TRANSCRIPTIONAL REGULATORY LOGIC OF CILIUM FORMATION IN *C. ELEGANS*

Doctoral Thesis by Rebeca Brocal Ruiz

Thesis Supervisor: Dr. Nuria Flames Bonilla

University Advisor: Dr. María Desamparados Pascual-Ahuir Giner

Doctoral Programme in Biotechnology

Universitat Politècnica de València

Instituto de Biomedicina de Valencia

Consejo Superior de Investigaciones Científicas (IBV-CSIC)

November 2021

To my beloved ones.

You have evolved from worm to man,

but much within you is still worm.

Thus Spoke Zarathustra - Friedrich Nietzsche

1883-1885

SUMMARY

Cilia are complex evolutionary conserved eukaryotic structures that, projecting from cell surfaces, perform a variety of biological roles. Cilia are traditionally classified into motile or sensory and hundreds of proteins take part in their composition. This set of genes coding for ciliary components is known as the ciliome. Mutations in the ciliome underlie an ever-growing group of highly pleiotropic multisystemic diseases globally termed as ciliopathies. These diseases are characterized, among other symptoms, by mental retardation, sensory defects and/or metabolic disorders. Despite an estimated 1 in 1,000 people affected by these diseases, the molecular bases of the ciliopathies are still poorly understood.

Proper cilium assembly and functionality requires the tightly co-regulated expression of ciliary components; however, little is known about the regulatory logic controlling ciliome transcription. Most ciliome genes are shared between motile and sensory cilia. RFX transcription factors (TFs) have an evolutionarily conserved role in the transcriptional regulation of both motile and sensory ciliome. In vertebrates, transcription of motile ciliome is also directly regulated by FoxJ1, a Forkhead (FKH) TF. However, to date, TFs working together with RFX in the transcription of the sensory ciliome are unknown in any organism.

In this work, we have identified FKH-8, a FKH TF, as a terminal selector of the sensory ciliome in *C. elegans*. *fkh-8* is consistently expressed within the sixty ciliated sensory neurons of *C. elegans*, it binds the regulatory regions of the sensory ciliome genes, it is also required for correct ciliome gene expression and acts synergistically with the known master regulator of the ciliogenesis DAF-19/RFX. Accordingly, *fkh-8* mutants display a wide range of behavioural defects in a plethora of sensory mediated paradigms, including olfaction, gustation, and mechano-sensation.

Thus, we have identified, for the first time, a TF that acts together with RFX TFs in the direct regulation of the sensory ciliome. Moreover, our results, together with previous work, show that FKH and RFX TFs act together in the regulation of both motile and sensory cilia, suggesting this regulatory logic could be an ancient trait pre-dating functional sub-specialization of cilia. Finally, we hope our results could help better understand the biological basis of orphan ciliopathies.

Resumen

Los cilios son estructuras eucariotas complejas conservadas evolutivamente que, proyectando desde la superficie de las células, desempeñan un gran número de funciones biológicas. Los cilios se clasifican tradicionalmente en móviles o sensoriales y en su composición intervienen cientos de proteínas. Este conjunto de genes que codifican para los componentes ciliares se conoce como cilioma. Las mutaciones en el cilioma subyacen a un grupo cada vez mayor de enfermedades multisistémicas altamente pleiotrópicas denominadas globalmente como ciliopatías. Estas enfermedades se caracterizan, entre otros síntomas, por retraso mental, defectos sensoriales y/o trastornos metabólicos. A pesar de que se estima que 1 de cada 1.000 personas está afectada por estas enfermedades, las bases moleculares de las ciliopatías son todavía poco conocidas.

El adecuado ensamblaje y funcionalidad del cilio requieren de la expresión estrechamente coordinada de los componentes del cilio; sin embargo, se sabe poco sobre la lógica reguladora que controla la transcripción del cilioma. La mayoría de los genes del cilioma son compartidos tanto por cilios móviles como sensoriales. Los factores de transcripción (FTs) de la familia RFX tienen un papel evolutivamente conservado en la regulación transcripcional del cilioma tanto móvil como sensorial. En los vertebrados, la transcripción del cilioma móvil también está regulada directamente por FoxJ1, un FT de la familia *forkhead* (FKH). Sin embargo, hasta la fecha, se desconocen los FTs que actúan junto a RFX en la transcripción del cilioma sensorial en cualquier organismo.

En este trabajo, hemos identificado a FKH-8, un FT de la familia FKH, como selector terminal del cilioma sensorial de *C. elegans*. *fkh-8* se expresa de forma consistente en las sesenta neuronas sensoriales ciliadas de *C. elegans*, se une a las regiones reguladoras de los genes del cilioma sensorial, también es necesario para la correcta expresión de los genes del cilioma y actúa de forma sinérgica con el conocido regulador maestro de la ciliogénesis DAF-19/RFX. En consecuencia, los mutantes para *fkh-8* muestran una amplia gama de defectos de comportamiento en una plétora de paradigmas sensoriales, incluyendo la olfacción, la gustación y la mecano-sensación.

Así, hemos identificado, por primera vez, un FT que actúa junto con los FTs de la familia RFX en la regulación directa del cilioma sensorial. Además, nuestros resultados, junto con trabajos anteriores, muestran que los FTs FKH y RFX actúan conjuntamente en la regulación de los cilios tanto móviles como sensoriales, lo que sugiere que esta lógica reguladora podría ser un rasgo evolutivo antiguo anterior a la subespecialización funcional de los cilios. Finalmente, esperamos que los resultados de nuestro trabajo ayuden a entender mejor las bases biológicas de las ciliopatías huérfanas.

RESUM

Els cilis són estructures eucariotes complexes conservades evolutivament que, projectant des de la superfície de les cèl·lules, exerceixen un gran nombre de funcions biològiques. Els cilis es classifiquen tradicionalment en mòbils o sensorials i en la seua composició intervenen centenars de proteïnes. Aquest conjunt de gens que codifiquen per als components ciliars es coneix com el cilioma. Les mutacions en el cilioma subjauen a un grup cada vegada major de malalties multisistèmiques altament pleiotròpiques denominades globalment com ciliopaties. Aquestes malalties es caracteritzen, entre altres símptomes, per retard mental, defectes sensorials i/o trastorns metabòlics. A pesar que s'estima que 1 de cada 1.000 persones està afectada per aquestes malalties, les bases moleculars de les ciliopaties són encara poc conegudes.

L'adequat assemblatge i funcionalitat del cili requereixen de l'expressió estretament coordinada dels components del cili; no obstant això, se sap poc sobre la lògica reguladora que controla la transcripció del cilioma. La majoria dels gens del cilioma són compartits tant per cilis mòbils com sensorials. Els factors de transcripció (FTs) de la família RFX tenen un paper evolutivament conservat en la regulació transcripcional del cilioma tant mòbil com sensorial. En els vertebrats, la transcripció del cilioma mòbil també està regulada directament per FoxJ1, un FT de la família *forkhead* (FKH). No obstant això, fins hui, es desconeixen els FTs que actuen al costat de RFX en la transcripció del cilioma sensorial en qualsevol organisme.

En aquest treball, hem identificat a FKH-8, un FT de la família FKH, com a selector terminal del cilioma sensorial de *C. elegans*. *fkh-8* s'expressa de manera consistent en les seixanta neurones sensorials ciliades de *C. elegans*, s'uneix a les regions reguladores dels gens del cilioma sensorial, també és necessari per a la correcta expressió dels gens del cilioma i actua de manera sinèrgica amb el conegut regulador mestre de la ciliogènesi DAF-19/RFX. En conseqüència, els mutants per a *fkh-8* mostren una àmplia gamma de defectes de comportament en una plètora de paradigmes sensorials, incloent la olfacció, la gustació i la mecano-sensació.

Així, hem identificat, per primera vegada, un FT que actua juntament amb els FTs de la família RFX en la regulació directa del cilioma sensorial. A més, els nostres resultats, juntament amb treballs anteriors, mostren que els FTs FKH i RFX actuen conjuntament en la regulació dels cilis tant mòbils com sensorials, la qual cosa suggereix que aquesta lògica reguladora podria ser un tret evolutiu antic anterior a la subespecialització funcional dels cilis. Finalment, esperem que els resultats del nostre treball ajuden a entendre millor les bases biològiques de les ciliopaties òrfenes.

CONTENTS

INDEX OF FIGURES, TABLES & ANNEXES

ABBREVIATIONS & ACRONYMS

°C	Celsius degrees
μg	microgram/s
μL	microlitre/s
μm	micrometre/s
bp	base pair/s
CAP	catabolite activator protein
cDNA	Coding DNA
CGC	Caenorhabditis Genetics Center
ChIP-seq	chromatin immunoprecipitation sequencing
CI	chemotaxis index
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
ddH$_2$O	doble-distilled water
DEPC	diethyl pyrocarbonate treated doble distilled water
DIC	Differential interference contrast
DiD	1,1'-dioctadecyl-3,3,3',3'-tetramethylindodicarbocyanine, 4-chlorobenzenesulfonate salt
DNA	deoxyribonucleic acid
F1	Filial 1 (first generation)
F2	Filial 2 (second generation)
FITC	5-fluorescein isothiocyanate
FKH	forkhead (family of transcription factors)
FW	forward
g	gram
GFP	green fluorescent protein
IDE	integrated development environment
KEGG	Kyoto Encyclopedia of Genes and Genomes
M	molar

MCS	multicloning site
mg	milligram
MGC	Mammalian Gene Collection
mL	millilitres
mm	millimetres
mM	millimolar
NBRP	National BioResource Project
NLS	nuclear localization signal
nm	nanometres
PCR	polymerase chain reaction
PWM	position weight matrix
RFX	regulatory factor X (family of transcription factors)
RI	response index
RNA	ribonucleic acid
RNA-seq	RNA sequencing
rpm	revolutions per minute
RV	reverse
sci-RNA-seq	single-cell combinatorial indexing RNA sequencing
SDS	sodium dodecyl sulfate
ssODN	single stranded oligodeoxynucleotide
SV40	simian virus 40
tracrRNA	trans-activating crRNA
t-SNE	t-distributed stochastic neighbour embedding
URL	uniform resource locator
UTR	untranslated region
V/V	volume/volume
w/V	weight/volume
WT	wild type

Introduction

Achieving cell diversity

Regardless of its origin, the formation of a new pluricellular organism in the animal kingdom involves an intricate flow of continuous cell-state transitions that finally leads to the generation of different cell types. How this cell diversity is achieved starting from a single cell has long fascinated the human mind, far beyond the fields of raw biology. In the light of modern science, "the body plan of an animal, and hence its exact mode of development, is a property of its species and is thus encoded in the genome" (Davidson 2010). Since each cell from a pluricellular animal inherits a copy of the same genome, the extrapolation of the former statement implies that the display of the final cell diversity is nothing but the reflection of differential gene expression in a cell-type specific manner.

An overview of the eukaryotic gene expression.

From the most elementary point of view, the first step towards achieving gene expression requires the transcriptional machinery to transcribe that particular gene. This process starts when the enzyme that catalyses the transcription of DNA to RNA, the RNA-polymerase II or RNAP II, is in charge of mRNA and microRNA transcription among others. It binds to DNA nearby the so-called transcription start site (TSS). This binding is influenced by the presence and activity of the general transcription factors (GTFs), proteins that, together with RNAP II, make up the polypeptide pre-initiation complex that primers transcription (Roeder 1996). The binding of such GTFs to specific motifs in the DNA

is typically constrained in a region spanning ±50 base pairs around the TSS; that region is usually termed the "core promoter".

trans-regulation of gene expression: transcription factors.

Although core promoters contain sufficient information to trigger the initiation of transcription under specific conditions, this process is further modulated by the action of sequence-specific DNA-binding proteins. These proteins, the transcription factors (TFs), can directly or indirectly impact either on the recruitment of RNAP II to the core promoter, the initiation of the transcription or the elongation of the nascent RNA (Andersson and Sandelin 2020). TFs recognize and bind to short, degenerated DNA motifs that usually range from 6 to 12 base pairs in length (Spitz and Furlong 2012). The collection of degenerated, yet similar sequences bound by a specific transcription factor can be represented, once aligned, using nucleotide-frequency derived models such as the Position Weight Matrix (PWMs) (Berg 1988; Inukai, Kock, and Bulyk 2017; Stormo and Zhao 2010). Graphically, the patterns in a set of aligned sequences can be portrayed as sequence logos (see **Figure I1.1)** in which letters represent the nucleotides and letter size in each position informs about its probability of occurrence (Gao, Liu, and Ruan 2017; Schneider and Stephens 1990).

Transcription factors are able to bind their corresponding binding sites as a result of the interactions produced between the DNA and specific protein domains TFs have (DNA binding domains, DBDs). Thus,

particular features within these DDBs confer specificity toward the recognized motifs present in the DNA (Yang and Ramsey 2015). DBDs exhibit variable degrees of sequence conservation and, accordingly, they are used to classify TFs in different families (Wingender, Schoeps, and Dönitz 2013). In general, TFs from the same TF family recognize similar TFBS. However, transcription is a complex process that differs from the simplified scenario described here (see **Figure I.2A**). Many different players are involved (see **Figure I1.2B & C**), some molecules bound in close proximity to the transcriptional start site, some binding thousands of kilobases away and interacting with the promoter due to chromatin looping, and all of them being affected by their own regulatory processes. In the case of several TFs, an additional layer of complexity is added due to the presence of more than one DNA binding domain that allow TFs to widen their DNA-binding repertoire, thus increasing the number of DNA motifs they can recognize (see **Figure I1.3**). Additional complexity arises from TFs-TFs interactions as homo or heteromultimers. Examples exist (reviewed in Siggers and Gordân 2014) in which multiprotein complexes formed by the interactions of different TFs (see **Figure I1.4**) exhibited novel DNA-binding specificities that were not predictable from the isolated behaviour of the individual proteins. Finally, TFs activity can be modulated by post-translational modifications that change their conformation, location or interacting partners and thus their DNA binding specificities.

Transcription Factor Binding Sites(TFBSs)

	1			5					10					15
1	G	T	G	T	G	T	G	T	G	T	G	C	A	T	G
2	A	A	A	C	G	C	G	C	C	T	G	T	G	G	G
3	C	G	G	A	G	C	G	A	T	T	G	T	T	A	C
4	G	T	G	T	G	T	G	G	G	T	G	G	A	T	T
5	G	A	G	A	G	T	G	T	G	T	G	T	A	T	C
...							...							
156	G	A	G	C	G	T	G	T	C	T	G	C	G	C	T

Sequence Logo
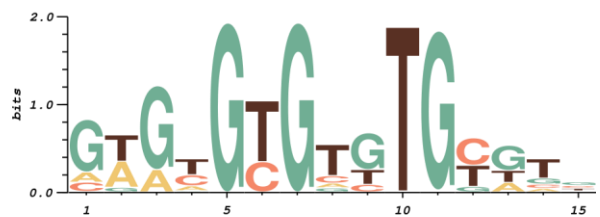
Position Frequency Matrix (PFM)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	22	71	33	23	0	0	0	15	0	1	0	0	32	17	17
C	20	0	0	52	0	52	0	25	20	0	1	79	0	20	43
G	113	13	123	6	156	0	156	14	97	0	154	16	84	28	56
T	1	72	0	75	0	104	0	102	39	155	1	61	40	91	40

Position Probability Matrix (PPM)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	0.14	0.46	0.21	0.15	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.00	0.21	0.11	0.11
C	0.13	0.00	0.00	0.33	0.00	0.33	0.00	0.16	0.13	0.00	0.01	0.51	0.00	0.13	0.28
G	0.72	0.08	0.79	0.04	1.00	0.00	1.00	0.09	0.62	0.00	0.99	0.10	0.54	0.18	0.36
T	0.01	0.46	0.00	0.48	0.00	0.67	0.00	0.65	0.25	0.99	0.01	0.39	0.26	0.58	0.26

Position Weight Matrix (PWM)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	-0.83	0.86	-0.24	-0.76	$-\infty$	$-\infty$	$-\infty$	-1.38	$-\infty$	-5.29	$-\infty$	$-\infty$	-0.29	-1.20	-1.20
C	-0.96	$-\infty$	$-\infty$	0.42	$-\infty$	0.42	$-\infty$	-0.64	-0.96	$-\infty$	-5.29	1.02	$-\infty$	-0.96	0.14
G	1.53	-1.58	1.66	-2.70	2.00	$-\infty$	2.00	-1.48	1.31	$-\infty$	1.98	-1.29	1.11	-0.48	0.52
T	-5.29	0.88	$-\infty$	0.94	$-\infty$	1.42	$-\infty$	1.39	0.00	1.99	-5.29	0.65	0.04	1.22	0.04

◄ Figure I1.1. Transcription factor binding sites can be represented in several ways.

Figure displays 5 different ways to portray the binding site for the *C. elegans* transcription factor *daf-12*. Real data from *daf-12* ChIP-seq experiment is used (see JASPAR entry MA0538.1). 156 binding site sequences were retrieved; upper left, 6 exemplary transcription factor binding sites (TFBSs) are shown. For the 156 *daf-12* binding sites, occurrence of each nucleotide in each of the 15 positions are summarized in a position frequency matrix (PFM). Considering the total number of events, a position probability matrix (PPM) can be built. Frequency data can be further modelled as a logarithmic ratio considering the probability of nucleotide occurrence (which is 1/4). A position weight matrix (PWM) is often expressed as the base 2 logarithm for the quotient between position probability and probability of nucleotide. For an adenine in position one, the score is calculated as $\log_2(0.14/0.25) = -0.83$. Notice that in this example no pseudo-counts are used to avoid the value of 0 in the PPM; hence minus infinite appears in the corresponding PWM entry. Application of Shannon's Information Theory (Shannon 1948) by (Schneider and Stephens 1990) upon nucleotides frequency allows for the creation of the sequence logo, a graphical representation informing about probability of nucleotide occurrence in each TFBS position

A)

B)

C)
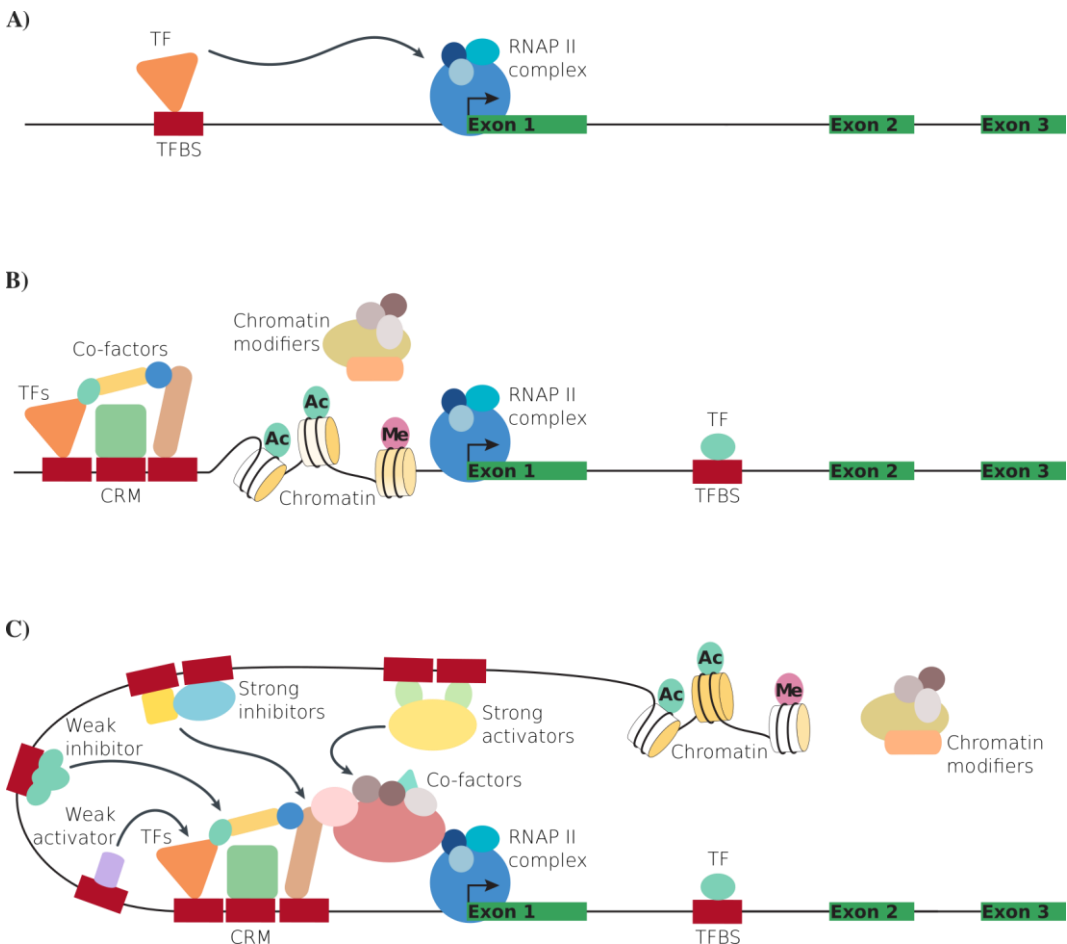
◄ **Figure I1.2. Transcription is a highly complex process controlled by several layers of regulation.**

Several layers of regulation acting over the transcription of a single gene, here portrayed by green boxes depicting 3 of its exons. **A)** Over-simplified view of transcription in which a single transcription factor (TF) triggers the activation of the RNAP II complex. **B)** Generalized view of transcription shows the main players acting over the transcriptional regulation of a single gene. Regulatory sequences appear upstream and downstream the transcriptional start site. Several TFs are needed to trigger transcription and an addition layer of regulation appears by the binding of co-factors. Chromatin modifiers controls accessibility to regulatory regions, imposing both positive and negative regulation. **C)** Complexity of the transcriptional process involve both *cis-* and *trans*-acting elements exerting their actions at close and long distances. DNA looping allows distal regions to interact with the transcriptional machinery. Final transcriptional rate of a gene results from the opposing contributions of inhibitor and activators. Ac: acetylation, Me: methylation. Adapted from (Nam and Reineke 2017).
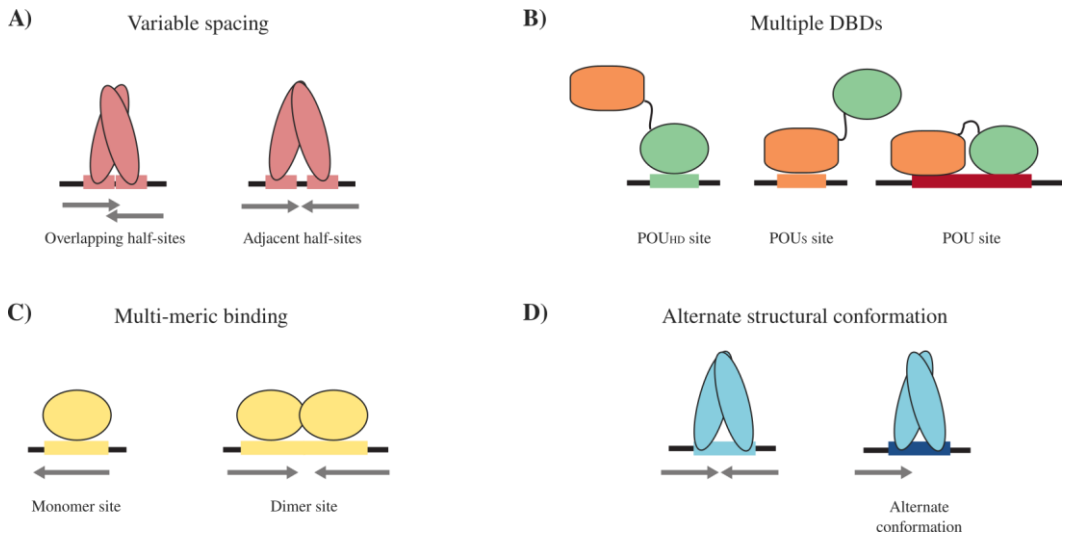
▲ **Figure I1.3. Transcription factors recognize different DNA binding sites via distinct structural dispositions.**

Several possibilities for TFs binding to DNA motifs are exemplified. Black thick lines represent DNA, coloured boxes depict different TFBSs. **A)** TFs dimers can bind to bipartite sites separated by a variable number of bases, either binding overlapping or adjacent TFBSs (illustrated through grey arrows). **B)** TFs can harbour different DNA binding domains, hence allowing the protein to bind distinct TFBSs. **C)** Multimerization further expand binding specificities of TFs. Several proteins have been shown to bind either as a monomer or as homo- or heterodimers. **D)** TFs with a single DNA binding domain can bind to different TFBSs by adopting an alternate structural conformation. Adapted from (Siggers and Gordân 2014).
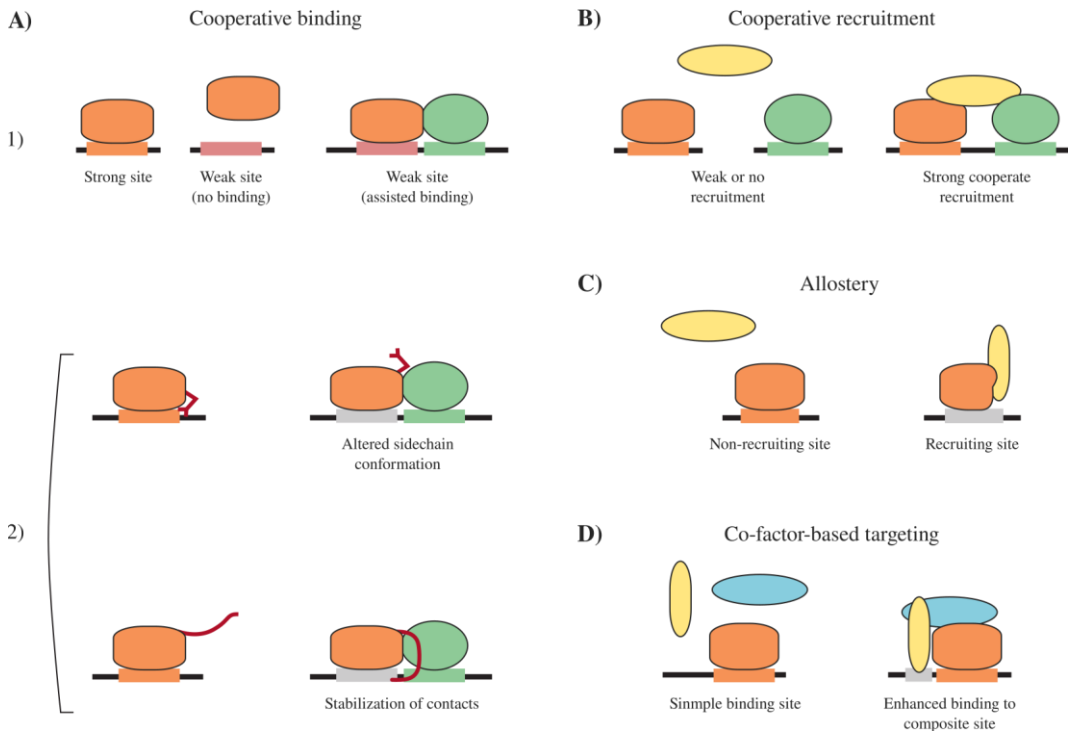
⤴ **Figure I1.4. Transcription factors form multi-protein complexes.**
Four mechanistic categories by which TFs form multi-protein complexes able to target novel DNA sites. Black thick lines represent DNA, coloured boxes depict different TFBSs. **A) 1)** A given TF shows strong affinity towards strong TFBSs (in orange) but is unable to bind the weaker ones (in pink). Cooperative binding stabilizes this weak interaction through the formation of a complex with another TF. **2)** Binding specificities (from orange to grey TFBSs) can be altered (top) or stabilized (bottom) through inter-protein interactions between TFs sidechains (in red). **B)** Isolated TFs are unable to recruit a given co-factor. Multiple TFs are needed to perform its cooperative recruitment. The genomic location of this co-factor is hence determined by the less frequent and coincident presence of multiple TFs. **C)** Through the allosteric control of cofactor recruitment, a TF bound to a given TFBS (in orange) is unable to recruit a specific cofactor while the same TF bound to a different TFBS (in grey) can, thus limiting cofactor recruitment to only a subset of the TFBSs. **D)** In contrast to the traditional view, recruited cofactors (here represented in yellow) have been shown to interact directly with DNA when being part of a larger, multi-protein complex. Through this mechanism, the non-DNA-binding cofactor (here represented in blue) mediate sequence-specific interactions that preferentially stabilize the binding to composite DNA sites containing specific auxiliary motifs (in grey). Adapted from (Siggers and Gordân 2014).

Since transcription is regulated by the action of TFs bound to DNA, TFBSs are considered the core building blocks of transcriptional regulation. Accordingly, important efforts have been made in the past decades to determine were in the genome TFs bind. Different *in vivo* and *in vitro* technologies have applied to characterize TFBSs (reviewed in Levo and Segal 2014). Commonly *in vivo* technologies employed to delineate TFs

occupancy are *chromatin immunoprecipitation followed by microarray* (ChIP–chip) or more recently, *ChIP followed by high-throughput sequencing* (ChIP–seq). Both technologies use antibodies to pull down specific TFs and the corresponding DNA-bound regions that are later identified by sequencing. Example extensions of these technologies include the use of *exonuclease trimming* following immunoprecipitation to increase the resolution of binding events being mapped (ChIP-exo) or the replacement of the initial steps of standard ChIP by a micrococcal nuclease digestion (MNase-seq) producing the so-called *occupied regions of genomes from affinity-purified naturally isolated chromatin* (ORGANIC) and more recently CUT&RUN and CUT&TAG modified versions of ChIP. Complementary techniques aimed to the identification of open chromatin regions at the genome-wide scale such as *DNase I hypersensitive site sequencing* (DNase-seq), the *assay for transposase-accessible chromatin using sequencing* (ATAC-seq) or the formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) have also been employed to delineate chromatin accessibility. A graphical summary for these *in vivo* techniques can be seen in **Figure I1.5.A.** Other than *in vivo* occupancy, a second approach use several *in vitro* technologies to measure affinity that chosen TFs exhibit towards many different short sequences, thus providing a more quantitative description for TFs specificities. Commonly used technologies include *protein binding microarrays* (PBMs) (see **Figure I1.5.B1**) or *high-throughput systematic evolution of ligands by exponential enrichment* (HT-SELEX) (see **Figure I1.5.B2**), among others. Independently of the approach or technology, the binding specificities collected by different studies have been gathered (and keep being gathered) in several TFBSs databases. Examples of such databases are TRANSFAC (Matys et al. 2006), CIS-BP (Weirauch et al. 2014), UniPROBE (Hume et al. 2015) or JASPAR (Fornes et al. 2020). Information collected by most of these databases is freely accessible, many through user-friendly on-line suites, and can be used to predict or analyse potential binding events both genome-wide or focused on putative regulatory sequences under analysis.

cis-regulation of gene expression: promoters, enhancers and other elements.

To exert their action, TFs need to recognize and bind to specific regions in the DNA. These loci, bound by several TFs, constitute functional regulatory regions that ultimately govern transcription. Although controversial, several definitions exist to classify them.

As previously stated, locations nearby the TSS in which RNAP II and GTFs bind and form the pre-initiation complex constitute the core promoter of a given gene. The fact that some TFs bind proximal but not within the core promoter defines a larger region of undefined length that engulf the core promoter and is often termed "proximal promoter" or just promoter. Since the size of promoters is arbitrarily defined, chances are that only some regions within their sequences have actual impact on transcription. The regions containing the biological information needed to drive expression can be identified through the so-called "promoter bashing". In this technique (see Andersson and Sandelin 2020), selected sub-regions of the promoter

are fused to a reporter gene and then reporter expression is analysed. Those sub-regions that are able to recapitulate the expression pattern observed for the whole promoter constitute the "minimal promoters".

Considering both motif type and directionality, a functional tripartition for the promoter sequences has been proposed (Lenhard, Sandelin, and Carninci 2012). This classification arises from the commonalities found in an exhaustive analysing performed on promoter sequences from both *D. melanogaster* and human (FitzGerald et al. 2006). Following Lenhard *et al* nomenclature, tissue-specific genes own Type I promoters with high enrichment for a TATA box (a T/A-rich sequence that is found upstream of the TSS) located at an appropriate distance from the so-called initiator element (which is a sequence pattern overlapping the TSS). Type II promoters are found in ubiquitous "housekeeping" genes and they are associated with either a DNA recognition element or a combination of novel TFBSs. Finally, developmentally regulated genes preferentially possess Type III promoters in which only an initiator element (or an initiator element plus a downstream promoter element) is (are) found. In mammalian, different promoters show different content of CpG islands (genomic sequences not depleted from CG dinucleotides

that are often found overlapping or located near the TSSs). Type I promoters tend to have a low content of CpG islands whereas Type II usually have only one short CpG island overlapping the TSS; by contrast, Type III promoters have several large CpG islands sometimes covering the whole gene (Akalin et al. 2009; Lenhard et al. 2012). Despite this classical view of promoter categories, it is not clear if specific TFs bind to specific types of promoters and how promoter activity is regulated.

Among the regulatory regions that locate at further distances from the TSS are the enhancers; segments of DNA that typically span a few hundred base pairs in length and act as recruiter platforms for TFs (Spitz and Furlong 2012). Enhancers usually locate in the non-coding genome within intergenic or intronic regions (Shlyueva, Stampfel, and Stark 2014), although examples of enhancers in exonic sequences have also been found (Birnbaum et al. 2014). A differential property of enhancers, one that has been used as the "gold standard" in their functional testing, is their capability to activate or enhance

transcription from a core promoter (Shlyueva et al. 2014). This activity is independent of the sequence context surrounding the enhancer or its orientation and can thus be assessed in heterologous reporter systems (see Andersson and Sandelin 2020). Bashing of regions distal to the TSS is also used to identify minimal enhancer regions driving specific gene expression.

In humans, relative contributions of promoter and enhancers to final gene regulation have been shown to follow the functional tripartition proposed for promoter sequences (Ernst et al. 2011). This way, Type I promoters found in tissue-specific genes tend to be regulated by proximal *cis*-regulatory modules (enhancers) formed by the clustering of TFBSs. Type II promoters from ubiquitous "housekeeping" genes are generally found in an active state and have few enhancers nearby. Finally, expression of developmental genes with Type III promoters is controlled at both promoter and enhancer levels since this class exhibits the highest number of associated enhancers and both active and inactive promoter states are found.

Classical distinction between promoter and enhancers has tacitly assumed that these regulatory elements belong essentially to two different functional classes. However, the advent of diverse *omics* technologies has shown that enhancers and promoters share, indeed, several properties and functions. As reviewed in (Andersson and Sandelin 2020), their chromatin and sequence architectures are highly similar, enhancer activity has been reporter for several promoters and, conversely, active enhancers have been shown to trigger transcription at their boundaries,

thus acting as promoters themselves. Accounting for this growing body of evidence, Andersson and Sandelin have proposed an updated model for regulatory elements in which any nucleosome-depleted region (a common feature found in promoters and enhancers due to the presence of binding sites for TFs, which exclude the nucleosomes) harbours the potential to act as a regulatory element. From this shared baseline, no enhancer/promoter dichotomy is imposed; rather, different features associate the region with a higher promoter or enhancer potential. Many of these features relate to the nature of TFBSs present in those regions. Accordingly, promoter potential associates with regions in which a high diversity, high overlapping and high coverage of TFBSs exist. Moreover, TFBSs for TFs with a clear promoter-driving effect appear, these including CREB, SP1, ETS and RFX TFs. Since these TFBSs tend to have a high GC content, potential promoter regions show a higher CpG density than those with enhancer potential. Furthermore, enhancers show fewer TFBSs per region that are often rich in AT sequences and might also be slightly imperfect. Additionally, in contrast with the more rigid configuration of promoters, many diverse configurations of TFs or recruited co-factors may lead to strong enhancer activity. Accounting for this model, the enhancer potential may correspond to the "base state" of the regulatory elements, explaining why enhancer regions are seldom highly evolutionarily conserved whereas promoter regions are often evolutionarily conserved.

Enhancers have demonstrated to act independent of distance and orientation, being able to regulate their target genes either

located in the same or in different chromosomes (Sanyal et al. 2012). However, it is known that internal architecture within enhancers sequences can actually play a very important role in their outcoming functionality. This concept of regulatory architecture accounts for the number (multiplicity), affinity, identity and arrangements (considering both for position and orientation) for all the TFBSs allocated in a given enhancer sequence (see **Figure I1.6**). In theory, the constrains that emerge due to this internal architecture should allow to envision for the existence of the so-called "grammatical rules" of regulatory sequences (Levo and Segal 2014). However, as seen in **Figure I1.6.D**, experimental evidence observed when analysing the influence of position and orientation of TFBSs produce very different outcomes, and so far it has been difficult to identify specific grammatic rules underlying enhancer function. This has led to the proposal of three different model (Spitz and Furlong 2012) explaining enhancer functionality (see **Figure I1.7**). One first model, the "enhanceosome", portrays a rigid distribution in position, spacing and orientation of TFBSs. Functional output is achieved only if all TFs bind the DNA at once, since they are meant to assemble a complex that ultimately demands a fixed protein interface, thus there is a unique grammar underlying enhancer function and only this particular disposition is functional. Although the enhanceosome model is based on functional data of the architecture of the IFN enhancer, it seems not to be the prevalent mode of action. A second, highly flexible scenario, is represented by the "billboard" model. TFBSs composition may be fixed, but no internal grammar rules among TFBSs are needed. As

in the case of the enhanceosome, binding of TFs to the DNA may follow a cooperative mechanism in which protein-protein interactions exist. Indirect mechanism of TFs cooperativity can also be found for the billboard model since the binding of a TF can facilitate the binding of another one. In contrast to the enhanceosome, mechanisms of additivity are found for this model, thus allowing for the modulation of enhancers functionality but also implying that the presence of all TFs in not necessary to achieve a functional output. A third model is the "TF collective" model. Binding to DNA follows a cooperative mechanism but, in contrast to the previous two models, TFs themselves can act as scaffolds for subsequent events of binding. As a direct consequence, greater variability for interactions among TFs exist. In this model, flexibility is allowed for both composition and position of TFBSs as well as their orientation, however, in contrast to the billboard model, due to complex protein-protein interactions, this flexibility is not unlimited, and some arrangements are predicted to have restrictions in distance and orientation, thus grammar rules are proposed to exist although TF action versatility makes it difficult to identify such grammatic constrains. For enhancers following this model, functional output may need the binding of most TFs, although it is not clear if there is a requirement for the binding of all of them.

Biological relevance of enhancer sequence composition is manifested by the plethora of phenotypes generated by enhancer mutations. Indeed, most genetic associations to human diseases locate in the non-coding genome and are predicted to have regulatory effects (see Mathelier, Shi, and

Wasserman 2015). Due to the capital role that enhancers play in transcriptional regulation, multiples efforts have been made in the past decades to predict, identify, and test these regulatory sequences. The different approaches (reviewed in Shlyueva et al. 2014) that have been used included the analyses of TFBSs, the *in vivo* binding sites of both TFs and cofactors, the presence of distinctive chromatin features (such as acetylation of lysine 27 on histone 3) as well as massively parallel enhancer assays in which thousands of synthetic sequences are tested for their enhancer activity *in vivo*. Undoubtedly, all these strategies have benefited enormously (or even were made possible) by the advent of deep sequencing.
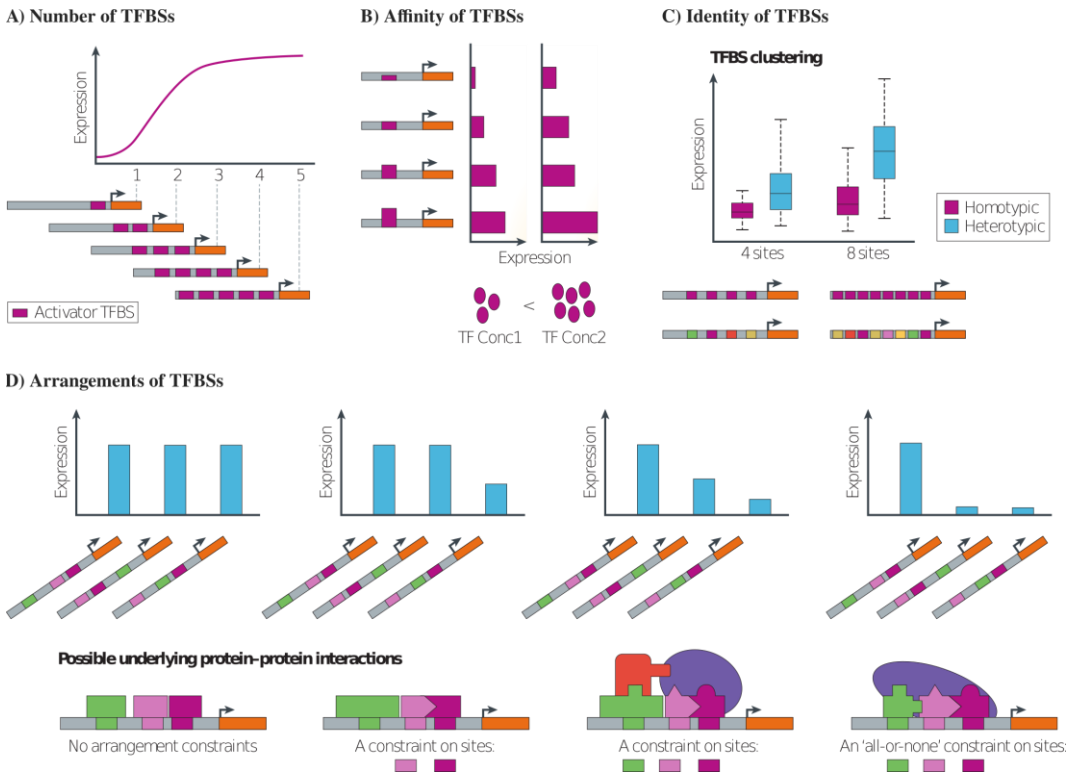
▲ **Figure I1.6. Internal transcriptional architecture of enhancers affects their functionality.**
A) In this example, a TF that acts as an activator increases gene expression with the addition of more TFBSs until reaching saturation. **B)** Affinity of TFBSs affects gene expression by working as "sensor" for TF concentration. In this example, a weak site (represent as the thinnest purple band over the grey regulatory sequence) ensures binding only above some threshold concentration (conc). **C)** In this example, heterogeneous (heterotypic) clustering of TFBSs results in higher expression levels than their corresponding homogenous (homotypic) clustering. **D)** Gene expression is sensitive to TFBSs arrangements, accounting for both their position and spacing. On the left, no constrains appear independently of TFBSs arrangements and robust expression is achieved in all 3 combinations represented. To the right, different scenarios for TFBSs arrangements lead to a several degrees of decreased gene expression. These constrains may arise due to specific protein-protein interactions being required among specific TFs. Adapted from (Levo and Segal 2014)

However, achieving specific expression of differential sets of genes that constitute the molecular signature behind cell diversity is not only a function of transcriptional activation. Also located at further distances from the TSS are the silencers, DNA sequences bound by repressor factors able to interact either with the basal transcriptional machinery or with enhancers to mediate silencing of promoter regions. Additionally, chromatin elements such as the insulators can act blocking the influence of the activation signals triggered from enhancers as well as the repressive signals received from silencers (or even from the heterochromatin). Integrating all these proximal and distal *cis*-regulatory modules are the tethering elements, *cis*-regulatory elements usually located near the promoter region of the target gene, capable to direct the activity of an enhancer towards that specific gene avoiding those in close proximity. Finally, although less explored, it is known that some TFs can act as activators or repressor depending on the specific cellular or enhancer contexts.
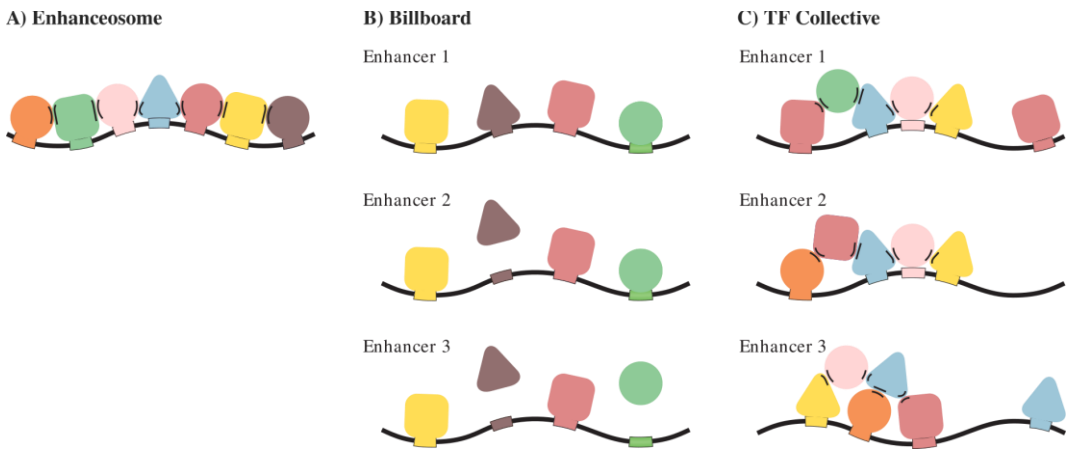
⬆ Figure I1.7. Models for enhancer functionality.
A) In the enhanceosome model functional output requires cooperative binding of all the TFs in a fixed position. **B)** The billboard model allows for both additive or cooperative mechanisms of binding to the DNA and only some TFBS may be occupied in a given moment, order or disposition of TFBS does not affect enhancer activity. **C)** In the TF collective model, the whole group of TFs that integrate the collective may be required to achieve functional output of the enhancer, although interactions among TFs can vary depending on the motif composition and position. Adapted from (Spitz and Furlong 2012).

Regulatory landscapes: 3D conformations in cell-fate decisions.

Ultimately, the establishment and maintenance of the gene-expression programs that define a specific cell type arise from the interplay between the *cis*- and *trans*-acting elements that govern transcriptional regulation. These elements thus conform a regulatory landscape that furthermore exists within a three-dimensional space. As previously stated, TFs need to bind to DNA to exert their regulatory actions. However, accessibility to TFBSs can be fully blocked by the

wrapping of the DNA in the nucleosomes. During development, when new expression programs need to be initialized, a special type of TFs termed "pioneer factors" (Zaret and Carroll 2011) have the capability to bind DNA that is inaccessible for other TFs. Once bound, these pioneer factors recruit chromatin-remodelling complexes that are able to relocate nucleosome distribution thus making DNA accessible for the rest of the transcriptional machinery (Shlyueva et al. 2014). Thus, the capability of a TFs to bind specific locations in the genome is not only determined by its DNA binding domain or its interactions with other TFs expressed at that particular time, but it is also greatly determined by the encountered chromatin landscape that has been modelled throughout the developmental history of the cell.

The pattern of chemical marks that induces and maintains modification in the structure of the chromatin not changing the sequence of the DNA constitute the so-called epigenome of the cell. Epigenetic marks, consequently, add another layer of complexity in the transcriptional regulation of gene expression. Among the most commonly analysed epigenetic marks are methylation and acetylation of both proteins and DNA. Two classical marks for active enhancers are acetylation of lysine 27 or monomethylation of lysine 4 of histone H3 (Shlyueva et al. 2014). On the other hand, methylation of cytosines in DNA has been shown to negatively affect the binding of TFs whereas methylation of lysine 9 of histone H3 is also enriched and required for transcriptional gene silencing (Bell et al. 2011). Despite this broad code of epigenetic marks, not all forms of epigenetic modifications are present in all organisms, for example DNA methylation is not present in *C. elegans* (Wenzel, Palladino, and Jedrusik-Bode 2011).

As it has been shown, control of gene expression requires the topological closeness of distal and proximal *cis*-regulatory elements, thus illustrating the role that the three-dimensional conformation of the genome plays in transcriptional regulation. The development of technologies such as the genome-wide chromosome conformation capture (Hi-C) (Lieberman-Aiden et al. 2009) that is able to systematically interrogate the interaction frequencies between genomic regions has resulted in the discovery of several principles of 3D genome organization (see **Figure I1.8**).

Chromosome territories | Interchromosomal hubs | TADs and loop domains | Intra-TAD interactions

A) Lamina-associated region   Nucleus   Chr. a   Chr. b

B) Speckle   Chr. a   Chr. b   Nucleolus

C) Compartment   A   B   Lamina-associated domain   Chr. a

D)

E) Gene   Enh.

A) Individual chromosomes (represented here as Chr. a and Chr. b) occupy distinct 3D territories inside the cell's nucleus. **B)** Chromosomes form the so-called interchromosomal hubs, in which active chromatin from different chromosomes locates near nuclear speckles whereas inactive chromatin is clustered around the nucleolus**. C)** Inside chromosome territories, active and inactive chromatin further segregates into distinct compartments. A-compartment domains, containing inactive euchromatin (yellow spheres), are positioned towards the nuclear interior. B-compartment domains containing active heterochromatin (blue spheres), are enriched at the nuclear lamina, forming the so-called lamina-associated domains (LADs). **D)** Within each compartment, chromatin is yet further organized into insulated spatial neighbourhoods referred to as topological-associated domains (TADs). **E)** Inside these domains, gene expression is controlled by dynamic interactions between regulatory regions (such as promoter and enhancers (Enh.)) that now locate in close proximity due to this 3D genome organization. Adapted from (Stadhouders, Filion, and Graf 2019).

All these overlapping layers of regulatory elements portray the complexity controlling gene expression and configures the regulatory landscape of a given cell at a given time (see **Figure I1.9**). During the continuous cell-state transitions that ultimately establishes new cell identities, the iterative interplay of 3D genome conformation, transcriptome and epigenome ensure the precise spatiotemporal and robust expression of the differential sets of genes that characterize them.

The making of a neuron.

The generation of cell diversity starts with the commitment of a progenitor cell towards a specific cell fate. In the case of neurons (see **Figure I1.10.A**), this flow of continuous cell-state transitions can be schematized by four steps (Hobert 2005). First, the progenitor needs to select a neuronal versus non-neuronal fate. Rather than a decision, this transition is marked by environmental cues triggered by morphogens and other signalling molecules that start specific gene regulatory networks (GRNs) mediated by signal-regulated TFs that commit progenitors towards neuronal fate but also towards specific neuronal types. These GRNs lead to, but are not limited, activation of proneuronal genes. Exiting from cell cycle, the cell becomes a postmitotic neuronal precursor, constituting a new-born neuron that, being phenotypically immature, stars to migrate and differentiate into a specific neuron type. In a final maturation step, axonal and dendritic morphogenesis, expression of neurotransmitter pathways and synaptogenesis occur. At this point, the mature neuron becomes integrated into a neural circuitry, exhibiting its final functional properties. Differentiation and maturations steps occur simultaneously and reflect the expression of the specific batteries of genes that define the neuron type and that allows for the specific functions of the mature neuron (**Figure I1.10.B**).

This set of cell-specific genes, which are expressed throughout the life of a mature, functional neuron, are the so-called effector genes or terminal features and represent the molecular signature shaping the identity of the neuron (Hobert 2016b). The nature of these neuron-type specific set of genes is

rather combinatorial: is the combinations of genes expressed in a neuron type what is specific of that neuron and not the presence of exclusively-expressed gene products.
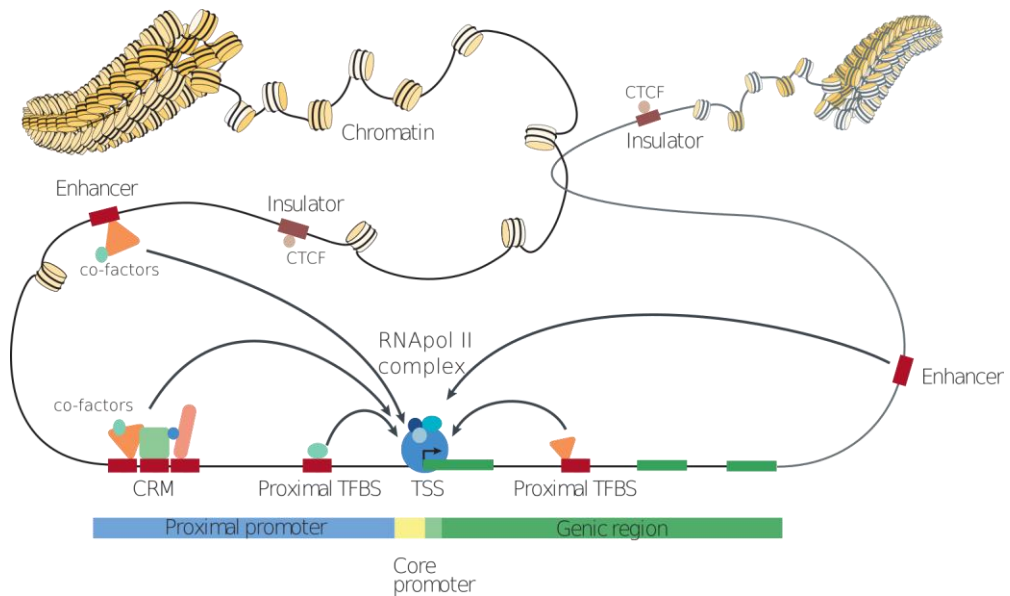
⬥ Figure I1.9. The transcriptional regulatory landscape of a cell integrates the inputs of *cis*-, *trans*- and 3D elements affecting gene expression.

Several layers of regulatory elements control gene expression. Proximal and distal *cis*-regulatory elements, such as promoters or enhancers, act as the scaffolds in *which* trans-acting elements such as transcription factor or cofactors bind to activate transcription triggering RNApol II activity. Three-dimensional elements also influence the outcome of transcription with insulator marking the chromatin boundaries of the territories accessible for the transcriptional machinery. The CTCF transcription factor helps to establish these boundaries (or TADs) by biding the CCCTC-binding motifs that are enriched in these domains. CRM: cis-regulatory module clustering TFs. Although nematodes have lost CTCF protein, the presence of TADs has been reported in *C. elegans*. Adapted from (Lenhard et al. 2012).

Neuronal terminal differentiation has been best studied in *C. elegans* where it has been shown that most effector genes expressed in a specific neuron type are co-regulated by the same combination of TFs and thus share a regulatory signature in their enhancers (REF). This was exemplified for the first time in a seminal study that showed that many effector genes expressed in the AIY interneuron, were co-regulated by the a heterodimer composed by two different transcription factors (*ttx-3* and *ceh-10*). (Wenick and Hobert 2004).

In the 1970s, Dr. Garcia-Bellido coined the term "selector gene" to designate genes controlling the identity of specific cellular domains in the developing *Drosophila* (Garcia Bellido 1975). Ectopic expression of these selector genes is sufficient to induce homeotic transformations sufficient to generate complex ectopic tissues or structures. Thus, building on this concept, Dr. Hobert proposed

the term "terminal selector" to describe those TFs directly controlling the expression of the terminal, effector genes (Hobert 2008). Within a neuron type, the *cis*-regulatory regions that control the expression of the effector genes that are active in that neuron harbour TFBSs for the specific combinations of TFs that act as the terminal selectors of that neuron (Hobert 2011). The combinatorial activity of the several TFs that are needed to activate the expression of those genes defines the so-called "terminal selector codes" (see **Figure I1.11.A**).

