CHARACTERIZATION OF NEMATODE DAF-19 AND HUMAN RFX6 TRANSCRIPTION FACTORS

by

Lucie Semenec
B.Sc., University of British Columbia, 2007

THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
Department of Molecular Biology & Biochemistry

SIMON FRASER UNIVERSITY

Summer 2010

Approval

Name: **Lucie Semenec**

Degree: **Master of Science**

Title of Thesis: **Characterization of Nematode DAF-19 and Human RFX6 Transcription Factors**

Examining Committee:

Chair: **Dr. M. Leroux**
Professor, Department of Molecular Biology & Biochemistry

_____

Dr. Jack Chen
Senior Supervisor
Associate Professor, Department of Molecular Biology & Biochemistry

_____

Dr. Christopher Beh
Supervisor
Associate Professor, Department of Molecular Biology & Biochemistry

_____

Dr. Marco Marra
Supervisor
Professor, Director, Genome Sciences Centre, BC Cancer Agency

_____

Dr. Glen Tibbits
Supervisor
Professor, Biomedical Physiology & Kinesiology

_____

Dr. David Baillie
Internal Examiner
Professor, Department of Molecular Biology & Biochemistry

Date Defended/Approved: July 16, 2010

Declaration of
Partial Copyright Licence

Abstract

Regulatory factor X (RFX) transcription factors play an important role in regulating expression of ciliary genes associated with ciliopathies. However, the annotation of RFX genes may be incomplete and their function is not well understood. Here I describe two novel tissue-specific RFX genes in humans: RFX6 and RFX7. To study RFX genes in the model organism *Caenorhabditis elegans*, I undertook examining the expression of all four known isoforms of *daf-19*, the only RFX gene in *C. elegans*, by using Mos1 mediated Single Copy transgene Insertion (MosSCI) method. I discovered that both *daf-19c* and *daf-19d* isoforms are expressed in ciliated neurons and that their promoters are modular. In particular, *daf-19c* is expressed in all ciliated neurons while *daf-19d* in all but labial neurons. My analysis helped select suitable promoters for driving expression of RFX6 in ciliated neurons of *C. elegans* for testing its function in cilia.

Keywords: Regulatory Factor-X; Cilia; Alternative promoter; Transcription; MosSCI; DAF-19; Isoforms

Dedication

To my parents for their limitless support,

To my siblings for always believing in me,

And to my best friend, for inspiring the best in me.

Acknowledgements

Many people have impacted the success of my thesis but first and foremost I would like to acknowledge my supervisor Dr. Jack Chen. I would like to thank him for his guidance and support and most importantly for instilling in me a very good work ethic. Dr. Maja Tarailo-Graovac from my lab has often felt like a secondary supervisor throughout my studies here and I greatly appreciate all the advice and technical help that I received from her. All the transgenic worms generated would not have been possible without the generous assistance of Domena Tu of the Baillie lab who performed all injections. Victoria Ng worked on her ISS under my supervision and was a great help in generating and doing analysis of the JNC65 strain. Jeffery Chu and Jun Wang have both been very helpful with their advice with troubleshooting protocols. Dr. David Baillie and Dr. Robert Johnson were very insightful and informative regarding aspects of my project and I would like to thank them for their support as well. I would like to thank Tao Luo for all of ther support with all general lab functions and for freezing down my strains.

All other members of the Chen lab, Ismael Vergara, Christian Frech, Bora Uyar, Tammy Wong, Matthew Nesbitt and others have been very supportive during lab meetings and have kept me happy and entertained. Also all members of the Baillie lab have been very helpful to me in particular Allan Mah and Carrie Sims have taught me a lot.

Table of Contents

List of Figures

List of Tables

Abbreviations

AD	Activation domain
B	B extended dimerization domain
BBS	Bardet Beidl syndrome
BLS	Bare lymphocyte syndrome
C	C extended dimerization domain
CSN	Ciliated sensory neuron
D	Dimerization domain
Daf-c	Dauer formation constitutive
DBD	DNA binding domain
DIC	Differential interference contrast
DIM	B, C and D dimerization domains
DiO	Dioctadecyloxacarbocyanine perchlorate
Dyf	Dye-filling defective
EST	Expressed sequence tag
GFP	Green fluorescent protein
HGNC	HUGO Genome Nomenclature Committee
HiMAP	Human interactome map
HMM	Hidden Markov Model

Abbreviations

IFT	Intraflagellar transport
MHC	Major histocompatibility complex
MKS	Meckel-Gruber syndrome
MosSCI	Mos1 mediated single copy insertion
NCBI	National Center for Biotechnology Information
ORF	Open reading frame
PCR	Polymerase chain reaction
PKD	Polycystic kidney disease
RNAP	RNA polymerase
RFX	Regulatory Factor X
SAGE	Serial analysis of gene expression
TF	Transcription factor
TSS	Translational start site
UTR	Untranslated region

Chapter 1: General Introduction

1.1 Ciliopathies and ciliary genes

Nearly all cells in the human body contain tiny hair-like protrusions called cilia (Rosenbaum and Witman, 2002). This includes lung cells, where cilia contribute to chemosensory functions (Shah *et al.*, 2009), cilia based rod cells in the eye for photoreception (Roepman and Wolfrum, 2007) and in sperm cells where they provide motility (Walt and Hedinger, 1983), amongst others. Cilia are cellular microtubule based structures that emanate from the cell body and can have either motile or sensory functions, thus defining the two major types of cilia: motile and sensory (also known as non-motile or primary). Motile cilia have a 9+2 microtubule structure, where 9 doublet microtubules encircle the ciliary axoneme and two singlet microtubules occupy the center of the axoneme (Figure 1). Sensory cilia on the other hand, have a 9+0 microtubule structure with no central microtubules (Pazour and Rosenbaum, 2002). The conserved mechanism by which all eukaryotic cilia are made is called intraflagellar transport (IFT) and was first discovered in the single celled green algae, *Chlamydomonas reinhardtii* (Kozminski *et al.*, 1993; Chu *et al.*, 2010). IFT consists of several proteins that are responsible for shuttling necessary components of cilia from its centriole base, called the basal body, to the tip of the cilia and back. Cilia are widespread throughout evolution and are present in most eukaryotes, from unicellular *Chlamydomonas* to multicellular eukaryotes such as mammals, however they are absent in plants and nearly all fungi

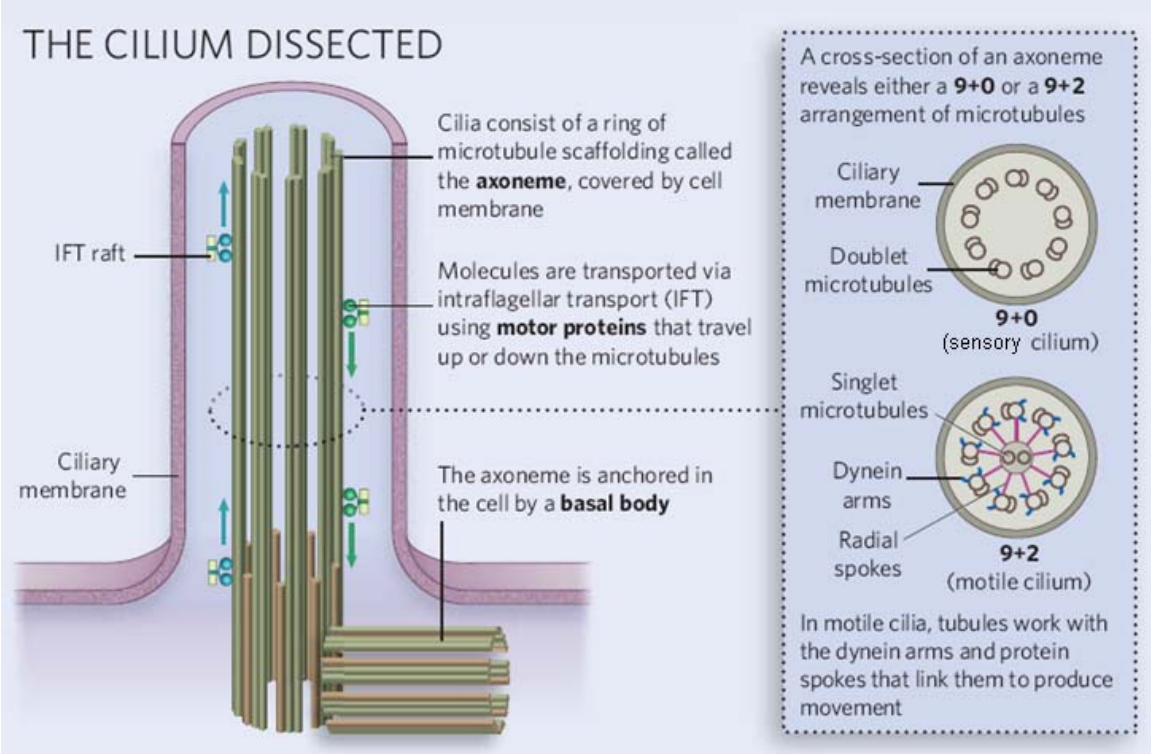


Figure 1 Diagramatic representation of cilia structure.

The left panel depicts the cilium architecture with the IFT machinery and the centriole based basal body at the base of the cilia. The right panel shows a cross-section of motile and non-motile (primary) cilia. This figure was adapted from (Ainsworth, 2007).

(Baker and Beales, 2009). There are only two known fungi in which the IFT genes are conserved, *Allomyces macrogynus* and *Batrachochytrium dendrobatidis* (Chu *et al.*, 2010).

Not surprisingly, defects in these conserved structures leads to devastating hereditary disease conditions, termed ciliopathies, including Bardet-Biedl syndrome (BBS), polycystic kidney disease (PKD) and Meckel-Gruber syndrome (MKS). Patients with PKD suffer from an overgrowth of cysts on their kidneys, ultimately leading to kidney failure; other organs such as the liver and brain can be affected as well (Diseases, November 2007). Several studies have shown that PKD is most likely caused by defective cilia formation due to mutations in IFT genes (Pazour, 2004). In the case of BBS, the symptoms have multiple forms and can include retinal dystrophy, obesity, polydactyly and cognitive impairment amongst others. These various phenotypes demonstrate this disorders pleiotropic nature (Katsanis *et al.*, 2001), which is likely due to the ubiquitous nature of cilia (Ansley *et al.*, 2003). In addition to being phenotypically heterogeneous, it was found that BBS is also caused by multiple different genes, demonstrating that it has genetic heterogeneity as well, where mutation in more than one allele can cause this same disease (Kwitek-Black *et al.*, 1993; Bruford *et al.*, 1997). Various studies haven shown that the BBS genes share a common feature in that they are all involved in cilia development and function (Mykytyn and Sheffield, 2004). Ciliopathies are a major burden on human health. For instance, the autosomal recessive form of PKD affects 1 in 10, 000 newborns of which 50% die within weeks of being born (Blyth and Ockenden, 1971; Cole *et al.*, 1987). Thus it is important to understand the

mechanisms by which ciliary genes are regulated, so that efficient drug strategies can be developed.

1.2 Regulation of Ciliary Genes by Regulatory Factor X

In 2000, Swoboda and colleagues gave new insight into the regulation of ciliary genes by discovering a connection between ciliary genes and the Regulatory Factor X (RFX) family of transcription factors (TFs) (Swoboda *et al.*, 2000). TFs in the RFX family are named as such because their conserved DNA binding domain (DBD) (Emery *et al.*, 1996a) binds to a conserved *cis*-regulatory element called the X-box motif, which consists of a 12-15 nucleotide imperfect palindromic sequence (Figure 2) (Reith *et al.*, 1988). Swoboda and colleagues found that DAF-19, the sole RFX TF in *C. elegans* (Emery *et al.*, 1996b), was expressed in ciliated sensory neurons, suggesting its potential role in ciliary gene regulation (Swoboda *et al.*, 2000). This role is supported by the observation that *daf-19* mutant worms, which have a truncated DAF-19 that lacks a DBD, are completely devoid of ciliated endings in their sensory neurons (Perkins *et al.*, 1986; Swoboda *et al.*, 2000). Swoboda and colleagues further supported this putative regulatory role of DAF-19 by showing that several ciliary genes containing a typical mammalian X-box motif, drove expression of reporter GFP in ciliated neurons but failed to do so in a *daf-19* mutant background (Swoboda *et al.*, 2000). This study established the first critical link between RFX TFs and ciliary gene regulation that has since set the foundation for future studies in cilia research.

Such studies included bioinformatics based genome wide searches of X-box motifs in *C. elegans*, to help identify novel ciliary genes (Efimenko *et al.*, 2005; Chen *et al.*, 2006). Also, RFX TFs in other species, such as the fruit fly (Dubruille *et al.*, 2002),

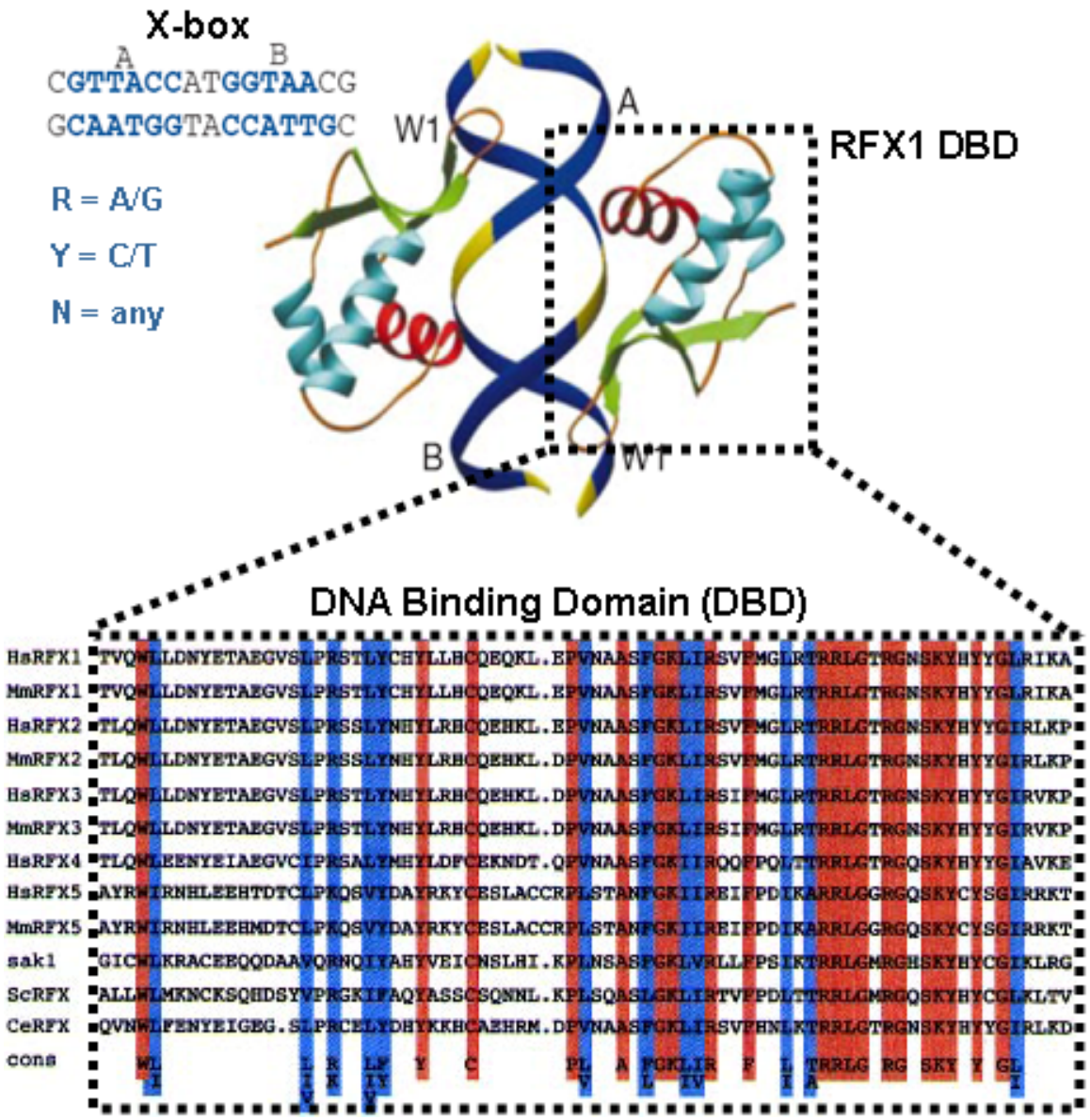


Figure 2 Interaction of the RFX DBD with the X-box motif.

Upper panel of the figure shows ribbon structure of RFX1 DBDs and is adapted from (Gajiwala *et al.*, 2000). Each ribbon structure depicts an RFX1 DBD binding to each half-site of the X-box palindromic sequence. The lower panel of the figure shows the conservation of the RFX DBD with red highlighted residues being most conserved, blue highlighted residues being less conserved and non-highlighted residues showing the least conservation. This lower panel figure is adapted from (Emery *et al.*, 1996a).

mice (Bonnafe *et al.*, 2004; Baas *et al.*, 2006) and more recently, humans (El Zein *et al.*, 2009; Purvis *et al.*, 2010), were subsequently found to have a role in cilia development. In the fruit fly, *Drosophila melanogaster,* RFX was shown to have a similar role to DAF-19 in that it is important for regulating ciliated sensory neuron differentiation (Dubruille *et al.*, 2002). In mice, RFX3 has been shown to regulate ciliated ependymal cell differentiation (Bonnafe *et al.*, 2004), nodal monocilia (Baas *et al.*, 2006), primary ciliary differentiation in pancreas (Ait-Lounis *et al.*, 2007) and also motile cilia differentiation (El Zein *et al.*, 2009). Additionally, mouse RFX4 has been shown to regulate ciliary genes in the central nervous system and is important for proper cilia development (Ashique *et al.*, 2009). Lastly, human RFX1 and RFX2 were shown to play a role in the regulation of the *ALMS1* gene which encodes a centrosomal protein that is used in the assembly of cilia (Purvis *et al.*, 2010). Although there is clearly a strong correlation between RFX TFs and the regulation of ciliary genes, there is an exception to this rule in the case of RFX5, which regulates the transcription of major histocompatibility complex II (MHCII) genes (Mach *et al.*, 1996). Thus, although the RFX TF family is largely associated with regulating cilia development and function it may also have more diverse functions.

As of 2008, there were five known RFX TFs, RFX1-5, and two genes that contain RFX DNA binding domain (RFXDC1 and RFXDC2), and only one RFX TF in *C. elegans*, DAF-19 (Flicek *et al.*, 2008). Although five human RFX genes were reported, could the putative RFX genes be bona fide RFX in the human genome? Now with the fully sequenced human genome, it is possible to carry out an extensive search for

additional RFX genes. Details about the identification of novel RFX TFs in human is described in Chapter 3.

1.3 Alternative RFX Isoforms

In addition to the diversity of individual RFX members, there is also diversity associated with each RFX gene due to alternative isoforms. Alternative protein isoforms can occur via alternative splicing (Matlin *et al.*, 2005), alternative translation initiation sites (Touriol *et al.*, 2003) or by usage of alternative promoters (Ayoubi and Van De Ven, 1996; Touriol *et al.*, 2003; Matlin *et al.*, 2005). Alternative promoter usage is found in ~50% of human genes (Baek *et al.*, 2007) and almost all RFX genes are alternatively spliced (Aftab *et al.*, 2008). Different isoforms of a gene can have drastically different function and expression. An example of a TF with multiple isoforms that are generated by alternative promoters and have distinct functions is that of the zebrafish *SCL* gene (Qian *et al.*, 2007). There are two isoforms of this gene, *scl-α* and *scl-β,* that are expressed at different levels and although they share function in initiation of hematopoiesis, *scl-β* has a distinct role in erythrocyte maturation and the development of hematopoietic stem cells (Qian *et al.*, 2007). Human RFX genes also exist in multiple isoforms, however it is not yet known whether these have different functions (Flicek *et al.*, 2008).

The phenomenon of multiple promoters to generate alternate isoforms is clearly seen in *daf-19*, the RFX TF gene in *C. elegans*. Initially, *daf-19* was found to be expressed only in ciliated sensory neurons and associated with its role in cilia development (Swoboda, Adler et al. 2000). However, a later study showed that *daf-19*

can generate several different isoforms. Senti and Swoboda reported evidence for three separate isoforms of *daf-19*, called *daf-19a*, *daf-19b* and *daf-19d* (Senti and Swoboda, 2008). In this study it was shown that the *daf-19d* isoform is specifically expressed in ciliated sensory neurons and regulates ciliary genes, while the *daf-19a/b* isoform is actually expressed in non-ciliated neurons and plays a role in sensory neuron synapse formation (Senti and Swoboda, 2008). In concordance with these observations, the target genes of DAF-19d are expressed in ciliated neurons and have an X-box in their promoters, while the target genes of DAF-19a/b do not seem to have an X-box, at least not one that is typical for DAF-19d regulated genes (Senti and Swoboda, 2008). Senti and colleagues proposed that this could either mean that DAF-19a/b targets are not regulated by an X-box or that they have one, but it differs from the DAF-19d recognized X-box. Also, a fourth isoform is called *daf-19c* and is described as male-specific in Wormbase (Harris *et al.*, 2010).

1.4 *C. elegans* as a Model to Study Cilia and RFX TFs

C. elegans was first established as a model organism in 1974 by Sydney Brenner (Brenner, 1974) and has since been a very useful tool for studying biology. This nematode has several features that make it ideal to work with in the laboratory. It has a fully transparent body, allowing for easy identification of cells. This feature has greatly facilitated the mapping of the entire cell lineage (Sulston and Horvitz, 1977; Kimble and Hirsh, 1979). It also was the first multicellular eukaryote to have its entire genome sequenced (Consortium, 1998) and has a very well annotated genome with a useful data repository available called WormBase (Harris *et al.*, 2010). Additionally, the genome of a

closely related species, *C. briggsae*, has been sequenced and annotated (Stein *et al.*, 2003).

Despite its small body, this nematode contains 959 somatic cells in the hermaphrodite, of which 302 are neurons and 60 of these are ciliated sensory neurons (CSN) (Ward *et al.*, 1975). It is very useful for studying cilia because a mutant in which all cilia are obliterated is available. This is the *daf-19(m86)* mutant, which contains a point mutation that causes a non-sense mutation upstream of the DBD coding sequence (Swoboda *et al.*, 2000). The phenotypes associated with this mutant include the Daf-c phenotype and the Dyf phenotype. In the Daf-c phenotype, the majority of worms enter a survival state called dauer and in the Dyf phenotype, the worms fail to fill the ciliated endings of their neurons with a fluorescent dye due to the lack of cilia (Swoboda *et al.*, 2000). Another useful feature is that due to a recent development by the Swoboda laboratory (Senti *et al.*, 2009), it is possible to study one single ciliated neuron at a time.

C. elegans is also a very suitable model for studying isoforms of RFX TFs because it consists of a simplified RFX network where only one RFX gene exists but in multiple isoforms. This is a simpler organization than in mammals, where multiple RFX genes are present and there are multiple isoforms of each gene. Thus *C. elegans* is very appealing because it is a conserved simple model that can be applied to studying the more complex functions of multiple RFX genes with multiple isoforms. Also, it is a good model for studying the role of RFX TFs in cilia development, as *C. elegans* DAF-19 was the first RFX to be associated with a role in cilia development (Swoboda *et al.*, 2000).

1.5 Specific Aims of Thesis

This thesis consists of two major specific aims. The first specific aim of this thesis was to establish *C. elegans* as a model organism for studying RFX genes. To do this, I examined the expression patterns of all known *daf-19* isoforms using fluorescence reporter genes (Chapter 2). Second, putative RFX members, RFXDC1 and RFXDC2, were characterized and identified as novel RFX genes in the human genome (Chapter 3). Additionally, I used the *C. elegans* model to initiate the study of the function of a newly identified human RFX gene (Appendix).

Chapter 2: Expression of *daf-19* isoforms

2.1 INTRODUCTION

2.1.1 Expression of *daf-19* isoforms

Four isoforms of *daf-19* have been described (Figure 8). Among them, three have been published and the fourth, *daf-19c*, has not yet been reported in publication but is described in WormBase (http://www.wormbase.org/) (Harris *et al.*, 2010). The longest known isoform of the *daf-19* gene, *daf-19b*, consists of 12 exons and 11 introns, while the second longest isoform, *daf-19a,* shares all of these exons as well, except one alternatively spliced exon 4. The *daf-19c* isoform is the shortest with only 8 exons and *daf-19d* is the second shortest with 9 exons (Figure 8).

The expression of *daf-19* in *C. elegans* was first probed by examining the expression profile of a green fluorescence protein (GFP) reporter gene driven by a construct containing 2.9 kb of the *daf-19* promoter and a ~10 kb genomic region of *daf-19* that contains all exons up to exon 10 and thus excluding the dimerization domains (DIMs) (Figure 9) (Swoboda, Adler et al. 2000). Because this transgene was made from genomic DNA and included nearly all exons and introns of *daf-19* (with the exception of the DIMs), the expression pattern indicated by the GFP profile likely represents all four of the *daf-19* isoforms. The reporter GFP was observed in all ciliated sensory neurons (Figure 9) and for a few worms, in non-ciliated neurons, hypoderm and gut cells. To illustrate that the expression of the reporter construct represents the endogenous

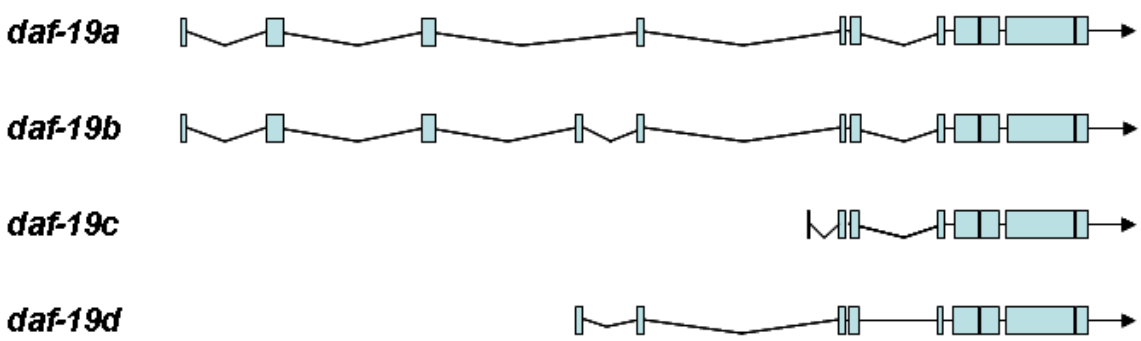


Figure 3 Four known isoforms of *C. elegans daf-19* gene.

Exons are represented as blue squares and introns as black connecting lines. The size and position of the introns and exons are proportional to their actual lengths. Figure was made based on gene models in WormBase (Harris *et al.*, 2010).

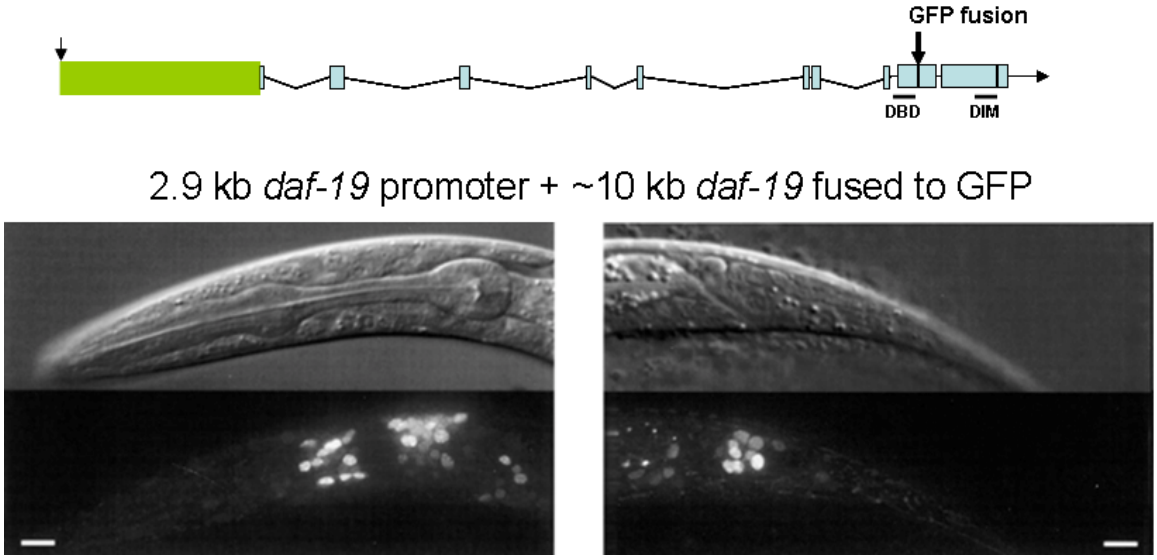


Figure 4 Reporter fusion of *daf-19* with GFP is expressed in ciliated neurons.

The reporter transgene is depicted above the expression pattern image. The thin arrow indicates the start of the reporter transgene and the thick arrow indicates the point at which GFP is fused with the *daf-19* gene. The putative promoter of *daf-19* is indicated by the green box. The sizes of introns, exons and promoter are proportional to their actual lengths. The microscope images show that there is expression in all ciliated sensory neurons, including those in the head (amphid and labial neurons) observed in the left panel and those in the tail (phasmid neurons) seen in the right panel. This figure is adapted from (Swoboda *et al.*, 2000). Bars represent 5 μm.

expression of *daf-19*, Swoboda and colleagues demonstrated that the construct was able to rescue both the Daf-c and Dyf phenotypes associated with the cilia defective *daf-19(m86)* mutant worms (Swoboda, Adler et al. 2000).

In a subsequent study, the Swoboda laboratory characterized the expression and function of three isoforms of *daf-19*: *daf-19a*, *daf-19b*, and *daf-19d* in *C. elegans* using transcriptional reporter constructs containing GFP as well as antibody staining methods (Senti and Swoboda, 2008). They reported that while the *daf-19d* isoform was found to be expressed exclusively in ciliated neurons, the *daf-19a/b* isoforms were found to be expressed in non-ciliated neurons. The *daf-19d* isoform expression was first determined by creating a transgene containing the GFP open reading frame (ORF) driven by a genomic fragment consisting of a specific genomic region of *daf-19*. This genomic fragment, which starts at intron 3 and ends at the end of the *daf-19* gene, thus reprensenting *daf-19d*, was named pGG14. The pGG14 construct was found to be expressed in neurons in the head and tail that are characteristic of ciliated neurons, suggesting a role of *daf-19d* in cilia development. This role was confirmed when it was shown that this transgene was able to rescue both Daf-c and Dyf phenotypes of the *daf-19* worm, suggesting that the *daf-19a* and *daf-19b* isoforms are inessential for both Daf-c and Dyf phenotypes. Furthermore, this transgene was able to activate expression of *osm-5* and *bbs-7* that are known targets of DAF-19 and are also expressed in ciliated sensory neurons (Senti and Swoboda 2008).

To further define the regulatory elements important for expression of the *daf-19d* isoform, they engineered a reporter construct, named pGG20, containing the last 250 bp

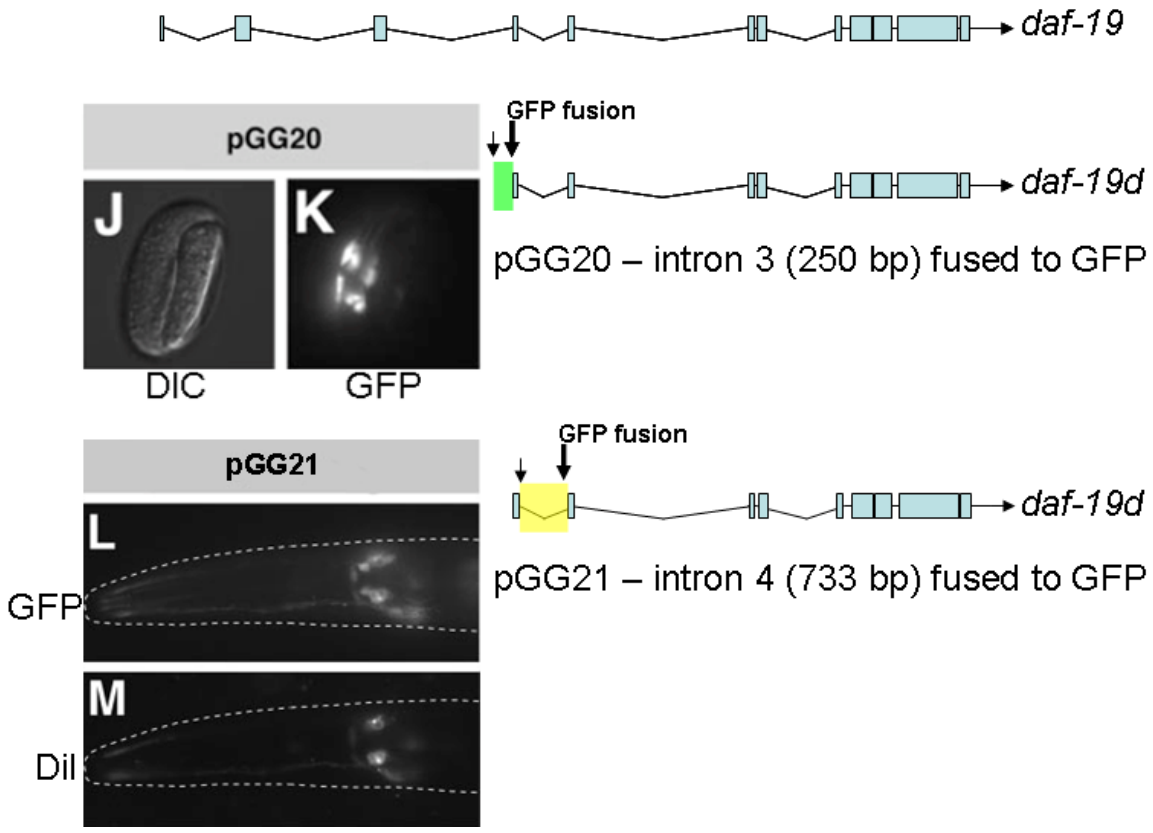


Figure 5 Reporter fusions of *daf-19d* promoters with GFP are expressed in ciliated neurons.

The gene model for the longest isoform *daf-19b* is depicted at the top of the figure as a reference to exons and introns. The expression analysis and microscope images are from (Senti and Swoboda, 2008) and the gene models illustrating the reporter genes made in this study were made by me and thus this figure is adapted from (Senti and Swoboda, 2008). The pGG20 reporter contains 250 bp of intron 3 (relative to *daf-19b*) fused to GFP. The thin arrow indicates the start of the promoter and the thick arrow indicates the end of the promoter where GFP was fused to. The green bar represents the promoter used in this reporter. Its expression is in ciliated neurons from mid-embryo to hatching as shown in (K) from (Senti and Swoboda, 2008). The pGG21 reporter contains the entire intron 4 (relative to *daf-19b*), indicated by the yellow bar, fused with GFP and is expressed in ciliated neurons from mid-embryo to adult. (L) shows the GFP expression in ciliated neurons and (M) shows the dye-filling with DiI in ciliated neurons.

of intron 3 (putative *daf-19d* isoform promoter), which drove expression in ciliated sensory neurons from mid-embryonic to hatching stage (Figure 10). A second reporter construct (pGG21), which contains the entire intron 4 (relative to the longest gene model: *daf-19b*), also drove expression in ciliated neurons, however from mid-embryo to adult stage (Figure 10) (Senti and Swoboda 2008).

These expression studies of *daf-19* isoforms provide important insight into their functions. However, many questions about the expression of *daf-19* isoforms remain to be addressed. First, the expression of a fourth isoform, *daf-19c*, which was recently identified and annotated in WormBase (Harris *et al.*, 2010), has not been reported. It is the shortest isoform among the four isoforms identified (Figure 8). The pGG14 construct used by Senti and Swoboda (2008) actually contains the entire length of this isoform. Thus the expression of *daf-19c* can potentially "contaminate" their reported expression pattern of *daf-19d*. More importantly, the relative expression intensity of these isoforms is unknown. All previous expression assays that applied transgenic *C. elegans* either did not have the transgene integrated into the genome (Swoboda *et al.*, 2000) or had it integrated into the genome in a random manner and not in a single copy (Senti and Swoboda 2008) (Mello *et al.*, 1991), which means a variable number of copies of transgenes per worm and in different locations of the genome. This situation makes it hard to compare the relative expression intensity between different constructs and different strains.

The goal of this project was to study the relative expression of *C. elegans daf-19* so that we can establish *C. elegans* as a model for studying RFX genes and RFX TF-mediated transcription. To achieve this goal, I have employed the newly established

method Mos1 mediated single copy insertion (MosSCI), which allows for integration of a single copy of a transgene into a precisely defined genomic region in the *C. elegans* genome (Frokjaer-Jensen *et al.*, 2008). Using this method, I compared the expression of all four isoforms. Additionally, I examined and compared reporter expression driven by different constructs including transcriptional as well as translational constructs. In particular I examined the expression pattern of the *daf-19c* isoform, whose expression and function has not yet been reported (Harris *et al.*, 2010).

