Control of *Caenorhabditis elegans* sexual development by TRA-1 and its targets

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Matthew Berkseth

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

David Zarkower, Ph.D., Advisor

December 2013

Acknowledgments

I thank my advisor David Zarkower for his guidance and support. I thank Vivian Bardwell, Connie Corcoran, Micah Gearhart, Andrea Kalis, Tony Krentz, Mary Kroetz, Robyn Lindeman, Clint Matson, Anna Minkina, Mark Murphy, and Teng Zhang for helpful scientific discussions and advice. I thank David Greenstein, Tom Neufeld, Ann Rougvie, Michel Sanders, and Howard Towle for serving on my thesis committee. Most importantly I thank my family, especially my wife Cathy, for their love and support throughout my graduate career.

Table of Contents

Acknowledgments **i**

Table of Contents **ii**

List of Tables **iii**

List of Figures **iv**

Chapter 1: Background **1**

Introduction 2

Animal sex-determination mechanisms are diverse 2

Common themes in sexual development 5

Sex determination in *C. elegans* 7

Sexual differentiation downstream of *tra-1* 16

Identifying regulators of sexual differentiation; a new approach 21

Chapter 2: TRA-1 ChIP-seq reveals regulators of sexual differentiation and multilevel feedback in nematode sex determination **23**

Summary 24

Introduction 24

Results 29

Discussion 76

Materials and Methods 81

Chapter 3: Conclusions and Future Directions **88**

Introduction 89

Important questions concerning TRA-1 regulation of sexual development 89

New approaches to understanding TRA-1 function 103

Conclusion 108

Bibliography **108**

List of Tables

Table 2.1: 184 TRA-1 binding sites in the *C. elegans* genome. 32

Table 2.2: 45 germ line-specific TRA-1 binding sites. 52

Table 2.3: 78 TRA-1 binding sites in the *C. briggsae* genome. 68

Table 2.4: 19 conserved TRA-1 binding sites. 73

List of Figures

Figure 1.1: Somatic sex determination in *C. elegans*. 9

Figure 1.2: Sexual dimorphisms in *C. elegans*. 19

Figure 2.1: Model for genetic control of *C. elegans* somatic sex determination. 27

Figure 2.2: Analysis of TRA-1 binding sites in the *C. elegans* genome. 48

Figure 2.3: Newly identified TRA-1 target genes exhibit male-specific expression patterns and reduction-of-function phenotypes. 56

Figure 2.4: TRA-1 binding sites near four heterochronic regulatory genes drive male-specific expression. 59

Figure 2.5: TRA-1 feeds back on the sex-determination pathway at multiple levels. 63

Figure 2.6: TRA-1 regulates *fem-3* expression. 65

Figure 2.7: Model for TRA-1 regulation of nematode sexual development. 79

Figure 2.8: Characterization of anti-TRA-1 antibody UMN163 used in this study. 86

Figure 3.1: Motif searching identifies a GAGA motif enriched at TRA-1 binding sites. 96

Figure 3.2: A comparison of TRA-1 occupancy and chromatin mark occupancy. 99

Figure 3.3: A comparison of TRA-1 occupancy and transcription factor occupancy. 101

Chapter 1

Background

Introduction

The vast majority of multicellular animals reproduce sexually, using the fusion of two gametes into a zygote to combine the genetic material of two parents into a single offspring. In most cases, the two distinct gametes, the sperm and the egg, are produced by two distinct sexes, the male and the female. These two sexes undergo distinct programs of sexual development, which are necessary to generate the two gamete types as well as the anatomical, physiological, and behavioral differences required for successful sexual reproduction.

Sexual development can be broken down into two sequential steps. First, initial sex-determining signals, either genetic or environmental, are interpreted by a sex-determination pathway that sets the global sex of the animal. Then, sexual differentiation manifests this fate through sex-specific developmental events, such as cell divisions, cell migrations, and cell deaths that create the dimorphisms we observe between the sexes. This chapter will review what is known about sexual development in diverse animal lineages, examine this process in detail in the nematode worm *Caenorhabditis elegans*, highlight how the regulation of sexual differentiation represents a major gap in current knowledge of sexual development, and finally introduce a biochemical approach aimed at improving our understanding of this fundamental biological process.

Animal sex-determination mechanisms are diverse

While sexual development is widespread in animals, the molecular mechanisms used to regulate this fundamental process are surprisingly diverse. Initial sex-determining signals include a wide variety of genetic and environmental cues which are interpreted by sex-

determination pathways that employ a wide variety of signaling and gene regulatory molecules.

Initial sex-determination cues can be broken down into two types, environmental and genetic. For animals with genetic sex determination, the initial sex-determining cue is a heritable genetic factor. Perhaps the most familiar example of this type of system is the XY scheme utilized by humans and most other mammals. In this system, there are two distinct sex chromosomes, the X and Y, and an individual's set of sex chromosomes determines whether it will develop as male or female. XY individuals develop as males due to the presence of a Y-encoded gene, *SRY* (Sex-determining region on Y) in humans (Berta et al., 1990; Sinclair et al., 1990), which initiates male-specific testis development in a genetically dominant fashion. In mammals, hormonal signaling from the testis directs male development throughout the animal, and individuals that lack male-specific hormonal signaling develop as females. Thus mammalian sex determination is cell non-autonomous, with gonad sexual fate determining sex throughout the body. In some other species, such as the fruit fly *Drosophila*, the process is at least partially cell autonomous, with each cell determining sexual fate on its own (Spencer, 1927). In the XY system, offspring receive one sex chromosome from their father (either his X or Y) and one from their mother (one of her two Xes), so this system results in a roughly even sex ratio.

In the ZW system, homogametic individuals (ZZ) develop as males, while heterogametic individuals (ZW) develop as females. This can be due to a dominantly female-promoting gene encoded on the W chromosome, such as in some amphibians (Yoshimoto et al., 2008), or due to the dosage effects of Z-linked sex-determining genes, as in birds (Smith

et al., 2009). In this system, all offspring receive a Z from their father and either a Z or a W from their mother, which again results is a roughly even sex ratio.

Genetic sex-determination systems are not always based on distinct sex chromosomes. For example, some insects have a single sex chromosome, which determines sex based on its copy number relative to the autosomes, such that animals with two X chromosomes develop as female, while animals with a single X develop as males. Bees utilize copy number of the entire genome to determine sex; unfertilized haploid eggs develop as males, while fertilized diploid eggs develop as females (Flanders, 1946). These systems can result in sex ratios that are far from even, and that can vary based on the method of reproduction, allowing these species to modify their sex ratios in response to environmental conditions or population dynamics.

Chromosome-level genetic sex-determining mechanisms, such as the XY, ZW, and XO systems discussed above, include entire chromosomes that are present at different copy numbers in the two sexes. A two-fold expression difference in all the genes on an entire chromosome would be lethal in most cases, but such organisms have evolved dosage compensation mechanisms to equalize the expression of sex-linked genes between the sexes. In mammals, dosage compensation is achieved via complete inactivation of one of the two X chromosomes in female animals (Plath, Mlynarczyk-Evans, Nusinow, & Panning, 2002). In *Drosophila*, males double the rate of transcription from their single X chromosome (Meller & Kuroda, 2002), while *C. elegans* hermaphrodites halve the transcriptional output of both of their X chromosomes (Meyer, 2005).

Environmental cues can also serve as initial sex-determining signals. Many reptiles use incubation temperature as such a signal, with eggs incubated a certain temperatures developing as males, and eggs raised at other temperatures developing as females (Bull & Vogt, 1979). The amphipod *Gammarus duebeni* uses photoperiod to influence sex, with individuals reared during longer days developing mostly as males, and individuals reared during shorter days developing mostly as females (Bulnheim, 1966). Some fish rely on social cues, with the dominant individual in a group developing as female while subordinate animals develop as male, or vice versa (Robertson, 1972). In these examples, sex ratios obviously vary greatly based on environmental conditions, often in ways that confer a clear adaptive benefit.

These varied initial sex-determining signals are interpreted by genetic sex-determination pathways that also show great diversity, utilizing a wide variety of signaling mechanisms and molecules to set the sex of the animal. The *Drosophila* sex-determination pathway relies heavily of regulators of alternative splicing (Salz, 2011), while transcription factors and hormonal signaling figure prominently in mammalian systems (Sekido & Lovell-Badge, 2013). *C. elegans* utilizes a wide variety of intracellular and extracellular signaling molecules in sex determination, which will be described in detail later in this chapter. Sex-determination mechanisms evolve rapidly, and there seems to be little restriction on what types of genes can be recruited into a sex-determining role, resulting in a remarkable diversity of sexual regulators.

Common themes in sexual development

Despite all this diversity, certain conserved features of sexual development have emerged. One common theme is that these diverse systems usually culminate in a single global sexual regulator that translates the upstream sex-determination pathway into a binary male-female decision. While these terminal sexual regulators are diverse, they typically encode a transcriptional factor that directs the coordinated execution of sex-specific differentiation events via direct transcriptional regulation of target genes.

One particular group of transcription factors, the DM-domain family, has emerged as a conserved component of sexual development in diverse animal lineages (Raymond et al., 1998). DM-domain proteins play roles in the sexual development of all organisms in which they have been comprehensively examined, and may represent an ancestral family of sexual regulators. Members of this family have been shown to function at all levels of sexual development. The Y-linked DMY (DM domain gene on Y) of *Medaka* (Matsuda et al., 2002) and the W-encoded DMW of some amphibians (Yoshimoto et al., 2008) function as initial sex-determining cues. In *Drosophila*, the DM domain protein DSX (Doublesex) lies downstream of initial sex-determining cues and functions as the terminal regulator of sex, directing sex-specific differentiation throughout the animal (Hildreth, 1965). *C. elegans* has several members of this gene family, and though none of them function in sex determination, at least three of them, *mab-3*, *dmd-3*, and *mab-23*, play important roles in the regulation of particular sex-specific differentiation events (Lints & Emmons, 2002; Mason, Rabinowitz, & Portman, 2008; Yi, Ross, & Zarkower, 2000).

It has been hypothesized that while upstream components of sex-determination pathways may be easily gained and lost through evolution, downstream components such as

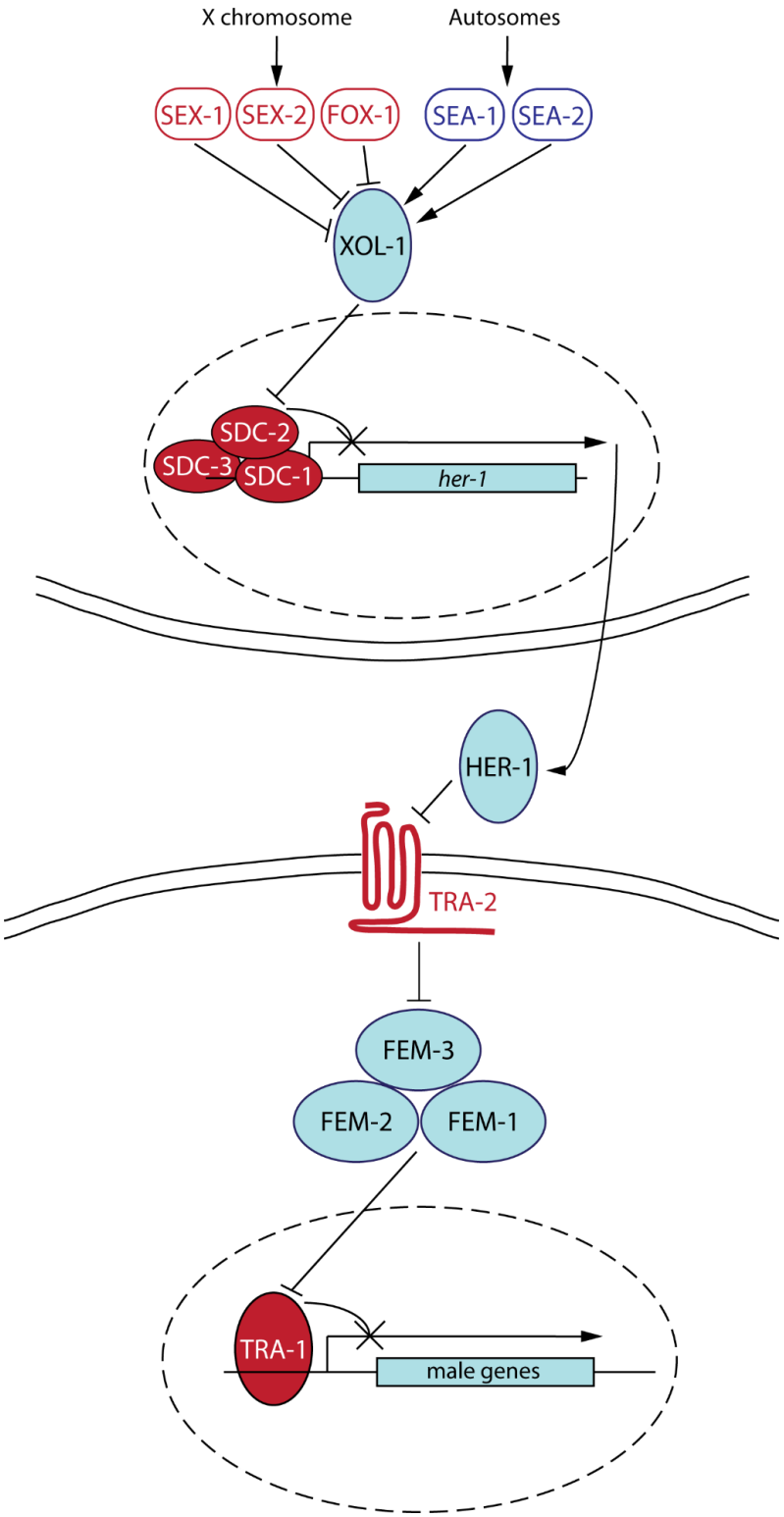
regulators of sexual differentiation are more constrained and therefore more likely to be conserved (Wilkins, 1995). The DM-domain proteins are the only known example of broadly conserved sexual regulators, but the degree to which other downstream sexual regulators may be conserved is largely unknown. This is due in part to the fact that few such regulators have been identified, even in species in which sex-determination mechanisms are fairly well understood, limiting the opportunity to systematically compare regulators of sexual differentiation in diverse lineages.

Sex determination in *C. elegans*

C. elegans has perhaps the best-understood sex-determination pathway of any animal, making it an excellent system in which to study sexual development. *C. elegans* determines sex genetically with an XO system (Nigon, 1951). XX animals develop as hermaphrodites, which are somatically female but capable of producing both eggs and sperm. *C. elegans* hermaphrodites produce and store a limited number of sperm in late larval development, allowing them to produce a few hundred progeny via self-fertilization as adults without receiving sperm from a male. XO animals develop as males, which produce only sperm and must mate with a hermaphrodite in order to reproduce. All hermaphrodite sperm contain an X chromosome, with rare exceptions (~1/500) lacking an X due to errors in meiosis. Since males are XO, only half of their sperm contain an X, so half of the progeny resulting from a mating will also be male. This results in a system in which males are produced quite rarely in wild-type populations of purely hermaphrodite animals, but at close to 50% in populations where a significant number of males are already present.

The sex-chromosome complement is interpreted by a global sex-determination pathway, whose activity determines the sexual fate of all somatic cells in the animal (Figure 1.1). Core members of this pathway were identified in genetic screens for sex reversals and other major defects in sexual development (J. Hodgkin, 1980; J. A. Hodgkin & Brenner, 1977), and subsequent epistasis analysis has placed these core genes in a linear pathway with alternating male-promoting and female-promoting activities, with each gene functioning to repress the gene immediately downstream of it (J Hodgkin, 1983; J. Hodgkin, 1984; Plenefisch, DeLong, & Meyer, 1989). This results in a back-and-forth regulatory cascade of alternating on and off genes which is in opposite phase in the two sexes (Figure 2.1).

Figure 1.1. Somatic sex determination in *C. elegans*. The sex-determination pathway in *C. elegans* in comprised of a negative regulatory cascade that transduces the X:Autosome ratio into the activity of the global master regulator TRA-1, a transcription factor that directs all aspects of sexual differentiation in the worm, primarily by repressing the transcription of genes that promote male development. Female promoting factors are shown in red, male promoting factors are shown in blue.


X chromosome
Autosomes
SEX-1
SEX-2
FOX-1
SEA-1
SEA-2
XOL-1
SDC-2
SDC-3
SDC-1
her-1
HER-1
TRA-2
FEM-3
FEM-2
FEM-1
TRA-1
male genes

The X:A ratio sets the activity of *xol-1*

The most upstream gene in this pathway is *xol-1* (XO lethal-1) (Miller, Plenefisch, Casson, & Meyer, 1988; Rhind, Miller, Kopczynski, & Meyer, 1995), which is a male-promoting gene encoding an acidic protein structurally related to members of the GHMP family of small molecule metabolic kinases (Luz et al., 2003). XOL-1, however, does not bind ATP and its molecular function remains unknown. *xol-1* activity is regulated by the X:A ratio in a complex manner, involving several X-signaling elements (XSEs) that repress *xol-1* activity acting in opposition to several autosomal-signaling elements (ASEs) that promote *xol-1* activity.

Genetic experiments indicate that there are at least four distinct XSEs (Akerib & Meyer, 1994). Three of these XSEs have been cloned and characterized and have been shown to repress *xol-1* activity at two distinct levels. The nuclear hormone receptor SEX-1 (Signal element on X-1) (Carmi, Kopczynski, & Meyer, 1998) and the ONECUT homeodomain protein CEH-39 (*C. elegans* homeobox-39) (Gladden & Meyer, 2007) repress *xol-1* at the transcriptional level, by directly binding to multiple sites in the *xol-1* promoter. The RNA binding protein FOX-1 (Feminizing gene on X-1) directs alternative splicing of the *xol-1* transcript that results in the formation of an inactive splice variant (J. Hodgkin, Zellan, & Albertson, 1994; Nicoll, Akerib, & Meyer, 1997; Skipper, Milne, & Hodgkin, 1999). The two ASEs that have been characterized, the T-box transcription factor SEA-1 (Signal element on autosome-1) (Powell, Jow, & Meyer, 2005) and the zing-finger protein SEA-2 (Huang, Zhang, & Zhang, 2011), both function to activate *xol-1* transcription, again by directly binding to the *xol-1* promoter. Thus the *xol-1* promoter serves as a site of

competition between X-linked transcriptional repressors and autosome-linked transcriptional activators (Farboud, Nix, Jow, Gladden, & Meyer, 2013).

This system is extremely sensitive to altered X:A ratios. In the wild type situation, it distinguishes between X:A ratios of 0.5 (1X:2A) and 1.0 (2X:2A), but experiments with polyploid animals have shown that it can reliably distinguish ratios of 0.67 (2X:3A), and 0.75 (3X:4A) (Madl & Herman, 1979). This sort of ratiometric mechanism of sex determination is not universal in all species with XO sex-determination systems. For example, *Drosophila* appears to rely primarily on X chromosome counting (XSEs), without significant contributions from autosomal chromosomes (ASEs), to determine sex (Erickson & Quintero, 2007). The relative contributions of XSEs and ASEs in other species with XO sex-determination systems are currently unknown.

xol-1 **represses the DCC via repression of** ***sdc-2***

Since males and hermaphrodites have different numbers of X chromosomes, a dosage compensation mechanism has evolved to equalize the expression of X-linked genes between the sexes. The upstream portion of the sex-determination pathway, from *xol-1* to *sdc-2*, regulates both sex determination and dosage compensation, thus ensuring that both processes are functioning in the same mode.

When active, as in wild-type XO animals, *xol-1* functions to repress *sdc-2* (sex determination and dosage compensation defect-2), leading to XX-specific expression of the SDC-2 protein (Dawes et al., 1999), although the molecular nature of this regulatory interaction is unknown. SDC-2 forms a dosage compensation complex (DCC) with a pair of zinc-finger proteins called SDC-1 and SDC-3. The DCC, which is a variant form of

the condensin complex and is normally active only in XX animals, functions to reduce the expression level of all X-linked genes by 2-fold, thus equalizing the expression of X-linked genes in XX and XO animals (Meyer, 2005). It accomplishes this by directly binding to both X chromosomes and reducing their transcriptional output to half that of non-DCC bound X chromosomes.

SDC-2 is the only sex-specifically expressed member of this complex, and ectopic expression of SDC-2 in XO animals is sufficient to initiate dosage compensation of the single X chromosome, causing lethality (Dawes et al., 1999). Thus *xol-1* mediated repression of *sdc-2* is responsible for the sex-specific activity of the entire complex.

A DCC containing SDC-2 represses *her-1*

The DCC also functions to repress a single autosomally-encoded gene, *her-1* (hermaphroditization of XO-1), which is required for normal sex determination (J. Hodgkin, 1980). DCC repression of *her-1* occurs through direct binding to the *her-1* locus and transcriptional repression, but is mechanistically distinct from X chromosome repression. The DCC is targeted to the *her-1* locus by SDC-3 rather than SDC-2, and the *her-1*-bound complex does not include the DPY-21 subunit found in the X chromosome-bound complex (Yonker & Meyer, 2003). The complex induces 20-fold transcriptional repression of *her-1* (Trent et al., 1991), much stronger than the 2-fold repression induced on the X chromosomes.

her-1* represses *tra-2

her-1 encodes a small secreted protein (Perry et al., 1993), and functions in a cell non-autonomous manner to promote male development (Hunter & Wood, 1992). This cell non-autonomous step may help coordinate the sexual fate of cells throughout the animal, helping to avoid reproductively incompetent intersex states (Hunter & Wood, 1992). The only described function of *her-1* is to repress *tra-2* (transformer-2), which encodes the patched-related transmembrane protein TRA-2A (Kuwabara, Okkema, & Kimble, 1992). HER-1 binds directly to the transmembrane protein TRA-2A (Hamaoka, Dann, Geisbrecht, & Leahy, 2004; Kuwabara, 1996), preventing it from performing its female-promoting function.

tra-2 represses the *fem* genes

TRA-2A represses the function of three *fem* (feminization of XX and XO) genes, *fem-1, -2,* and *-3*. Similar expression of *fem* gene mRNAs in both sexes suggests that their repression by TRA-2A is post transcriptional (Ahringer, Rosenquist, Lawson, & Kimble, 1992; Gaudet, VanderElst, & Spence, 1996; Pilgrim, McGregor, Jackle, Johnson, & Hansen, 1995), and TRA-2A has been shown to physically interact with FEM-3 (Mehra, Gaudet, Heck, Kuwabara, & Spence, 1999), suggesting that a protein-protein interaction between TRA-2A and FEM-3 is likely responsible for this repression.