A)

B)
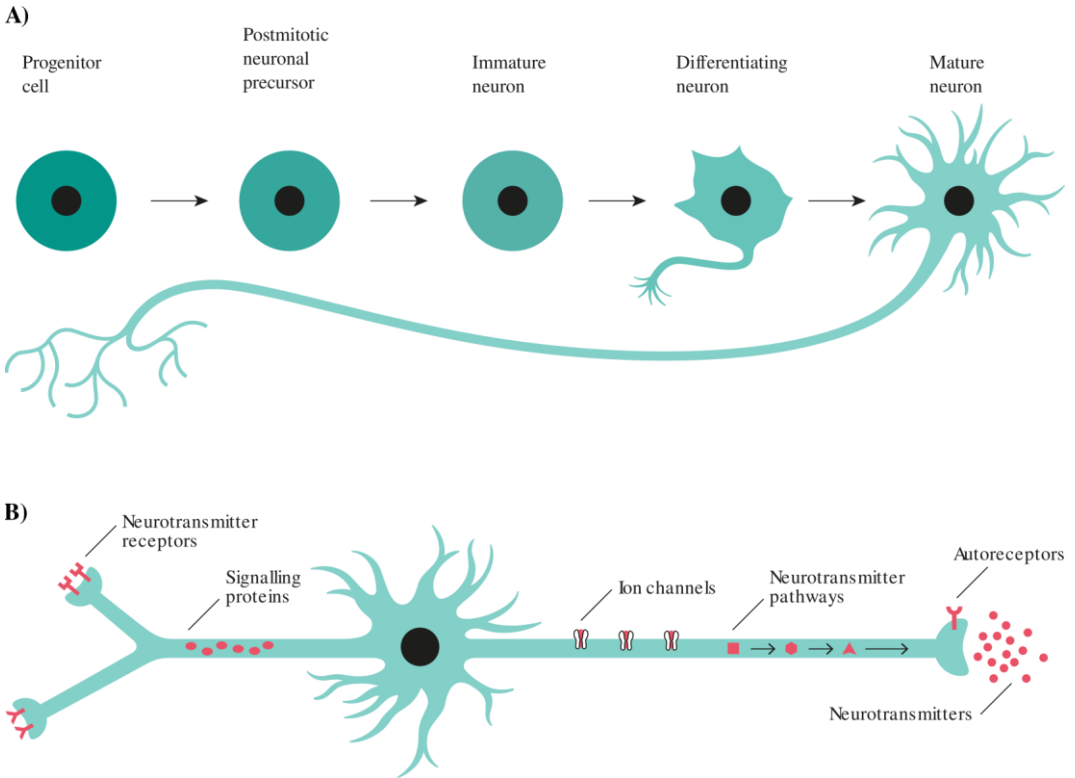
⏶ **Figure I1.10. Stages of neuronal development.**
A) Four continuous steps schematize the cell-state transitions that carry from a progenitor cell to a fully mature, functional neuron. **B)** Terminal features of a mature neuron reflect the expression of the differential sets of genes that allows for its specific functions and defines it cell-specific molecular signature. Adapted from Carla Lloret-Fernández PhD manuscript.

The terminal selector model explains why an individual neuron sub-type may not entail a large number of TFs to regulate effector genes, since the specific combination of TFs is the one directly regulating most neuron-type specific terminal features. This implies that the same TF can regulate different sub-routines by acting in combinations with different TFs. For example, it has been determined that TTX-3 acts as terminal selector for the AIY cholinergic interneuron together with CEH-10, but it also acts as a terminal selector for

the NSM serotonergic fate together with a different TFs, UNC-86 (Zhang et al. 2014). Conversely, shared effector genes contain different *cis*-regulatory modules in their sequences that are activated by the different combinations of terminal selectors being expressed in each cell type (see **Figure I1.11.C**).

However, it must be noted that terminal selectors controlling the expression of a neuron-type-specific set of genes do not control the expression of all the identity-defining genes present in a terminally differentiated neuron type. This is better illustrated when considering that removal of terminal selectors controlling neuron-type-specific features do not usually affect the expression of features shared by many neurons (such as those present in all 60 sensory neurons) or all neuron types (such as the pan-neuronal genes) (Stefanakis, Carrera, and Hobert 2015). Consequently, it has been proposed that in addition to terminal selector TFs other parallel regulatory programs controlling other neuronal or morphological features must be running along the ones orchestrated by terminal selectors (Hobert 2011) (**see Figure I1.11.B**). Currently, we do not understand how these parallel routines are activated and if there is any regulatory connection among them. A clear example from these parallel routines is found in the establishment of the ciliated identity seen in the sensory neurons of *C. elegans*, where the coordinated expression of genes required for the building of the sensory cilium is directly controlled by the RFX-type transcription factor *daf-19* (Senti and Swoboda 2008; Swoboda, Adler, and Thomas 2000a) (**see Figure I1.11.B**). Therefore, *daf-19* could be considered a terminal selector of the parallel routine of the ciliated fate. Nevertheless, at least for pan-neuronal gene expression, a high degree of redundancy exists due to the presence of multiple enhancers acting in parallel. Thus, although deletion of terminal selectors does not affect pan-neuronal gene expression, activity of specific pan-neuronal enhancers may be modulated by their direct action.

Of note, expression of terminal differentiation genes is usually required throughout the life of a mature neuron, since they define their final, functional state. Accordingly, it has been determined that terminal selectors are not only required for initiation of effector genes expression but also for the maintenance of their expression (Deneris and Hobert 2014).

Finally, it is important to mention that transcriptional regulation is only the first step towards the generation of specific neuron types. Additional post-transcriptional mechanisms such as regulation of translation and post-translational modifications are also necessary for the generation of specific neuronal types. However, these additional mechanisms are out of the scope of this Thesis and hence will be not further discussed in this Chapter.
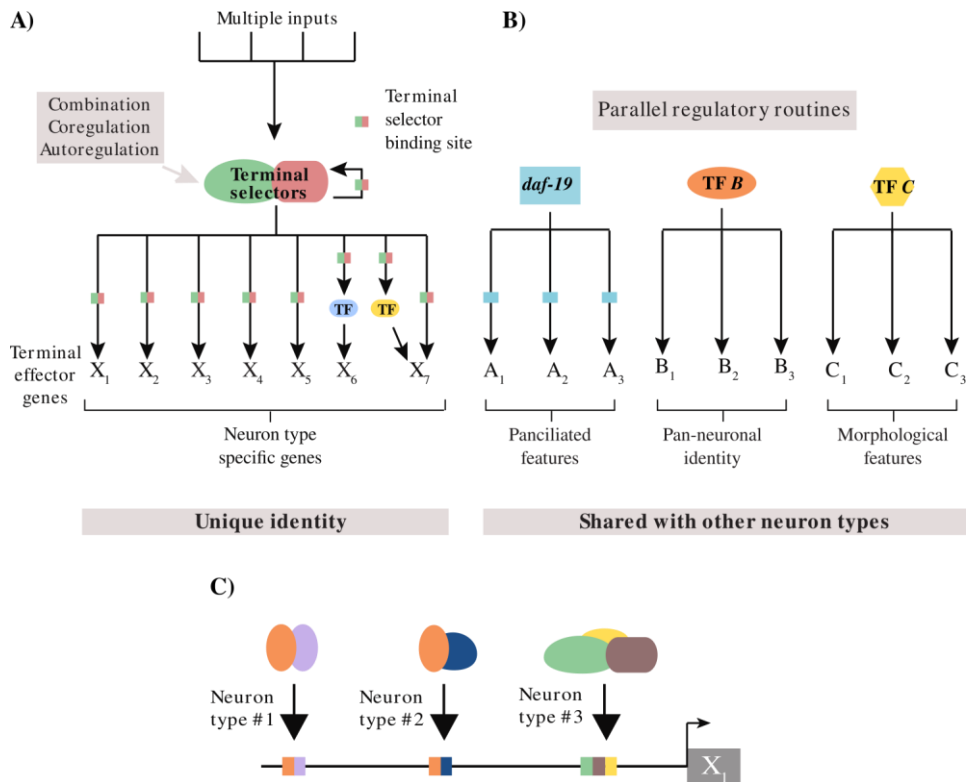
⬆ **Figure I1.11. The terminal selector model and the transcriptional regulation of the cell-type specific genes.**

A) Terminal selector codes regulate batteries of terminal effector genes in specific neuronal sub-types. They do so through the recognition of the terminal selector motif (green/pink rectangle). Sustained expression of terminal selectors is often, but not always, ensured by autoregulation. Xn represents the effector genes. Sometimes, terminal selector codes further activate additional TFs that control other sub-routines (blue and yellow circles). **B)** Parallel regulatory routines such as those regulating panneuronal features (via TF *B* in the scheme) or morphological features (via TF *C* in the scheme) are controlled by transcription factors that do not assign unique identities. **C)** Modular organisation of the regulatory sequence of effector gene X1 allow for its regulation by different combinations of terminal selectors. Adapted from (Hobert 2016b).

The cilium - a sensory organelle

In multicellular animals, sensory perception relies on the function of specific types of neurons that possess specialized sensory structures. One of those structures is the cilium. Cilia are eukaryotic organelles, morphologically resembling a lash, that may either protrude as an appendage from the cellular surface (see **Figure I2.1**) or be embedded within the cytosolic space.

Due to the size that some types of cilia are able to reach, they were the first organelles to ever be discovered. Antonie van Leeuwenhoek, known for his pioneering work in the field of microscopy, originally described them as "incredibly thin little feet, or little legs, which were moved very nimbly" (Leewenhoeck 1676). This description probably due to him observing aquatic free-moving ciliated protozoa. Otto Friedrich Muller is often credited to have proposed, more than a century later, the term "cilium" (meaning hair or eye lash) to name such structures described by van Leeuwenhoek. By the mid-19th century, French biologist Félix Dujardin is thought to have introduce the term "flagellum" (meaning whip) (see Bloodgood 2009 for an extensive review on the highlights of cilia history). In contemporary literature those terms are often used in a somewhat indistinguishable manner; however, it is common to stablish a difference considering basic functions and length. Accordingly, "flagellum" is often used to define a single (or few) relatively long motile organelle whereas "cilium" is used when a cell has many similar organelles comparatively shorter than flagella, either motile or non-motile. A further discrimination is sometimes done to define solitary, non-motile cilia by using the term "sensory primary cilium" or just "primary cilium". As long as length is concerned, vertebrate cilia typically range from 3 to 10 μm for many cilia and from 50 to 150 μm for sperm flagella; although lengths as long as 200 μm have been found for sensory olfactory cilia (Silverman and Leroux 2009).

The evolutionary emergence of cilia has been traced in the interval between divergence of eukaryotes from archaeal prokaryotes and divergence of extant eukaryotic clades about one billion years ago (Mitchell 2017). The last eukaryotic common ancestor (LECA) has been proposed to have hybrid motile/sensory cilia with gliding and beating motility as well as sensory receptors. Most of the proteins present in cilia today were already present in the LECA (as assessed through genomic, metagenomic, and transcriptomic analyses), thus making cilia a highly evolutionary conserved organelle. Later during evolution, a secondarily loss of cilia took place in a few eukaryotic lineages, such as most land plants and fungi (Richards and Cavalier-Smith 2005).

A)

B)

C)

D)

E)

F)
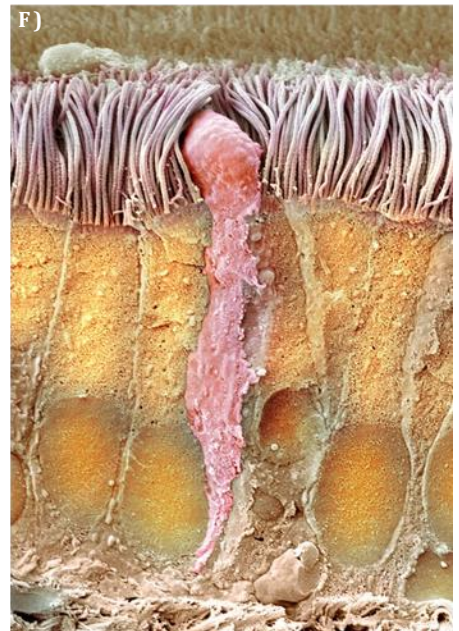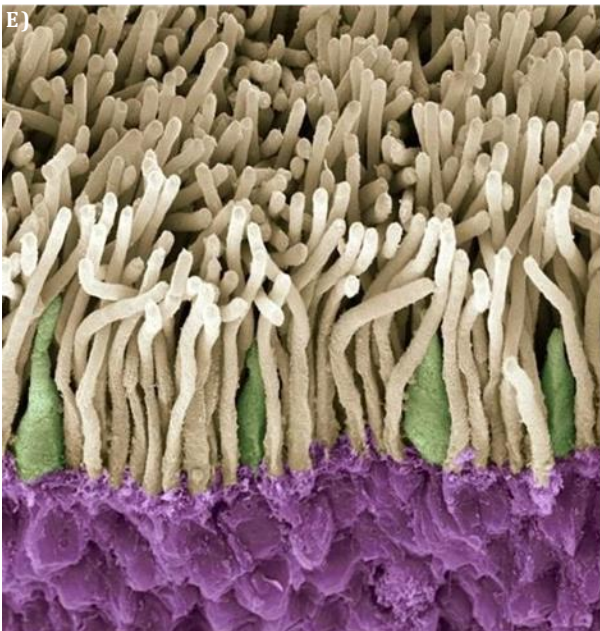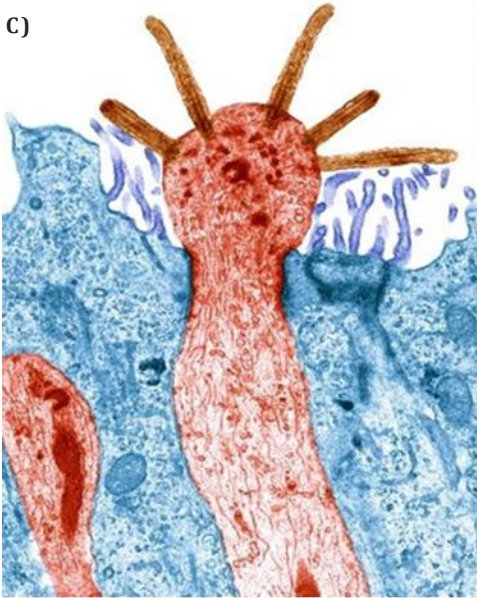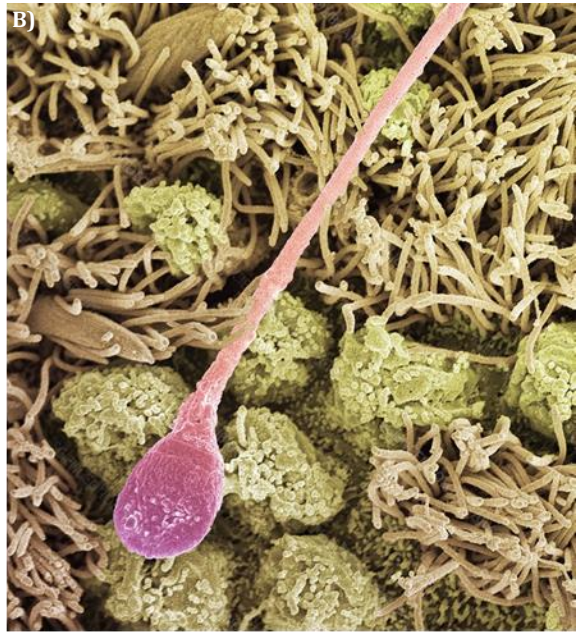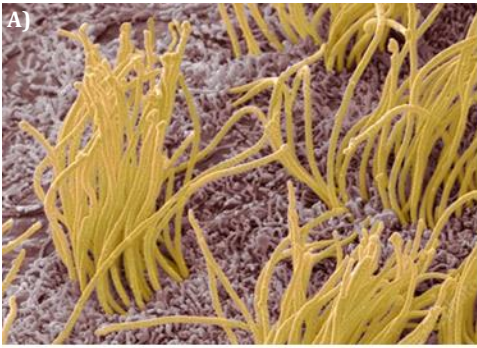
◄ **Figure I2.1. Colorized micrographs of mammalian ciliated cells.**
A) Coloured scanning electron micrograph (SEM) of the lining of the brain, showing the ciliary hairs (yellow) of ependymal cells. Ciliary hairs are motile cilia used to help circulate the cerebrospinal fluid ependymal cells produce. **B)** SEM of a sperm cell (pink) in a fallopian tube. Sperm cells possess a long motile cilium (termed flagellum) that they used for propulsion. Motile cilia at the fallopian tube move the egg through the tube into the uterus. **C)** Transmission electron micrograph of a section through the olfactory epithelium, showing an olfactory neuron with 5 protruding sensory cilia (red). Co-ordinated, wave-like beating of the motile cilia from surrounding cells (blue) propels mucus to the back of the nose. **D)** SEM of outer sensory hair cells (stereocilia) from the organ of Corti in the cochlea of the inner ear. The V-shaped arrangement of cilia belongs to a single cell. **E)** SEM of sensory cilia of human rod (pale yellow) and cone cells (green), adjacent to the outer nuclear layer (purple) of the retina. **F)** SEM of a section through the wall of a trachea. Mucus-secreting goblet cells (seen in the center in pink) is surrounded by multiciliated epithelial cells (orange). Cilia (light purple) move mucus and particles out of the respiratory tract. Credit STEVE GSCHMEISSNER / SCIENCE PHOTO LIBRARY

Basic morphological features of the cilium.

Cilia are highly organized organelles with a very defined morphology (see **Figure I2.2**). The core scaffold of a cilium is a cylindrical structure formed by 9 doublets of microtubules surrounded by a membrane with specific proteins and lipid composition (see Garcia, Raleigh, and Reiter 2018 for a review on ciliary membrane organization) in order to differentiate from the cell membrane. Some types of cilia possess an additional pair of microtubules at the centre of this structure. These alternative conformations are respectively denoted as 9+0 and 9+2. Although not universal, these conformations correlate well with the two main functions of cilia: sensory (9+0) and motile (9+2).

Functions		
Cell signalling	Chemosensation (e.g. for noxious compounds)	• Hedgehog • Receptor tyrosine kinase (RTK) • Cyclic nucleotide (cAMP, cGMP) • Calcium (PKD1, PKD2, etc.)
Motility	• Sperm locomotion • Cerebrospinal fluid flow • Respiratory airway clearance • Nodal flow • Fallopian tube flow	

Typical structures of the motile (left) and sensory non-motile cilia (right) appear in longitudinal (left) and transversal (right) views. Basic morphologies of the basal body and the transition zone are shared for all types of cilia and present a cylindrical conformation built on 9 triplets of microtubules denoted as A, B and C. Due to functional requirements, cilia may contain additional subdomains, such as regions rich in inversin domains (INV) involved in signalling that may appear on sensory cilia. The axoneme of both sensory and motile cilia are formed by the extension of the A and B microtubules. Owing on their functionality, axonemes of motile cilia appear decorated with several components required for motility. On the bottom, main cellular functions associated with both types of cilia are mentioned. Adapted from (Reiter and Leroux 2017).

The centrosome, the main microtubule organizing centre of animals' cells, constitute the organelle from which cilia emerge (see **Figure I2.3**). The centrosome is formed by two cylindrical structures, the centrioles, that are built on 9 triplets of microtubules and arranged in an orthogonal manner surrounded by the pericentriolar material. The two centrioles differ in age, functionality and presence of additional structures. The older centriole, the so-called mother centriole, arose at least one cell cycle earlier than the second centriole, which is consequently labelled as the daughter centriole. The base of the cilium (or basal body) projects from the distal part of the mother centriole and thus conserves a conformation containing the triplets of microtubules. Two types of appendages can be found around the mother centriole that do not appear in the daughter centriole. Considerable confusion and misconceptions exist in the literature regarding the definition of their exact location and basic properties. In this work we make use of the conclusions exposed by (Uzbekov and Alieva 2018). Nine distal appendages (see **Figure I2.3**) appear attached surrounding (in the same plane) the distal part of the mother centriole, fixing it to the cell membrane starting at cilium formation. A variable number of subdistal appendages, ranging from 0 to 14, can appear located along the entire length of the mother centriole and work as centres of microtubule nucleation. The proximal region of the mother centriole connects with the proximal region of the daughter centriole through thin filaments called connecting fibers. In some type of cilia, these filaments further extend to the interior of the cell, forming the so-called striated rootlet (Pedersen et al. 2012).

Immediately distal to the basal body is the transition zone. The transition zone is formed by the extension of the mother centriolar A and B microtubules, thus conferring the characteristic pattern of microtubule doublets that is connected to the ciliary membrane via Y-shaped structures (see **Figures I2.2 and I2.3**). These structures, along with other structural and functional features, transform the transition zone into a diffusion barrier or gate precisely controlling the access of components into the ciliary compartment.

The distal part of the transition zone joins to the core structure of the cilium, the axoneme, integrated by the extension of the plus end of microtubule doublets. B

microtubules are generally shorter than A microtubules, thus generating a distal structure of singlet microtubules that is termed the distal segment. Length of these structures varies substantially among species or are even inexistent (see Prevo, Scholey, and Peterman 2017). In motile cilia, axonemes appear decorated by axonemal dyneins and other formations required for motility that connect the peripheral doublets with the central pair of microtubules. Nodal cilia, a specialized form of motile cilia that appear in the embryonic ventral node and are crucial for the establishment of the left-right asymmetry (Nonaka et al. 1998), constitute the notable exception since they possess a 9+0 axoneme and not the 9+2 . structure typical of motile

cilia. The region between the microtubules end and the most distal part of the ciliary membrane constitutes the ciliary tip. This region limits the length of the cilium and is implicated in several signalling processes (Malicki and Johnson 2017), including the release of extraciliary vesicles and ectosomes (Wang and Dynlacht 2018).

Importantly, basic morphologies and protein composition of the basal body, the transition zone and axoneme are shared between both sensory and motile types of cilia. The main difference is the presence of structural dyneins in motile cilia and localization of specific receptors and signalling molecules in the sensory cilia.

The intraflagellar transport machinery.

Since the ciliary matrix is deployed of ribosomes and thus no proteins are synthesized inside the organelle, assembly, maintenance and disassembly of cilia fully relies on the action of the so-called intraflagellar transport (IFT), a form of motor-dependent cargo transport that functions along microtubular structures of cilia. IFT was first discovered in *Chlamydomonas*, when polystyrene beads were observed to move up and down the motile cilia independently of ciliary beating (Bloodgood 1977).

Although poorly understood, a six-steps model was proposed in 2008 to explain IFT (Pedersen et al. 2008). The first step involves the recruitment and assembly of the IFT machinery and its cargo. This process happens at the transition zone, but the required components are trafficked from several organelles in the cell, even directly from the nucleus (Satir and Satir 2019). In the second step, anterograde transport of IFT particles occur from the base of the cilium to its distal tip. Kinesins are the molecular motors triggering this transport, carrying their cargos along the microtubules. These IFT particles are in turn integrated by two sub-particles, namely IFT-A and IFT-B. IFT-A is usually associated with heterotrimeric kinesin-2 whereas IFT-B is typically linked to homodimeric kinesin-2 (see Prevo et al. 2017). To date, the specific cargos that associate with IFT-A or IFT-B proteins are poorly characterized. A second class of molecular motors, the dyneins, are also transported during anterograde IFT, usually attached as an inactive form to IFT-A particles. In those cases in which a distal segment exists, IFT is carried there by homodimeric kinesin-2.

Ciliary membrane biogenesis is promoted by increasing levels of Rab8/GTP, a small GTPase involved in several processes related to membrane and vesicular trafficking (Nachury et al. 2007). In addition, the IFT-B component IFT-54 directly interacts with Rabaptin-5 (a guanosyl exchange factor), which in turn also interacts with Rab8/GTP (Omori et al. 2008). Activation of Rab8 depends on its

own guanosyl exchange factor Rabin8 which in turn is associated to BBS1, one of the proteins integrating the complex termed BBSome. Although some authors consider the BBSome to be a third IFT particle, others consider this debatable (see Hao and Scholey 2009; Prevo et al. 2017). In any case, the BBSome has been shown to play a crucial role in the proper assembly of the IFT machinery (Wei et al. 2012) and is thought to be conserved among ciliated organisms (Hao and Scholey 2009). The detailed composition of the IFT particles and a description for the two types of molecular motors involved in IFT appear in **Figure I2.4.**

A)

B)

▲ **Figure I2.4. Particles and molecular motors integrating the intraflagellar transport machinery. A)** Known components for the IFT-A, IFT-B and BBSome complexes as they appear in (Taschner and Lorentzen 2016). **B)** Sub-units integrating the molecular motors responsible for the IFT in several species. Adapted from (Prevo et al. 2017)

Once the IFT train has reach the distal-most part of the axoneme, a series of poorly understood events globally termed "the turn-around" take place. Consequently, the third step of the IFT process involves the inactivation of the anterograde motors, the release of the cargo and the dissociation of the IFT particles. Next, in the fourth step, activation of the retrograde dynein motor, assembly of the retrograde IFT machinery and uploading of retrograde cargo occur. During the fifth step, retrograde IFT of ciliary turnover proteins and inactive anterograde motors are transported from the distal part of the axoneme back to the base of the cilium. At that point, in a sixth step, the IFT machinery is disassembled and their components are recycled. A whole scheme for the IFT process can be seen in **Figure I2.5**. Of note, IFT-A, IFT-B, BBS proteins and transport dyneins are expressed both in sensory and motile cilia.
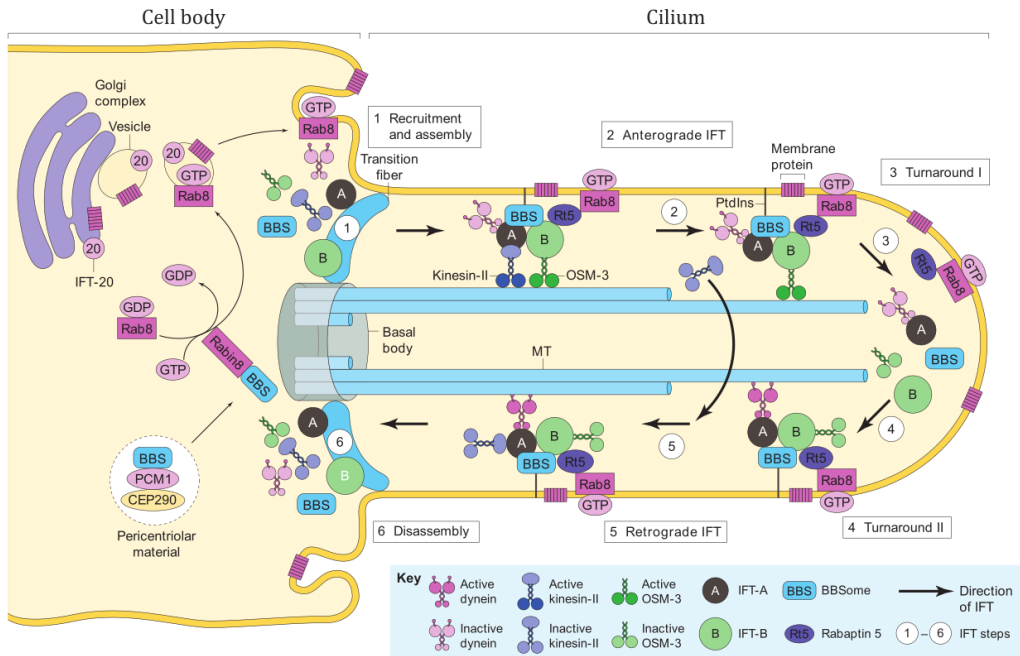
⬈ Figure I2.5. Intraflagellar transport at a glance.
Intraflagellar transport involves the recruitment of several components from the cell body to the transition zone, either arising from the Golgi (as IFT-20) or from pericentriolar vacuoles. Once the required cargo has been trafficked to the base of the cilium, IFT machinery is assembled and cargo transported into the ciliary space. The IFT machinery undergoes a process that can be divided in six steps (see text for details), ending with its disassembly. Notice that in those systems in which two types of kinesins exist, only the homodimeric kind (here represented through *C. elegans* OSM-3) proceeds along the distal segment. Adapted from (Hao and Scholey 2009)

Key stages of ciliogenesis.

Pioneering work by Sorokin already described, back in the 1960s, the two alternative pathways by which a cilium can be assembled (Sorokin 1968).

The intracellular pathway initiates when small vesicles (probably arising from the Golgi) localize near the distal end of the mother centriole, fusing together to form the so-called ciliary vesicle, which then attaches to the mother centriole through the distal appendages (**Figure IB.6**). In a second intracytosolic event, the mother centriole becomes the basal body and a (probably immature) transition zone begins to emerge. This forces the invagination from the side of the ciliary vesicle closest to the mother centriole, thus covering the structure with an inner and an outer membrane. Next, nearby vesicles accumulate at this region and fuse to this newly formed membrane to help accommodate the growing cilium. At this stage, depending on the cell type, the whole structure will migrate toward its final position in the cell. Eventually, the plasma membrane is reached and the outer membrane, forming a sheath surrounding the axonemal shaft, fuse to the cellular membrane. In this step, the original inner membrane of the ciliary vesicle becomes the ciliary membrane and gets exposed to the extracellular milieu. In some cell types, the forming cilium gets surrounded by an invagination of the cell membrane that is termed the ciliary pocket that acts as a hub for actin dynamics and endocytosis (Molla-Herman et al. 2010). At this point, the maturation of the transition zone forms the proper ciliary gate and IFT is initialized. Axonemal microtubules are then extended from the mother centriole until they reach the appropriate length for the cilium.

This intracellular pathway for ciliogenesis was first described by Sorokin in fibroblasts and smooth muscle cells. However, he also observed an alternative path being employed by lung epithelial and inner medullar collecting duct cells. In this extracellular pathway, the mother centriole migrates toward the cell membrane and the basal body directly docks through the distal appendages. A transition zone might start to develop just before, during or after this step and most of the resulting cilium will protrude out into the extracellular milieu. A modified version of this process is also used by multiciliated epithelial cells (see Spassky and Meunier 2017). Although the mode of ciliogenesis (see scheme in **Figure I2.6**) seems to be a specific feature of cell and tissue types, the reasons behind the choice of a specific pathway are not well understood. However, functional needs may correlate with the choice of the ciliogenic pathway. For example (and as seen by Sorokin (Sorokin 1968)) epithelial cells in charge of sensing extracellular fluids follow an extracellular pathway, thus creating an apically-positioned cilium fully protruded to the extracellular milieu. By contrast, cells following the intracellular pathway usually create cilia that remain almost fully submerged in their ciliary pockets. Much interestingly, a recent study in which this idea was tested found that the conversion from "submerged" to "surfaced" promoted flow sensing in cilia but dysregulated ciliary Hedgehog signalling (Mazo et al. 2016).
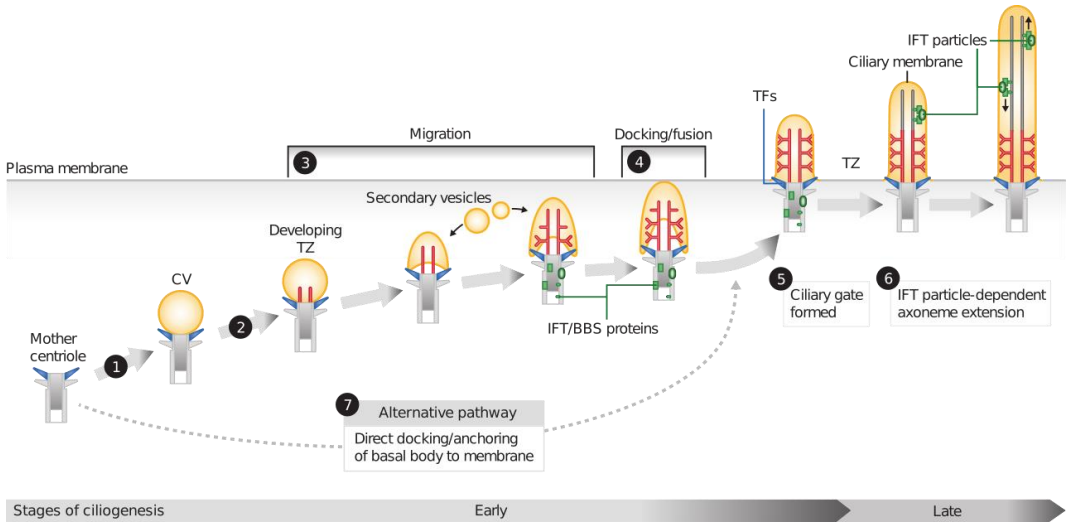
▲ Figure I2.6. Pathways for ciliogenesis and their key stages.
From (1) to (6), the intracellular pathway for ciliogenesis is schematized, alternative pathway starts at step 7. In the mother centriole, distal appendages are represented in blue whereas sub-distal appendages remain in grey. Ciliary vesicle (CV) and secondary vesicles are depicted in yellow. The developing transition zone (TZ) and its Y-shaped structures appear un red. IFT particles and the BBSome, which appear prior to docking of the basal body to the cell membrane, are portrayed in green. The growing microtubules in (6) are shown in grey. The alternative extracellular pathway starts by the direct docking of a basal body to the cell membrane (7), thus skipping the ciliary vesicles. Although not represent here, intracellular pathways usually lead to the creation of an invagination surrounding the final cilium due to the fusion between the outer membrane of the ciliary vesicle (the sheath) and the cell membrane. Extracellular pathways generally generate fully protruded cilia as the one portrayed in this scheme. TF: transition fibres. Adapted from (Reiter, Blacque, and Leroux 2012).

In terms of cell cycle, ciliated cells are found either on a quiescent state or in the G1 phase of the cycle. For those cells in which cellular division still occurs, assembly and disassembly of the cilia are inherently tied to cell cycle progression and thus constitute a checkpoint for the control of ciliogenesis. A newly post-mitotic cell lacks cilia. During the G1 phase, the mother centriole docks at the cell membrane and nucleates a cilium. During the S-phase, the mother and daughter centrioles each duplicate forming new daughter centrioles (or procentrioles). The cilium is disassembled prior to mitosis, thus realising the mother centriole from the cell membrane, allowing each of the newly created pair or centrioles to move to the spindle poles. Cell division occurs and cilia are then re-formed in the two daughter cells once they re-enter into the G1 phase of the cycle (see Breslow and Holland 2019 for a detailed explanation of the whole process). Since multiciliated cells are terminally differentiated, they do not undergo cell division and thus the previous mechanism do not apply. However, multiciliated precursors, which only possess one centrosome, must amplify between 30 to 300 centrioles to support cilia nucleation. To do

so they rely on intermediate structures called deuterosomes. Although controversial (see Rayamajhi and Roy 2020), those ring-shaped structures have been shown to support massive centriole formation by using shared elements from the cell cycle centrosome duplication program (see Spassky and Meunier 2017).

Just as cilia assembly has been found to follow different paths, so does cilia disassembly. Although comparatively less is known about this process, experimental evidence gathered from different organisms (see Breslow and Holland 2019) has led to the proposals of several models explaining cilia disassembly. These include cleavage of all or some portion of the protrudin cilium, retraction of the axoneme into the cell body sometimes associated with the discard of the ciliary membrane, and also active mechanisms in which the cilium is progressively shortened followed by either the resorption of the ciliary remnants or release of the basal body from the cell membrane. These methods have been seen to occur simultaneously and in different combinations. When the disassembly is related to the progression of the cell cycle it proceeds in a biphasic way: first major wave takes place in the G1 phase, shortly after the mitogenic stimulation of the quiescent cells, then a minor second wave occurs just before mitosis. Between these two edges of full ciliary assembly and disassembly an active cilium, which is a highly dynamical structure, needs to maintain itself. In fact, both opposing processes take place continually in a ciliated cell at steady-state, since ciliary length can only be maintained when the rates of assembly and disassembly are even (Ishikawa and Marshall 2011).

Functional diversity of sensory cilia.

Preceding sections have shown the commonalities found across cells and organisms for the two major types of cilia: motile and non-motile sensory cilia. However, an inherent source of variation modifying the final appearance of a cilium arises from the distinctive requirements that different functions demand. In addition to the main distinction between motile and sensory, sensory cilia itself displays the highest degree of variation. Accounting for the highly diverse nature of the cues they must perceive, cilia from different sensory systems must specialize in composition and shape to adequately perform their corresponding tasks.

In eukaryotes, the two principal categories of sensory receptors are the ion channels and the G-protein-coupled receptors (GPCRs) (Julius and Nathans 2012). Both GPCRs and ion channels contribute to downstream signalling cascades when they are activated by small molecules, mechanical perturbations, or radiation. In vertebrates, GPCRs are the main detectors for visual and olfactory stimuli whereas ion channels predominate in the detection of auditory and somatosensory cues. Depending on the particular taste, both GPCRs and ion channels serve as detectors in gustatory systems.

The need for specialized receptor composition in the ciliary membrane is well illustrated by olfactory systems. In animals, the sense of smell relies on the action of olfactory sensory neurons, bipolar neurons possessing sensory cilia located at their dendritic endings (see **Figure I2.1.C**). In vertebrates, olfactory sensory neurons possess a number of

cilia that range from 10 to 30 and they possess the 9+2 microtubule configuration that is usually found in motile cilia (see Falk et al. 2015). However, sensory cilia from these olfactory neurons lack the dynein arms necessary for movement; thus, they remain immotile. Olfaction signalling is initiated when an odorant, acting as a ligand, contacts a specific GPCR located on the ciliary membrane. In vertebrates, each olfactory sensory neuron express a single type of GPCRs (see Niimura and Nei 2003 as an example). However, in *C. elegans* that encodes for an extremely large class of GPCRs genes (1.200, approximately) and contains only 32 chemosensory neurons, each olfactory neuron in the worm co-expresses dozens of chemosensory receptors at once thus being able to detect a broad range of odorants each (Milanetti et al. 2019).

From the basic rod shape seen in many primary cilia to the convoluted structure observed in photoreceptors, very little is known about the mechanisms governing the acquisition of the highly divergent membrane morphologies that sensory cilia possess. Ciliary membranes specializes to differ from the cell membrane also in lipid composition and it has been proposed that the chemical nature and arrangement of different kind of lipids along the inner and the outer leaflets of the ciliary membrane may determine membrane curvature (Garcia et al. 2018). Analogously, proteins that localize to the ciliary or periciliary membrane also help determine membrane morphology, as exemplified in *C. elegans* by the membrane protein OSTA-1, an phylogenetically conserved organic solute transporter (Olivier-Mason et al. 2013). In *osta-1* mutants, the ciliary membrane exhibits a distorted branching and, interestingly,

selected ciliary GPCRs aberrantly enrich in the periciliary membrane compartment (a specialized domain surrounding the base of the nematode's cilia that remains in the absence of the basal bodies, which degenerate in this organism (at least for a subset of its ciliated neurons) after ciliogenesis (Nechipurenko and Sengupta 2017)). The aberrant location of GPCRs found in *osta-1* mutants illustrates how ciliary membrane morphology is also determined by the trafficking of specific components transported by the IFT machinery. Moreover, examples exist (reviewed in Silverman and Leroux 2009) in which cilia differing in their morphologies make use of distinct IFT mechanisms, coupling or un-coupling the same molecular motors and thus conditioning the type of cargos that enter the cilium (Mukhopadhyay et al. 2007). Importantly, it should be noted that transport of ciliary components from other organelles such as the Golgi or even the nucleus is performed through carrier vesicles that ultimately need to fuse to either the ciliary or the surrounding cell membrane. Although several models have been proposed for this kind of trafficking (Nachury, Seeley, and Jin 2010), disruption of endocytosis effectors is known to affect ciliary morphology (Kaplan et al. 2012). Finally, a role for cilia unrelated proteins such as kinases have been shown to shape ciliary membrane morphology, as exemplified by mammalian cyclin-dependent kinase 20 (*Cdk20*) (Snouffer et al. 2017). More importantly, the role of CDK20 in the control of ciliary length and morphology seems to be evolutionary conserved, since homolog *Chlamydomonas* mutants also display uncharacteristic bulbous cilia (Tam, Wilson, and Lefebvre 2007). Of note, *Cdk20*

mutant mouse embryonic fibroblasts present reduced ciliary Smothen (SMO) levels the receptor of Hedgehog signalling, thus exhibiting attenuated signalling of this pathway(Snouffer et al. 2017).

When cilia fail: the ciliopathies.

Due to the astonishing number of processes that cilia mediate, absence or impairment of these organelles are related to a plethora of human diseases globally termed as ciliopathies (see **Figure I2.7**). Motile cilia, other than providing propulsion for sperm locomotion, are responsible for the movement of extracellular fluids. As such, they can be found on airway epithelial cells clearing mucus, on fallopian tube cells helping eggs to reach the uterus or on ependymal cells in the brain making the cerebrospinal fluid to circulate. Impaired ciliary motility thus leads to a motile ciliopathy that is generally termed as primary ciliary dyskinesia (PCD). Depending on the affected gene or genes the degree of pleiotropy for PCD greatly varies (as is the norm for the rest of ciliopathies) and the phenotypic manifestations may include from *situs inversus* (incorrect laterality disposition of internal organs) to decreased fertility (Horani et al. 2016) (Pedersen et al. 2012). Although motile cilia can also harbour sensory functions, the sensory primary cilium has specialized as the major organelle responsible for the detection of mechano-, osmo-, and chemosensory cues. As a result, it is implicated in the regulation of several key embryonic processes such as cell cycle, differentiation, polarization or migration, but also participates in the maintenance of the cellular functions in adult tissues and organs. Several diseases were described long before

they could be attributed to defects in primary cilia. Those were the cases of the Bardet-Biedl syndrome (BBS), nephronophthisis (NPHP), Senior-Löken syndrome (SNLS), Alström syndrome (ALMS), Meckel syndrome (MKS), Joubert syndrome (JBTS), the polycystic kidney diseases (PKD) (see **Table I2.1**) and many others (see Tobin and Beales 2009). However, it was not until the 2000s that the primary cilium was discovered as a hub for signalling processes. Since then, requirements for primary cilia have been observed for all main intercellular signalling pathways, namely Hedgehog, WNT, NOTCH, planar cell polarity, receptor tyrosine kinases, G protein coupled receptors and the transforming growth factor-β signalling (reviewed in Nachury and Mick 2019). Since virtually every single cell in the human body has (or proceeded from a cell possessing) a primary cilium, the range of diseases associated with the malfunction of primary cilia constitutes a still growing cohort (see Reiter and Leroux 2017 for a review).

Upon the realisation of the key role of sensory cilia in mediating signalling processes, great efforts have been made to elucidate gene cilia composition and function, the so-called "ciliome". Indeed, many of the cilia-related genes that were identified during those studies were named after the diseases they were associated to. In 2013, the SysCilia consortium published what is considered the gold standard list of manually curated cilium-associated components (Van Dam et al. 2013). That first version of the SysCilia accounted for 303 genes; since then, many more genes have been (and keep being) added. This massive amount of data gathered during the past decades has prompted the

creation of several databases, such as CilDB (Arnaiz et al. 2009, 2015), in which the results of those studies are categorized and made easily available. Moreover, efforts aimed at the integration of all that information has led to initiatives such as the CiliaCarta, an integrated and validated compendium covering 956 ciliary genes (Van Dam et al. 2019).
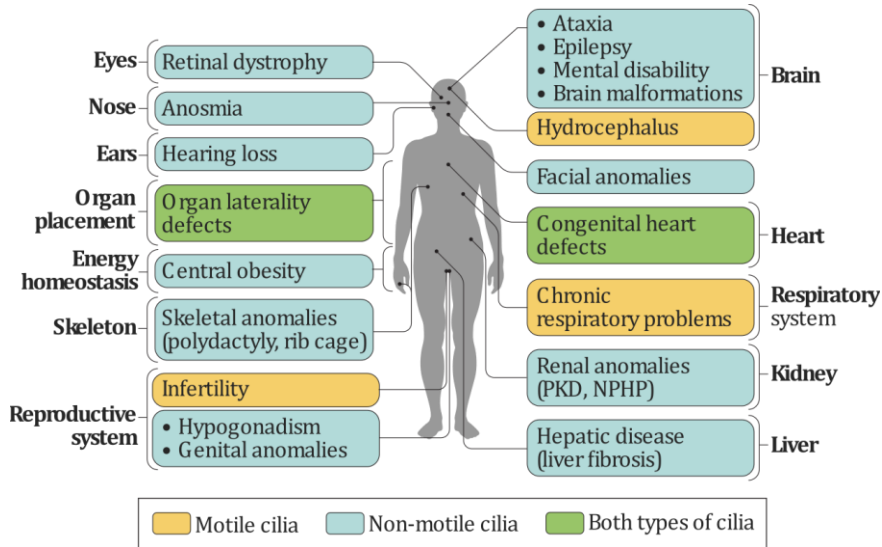
⌃ Figure I2.7. Ciliopathies impact most human systems and organs.
Coloured squares gather the phenotypic manifestation for several cilia-associated human diseases, indicating the organ or system affected. A colour code indicates the type of cilium being affected: primary non-motile (blue), motile (red) or both (green). NPHP, nephronophthisis; PKD, polycystic kidney disease. Adapted from (Reiter and Leroux 2017).

The complexity of diagnosing a ciliopathy is best illustrated when considering that, prior to the 2000s, only two such diseases were well-established. Those were the cases of the primary ciliary dyskinesia (formerly known as the Siewert's/Kartagener's syndrome) and a certain kind of retinitis pigmentosa labelled as the Usher syndrome (Bloodgood 2009), both conditions impacting the motile cilium and genetically linked to mutations in a number of different genes. Several already described diseases stayed long undiagnosed as ciliopathies due to the impossibility to find their genetic aetiology.

Whole exome sequencing technologies has help in the identification of rare mutations associated to ciliopathies. In other instances, even when the mutated gene was known, lack of understanding of cilia composition and function precluded from assigning the disease as ciliopathy. One example of such case was PKD: the mutation in two genes PKD1 and PKD2 was already linked to this disease; however it was not until the characterization of the *C. elegans* orthologs of these genes as cilia components that PKD was categorized as ciliopathy (Barr and Sternberg 1999).

▼ **Table I2.1. Common clinical features of assorted non-motile ciliopathies.**
Due to shared impairments in ciliary components and processes, non-motile ciliopathies display several common clinical features. Bardet-Biedl syndrome (BBS), nephronophthisis (NPHP), Senior-Löken syndrome (SNLS), Alström syndrome (ALMS), Meckel syndrome (MKS), Joubert syndrome (JBTS), the polycystic kidney diseases (PKD). Adapted from (Tobin and Beales 2009).

Feature	BBS	NPHP	SNLS	ALMS	MKS	JBTS	PKD
Renal cysts	✓	✓	✓		✓	✓	✓
Hepatobiliary disease	✓	✓	✓	✓	✓	✓	✓
Laterality defects	✓	✓			✓		
Polydactyly	✓				✓	✓	
Agenesis of corpus callosum	✓				✓	✓	
Cognitive impairment	✓				✓	✓	
Retinal degeneration	✓		✓	✓	✓	✓	
Posterior fossa defects/encephalocoele	✓				✓	✓	
Skeletal bone defects							
Obesity	✓			✓			

Despite the considerable progress achieved in the past decade, the ciliome remains still uncomplete and, accordingly, a number of conditions scape being diagnoses as ciliopathies. More importantly, an additional layer of complexity arises when considering the contribution of non-coding mutations, that affect gene expression, and are estimated to account for the vast majority of SNPs associated to disease. This has been exemplified by patients diagnosed with already well-stablished ciliopathies such as the *situs inversus* in which no obvious monogenic basis could be identified (Postema et al. 2020). Assessment of pathogenic non-coding variants in human diseases is just starting to be elucidated (Caron, Luo, and Rausell 2019) but some examples were already found within the ciliopathies field. Indeed, a link between ciliary proteins and obesity was established in 2006 when non-coding variants in the BBS2, BBS4, and BBS6 genes associated with obesity in a French–Caucasian population (Benzinou et al. 2006). More recently, whole genome sequencing demonstrating aberrant RNA splicing as a result of deep intronic variants in the DNAH11 (**D**ynein **A**xonemal **H**eavy **C**hain **11**) and CFTR (**C**ystic **F**ibrosis **T**ransmembrane conductance **R**egulator) genes has been used for the definitive diagnosis of cystic fibrosis and primary ciliary dyskinesia, respectively (Ellingford et al. 2018).

When considering the influence that mutations in the non-coding genome may impose on cilia-related conditions, one can envision a scenario in which such mutations modify the binding sites for transcription factors involved in the regulation of cilia-related genes. Consequently, deciphering the transcriptional regulatory logic behind cilia

establishment, maintenance and functionality will first help us complement the ciliome through the identification of target genes for such TFs and will also allow us to assess the biological meaning underlying the non-coding mutations in the regulatory genome.

Caenorhabditis elegans and its ciliated system

Animal survival strongly relies on the nervous system capacity to reliably perceive and integrate a myriad of environmental and physiological cues. Despite its relatively small and compact nervous system, the nematode *C. elegans* is able to elicit complex behavioural and physiological responses, including locomotion; foraging; feeding; defecation; egg laying and the acquisition of metabolic resistance states. Sensory responses to mechanical and chemical cues, among others, are also displayed by these nematodes, as well as much complex and elaborated behaviours such as male mating; social behaviours; learning; memory and even choice based on previous experience (Rankin 2002).

C. elegans as a model organism.

Outstanding contributions by South African biologist Sydney Brenner (13 January 1927 – 5 April 2019) tend to fade the records for the pioneering works that were performed into the *C. elegans* field almost a century before he did (see Nigon and Félix 2017 for a historical perspective on *C. elegans* research). Still, it was Brenner's foresight that led to the status of a premier model organism that "worms" hold today.

C. elegans is a small and transparent free-living soil nematode (roundworm) that distributes worldwide, predominantly in humid temperate areas (Frézal and Félix 2015). It survives by feeding on microbes, primarily bacteria, originated by decomposing plant material. *C. elegans* reference strain N2, nowadays widely used as a model organism, can be considered a laboratory animal. It was originally isolated in Bristol, England, and was continuously cultured in the laboratory for many years before appropriate freezing methods were developed (Nigon and Félix 2017). In the laboratory, *C. elegans* is cultivated on agar-filled plates seeded with a bacterial lawn of *Escherichia coli*. This food source was formerly used because it was already available in many laboratories, not because its natural association with wild *C. elegans*.

C. elegans has an unsegmented, cylindrical body shape that is tapered at the head and tail ends. Its body plan consists of an outer tube (or body wall) composed by the cuticle, hypodermis, excretory system, neurons and muscles and an inner tube comprising the pharynx, intestine and the adult gonad. Both tubes are separated by the pseudocoelom in which the pseudocoelomic fluid lies (see Altun and Hall 2009 for detailed anatomy). Wild type worms have two sexual forms: self-fertilizing hermaphrodites and males (see **Figure I3.1**). Both sexes are diploid for five autosomal chromosomes and differ in that hermaphrodites have two X chromosomes while males have only one (hence being referred as XO). Males are produced through a rare meiotic non-disjunction of the X chromosome that happens in 0.1 to 0.2% of the self-fertilized progeny of a hermaphrodite and is induced under adverse conditions such as increased temperature. Body plan of hermaphrodites arise from an invariant cell lineage finally leading to 959 somatic nuclei in the adult. Adult male body plan consists of 1031 somatic nuclei of which about 650 are identical to those in the hermaphrodite (Sulston, Albertson, and Thomson 1980;

Sulston and Horvitz 1977).

XX hermaphrodite
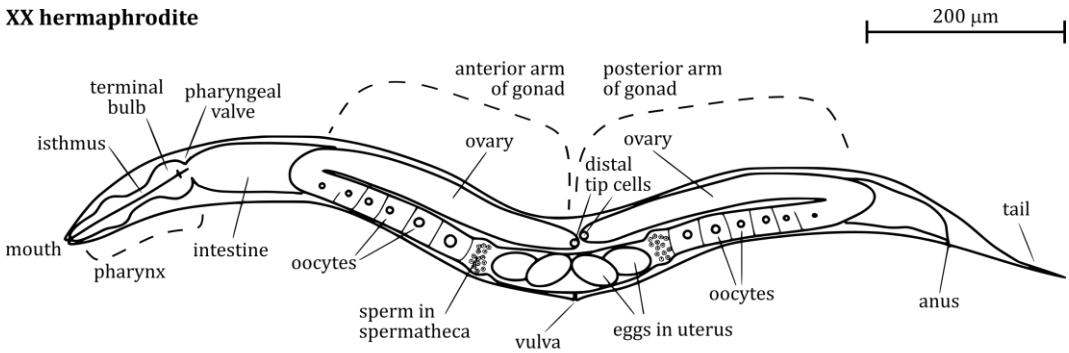
X0 male
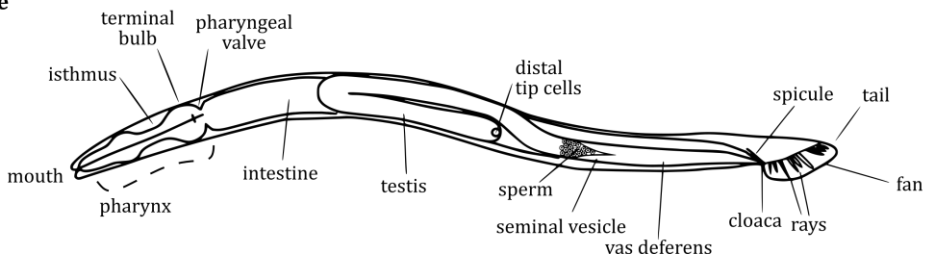
⌃ Figure I3.1. Overview of the anatomical features of adult *Caenorhabditis elegans*.
Major anatomical features of an adult hermaphrodite (top) and male (bottom) of *C. elegans*, viewed from the left side. Notice that gonads of both sexes are portrayed in the foreground merely on illustrative purposes. In the animal, the anterior arm of the gonad runs along the right side with the intestine running on the left. Adapted from (Zarkower 2006).

Newly hatched *C. elegans* larvae are 0.25 millimetres long and adults are 1 millimetre long, with males being slightly smaller than hermaphrodites. Animals capable to access food resources progress to adulthood through 4 larval stages, commonly referred as L1, L2, L3 and L4, reaching adulthood after 2 days at 22 Celsius degrees. At this stage, hermaphrodites are capable to produce offspring; eggs develop *in utero* for 150 hours and then take approximately 9 hours *ex utero* before hatching. However, under stressful conditions such as overpopulation, limited food resources or high temperatures, an alternative developmental life cycle occurs.

From the L1 stage animals shift to a stage termed *predauer* (or L2d), that is then followed by a metabolic resistance diapause stage called *dauer*. This alternative L3 stage allows the animals to survive in the absence of food up to 4 months. If favourable conditions return, *dauer* larvae resume feeding and development (Byerly, Cassada, and Russell 1976).

Genetic and environmental amenability of *C. elegans* has allowed the achievement of several milestones further positioning the model as an attractive choice for the dissection of neuronal circuitries. Starting in the

late 1970s, the whole developmental cell lineage of the worm was resolved (Sulston et al. 1983; Sulston and Horvitz 1977); this allowed for the study of cell-fate establishment and/or effects of mutations and environmental factors at the single cell resolution. A second milestone was achieved by the reconstruction by electron microscopy of the nervous system and its connectivity (White, J.G., Southgate, E., Thomson, J.N. and Brenner 1986). Years later, *C. elegans* became the first multicellular organism for which its genome was both fully sequenced and annotated (*C. elegans* sequencing consortium 1998). This was (in part) made possible thanks to the compact size of the worm's genome: 100 megabases compared to roughly 3,100 megabases of the human genome. Importantly, about 83% of *C. elegans*' genes have a human homolog and, conversely, 70% of human genes find an ortholog in the worm (Lai et al. 2000).

General architecture of the ciliated system.

In cells of higher vertebrates, cilia are nearly ubiquitous organelles but in the case of *C. elegans* cilia are only present in a sub-population of sensory neurons. The nervous system of an adult *C. elegans* hermaphrodite is integrated by its somatic system, containing 282 neurons, plus its pharyngeal system, containing 20 neurons embedded in the pharyngeal muscle (Ward et al. 1975; Sulston and Horvitz 1977; Sulston et al. 1983; White et al. 1986). The nervous system of the worm is thus composed by a total number of 302 neurons, mostly located in the head of the animal. Considering anatomical features and synaptic connectivity, neurons in *C. elegans*

are classified as sensory neurons, interneurons, motor neurons or polymodal neurons (when they perform more than just one of the aforementioned three functions) (Altun and Hall 2011). From those 302 neurons, males and hermaphrodites share 294 cells; consequently, hermaphrodites possess 8 sex-specific neurons. Males, on the other hand, contain 93 additional sex-specific neurons distributed along the ventral part of the body and tail (see **Figure I3.2**) (Serrano-Saiz et al. 2017). Each cell in *C. elegans* has its own name; in the case of neurons, it consists of either two or three uppercase letters indicative of their class. In certain cases, a number is added after the neuron name to indicate neuron number within a given class. Most neurons in *C. elegans* are bilateral pairs, noted by L (left) and R (right). When the same neuron class is located dorso-ventrally a D (dorsal) or V (ventral) is added following its name.

Sixty of the sex-shared sensory neurons in the adult *C. elegans* project a primary non-motile sensory cilium from their dendritic endings (see **Table I3.1**), and all except two (AQR and PQR) belong to bilateral pairs (Ward et al. 1975; Perkins et al. 1986; Doroquez et al. 2014). Some authors list the non-bilateral neuron PVR as ciliated, but the presence of a cilium is not consistent among different animals (Hall and Russell 1991). 53 of these ciliated neurons locate in the head, 2 neurons (PDE) are found in the posterior part of the midbody and 5 neurons (the PHA and PHB pairs plus PQR) are located in the tail. Adult males possess 48 additional ciliated neurons, and all but two (HOA and HOB) belong to bilateral pairs. From those 48 additional ciliated neurons, four (CEM) locate in the head of the animal and the rest of them

are found in the tail (RnA, RnB, HOA, HOB, PCA, SPD and SPV, with "n" spanning from 1 to 9) (Sulston et al. 1980).