2.1.2 Function of DAF-19 in transcriptional regulation

From the expression profile of *daf-19* in cilia, it was not a surprise when it was found that this gene is important for cilia development. Several ciliary target genes have been found for DAF-19, many of which are associated to disease. In a genome-wide search for X-box motifs in *C. elegans,* 750 putatitive target genes of DAF-19 were found. Amongst these putative genes a group of *bbs* genes, which are associated with the ciliopathy Bardet Biedl Syndrome, were retrieved (Efimenko *et al.*, 2005). Currently, there are 14 *bbs* genes that have been cloned in human (Tobin and Beales, 2007; Leitch *et al.*, 2008) for which 9 have known orthologs in *C. elegans,* excluding *bbs-6, bbs-10, bbs-12, bbs-13* and *bbs-14* (Flicek *et al.*, 2008). Mutations in the *C. elegans bbs* genes have been observed to result in ciliary defects and it has been shown that these genes are localized only in ciliated neurons (Ansley *et al.*, 2003). To determine what function these genes had in cilia formation, a study was done where the specific localization of the *bbs* genes was found in the basal bodies of cilia and that these BBS proteins had IFT motility. This suggested that the *C. elegans* BBS proteins function in the conserved process of IFT (Blacque *et al.*, 2004). This was later confirmed in a study that observed the phenotype

of *bbs* mutants, in particular *bbs-7* and *bbs-8,* and found that there was a breakdown of IFT particles in these mutants, thus demonstrating the role of BBS proteins in IFT (Ou *et al.*, 2005).

Other studies have also found that the target genes of DAF-19 are ciliary genes that are associated to IFT. For example in a study by Yu and colleagues, it was found that DAF-19 regulates *lov-1* and *pkd-2* genes which are orthologs to human *pkd-1* and *pkd-2* genes respectively (Yu *et al.*, 2003). The human *pkd* genes are associated with human polycystic kidney disease (PKD) since mutations in these genes lead to PKD. Another study showed that the *xbx-1* gene in *C. elegans,* involved in retrograde IFT, is also regualated by DAF-19 (Schafer *et al.*, 2003). Additionally, the DYF-2 protein, which is required for maintaining the structure of IFT, was shown to be regulated by DAF-19 since the expression of *dyf-2::*GFP was greatly reduced in *daf-19* mutants (Efimenko *et al.*, 2006). Hence, it is evident that DAF-19 is an important regulator of the IFT process and that *C. elegans* is a very suitable model to study ciliary genes.

2.1.3 Construction of reporter constructs

The use of reporter fusions has been a widely used method for studying the expression pattern of genes in both prokaryotes and eukaryotes (Slauch and Silhavy, 1991). Reporter fusions are comprised of two components, a *cis*-regulatory element fused (via PCR stitching (Hobert, 2002)) to a reporter gene. It is important to define these two components, which reporter to use as well as which regulatory region to use, before constructing fusion transgenes.

The first reporter gene in *C. elegans* for observing gene expression was the *LacZ* gene which encodes β-galactosidase (Fire *et al.*, 1990). This method required that the worms be fixed and stained with X-Gal so that the localization of β-galactosidase could be observed in the worms (Lis *et al.*, 1983). Following the discovery of a green fluorescent protein (GFP) isolated from the jellyfish, *Aequorea victoria* (Morin and Hastings, 1971), Martin Chalfie and colleagues were the first to utilize GFP reporter fusions in bacteria, *Escherichia coli* and in the nematode, *C. elegans* (Chalfie *et al.*, 1994). GFP as a reporter has an advantage over *LacZ* in that its expression can be observed in live worms. Recently, there have been various other fluorescent proteins discovered that also serve as successful reporter genes (Stepanenko *et al.*, 2008). One example is the mCherry reporter which is the one that I used for all of my reporter constructs. It is a monomeric (and thus less toxic) fluorescent protein, that has high pH stability, high photostability and a fast maturation rate (Shaner *et al.*, 2004). The *cis*-regulatory element sequence also needs to be carefully chosen. In *C. elegans*, *cis*-regulatory elements are usually found within several kilobases upstream of the gene start, with some also found in introns (Wenick and Hobert, 2004), in the 3'-UTR region (Wightman *et al.*, 1993) and far upstream or downstream of the gene start (Conradt and Horvitz, 1999).

There are two main types of reporter fusions, transcriptional and translational. In transcriptional fusions, only the region 5' upstream of the translational start site (TSS) is fused to the reporter gene. This method is technically very straightforward but may not fully represent the expression pattern since any *cis*-regulatory information found in introns, 3'-UTRs or highly distant regions would be excluded. Translational fusions are

more representative of the true expression pattern of the gene since they usually contain the promoter as well as the entire gene including introns and 3'UTR (Boulin *et al.*, 2006). However, it is technically challenging to engineer constructs with genes that are long (>10 Kb). Additionally, using the entire gene in a translational fusion construct does not allow the dissection of different expression patterns for different isoforms.

In my thesis project, I chose to use transcriptional fusions to respresent each of the *daf-19* isoforms. This strategy was chosen to probe the differential expression patterns of different isoforms because a translational fusion would not give resolution of the expression for each of the different isoforms. The downside of this strategy is that it is possible that the promoters chosen for each isoform do not necessarily include all *cis*-regulatory elements involved in transcription. To evaluate the effect of possible loss of important *cis*-elements, I tested the significance of using different promoter sizes by using promoters of different lengths and positions.

2.1.4 Single copy transgene integration into the genome

Once suitable promoters and a reporter gene were chosen, the next critical step was to introduce them into the worm genome. Standard injection of exogenous DNA into *C. elegans* was first introduced in 1985 and it was found that these transgenic DNA would form high molecular weight extrachromosomal arrays within the cells they were injected (Stinchcomb *et al.*, 1985). Although injection of transgenic extrachromosomal arrays has been a popular method of choice for many years, it has several drawbacks. First, the extrachromosomal array does not always get transferred to both daughter cells during mitosis and meiosis, which can lead to mosaicism in expression. Second, arrays are silenced in some tissues such as the germline (Kelly *et al.*, 1997) and in muscle

(Hsieh and Fire, 2000). Furthermore, arrays have a tendency to "drift" over generations which means that the extrachromosomal array is eventually lost from some cells causing the expression pattern to change over time (Sha and Fire, 2005). Lastly, arrays contain the transgene in hundreds of copies causing it to be over-expressed and can lead to toxic effects (Thellmann *et al.*, 2003). A method that increases the ability of the transgenic DNA to integrate randomly into the worm genome was discovered by adding single stranded oligonucleotides into the injection mix (Mello *et al.*, 1991). It was also found that homologous recombination of the array into the genome can occur after gene bombardment or after injection into nuclei instead of into the cytoplasm (Broverman *et al.*, 1993; Berezikov *et al.*, 2004). These methods were an improvement to the issues with extrachromosomal arrays however they still did not allow for a single copy of the transgene to be inserted into a specific intergenic section of the genome.

A new approach called MosSCI has been developed for obtaining near-endogenous expression of transgenes (Frokjaer-Jensen *et al.*, 2008). In this method, the transgene is injected into a specially constructed *C. elegans* strain that contains a *Drosophila* Mos1 transposon inserted into an intergenic region within the genome. A vector containing sites homologous to the region flanking the Mos1 transposon is used for cloning the transgene of interest into it in such a way that the homologous arms will flank the insert (Figure 11). A Mos1 transposase is co-injected with the transgene so that it mobilizes the Mos1 transposon out of the genome, creating double stranded breaks. These breaks are then repaired by insertion of the transgene into the emptied Mos1 site, through homologous recombination between the homologous arms. In order to detect strains carrying the array or insertion, the Mos1 carrying strain is also an *unc-119(ed3)* mutant (is uncoordinated in

movement) which is then rescued by a positive selection marker, a wild-type *unc-119* gene. The *unc-119* rescue gene is incorporated into the Mos vector and ends up being directly upstream of the inserted gene of interest, within the two homology arms, so that it is also inserted into the Mos1 site. Thus, all worms that move as wild-type worms, contain the array or an insertion. In order to distinguish between worms carrying the array from those that have an insertion, three mCherry vectors are used as a negative selection marker (Figure 11). Since all constructs that are injected form an extrachromosomal array so that all constructs are passed on together, worms that have lost the array and thus have lost mCherry fluorescence but remain wild-type moving, are those that contain a direct insertion of the transgene.

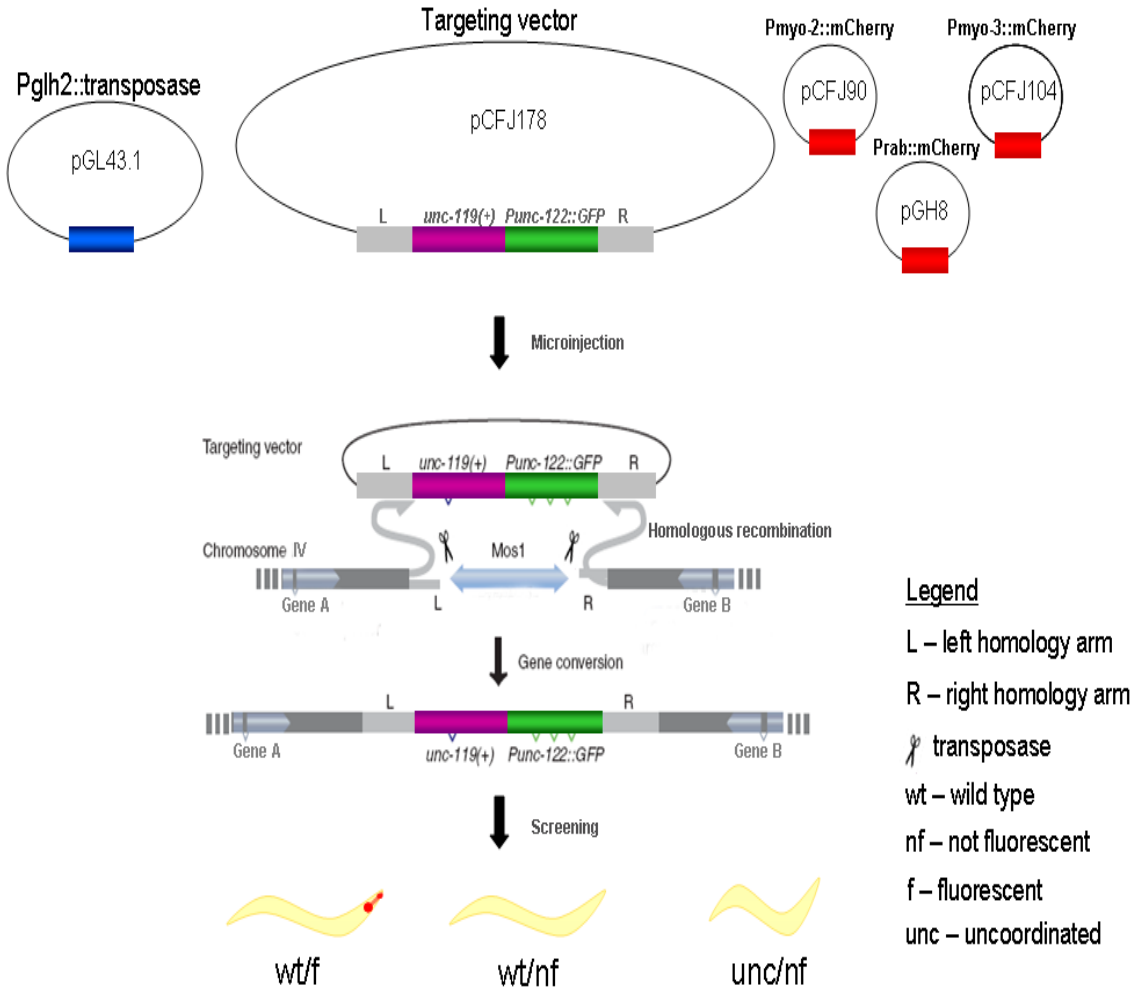


Figure 5 MosSCI Direct Insertion Method.

This figure is adapted from (Frokjaer-Jensen *et al.*, 2008). The five vectors that are microinjected are depicted. Pglh2::transposase containing vector (pGL43.1) has a germ-line induced promoter driving Mos1 transposase transcription. The targeting vector (pCFJ178) contains the transgene of interest cloned in front of an *unc-119* rescue gene and in between left (L) and right (R) homology arms. The three mCherry containing vectors, pCFJ90, pCFJ104 and pGH8 have mCherry under the control of Pmyo-2 (pharyngeal), Pmyo-3 (body wall) and Prab (pan-neuronal) promoters respectively. Injection into the Mos strain with Mos1 transposon in an intergenic region between Gene A and Gene B followed by transposase excision of the Mos1 element and homologous recombination of homology arms between the targeting vector and the chromosome insertion site is depicted. Gene conversion of the transgene of interest into the Mos1 site occurs. Screening for worms with direct insertions follows, where wild-type moving and non-fluorescing worms represent those that have lost the extrachromosomal array (containing mCherry vectors) and contain an insertion of the *unc-119* gene along with the transgene of interest.

2.2 MATERIALS AND METHODS

2.2.1 Selection of promoters

Translational reporters are commonly used for expression studies because they are more likely to contain all *cis*-regulatory elements as opposed to transcriptional reporters (Boulin *et al.*, 2006). In this study, I first chose to use transcriptional reporters to respresent each of the *daf-19* isoforms. It is possible that the promoters chosen do not include all *cis*-regulatory elements used in transcription; however they allowed me to define the expression of the individual isoforms, which would not be possible with a translational reporter. For the *daf-19a/b* isoform expression I used two different promoters, ***p****daf-19a/b* (long) (where the ***p*** stands for **p**romoter of *daf-19a/b*) and *pdaf-19a/b* (short). The long version contains ~2 kb upstream of the TSS of the *daf-19* gene and the short version contains ~400 bp upstream of the TSS (Table 2 & Figure 7). These two versions of the *daf-19a/b* promoter allowed me to test the significance of using different promoter sizes. The choice for the long *daf-19a/b* isoform promoter was based on the previous transcriptional fusion done by our lab (Figure 17) (Tarailo-Graovac M., unpublished). The short isoform promoter was chosen so that it still contains the two predicted X-box motifs found ~100 bp upstream the TSS of *daf-19a/b* (Figure 36). The choice for the *daf-19c* isoform promoter was not based on previous studies and I wanted to include the entire intron sequence upstream of the *daf-19c* TSS (Table 2 & Figure 7). For the *daf-19d* isoform, I chose two different promoters which were based on the Senti and Swoboda study. These two promoters, pGG20 and pGG21 (Table 2 & Figure 7), were described as being sufficient to drive expression in ciliated sensory neurons (Senti and Swoboda, 2008).

Table 1 – List of promoters selected and where they have been used previously.

Promoter	Length (bp)	Reference
pdaf-19a/b (long)	1924	(Tarailo-Graovac M., unpublished)
pdaf-19a/b (short)	396	this study
pdaf-19c	1985	this study
pdaf-19d (pGG20)	268	(Senti & Swoboda, 2008)
pdaf-19d (pGG21)	733	(Senti & Swoboda, 2008)

2.2.2 Generating transcriptional fusions

The primers used for amplifying the short *daf-19a/b* isoform promoter, *pdaf-19a/b* (short), from *C. elegans* N2 genomic DNA (isolated by Dr. A. Mah) were pd19abs_f (with AvrII site) and pd19ab_r (Table 3), of which the first 25 bp are the reverse complement of the first 25 bp of mCherry. The pd19ab_r primer is designed in this way so that the end of the promoter can be stitched to the start of mCherry by the PCR stitching method (Hobert, 2002). All of the reverse primers for amplifying the different *daf-19* isoform promoters were designed in this same manner. Primers used for amplifying the long *daf-19a/b* isoform promoter, *pdaf-19a/b* (long), were pd19ab_f and pd19ab_r (Table 3). Primers used for amplifying the *daf-19c* isoform promoter, *pdaf-19c*, were pd19c_f and pd19c_r (Table 3). Primers used for amplifying the *daf-19d* isoform promoter, *pdaf-19d*, were A*(pd19c_f) and pd19d_r (Table 3). Primers for amplifying

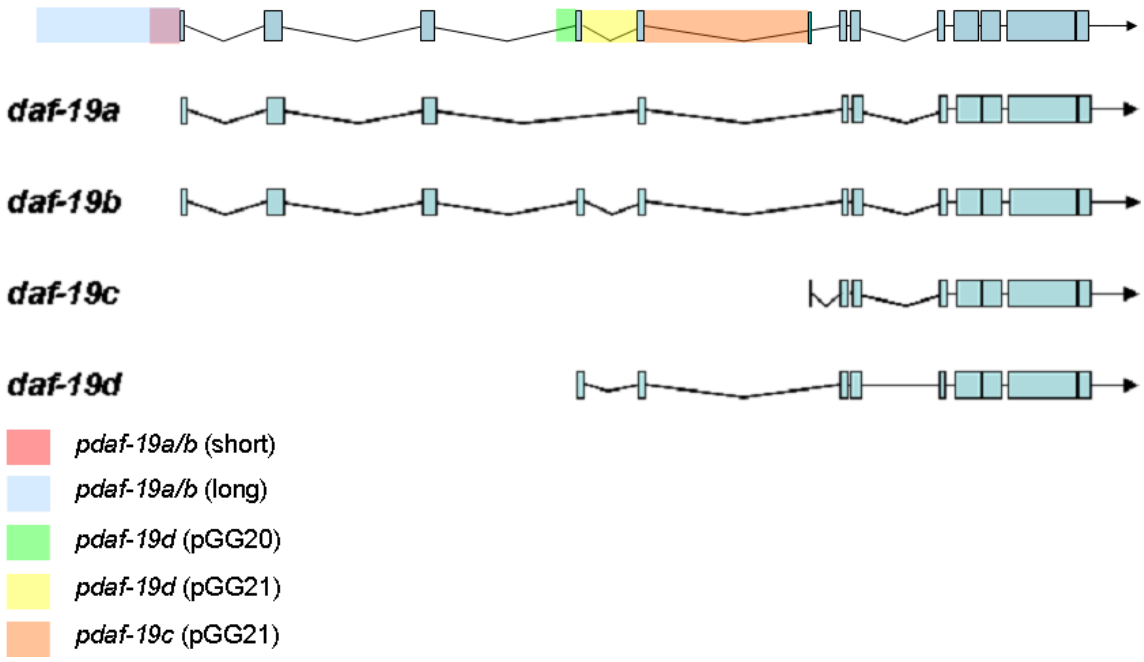


Figure 6 *daf-19* isoform promoters selected for creating transcriptional fusions.

Exons are represented as blue squares and introns as black connecting lines. The size and position of the introns and exons are proportional to their actual lengths. Promoters are represented by coloured blocks and they are also proportional to their actual lengths. Figure was made based on gene models in WormBase (Harris *et al.*, 2010).

mCherry, including the Unc-54 3'-UTR, from the pCFJ90 vector (a kind gift from the Jorgensen lab) were Rfp_f and Rfp_r (Table 3). The DNA was amplified using Finnzymes Phusion DNA Polymerase and the phusion 58 PCR program (see Table 2A in appendix).

Primers used to stitch *pdaf-19a/b* (short) to mCherry were pabs_f* (AvrII) with an AvrII restriction site and GGATAA filler sequence (to aid in digestion) and Rfp_r* (SbfI) with SbfI restriction site and AGGCGG filler sequence (Table 3). These filler sequences were used in all primers for stitching the promoters to mCherry. Primers for stitching *pdaf-19a/b* (long) to mCherry were pD19ab_f* (SbfI) and Rfp_r* (XhoI) (Table 3). Primers for stitching *pdaf-19c* to mCherry were pd19c_f* (AvrII) and Rfp_r* (SbfI) (Table 3). Primers for stitching *pdaf-19d* (pGG20) to mCherry were pd19ds_f* (AvrII) and Rfp_r* (SbfI) (Table 3). Primers for stitching *pdaf-19d* (pGG21) to mCherry were Intron4_F* (XhoI) and Rfp_r* (SpeI) (Table 3). The program used for all promoter fusions to mCherry was the Stitch 1 program (Table 2A in appendix). All primers were ordered from Integrated DNA Technologies (IDT).

2.2.3 Generating translational fusion of the *daf-19c* isoform

Primers used for amplifying the *daf-19c* isoform promoter, *pdaf-19c* (Table 2), up to the end of the *daf-19c* gene but excluding the stop codon, were pd19c_f and d19r_mCherry (Table 4). The pd19c_f primer is designed so that the last 25 bp are the

Table 2 - Transcriptional fusion primers to generate fusions of *daf-19* isoform promoters stitched to mCherry.

PCR Product	Primer Name	Primer Sequence
pdaf-19a/b (short)	pd19abs_f (avrII)	GGATAACCTAGGCACACACACATATCTCCTTT
	pd19ab_r	TATCTTCTTCACCCTTTGAGACCATGACTTTCTTCTCTGCCGCA
pdaf-19a/b (long)	pd19ab_f	TGATTCCGACGTTGGCTTTC
	pd19ab_r	TATCTTCTTCACCCTTTGAGACCATGACTTTCTTCTCTGCCGCA
pdaf-19c	pd19c_f	CGCGAGAGGAATTCGACTAT
	pd19c_r	TATCTTCTTCACCCTTTGAGACCATGATTGTAAGAGAATTAAGCT
pdaf-19d (pGG20)	A*(pd19c_f)	TTCCGGTGCCATTAGGTATC
	pd19d_r	TATCTTCTTCACCCTTTGAGACCATCTAAATGGAAGATGGTCATAGTTG
pdaf-19d (pGG21)	intron4_F	TTGCCTATGGAGAGGAGTCG
	intron4_R	TATCTTCTTCACCCTTTGAGACCATCTGAAAATTTTCGAAATTTA
mCherry	Rfp_f	ATGGTCTCAAAGGGTGAAGA
	Rfp_r	GGCCTCTTCGCTATTACGC
pdaf-19da/b (short)::mCherry	pabs_f* (avrII)	GGATAACCTAGGCACACACACATATCTCCTTT
	Rfp_r*(SbfI)	AGGCGGCCTGCAGGACGACGGCCAGTGAATTATC
pdaf-19a/b (long)::mCherry	pD19ab_f*(SbfI)	GGATAACCTGCAGGTTGACGGAAGATTCACAAGAA
	Rfp_r*(XhoI)	AGGCGGACTAGTacgacggccagtgaattatc
pdaf-19c::mCherry	pd19c_f*(AvrII)	GGATAACCTAGGGTTGTGAAATATAATTGGGGAG
	Rfp_r*(SbfI)	AGGCGGCCTGCAGGACGACGGCCAGTGAATTATC
pdaf-19d (pGG20)::mCherry	pd19ds_f*(AvrII)	GGATAACCTAGGCAGTGCCCTAACGACTCACA
	Rfp_r*(SbfI)	AGGCGGCCTGCAGGACGACGGCCAGTGAATTATC
pdaf-19d (pGG21)::mCherry	Intron4_F*(XhoI)	GGATAACTCGAGgtaagtgaattggtttgtta
	Rfp_r*(SpeI)	AGGCGGACTAGTACGACGGCCAGT

reverse complement of the first 25 bp of mCherry, enabling the end of the promoter to be stitched to mCherry. The Phusion 58 program (Table 2A in appendix) was used with a 4 minute elongation time to amplify the *daf-19c* promoter continuous with the *daf-19c* gene, *pdaf-19c_daf-19c*. Primers used for fusing this product, *pdaf-19c_daf-19c*, to mCherry were pd19c_f* (with AvrII site), with GGATAA as the filler sequence, and Rfp_r* (with SbfI site), with AGGCGG as the filler sequence (Table 4). The program used for the fusion PCR was the Stitch 2 program (Table 2A in appendix), with 4 min. elongation times.

Table 3 – Translational fusion primers to generate *pdaf-19c_daf-19c*::mCherry.

PCR Product	Primer Name	Primer Sequence
Pdaf-19c_daf-19c	pd19c_f	CGCGAGAGGAATTCGACTAT
	d19r_mCherry	TATCTTCTTCACCCTTTGAGACCATCAGAAGACCTGCTTTCTCGA
Pdaf-19c_daf-19c ::mCherry	pd19c_f*	GGATAACCTAGGGTTGTGAAATATAATTGGGGAG
	Rfp_r*	AGGCGGCCTGCAGGACGACGGCCAGTGAATTATC

2.2.4 Cloning into the pCFJ178 Mos vector

To insert the transcriptional fusion reporters into the pCFJ178 vector (Addgene), which has homology arms specific to homology arms on chromosome IV, both the vector and the transgenes were digested with the same restriction enzymes. For all transgenes except for *pdaf-19a/b* (long)::mCherry, New England BioLabs Inc.® (NEB) AvrII (#R0174S) and SbfI (#R0642S) were used for double digest. The restriction enzymes used for *pdaf-19a/b* (long)::mCherry were SbfI (#R0642S) and XhoI (#R0146S). The pCFJ178 vector was digested with the same enzymes as the transgene that would be ligated into it. Conditions for AvrII with SbfI double digest and SbfI with XhoI double

digest were to digest ~0.8 µg of transgene and ~1.0 µg of vector (pCFJ178) separately for 1.5 hr at 37˚C. Digested vector was de-phosphorylated with Roche alkaline phosphatase (catalog# 28-9034-71) to prevent re-ligation of vector, in the case where only one enzyme successfully cuts the vector. De-phosphorylation was done for 1 hour at 37˚C. After de-phosphorylation of the digested vector, the band corresponding to the size of the pCFJ178 vector was gel purified with the Illustra GFX^{TM} PCR DNA and Gel Band Purification Kit (catalog# 374804). The digested transgenes were also PCR purified with this kit. The transgene and vector were ligated together with Fermentas T4 DNA ligase (catalog# K1214) using a 1:8, vector:insert ratio. The ligated constructs were then transformed via electroporation into DH5α electrocompetent cells. Colonies were tested via single colony PCR using primers to amplify a segment of the vector up to a segment of the insert. Positive colonies were then cultured in LB media containing 100 µg/mL ampicillin overnight at 37˚C with shaking. Plasmid DNA from the overnight cultures was miniprepped with Fermentas GeneJET Plasmid Miniprep Kit (catalog #K0503) and a quality check PCR of the mini-preps was done to ensure the plasmid DNA was successfully obtained.

To insert the translational fusion reporter, *pdaf-19c_daf-19c*::mCherry, into the pCFJ178 vector, both the pCFJ178 vector and the transgene were digested with NEB SbfI (#R0642S) and SpeI (#R0133S) restriction enzymes. Conditions for SbfI and SpeI double digest were to digest ~0.8 µg of transgene and ~1.0 µg of pCFJ178 vector separately for 1.5 hr at 37˚C. Digested vector was de-phosphorylated with Roche alkaline phosphatase for 1 hour at 37˚C. The rest of the steps were the same as those described above for the transcriptional fusions.

2.2.5 Microinjection

Mos strain EG5003 uncoordinated (*unc-119*) worms (*Caenorhabditis* Genetics Center), containing a Mos1 transposon element within an intergenic region on chromosome IV, were injected with the transcriptional and translational fusion transgenes. Components of the injection mix included, the fusion transgenes in the pCFJ178 vector, at 50 ng/µL along with three mCherry markers (kind gifts from the Jorgensen lab), pCFJ90 (pharynx muscle), pCFJ104 (body muscle) and pGH8 (pan-neuronal) at the following concentrations, 2.5 ng/µL, 5 ng/µL and 10 ng/µL, respectively and a transposase vector, pJL43.1 (a kind gift from the Jorgensen lab) at 50 ng/µL. The MosSCI protocol for direct insertions was followed thereafter (Frokjaer-Jensen *et al.*, 2008). Briefly, injected worms were individually plated after a 3 hour recovery period at room temperature. These plates were left at room temperature for ~3 days after which the progeny were checked for successful injection. Successful injection was identified by a rescue of the uncoordinated phenotype (Unc). The plates with rescued F1 progeny were kept at 25˚C until starved. This preferentially selected for rescued worms over Unc worms, since Unc worms are at a disadvantage for obtaining food due to their impaired mobility. These plates were then screened for direct insertion by looking for non-fluorescing, but wild-type moving worms. If no direct insertion candidate worms were found at this point, the plates were chunked in quarters onto new plates and kept at 25˚C (to promote loss of extrachromosomal array) until starved and then these were screened for direct insertion again. All injections described in this thesis were generously performed by Domena Tu of the Baillie lab.

2.2.6 Lysis of worms

Single worm lysis was done using 10 mg/mL Roche proteinase K and 1x lysis buffer. A single worm was placed into 5 μL of lysis buffer with 3 μL proteinase K and incubated at -80˚C for 15 min. The tubes were then put in a lysis PCR cycle. The lysis PCR program was 60˚C for 1 hr followed by 95˚C for 15 min (Table 2A in appendix).

2.2.7 Screening for and confirming direct insertions

Direct insertion candidate worms, of the *daf-19* isoform fusion transgenes into the EG5003 strain (Table 5), were identified as non-fluorescing and wild-type moving worms on a Zeiss fluorescence dissection microscope (Stemi SV11). PCR genotyping was done to confirm a single homozygous insertion of the transgene into the Mos-1 site of chromosome IV. To confirm homozygous insertion, the following three primers were used in one PCR reaction: a forward primer in chromosome IV just before the Mos flanking homologous arms (ChIVgenoF1): ggagaccagggagacaagg (blue arrow in Figure 12), a reverse primer specific to the *unc-119* portion of the vector that gets inserted along with the transgene into the Mos1 site (Unc119R2): gtgtgctgctcggttaagag (green arrow in Figure 12) and a reverse primer inside the Mos right homologous recombination arm (mos178genoR): ccaagatcaaatgcacagga (purple arrow in Figure 12). In the event of a direct insertion, only a 2.2 kb band amplifies and corresponds to the product of amplification by the ChIVgenoF1 and Unc119R2 primers. In the absence of a direct insertion, a 3.2 kb band is amplified by the ChIVgenoF1 and mos178genoR primers (Figure 12). Not only does this PCR genotyping strategy determine whether a direct insertion occurred, it can also distinguish whether this insertion occurred on both chromosome pairs. In the event of a homozygous insertion, only a 2.2 kb band would be

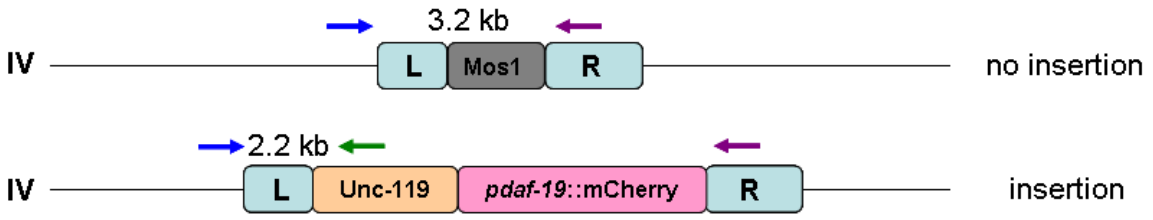
present and in the event of a heterozygous insertion, both a 2.2 kb and a 3.2 kb band are amplified corresponding to the insertion on one chromosome and the presence of the Mos1 transposon in the other chromosome. To confirm that this homozygous insertion was a single insertion and not a duplicate insertion, phusion polymerase was used to amplify the DNA between the two chromosomal arms that flank the insert. The presence of a band indicated a single insertion whereas the absence of a band would suggest that a duplicate insertion had occurred and was too large to amplify. All transgenic strains made are in Table 5.

2.2.8 Dye filling assay

Fluorescent dye filling with Invitrogen Molecular Probes 3, 3'-dioctadecyloxacarbocyanine perchlorate, DiO (Lot# 29097W) was done by adding 5 µl of 2 mg/mL stock of DiO to 1 mL of worms suspended in M9 media (giving a 10 µg/mL final concentration of DiO). Worms were incubated for 2 hr with dye and then left to feed for 1 hr on plates seeded with OP50. For labelling labial neurons in addition to amphid neurons, worms were incubated in 1 mL of 50 mM calcium acetate (diluted in ddH_2O) plus 10 µg/mL DiO and incubated for 2 hr followed by 1 hr of feeding on OP50 plates.

2.2.9 Expression analysis

The transgenic worms were observed on a Zeiss confocal spinning disc microscope (Observer Z1) using a net 60x magnification in oil. This microscope has a 2.487 µm/10 pixel resolution. The worms were prepared for viewing by washing plates of



Insertion Genotype	Expected Bands
Homozygous Insertion	2.2 kb
Heterozygous Insertion	2.2 kb & 3.2 kb
No Insertion	3.2 kb

Figure 7 PCR genotyping strategy for direct insertions.

Three primers (blue, green and purple arrows) were used in one PCR to detect direct insertions. In the case of no insertion, a 3.2 kb band amplifies and represents the region between the left side of the chromosome IV insertion site (blue arrow) and the right homology arm (purple arrow) with the Mos1 transposon in the insertion site. If there is an insertion of the transgene, a 2.2 kb band will amplify and corresponds to the region between the left side of the chromosome IV insertion site (blue arrow) and the *unc-119* rescue gene (green arrow). If there is an insertion but only in one of the pairs of chromosome IV (referred to as a heterozygous insertion), both a 2.2 kb and a 3.2 kb band are amplified.

Table 4 – List of strains used and generated with description of genotype.

Strain Name	Transgene direct insertion	Description
EG5003	Mos1 element	Mos1 insertion in chromosome IV
daf-19(m86)	N/A	C to T stop mutation
JNC60	*pdaf-19c*::mCherry	In EG5003 background
JNC61	*pdaf-19a/b* (long)::mCherry	In EG5003 background
JNC62	*pdaf-19a/b* (short)::mCherry	In EG5003 background
JNC63	*pdaf-19d* (pGG20)::mCherry	In EG5003 background
JNC65	*pdaf-19d* (pGG21)::mCherry	In EG5003 background
JNC66	*pdaf-19c_daf-19c*::mCherry	In EG5003 background
JNC64*	*pdaf19d* (pGG20)::RFX6::GFP	In EG5003 background

*described in Chapter 4.

worms to be observed with water and placing the wash into a 1.5 mL tube. These were then centrifuged at 4000 rpm for 1 min. in an Eppendorf tabletop centrifuge (model 5418). The supernatant was discarded and 3 µL of the pelleted worms was placed on top of 3% agarose mounted on a glass slide. To paralyze the worms, 3 µL 2% sodium azide was added to the worms on the agarose and a coverslip was place on top of this prepared slide. A minimum of 10 different hermaphrodites and 10 different males were observed for each transgenic worm observation and at least 10 different worms were observed for each of the stages.

2.3 RESULTS & DISCUSSION

2.3.1 Using MosSCI to generate integrated transgenes

The **Mos**1-mediated **s**ingle **c**opy gene **i**nsertion (MosSCI) method has recently emerged as a useful tool for obtaining stably integrated transgenic strains that contain a

single copy insertion of the transgene within an intergenic region of the worm genome (Frokjaer-Jensen *et al.*, 2008). This method involves three major steps: 1) generating the transgene, 2) cloning the transgene into the Mos vector and 3) screening worms transformed with the vector for direct insertion. I used this method for generating stable integrated lines containing transgenes for each of the *daf-19* isoform promoters.

For the first step, five transcriptional fusions corresponding to the three *daf-19* isoform promoters, *pdaf-19a/b* (where *pdaf-19a/b* means promoter of *daf-19a/b*), *pdaf-19c* and *pdaf-19d* were generated via PCR stitching (Hobert, 2002) (Figure 13, Figure 1A-3A). There were five transgenes for the three promoters because two different versions of the *daf-19a/b* promoter and two different versions of the *daf-19d* promoter were chosen (see Materials and Methods for details). In the second step, the fusion transgenes were cloned into the pCFJ178 vector, which has homology arms that match within an intergenic region of chromosome IV (Figure 11). These constructs were transformed into bacteria after which the colonies were PCR tested to confirm the presence of the reporter transgene in the vector (Figure 14, Figure 4A-6A). The successful constructs were then used for injection into EG5003 worms, which have left and right homologous arms (homologous to arms in pCFJ178) flanking a Mos1 transposon in an intergenic region of chromosome IV. The last step of finding candidate

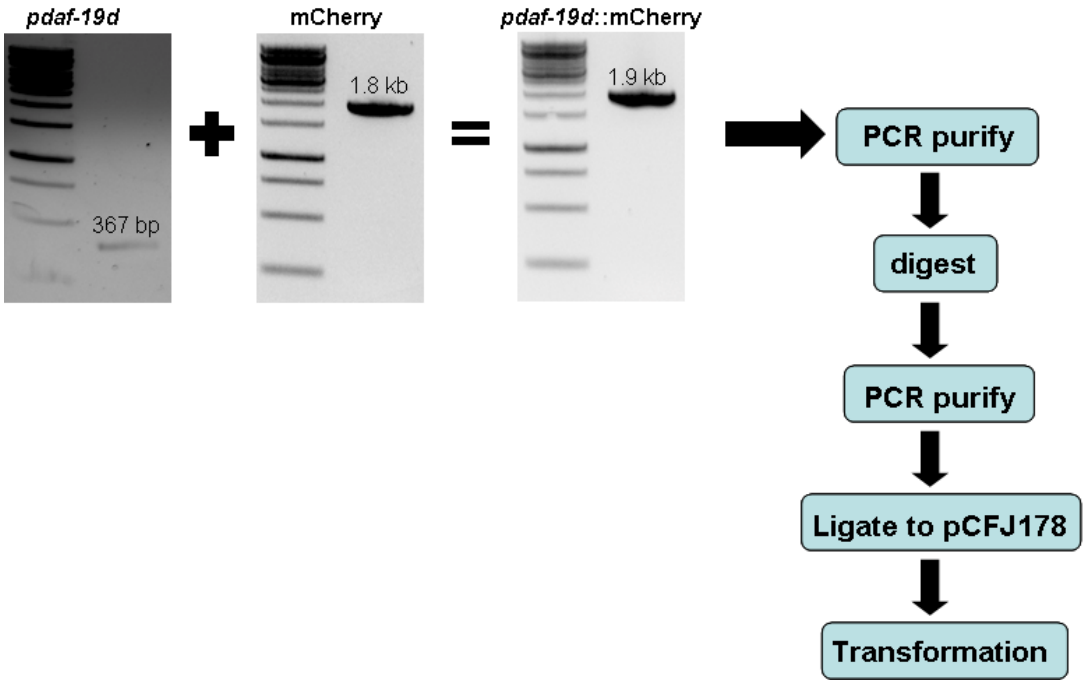


Figure 8 Flow diagram of transcriptional fusion PCR and cloning into the pCFJ178 vector.