The *fem* genes repress *tra-1*

The major function of the *fem* genes is the repression of *tra-1*, which occurs primarily at the post-translational level. FEM-1, an ankyrin repeat containing protein (Spence, Coulson, & Hodgkin, 1990), FEM-2, a serine/threonine phosphatase (Chin-Sang & Spence, 1996; Pilgrim et al., 1995), and the novel protein FEM-3 (Ahringer et al., 1992)

form a complex with the E3 ubiquitin ligase CUL-2 (Cullin-2) that targets the TRA-1 protein for proteasomal degradation (Starostina et al., 2007).

tra-1 regulates sex throughout the animal

tra-1 is the most downstream gene in the sex-determination pathway, and functions as the terminal global sexual regulator in *C. elegans* (J. Hodgkin, 1987). Like most terminal sexual regulators, *tra-1* encodes a transcription factor, specifically a C2H2 zinc-finger transcription factor related to the vertebrate GLI proteins (Zarkower & Hodgkin, 1992). The *tra-1* locus produces two transcripts, encoding a larger protein called TRA-1A and a smaller protein known as TRA-1B. TRA-1B does not bind DNA and has no described biological activity (Zarkower & Hodgkin, 1993). TRA-1A is C-terminally cleaved to generate the functional transcriptional regulator (Schvarzstein & Spence, 2006), which will simply be called TRA-1 in this thesis. Genetic experiments have shown that *tra-1* activity is both necessary and sufficient to promote female somatic development, indicating that TRA-1 and its downstream targets are responsible for generating all sexual dimorphism in the worm (J. Hodgkin, 1987).

A number of other players impact the core sex-determination pathway

In addition to the core pathway members described above, a number of other genes have been shown to impact the sex-determination decision. Many of these genes were identified in the same genetic screens that identified the core components, or in follow-up screens looking for enhancers or suppressors of particular sex-determination alleles. One such gene encodes the calpain protease TRA-3 (Barnes & Hodgkin, 1996), which functions with *tra-2* to repress the *fem*s, likely by cleaving TRA-2A to generate an

intracellular fragment with feminizing activity (Sokol & Kuwabara, 2000). Another encodes the RNA-binding protein SUP-26 (Suppresor-26), which represses *tra-2* at the translational level by binding to a specific site in the 3' UTR of the *tra-2* mRNA. These genes and other peripheral components of the sex-determination pathway may serve to stabilize the pathway, ensuring coordinated sexual development throughout the entire animal and the avoidance of deleterious intersex states.

The core sex-determination pathway also functions in the germ line, although with the addition of a few features not present in the soma. The pathway functions through the *fog* (feminization of germ line) genes in the germ line, which function downstream of *tra-1* to control germ cell sexual fate but play no role in somatic sex determination. In addition, the *fem* genes appear to act both upstream and downstream of *tra-1* to control germ line sex (J. Hodgkin, 1986), while they appear to function solely through *tra-1* in the soma. Finally, germ line sex switching in hermaphrodites requires translational repression of *tra-2* to allow for hermaphrodite spermatogenesis (Doniach, 1986; Goodwin, Okkema, Evans, & Kimble, 1993; Jan, Motzny, Graves, & Goodwin, 1999), as well as translational repression of *fem-3* for sustained oogenesis (Ahringer & Kimble, 1991; Ahringer et al., 1992; Rosenquist & Kimble, 1988).

Sexual differentiation downstream of *tra-1*

As described above, *C. elegans* naturally occurs as two sexes, the XX hermaphrodite and the XO male. These two sexes exhibit a great deal of sexual dimorphism, much of which has been well characterized at the anatomic level. The complete cell lineage has been mapped for both the hermaphrodite and the male, meaning sexual dimorphisms can be

described in terms of specific cells and the presence or absence of specific cell division and cell death events (Kimble & Hirsh, 1979; Sulston, Albertson, & Thomson, 1980; Sulston & Horvitz, 1977; Sulston, Schierenberg, White, & Thomson, 1983). Many sexual dimorphisms are readily apparent (Figure 1.2), including differences in overall size, gonad shape, and tail morphology. Beyond these obvious dimorphisms are a multitude of more subtle sex differences; roughly a third of somatic cells in the adult exhibit some form of observable sexual dimorphism.

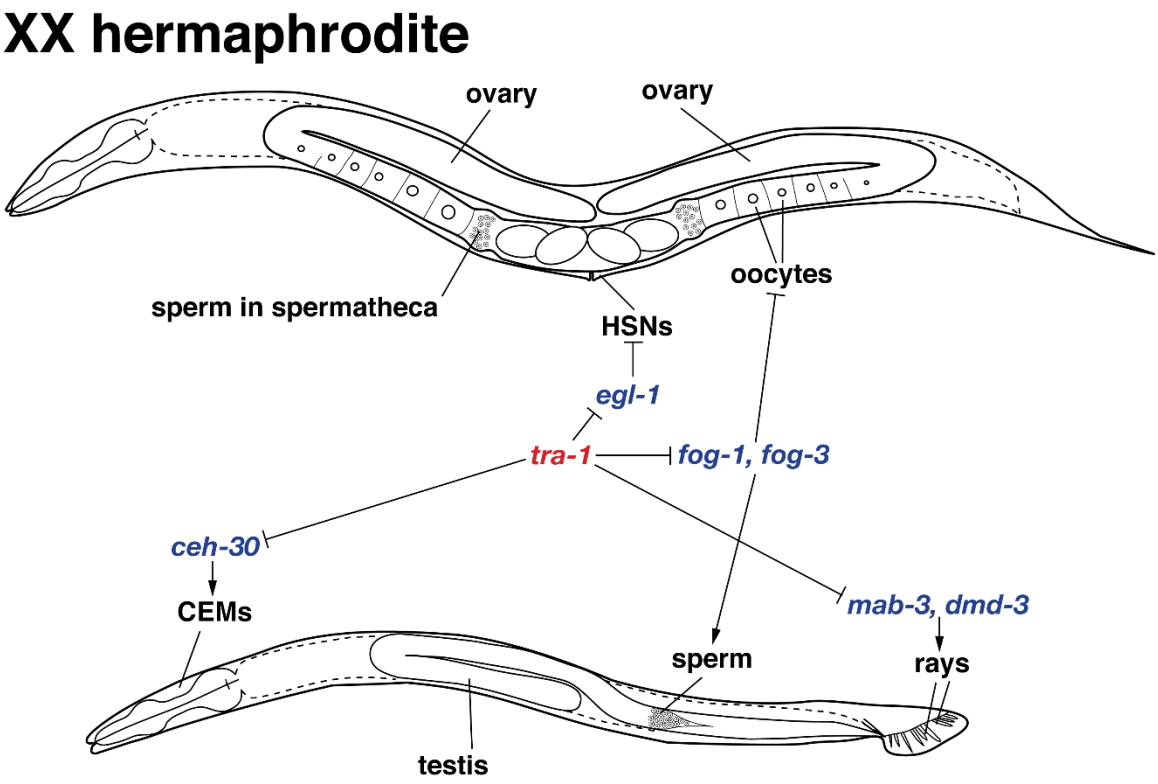
Genetic screens have identified genes that regulate a subset of these sexual dimorphisms (J. Hodgkin, 1983a). Some of these regulators of sexual differentiation were later shown to be direct targets of TRA-1, directly connecting the global sex-determination pathway with particular sex-specific differentiation events. However, for many sexual dimorphisms no genetic regulators have been identified, and for others the connection between TRA-1 and their known regulators remains unclear.

All known TRA-1 target genes regulate sexual differentiation in some way. The roles these genes play in sexual development were first identified based on observable loss-of-function phenotypes, and their direct regulation by TRA-1 was determined later. Two of these genes, *fog-1* and *fog-3*, function in the germ line to promote spermatogenesis (P. Chen & Ellis, 2000; Jin, Kimble, & Ellis, 2001). The other known TRA-1 targets direct specific somatic sexual differentiation events. *egl-1* promotes the male-specific cell death of the hermaphrodite specific HSN neurons (Conradt & Horvitz, 1999), while *ceh-30* promotes the male-specific survival of the CEM neurons (Peden, Kimberly, Gengyo-Ando, Mitani, & Xue, 2007). *mab-3* functions to repress yolk-protein synthesis in the

male (Shen & Hodgkin, 1988; Yi et al., 2000), and also functions with *dmd-3* to promote male-specific differentiation in the tail (Mason et al., 2008; Yi et al., 2000). All of these targets promote specific aspects of male development, suggesting that TRA-1 promotes female development primarily by repressing male-promoting genes.

While simple genetic screens have been successful in identifying a subset of sexual regulators, the genetic control of a great deal of the sexual dimorphism in the worm remains uncharacterized, highlighting the need for new approaches aimed at identifying regulators of sexual differentiation.

Figure 1.2. Sexual dimorphisms in *C. elegans*. *C. elegans* exists as two distinct sexes, which exhibit extensive sexual dimorphism. A few of the most obvious dimorphisms and the *tra-1* target genes that regulate their sex specific development are highlighted.


XX hermaphrodite
ovary
ovary
oocytes
sperm in spermatheca
HSNs
egl-1
tra-1
fog-1, fog-3
ceh-30
CEMs
mab-3, dmd-3
sperm
rays
testis
XO male

Identifying regulators of sexual differentiation; a new approach

Biochemical techniques can provide a complementary approach for the identification of genes that function in the regulation of sexual differentiation. They have the advantage of being able to identify certain types of genes that cannot be found with simple genetic screens, such as genes that lack strong loss-of-function phenotypes due to redundancy, or pleiotropic genes with very strong loss-of-function phenotypes in early development that mask a later role in sexual development.

Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) is one such biochemical approach that can be used when an important transcriptional regulator, such as TRA-1, is known. This approach identifies regions of the genome where that protein binds and implicates nearby genes as putative targets worthy of closer examination. As described above, several known TRA-1 targets function as regulators of sexual differentiation, and the *tra-1* null phenotype of nearly complete sex reversal suggests that essentially every sexual differentiation event involves a TRA-1 target at some level. Thus identifying the complete set of TRA-1 targets would presumably identify regulators of sexual differentiation events throughout the worm, providing a valuable list of candidate genes to examine for roles in sexual development, as well as a paradigm for understanding how sex-determination pathways regulate sexual differentiation on the scale of an entire organism.

The following chapters will describe efforts to expand our understanding of sexual development through the identification of TRA-1 binding sites throughout the genome of the worm. Chapter 2 describes the use of ChIP-seq to identify a substantial fraction of

sites bound by TRA-1, as well as the use of reporter transgenes and RNAi to examine the function of the associated genes that serve as candidates to mediate its sex-determining roles in different tissues. Chapter 3 discusses the implications of these findings, and describes future investigations aimed at broadening our understanding of how TRA-1 and its targets regulate sexual development.

Chapter 2

TRA-1 ChIP-seq reveals regulators of sexual differentiation and multilevel feedback in nematode sex determination

Reprinted with permission from:

Proceedings of the National Academy of Sciences USA

2013 Oct 1;110(40):16033-8

Berkseth M, Ikegami K, Arur S, Lieb JD, Zarkower D.

Author contributions: M.B., K.I., S.A., J.D.L., and D.Z. designed research; M.B., K.I., and S.A. performed research; M.B., K.I., and S.A. contributed new reagents/analytic tools; M.B., K.I., S.A., J.D.L., and D.Z. analyzed data; and M.B., K.I., J.D.L., and D.Z. wrote the paper.

Summary

How sexual regulators translate global sexual fate into appropriate local sexual differentiation events is perhaps the least understood aspect of sexual development. Here we have used Chromatin Immunoprecipitation followed by deep sequencing (ChIP-seq) to identify direct targets of the nematode global sexual regulator TRA-1, a transcription factor acting at the interface between organism-wide and cell-specific sexual regulation to control all sex-specific somatic differentiation events. We identified 184 TRA-1 binding sites in *C. elegans*, many with temporal- and/or tissue-specific TRA-1 association. We also identified 78 TRA-1 binding sites in the related nematode *C. briggsae*, 19 of which are conserved between the two species. Some DNA segments containing TRA-1 binding sites drive male-specific expression patterns, and RNAi depletion of some genes adjacent to TRA-1 binding sites results in defects in male sexual development. TRA-1 binds to sites adjacent to a number of heterochronic regulatory genes, some of which drive male-specific expression, suggesting that TRA-1 imposes sex specificity on developmental timing. We also found evidence for TRA-1 feedback regulation of the global sex-determination pathway: TRA-1 binds its own locus and those of multiple upstream masculinizing genes, and most of these associations are conserved in *C. briggsae*. Thus TRA-1 coordinates sexual development by reinforcing the sex-determination decision and directing downstream sexual differentiation events.

Introduction

Sexual development requires an initial sex-determination decision followed by sex-specific differentiation events that establish anatomical, physiological, and behavioral differences between the sexes. While the sex-determination switch has been well studied

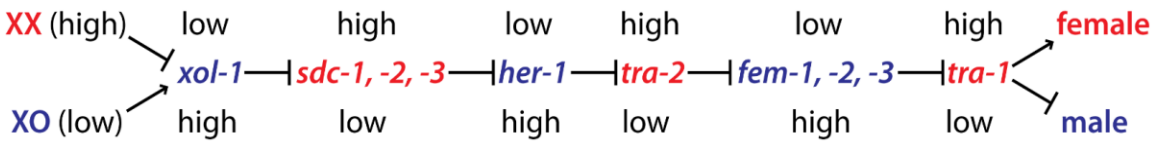
in many animals, the molecular and genetic mechanisms by which sexual fate subsequently is manifested represent a major gap in our understanding of sexual development. Here we address this question in the nematode *C. elegans*.

Sex is determined genetically in *C. elegans*: embryos with two X chromosomes (XX) develop as self-fertile hermaphrodites - females that transiently make sperm - whereas embryos with one X chromosome (XO) develop as males (Nigon, 1951). The sex chromosome complement sets the activity of TRA-1, a transcription factor that directs sexual differentiation throughout the animal (Zarkower & Hodgkin, 1992) (Fig. 2.1). High TRA-1 activity in XX animals promotes female differentiation, while low TRA-1 activity in XO animals allows male differentiation (J. Hodgkin, 1987). Sexual dimorphism is extensive, with roughly a third of adult somatic cells showing obvious sex differences (Sulston & Horvitz, 1977), and depends on *tra-1* activity. Null mutations in *tra-1* cause XX animals to develop as fertile pseudomales, whereas *tra-1* gain-of-function mutations cause XO animals to develop as fertile females (J. Hodgkin, 1987). Indeed, stable strains have been constructed in which the genotype at the *tra-1* locus determines sex independent of the sex chromosomes (J. Hodgkin, 1983b). Genetic mosaic analysis showed that *tra-1* determines sex autonomously in most cells (Hunter & Wood, 1990), although cell signaling does play a role in coordinating the sex-determination decision among cells (Hunter & Wood, 1992) and in local inductions of sexual cell fate (Kimble, 1981).

tra-1 performs several distinct functions in sexual development. First, it directs female somatic development via transcriptional repression of genes that promote male differentiation. Examples include *mab-3* (Yi et al., 2000) and *dmd-3* (Mason et al., 2008),

which jointly control specific aspects of male tail development, and *ceh-30* (Peden et al., 2007) and *egl-1* (Conradt & Horvitz, 1999), which regulate sex-specific cell deaths in the nervous system. Second, *tra-1* regulates germ line sex determination by repressing *fog-1* and *fog-3* to promote oogenesis (P. Chen & Ellis, 2000; P. J. Chen, Cho, Jin, & Ellis, 2001). Additionally, *tra-1* is required for proper cell organization in the early male somatic gonad (J. Hodgkin, 1987), and modulates membrane synthesis in the female germ line to facilitate oogenesis (Arur et al., 2011), though the TRA-1 targets functioning in these processes remain unknown. Identifying the suite of TRA-1 target genes and determining how they collectively guide sexual differentiation in *C. elegans* will provide a paradigm for understanding how sexual regulators translate global sexual fate into specific differentiation events.

Figure 2.1. Model for genetic control of *C. elegans* somatic sex determination. Minor genetic interactions and components have been omitted for simplicity.


XX (high)
XO (low)
low
high
xol-1
high
low
sdc-1, -2, -3
low
high
her-1
high
low
tra-2
low
high
fem-1, -2, -3
high
low
tra-1
female
male

Results

TRA-1 chromatin association exhibits temporal and tissue specificity

To identify TRA-1 binding sites genome-wide, we performed ChIP-seq using an antibody that detects both major TRA-1 isoforms and binds DNA-bound TRA-1 (Fig. 2.8) (Arur et al., 2011). We performed ChIP-seq in L2 larvae, L3 larvae, and *spe-11* young adults, which fail to produce embryos due to non-functional sperm (L'Hernault, Shakes, & Ward, 1988). We chose these timepoints, during which many TRA-1 dependent sex-specific differentiation events occur, in an effort to identify as many TRA-1 binding sites as possible. We also analyzed *glp-4* young adults deficient for germ cells (Beanan & Strome, 1992) to examine tissue-specificity of TRA-1 binding. We then analyzed *tra-1(e1834)* null mutant (J. Hodgkin, 1987, 1993) young adult pseudomales to identify sites enriched independently of TRA-1, which were excluded from further analysis. We identified a total of 184 sites that were specifically bound by TRA-1 in both biological replicates of at least one developmental condition (Table 2.1, Methods). The 184 TRA-1 binding sites identified are fewer than typical for site-specific transcription factors. However, they likely include about half of the TRA-1 binding sites in *C. elegans*, because binding sites were identified near five of eight previously described TRA-1 targets, *dmd-3, mab-3, xol-1, fog-1,* and *fog-3*. Binding was not observed near the previously described TRA-1 targets *egl-1, ceh-30*, and *lin-39*. The five previously described targets that were bound are required for sex-specific development in a greater number of cells than the three targets not bound, suggesting that binding may be cell type-specific and whole-animal ChIP-seq may not identify targets bound only in a small

number of cells. Alternatively, some targets may be bound at developmental timepoints we did not examine.

TRA-1 binds *in vitro* to a nonamer DNA motif very similar to that recognized by its mammalian homolog GLI (Zarkower & Hodgkin, 1993) (Fig. 2.2*A*). The most significantly enriched motif at TRA-1 binding sites was a nearly exact match to this *in vitro* defined motif (Fig. 2.2*A*), found at 169 of 184 (92%) TRA-1 binding sites, indicating that most, but not all, sites are bound through this motif. TRA-1 may bind other sites via more divergent motifs or indirect recruitment by other factors. While most TRA-1 binding sites had this motif, only about 4% of the best matches to this motif in the *C. elegans* genome were bound by TRA-1 (Fig. 2.2*B*). In general, TRA-1 ChIP enrichment was stronger for sites with closer matches to the consensus motif, but many perfect matches were not detectably bound by TRA-1, and many bound sites had imperfect matches (Fig. 2.2*B*). Thus like many transcription factors (Lieb, Liu, Botstein, & Brown, 2001), TRA-1 binds DNA at most of its sites via the consensus binding sequence, but additional factors help determine whether TRA-1 binds this motif.

To examine how TRA-1 DNA occupancy changes during development, we assessed ChIP enrichment at each of the 184 TRA-1 binding sites at each of the developmental conditions tested, counting significant enrichment in either of two replicates as an indication of TRA-1 binding (Fig. 2.2*C*). By this criterion, 32/184 sites were bound by TRA-1 in L2, 105 in L3, and 164 in *spe-11* young adults, indicating an increase in TRA-1 occupied sites as development progresses. Thus the previous suggestion that TRA-1 might bind its target sites at all times and in all tissues, only causing sex-specific expression in the cells in which nearby regulatory elements are active (Zarkower, 2001),

appears unlikely at least for many of the sites identified here. To ask whether TRA-1 associates tissue-specifically with its binding sites, we compared TRA-1 binding in *glp-4* (severely reduced germ line) and *spe-11* (normal germ line but no fertilized embryos) young adults. Of 164 sites with significant enrichment in *spe-11* animals, 45 were not bound in *glp-4* (Table 2.2, Methods), suggesting that TRA-1 binds to these sites primarily in the germ line. These germ line-specific TRA-1 targets provide candidates for future investigation of how TRA-1 regulates oogenic membrane organization (Arur et al., 2011).

Table 2.1. 184 sites in the *C. elegans* genome show significant TRA-1 ChIP enrichment in both biological replicates of at least one developmental condition. Sites that showed significant TRA-1 ChIP enrichment in *tra-1(e1834)* animals were excluded from this list. Sites are ordered by chromosomal position (column A). Columns B-E indicate the number of replicates that showed significant enrichment for each developmental condition at each site; 2 indicates significant enrichment was detected in both replicates, 1 indicates significant enrichment in 1 of 2 replicates, and 0 indicates significant enrichment in neither replicate. Column F indicates whether each site shares an orthologous TRA-1 binding site in *C. briggsae*. Column G indicates the coding gene with the transcriptional start site nearest to the TRA-1 binding site. Column H indicates all genes within 10 kilobases (kb) of the TRA-1 binding site. Column I indicates the type of expression pattern observed for reporters containing this site. Column J lists all genes associated with a given site that were examined by RNAi, and any sex-specific phenotypes that were observed. No phenotype listed indicates no sex-specific phenotypes observed.