XX hermaphrodite

200 μm

Ventral nerve cord: 6 hermaphrodite-specific neurons
Midbody: 2 hermaphrodite-specific neurons

X0 male

181 sex-shared neurons

Ventral nerve cord: 69 sex-shared neurons
Midbody: 12 sex-shared neurons

32 sex-shared neurons

6 male-specific neurons

Ventral nerve cord: 19 male-specific neurons

68 male-specific neurons
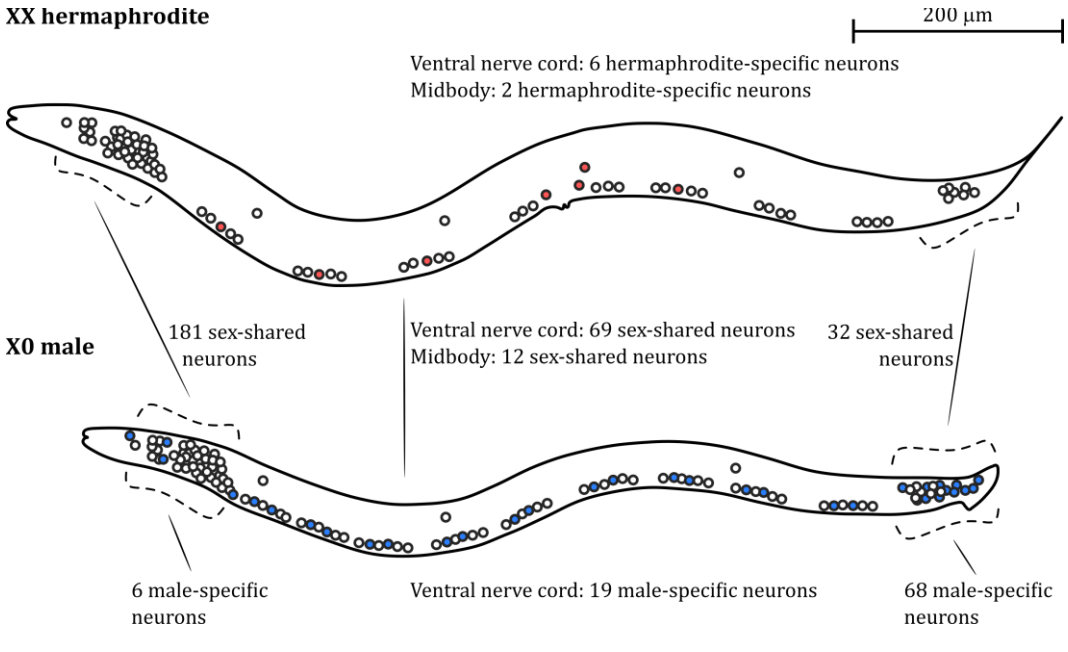
⬆ **Figure I3.2. Overview of the nervous system of adult *Caenorhabditis elegans*.**
Sex-shared nervous systems of adult *C. elegans* consist of 294 neurons; colourless circles portray a left lateral view of these shared neurons in the hermaphrodite (top) and the male (bottom). Adult hermaphrodites contain 302 neurons; red circles indicate 8 hermaphrodite-specific neurons. Adult males possess 387 neurons; blue circles depict 93 male-specific neurons. Adapted from (Serrano-Saiz et al. 2017)

C. elegans perceive environmental and internal cues by employing a combination of 24 sensilla and various isolated sensory neurons (**Figure I3.3**. The sensillum, the sensory organ of the worm, is composed by the sheath and the socket cells (two types of glia-like structural cells) plus one or more ciliated neurons (Ward et al. 1975; Perkins et al. 1986).

The major sensory organs of the nematodes are the amphids, a lateral pair of sensilla located in the tip of the head. In *C. elegans*, each amphid contains the dendrites of the 12 amphid neurons, namely: ASK, ASL,

ASI, AWB, ASG, AWA, ASE, AFD, ADF, ASH, AWC and ASJ. The soma of these neurons are located surrounding the second bulb of the pharynx and from there they project their dendrites to the tip of the head (see **Figure I3.3**). As the projections progress anteriorly, a sheath cell creates a channel engulfing and protecting the endings. The AFD dendrites remains embedded within their corresponding sheath cells and so it does its small cilium and the approximately 50 microvilli emerging from it (see **Figure I3.4**). Dendrites for the rest of the amphid neurons go through a matrix-filled invagination in the most distal part of the sheath cell (see **Figure I3.4**).

Projections from amphid winged neurons AWA, AWB and AWC exit from the invagination, returning to the sheath cell, finally embedding their winged-shaped membranous cilia into it (see **Figure I3.4**). Cilia for the rest of the amphid neurons pass thru a ring-like structure created by the socket cell, thus making direct contact to the external environment. These channel amphid neurons possess cilia shaped as single rods in the case of the ASE, ASG, ASH, ASI, ASJ and ASK neurons or pairs of rods in the ADF and ADL. Length of amphid cilia range from the 7.5 μm found in channel amphid neurons to the 1.5 seen in the short cilium of the AFD (Ward et al. 1975; Ware et al. 1975; Perkins et al. 1986).
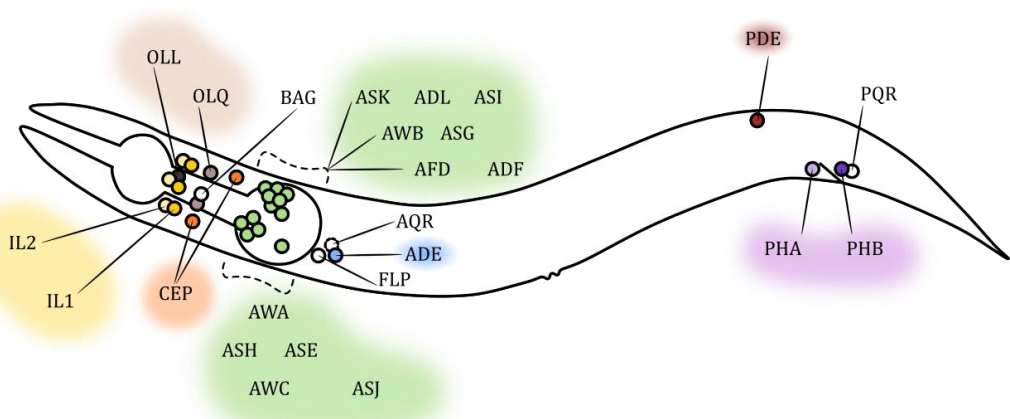
▲ **Figure I3.3. Overview of the sex-shared ciliated nervous system of adult *Caenorhabditis elegans*.**

Left lateral view of the sex-shared ciliated nervous systems of the adult *C. elegans* hermaphrodite. Circles indicate the position of neurons' soma whereas colours depict their sensilla classification in the worm. Inner labial sensilla (in yellow) are composed by the IL2 and IL1 classes, both constituting 3 bilateral pairs integrated by 6 neurons each. IL2 locate anteriorly to IL1. Outer labial sensilla (in brown) are composed by the OLL and OLQ classes, both constituting bilateral pairs integrated by 2 and 4 neurons each, respectively. Cephalic sensilla are composed by the CEP class, integrated by 4 neurons in two bilateral pairs. Amphid neurons (in green) are distributed in 12 different classes, all composed by bilateral pairs located surrounding the second bulb of the pharynx. Anterior (in blue) and posterior (in red) deirid sensilla are composed, respectively, by the bilateral class pairs ADE and PDE. The phasmids (in purple) are composed by the PHA and PHB classes, both constituting bilateral pairs integrated by 2 neurons each. Uncoloured neurons do not associate with any of the worm' sensilla. Those classes are the bilateral pairs BAG and FLP and the non-bilateral neurons AQR in the head and PQR in the tail. *C. elegans* possess 24 sensilla made up by the following classes: IL2 and IL1 (6), OLL (2), OLQ (4), CEP (4), Amphids (2), ADE (2), PDE (2) and Phasmids (2).

Outer labial, inner labial and cephalic sensilla also locate in the tip of the head, arranged in radial symmetry (see **Figure I3.5**). Soma of inner labial neurons IL1 and

IL2 locate anteriorly to those of the amphids, surrounding the posterior part of the first bulb in young adult animals, and both classes are composed by 6 cells distributed in 3 bilateral pairs (dorsal, lateral and ventral). In each set, projections from both classes progress together anteriorly and are engulfed by the same sheath cell. Once their cilia go through the socket cell, IL1 cilium results embedded in the subcuticle whereas IL2's is exposed to the environment. These six final inner labial sensilla terminate each in the lips-like protuberances seen in the tip of the worm's head (Ward et al. 1975; Ware et al. 1975; Perkins et al. 1986). Cell bodies for the 4 outer labial neurons OLQ locate in symmetrical quadrants surrounding the isthmus of the pharynx. Their dendrites progress anteriorly, get engulfed by their corresponding sheath cells and their cilia trespass their socket cells to terminate embedded in the cuticle (Ward et al. 1975; Ware et al. 1975; Perkins et al. 1986). In the cephalic sensilla, cephalic neurons CEP proceeds analogously. This neuron class is composed by 4 neurons arranged in two bilateral pairs, dorsal (located close to the second bulb of the pharynx) and ventral (found in the middle part of the isthmus). Their cilia end in a parallel position of those from the OLQ, also embedded in the cuticle (Ward et al. 1975; Ware et al. 1975; Perkins et al. 1986). Last two sensilla found in the head are formed by the outer labial neurons OLL. These two bilateral cells locate close to the IL classes and project dendrites whose cilia terminate in sensilla placed ventrally to those of the amphids and embedded in the cuticle (Ward et al. 1975; Ware et al. 1975; Perkins et al. 1986).

Two anterior deirid sensilla locate laterally, along the alae, at the posterior part of the head. Somas of the ADE neurons lie posteriorly and ventrally to the terminal bulb (see **Figure I3.3**). From there they project their dendrites perpendicular to the body wall. The distal part of the dendrites become protected by the corresponding sheath cell and the rod-shaped cilia trespass the socket cell to make direct contact with the subcuticle just below the alae (Ward et al. 1975; Perkins et al. 1986).

Two posterior deirid sensilla are found in the posterior part of the mid body. Cell bodies of the bilateral PDE neurons - born post-embryonically at the L2 stage - lie halfway between the vulva and the tail (see Figure B3.3), dorsally to the alae. The sort dendrite they project gets engulfed by a sheath cell and their cilium goes through the socket cell to terminate embedded in the subcuticle (Sulston, Dew, and Brenner 1975; Ward et al. 1975; Ware et al. 1975).

Two phasmid sensilla locate laterally at the end of the animal's tail. Cell bodies for the left and right pairs of PHA and PHB neurons locate adjacent to each other just posterior to the anus (see Figure B3.3). From there they project their dendrites posteriorly, being engulfed (similar to channel amphid cilia) by the same sheath cell but trespassing two socket cells (one wrapping the other). Phasmid cilia, also shaped as single rods, terminate exposed to the outer environment (Sulston et al. 1980; Hall and Russell 1991). Phasmid and certain amphid neurons share the capability to uptake fluorescent dyes when living animals are place in the corresponding solutions. The mechanism by which the neurons stain is not fully understood,

since not all the amphid neurons uptake the dye, but cell geometry, membrane properties and different rates of dye uptake and leakage depending on membrane potential or intracellular pH have been proposed (Perkins et al. 1986). 5-fluorescein isothiocyanate (FITC) is known to label the amphid ADF, ASH, ASI, ASJ, ASK, and ADL channel neurons plus the phasmid PHA and PHB. When using lipophilic fluorescent dyes such as the DiD (see full compound name in the Abbreviations Section), all same neurons but the ADF get stained. Instead, the AWB neuron uptake those dyes **(Table I3.1)** (Collet et al. 1998)
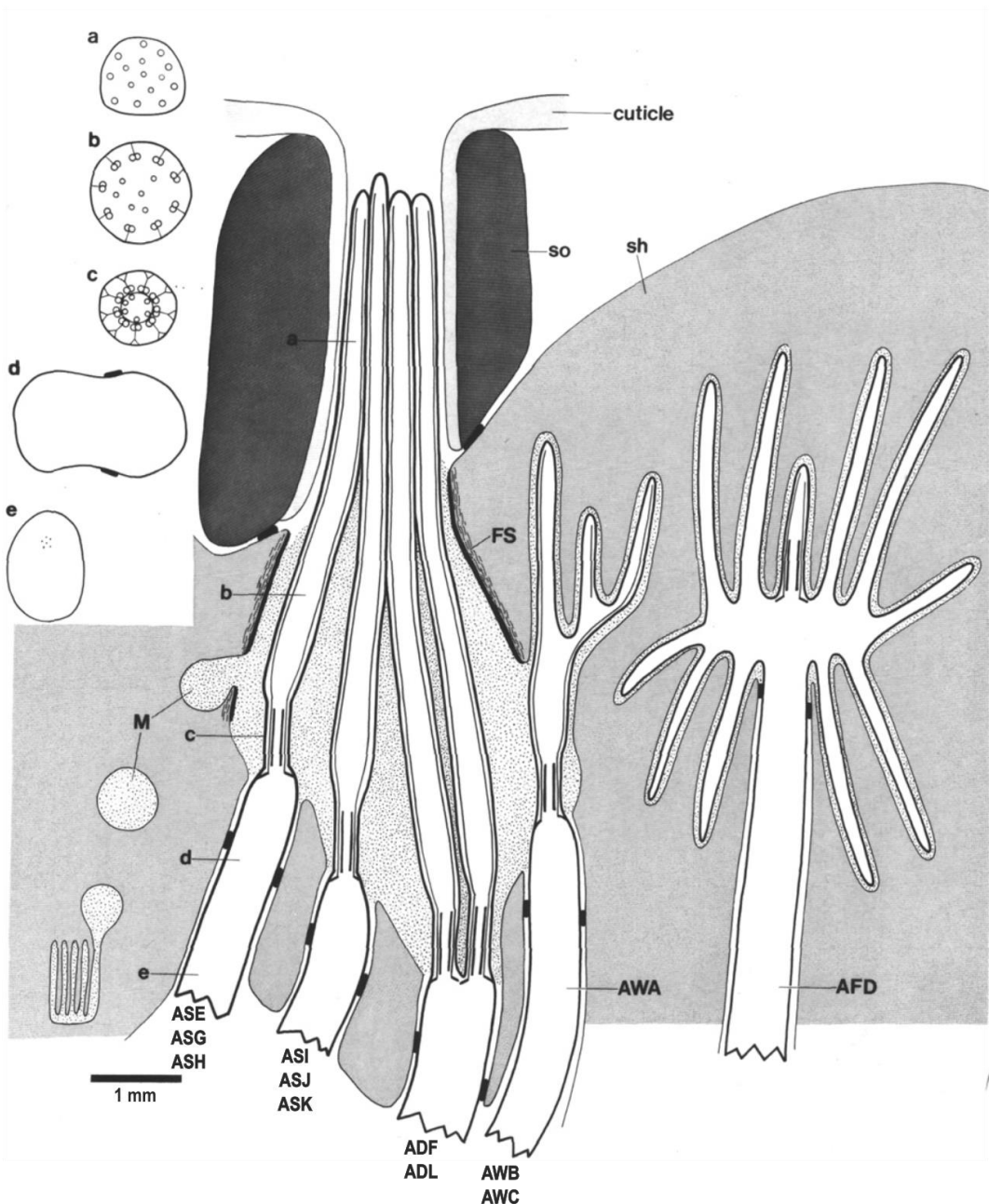
◀ **Figure I3.4. Longitudinal section of the amphid sensillum of *C. elegans*.**

Illustration of the amphid sheath cell (light grey) engulfing the dendrites of the amphid ciliated neurons. To clarify the view, only 3 amphid channel dendrites are portrayed, but corresponding trajectories of all 8 neurons is indicated by name. Analogously, trajectories of amphid winged neurons are indicated, although only AWA cilium is represented. AFD ending, fully embedded in the sheath cell, shows a lateral view of 10 microvilli. All amphid neurons except the ADF trespass an invagination of the sheath cell filled with matrix fluid coming from matrix-filled vesicles (M) originated in the Golgi. In the left part of the figure, cross sections from different parts of the ciliated endings portray the structure of: distal segment (a), middle segment (b), transition zone (c), neuron/sheath junction (d), and main dendrite in the papillary nerve (e). so: socket cell; sh: sheath cell; FS: filamentous scaffold. Adapted from (Perkins et al. 1986)
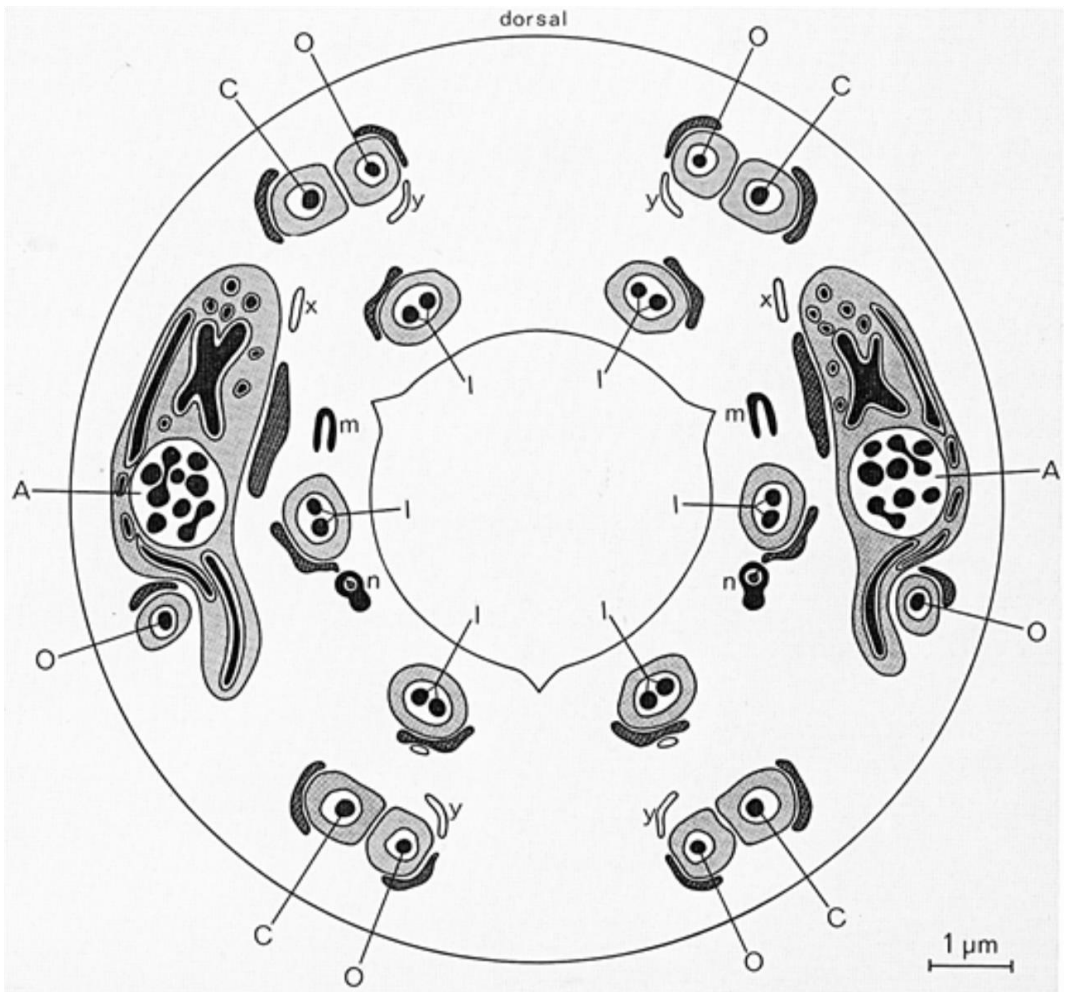
◄ **Figure I3.5. Arrangement of the 18 sex-shared head sensilla of *C. elegans*.**
Çross section of the *C. elegans* head just prior cilia (in black), engulfed by their sheath cells (in light grey), trespass their socket cells (in dark grey). External cercle represent the limits of the animal's body; internal cercle represents the external limit of the pharynx. **A: amphid sensilla**. Central cercle shows the endings of channel amphids neurons inside the invagination of the sheath cell. Already embedded in the body of the sheath cell, parts of the membranous wings of AWA, AWB and AWC cilia appear as well as microvilli from the AFD neuron. Amphid socket cells progress internally in parallel. **I: inner labial sensilla**. Dorsal, lateral and ventral sensilla distribute radially surrounding the limits of the pharynx. Each sheath cell protects the endings of one IL1 and one IL2 neurons. Corresponding socket cells progress externally in parallel. Processes of accessory cells, here termed as m and n, are also represented. **C: cephalic sensilla.** Dorsal and ventral sensilla progress anteriorly between inner labial processes and the external limits of the body. **O: outer labial sensilla.** Dorsal and ventral sensilla from the OLQ neurons run in parallel of those from the CEP neurons. Sensilla from OLL neurons locate laterally, ventral to the amphids. x and y represent non-sensillar cells. Modified from (Ward et al. 1975)

.

Four additional ciliated classes are found that do not associate with sensillar structures: AQR, PQR, BAG and FLP. Very little is known about the ultrastructure of the unique non-bilateral ciliated classes AQR and PQR, both born post-embryonically in the mid L1 stage. AQR locates near the pharynx, in the right side of the animal (see **Figure I3.3**), and its ciliated ending is free and directly exposed to the pseudocoelom (Altun and Hall 2011). PQR cell body locates in the left part of the animal's tail, close to PHBL (see Figure B3.3). Its dendrites are wrapped around one the left phasmid socket cells, but its cilium is also freely and directly exposed to the pseudocoelomic body fluid (Altun and Hall 2011; Hall and Russell 1991). Soma of the bilateral BAG neurons locate at both sides close to the pharynx's isthmus (see **Figure I3.3**). Their dendrites project anteriorly, along the processes of IL and OLL sensilla. Their ciliated endings terminate in a characteristic bag-like shape (**Figure I3.6**.) that remains embedded close to the nose behind the cuticle (Doroquez et al. 2014; Perkins et al. 1986; Ward et al. 1975; Ware et al. 1975). Cell bodies of the bilateral FLP neurons are found posterior and ventrally to the second bulb of the pharynx (see **Figure I3.3**). Their dendrites also progress anteriorly, along the processes of IL and OLL sensilla. Uniquely among ciliated neurons, this anterior process extensively branches covering the whole segment of the pharynx (**Figure I3.6**.). One of the branches terminates in a ciliated ending that remains embedded close to the nose behind the cuticle (Doroquez et al. 2014; Perkins et al. 1986; Ward et al. 1975; Ware et al. 1975). The diverse morphologies *C. elegans* cilia display appear represented in **Figure I3.6**.
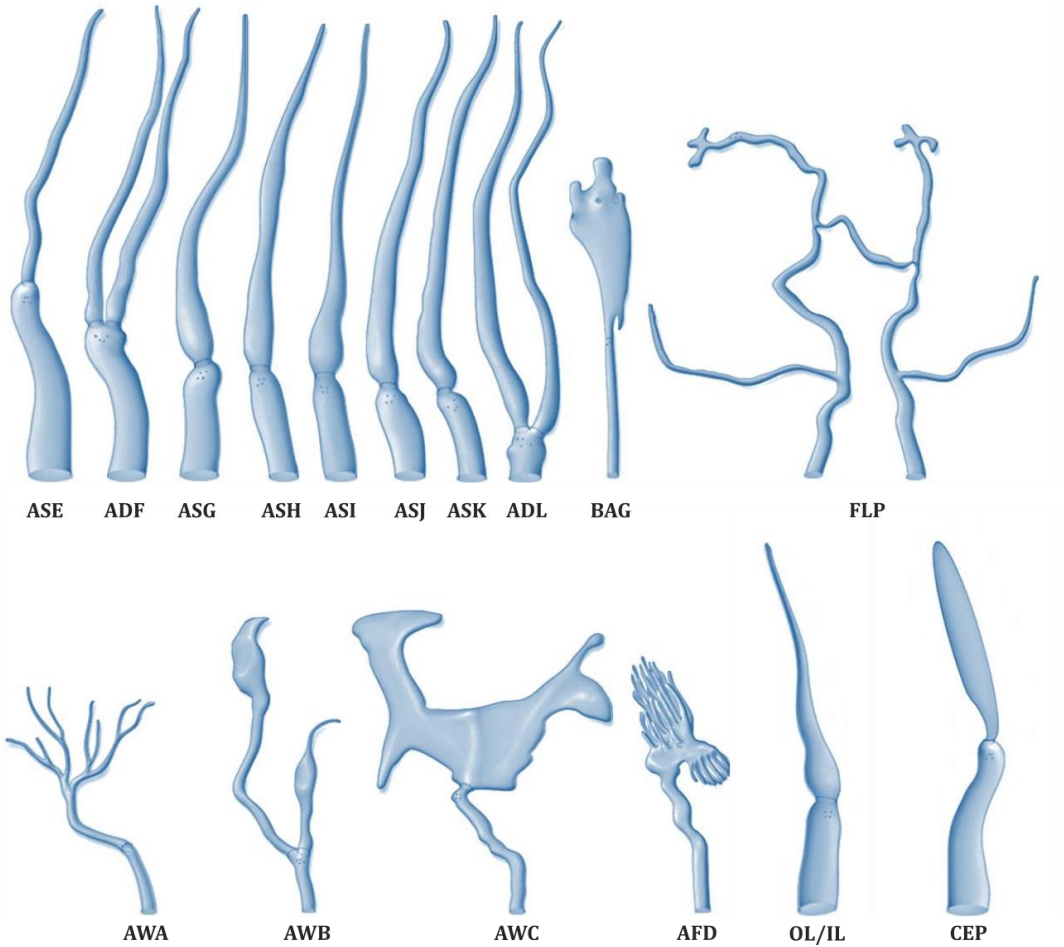
▲ **Figure I3.6. Cilia morphology in *C. elegans*.**
Amphid channel neurons (ASE, ADF, ASG, ASH, ASI, ASJ, ASK, ADL) display a rod-like shaped cilium, double in the case of ADF and ADL. Gas-sensing BAG neurons exhibit a characteristic bag-like cilia ending. Anterior processes of non-sensillar neurons FLP extensively branches covering the whole segment of the pharynx, terminating in a ciliated ending close to the nose. Olfactory amphid neurons AWA, AWB and AWC possess winged-shaped membranous cilia. The thermosensory neurons AFD own a very short cilium, but up to 50 microvilli emerge from it. Outer and inner sensory labial neurons (OLL, OLQ, IL1 and IL2) have rod-like cilia very similar to those of the amphid channel neurons. Cephalic neurons CEP show a flattened shovel-shaped cilium. Modified from (Altun and Hall 2010); ultimately based on (Ward 1973).

➤ **Table I3.1. *C. elegans* ciliated sensory system.**
Overview of key features for the 60 ciliated neurons shared between adult males and hermaphrodites. Number of neurons per class and anatomical placement is denoted between parentheses. Direct exposition of neuronal cilia to the outer environment is indicated; when cilia are instead exposed to the inner pseudocoelom a hash symbol (#) appears (Yes#). Occasional dye filling is indicated by and asterisk (*). Ciliated neurons able to uptake DiI normally uptake other analogous dyes such as DiO or DiD. Data from

this table have been gathered from (Inglis et al. 2007) and (Pereira et al. 2015).

Neuron class	Sensilla	Neurotransmitter	Exposed	Embedded	Dye filling
IL2 (6, head)	Inner labial	Acetylcholine	Yes	No	DiI
OLL (2, head)	Outer labial	Glutamate	No	Cuticle	No
IL1 (6, head)	Inner labial	Glutamate	No	Subcuticle	No
OLQ (4, head)	Outer labial	Glutamate	No	Cuticle	No
BAG (2, head)	None	Glutamate	No	Behind cuticle	No
CEP (4, head)	Cephalic	Dopamine	No	Cuticle	FITC*
ASK (2, head)	Amphid	Glutamate	Yes	No	FITC, DiI
ADL (2, head)	Amphid	Glutamate	Yes	No	FITC, DiI
ASI (2, head)	Amphid	Unknown	Yes	No	FITC, DiI
AWB (2, head)	Amphid	Acetylcholine	No	Sheath cell	DiI
ASG (2, head)	Amphid	Glutamate	Yes	No	No
AWA (2, head)	Amphid	Unknown	No	Sheath cell	No
ASE (2, head)	Amphid	Glutamate	Yes	No	No
AFD (2, head)	Amphid	Glutamate	No	Sheath cell	No
ADF (2, head)	Amphid	Acetylcholine/Serotonin	Yes	No	FITC
ASH (2, head)	Amphid	Glutamate	Yes	No	FITC, DiI
AWC (2, head)	Amphid	Glutamate	No	Sheath cell	No
ASJ (2, head)	Amphid	Acetylcholine	Yes	No	FITC, DiI
FLP (2, head)	None	Glutamate	No	Behind cuticle	No
ADE (2, head)	Anterior deirid	Dopamine	No	Subcuticle	FITC*
AQR (1, head)	None	Glutamate	Yes#	No	No
PDE (2, body)	Posterior deirid	Dopamine	No	Subcuticle	FITC*
PHA (2, tail)	Phasmid	Glutamate	Yes	No	FITC, DiI
PHB (2, tail)	Phasmid	Glutamate	Yes	No	FITC, DiI
PQR (1, tail)	Phasmid	Glutamate	Yes#	No	No

Overview of sensory perception mediated by ciliated neurons.

Downstream processes triggered by the action of sensory neurons ultimately determine the complex behaviours exhibited by *C. elegans* in response to environmental or internal cues. Proper functionality of a complex circuitry involves first the sensory neurons, then the connecting interneurons and finally the motor neurons. All components are crucial to exert the suitable response to a given stimulus. Sensory perception allows

the worm to appropriately react to mechanical, chemical, thermal, gaseous and even social stimuli, among others. *C. elegans* is able to navigate its world by employing a compact and relatively simple sensory system. In this system, each specific class of neurons is responsible for the detection of a specific set of cues (see **Table I3.2**).

Perception of mechanical cues.

In *C. elegans*, mechanosensation is detected by both ciliated and non-ciliated sensory neurons, depending on the given stimulus (see (Goodman and Sengupta 2019) for an extensive review on mechanosensation in *C. elegans*). Ablation of specific neurons coupled to the performance of classical paradigms of mechanosensation, in which direct physical stimulation is delivered to the worm, has allowed for the identification of the individual neurons mediating the scape responses elicited by the animals. Initially, harsh touch assays were performed manually by stroking the worms with a metal wire. For gentle touch, an eyelash/eyebrow was used instead (Chalfie and Sulston 1981). Although these classical paradigms are still in use and are regularly applied in the performance of behavioural tests, much more elaborated and automated methods are being employed nowadays (Cho et al. 2017). Direct physical stimulation can now be applied in a device-assisted fashion, thus allowing to fully control both intensity, duration, and exact anatomical place of stimulus. This distinction is not trivial, since it has been long known that changes in these parameters (difficult or impossible to assess when performing these assays manually) activate different sensory neurons (see Chalfie et al. 2014).

Gentle touch to the body is mainly sensed by non-ciliated sensory neurons. ALML, AMLR and AVM elicit an escape response if the animal is touched in the anterior part of its body. If this gentle touch is delivered to its posterior part, PLML and PLMR trigger the worm to move forward (Chalfie and Sulston 1981). Although initially discarded from this response, the non-ciliated sensory PVM neuron has been shown to get activated when the worm is gently touch in a microfluidics chip (Cho et al. 2017). A different modality of gentle touch, specifically delivered along the ventral side of the animal's nose evoke a head withdrawal response mediated by the ciliated OLQ, FLP, and IL1 neurons (Hart, Sims, and Kaplan 1995).

Scape response to body harsh touch was initially attributed to the action of the non-ciliated PVD. However, more recent evidence has shown also the activation of the ciliated FLP neurons in response to this stimulus (Chatzigeorgiou et al. 2010). Moreover, laser-mediated ablation studies have revealed that harsh touch specifically delivered to the nose is mediated by ciliated ASH, CEP, and FLP neurons. Although acting in parallel, each neuron has been found to mediate a fraction of the final behavioural output; those being: ASH 45%, FLP 29%, and OLQ 5% (Kaplan and Horvitz 1993). More recent laser ablation experiments (Li et al. 2011), also show the implication in body harsh touch avoidance responses for the ADE, AQR, PHA, PHB, PQR, and PDE ciliated neurons, although the mechanism has not being clarified yet.

▼ **Table I3.2. Main functions of *C. elegans* ciliated sensory neurons.**

Overview of main functions exerted by the 60 sex-shared ciliated neurons of *C. elegans*. Information gathered in this table summarize the data appearing in individual neurons entries from WormAtlas further complemented with (Goodman and Sengupta 2019).

Neuron class	Functions
IL2	Nictation. Putative polymodal (chemosensory and/or mechanosensory)
OLL	Pathogen avoidance. Putative mechanosensory.
IL1	Polymodal (mechanosensory, motor neuron and interneuron). Mediate harsh touch (nose tip), controlling aversive head-withdrawal reflex.
OLQ	Polymodal (mechanosensory and interneuron). Mediate harsh touch (nose tip), controlling aversive head-withdrawal reflex.
BAG	Oxygen and CO_2 sensing. Lifespan regulation.
CEP	Harsh touch (nose tip). Basal slowing response. Texture.
ASK	Chemotaxis. Avoidance. Local search behaviour. Pheromone sensing. Lightsensation (350-470 nm). Electrosensory navigation.
ADL	Repellents detection. NaCl response. Social feeding behaviour. Pheromone sensing.
ASI	Local search behaviour. Chemotaxis. Termotaxis. Pheromone sensing. Modulation of innate immune response.
AWB	Olfactory avoidance. Lightsensation (350-470 nm). Electrosensory navigation.
ASG	Gustatory attraction. *Dauer* entry. Lifespan regulation.
AWA	Olfactory attraction. Sexual attraction (males).
ASE	Gustatory attraction and avoidance. Adaptative food-leaving behaviour. Primary CO_2 sensing.
AFD	Termotaxis. Thermonociception. Locomotion. Primary CO_2 sensing. Social feeding behaviour. Magnetic sensing. Hygrotaxis.
ADF	Gustatory response. *Dauer* entry. Oxygen sensing. Locomotion.
ASH	Main nociceptor. Avoidance. Lightsensation (350-470 nm). Electrosensory navigation. Social feeding behaviour. Modulation of innate immune response.
AWC	Olfaction. Local search behaviour. Turns promotion. Termotaxis. Thermonociception. Electrosensory navigation. Sexual attraction (males).
ASJ	*Dauer* entry and exit. Termotaxis. Lightsensation (350-470 nm). Electrosensory navigation.
FLP	Harsh and gentle touch (nose tip). Termonociception. Hygrotaxis.
ADE	Harsh touch (body wall). Basal slowing response. Texture.
AQR	Oxygen- and minor CO_2- sensing. Social feeding (aggregation and bordering). Suppression of innate immunity.
PDE	Harsh touch (body wall). Basal slowing response. Texture.
PHA	Chemorepulsion. Harsh touch (posterior part). Mate-searching behaviour (males).
PHB	Chemorepulsion. Harsh touch (posterior part).
PQR	Oxygen- and minor CO_2- sensing. Social feeding (aggregation and bordering). Suppression of innate immunity. Mate-searching behaviours in males.

Ciliated dopaminergic neurons CEP, ADE and PDE are responsible for the detection of texture sensation in the worm. This capability to perceive small object assist the animals into food recognition and is the reason behind the slowing of locomotion observed in the worms when entering a bacterial lawn (the so-called basal slowing response) (Sawin, Ranganathan, and Horvitz 2000).

Gustatory and olfactory detection.

C. elegans display a fine taste for a wide range of both water-soluble (gustatory) and volatile (olfactory) chemical compounds. Indeed, 5% of its genome encodes genes devoted to chemical recognition (Bargmann 1998, 2006). Ciliated amphid and phasmid neurons mediate these chemosensory responses, directing either attraction or aversion to specific cues and even modulating the outcoming responses depending on chemical's concentration (Bargmann, Hartwieg, and Horvitz 1993; Bargmann and Horvitz 1991; Troemel, Kimmel, and Bargmann 1997).

Among the gustatory amphid neurons, the ASE pair is the only known class to act bilaterally in a differentially functional fashion: while the ASER (right) neuron preferentially detects chloride and potassium ions, the ASEL (left) neuron detects sodium ions (Pierce-Shimomura et al. 2001). Laser ablation experiments have exposed the main role that the ASE neurons play in the chemotaxis towards water-soluble attractants. Indeed, in the absence of the ASE neurons, the ciliated amphid neurons ADF, ASG, ASI, ASK, and ASJ can perform only a weak residual response to attractive compounds such as Na^+, Cl^-, cAMP, biotin, lysine, and serotonin (Bargmann and

Horvitz 1991).

In addition to its mechanosensory role, the polymodal ASH neurons also modulate chemosensation directing rapid avoidance responses toward a huge range of cues, including high osmolarity, heavy metals, detergents, bitter alkaloids, low pH, and some organic odours In the absence of ASH, minor roles for the detection of some of these chemical repellents can be observed in the ciliated ADL, ASK and ASE neurons (Bargmann, Thomas, and Horvitz 1990; Hilliard et al. 2004; Hilliard, Bargmann, and Bazzicalupo 2002; Kaplan and Horvitz 1993). In the case of heavy metal such as copper or cadmium, appropriate avoidance response is redundantly coordinated by the ciliated amphid pairs ASH, ASK and ADL (Sambongi et al. 1999). Ciliated phasmid neurons PHA and PHB have been shown to modulate chemorepulsion behaviours acting antagonistically to the ASK, ADL and ASH neurons. Together, these amphid and phasmid neurons constitute a circuit in which the integration of head and tail cues elicits the appropriate avoidance response to noxious chemicals such as SDS (Hilliard et al. 2002).

In *C. elegans*, olfaction is mainly performed by the amphid winged ciliated neurons AWA, AWB and AWC. Worms are able to discriminate a wide range of volatile compounds, including alcohols, ketones, esters, pyrazines, thiazoles, aromatic compounds, aldehydes, amines and others (Bargmann et al. 1993). The AWA and AWC neurons mediate responses to attractive volatile odorants; some being detected by only one of the neuronal pairs, others being redundantly detected by both (Bargmann et al. 1993). As in

the case of the ASE neurons, the bilateral AWCL and AWCR neurons are also functionally distinct from each other. However, AWC difference is stablished by the stochastic expression of the *str-2* receptor in one of the neurons (hence denoted as AWC[ON]), which is never expressed in both (determining the so-called AWC[OFF] neuron) (Troemel, Sagasti, and Bargmann 1999). In the worm, individual olfactory neurons are able to sense several distinct volatile compounds, (even discriminating one odour in the homogeneous presence of another), but no neuron is known to detect both attractive and repulsive stimuli at the same time. When it comes to the detections of aversive volatile compounds, the AWB olfactory neurons mediate the long-range avoidance response (Troemel et al. 1997).

Thermosensation

In the wild, *C. elegans* must confront large diurnal and seasonal fluctuations in temperature that are specially challenging for an ectotherm organism able to tolerate a limited temperature range between 12 and 27 Celsius degrees (Hedgecock and Russell 1975). Three pairs of amphid sensory neurons have been implicated in regulating thermosensory navigation behaviours (see also Goodman and Sengupta 2019). AFD neurons are the main players regulating thermotaxis, while AWC and ASI play minor modulatory roles.

In addition to physiological temperatures, *C. elegans* also reacts to noxious heat or cold employing an alternative neuronal circuit in which both ciliated and non-ciliated neurons participate. Ciliated AFD and FLP neurons in the head and non-ciliated PHC neurons in the tail sense noxious heat (Chatzigeorgiou et al. 2010); whereas non-ciliated PVD neurons in

the body respond to acute cold (Chatzigeorgiou et al. 2010).

Sensing of magnetic and electrical fields.

The capability to perceive the Earth's magnetic field is present in several organisms from distinct phyla and allows for the orientation and navigation of long distances. In *C. elegans*, the amphid AFD neuronal pair has been proposed to be magneto-sensory (Vidal-Gadea et al. 2015). Although controversial (Landler et al. 2018; Vidal-Gadea et al. 2018), both the cilium and microvilli of AFD neurons have been shown necessary for the correct orientation of *C. elegans* in a magnetic field (Bainbridge et al. 2016).

Even if the nature of the electro-sensory transduction pathways is yet to be discovered, *C. elegans* exhibits robust responses to external imposed electrical fields. In the presence of a constant linear electrical field, animals move toward the anode (Chrisman et al. 2016; Sukul and Croll 1978). This response seems to be primarily directed by the functionally asymmetric AWC[ON] member of the amphid AWC neurons (Chrisman et al. 2016). In contrast, in the presence of a slowly rotating electrical field, worms move in circles mainly directed by the action of the ASJ neurons. Calcium responses are detected in other ciliated neurons when presenting such stimuli, although weak (ASH) to no defects (AWB, AWC or ASK) in electro-sensory steering are observed when ablating those neurons (Gabel et al. 2007).

Light sensation and higrotaxis.

Even in the absence of eyes, ocellus, or any other light-processing structure, *C. elegans* can detect and avoidantly react to light.

This response is proposed to be elicited to prevent potential damage or severe dehydration due to sunlight exposure in the wild. Evasive behaviour has been found to peak in the short-wavelength light-range covering ultraviolet, blue and violet lights. In fact, long-term illumination at these wavelengths causes first the paralysis of the animals leading to death after approximately 30 minutes (Edwards et al. 2008). Although the exact sensory mechanism is unknown, as well as the identity of the neurons driving the aversive response, ciliated ASJ and ASH neurons have been implicated in the avoidance behaviour to ultraviolet light (Liu et al. 2010).

C. elegans also display higrotactic behaviours, a rather plastic response depending on satiety state and growing conditions (Russell et al. 2014). Migration towards a dry side of a humidity gradient has been shown to depend on the action of both mechanosensory FLP or thermosensory AFD ciliated neurons (Russell et al. 2014).

Oxygen and carbon dioxide detection.

Despite lacking a specialized respiratory system, *C. elegans* is capable of employing oxygen thanks to its diffusion through the pseudo-coelomic fluid that ultimately bathe all the animal's tissues. If allowed to choose, worms prefer oxygen concentrations ranging from 5 to 12% and elicit avoidance responses in environment diverging from those values (Van Voorhies and Ward 2000). As in the case of higrotactic behaviours, preferences towards oxygen concentrations is also a plastic response depending on growing/cultivating conditions (Cheung et al. 2005). In the wild, fluctuations of oxygen levels may reflect bacterial metabolic rates hence indicative of food

sources. Moreover, social behaviour such as aggregation of large numbers of animals is also regulated by oxygen concentration (Gray et al. 2004). In fact, clumping of cultivated animals in the borders of a bacterial lawn is suppressed when shifting from a high oxygen concentration (21%) to a preferred lower one (8%) (Coates and De Bono 2002). Primary oxygen sensing is performed by both ciliated (AQR and PQR) and non-ciliated sensory neurons (URX) (Chang et al. 2006; Cheung et al. 2005; Gray et al. 2004). However, social feeding behaviours in which oxygen levels are implicated are also stimulated by the action of ciliated ASH and ADL and mainly antagonized by the ASI neurons (Chang et al. 2006; Coates and De Bono 2002).

As in the case of other animals, carbon dioxide is also toxic for *C. elegans*. Worms respond to fluctuations in ambient CO_2 moving to places of lower concentration and well-fed animals elicit acute avoidance responses if CO_2 levels exceed 0.5% (Bretscher, Busch, and De Bono 2008). These behaviours are primary directed by the ciliated neurons AFD, BAG and ASE (Bretscher et al. 2011).

Sensory neuroendocrine control.

The role of *C. elegans*' ciliated sensory neurons expand beyond controlling behavioural responses to internal and/or external cues. Indeed, many of the outputs elicited by these neurons are physiological rather than behavioural, since chemosensory neurons secrete several different peptides with the capability to act at long distances thus affecting both neuronal and non-neuronal cells (see Allen et al. 2015; Bargmann 2006).

One key aspect regulated by ciliated

sensory neurons is the entry and exit from the metabolic resistance *dauer* state. During their lives, *C. elegans* secrete a pheromone mixture (the so-called *dauer* pheromone) indicative of population density. Low pheromone concentration coupled with high food resources trigger ADF, ASG and ASI ciliated neurons to block *dauer* entry. Conversely, a high ratio between pheromone and alimentary cues drives ASJ and ASK neurons to promote *dauer* entry. Once the animal stands in the *dauer* state, food-derived stimulus triggers either IL2 neurons to maintain it or both ASJ and AWC neurons to promote its exit. Analogously, the food-type derived cues sensed by different classes of ciliated sensory neurons have been shown to affect both lifespan and developmental rates in the worm (see Allen et al. 2015 for especific references).

Exemplifying its evolutionary conservation, ciliary function and obesity is also link in *C. elegans*. *tub-1*, a TUBBY homolog that affects fat storage across species, is expressed in the worm's ciliated system and also undergoes intraflagellar transport along cilia. Moreover, by employing a parallel neuroendocrine signalling pathway, *tub-1* is also able to regulate lifespan (Mukhopadhyay et al. 2005).

In recent years, an increasing body of evidence has shown that the nervous system in general, and the chemosensory system in particular, play crucial roles in pathogen sensing and regulation of the innate immunity in *C. elegans*. Although the molecular mechanisms orchestrating such connections are only starting to be elucidated (see (Foster et al. 2020) (X. Cao et al. 2017)), multiple neuronal pathways involving several ciliated sensory neurons have been found to regulate intestinal host defence in the worm (Wani, Goswamy, and Irazoqui 2020).

Phenotypes & behaviours of ciliary mutants.

Genetic amenability was one of the key aspects Brenner observed when choosing *C. elegans* as a model system (S. Brenner 1974). Mutant animals can be obtained either by exposure to chemical mutagens or through ionizing radiation. Nowadays, and benefiting from the whole sequenced and annotated genome, these random mutagenic processes can be substituted by the targeted CRISPR technology. Moreover, and thanks to self-fertilization, a mutagenized hermaphrodite can provide mutant homozygous offspring as fast as a week's time. Since the model was originally stablished, a plethora of mutant strains displaying behavioural and structural defects have been isolated. Many of them have been shown to affect cilia functionality and/or differentiation of ciliated neurons (Inglis et al. 2007). *C. elegans* has become since a bidirectional system in which the analysis of defects in ciliated cells has been tracked to the gene level whereas the identification of mutated genes has revealed their implication in the ciliated fate.

One of the most widely and simplest method used to assess cilia structural integrity benefits for the capability of certain amphid and phasmid neurons to uptake fluorescent dyes. Mutant animals displaying dye uptake defects are labelled as dye-filling defective; hence stablishing the so-called dyf phenotype. Mutants for genes coding for BBS or IFT components exhibit this phenotype, but

also alterations in sheath and socket cells produce it (Perens and Shaham 2005). However, not all ciliary mutants display this phenotype, as is the case of *ifta-2* mutants (Schafer et al. 2006).

Although mechanosensation is known to be mediated by the action of ciliated and non-ciliated neurons, <u>mec</u>hanosensory mutants (mec) exist also displaying a Dyf phenotype. That is the case of *mec-1* and *mec-8* mutants (Perkins et al. 1986). Conversely, and as expected for a behavior controlled by ciliated neurons, mutant animals for key structural cilia genes such as *che-2*, *che-3*, *che-13*, *osm-6*, *che-12* and *osm-3* display nose touch defects (Kaplan and Horvitz 1993).

The <u>osm</u>otic avoidance abnormal (osm) phenotype, in which animals fail to avoid regions of high osmolarity, is frequently associated with cilia defects (Perkins et al. 1986). Correspondingly, an abnormal <u>che</u>motaxis (che) phenotype exists, in which animals fail to sense both volatile and non-volatile compounds (Bargmann et al. 1993; Ward 1973). Many of these *osm* and *che* mutants have been found to encode for genes crucial for cilia structure and/or functionality.

The relevance of proper cilia functionality for the achievement of a normal development in the worm is manifested when animals completely lacking sensory cilia constitutively enter into the *dauer* state, stablishing the so-called Daf-constitutive or daf-c phenotype. Much strikingly, other mutants with affected cilia structure (such as *daf-6* and *daf-10*) fail to enter the *dauer* state; hence constituting a Daf-defective or daf-d phenotype (Bell et al. 2006; Perens and Shaham 2005). Related with lifespan control, several ciliary mutants have been found to be long-lived and laser ablation of certain ciliated neurons produces an increased in the lifespan (Apfeld and Kenyon 1999). Since those animal display altered sensory perception, caloric restriction derived for deficient food foraging could explain such phenotype. Further neuroendocrine phenotypes are found connecting sensory cilia and lipid accumulation, since *tub-1* mutants exhibit a two-fold increase in fat content as assessed via the Nile Red analysis (Ashrafi et al. 2003).

Although this introductory block has focused on the sex-shared ciliated system, the 48 male-specific ciliated neurons should not be ignored. Accordingly, males of ciliary mutants display several phenotypes specific of their sex, mainly related to impaired mating behaviors. These defects range from their inability to first locate and then sense the hermaphrodite to a compromised capability for vulvar location, spicule insertion and sperm transfer (Barr and Sternberg 1999). For and extended treatment on male sensory neurons associated with mating behaviours see (Barr and Garcia 2006).

In summary, *C. elegans* ciliated system controls a wide range of sensory-mediated behaviours that are essential for its adaptation to the ever-changing environment.

Transcriptional regulatory logic of cilium formation

Cilia constitute dynamical structures that can undergo assembly or reabsorption in response to the ever-changing physiological conditions that arise during ordinary cell cycle, throughout the developmental processes of an organism or in response to environmental change. Appropriate regulatory mechanisms must then exist to guarantee ciliogenesis in a time- and tissue-specific manner. These regulatory mechanisms consist not only in the correct transport and assembly of the hundreds of cilia components but also the transcriptional activation of the genes coding for the ciliome. Correlations between increased transcription of genes encoding for cilia (or flagella) components and cilia (or flagella) growth were first documented in the flagellated *Chlamydomonas* (Lefebvre et al. 1978) and the ciliated sea urchin (Harlow and Nemer 1987). However, the nature for the signals triggering protein synthesis was then unknown. Years later, several genetic studies performed in metazoans identified the two major families of transcription factors controlling the transcriptional regulatory programs behind cilia assembly.

The RFX family of transcription factors.

Transcription factors of the Regulatory Factor X family (RFXs) were initially identified thanks to their capability to bind a conserved *cis*-element motif –termed motif X – found in the promoter sequences of major histocompatibility complex (MHC) class II genes (Reith et al. 1990). That pioneering work revealed the existence of several characteristic domains within the coding sequences of human RFXs; namely, the activation, the DNA-binding and the dimerization domains (see **Figure I4.1.A**). RFXs were shown to bind DNA either as monomers or dimers forming both homo- and heterodimers. Dimerization was seen to occur prior to DNA binding, although both binding and dimerization domains proved to be fully independent at the functional level (see **Figure I4.1.B**). Truncated polypeptides not containing the dimerization domain still retained their binding ability whereas polypeptides lacking the DNA-binding domain could still dimerize. Reith and collaborators also noted that both subunits of RFXs homodimers were able to bind simultaneously separated X boxes, thus proposing RFX homodimers as crosslinking agents between regulatory sequences located far apart in the DNA.

In the following years, profiting from the strong conservation in the sequence coding for the 76 amino acid DNA-binding domain, new members of the RFX family were identified in several organisms (Emery et al. 1996; Chu, Baillie, and Chen 2010) . This highly conserved DNA-binding domain was shown to belong to the winged-helix family of transcription factors. It was composed by three α-helices (H), three β-strands (S) and three connecting loops (L) arranged in the order H1-S1-H2-L1-H3-L2-S2-W1-S3 (see **Figure I4.2.A**). In this structure, the third loop connected the β-strands S2 and S3 forming the wing W1 of the winged-helix motif (Gajiwala et al. 2000). In contrast with other helix-turn-helix proteins, these authors found that RFX factors employed the so-called wing to

recognize DNA (see **Figure I4.2.B**), instead of the recognition helix typically used by transcription factors of the winged-helix family for that purpose.

A)

Activation   DNA binding   Dimerization

B)
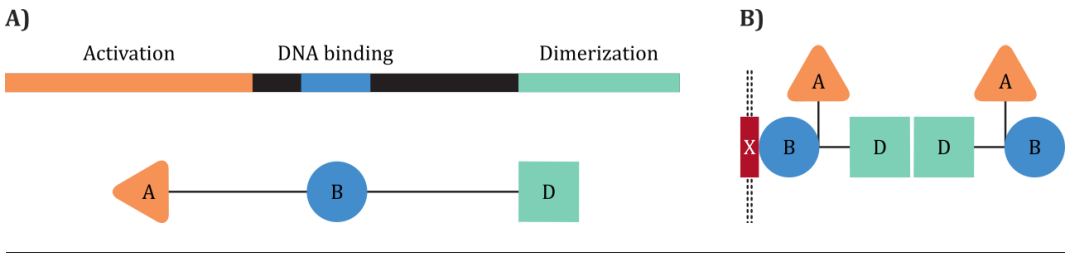
⮝ **Figure I4.1 Original model proposed for the structure and binding of human RFX5.**
A) The cloning of an RFX cDNA originally allowed to identify three structurally and functionally separated domains: the activation domain (A, in orange), the DNA binding domain (B, in blue) and the dimerization domain (D, in turquoise). A linear map of such cloned RFX (top) and a schematic representation of its resulting protein (bottom) are represented. **B)** Independence of the DNA-binding and dimerization domains led to a binding model in which the two subunits of RFX homodimers were proposed not to bind the same X box motif (X, in red), implying that a single subunit was sufficient to perform the binding. Adapted from (Reith et al. 1990).

A)

B)

A
CGTTACCATGGTAACG
GCAATGGTACCATTGC
B
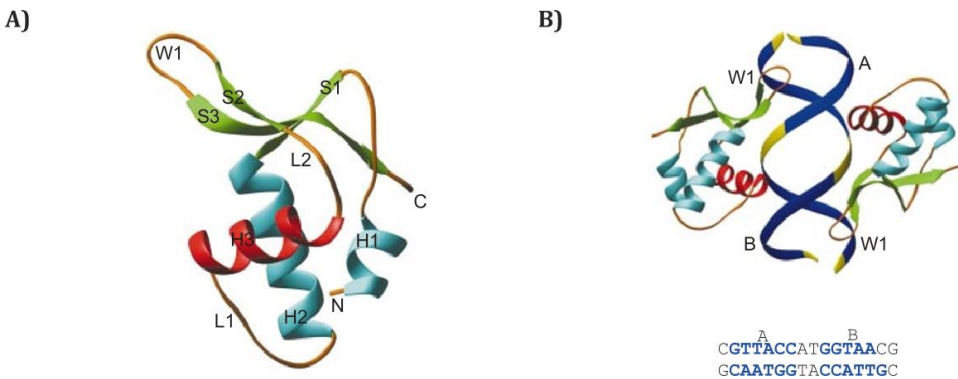
⮝ **Figure I4.2. Crystallographic structure of the DBD of human winged-helix protein RFX1.**
A) Human RFX1 DBD is composed by three α-helices (H), three β-strands (S, in green) and three connecting loops (L) arranged in the order H1-S1-H2-L1-H3-L2-S2-W1-S3. The third loop connecting β-strands S2 and S3 forms the wing (W1) of the winged-helix motif. As an RFX, RFX1 employs W1 to contact DNA, instead of the recognition helix (depicted in red) typically used by winged-helix family TFs. **B)** Diagram of the 2:1 complex formed by human RFX1 and crystallization oligonucleotide. The two halves of the X-box motif are indicated as A and B. W1 contacts with the major groove of one half-site of the symmetric X-box whereas a single side chain from H3 interacts with the minor groove of the other half-site. Adapted from (Gajiwala et al. 2000).