The first gel image shows the amplification product of the *daf-19d* (pGG20) isoform promoter. The second gel image shows the band amplified corresponding to mCherry from the pCFJ90 vector. The + sign between these two images indicates a PCR stitching step. The third gel image shows the product of the PCR stitching, *daf-19d* (pGG20) promoter stitched in front of the mCherry gene (*pdaf-19d* (pGG20)::mCherry). The steps following generation of the transgene are indicated in the flow diagram. After each reporter transgene is made, it is PCR purified, followed by digestion with the appropriate restriction enzymes so that it can be cloned into the pCFJ178 vector. A second PCR purification is done after restriction digestion and then the transgene is ligated into the pCFJ178 vector. Following ligation, a transformation of the vector containing the transgene into bacteria is performed.

direct insertion worms involved PCR genotyping the candidates to confirm direct insertion. Three primers (Figure 12) were used in this PCR to determine whether an insertion was found and whether it was homozygous, since a strain that is homozygous for direct insertion is desired in the end. One forward primer (blue arrow, Figure 12) was specific to a region within chromosome IV, a second reverse primer (purple arrow, Figure 12) was specific to a region within right homologous recombination arm and a third reverse primer (green arrow, Figure 12) was specific to the *unc-119* rescue gene, which is a part of the inserted transgene. In the case of a homozygous direct insertion only a 2.2 kb band is observed which corresponds to the region between the forward primer in chromosome IV and the reverse primer in the *unc-119* gene. Successful direct insertions were achieved for all isoform constructs (Figure 15).

2.3.2 Confirming previously observed expression patterns using MosSCI

The *daf-19a/b* isoform expression has been shown to be in non-ciliated neurons by using antibodies specific to these two isoforms (Senti and Swoboda, 2008). However, the expression appears in more tissues for translational reporters of *daf-19a/b*, appearing in non-ciliated neurons, gut and hypodermis (Swoboda *et al.*, 2000). In agreement with these results, my transgenic worms containing the long *pdaf-19a/b* reporter construct also have a more wide-spread expression pattern as well as what appears to be expression in non-ciliated neurons, however individual cells cannot be resolved (Figure 16). Transgenic adult hermaphrodites express mCherry in the hypodermis, pharynx, gut and non-ciliated neurons. Expression appears from the two-fold embryonic stage and is expressed at the

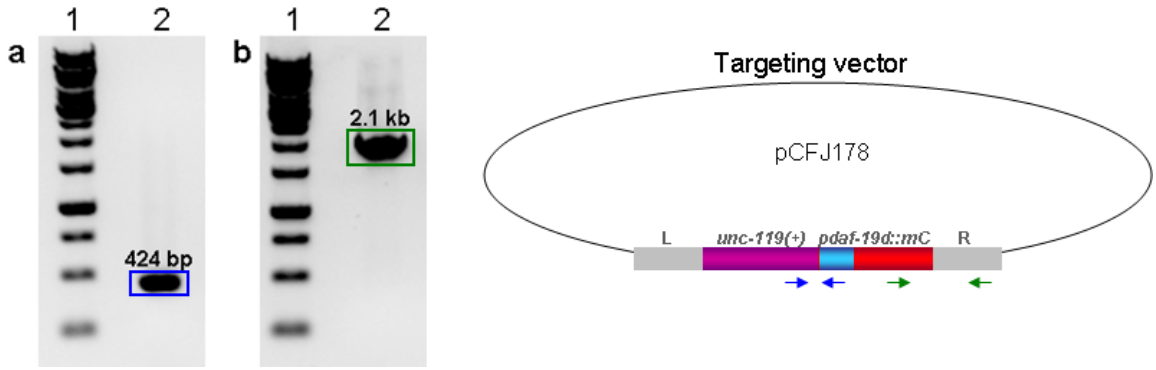


Figure 9 Confirmation of successful cloning of reporter gene into the pCFJ178 targeting vector.

Single-colony PCR genotyping of transformed bacteria was done for each reporter transgene cloned into the targeting vector, pCFJ178, to ensure it was properly ligated into the vector. **(**A) Lane 1 is Fermentas O'GeneRuler 1 kb ladder. Lane 2 is amplification of a single colony using primers (blue arrows) to amplify the left junction site containing pCJF178 *unc-119* sequence and the *daf-19d* (pGG20) promoter sequence. The expected 424 bp band was amplified. (B) Lane 1 is Fermentas O'GeneRuler 1 kb ladder. Lane 2 is amplification of the right junction site with primers (green arrows) specific to mCherry sequence and the pCFJ178 right homology arm sequence. The expected 2.1 kb band was amplified.

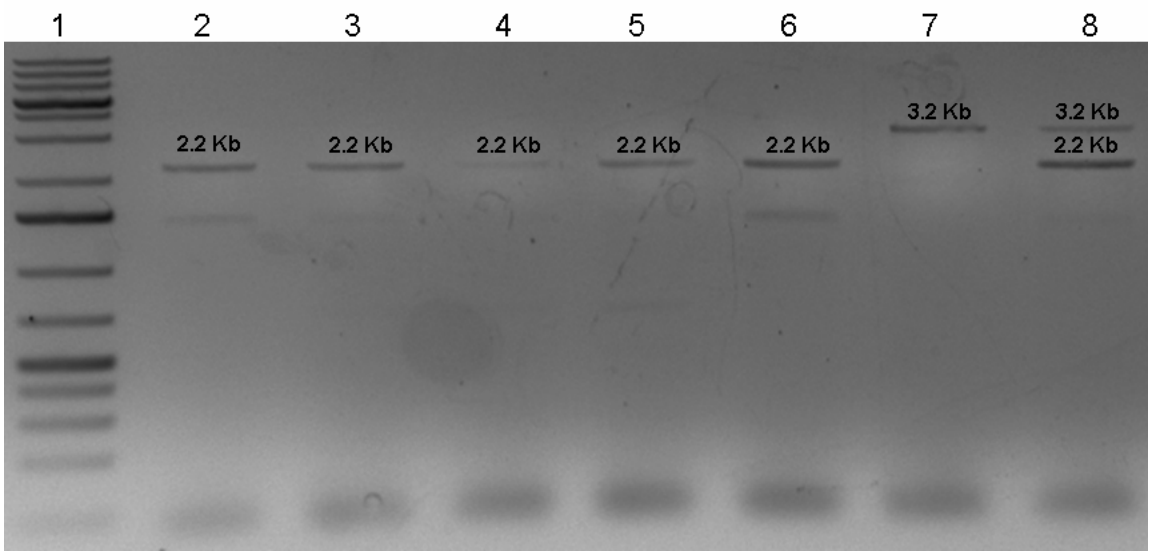


Figure 10 PCR genotyping results for direct insertion of the *daf-19* isoform reporter transgenes into EG5003 Mos strain.

Lane 1 is Fermentas GeneRuler 1 kb Plus ladder. Lane 2 is worm lysate of the JNC62 strain, containing *pdaf-19a/b*(short)::mCherry, Lane 3 is worm lysate of JNC61 containing *pdaf-19a/b*(long)::mCherry, Lane 4 is worm lysate of JNC60 containing *pdaf-19c*::mCherry, Lane 5 is JNC63 containing *pdaf-19d* (pGG20)::mCherry, Lane 6 is JNC65 containing *pdaf-19d* (pGG21)::mCherry, Lane 7 is a negative control of Mos strain, EG5003, worm lysate and Lane 8 is JNC65 plus EG5003 worms as a heterozygous control to ensure both 2.2 kb and 3.2 kb bands could be amplified under the given PCR conditions.

same intensity through to adulthood, since the same exposure settings give rise to the same intensity of fluorescence (Figure 16). The exact cells with expression in two-fold stage embryos is not easily seen, however it is clear that there is expression in the hypodermis and expression in other tissues throughout the body. This expression pattern is also in good agreement to one seen previously by our lab in a non-integrated strain carrying a *pdaf-19a/b* (same promoter as my *pdaf-19a/b* (long) construct) GFP reporter construct (Figure 17). This strain also had expression in hypodermis, pharynx and non-ciliated neurons, however, no expression was seen in the gut as in my strain. Additionally, a shorter version of the *daf-19a/b* promoter was used (see Materials & Methods for detail) and it gave rise to the same expression pattern as the long version of the promoter (Figure 18). This shows that there are no *cis*-regulatory elements missing in the short version of the *daf-19a/b* promoter.

The *daf-19d* isoform expression has also been previously characterized and I was able to confirm it using the MosSCI method. Expression of *daf-19d* is in ciliated sensory neurons, which has been proven through transcriptional reporters and through use of antibodies (Senti and Swoboda, 2008). Two versions of the *daf-19d* promoter used for transcriptional reporters, pGG20 and pGG21 (Figure 10), both drove expression in ciliated sensory neurons; however the former from mid-embryonic to hatching and the latter from mid-embryo to adult stage (Senti and Swoboda, 2008). Using MosSCI for my *pdaf-19d* mCherry reporter transgenes, that have the same promoter as these two constructs, I was able to confirm their expression in ciliated sensory neurons. With this

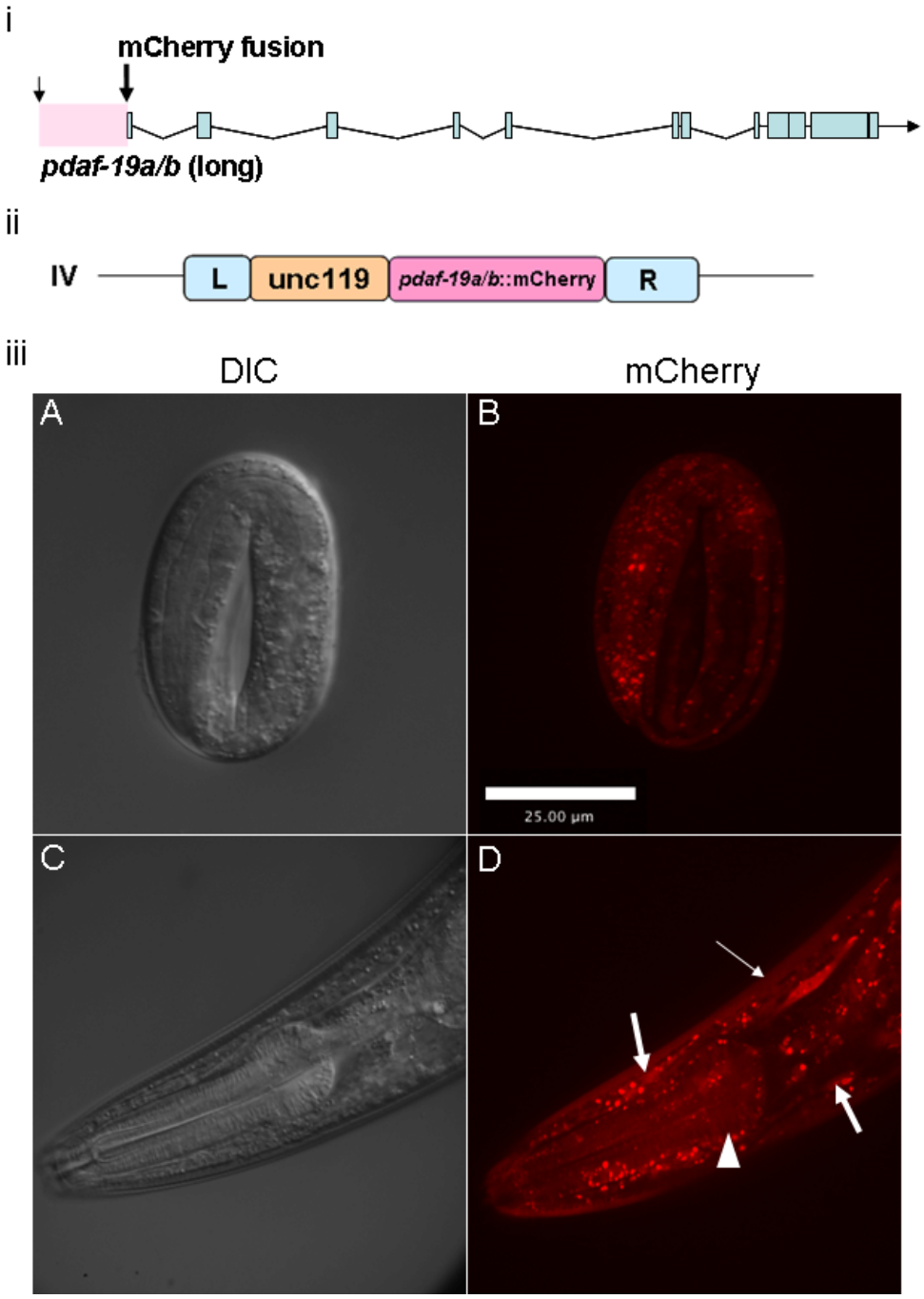


Figure 11 The long *daf-19a/b* isoform promoter drives expression in non-ciliated neurons, hypodermis, pharynx and gut from embryo to adult.

(A) DIC of 2-fold embryo transgenic worm, JNC61, that contains the *pdaf-19a/b* (long)::mCherry reporter. (B) Expression of transgenic embryo shows expression in many cells during the 2-fold embryo stage. (C) DIC of transgenic adult hermaphrodite. (D) Expression of transgenic adult hermaphrodite is found in non-ciliated neurons, hypodermis, pharynx and gut. Thin arrows indicate hypodermal cells, arrowheads indicate pharynx, and thick arrows indicate non-ciliated neurons. All exposures were taken at 3 sec. Bar represents 25 µm.

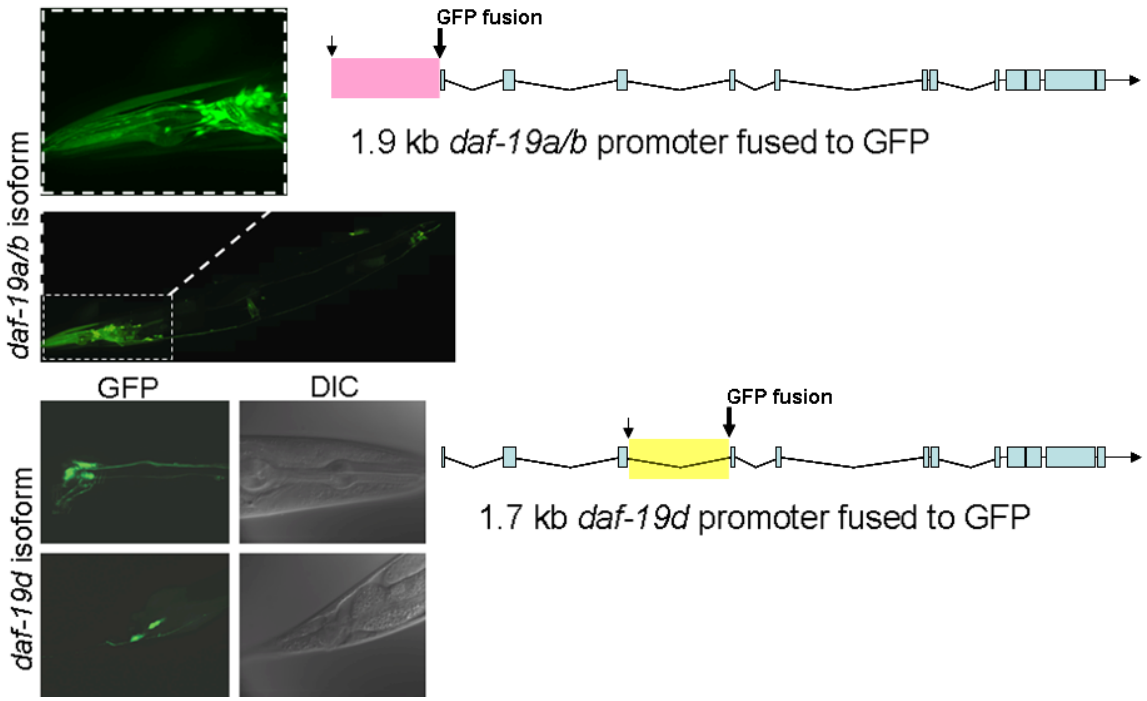


Figure 12 Reporter fusions of *daf-19a/b* promoter and *daf-19d* promoter drive differential expression.

The expression analysis and microscope images are from (Tarailo-Graovac M., unpublished) and the gene models illustrating the reporter genes made in this experiment were made by me and thus this figure is adapted from (Tarailo-Graovac M., unpublished). The *daf-19a/b* isoform promoter reporter contains a 1.9 kb sequence upstream the *daf-19* TSS, fused to GFP and is represented by the pink bar. The thin arrow indicates where the promoter starts and the thick arrow indicates where the promoter ends and the point at which it is fused to GFP. Its expression is in non-ciliated neurons, pharynx and hypoderm. The *daf-19d* isoform promoter reporter contains a 1.7 kb intron 3 sequence, fused to GFP and is expressed in ciliated neurons, including head, amphid neurons and tail, phasmid neurons. This promoter is indicated by the yellow bar.

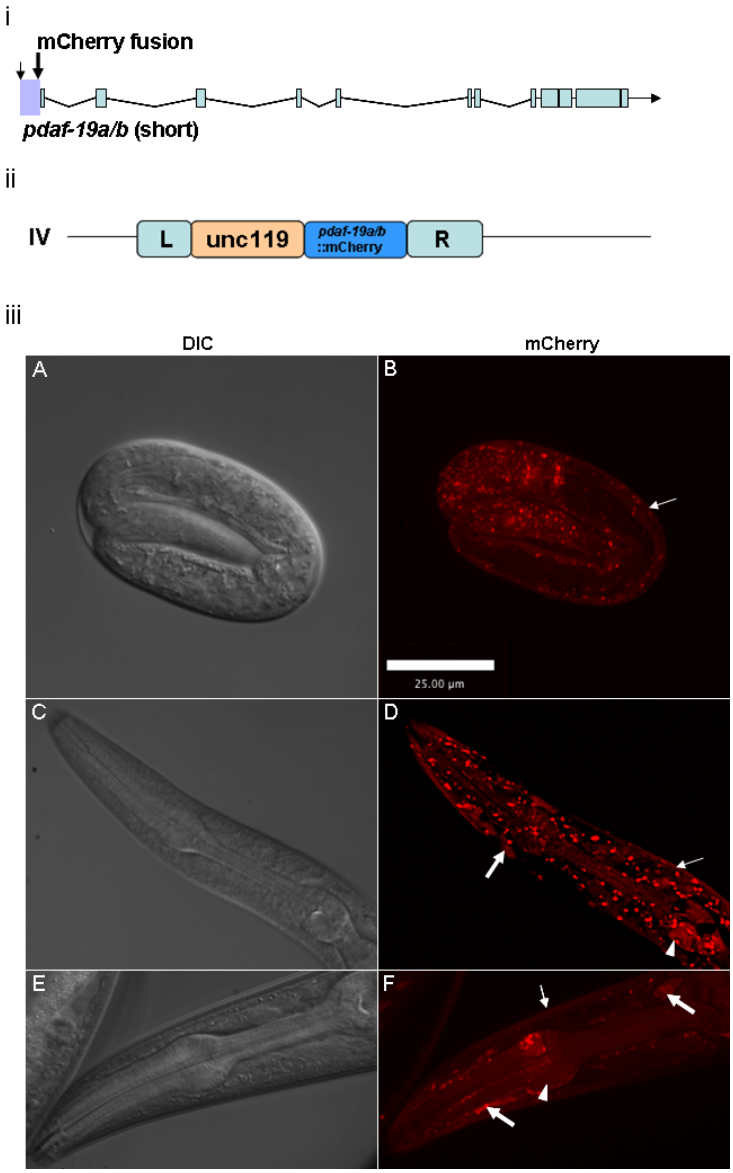


Figure 13 The short *daf-19a/b* isoform promoter drives expression in non-ciliated neurons, hypodermis, pharynx and gut from embryo to adult.

(A) DIC of 3-fold embryo transgenic worm, JNC62, that contains the *pdaf-19a/b* (short)::mCherry reporter. (B) Expression of transgenic embryo shows expression in many cells during 3-fold embryo stage. (C) DIC of transgenic larva hermaphrodite. (D) Expression of transgenic larva hermaphrodite is found in the non-ciliated neurons, hypodermis, pharynx and gut. (E) DIC of transgenic adult hermaphrodite. (F) Expression also in non-ciliated neurons, hypodermis, pharynx and gut. Thin arrows indicate hypodermal cells, arrowheads indicate pharynx, and thick arrows indicate non-ciliated neurons. All exposures were taken at 3 sec. Bar represents 25 µm.

method, I have achieved a better resolution of cells and my result showed that each of the *cis*-elements examined, controlled expression in different sets of amphid ciliated neurons. The better resolution of my study using the MosSCI method could be due to the the fact that a single copy of the transgene is present, thus reducing noise caused by overexpression of multiple copies of the transgene. Also, I observed expression from the 2-fold embryonic stage to adult stages for both constructs but at different intensities which is in contrast to the previous result. Expression of the pGG20-like construct appeared from the 2-fold embryonic to adult stage, however the expression in larva versus adults appeared to be slightly different, with larva showing some additional expression near the first pharyngeal bulb, whereas in adults expression only appeared to be in amphid neurons (Figure 19). Dye-filling was done to confirm the expression was in the amphid ciliated neurons. As seen in Figure 20, there was no overlap between the DiO staining of amphid neurons with mCherry expression. In the merged image of DiO staining with mCherry fluorescence, it is clear that the expression of mCherry is situated in between the DiO stained neurons (Figure 20). The mCherry expression corresponds well with the amphid neurons that are not able to stain with fluorescent dyes, AWA, AWB, AWC and AFD which are located right next to those that do fill with dye, ADF, ASH, ASI, ASJ, ASK and ADL (Figure 21). The *daf-19d* isoform promoter corresponding to the pGG21 construct showed expression in all amphid neurons that fill with dye and also appeared to be in those that do not fill with dye as well as in the phasmid ciliated neurons, PHA and PHB (Figure 22). These results suggest that the

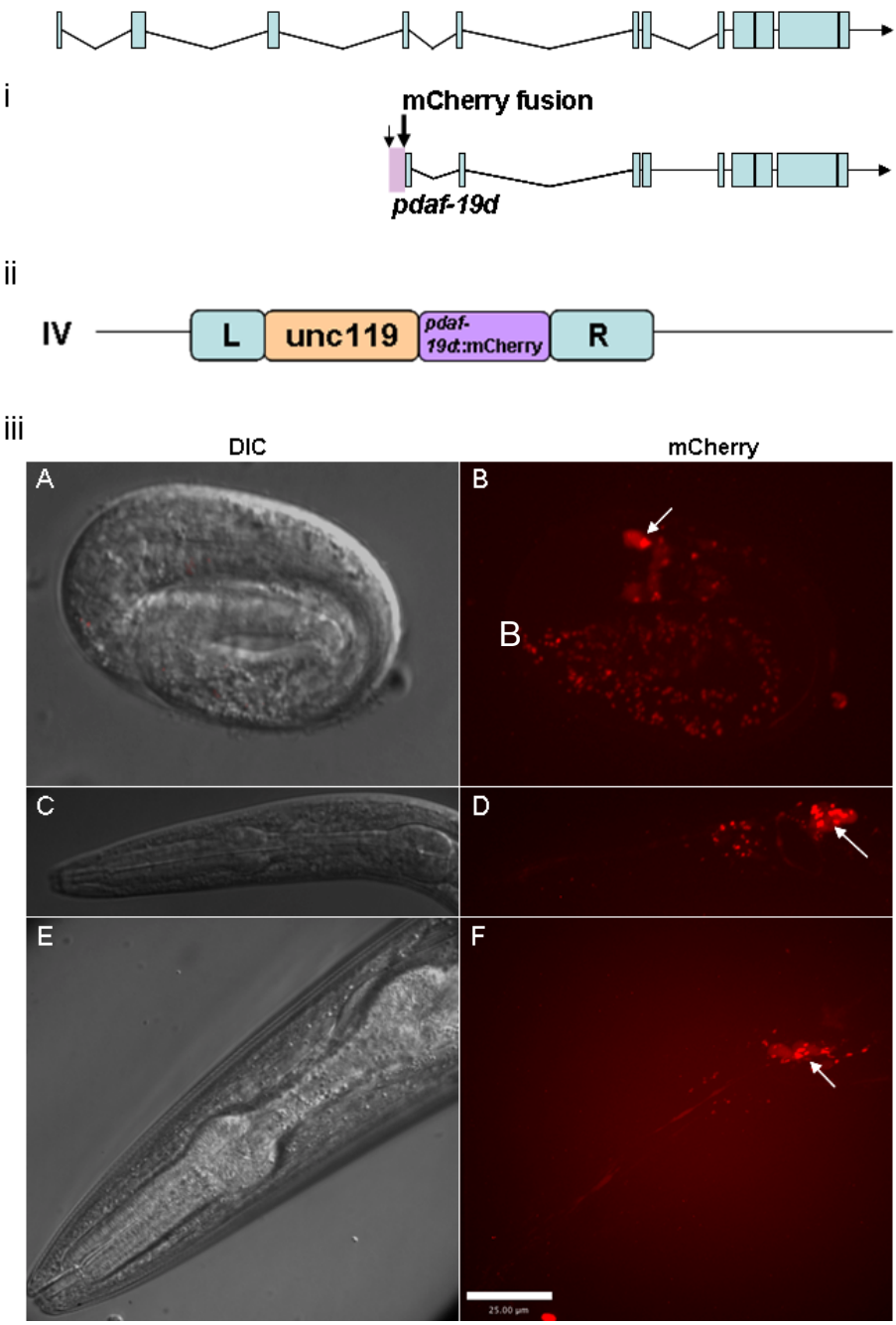


Figure 14 The pGG20 *daf-19d* isoform promoter drives expression in ciliated sensory head neurons in hermaphrodites starting at the 3-fold stage.

The gene model for the longest isoform *daf-19b* is depicted at the top of the figure as a reference to exons and introns. (A) DIC of 3-fold embryo transgenic worm, JNC63, that contains the *pdaf-19d* (pGG20)::mCherry reporter. (B) Expression of transgenic embryo with arrow pointing to ciliated head neurons. (C) DIC of transgenic larva hermaphrodite. (D) Expression of transgenic larva appears in ciliated head neurons, with arrow pointing to putative amphid neurons. (E) DIC of transgenic adult hermaphrodite. (F) Expression in transgenic adult appears to be in amphid neurons indicated by arrow. All exposures were taken at 10 sec. Bar represents 25 µm.

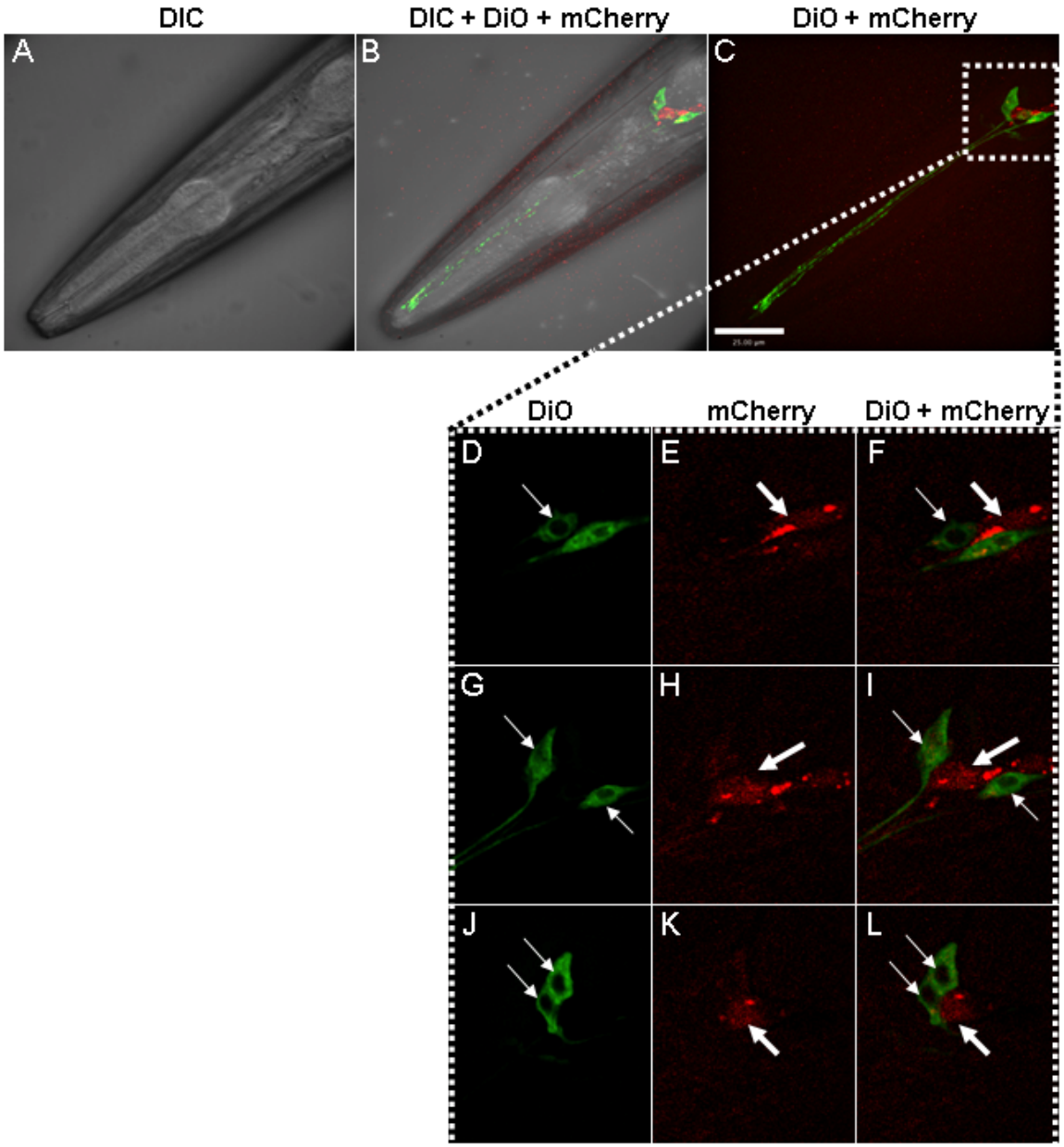


Figure 15 DiO staining reveals that the pGG20 *daf-19d* isoform promoter may drive expression in non-dye filling amphid neurons.

(A) DIC of JNC63 transgenic adult hermaphrodite (B) Merge of DIC, DiO staining and mCherry expression. (C) Merge of DiO staining and mCherry expression. (D, G, J) DiO staining of layers 1, 2 and 3 respectively. (E, H, K) mCherry expression of layers 1, 2 and 3 respectively. (F, I, L) Merge of DiO staining and mCherry expression of layers 1, 2 and 3 respectively. There is no overlap between DiO staining and mCherry expression. Thin arrows point to dye-filling amphid neurons and thick arrows point to non-dye filling amphid neurons. All exposures were taken at 2 sec. Bar represents 25 μm.

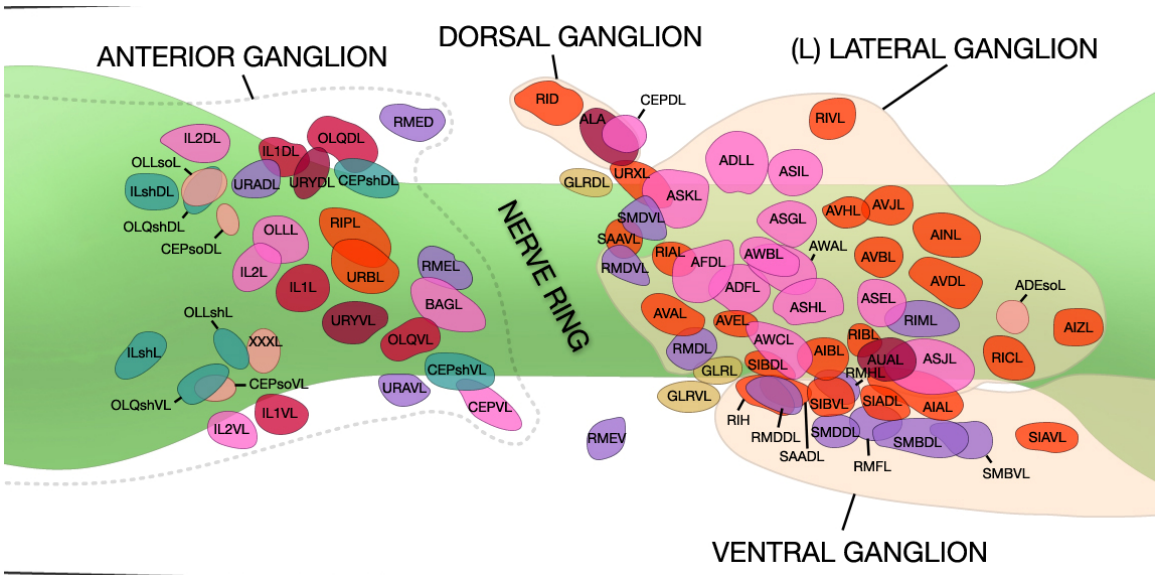


Figure 16 Amphid ciliated sensory neurons in *C. elegans.*

Figure adapted from WormAtlas (http://www.wormatlas.org/images/HeadNeurons.jpg) (Altun and Hall, 2002-2006). The pink coloured cells indicated ciliated sensory neurons.

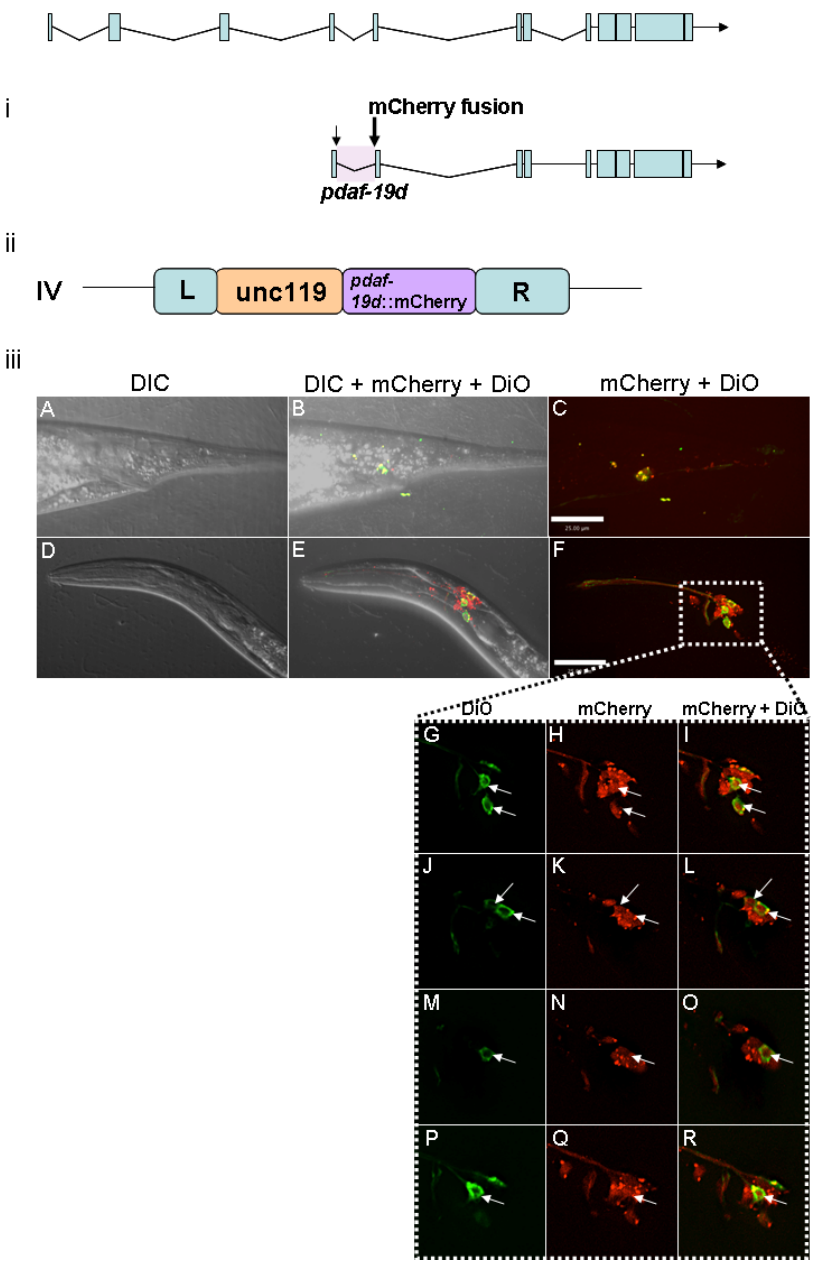


Figure 17 The pGG21 *daf-19d* isoform promoter drives expression in ciliated sensory head & tail neurons in hermaphrodites.