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrI : 10016505	fog-3	2	1	2	1		C03C11.1, F10G8.2, eak−6, fog−3, npp−17		
chrI : 10517086	nlp-4	2	2	2	0		F59C6.12, F59C6.14, F59C6.16, F59C6.2, F59C6.3, F59C6.5, F59C6.8, che−13, exos−3, nlp−4		
chrI : 10565280	pas-5	2	2	1	0		F25H2.12, F25H2.6, F25H2.7, pas−5, rla−0, rtel−1, tct−1, ubc−25		
chrI : 10761847	daf-16	0	0	2	2		daf−16		
chrI : 10769587	F55A3.5	0	0	2	2		daf−16		
chrI : 10798821	marc-6	0	0	1	2		C35E7.8, C35E7.9, F55A3.2, F55A3.3, marc−6		
chrI : 11202385	ztf-6	0	1	2	2		ins−36, ztf−6	male-specific expression in the developing tail	ztf-6, male tail defects - fused rays and reduced fan
chrI : 11512609	F32B4.2	0	1	2	0		F32B4.1, F32B4.2, F32B4.4, F32B4.5, F32B4.6, F32B4.8		
chrI : 11916110	hda-3	0	1	2	1		R09B3.2, R09B3.3, exo−3, fdps−1, hda−3, mag−1, ubc−12, wve−1		
chrI : 12039858	F58D5.5	0	1	2	2	yes	F58D5.5, F58D5.6, F58D5.7, F58D5.8, F58D5.9, ksr−2	sex non-specific expression	F58D5.5
chrI : 12520488	K11D2.4	1	2	2	2		K11D2.4, K11D2.5, pri−2, unc−101	sex non-specific expression	K11D2.4

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrI : 13325048	gcy-35	0	0	2	0		T04D3.5, W08E3.2, gcy−35, pde−1		
chrI : 13331055	T04D3.5	0	1	2	0		T04D3.5, W08E3.2, W08E3.4, gcy−35, snr−2, tag−210		
chrI : 13411282	C01A2.1	0	0	1	2		C01A2.1, W05H12.1, W05H12.2, nlp−38		
chrI : 14030225	ZC334.12	0	1	2	2		ZC334.12, ins−24, ins−25, ins−26, ins−27, ins−29, ins−30		
chrI : 14162185	C37A5.1	0	1	2	1		C37A5.1, C37A5.11, C37A5.3, C37A5.7, fipr−22, fipr−23, fipr−24	sex non-specific expression	
chrI : 1836861	egl-30	0	1	2	0		C53H9.2, C53H9.3, Y71G12B.28, egl−30, emr−1, nlp−12, rpl−27, tag−96		egl-30
chrI : 2299756	epg-2	0	1	2	2		Y39G10AR.11, Y39G10AR.21, Y39G10AR.9, cdk−7, epg−2, icp−1, tpxl−1	sex non-specific expression	
chrI : 2727709	Y71F9B.6	0	0	1	2		Y71F9B.2, Y71F9B.6, Y71F9B.8, dnj−30, lin−17, snr−7, yop−1		
chrI : 2864169	Y71F9AL.11	0	0	2	2		W10C8.5, W10C8.6, Y71F9AL.10, Y71F9AL.11, Y71F9AL.12		
chrI : 3210729	fog-1	2	2	2	2		Y54E10A.13, Y54E10A.17, fog−1, rom−5		
chrI : 4032639	unc-73	0	0	2	1		unc−73, unc−89		
chrI : 4933144	F28B3.4	0	0	2	0		F28B3.1, F28B3.4, F28B3.5, acl−11, imb−1		

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrI : 5033358	C46H11.3	0	0	0	2		C46H11.3, C46H11.6, C46H11.7, lfe−2		
chrI : 5843045	col-54	0	1	1	2		F33D11.1, F33D11.2, F33D11.5, F33D11.6, F33D11.7, ceh−12, col−54, obr−4		
chrI : 6007736	C06A5.12	0	1	2	0		C06A5.1, C06A5.12, C06A5.2, C06A5.8, rnf−1, unc−94		
chrI : 7214885	nhr-23	0	0	2	0		C01H6.2, C01H6.3, C01H6.4, C01H6.6, nhr−23, tag−298		
chrI : 7475444	T28F4.1	0	1	2	2	yes	T28F4.1, T28F4.3, ZK524.4, asic−2		
chrI : 7500597	C26C6.8	0	0	2	0		C26C6.6, C26C6.8, T28F4.4, T28F4.5, T28F4.6, glb−8, pbrm−1		
chrI : 7715772	lsy-22	0	0	1	2		F27D4.4, F27D4.7, lsy−22		
chrI : 824241	ZC123.1	0	0	2	0		ZC123.3		
chrI : 8409884	lin-28	0	1	1	2		F02E9.1, F02E9.3, F02E9.7, T19A6.3, T19A6.4, lin−28, sin−3	male-specific expression in body wall muscle	
chrI : 851609	Y95B8A.8	0	0	2	2		Y95B8A.7, Y95B8A.8, ZC123.1, ZC123.4		
chrI : 8709006	cdk-8	0	0	2	0		cdk−8, cle−1, tlf−1		
chrI : 8792355	sup-17	0	1	2	2		F36A2.2, F36A2.3, mfb−1, sup−17	sex non-specific expression	
chrI : 9854145	K10C3.2	0	1	1	2	yes	K10C3.2, K10C3.4, K10C3.5, nhr−62, zig−1		

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrII : 10502785	pqn-47	0	1	2	1		F59B10.2, F59B10.3, ZK666.12, ZK666.14, ZK666.8, pqn−47		
chrII : 10528184	mab-10	0	1	2	1		F59B10.4, F59B10.5, F59B10.6, R166.2, R166.3, mab−10	no expression	mab-10
chrII : 11103267	C09H10.5	0	2	1	2		C09H10.10, C09H10.5, C09H10.7, C09H10.9, glb−4, mat−2, nasp−1, nuo−1, rpl−41, rpl−42		
chrII : 11688152	kel-1	0	0	1	2		C47D12.2, C47D12.4, C47D12.5, C47D12.8, VF13D12L.1, kel−1, sfxn−1.4, tars−1		
chrII : 1236838	lin-42	0	0	2	0		F53A10.2, lin−42	male-specific expression in the developing tail	
chrII : 12549783	F49C5.10	0	1	2	1		F49C5.10, F49C5.11, F49C5.12, F49C5.4, F49C5.5, F49C5.9		
chrII : 13291657	ztf-22	0	0	2	0		Y48C3A.5, ztf−22	sex non-specific expression	
chrII : 13738123	R06B9.5	0	1	2	2		R06B9.5, Y54G9A.9, bub−3, mig−14	no expression	mig-14
chrII : 13836026	ins-37	2	0	0	0		F08G2.4, F08G2.5, F08G2.7, his−42, his−43, his−44, ins−37		
chrII : 13852762	F08G2.8	2	1	0	0		F08G2.8, Y51H1A.1, Y51H1A.2		
chrII : 14244807	jmjd-2	0	0	2	2		Y54G11A.2, hmg−1.1, jmjd−2	sex non-specific expression	

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrII : 1470737	col-70	0	0	2	0		H17B01.5, Y51H7C.1, Y51H7C.13, Y51H7C.15, col−70		
chrII : 15209387	spsb-1	0	0	1	2		spsb−1		
chrII : 162336	etr-1	0	2	0	0		T01D1.7, T25D3.2, T25D3.3, T25D3.4, etr−1		
chrII : 1851935	C52E2.3	0	0	0	2		C16C4.18, C52E2.2, C52E2.3, C52E2.4, C52E2.5, C52E2.8, fbxb−95, fbxb−96, fbxb−97		
chrII : 3376793	bath-9	0	1	1	2		T06D4.1, T06D4.2, Y49F6C.2, Y49F6C.7, Y49F6C.8, bath−10, bath−23, bath−9	sex non-specific expression	
chrII : 3439757	C16C8.17	0	1	1	2		C16C8.14, C16C8.16, C16C8.17, C16C8.18, C16C8.2, C16C8.20, C16C8.21, C16C8.22, C16C8.4, C16C8.5		
chrII : 3786698	T24E12.13	0	1	0	2		T24E12.1, T24E12.11, T24E12.13, T24E12.2, T24E12.3, Y8A9A.2, Y8A9A.3, Y8A9A.4, fbxb−99		
chrII : 4376731	ntl-2	0	0	2	2		B0286.1, B0286.3, fkh−6, lat−2, ntl−2, try−9		
chrII : 5336805	F09E5.14	0	0	2	2		F09E5.14, F09E5.9, vhp−1		
chrII : 5599784	C17G10.10	2	2	1	0		C17G10.1, C17G10.10, C17G10.2, C17G10.6, C17G10.7, cdc−14		

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrII : 5901848	tsp-18	0	0	2	2		F55C12.1, lin-4, F59G1.4, F59G1.8, cgt−3, tsp−18, vps−35	male-specific expression in the developing tail	
chrII : 6935045	F22D3.5	0	0	2	1		F22D3.2, F22D3.4, F22D3.5		
chrII : 8091653	ltd-1	2	2	1	0		K02C4.3, ltd−1, sfxn−5, sre−2		
chrII : 915308	fbxc-29	0	0	0	2		clec−44, fbxa−200, fbxc−24, fbxc−27, fbxc−28, fbxc−29, fbxc−30, fbxc−31, fbxc−33		
chrII : 9301031	C01G6.9	0	0	0	2		C01G6.9, acs−7, cam−1		
chrII : 9726527	cutl-2	0	1	2	2		Y53C12A.11, Y53C12A.6, cutl−2, mab−3, wee−1.3		
chrII : 9728870	cutl-2	2	2	2	2	yes	Y53C12A.11, Y53C12A.6, Y53C12B.2, cutl−2, mab−3		
chrII : 9746211	Y53C12B.1	0	0	2	1		Y53C12B.1, Y53C12B.2, Y53C12B.7, ZK1321.1, mab−3, nos−3		
chrIII : 10096662	bath-43	0	0	2	2		T16H12.11, T16H12.2, T16H12.4, bath−43, kel−10, srt−55		
chrIII : 10632610	Y39A1A.9	0	2	2	2	yes	Y39A1A.10, Y39A1A.12, Y39A1A.27, Y39A1A.8, Y39A1A.9, dhs−11, hpr−9	sex non-specific expression	Y39A10A.10
chrIII : 10634106	Y39A1A.9	0	1	2	1	yes	Y39A1A.10, Y39A1A.12, Y39A1A.27, Y39A1A.8, Y39A1A.9, dhs−11, hpr−9	sex non-specific expression	Y39A10A.9

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrIII : 10759189	K01G5.9	0	1	2	0		K01G5.10, K01G5.3, K01G5.5, K01G5.9, dpy−28, hpl−2, rnf−113		
chrIII : 11185869	tra-1	0	2	2	2	yes	tra−1		
chrIII : 11188157	tra-1	0	2	1	0		tra−1		
chrIII : 11243471	rsks-1	0	0	1	2		Y47D3A.14, Y47D3A.32, aakb−2, elo−8, rsks−1		
chrIII : 11560690	ztf-29	0	1	2	1		Y66D12A.10, Y66D12A.11, Y66D12A.13, Y66D12A.9, ztf−29		
chrIII : 1226441	pes-4	0	1	2	1		fbxa−21, pes−4		
chrIII : 12599056	Y111B2A.10	0	0	2	1		Y111B2A.10, Y111B2A.12, epc−1		
chrIII : 12613574	Y111B2A.12	0	0	1	2		Y111B2A.12, epc−1		
chrIII : 12919983	Y37D8A.16	0	1	2	1		Y37D8A.16, Y37D8A.17, Y37D8A.18, Y37D8A.19, Y37D8A.21, flp−14		
chrIII : 13068332	Y39E4B.6	0	1	1	2		Y39E4B.6		
chrIII : 13734003	W06F12.3	0	2	1	1		W06F12.2, W06F12.3, lit−1, tyr−2		W06F12.2, W06F12.3, lit-1
chrIII : 147792	fbxb-84	0	0	0	2		T17A3.10, T17A3.12, T17A3.2, T17A3.9, fbxb−81, fbxb−82, fbxb−83, fbxb−84, ver−1, ver−2		
chrIII : 154398	T17A3.9	0	0	0	2		F40G9.8, T17A3.10, T17A3.12, T17A3.9, fbxb−83, fbxb−84, ver−1, ver−2		
chrIII : 3786572	acy-3	0	1	2	1		acy−3, mtm−3		

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrIII : 3953431	F37A8.5	0	0	2	2		F37A8.5, R10E9.2, R10E9.3		
chrIII : 4282668	tag-310	0	2	2	2		R10E4.1, R10E4.3, R10E4.6, mcm−5, nth−1, tag−310	sex non-specific expression	R10E4.3, tag-310
chrIII : 4667049	T23F11.4	0	0	1	2		T23F11.4, cdka−1, srg−13		
chrIII : 4748760	col-89	0	0	2	0		B0393.3, B0393.8, F17C8.6, VB0393L.2, col−89, ras−2, rbg−3, rps−0, twk−6		
chrIII : 4796195	tag-325	0	1	2	1		C38D4.4, C38D4.9, mel−28, pal−1, tag−325		
chrIII : 4950080	flh-2	0	1	1	2	yes	C26E6.1, C26E6.3, C27F2.4, flh−2, fsn−1, rpb−2, vps−22	sex non-specific expression	C26E6.3, flh-2
chrIII : 5423918	dnj-24	0	1	2	0		W03A5.4, dnj−24, grl−22		
chrIII : 5463824	F48E8.2	0	1	2	0		F48E8.2, F48E8.3, F48E8.4		
chrIII : 5622659	daf-4	0	0	1	2		C05D2.3, bas−1, daf−4		
chrIII : 5642861	ckk-1	0	1	2	1		ckk−1		
chrIII : 5653892	sdz-21	2	1	1	0		F54E7.6, ckk−1, sdz−21		
chrIII : 5953996	psf-1	0	0	2	0		F25B5.2, F25B5.3, F25B5.5, F25B5.6, psf−1, ubq−1		
chrIII : 6421140	C05D11.7	0	1	2	0		C05D11.10, C05D11.5, C05D11.7, C05D11.8, C05D11.9, nas−4	sex non-specific expression	
chrIII : 6464203	dlc-1	0	1	1	2		T26A5.2, T26A5.4, T26A5.5, T26A5.8, dlc−1, nduf−2.2, wht−6		

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrIII : 6603017	let-765	0	0	1	2		F20H11.4, Y40D12A.2, let−765, mdh−2, srh−40		
chrIII : 7659500	ZK783.7	0	1	1	2		C18H2.1, ZK783.6, ZK783.7, flt−1		
chrIII : 8425205	cec-1	0	0	2	1		ZK1236.1, ZK1236.5, ZK1236.9, cec−1, sor−1		
chrIII : 8658334	F22B7.9	0	0	1	2		F22B7.1, F22B7.3, F22B7.9, dpy−19, flp−23		F22B7.9, dpy-19
chrIII : 9765617	cbp-1	1	1	2	1		T05G5.9, cbp−1, rmd−1, vps−53		
chrIV : 1064617	Y55F3AM.14	0	0	2	2	yes	Y55F3AM.1, Y55F3AM.13, Y55F3AM.14, dcap−1	sex non-specific expression	Y55F3AM.14
chrIV : 1070629	Y55F3AM.14	0	1	1	2	yes	Y55F3AM.14, csn−4	no expression	
chrIV : 12891986	K08D8.1	0	0	2	0		K08D8.1, K08D8.10, K08D8.11, K08D8.12, K08D8.6, K08D8.7, K08D8.9, ZK896.4		
chrIV : 13395155	Y62E10A.13	0	1	2	1		Y62E10A.13, Y62E10A.14, Y62E10A.19, Y62E10A.20, lsm−3, mdt−9		
chrIV : 1358226	col-105	0	0	2	0		col−104, col−105, rfc−2		
chrIV : 15719080	Y105C5A.17	0	0	2	0		Y105C5A.15, Y105C5A.17		
chrIV : 3016585	lin-22	0	1	2	1		Y54G2A.27, Y54G2A.28, Y54G2A.29, Y54G2A.3, lin−22	sex non-specific expression	
chrIV : 3079231	Y67D8C.1	0	0	2	0		Y67D8C.1, Y67D8C.2, Y67D8C.9, cpg−9		

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrIV : 5063480	msp-57	2	1	0	0		C02B10.6, C09B9.7, R13H9.5, R13H9.6, msp−53, msp−55, msp−57, rmd−6		
chrIV : 8782332	C33D9.13	0	1	1	2		C33D9.10, C33D9.13, C33D9.3, C33D9.6, C33D9.8, F20C5.6, sru−35	sex non-specific expression	C33D9.13
chrIV : 9105139	fem-3	2	2	2	2	yes	C01F6.2, C01F6.9, aly−1, cpna−3, fem−3, icl−1, lpl−1		
chrIV : 9130782	F23B2.3	2	1	0	0		F23B2.3, daf−10, tag−60		
chrIV : 9525966	elo-1	0	0	0	2		C33A12.1, C33A12.19, F56H11.2, F56H11.6, elo−1, elo−7, nlp−35		
chrIV : 991285	egrh-2	2	2	2	0		Y55F3AM.5, Y55F3AM.6, egrh−2, plx−1		Y55F3AM.6
chrV : 10032506	acl-2	0	0	2	2	yes	F57A8.1, K08H10.2, T06E8.2, acl−2, lea−1	sex non-specific expression	
chrV : 10630912	mab-23	0	1	1	2		C32C4.1, mab−23	male-specific expression in the adult tail	
chrV : 10717312	pph-1	0	1	2	2		T28B11.1		
chrV : 11175638	gpa-1	2	1	0	0		T19C4.5, cnc−10, gpa−1		
chrV : 11403414	nas-19	0	1	2	1		K03B8.4, K03B8.6, K03B8.8, deg−3, grl−3, nas−16, nas−17, nas−18, nas−19	sex non-specific expression	nas-19
chrV : 11585317	srx-89	2	1	1	0		nhr−68, srx−80, srx−82, srx−88, srx−89, srx−90		

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrV : 12465610	egl-10	0	0	2	1		F28C1.3, egl−10		
chrV : 13084117	C34B4.2	0	0	2	2		C34B4.2, C34B4.4, C34B4.7		
chrV : 14028579	T08G5.3	2	2	1	0		T08G5.1, T08G5.15, T08G5.2, T08G5.3, ins−10, mtl−2, vps−39		
chrV : 14177804	ZC376.4	0	1	2	1		ZC376.1, ZC376.2, ZC376.3, ZC376.4, ZC376.6, ZC376.8, ocam−1, trm−1		
chrV : 14358764	Y49A3A.4	2	1	0	0		Y49A3A.1, Y49A3A.3, Y49A3A.4, cyn−1, vha−13		
chrV : 14463270	F08H9.3	2	2	1	0		F08H9.2, F08H9.3, F08H9.4, bmk−1, clec−227, clec−57, coh−3, srz−97		
chrV : 14468666	srz-97	2	2	0	0		F08H9.2, F08H9.3, F08H9.4, clec−227, clec−54, clec−56, clec−57, coh−3, srz−97		
chrV : 15029059	F14H8.8	2	1	0	0		C25D7.1, C25D7.12, F14H8.2, F14H8.4, F14H8.5, F14H8.8, cng−1		
chrV : 15065357	C25D7.9	0	0	1	2		C25D7.10, C25D7.15, C25D7.16, C25D7.5, C25D7.9, dnj−3, mcm−3, otub−1, rap−2, unc−76		
chrV : 16225834	F11A5.13	2	1	0	0		F11A5.13, F11A5.15, F11A5.9, F21H7.12, F21H7.2, F21H7.3, clec−233, glc−1, gst−22, srh−98, stdh−2		

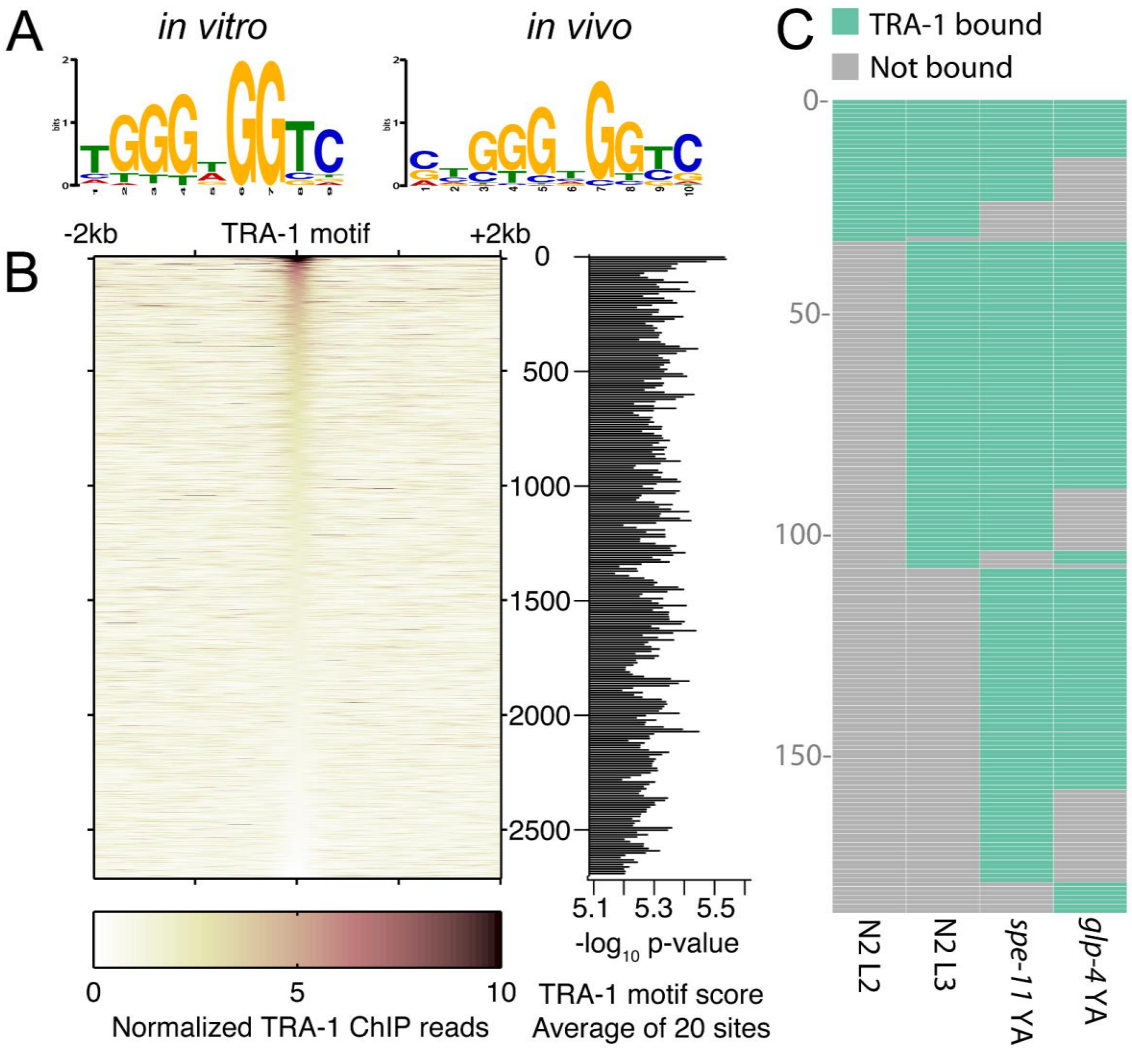
Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrV : 17951199	Y59A8B.24	0	0	1	2		Y59A8B.24, dpy−21, gck−3		
chrV : 18346209	C08E8.3	0	1	0	2		C08E8.10, C08E8.11, C08E8.2, C08E8.3		
chrV : 19212751	nhr-79	0	0	2	0		Y39B6A.1, nhr−79, oac−49, srr−2		
chrV : 19666471	Y43F8C.11	1	1	2	2		Y43F8C.11, mrp−7		
chrV : 20239880	Y113G7B.24	0	1	2	1		Y113G7B.15, Y113G7B.24, psa−1		
chrV : 20776772	F38A6.4	0	0	1	2		F38A6.4, F38A6.5, elp−1, hif−1		
chrV : 3703152	lin-40	0	1	2	0		T27C4.1, W08A12.1, lin−40		
chrV : 3714944	T27C4.1	0	1	2	0		T27C4.1, lin−40, str−30		
chrV : 5439212	F08F3.9	2	2	2	2	yes	F08F3.1, F08F3.8, F08F3.9, acl−6, glc−3, rhr−1	no expression	F08F3.9
chrV : 6224451	W06H8.6	0	0	2	2	yes	W06H8.6, rme−1, str−206		
chrV : 6355484	F13H6.1	0	0	2	0		F13H6.1, F13H6.3		
chrV : 9273768	ZK682.2	0	0	1	2		ZK682.2, ZK682.5, hlh−10		
chrX : 10079404	col-176	0	0	2	1		W07E11.1, ZC373.4, ZC373.5, col−176, dao−4, flp−28		
chrX : 10188704	F21A10.4	0	1	2	2	yes	F21A10.1, F21A10.2, F21A10.4, nucb−1	sex non-specific expression	F21A10.2, F21A10.4, nucb-1
chrX : 10435644	mig-15	0	0	2	1		C39B10.3, C39B10.4, mig−15		
chrX : 10480676	elt-2	0	2	0	1		C33D3.3, C33D3.4, C39B10.7, elt−2, elt−4		elt-2
chrX : 10637590	daf-12	0	1	2	1	yes	M79.2, daf−12		daf-12