Although previous definitions of consensus binding sites were already proposed for RFX transcription factors, the first systematic, experimental determination for X-boxes was performed by Emery and collaborators (P Emery et al. 1996). These authors found that the recognized site was an imperfect palindromic sequence consisting of two half-sites of 6 base pairs. Although spacers were found ranging from 0 to 3 nucleotides, these half-sites were preferentially separated by 1 or 2 nucleotides. A strong dependency was seen on at least one of the half-sites to match the sequence [(G/A) G (T/C) A A C]. A more degenerated sequence defines the second half-site, thus rendering the final consensus binding site [G T N (G/A) C (C/N)-N_{0-3}-(G/A) G (T/C) A A C].

The first link between RFX TFs and the transcriptional regulation of cilia components was made in the nematode *C. elegans*, in a seminal work from Swoboda and collaborators (Swoboda, Adler, and Thomas 2000). Defective or fully absent cilia was a long-known phenotype for *daf-19* mutant worms; however, the molecular nature of this mutation was unknown (Perkins et al. 1986). The presence of an RFX TF had been already identified in the genome of *C. elegans* (Patrick Emery et al. 1996), but it was not until 2000 when Swoboda identified *daf-19* as the unique RFX member in the nematode. Moreover, Swoboda and collaborators described the presence of X-box sites in the regulatory regions of several cilia component genes, assigning for the first time direct functions to DAF-19/RFX as a terminal selector of the sensory ciliome (Swoboda, Adler, and Thomas 2000).

Soon after, mutant analysis of RFX in *Drosophila* revealed behavioural phenotypes, loss of sensory transduction and defects in ciliary morphology and ultrastructure demonstrated an evolutionary conserved role for RFX in the establishment of sensory cilium differentiation (Dubruille et al. 2002). The pioneering work by Swoboda and collaborators opened the way for successive studies in which the presence of the RFX regulatory signature (the X-box site) was used to identify genes coding for novel components of the cilium. Notable examples include (Li et al. 2004), (Blacque et al. 2005a), (Efimenko et al. 2005) and (Chen et al. 2006) in *C. elegans*, but this methodology was also employed by (Avidor-Reiss et al. 2004) and (Laurençon et al. 2007) in *D. melanogaster*. Despite initial reports assigned a role for RFX in sensory cilia differentiation, later studies in vertebrates (reviewed in Choksi, Lauter, et al. 2014) revealed the essential role that RFX transcription factors play in the regulation of transcription also in motile cilia. Thus, RFX TFS have conserved role as terminal selectors of ciliome most likely through the transcriptional regulation of core components that are key for the assembly of both types of cilia.

RFX members are found in different opisthokonta groups, hence preceding multicellularity and metazoans (de Mendoza and Sebé-Pedrós 2019). Bioinformatics approaches and molecular characterizations (Chu et al. 2010; Piasecki, Burghoorn, and Swoboda 2010) have revealed that RFX TFs are restricted to both ciliated and non-ciliated Unikonta (comprising animals, fungi and amoebozoa). Conversely, no RFX TFs have been found in the genomes of algae, plants,

and various protozoan branches, whether they are ciliated or not (see **Figure I4.3**). Consequently, RFX and ciliary genes are thought to have evolved independently. Since RFX TFs originated early in the unikont lineage (distinctly after cilia had evolved), it has been proposed that they gained secondarily control over ciliary genes probably in the early metazoans (Piasecki et al. 2010). Accordingly, RXFs TFs have additional roles not related to transcription regulation of the ciliome; such as in yeast, where cilia has been secondarily lost and RFXs are involved in cell cycle regulation (Lubelsky, Reuven, and Shaul 2005). Very little is known about which TFs are responsible for the transcriptional regulation of the ciliary genes outside the animal kingdom. Recently, XAP5, a new class of TF that is also present in ciliated and non-ciliated organisms, was shown to regulate flagellar assembly in *Chlamydomonas* (Li et al. 2018).

Species	Group	RFX #
Homo sapiens*	Metazoans	8
Mus musculus*	Metazoans	7
Gallus gallus*	Metazoans	7
Xenopus tropicalis*	Metazoans	7
Danio rerio*	Metazoans	9
Ciona intestinalis*	Metazoans	4
Drosophila melanogaster*	Metazoans	4
Strongylocentrotus purpuratus*	Metazoans	6
Caenorhabditis elegans*	Metazoans	1
Nematostella vectensis*	Metazoans	5
Monosiga brevicollis*	Metazoans	2
Saccharomyces cerevisiae	Fungi	1
Schizosaccharomyces pombe	Fungi	1
Cryptococcus neoformans grubii	Fungi	0
Batrachochytrium dendrobatidis*	Fungi	0
Allomyces macrogynus*	Fungi	1
Physarum polycephalum*	Protists	0
Paramecium tetraurelia*	Protists	0
Tetrahymena thermophila*	Protists	0
Phytophthora ramorum*	Protists	0
Plasmodium falciparum*	Protists	0
Leishmania infantum*	Protists	0
Trypanosoma brucei*	Protists	0
Giardia lamblia*	Protists	0
Chlamydomonas reinhardtii*	Plants & Algae	0
Volvox carteri*	Plants & Algae	0
Arabidopsis thaliana	Plants & Algae	0
Methaococcus jannaschii	Prokaryotes	0
Sulfolobus solfataricus	Prokaryotes	0
Escherichia coli	Prokaryotes	0
Bacillus subtilis	Prokaryotes	0

◄ Figure I4.3. Distribution of RFX members across several phyla.
Number of RFX transcription factors present in the genome of 31 selected species in indicated on the right. Species were selected to provide a wide sampling of the "tree of life" and phylogenetic relationship between each was derived from the "Tree of Life Web Project" (Maddison, Schulz, and Maddison 2007). Organisms marked with an asterisk have been reported to be ciliated. Greenish background for cells in the right column depicts organisms in which putative RFX TFs have been reported. Adapted from (Chu et al. 2010) and updated with (De Stasio et al. 2018) for and eighth human RFX member.

When performing phylogenetic analysis of their DNA binding domains (Aftab et al. 2008; Chu et al. 2010), three major groups of RFX TFs can be distinguished. These groups were first stablished for human RFXs, for which seven members had been identified (Aftab et al. 2008). The first group is composed by RFX1, RFX2 and RFX3; the second by RFX4 and RFX6, while the third contained RFX5 and RFX7. Today, those three groups are commonly referred in the literature by their human name as RFX1-3, RFX4-6 and RFX5-7, respectively. Phylogenetic relationships between the DBD of *M. brevicollis* and previously identified RFX TFs suggest that these three RFXs groups were established before the split between choanoflagellates and metazoans (Chu et al. 2010).

Although RFXs TFs are defined by their highly conserved DNA binding domain, several other functional domains have been described. Those are the cases of the activation and dimerization domains already mentioned, but also the so-called B and C domains (first described by Katan-Khaykovich and Shaul 1998 in human RFX1), which are usually referred as extended dimerization domains. Interestingly, the phylogenetical clustering of the RFX TFs that results from analysing the similarities between their DNA binding domains is consistent with the

appearance of the different functional domains in the different proteins. This way, members from the RFX5-7 usually only contain the DBDs, whereas members from the RFX1-3 contain all 5 different functional domains and members from the RFX4-6 contain all but the activation domain (see **Figure I4.4**). Moreover, the presence of these different functional domains seems to correlate with the processes that the different RFX TFs regulate. As a result, members from the RFX5-7 group that usually only possess the DBD, are generally related to the control of transcriptional cascades not associated with cilia. Those are the cases of *S. cerevisiae* CRT1, controlling cell cycle progression (Huang, Zhou, and Elledge 1998) or human RFX5, related to the immune system through the regulation of the MHC class II genes (Steimle et al. 1995). In contrast, RFXs members from the other two groups have been related to the transcriptional regulation of the ciliary differentiation. Those are the cases (reviewed in Choksi, Lauter, et al. 2014) of *C. elegans* DAF-19, *Drosophila* RFX, and some of the vertebrate RFXs (notably, mammalian RFX2/3/4). However, this correlation between functional domains and processes controlled has its outsiders. For instance, although human RFX1 has been loosely related to cilia (see Choksi, Lauter, et al. 2014), an analogous function to yeast CRT1 in response to DNA

damage is also known (Lubelsky et al. 2005). Conversely, *Xenopus* RFX4 and RFX7 (which cluster with human RFX5-7), are required for the formation of cilia in the neural tube via a mechanism in which RFX7 (member lacking additional domains) regulate RFX4 which, in turns, regulate ciliary genes (Manojlovic et al. 2014).
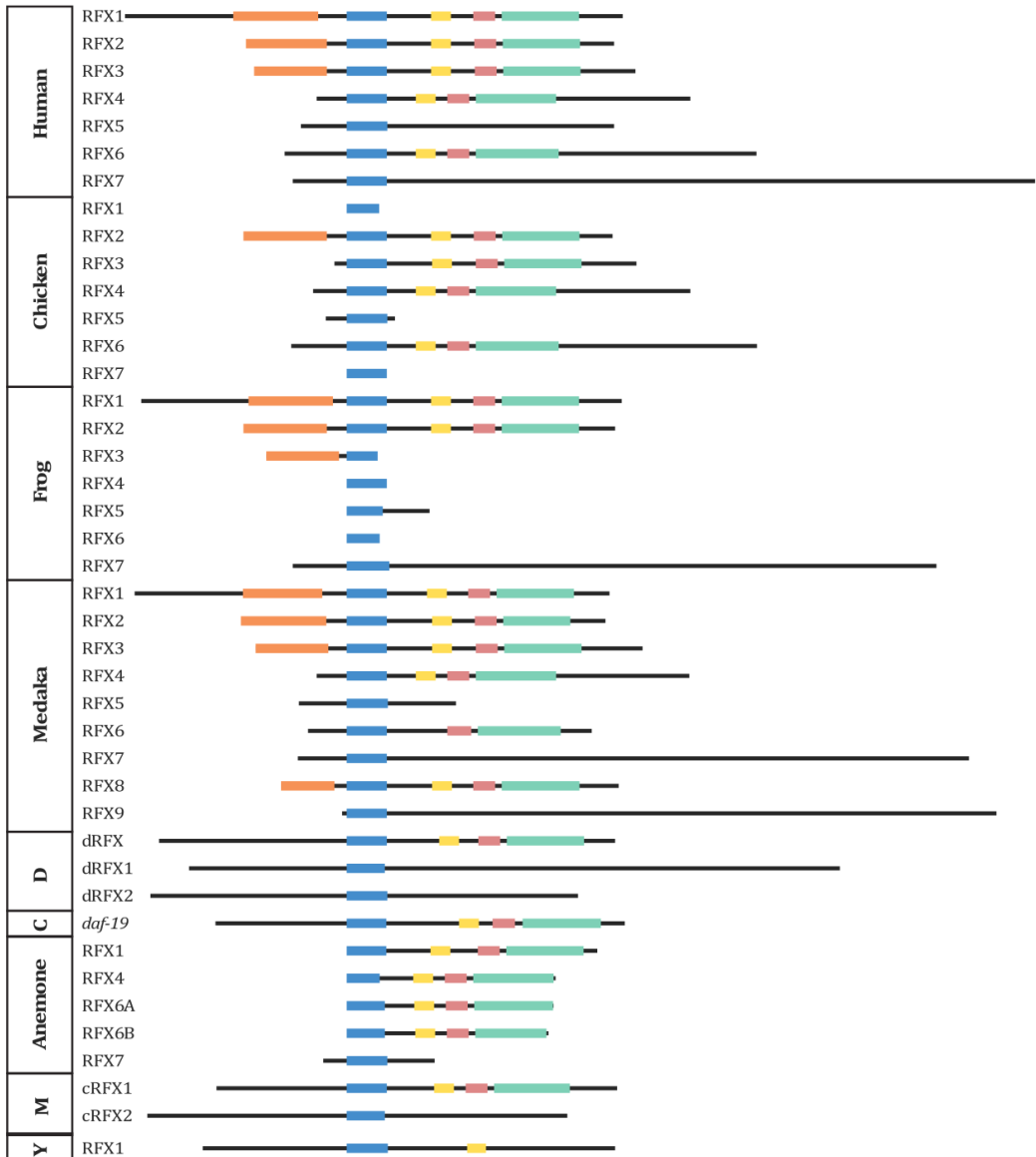
⬆ **Figure I4.4. Predicted protein domains of RFX TFs in representative species.**
Colour boxes represent predicted functional domains for RFX TFs from several organisms. Proteins are depicted as horizontal black lines proportional to length. Domain lengths and location are also proportional to their actual lengths. The defining domain of all RFX TFs is the DNA binding domain, here portrayed in blue. Other domains include: the activation domain (in orange), the B domain (in yellow), the

C domain (in pink), and the dimerization domain (in turquoise). Phylogenetic analysis based on the conservation of the DBD have stablished three groups that are consistent with protein domain composition: (1) TFs in which only the DBD appears, (2) TFs in which all functional domains appear and (3) TFs in which all but the activation domain appear. Y: *S. cerevisiae*, M: *Monosiga brevicollis*, C: *C. elegans*, D: *D. melanogaster*. Adapted from (Chu et al. 2010).

FOX genes, FOXJ1 and the regulation of the motile ciliogenic program.

The first *fork head* gene (*fkh*) was originally identified in a random mutagenesis screen performed in *Drosophila,* where its activity was seen to be essential for the proper formation of terminal structures at both ends of the embryo. In such separated domains, *fkh* mutations caused homeotic transformation forcing the foregut and hindgut to be replaced by ectopic head structures (Weigel et al. 1989). Soon after, the structure of the fork head DNA binding domain was solved by X-ray crystallography (Clark et al. 1993), showing a shape reminiscent of helix-turn-helix proteins with three α-helices and three β-sheets connected to a pair of loops or wings (see **Figure I4.5**). Several secondary structure elements led to a three-dimensional form resembling the shape of a butterfly in which a core originated from the α-helices and the β-sheets were flanked by the two wings. This fact coined the term "winged helix" as a synonym for the forkhead domain.

Profiting from the strong conservation exhibited by the nearly 100 amino acids integrating the DNA-binding domain, several additional members for the *fork head box* family –then termed as the Fox genes– were soon discovered in multiple organisms (early reviewed in Kaufmann and Knöchel 1996). This fast expansion in the number of newly identified genes highlighted the need for an established nomenclature (Kaestner, Knöchel, and Martínez 2000). Accounting for sequence homology within and outside the forkhead domain, Fox genes were categorized into 19 subclasses, ranging from FoxA to FoxS (Jackson et al. 2010). In 2016, documented Fox genes (reviewed in Golson and Kaestner 2016) were found from yeast to humans and varied widely in number among different species. Just to name a few, 44 Fox members were identified in mouse, 50 in human (see Jackson et al. 2010 for orthologous relathionships), 17 in *Drosophila* (Lee and Frasch 2004), 15 in *C. elegans* (Hope et al. 2003) and more than 30 in *Xenopus* (Pohl and Knöchel 2005). Interestingly, no Fox genes were discovered in plants, suggesting that this gene family evolved from a common clade that diversified from plants around 1.5 billion years ago (Wang, Kumar, and Hedges 1999; Lalmansingh et al. 2012). However, more recent studies have revealed the presence of FKH members in organisms belonging to more ancestral clades (such as the Cryptista clade but also in the yet more ancient Metamonada phylum), thus tracing the origin of this family of TFs back to the eukaryotic root and secondarily lost in plants (de Mendoza and Sebé-Pedrós 2019).
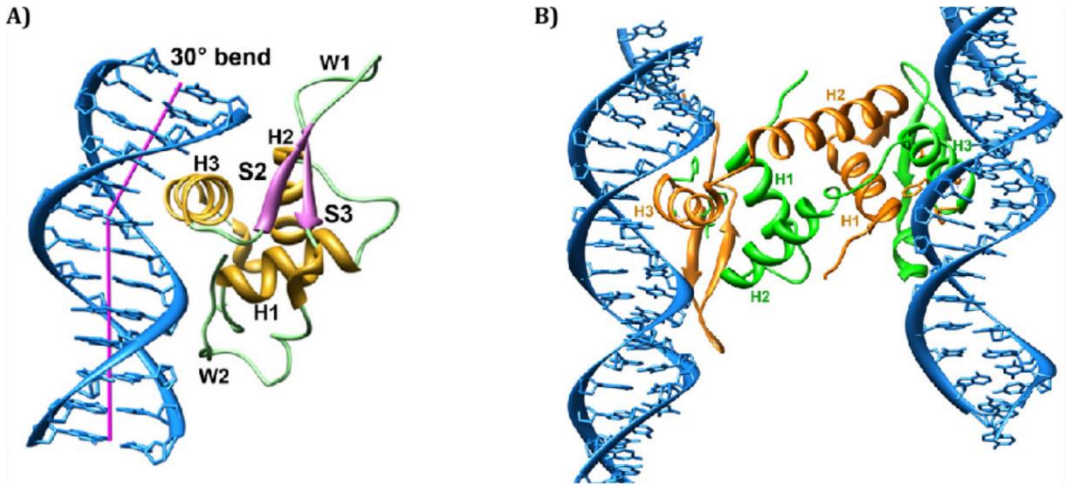
⌃ Figure I4.5. Crystallographic structure of the DBD of FKH proteins.
A) Interaction between FOXA3 DBD (to the right) and DNA (to the left, in blue). The forkhead domain of the FOXA3 TF contains 3 helices (H1–H3) arranged in a helix–turn–helix fold. Helix H3 (in gold) fits into the major groove of the DNA. The two flanking wings W1 and W2 (in green) interact with the DNA minor grooves on either side. FOXA3 binding to DNA induces an overall axial bending of approximately 30°. **B)** FOXP3 DBD dimer (in the center) bound to two different fragments of DNA (at the edges, in blue). The FOXP3 DBD monomer also contains three alpha-helices (H1–H3) arranged in a helix–turn–helix fold. Helices H1 and H2 of one FOXP3 monomer (to the right, in gold) fold onto helix H3 of the second FOXP3 monomer (to the right, in green), and *vice versa*, to form the domain-swapped dimer. Each helix H3 of the dimer fits into the major groove to bridge interactions with two separate DNA molecules. Adapted from (Lalmansingh et al. 2012).

As a consequence of the shared similarities within their DNA-binding domain, Fox members have been found to recognize and bind to a similar motif in the DNA. The first recognition consensus sequence was initially proposed in 1994 (Overdier, Porcella, and Costa 1994). This and several other independent studies all rendered the now common 7 base pairs recognition motif 5' [(G/A) (T/C) (C/A) A A (C/T) A] 3' (see Kaufmann and Knöchel 1996). This consensus sequence is now frequently depicted in the literature through its IUPAC degenerated nucleotide code as the word RYMAAYA. Despite this common recognition motif, Fox genes have been shown to bind DNA with different

affinities and divergent sequences outside their conserved DNA-binding domain have been proposed to provide the binding specificity (see **Figure I4.6**). In addition, other studies have shown that FKH TFs are able to bind very divergent motifs than those traditionally deemed as the canonicals. Indeed, the *in vitro* analysis of the DNA-binding specificities of several Fox proteins assayed through the PBM technology (see **Figure I1.5**) has revealed up to seven different classes of binding sites being bound by these TFs (see **Figure I4.7**). Strikingly, several members of the Fox family have been shown to bind indistinctly to sequences as different as the canonical RYMAAYA and the alternative

motif GACGC using the same DBD (Nakagawa et al. 2013). Recently, crystallographic evidence has shown that this bi-specificity observed in Fox TFs results as a consequence of the different structures adopted by the DNA in each of the complexes formed. This was exemplified by co-crystal structures of FoxN3 DBD, either bound to the canonical or the alternative binding motif, in which the different DNA structures allowed the protein to contact different DNA bases using the same amino acid residues (Rogers et al. 2019). Consequently, as a result of all these variant binding specificities, different FOX proteins display differentiated modes of action, allowing them to control gene expression by acting as pioneer factors, as transcription activators (or repressors) or even as both (see Lalmansingh et al. 2012).
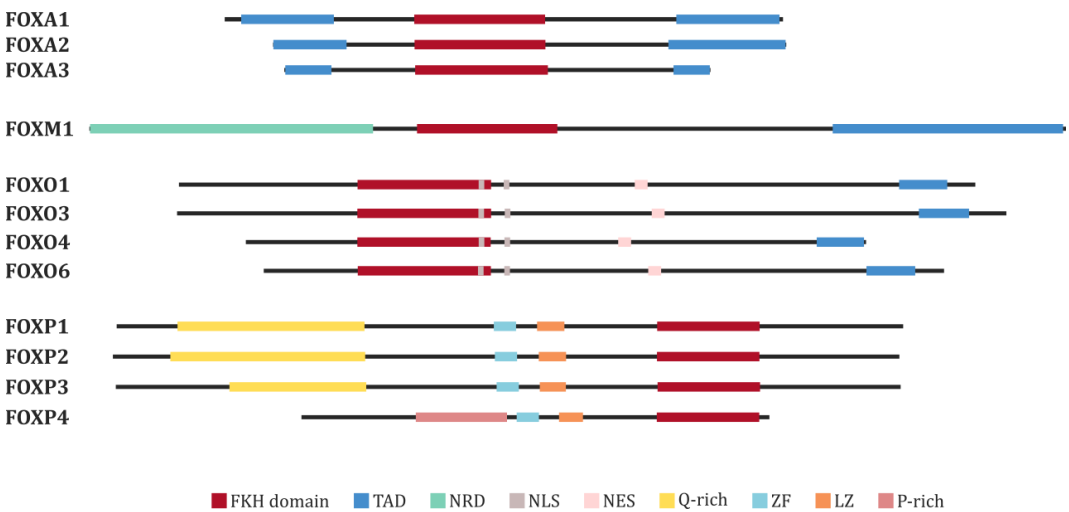
⬑ **Figure I4.6. The domain structure of selected Fox family members.**
Colour boxes represent functional domains found in mouse Fox TFs. Proteins are depicted as horizontal black lines proportional to the protein lengths. Domain lengths and location are also proportional to their actual lengths. TAD, transactivation domain; NRD, N-terminal repressor domain; NLS, nuclear localization signal; NES, nuclear export signal; ZF, zinc finger; LZ, leucine zipper. Adapted from (Golson and Kaestner 2016).

Due to the wide range of biological processes that Fox genes regulate, mutations or malfunctioning of these proteins have been associated to a plethora of developmental defects, metabolic disorders, and other diseases (see **Table I4.1**). Since the 1990s, strong experimental evidence has established FOXJ1 as the master regulator of the motile ciliogenic program in several organisms (reviewed in Thomas et al. 2010 and Choksi et al. 2014). First hints were based on the expression pattern of FOXJ1, which was found to be spatially restricted to mammalian tissues in which motile cilia differentiated. However, the final link was established in murine models using FOXJ1 null mutant animals. Staining of typically motile ciliated tissues such as the proximal respiratory

epithelium, the oviduct or the choroid plexus revealed the complete absence of cilia (or flagella when the sperm was analysed) in the mutant animals. Moreover, and consistent with an impairment in the development of ciliated cells, random left-right asymmetry was also observed (Chen et al. 1998). Soon later, transmission electron microscopy (TEM) of airway cells revealed that targeted deletion of Foxj1 specifically disrupted the characteristic 9+2 axonemal morphology of motile cilia. Interestingly, the 9+0 structure of primary sensory cilia remained intact in

these mutants (Brody et al. 2000). Based on their findings, these authors proposed a classification for the three types of cilia present in the mammalian ciliated cells: (1) non-motile; sensory cilia with a 9+0 axoneme in cells not expressing FOXJ1; (2) motile; 9+2 axoneme, multiciliated cells with a wavelike beating pattern, expressing FOXJ1 and requiring FOXJ1 for axoneme growth; and (3) motile, monocilium with a 9+0 axoneme and dynein arms providing a rotary beating pattern to a cell expressing FOXJ1, but not requiring FOXJ1 for axoneme growth.
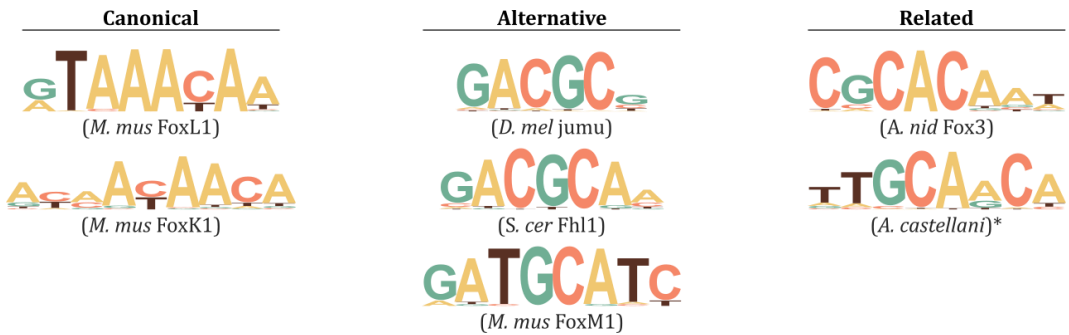
⬆ Figure I4.7. DNA binding-site motifs bound by FKH proteins.
Despite their highly conserved DBD, TFs for the FKH family may display very different DNA binding specificities. A representative member for each reported class of binding is shown. Canonical DNA binding site matching the RYMAAYA consensus is represented by the motif bound by *M. musculus* FoxL1. A common similar variant, represented by the AHAACA consensus, is recognized by mouse FoxK1. Fully divergent binding motifs are found for several members of the FKH family, even simultaneously along the canonical one. The three different classes reported for this alternative binding specificity containing the core motif GAYGC are represented by *D. melanogaster* jumu, *S. cerevisiae* Flh1 and mouse FoxM1. Yet more divergent classes of binding motifs have been found, represented here by the sequence bound by *A. nidulans* Fox3 and a forkhead-related DBD from *A. castellanii*. Modified from (Nakagawa et al. 2013).

At the time of writing, the key role of FOXJ1 in the specification of the motile ciliated fate has been so well established that the term *FIG* has been coined to specify the Foxj1-induced genes (Choksi, Babu, et al. 2014). Today, and similar to what was done

with the presence of X-boxes in the regulatory regions of cilia-related genes, *FIGs* are being employed to decipher both the transcriptional networks associated with ciliogenesis and ciliary function (Mukherjee, Roy, and Chakrabarti 2019) as well as the

transcriptional signature associated with human motile cilia (Patir et al. 2020). Importantly, and in contrast to RFX TFs, the reported role of FKH TFs in directing ciliogenesis has been strictly restricted to motile cilia.

▼ **Table I4.1. Summarized functions of murine Fox family members and their role in human diseases.**

Comprehensive overview for the 44 members of the Fox family of transcription factors found in mouse, including expression patterns and association with developmental disorders and human diseases. Original content taken from (Golson and Kaestner 2016) was assembled using Mouse Genome Informatics (http://www.informatics.jax.org), Online Mendelian Inheritance in Man (www.omim.org) and the Genetics Home Reference (https://ghr.nlm.nih.gov).

Gene	Null mouse phenotypes	Developmental roles	Role in human disease
Foxa1	Early postnatal (P2-P14) lethality; hypoglycaemia; dehydration	Establishes and maintains cellular identity in multiple endodermal-derived tissues including lung, liver, kidney, pancreas and prostate.	Mutated or lost in prostate cancer; presence in ER-negative breast cancer correlates with a better prognosis but presence in ER-positive breast cancer correlates with a worse prognosis
Foxa2	Early embryonic lethality (E9-E10) due to lack of node; somite, neural tube, floor plate and motor neuron defects; heterozygote exhibits Parkinson's-like phenotype due to loss of dopaminergic neuron maintenance	Establishes and maintains cellular identity in multiple endodermal-derived tissues	Downregulated or mutated in multiple cancers of epithelial origin
Foxa3	Hypoglycemia after long fast	Liver-specific gene expression maintenance	
Foxb1	Variable embryonic lethality; high penetrance of postnatal lethality among survivors; motor weakness; midbrain abnormalities; lactation defect among female survivors	Neural tube and neuron process development	
Foxb2	Uncharacterized		
Foxc1	Perinatal lethality; hydrocephalus, edema, and eye, skull, cardiovascular system, kidney and skeletal defects	Skull, tooth, eye, cardiovascular, kidney, and hematopoietic stem cell development	Axenfeld-Rieger syndrome type 3; iris hypoplasia
Foxc2	Embryonic or neonatal lethality; thinned myocardium; abnormal spine and skull; distichiasis; increased brown fat; heterozygotes have lymphatic node and vessel hyperplasia	Cardiac muscle, skeletal, iris and lymphatic system development	Lymphedema-distichiasis syndrome; promotes epithelial-to-mesenchyme transition and metastasis in breast, prostate and colon cancer
Foxd1	Embryonic lethality within 24 h of birth; defects in renal system development	Renal and eye development	
Foxd2	Incompletely penetrant renal hypoplasia	Renal development	

Gene	Null mouse phenotypes	Developmental roles	Role in human disease
Foxd3	Embryonic (E6.5) lethality; lack of primitive streak; failure of gastrulation; expansion of extraembryonic tissue	Neural crest cell development and melanoblast differentiation; embryonic stem cell pluripotency	Autosomal dominant vitiligo
Foxd4	Uncharacterized		
Foxe1	Lethality within 48 h, with cleft palate, absent thyroid, and sparse, kinked hair	Thyroid development	Some mutations associated with thyroid dysgenesis, spiky hair and cleft palate; other mutations associated with thyroid cancer
Foxe3	Fusion of lens and cornea; other lens abnormalities	Lens development; proliferation, cell cycle, and apoptosis in lens	Anterior segment dysgenesis and congenital absence of a lens
Foxf1	Embryonic lethality at ~E9; abnormalities in somite and posterior development as well as in extraembryonic tissues and lateral plate mesoderm; haploinsufficient mice on certain genetic backgrounds die perinatally with pulmonary defects	Pulmonary and gut development; liver stellate cell activation	Misalignment of pulmonary veins; gastrointestinal abnormalities
Foxf2	Lethality within first 2 h of life due to cleft palate		
Foxg1	Perinatal lethality due to lung defects; reduced cerebral size; eye defects	Neuronal differentiation; cell cycle progression	Rett syndrome and other forms of mental retardation; microcephaly and other brain abnormalities
Foxh1	Perinatal lethality with variable patterning defects: variable right isomerism leading to defects in heart, lungs and stomach and asplenia; cyclopia; microcephaly; absent jaw	Regulation of body patterning, including leftright asymmetry	Congenital heart disease and ventricular septal defects
Foxi1	50% perinatal lethality; inner ear and renal defects; 25% of heterozygotes exhibit perinatal lethality	Renal and inner ear development	Pendred syndrome
Foxi2	Decreased circulating glucose		
Foxi3	Lethality starting at E9.5 through perinatal life; lack whiskers and a mouth; absence of inner, middle and outer ear; increased cranial neural crest apoptosis	Differentiation of branchial arch-derived skeletal structures	
Foxj1	Perinatal lethality; defective ciliogenesis and random left-right asymmetry; growth failure	Regulation of left-right asymmetry; Tlymphocyte quiescence	Allergic rhinitis
Foxj2	Uncharacterized		
Foxj3	Decreased muscle mass and recovery after injury	Mitochondrial biogenesis	

Gene	Null mouse phenotypes	Developmental roles	Role in human disease
Foxk1	Reduced numbers of myogenic progenitor cells, with cell cycle arrest; reduced muscle recovery after injury	Muscle cell and progenitor proliferation	
Foxk2	Uncharacterized		
Foxl1	Low level of postnatal lethality; gastric mucosa hyperplasia with disorganized glands; decreased acid secretion from stomach parietal cells; abnormal crypt structure with an abnormal distribution of Paneth cells; decreased dextrose uptake in intestine	Gastrointestinal development and function	
Foxl2	Inpenetrant postnatal lethality; female sterility	Ovary differentiation and maintenance, including prevention of transdifferentiation into testis; eye development	Adult granulosa cell tumor; blepharophimosis, ptosis and epicanthus inversus syndrome (BPES; a syndrome that is characterized by the inability to fully open the eyes and in some cases early loss of ovarian function); adult granulosa cell tumor
Foxm1	Lethality beginning as early as E15.5, with some alleles displaying lethality perinatally; reduced proliferation in multiple tissues	Regulates cell cycle progression, karyokinesis and cytokinesis	Upregulated in most cancers
Foxn1	Premature mortality resulting from the lack of a thymus; lack of hair; heterozygotes have enlarged thymus	Thymus development	A form of severe combined immunity deficiency (SCID): T-cell immunodeficiency, congenital alopecia and nail dystrophy
Foxn2	Uncharacterized		
Foxn3		Inhibition of proliferation	
Foxn4	Early embryonic to postnatal lethality with reduced growth	Retinal development; neural development	
Foxo1	Embryonic lethality (E10.5-E11.5) due to incomplete vascular development; heterozygotes have elevated glycogen	Vasculogenesis; suppressor of proliferation and apoptosis; blood vessel, yolk sac and diaphragm development; gluconeogenesis; glycogenolysis; adipogenesis; angiogenesis; osteogenesis; T-cell regulation; embryonic stem cell pluripotency	Rhabdomyosarcoma 2 (RMS2)
Foxo3	Ovarian defect causing progressive female sterility; defects in immune system function and hematopoiesis	Follicular growth activation	Fused to MLL gene in acute leukemias
Foxo4	No gross abnormalities	Contributes to myeloid lineage regulation	Fused to MLL gene in acute leukemias

Gene	Null mouse phenotypes	Developmental roles	Role in human disease
Foxo6	Abnormal dendritic spine morphology and defects in memory consolidation	Neuronal development	
Foxp1	Embryonic (E14.5) lethality due to defects in cardiac morphogenesis; edema	Neuronal differentiation; cardiomyocyte proliferation; language acquisition; B-cell differentiation	Mental retardation and language defects with or without autistic features; fusion with PAX6 in acute lymphoblastic leukemia
Foxp2	Postnatal lethality; reduced vocalization; abnormal neural development; delayed eye opening and ear emergence; impaired motor activity; hypoactivity	Neuronal differentiation; language acquisition	Speech-language disorder 1 (SPCH1)
Foxp3	Hemizygous male and homozygous female lethality at ~P21; scaly skin; reduced Tregulatory cells; reddening and swelling of genitals; micro- and cryptoorchidism	T-regulatory cell development	Multiple autoimmune disorders including type 1 diabetes and immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome
Foxp4	Embryonic (by E12.5) lethality; delay of foregut closure leading to development of two hearts; esophagus and trachea fail to develop	Gastrointestinal development; T-cell response	
Foxq1	Silky coat		
Foxr1	Uncharacterized		
Foxr2	Uncharacterized		
Foxs1	Uncharacterized		

Interplays between ciliary transcriptional networks

The experimental evidence collected during the past 30 years has shown how some members of the RFX family direct the transcriptional regulation of both motile and sensory cilia components needed to create the basic ciliary template. Consistently, most of the X-box-containing target genes that were identified in the pioneering works performed in *C. elegans* and *D. melanogaster* encode for core structural features that are essential for cilia formation. This category of genes includes those coding for the components of the IFT machinery, the BBSome or the transition zone. On the other hand, FOXJ1 has been shown to be required for the establishment of the motile ciliated fate. Consequently, genes regulated by FoxJ1 include those encoding for components needed for the generation, assembly, trafficking and docking of the inner and outer dynein arms, the radial spokes and the central pair of microtubules that are characteristic of motile cilia. However, despite this generalization, examples exist in motile cilia in which FOXJ1 directly regulates core cilia genes whereas RFX regulates the expression of ciliary motility components (**see Figure I4.8**).
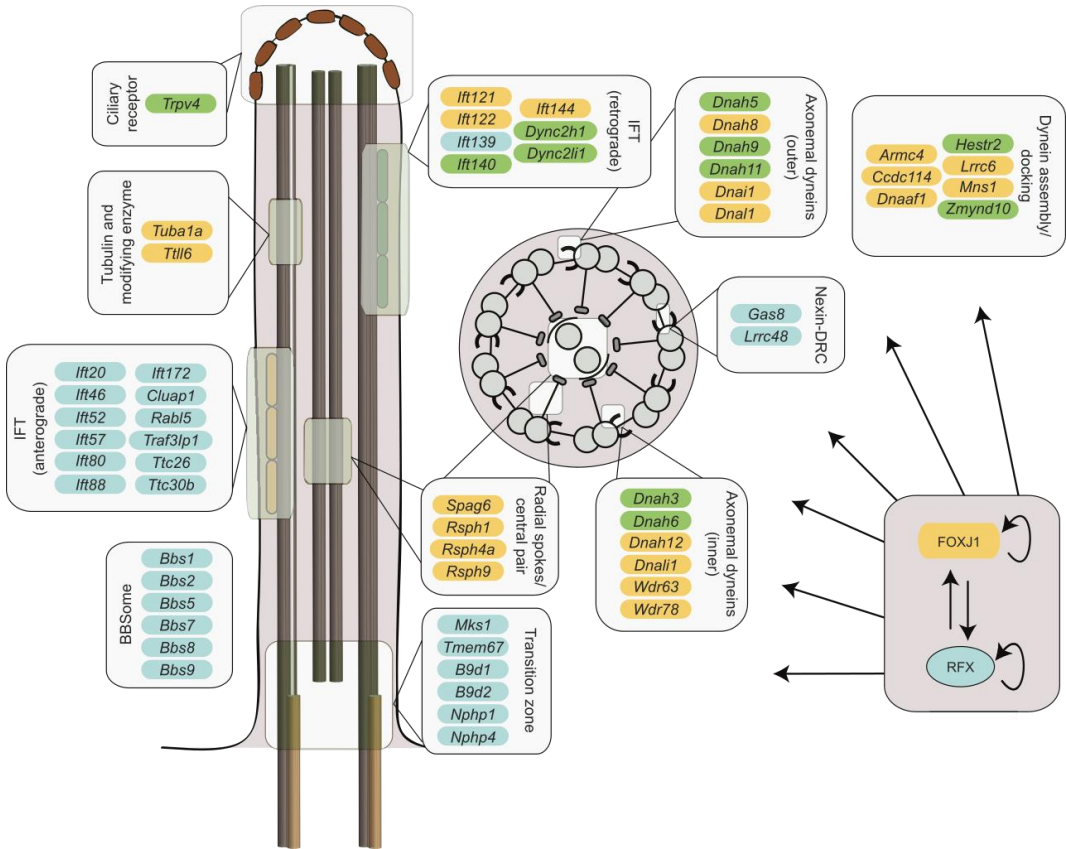
▲ **Figure I4.8. Direct and indirect targets of RFX and FOXJ1 ciliary transcription factors.**
On the left, a longitudinal section of a canonical cilium is represented, showing only two of the nine radial set of microtubules plus its central pair. A cross section of the cilium appears in the centre, showing the whole microtubular structure and the dynein arms that decorate the axoneme of typical motile cilia. Target genes are collected from previously known FOXJ1, FD3F (not shown) and RFX targets found in *Drosophila*, *C. elegans* or vertebrates. Target genes are organized in grey boxes by their associations/functions with respect to ciliary structures. Genes in blue are targets of RFX transcription factors, genes in yellow are targets of FOXJ1 (or FD3F), whereas targets of both TFs are depicted in green. Adapted from (Choksi, Lauter, et al. 2014).

The requirement for transcription factors of the RFX and FKH families in the establishment of the ciliated fate has been further highlighted through the phenotypes observed in several model organisms in which loss-of-function mutants were analysed (see **Table I4.2**). Since both types of transcription factor must co-exist in cells possessing motile cilia, considerations about the interplay, cooperation and hierarchy between both transcriptional networks arise. Interestingly, examples illustrating the two possibilities for positive cross-regulation have been found between members of RFX and FKH families. RFX3 is partially responsible for the transcriptional regulation of FoxJ1 in mouse

ependymal cells (El Zein et al. 2009). Conversely, FOXJ1 can induce the expression of Rfx2 in zebrafish (Yu et al. 2008) and Rfx3 in the human airway epithelium (Didon et al. 2013). This later study also exemplifies the case of cooperativity in the establishment of the motile cilia fate between members of the RFX and FKH families. Didon *et al.* found that the combined transfection of both FOXJ1 and RFX3 enhanced the promoter activity and messenger expression of cilia-related genes beyond those due to FOXJ1 alone. Another example was found in the chordotonal neurons of *Drosophila*, in which a long mechanosensory cilium also possesses a motile part. In this system, the promoter regions of specific genes needed for motile ciliogenesis were shown to harbour both *Rfx* (RFX) and *fd3F* (FKH) binding sites in close proximity. Expression of their target genes was severely reduced by either the individual mutation of their binding sites or in the genetic background of each individual mutant. Further supporting a coregulatory role for these factors in the specification of the motile ciliated fate of *Drosophila* chordotonal neurons, no evidence of a regulatory hierarchy was found, since transcription of each TF was seen independent from each other (Newton et al. 2012). Yet another recent example for co-binding, coordination and cooperativity was found in the multiciliated cells of *Xenopus* (Quigley and Kintner 2017). In that context, Foxj1 and Rfx2 were seen to play complementary roles. Although Foxj1 was responsible for the activation of the target genes, it was found to bind at distal sites in the genome. Rfx2, which in turn was unable to activate the target genes by itself, recruited Foxj1 to proximal regions via a loop created through its own binding to the promoter regions of the target genes. This mechanism is reminiscent of the crosslinking role already proposed for RFX homodimers by (Patrick Emery et al. 1996). Finally, models of redundancy have also been found in the interplay between members of the RFX and FKH families. As an example, in the floor plate of mouse embryos, *Rfx3* has been suggested to act in a parallel manner with *Foxj1* to induce cilia lengthening of long 9+0 cilia that are presumed to be motile. Supporting this idea, floor plate identity and ciliogenesis are unaffected in mouse embryos lacking Foxj1 but individual overexpression of either TF induces lengthening in neuronal structures of transfected chick embryos (Cruz et al. 2010).

In summary, solid evidence exists for a cooperative action between RFX and FOXJ1 acting as terminal selectors for motile cilia. In contrast, for sensory cilia, no other TFs in addition to RFX TFs are known to directly activate the ciliome of sensory cells.

➤ **Table I4.2. Transcription factors required for cilia specification in selected organisms.**
Members of the RFX and FKH families of transcription factors (TFs) have been found to control the specification of the ciliated fate in several organism. This table gathers those previously reported for both motile and primary cilia. Adapted from (Choksi, Lauter, et al. 2014).

	Cilia type	Organism	TFs	References
MOTILE CILIA	**Airway motile multicilia**	*M. musculus*	FOXJ1	(Chen et al. 1998; Brody et al. 2000)
	Epidermal motile multicilia	*X. laevis*	FOXJ1 RFX2	(Stubbs et al. 2008) (Chung et al. 2012)
	Sperm flagellum	*M. musculus*	FOXJ1	(Chen et al. 1998)
	Oviduct motile multicilia	*M. musculus*	FOXJ1	(Chen et al. 1998; Brody et al. 2000)
	Brain ependymal multiple motile cilia	*M. musculus*	RFX3 FOXJ1	(El Zein et al. 2009) (Chen et al. 1998; Brody et al. 2000)
	Brain ependymal monocilia/multicilia	*X. laevis*	FOXJ1	(Hagenlocher et al. 2013)
	Spinal canal ependymal motile cilia	*D. rerio*	FOXJ1A	(Yu et al. 2008)
	Nodal motile monocilia	*M. musculus*	RFX3 FOXJ1	(Bonnafe et al. 2004) (Alten et al. 2012)
	Kupffer's vesicle motile monocilia	*D. rerio*	RFX2 FOXJ1A	(Bisgrove et al. 2012) (Stubbs et al. 2008; Yu et al. 2008)
	Gastrocoel roof-plate motile monocilia	*X. laevis*	RFX2 FOXJ1	(Chung et al. 2012) (Stubbs et al. 2008)
	Pronephric motile multicilia an monocilia	*D. rerio*	RFX2 FOXJ1A FOXJ1B	(Liu et al. 2007) (Yu et al. 2008) (Hellman et al. 2010)
	Otic vesicle kinocilia	*D. rerio*	FOXJ1B	(Yu et al. 2011)
	Chordotonal organ sensory motile cilia	*D. melanogaster*	FD3F	(Cachero et al. 2011; Newton et al. 2012)
	Olfactory motile cilia	*D. rerio*	FOXJ1A	(Hellman et al. 2010)
PRIMARY CILIA	**Immotile signalling cilia**	*M. musculus* *D. rerio* *X. laevis*	RFX4 RFX2 RFX2	(Ashique et al. 2009) (Yu et al. 2008) (Chung et al. 2012)
	Sensory neurons	*D. melanogaster* *C. elegans*	RFX DAF-19	(Dubruille et al. 2002) (Swoboda, Adler, and Thomas 2000b)

The making of specialised cilia

Despite the basic common genes and structure that define sensory and motile cilia, the cilium in each cell type contains specific components that allow for its specialized functions. Although the regulatory logic behind the generation of the basic type of cilium has been shown to clearly require the activation of the RFX/FOXJ1 module, recent studies

demonstrate that the achievement of ciliary diversity entails other molecular mechanisms.

Several examples (reviewed in Thomas et al. 2010 and Choksi et al. 2014) gathered from zebrafish, mouse and *Xenopus* have linked differentiation of specific types of motile cilia with the main intercellular signalling pathways: hedgehog (HH), fibroblast growth factor (FGF), WNT and NOTCH signalling. Additionally, cell type-specific transcription factors such as Atonal (ATO), MCIDAS, MYB or the notochord homeobox (NOTO) have also been found to regulate the acquisition of motile cilia diversity , however, the direct action of these TFs on transcriptional regulation of cilia structural components has not been demonstrated(see **Figure I4.9**).
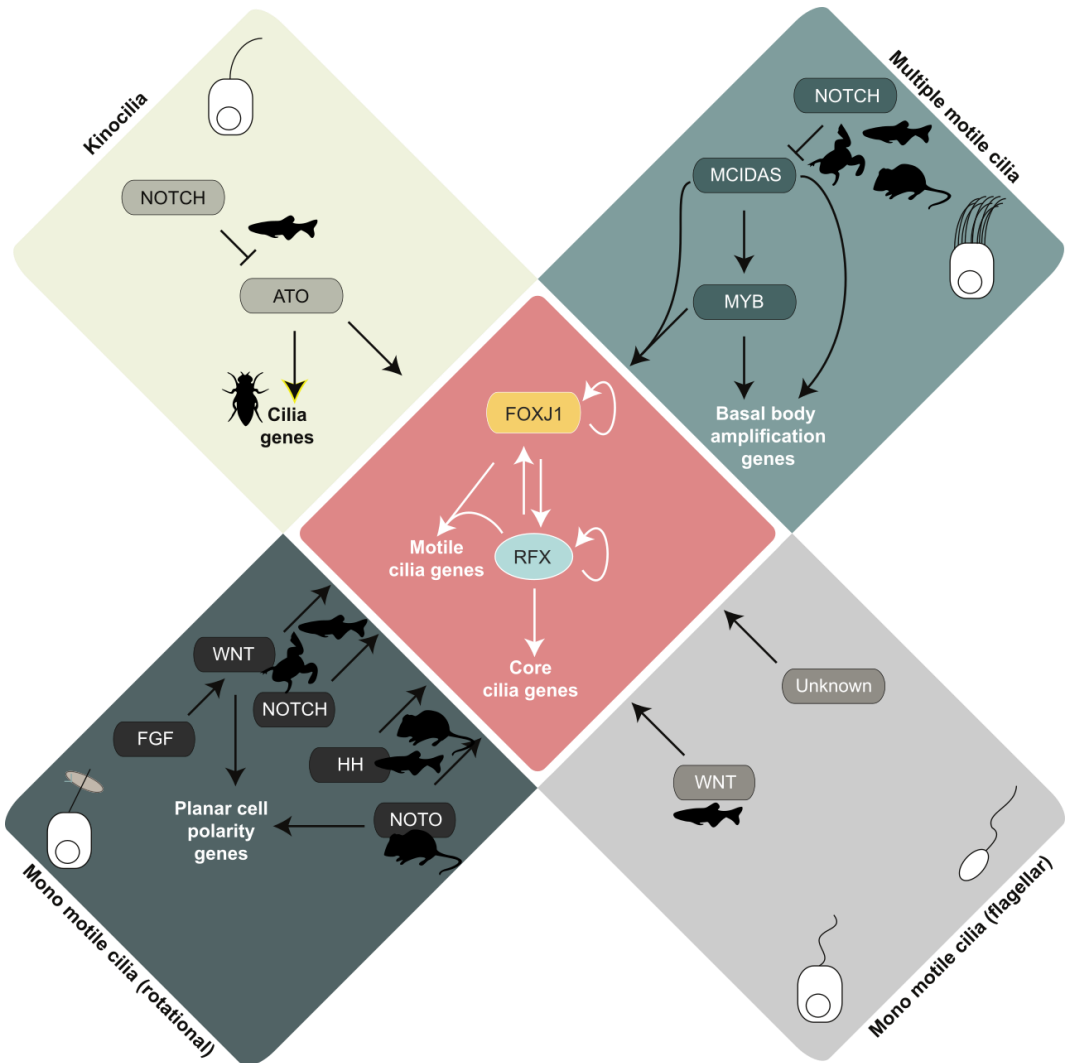
◄ **Figure I4.9. Transcriptional regulatory logic of motile ciliogenesis requires the RFX/FOXJ1 module.**

Transcriptional regulation of motile cilia components requires a module composed by RFX, generally controlling the expression of structural core genes, and FOXJ1, which generally activates the expression of genes involved in ciliary motility. In turn, and depending on the specific type of motile cilium, several signalling pathways and TFs are known to activate both RFX, FOXJ1 or even cilia genes. **Rotational mono motile cilia.** Those in the organs of laterality from zebrafish and *Xenopus* are induced by NOTCH and WNT (the latter acting under the control of FGF). In the mouse node, the notochord homeobox (NOTO) transcriptional regulator controls the activation of FOXJ1 and an RFX factor. In the floor plate of *D. rerio* and *M. musculus*, hedgehog (HH) signalling is involved. **Kinocilia**. The proneural transcription factor Atonal ATO, which is repressed by NOTCH signalling in *D. rerio*, is responsible for the generation of the kinocilia in the developing ear through the activation of FOXJ1. Moreover, in *Drosophila* ATO can also independently activate cilia-related genes. **Multiple motile cilia**. Multiciliated cells of Xenopus, mouse and zebrafish employ a transcriptional cascade involving MCIDAS and MYB to activate both the RFX/FOXJ1 module and the genes needed for basal body synthesis and docking. **Flagellar mono motile cilia.** Pathways controlling the establishment of these cilia are largely unknown. Nevertheless, in the zebrafish pronephric duct, FOXJ1 expression is activated by WNT signalling. Adapted from (Choksi, Lauter, et al. 2014).

Regulatory programs involved in the generation of sensory cilia subtype diversity have also been explored. In mouse, mutants for the HNF1β (hepatocyte nuclear factor 1β) TF show severe polycystic kidneys, a hallmark of many ciliopathies in humans. Among direct targets of HNF1β, many ciliary genes, including Pkhd1, Pkd2 or Ift88 have been found. (Gresh et al. 2004; Hiesberger et al. 2005). Posterior indirect data led to the suggestion that HNF1β may also participate in the genetic cascade controlling ciliogenesis in the pancreatic ducts. *Hnf6* (hepatocyte nuclear factor 6) deficient mice show kidney and pancreatic duct cysts associated with lack of cilia, lack of *Hnf1β* expression in the cells lining the cysts and decreased *Pkhd1* expression. Since the *Hnf1β* gene promoter harbours HNF6-binding sites, it was proposed that *Pkhd1* down-regulation was due to the lack of HNF1β in the pancreas (Pierreux et al. 2006). In *C. elegans*, the different isoforms of

daf-19, the master regulator of ciliogenesis, are expressed in a cell-type specific manner (see **Figure I4.10**). All five isoforms share the DNA binding domain while differ in its Nt. The long isoforms *DAF-19A/B* are mainly expressed in non-ciliated neurons of the worm and regulate basic synaptic maintenance as well as neuronal outputs such as *dauer* formation, foraging or locomotion (Senti and Swoboda 2008; De Stasio et al. 2018). DAF-19D, (following Wormbase nomenclature but originally termed DAF-19C), which is expressed in virtually all ciliated sensory neurons of the animal, primarily activates cilia-related genes (Senti and Swoboda 2008). The shortest of the known isoforms, DAF-19C (following Wormbase nomenclature but originally termed DAF-19M), is activated by the transcription factor EGL-46 to control the expression of the so-called PDK (polycystic kidney disease) genes in the male-specific ciliated neuron HOB (Yu et al. 2003; Wang et al.

2010). These genes include *lov-1* and *pkd-2*, the kinesin-like protein gene *klp-6*, and five "coexpressed with polycystin" (*cwp*) genes, which are expressed and required for correct cilia function in HOB neurons but not in hermaphrodite ciliated neurons. However, *egl-46* plays no apparent role in the rest of the male-specific ciliated neurons of the worm. Much interestingly, the PKD battery genes do not contain X-boxes in their promoter sequences, illustrating how *daf-19* is able to regulate both core ciliary genes (through the direct binding of DAF-19D to the X-box motifs) and ciliary specialization (here exemplified by the action of DAF-19C) maybe through indirect actions (Wang et al. 2010). Despite this isoform-specific control, how *daf-19* isoforms achieve the different and specific activation of its target genes is largely unknown and the contribution of transcriptional co-factors for *daf-19* (that are yet to be discovered) has been proposed. Favouring this view is the experimental evidence showing that the promoter regions of genes broadly expressed within the ciliated system of *C. elegans*, *C. briggsae* and *C. remanei* harbour, in close proximity to the X-box motif (less than 60 base pairs away), a novel C/T rich motif (therein termed as the C-box) that is not found in cilia genes whose expression is restricted to subsets of ciliated neurons, however, the molecule recognize this motif, if any, has not been identified yet (Burghoorn et al. 2012). Related with the additional contribution of novel transcription factors but also reminiscent of the examples in which RFX factors activate FKH members is the case of *fkh-2*. In the subpopulation of chemosensory AWB neurons in *C. elegans*, *daf-19* activates *fkh-2*, which in turns

activates the kinesin II subunit *kap-1*, which then contributes to the acquisition of the characteristic winged shaped cilium that AWB neurons possess (Mukhopadhyay et al. 2007). Thus *fkh-2* would be responsible for cilia subtype gene expression in this particular neuron; interestingly, broad *kap-1* expression in non AWB ciliated neurons is unaffected in *fkh-2* mutants, suggesting a subtype specific effect for this TF.

Modulation of gene expression levels illustrate another mechanism employed in the acquisition of cilia diversity. In *Drosophila*, low levels of RFX correlate with the expression of core cilia-related genes in all ciliated types of sensory neurons. However, high levels of RFX are responsible for the expression of the genes required for the generation of the specialized motile cilia that chordotonal neurons possess (Cachero et al. 2011; Newton et al. 2012).

Morphological diversity induced by differential gene expression has also been found for immotile cilia. In *C. elegans*, the increasing membrane elaborations of the winged-shaped cilia that AWB, AWC and AWA neurons possess are instructed by the increased levels of the immunoglobulin domain protein OIG-8. Much interestingly, the expression of *oig-8* is controlled by neuron-type-specific terminal selectors that act independently of the main ciliary transcription module operated by *daf-19* (Howell and Hobert 2017). This transcriptional path, in which cell-type-specific factors modulate the acquisition of ciliary diversity by running in parallel to the main ciliogenic programme, has been seen to act also in higher vertebrates (see Gresh et al. 2004 and Kiselak et al. 2010).
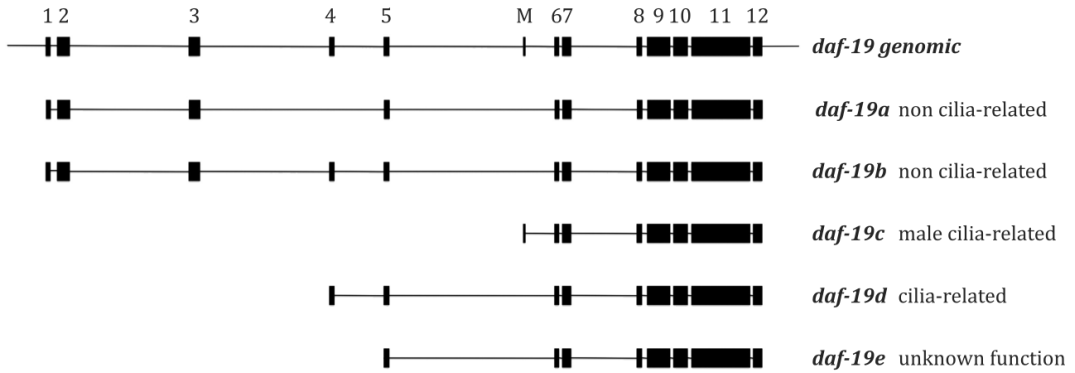
⌃ Figure I4.10 Different isoforms of *daf-19* control sensory cilia diversity in *C. elegans*.
The *daf-19* locus encodes 5 different isoforms through various exon-skipping or alternative promoter events. Exons are represented here as black boxes and exon number within the locus is indicated. Isoform names vary in the bibliography; here, the WormBase nomenclature is shown. Due to its cilia-related functions, *daf-19d* is cited in several publications as *daf-19c* (for function in "cilia"). Analogously, *daf-19c*, which control male-specific cilia features, is usually termed *daf-19m* (for function in "male mating"). Adapted from (Wells, Rowneki, and Killian 2015).