The gene model for the longest isoform *daf-19b* is depicted at the top of the figure as a reference to exons and introns. (A) & (D) DIC of adult transgenic worm, JNC65, that contains the *pdaf-19d* (pGG21)::mCherry reporter. (B) & (E) Merge of DIC, DiO staining and mCherry expression. (C) & (F) Merge of DiO staining and mCherry expression. (G, J, M, P) DiO staining of layers 1, 2, 3 and 4 respectively. (H, K, N, Q) mCherry expression of layers 1, 2, 3 and 4 respectively. (I, L, O, R) Merge of DiO staining and mCherry expression of layers 1, 2, 3 and 4 respectively. Arrows point to dye-filling amphid neurons. Overlap between DiO staining and mCherry expression is seen in the amphid neurons that fill with dye. All exposures were taken at 1 sec. Bar represents 25 µm.

expression of *daf-19d* in ciliated neurons is regulated with modular promoter elements. Modular promoters of other genes in *C. elegans* have been observed previously (McGhee and Krause, 1997), for instance the *unc-54* gene also has a predicted enhancer element within its third intron (Okkema, Harrison et al. 1993).

In summary, MosSCI proved to be useful for generating stable *C. elegans* strains whose expression agreed well with previous reports. These results established that this method is a robust tool for studying expression with reporters in *C. elegans*.

2.3.3 Using MosSCI to observe *daf-19c* isoform expression

The expression of the *daf-19c* isoform has not been reported, although it is referred to as male-specific in the current version of WormBase (Harris *et al.*, 2010). In contrast to this annotation, my results show that *daf-19c* is expressed with unique patterns in hermaphrodites, from early embryos through to adults. In worms that carry the transcriptional construct shown in Figure 23, expression appears as early as the 2-fold stage and maintains the same intensity and localization throughout the larval stages and into adulthood (Figure 23). Expression in hermaphrodites is seen specifically in head neurons that exist as three pairs with a six-fold symmetry located near the first pharyngeal bulb. This pattern is reminiscent to the location of the inner and outer labial ciliated neurons (Figure 24). To validate that the expression was in the labial neurons, dye-filling was done to see whether the expression would overlap with the DiO staining of the IL2 neurons. In dye-filling, the ciliated sensory neurons can fill with dye via their ciliated endings and allow for visualization of ciliated bodies and their extensions (Hedgecock, Culotti et al. 1985; Perkins, Hedgecock et al. 1986). However, only ciliated

neurons whose cilia extend to the external environment (Figure 25) are able to dye-fill. As a result, DiO fills 6 of the 12 pairs of ciliated amphid neurons, ADF, ASH, ASI, ASJ, ASK and ADL, two pairs of ciliated tail neurons, PHA and PHB (Hedgecock *et al.*, 1985) and in the presence of calcium acetate, can stain IL2 ciliated labial neurons (Burket *et al.*, 2006). There was a clear overlap in the dye-filling of labial neurons with the mCherry expression and thus it could be concluded that the expression of *daf-19c* is indeed in the IL2 labial neurons (Figure 26). The fact that another promoter, *pdaf-19c,* encodes for expression of an isoform in a different set of ciliated neurons from the amphids seen by *daf-19d* expression, suggests a modularity in the promoters driving expression in ciliated sensory neurons.

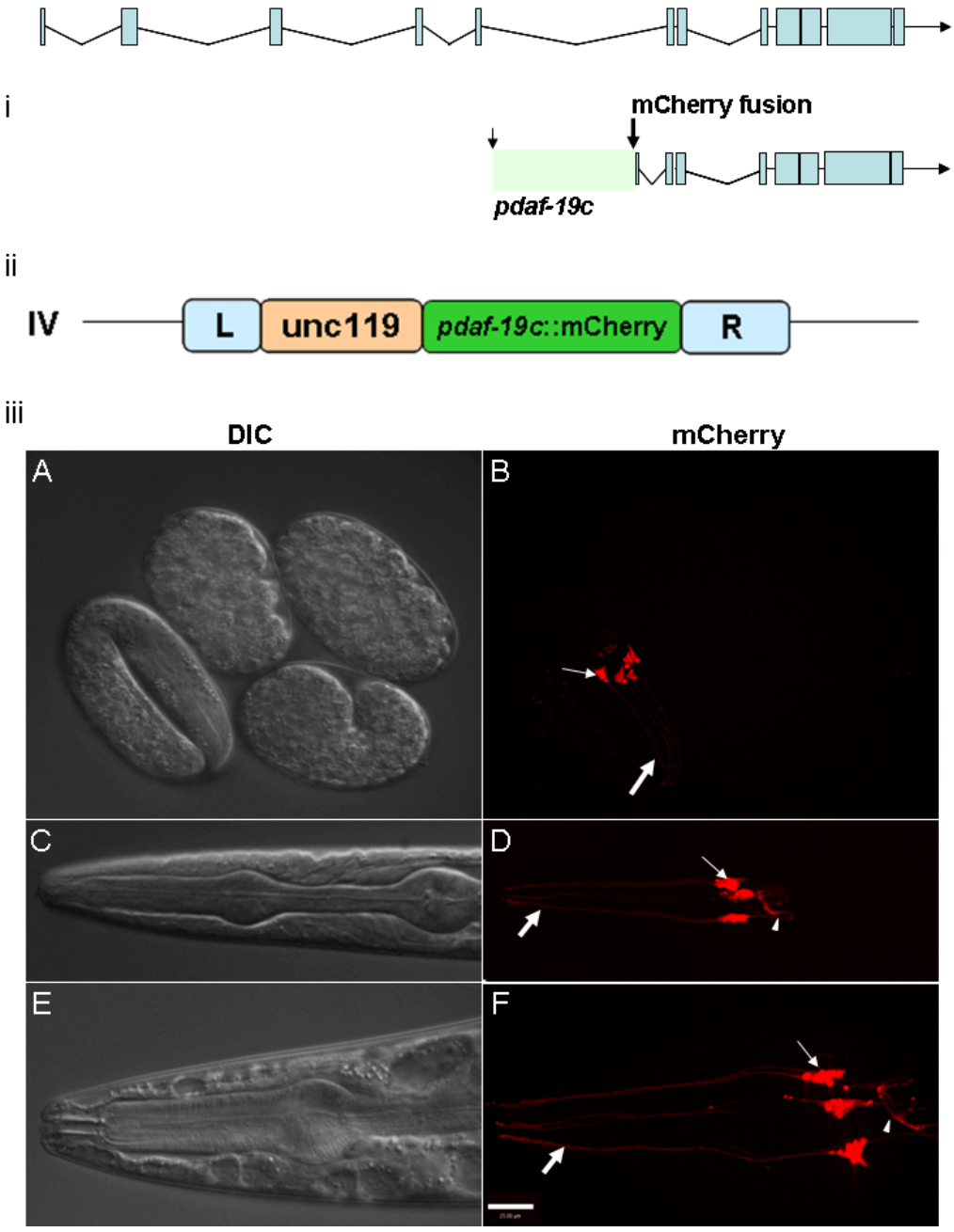


Figure 18 The *daf-19c* isoform promoter drives expression in labial neurons in hermaphrodites starting at the 2-fold stage.

The gene model for the longest isoform *daf-19b* is depicted at the top of the figure as a reference to exons and introns. (A) DIC of transgenic embryos, JNC60, that contain the *pdaf-19c*::mCherry reporter. (B) Expression of transgenic embryos shows that expression starts at the 2-fold stage. (C) DIC of transgenic larva hermaphrodite. (D) Expression of transgenic larva appears in labial neurons. (E) DIC of transgenic adult hermaphrodite. (F) Expression in transgenic adults also appears in labial neurons. Thin arrows point to labial neurons, arrowheads point to axons of these neurons and thick arrows point to dendrites of these neurons. All exposures were taken at 1 sec. Bar represents 25 μm.

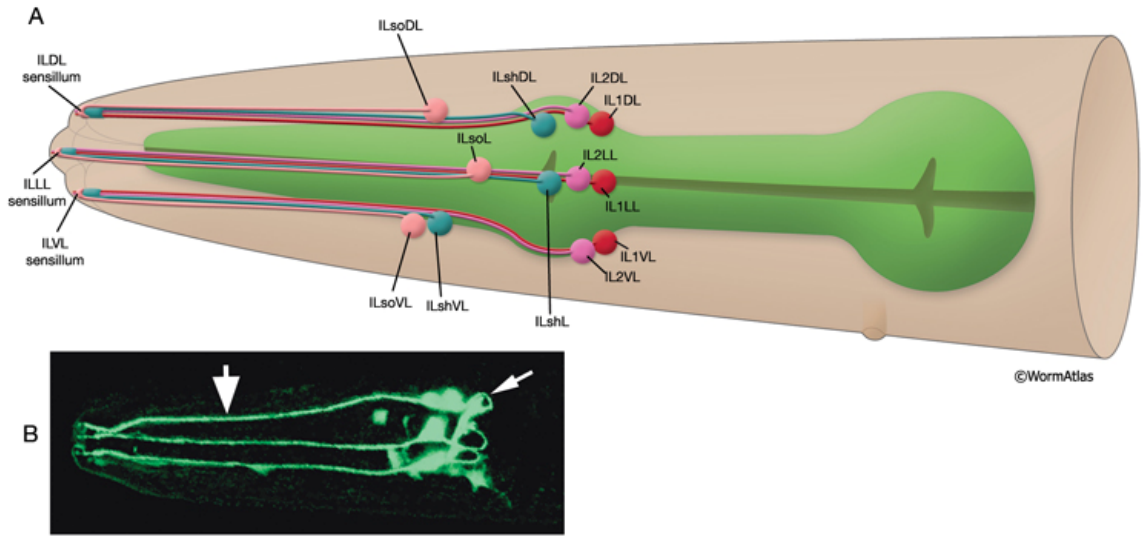


Figure 19 Schematic representation of the labial ciliated neurons in *C. elegans*.

(A) Illustration of the anatomic position of the inner labial neurons. (B) Epifluorescent image of IL1 labial neurons from a transgenic animal with the reporter gene, *Y111B2A.8::GFP*. (Strain source: The Genome BC *C. elegans* gene expression consortium (McKay et al., 2004)). Thick arrows point to dendrite of neuron and thin arrows indicate the axon of neurons. Bar represents 20 µm. (Image source: L. Ryder). Figure adapted from WormAtlas (http://www.wormatlas.org/hermaphrodite/neuronalsupport/Neurosupportframeset.html) (Altun and Hall, 2002-2006).

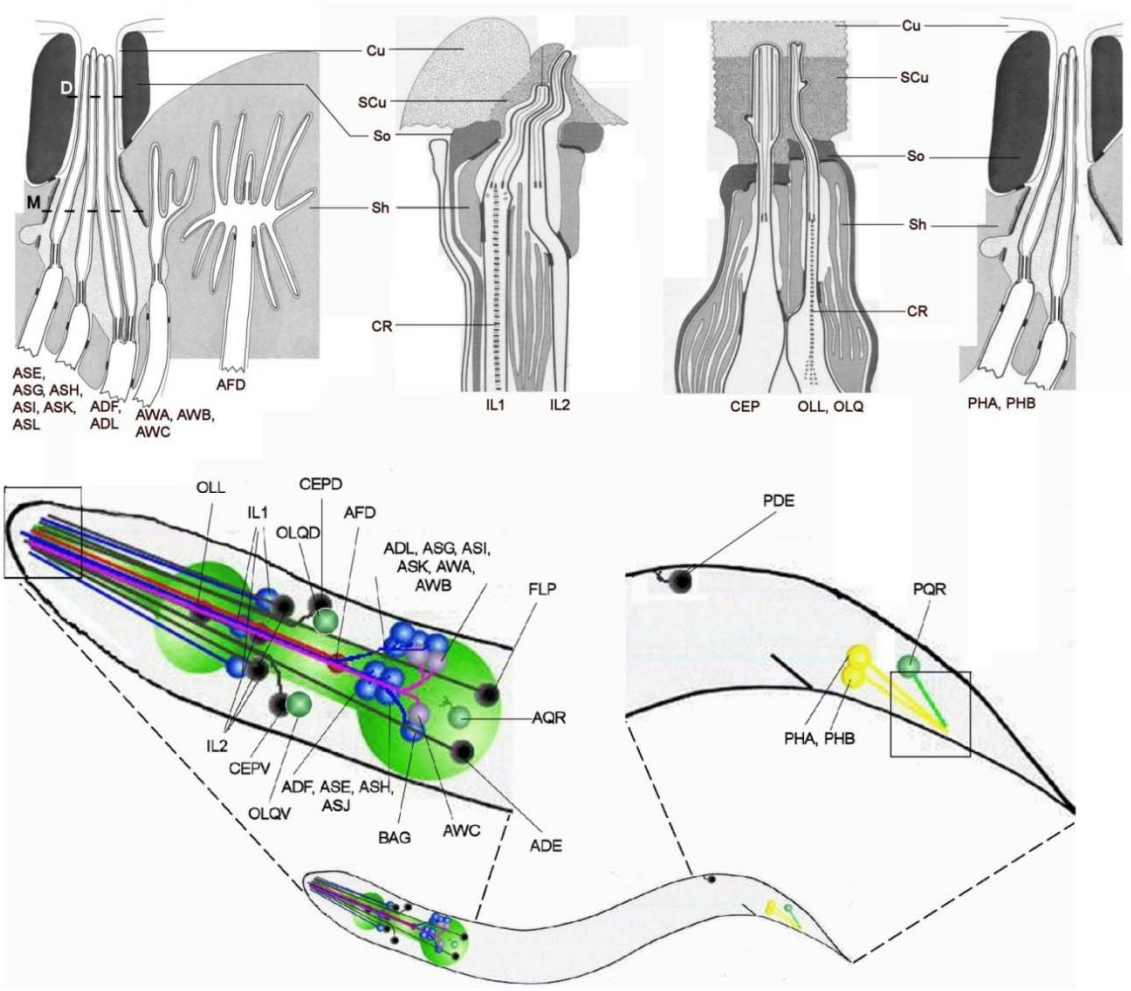


Figure 20 Structures and relative positions for sensory ciliated neurons in *C. elegans.*

The four insets shown at the top are schematic diagrams of electron micrograph images of ciliated neuron endings (figure adapted from (Inglis *et al.*, 2006)). Cu, cuticle; CR, ciliary rootlet; SCu, subcuticle; So, socket cell; Sh, sheath cell. ASE, ASG, ASI, ASL, ASH, ASK, ADF, ADL, AWA, AWB AWC and AFD are amphid ciliated neurons; IL1, IL2, OLL, OLQ are labial ciliated neurons; CEP are cephalic ciliated neurons; PHA, PHB, PQR are phasmid ciliated neurons.

2.3.4 Translational fusion for functional analysis of *daf-19c* using MosSCI

Since the relationship of expression with function of a gene has been demonstrated previously (Senti and Swoboda, 2008) and the function of the *daf-19c* isoform has not yet been revealed, it was my goal to determine whether its labial specific expression corresponded to a role in cilia development or function. A translational fusion was made to allow for probing into the function of this isoform (Figure 27). The steps taken with the transcriptional fusion were followed and after confirming successful insertion of the transgene in the pCFJ178 vector (Figure 28), the construct was injected into worms which were screened and confirmed for direct insertion (Figure 29). The expression of translational fusions also allow for detection of subcellular localization. Because DAF-19 is a TF, its subcellular location is in the nucleus. Based on the labial-specific expression of the *daf-19c* promoter I chose, I expected the *daf-19c* translational reporter to be expressed in the nucleus of labial neurons. To my surprise, this was not the case. The reporter actually expressed in the nucleus of all ciliated sensory neurons very similar to the *daf-19* translational fusion (Swoboda *et al.*, 2000) expression (Figure 30 & Figure 9). The expression in labial and amphid neurons was confirmed with dye-filling and expression in a tail neuron that is not a ciliated phasmid neuron was also observed (Figure 31).

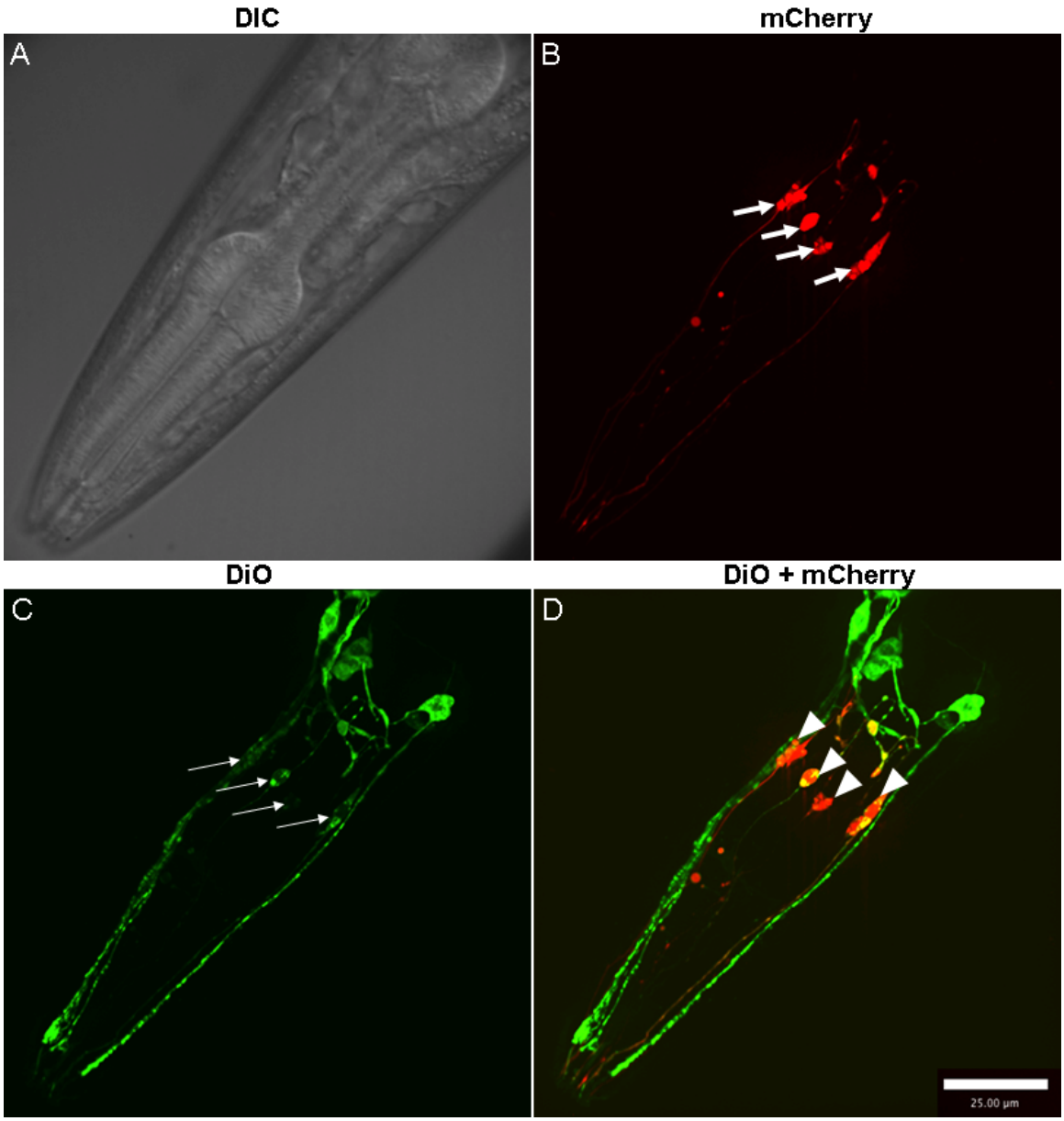


Figure 21 DiO staining confirms that the *daf-19c* isoform promoter drives expression in labial neurons in hermaphrodites.

(A) DIC of transgenic adult hermaphrodite, JNC60, that contains the *pdaf-19c*::mCherry reporter. (B) mCherry expression in putative labial neurons, indicated by thick arrows. (C) DiO staining, where thin arrows indicate labial neurons. (D) Merge of DiO staining and mCherry expression where the areas of overlap between DiO staining and mCherry expression are confirmed labial neurons and are indicated by arrowheads. All exposures were taken at 1 sec. Bar represents 25 µm.

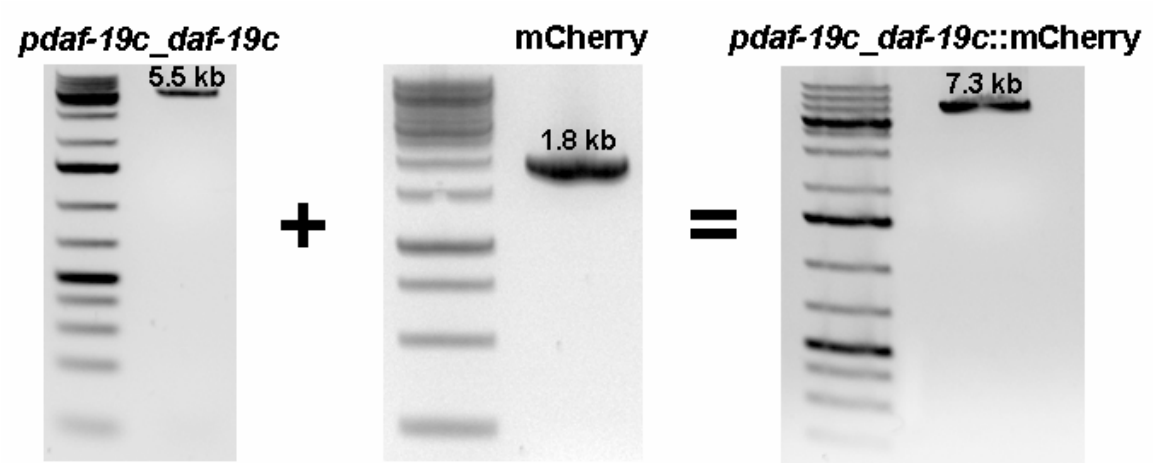


Figure 22 Translational fusion PCR of genomic *daf-19c* with mCherry.

The first gel image shows the 5.5 kb amplification product of the *daf-19c* isoform including the 2.0 kb promoter, used in the *pdaf-19c*::mCherry reporter, and the entire rest of the *daf-19* gene. This promoter sequence continuous with the gene sequence is referred to as *pdaf-19c_daf-19c.* The second gel image shows the band amplified corresponding to mCherry from the pCFJ90 vector. The + sign between these two images indicates a PCR stitching step. The third gel image shows the product of the PCR stitching, *pdaf-19c_daf-19c* stitched in front of the mCherry gene giving the final product, *pdaf-19c_daf-19c*::mCherry)

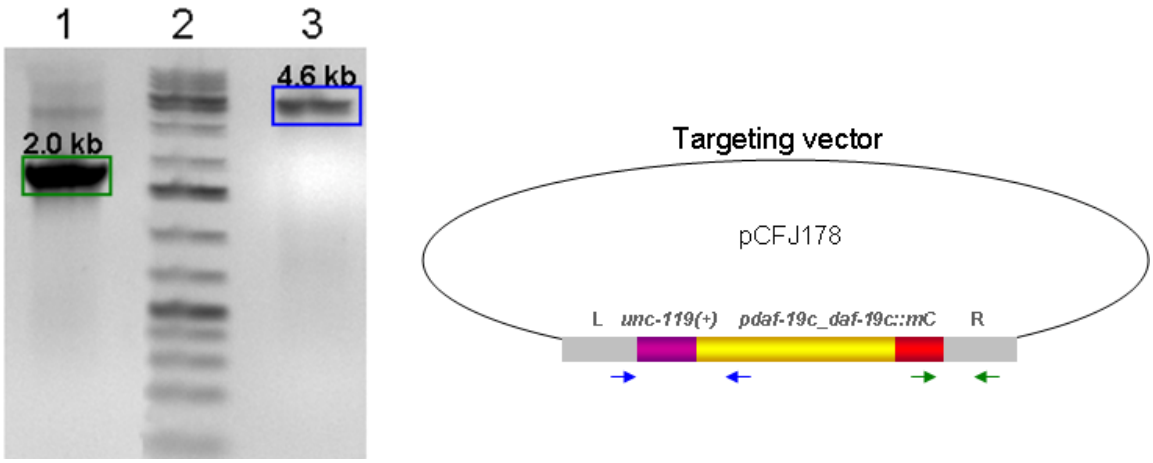


Figure 23 Confirmation of successful cloning of translational *daf-19c* reporter gene into the pCFJ178 targeting vector.

Lane 1 is amplification of a single colony using primers to amplify the right junction site using primers (green arrows) specific to mCherry sequence and the pCFJ178 right homology arm sequence. The expected 2.0 kb band was amplified. Lane 2 is Fermentas GeneRuler 1 kb Ladder Plus. Lane 3 is the same single colony as template, but with primers (blue arrows) specific to the pCJF178 left homology arm sequence and the *daf-19c* promoter sequence. The expected 4.6 kb band was amplified.

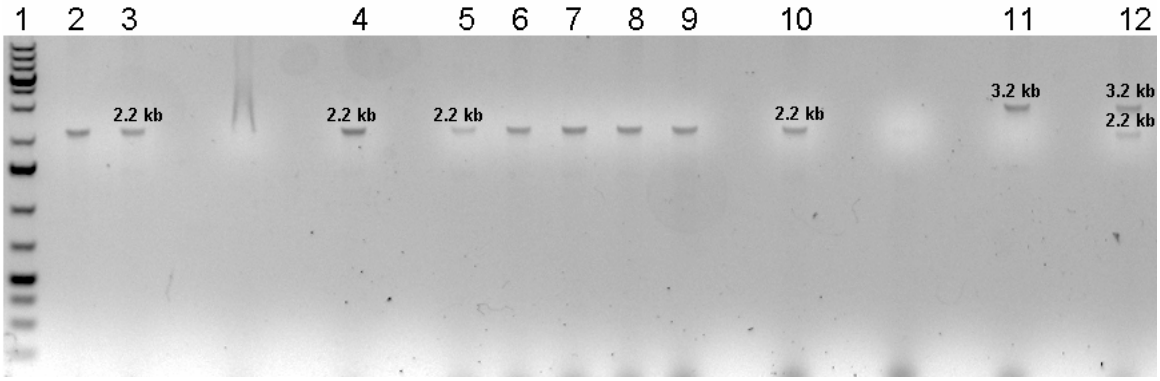


Figure 24 PCR genotyping results for direct insertion of *pdaf-19c_daf-19c*::mCherry into EG5003 Mos strain.

Lane 1 is Fermentas GeneRuler 1 kb Ladder Plus. Lane 2-10 are lysed worm candidates for direct insertion. Lane 11 is EG5003 as a negative control; the expected 3.2 kb band is present. Lane 12 is a heterozygous control of lysed EG5003 with lysed positive control for direct insertion (JNC61); the expected 3.2 kb and 2.2 kb bands are present. Lanes 2, 3, 4, 5-9 and 10 all contain worms with a direct insertion.

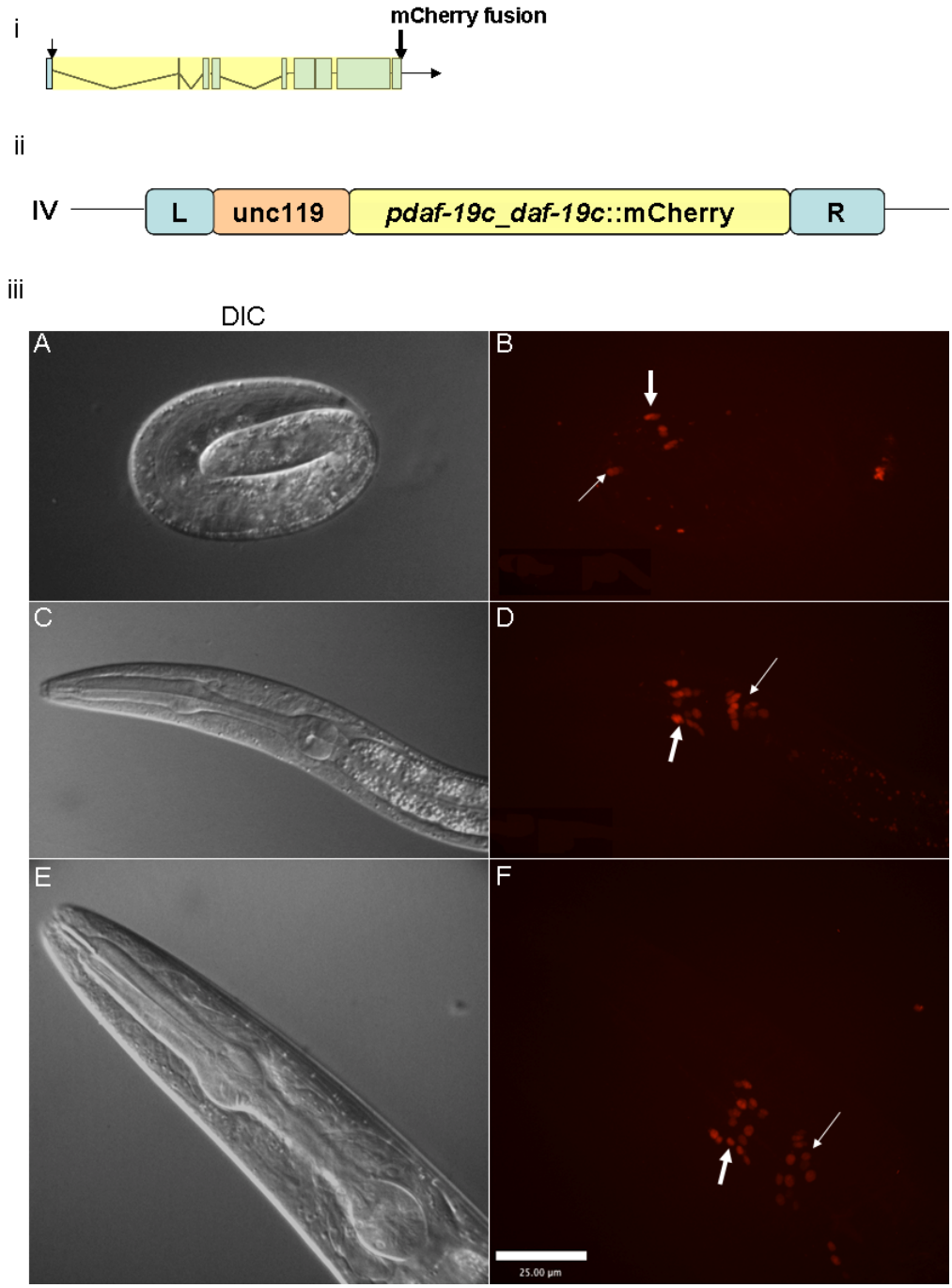


Figure 25 Translational reporter of the *daf-19c* isoform is expressed in all ciliated sensory head neurons in hermaphrodites starting at the 3-fold stage.

(A) DIC of transgenic, JNC66, 3-fold embryo containing the *pdaf-19c_daf-19c*::mCherry transgene. (B) Expression of mCherry in 3-fold embryo. (C) DIC of transgenic larva hermaphrodite worm. (D) Expression of mCherry in transgenic larva. (E) DIC of transgenic adult hermaphrodite transgenic worm. (F) Expression of mCherry in transgenic adult. Thin arrows point to putative amphid neurons and thick arrows point to putative labial neurons. All exposures were taken at a 1 sec. Bar represents 25 µm.

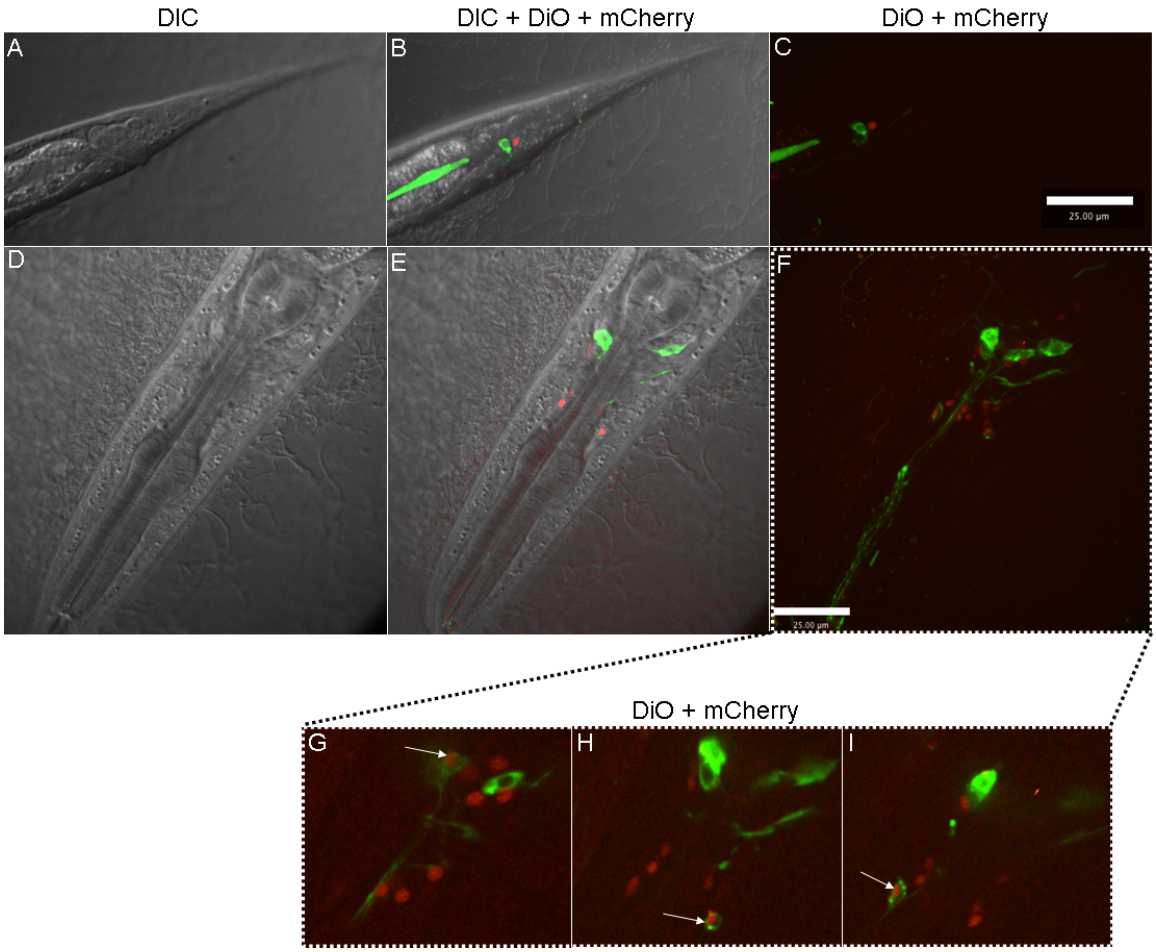


Figure 26 DiO staining shows that the *daf-19c* isoform translational reporter is expressed in both amphid and labial ciliated sensory head neurons in hermaphrodites.

(A) & (D) DIC of transgenic, JNC66, adult hermaphrodite containing the *pdaf-19c_daf-19c*::mCherry transgene. (B) & (E) DIC merge with DiO staining and mCherry expression. (C) Merge of DiO staining and mCherry, showing that there is no overlap of mCherry expression with DiO in the tail. (F) Merge of DiO staining and mCherry expression, showing overlap of DiO staining and mCherry expression in both labial and amphid neurons. (G-I) Amplified image of three different layers of DiO staining and mCherry expression; showing the overlap of DiO staining with mCherry in the labial and amphid neurons more clearly. Arrows indicate the amphid and labial neurons that have overlap with DiO and mCherry fluorescence. All exposures of mCherry were taken at 1 sec. Bar represents 25 µm.

This result suggests that the sequence I chose to represent the *daf-19c* isoform promoter may not be a completely representative of the endogenous promoter. This result also confirms the well known problem of promoter selection and that many sequence elements, including exons, introns and UTRs, may contribute to the regulation of a gene.

A successful direct insertion for *pc_daf-19c*::mCherry has been obtained and is currently being crossed into the *daf-19(m86)* mutant background, after which screening for rescue of cilia development will be done using a dye-filling assay.