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrX : 11465227	lin-14	0	1	2	1	yes	lin−14	no expression	
chrX : 11474368	lin-14	0	1	2	1		lin−14		
chrX : 12175862	F57G12.1	0	0	2	1		F57G12.1		
chrX : 12629406	tag-147	0	0	2	1		R09A8.2, R09A8.5, col−182, tag−147		
chrX : 1319933	H42K12.2	0	2	2	2		F21E9.2, H42K12.2, H42K12.3, pdk−1		H42K12.3, pdk-1
chrX : 1327454	F21E9.2	0	1	2	0		F21E9.2, H42K12.2, pdk−1, ttr−37		H42K12.2
chrX : 13687166	adm-2	0	1	2	1		adm−2, gck−4		
chrX : 14016867	C04C11.1	0	0	1	2		C04C11.1, tbc−19		
chrX : 15070232	H03G16.2	0	1	2	0		H03G16.2, H40L08.2, H40L08.3, fbxa−96		
chrX : 15189293	T10B10.4	0	0	1	2		T10B10.3, T10B10.4, phat−6, snt−7		
chrX : 15275776	kin-20	0	1	2	2	yes	F46F2.5, jud−4, kin−20, ser−2	male-specific expression in the developing tail	
chrX : 16128908	ZK1073.1	0	1	1	2		ZK1073.1, ZK1073.2		
chrX : 16286501	F52E10.3	0	0	2	0		F52E10.2, F52E10.3, F52E10.4, sdc−1		
chrX : 16961323	F52G3.1	0	1	2	2	yes	F52G3.1, F52G3.4	sex non-specific expression	F52G3.1, F52G3.4

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrX : 17314852	ptr-5	1	2	2	2		C33E10.1, C53C11.4, ptr−5	male-specific expression in body wall muscle	ptr-5
chrX : 17416863	crb-1	0	0	2	1		crb−1, osm−11		
chrX : 1840193	T26C11.3	0	0	2	1		T26C11.2, T26C11.3, T26C11.4, ceh−21, ceh−41		
chrX : 1961789	F52D2.7	2	2	2	2		F52D2.7, fbxa−17, rgs−8.1, str−79	sex non-specific expression	F52D2.7
chrX : 2329010	T01B6.1	2	1	1	0		T01B6.1, T01B6.4, aakg−4		
chrX : 2550805	F52H2.4	0	0	1	2		F52H2.3, F52H2.4, F52H2.5, F52H2.6, aat−3		
chrX : 2899548	F52B10.3	0	0	1	2		B0410.1, B0410.3, F52B10.3, nmy−1, vang−1	sex non-specific expression	
chrX : 3945373	prl-1	1	1	1	2		T19D2.1, T19D2.3, prl−1		
chrX : 3970177	C16B8.3	0	0	2	0		C16B8.2, C16B8.3, C16B8.4, lin−18		
chrX : 4307659	pqn-15	0	1	2	2		C52B9.8, pqn−15		
chrX : 4635878	C46C11.1	0	1	1	2		C46C11.1, C46C11.2, C46C11.3, C46C11.4		
chrX : 5336701	C26B9.6	0	1	2	1		C26B9.1, C26B9.2, C26B9.3, C26B9.5, C26B9.6, C26B9.7		
chrX : 6250602	W05H9.3	0	0	1	2		C14F11.1, W05H9.3		
chrX : 6279663	col-166	0	0	2	0		T07H6.4, W05H9.1, W05H9.2, col−166, lev−9		

Peak center	Nearest Gene	N2 L2	N2 L3	spe-11 YA	glp-4 YA	Conserved in *C. briggsae*?	All genes within 10 kb of site	Reporter expression	Associated genes examined by RNAi
chrX : 768330	ile-2	1	2	2	2		T04G9.4, T04G9.7, aex−3, dyf−11, ile−2, nas−15, trap−2	sex non-specific expression	
chrX : 8041025	xol-1	2	2	2	2		C18A11.4, C18A11.6, C34D10.1, dim−1, xol−1		
chrX : 8059886	C18A11.3	0	0	2	0		C18A11.1, C18A11.2, C18A11.3, dim−1		
chrX : 806669	F25E2.1	0	2	1	0		F25E2.1, F25E2.2, F25E2.3, ifd−2		
chrX : 8286727	F13B9.1	0	1	2	1		F13B9.1, F13B9.2, R09F10.3, fis−2, ksr−1		
chrX : 8568916	tag-279	0	0	1	2		F18E9.6, F18E9.8, tag−279, utx−1		
chrX : 9549120	F49E2.5	0	0	1	2		F49E2.5, nuc−1		
chrX : 9999508	grk-1	0	0	2	1		F19C6.2, F19C6.3, grk−1		

Figure 2.2. Analysis of TRA-1 binding sites in the *C. elegans* genome. (*A*) Comparison of *in vitro* and *in viv*o derived TRA-1 binding motifs. (*B*) Analysis of TRA-1 occupancy at the 2713 closest matches to the TRA-1 motif found in the *C. elegans* genome. Motifs ordered by level of TRA-1 ChIP enrichment at the L3 stage. Left: Normalized TRA-1 ChIP reads within a 4 kb window centered on each TRA-1 motif. Right: Strength of match to TRA-1 binding site for TRA-1 motifs, each bar represents an average score ($-\log_{10}$ of p-value) for 20 motifs, sliding every 10 motifs. (*C*) Analysis of stage- and tissue-specific TRA-1 occupancy at 184 *in vivo* TRA-1 binding sites. Green boxes indicate a site was significantly bound (p-value $< 10^{-5}$; fold enrichment >4) in at least one of two replicates performed for a given developmental condition.


A
in vitro
in vivo
B
-2kb
TRA-1 motif
+2kb
0
500
1000
1500
2000
2500
5.1
5.3
5.5
-log10 p-value
TRA-1 motif score
Average of 20 sites
0
5
10
Normalized TRA-1 ChIP reads
C
TRA-1 bound
Not bound
0
50
100
150
N2 L2
N2 L3
spe-11 YA
glp-4 YA

TRA-1 target genes are involved in male-specific sexual development

We performed a pilot screen to evaluate whether our list of 184 TRA-1 binding sites can be used to find new sexual regulators. We used reporter analysis to identify TRA-1 bound DNA segments capable of driving sex-specific expression (Fig. 2.3*A*), and RNAi to identify genes adjacent to TRA-1 binding sites that are required for sexual differentiation (Fig. 2.3*B*). We chose 35 sites to examine with reporter analysis (Table 2.1), based on the strength of TRA-1 ChIP enrichment, enrichment in multiple stages, and the presence of strong TRA-1 motifs. Six TRA-1 bound segments drove obvious male-specific expression when fused to a minimal promoter and GFP (Fig. 2.3*A*, Fig. 2.4*A*). One was adjacent to the zinc-finger transcription factor *ztf-6*, and drove expression in male-specific tail lineages. Another was adjacent to *ptr-5*, which encodes a protein distantly paralogous to *Drosophila* PATCHED. A *ptr-5* reporter expressed male-specifically in body wall and stomatointestinal muscle as well as in sex muscles and some neurons of both sexes (Fig. 2.3*A*). Ablating the TRA-1 binding motif in this reporter activated expression in hermaphrodite body wall and stomatointestinal muscle cells (Fig. 2.3*A*). Another site was adjacent to *mab-23*, which encodes a DM-domain transcription factor that regulates male sexual differentiation (Lints & Emmons, 2002). A short region containing this site drove male-specific reporter expression in the tail (Fig. 2.3*A*). Ablation of the TRA-1 motif in this reporter caused ectopic expression in the hermaphrodite tail, but the expression was weak and variable, suggesting that TRA-1 may act redundantly to repress *mab-23*, or that hermaphrodites lack an activator required for robust expression. *mab-23* is one of three DM-domain transcription factors (along with *mab-3* and *dmd-3*), implicated in male sexual differentiation and directly regulated

by TRA-1. Thus transcriptional regulation of this protein family, whose members also regulate sex in *Drosophila* and in vertebrates (Raymond et al., 1998), is a major mechanism by which TRA-1 regulates somatic sexual differentiation.

Since previously described TRA-1 binding sites generally lie within or adjacent to the gene that is being regulated, we speculate that most of the 184 TRA-1 binding sites will regulate an overlapping gene or one of the two immediately adjacent non-overlapping genes. We identified 444 genes either containing or adjacent to a TRA-1 binding site, and chose 34 to examine by RNAi (Table 2.1), based on their proximity to strong TRA-1 binding sites and predicted molecular functions consistent with a role in sexual differentiation. RNAi depletion of one of these genes, *ztf-6,* caused male tail defects (Fig. 2.3*B*), including fused rays and a reduced fan. Thus *ztf-6*, *ptr-5*, and *mab-23* are likely direct TRA-1 target genes with roles in male sexual differentiation. Since this pilot screen examined expression patterns for only 20% of the 184 TRA-1 binding sites, and RNAi phenotypes for only 8% of the 444 genes near these sites, it is highly likely that the ChIP data presented above can be used to find additional regulators of male differentiation.

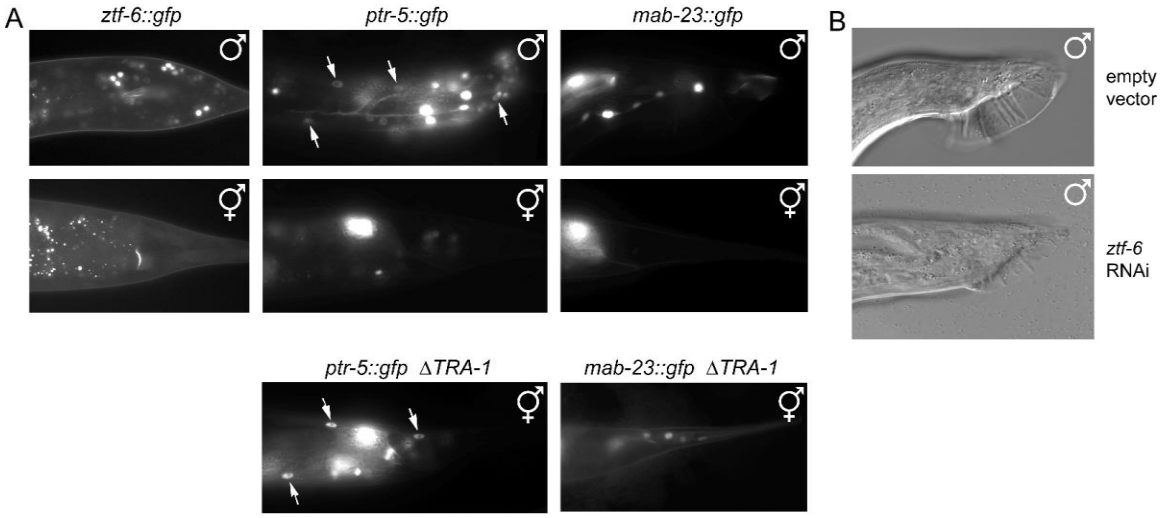
Table 2.2. 45 TRA-1 binding sites show significant enrichment in at least one replicate of *spe-11* animals but do not show significant enrichment in either *glp-4* replicate, suggesting that TRA-1 binds to these sites primarily in the germ line. Sites are ordered by chromosomal position (column A). Columns B-E indicate the number of replicates that showed significant enrichment for each developmental condition at each site; 2 indicates significant enrichment was detected in both replicates, 1 indicates significant enrichment in 1 of 2 replicates, and 0 indicates significant enrichment in neither replicate. Column F indicates whether each site shares an orthologous TRA-1 binding site in *C. briggsae*. Column G indicates the coding gene with the transcriptional start site nearest to the TRA-1 binding site. Column H indicates all genes within 10 kb of the TRA-1 binding site.

Peak center	**Nearest gene**	**N2 L2**	**N2 L3**	**spe-11 YA**	**glp-4 YA**	**Conserved in *C. briggsae*?**	**All genes within 10 kb of site**
chrI : 10517086	nlp-4	2	2	2	0	no	F59C6.12, F59C6.14, F59C6.16, F59C6.2, F59C6.3, F59C6.5, F59C6.8, che-13, exos-3, nlp-4
chrI : 10565280	pas-5	2	2	1	0	no	F25H2.12, F25H2.6, F25H2.7, pas-5, rla-0, rtel-1, tct-1, ubc-25
chrI : 11512609	F32B4.2	0	1	2	0	no	F32B4.1, F32B4.2, F32B4.4, F32B4.5, F32B4.6, F32B4.8
chrI : 13325048	gcy-35	0	0	2	0	no	T04D3.5, W08E3.2, gcy-35, pde-1
chrI : 13331055	T04D3.5	0	1	2	0	no	T04D3.5, W08E3.2, W08E3.4, gcy-35, snr-2, tag-210
chrI : 1836861	egl-30	0	1	2	0	no	C53H9.2, C53H9.3, Y71G12B.28, egl-30, emr-1, nlp-12, rpl-27, tag-96
chrI : 4933144	F28B3.4	0	0	2	0	no	F28B3.1, F28B3.4, F28B3.5, acl-11, imb-1
chrI : 6007736	C06A5.12	0	1	2	0	no	C06A5.1, C06A5.12, C06A5.2, C06A5.8, rnf-1, unc-94
chrI : 7214885	nhr-23	0	0	2	0	no	C01H6.2, C01H6.3, C01H6.4, C01H6.6, nhr-23, tag-298
chrI : 7500597	C26C6.8	0	0	2	0	no	C26C6.6, C26C6.8, T28F4.4, T28F4.5, T28F4.6, glb-8, pbrm-1
chrI : 824241	ZC123.1	0	0	2	0	no	ZC123.3
chrI : 8709006	cdk-8	0	0	2	0	no	cdk-8, cle-1, tlf-1
chrII : 1236838	lin-42	0	0	2	0	no	F53A10.2, lin-42
chrII : 13291657	ztf-22	0	0	2	0	no	Y48C3A.5, ztf-22
chrII : 1470737	col-70	0	0	2	0	no	H17B01.5, Y51H7C.1, Y51H7C.13, Y51H7C.15, col-70
chrII : 5599784	C17G10.10	2	2	1	0	no	C17G10.1, C17G10.10, C17G10.2, C17G10.6, C17G10.7, cdc-14
chrII : 8091653	ltd-1	2	2	1	0	no	K02C4.3, ltd-1, sfxn-5, sre-2
chrIII : 10759189	K01G5.9	0	1	2	0	no	K01G5.10, K01G5.3, K01G5.5, K01G5.9, dpy-28, hpl-2, rnf-113
chrIII : 11188157	tra-1	0	2	1	0	no	tra-1

Peak center	**Nearest gene**	**N2 L2**	**N2 L3**	**spe-11 YA**	**glp-4 YA**	**Conserved in *C. briggsae*?**	**All genes within 10 kb of site**
chrIII : 4748760	col-89	0	0	2	0	no	B0393.3, B0393.8, F17C8.6, VB0393L.2, col-89, ras-2, rbg-3, rps-0, twk-6
chrIII : 5423918	dnj-24	0	1	2	0	no	W03A5.4, dnj-24, grl-22
chrIII : 5463824	F48E8.2	0	1	2	0	no	F48E8.2, F48E8.3, F48E8.4
chrIII : 5653892	sdz-21	2	1	1	0	no	F54E7.6, ckk-1, sdz-21
chrIII : 5953996	psf-1	0	0	2	0	no	F25B5.2, F25B5.3, F25B5.5, F25B5.6, psf-1, ubq-1
chrIII : 6421140	C05D11.7	0	1	2	0	no	C05D11.10, C05D11.5, C05D11.7, C05D11.8, C05D11.9, nas-4
chrIV : 12891986	K08D8.1	0	0	2	0	no	K08D8.1, K08D8.10, K08D8.11, K08D8.12, K08D8.6, K08D8.7, K08D8.9, ZK896.4
chrIV : 1358226	col-105	0	0	2	0	no	col-104, col-105, rfc-2
chrIV : 15719080	Y105C5A.17	0	0	2	0	no	Y105C5A.15, Y105C5A.17
chrIV : 3079231	Y67D8C.1	0	0	2	0	no	Y67D8C.1, Y67D8C.2, Y67D8C.9, cpg-9
chrIV : 991285	egrh-2	2	2	2	0	no	Y55F3AM.5, Y55F3AM.6, egrh-2, plx-1
chrV : 11585317	srx-89	2	1	1	0	no	nhr-68, srx-80, srx-82, srx-88, srx-89, srx-90
chrV : 14028579	T08G5.3	2	2	1	0	no	T08G5.1, T08G5.15, T08G5.2, T08G5.3, ins-10, mtl-2, vps-39
chrV : 14463270	F08H9.3	2	2	1	0	no	F08H9.2, F08H9.3, F08H9.4, bmk-1, clec-227, clec-57, coh-3, srz-97
chrV : 19212751	nhr-79	0	0	2	0	no	Y39B6A.1, nhr-79, oac-49, srr-2
chrV : 3703152	lin-40	0	1	2	0	no	T27C4.1, W08A12.1, lin-40
chrV : 3714944	T27C4.1	0	1	2	0	no	T27C4.1, lin-40, str-30
chrV : 6355484	F13H6.1	0	0	2	0	no	F13H6.1, F13H6.3
chrX : 1327454	F21E9.2	0	1	2	0	no	F21E9.2, H42K12.2, pdk-1, ttr-37

Peak center	**Nearest gene**	**N2 L2**	**N2 L3**	**spe-11 YA**	**glp-4 YA**	**Conserved in *C. briggsae*?**	**All genes within 10 kb of site**
chrX : 15070232	H03G16.2	0	1	2	0	no	H03G16.2, H40L08.2, H40L08.3, fbxa-96
chrX : 16286501	F52E10.3	0	0	2	0	no	F52E10.2, F52E10.3, F52E10.4, sdc-1
chrX : 2329010	T01B6.1	2	1	1	0	no	T01B6.1, T01B6.4, aakg-4
chrX : 3970177	C16B8.3	0	0	2	0	no	C16B8.2, C16B8.3, C16B8.4, lin-18
chrX : 6279663	col-166	0	0	2	0	no	T07H6.4, W05H9.1, W05H9.2, col-166, lev-9
chrX : 8059886	C18A11.3	0	0	2	0	no	C18A11.1, C18A11.2, C18A11.3, dim-1
chrX : 806669	F25E2.1	0	2	1	0	no	F25E2.1, F25E2.2, F25E2.3, ifd-2

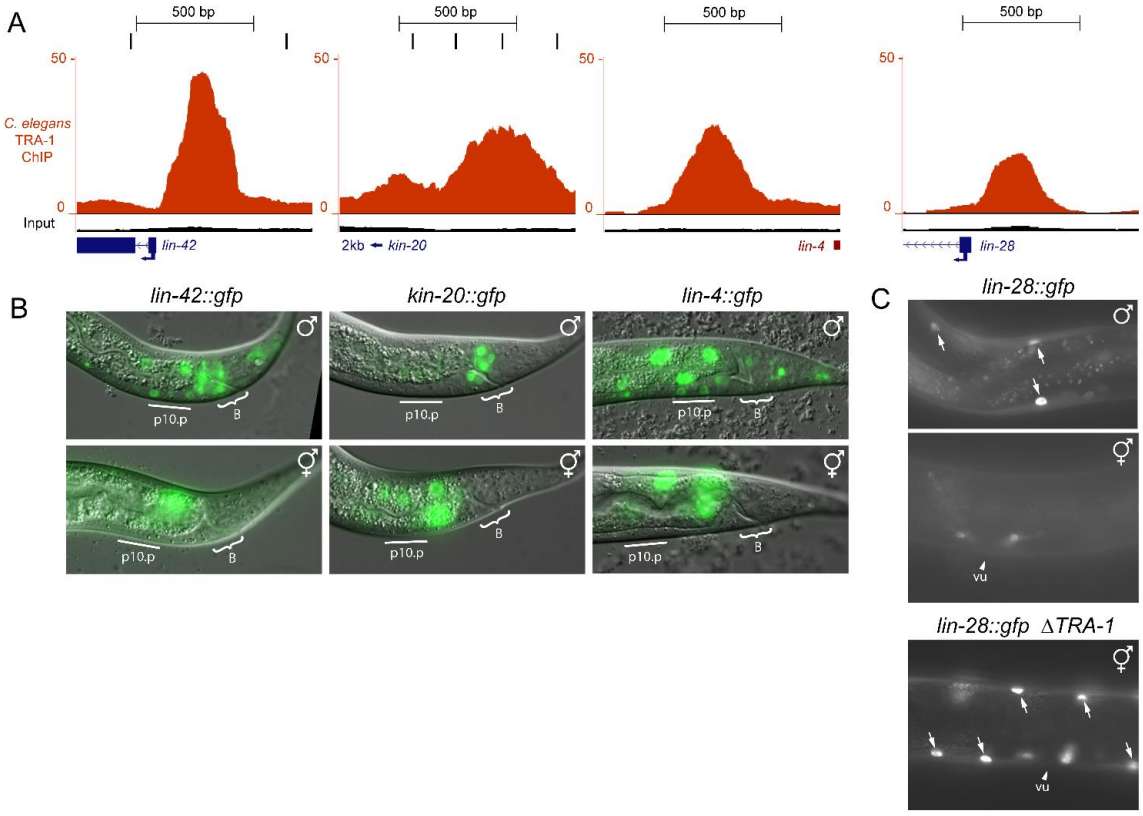
Figure 2.3. Newly identified TRA-1 target genes exhibit male-specific expression patterns and reduction-of-function phenotypes. (*A*) Genomic intervals containing TRA-1 binding sites drive male-specific reporter gene expression. *ztf-6::gfp* is expressed male-specifically in the developing tail (left). *ptr-5::gfp* is expressed male-specifically in body wall (arrows) and stomatointestinal muscle and the sex-specificity of this expression is dependent on the presence of a TRA-1 motif (center). *mab-23::gfp* is expressed male-specifically in the tail, and the sex-specificity of this expression is dependent on the presence of a TRA-1 motif (right). (*B*) RNAi of *ztf-6* results in defects in male tail development, including fused rays and a reduced fan.


A
ztf-6::gfp
ptr-5::gfp
mab-23::gfp
ptr-5::gfp ΔTRA-1
mab-23::gfp ΔTRA-1
B
empty vector
ztf-6 RNAi

TRA-1 binds to and regulates the expression of genes that control developmental timing

Developmental timing in *C. elegans* is controlled by a complex network of heterochronic regulatory genes (Moss, 2007). Heterochronic mutations cause defects in multiple aspects of male sexual development, probably by disrupting the timing of critical cell divisions (Nelson et al., 2011). Strikingly, among the roughly two-dozen heterochronic genes, TRA-1 bound adjacent to six, *daf-12*, *kin-20*, *lin-4*, *lin-14*, *lin-28*, and *lin-42* (Fig. 2.4*A*, Table 2.1). DNA segments containing three of these sites, those adjacent to *lin-42*, *kin-20*, and *lin-4*, drove strong expression in lineages that undergo male-specific developmental programs, particularly the P10.p and B lineages of the developing tail (Fig. 2.4*B*). A *lin-28* reporter showed male-specific expression in body wall muscle, a sexually specialized cell type found in both sexes (Fig. 2.4*C*), and ablating the TRA-1 binding motif activated expression in hermaphrodite body wall muscle (Fig. 2.4*C*). These results collectively indicate that TRA-1 directly represses transcription of heterochronic genes in hermaphrodites.

Figure 2.4. TRA-1 binding sites near four heterochronic regulatory genes drive male-specific expression. (*A*) TRA-1 associates with chromatin near *lin-42*, *kin-20*, *lin-4*, and *lin-28*. TRA-1-ChIP (red) and input DNA tracks (black) are shown at same scale. Close matches to the TRA-1 consensus binding motif are indicated by vertical black bars above the ChIP trace. (*B*) Regions surrounding TRA-1 binding sites near *lin-42*, *kin-20*, and *lin-4* drive reporter gene expression in male-specific cell lineages in the tail, including the P10.p lineage (bar) and B lineage (bracket). (*C*) The region surrounding the TRA-1 binding site near *lin-28* drives male-specific expression in body wall muscle (arrows), and the sex-specificity of this expression is dependent on the presence of a TRA-1 motif. Arrowheads indicate the position of the hermaphrodite vulva (vu).