The molecular mechanisms briefly explained in this section are aimed to illustrate the complicated scenario of interconnected, redundant and parallel pathways that ciliated cells must confront to achieve the coordinated expression of the proper set of genes that are needed to produce a functional cilium. Several gaps and unanswered question remain open about the transcriptional regulation of the ciliated fate. The study of gene regulatory networks underlying ciliogenesis will not only increase our understanding on how such an evolutionarily conserved organelle is generated but will also help us better understand the molecular mechanisms underlying the wide spectrum of diseases the ciliopathies integrate.

Motivation & Aims

Cilia are complex evolutionarily conserved subcellular structures that project from eukaryotic cell surfaces to perform a variety of biological roles, ranging from motion to sensation.

Ciliopathies comprise a group of multisystemic human disorders associated to mutations that lead to defects in the function or structure of cilia. With an estimated 1 in 1,000 people affected by these diseases, very little is known about the molecular bases of most ciliopathies.

Although it is known that in different animal groups RFX transcription factors are essential for the expression of most genes encoding for the components of the primary sensory cilium, no other transcription factors have been described to date to participate in this process. In *C. elegans*, DAF-19/RFX acts as terminal selector controlling the expression of ciliary components in sensory neurons, however, several pieces of evidence suggest that *daf-19* cannot be acting alone in the specification and maintenance of the ciliated fate.

The main objective of this thesis project is to identify *daf-19* co-operators and new regulators of sensory ciliogenesis in *C. elegans*.

To fulfil this general objective, the following specific aims have been established:

Aim 1

Identification and selection of transcription factors candidates to regulate the expression of ciliary features.

Aim 2
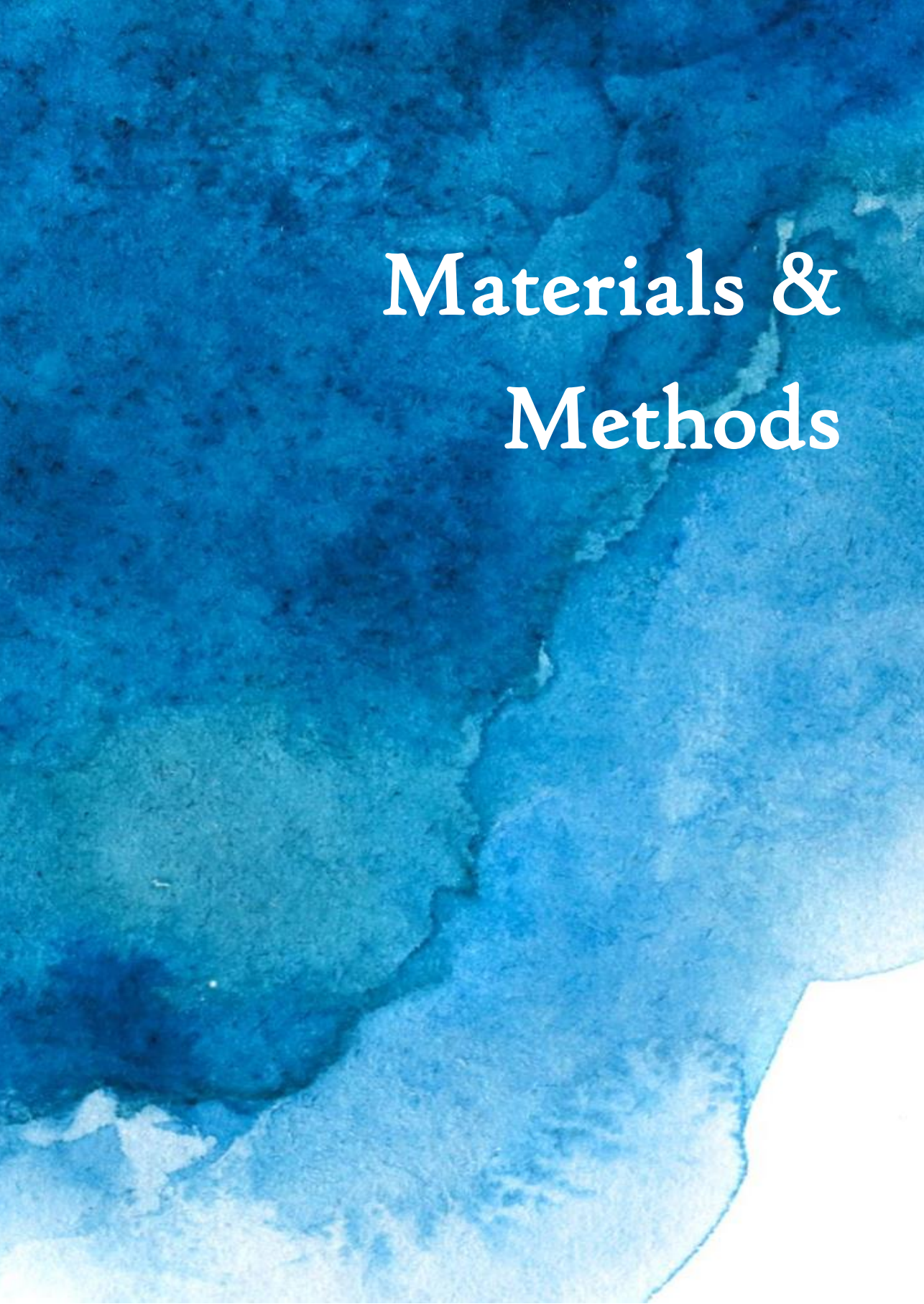
Characterization of candidates' role on ciliome gene expression through ChIP-seq, RNA-seq and *in vivo* reporter analysis.

Aim 3

Study of the sensory-mediated behaviours in mutant animals for selected candidates.

Aim 4

Analysis of the crosstalk and synergy between selected candidates and *daf-19*.

Materials & Methods

Experimental procedures

This section describes how the experiments were performed and the reagents used. Procedures are arranged considering work's development.

C. elegans strains and maintenance.

C. elegans maintenance and culture were performed following standard methods (S Brenner 1974). Briefly, worms were grown in Petri dishes (60 x 15 mm, no vents, Corning #BP50-01) filled with Nematode Growth Medium (NGM) agar (see Materials section). Plates were seeded with the uracil-requiring mutant of *Escherichia coli* OP50, obtained from the Caenorhabditis Genetic Center (CGC).

C. elegans N2 strain, also obtained from the CGC, was used as control. Routine handling of the worms was carried out under the dissection scope (Zeiss, Stemi 305 LAB). Animals were maintained at 15 °C for medium-term storage, at 20 °C for regular procedures and 25 °C for fluorescent reporter scoring. Strains were conserved by freezing as previously reported (S Brenner 1974). A list of all the strains used and/or generated for this work is shown in **Annex.1**.

Worm lysis for genomic DNA preparations.

Genomic DNA from single worms was obtained by transferring through picking individual animals into a 0.2 mL PCR tubes containing 10 μL of 2% Protease K lysis solution (see Materials section).

For genotyping of several animals, worm-containing plates were rinsed with 400 mL of M9 1X buffer (see Materials section) and 10 μL of this worm-containing mix was transferred into 0.2 mL PCR tubes. Volume was then adjusted to 50 μL with a final concentration of 2% Protease K.

In both cases, tubes were briefly vortexed, spun and placed at 65 °C for 1 hour followed by 30 minutes at 95 °C to inactivate Proteinase K. Genomic DNA was then ready to use as template for PCR procedures or to store by freezing.

Generation of fluorescent reporters: fusion PCR.

Fluorescent reporters for ciliary features were generated through fusion PCR (Hobert 2002). To facilitate the identification and scoring of reporter-expressing cells, the fluorescent protein was localized to the nucleus by tagging with a modified sequence of the classical SV40 large T antigen nuclear localizing signal (NLS) (Kalderon *et al.* 1984) (see **Figure M1.1**). NLS::tagRFP::*unc-54* 3' UTR amplicon was obtained from *otIs395(ift-20::NLS::tagRFP)III* genomic DNA using primers C and D and standardly cloned into pPD95·75 (Addgene #1494, see **Figure M1.2**) between the PstI and SpeI restriction sites, generating an intermediate NLS::tagRFP::*unc-54* 3' UTR plasmid. This plasmid was then used to replace tagRFP::*unc-54* 3' UTR by GFP::3' UTR *unc-54* from pPD95·75 between AgeI and SpeI restriction sites, hence creating a 2xNLS version of the pPD95·75 plasmid (pNF400).
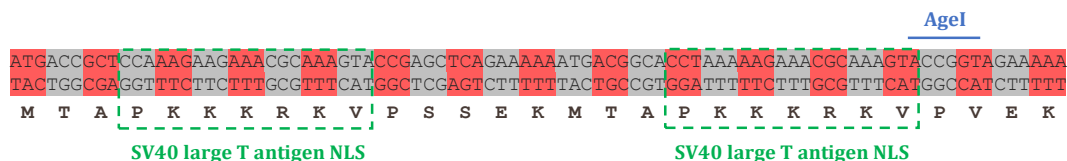
Figure M1.1. Modified SV40 large T antigen NLS sequence.
Alternating codons within the sequence are depicted with pink and grey colours. Green dotted box highlights the two monopartite nuclear localization signals recognized by importin-α mediated transport. The target sequence for the AgeI endonuclease used for cloning purposes is indicate in blue.
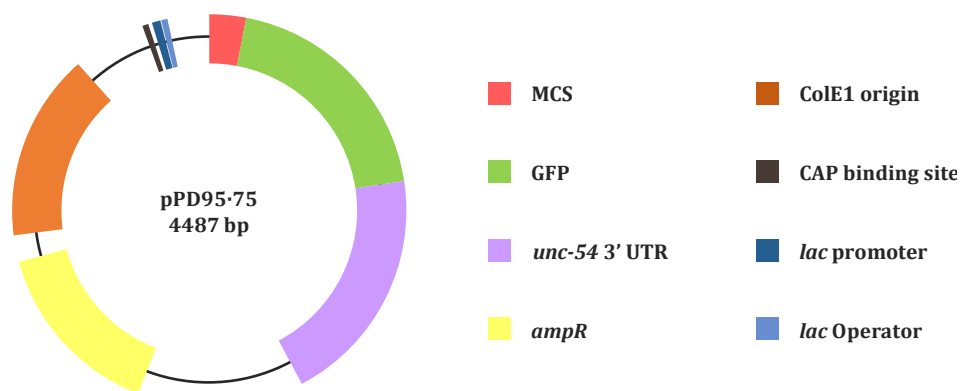
Figure M1.2. Schematic of the pPD95·75 plasmid.
Coloured arcs mark the location of pPD95·75 main features. GFP coding sequence was amplified from this plasmid to generate fluorescent reporters for cilia-related features.

For *cis*-regulatory analysis, regulatory sequences were amplified with custom oligonucleotides from N2 genomic DNA preparations (see PCR 1a in **Table M1.1** and **Figure M1.3**). An independent PCR was used to amplify the 2xNLS::GFP:: *unc-54* 3'UTR fragment from pNF400 (see PCR 1b in **Table M1.1** and **Figure M1.3**). A final fusion PCR was set up with a mix of these two reaction products (see PCR 2 in **Table M1.1** and **Figure M1.3**).

Fusion events were visualized through agarose gel electrophoresis by migration comparison with NLS::GFP:*unc-54* 3' UTR PCR. Successfully fused PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN, #28106) and resuspended in nuclease-free water (Sigma, #W4502). PCR program used for this procedure is shown in **Table M1.2**. Primers used to generate fluorescent reporters are listed in **Table M1.3**.

▼ **Table M1.1. Two-step fusion PCR mixes.**

Component	Volume (μL)	Final concentration
PCR1		
Expand Long Template Buffer 2	2.5	1X (2.75 mM MgCl$_2$)
PCR Nucleotide Mix	0.35	0.2 μM
PCR 1a FW primer (A or A*) or PCR 1b FW primer (oRR265)	0.3	0.5 μM
PCR 1a RV primer (B) or PCR 1b RV primer (D)	0.3	0.5 μM
Expand Long Template DNA Polymerase (5U/ μL)	0.2	1 U/25 μL
PCR 1a Template DNA (N2 genomic) or PCR 1b Template DNA (pNF400)	1	< 250 ng/25 μL
Nuclease-Free Water	20.35	(–)
MIX	25	
PCR2		
Expand Long Template Buffer 2	2.5	1X (2.75 mM MgCl$_2$)
PCR Nucleotide Mix	0.35	0.2 μM
FW primer (A*)	0.3	0.5 μM
RV primer (D*)	0.3	0.5 μM
Expand Long Template DNA Polymerase (5U/ μL)	0.2	1 U/25 μL
Template DNA (PCR 1a + PCR 1b)	1 + 1	< 500 ng/25 μL
Nuclease-Free Water	19.35	(–)
MIX	25	

Generation of reporter transgenic lines.

Simple-array transgenic strains were generated by intragonadal microinjection into the N2 strain. The injection mix was composed by 50 ng/μL of a given purified fusion PCR plus 100 ng/μL of pFR4 plasmid, *rol-6(su1006), as* co-marker (Mello *et al.* 1991). *su1006* is a dominant negative mutation of the structural constituent of collagen and cuticulin-based cuticle gene *rol-6*. It is expressed in the hypodermis and when inherited it confers the worms a helically twisted body (a phenotype colloquially termed as "roller") that allows for the easy identification of transgenic animals under the dissection scope.

Prior to microinjection, 2% agarose pads were prepared (Mello and Fire 1995). Prior to microinjections, DNA mixes were centrifuged at full speed for about 5 minutes to prevent needle clogging. 4 μL of supernatant was

transferred to a new tube and 1 final μL of the mix was aspirated with a microloader (Eppendorf, #930001007) and loaded into a 0.5 μm diameter Femtotip II needle (Eppendorf, #930000043). This needle was then adjusted to a micromanipulator (Narishigue, MN-151) and placed in the correct angle for gonadal microinjection (15° to 45°). Adult hermaphrodites with clearly visible gonads were selected, placed in groups of 5 over the agarose pad for immobilization and covered with a drop of halocarbon oil 700 (Sigma, #H-8898) to prevent from dehydration.

Microinjection process was assisted by a Femtojet (Eppendorf, #5274) coupled to an inverted microscope (Zeiss, Axio Vert.A1). Worms were microinjected at the syncytial arms of both gonads where DNA mixes were released. A drop of M9 1X buffer was then added over the injected animals that were allowed to recover for several minutes before being placed in individual NGM agar plates and incubated at 20 Celsius degrees. Typically, between 10 to 20 animals were injected per mix.

Three days after this process, the presence of the *rol-6(su1006)* phenotype was analysed on the F1 progeny. About 100 to 200 of the existing F1 "rollers" were individually placed in fresh NGM agar plates. Three days later transgenic lines are selected by mosaic stable transmission of the marker phenotype to the F2 progeny. 3 to 4 independent lines were selected and maintained.

Final fusion PCR product ready to inject into *C. elegans*
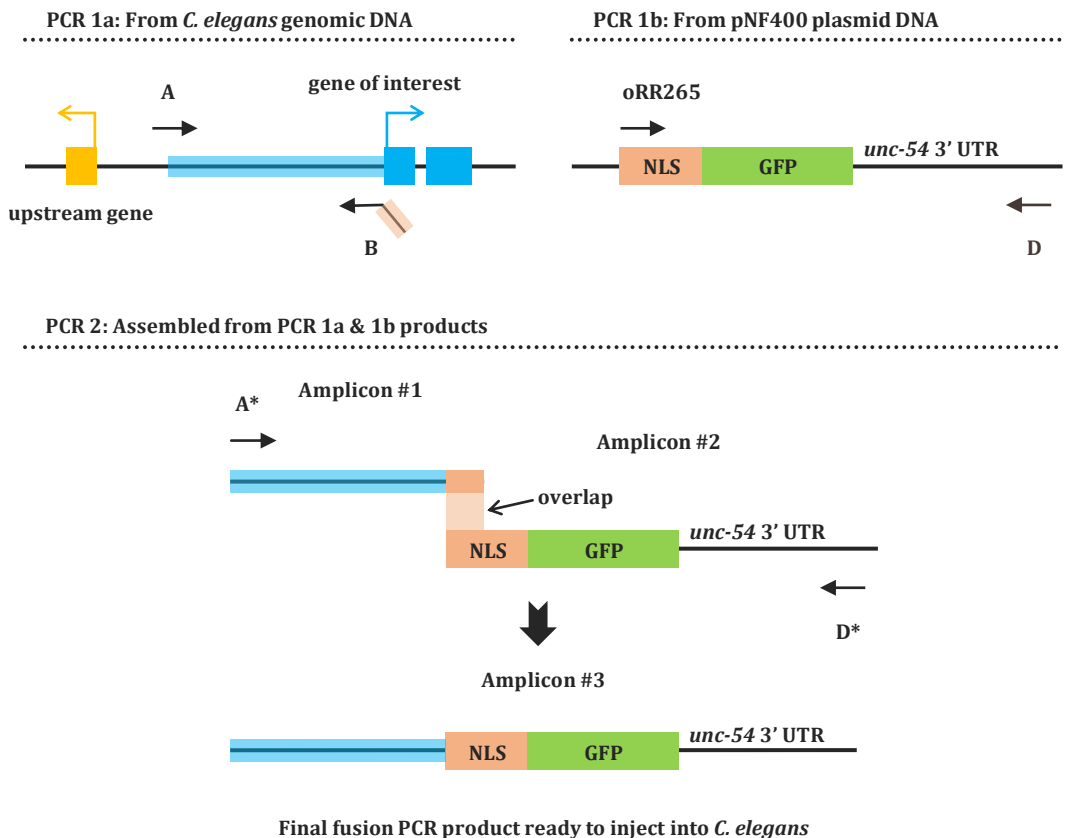
◀ **Figure M1.3. Fusion PCR protocol.**

Schematic of the two-steps protocol performed to generate GFP-tagged reporters from cilia related genes. In the first step, custom oligos are used to amplify independently the regulatory region of the gene of interest (PCR1a) and the coding sequence for a NLS-tagged GFP (PCR1b). In the second step (PCR2), the NLS-complementary tale introduced with primer B allows for the overlapping of both fragments, thus creating a fusion product that is further amplified using the corresponding nested primers (A* and D*)

▼ **Table M1.2. Two-step fusion PCR program.**

Step	Temperature (°C)	Time (s)	Cycles
Initial denaturation	92	120	1
Denaturation	92	10	
Annealing	58	10	25
Extension	68	60/kb	
Final extension	68	7	1
Storage	10	∞	1

▼ **Table M1.3. Primers used to create NLS fluorescent reporters.**

The symbol (*) represents nested primers; the original fusion PRC protocol used this strategy to ensure proper fragment amplification. Note that some of the promoters (*mks-1*, *osm-5*, *tmem-107* and *xbx-1*$_{(short)}$) were successfully amplified using only the edge nested primer in both PCR 1a and PCR 2. Lengths of the regulatory regions used to direct the expression of the NLS reporters are indicated in parenthesis. All of the promoters were generated from regions immediately upstream of the starting codon ATG. Following this logic, regulatory regions for both *peli-1* and *xbx-1* were generated in two different lengths, short and long.

Target	Type	Name	Sequence
pPD95·75 and its derivatives	FW	C	AGCTTGCATGCCTGCAGGTCG
	RV	D	AAGGGCCCGTACGGCCGACTA
	RV*	D*	GGAAACAGTTATGTTTGGTATA
SV40 NLS	FW	oRR265	ATGACCGCTCCAAAGAAGAAAC
che-11 (-208 bp from ATG)	FW	oRR121	CCGAATTTATCTAGATGTTTCCCCGA
	FW*	oRR122	TAGATGTTTCCCCGAAATT
	RV	oRR271	GTTTCTTCTTTGGAGCGGTCATTTATGTTAATAATTGAAGAAATCGGGAAGTAAGG
che-13 (-525 bp from ATG)	FW	oRR108	AAGTACTGCATAATAGGAACATACGG
	FW*	oRR109	AGGAACATACGGCGTTAT
	RV	oRR272	GTTTCTTCTTTGGAGCGGTCATTGTAATAATCAGATAATAATAATGAAAGTGAGCT

Target	Type	Name	Sequence
mks-1 **(-144 bp from ATG)**	FW*	oRR383	TCTTTTGTATTTTCTTCGAAT
	RV	oRR384	GTTTCTTCTTTGGAGCGGTCATATTGATGTTCTACTAGGGGT
osm-1 **(-1770 bp from ATG)**	FW	oRR212	TGCCCAGTTGTGACCAGTATTA
	FW*	oRR213	GTGACCAGTATTACAATCCTTCCCA
	RV	oRR280	GTTTCTTCTTTGGAGCGGTCATCCTCATGAACTGAATAACTGCAAGAAA
osm-5 **(-142 bp from ATG)**	FW*	oRR379	TGTTACACTCTCTCTCTTG
	RV	oRR380	GTTTCTTCTTTGGAGCGGTCATTAAGAAAAGTGTTCTCAGAAG
peli-1(short) **(-225 bp from ATG)**	FW	oRR221	GCTCATAGATCTCTAAGAATTTCG
	FW*	oRR222	TCTAAGAATTTCGACACTTTTC
peli-1(long) **(-1128 bp from ATG)**	FW	oRR224	TGTGATTCTAATCTTCATCAAGAT
	FW*	oRR225	AGATCTTTACACCTGAGTTC
peli-1	RV	oRR333	GTTTCTTCTTTGGAGCGGTCATTAATCTGAAACTCAACATACAAAT
tmem-107 **(-143 bp from ATG)**	FW*	oRR381	AAAATAAAATTTTTCATGTAAACC
	RV	oRR382	GTTTCTTCTTTGGAGCGGTCATTTCTTAAAGAAGCTTGCG
tub-1 **(-424 bp from ATG)**	FW	oRR105	ATGGATATTTGGATGTCATGCTAAAAATTATACA
	FW*	oRR106	GCTAAAAATTATACATTCATTTATGTTG
	RV	oRR282	GTTTCTTCTTTGGAGCGGTCATGATTACCTGGAAACTTTGAATAGTTTTTGAAC
xbx-1(short) **(-105 bp from ATG)**	FW*	oRR337	CTCTCCTCGTTTCTTGTTGAACG
xbx-1(long) **(-370 bp from ATG)**	FW	oRR94	CATGAAGAGAACATTTGGTTCTAGGAATATTT
	FW*	oRR95	GAATATTTATTTGAATTAATTCTTGTTTTT
xbx-1	RV	oRR283	GTTTCTTCTTTGGAGCGGTCATACTAATAAATTGAGCAATATTGGAGAAAAGGA
ocr-1 **(-636 bp from ATG)**	FW	oRR208	CTTCCTACTTAGCATTGTATGCAGTAAAC
	FW*	oRR209	TATGCAGTAAACTATTTTTGACACCTAACTG
	RV	oRR274	GTTTCTTCTTTGGAGCGGTCATTCTGAATCAAGAATTTACATTAAAAA
ocr-2 **(-859 bp from ATG)**	FW	oRR210	GATCTATCAAGTTTTCAGAAGTTTCTTGTACA
	FW*	oRR211	TTTTCAGAAGTTTCTTGTACAAACTCTTTTCA
	RV	oRR275	GTTTCTTCTTTGGAGCGGTCATCTTAATGATGTGATGTACTCTACTGATAAGATGA

Target	Type	Name	Sequence
ocr-4 (-360 bp from ATG)	FW	oRR226	CGGAGATGGCATATGATTT
	FW*	oRR227	AGTTTTAATAGCAATGCTCG
	RV	oRR279	GTTTCTTCTTTGGAGCGGTCATTAATACAAGTTAGATTCAGAGAATATTT
osm-9 (-680 bp from ATG)	FW	oRR111	CCTTCTTCACCCAATATTCTGTGAGT
	FW*	oRR112	CCAATATTCTGTGAGTTGAGT
	RV	oRR281	GTTTCTTCTTTGGAGCGGTCATGTTTGGTTTCTGAAAAAATTGGAAAATTAAAGAG

Strain crossing.

Strain crossing procedures were performed in order to achieve appropriate combinations of gene reporters or mutant backgrounds and/or to assess the influence of different genetic backgrounds over the expression of several fluorescent reporters.

For this aim, 4 L4 hermaphrodites and at least 12 or more males (if possible) were put together in fresh NGM agar plates. To maximize encounter probability, mating plates were seeded with small bacterial lawns placed on the centre of the plate. Plates were incubated at 20 Celsius degrees for 3 days approximately.

Whenever possible, cross progeny was selected under the fluorescent scope and about 6 L4 hermaphrodites from the F1 were singled into new NGM agar plates and incubated at 20 Celsius degrees for 3 days approximately. F2 progeny was then appropriately segregated in order to achieve the desired genetic backgrounds. For crosses with *daf-19(m86)* allele, F1 and F2 numbers were usually doubled as we found desired genotypes were harder to obtained probably due to lethality or worm loss in the edges of the plate.

The specific source of each mutant and the resulting mutant lines obtained by crossing are listed in **Annex 1**. A description of the allelic nature of the mutants used in this work can be found in **Table M1.4**.

▼ **Table M1.4. Description of the allelic nature of the mutant strains used in this work.**

Allele	Mutation	Reference
daf-19(m86)II	G to A substitution generates an early STOP codon, originally reported to be in exon 7, actually located in exon 8 due to a new isoform being discovered. This allele is considered to be null and animals harbouring this mutation are dauer-constitutive (Daf-c), osmotic avoidant (Osm) and dye-filling defective (Dyf).	(Swoboda et al. 2000a)
daf-19(of5)II	A 12 bp deletion remove the original ATG in exon 1, creating a new one and hence leading to a frameshift mutation. Mutated exon 1 generates a different protein fragment and a STOP codon hits exon 2. This mutation is meant to abolish expression of long *daf-19* isoforms a and b, leaving short cilia-related isoform c intact. Consistent with this hypothesis, *daf-19(of5)* animals are not Daf-c, not Osm, not Dyf and, also, not dauer-defective (Daf-d).	Dr. Swoboda, personal communication

Allele	Mutation	Reference
daf-12(sa204)X	G to A substitution generates an early STOP codon in exon 5 of the longest isoform (transcript F11A1.3a.1). This allele is considered to be null and abolishes the Daf-c phenotype of null *daf-19* mutants.	Dr. Swoboda, personal communication and (Senti *et al.* 2009)
fkh-8(tm292)II	A 531 bp deletion affecting exons 5, 6, 7 and 8. FKH DNA binding domain remains unaffected and transcript reads can be seen in RNA-seq experiments. This allele is meant to be hypomorph.	(Barstead *et al.* 2012) and this work
fkh-8(vlc43)II	An 1864 bp deletion removing the whole *fkh-8* locus. Mutation was confirmed through Sanger sequencing and mapped 79 bp upstream *fkh-8* ATG and 122 bp downstream its TAA STOP codon. This allele has to be considered null.	This work

Genotyping.

Newly generated mutant strains were genotyped to assess the nature of their genetic backgrounds. Accounting for the specific nature of a given mutation, appropriate primers were designed and ordinary PCRs performed using 1 to 3 μL of F2 lysates as DNA template prepared as described in the corresponding section. Point mutations were always confirmed through Sanger sequencing (Macrogen). Deletion alleles were confirmed through agarose gel electrophoresis by comparison of migration lengths against wild type and mutant controls. **Table M1.5** shows the primers used for genotyping procedures and the nature of the mutations that were assessed.

⌄ **Table M1.5. Primers used for genotyping of mutant strains.**

Target	Design	Type	Name	Sequence	Product
daf-19(m86)	Flanking point mutation	FW	oMM176	CAGTATTCAAACAACATGCAG	269 bp
		RV	oMM179	GAAGATACAACATGTTTCGG	
daf-19(of5)	External to deletion	FW	oRR428	ACTAAATGCAACATCGTCTGG	523 bp (wt) 511 bp (mut)
		RV	oRR430	GTAGTTGTTCTACGAAACGCA	
	Internal to deletion	FW	oRR429	GAAAGTCATGACCAATGAGGA	456 bp (wt) No band (mut)
		RV	oRR430	GTAGTTGTTCTACGAAACGCA	
daf-12(sa204)	Flanking point mutation	FW	oMM178	GATTTGGTTATTTCTAGCATC	505 bp
		RV	oMM179	AACATAGCAAATTATAGGGG	
fkh-8(tm292)	External to deletion	FW	oRR422	TTGAAGAGTGGCTGAAAGAACAGA	753 bp (wt) 222 bp (mut)
		RV	oRR423	GAGGATGGTGGGTTTGTGTAGA	
fkh-8(vlc43)	External to deletion	FW	oCC554	CTGCTTGCTGCATTCCTCGCTTC	2470 bp (wt) 606 bp (mut)
		RV	oCC555	TCTATTAACGTGTGGAGAGACGC	
	Internal to deletion	FW	oCC554	CTGCTTGCTGCATTCCTCGCTTC	609 bp (wt) No band (mut)
		RV	oCC556	CCAGATGTTGACCTCAGCCAC	

Selection of candidate plates for genotyping was usually guided by the corresponding phenotypes. Due to lack of cilia, homozygous mutant animals for the *daf-19(m86)* allele show complete dye filling defects in specific sub-populations of sensory neurons. Hence, dye filling assays were commonly used to assess the presence of such mutation. A second phenotype this allele confers is the larval metabolic resistance state known as *dauer*. However, this *dauer* phenotype is abolished in the presence of the mutant homozygous allele of *daf-12(sa204)*, a useful feature in order to assist the genotyping of both *daf-19* and *daf-12* alleles. Finally, fluorescent reporters of many cilia-related features show a severe reduction in the number of reporter-expressing cells in a *daf-19(m86)* background; *ift-20* among them. This fact was also advantageous, since *otIs395(ift-20::NLS::tagRFP)* - that was used as the panciliary expression pattern template strain - was usually present within the strains being crossed, thus facilitating the selection of F2 homozygous mutants for genotyping just by simply checking *ift-20* reporter expression under the fluorescence scope.

Microscopy & Scoring.

All scoring for this work were performed on young adult animals. At that stage, all ciliated neurons, including 4 born post-embryonically, are present. To guarantee maximum signal of reporter expression, animals were cultivated at 25 Celsius degrees. This proved to be especially important in the case of the *otIs395(ift-20::NLS::tagRFP)* strain, in which animals scored at 20 Celsius degrees showed an average of 40 reporter-positive neurons instead of the almost virtually panciliary expression pattern (60 neurons) that could be observed at 25 Celsius degrees. Thus, all other reporters were also scored at this temperature.

All *in vivo* scorings were performed under a Zeiss Axioplan 2 epifluorescence microscope with the 63X objective. To do so, worms were mounted on 4% agarose pads (diluted in distilled water) over a regular microscope glass slide (Rogo Sampaic, #11854782) using a drop of 0.5 M sodium azide (Sigma, #26628-22-8) over the dry pad as an immobilization agent. Once the animals stopped moving, this preparation was sealed by covering with a 24 x 60 mm coverslip (RS France, #BPD025).

To facilitate neuron identification and provide a standardized method, all scorings in this work were performed using anatomical landmarks. Animals were divided into five different anatomical regions as previously reported (Burghoorn et al. 2012), (see **Figure M1.4**). Sensory ciliated neurons belonging to the classical amphid and phasmid sensilla are grouped into Regions 1 and 2, respectively. Sensory ciliated neurons belonging to the cephalic, inner and outer labial sensilla plus the BAG neurons are grouped into Region 3. Region 4 contains the ciliated cells belonging to the anterior deirid sensilla plus the FLP and AQR neurons. Finally, the PDE neurons of the posterior deirid are located in Region 5. The number of ciliated sensory neurons present in Regions 1 to 5 is, 24, 5, 24, 5 and 2 respectively.

The anatomical regions defined in this way were also scored for every worm following a given order. Regions 2 and 5 (tail and PDEs, respectively) were first analysed *de*

visu. For scorings of regions 1, 3 and 4, where the expression of a panciliary reporter is expected in 53 ciliated sensory neurons, a stack of images was first acquired. For this aim, lateral views of worms' heads were positioned facing the left side of the image, with the grinder placed at the centre of the microscope field. Depending on the global positioning of the worm over the agarose pad, among 14 to 26 individual stack images separated by 1 μm were needed per head in order to cover the total volume of tissue in which reporter expression was present in a young adult hermaphrodite. Typically, a whole stack per channel was retrieved and, in the cases in which more than one fluorophore was being analysed within the same stack, DiD was first acquired, then tagRFP, then GFP and finally the DIC channel. Images were acquired with an AxioCam MRm camera (Zeiss, #426509-9901-000) coupled to a computer running the Zen System 2011 (Zeiss). Head scorings were performed manually analysing Z stacks using the open-source image processing package Fiji (ImageJ 1.50i) (Schindelin *et al.* 2012).
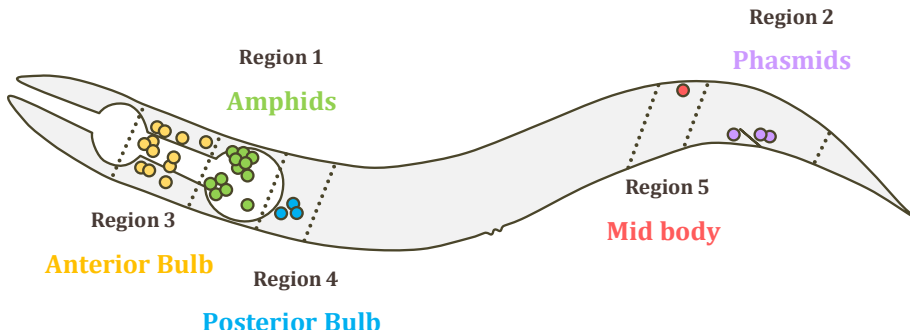
⬏ **Figure M1.4. Anatomical regions defined to score ciliated sensory neurons of *C. elegans*.**
Left lateral view of a young adult hermaphrodite showing the five anatomical regions used to facilitate ciliated neuron identification. Region 3 groups 24 ciliated neurons: 6 IL2s, 2 OLLs, 6 IL1s, 4OLQs, 2 BAGs, and 4CEPs. Region 1 gathers all 24 classical amphid neurons: 2 ASKs, 2 ADLs, 2 ASIs, 2 AWBs, 2ASGs, 2 AWAs, 2ASEs, 2AFDs, 2ADF, 2 ASHs, 2 AWC and 2ASJs. Behind the posterior bulb, Region 4 comprises 5 ciliated neurons: 2 FLPs, 2 ADEs and the non-bilateral neuron AQR. Region 5 contains 2 PDE neurons. Tail phasmid neurons are included in Region 2, accounting for 2 PHAs, 2 PHBs and the non-bilateral neuron PQR. Adapted from (Burghoorn et al. 2012).

Bioinformatic analysis of regulatory sequences.

Regulatory sequences of genes of interest were retrieved from the Ensembl BioMart website. Since most genes coding for ciliary component in *C. elegans* have been shown to be transcriptionally regulated by *daf-19* and DAF-19 binding sites (the so-called X-boxes) exhibit a positional bias in which they locate within 300 to 400 base pairs upstream the start codons of the target genes (Efimenko et al. 2005), promoter sequences from genes of interest were initially retrieved as 700 base pairs long sequences upstream their start codons. These putative promoters were used to assess the presence of common motifs across regulatory sequences of cilium-related genes.

To this aim, both Metazoa server from the Regulatory Sequence Analysis Tools (RSAT) (Defrance *et al*. 2008; Turatsinze *et al*. 2008), and the motif-based sequence analysis tools from the MEME suite (Bailey *et al*. 2015) were used (see URLs in **Annex 2**).

For mutagenesis purposes, presence of putative RFX and FKH binding sites within selected regulatory sequences was analysed through the single sequences scanning tool of the CIS-BP database (Weirauch *et al*. 2014). Within the options provided by such platform, known RFX and FKH motifs for which direct experimental evidence had been evaluated were used. Accounting for possible phylogenetical conservation and to increase the number of available motifs, worm, fly, mouse and human RFX and FKH motifs were selected. Since the single sequences scanning tool of the CIS-BP database provide several distinct models to perform the alignment between TFs motifs and analysed sequences, selection of putative RFX or FKH binding sites to be mutagenized were favoured whenever its finding was achieved with more than one of the models present in the platform.

Identification of candidate transcription factors.

The identification of candidate transcription factors involved in the regulation of ciliary features' expression was achieved by following an unbiased approach. We searched for transcription factors with enriched expression in ciliated neurons using GExplore1.4 (Hutter and Suh 2016), a free, on-line tool for large-scale mining of *C. elegans* data. This tool allows for the mining of tissue specific gene expression data from single-cell combinatorial indexing RNA sequencing (sci-RNA-seq)

(J. Cao et al. 2017) using 27 different anatomical categories, including ciliated sensory neurons. In our analysis we used the most restrictive options available: a 5-fold enrichment ratio (ratio of the gene's expression in the most highly versus the second most highly expressing category) and a false detection rate of 0.001.

To visualize transcription factor expression in the different ciliated neuron classes we used heatmaps of the sci-RNA-seq data, accessible from the "A Cell Atlas of Worm" website and performed within the integrated development environment (IDE) for R RStudio using the *pheatmap* library. Transcription factor's expression patterns from the sci-RNA-seq dataset were visualized through t-distributed stochastic neighbour embedding (t-SNE) graphics obtained by using the on-line tool SCope.

Further analysis of the expression patterns of candidate transcription factors was performed was employing an additional set of scRNAseq, the *C. elegans* Neuronal Gene Expression Network (CeNGEN) (Taylor et al. 2019) whose results are freely accessible through the on-line tool SCeNGEA.

URLs for all the on-line tools and resources named in this section can be found in **Annex 2**.

In-vivo analysis of transcription factors expression.

DAF-19 and FKH-8 *in vivo* expression was assessed through fosmid-based reporter strains.

fkh-8 fosmid strain was obtained from the CGC (OP652 strain, with genotype *unc-*

119(tm4063) III; wgIs652[fkh-8::TY1::EGFP::3xFLAG + unc-119(+)]). Wildtype developmental expression pattern of *wgIs652* was performed over embryos and worms cultivated at 25 °C also expressing the *otIs395(ift-20::NLS::tagRFP)* reporter to label ciliated neurons. FKH-8 expression within the *daf-19(m86)II; daf-12(sa204)X* mutant background was analysed using dopamine transporter *dat-1* reporter (*otIs181[dat-1::mCherry + ttx-3::mCherry] III*) to label the eight mechanosensory dopaminergic ciliated neurons (Sulston et al. 1975).

NFB672, a *daf-19* fosmid strain, was generated as previously explained by microinjecting 40 ng/μL of an EGFP-based fosmid (TransgeneOme, 38576411703852465 C10) plus 50 ng/μL each of the following co-markers: *rol-6(su1006)* coded within the plasmid pFR4 and a *ttx-3*::mCherry reporter labelling the AIY neurons coded in the plasmid pNF101 (Bertrand and Hobert 2009). The expression of the resulting *daf-19* reporter (*vlcEx361*) in *fkh-8* mutant backgrounds was also analysed focusing on specific sub-populations of ciliated neurons. The *otIs181 dat-1* reporter was used to identify dopaminergic neurons whereas DiD staining was used to label dye filling positive neurons. Generated strains are listed in **Annex 1**.

Dye-filling assays.

Lipophilic dye filling assays were used to identify certain sub-populations of amphid and phasmid neurons or to assess cilia integrity of mutant strains. This technique takes advantage of the capability that certain ciliated neurons possess to uptake a lipophilic dye present in the environment through their exposed cilia ends. The dye filling assay allows for the robust identification of ASK, ADL, ASI, AWB, ASH and ASJ amphid neurons in the head and the PHA and PHB phasmid neurons in the tail of N2 wild type animals.

DiD fluorescent dye (Thermofisher, #D7757) with red-shifted excitation and emission spectra (644 and 665 nm respectively) was used that allows for the non-overlapping detection of both GFP and tagRFP markers. DiD stock solution was prepared as a 2 mg/mL dilution in N,N-dimethyl formamide (Sigma, #D4551) and DiD staining solution was freshly prepared prior to every assay as a 1:200 dilution of this stock in M9 1X buffer.

Staining of small amounts of worms was achieved by transferring the animals into a 1.5 mL tubes containing 200 μL of the DiD staining solution. For larger amounts of animals, worm-containing plates were rinsed with 600 mL of M9 1X buffer and 20 μL of this worm-containing mix was transferred into 1.5 mL tubes. 180 μL of a volume adjusted DiD staining solution was then added to the tubes, making a final mix of 1:200 DiD in M9 1X buffer.

In both cases, tubes were then protected from the light with aluminium foil and incubated in horizontal position for 2 hours at room temperature in an orbital shaker (VWR, #444-2900 (stirring at position 5)). Past that time, tubes were vertically placed on ice for 10 minutes to allow for the worms to sit at the bottom. Animals were then collected with a glass Pasteur pipette suctioning the minimum necessary liquid. A fresh NGM plate was used to transfer the worms, distributed in small drops. Staining procedures carried out as a genotyping approximation profiting for

dye filling defects of mutant strains were finished at this point and animals were directly analysed under the fluorescence scope. In staining procedures performed for scoring purposes or a more detailed neuronal identification under the fluorescence microscope, worms were allowed to crawl for 1 hour into the bacterial lawn in order to clean the DiD out of their intestines.

fkh-8 ChIP-seq analysis.

fkh-8 ChIP-seq data was available from modERN (ENCODE project). Those ChIP-seq experiments were performed using anti-GFP antibody and the *C. elegans* strain OP652 (*unc-119(tm4063) III; wgIs652[fkh-8::TY1::EGFP::3xFLAG + unc-119(+)]*) in bleach-synchronized whole L1 larva starved for 6 hours following egg hatching.

Output *fkh-8* ChIP-seq bed narrowPeak file – containing the optimal thresholded peaks having passed the irreproducible discovery rate (IDR) analysis for both replicates – was used as the input file for the web-based analysis tool for ChIP data ChIPseek (T. W. Chen et al. 2014). Several peak-associated gene lists were created and used for further gene enrichment analysis. Visualization of individual peaks within the genomic context of *C. elegans* was performed through the Integrative Genomics Viewer (IGV) tool.

Identification of putative FKH binding motifs within the sequences of retrieved FKH-8 ChIP-seq peaks was performed by employing two different software packages. Juan Jesús Tena Aguilar, postdoctoral researcher from Centro Andaluz de Biología del Desarrollo (CSIC/UPO), applied the HOMER software (Heinz et al. 2010). Additionally, we decided to expand our analysis by employing the on-line RSAT peak-motif tool as previously reported (Thomas-Chollier et al. 2012).

URLs for all the on-line tools and resources named in this section can be found in **Annex 2**.

Enrichment analysis.

Statistical overrepresentation tests were performed over several gene lists. Gene Set Enrichment Analysis tool (Angeles-Albores *et al.* 2016, 2018) of the WormBase website was used, as it considers tissue and phenotype ontologies in addition to the Classification by Gene Ontology (GO) terms (Ashburner *et al.* 2000; Carbon *et al.* 2019). Classification through WormEnrichr, an Enrich-derived (Chen et al. 2013; Kuleshov et al. 2016) gene list enrichment analysis tool for *Caenorhabditis elegans* that makes use of 37 different term libraries, was also applied. URLs for all the on-line tools named in this section can be found in **Annex 2**.

Population synchronization.

Well-fed gravid hermaphrodites from 4 different NGM plates were collected into a 1.5 mL tube by rinsing with M9 1X buffer. Tubes were centrifuged at 2,000 rpm for 1 minute at 4 °C. Worms were washed twice with M9. Tubes were then placed on ice for 10 minutes to allow worms to sink to the bottom of the tube. Supernatant was carefully aspirated and replaced by 500 μL of an egg preparation solution (see **Materials** section for composition). Tubes were then vortexed for 2-3 minutes checking for bleaching disintegration of adult worm bodies under the scope. When most of the bodies were disintegrated, tubes were filled with ice-cold M9 1X buffer

in order to dilute the bleaching action. Tubes were centrifuged at 3,500 rpm for 2 minutes and 4 °C and supernatant was carefully aspirated. Eggs at the bottom of the tubes were washed 4 times with M9 1X buffer to remove all traces of bleaching solution. Tubes were finally filled with 1 mL of room temperature M9 1X buffer and incubated for 12 to 18 hours on a rocking nutator placed at the desired temperature before plating the newly hatched larvae on fresh NGM plates.

Small scale egg-prep protocol (Dropbleach) was also performed to achieve population synchronization of 100 to 200 worms. For this aim, 35 to 40 gravid hermaphrodites of a given genotype were transferred into a 14 μL drop of egg preparation solution placed outside the bacterial lawn of a fresh NGM agar plate. Plates were then incubated for 24 hours at 20 °C and resulting L1 animals were transferred to new NGM agar plates for further use.

RNA extraction.

Whole transcriptome sequencing was performed on several *C. elegans* strains. *fkh-8(tm292)II* strain was five times backcrossed with the N2 strain and both *fkh-8* mutant and isogenic wild type strains were prepared for sequencing. *daf-19(m86)II*; *daf-12(sa204)X* strain was crossed with the *fkh-8* backcrossed mutant and both triple *daf-19*, *daf-12*, *fkh-8* mutant and isogenic wild type *fkh-8* strains were also isolated.

All steps prior to the RNA extraction that involved alive animals were carried out at 20 °C. Whole RNA was extracted from synchronized young adults before any egg was visible inside the animals. Population synchronization was performed as previously described and egg-containing tubes were incubated on a rocking nutator for 17 hours. Triple *daf-19*, *daf-12*, *fkh-8* mutant animals and eggs proved to be especially sensitive to the bleaching treatment and times regarding animal disintegration were slightly reduced and the whole process carefully speeded up. Newly hatched larvae were plated on fresh NGM plates and animals were allowed to grow in the presence of enough food to avoid starvation. Time needed to reach the appropriate stage depended on the given genotype: control wild type *fkh-8* strain took 54 hours from larvae to young adults, whereas mutant *fkh-8* took 55 hours and both *daf-19* mutant genotypes took 69 hours.

Optimized RNA extraction from young adults was performed as follows. Animals from 3 or 4 plates were collected with M9 into one 1.5 mL tube. Tubes were centrifuged at 1,500 rpm for 1 minute and animals were washed twice to get rid of feeding bacteria. Placing the tubes on ice, worms were allowed to sit at the bottom so the most quantity of M9 buffer could be aspirated. 1 mL of TRIzol™ LS Reagent (Invitrogen, #10296010) was added to less than 100 μL of worm pellet. Tubes were then parafilm-sealed and attached to a vortex platform with adhesive type to allow for animal disaggregation for 10 minutes at low speed. Samples were allowed to sit for 5 additional minutes and centrifuged at 15,000 rpm for 10 minutes and 4 °C. From this step, precautionary measures aimed to prevent RNA degradation were strictly observed, including the use of gloves, RNase-free tubes, filter pipettes and the regular wiping of working material and surfaces with RNaseZap™ RNase Decontamination

Solution (Invitrogen, #AM9780). In hood, supernatant was transferred to a new tube and 200 μL of chloroform were added. Samples were then vortexed for 60 seconds and allowed to sit at room temperature for 3 additional minutes. Phases separation was achieved by centrifuging the samples at 15,000 rpm for 10 minutes at 4 °C. Top aqueous phase, usually less than 250 μL, was then carefully transferred to a new tube and a second identical chloroform extraction was performed. This measure proved to be mandatory in order to achieve good RNA purity ratios. New top aqueous phase, now usually below 100 μL, was carefully transferred to a new tube and 0.7 volumes of room temperature isopropanol were added and mixed by pipetting. Samples were incubated for 10 minutes at room temperature to allow for RNA precipitation and then pelleted by centrifuging at 15,000 rpm for 10 minutes and 4 °C. Pellets were usually clearly visible at the bottom of the tubes. Supernatant was then aspirated and 200 μL of room temperature 75% ethanol (diluted in diethyl pyrocarbonate (DEPC) treated ddH20) was added. Samples were briefly vortexed to clean the pellets and centrifuged again at 15,000 rpm for 5 minutes and 4 °C. Supernatant was finally aspirated and RNA pellets were allowed to air-dry in hood for 5 minutes at room temperature. Depending on pellet size, 25 to 50 μL of DEPC-ddH$_2$0 were added and pipetted to facilitate RNA re-suspension. Samples were finally incubated at 60 °C in a water bath for 10 minutes and placed on ice until final storage at -80 °C.

Preliminary RNA quantity and quality was measure using a Nanodrop 2000 spectrophotometer (TermoFisher, #ND-2000). Prior to RNA-seq, a DNase treatment was performed. To this aim, 5 μg of extracted RNA was mixed with 0.5 μL of DNase I (New England BioLabs, #M0303S) and 5 μL of DNase I buffer to a final volume of 50 μL in DEPC-ddH$_2$O. Samples were then incubated at 37 °C for 10 minutes. 450 μL of DEPC-ddH$_2$O and 500 μL of chloroform were added to perform RNA extraction. From this step, RNA recovery was performed as previously explained. Quantity and quality of final DNase-clean RNA samples were measured with the Agilent 2100 Bioanalyzer microfluidics-based platform according to company protocols.

Three independent RNA samples from all four strains previously mentioned were shipped to Macrogen. **Table M1.6** shows the detailed information regarding commercial kits, library preparation, sequencer and software/programs used for RNA sequencing. **Table M1.7** gathers a summary of the raw data produced within the sequencing process.

fkh-8 RNA-seq analysis.

Differentially expressed genes analysis was performed by Macrogen: functional annotation and gene-set enrichment analysis were performed using GO and KEGG databases. Macrogen datasheets including gene expression information from all samples were individually supervised upon arrival. Raw P-value, and not multiple testing corrected q-Values, plus absolute values for the fold changes expressed as their binary logarithms being greater or equal to 2 were the criteria used by Macrogen pipeline to

address differential expression for genes when comparing between two genotypes. Alternative differentially expressed gene lists were created by applying several different criteria based on q-Values, p-Values and fold changes. These lists were then used for gene enrichment analysis.

▼ **Table M1.6. RNA sequencing project materials.**

Item	Version
Library Kit	TruSeq Stranded mRNA LT Sample Prep Kit
Library Protocol	TruSeq Stranded mRNA Sample Preparation Guide, Part # 15031047 Rev. E
Reagent	TruSeq rapid SBS kit or Truseq SBS Kit v4
Sequencing Protocol	HiSeq 2500 System User Guide Document # 15035786 v02 HCS 2.2.70
Type of Sequencer	HiSeq 2500
Read Length	101 bp
Sequencing Control Software	HCS 2.2.70
Raw Sequences Quality Check Program	FastQC v0.11.7
Trimming Program	Trimmomatic 0.38
Alignment Programs	HISAT2 version 2.1.0, Bowtie2 2.3.4.1
Assembler Program	StringTie version 1.3.4d

▼ **Table M1.7. Summary of raw data statistics produced within the RNA sequencing process.**
Three independent RNA samples from four different *C. elegans* strains were shipped for whole RNA sequencing to Macrogen. Raw data from each replicate is shown. Q20 represent the ratio of bases that have phred quality score greater than or equal to 20. Phred quality score 20 means 99% accuracy in base sequencing in each cycle and reads over score of 20 are accepted as good quality.

Genotype	Replicate	Concentration (ng/µL)	Total Read Bases	Total Reads	Q20 (%)
wild type	1	255.0	4,556,029,806	45,109,206	95.80
	2	257.5	3,930,898,386	38,919,786	95.07
	3	195.7	3,964,667,938	39,254,138	95.27
fkh-8(tm292)II	1	91.1	4,595,951,470	45,504,470	96.12
	2	158.2	3,650,340,384	36,141,984	96.07
	3	194.4	4,522,725,864	44,779,464	95.50
daf-19(m86)II; daf-12(sa204)X	1	91.1	4,831,082,096	47,832,496	95.97
	2	158.2	4,403,529,502	43,599,302	95.44
	3	194.4	3,878,314,958	38,399,158	95.91

Genotype	Replicate	Concentration (ng/µL)	Total Read Bases	Total Reads	Q20 (%)
fkh-8(tm292)II; daf-19(m86)II; daf-12(sa204)X	1	450.8	4,750,214,830	47,031,830	95.58
	2	453.5	4,454,420,978	44,103,178	95.78
	3	491.6	3,634,891,020	35,989,020	95.82

CRISPR-based generation of a *fkh-8* null allele.

Whole deletion of the *fkh-8* locus was performed through a co-CRISPR strategy (Kim *et al.* 2014) using *dpy-10(cn64)* as conversion marker (Arribere *et al.* 2014).

Custom CRISPR RNAs (crRNAs) were ordered (IDT, Alt-R® CRISPR-Cas9 crRNA XT) targeting both sides of the desired deletion of *fkh-8* and at the 5' of the *dpy-10* site of mutation (**Table M1.8**). Single-stranded oligodeoxynucleotide (ssODNs) of approximately 100 base pairs overlapping each side of the genetic modifications were also ordered (IDT) and used as donor templates to achieve homology-directed repair. (**Table M1.8**). Cas9 nuclease (IDT, #1081058) and the universal trans-activating crRNA (tracrRNA) needed to initiate enzymatic activity (IDT, #1072532) were used.

▼ **Table M1.8. *fkh-8* co-CRISPR knock-out crRNAs and ssODNs.**

Oligo name	Oligo sequence	Oligo function
fkh-8(vlc-43)II; full deletion of *fkh-8*		
fkh-8 5' crRNA	GAGAATCGGTGGATGAATGG	5' crRNA
fkh-8 3' crRNA	TCATCTGAATCATTGTTAAC	3' crRNA
oc547	CATTCCATTTTCACAAACCTCATGCATTTCCAATCTCAATTATTCCTCCATAAATTGTTAATCTAAAGGTTCAAAAACTCACATATTTTTCACACAGTGT	Repair template
dpy-10(cn64)II; used for co-CRISPR		
dpy-10 crRNA	GCTACCATAGGCACCACGAG	5' homology arm
dpy-10 ssODN	CACTTGAACTTCAATACGGCAAGATGAGAATGACTGGAAACCGTACCGCATGCGGTGCCTATGGTAGCGGAGCTTCACATGGCTTCAGACCAACAGCCT	Repair template

Co-CRISPR injections were performed on young adult hermaphrodites expressing the pan-ciliated reporter *otIs395(ift-20::NLS::tagRFP)III*. Microinjection mix was freshly prepared with all 3 crRNAs plus the tracrRNA, ssODNs and Cas9 nuclease itself (see **Table M1.9** for the detailed recipe of the microinjection mix). Ribonucleoprotein complex formation was achieved by incubating this mix for 10 minutes at 37 Celsius degrees. Before use, the final mix was incubated on ice for 30 minutes. Microinjection procedures were performed as previously described.

Notice that the ratio between concentrations of *dpy-10* and *fkh-8* crRNAs is 1:7.