2.3.5 Determining sex-specific expression for *daf-19* isoforms

Since *daf-19* regulates ciliated sensory neuron development in hermaphrodites (Swoboda *et al.*, 2000), it is likely that it also regulates the male-specific ciliated sensory neurons. The male nervous system of *C. elegans* has 473 cells, 79 of which are male-specific neurons and 36 of which are support cells. Most of these additional neurons are located in the male tail (Peden and Barr, 2005) and are involved in mating behaviour (Sulston *et al.*, 1980). The expression of the different *daf-19* isoforms has not yet been determined in males, however the expression of the whole *daf-19* gene, corresponding to translational *daf-19*::GFP fusion transgene that was made previously (Swoboda *et al.*, 2000), has been observed in male-specific ciliated sensory neurons (Yu *et al.*, 2003). Thus it was of interest to determine the expression profile of each of the *daf-19* isoforms in males. This could shed some light on the potentially differing functions of the *daf-19* isoforms in males.

The expression of *pdaf-19a/b* in males appeared to be in as many tissues as in hermaphrodites, appearing in the hypodermis, pharynx and non-ciliated neurons.

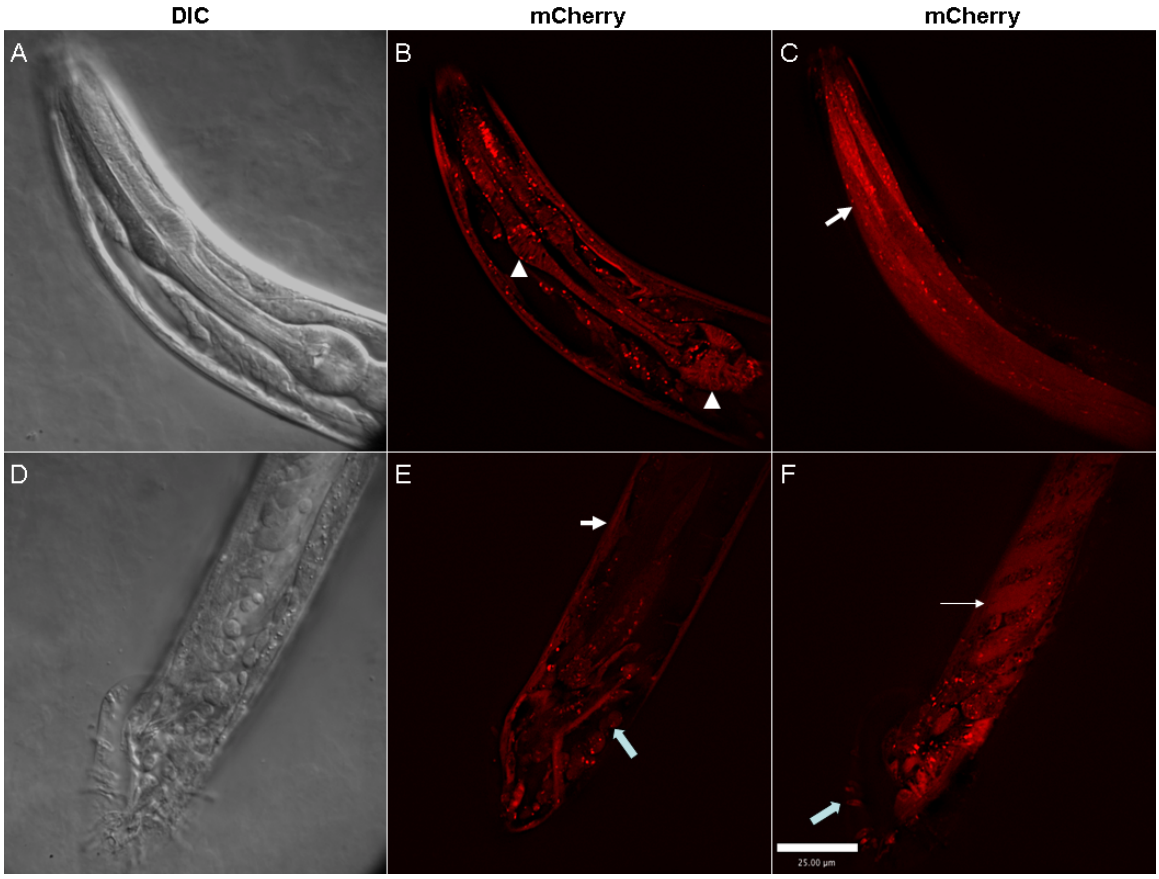


Figure 27 The long *daf-19a/b* isoform promoter in males drives expression in non-ciliated neurons, hypodermis, pharynx, male mesoderm and male specific neurons.

(A) & (D) DIC of transgenic, JNC61, male containing the long *pdaf-19a/b*::mCherry reporter. (B) Expression of mCherry in transgenic male appears in the pharynx (indicated by arrowheads) and non-ciliated neurons. (C) Expression of mCherry in layer 2 of the head is in the hypodermis, indicated by the thick arrow. (D) DIC of transgenic adult male tail. (E) Expression of mCherry in the transgenic adult male tail is found in hypoderm (indicated by thick arrow) and male specific neurons (indicated by block arrow). (F) Layer 2 of the male tail shows expression in the male mesoderm (indicated by thin arrow) and male specific neurons (indicated by block arrow). All exposures were taken at 3 sec. Bar represents 25 µm.

Additionally, there was distinct expression in male specific muscles in the tail region (Figure 32). Expression also occurs in what appear to be male neuronal support cells. As in hermaphrodites, the expression of *pdaf-19a/b* may suggest that the promoter chosen is not the complete *cis*-regulatory element of this isoform. It is possible that the downstream regions from exon 1-4 may still play roles as *cis*-regulatory elements for the *daf-19a/b* isoforms. The males containing the *pdaf-19c* reporter construct had a different expression pattern in the head from that seen in hermaphrodites. There are two major differences of this isoforms expression in males. First, expression in males was detected in both the labial and some neurons that were near the amphids (Figure 33) versus just the labial neuron expression seen in hermaphrodites. Second, expression appeared in male neuronal support cells in the tail, however the exact cell identity could not be determined. This is in contrast to hermaphrodites which did not have any expression detected in any other cells apart from the labial neurons. The expression in labial neurons and amphid neurons in males was confirmed by dye-filling as was done for hermaphrodites. The dye-filling confirmed the expression in labial neurons, however, the amphid neurons could not be confirmed (Figure 34). Given the position of the putative amphid neurons showing expression in males, it is most likely that these are one of the amphid neurons that are not capable of filling with fluorescent dyes. To obtain a more complete picture of where *pdaf-19d* is expressed, the expression was also observed in males. Intriguingly, the expression in males is also slightly different from that in hermaphrodites. It appears as though this promoter drives expression in more tissues in males, including expression in amphid neurons, as in hermaphrodites, as well as distinct

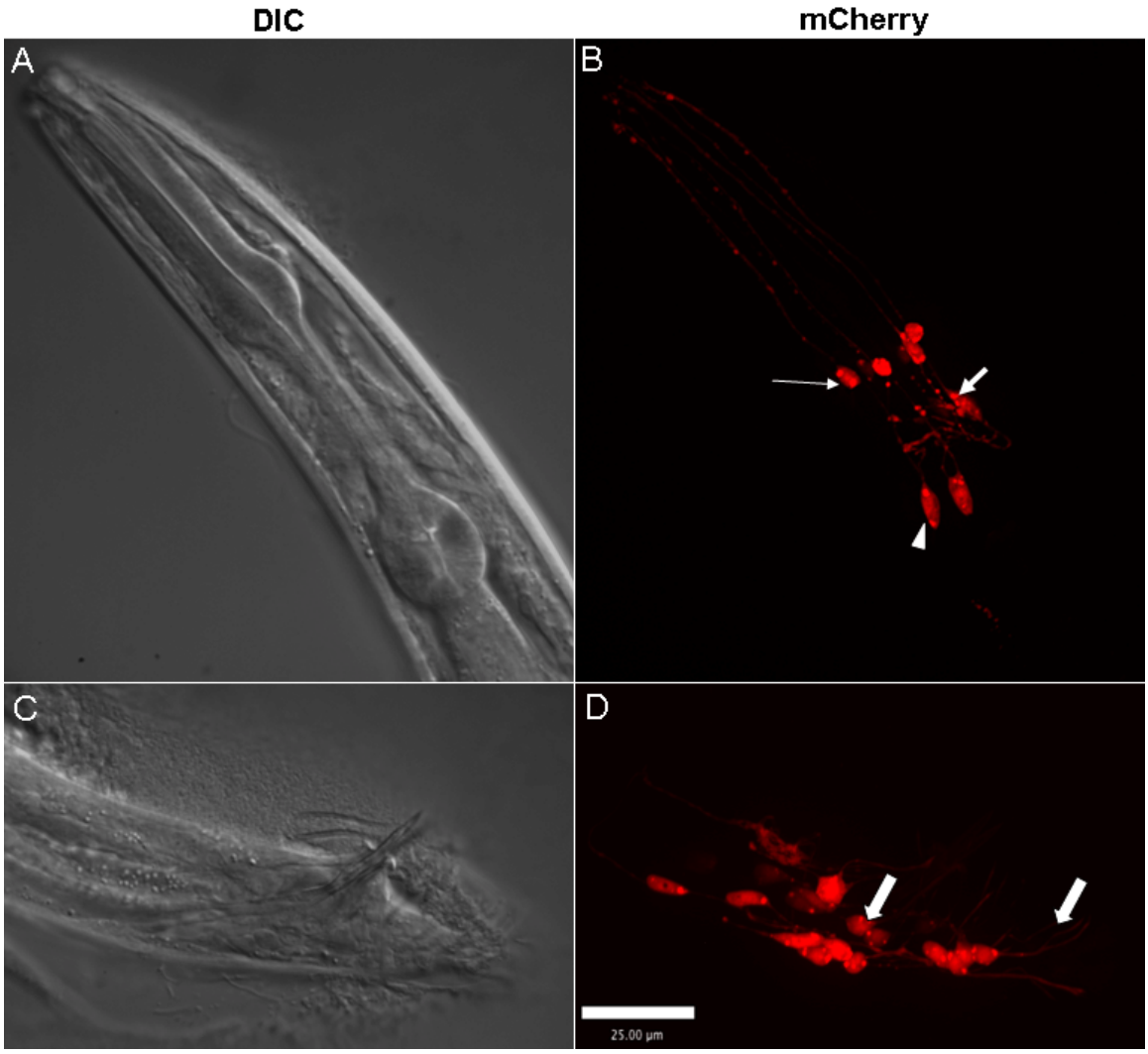


Figure 28 The *daf-19c* isoform promoter drives expression in labial and amphid neurons in males and in male neuronal support cells.

(A) DIC of transgenic, JNC60, adult male containing the *pdaf-19c*::mCherry reporter. (B) Expression of mCherry in the transgenic male appears to be in labial and amphid neurons. Putative labial neurons are indicated by thin arrow and putative amphids are indicated by arrowhead and thick arrow. (C) DIC of transgenic adult male tail. (D) Expression of mCherry in transgenic adult male tail appears to be in male neuronal support cells indicated by block arrows. All exposures were taken at 2 sec. Bar represents 25 µm.

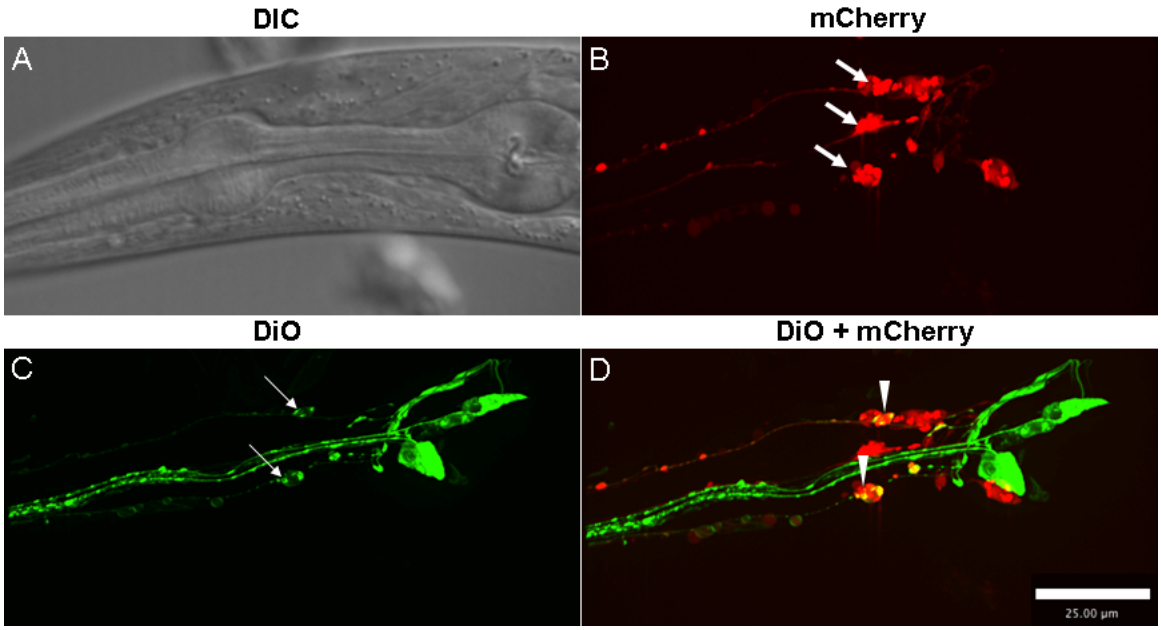


Figure 29 DiO staining confirms that the *daf-19c* isoform promoter drives expression in labial neurons in males.

(A) DIC of transgenic, JNC60, adult male containing the *pdaf-19c*::mCherry reporter. (B) Expression of mCherry in putative amphid and labial neurons. Thick arrows indicate putative labial neurons. (C) DiO staining of transgenic worm where thin arrows indicate labial neurons. (D) Merge of DiO staining and mCherry expression. Arrowheads indicate areas where DiO and mCherry overlap, which are confirmed labial neurons. All exposures were taken at 2 sec. Bar represents 25 µm.

expression in the spicules and in the fan and rays of the male tail that contain male specific ciliated sensory neurons (Figure 35).

The expression seen for the *daf-19c* isoform (Figure 33) seems to correspond best with the previously reported *daf-19::gfp* expression in males (Yu, Pretot et al. 2003), whereas, *daf-19a/b* and *daf-19d* do not seem to be represented in that expression analysis. A possible reason for the under-representation of *daf-19a/b* and *daf-19d* expression profiles in males in the Yu study could be that *daf-19c* is being expressed at higher levels than the other isoforms and thus its promoter out-competes the other isoforms for transcription. This similar reasoning can be applied to why the *daf-19* translational fusion by Swoboda (2000), which should represent all *daf-19* isoforms, predominantly shows expression patterns representative of *daf-19*c and *daf-19d* and not the non-ciliated neuron specific expression of *daf-19a/b*.

Although the precise mechanism of *daf-19* regulation is not known, the observed sex-specific expression of the *daf-19* isoforms suggests that the different *daf-19* isoform promoters recruit a male-specific set of co-activators of transcription for each of the isoforms, allowing for their observed tissue specific and sex-specific expression pattern.

These results also suggest that there is a sex-specific role of *daf-19* in males and thus a sex-specific phenotype is expected in *daf-19* mutant males. Since the various *daf-19* isoforms are expressed in male mesoderm (*daf-19a/b*), male-specific neurons (*daf-19c*) and the male spicules (*daf-19d*), the sex-specific phenotype should be related to these regions of expression. For instance, if *daf-19a/b* is involved in development or function of the male mesoderm as suggested by its expression profile, a mutant phenotype related to the male mesoderm would be expected in *daf-19(m86)* males. I observed the rays and

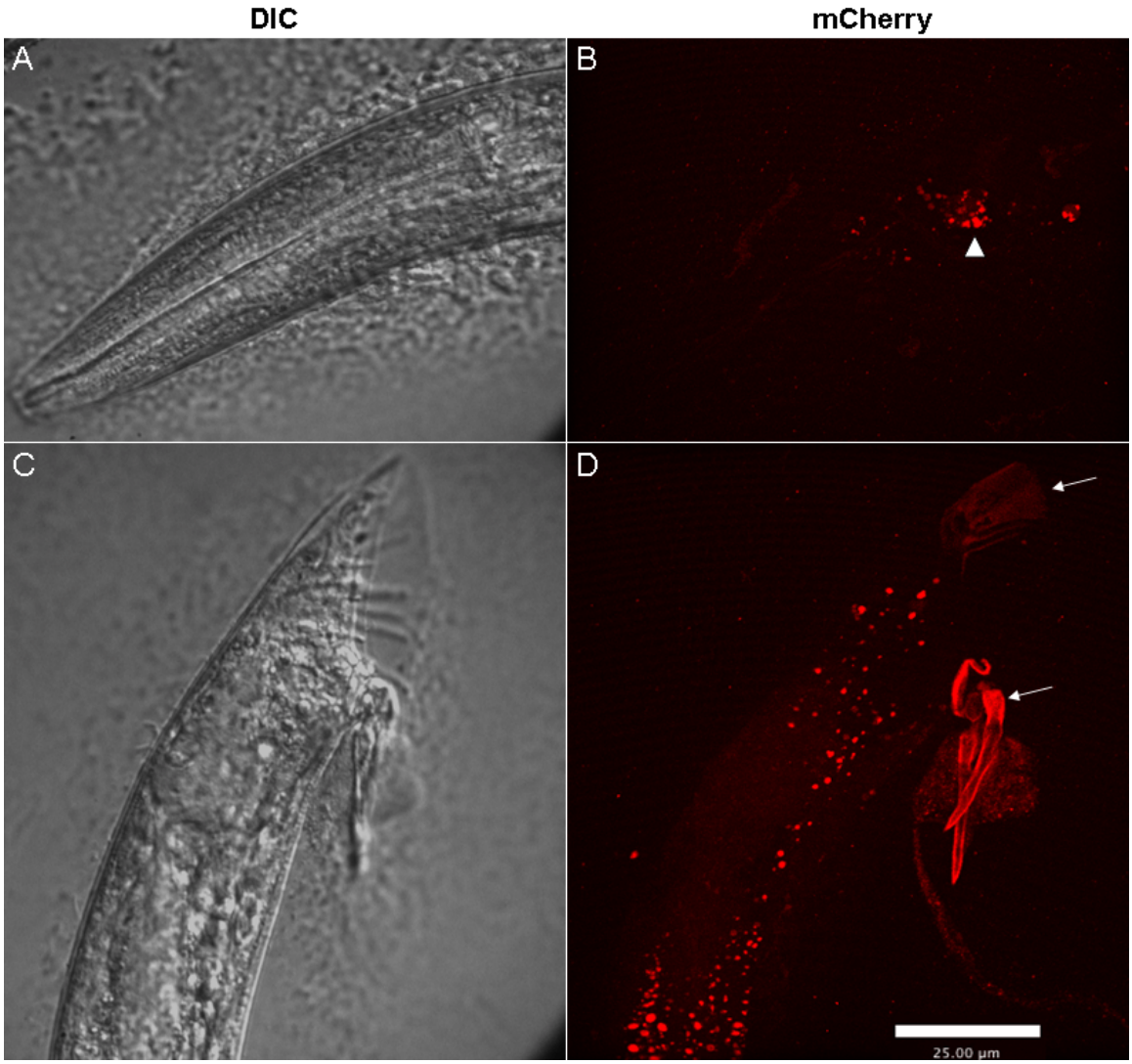


Figure 30 The *daf-19d* isoform promoter drives expression in ciliated sensory head neurons in males and in male neuronal support cells.

(A) DIC of transgenic, JNC63, adult male containing the *pdaf-19d*::mCherry reporter. (B) Expression of mCherry in transgenic male appears to be in ciliated sensory neurons indicated by arrowhead. (C) DIC of transgenic adult male tail. (D) Expression of mCherry in transgenic adult male tail is found in the spicules, rays and fan, indicated by arrows. All exposures were taken at 2 sec. Bar represents 25 µm.

spicules of *daf-19(m86)* males, a *him-5/daf-19(m86)* strain in particular (a kind gift from Dr. M. Tarailo-Graovac), under the microscope for any obvious aberrations to these structures but did not detect any (data not shown). I looked for any abnormalities in the number of rays and from the 10 worms observed by DIC, I did not see any deviations from the expected 18 rays on a wild-type male (Sulston *et al.*, 1980). Since *daf-19a/b* appears in muscle tissues in males, an Unc phenotype would be expected at least in males but is not observed. Unc phenotypes are characterized by the inability of the worm to move and also by resistance to the pharmacological paralyzing agents, aldicarb (Brenner, 1974) and levamisole (Lewis *et al.*, 1980). Although an obvious defect in movement is not visible in *daf-19(m86)* males, there is however, a slight Unc phenotype as shown by their moderate resistance to aldicarb and levamisole (Senti and Swoboda 2008).

2.3.6 Future studies to determine regulation of the DAF-19 transcription factor

Since *daf-19c* and *daf-19d* are expressed in a sub-set of ciliated neurons, this suggests that they may be self-regulated by DAF-19d, which is known for regulating cilia specific genes (Swoboda *et al.*, 2000; Chen *et al.*, 2006). If DAF-19d did regulate its own expression of these isoforms, it would be expected that all isoform promoters contain an X-box since RFX TFs normally recognize the X-box of the genes they regulate (Emery *et al.*, 1996a). When using HMMER to build a hmmprofile of the X-box motif (done by Chu J. S., unpublished), there is no typical X-box present in the two *daf-19d* (pGG20 and pGG21) promoters (Figure 36). However, when a more relaxed consensus X-box sequence was used, based on the average consensus, (RTHNYY WT RRNRAC) by Efimenko (Efimenko *et al.*, 2005), with the allowance of a 0-10 nt spacer,

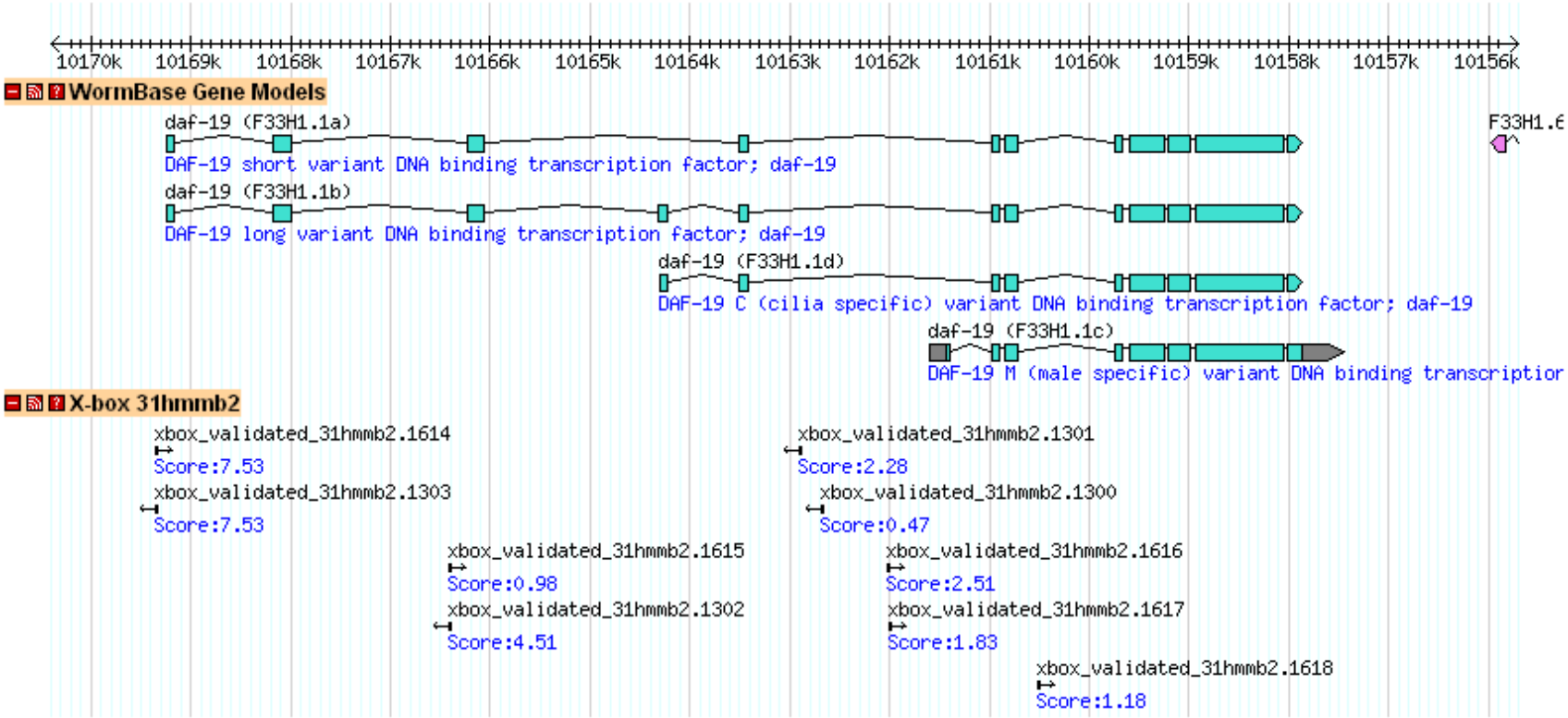


Figure 31 X-box motif predictions near *daf-19* gene (WS204 version) using HMMER.

HMMER prediction of X-box was done by Chu, J. S. (unpublished)

There were X-box motifs in these promoters (Figure 37). A good way to determine whether DAF-19 is regulating expression of each of the *daf-19* isoforms would be to cross the strains carrying the integrated transcriptional fusions of the promoters into a *daf-19* mutant background that is defective in DAF-19 and look at whether expression is still present. If expression is still observed, that would suggest DAF-19 is not regulating the expression from that promoter.

2.4 CONCLUSION

In this study, I was able to demonstrate the use of MosSCI as a successful method for generating *C. elegans* strains that allows comparative evaluation of gene expression. I successfully generated six direct insertion stable lines, which allowed me to probe the isoform specific expression of the *daf-19* gene in a comparative manner. The non-ciliated neuron expresson of *daf-19a/b* and the ciliated neuron specific expression of *daf-19d* were confirmed. Notably, the MosSCI approach offered more resolution than previous studies (Senti and Swoboda, 2008) for observing the *daf-19d* isoform expression which lead to the observation of the modular architecture of the promoter for this isoform. Additionally, the male specific and tissue specific expression patterns for each of the isoforms were identified and I was the first to discover the differential expression of *daf-19a/b* in the male mesoderm and expression of *daf-19d* in the male spicules, rays and fan. Furthermore, I discovered that the *daf-19c* isoform is also expressed in hermaphrodites, suggesting that it may not be a male-specific isoform, as currently reported in Wormbase (Harris *et al.*, 2010). More importantly, my research has revealed the presence of additional *cis*-regulatory elements in addition to the genomic sequence right upstream of the *daf-19c* isoform. This was revealed by the difference in expression

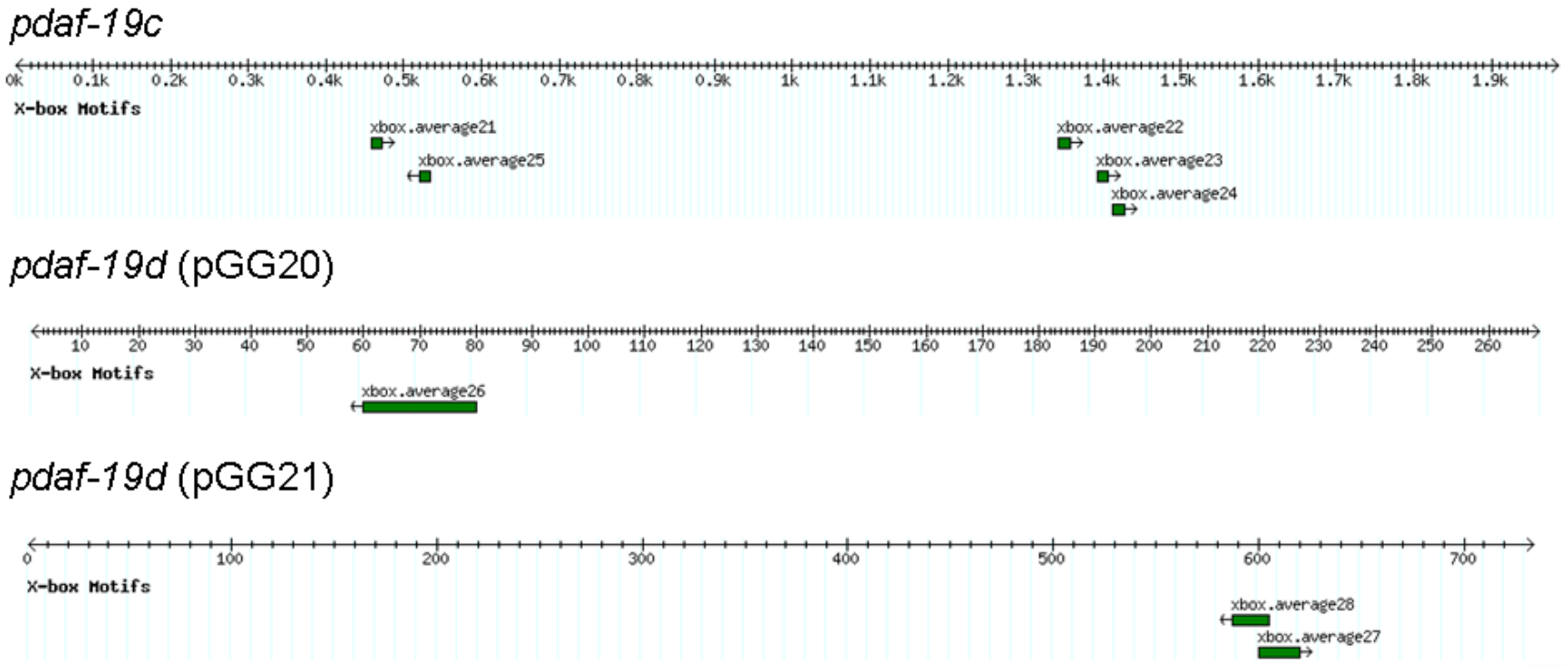


Figure 32 X-box motif predictions near to *pdaf-19c*, *pdaf-19d* (pGG20 and pGG21) using relaxed consensus X-box motif.
X-box prediction was done by Chu, J. S. (unpublished).

between the *daf-19c* transcriptional fusion and the *daf-19c* translational fusion. In the transcriptional fusion, that included the sequence upstream of the *daf-19c* TSS, the expression was limited to only labial ciliated neurons; whereas the *daf-19c* translational fusion, that included this same promoter as well as the entire downstream gene, was expressed in all ciliated neurons. This showed that the translational fusion contained extra *cis* regulatory elements that were allowing for expression in additional ciliated neurons. Thus, the *daf-19c* isoform is expressed in all ciliated neurons in hermaphrodites. This finding is very interesting because *daf-19d* was previously thought to be the sole *daf-19* isoform that is expressed in ciliated sensory neurons. It will be interesting to determine the function for the *daf-19c* isoform through rescue experiments that are currently underway in Dr. Chen's laboratory. It is expected that *daf-19c* will have similar function to *daf-19d* in the development and function of cilia since it is also expressed in ciliated sensory neurons.

Chapter 3: Bioinformatics identification of RFX transcription factors.

Note Regarding Contributions:

The following chapter is from the following publication:

Aftab, S., **L. Semenec**, J. S. Chu, N. Chen (2008). "Identification and characterization of novel human tissue-specific RFX transcription factors." BMC Evol Biol **8**: 226.

Chen N. conceived of the study. Aftab S. drafted the manuscript and the rest of the authors and I contributed to the writing of the manuscript. Aftab S. generated all tables in the manuscript. Aftab S. generated Figure 1 which involved the RFX DBD alignment. Chu J. S. and I constructed Figure 2 of the regulatory domain conservation of the different RFX family members. This involved making a script that made HMM profiles for the AD and DIM domains and searching these in RFX6 and RFX7. The coordinates for each domain were parsed and sorted into a table from which J. S. Chu constructed a figure, to scale, of the relative positions of the domains within the protein. Aftab S. and I generated Figure 3 of the protein interactions of RFX6. Aftab S. generated Figure 4 which was a phylogenetic analysis of the mammalian RFX1-7. Aftab S. and I generated Figure 5 which involved searching the mouse Atlas SAGE data for RFX expression.

3.1 ABSTRACT

Five regulatory factor X (RFX) transcription factors (TFs)–RFX1-5–have been previously characterized in the human genome, which have been demonstrated to be critical for development and are associated with an expanding list of serious human disease conditions including major histocompatibility (MHC) class II deficiency and ciliopathies. In this study, we have identified two additional RFX genes–RFX6 and RFX7–in the current human genome sequences. Both RFX6 and RFX7 are demonstrated to be winged-helix TFs and have well conserved RFX DNA binding domains (DBDs),

which are also found in winged-helix TFs RFX1-5. Phylogenetic analysis suggests that the RFX family in the human genome has undergone at least three gene duplications in evolution and the seven human RFX genes can be clearly categorized into three subgroups: (1) RFX1-3, (2) RFX4 and RFX6, and (3) RFX5 and RFX7. Our functional genomics analysis suggests that RFX6 and RFX7 have distinct expression profiles. RFX6 is expressed almost exclusively in the pancreatic islets, while RFX7 has high ubiquitous expression in nearly all tissues examined, particularly in various brain tissues. The identification and further characterization of these two novel RFX genes hold promise for gaining critical insight into development and many disease conditions in mammals, potentially leading to identification of disease genes and biomarkers.

3.2 INTRODUCTION

The regulatory factor X (RFX) gene family transcription factors (TFs) were first detected in mammals as the regulatory factor that binds to a conserved *cis*-regulatory element called the X-box motif about 20 years ago (Reith *et al.*, 1988). The X-box motifs, which are typically 14-mer DNA sequences, were initially identified as a result of alignment and inspection of the promoter regions of major histocompatibility complex (MHC) class II genes for conserved DNA elements (Dorn *et al.*, 1987; Sherman *et al.*, 1987). Further investigations revealed that the X-box motif is highly conserved in the promoter regions of various MHC class II genes (Kara and Glimcher, 1991). The first RFX gene (RFX1) was later characterized as a candidate major histocompatibility complex (MHC) class II promoter binding protein (Reith *et al.*, 1989). RFX1 was later found to function also as a transactivator of the hepatitis B virus enhancer (Siegrist *et al.*, 1993). Subsequent studies revealed that RFX1 is not alone. Instead, it became the

founding member of a novel family of homodimeric and heterodimeric DNA-binding proteins, which also includes RFX2 and RFX3 (Reith *et al.*, 1994). More members of this gene family were subsequently identified. A fourth RFX gene (RFX4) was discovered in a human breast tumor tissue (Dotzlaw *et al.*, 1992) and the fifth, RFX5, was identified as a DNA-binding regulatory factor that is mutated in primary MHC class II deficiency (bare lymphocyte syndrome, BLS) (Steimle *et al.*, 1995). The identification of RFX1-5 and RFX genes in other genomes including the genomes of lower eukaryote species *Saccharomyces cerevisiae* (Huang et al., 1998) and *Schizosaccharomyces pombe* (Wu and McLeod, 1995), and higher eukaryote species the nematode *Caenorhabdits elegans* (Swoboda et al., 2000) helped understand both the evolution of the RFX gene family and the DNA binding domains (Emery *et al.*, 1996b). Notably, while previous studies reported five RFX genes (RFX1-5) in human, only one RFX gene has been identified in most invertebrate animals and yeast. In contrast, the fruit fly (*Drosophila melanogaster*) genome has been found to have two RFX genes, dRFX (Dubruille *et al.*, 2002) and dRFX2 (Otsuki *et al.*, 2004). All of these RFX genes are transcription factors possessing a novel and highly conserved DNA binding domain (DBD) called RFX DNA binding domain (Emery *et al.*, 1996a), the defining feature of all members belonging to the RFX gene family, suggesting that these RFX TFs all bind to the X-box motifs.

In addition to the defining DBD domains in all of these RFX genes, most of these previously identified RFX genes also contain other conserved domains including B, C, and D domains (Emery *et al.*, 1996a). The D domain is also called the dimerization domain (Emery *et al.*, 1996a). The B and C domains also play a role in dimerization and are thus called the extended dimerization domains (Katan-Khaykovich and Shaul, 1998).

Another important domain found in many members of the RFX family is the RFX activation domain (AD). For instance, RFX1 contains a well defined AD (Katan-Khaykovich and Shaul, 1998). However, AD is not found in many other members of the RFX family including the human RFX5 and *C. elegans* DAF-19 (Emery *et al.*, 1996a). Outside of these conserved domains, RFX genes from different species or even from same species show little similarity in other regions, which is quite consistent with their diverse functions and distinct expression profiles.