A
500 bp
50
C. elegans
TRA-1
ChIP
0
Input
lin-42
2kb
kin-20
lin-4
lin-28
B
lin-42::gfp
kin-20::gfp
lin-4::gfp
p10.p
B
C
lin-28::gfp
vu
lin-28::gfp ΔTRA-1

TRA-1 feeds back onto the sex-determination pathway at multiple levels

We also identified TRA-1 binding sites representing two forms of feedback regulation on the sex-determination pathway. First, the *tra-1* locus contained three TRA-1 binding sites (Fig. 2.5*A*). Genetic evidence supports *tra-1* autoregulation: the feminizing effect of dominant *tra-1* gain-of-function alleles can be reduced by adding wild-type copies of the gene (J. Hodgkin, 1993), and many recessive hypomorphic *tra-1* alleles show equivalent masculinization when either homozygous or hemizygous (Schedl, Graham, Barton, & Kimble, 1989). Second, we observed apparent feedback on upstream sex-determination genes. TRA-1 can bind to a site in the *xol-1* promoter *in vitro*, and regulates *xol-1* reporter transgenes via this site (Hargitai et al., 2009). ChIP-seq confirmed *in vivo* TRA-1 binding to this site (Fig. 2.5*A*), and also identified TRA-1 binding sites adjacent to *fem-3* and *sup-26* (Mapes, Chen, Yu, & Xue, 2010), suggesting more extensive upstream feedback (Fig. 2.5*A*, Fig. 2.7, Table 2.1). A site near *her-1* also showed TRA-1 binding, but below the significance cutoff applied. All of these upstream targets normally promote male development, suggesting that TRA-1 transcriptional repression provides multi-level reinforcement of the feminizing mode of the sex-determination switch.

To further explore feedback regulation we focused on *fem-3*. Regulation of *fem-3* in the XX germ line involves post-transcriptional repression (Ahringer & Kimble, 1991), but expression of FEM-3 in somatic tissues has not been described. A FEM-3::GFP fosmid reporter was expressed in the vas deferens region of the male somatic gonad, but undetectable in the hermaphrodite gonad (Fig. 2.6*A*). Ablating the TRA-1 motif activated expression in the somatic gonad of the hermaphrodite, with robust expression restricted to the spermatheca (Fig. 2.6*A*), though this ectopic FEM-3 expression did not result in

any obvious developmental defects. We also examined endogenous FEM-3 expression by immunofluorescence, which confirmed stronger expression in male than hermaphrodite soma (Fig. 2.6*B*). We conclude that somatic expression of *fem-3* involves transcriptional feedback regulation by TRA-1, which potentially functions to reinforce the global sex-determination decision.

Figure 2.5. TRA-1 feeds back on the sex-determination pathway at multiple levels. (*A*) *C. elegans* TRA-1 binds near the *tra-1*, *xol-1*, *sup-26*, and *fem-3* loci. (*B*) *C. briggsae* TRA-1 binds near the *Cbr-tra-1*, *Cbr-xol-1*, and *Cbr-fem-3* loci. A site near *Cbr-sup-26* showed occupancy by CBR-TRA-1, but below the threshold we applied. (*C*) Alignments of TRA-1 motifs (underlined) identified near the *xol-1* and *fem-3* loci in *C. elegans*, *C. briggsae*, *C. brenneri*, and *C. remanei*.

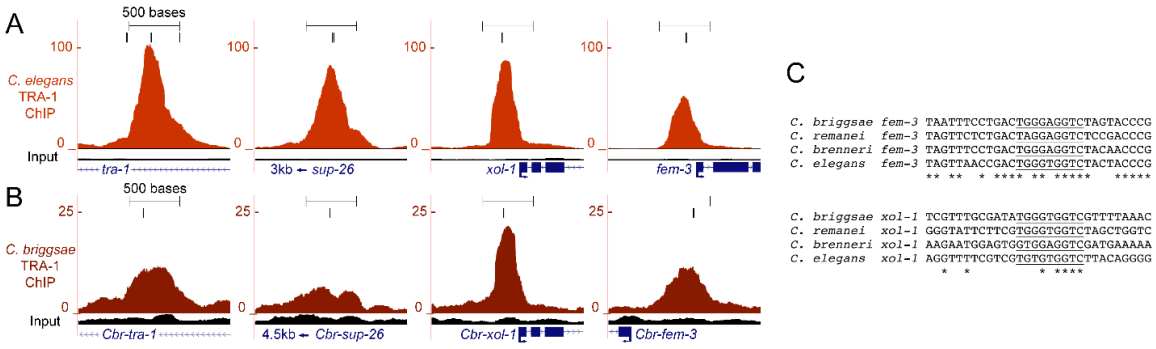
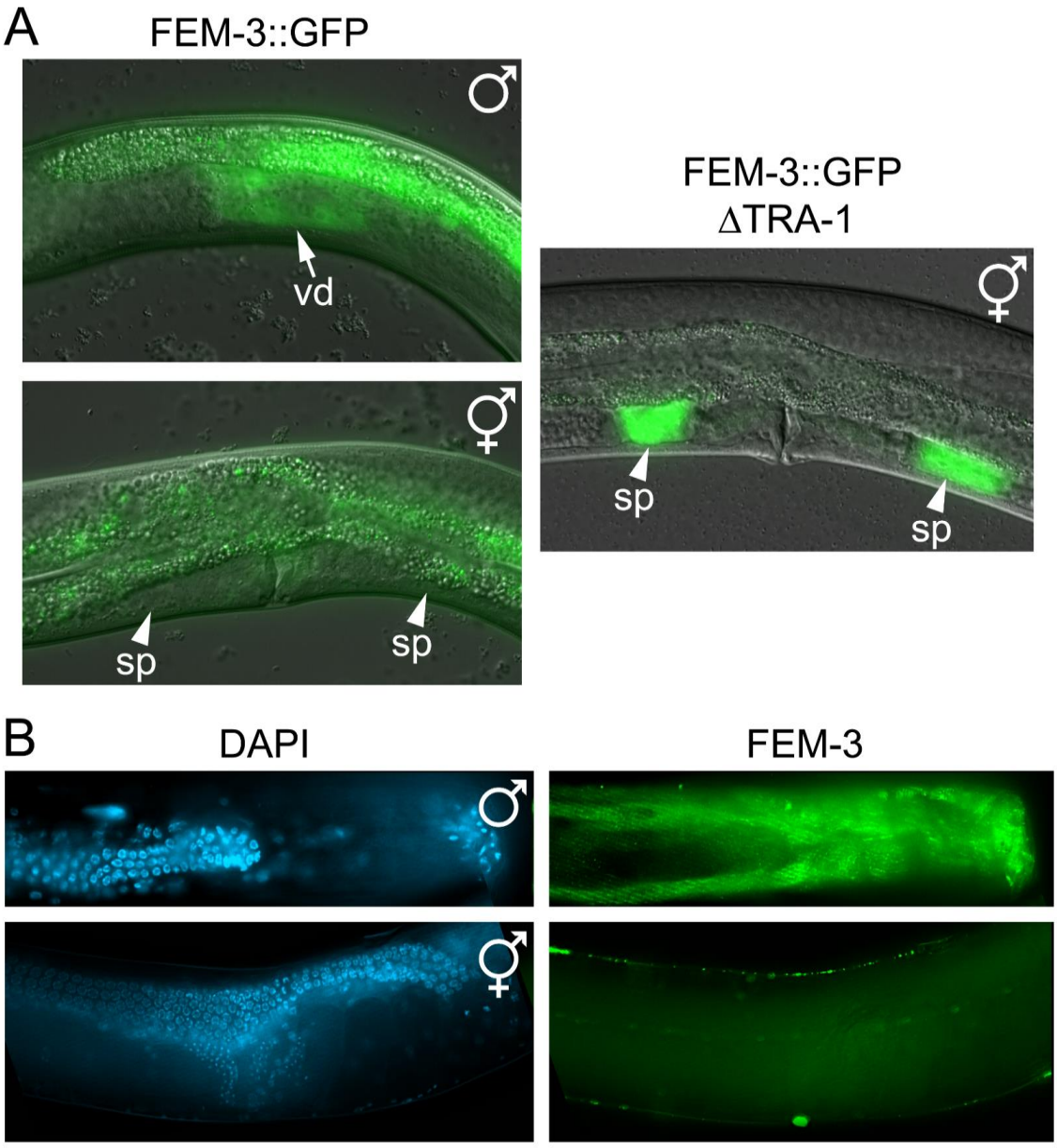

A
500 bases
100
C. elegans
TRA-1
ChIP
0
Input
tra-1
3kb ← sup-26
xol-1
fem-3
B
500 bases
25
C. briggsae
TRA-1
ChIP
Input
Cbr-tra-1
4.5kb ← Cbr-sup-26
Cbr-xol-1
Cbr-fem-3
C
C. briggsae fem-3 TAATTTCCTGACTGGGAGGTCTAGTACCCG
C. remanei fem-3 TAGTTCTCTGACTAGGAGGTCTCCGACCCG
C. brenneri fem-3 TAGTTTCCTGACTGGGAGGTCTACAACCCG
C. elegans fem-3 TAGTTAACCGACTGGGTGGTCTACTACCCG
C. briggsae xol-1 TCGTTTGCGATATGGGTGGTCGTTTTAAAC
C. remanei xol-1 GGGTATTCTTCGTGGGTGGTCTAGCTGGTC
C. brenneri xol-1 AAGAATGGAGTGGTGGAGGTCGATGAAAAA
C. elegans xol-1 AGGTTTTCGTCGTGTGTGGTCTTACAGGGG

Figure 2.6. TRA-1 regulates *fem-3* expression. (*A*) A FEM-3::GFP reporter is expressed in the vas deferens (arrow) of the male somatic gonad, but shows no expression in the hermaphrodite somatic gonad. FEM-3::GFP ΔTRA-1 is expressed in the spermatheca (arrowheads) of the hermaphrodite somatic gonad. (*B*) FEM-3 antibody staining shows somatic expression of FEM-3 is normally restricted to males.


A
FEM-3::GFP
vd
sp
sp
FEM-3::GFP
ΔTRA-1
sp
sp
B
DAPI
FEM-3

TRA-1 targets are evolutionarily conserved

TRA-1 homologs determine sex in other nematode species including *Pristionchus pacificus* (Pires-daSilva & Sommer, 2004) and *C. briggsae*, which are diverged from *C. elegans* by about 200 million years (MY) and 100 MY respectively. Despite the evolutionary stability of the TRA-1 sex-determining function, it is unknown whether any genes regulated by TRA-1 in *C. elegans* are also regulated in other species, although sequence analysis has shown that TRA-1 binding motifs are present near the *C. briggsae* orthologues of some identified *C. elegans* TRA-1 targets, including *mab-3* and *fog-3* (P. J. Chen et al., 2001; Yi et al., 2000). The TRA-1 antibody recognizes *C. briggsae* TRA-1 (Arur et al., 2011), so we performed ChIP-seq in *C. briggsae* L3 larvae, and identified 78 CBR-TRA-1 binding sites (Table 2.3). CBR-TRA-1 binding sites were highly enriched for the TRA-1 motif, which appears to be identical in the two species and was found at 54 of 78 (69%) CBR-TRA-1 binding sites. For 64 of 78 CBR-TRA-1 binding sites, syntenic alignments unambiguously identified a single region in the *C. elegans* genome corresponding to the bound region in *C. briggsae*. 19 of these 64 syntenic regions coincided with one of the 184 *C. elegans* TRA-1 binding sites (Table 2.4), suggesting substantial conservation of TRA-1 regulation through nematode evolution. Notably, sites representing feedback regulation of the sex-determination pathway, including those adjacent to *tra-1*, *fem-3*, and *xol-1*, showed TRA-1 occupancy in both species (Fig. 2.5*B*). A site near *Cbr-sup-26* also showed weak CBR-TRA-1 binding, but this was below the significance cutoff. Likewise, TRA-1 binding adjacent to the heterochronic genes *daf-12*, *kin-20* and *lin-14* is also conserved, suggesting that these binding events likely mediate biologically relevant regulation.

Table 2.3. 78 sites in the *C. briggsae* genome show significant CBR-TRA-1 ChIP enrichment in both replicates at the L3 stage. Sites are ordered by chromosomal position (column A). Column B indicates whether each site shares an orthologous TRA-1 binding site in *C. elegans*. Column C indicates the coding gene with the transcriptional start site nearest to the TRA-1 binding site. Column D indicates all genes within 10 kb of the TRA-1 binding site.

Peak center	**Nearest gene**	**Conserved in *C. elegans*?**	**All genes within 10 kb of site**
chrI : 13237778	F58D5.5	yes	F58D5.5, F58D5.9, Y47H9C.7, Y47H9C.8, hrp-2
chrI : 13798238	RITIcx1693		F22G12.4, RITIcx1693, RITIcx223, RITIcx2590, RITIcx2783, RITIcx829
chrI : 14827774	pbs-5		F31C3.6, K05C4.11, K05C4.2, RITIcx2013, RITIcx2132, Y54E5A.8, pbs-5
chrI : 15270102	C49G9.1		C49G9.1, C49G9.2, T07D10.2, dlk-1
chrI : 235622	T28F4.1	yes	RITIcx2888, T28F4.1, asic-2
chrI : 2730581	RITIcx2084		RITIcx2084, ppfr-1
chrI : 3080227	D1007.18		D1007.10, D1007.18, D1007.19, D1007.3, RITIcx440, RITIcx680, rpl-24.1
chrI : 5301075	K10C3.2	yes	K10C3.2, RITIcx648, zig-1
chrI : 6295162	D2005.6		D2005.6, R11A5.4, RITIcx2296
chrI : 717335	RITIcx1578		RITIcx1578, RITIcx1908
chrI : 82882	spr-4		F18C12.3, spr-4
chrII : 11560488	daf-5		RITIIcx720, daf-5
chrII : 13448967	RITIIcx2921		RITIIcx1492, RITIIcx2811, RITIIcx2921, RITIIcx3074
chrII : 15058944	str-52		RITIIcx2135
chrII : 3344987	ctns-1		F10B5.8, RITIIcx1624, ctns-1, ire-1, ulp-4
chrII : 3481527	mig-5		F21H12.7, RITIIcx1891, RITIIcx1935, RITIIcx2029, T05C12.8, T05C12.9, cct-1, dylt-3, mig-5
chrII : 4229168	rpl-5		F54C9.14, F54C9.3, F54C9.7, F54C9.9, RITIIcx741, bcs-1, col-38, rpl-5, stc-1
chrII : 4287420	pqn-73		RITIIcx1818, RITIIcx1928, RITIIcx228, pqn-73
chrII : 4340387	RITIIcx2671		R05H10.1, R05H10.7, RITIIcx2671, cdh-7

Peak center	**Nearest gene**	**Conserved in *C. elegans*?**	**All genes within 10 kb of site**
chrII : 4343058	R05H10.1		R05H10.1, R05H10.7, RITIIcx2671, cdh-7
chrII : 4357247	K02B7.3		F26H11.6, K02B7.3, R05H10.7, RITIIcx2968
chrII : 5236742	RITIIcx298		C18H9.5, RITIIcx1276, RITIIcx1653, RITIIcx298
chrII : 64392	cutl-2	yes	RITIIcx1849, Y53C12A.7, Y53C12B.7, cutl-2, mab-3
chrII : 8928962	ztf-27		T09F3.2, ztf-27
chrII : 9393776	C32D5.8		C32D5.10, C32D5.11, C32D5.7, C32D5.8, RITIIcx216, lgg-1
chrII : 949962	tag-165		C01G6.5, acs-7, tag-165
chrIII : 10206018	RITIIIcx523		F54F2.7, RITIIIcx523, RITIIIcx691, prx-19, ztf-1
chrIII : 11568504	glod-4		C16C10.9, RITIIIcx2687, glod-4, wht-3
chrIII : 12409078	ncl-1		RITIIIcx1084, RITIIIcx2739, ncl-1
chrIII : 12414616	RITIIIcx2739		RITIIIcx1581, RITIIIcx2739, RITIIIcx599, ncl-1
chrIII : 1500846	flp-14		Y37D8A.16, cco-2, flp-14
chrIII : 5174689	Y39A1A.9	yes	RITIIIcx1744, Y39A1A.12, Y39A1A.13, Y39A1A.9, dhs-11
chrIII : 5177473	RITIIIcx1744	yes	RITIIIcx1744, Y39A1A.12, Y39A1A.9, dhs-11
chrIII : 5690802	C26E6.3	yes	C26E6.3, C27F2.4, F09F7.3, RITIIIcx1344, RITIIIcx2289, flh-2
chrIII : 5842765	tra-1	yes	RITIIIcx1562, RITIIIcx2080, tra-1
chrIII : 5845101	tra-1		RITIIIcx1562, tra-1
chrIII : 5861049	RITIIIcx1917		
chrIII : 6456958	RITIIIcx1077		F38H4.6, RITIIIcx1077, RITIIIcx1725, RITIIIcx1791, RITIIIcx2347, unc-116
chrIII : 8306579	C14B1.8		C14B1.12, C14B1.6, C14B1.8, C14B1.9, F34D10.3, RITIIIcx2131, RITIIIcx2258, RITIIIcx437, syp-2

Peak center	**Nearest gene**	**Conserved in *C. elegans*?**	**All genes within 10 kb of site**
chrIV : 17096198	Y55F3AM.14	yes	K08C9.1, K08C9.2, RITIVcx1204, Y55F3AM.11, Y55F3AM.14
chrIV : 17102957	RITIVcx1425	yes	RITIVcx1204, RITIVcx1425, RITIVcx2235, RITIVcx2283, Y55F3AM.14
chrIV : 2584368	lst-4		RITIVcx3222, RITIVcx631, lst-4
chrIV : 4393894	fem-3	yes	RITIVcx421, aly-1, fem-3
chrIV : 503519	RITIVcx125		F26D10.11, RITIVcx125, RITIVcx1727, RITIVcx3209, atgp-1, clec-196
chrIV : 505965	RITIVcx125		F26D10.11, RITIVcx125, RITIVcx1641, RITIVcx1719, RITIVcx1727, RITIVcx2414, RITIVcx3209, atgp-1, clec-196
chrIV : 8666177	F40F11.3		F12F6.1, F40F11.2, F40F11.3, F40F11.4, RITIVcx128, RITIVcx664
chrIV : 8669739	F40F11.4		F12F6.1, F40F11.2, F40F11.3, F40F11.4, RITIVcx163
chrV : 10770915	nhr-232		C12D8.1, C12D8.9, akt-1, col-148, nhr-232, ric-4, rop-1
chrV : 1157238	RITVcx3257		F41E6.15, F41E6.5, F41E6.7, F41E6.9, RITVcx3257, RITVcx554, smk-1, tag-196
chrV : 11596809	F44E7.5		C28F5.4, F44E7.2, F44E7.5, F44E7.7, F44E7.9, nhr-142
chrV : 11962141	K08B12.3	yes	K08B12.3, W06H8.6, rme-1
chrV : 11965916	W06H8.6		K08B12.3, W06H8.6
chrV : 13362140	RITVcx788		K07B1.6, K07B1.7, RITVcx788, coq-6, try-5, ucp-4
chrV : 15147176	RITVcx1517		H39E23.2, RITVcx1517, par-1
chrV : 15420942	acl-2		acl-2, lea-1
chrV : 15436743	F57A8.1	yes	F57A8.1
chrV : 5174150	C39F7.5		C39F7.1, C39F7.5, rab-1
chrV : 6090376	sri-4		gei-7, sri-4, srp-1

Peak center	**Nearest gene**	**Conserved in *C. elegans*?**	**All genes within 10 kb of site**
chrV : 9388694	pyc-1		D2023.3, D2023.5, RITVcx1985, RITVcx2306, RITVcx603, RITVcx754, pyc-1
chrX : 10553020	H28G03.1		H28G03.1, H28G03.2, RITXcx193, pqn-62
chrX : 11149158	hbl-1		hbl-1
chrX : 12194961	mdl-1		R03E9.2, abts-4, mdl-1
chrX : 13770267	C34D10.1		C34D10.1, F45E1.1, RITXcx103
chrX : 13797945	RITXcx1939		RITXcx1872, RITXcx1939, RITXcx2332, xol-1
chrX : 1681119	F57G12.1		F57G12.1
chrX : 17184558	RITXcx1058		RITXcx1058, RITXcx1913, RITXcx2426, RITXcx272
chrX : 17710315	rab-14		K09E9.3, ent-2, erv-46, nipi-3, rab-14
chrX : 17711844	rab-14		K09E9.3, ent-2, erv-46, nipi-3, rab-14
chrX : 18067286	kin-20	yes	F46F2.5, RITXcx1358, jud-4, kin-20
chrX : 18811378	RITXcx2809		RITXcx1771, RITXcx2809, RITXcx3079, T24C2.5
chrX : 2819398	F52G3.1	yes	F52G3.1, RITXcx3687, RITXcx422
chrX : 4438001	RITXcx2925		RITXcx1138, RITXcx2925, lgc-40
chrX : 4843547	RITXcx1227		RITXcx1227, RITXcx2006, ftt-2
chrX : 5080456	lin-14	yes	lin-14
chrX : 5082986	lin-14		F48C5.2, lin-14
chrX : 6394175	daf-12	yes	M79.2, daf-12
chrX : 7108033	RITXcx754	yes	F21A10.1, F21A10.4, RITXcx1303, RITXcx3488, RITXcx754, nucb-1
chrX : 8165713	RITXcx1033		K10B3.5, K10B3.6, RITXcx1033, RITXcx173, RITXcx726

Table 2.4. 19 orthologous sites show binding by TRA-1 in the *C. elegans* genome and binding by CBR-TRA-1 in the *C. briggsae* genome. Sites are ordered by chromosomal position (column A). Columns B-E indicate the number of replicates that showed significant enrichment for each developmental condition at each site; 2 indicates significant enrichment was detected in both replicates, 1 indicates significant enrichment in 1 of 2 replicates, and 0 indicates significant enrichment in neither replicate. Column F indicates the coding gene with the transcriptional start site nearest to the TRA-1 binding site (in *C. elegans*). Column H indicates all genes within 10 kb of the TRA-1 binding site (in *C. elegans*).

