 directs the deposition of H3K4 trimethylation marks along with SET-2 protein. ERI-1 nuclease and ERI-6/7 are required for endogenous RNAi (thus, for expression of multicopy transgenes in somatic cells) but negatively regulate exogenous RNAi.

Cosuppression

Introduction of plasmids by injection into the *C. elegans* gonad leads to the formation of high copy number extrachromosomal Arrays. Genes that are normally expressed in the germline are frequently silenced when they reside in transgene arrays—a phenomenon that is immediately obvious when a ubiquitously-expressing promoter is used to drive GFP expression. In such a case, GFP expression will be observed in somatic cells, but not in germline tissue. Infrequently, the silencing phenomena will involve not only the transgene sequence, but also the endogenous, chromosomal genes that are homologous to the transgene DNA. This form of silencing, called co-suppression, has also been observed in plants (Jin & Guo, 2015), fungi and animals. In *C. elegans*, silencing from co-suppression has been described only for germline-expressed genes. However, one example of a co-suppression-like phenomenon, affecting the somatically expressed *unc-22* gene, was described (Fire, Albertson, Harrison, & Moerman, 1991). But, the *unc-22* co-suppression- like phenomenon was induced by the purposeful expression of anti-sense *unc-22* RNA strands, and the resulting reduction in UNC-22 protein was likely due to a late step in gene expression, as mRNA levels were not reduced. The more commonly observed co-suppression phenotypes associated with germline-expressed genes are distinct from classical RNAi mechanisms that are induced by experimental delivery of dsRNA, in that co-suppression does not require *rde-1* (encoding an Argonaute protein strictly required in RNAi experiments), yet co-suppression is dependent on *rde-2* and *mut-7. rde-2* and *mut-7* were shown to be essential for co-suppression of transgenes expressing *spo-11* and a truncated *him-14* (Dernburg, Zalevsky, Colaiacovo, & Villeneuve, 2000). *rde-2* and *mut-7* are part of a Mutator Complex of proteins, including MUT-2, MUT-8, MUT-9 and MUT-16, which are

required for RNAi and silencing of transposons as well as for co-suppression (Ketting & Plasterk, 2000). This indicates that RNA acts as the effector and target for silencing in this process.

RNAe or paramutation

Heritable epigenetic silencing refers to the stable silencing over several generations. Typically, heritable silencing involves small RNA silencing pathways elements. RNA-induced epigenetic silencing pathways have also been observed to permanently silence single-copy transgenes (Shirayama et al., 2012). This pathway has been termed RNAe, which resembles the paramutation phenomenon observed in plants. RNAe is associated with transcriptional silencing and chromatin modifications linked to gene repression. RNAe is shown to act both transcriptionally and post transcriptionally (Luteijn et al., 2012). In an experimental system that utilized a GFP reporter engineered with a target for piRNA binding, piRNAs, in association with the Argonaute protein PRG-1, is the initiator of RNAe (Ashe et al., 2012), but it is dispensable for maintenance of the silenced state (Shirayama et al., 2012). Maintenance of RNAe requires MUT-7 and the argonaute, WAGO-9. In an experiment to screen for factors essential for RNAe, it was shown that GFP transgene which was silenced by RNAe resurrected on the loss of *nrde-1 & 2* suggesting RNAe could act at the nuclear level (Luteijn et al., 2012). Strengthening this observation, the presence of H3K9 trimethylation marks in the transgenes was demonstrated, using ChIP-qPCR experiments (Mao et al., 2015). Silencing of transgenes by RNAe is gender-independently heritable. RNAe silencing can act in-trans where an epigenetically silenced allele causes the trans inactivation of a previously expressed allele, in which case, it is called paramutation (Hollick, 2017).

Transcriptional Gene Silencing

Silencing of repetitive elements and transposons in *C. elegans* occurs typically in its germline by RNAi dependent processes. Silencing of transgenic arrays also occurs in the germline and this silenced state is inherited by means of epigenetic marks, while the somatic cells retain the expression of transgenic arrays (Kelly et al., 1997). *mes-2, mes-6* and *hp1-2* (HP1 homolog in *C. elegans*) have been implicated in transgene silencing in the germline (Kelly & Fire, 1998) (Couteau, Guerry, Muller, & Palladino, 2002). Also, the presence of H3K9 methylation signs the occurrence of this silencing at the level of transcription. Apart from silencing by cosuppression, RNAi-induced Transcriptional Gene Silencing (TGS) also has been described in which a somatic transgene was silenced by dsRNA homologous to regions of the transgene and its repetitive arrays (Grishok, Sinskey, & Sharp, 2005). Since the silencing is associated with a decrease in RNA Pol II and acetylation marks, it is apt to refer to this as TGS*. rde-1, rde-4, rrf-1, hp1-3* but not *rrf-3* are required for this silencing (Grishok et al., 2005). This study also shows that inhibiting the activity of histone deacetylases prevents the silencing of somatic transgenes. Transgene silencing is enhanced in mutants that are hypersensitive to exogenous dsRNA dependent silencing (Simmer et al., 2002).

EXOGENOUS RNAi

On ingesting dsRNA containing HT115 (DE3) strain of *E. coli* cells, dsRNA is disseminated from intestinal cells to other tissues through dsRNA gated SID-1(Systemic RNAi Defective) channel protein. In the classical RNAi pathway, which is also called the post-transcriptional gene silencing (PTGS) pathway, the dsRNA is diced into primary siRNAs by the activity of Dicer complex. Dicer complex consists of proteins 1) Dicer – dsRNA specific RNAse III endoribonuclease - cleaves dsRNA into siRNAs, 2) RDE-4 – dsRNA binding protein, 3) RDE-1 – endoribonuclease activity, RNA helicase binding activity, PAZ+PIWI argonaute protein that interacts with primary siRNAs, converts ds siRNAs into ss siRNAs, part of RISC complex, 4) DRH-1 – DEAD/H box helicase – ATP binding activity (Grishok, 2005). In the second step of RNAi, roles of RNA dependent RNA polymerases (RdRP) are important. *rrf-1* is required for the production of secondary siRNAs in somatic cells while *ego-1* acts in siRNA amplification in the germline (Smardon et al., 2000). These RdRPs work in a primer-independent manner with the target mRNA as the template.

ENDOGENOUS RNAi

lin-4 miRNA is the first endogenous small RNA identified in *C. elegans.* Endogenous RNAi pathways function in silencing of transposons, anti-viral mechanisms, silencing of repetitive elements and maintaining heterochromatic regions (Sijen & Plasterk, 2003) (Grishok et al., 2005). Endogenous small RNAs include 5' guanosine 26nt RNAs (26G RNAs) and the downstream 22G RNAs. Enhanced RNAi is observed for exogenous dsRNA targets in *eri-1* and *rrf-3* mutants. But loss of endogenous siRNAs is seen in these two mutant animals

suggesting, there are common RNAi factors that act in competing exogenous and endogenous silencing pathways. ERI-1, RRF-3, RRF-1 and RDE-3 are factors involved in endogenous silencing as described by decreased abundance of siRNAs in mutants that are defective for these proteins. Additionally, an increase in the level of target transcripts was also observed in these mutants (Lee, Hammell, & Ambros, 2006). These target transcripts not only included mobile genetic elements and repetitive arrays, but also endogenous genes that have normal physiological functions. This strongly suggests that the endogenous silencing pathways play major roles in regulation of gene expression. Endogenous pathways could be involved in transcriptional gene silencing via chromatin modifications (Ambros, Lee, Lavanway, Williams, & Jewell, 2003; Grishok et al., 2005; Robert, Sijen, van Wolfswinkel, & Plasterk, 2005) or post transcriptional gene silencing.

It is possible that endogenous silencing happens through multiple pathways, as evident from non-coinciding targets in mutants of the above-mentioned factors. There is a convergence between diverse pathways, and this is consistent, for example, with the observation that *eri-1* mutants are hypersensitive to exogenous RNAi (RNAi from dsRNA by feeding or injecting) compared to wildtype animals. Besides, there is an increase in the level of transcription of endogenous genes in *eri-1* mutants. This contradictory function can be explained by the involvement of silencing factors in multiple pathways. In this case, *eri-1* mutants lack ERI-1 protein which 1) leads to loss of endogenous silencing of some genes, 2) releases limiting factors (for eg., DCR-1) that could participate in complementary silencing pathways viz exogenous RNAi, leading to Eri phenotype.

ARGONAUTES

Small RNA dependent silencing in animals require guiding proteins that form complex with other proteins in the silencing machinery. Argonaute proteins have gene silencing functions in plants, nematodes, flies and humans. There are 27 argonaute proteins in *C. elegans.* These proteins accommodate one of the three types of small RNAs, short interspersed RNAs (siRNAs), microRNAs (miRNAs) and PIWI-associated RNAs (piRNAs). Perfect homology between small RNAs and the complementary target sequence promotes siRNA mediated endonucleolytic cleavage, whereas imperfect homology with 3' UTR of target genes leads to miRNA dependent translational repression (Carthew & Sontheimer, 2009; Filipowicz, Bhattacharyya, & Sonenberg, 2008; Pillai, Bhattacharyya, & Filipowicz, 2007) and affect mRNA stability ((Bagga et al., 2005) by several mechanisms.

Argonaute proteins have two conserved domains – PAZ (Piwi/Argonaute/Zwille) domain required for siRNA interaction has a binding pocket that can hold 2 nt 3' overhang which is a product of Dicer activity. The second domain called the PIWI domain is important for endonucleolytic cleavage of mRNA referred to as the slicer activity.

MUT-7/RDE-2

C. elegans relies on endogenous RNAi pathways to silence transposons and foreign nucleic acids (Zhang et al., 2011). Activation of transposons is prevented by MUT (mutator class) proteins. Six mutator genes – *mut-2/rde-3, mut-7, mut-8/rde-2, mut-14, mut-15, mut-16* are identified (Ketting, Haverkamp, van Luenen, & Plasterk, 1999). Like the *eri* class mutants, mutations in *mut* genes also exhibit temperature sensitive sterility and increased X

chromosome nondisjunction. MUT-7(a 3'-5' exonuclease) is required for 22G RNA dependent silencing in the germline (Gu et al., 2009). *mut-14* encodes a helicase protein. MUT-7 is found in both the cytoplasm as well as the nucleus. MUT-7 interacts with RDE-2 and act as a complex in RNAi downstream of RDE-1 and RDE-4 (RDE-1 and RDE-4 are required for RNAi initiation), particularly in the siRNA amplification step. MUT-7 and RDE-2 are required for inheritance of RNAi while MUT-7 does not interact with RDE-2 in the nucleus. *mut-7 and rde-2* animals are Him (Tops et al., 2005).

MATERIALS

Primers

1418 TATATATATAAATTAATACGACTCACTATAGGGTTCACTGTTTCAAATGTTTTG

1419 TATATATATAAATTAATACGACTCACTATAGGGCTGAACCTTTTTTCATTTTCTAT

1420 TATATATATAAATTAATACGACTCACTATAGGGCATGGTTGTTTTCACGGACT

1421 TATATATATAAATTAATACGACTCACTATAGGGTTTACCGTTTCAAGCCTTTCATA

1422 TATATATATAAATTAATACGACTCACTATAGGGTTTTCCGTTTTACGAAAACGAGA

1423 TATATATATAAATTAATACGACTCACTATAGGGTAAAATTGTAAATCATGAGTGATAG

Plasmids

pLT 663 → pCR2.1 (TA vector) + PCR fragment of *flp-17* cDNA – primers 1424/1425

pLT 664 → pCR2.1 (TA vector) + PCR fragment of *flp-17* genomic DNA template using primers 1422/1423

pLT 665 → pCR2.1 (TA vector) + PCR fragment of *flp-17* genomic DNA template using primers 1420/1421

pLT 666 → pCR2.1 (TA vector) + PCR fragment of *flp-17* genomic DNA template using primers 1418/1419

pLT 667 → pCR2.1 (TA vector) + PCR fragment of *flp-17* genomic DNA template using primers 1420/1423

pLT 668 → pCR2.1 (TA vector) + PCR fragment of *flp-17* genomic DNA template using primers 1421/1418

pLT 669 → pCR2.1 (TA vector) + PCR fragment of *flp-17* genomic DNA template using primers 1418/1423

pLT 704 → pCFJ151 + pLT703 – Bgl2/Afl2 → *myo3 :: GFP* with *flp-17* introns (all 3 piRNAs) *:: let-858* 3'UTR with *unc-119* rescue sequence flanked by ttTi5605 Mos left and right homologous regions.

pLT 706 → pCFJ151 + pLT296 – Bgl2/Afl2 → *myo3 :: GFP* with introns, No NLS *:: let-858* 3'UTR with *unc-119* rescue sequence flanked by ttTi5605 Mos left and right homologous regions. (Control for RNAi)

Transgenic Strains

XX1869 - Single Copy Integrants of pLT704 [*unc-119*+ transgene in *unc-119* mutant] [*myo3 :: GFP* with *flp-17* introns (all 3 piRNAs) *:: let-858* 3'UTR in ttTi5605 MosSCI insertion vector injected into XX1327]

XX1870 - Single Copy Integrants of pLT706 [*unc-119*+ transgene in *unc-119* mutant] [*myo3 :: GFP* with introns *:: let-858* 3'UTR in ttTi5605 MosSCI insertion vector injected into XX1327]

METHODS

Construction of RNAi vectors

To construct RNAi vectors, cloning of PCR products into the kanamycin-resistant vector pCR 2.1 was adopted. Feeding plasmid IV-8M07 (plasmid # 87) was used as the template in PCR amplification of regions that correspond to RNAi targets with primers 1418-1425. The PCR reactions were carried out using Standard Taq Polymerase enzyme from NEB following its cycle recommendations. The PCR fragments were stored at -20 deg C and were cloned with pCR 2.1 vector using NEB Ligase enzyme at 15 deg C overnight. Calcium competent Top10 *E. coli* cells were transformed with the ligated plasmids and transformants were screened based on Kanamycin resistance.

	Fragment 1 pLT 666	Fragment 2 pLT 665	Fragment 3 pLT 664	Fragment 4 pLT 663	Fragment 5 pLT 668	Fragment 6 pLT 667	Fragment 7 pLT 669
Primers	1418/1419	1420/1421	1422/1423	1424/1425	1418/1421	1420/1423	1418/1423
Product Size (nt)	127	121	127	393 (cDNA)	503	310	692
AnnealingT (C)	55	55	55	59.8	55	55	55

Table 8 : Construction of feeding plasmids. pLT666 to pLT669 are feeding plasmids made from cloning PCR fragments (from *flp-17* locus) to TA vector pCR2.1. Rows show the primer pairs used to amplify the region in between, product size and annealing temperature for PCR reactions.

Culturing of *C. elegans*

All *C. elegans* strains were maintained on 6 cm NGM plates with *E. coli*(OP50).