Component	Volume (μL)	Final concentration
fkh-8 5' crRNA	1	35 μM
fkh-8 3' crRNA	1	35 μM
dpy-10 crRNA	1	5 μM
Cas9 trarnRNA	1.5	48 μM
Cas9 nuclease	1.64	10 μM
oc547	1	175 ng
dpy-10 ssODN	1	3.3 μM
nuclease-free water	1.86	(–)
MIX	10	

Within the co-CRISPR strategy, the crR-NAs for conversion markers are injected at a lower concentration than the ones corresponding to the desired mutation. As the marker and target modifications are independent events, if the effect of the visible conversion marker is detected then the target modification is highly probable to have occurred. Three days after the microinjection procedure, F1 offspring was checked for the presence of the dominant negative effect of the *dpy-10(cn64)* allele. *dpy-10* is a gene involved in the development of the cuticle and the R92C protein substitution (change from arginine to cysteine in the amino acid number 92) of the *cn64* allele confers the helically twisted body of the "roller" phenotype in heterozygous animals. In homozygous animals, this allele gives the worms a characteristic shorter and fatter phenotype termed "dumpy". From plates in which the "roller" phenotype was clearly visible within the F1 progeny, 457 non-roller hermaphrodites were isolated in new NGM agar plates. Non-roller worms were selected to prevent the possible linkage of the two modification events since *dpy-10* and *fkh-8* are both located in the chromosome II. Three days later, once the F2 had been produced, 100 F1 mothers were individually lysed and genotyped for the presence of the *fkh-8* deletion. 4 independent modification events were achieved, 1 of them being almost exactly as designed and thus generating the allele named *vlc43*. Worms from the F2 progeny of that plate were singled in new NGM agar plates and genotyped for the homozygous deletion of *fkh-8*. The isolated new strain was backcrossed 4 times into an N2 background to clean possible off-targets from the CRISPR method.

Co-CRISPR design and microinjection procedures were performed by Carlos Mora Martínez (then a PhD student at the Flames Lab). Ainara Esteve Serrano (then a Master student at the Flames Lab) aided during the genotyping process.

Rescuing experiments.

The coding DNA (cDNA) sequence for *fkh-8* (see **Table M1.10**) was synthetically generated (Biomatik). This cDNA was amplified by PCR and cloned into a pPD95.75-based plasmid under the control of a 400 base pairs long promoter of the *dat-1* gene (whose expression is unaffected in *fkh-8* mutants), hence directing its expression to the eight dopaminergic neurons of the worm. As an inner control for expression, the stop codon of *fkh-8* cDNA was removed and replaced by means of fusion PCRs by the DNA sequence of the self-cleaving peptide T2A (Ahier and Jarriault 2014), thus creating an autocatalytic peptide that generates independent FKH-8 and EGFP proteins (see **Figure M1.5**). See **Table M1.11** for details and primers used for the cloning process of these rescuing plasmids.

Table M1.10. Coding sequence of *fkh-8* used for rescuing experiments.

fkh-8 coding sequence was used to rescue null *fkh-8* mutant genotypes. Sequence was synthetized according to the indicated public data available in WormBase. Alternating exons within the coding sequence are depicted as bold / underlined letters.

Transcript	Coding Sequence	Protein Length
F40H3.4.1 (1104 bp) F40H3.4.2 (1152 bp)	F40H3.4 (987 bp)	310 aa

ATGTCAGATCTAAGCTCAAGCTTGTGCCAATTGAACTGGCTCATTGCCAAAGGTGGACTGAACACAGTACAAGAAATCGAAGTACCTGAAAACCCT AATATTGTTGGATCTGTGAATGTTTCGCCACTGGTCCTATCTCCTACAGTACCTGCAGAAACGTCAAAG<u>CCAGTGGCTCCTGCACACGAAGGAAAACG GAGGAGAATTTTTGATGGAGCGGATAAGCCACCATATTCATACTCTCAACTTATTCGTCTAGCTATCGAAGATACTCCCGATAAAAAATGTACTCTTGCTG AAATCTACTCTTTCATTGCTCACAACTTTCAATTCTATCGTGAAAACCGGAATTCAAGTTGGAAG</u>**AATTCCATTCGCCATAATCTGTCATTAAATAAACA ATTCAGCCGCATTGAGAAGACTGATGGAGATAGAA**<u>GAGGATGGTGGGTTTGTGTAGATCCGCCAGCCAAGAAACCACGAATTCTCAAAGGATCCCCAGT GCGAGTAAATCCAATTTACGAGCATCTATATCATA</u>**ATAAACAAGACATGCCAAATTTCACTCCTCCGCCAGAAGATCCATTCCTGTCAGAAATTAATG GAAATGTACATGAACTTACG**<u>GAACACGAAATAAGAGATTTGAATTTGTTCGAGTCATATGATCTAAACTCATCATTCCGTGATGTCTACAATCAAATTTT TGAGAAGTCCACTTCCCCCAATGGGAAGAAGCAA</u>**GCTGCTCAAATCGATTGGTTGAAAATAAGCTTAGAAGCTGCTGGACTTGATTATCACGATGAAC AAGAACTTCAAGAAGTTGATACGGATAAGTTGAAAG<u>ATTACGTCTACAATGGATTCCCAGCAGAATGCGACAGTGATGTCTCAACTCCAAGAACAGACT CTGGCCGTCATTCATCCTCAGATTCTGTTCTTTCAGCCACTCTTCAACCAGTTAACAATGATTCAGATGATGAGTATGACTGGGATAAATTGCTATAA</u>

otIs395(ift-20::tagRFP)III; fkh-8(vlc43)II
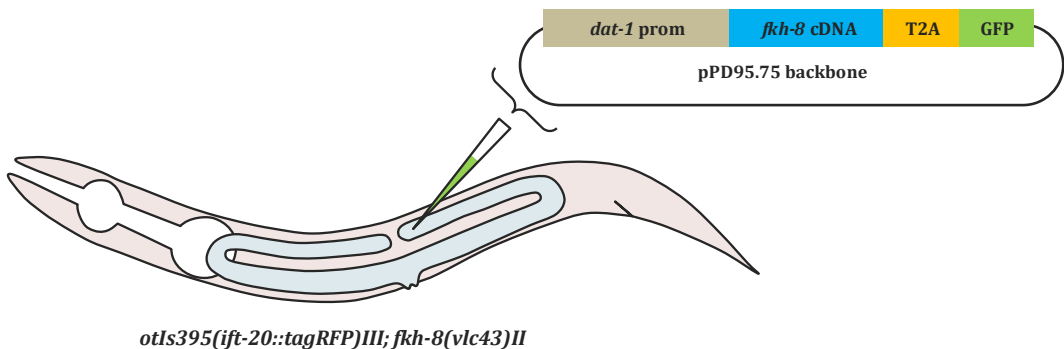
⌃ **Figure M1.5. Rescuing strategy for *fkh-8* null mutant worms.**

Rescuing plasmid carrying a construct expressing both the coding sequence of *fkh-8* and GFP separated by the viral autocatalytic T2A sequence and under the control of the *dat-1* promoter were injected at the syncytial arms of both gonads in null *fkh-8* mutant animals also expressing the panciliary reporter *ift-20*.

Target	Type	Name	Sequence
***fkh*-8 coding sequence**	FW	oRR455	GAGAGAGGTACCATGTCAGATCTAAGCTCAAGC
	RV	oRR456	CAGACTTCCTCTGCCCTCTAGCAATTTATCCCAGTCATACTC
	Forward primer included a KpnI restriction site for cloning purposes whereas reverse primer harboured a tail allowing for fusion to theT2A sequence.		
T2A	FW	oRR450	GAGGGCAGAGGAAGTCTGC
	RV*	oRR451	TGGGCCAGGATTCTCCTC
	Template	oRR452	GAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCA
	RV	oRR453	GAAAAGTTCTTCTCCTTTACTCATTGGGCCAGGATTCTCCTC
	oRR452 held the whole T2A sequence. A first PCR was set up with oRR452 and oRR453 to generate a T2A fragment with a tail allowing for fusion to the 5' end of the GFP coding sequence. oRR450 was used for that aim, oRR451 served for sequencing purposes.		
GFP	FW	oRR454	ATGAGTAAAGGAGAAGAACTTTTC
	RV	oRR45	TTAGTTAGTACCGAACTGTTTAAACTTACGT
	These primers were used to amplify 405 bp of the GFP coding sequence from pPD95·75. A fusion PCR was set up to create a T2A::GFP fragment. Once achieved, a second fusion PCR was used to created the *fkh-8*::T2A::GFP fragment. All three fragments were cloned between the KpnI and XhoI sites of the pPD95.75-based plasmid under the control of a *dat-1* promoter.		

This rescuing plasmid was injected into null *fkh-8* mutant young adult hermaphrodites also expressing the pan-ciliated reporter *otIs395(ift-20::NLS::tagRFP)*. Injection mixes were composed by 50 ng/μL of the purified rescuing plasmid plus 100 ng/μL of the *rol-6(su1006)* co-marker encoded by the pFR4 plasmid. Microinjection procedures and transformed lines were obtained as previously stated.

Site directed mutagenesis of *xbx-1 cis*-regulatory region.

To test for functionality, we designed specific mutations of predicted FKH binding sites found in a putative *cis*-regulatory region of the panciliary *xbx-1* gene as previously stated. The use of known FKH motifs from different species as well as the use of different alignment models provided overlapping putative FKH binding sites of different lengths located at defined positions within the analysed sequences. Accordingly, Clustal Omega multiple sequence alignment program (Madeira *et al*. 2019) (see **Annex 2** for URLs) was used to align such putative FKH binding sites and the common core of 9 base pairs were selected to be mutated.

Mutation criteria first accounted for the nature of the nitrogenous bases (purine to pyrimidine and vice versa) and number of hydrogen bonds they could form (2 for adenine (A) and thymine (T) and 3 for guanine (G) and cytosine (C)). Hence, A was first mutated to C and G was first mutated to T (and vice versa). Mutated sequence was checked to discard generation of new TF binding site motifs using both the motif scan tool of the CIS-BP database and, in order to take detailed

control over the motif database being used, the Tomtom tool (Gupta et al. 2007) from the MEME Suite website (**Annex 2**). When newly (mutated) created sites were matching known functional binding sites, second criterium related to number of bonds was used. If even then new predicted sites were created, convenient manual punctual mutations were applied.

This mutated version of the *xbx-1*(short) promoter was generated through PCR from N2 lysates, introducing the desire mutation (underlined) within the forward primer (oRR370: CTCTCCTCGTTT<u>CTGTGGTCC</u>CGCCCGT). Final fluorescent reporter was generated through fusion PCR as previously stated by using the corresponding primers listed in **Table M1.3**.

Behavioural assays.

To evaluate sensory-mediated behaviours, mechanosensory assays were performed over several *C. elegans* strains of different genetic backgrounds. All mechanosensory assays were performed over small-scale synchronized populations of young adult hermaphrodites. One day before the assays, synchronized L4 hermaphrodites were transferred to new OP50-seeded NGM agar plates. Three independent biological replicates per assay type were performed.

Both gentle and harsh touch mechanosensory tests were performed as previously described (Chalfie *et al.* 1985). Briefly, gentle touch assays were performed by alternatively stroking the animal just behind the pharynx and just before the anus with an eyebrow hair attached to a pipette tip for a total amount of 10 strokes (Hobert *et al.* 1999). To distinguish between touch-insensitive animals and those incapable of movement, harsh touch assays were also performed by stroking the worms across the posterior half of their bodies in a top-down manner with a platinum wire. Each worm was tested five times with a 2 minutes interval between each trial (Li *et al.* 2011).

Nose touch mechanosensory tests were performed as previously described (Kaplan and Horvitz 1993). Ten minutes before the assay, young adult hermaphrodites were transferred to non-seeded NGM agar plates and nose touch responses were elicited by placing an eyelash attached to a pipette tip in the path of an animal moving forward, hence causing a "nose-on" collision. With brief modifications from (Brockie *et al.* 2001), five consecutive nose touch trials were scored for each worm.

For all assays, positive responses were computed whenever the animals moved away from the stimulus, backward for anterior touches and forward for posterior ones. Individual responses of all trailed animals were recorded and a population response index (RI) was calculated for every replica of each type of assay as follows:

$$RI = \frac{\text{total number of positive responses}}{\text{total amount of strokes}}$$

Sensory-mediated behaviours towards environmental chemical cues were also tested. All chemosensory assays were performed over small-scale synchronized

populations of young adult hermaphrodites as previously explained. Chemotaxis towards diacetyl, 2-heptanone, NaCl and 2-nonanone were performed over freshly washed worms. For this aim, worms were rinsed three times with 1 mL of filtered, autoclaved CTX solution (see Materials for composition), aspirating the supernatant to a final volume of approximately 100 μL. To perform the assays, 2 μL of this worm-containing solution with no less than 25 animals were placed at the proper place of the assay plates (see **Figure M1.6**). During the assays, worms were allowed to freely crawl across the plates for 60 minutes at room temperature. Plates were then stored at 4 °C until the next day when worms' positions were scored and behavioural indexes were calculated.
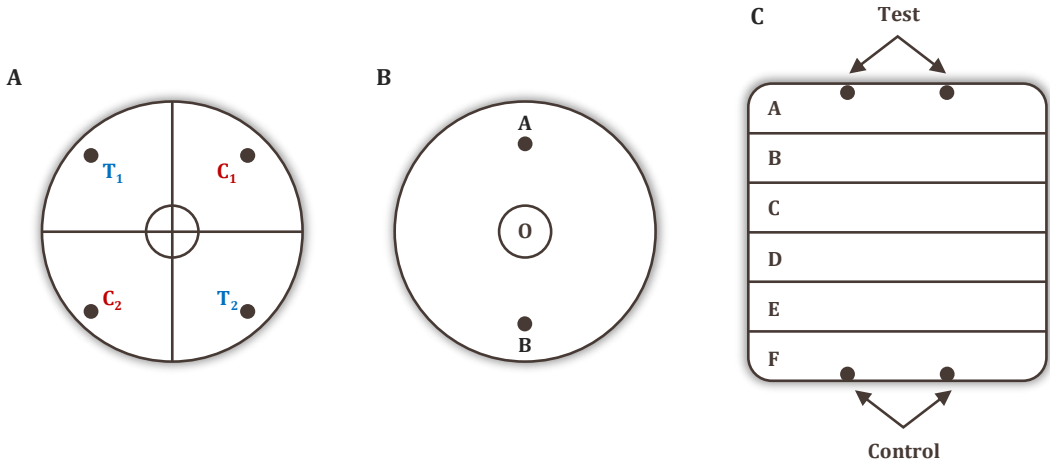
⬈ **Figure M1.6. Plate designs for chemosensory assays.**
A) A four-quadrant paradigm was used to test the response to diacetyl. Dots represent the approximate locations of either test (T in blue) or control (C in red) solutions over the agar. A central circle of 1 cm of diameter acts as the distance threshold that worms have to trespass in order to be scored. **B)** The two-halves paradigm was used to test the response to 2-heptanone and NaCl. Dots represent the locations at 0.5 cm from the edge of the wall plate of either test (A) or control (B) solutions over the agar. A central circle of 1 cm of diameter (O) acts as the distance threshold that worms have to trespass in order to be scored. **C)** Six-equal-sector paradigm was used to test the response to 2-nonanone. Worms were placed at the intersection between C and D sectors. Dots represent the approximate locations of either Test or Control solutions over the agar.

With few modifications, volatile diacetyl attraction assay was performed as described by (Margie, Palmer, and Chin-Sang 2013). A four-quadrant paradigm drawn at the base of non-seeded NGM agar plates was used, adding a circular central area that worms had to trespass to be scored (see **Figure M1.6.A**). Stock diacetyl (Sigma-Aldrich, #803528) test solution was prepared as a 0.5% V/V mix in absolute ethanol (Scharlau, #ET00101000). Absolute ethanol was used as control solution. Immediately after the worms were plated, 2 μL of a mix combining equal volumes of diacetyl stock solution and sodium azide 1M (dissolved in M9 1X buffer and used as an anaesthetic to prevent worm

movement upon reaching a given quadrant) were pipetted onto the T sites of the agar plate. Same procedure was then performed for the control sites. Once the assay was finished, a chemotaxis index (CI) was calculated as:

$$CI = \frac{\text{worms in } (T_1+T_2) \; - \; \text{worms in } (C_1+C_2)}{\text{total scored worms}}$$

Chemotaxis assay towards the attractive volatile organic compound 2-heptanone was performed as previously reported (Zhang *et al.* 2016). A two-halves paradigm drawn at the base of non-seeded NGM agar plates was used, adding the threshold distance by (Margie *et al.* 2013) to prevent immobile worms from skewing the data. (see **Figure M1.6.B**). 2-heptanone (Sigma Aldrich, #W254401) test solution was prepared as a 1:10 V/V mix in ethanol absolute. Ethanol was used as control solution. Immediately after the worms were plated, 3 µL of a mix combining equal volumes of 2-heptanone stock solution and sodium azide 1M were pipetted onto the A site of the agar plate and control solution in the B site. Once the assay was completed, a CI was calculated per genotype testes as:

$$CI = \frac{\text{worms in } (A) \; - \; \text{worms in } (B)}{\text{total scored worms}}$$

The two-halves paradigm was also used to assess chemotaxis toward NaCl. In this case, radial gradients of either test or control solutions were created prior to worm loading as originally stated (Ward 1973). Following (Frøkjær-Jensen, Ailion, and Lockery 2008), 10 µL of NaCl (Sigma, #S3014-1KG) 2.5 M (dissolved in double-distilled water (ddH$_2$O)) or ddH$_2$O itself were respectively pipetted onto the agar surface at A and B spots and allowed to diffuse for 12-14 hours at room temperature. To increase steepness of the gradients, 4 µL of NaCl 2.5 M or ddH$_2$O solutions were additionally added to the A and B spots respectively 4 hours prior to the chemotaxis assay. The population CI was calculated through the same expression used for the 2-heptanone attraction assay.

Avoidance assay of the volatile repellent 2-nonanone was performed as previously reported (Troemel et al. 1997). Briefly, six equal sectors were drawn on the base of squared plates (90 x 15 mm, Simport™, # 11690950) containing 15 mL of standard NGM agar (see **Figure M1.6.C**). Stock 2-nonanone (Sigma-Aldrich, #W278550) test solution was prepared as a 1:10 V/V mix in absolute ethanol. Ethanol was used as control solution. Immediately after the worms were plated on the centre of the plate, 2 µL of a mix combining equal volumes of 2-nonanone stock solution and sodium azide 1M were pipetted onto two spots within sector A. Same procedure was then performed for the ethanol control sites within sector F. A population avoidance index (AI) per genotype and replica was calculated as:

$$\text{AI} = \frac{\text{worms in (A+B)} - \text{worms in (E+F)}}{\text{total amount of worms}}$$

Avoidance responses to water-soluble compounds were evaluated by using the drop test as previously described (Hilliard et al. 2002). Briefly, glass capillaries (Harvard Apparatus, #W3 30-0045) were narrowed by means of a magnetic glass microelectrode horizontal puller (Narishige, #PN-30) set with the parameters: heater, 70; magnet sub, 59; magnet main, 59. The aim of this procedure was to create a capillary narrow enough to deliver a drop so small that could not disturb the on-going free movement of a worm when delivered near its tail. Following (Hilliard *et al*. 2004) with few modifications, well-fed synchronized young adult hermaphrodites were washed three times with M13 buffer (see Materials section for composition). 5 animals were then placed on unseeded NGM agar plates and allowed to rest for 10 minutes. Two test solutions were assayed: 0.1% w/V sodium dodecyl sulfate (SDS) (Sigma, #L3771-100G) and 0.1 mM $CuSO_4$ pentahydrate (Merck, #1027901000), both dissolved in the M13 buffer that acted as control solution. Each animal was tested, allowing for 2 minutes between each stimulus, with 4 single drops of the control solution and then with 4 single drops of the testing solution. Presence or absence of the avoidance backward movement towards the stimulus was then scored within 4 seconds after encountering the substance. A population avoidance index per genotype and replica was calculated as:

$$\text{AI} = \frac{\text{number of responses}}{\text{total amount of drops}}$$

Detailed population sizes, genotypes tested and number of replicas performed per behavioural assay type can be found in **Table 2.12**. All strains used for these behavioural studies appear in **Annex 1** too.

Statistical analysis.

The scoring of reporter-positive neurons implied the acquisition of a count data type in which observations took only non-negative integer values. This type of data is (should be) formally analysed through Poisson models (Guenni 2012; Hilbe 2011). However, as a consequence of the central limit theorem, Poisson distributions can be approximated to Gaussian models (Singer 2011). The goodness of this approximation increases as long as it does the rate parameter (λ) of the Poisson distribution. λ is a positive parameter representing the expected number of events meant to occur within a given interval. In the context of this work, it would represent the number of reporter-positive neurons within one animal. Since values of $\lambda \geq 15$ greatly approximate a Gaussian distribution (Boucher 2011), the majority of those scorings could be deemed as normally distributed, thus allowing the use of parametric statistics upon them. Mean number of reporter-expressing neurons was considered to be the best descriptor for the samples for most of the experiments presented in this thesis, hence parametric Student *t*-tests were performed to assess for statistical differences between

two groups. Some measures were taken when robustness of the *t*-test ("a test is said to be robust if it is relatively insensitive to violation of its underlying assumptions" – Kalpić *et al*. 2011) could be affected. Prior to the *t*-test, the variances of the two groups to be compared were calculated. In the case of equal variances, the standard *t*-test was performed. In the cases of unequal variances, in which the *t*-test remains robust only if the sample sizes are equal and both with $N \geq 5$ (Huck 2008), the Welch t-test approximation was used. In those cases in which whole animals were scored, what was meant to be rejected when performing hypothesis testing was if the true mean was different – and not just larger or smaller – than hypothesized; accordingly, those *t*-tests were calculated as two-tailed. However, once that difference had been assessed, corresponding one-tailed *t*-tests were performed when considering only the number of reporter-expressing neurons located in particular anatomical regions. However, in those cases in which the count (and not the mean) was considered as the best descriptor for the sample, appropriate Fisher or Chi-squared tests were performed.

Behavioural responses were ultimately analysed through the corresponding indexes ranging from 0 to 1 (or to -1 to 1 when avoidance responses were assayed), hence determining continuous parameters by themselves in which the mean was chosen as the best descriptor for the samples. For each type of assay, a population-based mean index was calculated for every replica and a final response index was then calculated as the mean of these means. Prior to hypothesis testing, the Shapiro-Wilk test (Shapiro and Martin 1965) was used to address the normality of these final indexes. Parametric Student *t*-tests were then performed to assess for statistical differences between two groups, taking into account all the considerations previously stated.

Multiple testing was adjusted by Bonferroni correction accounting for all possible pair comparisons within a given set. However, Bonferroni correction is a rather conservative test that prevents type I errors (false positives) to occur at the expense of increasing the probability of type II errors (false negatives) (Lee and Lee 2018). In the cases in which comparisons failed to pass the Bonferroni threshold, the Benjamini-Hochberg stepwise procedure was used to control the false discovery rate.

Means, standard deviations, F-tests for equality of variances and *t*-tests for equality of means were all calculated through in-built functions of the Microsoft Excel software (365 MSO (16.0.12527.20170) 32 bits). Multiple comparisons procedures, Shapiro-Wilk tests and two-proportions Z-tests were manually calculated following the corresponding formulas available in the aforementioned references.

➤ **Table M1.12. Population sizes used in behavioural assays.**
Number of worms per replica are indicated separated by vertical bars (|). Assays in which a threshold distance has to be trespassed indicate the number of worms that were assayed over (/) the total worms that were plated.

Strain	Genotype	Worms per replica	Population size
Gentle touch assay			
N2	wild type	20 | 20 | 20	60
(no name)	*fkh-8(tm292)II*	20 | 20 | 20	60
NFB2057	*fkh-8(vlc43)II*	20 | 20 | 20	60
CB1338	*mec-3(e1338) IV*	20 | 20 | 20	60
Harsh touch assay			
N2	wild type	30 | 30 | 30	90
(no name)	*fkh-8(tm292) II*	30 | 30 | 30	90
NFB2057	*fkh-8(vlc43) II*	30 | 30 | 30	90
CB1338	*mec-3(e1338) IV*	30 | 30 | 30	90
Nose touch assay			
N2	wild type	20 | 20 | 20	60
(no name)	*fkh-8(tm292) II*	20 | 20 | 20	60
NFB2057	*fkh-8(vlc43) II*	20 | 20 | 20	60
KP4	*glr-1(n2461) III*	20 | 20 | 20	60
Diacetyl assay			
N2	wild type	168/168 | 69/69 | 100/103	337/340
(no name)	*fkh-8(tm292) II*	55/57 | 76/85 | 110/110	241/252
NFB2057	*fkh-8(vlc43) II*	111/115 | 92/107 | 71/74	274/296
CB1033	*che-2(e1033) X*	99/104 | 87/89 | 79/79	265/272
2-heptanona assay			
N2	wild type	119/124 | 118/129 | 126/133	363/386
(no name)	*fkh-8(tm292) II*	65/68 | 81/94 | 92/102	238/264
NFB2057	*fkh-8(vlc43) II*	71/87 | 75/83 | 73/85	219/255
CX2065	*odr-1(n1936) X*	137/153 | 120/137 | 74/108	331/398
NaCl assay			
N2	wild type	54/62 | 76/78 | 70/72	200/212
(no name)	*fkh-8(tm292) II*	92/105 | 106/116 | 97/106	295/327
NFB2057	*fkh-8(vlc43) II*	97/111 | 49/52 | 67/78	213/241
PR678	*tax-4(p678) III*	42/63 | 67/82 | 54/66	163/211

Strain	Genotype	Worms per replica	Population size
2-nonanone assay			
N2	wild type	59 | 128| 114 | 165	466
(no name)	*fkh-8(tm292) II*	76 | 123| 129 | 209	537
NFB2057	*fkh-8(vlc43) II*	82 | 92 | 130 | 139	443
CB1033	*che-2(e1033) X*	22 | 79 | 126 | 39	273
Drop test (M13 & SDS 0.1%)			
N2	wild type	5 | 5 | 5 | 5 | 5 | 5	30
(no name)	*fkh-8(tm292) II*	5 | 5 | 5 | 5 | 5 | 5	30
NFB2057	*fkh-8(vlc43) II*	5 | 5 | 5 | 5 | 5 | 5	30
PR678	*tax-4(p678) III*	5 | 5 | 5 | 5 | 5 | 5	30
Drop test (M13 & Cu^{2+})			
N2	wild type	5 | 5 | 5 | 5 | 5 | 5	30
(no name)	*fkh-8(tm292) II*	5 | 5 | 5 | 5 | 5 | 5	30
NFB2057	*fkh-8(vlc43) II*	5 | 5 | 5 | 5 | 5 | 5	30
CB1033	*che-2(e1033) X*	5 | 5 | 5 | 5 | 5 | 5	30

Materials

The following section contains the detailed composition of reagents used that were not mentioned elsewhere.

Component	Concentration
NGM agar	
Agar	16 g/L
NaCl	0.05 M
Peptone	2.5 g/L
MgSO$_4$	1 M
Cholesterol	1 mL/L of a 5 mg/mL solution in ethanol
CaCl$_2$	1 M
KPO$_4$	1 M, pH 6.0
Nystanin	2.5 mL/L
Streptomycin	0.15 g/L
Worm lysis solution	
Proteinase K	2% V/V
GoTaq® Colorless Reaction Buffers for PCR	20% V/V
Nuclease-free water	78% V/V

Component	Concentration
M9 buffer	
Na$_2$HPO$_4$·12H$_2$O	80 mM
KH$_2$PO$_4$	22 mM
NaCl	8.5 mM
NH$_4$Cl	0.2 M
Egg preparation solution	
NaClO	50% V/V
KOH 5 M	20% V/V
ddH$_2$O	30% V/V
CTX buffer	
CaCl$_2$	10mM
MgSO$_4$	1 mM
KPO$_4$	5 mM
M13 buffer	
Tris-HCl	30 mM, pH 7.0
NaCl	100 mM
KCl	10 mM

Results

1. A regulatory-motif enrichment analysis of the *C. elegans* ciliome.

Subsets of ciliated neurons maintain the expression of ciliary genes in the absence of *daf-19*.

As stated in the introduction, *daf-19* acts as a terminal selector for the sensory ciliome in *C. elegans*. Accordingly, *daf-19* mutants display broad defects in the expression of genes encoding for cilium components (Efimenko et al. 2005; Phirke et al. 2011; Schafer et al. 2003; Swoboda et al. 2000a; Wang et al. 2010; Williams et al. 2008; Winkelbauer et al. 2005). *daf-19* directly regulates these genes through an RFX binding motif, the so called X-box, that is found near the ATG usually not more than 300 base pairs upstream the translational start site of ciliary genes. To date, no other TF has been involved in the direct regulation of gene expression for the sensory ciliome in any organism. However, also as stated in the introduction, terminal selectors are known to act in combinations controlling the expression of their target effector genes (Hobert 2008, 2011). Thus, we hypothesized that additional TFs should be working together with *daf-19* in the direct activation of genes encoding for sensory cilium components.

As a first approach, to test this hypothesis, we aimed to identify ciliary genes whose expression was not fully abolished in *daf-19* null mutant animals. Such circumstance would demonstrate that other TFs could compensate the absence of *daf-19* and still be sufficient to activate gene transcription of genes encoding for cilium components. Based on the literature, we selected ten structural ciliary genes known to be broadly expressed across the whole ciliated system of *C. elegans* and whose expression was reported to be partially affected in null *daf-19* mutant animals. Namely, those genes were: *che-11*, *che-13*, *ift-20*, *mks-1*, *oms-1*, *oms-5*, *peli-1*, *tmem-107*, *tub-1* and *xbx-1*. Although some remaining expression was assessed for these genes in the absence of *daf-19* both through reporter and microarray experiments (Burghoorn et al. 2012; Chen et al. 2006; Chu et al. 2012; Phirke et al. 2011; Schafer et al. 2003; Swoboda et al. 2000a), no detailed analysis on the remaining cells had been performed at that moment.

The pioneering work by Swoboda and collaborators (Swoboda et al. 2000a) already showed that *daf-19* acted as a key TF regulating the expression of those structural ciliary genes that are expressed in most or all ciliated neurons. The role of *daf-19* controlling the expression of subtype-specific ciliary genes is still controversial. While initially it was proposed that the role of *daf-19* was limited to the activation of core cilium-related genes (Swoboda et al. 2000a), other reports expanded the action of *daf-19* to non-structural ciliary genes (Wang et al. 2010). Thus, to complement our reporter expression study, we also analysed the expression of genes important for specific cellular functions that localize to subset of ciliated neurons, such as cilium-located receptors. To this aims, we decided to focus on the transient receptor potential cation channel

subfamily V (TRPV), for whom detailed expression pattern within the ciliated system of *C. elegans* was already known (Tobin et al. 2002). We selected 4 out of the 5 members of the family: *ocr-1*, *ocr-2*, *ocr-4* and *osm-9*, excluding *ocr-3* due to its expression in the rectal gland cells (Tobin et al. 2002). We built GFP reporters for 13 out of these 14 selected cilium-related genes and used the *ift-20* integrated reporter strain *otIs395(ift-20::tagRFP)III* (see **Figure R.1**). Interestingly, while the promoter sequences of all structural ciliary genes contained experimentally validated X-boxes (see **Table R.1**), we failed to identify predicted X-boxes in the promoter regions used to build our TRPV reporters when using three different consensus previously defined (namely, the "relaxed": RYYNYYWWRRNRAC, "refined": GTHNYYAT-RRNAAC and "average": RTHNYYWTRRNRA-C) consensus by (Efimenko et al. 2005)).

Loss or maintenance of reporter gene expression in *daf-19* mutants can be explained at least by two scenarios: first, for each construct, all ciliated sensory neurons are equally dependent on *daf-19* but the penetrance of the phenotype does not achieve 100%. If this is the case, cells expressing the reporter in *daf-19* mutants will be randomly distributed among cell types. On the contrary, it is possible that, for each gene reporter, specific ciliated neuron types are more sensitive to *daf-19* mutation. If this is the case, we expect reporter expression will be maintained in specific subpopulations. To facilitate further identification of ciliated neurons and provide a standardized method of scoring in this work, the ciliated system of the young adult *C. elegans* hermaphrodite was divided into 5 different anatomical

regions, similar to previous analysis of ciliome gene expression (Burghoorn et al. 2012). The results of this detailed scoring analysis are gathered in **Figure R.2** and **Annex 3**.

Accordingly, we first checked if wild type expression for our built reporters matched previously reported expression patterns. As expected, reporters for structural ciliary features showed a broadly distributed expression pattern across the whole ciliated system (see **Figure R.2**). On the contrary, TRPV reporters, as already reported, remained restricted to subsets of ciliated neurons; such was the case of the *ocr-4* reporter, which was exclusively expressed in the OLQ neurons (see **Figure R.2**). Then, we proceed to analyse reporter expression in a null *daf-19* mutant background. *daf-19* mutants are *dauer* constitutive; thus, to facilitate scoring and worm handling, we used the double *daf-19(m86)II*; *daf-12(sa204)X* mutant strain to suppress the *dauer* phenotype. Accordingly, the expression of all 14 cilium-related features was analysed in wild type, *daf-12* and *daf-19*; *daf-12* (for now on abbreviated as *daf-19*) mutant backgrounds.

Quantification of the total number of cells expressing the reporters showed that, unexpectedly, *daf-12* itself had a partial but significant effect on the expression of several structural ciliary reporters (namely, *che-13*, *ift-20*, *mks-1*, *osm-1* and *tmem-107*) as well as upon the reporter for the *ocr-2* TRPV channel (see **Figure R.2).** More precisely, all 6 affected reporters showed expression defects for neurons located in Region 1 (amphid neurons), whereas the reporters for *ift-20*, *osm-1* and *tmem-107* showed additional expression

defects for neurons located in Region 3 (labial neurons). This loss in the number of reporter-expressing cells suggest a possible role for *daf-12* in ciliome expression that, to our knowledge, has not been previously reported.

Regarding the effect of the *daf-19* mutation, we found that expression of the TRPV reporters showed no dependency towards *daf-19*. As explained above, *ocr-2* expression was affected by *daf-12* mutation, causing a loss in its expression in a pair of unidentified amphid neurons and, most probably, in the FLPs neurons based on anatomical position. In contrast, and as expected, all 10 structural ciliary reporters exhibited a severe reduction in the number of reporter-expressing neurons. However, only for 3 of these 10 reporters the expression was completely abolished in this null *daf-19* background: namely, *mks-1*, *osm-5* and *tmem-107* reporters. Of note, these three reporters shared two features: first, they were the shortest among all analysed reporters, spanning about 140 base pairs in length and being generated to cover about 70 base pairs flanking a central X-box motif and, second, even in the wildtype background, although these reporters were sufficient to drive expression in a subpopulation of ciliated neurons, they were expressed in less ciliated neurons when compared to the rest of reporters (see **Figure R.2)**. Altogether, these results suggest that other regions outside these small fragments also act as enhancers for the expression of these ciliary genes in other neurons and that in such small constructs the lack of *daf-19* cannot be compensated**.** In contrast, and despite the significant expression defects found for the rest of the structural ciliary reporters (namely: *che-11*, *che-13*, *ift-20*, *osm-1, peli-1*(long)*, tub-1 and xbx-1*(long)) in the null *daf-19* background, a variable number of neurons were able to maintain reporter expression in all the cases (**Figure R.2**).

These results suggests that, although small constructs baring X-boxes are sufficient to drive ciliated neuron gene expression, these constructs are not panciliary expressed and are totally dependent on *daf-19* while longer constructs show broader panciliary gene expression and are able to partially tolerate *daf-19* absence. To directly test this hypothesis, we selected two of the longer constructs that remained active in more neurons in the *daf-19* null mutant background and built shorter versions of them: namely, *peli-1*(short) and *xbx-1*(short) (**Figure R.1**). Expression analysis of these two short constructs matched the results previously obtained with other short reporters. First, in wildtype worms *peli-1* shorter construct was expressed in less neurons (see **Figure R.3**), suggesting that additional regulatory regions included in the long enhancer are also necessary to achieve panciliary expression. This effect was not so pronounced for *xbx-1*, as the short reporter was expressed in a similar number of neurons when compared to its longer counterpart (see **Figure R.3**). And secondly, the expression of both *peli-1*(short) and *xbx-1*(short) was much more severely affected in *daf-19* null mutants compared to their corresponding longer constructs (see **Figure R.3**). These results agree with our hypothesis that additional factors act with *daf-19* outside the X-box motif to induce ciliome gene expression in the ciliated sensory system.

che-11

ATCTCCATGGCAAC

che-11prom (-1/-208)

che-13

che-13prom (-1/-525)

GTTGCTATAGCAAC

ift-20

ift-20prom (-1/-271)

GTCTCTATAGCAAC

mks-1

mks-1prom (-1/-144)

GTCACCATAGGAAC

osm-1

osm-1prom (-1/-1770)

GCTACCATGGCAAC

osm-5

osm-5prom (-1/-143)

GTTACTATGGCAAC

peli-1

peli-1prom(short)(-1/-225)

peli-1prom(long)(-1/-1128)

GTTCTCTGAAAAC   GTCCTCACAAGTAAC   GTCTCCAATGGCAAC

tmem-107

tmem-107prom(-1/-143)

GTCGCCATGGCAAC

tub-1

tub-1prom(-1/-424)

ATCTCCATGACAAC

xbx-1

xbx-1prom(short) (-1/-105)

xbx-1prom(long) (-1/-370)

GTTTCCATGGTAAC

◄ ⅄ **Figure R.1. Genetic landscapes and reporter structure of analysed selected genes encoding for cilium-related features in *C. elegans*.**
These schematics illustrate the genetic landscape within the 2.000 base pairs regions (thick black line) surrounding the genes that were selected for this analysis. Coloured boxes represent coding exons. Exons for structural cilia genes are depicted in blue, exons for subtype-specific cilia genes appear in orange and non-cilia related genes are shown in grey. Sense of coding sequences is depicted with black arrow heads. Vertical dashed lines mark the genetic region used to build the reporters. Length of the promoters cloned, represented by white boxes, is indicated. Location and sequence of X-box motifs appear as dark blue boxes and text.

In summary, we found that *daf-12* has a small but significant effect in expression of some ciliary features, both structural or subtype-specific. On the other hand, expression of structural ciliary genes cannot be compensated in the absence of *daf-19* if small reporters are analysed but, in general, the lack of *daf-19* can be partially compensated when longer constructs are used. Moreover, this compensatory effect takes place in specific subpopulations of ciliated neurons, that depend on each of the constructs. Thus, with these experiments we demonstrate the existence of *daf-19* co-factors acting outside the X-box binding site.

Generation of a *bona-fide* list of ciliary genes.

Once we determined that additional factors act together with *daf-19* and that expression of structural ciliary genes in some ciliated neurons can be achieved even in the absence of *daf-19*, we next aimed to identify such factors.

One common feature of gene regulatory networks controlling neuron terminal differentiation is that effector genes are co-

regulated by the same TFs; this is most obvious with *daf-19*, that co-regulates expression of ciliary genes through the X-box binding motif. We reasoned that TFs acting as *daf-19* co-factors could also act broadly on cilium-related genes. If this was the case, in addition to the X-box motif, we should find enrichment of other motif/s corresponding to the binding site of such co-factor/s.

▼ **Table R.1. Experimentally validated X-box motifs found in the promoter regions of structural cilium-related genes.**
X-box position indicates 5' edge location of the binding site related to the translation starting codon of the corresponding gene.

Gene	Description	X-box sequence	X-box Position	Reference
che-11	IFT140	ATCTCCATGGCAAC	-86	(Efimenko et al. 2005)
che-13	IFT57	GTTGCTATAGCAAC	-75	(Efimenko et al. 2005)
ift-20	IFT20	GTCTCTATAGCAAC	-61	(Blacque et al. 2005a)
mks-1	MKS-1	GTCACCATAGGAAC	-70	(Efimenko et al. 2005)
osm-1	IFT172	GCTACCATGGCAAC	-86	(Efimenko et al. 2005)
osm-5	IFT88	GTTACTATGGCAAC	-116	(Efimenko et al. 2005)
peli-1	PELI1/2/3	GTCTCCAATGGCAAC	-149	(Chu et al. 2012)
		GTCCTCACAAGTAAC	-199	
		GTTCTCTGAAAAC	-1092	
tmem-107	TMEM107	GTCGCCATGGCAAC	-76	(Lambacher et al. 2016)
tub-1	TUB	ATCTCCATGACAAC	-183	(Efimenko et al. 2005)
xbx-1	DYNC2LI1	GTTTCCATGGTAAC	-78	(Schafer et al. 2003)

➤ **Figure R.2. Analysis of expression from fluorescent reporters of genes coding for cilium-related features of *C. elegans*.**
Top scheme shows a lateral left view of a *C. elegans* young adult hermaphrodite with coloured dots representing the neurons integrating its ciliated system. Non-bilateral AQR neuron locates only in the right side of the animal (R4 region) and hence appear depicted with dashes lines. Scoring of neurons for the five different anatomical regions is portrayed in five different colours. In each graph, coloured bars represent the average reporter-expressing neurons per region in three different genotypes: wildtype worms, *daf-12(sa204)* animals (*daf-12* mut.) and doble *daf-12(sa204)* and *daf-19(m86)* mutants (*daf-12/19* mut.). Error bars represent the standard deviation for the total mean number of neurons. Graphs corresponding to structural cilium reporters are marked with bluish background whereas subtype-specific cilium reporters appear in orange. Statistical significance, depicted by asterisks, was established through a *t* test once the variances of the two samples being compared had been analysed. For each reporter, a Bonferroni correction accounting for the 3 simultaneous hypotheses testing being performed was used. N= 5 to 15 animals per genotype and reporter. See **Annex 3** for detailed values.

Average reporter(+) neurons
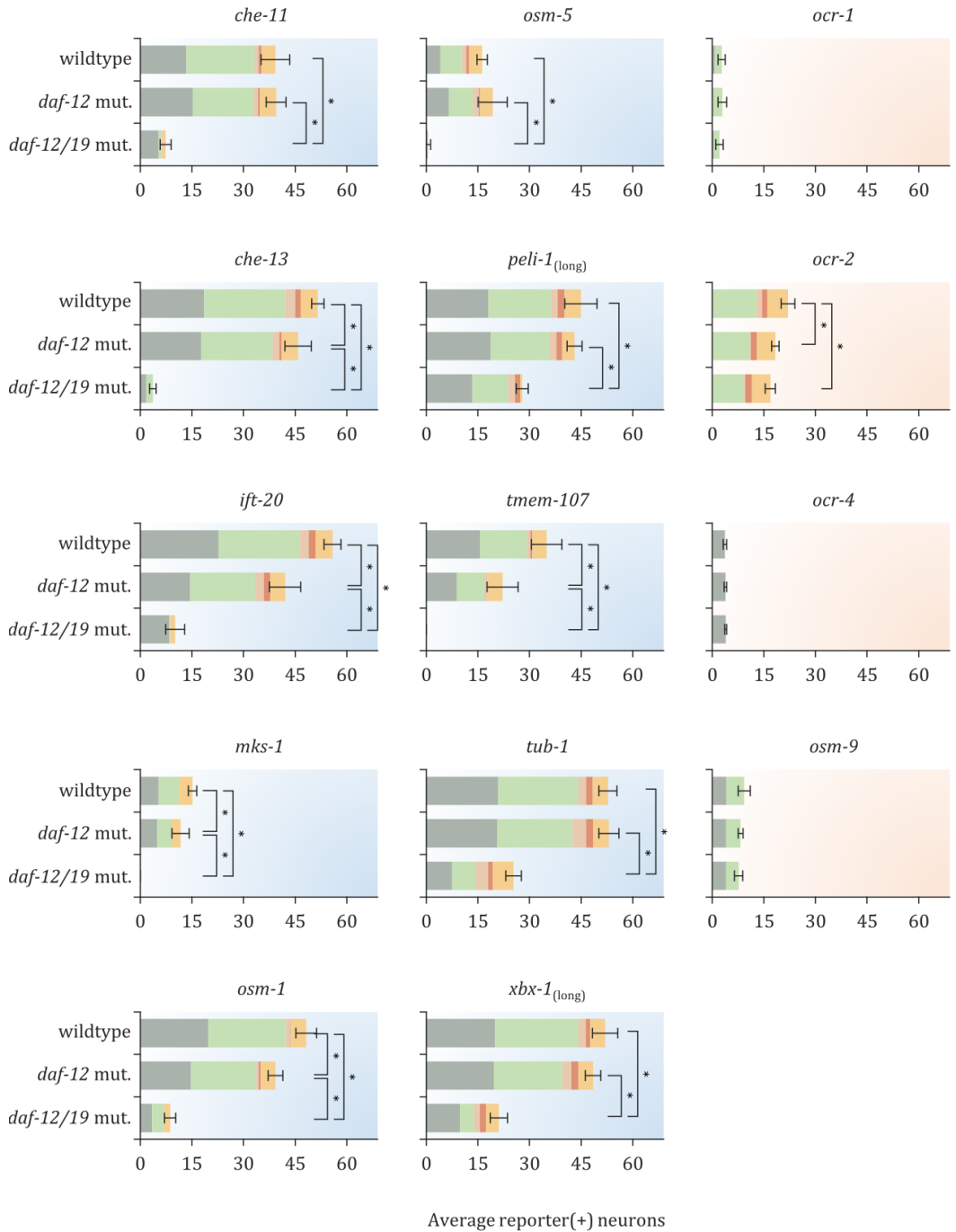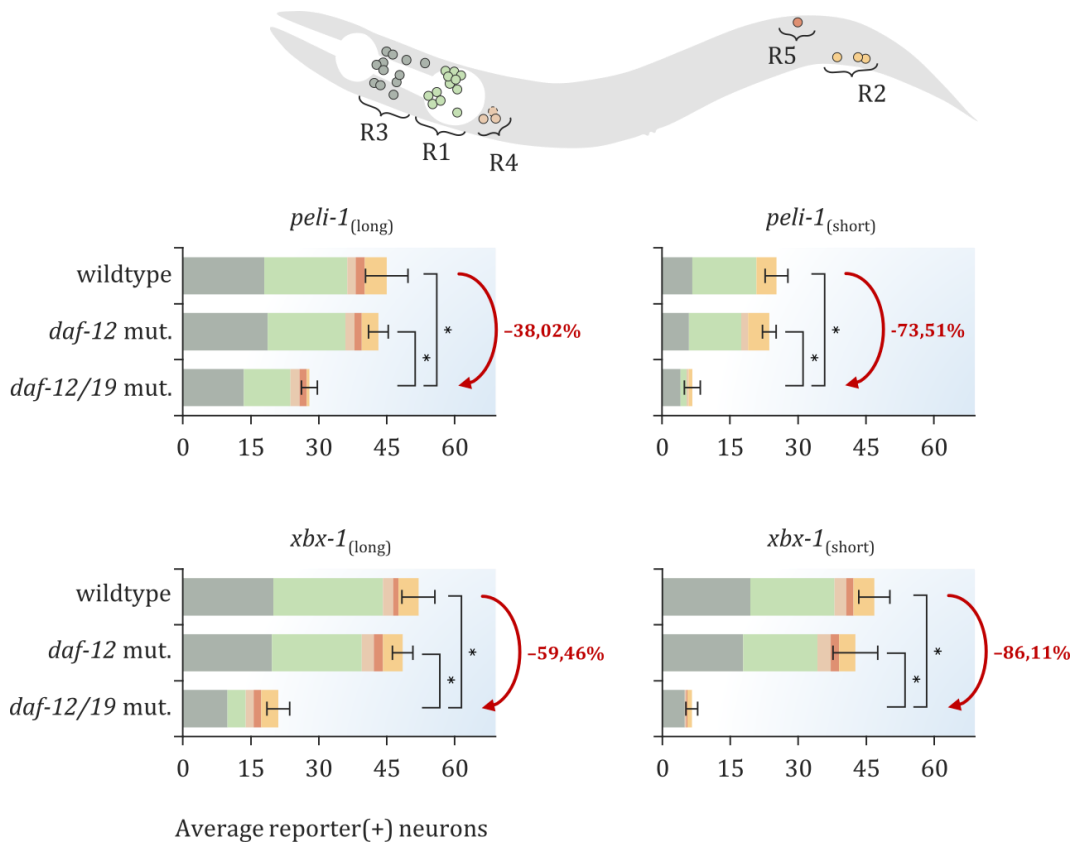
⌃ Figure R.3. Comparison of expression from fluorescent reporters of genes encoding for cilium-related features of *C. elegans*.

Top scheme shows here a lateral left view of a *C. elegans* young adult hermaphrodite with coloured dots representing the neurons integrating its ciliated system. Non-bilateral AQR neuron locates in the right side of the animal and hence the neuron appears depicted through dashes lines. Neurons gathered in the five different anatomical regions in which the ciliated system was divided are portrayed in five different colours. In each graph, coloured bars represent the average reporter-expressing neurons per region in three different genotypes: wildtype worms, *daf-12(sa204)* animals (*daf-12* mut.) and doble *daf-12(sa204)* and *daf-19(m86)* mutants (*daf-12/19* mut.). Error bars represent the standard deviation for the total mean number of neurons. Graphs corresponding to structural cilium reporters are marked with bluish background. Statistical significance, depicted by asterisks, was established through a *t* test once the variances of the two samples being compared had been analysed. For each reporter, a Bonferroni correction accounting for the 3 simultaneous hypotheses testing being performed was used. Red arrows and accompanying values indicate the percentage of neurons in which the expression of the reporter is lost when comparing the corresponding wild type and *daf-12/19* mut. means. N= 5 to 12 animals per genotype and reporter. See **Annex 3** for detailed values.

Thus, the first step was to create a curated list of genes encoding for cilia components that could be potential targets for the unknown *daf-19* co-factor/s. Two primary

sources were initially used to create this list: genes associated with the terms *cilium* or *cilia* within the Cellular Component and/or the Biological Process aspects from the Gene Ontology (time of consulting: December the 13rd, 2016) and genes with known functional X-boxes from (Burghoorn *et al.* 2012). This primary list was further refined through a manual curation process in which some genes for whom expression in ciliated neurons was reported in the WormBase were added. Because the main aim of this list was to collect genes encoding for ciliary components, we focused on effector genes and transcription factors were excluded. A total of 163 genes were included in the final list; for all the genes, expression within the ciliated system, ciliary defects associated to gene mutation or both were assessed through bibliographic research (see **Annex 4**).

This ciliome list was then divided into four ciliary sub-categories:

1) The "*Structural*" category included genes encoding for either intraflagellar transport (IFT) components (such as kinesins or dyneins) or elements related to the assembly of the cilium or IFT complexes (such as the BBSome particles). These genes are basic components present in all types of cilia and thus in all ciliated neurons in *C. elegans*.

2) Genes coding for channels or receptors, such as the TRPV channels, were classified into the "*Subtype-specific*" category. These genes are expressed in a neuron-type specific manner and are not broadly expressed in the sensory ciliated neurons in *C. elegans*.

3) Genes described as broadly expressed within the ciliated system but not associated with a well-defined structural ciliary function were classified within the "*Broad expression*" category.

4) Genes related to male-specific cilia functions were included in the "*Male*" category.

Figure R.4 shows the number of genes from this curated list that were assigned to all 4 different categories.
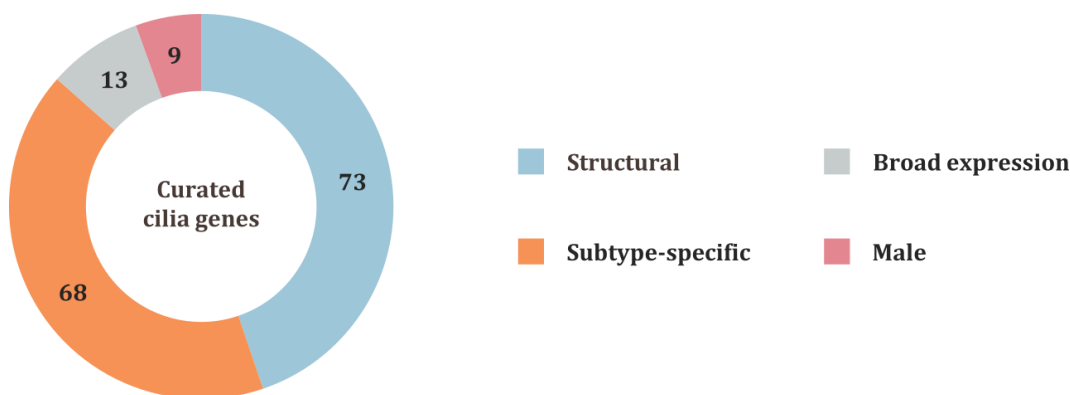
⌃ **Figure R.4. Functional categorization of ciliary genes used in this work.**
Pie chart of the distribution of 163 curated cilium genes classified into four different categories (See main text for explanation of each category).

De novo motif enrichment analysis of ciliary gene regulatory sequences.

Next, to gain some insight into the regulatory logic controlling the co-regulation of the genes in our ciliome list, we decided to follow an unbiased bottom-up approach performing a *de novo* motif discovery analysis using the RSAT *oligo-analysis* tool.

Considering the results that we obtained in our aforementioned *in vivo* reporter expression analysis, we decided to expand this *de novo* motif analysis further away from the location of the X-box motif. Accordingly, we decided to analyse putative regulatory sequences spanning 700 base pairs upstream the translational start site, aiming to achieve a region big enough to include additional motifs and enhancers. When a gene contained different coding isoforms, upstream region associated to each isoform was considered separately. However, as recommended by the RSAT protocol, any overlapping regions existing among different isoforms were conveniently purged to avoid an artificial enriched of particular motifs. Several nucleotide sequences (the so-called *oligos*), ranging from 6 to 8 base pairs in length, were found to be enriched within the set of analysed sequences (see **Table R.2**). Related *oligos* were ultimately clustered into 9 final position weight matrices (PWMs) or motifs whose sequence logos are presented in **Figure R.5**.

Interestingly, some of the motifs generated in this analysis were shown to target specific functional categories in our ciliome list (see **Figure R.6**). Motifs 1 to 6 were statistically enriched within the Structural ciliary genes compared to the Subtype-specific

category. In addition, Motif 1 to 4 were enriched within the Broad Expression category in comparison to the Subtype-specific ciliary genes. Motifs 1, 2 and 4 were also enriched within the Structural ciliary genes when compared against the Male category. Motif 6 was the only motif showing a differential enrichment within the Structural ciliary genes when compared to the genes integrating the Broad expression category. Finally, no enrichment was found in the distribution of Motifs 7 to 9 for any of the four different functional categories employed.

A schematic representing the localisation of all these PWMs within the putative regulatory sequences of each ciliome gene in our list can be seen in **Annex 5**.

RFX/DAF-19 binding sites are enriched in the promoters of structural ciliary genes.

Next, we wanted to assess if the 9 retrieved motifs matched any known experimentally validated transcription factor binding site using the CIS-BP *C. elegans* database. As expected from promoters of cilium-related genes, two of the motifs (Motif 3 & Motif 4) matched known binding sites for *daf-19* (see **Figure R.7**). We found that 85 and 89 out of the 163 ciliome genes in our list had matches for Motif 3 and Motif 4, respectively. In addition to DAF-19 binding sites corresponding to Motif 3 and Motif 4, Motif 7 showed similarity to binding sites for two bHLH transcription factors in *C. elegans*; matching to known *hlh-11* and *lin-32* binding motifs (see **Figure R.7**). The remaining 6 motifs that we retrieved (namely, Motifs 1, 2, 5, 6, 8, 9) did not match any known consensus

binding site for *C. elegans* TFs present in this database.

▼ **Table R.2. Enriched kmers found in putative regulatory regions of the *cilium-gene list*.**

268 putative regulatory sequences of 700 base pairs in length were used to feed an overrepresentation oligonucleotide algorithm (RSAT-oligo-analysis). Several motifs are enriched in different subsets of the analysed sequences. Observed (O) and expected (E) frequencies for each oligo are shown, as well as the binomial probability for its occurrence within the set (P) and the binomial statistical significance ($S \geq 0$) obtained when considering all non-coding sequences of *C. elegans*.