Peak center	**Nearest gene**	**N2 L2**	**N2 L3**	**spe-11 YA**	**glp-4 YA**	**Conserved in *C. Briggsae*?**	**All genes within 10 kb of site**
chrI : 12039858	F58D5.5	0	1	2	2	yes	F58D5.5, F58D5.6, F58D5.7, F58D5.8, F58D5.9, ksr-2
chrI : 7475444	T28F4.1	0	1	2	2	yes	T28F4.1, T28F4.3, ZK524.4, asic-2
chrI : 9854145	K10C3.2	0	1	1	2	yes	K10C3.2, K10C3.4, K10C3.5, nhr-62, zig-1
chrII : 9728870	cutl-2	2	2	2	2	yes	Y53C12A.11, Y53C12A.6, Y53C12B.2, cutl-2, mab-3
chrIII : 10632610	Y39A1A.9	0	2	2	2	yes	Y39A1A.10, Y39A1A.12, Y39A1A.27, Y39A1A.8, Y39A1A.9, dhs-11, hpr-9
chrIII : 10634106	Y39A1A.9	0	1	2	1	yes	Y39A1A.10, Y39A1A.12, Y39A1A.27, Y39A1A.8, Y39A1A.9, dhs-11, hpr-9
chrIII : 11185869	tra-1	0	2	2	2	yes	tra-1
chrIII : 4950080	flh-2	0	1	1	2	yes	C26E6.1, C26E6.3, C27F2.4, flh-2, fsn-1, rpb-2, vps-22
chrIV : 1064617	Y55F3AM.14	0	0	2	2	yes	Y55F3AM.1, Y55F3AM.13, Y55F3AM.14, dcap-1
chrIV : 1070629	Y55F3AM.14	0	1	1	2	yes	Y55F3AM.14, csn-4
chrIV : 9105139	fem-3	2	2	2	2	yes	C01F6.2, C01F6.9, aly-1, cpna-3, fem-3, icl-1, lpl-1
chrV : 10032506	acl-2	0	0	2	2	yes	F57A8.1, K08H10.2, T06E8.2, acl-2, lea-1
chrV : 5439212	F08F3.9	2	2	2	2	yes	F08F3.1, F08F3.8, F08F3.9, acl-6, glc-3, rhr-1
chrV : 6224451	W06H8.6	0	0	2	2	yes	W06H8.6, rme-1, str-206
chrX : 10188704	F21A10.4	0	1	2	2	yes	F21A10.1, F21A10.2, F21A10.4, nucb-1
chrX : 10637590	daf-12	0	1	2	1	yes	M79.2, daf-12
chrX : 11465227	lin-14	0	1	2	1	yes	lin-14

Peak center	**Nearest gene**	**N2 L2**	**N2 L3**	**spe-11 YA**	**glp-4 YA**	**Conserved in *C. Briggsae*?**	**All genes within 10 kb of site**
chrX : 15275776	kin-20	0	1	2	2	yes	F46F2.5, jud-4, kin-20, ser-2
chrX : 16961323	F52G3.1	0	1	2	2	yes	F52G3.1, F52G3.4

Discussion

Here we have used ChIP-seq to identify 184 binding sites of the nematode global sexual regulator TRA-1. A pilot analysis of genes adjacent to these sites identified seven new TRA-1 target genes that appear to promote male sexual differentiation based on male-specific expression patterns and/or male-specific RNAi phenotypes. However, none of the targets we examined showed a hermaphrodite-specific reduction-of-function phenotype or expression pattern. Likewise, ChIP-seq indicated that TRA-1 feedback onto the upstream sex-determination pathway occurs exclusively on male-promoting genes (Fig. 2.7), consistent with their repression by TRA-1 in hermaphrodites. Additionally, 12/256 (4%) of the genes adjacent to sites bound by TRA-1 at L2 or L3 showed male-biased expression in published L4 RNA-seq analysis (Gerstein et al., 2010). In contrast, none of these genes showed hermaphrodite-biased expression. We cannot exclude that TRA-1 can activate as well as repress transcription like some other GLI proteins. However, our data support the view that TRA-1 functions primarily as a transcriptional repressor, generating sexually dimorphic gene expression by restricting the expression of its targets in the hermaphrodite.

Our data and previous genetic analyses (Okkema & Kimble, 1991; Schedl et al., 1989) indicate that TRA-1 has two distinct roles in regulating sexual development: it controls sex-specific cell and tissue differentiation by regulating a suite of downstream effector genes, and it feeds back at multiple upstream steps to affect the primary sex-determination decision (Fig. 2.7). TRA-1 binding near *fem-3* and *xol-1* has been conserved between *C. elegans* and *C. briggsae*, and sequences containing TRA-1 binding motifs also are conserved near these genes in *C. brenneri* and *C. remanei* (Fig. 2.5*C*).

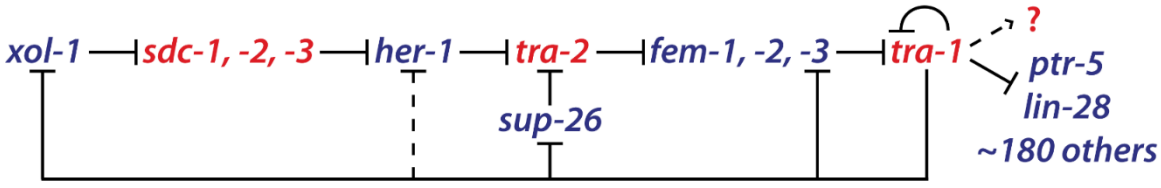
Reporter analysis and immunofluorescence confirmed the role of TRA-1 as a repressor of *fem-3* transcription in the hermaphrodite soma. Multicopy transgenes do not normally express in the germ line, so we have not yet examined whether this transcriptional feedback also occurs in germ cells; however the mutually exclusive germ line distribution of nuclear TRA-1 and FEM-3 (Arur et al., 2011) suggests that it may. Somatic FEM-3 expression has not been shown previously (Rosenquist & Kimble, 1988) but is supported by genetic data: somatic feminization of *tra-1(gf)* males is suppressed by gain-of-function mutations in the *fem-3* 3' UTR that do not affect *fem-3* transcript levels (Schedl et al., 1989). Thus FEM-3 activity is regulated through at least three separate means: the transmembrane receptor protein TRA-2 represses FEM-3 protein activity via physical interaction (Mehra et al., 1999), *fem*-3 mRNA is negatively regulated via the 3' UTR (B. Zhang et al., 1997), and we find that *fem-3* is also transcriptionally repressed by TRA-1. *fem-3* therefore appears to be an important node for control of the sex-determination pathway. Feedback from TRA-1 may serve at least two purposes. First, we speculate that it renders the global sex-determination decision more robust and helps ensure the coordinated execution of downstream sexual differentiation events. Second, feedback loops might provide additional inputs to control the temporary switch from female to male sex determination that occurs in hermaphroditic species during germ line development.

The TRA-1 binding sites we have identified provide a resource for identifying and functionally defining the suite of genes that control sexual differentiation in nematodes. The restricted male-specific expression patterns and limited male-specific developmental defects for putative target genes are consistent with the view that TRA-1 controls sexual

development by regulating many minor players in male development rather than a few major ones. To fully evaluate the roles these candidate target genes may play in sexual development it will be important to examine possible functional overlaps, such as those that occur between *mab-3* and *dmd-3* (Mason et al., 2008), as well as to perform phenotypic analysis in sensitized backgrounds and at finer scale.

Sex-determination pathways impose sex specificity on the spatial and temporal control of development in animals with diverse body plans, behaviors, and physiologies. The conservation of TRA-1 binding between *C. elegans* and *C. briggsae* suggests substantial evolutionary conservation of sexual development within nematodes. However, TRA-1 is among the fastest evolving genes in Caenorhabditids (de Bono & Hodgkin, 1996), and its orthologues do not appear to regulate sex determination outside nematodes. *DMRT* genes (orthologues of the TRA-1 targets *mab-3*, *mab-23*, and *dmd-3*) determine sex in many other animals including insects and vertebrates (Matson & Zarkower, 2012). Comparing how TRA-1 and *DMRT* genes overlay sex specificity on development therefore may shed considerable light on the evolution of sexual development in metazoans.

Figure 2.7. Model for TRA-1 regulation of nematode sexual development. TRA-1 represses sex-determination genes that globally promote male development, including *xol-1, sup-26, fem-3,* and possibly *her-1,* as well as sexual differentiation genes that promote tissue-specific male sexual development, including *lin-28 and ptr-5*.


xol-1
sdc-1, -2, -3
her-1
tra-2
fem-1, -2, -3
tra-1
sup-26
?
ptr-5
lin-28
~180 others

Materials and Methods

Worm strains and culture

The following strains and alleles were obtained from the *Caenorhabditis* Genetics Center: wild type *C. elegans* (N2), wild type *C. briggsae* (AF16), LG I: *spe-11(hc77)*, *glp-4(bn2),* LG II: *rrf-3(pk1426),* LG III: *tra-1(e1834),* LG IV: *him-8(e1489),* hT2[*bli-4(e937)* let-?(*q782*) qIs48] (I;III). Animals were maintained in nematode growth medium (NGM) plates with added bacteria (*Escherichia coli* strain OP50) at 22°C, unless otherwise indicated.

ChIP-seq

To generate material for ChIP, gravid hermaphrodites were bleached and their eggs were collected and allowed to hatch overnight in M9. Synchronized animals were then grown in standard liquid culture conditions (wormbook.org) using HB101 bacteria as a food source, and staged by gonad migration and time. N2, AF16, and *tra-1(e1834)*/hT2[*bli-4(e937)* let-?(*q782*) qIs48] animals were grown at 20°C; *spe-11(hc77)* and *glp-4(bn2)* animals were grown at 25°C. *tra-1(e1834)* homozygotes were sorted from *tra-1(e1834)*/hT2[*bli-4(e937)* let-?(*q782*) qIs48] siblings using a COPAS BIOSORT SELECT (Union Biometrica, 350-5000-000) to collect GFP-negative animals. Two consecutive rounds of sorting were used to ensure a pure population of *tra-1(e1834)* homozygotes. Collected animals were separated from residual bacteria by sucrose flotation and washes in M9, and then frozen in liquid nitrogen.

Frozen animals were physically disrupted by grinding in liquid nitrogen in a mortar and pestle, fixed in 1%PFA in PBS for 15 min at RT and post-fixed with 50mM glycine for 5 min at RT. All subsequent steps were performed at 4°C. Material was pelleted, washed

3x in PBS with protease inhibitors, and resuspended in 1%SDS lysis buffer. Fixed chromatin was sheared with a Diagenode Bioruptor Standard, on the high setting for 7.5 min total on time (15 min of 30 sec on, 30 sec off cycles). ChIP was performed as in (Murphy et al., 2010), using 2ug anti-TRA-1 antibody UMN163 per sample. Two biological replicates were performed for each developmental condition, using ~2 g frozen worms per replicate, with the exception of a single replicate of *tra-1(e1834),* which was performed on ~80,000 sorted animals.

DNA recovered from ChIP was ligated to sequencing adapters containing an 8-base multiplexing barcode sequence. DNA libraries were sequenced to 36 or 50 nucleotides by Illumina Genome Analyzer II or Hi-seq instruments. All *C. elegans* reads were subsequently trimmed to the first 27 bases and aligned to the ce10 reference genome using Bowtie (Langmead, Trapnell, Pop, & Salzberg, 2009), accepting only uniquely mapped reads with up to two mismatches. For *C. briggsae*, 48-base reads were aligned using the same Bowtie program settings to the latest cb4 reference genome for peak analyses and to the cb3 reference genome for signal visualization. To visualize ChIP-seq signals in the UCSC genome browser and in the heat map, aligned reads were extended to 200 bases, and the number of reads overlapping with each base of the genome was counted using Bedtools (Quinlan & Hall, 2010). The per-base read counts were then normalized to the genome-wide average in order to take the sequencing depth into account. Raw sequencing reads and processed data for genome browser visualization are available at Gene Expression Omnibus (GEO) website under accession ID GSE48917.

TRA-1 binding site identification

ChIP enriched regions were identified by the MACS2 program (Y. Zhang et al., 2008) using input DNA sequencing reads as a reference. For *C. elegans* data sets, regions fulfilling a MACS2 p-value threshold of less than 10^{-5} and a fold enrichment threshold of greater than 4 were extracted. This threshold identified 86 regions in L2-stage replicate 1 (DZ038); 111 regions in L2-stage replicate 2 (DZ039); 86 regions in L3-stage replicate 1 (DZ012); 337 regions in L3-stage replicate 2 (DZ015); 2480 regions in *spe-11* replicate 1 (DZ023); 176 regions in *spe-11* replicate 2 (DZ050); 129 regions in *glp-4* replicate 1 (DZ042); and 137 regions in *glp-4* replicate 2 (DZ043) (Supplemental Table 2). Regions that were identified in both biological replicates of a single condition (≥1 base pair (bp) overlap) were subsequently extracted and merged. This process narrowed the candidate regions to 72 regions for the L2 stage; 81 regions for the L3 stage; 163 regions for *spe-11* young adults; and 107 regions for *glp-4* young adults. Because some regions were identified in more than a single biological stage, we collected all regions in a single list and merged overlapping regions (≥ 1 bp). This yielded 257 candidate TRA-1-binding regions. Finally, we removed 73 candidate regions that overlap with regions with non-specific TRA-1 ChIP-seq signals (see below). We reported the remaining 184 regions (Supplemental Table 1) as high confidence TRA-1 binding sites in *C. elegans.*

To determine regions with non-specific TRA-1 ChIP-seq signal, we used TRA-1 ChIP-seq data generated in the *tra-1(e1834)* null mutant. We first made 500 bp windows across the *C. elegans* genome with a 250 bp offset and determined the *tra-1* mutant ChIP-seq read count for each window. These per-window counts followed a normal distribution with a mean count of 23.4 and standard deviation of 15.6. We identified 6439 windows whose read counts exceed 59.7 (p-value < 0.01) as regions with non-specific TRA-1

ChIP signals.

For *C. briggsae* data sets, ChIP-enriched regions fulfilling a MACS2 p-value threshold of less than 10^{-5} and a fold enrichment threshold of greater than 3 were extracted. This threshold identified 377 regions in replicate 1 (DZ040) and 82 regions in replicate 2 (DZ041). Seventy-eight regions were found in both replicates and identified as CBR-TRA-1-binding sites at the L3 stage.

Motif analysis

100 bp regions centered around TRA-1 binding sites were searched for DNA motifs using the MEME-chip program (Machanick & Bailey, 2011) with a background Markov model constructed with genome-wide di-nucleotide frequencies. 2713 TRA-1 motif sites in the *C. elegans* genome were identified by FIMO program (Grant, Bailey, & Noble, 2011) using a query DNA motif constructed by the 9-mer TRA-1 interacting sequences defined *in vitro* (Zarkower & Hodgkin, 1993), and a p-value threshold of 10^{-5}.

Expression analysis

500-1000 bp regions surrounding TRA-1 binding sites were cloned into pPD107.94 (Fire, Harrison, & Dixon, 1990). ΔTRA-1 reporters were generated by replacing the invariant GG in the gggtGGtc consensus binging motif with AA, using the Agilent QuikChange XL site-directed mutagenesis kit. FEM-3::GFP was generated by recombineering fosmid WRM067dB02 according to (Tursun, Cochella, Carrera, & Hobert, 2009), with *gfp* coding sequence replacing the *fem-3* stop codon. FEM-3::GFP ΔTRA-1 was also generated by recombineering according to (Tursun et al., 2009), making the same change as described above. Reporter constructs were injected with *str-1::gfp* as a co-injection marker (5-20 ng/uL reporter, 20 ng/uL *str-1::gfp*, sheared N2 genomic DNA to 100

ng/uL total DNA) into *him-8(e1489)* animals. Stable transgenic lines were assayed for *gfp* expression.

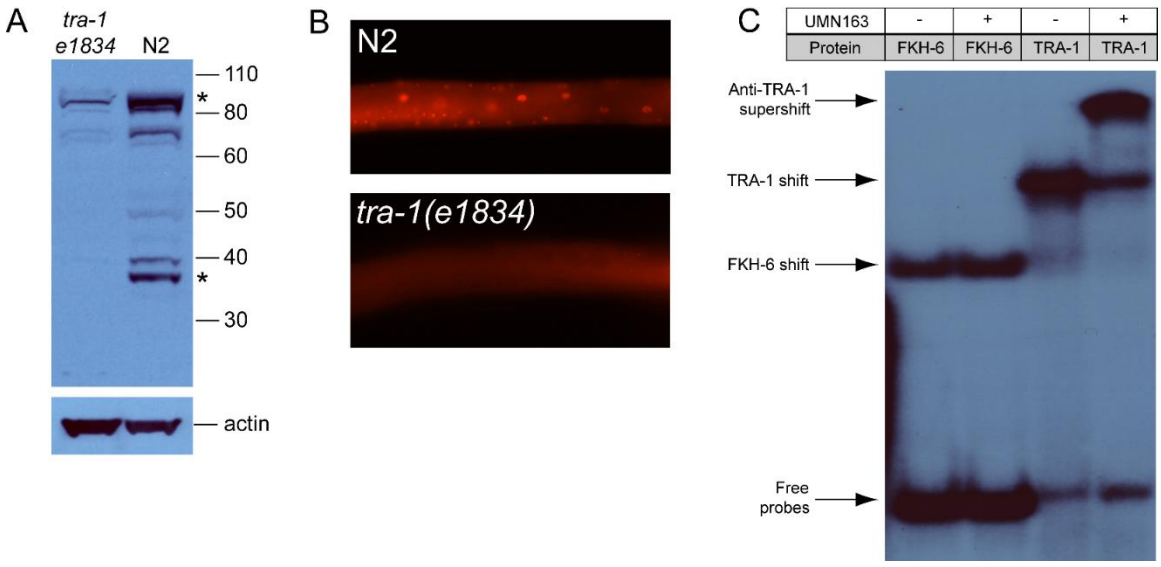
RNAi

him-8(e1489); *rrf-3(pk1426)* animals were used for all RNAi experiments, which were performed according to the standard *C. elegans* feeding RNAi protocol (wormbook.org). L4 hermaphrodite animals were placed on standard RNAi plates and their progeny were scored by DIC microscopy for defects in sexual development.

FEM-3 Antibody Staining

Antibody staining was as described in (Seydoux & Dunn, 1997). Wild type animals were analyzed as young adults. FEM-3 antibody was used at a 1:200 dilution.

Figure 2.8. Characterization of anti-TRA-1 antibody UMN163 used in this study. (A) Western blot: UMN163 recognizes bands corresponding in size to the two major species of TRA-1 in N2 animals but not in *tra-1(e1834)* mutant animals. Asterisks denote the approximate sizes of major TRA-1 isoforms. (B) Immunofluorescence: UMN163 shows nuclear staining in N2 animals but not in *tra-1(e1834)* mutant animals. (C) Gel-shift: UMN163 is able to supershift DNA-bound TRA-1, but not DNA-bound FKH-6.


A
tra-1
e1834
N2
110
80
60
50
40
30
actin
B
N2
tra-1(e1834)
C
UMN163
Protein
-
+
-
+
FKH-6
FKH-6
TRA-1
TRA-1
Anti-TRA-1
supershift
TRA-1 shift
FKH-6 shift
Free
probes

Chapter 3

Conclusions and Future Directions

Introduction

Chapter 2 described a biochemical approach designed to identify genes involved in sexual development in *C. elegans*. ChIP-seq identified a substantial fraction of the TRA-1 binding sites in the worm, generating a list of several hundred candidate sexual regulators that lie adjacent to the binding sites of this global sexual regulator. A preliminary examination of the function of these putative TRA-1 targets indicated that at least a few of them are expressed male-specifically and/or exhibit male-specific reduction-of-function phenotypes, consistent with transcriptional repression by TRA-1. This chapter will highlight a few important remaining questions regarding the role of TRA-1 and its targets in sexual development and describe future investigations aimed at addressing these questions.

Important questions concerning TRA-1 regulation of sexual development

What is the complete set of TRA-1 target genes?

The lists of 184 TRA-1 binding sites and 444 associated genes described in Chapter 2 likely only include about half of the total TRA-1 targets in *C. elegans*. Five of the eight previously identified TRA-1 targets are on this list, but the known targets *egl-1*, *ceh-30*, and *lin-39* are not. Not only did the sites near these genes fail to show the significant enrichment required to make the list of 184 TRA-1 binding sites, they showed essentially no enrichment at all in any of the TRA-1 ChIPs performed, despite compelling genetic and biochemical evidence that they are genuine TRA-1 binding sites (Conradt & Horvitz, 1999; Peden et al., 2007; Szabo et al., 2009).

The most plausible explanation for their absence in these experiments is that whole animal ChIP-seq using this particular TRA-1 antibody fails to identify all TRA-1 binding sites in the worm. It is conspicuous that the three TRA-1 targets not identified are each responsible for the sexually dimorphic development of only a few cells, particularly in light of the fact that experiments with germ line-less *glp-4* mutants suggest that TRA-1 exhibits at least some tissue-specific binding. This suggests that TRA-1 binding at some sites may be restricted to only a few cells, and it may not be possible to identify these sites with whole animal ChIP-seq using an antibody to TRA-1. Two possible approaches to circumvent this problem and identify additional TRA-1 targets will be discussed at the end of this chapter.

How do TRA-1 target genes control sexual development?

Although the list of 444 putative TRA-1 target genes appears to be incomplete, it likely represents a significant portion of the complete set of *C. elegans* TRA-1 targets. A handful of these putative TRA-1 target genes have loss-of-function phenotypes that include obvious defects in sexual development, including the previously known targets *mab-3* (J. Hodgkin, 1983a) and *dmd-3* (Mason et al., 2008). However only one of the 34 newly identified putative TRA-1 targets that were assayed, *ztf-6*, showed an obvious sexual differentiation defect when knocked down by RNAi.

This is likely due in part to the difficulty in assigning each TRA-1 binding site to a particular target gene. TRA-1 binding sites have been shown to regulate target genes at distances of at least a few kb (Conradt & Horvitz, 1999), and in the compact genome of *C. elegans*, there are often several genes within a few kb of a particular TRA-1 binding site. All of these nearby genes could be considered putative targets, but it is likely that

only a single gene is regulated per binding site, meaning the list of 444 genes associated with 184 TRA-1 binding sites likely includes a maximum of about 184 genuine TRA-1 targets.

Even taking this difficulty into account, it is still striking how few of the putative TRA-1 targets examined in Chapter 2 actually show defects in sexual development when knocked down by RNAi. Some of these genes may have subtle roles in sexual differentiation that are difficult to detect. Others may have roles that are masked by genetic redundancy, which has been observed in many developmental processes, including nematode sexual differentiation. The male-specific sexual regulators *mab-3* and *dmd-3* act in a partially redundant fashion to promote male tail tip morphogenesis (Mason et al., 2008). *mab-3* mutant males have minor tail tip retraction defects at low penetrance, and *dmd-3* mutant males have somewhat stronger retraction defects at higher penetrance, but *mab-3; dmd-3* double mutant males show a completely penetrant full tail tip retraction defect. It is possible that other TRA-1 targets have similarly redundant roles in promoting other aspects of male differentiation, which cannot be revealed in single-gene knockdown experiments.

Of course it is also possible that some of these TRA-1 binding sites do not function in sexual development, and are simply examples of transcription factor binding events that confer no significant benefit or detriment to the animal. Our cross-species ChIP data suggests that this is unlikely to be the case for at least a quarter of these sites, as roughly that proportion have been conserved through tens of millions of years of evolution, strongly suggesting that they are of biological significance. Determining the functional role of many of the putative TRA-1 targets on this list may be difficult, requiring

examinations at finer scale, in sensitized genetic backgrounds, or in compound mutants. However the presence of so many known sexual regulators on this list, including *mab-3*, *dmd-3*, *mab-23*, *ztf-6*, *fog-1*, *fog-3*, *tra-1*, *fem-3*, *sup-26*, and *xol-1*, suggests that it may be worth the effort, particularly for sites that are bound in both *C. elegans* and *C. briggsae*.

Do terminal sexual regulators in other systems control sexual differentiation via similar targets?