RNAi by feeding

RNAi feeding plates were prepared with NGM and TAI (Tetracycline Ampicillin IPTG) as described (Hull & Timmons, 2004; Timmons, Court, & Fire, 2001). These plates were seeded with HT115 (DE3) bacteria harboring each of plasmids pLT663 – pLT669 engineered to express dsRNA against specific regions of *flp-17* locus. Feeding plasmids were derived from L4440 vector (Timmons & Fire, 1998). Four L1 stage hermaphrodite worms were placed onto NGM plates and were incubated at 22.5 deg C and 25 deg C for 72 – 96 hours. These worms were allowed to lay eggs and progeny were scored for Him phenotype.

Construction of transgenic plasmids

Plasmids pLT704 and pLT706 were cloned by standard restriction digestion/ligation protocol. Transgenic vectors were verified by PCR and sequencing.

Microinjection

Transgene lines were established as single copy integrants by MosSCI method of injection as described (Frokjaer-Jensen et al., 2008). *unc-119* mutant *C. elegans* were injected with a mix of transgenic plasmid, plasmid containing *mos transposase* coding sequence and a selectable marker of transformation. Selection of single copy transformants was based on the absence of array markers – *mCherry* and *myo-2 :: GFP* (RFP and pharyngeal GFP signals) and the presence of muscular GFP in animals rescued of Unc phenotype. Transformants were verified for a single copy of insertion of the transgene by PCR amplification using a primer

complementary to the site of integration in the genome and a primer that is specific to the transgene.

qRT-PCR

Comparative quantitative RT-PCR was used to estimate the relative abundance of *eri-6/7* trans-spliced mRNA and *eIF3.c* mRNA in wildtype and *yy14* strain. Total cDNA was prepared by oligo(dT)-primed reverse transcription of total RNA from strains XX1683 and 32.1 (wildtype for *eri-6/7*), XX1999 , XX1327 (has *yy14*). PCR was then carried out with gene-specific primer pairs. The PCR products were resolved in a 1% agarose gel stained with ethidium bromide and imaged under UV.

RESULTS AND DISCUSSION

flp-17 (RNAi) affects X chromosome segregation in *C. elegans*

From our efforts to build better tools for the production of *C. elegans* males, we observed Him phenotype when the *flp-17* gene was knocked down using RNAi. We are not the first lab to observe this phenotype for *flp-17,* as our "male food" optimizations made use of existing information and data related to genes required for chromosome disjunction (Chapter2) (Fraser et al., 2000). RNAi against *flp-17* reproducibly resulted in increased number of male progeny (Figure 13).

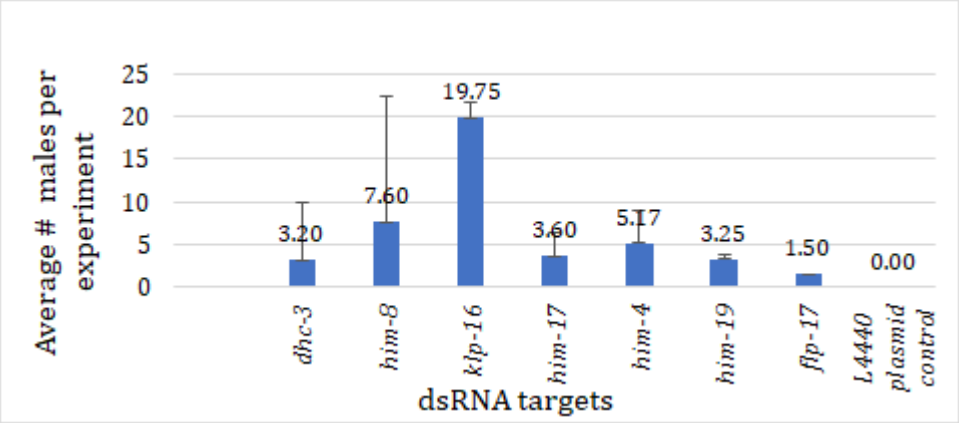


Figure 13: Him phenotypes in response to introduction of by feeding. (*y-axis* : Average number of males per experiment). *flp-17* is seen to exhibit Him phenotype. L4440 (empty RNAi vector) is negative control.

Based the fact that the *flp-17* gene encodes a FMRFamide-related neuropeptide, it is difficult to reconcile a role for *flp*-17 in chromosome disjunction. Neuropeptides are short peptides that function in signaling synaptic activity. These short sequences of amino acids are found in most organisms. *C. elegans* harbors over one hundred neuropeptide genes

which are categorized based on their composition into two families: insulin-related neuropeptides and FMRFamide – related peptides. FMRFamide (Phe-Met-Arg-Phe-NH2)–related peptides are a family of secreted proteins that contain FMRF sequences at their C termini (Li, Kim, & Nelson, 1999). Twenty-two *flp* genes endoce FMRFamide like proteins in *C. elegans*, with clusters of *flp* sequences found on Chromosomes IV and V. *flp-17* is located on chromosome IV. These *flp* genes function in motor and sensory pathways, in egg laying functions, in modulating pharyngeal muscle activity and in fat deposition (Cohen et al., 2009; Nelson, Rosoff, & Li, 1998; Ringstad & Horvitz, 2008; Stawicki, Takayanagi-Kiya, Zhou, & Jin, 2013). Chromosome disjunction is not an activity that has been associated with FLP protein function.

Presence of piRNAs in *flp-17* locus

On taking a closer look at the *flp-17* gene locus, we discovered that three piRNAs are found in the intronic region: 21ur 7573 (piRNA 1), 21ur 2713& 21ur 12283 (these two piRNAs have overlapping sequences that differ only by 3 bases and is considered as piRNA 2) and 21ur 3217 (piRNA3) lie between exons 1 and 2 of *flp-17* (Figure 14).

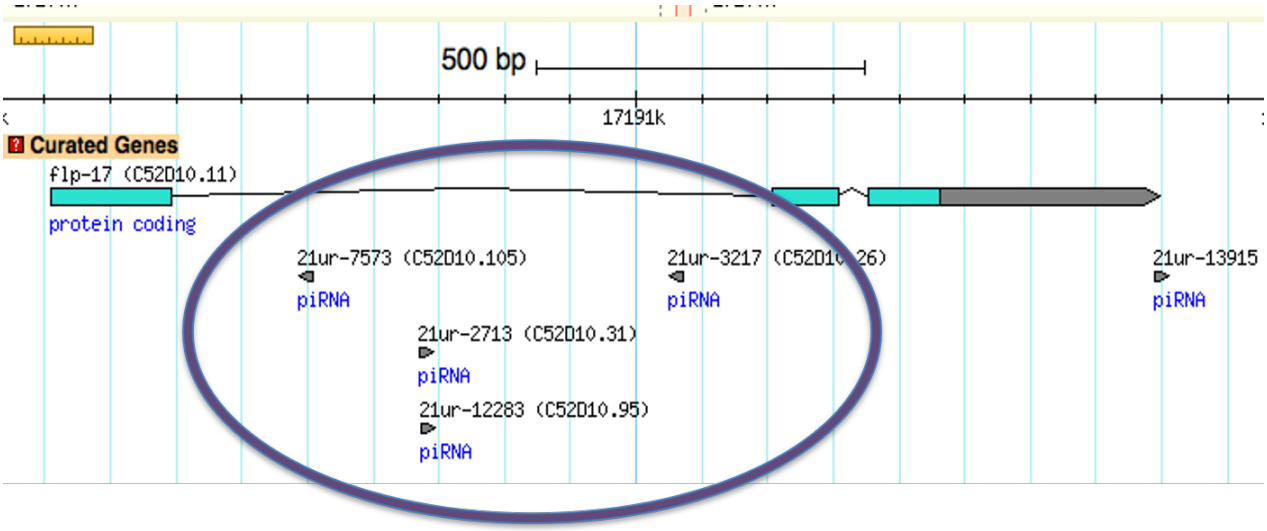


Figure 14: *flp-17* gene locus adapted from 'Wormbase'. piRNAs are in the intron region between exons 1 & 2 of the gene.

piRNAs are a group of small, non-coding RNAs transcribed from approximately 15,000 genes in *C. elegans* that are typically located in intronic regions or intergenic regions (Weick & Miska, 2014). piRNAs are also referred to as 21U-RNAs due to their unique 21 nt length and a U (uridine) at the 5' end. 21U-RNAs are expressed in the germ line where they are found in association with PRG-1 and PRG-2--piwi-related argonaute proteins. The Argonaute-piRNA complex recognizes the target by incomplete base pairing and functions in transposon silencing and regulation of mRNA. piRNAs in *C. elegans,* also act as mediators for initiation of H3K9 methylation and their functions can lead to stalling of RNA Polymerase II for genes transcribed in the germ line (Weick & Miska, 2014) (Guang et al., 2010) . piRNA/Argonaute functions cause the resulting gametes to stably inherit signals leading to transgenerational silencing in response to experimentally delivered dsRNAs—a phenomena that is also dependent on the Argonaute proteins NRDE or HRDE that shuttle silencing RNAs

into the nucleus (Weick & Miska, 2014). Research by J. J. Wang et al has shown that 3% of protein coding genes are upregulated in PRG-1 mutants, highlighting roles for this pathway in endogenous gene regulation (J. J. Wang et al., 2014). piRNAs in *C. elegans* play a protective function of safeguarding the genome against mobile sequences such as transposons. piRNAs are expressed in the germline tissues to protect and transmit genome to successive generations without defects caused by non-self-components.

piRNAs do not have perfect homology with their target RNAs, lending to the difficulty in deciphering specific targets, and thus functions, for any single piRNA. In *C. elegans*, the majority of piRNA genes are found in clusters on chromosome IV (Ruby et al., 2006). Forkhead transcription factors are believed to transcribe piRNA genes, acting on a conserved motif of 8 nucleotides -CTGTTTCA-. This recognition sequence, termed a Ruby motif, is 40 bp upstream of the piRNA genes (Ruby et al., 2006). Thus, the biogenesis of piRNAs in *C. elegans* is quite different from that in *Drosophila*. In *Drosophila*, piRNAs are generated using a ping-pong mechanism that relies on expression of long pre-piRNA precursors from specific sites in the genome that are configured as clusters of transposon insertions. The pre-piRNAs are processed in an mRNA target-dependent fashion by the slicing action of Argonaute proteins. By contrast, the thousands of piRNAs in *C. elegans* are believed to rise as independently transcribed genes; although the precise nature of their processing is yet to be elucidated.

In keeping with the more well established role in protection of the germ line from transposon mobilization defined in other organisms, at least one Piwi/piRNA in *C. elegans* is required to suppress the mobility or excision of Tc3 transposon (Das et al., 2008). Loss of PRG-1 is shown to be associated with decrease in the accumulation of piRNAs (Batista et al.,

2008), and is associated with germline defects and sterility (G. Wang & Reinke, 2008). The role of piRNAs in regulation of endogenous genes is still unknown.

The piRNA genes located in the *flp-17* intron have an intact forkhead promoter sequence ~40bp upstream. Forkhead proteins regulate the transcription of piRNA genes. *flp-17* gene is located on Chr IV and the majority of piRNA genes in *C. elegans* are found in clusters on the same chromosome.

Feeding induced Him phenotype in non-overlapping regions in *flp-17* locus

The *flp-17* plasmid used as dsRNA template in the previous RNAi experiments was obtained from Source Biosciences (Fraser et al., 2000). The sequence in the plasmid included genomic DNA corresponding to exons 1 and 2, including the piRNAs-containing intron, of *flp-17*. To distinguish whether the *flp-17* gene or the piRNA sequences contributed to the RNAi phenocopy, we divided the dsRNA template into smaller regions for the purposes of more specific gene targeting in our RNAi experiments. We achieved this by constructing plasmids, as indicated in Figure 15, and introducing them into the appropriate bacterial host for the purpose of expressing dsRNA.

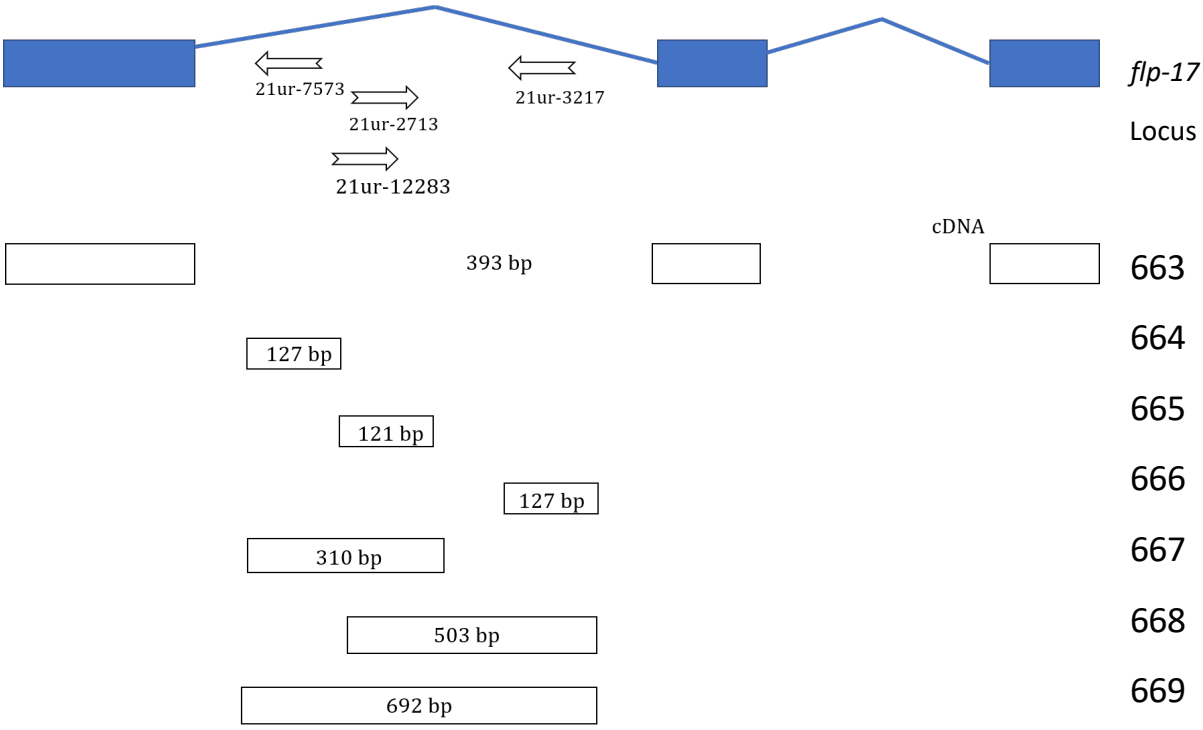


Figure 15 : Regions of *flp-17* targeted by RNAi. Top panel – blue rectangles represent exons and blue thin lines represent introns. Shown between first two exons are piRNAs with piRNA1 (21ur-7573) and piRNA3 (21ur-3217) transcribed opposite in direction to piRNA2 (2iur-2713 & 21ur-12283). The bottom panel shows regions that were targeted for RNAi by dsRNA expressed from feeding plasmids pLT663-pLT669. Bottom panel - regions that were amplified from genomic DNA and subsequently cloned to result in a set of feeding plasmids that targeted smaller regions in the locus.