	Oligo	O	E	P	S		Oligo	O	E	P	S
HEXAMERS	CCATGG	34	10.96	1.90E-08	4.39	**OCTAMERS**	ATGGCAAC	20	3.08	1.30E-10	5.38
	CATGGC	44	21.38	1.20E-05	1.60		CATGGCAA	19	2.96	4.60E-10	4.82
	CATGGA	63	37.33	7.90E-05	0.78		CCATGGCA	13	1.27	1.10E-09	4.45
	ATGGAG	78	50.22	1.70E-04	0.45		CCATGACA	15	2.17	1.10E-08	3.43
	GGCAAC	61	37.16	2.10E-04	0.37		CGTTGCCA	13	1.87	9.60E-08	2.50
	AGCTCA	81	53.12	2.20E-04	0.33		ATAGCAAC	16	2.98	1.10E-07	2.43
	GCTCAC	56	34.04	3.40E-04	0.15		CATGGAGA	15	2.72	2.00E-07	2.19
	ATGAGA	137	100.82	3.50E-04	0.13		CATGACAA	20	4.96	3.10E-07	1.99
	ATGCAC	70	45.48	4.40E-04	0.04		GCCATGGA	9	1.20	4.80E-06	0.80
	CATAGC	40	22.28	4.50E-04	0.03		CCATAGCA	9	1.35	1.20E-05	0.39
HEPTAMERS	CCATGGC	19	3.79	2.30E-08	3.73		ACCAGCTG	11	2.15	1.60E-05	0.27
	CGTTGCC	21	5.26	1.90E-07	2.82		ACGTTGTC	11	2.30	3.00E-05	0.01
	GTTGCCA	35	12.99	3.20E-07	2.58						
	CATGGCA	23	7.24	2.30E-06	1.72						
	CCATGGA	20	5.77	3.00E-06	1.61						
	ATGGAGA	40	18.84	1.50E-05	0.92						
	CATGGAG	20	6.92	3.80E-05	0.50						
	CGTTGTC	23	8.96	6.30E-05	0.29						
	CCATGAC	18	6.14	7.60E-05	0.21						

TFs from different organisms that belong to the same family of TFs usually bind similar motifs; thus, to identify the putative transactivators of these additional 6 motifs we decided to compare our PWMs against the JASPAR CORE non-redundant 2018 database that gathers information from many different organisms, including human and mouse. In addition to matches of Motif 3 and Motif 4 to RFX motifs (coincident with CIS-BP retrieval of DAF-19 binding), JASPAR CORE analysis identified Motifs 1 and 2 as consensus binding sites for 4 human RFX members (from RFX2 to RFX5) as well as with mouse Rfx1.

Sequence logos representing these five vertebrate RFX transcription factors binding sites can be seen in **Figure R.8**. Motif comparison for Motif 1 to Motif 4 with their most significant match to a vertebrate RFX matrix is shown in **Figure R.9**.

Motif 1: 74/163 genes

Motif 2: 80/163 genes

Motif 3: 85/163 genes

Motif 4: 89/163 genes

Motif 5: 46/163 genes

Motif 6: 45/163 genes

Motif 7: 45/163 genes

Motif 8: 47/163 genes

Motif 9: 37/163 genes
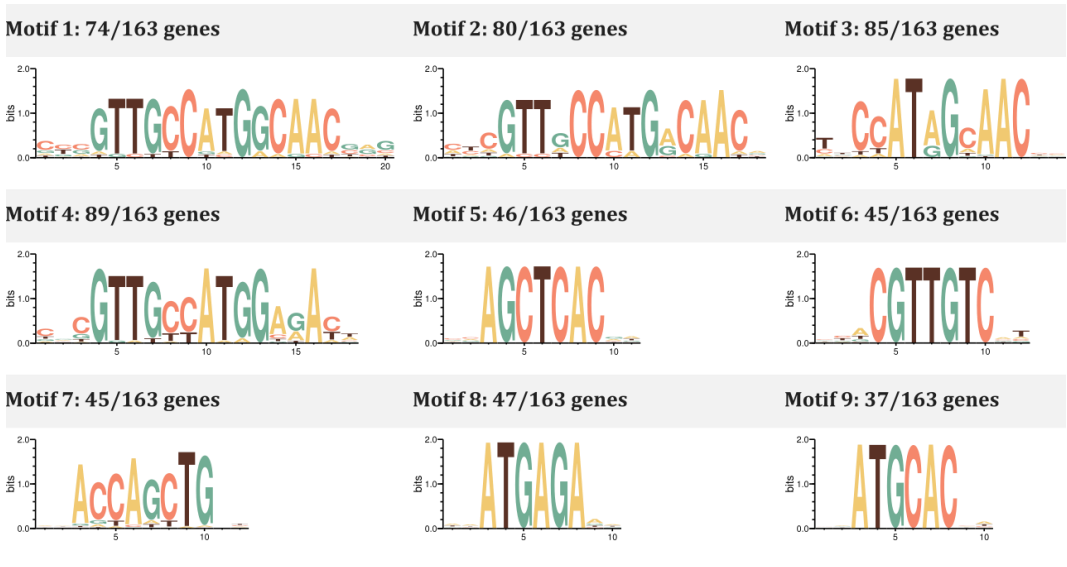
⤒ **Figure R.5. Sequence logos for PWMs generated from enriched oligonucleotides found in putative regulatory sequences of cilium-gene list.**

Enriched kmers are clustered into nine position weight matrices. Sequence logos represent the information content measured in bits. When a given position is defined by a unique nucleotide its information content is 2 bits. Headlines indicate the number of genes in which each motif was found over the total number of genes included in this analysis.

Thus, from the 9 motifs retrieved in this analysis, Motif 1 to Motif 4 corresponded to different matches to RFX binding sites. Specifically, Motif 1 and Motif 2 were very similar and almost palindromic sequences, Motif 4 was a 3' degenerated version of this palindrome whereas Motif 3 represented only one side of it. As a whole, 102 out of the 163 genes from our ciliome list contained at least one RFX binding motif (X-box) within their putative regulatory sequences. Data for all the statistically significant alignments stated in this paragraph are collected in **Annex 6.**

➤ **Figure R.6. Motifs generated from enriched oligonucleotides found in putative regulatory sequences of the cilium gene list are differentially enriched and distribute differentially among different functional categories.**

Ring charts indicate the percentage of total curated cilium genes targeted by a given motif whereas bar charts specify motif distribution per category. Statistical significance of motif distribution was assessed by a two-tailed Fisher test and is indicated through asterisks. The modified Bonferroni correction suggested by Keppel (Keppel 1991) was used to account for all 6 possible pair comparisons in each motif distribution analysis (*: $p \leq 4.42E\text{-}02$).
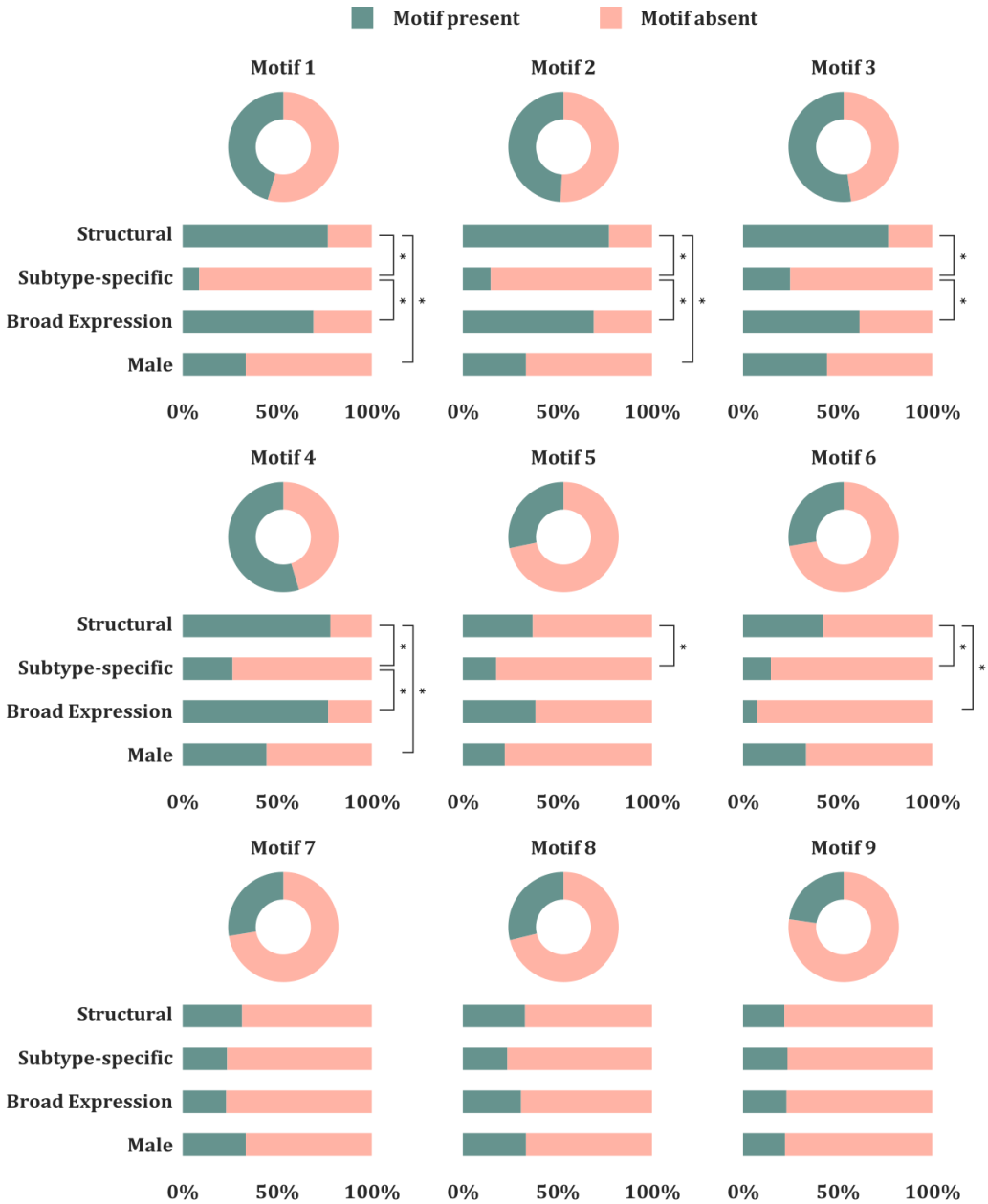
In agreement to published results (Burghoorn et al. 2012; Efimenko et al. 2005; Swoboda et al. 2000b), we found that RFX-related motifs were preferentially associated to *Structural* and *Broadly expressed* ciliome genes when compared against the *Subtype-specific* ciliary category (see **Figure R.10**). No trend was identified for the genes grouped into the *Male* category. Moreover, experimentally validated RFX binding sites are known to exhibit a positional bias, being enriched near the translation start site of their target genes (Efimenko et al. 2005). We wanted to assess if the matches for our RFX-

related motifs also exhibited such positional bias. We found that gene categories with enrichment of X-boxes in their regulatory regions (namely, the *Structural* and *Broad expression* categories) showed enrichment in X-box location near the initiation codon (less than 200 base pairs in distance). This locational bias was not observed in other gene categories that do not show enrichment of X-boxes such as *Function-specific* or *Male* cilia genes (see **Figure R.11**).

Motif 3 vs M1534_1.02 (*daf-19*)

Motif 3

M1534_1.02 (*daf-19*)

Motif 4 vs M1534_1.02 (*daf-19*)

Motif 4

M1534_1.02 (*daf-19*)

Motif 7 vs M0228_1.02 (*hlh-11*)

Motif 7

M0228_1.02 (*hlh-11*)

Motif 7 vs M2264_1.02 (*lin-32*)

Motif 7

M2264_1.02 (*lin-32*)
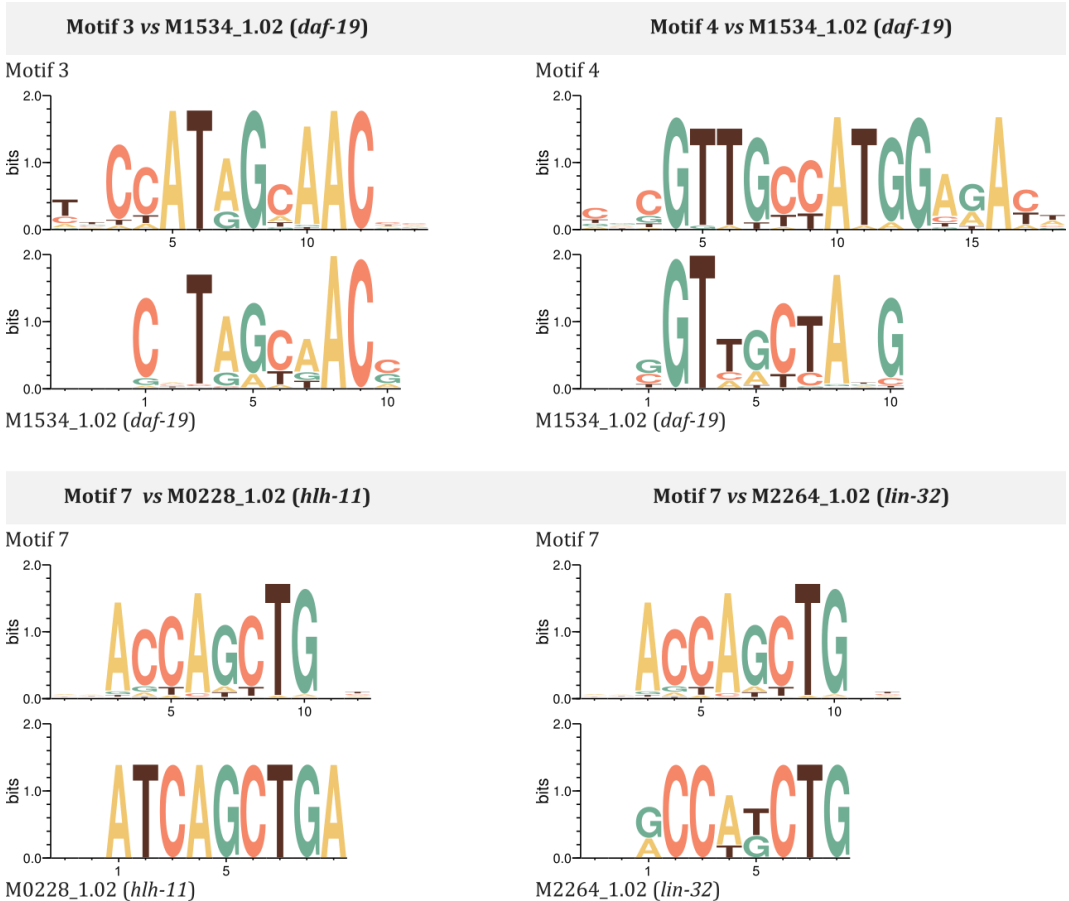
⬆ Figure R.7. Logo comparison between motifs associated with ciliary genes and known binding sites for *C. elegans* transcription factors.
Sequence logos of statistically significant optimal alignments between generated *de novo* motifs from cilia-related genes and binding sites for known transcription factors of the worm.

Thus, 4 of the 9 identified motifs corresponded to previously described X-boxes, preferentially located in structural or broadly expressed ciliary genes with a location bias for the motif close to the ATG. In addition to known RFX binding sites, we only retrieved a putative site for bHLH factors that is not enriched in any ciliary category. The presence of this motif could be explained by the known proneural actions of some bHLH TFs, including *lin-32*, which are not restricted to ciliated neurons.
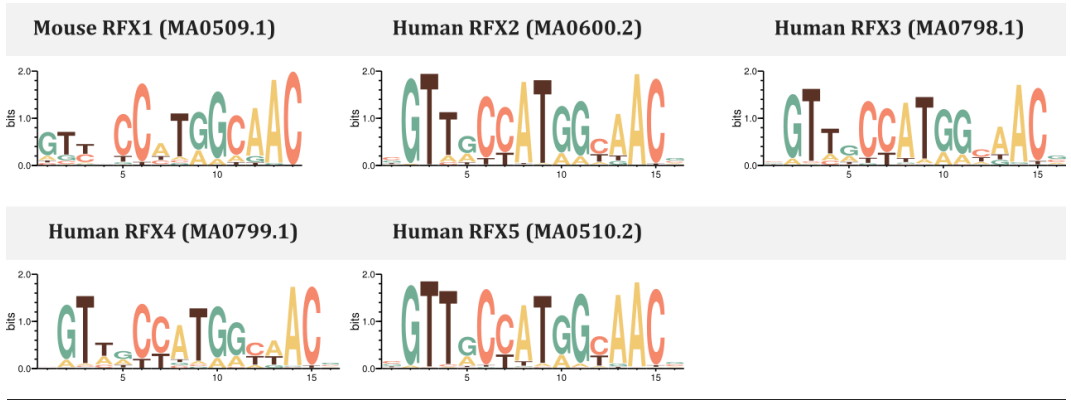
⌅ Figure R.8. Sequence logos for experimentally validated binding sites of vertebrate RFX transcription factors.

The PWM ID is indicated in parenthesis.
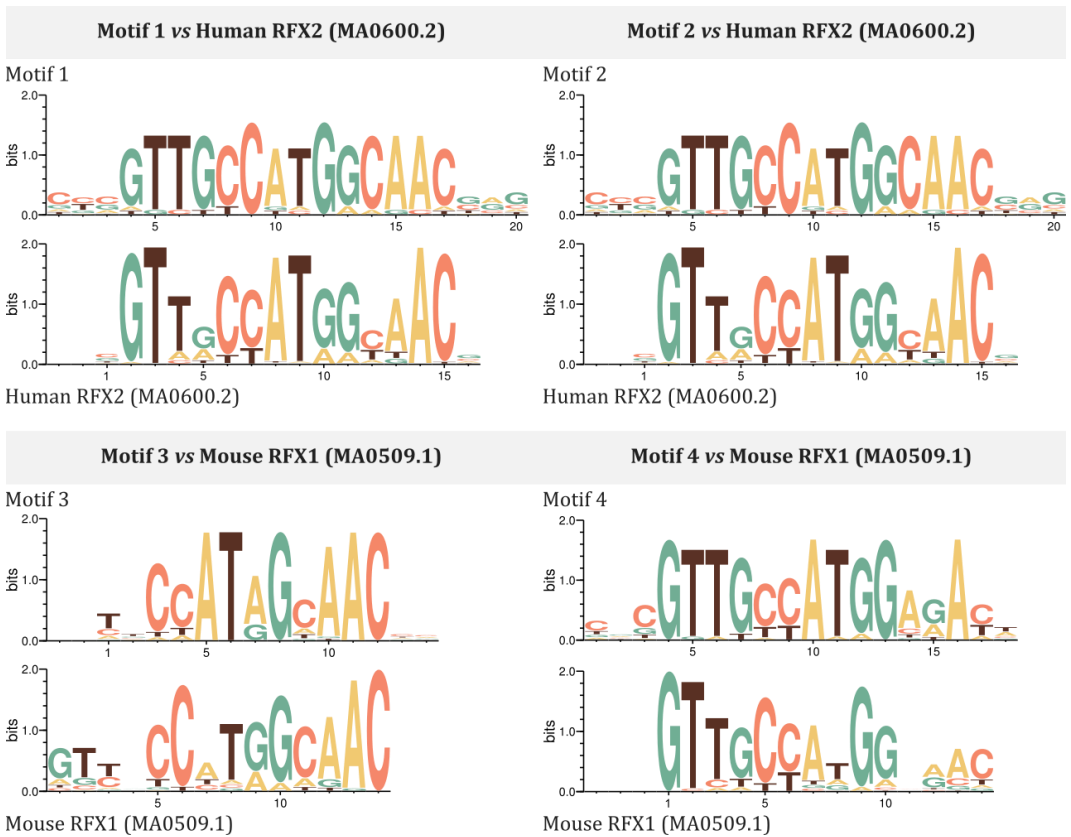
⌅ Figure R.9. Logo comparison between motifs associated with ciliary genes and known binding sites for vertebrate RFX transcription factors.

Sequence logos of statistically significant optimal alignments between generated *de novo* motifs from cilia-related genes and binding sites for RFX vertebrate transcription factors.
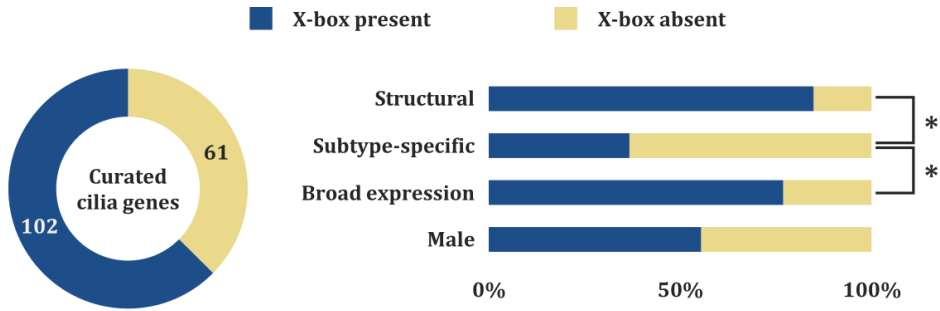
▲ Figure R.10. Analysis of RFX binding motif presence in the different categories of ciliary genes. 102 out of 163 cilia-related genes contain RFX- motif matches within their promoter sequences (-700 base pairs from ATG). Statistically significant enrichment for X-box is found in the structural components of the cilium and broadly expressed cilia genes. These differences were significant with two-tailed Fisher test considering the modified Bonferroni correction suggested by Keppel (Keppel 1991) accounting for all 6 possible pair comparisons. "Structural" vs "Subtype-specific" (p = 4,38E-09) and "Subtype-specific" vs "Broad expression" (p = 1.26E-02).
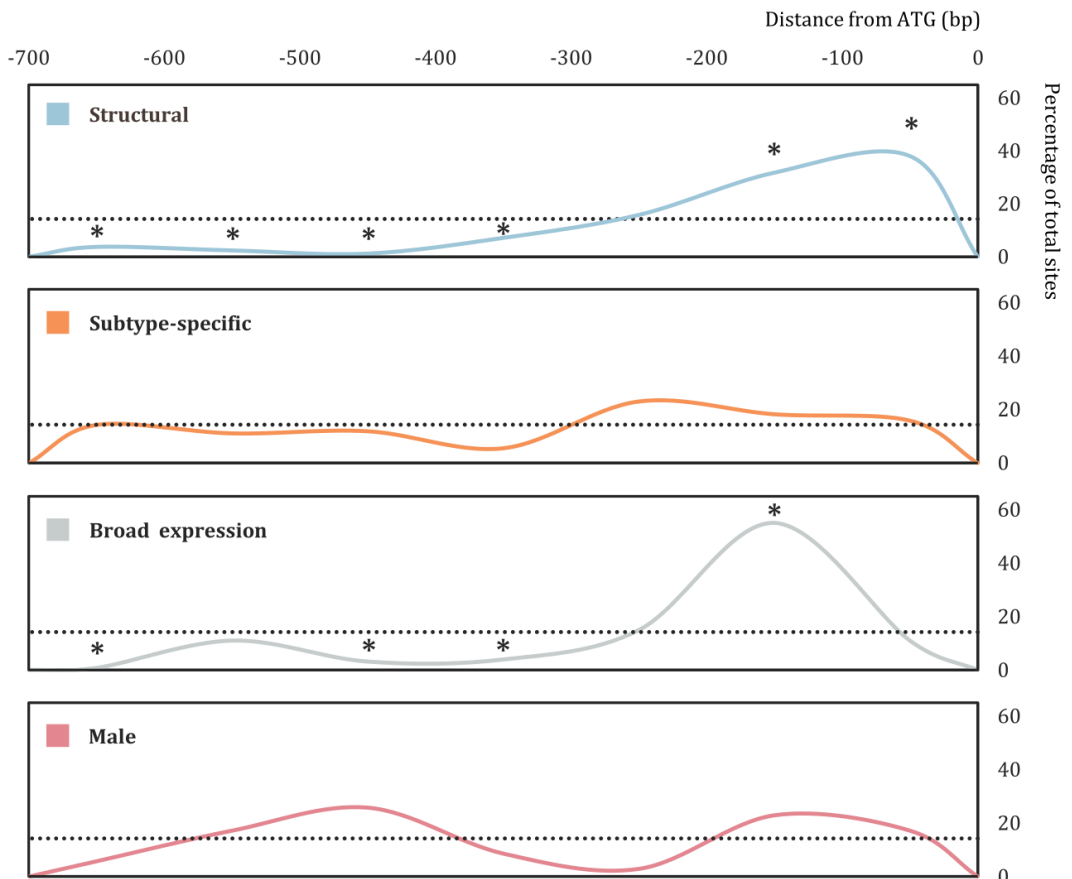
◄ **Figure R.11. RFX-related motifs exhibit a positional bias.**
Putative regulatory sequences of cilia-related genes were divided into 7 regions of 100 base pairs each. For each gene region, the percentage of RFX-related sites was calculated. Dotted lines represent the expected proportion of sites from a corresponding equally-sized distribution with no positional bias for site occurrence. Asterisks indicate the ranks in which observed site frequency significantly differs from the expected by the unbiased distribution as assessed through a two-proportions Z-test. Both structural cilia genes and broadly expressed cilia genes show a bias in RFX location near the ATG (<200bp), while RFX motifs found in the other categories are not preferentially distributed in any specific region.

The other motifs (namely, Motifs 5, 6, 8 and 9), which corresponded to non-degenerated sequences, did not match any known binding site present in either of the two databases used in this analysis. In addition, the percentage of appearance for these motifs within the whole set of curated ciliome genes was relatively low, ranging from less than 23% for Motif 9 to less than 29% for Motif 8, hence out of the scope for a good candidate TF broadly related to ciliome gene expression.

Additional TF binding motifs are enriched in structural ciliary genes with RFX binding motifs.

Our initial approach analysing the list of 163 curated ciliome effector genes (categorised as *Structural* (73), *Subtype-specific* (68), *Broad expression* (13) and *Male* (9)), failed to retrieve additional motifs other than RFX binding sites or general bHLH proneural factors. Thus, as an alternative to find motifs for *daf-19* co-factors, we restricted our oligonucleotide enrichment analysis to only those 102 genes in which RFX binding sites were present.

From the clustering of the enriched oligonucleotides found in the 173 putative regulatory sequences used in this analysis (see

Table R.3) we could retrieve 10 new different motifs that appear gathered in **Figure R.12**. Alignment of those new motifs against the matrices present in the same two databases previously used allowed us to relate 6 of them to experimentally validated RFX binding sites. Specifically, new Motifs 1 to 6 matched the same mouse and human RFX transcription factors than previous Motifs 1 to 4 and new Motifs 3 and 4 both matched the matrix provided for the *daf-19* binding site (see **Table R.4**). None of the remaining 4 new motifs were able to pass all three thresholds used to assess for the statistical significance of the alignments. On this occasion, two non-degenerated motifs were retrieved. New Motif 7 corresponded to the same PWM obtained for previous Motif 5 and the novel new Motif 8 was found to be an artifact created by the concatenated repetition of 16 of those motifs in the sequence of the putative *cpaf-36* promoter in its isoform a. Thus, with this new analysis, we retrieved two new degenerated assemblies: Motif 9 and Motif 10 that remained unidentified. Since the absence of matches could be simply due to the limited number of experimentally validated binding sites known for *C. elegans* (at the time of writing, MEME suite makes use of the *C. elegans* CIS-BP 1.02 database from which only 269 transcription factors are being considered,

hence accounting for roughly a 30% of the total), we decided to further analyse the results of their alignments in this database. Although the matches for those two assemblies failed to pass any multiple comparisons correction, the statistically significant alignment that were retrieved seemed to target specific families of transcription factors (see **Table R.5**). Those were the cases of new Motif 9, targeting members of the Forkhead family, and new Motif 10, matching to members of the Nuclear Hormone Receptor family. When we expanded this analysis of suboptimal alignments by employing the matrices included in the JASPAR CORE non-redundant 2018 database, we were unable to retrieve a particular targeted family as a match for Motif 10. However, new Motif 9 matched recognition sites for members of the FOX/Forkhead family as well as the Homeodomain family of transcription factors (see **Table R.6**).

As a whole, through this unbiased bottom-up approach in which putative regulatory sequences of curated ciliome genes were used to perform a *de novo* motif discovery analysis, we were able to retrieve the expected RFX-related motifs as well as matches for bHLH, including the proneural factor *lin-32*. On the other hand, suboptimal alignment seemed to target members of the Forkhead, Homeodomain and Nuclear Hormone Receptor families of TFs.

A schematic representing where these new 10 motifs localise within the putative regulatory sequences of all known isoforms of the 102 curated ciliome genes also harbouring X-box motifs that were used in this section can be seen in **Annex 7**.

▼ Table R.3. Enriched *kmers* found in putative regulatory sequences of ciliary genes also harbouring RFX-related motifs.

173 putative regulatory sequences of 700 base pairs in length were used to feed an overrepresentation oligonucleotide algorithm (RSAT-oligo-analysis). Observed (O) and expected (E) frequencies for each oligo are shown, as well as the binomial probability for its occurrence within the set (P) and the binomial statistical significance ($S \geq 0$) obtained when considering all non-coding sequences of *C. elegans*.

	Oligo	O	E	P	S		Oligo	O	E	P	S
HEXAMERS	CCATGG	31	7.19	4.3E-11	7.05	**HEPTAMERS**	GTTGCCA	32	8.52	6.0E-10	5.31
	CATGGC	37	14.03	2.5E-07	3.28		CATGGCA	23	4.75	1.5E-09	4.91
	GGCAAC	51	24.38	1.7E-06	2.46		CGTTGCC	19	3.45	5.3E-09	4.36
	CATGGA	51	24.49	1.9E-06	2.40		CCATGGC	16	2.49	9.9E-09	4.09
	CCATAG	35	15.54	1.5E-05	1.51		CCATGGA	19	3.79	2.2E-08	3.74
	CATAGC	33	14.61	2.4E-05	1.29		GTTGCTA	22	6.32	9.2E-07	2.12
	ATGGAG	59	32.94	2.7E-05	1.25		CCATGAC	17	4.03	1.2E-06	1.99
	TCTCCA	78	47.57	3.2E-05	1.18		CATGGAG	18	4.54	1.5E-06	1.92
	GCTCAC	44	22.33	3.3E-05	1.17		CCATAGC	13	2.63	4.1E-06	1.48
	ATGCCG	33	14.94	3.7E-05	1.11		ATGGCAA	30	11.62	4.8E-06	1.40
	AGCTCA	61	34.85	3.8E-05	1.10		ATGGAGA	31	12.36	6.0E-06	1.31
	CTCCAC	54	29.83	4.4E-05	1.04		CGTTGTC	19	5.88	1.3E-05	0.96
	GGAGAC	40	21.60	2.5E-04	0.28		CATAGCA	17	4.88	1.5E-05	0.92
	GACAAC	56	33.99	3.3E-04	0.16		CCATTTC	39	19.57	7.0E-05	0.24
	ATAGCA	49	28.66	3.4E-04	0.15						
	GAGAGC	45	26.06	4.7E-04	0.01						
OCTAMERS	ATGGCAAC	20	2.02	7.5E-14	8.61		ACGTTGTC	11	1.51	5.8E-07	1.72
	CATGGCAA	19	1.94	4.0E-13	7.88		ATGGAAAC	19	4.95	1.2E-06	1.41
	CCATGGCA	13	0.83	6.7E-12	6.66		ATGGTAAC	11	1.63	1.2E-06	1.39
	CCATGACA	15	1.42	4.0E-11	5.88		GCCATGGA	8	0.79	1.8E-06	1.23
	ATAGCAAC	16	1.95	3.4E-10	4.95		ATAGAAAC	20	5.74	2.7E-06	1.05
	CGTTGCCA	13	1.23	7.3E-10	4.62		CATAGCAA	11	1.98	7.6E-06	0.60
	CATGGAGA	13	1.78	5.7E-08	2.73		GGTTGCCA	9	1.35	1.2E-05	0.40
	CATGACAA	17	3.26	6.9E-08	2.65		GGTTTCCA	13	2.98	1.5E-05	0.31
	ATGACAAC	16	3.17	2.5E-07	2.08		GGTTACCA	8	1.09	1.9E-05	0.21
	CCATAGCA	9	0.89	4.2E-07	1.86						
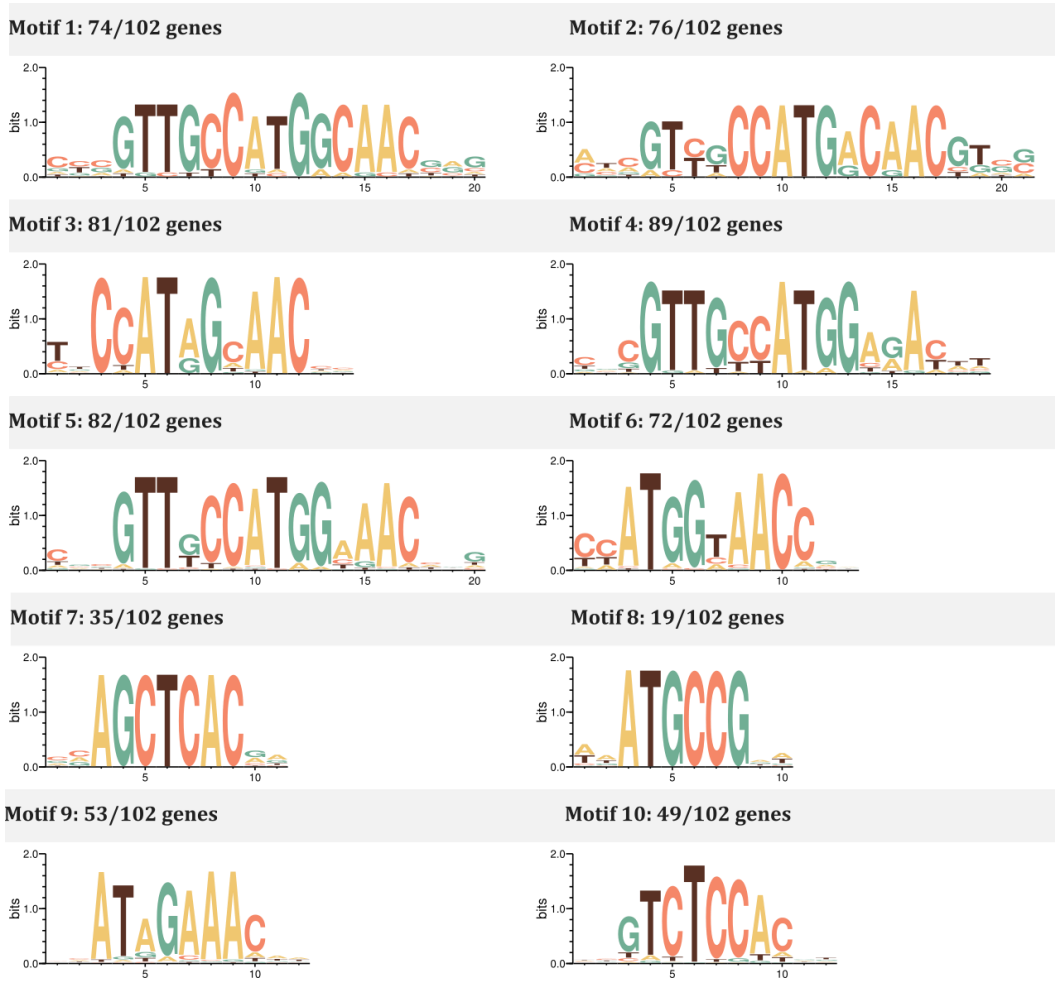
⬆ Figure R.12. Sequence logos for PWMs generated from enriched kmers found in putative regulatory sequences of genes in the cilia list with RFX-related sites.

Enriched kmers are clustered into ten position weight matrices. Sequence logos represent the information content measured in bits. When a given position is defined by a unique nucleotide its information content is 2 bits. Headlines indicate the number of genes in which each motif was found over the total number of genes included in this analysis.

➤ Table R.4. TF binding site matches for over-represented motifs in the set of cilium genes harbouring RFX-related sites.

6 out of the 10 motifs generated in this analysis matched experimentally validated binding sites of worm, mouse and human RFX transcription factors (Motifs 1 to 6). To reduce the number of false positives, three different statistical thresholds were used. p-value indicates the probability for a random motif of equivalent length to have a match score as good or better than the query motif. q-value implements a false discovery rate method correcting for the number of all the simultaneous significant matches for each query motif. E-value further corrects for multiple comparisons now accounting for all the motifs present in the target database being used.

Query	Target DB	Target ID	Gene	Organism	p-value	q-value	E-value
Motif 1	JASPAR	MA0600.2	RFX2	*H. sapiens*	3.63E-12	5.09E-09	5.10E-09
Motif 1	JASPAR	MA0510.2	RFX5	*H. sapiens*	2.05E-11	1.43E-08	2.87E-08
Motif 1	JASPAR	MA0798.1	RFX3	*H. sapiens*	3.99E-11	1.86E-08	5.60E-08
Motif 1	JASPAR	MA0799.1	RFX4	*H. sapiens*	1.31E-10	4.60E-08	1.84E-07
Motif 1	JASPAR	MA0509.1	Rfx1	*M. musculus*	5.08E-10	1.42E-07	7.13E-07
Motif 2	JASPAR	MA0600.2	RFX2	*H. sapiens*	8.86E-09	1.18E-05	1.24E-05
Motif 2	JASPAR	MA0799.1	RFX4	*H. sapiens*	1.59E-08	1.18E-05	2.24E-05
Motif 2	JASPAR	MA0509.1	Rfx1	*M. musculus*	1.70E-08	1.18E-05	2.39E-05
Motif 2	JASPAR	MA0798.1	RFX3	*H. sapiens*	2.24E-08	1.18E-05	3.15E-05
Motif 2	JASPAR	MA0510.2	RFX5	*H. sapiens*	3.50E-08	1.37E-05	4.91E-05
Motif 3	CIS-BP	M1534_1.02	*daf-19*	*C. elegans*	4.50E-06	2.42E-03	1.21E-03
Motif 3	JASPAR	MA0509.1	Rfx1	*M. musculus*	9.89E-06	1.29E-02	1.39E-02
Motif 3	JASPAR	MA0510.2	RFX5	*H. sapiens*	1.57E-05	1.29E-02	2.20E-02
Motif 3	JASPAR	MA0798.1	RFX3	*H. sapiens*	2.07E-05	1.29E-02	2.91E-02
Motif 3	JASPAR	MA0600.2	RFX2	*H. sapiens*	2.44E-05	1.29E-02	3.43E-02
Motif 4	CIS-BP	M1534_1.02	*daf-19*	*C. elegans*	6.70E-05	3.61E-02	1.80E-02
Motif 4	JASPAR	MA0509.1	Rfx1	*M. musculus*	3.31E-08	9.31E-05	4.65E-05
Motif 4	JASPAR	MA0600.2	RFX2	*H. sapiens*	7.40E-06	3.35E-03	1.04E-02
Motif 4	JASPAR	MA0798.1	RFX3	*H. sapiens*	7.40E-06	3.35E-03	1.04E-02
Motif 4	JASPAR	MA0799.1	RFX4	*H. sapiens*	8.59E-06	3.35E-03	1.21E-02
Motif 4	JASPAR	MA0510.2	RFX5	*H. sapiens*	1.07E-05	3.35E-03	1.51E-02
Motif 5	JASPAR	MA0509.1	Rfx1	*M. musculus*	2.56E-08	5.05E-05	3.59E-05
Motif 5	JASPAR	MA0510.2	RFX5	*H. sapiens*	5.19E-07	3.75E-04	7.29E-04
Motif 5	JASPAR	MA0798.1	RFX3	*H. sapiens*	6.19E-07	3.75E-04	8.69E-04
Motif 5	JASPAR	MA0600.2	RFX2	*H. sapiens*	8.02E-07	3.75E-04	1.13E-03
Motif 5	JASPAR	MA0799.1	RFX4	*H. sapiens*	2.20E-06	6.86E-04	3.09E-03
Motif 6	JASPAR	MA0510.2	RFX5	*H. sapiens*	2.75E-05	3.19E-02	3.87E-02

▼ Table R.5. Suboptimal alignments for PWMs generated from enriched k-mers found in putative regulatory sequences of curated cilia genes harbouring RFX-related sites target specific families of *C. elegans* transcription factors.

TF binding site predictions for two out of the ten motifs generated in this analysis surpass the first threshold delimited by the p-value but failed to pass a correction for multiple comparisons when aligned against the motifs present into the *C. elegans* CIS-BP 1.02 database. Matches seem to target specific families of transcription factors. In this table, p-value indicates the probability for a random motif of equivalent length to have a match score as good or better than the query motif, q-value implements a false discovery rate method correcting for the number of the simultaneous significant matches for each query motif and E-value further corrects for multiple comparisons now accounting for all the motifs present in the target database being used.

Query	Target ID	Gene	Family	p-value	q-value	E-value
Motif 9	M0382_1.02	*die-1*	C2H2 ZF	2.58E-02	1.00E+00	6.94E+00
Motif 9	M0622_1.02	*dmd-4*	DM	1.51E-03	8.13E-01	4.06E-01
Motif 9	M0624_1.02	*dmd-10*	DM	8.68E-03	8.21E-01	2.33E+00
Motif 9	M0643_1.02	*dmd-3*	DM	1.24E-02	8.21E-01	3.34E+00
Motif 9	M0639_1.02	*dmd-5*	DM	3.41E-02	1.00E+00	9.18E+00
Motif 9	M0741_1.02	*fkh-9*	Forkhead	7.64E-03	8.21E-01	2.05E+00
Motif 9	M6237_1.02	*unc-130*	Forkhead	8.35E-03	8.21E-01	2.25E+00
Motif 9	M0748_1.02	*daf-16*	Forkhead	1.44E-02	8.21E-01	3.88E+00
Motif 9	M0742_1.02	*lin-31*	Forkhead	1.52E-02	8.21E-01	4.08E+00
Motif 9	M0739_1.02	*let-381*	Forkhead	1.53E-02	8.21E-01	4.11E+00
Motif 9	M6235_1.02	*fkh-6*	Forkhead	2.39E-02	1.00E+00	6.44E+00
Motif 9	M4721_1.02	*php-3*	Homeodomain	1.33E-02	8.21E-01	3.58E+00
Motif 9	M1534_1.02	*daf-19*	RFX	1.20E-02	8.21E-01	3.23E+00
Motif 10	M0276_1.02	*crh-2*	bZIP	3.49E-02	1.00E+00	9.40E+00
Motif 10	M0457_1.02	*bnc-1*	C2H2 ZF	1.50E-02	1.00E+00	4.04E+00
Motif 10	M0644_1.02	*mab-3*	DM	2.32E-02	1.00E+00	6.24E+00
Motif 10	M0742_1.02	*lin-31*	Forkhead	2.86E-02	1.00E+00	7.68E+00
Motif 10	M1486_1.02	*nhr-86*	Nuclear receptor	3.41E-03	1.00E+00	9.17E-01
Motif 10	M1482_1.02	*nhr-79*	Nuclear receptor	1.13E-02	1.00E+00	3.04E+00
Motif 10	M1487_1.02	*nhr-122*	Nuclear receptor	2.61E-02	1.00E+00	7.03E+00
Motif 10	M1463_1.02	*nhr-182*	Nuclear receptor	3.30E-02	1.00E+00	8.88E+00
Motif 10	M1424_1.02	*nhr-154*	Nuclear receptor	3.49E-02	1.00E+00	9.40E+00
Motif 10	M1498_1.02	*egl-38*	Paired box	9.13E-03	1.00E+00	2.46E+00
Motif 10	M1636_1.02	*tbx-33*	T-box	4.45E-03	1.00E+00	1.20E+00

▼ **Table R.6. Suboptimal alignments for PWMs generated from enriched k-mers found in putative regulatory sequences of curated cilia genes harbouring RFX-related sites target specific families of eukaryotic transcription factors.**

Similar to CIS-BP analysis, two out of the ten motifs generated in this analysis failed to pass a correction for multiple comparisons when aligned against the motifs present into the JASPAR CORE non-redundant 2018 database. Motif 9 matches able to surpass the first threshold delimited by the p-value seem to target specific families of transcription factors that are partially coincident with CIS-BP analysis. In this table, p-value indicates the probability for a random motif of equivalent length to have a match score as good or better than the query motif, q-value implements a false discovery rate method correcting for the number of the simultaneous significant matches for each query motif and E-value further corrects for multiple comparisons now accounting for all the motifs present in the target database being used.

Query_ID	Target_ID	Gene	Family	Organism	p-value	q-value	E-value
Motif 9	MA0610.1	DMRT3	DMRT	*H. sapiens*	4.09E-03	9.99E-01	5.74E+00
Motif 9	MA1103.1	FOXK2	FOX	*H. sapiens*	2.21E-03	9.99E-01	3.10E+00
Motif 9	MA0148.3	FOXA1	FOX	*H. sapiens*	3.17E-03	9.99E-01	4.46E+00
Motif 9	MA0846.1	FOXC2	FOX	*H. sapiens*	5.31E-03	9.99E-01	7.46E+00
Motif 9	MA0032.2	FOXC1	FOX	*H. sapiens*	6.49E-03	1.00E+00	9.12E+00
Motif 9	MA0047.2	Foxa2	FOX	*M. musculus*	6.88E-03	1.00E+00	9.66E+00
Motif 9	MA0216.2	cad	HD	*D. melanogaster*	2.72E-03	9.99E-01	3.81E+00
Motif 9	MA0909.1	HOXD13	HD	*H. sapiens*	3.03E-03	9.99E-01	4.26E+00
Motif 9	MA0650.1	HOXA13	HD	*H. sapiens*	3.75E-03	9.99E-01	5.26E+00
Motif 9	MA0899.1	HOXA10	HD	*H. sapiens*	5.00E-03	9.99E-01	7.02E+00
Motif 9	MA0901.1	HOXB13	HD	*H. sapiens*	5.28E-03	9.99E-01	7.41E+00
Motif 9	MA0606.1	NFAT5	NFAT	*G. gallus*	6.86E-03	1.00E+00	9.64E+00
Motif 9	MA0509.1	Rfx1	RFX	*M. musculus*	1.67E-03	9.99E-01	2.35E+00
Motif 9	MA0378.1	SFP1	ZF - C2H2	*S. cerevisiae*	1.42E-03	9.99E-01	2.00E+00
Motif 10	MA0479.1	FOXH1	FOX	*H. sapiens*	5.56E-03	1.00E+00	7.80E+00
Motif 10	MA0795.1	SMAD3	SMAD	*H. sapiens*	1.47E-03	1.00E+00	2.06E+00
Motif 10	MA0130.1	ZNF354C	ZF - C2H2	*H. sapiens*	1.27E-03	1.00E+00	1.78E+00

2. FKH-8 is expressed in all sensory ciliated neurons of *C. elegans*.

As previously stated, our motif enrichment analysis modestly suggested that HD, FKH, NHRs or bHLH TFs could be involved in the direct activation of ciliary components. However, according to the CIS-BP 1.02 data base, the *C. elegans* genome encodes for 101 HD TFs, 16 FKH TFs, 271 NHR TFs and 40 bHLH TFs and our regulatory analysis did not point to any specific candidate. Thus, we decided to follow an alternative approach. Ciliated neurons express a specific isoform of *daf-19* (DAF-19C) that acts as a terminal selector of the ciliome. This isoform has been reported to be specifically expressed in ciliated neurons (Senti and Swoboda 2008); thus, we hypothesized that maybe *daf-19* co-operator's expression could also be enriched in ciliated neurons compared to other neuron types.

Ten transcription factors are highly and specifically enriched within the ciliated sensory neurons of *C. elegans*.

Available data from single-cell combinatorial indexing RNA sequencing (sci-RNA-seq) published in (J. Cao et al. 2017) allowed us to retrieve expression pattern information for 861 out of the 875 *C. elegans* transcription factors encoded in the *C. elegans* genome (Narasimhan *et al.* 2015). This single-cell data, which is publicly available (see **Annex 2**), is easily searchable thanks to the online GExplore$_{1.4}$ platform. Using the most stringent conditions, we found 10 transcription factors that were highly and specifically enriched within the ciliated sensory neurons category compared to other neuron types or non-neuronal tissues (see **Figure R.13**). These ten candidates belonged to only four different transcription factors families, namely: one Zinc Finger (ZF - C2H2) (*Y22D7AL.16*), one Forkhead (FKH) (*fkh-8*), seven nuclear hormone receptors (NHR) (*nhr-158*, *nhr-188*, *nhr-124*, *nhr-30*, *nhr-38*, *nhr-216*, *nhr-277*) and one homeodomain (HD) (*ceh-57*). Of note, binding motifs for all of those families were among the ones being targeted in our analysis for putative regulatory sequences of ciliary genes harbouring X-box motifs (see **Table R.5 and Table R.6**).

TFs acting as broad co-operators of *daf-19* need to meet two requirements: first, their expression must be enriched within the ciliated sensory neurons category, something all these ten TFs fulfil, and second, similar to the *daf-19c* isoform, all or most ciliated neurons have to express such TF. Since very little was known about most of the ten TFs we retrieved in this analysis, we decided to increase the resolution of our sci-RNA-seq analysis. Thus, we opted for visualizing TFs' expression over a bidimensional cell map of the worm in order to identify the specific neuron types that were expressing each TF. To that aim, we used SCope, a tool that allows for the visualization of data sets from large-scale single-cell RNA sequencing. SCope applies the visualization through the so-called t-SNE coordinates. Briefly, this technique allows for the modelling of each cell as a two- or three-dimensional point so cells that have

similar expression profiles are close together in the resulting graph. Thus, specific cell types appear as clusters of dots that can be identified by using the expression of known markers. Applying this technique over the sci-RNA-seq data, 16 different bidimensional neuronal clusters of sensory ciliated neurons were identified (see **Figure R.14**). Of those 16 clusters, 12 were shown to gather 15 different sensory neuronal classes (namely:

ASE, AQR, PQR, CEP, ADE, BAG, ASK, ASJ, ASI, AWA, AFD, AWB, AWC and ASG). The 4 additional clusters collected 4 different sensory neuronal classes (namely: ADF, ASH, PHA and PHB) in which the specific neuronal class per cluster could not be assessed. A cluster for the ADF neurons is suggested in this work based on the exclusive expression of the *srh-142* gene (see **Figure R.14**).
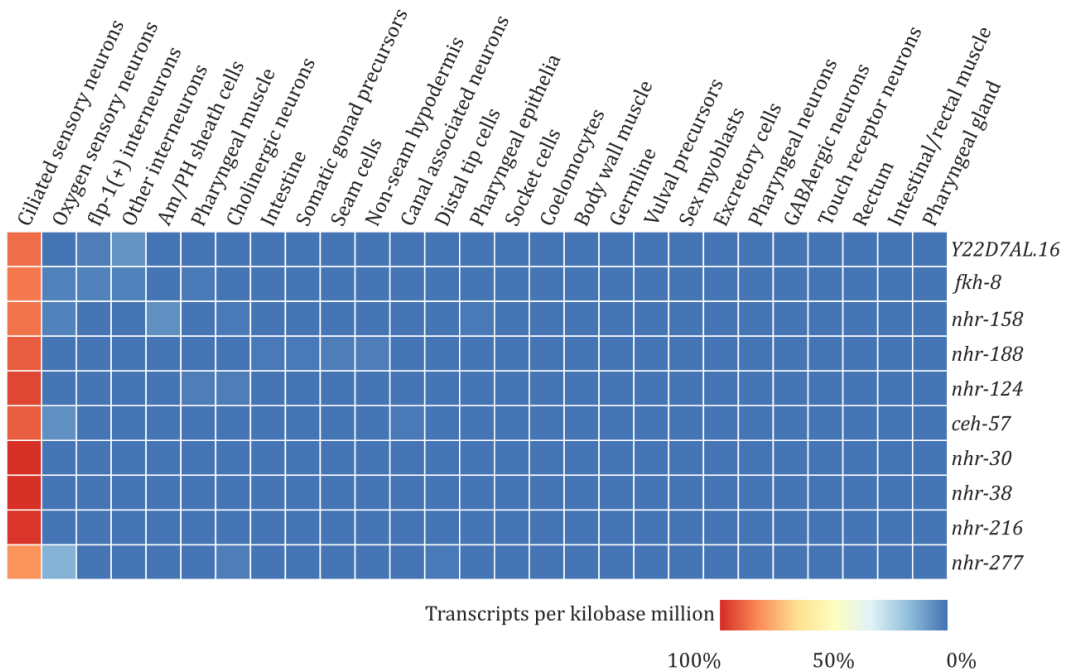
▲ **Figure R.13. sci-RNA-seq data analysis shows the expression of ten transcription factors is specifically enriched in ciliated sensory neurons in *C. elegans*.**
Expression levels, measured as transcript per kilobase million (TPM), are represented with a coloured heatmap. Highest percentage of expression is depicted in red whereas absence of transcript reads is indicated in blue. sci-RNA-seq data obtained from L2 worms' single cells were distributed in 27 different tissue categories (J. Cao et al. 2017). Ten transcription factors were retrieved from the whole dataset by setting a 5-fold enrichment ratio between ciliated sensory neurons category and the second most enriched tissue category, with a rate of false discoveries of 0.001.

We next mapped the expression from each of the 10 candidate TFs in the t-SNE plot to identify which clusters expressed each TF

(see **Figure R.15**). Four TFs showed expression in more than half of the sensory ciliated neuronal clusters. *nhr-277* was the broadest

expressed TF, as 12 out of the 16 possible ciliary clusters were positive for its expression (12/16). Next in the rank, *fkh-8* showed expression in 11 out of 16 clusters (11/16), followed by *nhr-158* (10/16) and *ceh-57* (9/16).

Finally, expression of the 6 additional TFs was detected in less than half of the clusters, namely: *nhr-216* (7/16), *nhr-124* (6/16), *Y22D7AL.16* (6/16), *nhr-30* (5/16), *nhr-188* (3/16) and *nhr-38* (2/16).
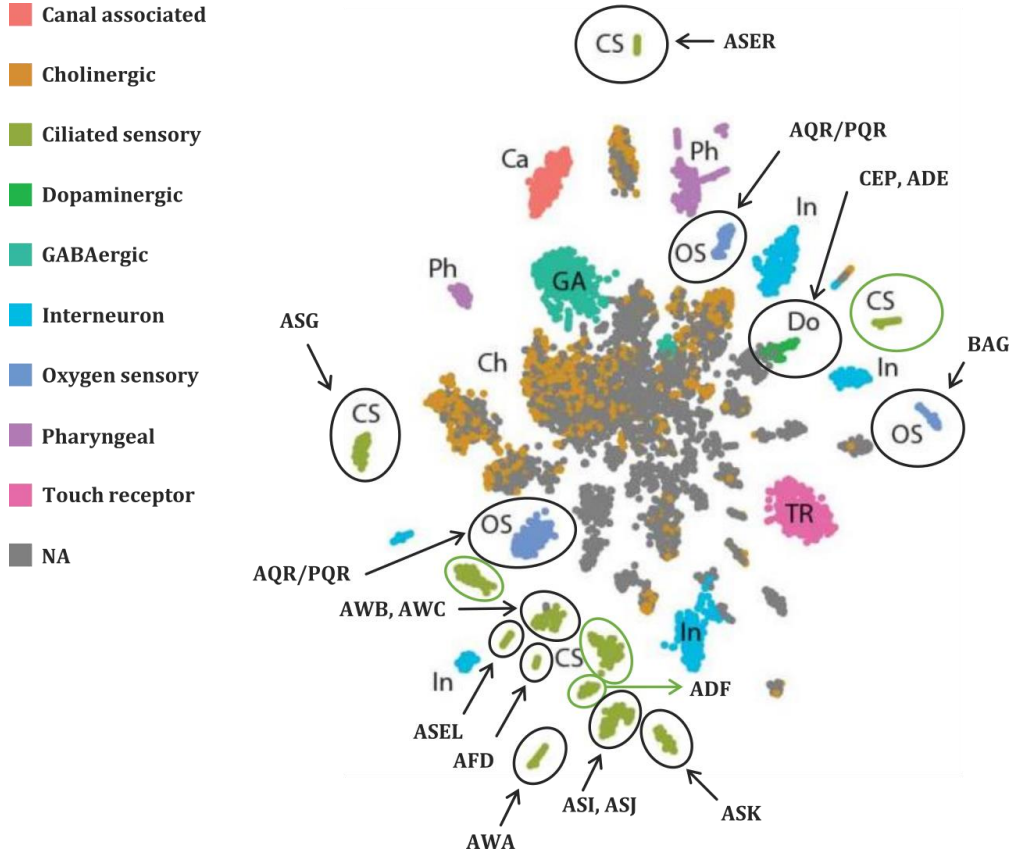
⮝ Figure R.14. t-SNE visualization of high-level neuronal subtypes from L2 animals.
Each dot represents a sequenced neuron from L2 sci-RNA-seq data (J. Cao et al. 2017). Different colours illustrate different neuronal sub-populations as identified through the expression of key genes. Ciliated neurons clusters (CS) are encircled and represented as green dots, or blue dots for oxygen sensing ciliated neurons. Clusters of ciliated neurons are circled and neuron identity is indicated by the name. Clusters from which the specific neuronal identities could not be assessed in the original publication (those for ADF, ASH, PHA and PHB) are circled in green. A cluster for the ADF neurons is suggested in this work based on the exclusive expression of the *srh-142* gene. Clusters in which several neuronal subpopulations were identified show the names of those neurons separated by commas. Clusters for which neuronal subpopulations could not be unequivocally assigned show the names of those neurons separated by slashes. As a whole, this representation allows to retrieve expression information from 34 out of the 60 ciliated neurons of the worm, hence accounting for roughly 57% of the system. Figure adapted from (J. Cao et al. 2017).