In addition to identifying the complete set of TRA-1 targets and determining what role they play in nematode sexual development, it will be interesting to see whether terminal sexual regulators in distantly related species control the homologues of these genes, and whether those homologues play similar roles in the sexual development of those species. Recent work in *Drosophila* has identified a partial set of targets for the terminal sexual regulator DSX (Luo, Shi, & Baker, 2011). None of the genes on this list of 23 DSX targets have a *C. elegans* homologue adjacent to a TRA-1 binding site, although this comparison is obviously limited by the small list of DSX targets. DSX ChIP-seq is currently being pursued (Mark Van Doren, personal communication) and will hopefully generate a more comprehensive list of DSX targets to use in such comparisons. Genome-wide binding site surveys will likely be performed for other terminal sexual regulators in the coming years, allowing for comparisons between regulators of sexual differentiation across diverse species that will greatly enhance our understanding of this important biological process and the degree to which its regulation is conserved. Recently evolved, clade-specific sexual differentiations are likely to be regulated by genes that have recently been recruited into a role in sexual differentiation and may not play similar roles in other clades. However, many sexual differentiation processes, such as sex-specific

neural development and sex-specific gonadogenesis, are likely evolutionarily ancient and may be controlled by conserved regulators of sexual differentiation that could be revealed in such comparisons.

What is the purpose of such extensive feedback in the C. elegans sex-determination pathway?

ChIP-seq identified TRA-1 binding sites near several genes known to function upstream of *tra-1* in the sex-determination pathway. In fact, evidence of TRA-1 binding was observed near masculinizing genes at every level of the core pathway, and these sites included some of the most convincing TRA-1 binding sites identified by ChIP-seq; they were among the most enriched sites in every ChIP performed (including in *C. briggsae*), they harbor excellent matches to the TRA-1 consensus binding sequence, and reporter analysis suggests that the sites adjacent to *xol-1* (Hargitai et al., 2009) and *fem-3* drive direct male-specific expression in a TRA-1 dependent manner. So these sites appear to be real and functional, which raises the question of why so many of them are present.

Feedback onto *fem-3*, which directly represses TRA-1 by targeting it for proteasomal degradation (Starostina et al., 2007), is likely a cell-autonomous form of feedback regulation. This may function to ensure that a cell expressing TRA-1 is incapable of expressing high levels of FEM-3 and cannot enter the male mode of sexual development even if upstream elements of the pathway or the cell's ability to interpret them are compromised.

As described in Chapter 1, HER-1 is a secreted factor that functions cell non-autonomously to promote male development, so TRA-1 binding sites near *her-1* and the upstream gene *xol-1* may represent cell non-autonomous forms of feedback regulation,

possibly functioning to coordinate the sex-determination decision between cells. *xol-1* controls both sex determination and dosage compensation, ensuring these processes are functioning in the same mode. Sex-inappropriate dosage compensation, either its ectopic activity in XO males or inactivity in XX hermaphrodites, is lethal. If the purpose of these feedback sites is to ensure faithful execution of the sex-determination pathway, feedback upstream of the DCC at the level of *xol-1* would ensure that this reinforcement of the female mode of sex determination also reinforces the female mode of dosage compensation activity.

Extensive feedback regulation may also be a relic of the evolutionary history of the pathway. If the pathway evolved in a bottom-up manner as has been hypothesized (Wilkins, 1995), TRA-1 feedback on every male-promoting gene in the pathway may simply represent a selective advantage for feedback from TRA-1 onto the most upstream male-promoting gene in the pathway throughout this evolutionary history.

How does TRA-1 regulate transcription?

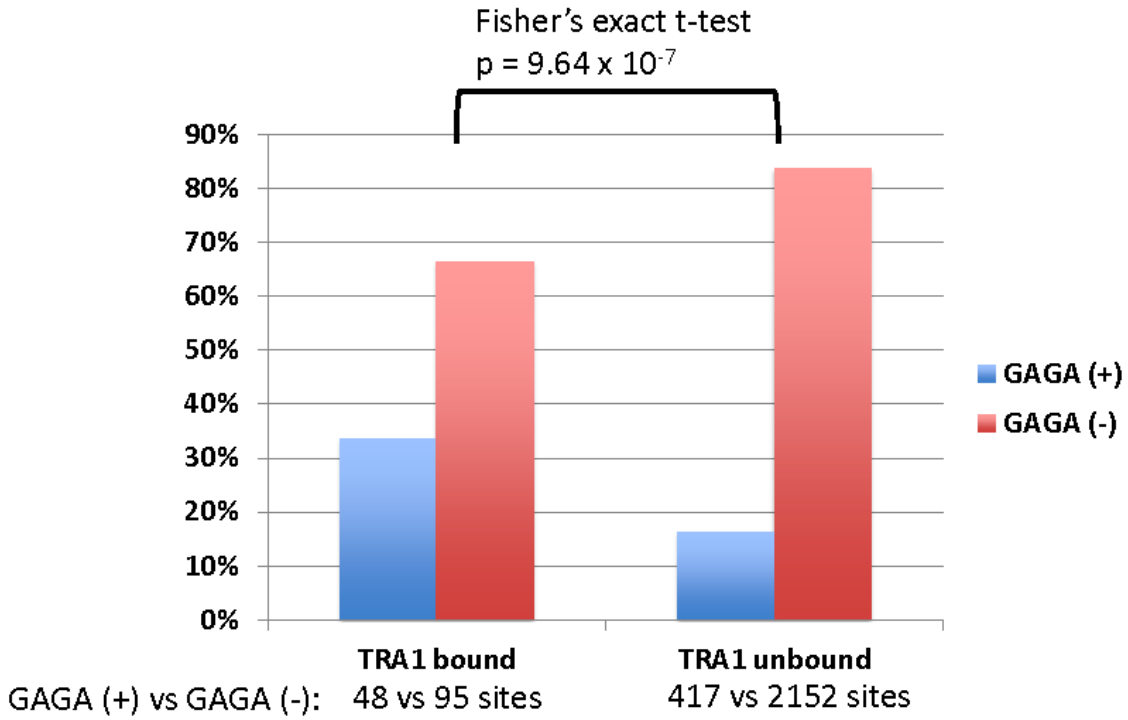
All TRA-1 target genes identified prior to this work are male-promoting genes that are transcriptionally repressed by TRA-1. Several of the new TRA-1 targets identified here show male-specific expression and/or male-specific developmental defects when knocked down, but none exhibited hermaphrodite-specific expression or hermaphrodite-specific reduction-of-function phenotypes. So TRA-1 appears to act primarily as a transcriptional repressor, but the molecular nature of this repression remains unknown.

Does TRA-1 repress transcription simply by physically binding to its targets and preventing the binding of positive regulators? Or does TRA-1 recruit other proteins to its binding sites, such as chromatin regulators that establish a repressive chromatin state?

Can *cis*-acting elements other than the TRA-1 consensus binding sequence influence TRA-1 binding?

The list of TRA-1 binding sites identified by ChIP-seq can be used to begin to address these questions of TRA-1 function. In Chapter 2, the MEME program was used to determine the *in vivo* TRA-1 consensus binding sequence, which was shown to be a close match to the *in vitro* motif (Zarkower & Hodgkin, 1993). MEME also identified a relatively unstructured GAGA motif as enriched at TRA-1 binding sites. This motif is roughly twice as likely to be found at TRA-1 consensus binding sequences that are bound by TRA-1 than at consensus binding sequences that are not bound by TRA-1 (Figure 3.1), suggesting that the presence of this GAGA motif may facilitate TRA-1 binding, either directly or through the recruitment of a TRA-1 co-regulator.

Figure 3.1. Motif searching identifies a GAGA motif enriched at TRA-1 binding sites. The 2713 closest matches to the TRA-1 consensus binding sequence in the *C. elegans* genome were grouped into two categories, TRA-1 bound and TRA-1 unbound, based on whether or not they showed significant TRA-1 occupancy in L3 stage TRA-1 ChIP. 48/143 (34%) of TRA-1 bound sites contain a GAGA motif, while 417/2569 (16%) of TRA-1 unbound sites contain this motif. This difference is statistically significant.


Fisher's exact t-test
p = 9.64 x 10-7
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
TRA1 bound
TRA1 unbound
GAGA (+)
GAGA (-)
GAGA (+) vs GAGA (-): 48 vs 95 sites
417 vs 2152 sites

The modENCODE consortium has profiled the occupancy of several chromatin marks using ChIP-seq (Gerstein et al., 2010). When our TRA-1 binding sites are compared to these marks, few correlations are evident, suggesting that recruitment of chromatin modifiers that regulate these particular marks is not a major mode of TRA-1 function (Figure 3.2). There is a weak association between TRA-1 binding and H3K4 tri-methylation, which is a chromatin mark associated with active promoter regions. However this may simply reflect the tendency of TRA-1 binding sites to lie in promoter regions rather than reveal a means of TRA-1 transcriptional repression. Perhaps more detailed mapping of the chromatin state of the *C. elegans* genome will reveal more meaningful correlations between TRA-1 binding and chromatin marks that will provide clues as to how TRA-1 represses transcription.

The modENCODE consortium also used ChIP-seq to identify binding sites for many site-specific transcription factors (Niu et al., 2011). A comparison of the occupancy of these transcription factors at TRA-1-bound vs. TRA-1-unbound TRA-1 binding motifs shows a modest relationship between TRA-1 occupancy and the occupancy of two of these transcription factors (Figure 3.3). TRA-1 bound sites are more likely to be bound by the homeobox protein CEH-9, and less likely to be bound by zinc-finger transcription factor EOR-1, than TRA-1-unbound sites. This suggests CEH-9 may serve to facilitate and EOR-1 binding to restrict TRA-1 binding to its recognition sequence, and that the presence or absence of other site specific transcription factors may regulate TRA-1 binding at other sites.

Figure 3.2. A comparison of TRA-1 occupancy and chromatin mark occupancy. The 2713 closest matches to the TRA-1 consensus binding sequence in the *C. elegans* genome were ordered based on their degree of enrichment in L3 stage TRA-1 ChIP. Normalized ChIP reads within a 4 kb window centered on each TRA-1 motif are shown for TRA-1 and 8 chromatin marks. TRA-1 and H3K4me3 show a weak positive correlation. No correlation is seen between TRA-1 and any of the other marks.

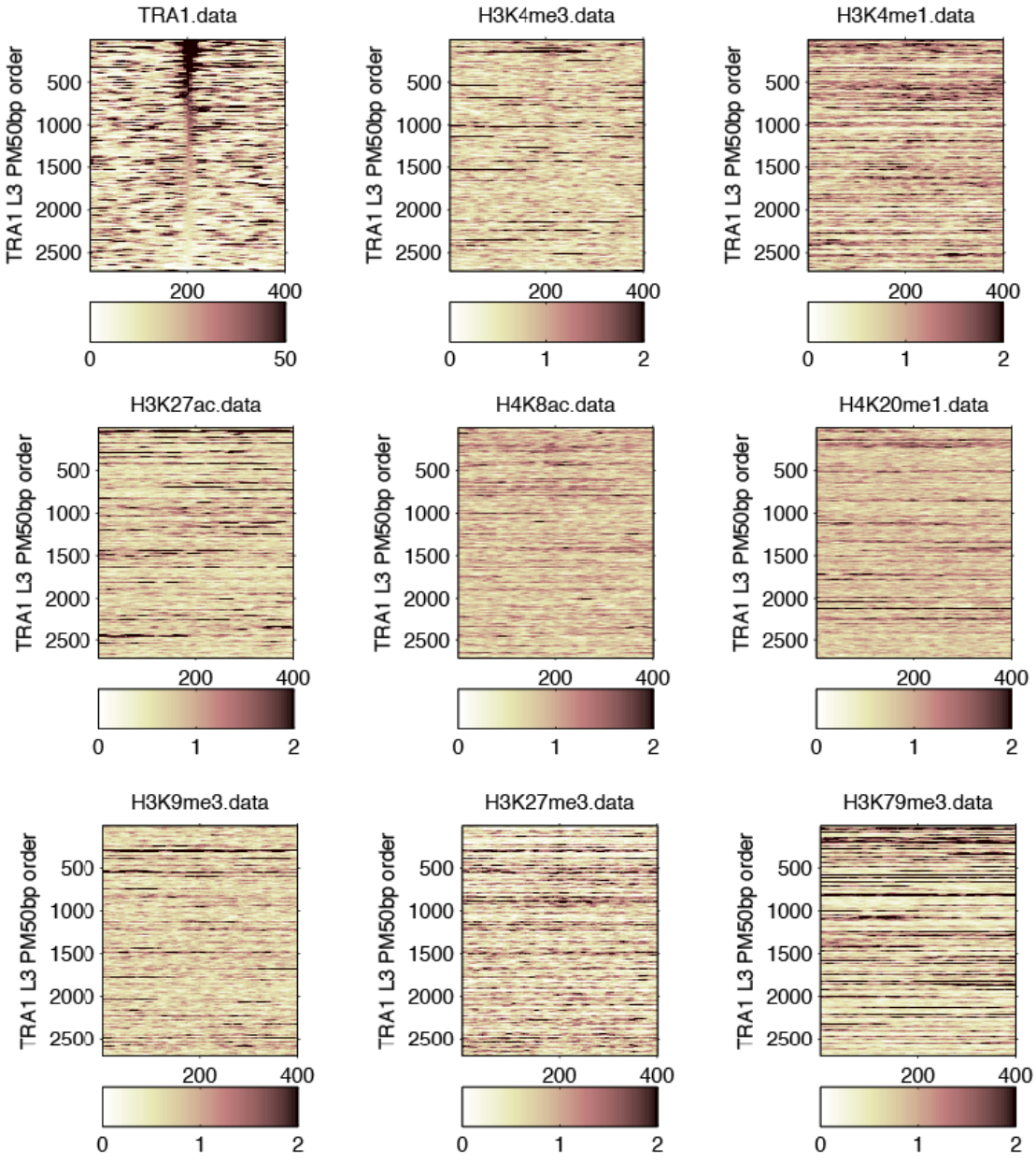
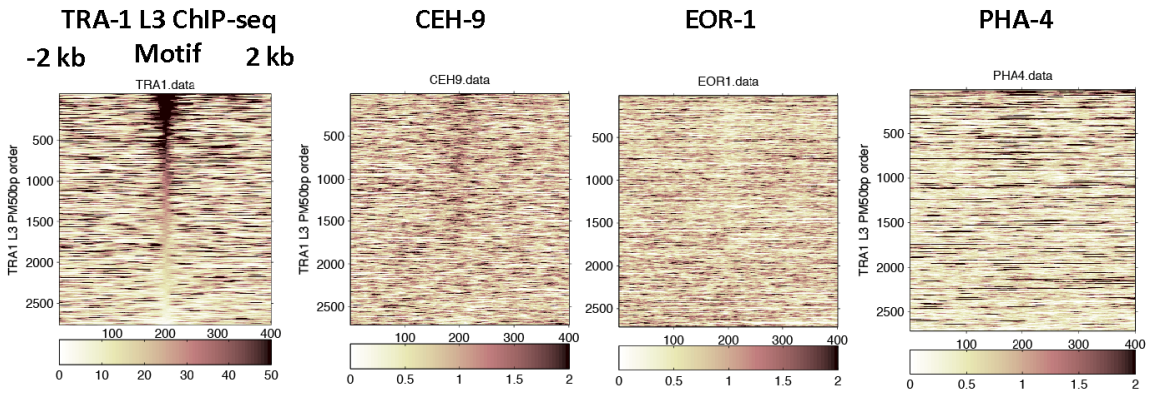

TRA1.data
H3K4me3.data
H3K4me1.data
H3K27ac.data
H4K8ac.data
H4K20me1.data
H3K9me3.data
H3K27me3.data
H3K79me3.data
TRA1 L3 PM50bp order
500
1000
1500
2000
2500
200
400
0
50
1
2

Figure 3.3. A comparison of TRA-1 occupancy and transcription factor occupancy. The 2713 closest matches to the TRA-1 consensus binding sequence in the *C. elegans* genome were ordered based on their degree of enrichment in L3 stage TRA-1 ChIP. Normalized ChIP reads within a 4 kb window centered on each TRA-1 motif are shown for TRA-1 and 3 other site specific transcription factors. Occupancies of CEH-9 and TRA-1 show a positive correlation. Occupancies of EOR-1 and TRA-1 show a negative correlation. PHA-4 and TRA-1 show no correlation.


TRA-1 L3 ChIP-seq
-2 kb
Motif
2 kb
TRA1.data
TRA1 L3 PM50bp order
CEH-9
CEH9.data
EOR-1
EOR1.data
PHA-4
PHA4.data

How is TRA-1 regulated?

Previous work has shown that the *tra-1* locus produces two transcripts (Zarkower & Hodgkin, 1992), and that the protein encoded by the larger transcript, TRA-1A, is C-terminally cleaved to generate the functional transcriptional regulator (Schvarzstein & Spence, 2006). However, it is unclear at exactly which residue this cleavage occurs, and what proteins mediate this cleavage event remain unknown. It is also unclear how TRA-1 protein expression may differ in different cells and at different times. DSX, the terminal regulator of sex in *Drosophila*, has recently been shown to exhibit dynamic expression through development, in many cases coming on in particular cells as they execute sex-specific differentiation events (Robinett, Vaughan, Knapp, & Baker, 2010). Thus spatial and temporal control of DSX expression appears to be an important component of *Drosophila* sexual development, and it is unclear whether a similar principle applies in *C. elegans*.

New approaches to understanding TRA-1 function

Cell type specific ChIP-seq

One limitation of the approach described in Chapter 2 is that all ChIPs were performed in large batches of whole animals and likely failed to detect sites that are bound by TRA-1 in only a few cells. One obvious potential solution to this problem is to simply perform TRA-1 ChIP-seq specifically in particular cells of interest. Fluorescence assisted cell sorting (FACS) can be used to generate very pure populations of cells expressing a fluorescent marker, and promoters capable of driving cell-type specific expression of such markers are known for many cell types in *C. elegans*. By coupling FACS and ChIP-

seq, cell-type specific TRA-1 binding sites could be detected. Of course this is a very technically challenging approach, but may be feasible for certain cell types.

A postdoctoral fellow in the lab, Mary Kroetz, has used FACS to generate large numbers (10^5-10^6) of highly pure gonadal precursor cells (GPCs) from *C. elegans* larvae, at a timepoint in which these cells represent less than 1% of total cells in the animal (Mary Kroetz, personal communication). The fewest cells used in a successful TRA-1 ChIP experiment thus far is ~8 x10^7, but other groups have performed successful site-specific transcription factor ChIP-seq experiments in a few as 10^4 cells (Zwart et al., 2013), suggesting that ChIP-seq of TRA-1 from 10^5-10^6 GPCs may be feasible. We are currently pursuing TRA-1 ChIP-seq using these sorted GPCs in an attempt to identify gonadal precursor-specific TRA-1 binding sites. As mentioned in Chapter 2, *tra-1* plays a unique role in the gonad, promoting gonadal proliferation in both sexes, but nothing is known about which TRA-1 target genes are responsible for this activity, making the GPCs a particularly interesting tissue in which to perform cell-type specific TRA-1 ChIP-seq.

Tagging TRA-1 at the endogenous locus

One of the difficulties in addressing many of the questions described in the first portion of this chapter is that the only means to follow the TRA-1 protein currently available is the TRA-1 antibody presented in Chapter 2, which does not perform well in many applications. A rescuing tagged version of the protein would provide another way to examine TRA-1 protein function, but *tra-1* is notoriously difficult to rescue with transgenes. The *tra-1* locus is extremely large for *C. elegans*, with over 40 kb of sequence between the nearest genes on either side. In the compact *C. elegans* genome, such large intergenic regions are quite rare, and often contain important regulatory elements for the

adjacent genes. The identification of TRA-1 binding sites in an intron of *tra-1* as well as several kb upstream of the coding region suggests that this may indeed be the case for *tra-1*. The *tra-1* locus is also not covered in current fosmid libraries, possibly owing to instability in *E. coli* (J. Hodgkin, 1993), which makes the generation of a full genomic clone of the locus challenging. However, recent advances in *C. elegans* transgene technology provide the exciting prospect of tagging the endogenous *tra-1* locus, which would avoid the problems of high copy number and missing regulatory elements associated with traditional transgenic methods, and potentially generate a strain bearing a tagged version of TRA-1 that is capable of rescuing a *tra-1* mutant.

CRISPR-Cas9-mediated genome editing technology is based on the bacterial Cas9 endonuclease, which generates DNA double-strand breaks (DSBs) at specific target sites based on complementary between DNA sequences at those sites and a guide RNA associated with the Cas9 enzyme. Guide RNAs can be engineered to target virtually any DNA sequence, allowing for the generation of site-specific DSBs that can then be repaired via homologous recombination from an exogenously supplied repair template. Several groups have demonstrated the utility of this system for engineering the *C. elegans* genome (Frokjaer-Jensen, 2013). I am currently attempting to use this system to generate a strain expressing an N-terminally tagged BIOtag::GFP::3XFLAG::TRA-1 protein under the control of all the regulatory elements of the endogenous *tra-1* locus, which would potentially be capable of rescuing a *tra-1* null mutation.

Such a strain would be of great utility in answering several of the questions posed above. GFP would allow for the visualization of TRA-1 protein dynamics in living animals, shedding light on questions of TRA-1 regulation as well as TRA-1 function. The BIOtag,

GFP, and 3XFLAG tags can all be used for protein purifications, either individually or in tandem. BIOtag is a sequence of amino acids that includes the substrate and recognition site required for biotinylation by the BirA protein. *C. elegans* does not normally express BirA, but stable transgenic strains have been constructed in which regulatory elements that direct ubiquitous expression drive BirA expression throughout the animal with no phenotypic effects (Ooi, Henikoff, & Henikoff, 2010). Biotinylated proteins can be extremely efficiently purified using streptavidin; the biotin-streptavidin interaction is among the strongest known non-covalent biomolecular interactions, with a K_d of roughly 10^{-14} mol/L (Green, 1975). This is roughly 10^6 fold higher than the 10^{-7}-10^{-9} mol/L K_d values typical of protein-antibody interactions, making it an excellent means by which to perform purifications. High-affinity TRA-1 pulldowns based on the biotin-streptavidin interaction could address several of the questions posed above. TRA-1 cleavage sites could be accurately mapped, and proteins that physically interact with TRA-1 could be identified in Co-IPs. This could potentially identify regulators of TRA-1 activity as well as transcriptional co-regulators that may shed light on how TRA-1 functions as a transcriptional repressor.

This strain could also provide a simpler alternative to the technically challenging cell-specific ChIP-seq experiments described earlier. Driving the *birA* gene using a cell-type specific promoter would generate biotinylated TRA-1 in only those cells in which the promoter is active, and subsequent ChIPs using streptavidin beads to pull down biotinylated TRA-1 would enrich for sites bound by TRA-1 specifically in those cells, potentially allowing for the identification of cell-type specific TRA-1 binding sites that cannot be identified in whole animals.

Conclusion

The list of TRA-1 binding sites and putative target genes presented here represents an important resource for further investigations into the sexual development of *C. elegans*. Determining what role these genes play in the worm will provide a paradigm for understanding how sexual differentiation is regulated on the level of an entire organism. In addition, this list can be used in comparisons between TRA-1 targets and the regulators of sexual differentiation in other species that will greatly enhance our understanding of this important biological process and the degree to which its regulation is conserved throughout animals.