Surprisingly, a similar phenotype was observed when two non-overlapping regions were targeted for silencing by RNAi. RNAi against cDNA region (includes only exon regions) as well as that against intronic region that included only the piRNAs, showed evidences of non-disjunction phenotype (scored as Him) (Table 9). It is known that feeding *C. elegans* with

bacteria that produce dsRNA triggers a homology dependent degradation of mRNA in the cytoplasm with the help of siRNAs and RISC. In our case, the silencing was scored by chromosomal non-disjunction phenotype (seen as presence of males in progeny).

Feeding plasmid	**Region targeted for non-disjunction by RNAi**	**Strength of phenotype(# of males / expt) Wildtype**	
		25 deg C	**22.5 deg C**
pLT663	cDNA	6/11	3/8
pLT669	All 3 piRNAs	6/11	3/8
pLT667	piRNAs 1 & 2	4/11	10/8
pLT668	piRNAs 2 & 3	0/11	0/8
pLT664	piRNA1	0/11	0/8
pLT665	piRNA2	0/11	0/8
pLT666	piRNA3	0/11	0/8

Table 9 : RNAi knockdown of specific regions in *flp-17* locus in wildtype *C. elegans*. RNAi was performed with plasmids pLT663 through pLT669 at two different temperatures. The numerator in the fraction denotes the number of males while the denominator represents the total number of experiments. Each experiment had 4 L1s added at the beginning of the experiment. piRNAs 2 and 3 did not show Him phenotype when targeted individually or in combination and this is attributed to the relatively small size of the target region.

Region targeted for RNAi	Non-disjunction ? (Males) Wildtype
Genomic DNA	Yes
cDNA	**Yes**
All 3 piRNAs	**Yes**
piRNA 1 & piRNA 2	Yes
piRNA 2 & piRNA 3	No
piRNAs 1 or 2 or 3	No

Table 10 : Summary of results from RNAi assay for non-disjunction. Two non-overlapping regions in *flp-17* – (shown in bold) exhibit the same phenotype.

The fact that non-overlapping dsRNA triggers can elicit the same phenotype led us to hypothesize that 1) Both, *flp-17* gene and one or more piRNAs independently are involved in chromosomal disjunction during meiosis, or 2) Perhaps dsRNA directed towards this locus silences multiple genes simultaneously. Since the two regions do not share sequence identity, the best model for the latter hypothesis, would involve a mechanism in which the dsRNA acts in the nucleus to trigger heterochromatin formation. The tendency of heterochromatin to spread would lead to silencing of adjacent regions simultaneously. In order to test whether feeding-induced silencing occurs in the nucleus, we performed the same RNAi experiments in a *nrde-3* background at two different temperatures.

NRDE-3 is an argonaute protein identified from screening of genes required for transcriptional gene silencing (TGS). Unlike PTGS, this form of silencing happens in the nucleus and requires factors that are collectively called nuclear RNAi deficient (NRDE)

proteins (Guang et al., 2010). NRDE-3 localizes to the nucleus in an siRNA-dependent fashion (Zhuang, Banse, & Hunter, 2013). The prevailing model for NRDE-3 function in RNAi is that NRDE-3 shuttles silencing RNAs into the nucleus where the RNAs direct sequence-specific activities that affect gene expression. siRNA/NRDE-3 complexes with NRDE-1, 2 and 4 in the nucleus and facilitates silencing by identifying homologous nascent mRNA targets. NRDE-3 protein lacks the domain associated with slicer activity (DDH amino acids), suggesting NRDE-3 dependent nuclear silencing circumvents the slicing of pre mRNAs. Pausing of RNA polymerase II and histone modifications are some activities associated with NRDE-3-dependent silencing (Mao et al., 2015) (Shiu & Hunter, 2017). Animals that are mutant for *nrde-3* are defective for nuclear RNAi (Guang et al., 2008). When we targeted the *flp-17* locus for silencing in a *nrde-3* background, we did not see any males in the progeny (Figures 16 & 17). The fact that our RNAi experiments elicit silencing when exon or intron sequences are used as trigger, and that the silencing is *nrde-3* dependent helps confirm our hypothesis that the *flp-17* region is particularly amenable to silencing in the nucleus.

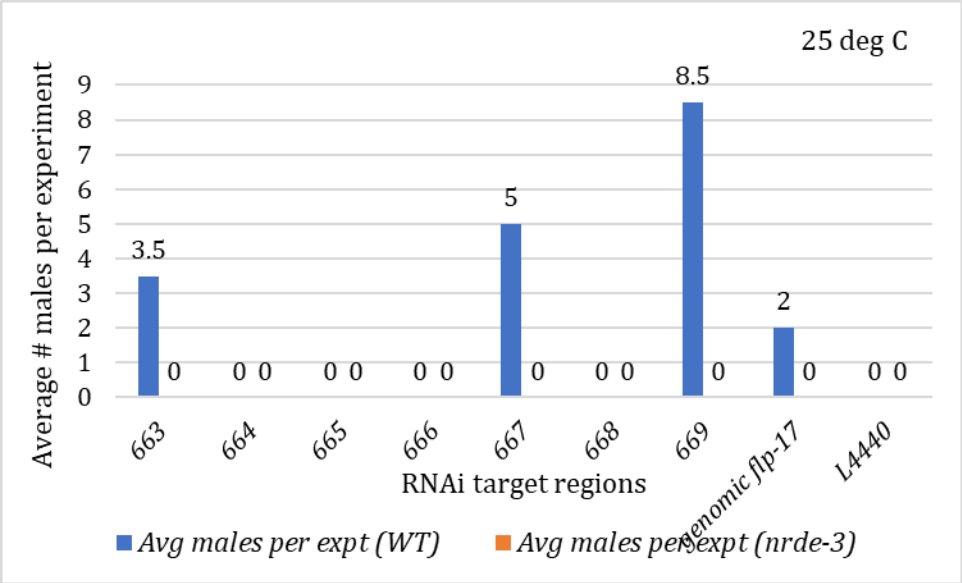


Figure 16: Feeding-induced Him phenotype at *flp-17* requires NRDE-3. RNAi assay includes feeding plasmids pLT663-pLT669 to target specific regions of *flp-17* locus to elicit Him phenotype. Each assay was performed in wildtype and *nrde-3* mutant animals. In every case, no males were seen in the progeny of *nrde-3* mutants. The data is collected from 11 experiments. Each experiment had 4 L1s added at the beginning of the experiment. Each L1 produces ~300 progeny. Refer to Table 9 for data.

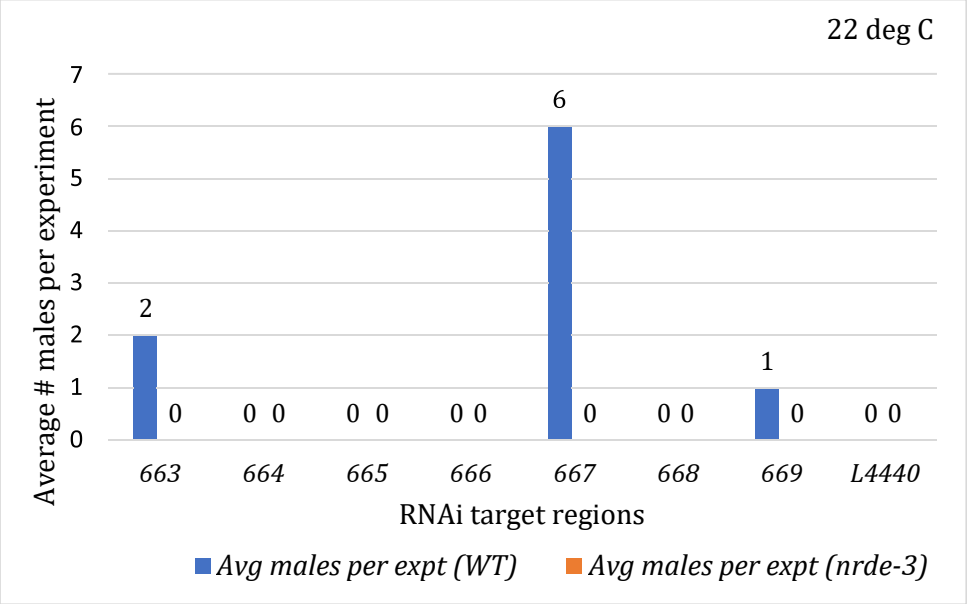


Figure 17: Feeding-induced Him phenotype at *flp-17* requires NRDE-3. RNAi assay as done in Figure 16 was repeated at 22 deg C. The results reiterate that NRDE-3 is required for RNAi induced non-disjunction phenotype at *flp-17* locus. Data is collected from 8 replicate plates. Each L1 produces ~300 progeny. Refer to Table 9 for data.

Deletion mutants did not display chromosomal disjunction defects

In order to determine if the piRNAs in the *flp-17* region contribute to the chromosomal non-disjunction phenotype, *flp-17* deletion mutants were procured from *Caenorhabditis* Genetics Center (CGC). Mutant strains, *ok3587* (about 700bp deletion) and *ok3614* (about 300bp deletion) had partial deletions of *flp-17* intronic region, as shown in Figure 18. The deletion breakpoints were verified by Sanger sequencing. In both strains, exon regions are deleted, producing frameshifts in the *flp*-17 coding region. While neither deletion strain is predicted to produce a functional FLP-17 protein, the strains did not

eliminate all of the intronic piRNAs. piRNA 1 is intact, including its upstream promoter region, in both strains.

The *flp-17* deletion strains did not display a Him phenotype, which helps to rule out a function for this protein in chromosome disjunction mechanisms. However, when RNAi targeted the remaining intronic region (piRNA 1) in the mutant background of strain *ok3587*, it elicited a Him phenotype. Interestingly, the *flp-17* deletion background was more sensitized for this assay, as this small dsRNA trigger was not potent for RNAi in a wild type background. These results may indicate a role for piRNA 1 in meiotic chromosomal disjunction; although a more complex response to dsRNA in this region might also be elicited.

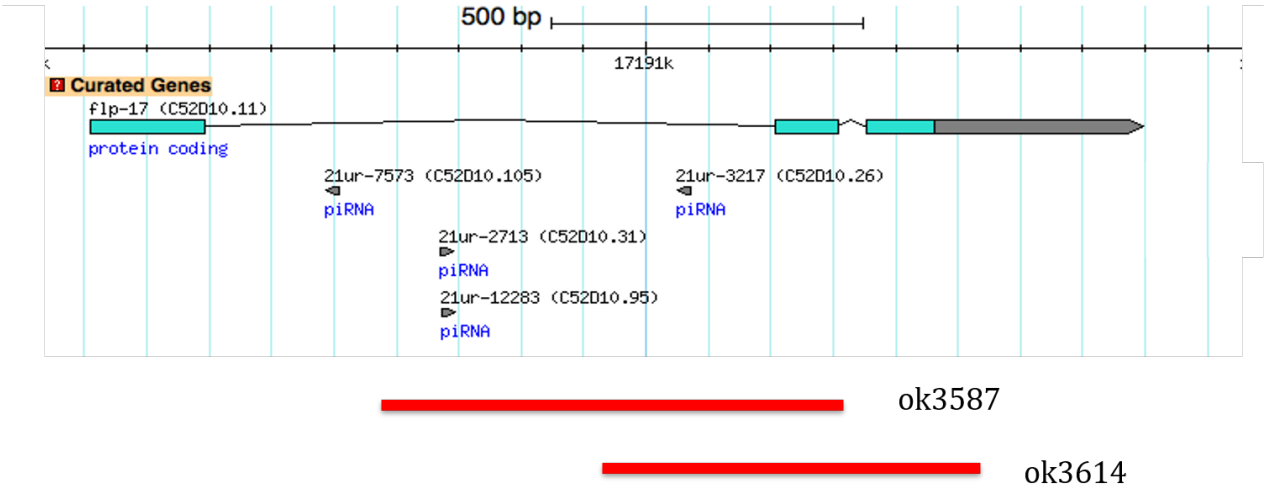


Figure 18 : Deletion mutants lacking piRNAs. The red bars indicate the span of deletion within the *flp-17* locus in the two strains *ok3587* and *ok3614*.

A novel nuclear silencing mechanism is active on the *flp-17* locus in wild-type *C. elegans*

eri-1 mutants are commonly used in RNAi experiments, as these mutants display hyperactive RNAi responses. *eri-1* mutants show an increased abundance of secondary siRNAs in the cytoplasm. The genetic screen that was used to identify NRDE-3 was performed in an *eri*-1 mutant background, and all the functions of *nrde*-3 with regard to silencing in the nucleus have been elicited in this context. In our case, the silencing we observe for the *flp*-17 locus occurred in a wild-type background. While evidence of RNAi in the nucleus in response to experimental delivery of dsRNA has been observed (in the form of histone methylation in the targeted region, for example), experimental delivery of dsRNA mostly elicits a RISC-based mRNA degradation response in the cytoplasm. Evidence for this comes from observations that RNAi by feeding requires the cytoplasmic Argonaute *rde-1,* but not the nuclear/shuttling Argonaute *nrde-3. rde-1* mutants are RNAi defective, but *nrde-3* mutants are not—*nrde-3*-related RNAi defects are only observed in an *eri-1* background. The nuclear silencing, we observe when targeting the *flp-17* locus is an example of an unusual form of nuclear silencing in a wild-type background.

ERI-1 is an RNA exonuclease with a nucleic acid binding domain, ERI-1 negatively regulates RNAi by degrading siRNAs. ERI-1 is found in the cytoplasm of wildtype *C. elegans* and its mutation causes the accumulation of siRNAs derived from exogenous dsRNA triggers. The enhanced RNAi phenotype observed in *eri-1* mutants can be explained, in part, by the accumulation of larger amounts of siRNAs. Accumulation of siRNAs in the cytoplasm acts as a trigger for NRDE-3 to transport the siRNAs to the nucleus resulting in knockdown of target

in the nucleus (Kennedy, Wang, & Ruvkun, 2004; Simmer et al., 2002). However, in *flp-17* region, we were presented with an unusual nuclear RNAi response in wildtype animals.

To test the possibility that this is a general phenomenon when piRNAs are present in the introns of any protein-coding gene, we identified genes with an architecture similar to that of *flp-17* locus--having piRNAs in introns. We further selected genes with an easy-to-score loss-of-function phenotype. We performed a control RNAi experiment using bacterial feeding methodology targeting the *unc-22* gene. The dsRNA produced includes genomic regions that include piRNAs located within *unc-22* introns. In our experiments, we scored Unc phenotypes and looked for Him phenotypes in wild type and in n*rde-3* mutants. We did not observe *nrde-3* dependent silencing in this region (Table 11). Not only did we fail to observe Him phenotpyes, but the Unc phenotype was independent of *nrde*-3 function. Thus, the presence of a piRNA in a locus being targeted by RNAi is not sufficient to elicit a Him phenotype. (This is expected; if this were not the case, Him phenotypes would be much more commonly observed in RNAi screens using the Source Biosciences RNAi library reagents as these were build using genomic DNA—many genes on chromosome IV contain intronic piRNA sequences.). These results also highlight the fact that *nrde*-3 is not essential for RNAi—the Unc phenotype we observed in these RNAi experiments may be completely accounted for by RDE-1/RISC-mediated mRNA degradation in the cytoplasm. These results also highlight the novel, *nrde-3* dependent silencing that we observe in wildtype *C. elegans* when the *flp-17* locus is the target of RNAi.