In humans, RFX1 is primarily found in the brain with high expression in cerebral cortex and Purkinje cells (Ma *et al.*, 2006). RFX2 (Wolfe *et al.*, 2006) and RFX4 (Morotomi-Yano *et al.*, 2002) are found to be heavily expressed in the testis. RFX4 is also expressed in the brain (Blackshear *et al.*, 2003). RFX3 is expressed in ciliated cells and is required for growth and function of cilia including pancreatic endocrine cells (Ait-Lounis *et al.*, 2007), ependymal cells (Baas *et al.*, 2006), and neuronal cells (Bonnafe *et al.*, 2004). RFX3-deficient mice show left-right (L-R) asymmetry defects (Bonnafe *et al.*, 2004), developmental defect, diabetes (Ait-Lounis *et al.*, 2007), and congenital hydrocephalus in mice (Baas *et al.*, 2006). RFX5 is the most extensively studied RFX gene so far primarily since it serves as a transcription activator of the clinically important MHC II genes (Villard *et al.*, 2000) and mediates a enhanceosome formation, which results in a complex containing RFXANK (also known as RFX-B), RFXAP, CREB, and CIITA (Reith and Mach, 2001). Mutation in any one of these complex members leads to bare lymphocyte syndrome (BLS) (Reith and Mach, 2001). In *C.elegans* and *S.cerevisae* only one copy of the RFX gene exists. In *C. elegans* it is called DAF-19 and in *S.cerevisae* it is called Crt1. DAF-19 is involved in regulation of sensory neuron cilium

whereas Crt-1 is involved in regulating DNA replication and damage checkpoint pathways (Huang *et al.*, 1998; Swoboda *et al.*, 2000). In *D.melanogaster*, two of RFX genes have been identified, one is called dRFX and the other is called dRFX2. dRFX is expressed in the spermatid and brain and is necessary for ciliated sensory neuron differentiation (Vandaele *et al.*, 2001; Dubruille *et al.*, 2002). dRFX2 has not been studied extensively and as such its function in Drosophila still remains unclear; however, there is evidence suggesting that dRFX2 plays a role in cell-cycle of the eye imaginal discs (Otsuki *et al.*, 2004).

In this project, we have identified and characterized two novel RFX genes in genomes of human and many other mammals, which have now been sequenced, annotated, and analyzed.

3.3 MATERIALS AND METHODS

3.3.1 Data source and data mining

Gene sets were obtained from the FTP site of the ENSEMBL database http://www.ensembl.org/index.html (Flicek *et al.*, 2008). In this project, the genomes of six mammals were analyzed. They are human (*Homo sapiens*, NCBI36.44), chimpanzee (*Pan troglodytes*, CHIMP2.1.44), dog (*Canis familiaris*, BROADD2.44), monkey (*Macaca mulatta*, MMUL_1.44), mouse (*Mus musculus*, NCBIM36.44), and rat (*Rattus norvegicus*, RGSC3.4.44). DBD sequences in human RFX1-5 were manually identified and extracted to a file. The sequences were aligned using ClustalW (Chenna *et al.*, 2003). The alignment was used as input to the profile building program hmmbuild, which is a

program in the HMMER package http://hmmer.janelia.org (Durbin *et al.*, 1998). The resulting profile was used for searching curated proteomes of the six mammals described above using hmmsearch, another program in the HMMER package.

3.3.2 Gene model improvement

All RFX genes except one–dog (Cfa) RFX7–show good alignment with their corresponding orthologs. Dog RFX7 gene is truncated at the N-terminus, missing 37 residues compared to other RFX7 genes. We attempted to use GeneWise http://www.ebi.ac.uk/Wise2/ (Birney and Durbin, 2000; Birney *et al.*, 2004) to remodel this RFX gene. Using human (Hsa) RFX7 as the reference protein sequence and GeneWise, we recovered the missing residues. However, the first codon so identified was not the typical Met. Extending the coding sequence upstream did not help. This is likely due to a sequencing error.

3.3.3 Protein domain analysis

We retrieved DBDs and ADs from RFX genes using InterProScan (version 4.3.1). To identify B, C, D domains, we used the HMMER program (Durbin *et al.*, 1998) as described above. Briefly, for HMMER searches, we used sequences of B, C, and D domains from known RFX genes (RFX1-3) to generate profiles for these domains respectively. We then searched for candidate B, C, and D domains in RFX6 and RFX7 using these profiles.

3.3.4 RFX interactome network analysis

Data were obtained at the HiMAP http://www.himap.org/ database (Rhodes *et al.*, 2005) following online search instructions. All types of interactions were selected for searching. All seven interactions between RFX6 and other genes (DAT1, DTX2, FHL3, SS18L1, CCNK, RFX2, and RFX3) were previously reported by Rual *et al* (Rual *et al.*, 2005).

3.3.5 Sequence alignment and phylogenetic analysis

Multiple-sequence alignment was carried out using the program ClustalW (version 1.83) (Chenna *et al.*, 2003). Phylogenetic tree construction was performed using PHYLIP http://evolution.genetics.washington.edu/phylip.html (Version 3.66). Briefly, sequence alignment in PHYLIP format was first created using ClustalW (Version1.83) (Chenna *et al.*, 2003). The alignment was used as input for PHYLIP. Programs utilized in the PHYLIP, in their respective order, were seqboot, protdist, neighbor, and consense. The phylogenetic tree file was visualized using Tree View http://taxonomy.zoology.gla.ac.uk/rod/treeview.html.

3.3.6 Expression profile of mammalian RFX genes using ESTs and SAGE libraries

The EST database from NCBI was used to perform tblastn. The queries used for this tblastn were RFX1-7 of *H. sapiens, M. musculus*, and *R. norvegicus*. Hits with identity greater than or equal to 95% were selected.

3.4 RESULTS & DISCUSSION

With the current version of the human genome, originally reported in 2001, (Lander *et al.*, 2001; Venter *et al.*, 2001), we explored whether additional members of the RFX TF family could be identified and characterized in the human genome. We applied a Hidden Markov Model (HMM) based search method (Durbin *et al.*, 1998) and used DBD domain sequences of known human RFX TFs to search the entire human proteome. In addition to retrieving all known human RFX genes–RFX1-5, we identified two additional genes in the human genome that contain well conserved RFX DBDs. These two genes were previously assigned as RFXDC1 and RFXDC2 by the HUGO Gene Nomenclature Committee (HGNC, http://www.genenames.org/); this nomenclature was based solely on an initial bioinformatic analyses. There are no previous publications describing these two genes. Here, we demonstrate that these two genes are also RFX gene family members closely related to RFX1-5, and our phylogenetic analysis suggests two separate recent gene duplications leading to the generation of these two genes. Thus, we proposed new gene nomenclature of RFX6 and RFX7 (Table 1), respectively. Our proposal has been accepted by the HGNC.

Because all known human RFX genes–RFX1-5–are well conserved and have been identified in other mammalian genomes, we hypothesized that orthologs of RFX6 and RFX7 also exist in other mammalian genomes. As expected, we have retrieved all seven RFX genes in the genomes of five other mammalian species including chimpanzee (*Pan troglodytes*), monkey (*Macaca mulatta*), dog (*Canis familiaris*), mouse (*Mus musculus*), and rat (*Rattus norvegicus*) with only one exception. In the rat genome, all except RFX2 were found despite extensive searches (Table 1A). Most identified RFX genes are

expressed and their transcripts can be found in existing EST libraries. Interestingly, existing EST evidence suggests that RFX6 and RFX7 have no or very few alternative isoforms similar to RFX1. In contrast, RFX2-4 usually have more alternative isoforms (Table 1A).

To confirm that the two novel human RFX genes–RFX6 and RFX7 are indeed RFX TFs, we further examined their DBDs by aligning them with DBDs from RFX1-5 protein sequences. As expected, the DBDs of RFX6 and RFX7 align well with those of RFX1-5 (Figure 3). RFX TFs belong to the winged-helix family of DNA binding proteins because their DBDs are related in structure and function to the helix-turn-helix bacterial transcriptional regulatory proteins (Wolberger and Campbell, 2000). DBDs from RFX6 and RFX7 each contain one wing (W1), which is the same as DBDs from RFX1-5. W1 interacts with the major groove and another conserved fold H3 (helix 3) interacts with the minor groove of DNA. In particular, the nine residues in DBDs (Figure 3, indicated with arrowheads). that make direct or water-mediated DNA contacts (Gajiwala *et al.*, 2000) are almost entirely conserved in RFX6 and RFX7 (Figure 3 with a couple of minor exceptions. Of the nine residues, the human RFX7 DBD has two residues different from most of the other RFX DBDs. The first different residue is the first of the nine indicated residues. It is Lys in RFX7 DBD and RFX5 DBD, compared to Arg in DBDs of other RFX genes. Thus this difference is shared with the RFX5 DBD. The other different residue is the third of the nine residues. It is Lys in RFX7, compared to Arg at this site for DBDs of all other RFX genes. Because both Lys and Arg are basic amino acids, such substitutions are not expected to have dramatic impacts on the binding between the DBDs

Table 5 – Names and Protein IDs of representative RFX genes.

Gene names	Accession Number (RefSeq)	ENSEMBL protein ID	Genomic coordinates				Protein lengths	Number of exons	Number of isoforms
			chromosome	Start	end	strand			
RFX1	NM_002918	ENSP00000254325	19	13933353	13978097	-1	979	21	1
RFX2	NM_000635	ENSP00000306335	19	5944175	6061554	-1	723	18	2
RFX3	NM_134428	ENSP00000371434	9	3208297	3515983	-1	749	18	8
RFX4	NM_213594	ENSP00000350552	12	105501163	105680710	1	744	18	4
RFX5	NM_000449	ENSP00000357864	1	149581060	149586457	-1	616	11	3
RFX6	NM_173560	ENSP00000332208	6	117305068	117351384	1	928	19	2
RFX7	NM_022841	ENSP00000373793	15	54166958	54222377	-1	1281	7	1

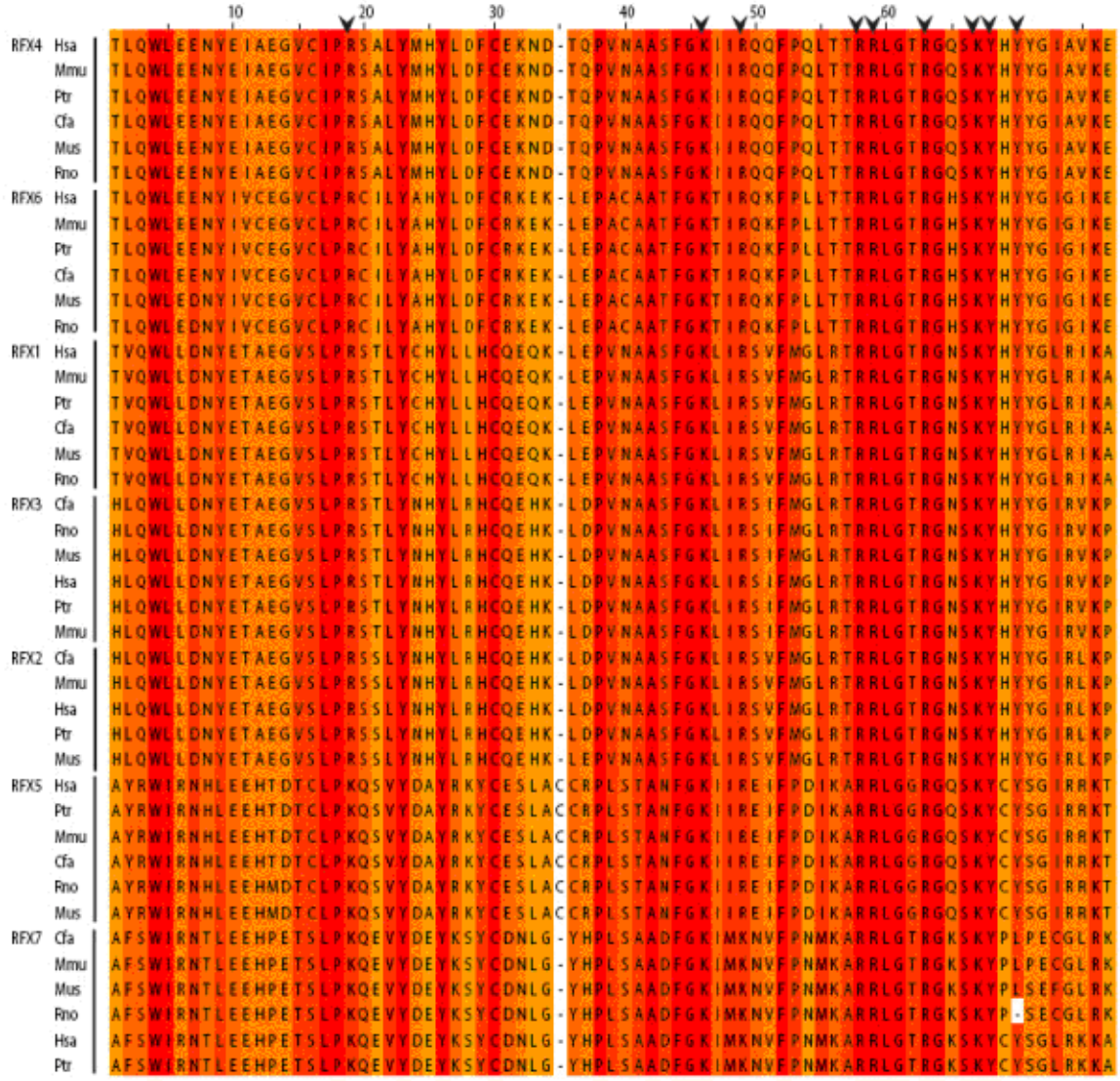


Figure 33 Mammalian RFX DBDs are highly conserved.

DBDs from six mammalian RFX genes were aligned using ClustalW. The conservation of amino acid is depicted by a colour gradient from the colour yellow, which indicates low conservation, to red, which indicates high conservation. Nine residues that make direct or water-mediated DNA contacts are indicated with arrow heads. The species names included in this figure are abbreviated. They are: Mus-mouse (*Mus musculus*); Rno-Rat (*Rattus norvegicus*); Cfa-dog (*Canis familiaris*); Ptr-chimpanzee (*Pan troglodytes*); Mmu-monkey (*Macaca mulatta*) and Hsa-human (*Homo sapiens*).

and their cognate binding sites. This high degree of conservation suggests that RFX6 and RFX7 may bind to similar if not identical *cis*-regulatory elements, i.e., the X-box motif (Reith *et al.*, 1988). Hence RFX6 and RFX7 are new members of the human RFX gene family with conserved DBDs.

In addition to the highly conserved DBDs, other domains including ADs, B, C, and D domains (also known as dimerization domain) (Emery *et al.*, 1996a) have been described in human RFX1-3 (Figure 4). Among these functional domains, ADs have been identified in RFX1-3. However, ADs have not been identified RFX4-5. The B and C domains, which are usually called extended dimerization domains, play supporting roles in dimerization (Katan-Khaykovich *et al.*, 1999). B, C, and D domains have also been identified in RFX4 but are missing from RFX5. Using InterProScan (Mulder and Apweiler, 2007) and HMMER (Durbin *et al.*, 1998), we have found that RFX6 possesses B, C, and D domains, but not AD (Figure 4). The motif composition of RFX6 is similar to RFX4, which also has B, C, and D domains but lacks AD. In contrast, we failed to identify B, C, and D domains or AD in RFX7. None of these domains can be found in RFX5 as well. Because these C-terminal domains–B, C, and D domains–have been shown to mediate dimerization as well as transcriptional repression (Katan-Khaykovich *et al.*, 1999), RFX6, which contains B, C, D domain, and RFX7, which does not possess B, C, or D domains, may therefore play different role in transcriptional regulation.

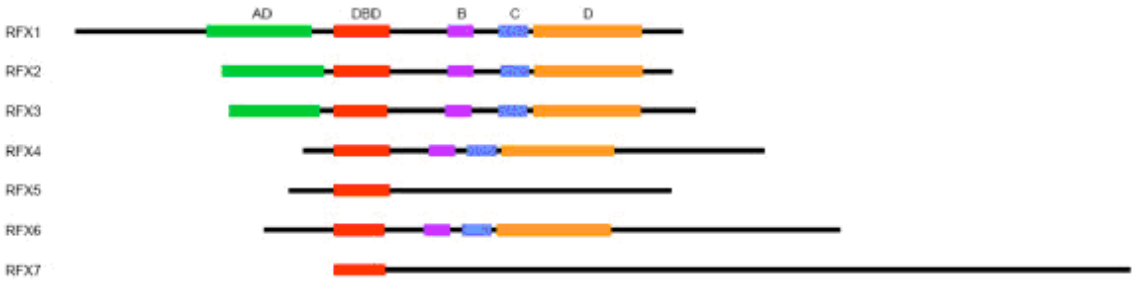


Figure 34 Functional domains in the known and novel human RFX genes.

The functional domains, AD, DBD, B, C, and D are indicated using color-coded boxes. Genes are represented using horizontal lines, which are proportional to the protein lengths. The domain lengths and positions are also proportional to their actual lengths. The graphs are aligned based on the position of the DBDs.

Characterization of the functional domain composition of RFX genes will provide insights into how different RFX TFs function. In particular, how do RFX6 and RFX7, as well as RFX4 and RFX5, function in transcription considering that they do not have identified ADs? There are two possible mechanisms. First, because RFX TFs are known to form dimers and bind to same or similar binding sites (the X-box motifs) in DNA (Gajiwala *et al.*, 2000), they may function together with RFX genes (RFX1-3) that do have ADs. Examination of a recently available proteome-scale map of the human protein-protein interaction network (Rual *et al.*, 2005), which was constructed using yeast-two-hybrid technique, has shown that RFX6 and RFX1-4 interact with each other and also interact with many other genes (Figure 5). RFX6 interacts directly with RFX2 and RFX3, the latter of which has been shown to be expressed and to function in the pancreas (Ait-Lounis *et al.*, 2007), as well as many other tissues. The interaction between RFX6 and other RFX TFs provides further supporting evidence that RFX6 is indeed a member of the RFX gene family. Interactions between RFX7 and other genes were not observed, which is likely due to the incomplete coverage of the human protein-protein interactions analyzed in this study. Second, RFX TFs may function by interacting with many other non-RFX TFs. For example, it has been demonstrated that mammalian RFX 5 forms a complex ("enhanceosome") with RFXANK (also known as RFX-B), RFXAP, CREB, and CIITA to regulate expression of MHC class II genes (Reith and Mach, 2001). Notably, all of the five genes shown to interact with RFX6 (DTX1, DTX2, FHL3, CCNK, and SS18L1) (Figure 5) except only one–SS18L1–are also putative TFs.

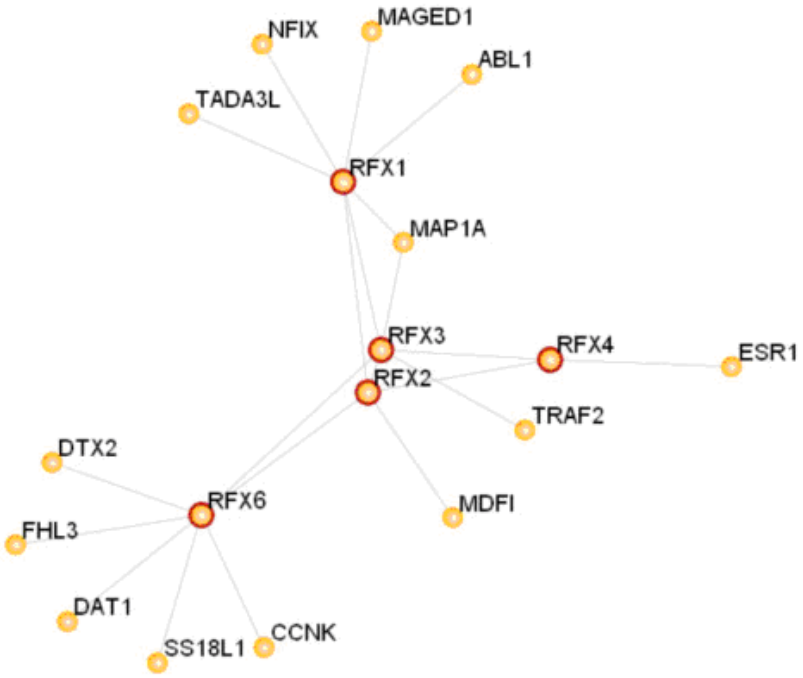


Figure 35 RFX interactome.

Circles depict gene products and lines depict protein-protein interactions. The interactions between RFX6 and its direct interactors were obtained using yeast-two-hybrid method in a large-scale human protein-protein interaction study (Rual *et al.*, 2005). Additional interactions were constructed by Rhodes *et al* (Rhodes *et al.*, 2005). The network was generated using program available at the HiMap website http://www.himap.org/ (Rhodes *et al.*, 2005).

To explore the relationship between RFX6 and RFX7 and the known RFX family members RFX1-5, we have constructed a phylogenetic tree that contains all mammalian RFX genes described above (Table 1A, Figure 6), as well as *C. elegans* RFX gene *daf-19* product DAF-19 (Swoboda *et al.*, 2000), which has been extensively studied, for comparison. We used the DBD sequence of the yeast *Saccharomyces cerevisiae* RFX gene Crt-1 (Huang *et al.*, 1998) as an out group in the phylogenetic tree construction. From the phylogenetic tree (Figure 6) all seven genes show perfect one-to-one orthologous relationships between different mammalian genomes. It is clear that the seven mammalian RFX genes fall into three subgroups (Figure 6). The first subgroup contains RFX1-3; the second RFX4 and RFX6; while the third RFX5 and RFX7. It is likely that RFX4 and RFX6 resulted from one gene duplication that predated the split of these mammalian species, while RFX5 and RFX7 resulted from another similar independent duplication. This hypothesis is generally consistent with the gene models of these RFX genes (Figure 8A). RFX6 has 19 exons, which is similar to the number of exons contained in RFX4 (18 exons); while RFX7 has 6 exons, which is similar to the number of exons contained in RFX5 (9 exons). The *C. elegans* RFX gene, DAF-19 clusters together with RFX1-3 genes, supporting a previously proposed hypothesis that the divergence of the subgroup RFX1-3 from other two subgroups likely predated the divergence between mammals and the nematodes (Emery *et al.*, 1996a). This hypothesis predicts that *C. elegans* should have orthologous RFX TFs to RFX4-7 (Emery *et al.*, 1996b). However, only one *C. elegans* RFX gene–*daf-19*–has been reported so far and our extensive search has concluded that *daf-19* is the only RFX TF in *C. elegans*. One possible explanation is that additional RFX TFs were lost in evolution. Alternatively,

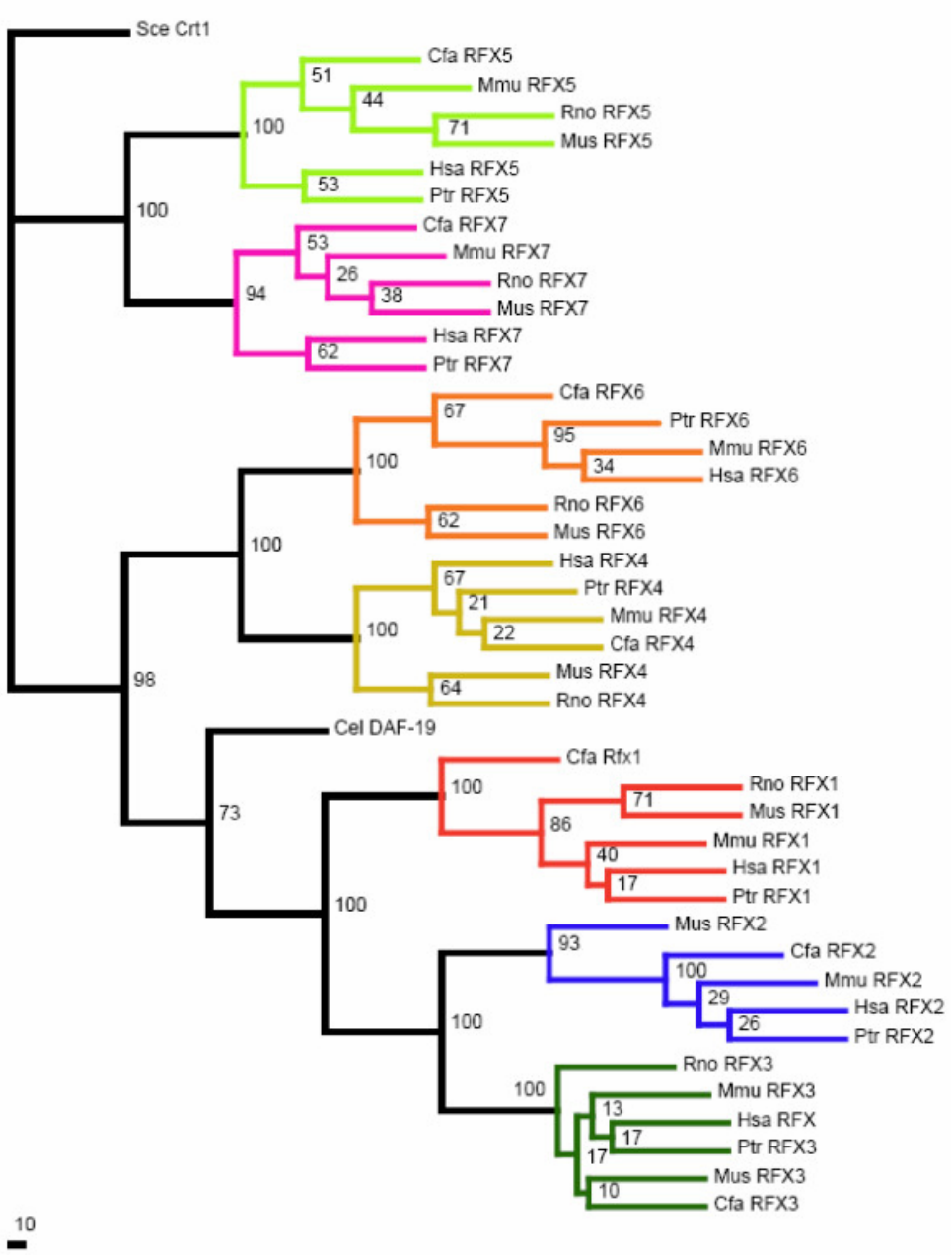


Figure 36 Phylogenetic analysis of mammalian RFX genes.

This phylogenetic tree was constructed based on DBDs of RFX genes for six mammalian species and *C. elegans* using yeast RFX gene product Crt1 as the out-group. The phylogenetic tree was bootstrapped for 100 times with the numbers at each internal node being the bootstrap values. Each ortholog group is colored differently. The species names included in this figure are abbreviated. They are: Mus-mouse (*Mus* musculus); Rno-Rat (*Rattus norvegicus*); Cfa-dog (*Canis familiaris*); Ptr-chimpanzee (*Pan troglodytes*); Mmu-monkey (*Macaca mulatta*) and Hsa-human (*Homo sapiens*).

RFX4-7 may have undergone positive selection in mammals to accommodate additional functional complexity in mammalian gene regulation, while RFX1-3 and *daf-19* remained highly conserved due to purifying evolution. Interestingly, although the phylogenetic tree was constructed based only on DBDs, the grouping of these mammalian RFX genes is also consistent with the composition of other conserved domains. In particular, RFX1-3 all contain DBDs, ADs, Bs, Cs and Ds, while RFX4 and RFX6 have all of these domains except ADs, and RFX5 and RFX7 have only DBDs (Figures 4 and 6). To gain insight into the function of these two newly identified RFX genes, we explored the expression profiles of RFX6 and RFX7 and compared them to those of RFX1-5. We analyzed two independent datasets. First, we searched the dbEST database in genBank http://www.ncbi.nlm.nih.gov/dbEST/ (Rodriguez-Tome, 1997) to examine which EST libraries express transcripts of these RFX genes. The results indicate that the expression profile of RFX1-5 matches well with previously published data (see INTRODUCTION): RFX1 is found in many different tissue types including white blood cells, heart, eye, testis, and cancerous cell; RFX2 appears to be expressed in testis and brain; RFX3 appears to be expressed in the placenta and brain (*i.e.*, medulla); RFX4 is found in the brain, as well as in testis as RFX2; and RFX5 expression has been observed in various different tissues including thymus, T-cells, kidney, brain, and lymph. The consistency of expression for RFX1-5 obtained from the dbEST database with previous observations suggests that dbEST provides good estimations of RFX genes' expression profiles. Using the same method, we found that RFX6 is primarily expressed in pancreas, with minor expression in liver, while RFX7 is widely and heavily expressed in many different tissue types including kidney (tumor tissues), thymus, brain, and placenta.

Second, to gain a quantitative understanding of the expression of RFX genes, we took advantage of the recent availability of serial analysis of gene expression (SAGE) libraries constructed by the Mouse Atlas of Gene Expression Project http://www.mouseatlas.org/ (Siddiqui *et al.*, 2005). To start with, we tested the hypothesis that the expression of mouse RFX TFs approximates the expression of human RFX TFs. We analyzed 196 mouse SAGE libraries, each of which was produced by using a RNA library prepared from different tissue types (some of which are duplicates). Different SAGE libraries contain slightly different number of total SAGE tags. To ensure that SAGE tags and tag counts were comparable between different SAGE libraries all the libraries were normalized to 1,000,000 SAGE tags. Qualitatively, expression profiles of mouse RFX genes obtained from SAGE analysis are consistent with the expression profiles of human RFX genes obtained from the dbEST database analysis, as well as previous publications about human RFX gene expressions (Figure 7). In contrast to all other RFX genes–RFX1-5 and RFX7, which are heavily expressed in the brain, RFX6 is clearly absent from all types of brain tissues (Figure 7). RFX6 is primarily found in the pancreas (Figure 7) which is consistent with results obtained from analyzing dbEST. Low level expression of RFX6 is found in liver (also detected in dbEST) and heart. In addition to the high tissue-specificity, RFX6 has the lowest overall expression level among all seven RFX genes, suggesting that RFX6 may be under tighter regulatory control. In contrast, RFX7 has the highest relative expression level among all seven mouse RFX genes. Similar to RFX1 and RFX5, RFX7 is found in essentially all types of tissues that were examined (Figure 7). Examining additional gene expression databases, including

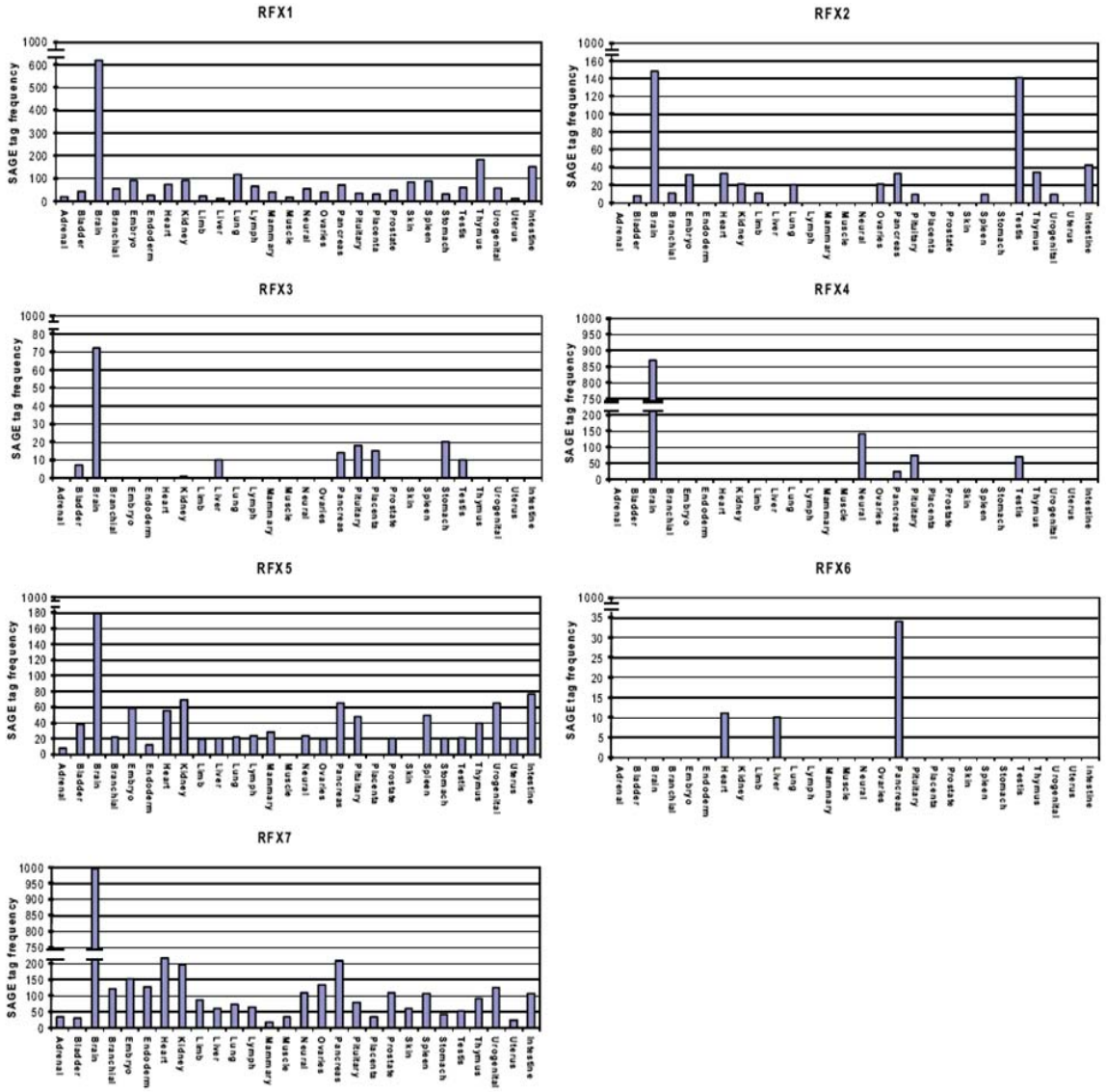


Figure 37 Relative expression of human RFX genes revealed by SAGE.

Original SAGE libraries were generated by the Mouse Atlas Project (Siddiqui *et al.*, 2005). X-axis shows different tissue types, while Y-axis shows relative SAGE tag frequency.

publicly available Genomics Institute of the Novartis Research Foundation (GNF) Gene Expression Database http://symatlas.gnf.org/SymAtlas/, revealed very similar results.

3.5 CONCLUSION

Our results show that we have identified two novel RFX genes in the human genome, RFX6 and RFX7, thus expanding the human RFX gene family from five members (RFX1-5) to seven members (RFX1-7). In addition to their possession of highly conserved DBDs, RFX6 and RFX7 show similarity to known human RFX TFs in their functional domains. In particular, RFX6 and RFX4 all have B, C, and D domains, while RFX7 and RFX5 only have DBDs. Studies carried out over the past 20 years have demonstrated that RFX1-5 are critical for development and many additional biological processes and play an important role in various devastating disease conditions. For example, RFX3-deficient mice show left-right (L-R) asymmetry defects (Bonnafe *et al.*, 2004), developmental defects, diabetes (Ait-Lounis *et al.*, 2007), and congenital hydrocephalus (Baas *et al.*, 2006). RFX3 may regulate the transcription of many genes that, when mutated, cause cilia defects and many disease conditions collectively called ciliopathies (Badano *et al.*, 2006). Many known ciliopathy genes, including Bardet-Biedl syndrome (BBS) genes, are well conserved and the transcription of their *C. elegans* orthologs are regulated by the only RFX gene in *C. elegans*–DAF-19 (Swoboda *et al.*, 2000; Blacque *et al.*, 2005; Efimenko *et al.*, 2005; Chen *et al.*, 2006). Mutation in any one of the RFX5 enhanceosome members–RFXANK, RFXAP, CREB, and CIITA–leads to bare lymphocyte syndrome (BLS) (Reith and Mach, 2001). We hypothesize that RFX6

and RFX7 are equally important as RFX1-5. The fact that RFX6 is primarily expressed in pancreatic tissues and is expressed at a low level compared to all other RFX genes (Figure 7) is particularly interesting. RFX6 may function as a key component of a transcriptional regulatory complex that regulates pancreas development and function.

Chapter 4: General Conclusions

4.1 Finding novel RFX TFs

Probing the newly fully sequenced human genome with the conserved DBD from known RFX TFs, RFX1-5, allowed us to reveal two novel RFX TFs in humans. Upon further investigation of the domains present in these TFs, we were able to demonstrate that RFX6 also contains conserved DIMs but no conserved AD. RFX7 did not have any of these other conserved domains. A phylogenetic analysis of the DBD domains from RFX1-7 showed that RFX6 forms a subgroup with RFX4 and RFX7 forms a subgroup with RFX5. This observation suggested that the two pairs of RFX, RFX4 with RFX6 and RFX5 with RFX7, underwent gene duplication events that predated the split of mammalian species represented in the phylogenetic tree. Also, a search of HiMAP, which is based primarily on high-throughput yeast-two-hybrid screens, revealed that RFX6 has direct interactions with RFX2 and RFX3 and indirectly with RFX1 and RFX4. We further characterized these TFs by aligning the RFX6 and RFX7 sequences with ESTs from the dbEST database and with SAGE tags from MouseAtlas SAGE library. The results from both alignments corresponded well with one another and revealed the highly specialized expression of RFX6 in the pancreas with some expression also in heart and liver, and the high expression of RFX7 in the brain with ubiquitous expression in other tissues. Using the same approach, my laboratory has recently searched for RFX TFs in all sequenced genomes and has found multiple RFX genes in all metazoans except in nematodes (Chu *et al.*, 2010). Based on their evolutionary relationship with the ciliary

genes in organisms ranging from unicellular species including yeast and green algae to mammals, it has been hypothesized that the convergent evolution of RFX genes and ciliary genes played a pivatol role in the establishment of metazoans (Chu *et al.*, 2010).