Bibliography

Ahringer, J., & Kimble, J. (1991). Control of the sperm-oocyte switch in Caenorhabditis elegans hermaphrodites by the fem-3 3' untranslated region. *Nature, 349*(6307), 346-348. doi: 10.1038/349346a0

Ahringer, J., Rosenquist, T. A., Lawson, D. N., & Kimble, J. (1992). The Caenorhabditis elegans sex determining gene fem-3 is regulated post-transcriptionally. *EMBO J, 11*(6), 2303-2310.

Akerib, C. C., & Meyer, B. J. (1994). Identification of X chromosome regions in Caenorhabditis elegans that contain sex-determination signal elements. *Genetics, 138*(4), 1105-1125.

Arur, S., Ohmachi, M., Berkseth, M., Nayak, S., Hansen, D., Zarkower, D., & Schedl, T. (2011). MPK-1 ERK controls membrane organization in C. elegans oogenesis via a sex-determination module. *Dev Cell, 20*(5), 677-688. doi: 10.1016/j.devcel.2011.04.009

Barnes, T. M., & Hodgkin, J. (1996). The tra-3 sex determination gene of Caenorhabditis elegans encodes a member of the calpain regulatory protease family. *EMBO J, 15*(17), 4477-4484.

Beanan, M. J., & Strome, S. (1992). Characterization of a germ-line proliferation mutation in C. elegans. *Development, 116*(3), 755-766.

Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N., & Fellous, M. (1990). Genetic evidence equating SRY and the testis-determining factor. *Nature, 348*(6300), 448-450. doi: 10.1038/348448a0

Bull, J. J., & Vogt, R. C. (1979). Temperature-dependent sex determination in turtles. *Science, 206*(4423), 1186-1188.

Bulnheim, H. P. (1966). [Photoperiodic influence of the sex ratio in Gammarus duebeni (Crustacea, Amphipoda)]. *Naturwissenschaften, 53*(24), 709.

Carmi, I., Kopczynski, J. B., & Meyer, B. J. (1998). The nuclear hormone receptor SEX-1 is an X-chromosome signal that determines nematode sex. *Nature, 396*(6707), 168-173.

Chen, P., & Ellis, R. E. (2000). TRA-1A regulates transcription of fog-3, which controls germ cell fate in C. elegans. *Development, 127*(14), 3119-3129.

Chen, P. J., Cho, S., Jin, S. W., & Ellis, R. E. (2001). Specification of germ cell fates by FOG-3 has been conserved during nematode evolution. *Genetics, 158*(4), 1513-1525.

Chin-Sang, I. D., & Spence, A. M. (1996). Caenorhabditis elegans sex-determining protein FEM-2 is a protein phosphatase that promotes male development and interacts directly with FEM-3. *Genes Dev, 10*(18), 2314-2325.

Conradt, B., & Horvitz, H. R. (1999). The TRA-1A sex determination protein of C. elegans regulates sexually dimorphic cell deaths by repressing the egl-1 cell death activator gene. *Cell, 98*(3), 317-327.

Dawes, H. E., Berlin, D. S., Lapidus, D. M., Nusbaum, C., Davis, T. L., & Meyer, B. J. (1999). Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate. *Science, 284*(5421), 1800-1804.

de Bono, M., & Hodgkin, J. (1996). Evolution of sex determination in caenorhabditis: unusually high divergence of tra-1 and its functional consequences. *Genetics, 144*(2), 587-595.

Doniach, T. (1986). Activity of the sex-determining gene tra-2 is modulated to allow spermatogenesis in the C. elegans hermaphrodite. *Genetics, 114*(1), 53-76.

Erickson, J. W., & Quintero, J. J. (2007). Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila. *PLoS Biol, 5*(12), e332. doi: 10.1371/journal.pbio.0050332

Farboud, B., Nix, P., Jow, M. M., Gladden, J. M., & Meyer, B. J. (2013). Molecular antagonism between X-chromosome and autosome signals determines nematode sex. *Genes Dev, 27*(10), 1159-1178. doi: 10.1101/gad.217026.113

Fire, A., Harrison, S. W., & Dixon, D. (1990). A modular set of lacZ fusion vectors for studying gene expression in Caenorhabditis elegans. *Gene, 93*(2), 189-198.

Flanders, S. E. (1946). The mechanism of sex control in the honey bee. *J Econ Entomol, 39*, 379.

Frokjaer-Jensen, C. (2013). Exciting Prospects for Precise Engineering of Caenorhabditis elegans Genomes with CRISPR/Cas9. *Genetics, 195*(3), 635-642. doi: 10.1534/genetics.113.156521

Gaudet, J., VanderElst, I., & Spence, A. M. (1996). Post-transcriptional regulation of sex determination in Caenorhabditis elegans: widespread expression of the sex-determining gene fem-1 in both sexes. *Mol Biol Cell, 7*(7), 1107-1121.

Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., . . . Waterston, R. H. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science, 330*(6012), 1775-1787. doi: 10.1126/science.1196914

Gladden, J. M., & Meyer, B. J. (2007). A ONECUT homeodomain protein communicates X chromosome dose to specify Caenorhabditis elegans sexual fate by repressing a sex switch gene. *Genetics, 177*(3), 1621-1637. doi: 10.1534/genetics.106.061812

Goodwin, E. B., Okkema, P. G., Evans, T. C., & Kimble, J. (1993). Translational regulation of tra-2 by its 3' untranslated region controls sexual identity in C. elegans. *Cell, 75*(2), 329-339.

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics, 27*(7), 1017-1018. doi: 10.1093/bioinformatics/btr064

Green, N. M. (1975). Avidin. *Adv Protein Chem, 29*, 85-133.

Hamaoka, B. Y., Dann, C. E., 3rd, Geisbrecht, B. V., & Leahy, D. J. (2004). Crystal structure of Caenorhabditis elegans HER-1 and characterization of the interaction between HER-1 and TRA-2A. *Proc Natl Acad Sci U S A, 101*(32), 11673-11678. doi: 10.1073/pnas.0402559101

Hargitai, B., Kutnyanszky, V., Blauwkamp, T. A., Stetak, A., Csankovszki, G., Takacs-Vellai, K., & Vellai, T. (2009). xol-1, the master sex-switch gene in C. elegans, is a transcriptional target of the terminal sex-determining factor TRA-1. *Development, 136*(23), 3881-3887. doi: 10.1242/dev.034637

Hildreth, P. E. (1965). DOUBLESEX, RECESSIVE GENE THAT TRANSFORMS BOTH MALES AND FEMALES OF DROSOPHILA INTO INTERSEXES. *Genetics, 51*, 659-678.

Hodgkin, J. (1980). More sex-determination mutants of Caenorhabditis elegans. *Genetics, 96*(3), 649-664.
Hodgkin, J. (1983a). Male Phenotypes and Mating Efficiency in CAENORHABDITIS ELEGANS. *Genetics, 103*(1), 43-64.
Hodgkin, J. (1983b). Two types of sex determination in a nematode. *Nature, 304*(5923), 267-268.
Hodgkin, J. (1983). X chromosome dosage and gene expression in Caenorhabditis elegans: two unusual dumpy genes. *Molecular and General Genetics MGG, 192*(3), 452-458.
Hodgkin, J. (1984). Switch genes and sex determination in the nematode C. elegans. *J Embryol Exp Morphol, 83 Suppl*, 103-117.
Hodgkin, J. (1986). Sex determination in the nematode C. elegans: analysis of tra-3 suppressors and characterization of fem genes. *Genetics, 114*(1), 15-52.
Hodgkin, J. (1987). A genetic analysis of the sex-determining gene, tra-1, in the nematode Caenorhabditis elegans. *Genes Dev, 1*(7), 731-745.
Hodgkin, J. (1993). Molecular cloning and duplication of the nematode sex-determining gene tra-1. *Genetics, 133*(3), 543-560.
Hodgkin, J., Zellan, J. D., & Albertson, D. G. (1994). Identification of a candidate primary sex determination locus, fox-1, on the X chromosome of Caenorhabditis elegans. *Development, 120*(12), 3681-3689.
Hodgkin, J. A., & Brenner, S. (1977). Mutations causing transformation of sexual phenotype in the nematode Caenorhabditis elegans. *Genetics, 86*(2 Pt. 1), 275-287.
Huang, X., Zhang, H., & Zhang, H. (2011). The zinc-finger protein SEA-2 regulates larval developmental timing and adult lifespan in C. elegans. *Development, 138*(10), 2059-2068. doi: 10.1242/dev.057109
Hunter, C. P., & Wood, W. B. (1990). The tra-1 gene determines sexual phenotype cell-autonomously in C. elegans. *Cell, 63*(6), 1193-1204.
Hunter, C. P., & Wood, W. B. (1992). Evidence from mosaic analysis of the masculinizing gene her-1 for cell interactions in C. elegans sex determination. *Nature, 355*(6360), 551-555. doi: 10.1038/355551a0
Jan, E., Motzny, C. K., Graves, L. E., & Goodwin, E. B. (1999). The STAR protein, GLD-1, is a translational regulator of sexual identity in Caenorhabditis elegans. *EMBO J, 18*(1), 258-269. doi: 10.1093/emboj/18.1.258
Jin, S. W., Kimble, J., & Ellis, R. E. (2001). Regulation of cell fate in Caenorhabditis elegans by a novel cytoplasmic polyadenylation element binding protein. *Dev Biol, 229*(2), 537-553. doi: 10.1006/dbio.2000.9993
Kimble, J. (1981). Alterations in cell lineage following laser ablation of cells in the somatic gonad of Caenorhabditis elegans. *Dev Biol, 87*(2), 286-300.
Kimble, J., & Hirsh, D. (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. *Dev Biol, 70*(2), 396-417.
Kuwabara, P. E. (1996). A novel regulatory mutation in the C. elegans sex determination gene tra-2 defines a candidate ligand/receptor interaction site. *Development, 122*(7), 2089-2098.

Kuwabara, P. E., Okkema, P. G., & Kimble, J. (1992). tra-2 encodes a membrane protein and may mediate cell communication in the Caenorhabditis elegans sex determination pathway. *Mol Biol Cell, 3*(4), 461-473.

L'Hernault, S. W., Shakes, D. C., & Ward, S. (1988). Developmental genetics of chromosome I spermatogenesis-defective mutants in the nematode Caenorhabditis elegans. *Genetics, 120*(2), 435-452.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol, 10*(3), R25. doi: 10.1186/gb-2009-10-3-r25

Lieb, J. D., Liu, X., Botstein, D., & Brown, P. O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet, 28*(4), 327-334. doi: 10.1038/ng569

Lints, R., & Emmons, S. W. (2002). Regulation of sex-specific differentiation and mating behavior in C. elegans by a new member of the DM domain transcription factor family. *Genes Dev, 16*(18), 2390-2402. doi: 10.1101/gad.1012602

Luo, S. D., Shi, G. W., & Baker, B. S. (2011). Direct targets of the D. melanogaster DSXF protein and the evolution of sexual development. *Development, 138*(13), 2761-2771. doi: 10.1242/dev.065227

Luz, J. G., Hassig, C. A., Pickle, C., Godzik, A., Meyer, B. J., & Wilson, I. A. (2003). XOL-1, primary determinant of sexual fate in C. elegans, is a GHMP kinase family member and a structural prototype for a class of developmental regulators. *Genes Dev, 17*(8), 977-990. doi: 10.1101/gad.1082303

Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics, 27*(12), 1696-1697. doi: 10.1093/bioinformatics/btr189

Madl, J. E., & Herman, R. K. (1979). Polyploids and sex determination in Caenorhabditis elegans. *Genetics, 93*(2), 393-402.

Mapes, J., Chen, J. T., Yu, J. S., & Xue, D. (2010). Somatic sex determination in Caenorhabditis elegans is modulated by SUP-26 repression of tra-2 translation. *Proc Natl Acad Sci U S A, 107*(42), 18022-18027. doi: 10.1073/pnas.1004513107

Mason, D. A., Rabinowitz, J. S., & Portman, D. S. (2008). dmd-3, a doublesex-related gene regulated by tra-1, governs sex-specific morphogenesis in C. elegans. *Development, 135*(14), 2373-2382. doi: 10.1242/dev.017046

Matson, C. K., & Zarkower, D. (2012). Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. *Nat Rev Genet, 13*(3), 163-174. doi: 10.1038/nrg3161

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., . . . Sakaizumi, M. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature, 417*(6888), 559-563. doi: 10.1038/nature751

Mehra, A., Gaudet, J., Heck, L., Kuwabara, P. E., & Spence, A. M. (1999). Negative regulation of male development in Caenorhabditis elegans by a protein-protein interaction between TRA-2A and FEM-3. *Genes Dev, 13*(11), 1453-1463.

Meller, V. H., & Kuroda, M. I. (2002). Sex and the single chromosome. *Adv Genet, 46*, 1-24.

Meyer, B. J. (2005). X-Chromosome dosage compensation. *WormBook*, 1-14. doi: 10.1895/wormbook.1.8.1

Miller, L. M., Plenefisch, J. D., Casson, L. P., & Meyer, B. J. (1988). xol-1: A gene that controls the male modes of both sex determination and X chromosome dosage compensation in C. elegans. *Cell, 55*(1), 167-183.

Moss, E. G. (2007). Heterochronic genes and the nature of developmental time. *Curr Biol, 17*(11), R425-434. doi: 10.1016/j.cub.2007.03.043

Murphy, M. W., Sarver, A. L., Rice, D., Hatzi, K., Ye, K., Melnick, A., . . . Bardwell, V. J. (2010). Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian Doublesex homolog DMRT1 in the juvenile testis. *Proc Natl Acad Sci U S A, 107*(30), 13360-13365. doi: 10.1073/pnas.1006243107

Nelson, M. D., Zhou, E., Kiontke, K., Fradin, H., Maldonado, G., Martin, D., . . . Fitch, D. H. (2011). A bow-tie genetic architecture for morphogenesis suggested by a genome-wide RNAi screen in Caenorhabditis elegans. *PLoS Genet, 7*(3), e1002010. doi: 10.1371/journal.pgen.1002010

Nicoll, M., Akerib, C. C., & Meyer, B. J. (1997). X-chromosome-counting mechanisms that determine nematode sex. *Nature, 388*(6638), 200-204.

Nigon, V. (1951). Polyploidie experimentale chez un Nematode libre, Rhabditis elegans. *Maupas. Bull. Biol. Fr. Belg.*(85), 187–225.

Niu, W., Lu, Z. J., Zhong, M., Sarov, M., Murray, J. I., Brdlik, C. M., . . . Reinke, V. (2011). Diverse transcription factor binding features revealed by genome-wide ChIP-seq in C. elegans. *Genome Res, 21*(2), 245-254. doi: 10.1101/gr.114587.110

Okkema, P. G., & Kimble, J. (1991). Molecular analysis of tra-2, a sex determining gene in C.elegans. *EMBO J, 10*(1), 171-176.

Ooi, S. L., Henikoff, J. G., & Henikoff, S. (2010). A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans. *Nucleic Acids Res, 38*(4), e26. doi: 10.1093/nar/gkp1090

Peden, E., Kimberly, E., Gengyo-Ando, K., Mitani, S., & Xue, D. (2007). Control of sex-specific apoptosis in C. elegans by the BarH homeodomain protein CEH-30 and the transcriptional repressor UNC-37/Groucho. *Genes Dev, 21*(23), 3195-3207. doi: 10.1101/gad.1607807

Perry, M. D., Li, W., Trent, C., Robertson, B., Fire, A., Hageman, J. M., & Wood, W. B. (1993). Molecular characterization of the her-1 gene suggests a direct role in cell signaling during Caenorhabditis elegans sex determination. *Genes Dev, 7*(2), 216-228.

Pilgrim, D., McGregor, A., Jackle, P., Johnson, T., & Hansen, D. (1995). The C. elegans sex-determining gene fem-2 encodes a putative protein phosphatase. *Mol Biol Cell, 6*(9), 1159-1171.

Pires-daSilva, A., & Sommer, R. J. (2004). Conservation of the global sex determination gene tra-1 in distantly related nematodes. *Genes Dev, 18*(10), 1198-1208. doi: 10.1101/gad.293504

Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A., & Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. *Annu Rev Genet, 36*, 233-278. doi: 10.1146/annurev.genet.36.042902.092433

Plenefisch, J. D., DeLong, L., & Meyer, B. J. (1989). Genes that implement the hermaphrodite mode of dosage compensation in Caenorhabditis elegans. *Genetics, 121*(1), 57-76.

Powell, J. R., Jow, M. M., & Meyer, B. J. (2005). The T-box transcription factor SEA-1 is an autosomal element of the X:A signal that determines C. elegans sex. *Dev Cell, 9*(3), 339-349. doi: 10.1016/j.devcel.2005.06.009

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. doi: 10.1093/bioinformatics/btq033

Raymond, C. S., Shamu, C. E., Shen, M. M., Seifert, K. J., Hirsch, B., Hodgkin, J., & Zarkower, D. (1998). Evidence for evolutionary conservation of sex-determining genes. *Nature, 391*(6668), 691-695. doi: 10.1038/35618

Rhind, N. R., Miller, L. M., Kopczynski, J. B., & Meyer, B. J. (1995). xo1-1 acts as an early switch in the C. elegans male/hermaphrodite decision. *Cell, 80*(1), 71-82.

Robertson, D. R. (1972). Social control of sex reversal in a coral-reef fish. *Science, 177*(4053), 1007-1009. doi: 10.1126/science.177.4053.1007

Robinett, C. C., Vaughan, A. G., Knapp, J. M., & Baker, B. S. (2010). Sex and the single cell. II. There is a time and place for sex. *PLoS Biol, 8*(5), e1000365. doi: 10.1371/journal.pbio.1000365

Rosenquist, T. A., & Kimble, J. (1988). Molecular cloning and transcript analysis of fem-3, a sex-determination gene in Caenorhabditis elegans. *Genes Dev, 2*(5), 606-616.

Salz, H. K. (2011). Sex determination in insects: a binary decision based on alternative splicing. *Curr Opin Genet Dev, 21*(4), 395-400. doi: 10.1016/j.gde.2011.03.001

Schedl, T., Graham, P. L., Barton, M. K., & Kimble, J. (1989). Analysis of the role of tra-1 in germline sex determination in the nematode Caenorhabditis elegans. *Genetics, 123*(4), 755-769.

Schvarzstein, M., & Spence, A. M. (2006). The C. elegans sex-determining GLI protein TRA-1A is regulated by sex-specific proteolysis. *Dev Cell, 11*(5), 733-740. doi: 10.1016/j.devcel.2006.09.017

Sekido, R., & Lovell-Badge, R. (2013). Genetic control of testis development. *Sex Dev, 7*(1-3), 21-32. doi: 10.1159/000342221

Seydoux, G., & Dunn, M. A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. *Development, 124*(11), 2191-2201.

Shen, M. M., & Hodgkin, J. (1988). mab-3, a gene required for sex-specific yolk protein expression and a male-specific lineage in C. elegans. *Cell, 54*(7), 1019-1031.

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., . . . Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature, 346*(6281), 240-244. doi: 10.1038/346240a0

Skipper, M., Milne, C. A., & Hodgkin, J. (1999). Genetic and molecular analysis of fox-1, a numerator element involved in Caenorhabditis elegans primary sex determination. *Genetics, 151*(2), 617-631.

Smith, C. A., Roeszler, K. N., Ohnesorg, T., Cummins, D. M., Farlie, P. G., Doran, T. J., & Sinclair, A. H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature, 461*(7261), 267-271. doi: 10.1038/nature08298

Sokol, S. B., & Kuwabara, P. E. (2000). Proteolysis in Caenorhabditis elegans sex determination: cleavage of TRA-2A by TRA-3. *Genes Dev, 14*(8), 901-906.
Spence, A. M., Coulson, A., & Hodgkin, J. (1990). The product of fem-1, a nematode sex-determining gene, contains a motif found in cell cycle control proteins and receptors for cell-cell interactions. *Cell, 60*(6), 981-990.
Spencer, W. P. (1927). A Gynandromorph in Drosophila funebris. *AMER NATURALIST, 01*(61).
Starostina, N. G., Lim, J. M., Schvarzstein, M., Wells, L., Spence, A. M., & Kipreos, E. T. (2007). A CUL-2 ubiquitin ligase containing three FEM proteins degrades TRA-1 to regulate C. elegans sex determination. *Dev Cell, 13*(1), 127-139. doi: 10.1016/j.devcel.2007.05.008
Sulston, J. E., Albertson, D. G., & Thomson, J. N. (1980). The Caenorhabditis elegans male: Postembryonic development of nongonadal structures. *Dev Biol, 78*(2), 542-576. doi: http://dx.doi.org/10.1016/0012-1606(80)90352-8
Sulston, J. E., & Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev Biol, 56*(1), 110-156.
Sulston, J. E., Schierenberg, E., White, J. G., & Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol, 100*(1), 64-119.
Szabo, E., Hargitai, B., Regos, A., Tihanyi, B., Barna, J., Borsos, E., . . . Vellai, T. (2009). TRA-1/GLI controls the expression of the Hox gene lin-39 during C. elegans vulval development. *Dev Biol, 330*(2), 339-348. doi: 10.1016/j.ydbio.2009.04.005
Trent, C., Purnell, B., Gavinski, S., Hageman, J., Chamblin, C., & Wood, W. B. (1991). Sex-specific transcriptional regulation of the C. elegans sex-determining gene her-1. *Mech Dev, 34*(1), 43-55.
Tursun, B., Cochella, L., Carrera, I., & Hobert, O. (2009). A toolkit and robust pipeline for the generation of fosmid-based reporter genes in C. elegans. *PLoS One, 4*(3), e4625. doi: 10.1371/journal.pone.0004625
Wilkins, A. S. (1995). Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays, 17*(1), 71-77. doi: 10.1002/bies.950170113
Yi, W., Ross, J. M., & Zarkower, D. (2000). Mab-3 is a direct tra-1 target gene regulating diverse aspects of C. elegans male sexual development and behavior. *Development, 127*(20), 4469-4480.
Yonker, S. A., & Meyer, B. J. (2003). Recruitment of C. elegans dosage compensation proteins for gene-specific versus chromosome-wide repression. *Development, 130*(26), 6519-6532. doi: 10.1242/dev.00886
Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., . . . Ito, M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in Xenopus laevis. *Proc Natl Acad Sci U S A, 105*(7), 2469-2474. doi: 10.1073/pnas.0712244105
Zarkower, D. (2001). Establishing sexual dimorphism: conservation amidst diversity? *Nat Rev Genet, 2*(3), 175-185. doi: 10.1038/35056032
Zarkower, D., & Hodgkin, J. (1992). Molecular analysis of the C. elegans sex-determining gene tra-1: a gene encoding two zinc finger proteins. *Cell, 70*(2), 237-249.

Zarkower, D., & Hodgkin, J. (1993). Zinc fingers in sex determination: only one of the two C. elegans Tra-1 proteins binds DNA in vitro. *Nucleic Acids Res, 21*(16), 3691-3698.

Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., & Wickens, M. P. (1997). A conserved RNA-binding protein that regulates sexual fates in the C. elegans hermaphrodite germ line. *Nature, 390*(6659), 477-484. doi: 10.1038/37297

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol, 9*(9), R137. doi: 10.1186/gb-2008-9-9-r137

Zwart, W., Koornstra, R., Wesseling, J., Rutgers, E., Linn, S., & Carroll, J. S. (2013). A carrier-assisted ChIP-seq method for estrogen receptor-chromatin interactions from breast cancer core needle biopsy samples. *BMC Genomics, 14*(1), 232. doi: 10.1186/1471-2164-14-232