Bacterial clone (6K06) – RNAi food	**% Unc worms in F1 progeny**	
	Wildtype	***nrde-3***
Experiment 1	100 %	100%
Experiment 2	100 %	75 %
Experiment 3	60 %	75 %

Table 11 : *nrde-3* dependent silencing is unique to *flp-17* locus. RNAi experiments against piRNAs in *unc-22* locus in wildtype and *nrde-3* mutants exhibited similar phenotypes showing there is no nuclear dependent silencing for this locus.

Generating a Nuclear Silencing - sensitive target

The *flp-17* genomic locus, with its intronic piRNA genes, is particularly amenable to nuclear silencing. We hypothesize that this behavior may be due to a) a specific function of one or more piRNAs in the targeted region, or b) specific chromatin-related features of the locus, or a three-dimensional architecture. To address these hypotheses, we designed transgenes, intending to move the piRNAs to a new genic and genomic context – that is, the intronic region of a reporter gene. The reporter gene used was GFP that was designed to express in muscle cells. The transgenic vector pLT704 was constructed to include *flp-17* region encoding all the 3 piRNAs (approx. 375 bp), inserted within the introns of GFP reporter under the transcriptional control of *myo-3* promoter. As a control, in pLT706, GFP lacked this intron that consists of piRNAs from *flp-17* gene region (Figure 19). This allows

for an artificial target in our RNAi experiments and will explain if the susceptibility to nuclear silencing is attributed to the genomic interval containing the piRNAs.

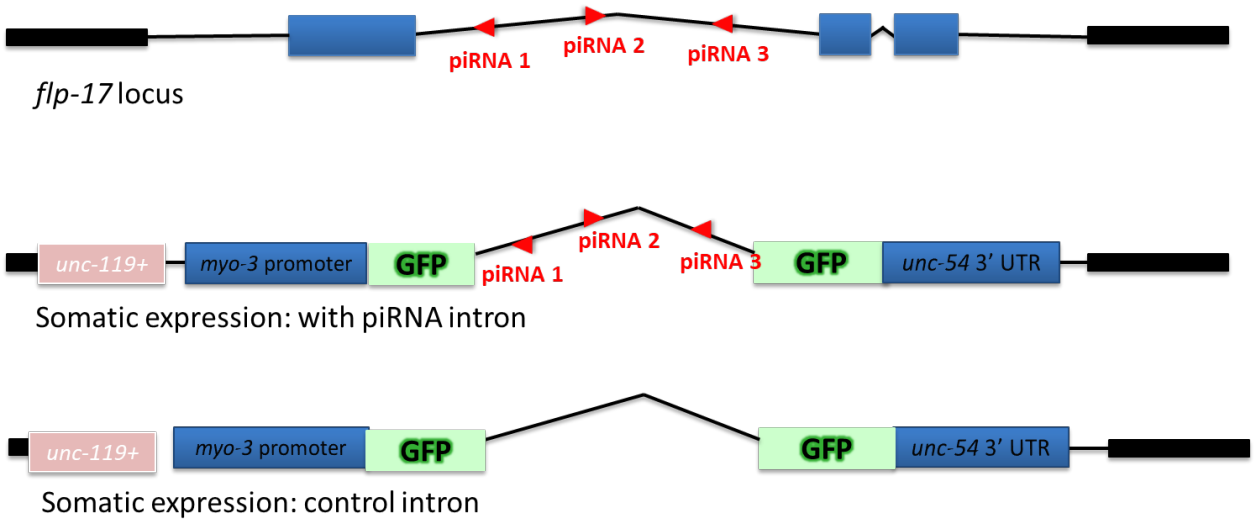


Figure 19 : Artificial 'nuclear silencing – sensitive target'. The top panel shows the schematic of *flp-17* gene with thick black bars representing homologous sequences for Mos mediated recombination, blue rectangles representing exons, and introns are seen as thin black bars between exons. piRNAs are found between exons 1 & 2. The middle panel shows the structure of a transgene construct with piRNAs included in the intronic region of a reporter gene, GFP under the control of *myo-3* promoter (pLT704). The bottom panel represents a control transgene which lacks the piRNAs (pLT706).

Transgenic single copy integrants for both constructs were obtained by the MosSCI method of transformation of *unc-119* mutant animals. The presence of transgene integrants was confirmed by PCR amplification. Once transgenic worms were obtained, the goal was to

perform RNAi experiments with dsRNA corresponding to the intron/ piRNAs insertion and the reporter sequence(GFP). In these cases, it was decided to assess for a Him phenocopy, and for loss of GFP expression. Similar to previous RNAi assays, these RNAi experiments were designed to include *nrde-3* controls to test for *nrde-3* dependence that confirms nuclear silencing. The speculation was that, an easier readout with a reporter like GFP will help us gauge the extent to which this form of nuclear silencing is prevalent. Furthermore, recapitulating the *flp-17* architecture in a different gene set-up will allow us to determine if the amenability to nuclear silencing is context-specific (the genomic region including *flp-17*) or if it is a feature of including small non-coding RNAs in the intronic region of any gene.

Once pLT704 and pLT706 transgenes were injected in *unc-119 C. elegans* (XX1327) by MosSCI method of transformation, we obtained 22 independent lines from pLT704 and 3 lines from pLT706. In addition to these, we obtained a number of array lines for each transgene. We hypothesize that RNAi against the piRNAs will silence GFP expression, lining with our previous observation of nuclear silencing.

Unexpected silencing of transgenes

When the above two previously discussed transgene constructs were injected, we were faced with a unique phenomenon. In transgenic experiments it is not surprising to observe silencing in germline tissue in *C. elegans (Kelly et al., 1997)*. We know that the germline of *C. elegans* has potent anti-foreign genome responses. But in our case, silencing of transgene is repeatedly seen in somatic tissues such as muscle cells, in our independent single copy integrants, in the absence of dsRNA triggers. This implies the silencing is novel

and unprecedented in *C. elegans*. In a previous research in the field, silencing of a transgene *elt2 :: gfp/LacZ* was observed in the soma of *C. elegans* (Grishok et al., 2005). But it was shown to involve RNAi pathway genes and hence referred to it as RNAi-induced TGS. Transgene silencing in soma is observed in *pals-22* mutants. In this case, highly repetitive array transgene, but not single copy reporters, is silenced at the loss of PALS-22 (Leyva-Diaz et al., 2017). Contrary to all these observations, transgenic silencing in our experiments occurred spontaneously and in wildtype background.

GFP transgene is capable of expression in muscle cells of transformed *C. elegans*

Transgenic plasmids pLT704 and pLT706 were injected into *unc-119 C. elegans* mutants. Transgenes that are designed for single copy integration into the worm genome initially formed extrachromosomal arrays of the injected DNA. These worms which carried the arrays, expressed *myo-3* driven GFP. Secondly, in an attempt to desilence the transgene, we conducted gene knockdown experiments using RNAi by feeding against potential effectors of transgene silencing. We selected genes that were shown to be implicated in desilencing of *let858::GFP* transgenic array as shown by Towbin et al (Towbin et al., 2012). It is to be noted that the L4440 RNAi vector also resulted in minimal expression of GFP from in single copy integrants of transgenes that were earlier silenced. Among the targets that were knocked down, *ubl-5* RNAi consistently resulted in re-expression (or desilencing) of *myo-3::GFP*. Thirdly, spontaneous reversal of transgene silencing occurred with all independent lines that were previously lacking GFP expression. Maintenance of population of single copy integrants of pLT704 and pLT706 lines over several generations resulted in

desilencing of the GFP-expressing transgene (Figure 20). These lines were now seen to express *myo-3 :: GFP* in a heritable manner. Collectively, these results indicate that our transgenes pLT704 and pLT706 are capable of expressing GFP in muscle cells *(myo-3::GFP)*.

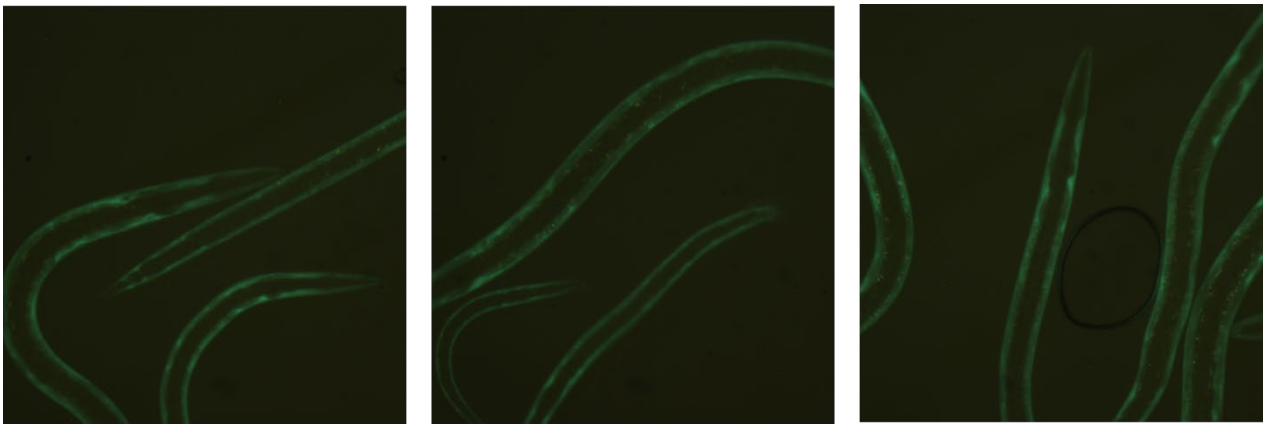


Figure 20: Spontaneous reversal of transgene silencing. Three representative single copy integrant lines transformed with plasmid pLT704 (transgenic plasmid with *myo-3::GFP* with piRNAs included in the intron of GFP) shows somatic expression of transgene.

Investigating mechanisms of silencing

To explore the nature of desilencing of transgene expression, whole genome sequencing was performed on previously silenced integrant animals. Surveying for mutations associated with RNAi functions revealed a previously uncharacterized 88 bp deletion in a RNA helicase gene. The deletion is mapped to a regulatory region (Untranslated region) of *eri-6* gene on chromosome 1 (Figure 21). The 88bp deletion mutation is designated as *yy14*.

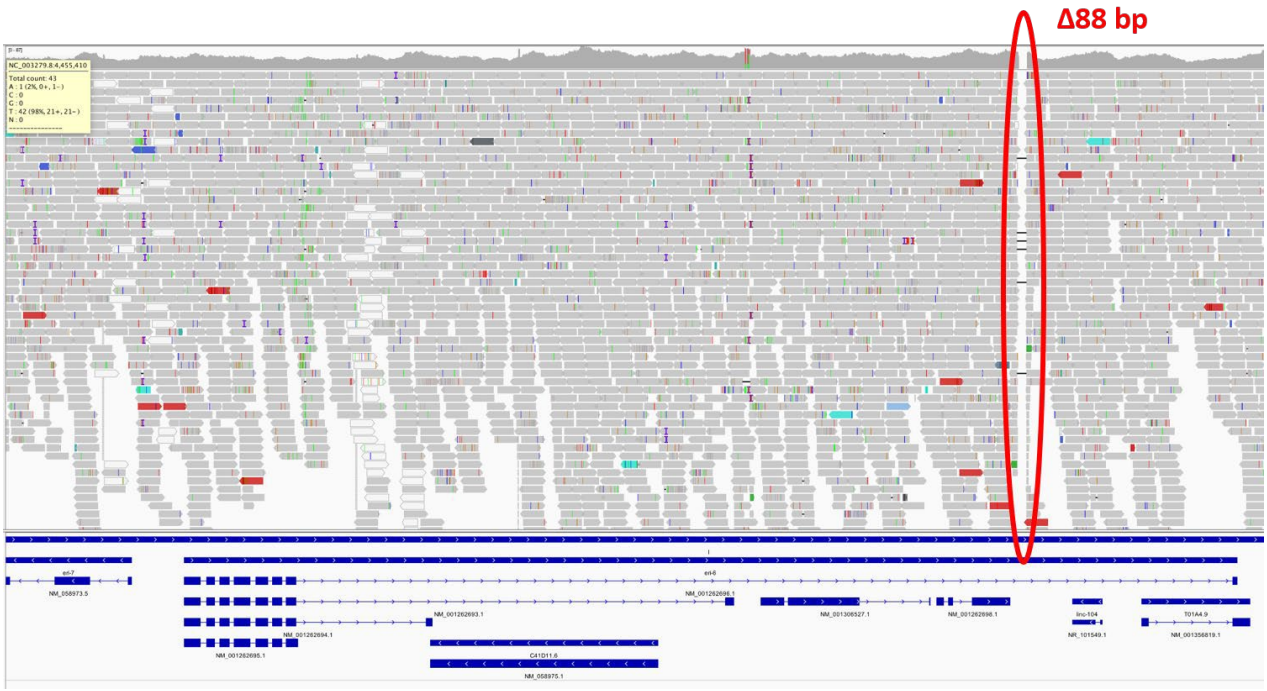


Figure 21 : IGV window showing the deletion of 88 bases in *eri-6*.

eri-6 and *eri-7* genes are encoded on opposite strands on chromosome 1 of *C. elegans*. Independent transcription of the two genes is achieved by the presence of a bidirectional promoter (Figure 22). The pre mRNAs of the two genes are found to trans-splice to form a single mRNA where the 3' end of mRNA corresponds to *eri-7* sequence. This is the first case of trans-splicing in *C. elegans* between exons of two genes. In a closely related species, *C. briggsae*, these are encoded by a single gene that lies on the same chromosome as *eri-7* lies in *C. elegans*. In *C. elegans* wildtype strain N2, the *eri-6* sequence lies between 930 bp direct repeats on either side (Fischer et al., 2008). A 25-bp inverted repeat lies within each of the direct repeat sequences. The trans-splicing was shown by Reverse Transcription-PCR indicating the presence of one single mRNA (Fischer et al., 2008) .