4.2 *daf-19* expression analysis

By using MosSCI, which allows comparative analysis of gene expression at a near-endogenous level, I was able to characterize ciliary expression of the *daf-19c* and *daf-19d* isoforms. The expression of the *daf-19d* isoform was found to be regulated by a modular promoter. The two promoters chosen, *pdaf-19d* (pGG20) and *pdaf-19d* (pGG21), were based on a previous study where these two promoters were found to express in ciliated neurons but at different stages. In my MosSCI stable transgenic worms, I observed a similar but more modular expression pattern in that each promoter expressed in a different subset of amphid ciliated neurons. The expression from *pdaf-19d* (pGG20) was specific to amphids that are not capable of filling with the fluorescent dye, DiO, and *pdaf-19d* (pGG21) drove a higher level of expression in both amphids that fill with dye and those that do not. This result illustrates the modular nature of this promoter, suggesting that the endogenous expression of the *daf-19d* isoform depends on both modules of the entire promoter.

Furthermore, the *daf-19c* promoter expression was probed using both a transcriptional fusion construct and a translational fusion construct. The transcriptional reporter, containing only the selected *daf-19c* promoter, drove expression very specifically in labial ciliated neurons in both hermaphrodites and males which is in contrast to the current annotation in WormBase that states this isoform is male-specific. Surprisingly, the translational construct expressed in both labial and amphid ciliated

neurons as well as some additional head neurons. This difference in expression between the two *daf-19c* constructs indicates that the regulation of this isoform may also be by modular *cis*-regulatory elements.

The expression of *daf-19a/b* isoforms may not reflect their endogenous expression profiles. In this project, I found that both of the *daf-19a/b* promoters chosen, *pdaf-19a/b* (long) and *pdaf-19a/b* (short), drove expression in non-ciliated neurons, hypodermis, pharynx and gut. The expression in non-ciliated neurons could not be resolved to individual cells. The expression in non-ciliated neurons is generally consistent with a previous study where antibodies were used to confirm the expression of the *daf-19a/b* isoform (Senti and Swoboda, 2008). The expression in the other tissues was also observed previously in some worms containing a translational fusion of the *daf-19* gene (Swoboda *et al.*, 2000) and in a transcriptional fusion of *pdaf-19a/b* (long) in a non-stable strain (Tarailo-Graovac M., unpublished). The expression of *daf-19a/b* isoforms has been observed in non-ciliated neurons using immunostaining with *daf-19a/b* specific antibodies. In contrast, it has not been confirmed with antibodies whether the *daf-19a/b* isoform is in fact expressed in these other tissues, thus the promoter for the *daf-19a/b* isoform that was chosen may not be a complete promoter.

Additionally, the expression of the *daf-19* isoform promoters was observed in males and interestingly there was expression in male-specific cells that was not observed in hermaphrodites. This sex-specific expression suggests that there are male specific co-factors of transcription that allow for expression in male specific cells.

Thus, MosSCI proved to be a valuable tool for confirming and better defining previous expression patterns as well as revealing some novel ones for the known

isoforms of *daf-19*. Understanding *daf-19* and its complex transcriptional regulation will give insight into the regulation of the the various human RFX TFs which may have conserved mechanisms of regulation. Also, it will set the foundation for studies of human RFX TFs in the model organism *C. elegans*.

Appendix I: Functional analysis of human RFX6

INTRODUCTION

Through bioinformatics searches described in Chapter 3, two new RFX TFs, RFX6 and RFX7, were identified in mammals (Aftab *et al.*, 2008). As discussed in Chapter 2, RFX6 was found to be mainly expressed in pancreas, with low levels of expression in the heart and liver. This observation suggests that RFX6 may be playing an important role in the development and function of pancreatic islets. The specialized expression of RFX6 in pancreatic islets may provide a unique opportunity for understanding the transcriptional regulation of genes in pancreatic islet cells. Pancreatic islets are important for the production of insulin, a hormone that is secreted when food is ingested and signals for the breakdown of sugars so that they can be utilized for energy production. Defects in insulin production lead to diabetes.

In order to examine whether RFX6 plays a role in cilia development, I took advantage of the model organism *C. elegans*. In this nematode, a well established mutant of the RFX TF, *daf-19,* exists and is called *daf-19(m86)*. This mutant is known to be completely devoid of cilia in the ciliated sensory neurons (Swoboda *et al.*, 2000). Through the robust expression analysis of the *daf-19* isoforms described in Chapter 3, I can choose a suitable promoter to drive RFX6 expression in ciliated neurons and test its function in cilia development. The *pdaf-19d* (pGG20) promoter drove expression in a subset of ciliated sensory neurons and was chosen to drive human RFX6 expression in *C.*

elegans. Later, other promoters characterized in Chapter 3 can be used to elucidate RFX6 function in different subsets of cells.

Rescue of cilia defective phenotypes: Dyf and Daf-c

The *daf-19(m86)* mutant contains a nonsense mutation just before all functional RFX domains, DBD and DIMSs and is characterized by two phenotypes: 1) dye-filling defective (Dyf) and 2) dauer formation constitutive (Daf-c) (Swoboda *et al.*, 2000). The Dyf phenotype is characterized by the inability of worms to take up a fluorescent dye into their ciliated neurons. This is because there are no cilia extensions in the neurons of these mutants and thus no dye is able to fill these cells (Perkins *et al.*, 1986). In Daf-c, ~90% of worms go into a survival state called dauer, instead of continuing into the third larval molt (Figure 38) and eventually becoming adults (Swoboda *et al.*, 2000). The dauer state is entered in response to three signals in particular: high population density, depletion of food supply and high temperatures (Golden and Riddle, 1982). The dauer pheromone is secreted by worms and reflects the population density. ADF, ASI and ASG ciliated sensory neurons are responsible for sensing the dauer pheromone and when ablated, as in *daf-19(m86)* mutants, worms enter dauer constitutively (Bargmann and Horvitz, 1991). If RFX6 has a role in cilia development, then it may be able to rescue these cilia related mutant phenotypes in *C. elegans.*

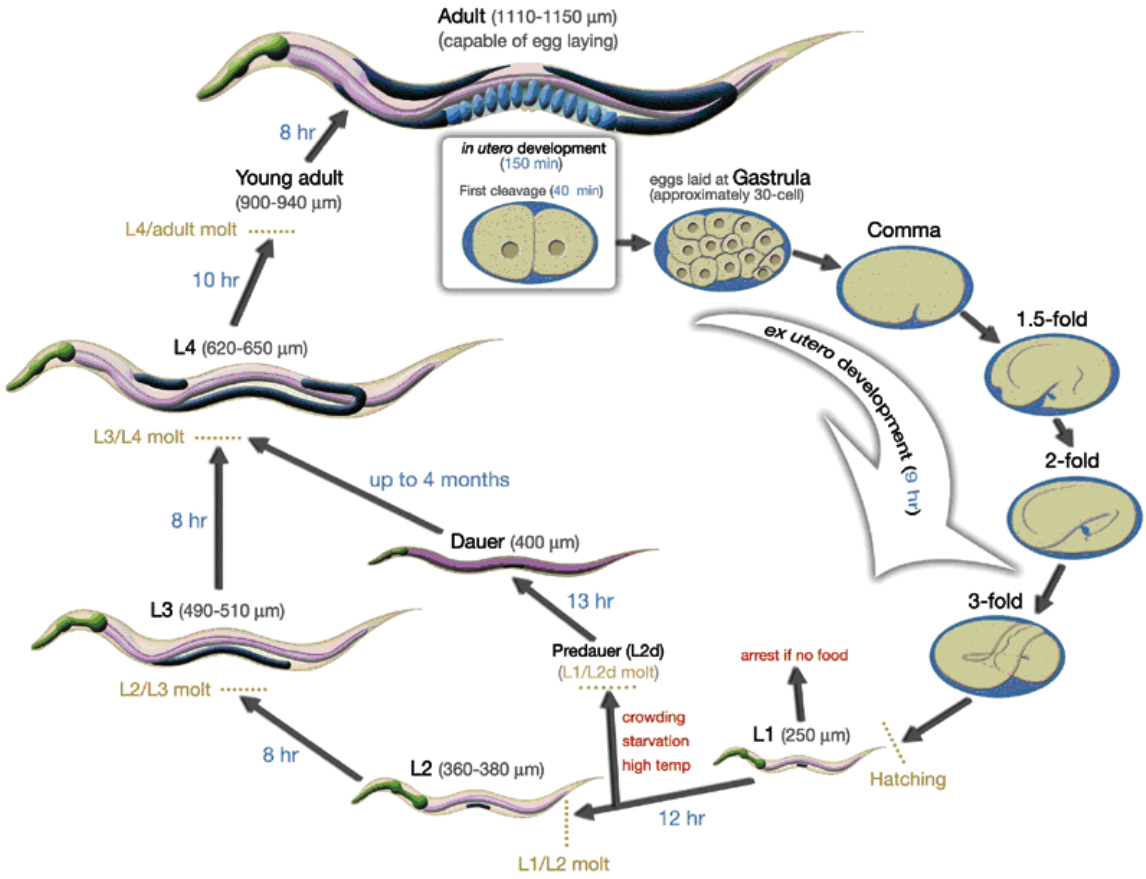


Figure IA Schematic diagram of the life cycle of *C. elegans.*

Figure from WormAtlas (http://www.wormatlas.org/images/HeadNeurons.jpg) (Altun and Hall, 2002-2006).

MATERIALS AND METHODS

Generating translational fusion of RFX6 with *daf-19* promoter

Primers used for amplifying the *daf-19d* (pGG20) isoform promoter were A and B (Table 6). Primers for amplifying human RFX6 cDNA (from openBiosystems), where the first 25 nt are the last 25 nt of the *daf-19d* isoform promoter so that the promoter can be stitched to the start of RFX6, and H (Table 6). Once these products were amplified they were both used as template in the next stitching PCR where the primers to stitch these two products together were A* and H* (Table 6), where H* has the reverse complement of the first 25 nt of the GFP sequence so that the *pdaf-19d* (pGG20)::RFX6 fusion product can be stitched to GFP. The program used for this fusion was the Stitch 1 program (Table 2A in appendix). GFP was amplified from the pPD95.67 vector using primer E and F (Table 6). The GFP was fused to *pdaf-19d* (pGG20)::RFX6 with primer A** and primer F* (Table 6) by PCR stitching. The program used for this PCR stitching was the Stitch 2 program (Table 2A in appendix). This entire product, *pdaf-19d* (pGG20)::RFX6::GFP, was then amplified with primers specific for cloning into the pCFJ178 vector with primer Pd6GFP178_F, where AGGCGG is the filler sequence followed by a BsiWI site and Pd6GFP178_R, where GGATAA is a filler sequence followed by a BssHII site.

Table IA – Translational fusion primers to generate *pdaf-19d* (pGG20)::RFX6::GFP.

PCR Product	Primer Name	Primer Sequence
pd19d	A	TGCCTCCGTAAGATTTGAGG
	B	CTAAATGGAAGATGGTCATAGTTG
RFX6	G	CCAACTATGACCATCTTCCATTTAGATGGCCAAGGTCCCGAAG
	H	TCGGAATGCTGCTATGTGAC
pd19d::RFX6	A*	TTCCGGTGCCATTAGGTATC
	H*	GAGTCGACCTGCAGGCATGCAAGCTAGTGCCTCCAGCTGCTGT
GFP	E	AGCTTGCATGCCTGCAGGTCGACT
	F	AAGGGCCCGTACGGCCGACTAGTAGG
pd19d::RFX6::GFP	A**	CAGTGCCCTAACGACTCACA
	F*	GGAAACAGTTATGTTTGGTATATTGGG
pd19d::RFX6::GFP	Pd6GFP178_F	AGGCGGCGTACGCAGTGCCCTAACGACTC
	Pd6GFP178_R	GGATAAGCGCGCGGAAACAGTTATGTTTGG

Cloning into Mos vector, injection and direct insertion screening

To insert the translational fusion transgenes into the pCFJ178 vector, both the pCFJ178 vector and the transgene were digested with BsiWI (#R0553S) and BssHII (#R0199S) restriction enzymes. Conditions for the BsiWI and BssHII double digest were to digest ~0.8 µg of transgene and ~1.0 µg of vector (pCFJ178) separately for 1.5 hr at 50˚C with BsiWI only and then to raise the temperature to 55˚C and add BssHII and digest for an additional 1.5 hr. Digested vector was de-phosphorylated with Roche alkaline phosphatase for 1 hour at 37˚C. The rest of the steps were the same as those described in section 3.2.2. After cloning microinjection was done by Domena Tu and the injection mix was made according to the MosSCI direct insertion protocol as was described in section 3.2.3. Direct insertion screening followed injection and followed the same protocol as described in section 3.2.4.

RESULTS & DISCUSSION

Using MosSCI to generate integrated strains

As described in previous chapter, the MosSCI method involves three major steps: 1) generating the transgene, 2) cloning the transgene into the Mos vector and 3) screening worms transformed with the vector for direct insertion. I used this method to generate a stable integrated line containing the RFX6 rescue transgene. This rescue transgene contains a *daf-19d* promoter (pGG20) that is PCR stitched (Hobert, 2002) to human RFX6 cDNA and driving a GFP reporter (Figure 39). The pGG20 *daf-19d* promoter was chosen because it drives expression in amphid ciliated neurons (Figure 20). The RFX6 rescue transgene was also cloned into the pCFJ178 vector, which was confirmed by PCR (Figure 40). The successful construct was injected into the EG5003 strain and screened and confirmed for direct insertion, using the same genotyping primers as previously (Figure 41).

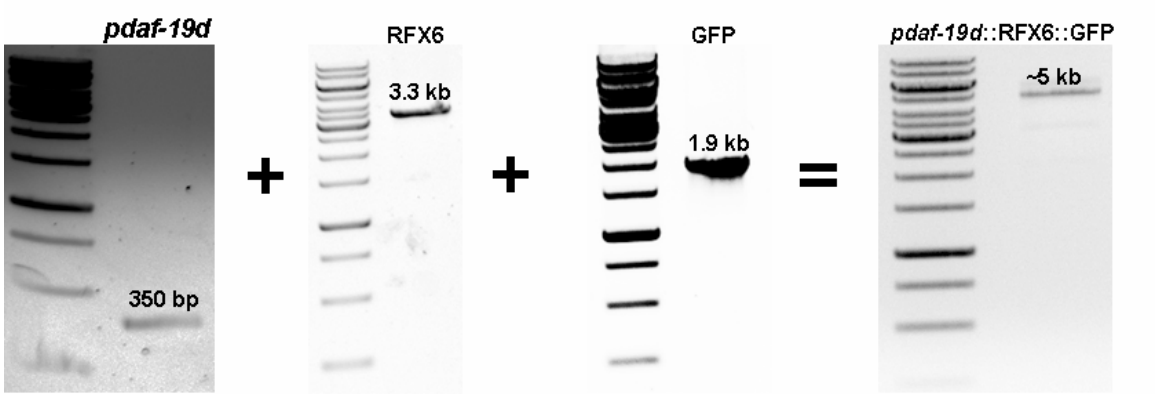


Figure IB Translational fusion PCR of *pdaf-19d* (pGG20)::RFX6::GFP.

The first gel image shows the 350 bp amplification product of the *pdaf-19d* (pGG20) isoform promoter. The second gel image shows the 3.3 kb band amplified, corresponding to human RFX6 cDNA. The + sign between these two images indicates a PCR stitching step. The third gel image shows the 1.9 kb band amplified, corresponding to GFP product (amplified from pPD95.67 vector). The fourth gel shows the product of the PCR stitching, *pdaf-19d* (pGG20) stitched to RFX6 and then *pdaf-19d* (pGG20)::RFX6 was further stitched to GFP to give the final ~5 kb product, *pdaf-19d* (pGG20)::RFX6::GFP. All ladders are O'GeneRuler 1 kb Ladder.

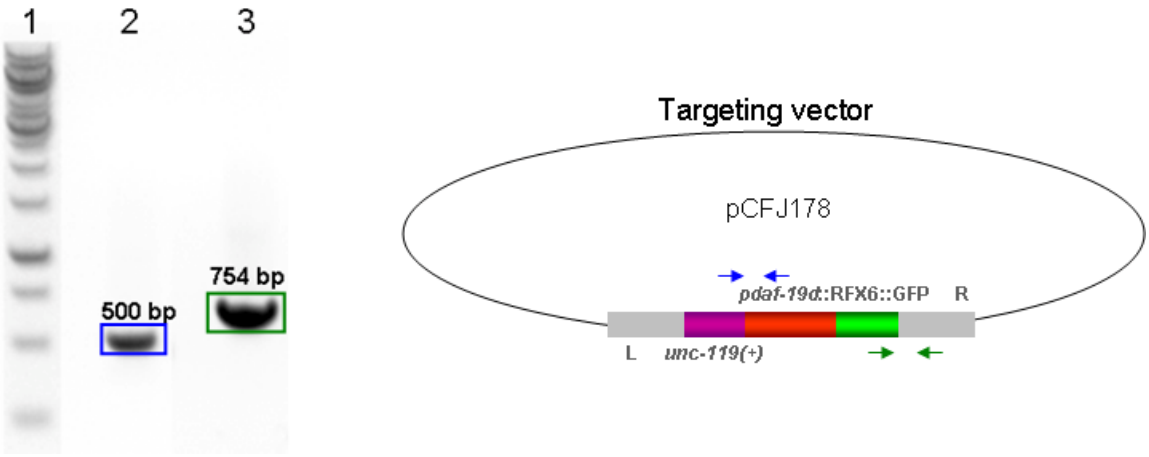


Figure IC Confimation of successful cloning of *pdaf-19d* (pGG20)::RFX6::GFP transgene into the pCFJ178 targeting vector.

Lane 1 is Fermentas GeneRuler 1 kb Ladder Plus. Lane 2 is a PCR using a single colony transformed with *pdaf-19d* (pGG20)::RFX6::GFP as template and using primers (blue arrows) specific to the *unc-119* rescue gene and the *pdaf-19d* (pGG20) promoter to amplify a 500 bp band . Lane 3 contains the same template as Lane 2 but primers (green arrows) specific to GFP and the right homology arm of the pCFJ178 vector to amplify the 754 bp band. These PCRs confirms that the full transgene, including left and right arms, were successfully cloned into the pCFJ178 vector.

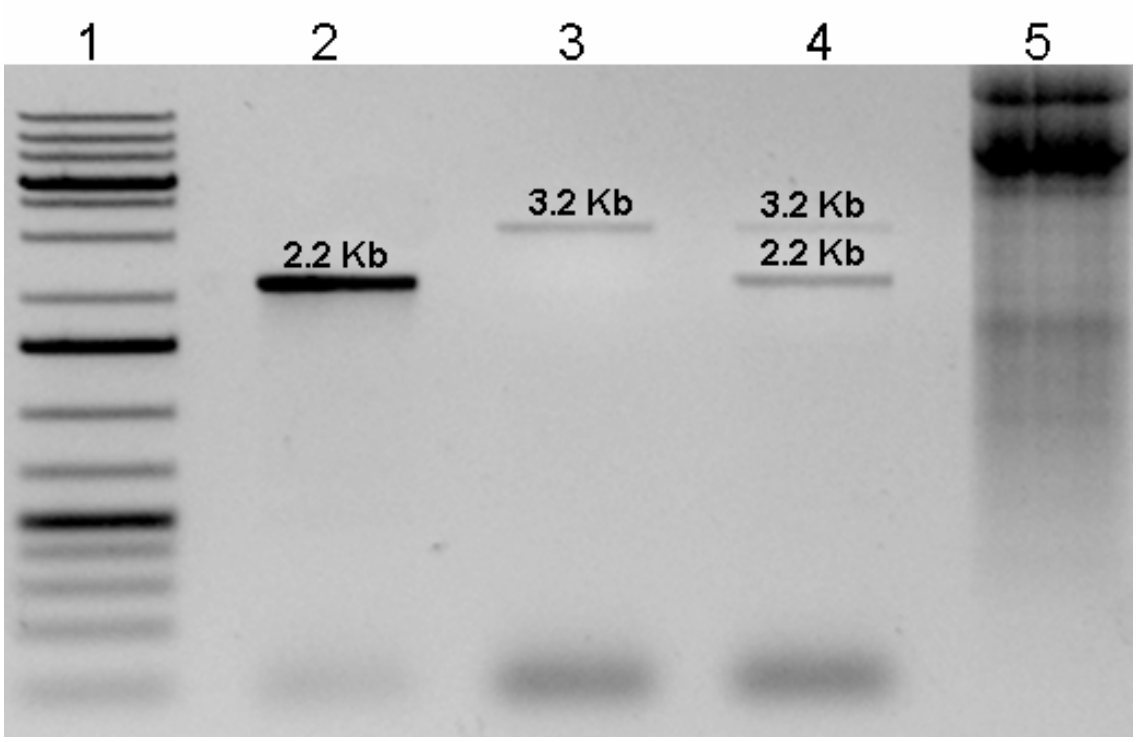


Figure ID PCR genotyping results for direct insertion of *pdaf-19d* (pGG20)::RFX6::GFP into EG5003 Mos strain.

Lane 1 is Fermentas GeneRuler 1 kb ladder Plus. Lane 2 is worm lysate of transgenic worm, JNC66, containing *pdaf-19d*::RFX-6::GFP in the EG5003 background and the 2.2 kb band is amplified corresponding to a direct insertion. Lane 3 is a negative control of EG5003 worm lysate and the 3.2 kb band is amplified corresponding to no insertion. Lane 4 is JNC64 (positive control) plus EG5003 worms to act as a heterozygous control; the expected 3.2 kb and 2.2 kb bands are present. Lane 5 is pCFJ178 vector as a negative control.

Human RFX6 transgene in *C. elegans daf-19(m86)*

The *pdaf-19d*(pGG20)::RFX6::GFP translational fusion rescue construct is currently being crossed into the *daf-19(m86)* homozygous background and will then be assessed for its ability to rescue the Dyf phenotype of this *daf-19* mutant strain. Microscope images will be taken of the dye-filling results. Also, a screen for Daf-c rescue will be done where the percentage of dauer formation will be counted. This result will help elucidate the potential role of RFX6 in cilia development.

Recently, two studies in mice gave conclusive *in-vivo* evidence that RFX6 is expressed in and important for the development of pancreatic islets and for the production of insulin, thus confirming its function in the pancreas that was proposed in Chapter 2 (Smith *et al.*, 2010; Soyer *et al.*, 2010). Although the function of RFX6 has been elucidated, the mechanism by which it regulates insulin production and pancreatic islet development is still not known. A study by Bockman and colleagues found that cilia in the pancreas have an anatomy similar to the 9+0 pattern present in sensory cilia (Bockman *et al.*, 1986). Interestingly, these pancreatic cilia contain insulin receptors on their membranes (Mossner *et al.*, 1984). Since RFX6 is important for insulin production and insulin receptors are found on pancreatic cilia, RFX6 may have a role in cilia development or function. Also, since RFX3 has a role in the development of pancreatic cilia (Ait-Lounis *et al.*, 2007) and RFX6 is proposed to interact with RFX3 (Aftab *et al.*, 2008), this further supports the hypothesis that RFX6 may have a role in cilia development or function.

Appendix II

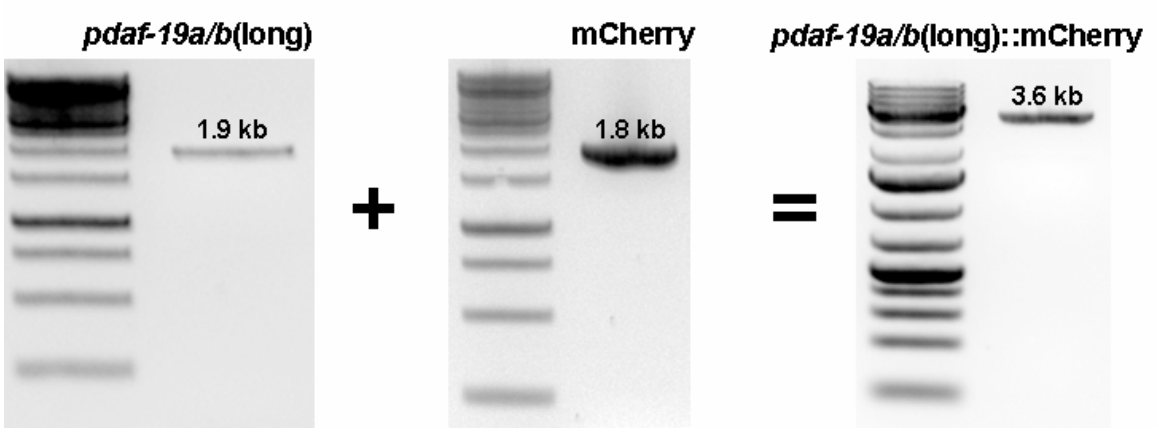


Figure IIA transcriptional fusion of *pdaf-19a/b*(long)::mCherry.
All ladders are O'GeneRuler 1 kb Ladder.

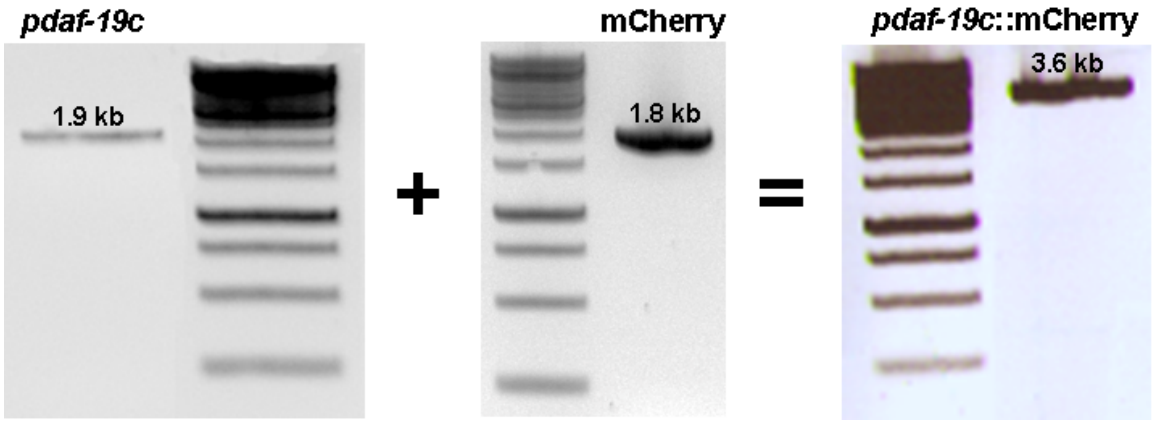


Figure IIB transcriptional fusion of *pdaf-19c*::mCherry.
All ladders are O'GeneRuler 1 kb Ladder.

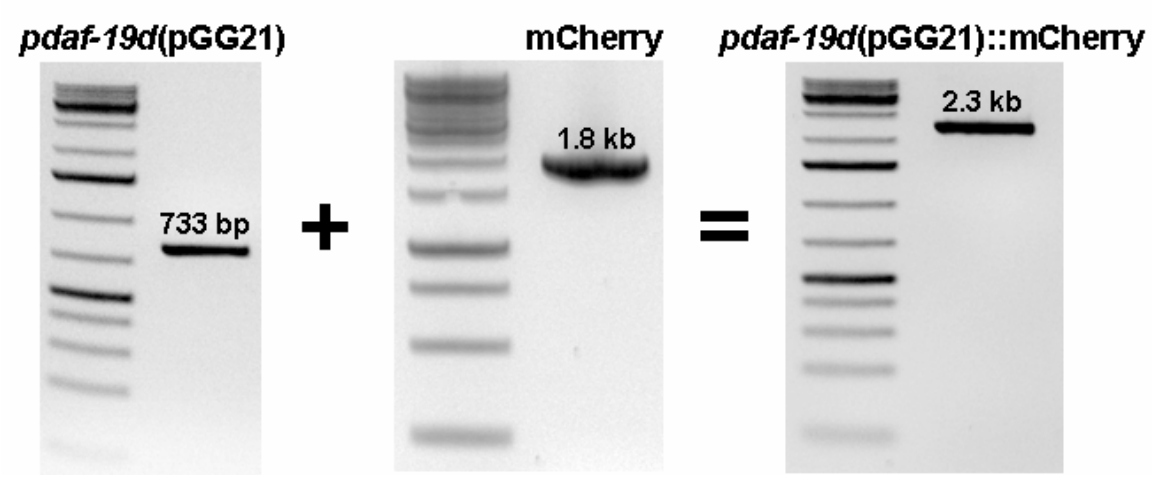


Figure IIC transcriptional fusion of *pdaf-19d*(pGG21)::mCherry.
All ladders are O'GeneRuler 1 kb Ladder.

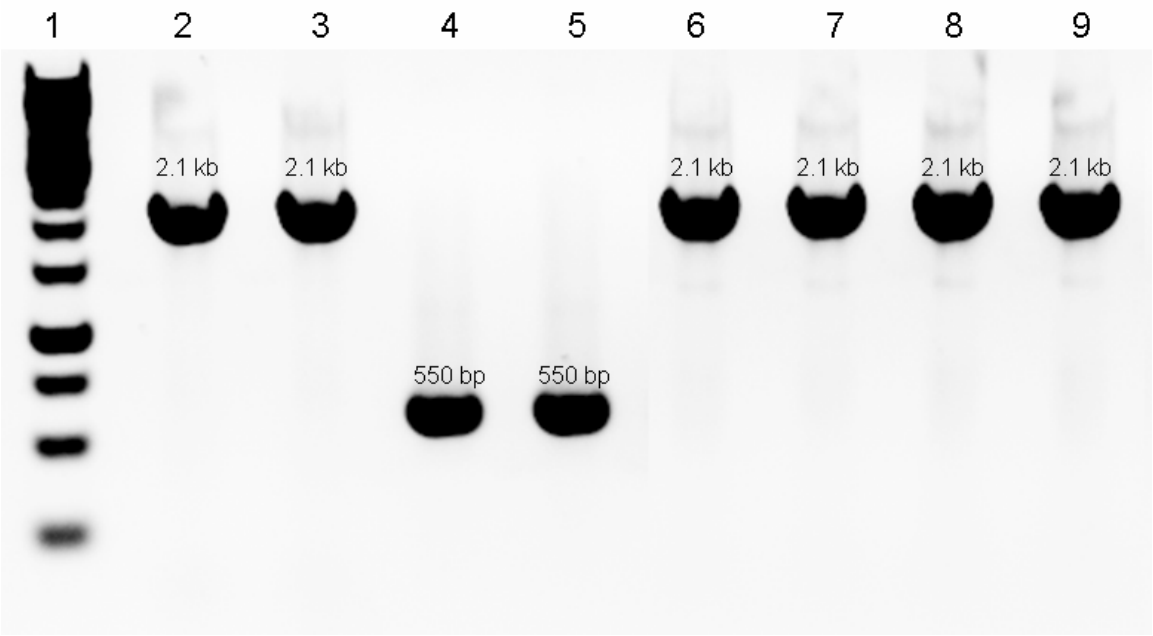


Figure IID PCR genotyping for insertion of *pdaf-19c*::mCherry into pCFJ178 and insertion of *pdaf-19a/b*(short)::mCherry into pCFJ178.

Lane 1 is Fermentas O'GeneRuler 1 kb Ladder. Lane 2 and Lane 3 contain mini-prepped DNA from colonies 1 and 2, respectively, transformed with *pdaf-19c*::mCherry transgene as template and Mos6F15 and P19c_r as primers. Lane 4 and 5 contain mini-prepped DNA from colonies 1 and 2, respectively, transformed with *pdaf-19a/b*(short)::mCherry transgene as template and Mos6F15 and P19ab_r as primers. Lane 6 and 7 have same template as Lane 2 and 3 but using RFP_f and 178_MosR as primers. Lane 8 and 9 have same template as Lane 4 and 5 respectively, but using RFP_f and 178_MosR as primers.

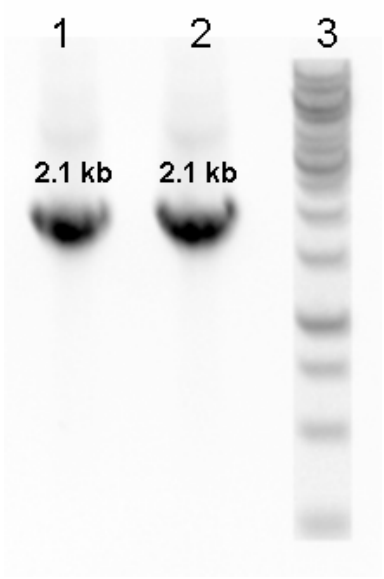


Figure IIE PCR genotyping for insertion of *pdaf-19a/b*(long)::mCherry into pCFJ178.

Lane 1 contains mini-prepped DNA from transformant with *pdaf-19a/b*(long)::mCherry transgene as template and Mos6F15 and P19a/b_r as primers. Lane 2 contains the same template as Lane 1 but using RFP_f and 178_MosR as primers. Lane 3 is Fermentas O'GeneRuler 1 kb Ladder.

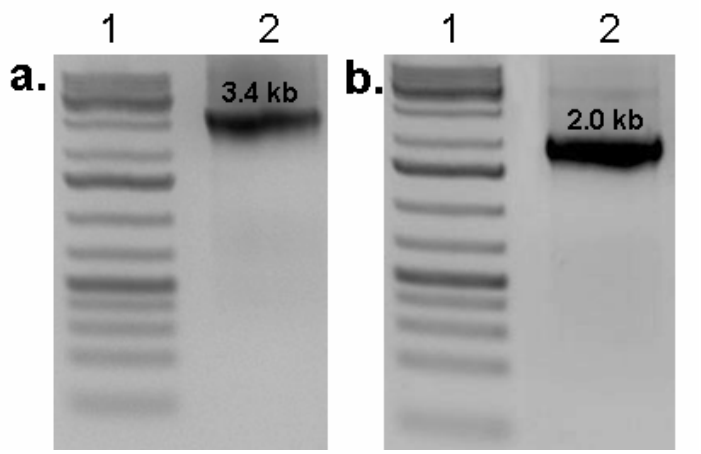


Figure IIF PCR genotyping for insertion of *pdaf-19d*(int4)::mCherry into pCFJ178. A.

Lane 1 is Fermentas GeneRuler 1 kb Ladder Plus. Lane 2 contains mini-prepped DNA from transformant with *pdaf-19d*(int4)::mCherry transgene as template and Mos178genoF and Intron4_R as primers. B. Lane 1 is Fermentas GeneRuler 1 kb Ladder Plus. Lane 2 contains the same template as Lane 1 from gel A, but using RFP_f and Mos178genoR as primers.

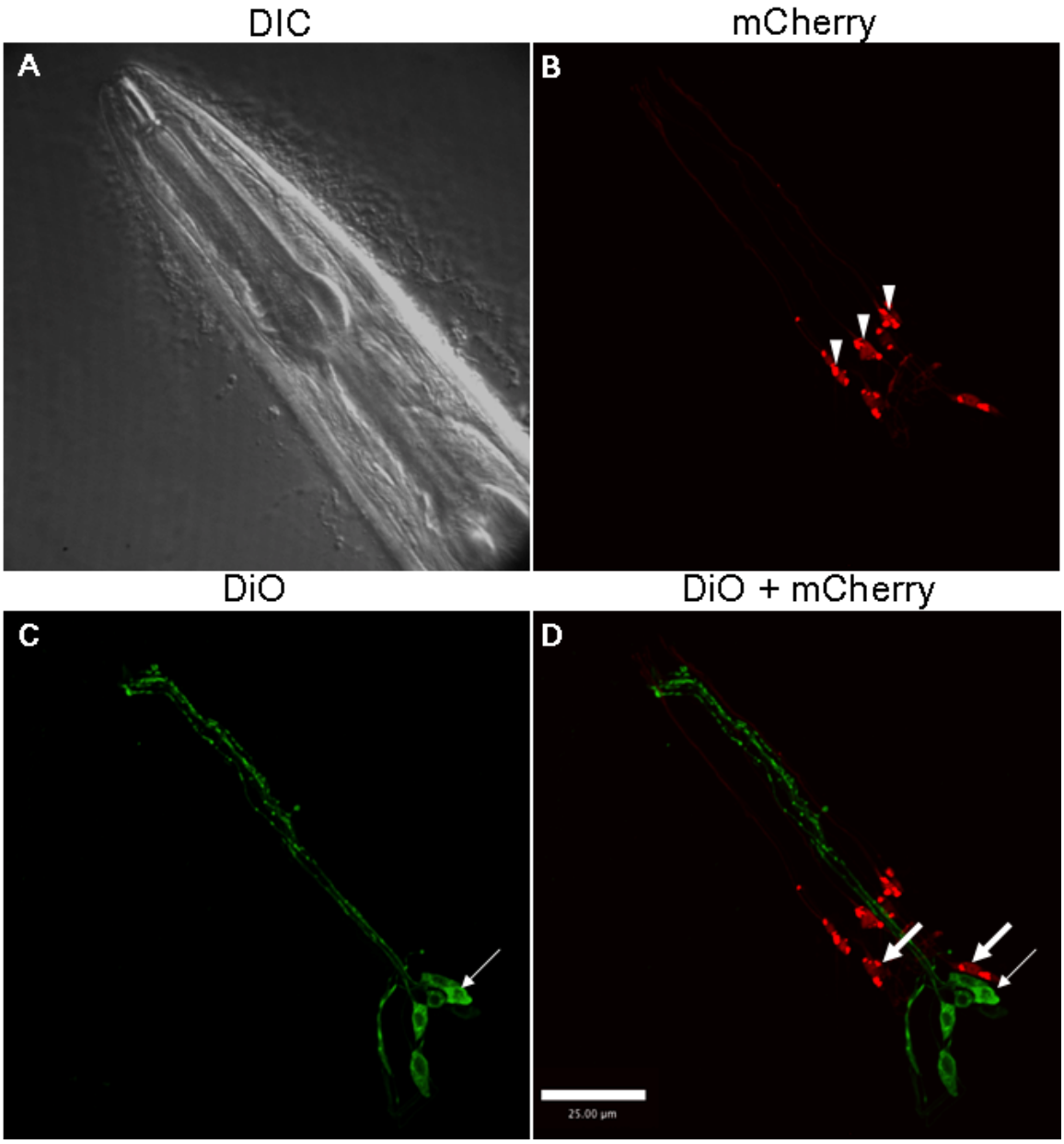


Figure IIG DiO confirms the *daf-19c* isoform is expressed in labial neurons in males.