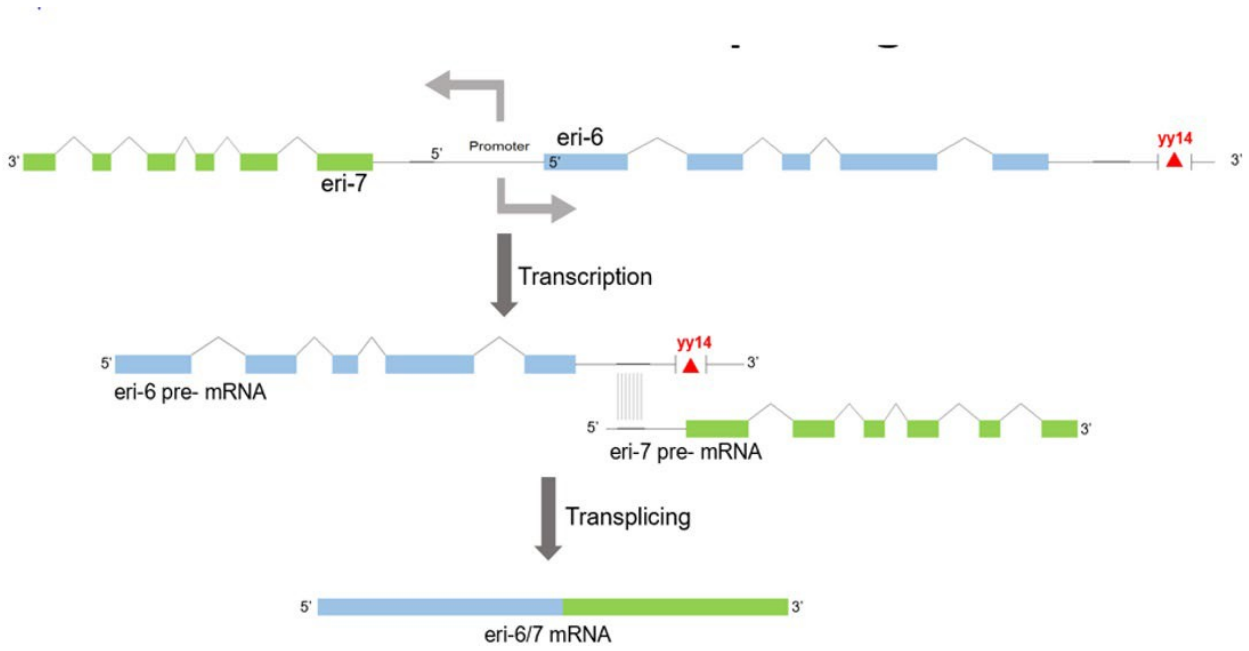


Figure 22 : Trans-splicing of *eri-6 and 7* occurs via the homologous complementary region. Exons of *eri-6* and *eri-7* are denoted by blue and green rectangles respectively. The genes are transcribed by a common promoter that acts bidirectionally. *yy14* is present in a regulatory region of *eri-6*.

The chimeric protein ERI-6/7 belongs to superfamily I helicases. Superfamily I helicases can act on DNA or RNA. But the presence of ERI-6/7 in cytoplasm indicates it to act on RNAs. ERI-6/7 is required in endogenous small RNA pathways. ERGO-1 is an argonaute protein that associates with a subset of ERI class of small RNAs. 26G small RNAs (monophosphorylated at 5' end) and their downstream 22G siRNAs are the ERI class small RNAs. In *eri-6/7* mutants, the ERGO-1 dependent 26G RNAs corresponding to about one hundred gene loci (which includes, pseudogenes and transposons) are absent as shown by qRT-PCR experiment. As a consequence, it was shown that the mRNAs of these genes were

up regulated (Fischer et al., 2011). These target genes are non-conserved between *C. elegans* and *C. briggsae* which suggests that ERI-6/7 is required for antiviral like functions protecting genome from the expression of transposons, pseudogenes and repetitive sequences.

Eri phenotypes from other *eri* genes are associated with temperature dependent sterility and nondisjunction of X chromosomes which are absent in *eri-6/7* mutants (Fischer et al., 2008). ERI-6/7 is a negative regulator of exogenous RNAi and promotes endogenous silencing of pseudogenes, repetitive arrays by an ERGO-1/26G RNA dependent small RNA silencing pathway. The opposite roles of ERI-6/7 in exogenous and endogenous RNAi implies competition between the two, in terms of common ERI-6/7 associated factors. From the nature of targets of ERI-6/7, it is speculated that ERI-6/7 has implications in antiviral like defense mechanisms.

yy14 regulates trans-splicing of *eri-6* and *eri-7*

As trans-splicing is a feature unique to this region, we hypothesized that the deletion will affect trans-splicing. To determine if the deletion mutation regulates the inter-genic trans-splicing, we performed qRT-PCR to quantify trans-splicing in *yy14* mutants relative to wildtype animals. In these experiments, cDNA from various regions viz. *eri-6/7* junction exon in *eri-6,* trans-splice region and exon from *eri-7* were amplified to investigate the effect of *yy14* mutation on trans-splicing. Reproducibly, the levels of trans-spliced mRNA increased two-folds in the presence of *yy14* mutation (Figure 23). This is consistent with a role for ERI-6/7 in silencing of our transgene.

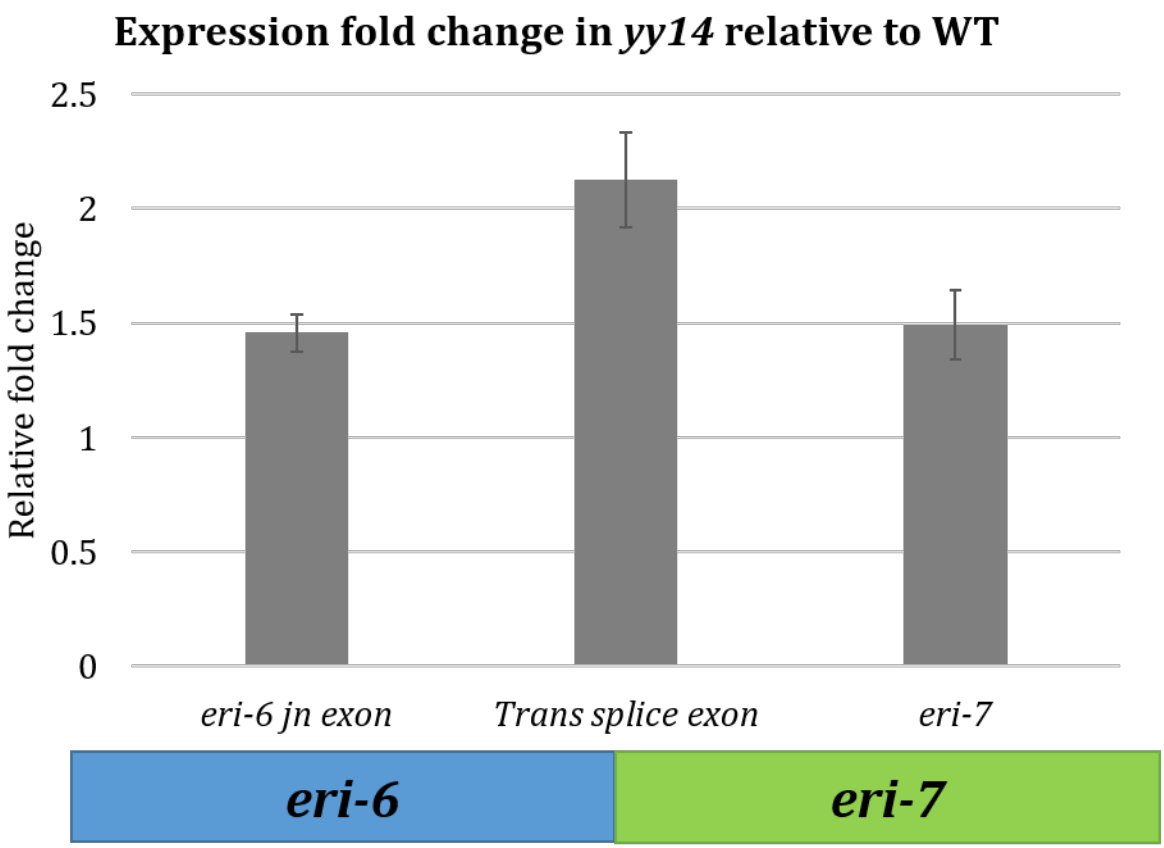


Figure 23: *yy14* affects trans-splicing. Along the x-axis are the regions amplified in the PCR. Y-axis denotes expression levels. Wildtype expression level is taken as 1. The middle bar represents expression level when trans-splice exon in *eri-6/7* is amplified and showed a 2-fold change in *yy14* with respect to wildtype levels. Number of experiments = 2; Number of plates in each experiment = 3; Total number of worms = >2000.

The hypothesis is that ERI-6/7 protein promotes the biogenesis of secondary siRNAs, the 26G RNAs corresponding to our transgene sequence. These secondary small RNAs mediate silencing of our transgene in an ERGO-1 dependent manner.

eri mutations affect inheritance of extrachromosomal arrays

Introduction of DNA by microinjection results in the formation of transgenic arrays in *C. elegans.* The DNA becomes concatenated into extrachromosomal arrays by non-homologous recombination, resulting in repetitive, presumably tandemly arranged plasmid sequences, each present in the megabase array in multi-copy. These arrays are mitotically and meiotically unstable in that they are not always distributed to daughter cells. The transmission frequency with which an array is meiotically inherited ranges from 1-99%, depending on the array. Therefore, there is a possibility of extrachromosomal arrays being lost after several generations (Stinchcomb, Shaw, Carr, & Hirsh, 1985). Array formation is likely a response to the invasion of foreign DNA. Arrays are targeted for silencing using RNAi and chromatin machineries that lead to heterochromatin formation. These silencing mechanisms are particularly robust in the germ line; indeed, even though the Arrays harbor multiple copies of introduced genes, they are often transcriptionally silenced in the germ line. This provides a further line of protection against foreign DNA, as invading viruses or transposons would not transcribe the necessary genes that would allow them to establish a foothold in the organism. The fact that Arrays are non-uniformly inherited likely provides some evolutionary advantages. As influx of genes leads to evolution, mechanisms that completely disallow such influx would lead to an evolutionary dead end, which might prove detrimental to the species in adverse environments. The non-uniform nature of inheritance of Arrays would allow for the possibility of some individuals to potentially acquire new traits, providing an evolutionary advantage to the species.

We observed that the frequency of transmission of GFP hairpin arrays (TFA phenotype) decreases in *eri-1* mutants compared to wild type. In *eri-1* mutants, the

frequency at which Arrays are inherited is sufficiently low that it can be difficult to maintain the stock. To test if this phenotype extends to other *eri* genes, we performed RNAi experiments that knocked down *eri* genes, including *eri-6* and *eri-7*. Interestingly, we observed that RNAi of all *eri* genes tested exhibited a decrease in frequency of transmission of an Array expressing a GFP hairpin in comparison to wild type (Figure 24).

Arrays are high molecular weight repeat sequences that acquire centromeres de novo and are segregated autonomously of the genomic DNA at different frequencies (Yuen, Nabeshima, Oegema, & Desai, 2011). These extrachromosomal arrays are not monocentric. Monocentromeres refer to single centromeric region on the chromosome with a localized kinetochore. Such centromeres are found in vertebrates including humans, and fungi. There are some organisms that have delocalized or dispersed centromeres in which the kinetochore complex is known to diffuse along the length of the chromosome. *C. elegans* is a well-known example of an organism with holocentric chromosomes. In this case, the microtubule spindles attach at positions along the entire length of the chromosomes (Maddox, Oegema, Desai, & Cheeseman, 2004). Thus, for the purposes of segregation, array chromosomes are a model of endogenous chromosomes *in C. elegans*. With this understanding, we predict a role for *eri* genes in chromosomal segregation functions of centromere.

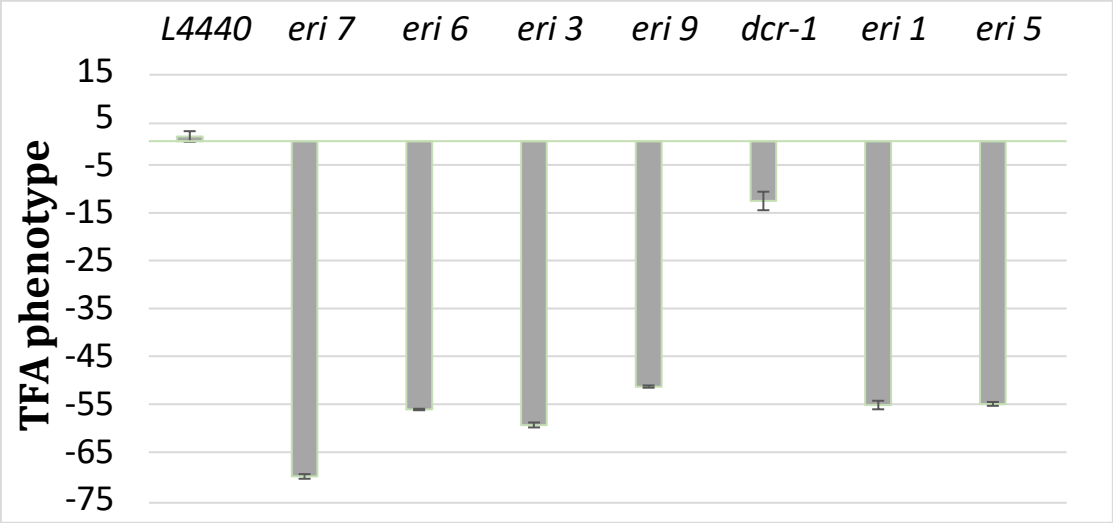


Figure 24: *eri* affects centromere function. Along the x-axis are genes that are knocked down by RNAi. Wild type worms harboring an extrachromosomal Array marked with Roller phenotype (caused by a dominant mutation in a collagen gene) were placed on gene -specific dsRNA-expressing bacteria, or control bacteria (L4440). The meiotic transmission of the y frequency (TFA) was assessed by scoring all the F1 progeny produced from a single Roller adult (Roller frequency). The Y-axis denotes the percentage difference in TFA normalized to animals reared on normal food. *eri* mutants show decrease in TFA compared to wildtype seen as greater % difference in TFA. *eri-7* RNAi resulted in the highest decrease in TFA among *eri* genes.

The centromeres are occupied by histone variant of H3 , cenH3 in *C. elegans*, referred to as CENP-A in many organisms. Research led by Steven Henikoff has shown that the holocentromeres in *C. elegans* are polycentromeric in nature, localized to discrete positions along the entire length. There are about 100 centromeric sites in each chromosome but only 100 microtubule attachments are found, in total, for all six C. *elegans* chromosomes (Steiner

& Henikoff, 2014). During the cell divisions in the developing organism, the cenH3 is lost from somatic cells as they cease cell divisions, and the cenH3 is not replaced. In the germline, sites of active transcription anti-correlates with cenH3 deposition (Gassmann et al., 2012). Moreover, non-transcribed regions at the embryonic stage were correlated with cenH3 incorporation. (The presence of cenH3 correlated with the presence of the inner kinetochore protein CENP-C/HCP-4, indicating that these are kinetochore attachment sites.) The cenH3 binding sites correlated with a motif that, through ChIP analysis, correlates with binding of 22 transcriptional factors. Thus, HOT sites and centromeric sites share common chromatin features and are targeted by both cenH3 nucleosomes and transcription factors, depending on the stage of development. It is postulated that the replacement of centromeric sites by transcription factors occurs upon exit from the cell cycle, and that cenH3, deposited in non-transcribed regions, essentially marks these regions for transcription factor incorporation during later stages of development in differentiated cells that now transcribe RNAs from the region (Steiner & Henikoff, 2014).

The mechanisms that lead to establishment and maintenance of heterochromatin as well as placement of cenH3 is better understood for yeast centromeres, which can be partially defined based on DNA sequence (Lejeune, Bayne, & Allshire, 2010). In yeast, transcription of the centromere outer repeats generates dsRNA, which can be processed into siRNAs by Dicer1. The siRNAs are loaded onto Ago1, targeting the effector complex RITS (RNA-induced Transcriptional Silencing) to the outer repeat region by a base-pairing mechanism. RITS interacts with the RDRC **(**RNA-dependent RNA Polymerase Complex), increasing the pool of siRNAs via Rdp1 activity. A subset of splicing factors also contributes to this amplification process, in an unknown fashion. The activity of the RNAi machinery

results in recruitment of the CLRC histone-modifier complex to chromatin, allowing the CLRC subunit Clr4 to methylate histone H3 on lysine 9. This creates a mark that is recognized by the chromodomain protein Swi6 that, in turn, recruits cohesin, which is required for proper biorientation of centromeres during mitosis. All these silencing complexes and silencing marks might function as boundary elements, allowing for the correct placement and incorporation of CENP-A (CenH3) (Nakagawa & Okita, 2019). Indeed, for the more permanent yeast centromeres, transcription from centromeric regions is detrimental, leading not only to chromosome non-disjunction, but also to chromosomal rearrangements.

It is interesting to speculate that RNAi mechanisms might contribute to heterochromatin formation around C. *elegans* cenH3 sites, allowing for proper placement of cenH3 in early development akin to the yeast mechanism, and also preventing gene expression and transcription factor occupancy in the region that time in development. Early evidence that this might be the case came from analysis of *eri-1* mutants. Not only are *eri-1* mutants hyperactive for RNAi, but these mutants also display a Him phenotype and spindle defects in early embryos (Pavelec, Lachowiec, Duchaine, Smith, & Kennedy, 2009). ERI-1 is a potent modulator of exogenous and exogenous RNAi pathways, and is found in a complex with the RNA-dependent RNA polymerase RRF-3, the novel protein ERI-3 and the Tudor domain-containing protein ERI-5 (Lee et al., 2006). *eri-1* and *rrf-3* mutants fail to accumulate 26G endo-siRNAs that regulate spermatogenesis(Han et al., 2009) and zygotic development and give rise to 22G RNAs(Gu et al., 2009) . Although *rrf-3* and *eri-1* mutants are similar in many ways (both are hyperactive for exogenous RNAi, both display Him phenotypes, both reside in similar complexes, both are involved in 26G- and 22G-endo- siRNA production), yet biochemically the proteins are quite different. *eri-1* encodes an

RNaseT-like nuclease while RRF-3 encodes an RNA-dependent RNA polymerase. There are few examples of non-overlapping functions for these genes.

When we performed the TFA analysis on *rrf-3* mutants, we observed a phenotype opposite from that observed in *eri-1* mutants: the same extrachromosomal Array crossed into an *rrf-3* mutant background was transmitted to a greater number of progeny than the same Array in a wild-type background. As silencing RNAs can spread in *C. elegans* from somatic cells to germ line, it is interesting to speculate that trafficking of RNAs requiring ERI-1/RRF-3 for function can influence epigenetic modifications such as histone replacements in a multi-generational fashion. For example, active transcription at HOT sites in nondividing somatic cells may lead to the production of non-coding RNAs that traffick to the germline and influence the placement of cenH3 in an RNAi/heterochromatin- and sequence-dependent fashion. Disruptions in the amount of silencing RNAs may influence how many germline cells receive the epigenetic mark. Opposite effects observed in *eri-1* versus *rrf-3* mutations might reflect differential influences on the status of chromatin in Array centromeres at cenH3 binding sites versus surrounding heterochromatin.

CONCLUSION

We have shown here that *eri-6/7(yy14)* results in increase in the trans-spliced mRNA, along with nearly a 1.5-fold increase in expression of *eri*-6 and *eri-7* mRNA. ERI-6/7 protein promotes the biogenesis of secondary siRNAs, the 26G RNAs. In our case, we postulate that ERI-6/7 is important for the biogenesis of 26G RNAs corresponding to our transgene sequence. These secondary small RNAs may mediate the silencing of our transgene in an ERGO-1 dependent manner.

Our model reasons out that increase in trans-spliced mRNA results in upregulation of small RNAs (26G RNAs) that efficiently function in the silencing of our transgene (Figure 25).

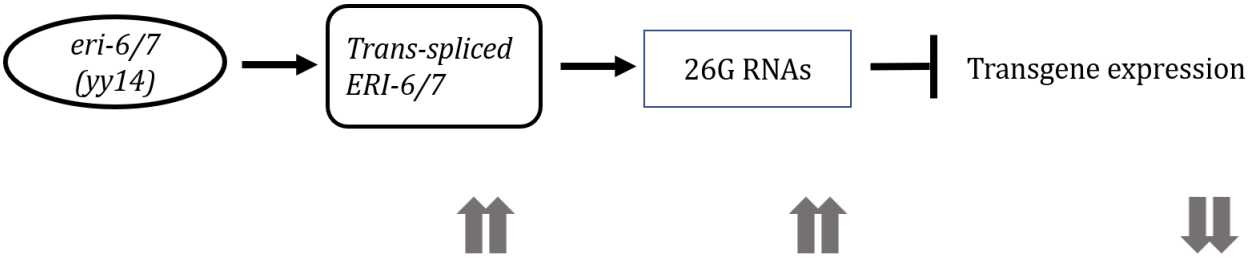


Figure 25: *eri-6/7(yy14)* results in silencing of the transgene. A hypothetical model explains ERI-6/7 dependent 26G-RNAs are required for targeting the transgenic DNA for silencing in germ line and somatic tissues.

This needs to be validated by performing deep sequencing of cDNA amplicons corresponding to 26G small RNAs from yy14 mutant animals compared to that of wildtype. In addition, a quantitative RT-PCR is required to determine the effect of *yy14* on the levels of these 26G-RNAs. To determine if the transgene sequences are targets of silencing by small

RNA pathways, the deep sequencing datasets will be assessed. Additionally, to test this hypothesis, single copy integrants of our transgene will be obtained in *ergo-1* mutants.

On the other hand, to explain how *yy14* might regulate trans-splicing levels, we hypothesize 88 bases in the regulatory region is essential for the binding of a repressor protein. This can be tested by bioinformatics analysis of the 88bp region and looking for homologous regulatory regions and proteins associated with them. The protein can be found using Enzyme mobility shift assays or other DNA binding and pulldown assays and the function of the regulatory protein can be confirmed by β-galactosidase assay, qPCR or other transcriptomic analysis tools like RNAseq. Alternatively, loss of 88 bases may increase the homology of reverse complementary site between *eri-6* and *eri-7,* thus upregulating trans-splicing of mRNAs. This can be investigated by using fluorescent resonance energy transfer (FRET) based approach where the intensity of fluorescent emission corresponds to the complementarity of the region.

REFERENCES

1. Allen, M. A., Hillier, L. W., Waterston, R. H., & Blumenthal, T. (2011). A global analysis of C. elegans trans-splicing. *Genome Res, 21*(2), 255-264. doi:10.1101/gr.113811.110
2. Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., & Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in C. elegans. *Curr Biol, 13*(10), 807-818.
3. Ashe, A., Sapetschnig, A., Weick, E. M., Mitchell, J., Bagijn, M. P., Cording, A. C., . . . Miska, E. A. (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. *Cell, 150*(1), 88-99. doi:10.1016/j.cell.2012.06.018
4. Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., & Pasquinelli, A. E. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell, 122*(4), 553-563. doi:10.1016/j.cell.2005.07.031
5. Batista, P. J., Ruby, J. G., Claycomb, J. M., Chiang, R., Fahlgren, N., Kasschau, K. D., . . . Mello, C. C. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol Cell, 31*(1), 67-78. doi:10.1016/j.molcel.2008.06.002
6. Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell, 136*(4), 642-655. doi:10.1016/j.cell.2009.01.035
7. Cohen, M., Reale, V., Olofsson, B., Knights, A., Evans, P., & de Bono, M. (2009). Coordinated regulation of foraging and metabolism in C. elegans by RFamide neuropeptide signaling. *Cell Metab, 9*(4), 375-385. doi:10.1016/j.cmet.2009.02.003

8. Couteau, F., Guerry, F., Muller, F., & Palladino, F. (2002). A heterochromatin protein 1 homologue in Caenorhabditis elegans acts in germline and vulval development. *EMBO Rep, 3*(3), 235-241. doi:10.1093/embo-reports/kvf051
9. Das, P. P., Bagijn, M. P., Goldstein, L. D., Woolford, J. R., Lehrbach, N. J., Sapetschnig, A., . . . Miska, E. A. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol Cell, 31*(1), 79-90. doi:10.1016/j.molcel.2008.06.003
10. Dernburg, A. F., Zalevsky, J., Colaiacovo, M. P., & Villeneuve, A. M. (2000). Transgene-mediated cosuppression in the C. elegans germ line. *Genes Dev, 14*(13), 1578-1583.
11. Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet, 9*(2), 102-114. doi:10.1038/nrg2290
12. Fire, A., Albertson, D., Harrison, S. W., & Moerman, D. G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. *Development, 113*(2), 503-514.
13. Fischer, S. E., Butler, M. D., Pan, Q., & Ruvkun, G. (2008). Trans-splicing in C. elegans generates the negative RNAi regulator ERI-6/7. *Nature, 455*(7212), 491-496. doi:10.1038/nature07274
14. Fischer, S. E., Montgomery, T. A., Zhang, C., Fahlgren, N., Breen, P. C., Hwang, A., . . . Ruvkun, G. (2011). The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications. *PLoS Genet, 7*(11), e1002369. doi:10.1371/journal.pgen.1002369

15. Fischer, S. E., Pan, Q., Breen, P. C., Qi, Y., Shi, Z., Zhang, C., & Ruvkun, G. (2013). Multiple small RNA pathways regulate the silencing of repeated and foreign genes in C. elegans. *Genes Dev, 27*(24), 2678-2695. doi:10.1101/gad.233254.113
16. Flouriot, G., Brand, H., Seraphin, B., & Gannon, F. (2002). Natural trans-spliced mRNAs are generated from the human estrogen receptor-alpha (hER alpha) gene. *J Biol Chem, 277*(29), 26244-26251. doi:10.1074/jbc.M203513200
17. Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., & Ahringer, J. (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature, 408*(6810), 325-330. doi:10.1038/35042517
18. Frokjaer-Jensen, C., Davis, M. W., Hopkins, C. E., Newman, B. J., Thummel, J. M., Olesen, S. P., . . . Jorgensen, E. M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat Genet, 40*(11), 1375-1383. doi:10.1038/ng.248
19. Gassmann, R., Rechtsteiner, A., Yuen, K. W., Muroyama, A., Egelhofer, T., Gaydos, L., . . . Desai, A. (2012). An inverse relationship to germline transcription defines centromeric chromatin in C. elegans. *Nature, 484*(7395), 534-537. doi:10.1038/nature10973
20. Grishok, A. (2005). RNAi mechanisms in Caenorhabditis elegans. *FEBS Lett,579*(26), 5932-5939. doi:10.1016/j.febslet.2005.08.001
21. Grishok, A., Sinskey, J. L., & Sharp, P. A. (2005). Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. *Genes Dev, 19*(6), 683-696. doi:10.1101/gad.1247705
22. Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P. J., Claycomb, J. M., . . . Mello, C. C. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome

surveillance in the C. elegans germline. *Mol Cell, 36*(2), 231-244. doi:10.1016/j.molcel.2009.09.020

23. Guang, S., Bochner, A. F., Burkhart, K. B., Burton, N., Pavelec, D. M., & Kennedy, S. (2010). Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature, 465*(7301), 1097-1101. doi:10.1038/nature09095
24. Guang, S., Bochner, A. F., Pavelec, D. M., Burkhart, K. B., Harding, S., Lachowiec, J., & Kennedy, S. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science, 321*(5888), 537-541. doi:10.1126/science.1157647
25. Han, T., Manoharan, A. P., Harkins, T. T., Bouffard, P., Fitzpatrick, C., Chu, D. S., . . . Kim, J. K. (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. *Proc Natl Acad Sci U S A, 106*(44), 18674-18679. doi:10.1073/pnas.0906378106
26. Hollick, J. B. (2017). Paramutation and related phenomena in diverse species. *Nat Rev Genet, 18*(1), 5-23. doi:10.1038/nrg.2016.115
27. Hull, D., & Timmons, L. (2004). Methods for delivery of double-stranded RNA into Caenorhabditis elegans. *Methods Mol Biol, 265*, 23-58. doi:10.1385/1-59259-775-0:023
28. Jin, Y., & Guo, H. S. (2015). Transgene-induced gene silencing in plants. *Methods Mol Biol, 1287*, 105-117. doi:10.1007/978-1-4939-2453-0_7
29. Kelly, W. G., & Fire, A. (1998). Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. *Development, 125*(13),2451-2456.

30. Kelly, W. G., Xu, S., Montgomery, M. K., & Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. *Genetics, 146*(1), 227-238.

31. Kennedy, S., Wang, D., & Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. *Nature, 427*(6975), 645-649. doi:10.1038/nature02302

32. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G., & Plasterk, R. H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell, 99*(2), 133-141.

33. Ketting, R. F., & Plasterk, R. H. (2000). A genetic link between co-suppression and RNA interference in C. elegans. *Nature, 404*(6775), 296-298. doi:10.1038/35005113

34. Kuersten, S., Lea, K., MacMorris, M., Spieth, J., & Blumenthal, T. (1997). Relationship between 3' end formation and SL2-specific trans-splicing in polycistronic Caenorhabditis elegans pre-mRNA processing. *RNA, 3*(3), 269-278.

35. Lee, R. C., Hammell, C. M., & Ambros, V. (2006). Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. *RNA, 12*(4), 589-597. doi:10.1261/rna.2231506

36. Lejeune, E., Bayne, E. H., & Allshire, R. C. (2010). On the connection between RNAi and heterochromatin at centromeres. *Cold Spring Harb Symp Quant Biol, 75*, 275-283. doi:10.1101/sqb.2010.75.024

37. Leyva-Diaz, E., Stefanakis, N., Carrera, I., Glenwinkel, L., Wang, G., Driscoll, M., & Hobert, O. (2017). Silencing of Repetitive DNA Is Controlled by a Member ofan

Unusual Caenorhabditis elegans Gene Family. *Genetics, 207*(2), 529-545. doi:10.1534/genetics.117.300134

38. Li, C., Kim, K., & Nelson, L. S. (1999). FMRFamide-related neuropeptide gene family in Caenorhabditis elegans. *Brain Res, 848*(1-2), 26-34.
39. Luteijn, M. J., van Bergeijk, P., Kaaij, L. J., Almeida, M. V., Roovers, E. F., Berezikov, E., & Ketting, R. F. (2012). Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. *EMBO J, 31*(16), 3422-3430. doi:10.1038/emboj.2012.213
40. Maddox, P. S., Oegema, K., Desai, A., & Cheeseman, I. M. (2004). "Holo"er than thou: chromosome segregation and kinetochore function in C. elegans. *Chromosome Res, 12*(6), 641-653. doi:10.1023/B:CHRO.0000036588.42225.2f
41. Mao, H., Zhu, C., Zong, D., Weng, C., Yang, X., Huang, H., . . . Guang, S. (2015). The Nrde Pathway Mediates Small-RNA-Directed Histone H3 Lysine 27 Trimethylation in Caenorhabditis elegans. *Curr Biol, 25*(18), 2398-2403. doi:10.1016/j.cub.2015.07.051
42. McManus, C. J., Duff, M. O., Eipper-Mains, J., & Graveley, B. R. (2010). Global analysis of trans-splicing in Drosophila. *Proc Natl Acad Sci U S A, 107*(29), 12975-12979. doi:10.1073/pnas.1007586107
43. Mello, C. C., Kramer, J. M., Stinchcomb, D., & Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO J, 10*(12), 3959-3970.
44. Nakagawa, T., & Okita, A. K. (2019). Transcriptional silencing of centromere repeats by heterochromatin safeguards chromosome integrity. *Curr Genet*. doi:10.1007/s00294-019-00975-x

45. Nelson, L. S., Rosoff, M. L., & Li, C. (1998). Disruption of a neuropeptide gene, flp-1, causes multiple behavioral defects in Caenorhabditis elegans. *Science, 281*(5383), 1686-1690.