(A) DIC of adult male *pdaf-19c*::mCherry transgenic worm head. (B) mCherry expression of putative amphid and labial neurons. Arrowheads indicate labial neurons. (C) DiO staining of transgenic worm. Thin arrows indicate amphid, DiO filling neurons. (D) Merged image of DiO staining and mCherry expression. Thick arrows indicate unidentified but putative amphid, non-DiO filling neurons. All exposures were taken at 2 sec. Bar represents 25 µm.

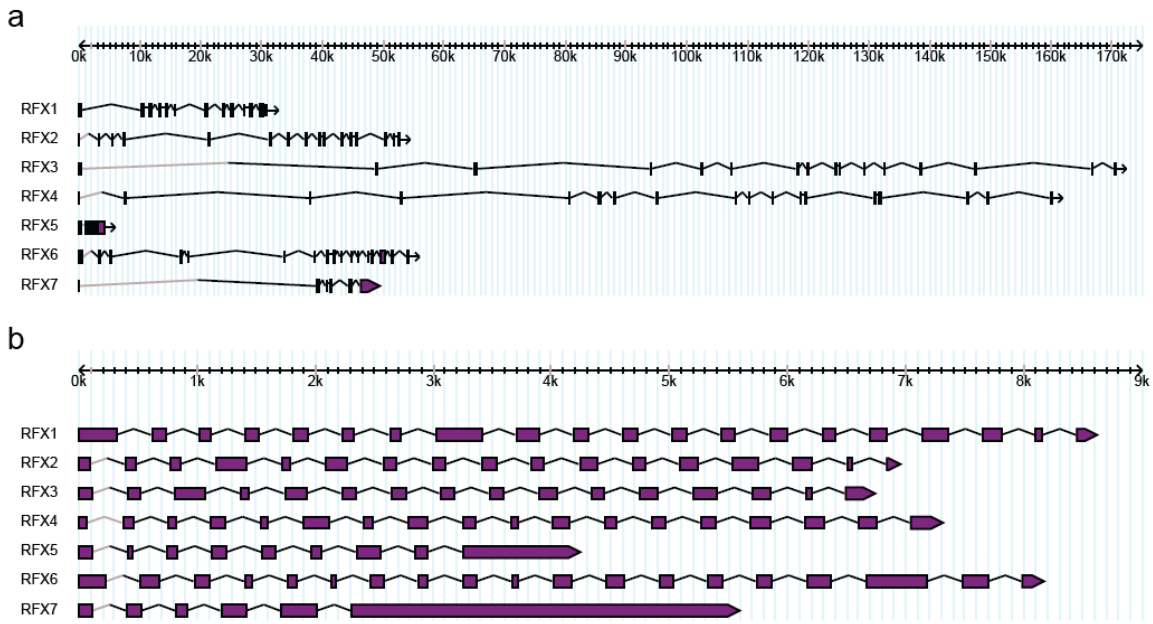


Figure IIH Gene models of human RFX genes, including RFX1-5 and newly identified RFX6-7.

(A) Exon-intron structures of human RFX genes. Exons are represented using boxes, while introns are represented using lines. Both exons and introns shown in this panel are proportional to their real lengths. (B) Illustrations of exon-intron structures of human RFX-genes. In this panels, while exons are proportional to their real lengths, for better visualization, introns are represented using lines of same lengths, regardless of their real lengths.

Table IIA: Gene names and Protein ID of mammalian RFX genes. Data taken from Ensembl (http://www.ensembl.org/)

Species	Gene Name	Accession Number (RefSeq DNA)	Protein id	Genomic Coordinates				Protein length	Number of isoforms	Number of exons
				chr	start	end	strand			
Humans (*Homo sapiens*)	Rfx1	NM_002918	ENSP00000254325	19	13933353	13978097	-1	979	1	21
	Rfx2	NM_000635	ENSP00000306335	19	5944175	6061554	-1	723	2	18
	Rfx3	NM_134428	ENSP00000371434	9	3208297	3515983	-1	749	8	18
	Rfx4	NM_213594	ENSP00000350552	12	105501163	105680710	1	744	4	18
	Rfx5	NM_000449	ENSP00000357864	1	149581060	149586457	-1	616	3	11
	Rfx6	NM_173560	ENSP00000332208	6	117305068	117351384	1	928	2	19
	Rfx7	NM_022841	ENSP00000373793	15	54166958	54222377	-1	1281	1	7
Chimpanzee (*Pan troglodytes*)	Rfx1	XM_524133	ENSPTRP00000018049	19	14361897	14409186	-1	979	1	22
	Rfx2	XM_512310	ENSPTRP00000017613	19	6101537	6159032	-1	699	2	17
	Rfx3	n/a	ENSPTRP00000043155	9	3260630	3528579	-1	715	5	18
	Rfx4	XP_001161826	ENSPTRP00000040381	12	107770229	107951773	1	735	3	18
	Rfx5	XM_001171715	ENSPTRP00000002192	1	130402119	130408780	-1	616	1	11
	Rfx6	XM_527584	ENSPTRP00000031653	6	118855615	118910695	1	932	1	19
	Rfx7	XM_001171418	ENSPTRP00000012142	15	53515170	53604052	-1	1281	1	9
Dog (*Canis familiaris*)	Rfx1	n/a	ENSCAFP00000024289	20	51497212	51519588	1	954	1	24
	Rfx2	XM_533937	ENSCAFP00000027631	20	57014252	57054444	1	727	3	17
	Rfx3	XM_533540	ENSCAFP00000003031	1	94845887	95002161	-1	755	1	16
	Rfx4	XM_845927	ENSCAFP00000002647	10	35119527	35257965	-1	745	3	18
	Rfx5	XM_540315	ENSCAFP00000018636	17	63473796	63477569	-1	619	1	9
	Rfx6	XM_541213	ENSCAFP00000001307	1	60291581	60346011	1	923	1	21
	Rfx7	XM_544696	ENSCAFP00000023611	30	24303138	24347785	-1	1365	2	6
Monkey (*Macaca mulatta*)	Rfx1	n/a	ENSMMUP00000024954	19	13655404	13685077	-1	738	1	16
	Rfx2	n/a	ENSMMUP00000015695	19	5896888	5955543	-1	726	2	17
	Rfx3	n/a	ENSMMUP00000001093	15	73743838	74012735	1	749	6	18
	Rfx4	n/a	ENSMMUP00000029123	11	107675355	107834734	1	747	2	18
	Rfx5	n/a	ENSMMUP00000017303	1	129774114	129778573	-1	620	1	9
	Rfx6	n/a	ENSMMUP00000022612	4	147022168	147078490	-1	932	1	19
	Rfx7	n/a	ENSMMUP00000033587	7	34408030	34456302	-1	1364	2	6
Mouse	Rfx1	NM_009055	ENSMUSP00000005600	8	86956971	86987107	1	963	1	21

(*Mus musculus*)	Rfx2	n/a	ENSMUSP00000002444	17	56461024	56516132	-1	717	2	18
	Rfx3	NM_011265	ENSMUSP00000038760	19	27834136	28077137	-1	749	3	17
	Rfx4	NM_001024918	ENSMUSP00000051107	10	84185847	84336335	1	735	3	18
	Rfx5	NM_017395	ENSMUSP00000029772	3	95039605	95046752	1	658	6	10
	Rfx6	n/a	ENSMUSP00000020054	10	51366225	51418218	1	931	3	19
	Rfx7	NM_001033536	ENSMUSP00000091338	9	72330807	72413719	1	1269	1	6
Rat (*Rattus norvegicus*)	Rfx1	NM_001105944	ENSRNOP00000008104	19	25745412	25776701	-1	964	1	21
	Rfx3	n/a	ENSRNOP00000019515	1	231342761	231601324	-1	749	5	17
	Rfx4	XM_576205	ENSRNOP00000056437	7	20977074	21115616	-1	725	3	17
	Rfx5	NM_001107694	ENSRNOP00000028531	2	189856029	189860951	1	652	1	12
	Rfx6	NM_001106388	ENSRNOP00000029472	20	30335024	30390585	1	931	2	19
	Rfx7	XM_001053787	ENSRNOP00000008644	8	77120962	77207044	1	1268	2	7

Table IIB - Description of PCR programs

PCR Program Name	PCR Program			
	Temperature	Time	Repeat	Notes
Phusion 58	98°C	30 sec.	1x	
	98°C 58°C 72°C	10 sec. 30 sec. 3.5 min.	35x	
	72°C	10 min.	1x	
	4°C	∞		
Stitch 1	98°C	1 min.	1x	
	98°C 56°C 72°C	5 sec. 15 sec. 2.5 min.	20x	
	98°C 56°C 72°C	5 sec. 15 sec. 2.5 min.	11x	↑ 10 sec/repeat
	72°C	5 min.	1x	
	4°C	∞		
Stitch 2	98°C	1 min.	1x	
	98°C 56°C 72°C	5 sec. 3.5 min. 2.5 min.	20x	
	98°C 56°C 72°C	5 sec. 3.5 min. 2.5 min.	11x	↑ 10 sec/repeat
	72°C	5 min.	1x	
	4°C	∞		
Htaq	94°C	10 min.	1x	
	94°C 58°C 72°C	30 sec. 30 sec. 3.5 min.	30x	
	72°C	10 min.	1x	
	4°C	∞		
Lysis 60	60°C 95°C 4°C	1 hr 15 min. ∞	1x 1x	

REFERENCE LIST

Aftab, S., Semenec, L., Chu, J.S., and Chen, N. (2008). Identification and characterization of novel human tissue-specific RFX transcription factors. BMC evolutionary biology *8*, 226.

Ainsworth, C. (2007). Cilia: tails of the unexpected. Nature *448*, 638-641.

Ait-Lounis, A., Baas, D., Barras, E., Benadiba, C., Charollais, A., Nlend Nlend, R., Liegeois, D., Meda, P., Durand, B., and Reith, W. (2007). Novel function of the ciliogenic transcription factor RFX3 in development of the endocrine pancreas. Diabetes *56*, 950-959.

Altun, Z.F., and Hall, D.H. (2002-2006). WormAtlas: ed.s.

Ansley, S.J., Badano, J.L., Blacque, O.E., Hill, J., Hoskins, B.E., Leitch, C.C., Kim, J.C., Ross, A.J., Eichers, E.R., Teslovich, T.M., Mah, A.K., Johnsen, R.C., Cavender, J.C., Lewis, R.A., Leroux, M.R., Beales, P.L., and Katsanis, N. (2003). Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. Nature *425*, 628-633.

Ashique, A.M., Choe, Y., Karlen, M., May, S.R., Phamluong, K., Solloway, M.J., Ericson, J., and Peterson, A.S. (2009). The Rfx4 transcription factor modulates Shh signaling by regional control of ciliogenesis. Science signaling *2*, ra70.

Ayoubi, T.A., and Van De Ven, W.J. (1996). Regulation of gene expression by alternative promoters. Faseb J *10*, 453-460.

Baas, D., Meiniel, A., Benadiba, C., Bonnafe, E., Meiniel, O., Reith, W., and Durand, B. (2006). A deficiency in RFX3 causes hydrocephalus associated with abnormal differentiation of ependymal cells. The European journal of neuroscience *24*, 1020-1030.

Badano, J.L., Mitsuma, N., Beales, P.L., and Katsanis, N. (2006). The ciliopathies: an emerging class of human genetic disorders. Annual review of genomics and human genetics *7*, 125-148.

Baek, D., Davis, C., Ewing, B., Gordon, D., and Green, P. (2007). Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. Genome research *17*, 145-155.

Baker, K., and Beales, P.L. (2009). Making sense of cilia in disease: the human ciliopathies. American journal of medical genetics *151C*, 281-295.

Bargmann, C.I., and Horvitz, H.R. (1991). Control of larval development by chemosensory neurons in Caenorhabditis elegans. Science *251*, 1243-1246.

Berezikov, E., Bargmann, C.I., and Plasterk, R.H. (2004). Homologous gene targeting in Caenorhabditis elegans by biolistic transformation. Nucleic acids research *32*, e40.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome research *14*, 988-995.

Birney, E., and Durbin, R. (2000). Using GeneWise in the Drosophila annotation experiment. Genome research *10*, 547-548.

Blackshear, P.J., Graves, J.P., Stumpo, D.J., Cobos, I., Rubenstein, J.L., and Zeldin, D.C. (2003). Graded phenotypic response to partial and complete deficiency of a brain-specific transcript variant of the winged helix transcription factor RFX4. Development (Cambridge, England) *130*, 4539-4552.

Blacque, O.E., Perens, E.A., Boroevich, K.A., Inglis, P.N., Li, C., Warner, A., Khattra, J., Holt, R.A., Ou, G., Mah, A.K., McKay, S.J., Huang, P., Swoboda, P., Jones, S.J., Marra, M.A., Baillie, D.L., Moerman, D.G., Shaham, S., and Leroux, M.R. (2005). Functional genomics of the cilium, a sensory organelle. Curr Biol *15*, 935-941.

Blacque, O.E., Reardon, M.J., Li, C., McCarthy, J., Mahjoub, M.R., Ansley, S.J., Badano, J.L., Mah, A.K., Beales, P.L., Davidson, W.S., Johnsen, R.C., Audeh, M., Plasterk, R.H., Baillie, D.L., Katsanis, N., Quarmby, L.M., Wicks, S.R., and Leroux, M.R. (2004). Loss of C. elegans BBS-7 and BBS-8 protein function results in cilia defects and compromised intraflagellar transport. Genes & development *18*, 1630-1642.

Blyth, H., and Ockenden, B.G. (1971). Polycystic disease of kidney and liver presenting in childhood. Journal of medical genetics *8*, 257-284.

Bockman, D.E., Buchler, M., and Beger, H.G. (1986). Structure and function of specialized cilia in the exocrine pancreas. Int J Pancreatol *1*, 21-28.

Bonnafe, E., Touka, M., AitLounis, A., Baas, D., Barras, E., Ucla, C., Moreau, A., Flamant, F., Dubruille, R., Couble, P., Collignon, J., Durand, B., and Reith, W. (2004). The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. Molecular and cellular biology *24*, 4417-4427.

Boulin, T., Etchberger, J.F., and Hobert, O. (2006). Reporter gene fusions. WormBook, 1-23.

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

Broverman, S., MacMorris, M., and Blumenthal, T. (1993). Alteration of Caenorhabditis elegans gene expression by targeted transformation. Proceedings of the National Academy of Sciences of the United States of America *90*, 4359-4363.

Bruford, E.A., Riise, R., Teague, P.W., Porter, K., Thomson, K.L., Moore, A.T., Jay, M., Warburg, M., Schinzel, A., Tommerup, N., Tornqvist, K., Rosenberg, T., Patton, M., Mansfield, D.C., and Wright, A.F. (1997). Linkage mapping in 29 Bardet-Biedl syndrome families confirms loci in chromosomal regions 11q13, 15q22.3-q23, and 16q21. Genomics *41*, 93-99.

Burket, C.T., Higgins, C.E., Hull, L.C., Berninsone, P.M., and Ryder, E.F. (2006). The C. elegans gene dig-1 encodes a giant member of the immunoglobulin superfamily that promotes fasciculation of neuronal processes. Developmental biology *299*, 193-205.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science (New York, N.Y *263*, 802-805.

Chen, N., Mah, A., Blacque, O.E., Chu, J., Phgora, K., Bakhoum, M.W., Newbury, C.R., Khattra, J., Chan, S., Go, A., Efimenko, E., Johnsen, R., Phirke, P., Swoboda, P., Marra, M., Moerman, D.G., Leroux, M.R., Baillie, D.L., and Stein, L.D. (2006). Identification of ciliary and ciliopathy genes in Caenorhabditis elegans through comparative genomics. Genome biology *7*, R126.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic acids research *31*, 3497-3500.

Chu, J.S., Baillie, D.L., and Chen, N. (2010). Convergent evolution of RFX transcription factors and ciliary genes predated the origin of metazoans. BMC evolutionary biology *10*.

Cole, B.R., Conley, S.B., and Stapleton, F.B. (1987). Polycystic kidney disease in the first year of life. The Journal of pediatrics *111*, 693-699.

Conradt, B., and 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*, 317-327.

Consortium, T.*C.e.*S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science (New York, N.Y *282*, 2012-2018.

Diseases, N.I.o.D.a.D.a.K. (November 2007). Polycystic Kidney Disease. In: Kidney and Urologic Diseases.

Dorn, A., Durand, B., Marfing, C., Le Meur, M., Benoist, C., and Mathis, D. (1987). Conserved major histocompatibility complex class II boxes--X and Y--are transcriptional control elements and specifically bind nuclear proteins. Proceedings of the National Academy of Sciences of the United States of America *84*, 6249-6253.

Dotzlaw, H., Alkhalaf, M., and Murphy, L.C. (1992). Characterization of estrogen receptor variant mRNAs from human breast cancers. Molecular endocrinology (Baltimore, Md *6*, 773-785.

Dubruille, R., Laurencon, A., Vandaele, C., Shishido, E., Coulon-Bublex, M., Swoboda, P., Couble, P., Kernan, M., and Durand, B. (2002). Drosophila regulatory factor X is necessary for ciliated sensory neuron differentiation. Development (Cambridge, England) *129*, 5487-5498.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). probabilistic models of protein and nucleic acids. In: Biological Sequence Analysis, ed. C.U. Press, Cambridge, UK, 356.

Efimenko, E., Blacque, O.E., Ou, G., Haycraft, C.J., Yoder, B.K., Scholey, J.M., Leroux, M.R., and Swoboda, P. (2006). Caenorhabditis elegans DYF-2, an orthologue of human WDR19, is a component of the intraflagellar transport machinery in sensory cilia. Molecular biology of the cell *17*, 4801-4811.

Efimenko, E., Bubb, K., Mak, H.Y., Holzman, T., Leroux, M.R., Ruvkun, G., Thomas, J.H., and Swoboda, P. (2005). Analysis of xbx genes in C. elegans. Development (Cambridge, England) *132*, 1923-1934.

El Zein, L., Ait-Lounis, A., Morle, L., Thomas, J., Chhin, B., Spassky, N., Reith, W., and Durand, B. (2009). RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies. Journal of cell science *122*, 3180-3189.

Emery, P., Durand, B., Mach, B., and Reith, W. (1996a). RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. Nucleic acids research *24*, 803-807.

Emery, P., Strubin, M., Hofmann, K., Bucher, P., Mach, B., and Reith, W. (1996b). A consensus motif in the RFX DNA binding domain and binding domain mutants with altered specificity. Molecular and cellular biology *16*, 4486-4494.

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

Flicek, P., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S.C., Eyre, T., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K.L., Howe, K., Johnson, N., Jenkinson, A., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A.J., Vogel, J., White, S., Wood, M., Birney, E., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X.M., Herrero, J., Hubbard, T.J., Kasprzyk, A., Proctor, G., Smith, J., Ureta-Vidal, A., and Searle, S. (2008). Ensembl 2008. Nucleic acids research *36*, D707-714.

Frokjaer-Jensen, C., Davis, M.W., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nat Genet *40*, 1375-1383.

Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature *403*, 916-921.

Golden, J.W., and Riddle, D.L. (1982). A pheromone influences larval development in the nematode Caenorhabditis elegans. Science (New York, N.Y *218*, 578-580.

Harris, T.W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W.J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., Fernandes, J., Han, M., Kishore, R., Lee, R., Muller, H.M., Nakamura, C., Ozersky, P., Petcherski, A., Rangarajan, A., Rogers, A., Schindelman, G., Schwarz, E.M., Tuli, M.A., Van Auken, K., Wang, D., Wang, X., Williams, G., Yook, K., Durbin, R., Stein, L.D., Spieth, J., and Sternberg, P.W. (2010). WormBase: a comprehensive resource for nematode research. Nucleic acids research *38*, D463-467.

Hedgecock, E.M., Culotti, J.G., Thomson, J.N., and Perkins, L.A. (1985). Axonal guidance mutants of Caenorhabditis elegans identified by filling sensory neurons with fluorescein dyes. Dev Biol *111*, 158-170.

Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. BioTechniques *32*, 728-730.

Hsieh, J., and Fire, A. (2000). Recognition and silencing of repeated DNA. Annu Rev Genet *34*, 187-204.

Huang, M., Zhou, Z., and Elledge, S.J. (1998). The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. Cell *94*, 595-605.

Inglis, P.N., Ou, G., Leroux, M.R., and Scholey, J.M. (2006). The sensory cilia of *Caenorhabditis elegans*. In: WormBook, ed. T.C.e.R. Community, WormBook doi/10.1895/wormbook.1.126.1, http://www.wormbook.org.

Kara, C.J., and Glimcher, L.H. (1991). Regulation of MHC class II gene transcription. Current opinion in immunology *3*, 16-21.

Katan-Khaykovich, Y., and Shaul, Y. (1998). RFX1, a single DNA-binding protein with a split dimerization domain, generates alternative complexes. The Journal of biological chemistry *273*, 24504-24512.

Katan-Khaykovich, Y., Spiegel, I., and Shaul, Y. (1999). The dimerization/repression domain of RFX1 is related to a conserved region of its yeast homologues Crt1 and Sak1: a new function for an ancient motif. Journal of molecular biology *294*, 121-137.

Katsanis, N., Lupski, J.R., and Beales, P.L. (2001). Exploring the molecular basis of Bardet-Biedl syndrome. Human molecular genetics *10*, 2293-2299.

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

Kimble, J., and Hirsh, D. (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. Developmental biology *70*, 396-417.

Kozminski, K.G., Johnson, K.A., Forscher, P., and Rosenbaum, J.L. (1993). A motility in the eukaryotic flagellum unrelated to flagellar beating. Proceedings of the National Academy of Sciences of the United States of America *90*, 5519-5523.

Kwitek-Black, A.E., Carmi, R., Duyk, G.M., Buetow, K.H., Elbedour, K., Parvari, R., Yandava, C.N., Stone, E.M., and Sheffield, V.C. (1993). Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic genetic heterogeneity. Nature genetics *5*, 392-396.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee,

C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Leitch, C.C., Zaghloul, N.A., Davis, E.E., Stoetzel, C., Diaz-Font, A., Rix, S., Alfadhel, M., Lewis, R.A., Eyaid, W., Banin, E., Dollfus, H., Beales, P.L., Badano, J.L., and Katsanis, N. (2008). Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. Nature genetics *40*, 443-448.

Lewis, J.A., Wu, C.H., Berg, H., and Levine, J.H. (1980). The genetics of levamisole resistance in the nematode Caenorhabditis elegans. Genetics *95*, 905-928.

Lis, J.T., Simon, J.A., and Sutton, C.A. (1983). New heat shock puffs and beta-galactosidase activity resulting from transformation of Drosophila with an hsp70-lacZ hybrid gene. Cell *35*, 403-410.

Ma, K., Zheng, S., and Zuo, Z. (2006). The transcription factor regulatory factor X1 increases the expression of neuronal glutamate transporter type 3. The Journal of biological chemistry *281*, 21250-21255.

Mach, B., Steimle, V., Martinez-Soria, E., and Reith, W. (1996). Regulation of MHC class II genes: lessons from a disease. Annual review of immunology *14*, 301-331.

Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol *6*, 386-398.

McGhee, J.D., and Krause, M.W. (1997). Transcription factors and transcriptional regulation: IV analysis of *C. elegans* promoters. In: *C. elegans* II, eds. D.L. Riddle, T. Blumenthal, B.J. Meyer, and J.R. Priess, New York: Cold Spring Harbor Press, 147-184.

Mello, C.C., Kramer, J.M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. The EMBO journal *10*, 3959-3970.

Morin, J.G., and Hastings, J.W. (1971). Biochemistry of the bioluminescence of colonial hydroids and other coelenterates. Journal of cellular physiology *77*, 305-312.

Morotomi-Yano, K., Yano, K., Saito, H., Sun, Z., Iwama, A., and Miki, Y. (2002). Human regulatory factor X 4 (RFX4) is a testis-specific dimeric DNA-binding protein that cooperates with other human RFX members. The Journal of biological chemistry *277*, 836-842.

Mossner, J., Logsdon, C.D., Goldfine, I.D., and Williams, J.A. (1984). Regulation of pancreatic acinar cell insulin receptors by insulin. The American journal of physiology *247*, G155-160.

Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. Methods in molecular biology (Clifton, N.J *396*, 59-70.

Mykytyn, K., and Sheffield, V.C. (2004). Establishing a connection between cilia and Bardet-Biedl Syndrome. Trends in molecular medicine *10*, 106-109.

Otsuki, K., Hayashi, Y., Kato, M., Yoshida, H., and Yamaguchi, M. (2004). Characterization of dRFX2, a novel RFX family protein in Drosophila. Nucleic acids research *32*, 5636-5648.

Ou, G., Blacque, O.E., Snow, J.J., Leroux, M.R., and Scholey, J.M. (2005). Functional coordination of intraflagellar transport motors. Nature *436*, 583-587.

Pazour, G.J. (2004). Intraflagellar transport and cilia-dependent renal disease: the ciliary hypothesis of polycystic kidney disease. J Am Soc Nephrol *15*, 2528-2536.

Pazour, G.J., and Rosenbaum, J.L. (2002). Intraflagellar transport and cilia-dependent diseases. Trends in cell biology *12*, 551-555.

Peden, E.M., and Barr, M.M. (2005). The KLP-6 kinesin is required for male mating behaviors and polycystin localization in Caenorhabditis elegans. Curr Biol *15*, 394-404.

Perkins, L.A., Hedgecock, E.M., Thomson, J.N., and Culotti, J.G. (1986). Mutant sensory cilia in the nematode Caenorhabditis elegans. Developmental biology *117*, 456-487.

Purvis, T.L., Hearn, T., Spalluto, C., Knorz, V.J., Hanley, K.P., Sanchez-Elsner, T., Hanley, N.A., and Wilson, D.I. (2010). Transcriptional regulation of the Alstrom syndrome gene ALMS1 by members of the RFX family and Sp1. Gene.

Qian, F., Zhen, F., Xu, J., Huang, M., Li, W., and Wen, Z. (2007). Distinct functions for different scl isoforms in zebrafish primitive and definitive hematopoiesis. PLoS biology *5*, e132.

Reith, W., Barras, E., Satola, S., Kobr, M., Reinhart, D., Sanchez, C.H., and Mach, B. (1989). Cloning of the major histocompatibility complex class II promoter binding protein affected in a hereditary defect in class II gene regulation. Proceedings of the National Academy of Sciences of the United States of America *86*, 4200-4204.

Reith, W., and Mach, B. (2001). The bare lymphocyte syndrome and the regulation of MHC expression. Annual review of immunology *19*, 331-373.

Reith, W., Satola, S., Sanchez, C.H., Amaldi, I., Lisowska-Grospierre, B., Griscelli, C., Hadam, M.R., and Mach, B. (1988). Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RF-X. Cell *53*, 897-906.

Reith, W., Ucla, C., Barras, E., Gaud, A., Durand, B., Herrero-Sanchez, C., Kobr, M., and Mach, B. (1994). RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins. Molecular and cellular biology *14*, 1230-1244.

Rhodes, D.R., Tomlins, S.A., Varambally, S., Mahavisno, V., Barrette, T., Kalyana-Sundaram, S., Ghosh, D., Pandey, A., and Chinnaiyan, A.M. (2005). Probabilistic model of the human protein-protein interaction network. Nature biotechnology *23*, 951-959.

Rodriguez-Tome, P. (1997). Searching the dbEST database. Methods in molecular biology (Clifton, N.J *69*, 269-283.

Roepman, R., and Wolfrum, U. (2007). Protein networks and complexes in photoreceptor cilia. Sub-cellular biochemistry *43*, 209-235.

Rosenbaum, J.L., and Witman, G.B. (2002). Intraflagellar transport. Nat Rev Mol Cell Biol *3*, 813-825.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P., and Vidal, M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.

Schafer, J.C., Haycraft, C.J., Thomas, J.H., Yoder, B.K., and Swoboda, P. (2003). XBX-1 encodes a dynein light intermediate chain required for retrograde intraflagellar transport and cilia assembly in Caenorhabditis elegans. Molecular biology of the cell *14*, 2057-2070.

Senti, G., Ezcurra, M., Lobner, J., Schafer, W.R., and Swoboda, P. (2009). Worms with a single functional sensory cilium generate proper neuron-specific behavioral output. Genetics *183*, 595-605, 591SI-593SI.

Senti, G., and Swoboda, P. (2008). Distinct isoforms of the RFX transcription factor DAF-19 regulate ciliogenesis and maintenance of synaptic activity. Molecular biology of the cell *19*, 5517-5528.

Sha, K., and Fire, A. (2005). Imprinting capacity of gamete lineages in Caenorhabditis elegans. Genetics *170*, 1633-1652.

Shah, A.S., Ben-Shahar, Y., Moninger, T.O., Kline, J.N., and Welsh, M.J. (2009). Motile cilia of human airway epithelia are chemosensory. Science (New York, N.Y *325*, 1131-1134.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nature biotechnology *22*, 1567-1572.

Sherman, P.A., Basta, P.V., and Ting, J.P. (1987). Upstream DNA sequences required for tissue-specific expression of the HLA-DR alpha gene. Proceedings of the National Academy of Sciences of the United States of America *84*, 4254-4258.

Siddiqui, A.S., Khattra, J., Delaney, A.D., Zhao, Y., Astell, C., Asano, J., Babakaiff, R., Barber, S., Beland, J., Bohacec, S., Brown-John, M., Chand, S., Charest, D., Charters, A.M., Cullum, R., Dhalla, N., Featherstone, R., Gerhard, D.S., Hoffman, B., Holt, R.A., Hou, J., Kuo, B.Y., Lee, L.L., Lee, S., Leung, D., Ma, K., Matsuo, C., Mayo, M., McDonald, H., Prabhu, A.L., Pandoh, P., Riggins, G.J., de Algara, T.R., Rupert, J.L., Smailus, D., Stott, J., Tsai, M., Varhol, R., Vrljicak, P., Wong, D., Wu, M.K., Xie, Y.Y., Yang, G., Zhang, I., Hirst, M., Jones, S.J., Helgason, C.D., Simpson, E.M., Hoodless, P.A., and Marra, M.A. (2005). A mouse atlas of gene expression: large-scale digital gene-expression profiles from precisely defined developing C57BL/6J mouse tissues and cells. Proceedings of the National Academy of Sciences of the United States of America *102*, 18485-18490.

Siegrist, C.A., Durand, B., Emery, P., David, E., Hearing, P., Mach, B., and Reith, W. (1993). RFX1 is identical to enhancer factor C and functions as a transactivator of the hepatitis B virus enhancer. Molecular and cellular biology *13*, 6375-6384.

Slauch, J.M., and Silhavy, T.J. (1991). Genetic fusions as experimental tools. Methods in enzymology *204*, 213-248.

Smith, S.B., Qu, H.Q., Taleb, N., Kishimoto, N.Y., Scheel, D.W., Lu, Y., Patch, A.M., Grabs, R., Wang, J., Lynn, F.C., Miyatsuka, T., Mitchell, J., Seerke, R., Desir, J., Eijnden, S.V., Abramowicz, M., Kacet, N., Weill, J., Renard, M.E., Gentile, M., Hansen, I., Dewar, K., Hattersley, A.T., Wang, R., Wilson, M.E., Johnson, J.D., Polychronakos, C., and German, M.S. (2010). Rfx6 directs islet formation and insulin production in mice and humans. Nature *463*, 775-780.

Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M.L., Mellitzer, G., and Gradwohl, G. (2010). Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development

Development (Cambridge, England) *137*, 203-212.

Steimle, V., Durand, B., Barras, E., Zufferey, M., Hadam, M.R., Mach, B., and Reith, W. (1995). A novel DNA-binding regulatory factor is mutated in primary MHC class II deficiency (bare lymphocyte syndrome). Genes & development *9*, 1021-1032.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Mardis, E.R., Marra, M.A., Miner, T.L., Minx, P., Mullikin, J.C., Plumb, R.W., Rogers, J., Schein, J.E., Sohrmann, M., Spieth, J., Stajich, J.E., Wei, C., Willey, D., Wilson, R.K., Durbin, R., and Waterston, R.H. (2003). The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS biology *1*, E45.

Stepanenko, O.V., Verkhusha, V.V., Kuznetsova, I.M., Uversky, V.N., and Turoverov, K.K. (2008). Fluorescent proteins as biomarkers and biosensors: throwing color lights on molecular and cellular processes. Current protein & peptide science *9*, 338-369.

Stinchcomb, D.T., Shaw, J.E., Carr, S.H., and Hirsh, D. (1985). Extrachromosomal DNA transformation of Caenorhabditis elegans. Molecular and cellular biology *5*, 3484-3496.

Sulston, J.E., Albertson, D.G., and Thomson, J.N. (1980). The Caenorhabditis elegans male: postembryonic development of nongonadal structures. Developmental biology *78*, 542-576.

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

Swoboda, P., Adler, H.T., and Thomas, J.H. (2000). The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans. Molecular cell *5*, 411-421.

Thellmann, M., Hatzold, J., and Conradt, B. (2003). The Snail-like CES-1 protein of C. elegans can block the expression of the BH3-only cell-death activator gene egl-1 by antagonizing the function of bHLH proteins. Development (Cambridge, England) *130*, 4057-4071.

Tobin, J.L., and Beales, P.L. (2007). Bardet-Biedl syndrome: beyond the cilium. Pediatric nephrology (Berlin, Germany) *22*, 926-936.

Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A.C., and Vagner, S. (2003). Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biology of the cell / under the auspices of the European Cell Biology Organization *95*, 169-178.

Vandaele, C., Coulon-Bublex, M., Couble, P., and Durand, B. (2001). Drosophila regulatory factor X is an embryonic type I sensory neuron marker also expressed in spermatids and in the brain of Drosophila. Mechanisms of development *103*, 159-162.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C.,

Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The sequence of the human genome. Science (New York, N.Y *291*, 1304-1351.

Villard, J., Peretti, M., Masternak, K., Barras, E., Caretti, G., Mantovani, R., and Reith, W. (2000). A functionally essential domain of RFX5 mediates activation of major histocompatibility complex class II promoters by promoting cooperative binding between RFX and NF-Y. Molecular and cellular biology *20*, 3364-3376.

Walt, H., and Hedinger, C. (1983). Motile components in spermatids as related to transport of spermatids and spermatozoa. Andrologia *15*, 34-39.

Ward, S., Thomson, N., White, J.G., and Brenner, S. (1975). Electron microscopical reconstruction of the anterior sensory anatomy of the nematode Caenorhabditis elegans.?2UU. J Comp Neurol *160*, 313-337.

Wenick, A.S., and Hobert, O. (2004). Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in C. elegans. Developmental cell *6*, 757-770.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Wolberger, C., and Campbell, R. (2000). New perch for the winged helix. Nature structural biology *7*, 261-262.

Wolfe, S.A., van Wert, J., and Grimes, S.R. (2006). Transcription factor RFX2 is abundant in rat testis and enriched in nuclei of primary spermatocytes where it appears to be required for transcription of the testis-specific histone H1t gene. Journal of cellular biochemistry *99*, 735-746.

Wu, S.Y., and McLeod, M. (1995). The sak1+ gene of Schizosaccharomyces pombe encodes an RFX family DNA-binding protein that positively regulates cyclic AMP-dependent protein kinase-mediated exit from the mitotic cell cycle. Molecular and cellular biology *15*, 1479-1488.

Yu, H., Pretot, R.F., Burglin, T.R., and Sternberg, P.W. (2003). Distinct roles of transcription factors EGL-46 and DAF-19 in specifying the functionality of a polycystin-expressing sensory neuron necessary for C. elegans male vulva location behavior. Development (Cambridge, England) *130*, 5217-5227.