46. Pavelec, D. M., Lachowiec, J., Duchaine, T. F., Smith, H. E., & Kennedy, S. (2009). Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in Caenorhabditis elegans. *Genetics, 183*(4), 1283-1295. doi:10.1534/genetics.109.108134

47. Pillai, R. S., Bhattacharyya, S. N., & Filipowicz, W. (2007). Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol, 17*(3),118-126. doi:10.1016/j.tcb.2006.12.007

48. Ringstad, N., & Horvitz, H. R. (2008). FMRFamide neuropeptides and acetylcholine synergistically inhibit egg-laying by C. elegans. *Nat Neurosci, 11*(10), 1168-1176. doi:10.1038/nn.2186

49. Robert, V. J., Sijen, T., van Wolfswinkel, J., & Plasterk, R. H. (2005). Chromatin and RNAi factors protect the C. elegans germline against repetitive sequences. *Genes Dev, 19*(7), 782-787. doi:10.1101/gad.332305

50. Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., . . . Bartel, D. P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell, 127*(6), 1193-1207. doi:10.1016/j.cell.2006.10.040

51. Shirayama, M., Seth, M., Lee, H. C., Gu, W., Ishidate, T., Conte, D., Jr., & Mello, C. C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. *Cell, 150*(1), 65-77. doi:10.1016/j.cell.2012.06.015

52. Shiu, P. K., & Hunter, C. P. (2017). Early Developmental Exposure to dsRNA Is Critical for Initiating Efficient Nuclear RNAi in C. elegans. *Cell Rep, 18*(12), 2969-2978. doi:10.1016/j.celrep.2017.03.002

53. Sijen, T., & Plasterk, R. H. (2003). Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature, 426*(6964), 310-314. doi:10.1038/nature02107

54. Simmer, F., Tijsterman, M., Parrish, S., Koushika, S. P., Nonet, M. L., Fire, A., . . . Plasterk, R. H. (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi. *Curr Biol, 12*(15), 1317-1319.

55. Smardon, A., Spoerke, J. M., Stacey, S. C., Klein, M. E., Mackin, N., & Maine, E. M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. *Curr Biol, 10*(4), 169-178.

56. Stawicki, T. M., Takayanagi-Kiya, S., Zhou, K., & Jin, Y. (2013). Neuropeptides function in a homeostatic manner to modulate excitation-inhibition imbalance in C. elegans. *PLoS Genet, 9*(5), e1003472. doi:10.1371/journal.pgen.1003472

57. Steiner, F. A., & Henikoff, S. (2014). Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *Elife, 3*, e02025. doi:10.7554/eLife.02025

58. Stinchcomb, D. T., Shaw, J. E., Carr, S. H., & Hirsh, D. (1985). Extrachromosomal DNA transformation of Caenorhabditis elegans. *Mol Cell Biol, 5*(12), 3484-3496. doi:10.1128/mcb.5.12.3484

59. Sutton, R. E., & Boothroyd, J. C. (1986). Evidence for trans splicing in trypanosomes. *Cell, 47*(4), 527-535.

60. Timmons, L., Court, D. L., & Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. *Gene, 263*(1-2), 103-112.

61. Timmons, L., & Fire, A. (1998). Specific interference by ingested dsRNA. *Nature, 395*(6705), 854. doi:10.1038/27579

62. Tops, B. B., Tabara, H., Sijen, T., Simmer, F., Mello, C. C., Plasterk, R. H., & Ketting, R. F. (2005). RDE-2 interacts with MUT-7 to mediate RNA interference in Caenorhabditis elegans. *Nucleic Acids Res, 33*(1), 347-355. doi:10.1093/nar/gki183

63. Towbin, B. D., Gonzalez-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., . . . Gasser, S. M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell, 150*(5), 934-947. doi:10.1016/j.cell.2012.06.051

64. Vandenberghe, A. E., Meedel, T. H., & Hastings, K. E. (2001). mRNA 5'-leader trans-splicing in the chordates. *Genes Dev, 15*(3), 294-303. doi:10.1101/gad.865401

65. Wang, G., & Reinke, V. (2008). A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol, 18*(12), 861-867. doi:10.1016/j.cub.2008.05.009

66. Wang, J. J., Cui, D. Y., Xiao, T., Sun, X., Zhang, P., Chen, R., . . . Huang, D. W. (2014). The influences of PRG-1 on the expression of small RNAs and mRNAs. *BMC Genomics, 15*, 321. doi:10.1186/1471-2164-15-321

67. Weick, E. M., & Miska, E. A. (2014). piRNAs: from biogenesis to function. *Development, 141*(18), 3458-3471. doi:10.1242/dev.094037

68. Yuen, K. W., Nabeshima, K., Oegema, K., & Desai, A. (2011). Rapid de novo centromere formation occurs independently of heterochromatin protein 1 in C. elegans embryos. *Curr Biol, 21*(21), 1800-1807. doi:10.1016/j.cub.2011.09.016

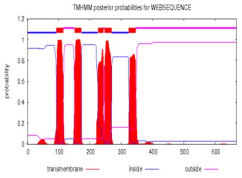
69. Zaphiropoulos, P. G. (2011). Trans-splicing in Higher Eukaryotes: Implications for Cancer Development? *Front Genet, 2*, 92. doi:10.3389/fgene.2011.00092

70. Zhang, C., Montgomery, T. A., Gabel, H. W., Fischer, S. E., Phillips, C. M., Fahlgren, N., . . . Ruvkun, G. (2011). mut-16 and other mutator class genes modulate 22G and 26G siRNA pathways in Caenorhabditis elegans. *Proc Natl Acad Sci U S A, 108*(4), 1201-1208. doi:10.1073/pnas.1018695108

71. Zhuang, J. J., Banse, S. A., & Hunter, C. P. (2013). The nuclear argonaute NRDE-3 contributes to transitive RNAi in Caenorhabditis elegans. *Genetics, 194*(1), 117-131. doi:10.1534/genetics.113.149765

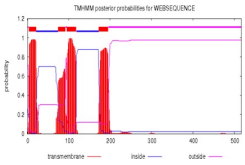
HAF-6A

inside	1	92
TMhelix	93	115
outside	116	150
TMhelix	151	173
inside	174	225
TMhelix	226	243
outside	244	246
TMhelix	247	269
inside	270	323
TMhelix	324	346
outside	347	668



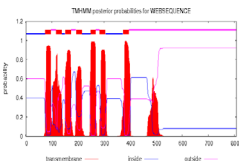
HAF-6B

outside	1	3
TMhelix	4	21
inside	22	74
TMhelix	75	92
outside	93	95
TMhelix	96	118
inside	119	172
TMhelix	173	195
outside	196	517



ABCB2/TAP1 isoform NP_000584.2

inside	1	73
TMhelix	74	96
outside	97	115
TMhelix	116	138
inside	139	150
TMhelix	151	173
outside	174	192
TMhelix	193	215
inside	216	248
TMhelix	249	268
outside	269	282
TMhelix	283	305
inside	306	372
TMhelix	373	395
outside	396	808



HAF-6A
ABCB2/TAP1 isoform 1

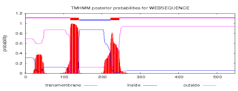
```
# Identity:       33/436 ( 7.6%)
# Similarity:     71/436 (16.3%)
# Gaps:          296/436 (67.9%)
# Score: 72.0
```

HAF-6B
ABCB2/TAP1 isoform 1

```
# Identity:       18/413 ( 4.4%)
# Similarity:     45/413 (10.9%)
# Gaps:          312/413 (75.5%)
# Score: 62.0
```

ABCB2/TAP1 isoform NP_001278951.1

outside	1	117
TMhelix	118	140
inside	141	222
TMhelix	223	245
outside	246	547



HAF-6A
ABCB2/TAP1 isoform 2

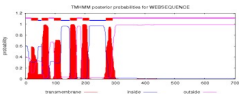
```
# Identity:       53/284 (18.7%)
# Similarity:   113/284 (39.8%)
# Gaps:          69/284 (24.3%)
# Score: 154.0
```

HAF-6B
ABCB2/TAP1 isoform 2

```
# Identity:       42/247 (17.0%)
# Similarity:     94/247 (38.1%)
# Gaps:          57/247 (23.1%)
# Score: 147.5
```

ABCB3/TAP2 isoform NP_000535.3

outside	1	19
TMhelix	20	42
inside	43	54
TMhelix	55	77
outside	78	96
TMhelix	97	119
inside	120	149
TMhelix	150	172
outside	173	191
TMhelix	192	214
inside	215	269
TMhelix	270	292
outside	293	703



HAF-6A
ABCB3/TAP2 isoform 1

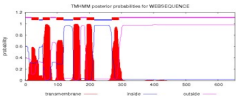
```
# Identity:       42/417 (10.1%)
# Similarity:     68/417 (16.3%)
# Gaps:          288/417 (69.1%)
# Score: 114.0
```

HAF-6B
ABCB3/TAP2 isoform 1

```
# Identity:       28/391 ( 7.2%)
# Similarity:     50/391 (12.8%)
# Gaps:          298/391 (76.2%)
# Score: 107.5
```

ABCB3/TAP2 isoform NP_061313.2

outside	1	19
TMhelix	20	42
inside	43	54
TMhelix	55	77
outside	78	96
TMhelix	97	119
inside	120	149
TMhelix	150	172
outside	173	191
TMhelix	192	214
inside	215	269
TMhelix	270	292
outside	293	653



HAF-6A
ABCB3/TAP2 isoform 2

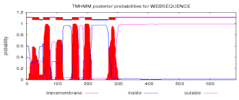
```
# Identity:       42/417 (10.1%)
# Similarity:     68/417 (16.3%)
# Gaps:          288/417 (69.1%)
# Score: 114.0
```

HAF-6B
ABCB3/TAP2 isoform 2

```
# Identity:       28/391 ( 7.2%)
# Similarity:     50/391 (12.8%)
# Gaps:          298/391 (76.2%)
# Score: 107.5
```

ABCB3/TAP2 isoform NP_001276972.1

outside	1	19
TMhelix	20	42
inside	43	54
TMhelix	55	77
outside	78	96
TMhelix	97	119
inside	120	149
TMhelix	150	172
outside	173	191
TMhelix	192	214
inside	215	269
TMhelix	270	292
outside	293	686



HAF-6A
ABCB3/TAP2 isoform 3

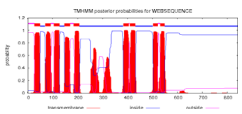
```
# Identity:      42/417 (10.1%)
# Similarity:    68/417 (16.3%)
# Gaps:         288/417 (69.1%)
# Score: 114.0
```

HAF-6B
ABCB3/TAP2 isoform 3

```
# Identity:      28/391 ( 7.2%)
# Similarity:    50/391 (12.8%)
# Gaps:         298/391 (76.2%)
# Score: 107.5
```

ABCB6 isoform 1 NP_005680.1

outside	1	26
TMhelix	27	49
inside	50	69
TMhelix	70	92
outside	93	104
TMhelix	105	127
inside	128	146
TMhelix	147	169
outside	170	183
TMhelix	184	206
inside	207	382
TMhelix	383	405
outside	406	408
TMhelix	409	431
inside	432	501
TMhelix	502	521
outside	522	530
TMhelix	531	548
inside	549	842



HAF-6A
ABCB6 isoform 1

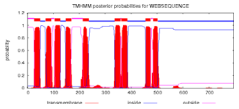
```
# Identity:      58/577 (10.1%)
# Similarity:   104/577 (18.0%)
# Gaps:         352/577 (61.0%)
# Score: 57.5
```

HAF-6B
ABCB6 isoform 1

```
# Identity:      46/563 ( 8.2%)
# Similarity:    82/563 (14.6%)
# Gaps:         386/563 (68.6%)
# Score: 57.5
```

ABCB6 isoform 2 NP_001336757.1

outside	1	26
TMhelix	27	49
inside	50	69
TMhelix	70	92
outside	93	104
TMhelix	105	127
inside	128	147
TMhelix	148	170
outside	171	213
TMhelix	214	236
inside	237	336
TMhelix	337	359
outside	360	362
TMhelix	363	385
inside	386	455
TMhelix	456	475
outside	476	484
TMhelix	485	502
inside	503	796



HAF-6A
ABCB6 isoform 2

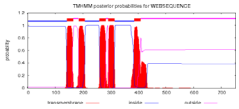
```
# Identity:      58/531 (10.9%)
# Similarity:   104/531 (19.6%)
# Gaps:         306/531 (57.6%)
# Score: 57.5
```

HAF-6B
ABCB6 isoform 2

```
# Identity:      46/517 ( 8.9%)
# Similarity:    82/517 (15.9%)
# Gaps:         340/517 (65.8%)
# Score: 57.5
```

ABCB7 isoform 1 NP_004290.2

inside	1	142
TMhelix	143	165
outside	166	184
TMhelix	185	207
inside	208	259
TMhelix	260	282
outside	283	291
TMhelix	292	314
inside	315	385
TMhelix	386	408
outside	409	753



HAF-6A
ABCB7 isoform 1

```
# Identity:      57/317 (18.0%)
# Similarity:   104/317 (32.8%)
# Gaps:         114/317 (36.0%)
# Score: 63.5
```

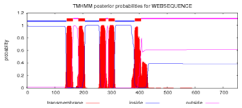
HAF-6B
ABCB7 isoform 1

```
# Identity:      47/293 (16.0%)
# Similarity:    86/293 (29.4%)
# Gaps:         128/293 (43.7%)
# Score: 62.5
```

ABCB7 isoform 2

inside	1	141
TMhelix	142	164
outside	165	183
TMhelix	184	206
inside	207	258
TMhelix	259	281
outside	282	290
TMhelix	291	313
inside	314	384
TMhelix	385	407
outside	408	752

NP_001258625.1



HAF-6A
ABCB7 isoform 2

```
# Identity:      57/317 (18.0%)
# Similarity:   104/317 (32.8%)
# Gaps:         114/317 (36.0%)
# Score: 63.5
```

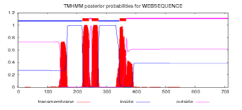
HAF-6B
ABCB7 isoform 2

```
# Identity:      47/293 (16.0%)
# Similarity:    86/293 (29.4%)
# Gaps:         128/293 (43.7%)
# Score: 62.5
```

ABCB7 isoform 3

inside	1	218
TMhelix	219	241
outside	242	250
TMhelix	251	273
inside	274	344
TMhelix	345	367
outside	368	712

NP_001258626.1



HAF-6A
ABCB7 isoform 3

```
# Identity:      36/286 (12.6%)
# Similarity:    63/286 (22.0%)
# Gaps:         169/286 (59.1%)
# Score: 41.5
```

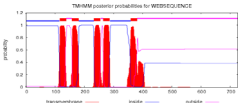
HAF-6B
ABCB7 isoform 3

```
# Identity:      36/224 (16.1%)
# Similarity:    63/224 (28.1%)
# Gaps:         107/224 (47.8%)
# Score: 41.5
```

ABCB7 isoform 4

inside	1	115
TMhelix	116	138
outside	139	157
TMhelix	158	180
inside	181	232
TMhelix	233	255
outside	256	264
TMhelix	265	287
inside	288	358
TMhelix	359	381
outside	382	726

NP_001258627.1



HAF-6A
ABCB7 isoform 4

```
# Identity:      57/317 (18.0%)
# Similarity:   104/317 (32.8%)
# Gaps:         114/317 (36.0%)
# Score: 63.5
```

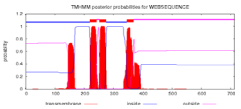
HAF-6B
ABCB7 isoform 4

```
# Identity:      47/293 (16.0%)
# Similarity:    86/293 (29.4%)
# Gaps:         128/293 (43.7%)
# Score: 62.5
```

ABCB7 isoform 5

inside	1	219
TMhelix	220	242
outside	243	251
TMhelix	252	274
inside	275	345
TMhelix	346	368
outside	369	713

NP_001258628.1



HAF-6A
ABCB7 isoform 5

```
# Identity:      36/286 (12.6%)
# Similarity:    63/286 (22.0%)
# Gaps:         169/286 (59.1%)
# Score: 41.5
```

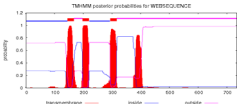
HAF-6B
ABCB7 isoform 5

```
# Identity:      36/224 (16.1%)
# Similarity:    63/224 (28.1%)
# Gaps:         107/224 (47.8%)
# Score: 41.5
```

ABCB8 isoform a

inside	1	144
TMhelix	145	167
outside	168	197
TMhelix	198	220
inside	221	293
TMhelix	294	316
outside	317	735

NP_001269220.1



HAF-6A
ABCB8 isoform a

```
# Identity:      77/255 (30.2%)
# Similarity:   108/255 (42.4%)
# Gaps:          84/255 (32.9%)
# Score: 357.0
```

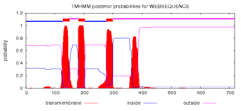
HAF-6B
ABCB8 isoform a

```
# Identity:      56/249 (22.5%)
# Similarity:    76/249 (30.5%)
# Gaps:         134/249 (53.8%)
# Score: 280.0
```

ABCB8 isoform b

inside	1	127
TMhelix	128	150
outside	151	180
TMhelix	181	203
inside	204	276
TMhelix	277	299
outside	300	718

NP_009119.2



HAF-6A
ABCB8 isoform b

```
# Identity:      77/255 (30.2%)
# Similarity:   108/255 (42.4%)
# Gaps:          84/255 (32.9%)
# Score: 357.0
```

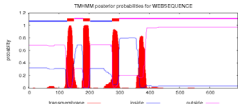
HAF-6B
ABCB8 isoform b

```
# Identity:      56/249 (22.5%)
# Similarity:    76/249 (30.5%)
# Gaps:         134/249 (53.8%)
# Score: 280.0
```

ABCB8 isoform c

inside	1	127
TMhelix	128	150
outside	151	180
TMhelix	181	203
inside	204	276
TMhelix	277	299
outside	300	693

NP_001269221.1



HAF-6A
ABCB8 isoform c

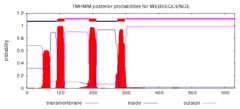
```
# Identity:      77/255 (30.2%)
# Similarity:   108/255 (42.4%)
# Gaps:          84/255 (32.9%)
# Score: 357.0
```

HAF-6B
ABCB8 isoform c

```
# Identity:      56/249 (22.5%)
# Similarity:    76/249 (30.5%)
# Gaps:         134/249 (53.8%)
# Score: 280.0
```

ABCB8 isoform d		
inside	1	92
TMhelix	93	115
outside	116	188
TMhelix	189	211
inside	212	274
TMhelix	275	297
outside	298	630

NP_001269222.1



HAF-6A
ABCB8 isoform d

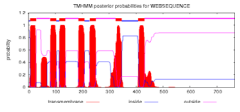
```
# Identity:      96/263 (36.5%)
# Similarity:   129/263 (49.0%)
# Gaps:          67/263 (25.5%)
# Score: 466.0
```

HAF-6B
ABCB8 isoform d

```
# Identity:      94/205 (45.9%)
# Similarity:   126/205 (61.5%)
# Gaps:          13/205 ( 6.3%)
# Score: 457.0
```

ABCB9 isoform 1		
inside	1	6
TMhelix	7	29
outside	30	80
TMhelix	81	103
inside	104	114
TMhelix	115	137
outside	138	185
TMhelix	186	208
inside	209	227
TMhelix	228	250
outside	251	323
TMhelix	324	346
inside	347	407
TMhelix	408	430
outside	431	766

NP_062571.1



HAF-6A
ABCB9 isoform 1

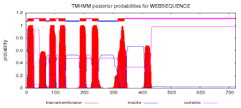
```
# Identity:      54/439 (12.3%)
# Similarity:   120/439 (27.3%)
# Gaps:         194/439 (44.2%)
# Score: 194.0
```

HAF-6B
ABCB9 isoform 1

```
# Identity:      42/430 ( 9.8%)
# Similarity:    96/430 (22.3%)
# Gaps:         238/430 (55.3%)
# Score: 166.0
```

ABCB9 isoform 2		
inside	1	6
TMhelix	7	27
outside	28	46
TMhelix	47	69
inside	70	81
TMhelix	82	104
outside	105	113
TMhelix	114	136
inside	137	185
TMhelix	186	208
outside	209	227
TMhelix	228	250
inside	251	316
TMhelix	317	339
outside	340	723

NP_062570.1



HAF-6A
ABCB9 isoform 2

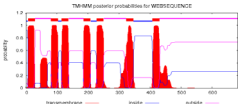
```
# Identity:      35/436 ( 8.0%)
# Similarity:    74/436 (17.0%)
# Gaps:         279/436 (64.0%)
# Score: 115.0
```

HAF-6B
ABCB9 isoform 2

```
# Identity:      23/427 ( 5.4%)
# Similarity:    50/427 (11.7%)
# Gaps:         323/427 (75.6%)
# Score: 87.0
```

ABCB9 isoform 4		
inside	1	6
TMhelix	7	29
outside	30	80
TMhelix	81	103
inside	104	114
TMhelix	115	137
outside	138	185
TMhelix	186	208
inside	209	227
TMhelix	228	250
outside	251	323
TMhelix	324	346
inside	347	407
TMhelix	408	430
outside	431	683

NP_982269.2



HAF-6A
ABCB9 isoform 4

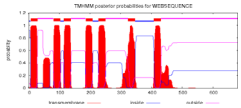
```
# Identity:      54/439 (12.3%)
# Similarity:   120/439 (27.3%)
# Gaps:         194/439 (44.2%)
# Score: 194.0
```

HAF-6B
ABCB9 isoform 4

```
# Identity:      42/430 ( 9.8%)
# Similarity:    96/430 (22.3%)
# Gaps:         238/430 (55.3%)
# Score: 166.0
```

ABCB9 isoform 5		
inside	1	6
TMhelix	7	29
outside	30	80
TMhelix	81	103
inside	104	114
TMhelix	115	137
outside	138	185
TMhelix	186	208
inside	209	227
TMhelix	228	250
outside	251	323
TMhelix	324	346
inside	347	407
TMhelix	408	430
outside	431	681

NP_001229943.1



HAF-6A
ABCB9 isoform 5

```
# Identity:      54/439 (12.3%)
# Similarity:   120/439 (27.3%)
# Gaps:         194/439 (44.2%)
# Score: 194.0
```

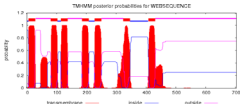
HAF-6B
ABCB9 isoform 5

```
# Identity:      42/430 ( 9.8%)
# Similarity:    96/430 (22.3%)
# Gaps:         238/430 (55.3%)
# Score: 166.0
```

ABCB9 isoform 6

inside	1	6
TMhelix	7	29
outside	30	80
TMhelix	81	103
inside	104	114
TMhelix	115	137
outside	138	185
TMhelix	186	208
inside	209	227
TMhelix	228	250
outside	251	323
TMhelix	324	346
inside	347	407
TMhelix	408	430
outside	431	703

NP_001229942.1



HAF-6A
ABCB9 isoform 6

```
# Identity:      54/439 (12.3%)
# Similarity:   120/439 (27.3%)
# Gaps:         194/439 (44.2%)
# Score: 194.0
```

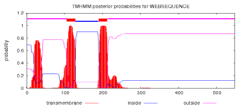
HAF-6B
ABCB9 isoform 6

```
# Identity:      42/430 ( 9.8%)
# Similarity:    96/430 (22.3%)
# Gaps:         238/430 (55.3%)
# Score: 166.0
```

ABCB9 isoform X2

outside	1	105
TMhelix	106	128
inside	129	189
TMhelix	190	212
outside	213	548

XP_005253615.1



HAF-6A
ABCB9 isoform X2

```
# Identity:      44/275 (16.0%)
# Similarity:    99/275 (36.0%)
# Gaps:          84/275 (30.5%)

# Score: 159.5
```

HAF-6B
ABCB9 isoform X2

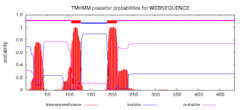
```
# Identity:      42/212 (19.8%)
# Similarity:   100/212 (47.2%)
# Gaps:          20/212 ( 9.4%)

# Score: 163.0
```

ABCB9 isoform X3

outside	1	105
TMhelix	106	128
inside	129	189
TMhelix	190	212
outside	213	485

XP_024304676.1



HAF-6A
ABCB9 isoform X3

```
# Identity:      44/275 (16.0%)
# Similarity:    99/275 (36.0%)
# Gaps:          84/275 (30.5%)

# Score: 159.5
```

HAF-6B
ABCB9 isoform X3

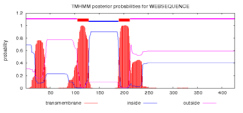
```
# Identity:      42/212 (19.8%)
# Similarity:   100/212 (47.2%)
# Gaps:          20/212 ( 9.4%)

# Score: 163.0
```

ABCB9 isoform X4

outside	1	105
TMhelix	106	128
inside	129	189
TMhelix	190	212
outside	213	426

XP_024304677.1



HAF-6A
ABCB9 isoform X4

```
# Identity:      44/275 (16.0%)
# Similarity:    99/275 (36.0%)
# Gaps:          84/275 (30.5%)
# Score: 159.5
```

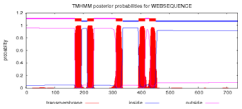
HAF-6B
ABCB9 isoform X4

```
# Identity:      42/212 (19.8%)
# Similarity:   100/212 (47.2%)
# Gaps:          20/212 ( 9.4%)
# Score: 163.0
```

ABCB10

outside	1	170
TMhelix	171	193
inside	194	213
TMhelix	214	236
outside	237	313
TMhelix	314	336
inside	337	392
TMhelix	393	415
outside	416	429
TMhelix	430	452
inside	453	738

NP_036221.2



HAF-6A
ABCB10

```
# Identity:      68/293 (23.2%)
# Similarity:   125/293 (42.7%)
# Gaps:          50/293 (17.1%)
# Score: 272.
```

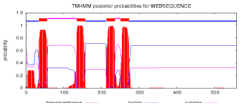
HAF-6B
ABCB10

```
# Identity:      55/282 (19.5%)
# Similarity:   101/282 (35.8%)
# Gaps:          90/282 (31.9%)
# Score: 243.0
```

ABCB10 isoform X1

inside	1	34
TMhelix	35	57
outside	58	134
TMhelix	135	157
inside	158	213
TMhelix	214	236
outside	237	250
TMhelix	251	273
inside	274	559

XP_011542437.1



HAF-6A
ABCB10 isoform X1

```
# Identity:      65/293 (22.2%)
# Similarity:   122/293 (41.6%)
# Gaps:          59/293 (20.1%)
# Score: 269.5
```

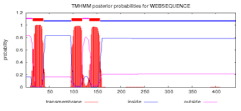
HAF-6B
ABCB10 isoform X1

```
# Identity:      55/273 (20.1%)
# Similarity:   101/273 (37.0%)
# Gaps:          81/273 (29.7%)
# Score: 243.0
```

ABCB10 isoform X2

outside	1	14
TMhelix	15	37
inside	38	96
TMhelix	97	119
outside	120	133
TMhelix	134	156
inside	157	442

XP_011542438.1



HAF-6A
ABCB10 isoform X2

```
# Identity:      33/293 (11.3%)
# Similarity:    59/293 (20.1%)
# Gaps:         176/293 (60.1%)
# Score: 125.0
```

HAF-6B
ABCB10 isoform X2

```
# Identity:      33/231 (14.3%)
# Similarity:    59/231 (25.5%)
# Gaps:         114/231 (49.4%)
# Score: 125.0
```

Supplementary information 1 : Amino acid similarity between human transporters and *C. elegans* HAF-6 protein. (Top two panels) Prediction of transmembrane helices in *Caenorhabditis elegans* HAF-6 protein isoforms using TMHMM, based on a hidden Markov model (http://www.cbs.dtu.dk/services/TMHMM-2.0/). (Krogh, Larsson et al. 2001). (Remainder of panels) Regions highlighted in yellow were used to calculate % amino acid identity with human half transporters from the B-subfamily using MUSCLE (multiple sequence comparison by log-expectation) (Edgar 2004) and (Edgar 2004). Hydrophobicity plots of human half transporter members of the B subfamily are depicted at left, with predicted transmembrane regions used in alignments highlighted in yellow. Alignments were made to *C. elegans* HAF-6A isoform (middle section) and to the HAF-6B isoform (right section). An overall alignment score, which includes a maximized SP score (sum of all pairs) that is NP-complete is produced by MUSCLE, with larger values representing greater similarity. For reference, a self-alignment of HAF-6A transmembrane regions produces a score of 1269; a self-alignment of HAF-6B transmembrane regions produces a score of 946. The best alignment was observed for human ABCB8 isoform d with a score of 466 for HAF- 6A alignment; 457 for HAF-6B. (Gap penalty: 10.0; Extend penalty 0.5)