Thus, we next decided to focus on *nhr-277, fkh-8, nhr-158* and *ceh-57* to confirm broad and specific expression in sensory ciliated neurons. To that aim, we used an additional set of sc-RNA-seq data, the *C. elegans* Neuronal Gene Expression Network (CeNGEN) (see **Figure R.16**), (Taylor et al. 2019), whose results are freely accessible through the online tool SCeNGEA. Unlike the (J. Cao et al. 2017) data set, that employs an L2 larval stage, the CeNGEN data set profiles L4 animals. L4 expression of all four TFs was consistent with data from L2 larvae; however, in contrast to the L2 data, the most broadly expressed TF at the L4 stage was *fkh-8*. We found *fkh-8* expression in 24 out of the 25 different sensory neuron classes composing the ciliated system of *C. elegans* (corresponding to 26 clusters), while *nhr-158* was detected in 20 neuron types (22 clusters), *nhr-277* in 19 neuron types (21 clusters) and *ceh-57* in 16 neuron types (18 clusters) (see **Figure R.17**).

Altogether, considering expression in both data sets and the presence of FKH motifs that we found in putative regulatory sequences of ciliome genes, we decided to further focus on *fkh-8*. Moreover, *fkh-8* is a known evolutionary conserved TF, and vertebrate FoxJ1, the master regulator of motile cilia, belongs to the same TF family. In contrast, *nhr-158* and *nhr-227* belong to a family of TFs that has suffered a massive expansion in *C. elegans* and they could stand as nematode-specific NHRs likely having their functions being nematode-specific as well.

FKH-8 is broadly expressed across the whole ciliated system of *C. elegans*

Single-cell RNA-seq data suggest FKH-8 is specifically expressed in ciliated neurons; thus, our next aim was to corroborate this finding using fluorescent reporters. Transcriptional regulation of TFs is often complex and regulatory modules can be found at long distances from the transcriptional start site. Accordingly, to accurately recapitulate *fkh-8* endogenous expression, we used an integrated FKH-8 fosmid reporter strain available from the CGC (*wgIs652*). This strain contains the fosmid WRM0630bB04, which includes the whole *fkh-8* locus plus several genes up and downstream spanning approximately 35 kilobases of chromosome II. In this strain, FKH-8 is tagged in its Carboxyl terminus with the TY1 epitope, enhanced GFP and three copies of the FLAG epitope (*fkh-8::TY1::EGFP::3xFLAG*) (see **Figure R.18**). To unequivocally identify sensory ciliated neurons, we used an integrated transcriptional reporter strain for the panciliary gene *ift-20* (*otIs395*), for which we detect expression, as a mean, in roughly 56 out of the 60 ciliated neurons within a wild type genetic background.

We first analysed young adult animals. FKH-8 expression was found in the head, midbody and tail regions, where ciliated neurons locate. Co-localization analysis with the *ift-20* reporter showed that FKH-8 expression almost perfectly matched with the ciliated sensory system (**Figure R.19**). Quantification of co-expression showed 93% of *ift-20* expressing neurons were also *fkh-8* positive while 83% of *fkh-8* positive neurons were *ift-20* positive. Unexpectedly, we noticed that overexpression of *fkh-8* due to the presence of the *fkh-8* fosmid significantly reduced the mean number of *ift-20* positive neurons from

roughly 56 (as assessed over a population of 10 worms) to roughly 47 (when analysing 8 animals from the co-localizing strain). These defects were sparse across the whole ciliated system and varied from animal to animal. Nevertheless, as in preceding sections, five anatomical regions were used for the classification of the ciliated neurons expressing *fkh-8* (see **Figure R.20**). 95% of the neurons expressing *ift-20* were also positive for *fkh-8* in Region 3 (that is, neurons located behind the first bulb), whereas this co-localization percentage was 94% in Region 1 (amphid neurons), 66% in Region 4 (neurons located behind the second bulb corresponding to the ADE, AQR and FLP neurons), 66% in Region 5 (PDE neurons) and 100% in Region 2 (where phasmid neurons locate). Different levels for the intensity of the *fkh-8* reporter could also be observed. This fact was especially evident in the head when comparing the signals from neurons located in Region 3, in which a lower intensity was retrieved, and those located in Region 1.

Based on both location and *ift-20* co-localization we could easily identify expression of FKH-8 in the PHA, PHB and PQR neurons located in the tail, PDE neurons located in the midbody and ADE and AQR neurons located behind the second bulb of the pharynx. Through DiD staining we assessed FKH-8 expression in the ASK, ADL, ASI, AWB, ASH and ASJ amphid neurons. Thus, considering the mean number of neurons expressing *fkh-8* in Region 1 (roughly 24), the ASE, ADF, ASG, AWA, AWC and AFD amphid neurons were most certainly also expressing FKH-8. Similarly, and although no additional co-localization was performed, considering the mean number of neurons expressing *fkh-8* in Region 3 (roughly 21) plus overlapping *ift-20* expression (95%), we suggest *fkh-8* to be expressed in virtually all ciliated neurons located in that area; namely: IL2, IL1, OLL, OLQ, BAG and CEP neurons. Analogously, through *dat-1* co-localization, we could assess FKH-8 expression within the ADE neurons, as well as in the non-bilateral AQR neuron due to its closeness to ADER. Lack of FKH-8 expression within the FLP neurons was in full agreement with the CeNGEN data, in which no expression was detected for this neuron class.

➤ **Figure R.15. t-SNE visualization of transcription factor expression in the different neuronal clusters.**

t-SNE neuronal clustering from (J. Cao et al. 2017) is used here to visualise expression patterns of the ten transcription factors enriched in ciliated sensory neurons. Individual neurons expressing a given transcription factor are depicted as red dots. As expected from enrichment analysis, expression in non-ciliated neuron clusters is almost absent. Clusters of ciliated neurons in which red dots are found are circled in green, thus representing a cluster being "ON". Cluster of ciliated neurons in which no expression has been found are circled in red, thus representing a cluster being "OFF". Broadness of expression across the ciliated system among the ten transcription factors is ordered up to bottom and left to right.

nhr-277 · *fkh-8* · *nhr-158* · *ceh-57* · *nhr-216* · *nhr-124* · *Y22D7AL.16* · *nhr-30* · *nhr-188* · *nhr-38*

⬆ Figure R.16. UAMP visualization of high-level neuronal subtypes.
The Uniform Manifold Approximation and Projection (UMAP) algorithm was used by (Taylor et al. 2019) to reduce dimensionality in the clustering of the original expression data from L4 animals. Each dot represents here a sequenced neuron from such single-cell RNA-seq dataset. Different colours illustrate different neuronal sub-populations as identified through the expression of key genes. Neuronal identity is indicated by name and clusters of ciliated neurons are shown as grey boxes. The whole ciliated system of *C. elegans* is represented in this visualization. Figure adapted from (Taylor et al. 2019).

The expression pattern that we could assess for the *fkh-8* fosmid-based reporter analysed in this work was generally restricted to the ciliated system. However, some consistent expression could also be detected in other neurons of the young adult hermaphrodite. Specifically, we found FKH-8 expression in VC4 and VC5 neurons, located perpendicularly at both sides of the vulval lips (see **Figure R.21.A**). This expression was also confirmed by the CeNGEN data, showing expression levels close to the ciliated CEP and AFD neurons. A pair of additional unidentified cells, anterior to the PDE neurons also expressed FKH-8. These are likely to be the PVD neurons, as that is the only bilateral pair in addition to the PDE in such anatomical region (see **Figure R.21.B**). This identity was also confirmed by CeNGEN data, which shows strong FKH-8 expression in the PVD neurons at levels close to those observed for the ciliated PHA and PHB classes.

⌃ Figure R.17. UMAP visualization of transcription factor expression in different clusters of ciliated neurons.

UMAP neuronal clustering from (Taylor et al. 2019) is used here to visualise expression patterns of 4 transcription factors enriched in ciliated sensory neurons. Neurons expressing a given transcription factor are depicted as red dots. Clusters gathering different classes of ciliated neurons in which red dots are found are circled in green. Broadness of expression across the ciliated system among the 4 transcription factors is ordered up to bottom and left to right.
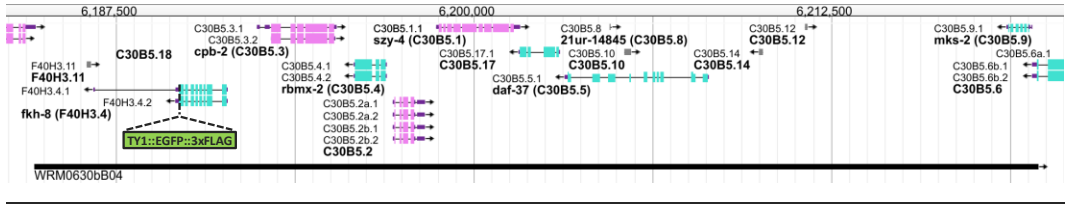
▲ **Figure R.18. Structure of the fosmid reporter for *fkh-8*.**

FKH-8 expression was assessed through a fosmid-based reporter strain carrying the fosmid WRM0630bB04. In the context of approximately 35 kilobases fosmid containing several genes upstream and downstream from *fkh-8*, the C-terminal end of the *fkh-8* locus was engineered to contain in frame the TY1 epitope, enhanced GFP and three copies of the FLAG epitope.
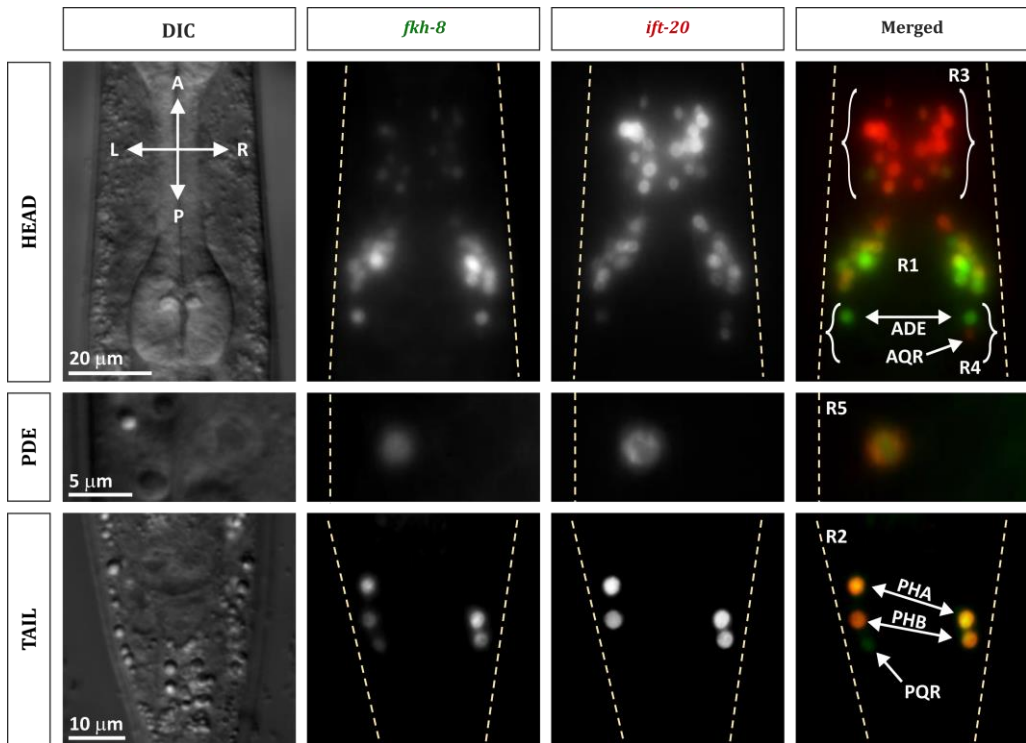
▲ **Figure R.19. *fkh-8* is expressed across the whole ciliated system of *C. elegans*.**

Representative dorsoventral views of a *C. elegans* young adult showing EGFP wild type expression of an integrated fosmid-based reporter for *fkh-8* (*wgIs652*) and RFP panciliary expression for the *otIs39*5 reporter of *ift-20*. First column displays the DIC images, second and third columns show the expression for the fluorescent reporters of *fkh-8* and *ift-20* respectively whereas the final column shows colocalization between both reporters with *fkh-8* being represented in green and *ift-20* in red. For all pictures, spatial positioning is as depicted in the first image of the first panel where A stands for anterior, P for posterior, R for right and L for left. Body limits of the worm are represented as dashed lines in fluorescent images. Final fluorescent images correspond to Z stacks projections of images at 1 micron distance taken under a vertical fluorescence microscope. Notice that for this particular animal, PQR (last neuron in the left side of the worm) is positive for *fkh-8* but does not appear to express *ift-20* at detectable levels.
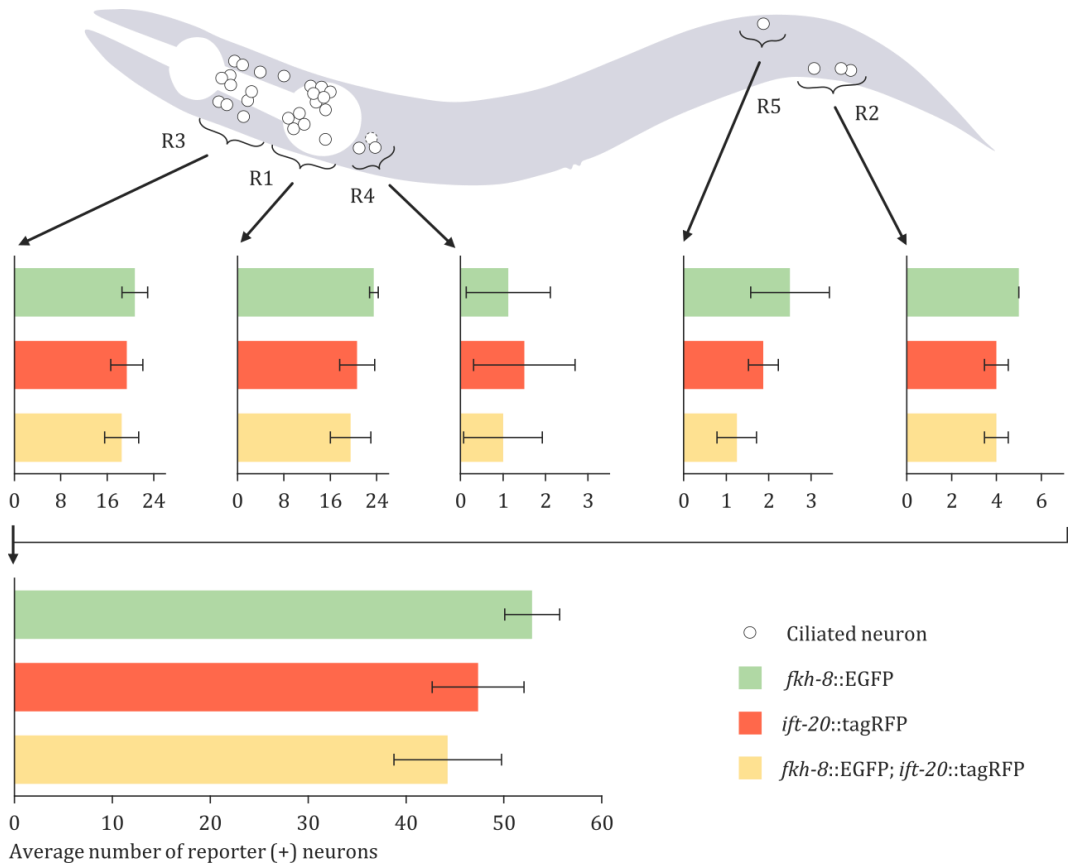
Figure R.20. Adult *fkh-8* expression highly colocalizes with the panciliary reporter *ift-20*.
Top scheme shows a lateral left view of a *C. elegans* young adult hermaphrodite portraying the location of the ciliated system neurons. Non-bilateral AQR neuron in the head locates in the right side of the animal and hence appears depicted in dashed lines. For each of the five anatomical regions, corresponding graphs show the average number of reporter-expressing neurons for both *fkh-8* (in green) and *ift-20* (in red) reporters, as well as their colocalization (in yellow). Bottom graph gathers the scorings for the whole animals. Error bars represent here the standard deviation of the mean. N = 8 animals.

Next, we performed FKH-8 expression analysis throughout embryonic development. Embryonically generated ciliated neurons are born at 300 to 400 minutes post-fertilization, during bean and comma stages, and they differentiate shortly after. First *fkh-8* expression was already visible in a pair of neurons located in the tail at early bean stage, which are presumably the PHA ciliated neurons born at around 300 minutes (**Figure R.22**). Shortly after, at early comma stage, GFP expression was detected also in the head (**Figure R.22**). The number of GFP expressing cells in the head and tail continued to increase during comma stage and already at the two-fold embryonic stage *fkh-8* reached an apparent panciliary expression pattern as determined by co-localization with *ift-20* (**Figure R.22**). Of note, although ciliated neurons are born at bean and comma stages, *ift-*

20:tagRFP signal was not visible until the two-fold stage. This is likely due to the fact that RFP-derived fluorescent reporters require longer maturation times compared to those based on GFP (Balleza, Kim, and Cluzel 2018). Indeed, using an *ift-20*::GFP reporter we were able to detect GFP expression already at the early coma stage (see **Figure R.23**), thus starting shortly after the onset of detectable *fkh-8*::EGFP expression.
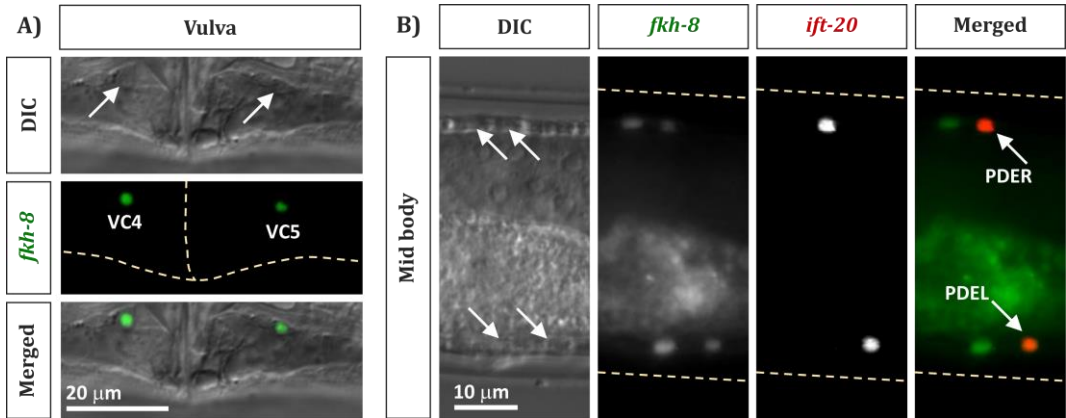
⬏ **Figure R.21.** *fkh-8* **is expressed in a few non-ciliated neurons of the worm.**
A) Representative lateral view of the vulval region of a young adult hermaphrodite. In the first panel, white arrows pinpoint at the nuclei of the VC4 and VC5 neurons, in which fkh-8 expression is clearly seen and represented in green on the following images. **B)** Representative top view of a section of the mid body of a young adult hermaphrodite. In the first panel, white arrows pinpoint at the nuclei of the bilateral cells in which *fkh-8* expression is observed (in panel two). Clear expression of *ift-20* (panel two) allows for the identification of the PDE neurons. In panel four, colocalization of both reporters revealed the anterior location of the two unidentified cells when compared against the PDE neurons. For all images showing fluorescent signal, body limits of the worm are represented as dashed lines.

In summary, *in vivo* reporter analysis performed in this section demonstrates *fkh-8* expression is restricted to the whole ciliated system. FKH-8 expression onset in development coincides with neurogenesis and early differentiation of ciliated neurons and its expression is maintained throughout the life of the animal. These results prompted us to select *fkh-8* as a strong candidate controlling transcriptional regulation of ciliated features in *C. elegans*.
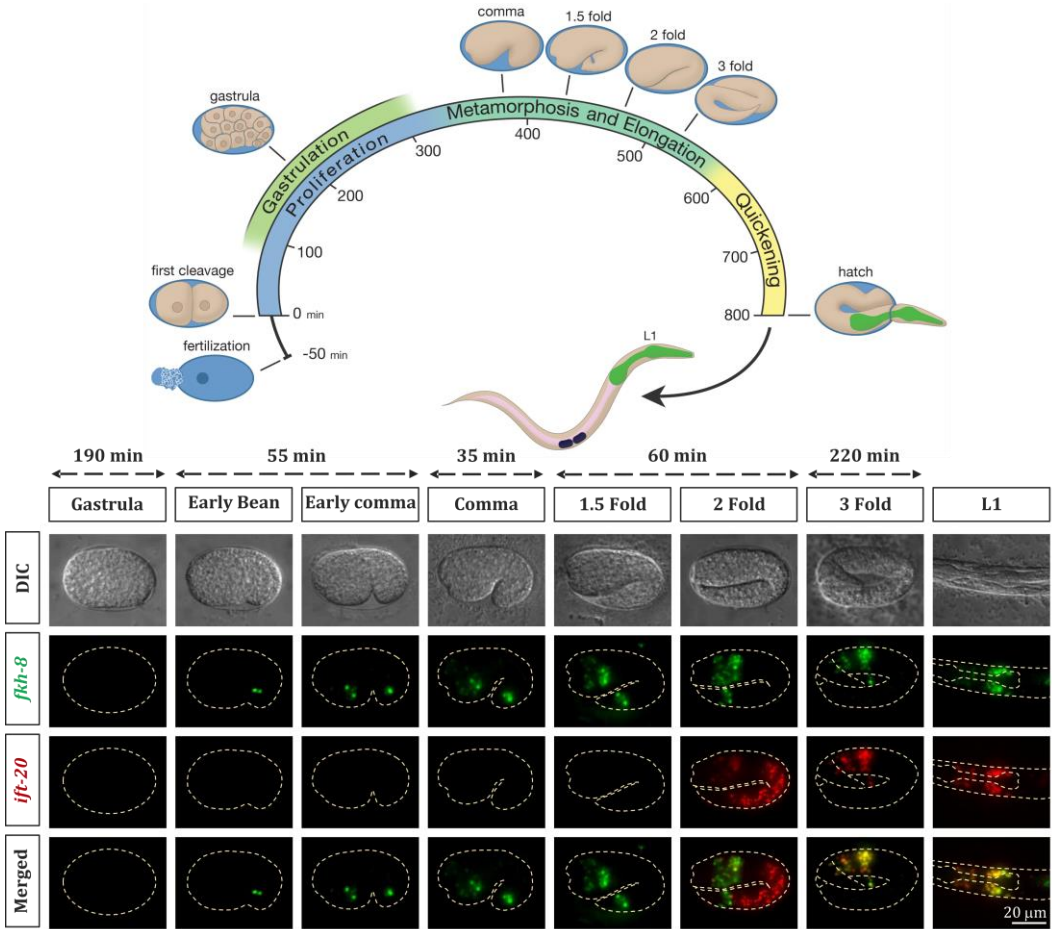
⌃ Figure R.22. Developmental expression pattern of *fkh-8*.

Upper scheme represent the major landmarks in the *C. elegans* embryonic development until hatching as a L1 worm at 22 degrees Celsius. Taken from WormAtlas. Bottom grid shows representative images for the different embryonic stages of *C. elegans* until its hatching. First row displays the DIC images exposing the morphology of the developing animals, second and third rows show the expression for the fluorescent reporters of *fkh-8* (in green) and *ift-20* (in red) whereas the final row shows the colocalization between both reporters. For all fluorescent images body limits of the worm are represented as dashed lines. Final fluorescent images correspond to Z stacks projections of images at 1 micron distance taken under a vertical fluorescence microscope. Time frames for each developmental stage were extracted from (Cornaglia *et al.* 2015) and appear indicated at the top of the figure with dashes lines.
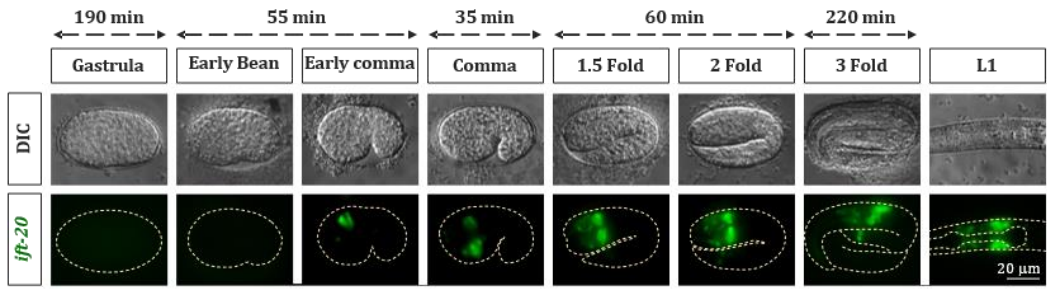
▲ **Figure R.23. Developmental expression pattern of *ift-20::gfp reporter*.**
First row displays the DIC images exposing the morphology of the developing animals whereas the second row reveals the expression of *ift-20* tagged with GFP (*sIs13113*). For all fluorescent images body limits of the worm are represented as dashed lines. Final fluorescent images correspond to Z stacks projections of images at 1 micron distance taken under a vertical fluorescence microscope. Time frames for each developmental stage were extracted from (Cornaglia *et al.* 2015) and appeared indicated at the top of the figure with dashes lines.

3. FKH-8 preferentially binds regulatory regions of ciliome genes

FKH-8 preferentially binds to ciliome genes when compared to the binding of other TFs.

Our candidate approach allowed us to identify FKH-8 as a TF specifically expressed in ciliated neurons, with a developmental expression similar to the DAF-19C isoform. Thus, we next aimed to analyse if FKH-8 binds in the genome near ciliome effector genes.

Two major consortia, modENCODE and modERN, are aimed to generate and provide access to high volumes of genomic information including several hundred ChIP-seq experiments. At the time of consulting (January the 10th, 2019), 446 of such *C. elegans* ChIP-seq data sets from different developmental stages of the worm allowed us to retrieve information from 259 different transcription factors, including FKH-8 but not DAF-19. We used our list of 163 ciliome genes and asked whether ChIP-seq peaks from any of those 446 datasets were located between 2,000 base pairs upstream and 1,000 base pairs downstream of their putative translational start site. From the distribution generated by the number of targets that those 446 datasets had, 6 transcription factors, namely, *fkh-8*, *dsc-1*, *C34F6.9*, *elt-2*, *nhr-71* and *lsy-2*, behaved as statistical outliers, with a higher number of peaks targeting ciliome genes when compared to the rest of the data sets (see **Figure R.24A**). Remarkably, *fkh-8* ChIP-seq peaks targeted considerably more ciliome genes than any other TF (see **Figure R.24A**). However, this could just be the reflection of a high number of total peaks retrieved from the *fkh-8* ChIP-seq experiment. To discard this possibility we plotted, for each of the 446 data sets, the number of targeted ciliome genes as a function of the total number of peaks in each data set. As expected, we found a positive correlation between the total number of peaks in each data set and the number of ciliome genes targeted by peaks for the same TF (Pearson's correlation = 0.6905) (**Figure R.24B**). Finally, we compared the FKH-8 data set individually against the 33 ChIP-seq experiments in which the number of total peaks was greater than those of *fkh-8*. As expected, the proportion of ciliome genes targeted by *fkh-8* ChIP-seq peaks was significantly higher than those corresponding to any of the 33 ChIP-seq data sets used for comparison (**Table R.7**). Therefore, among all available ChIP-seq data, we found FKH-8 preferentially binds to curated cilium-related genes.

Next, we aimed to analyse if FKH-8 binding was equally distributed in the four functional categories we used to classify the genes in our ciliome gene list. Interestingly, we found *fkh-8* ChIP-seq peaks associated most significantly to Structural ciliary genes, since roughly a 75% of those genes were targeted by a FKH-8 peak in comparison with the 22% of Subtype-specific ciliary genes being bound by *fkh-8*. No trend was observed for the presence of FKH-8 peaks in Male ciliary components or genes under the Broad expression category (see **Figure R.25A**).
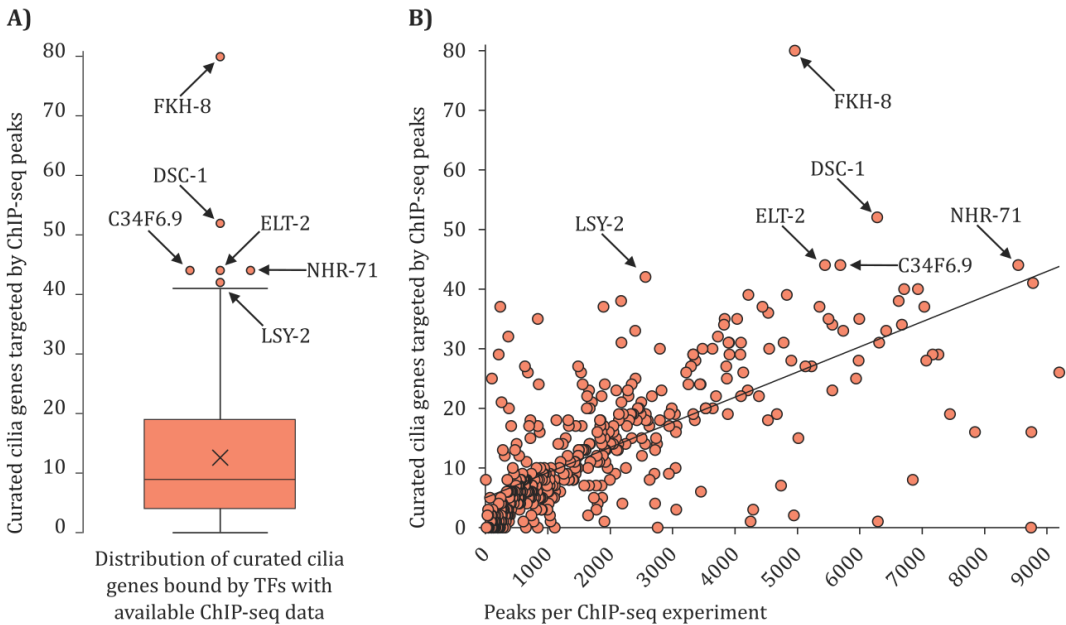
⬆ Figure R.24. *fkh-8* **ranks first targeting cilia genes among 446 available ChIP-seq experiments.**
A) Box plot for the distribution generated by the number of curated cilia genes targeted by each of the 446 ChIP-seq experiments analysed. Six transcription factors behave as outliers since they locate 1.5 times the interquartile range above the third quartile. **B)** Correlation between total number of peaks and number of cilia genes with an associated peak. For the whole dataset of 446 ChIP-seq experiments analysed, a positive correlation exists between the number of cilia genes being targeted and the number of peaks in a ChIP-seq experiment. Pearson correlation coefficient between the two variables was found to be 0,6905. Student's t-distribution was used to infer statistical significance for the correlation. With a p-value of 1,61E-64, null hypothesis for absence of correlation was rejected. ChIP-seq data analysis was performed with the help of Carlos Mora Martínez.

➤ Table R.7. *fkh-8* **ChIP-seq peaks target a higher proportion of cilia genes even when compared against ChIP-seq experiments for other TFs with more assigned peaks.**
ChIP-seq experiments for 33 transcription factors show a higher number of peaks than those retrieved by *fkh*-8 datasets. Proportion (p) of curated cilia genes targeted per peak was calculated for each experiment and compared against that of *fkh-8*. Score from z-tests for proportions (z) and their associated p-values are indicated for each comparison. All comparisons were statistically significant after Bonferroni correction considering the 33 hypotheses being tested simultaneously (p-value ≤ 0.05/33). modEN-CODE/modERN identifiers are shown, as well as worm's developmental stage in which each experiment was performed. LE: late embryo, L1: larval 1, L2: larval 2, L3: larval 3, L4: larval 4, YA: young adult, S1: starved L1.

ChIP-seq gene	Stage	ChIP-seq ID	Peaks	Targets	p	z	p-value
fkh-8	L1	ENCFF616NLH	4963	80	0.0161		
elt-2	L3	ENCFF417YEF	9198	26	0.0028	8.7560	0.00E+00
daf-16	L4	ENCFF387SWS	8777	41	0.0047	6.8993	2.61E-12
blmp-1	L1	ENCFF786OZI	8751	16	0.0018	9.6461	0.00E+00
nhr-28	L4	ENCFF049VYW	8745	0	0.0000	11.9076	0.00E+00
nhr-71	L1	ENCFF880YTO	8540	44	0.0052	6.4415	5.92E-11
pha-4	L3	ENCFF731VKY	7847	16	0.0020	9.0019	0.00E+00
fos-1	L2	ENCFF401LJF	7446	19	0.0026	8.3226	0.00E+00
dve-1	L4	ENCFF560YTQ	7262	29	0.0040	7.0040	1.24E-12
hlh-30	L4	ENCFF793PTZ	7168	29	0.0040	6.9287	2.12E-12
sma-9	L2	ENCFF220NCO	7065	28	0.0040	6.9577	1.73E-12
nhr-80	YA	ENCFF956VXA	7038	37	0.0053	5.9641	1.23E-09
nfya-1	L3	ENCFF797WBT	6932	40	0.0058	5.5694	1.28E-08
lsy-2	S1	ENCFF693SQO	6850	8	0.0012	9.3279	0.00E+00
lin-35	L1	ENCFF638REK	6712	40	0.0060	5.3807	3.71E-08
snu-23	L1	ENCFF468IRP	6675	34	0.0051	5.9729	1.17E-09
nhr-129	L2	ENCFF266DSF	6626	38	0.0057	5.5101	1.79E-08
nhr-20	L1	ENCFF390OCN	6423	33	0.0051	5.8618	2.29E-09
lsy-2	L2	ENCFF685EPU	6312	31	0.0049	5.9837	1.09E-09
ces-1	LE	ENCFF884HVS	6290	1	0.0002	9.9441	0.00E+00
dsc-1	L1	ENCFF628DQW	6283	52	0.0083	3.8345	6.29E-05
unc-55	L2	ENCFF126HEZ	5993	35	0.0058	5.2552	7.39E-08
cebp-1	YA	ENCFF828QGA	5982	28	0.0047	6.0271	8.35E-10
lsy-2	L4	ENCFF637JNX	5944	25	0.0042	6.3453	1.11E-10
C27D6.4	L2	ENCFF402SFY	5735	33	0.0058	5.2297	8.49E-08
C34F6.9	L2	ENCFF747FCY	5692	44	0.0077	4.0277	2.82E-05
rnt-1	L1	ENCFF650RDJ	5560	34	0.0061	4.9488	3.73E-07
pqm-1	L3	ENCFF274XKV	5560	23	0.0041	6.2327	2.29E-10
mes-2	L4	ENCFF587ZXU	5501	35	0.0064	4.7803	8.75E-07
elt-2	L1	ENCFF848ZRN	5443	44	0.0081	3.7731	8.06E-05
nhr-47	L1	ENCFF043OZU	5357	37	0.0069	4.4166	5.01E-06
mep-1	L4	ENCFF252LNY	5229	27	0.0052	5.4241	2.91E-08
F13H6.1	L2	ENCFF501VXJ	5133	27	0.0053	5.3269	4.99E-08
nhr-25	L3	ENCFF452LBG	5017	15	0.0030	6.7539	7.20E-12
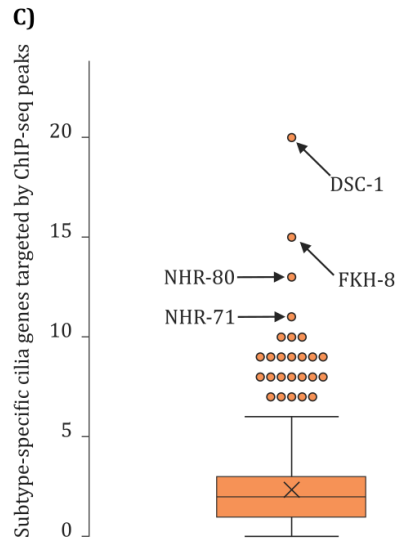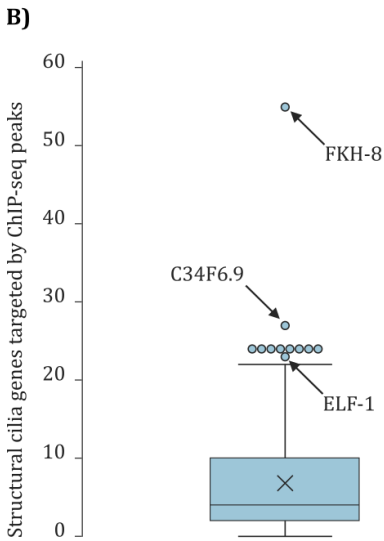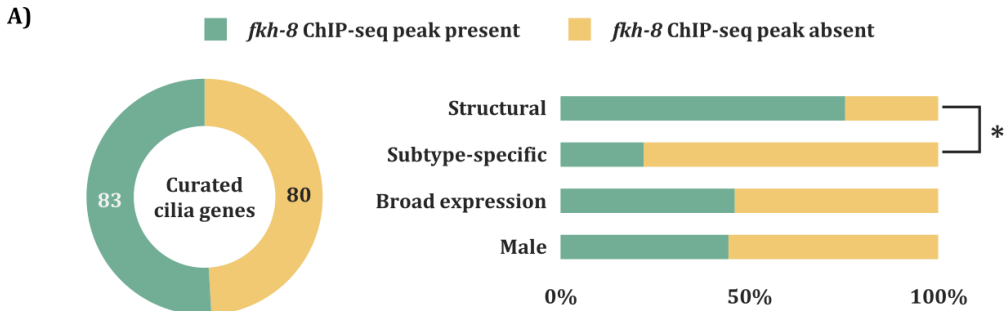
⌂ Figure R.25. FKH-8 binds preferentially to structural cilia genes.
A) 80 out of the 163 genes of our cilia list contain *fkh-8* ChIP seq peaks in sequence windows between 2,000 base pairs upstream and 1,000 base pairs downstream their transcription start site. FKH-8 binding is biased towards structural cilia genes. A Chi-squared test for all 4 categories accounting for the number of genes in which *fkh-8* ChIP-seq peaks were present or absent concluded that those differences were significant (p = 9.98E-09). Further *post hoc* analysis through the Fisher exact test allowed to determine that these differences could be established between "Structural" *vs* "Subtype-specific" cilia genes (p = 1.89E-10) Differences were significant after the modified Bonferroni correction proposed by Keppel (Keppel 1991) accounting for all 6 possible pair comparisons. **B)** Box plot for the distribution generated by the number of Structural cilia genes targeted by each of the 446 ChIP-seq experiments analysed. Nine transcription factors behave as statistical outliers, since they located 1.5 times the interquartile range above the third quartile. Two different ChIP-seq datasets for each TFs *elf-1* and *lsy-2* appear among the eleven dots representing the outliers. FKH-8 ranks first in this distribution with 55 of the 73 Structural genes being targeted by its peaks. **C)** Box plot for the distribution generated by the number of Subtype-specific cilia genes targeted by each of the 446 ChIP-seq experiments analysed. Twenty-three transcription factors behave as statistical outliers, since they locate 1.5 times the interquartile range above the third quartile. Two different ChIP-seq datasets for each TFs *let-2, pqm-1* and *lsy-2* appear among the twenty-six dots representing the outliers. DSC-1 ranks first in this distribution with 20 of the 68 Subtype-specific genes being targeted by its peaks while FKH-8 ranks second.

Association of *fkh-8* with Structural ciliary genes was further exposed when analysing the presence of ChIP-seq peaks from any of the 446 data sets previously used. From the distribution generated by the number of targets those 446 ChIP-seq data sets had within the 73 Structural ciliary genes present in our curated ciliome list, 9 transcription factors (namely, *fkh-8, C34F6.9, dpl-1, dsc-1, lin-35, lsy-2, nfya-1, pha-4* and *efl-1*) behaved as statistical outliers, with a higher number of peaks targeting such type of genes when compared to the number of targets from the rest of the TFs in the set (see **Figure R.25B**). Again, *fkh-8* ChIP-seq peaks targeted considerably more Structural ciliary genes than any other TF (see **Figure R.25B**).

Next, we analysed the binding profile of the 446 ChIP-seq data sets within the 68 Subtype-specific ciliary genes included in our curated ciliome list. On this occasion, up to 23 TFs (namely, *dsc-1, fkh-8, nhr-80, nhr-71, daf-16, elt-2, nhr-28, blmp-1, C34F6.9, fos-1, nhr-47, pqm-1, rnt-1, C27D6.4, efl-1, lsy-2, nfya-1, nhr-129, nhr-25, snpc-4, dve-1, F13H6.1* and *sma-9*) behaved as statistical outliers, targeting a higher number of Subtype-specific ciliary genes when compared to the number of targets from the rest of the TFs in the set (see **Figure R.25C**). Interestingly, *dsc-1*, a TF known to be expressed in subsets of amphid neurons (namely the ASE and AWC neurons, as assessed through the CenGEN data), targeted considerably more Subtype-specific ciliary genes than any other TF (see **Figure R.25C**). Of note, *fkh-8* ranked second in this analysis, targeting 15 of the 68 genes included within the Subtype-specific category.

Thus, similar to the X-box, *fhk-8* binding seems preferentially associated to Structural ciliary genes although an association of *fkh-8* with a subset of Subtype-specific ciliary genes is also suggested when comparing its binding profile against 258 different TFs.

Genomic binding analysis of FKH-8.

Once assessed the preferential binding of FKH-8 to the genes integrating our curated ciliome list, we decided to expand our analysis for the FKH-8 binding profile to the whole genome. To retrieve putative FKH-8 target genes we applied the commonly used annotation criteria employed by HOMER (Heinz *et al.* 2010), in which peaks are assigned to the gene whose transcriptional start site is closer. In this way, FKH-8 ChIP-seq peaks were annotated to 3,987 final unique genes. Analysis based on the GO Biological Process 2018 ontology (see **Annex 2**) revealed a statistically significant enrichment in several cilium-related categories (such as cilium organization and assembly) or cilia functionality (such as those associated with *dauer* development), as well as an enrichment in neuronal, transport or signalling categories (see **Figure R.26**). Moreover, analysis of the genes assigned to FKH-8 ChIP-seq peaks on the basis of the Anatomy Association Worm-Base 2018 ontology (see **Annex 2**) showed a strong association to neuronal categories, especially to ciliated sub-types (**Figure R.27**). Finally, KEGG 2019 ontology analysis showed enrichment in both mTOR signalling (p-value = 1.20E-04), a key pathway that has been found to be regulated by primary cilia to control cellular size (Boehlke *et al.* 2010), as well as the phosphatidylinositol signalling (p-value = 1.01E-03), a system that has been

recently found to regulate ciliary protein trafficking involved in the modulating of Hedgehog signals (Garcia-Gonzalo *et al.* 2015).
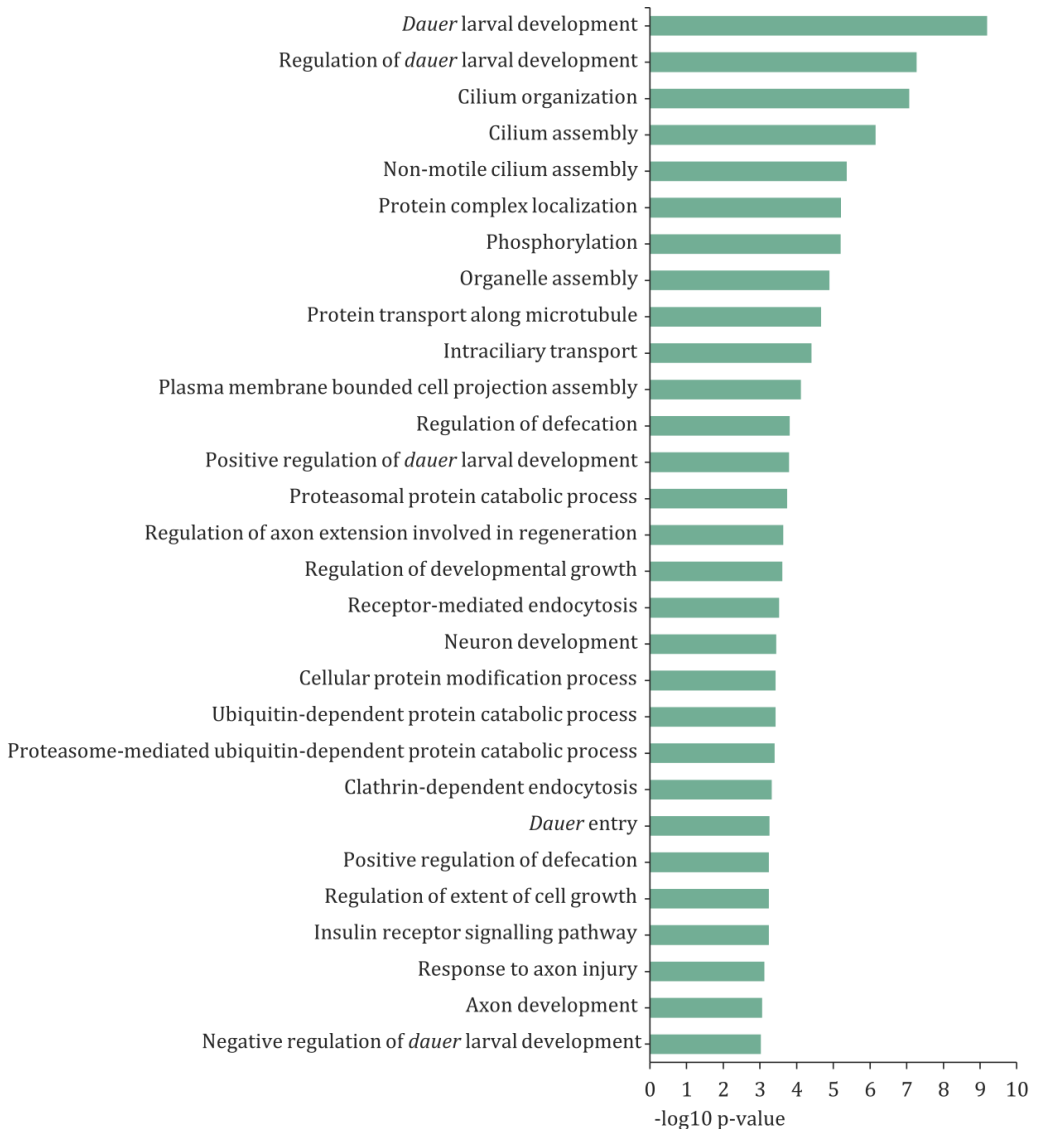
⌃ Figure R.26. Genes assigned to FKH-8 ChIP-seq peaks enrich cilia-related categories for biological processes.

Bars represent the statistical significance for the enrichment of terms in the GO Biological Process 2018 ontology when comparing the set of 3.987 unique genes assigned to FKH-8 ChIP-seq peaks against the *C. elegans* reference. Benjamini-Hochberg procedure was performed to correct for multiple hypotheses.
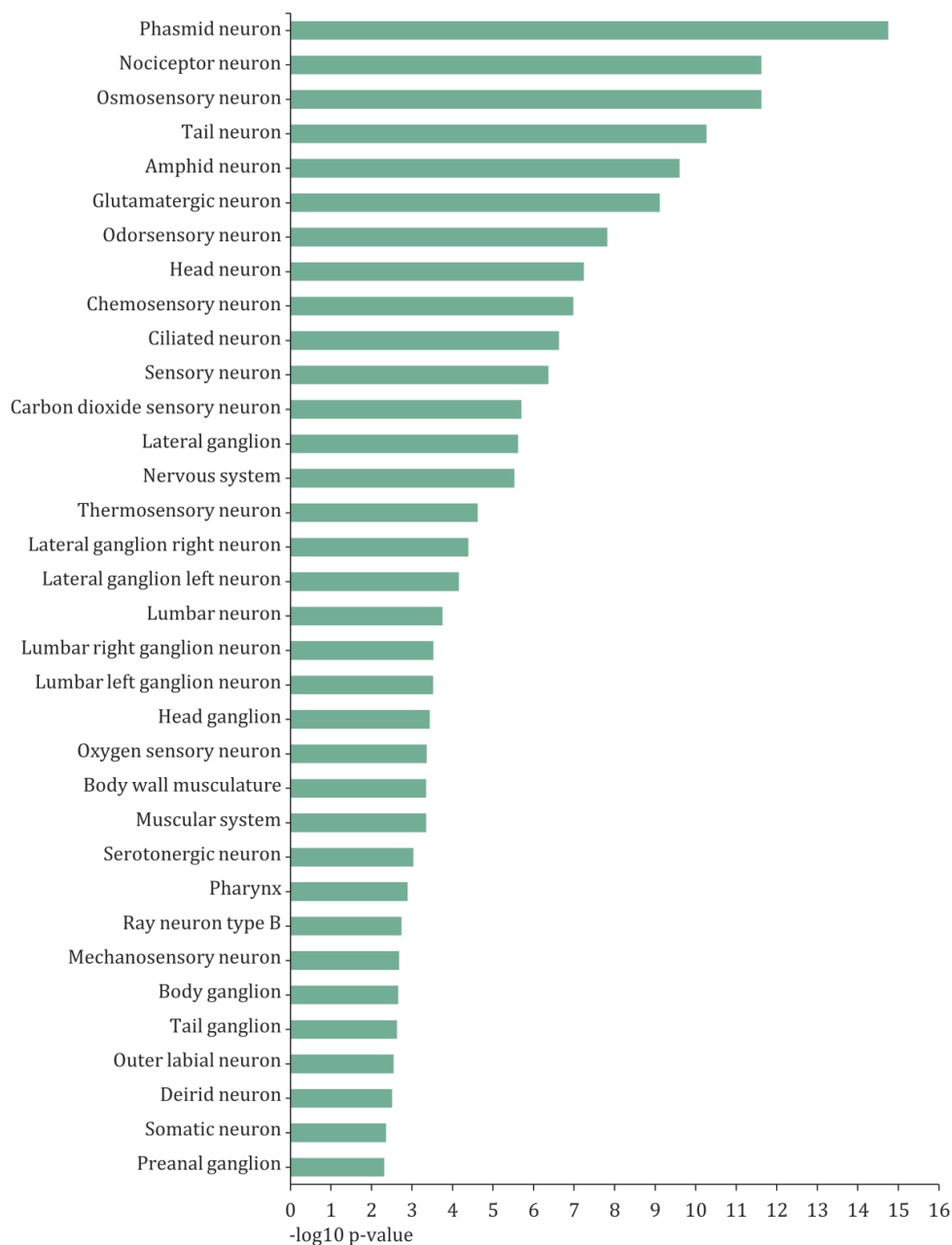
⏶ Figure R.27. Genes assigned to FKH-8 ChIP-seq peaks enrich for anatomic associations related to cilia.

Bars represent here the statistical significance for the enrichment of selected terms in the Anatomy Association WormBase 2018 ontology when comparing the set of 3.987 unique genes assigned to *fkh-8* ChIP-seq peaks against the *C. elegans* reference. Benjamini-Hochberg procedure was performed to correct for multiple hypotheses. To avoid redundancy, all statistically significant enriched terms for individual neuronal classes were deliberately set aside this graph. Those terms include: ADE, ADF, ADL, AFD, ASE, ASG, ASH, ASI, ASJ, ASK, AWA, AWB, AWC, OLQ, PDE, PHA, PHB, PQR, R2B and R6B.

We next aimed to identify putative FKH binding motifs within the sequences of the peaks retrieved in the FKH-8 ChIP-seq experiment that would suggest direct FKH-8 binding to DNA. At the time of writing, no known binding site motif has been experimentally determined for *fkh-8*. We decided to split the set of FKH-8 ChIP-seq peaks considering their relative position towards the closest transcriptional start site (TSS). Accordingly, we created 4 different categories:

1. Proximal peaks centred within 50 base pairs upstream of a TSS (2045 sequences).
2. Proximal peaks centred within 500 base pairs upstream of a TSS (3027 sequences in which those from category 1 were also included).
3. Distal peaks centred beyond 500 base pairs upstream of a TSS (1936 sequences in which those from category 4 were also included).
4. Distal peaks centred beyond 2,000 base pairs upstream of a TSS (630 sequences).

As a first approach aimed to retrieve a general overview for the distribution of FKH-related motifs within the set of FKH-8 ChIP-seq peak sequences, we analysed the presence of HOMER known motifs for members of the FKH family. All 4 categories analysed showed enrichment of FKH-related motifs when compared against their corresponding background controls (see **Table R.8**). The known binding motif for the *C. elegans* TF *pha-4* was the most commonly found motif for a member of the FKH family, being present in roughly 64% of all peak sequences. However, no statistical enrichment for the presence of such motif was found among all four different categories analysed. This is likely due to the small and degenerate consensus of the FKH binding sites. Interestingly, proximal peaks showed enrichment in binding motifs for FKH members of the FoxA and FoxM classes that were absent within the sequences of the distal peaks (see **Table 3.8**).

➤ **Table R.8. FKH-8 ChIP-seq peaks harbour sites matching HOMER known motifs for members of the FKH family.**
Sequences retrieved from a FKH-8 ChIP-seq experiment (ENCSR435RXY) were split into 4 different categories attending to their relative position to the closest transcriptional start site (TSS). Proximal peaks were classified considering the location of their centres within 50 (p50) or 500 (p500) base pairs upstream the closest TSS. Distal peaks were classified considering the location of their centres beyond 500 (d500) or 2,000 (d2000) base pairs upstream the closest TSS. Table gathers the percentage of FKH-8 ChIP-seq peak sequences from each category in which sites matching HOMER known motifs for members of the FKH family were found. Presence of such motifs was significantly higher than that observed in the corresponding control background sequences. The Benjamini-Hochberg method was applied to correct for multiple comparisons. HOMER known motif analysis was performed by Juan J. Tena.

Motif name	Consensus	p50	p500	d500	d2000
Fox:ebox(forkhead,bhlh)/panc1-foxa2-chip-seq(gse47459)/homer	NNNVCTGWGYAAACASN	15.40	15.00		
Foxa1(forkhead)/lncap-foxa1-chip-seq(gse27824)/homer	WAAGTAAACA	28.90	28.18		
Foxa1(forkhead)/mcf7-foxa1-chip-seq(gse26831)/homer	WAAGTAAACA	24.25	23.42		
Foxa1:ar(forkhead,nr)/lncap-ar-chip-seq(gse27824)/homer	AGTAAACAAAAAAGAACAND		1.68		
Foxa2(forkhead)/liver-foxa2-chip-seq(gse25694)/homer	CYTGTTTACWYW	15.21	15.16	16.22	
Foxa3(forkhead)/liver-foxa3-chip-seq(gse77670)/homer	BSNTGTTTACWYWGN	6.11	5.68		
Foxf1(forkhead)/lung-foxf1-chip-seq(gse77951)/homer	WWATRTAAACAN	24.84	24.22	23.19	26.35
Foxk1(forkhead)/hek293-foxk1-chip-seq(gse51673)/homer	NVWTGTTTAC	24.94	24.18	26.60	28.89
Foxk2(forkhead)/u2os-foxk2-chip-seq(e-mtab-2204)/homer	SCHTGTTTACAT	12.91	12.88	14.00	14.60
Foxl2(forkhead)/ovary-foxl2-chip-seq(gse60858)/homer	WWTRTAAACAVG	21.52	21.21	21.07	23.81
Foxm1(forkhead)/mcf7-foxm1-chip-seq(gse72977)/homer	TRTTTACTTW	23.67	23.06		
Foxo1(forkhead)/raw-foxo1-chip-seq(fan_et_al.)/homer	CTGTTTAC	30.27	29.77	31.87	33.33
Foxo3(forkhead)/u2os-foxo3-chip-seq(e-mtab-2701)/homer	DGTAAACA	15.79	15.59	17.72	20.16
Foxp1(forkhead)/h9-foxp1-chip-seq(gse31006)/homer	NYYTGTTTACHN	9.93	9.81	10.07	12.38
Pha-4(forkhead)/celegans-embryos-pha4-chip-seq(modencode)/homer	KTGTTTGC	66.01	64.19	64.67	65.24

As a second approach aimed to discover putative binding sites for FKH-8, we performed a HOMER *de novo* motif discovery analysis. In contrast with the results obtained from the previous HOMER known motif analysis, enrichment for *de novo* sites matching FKH-related motifs was only found in 1 of the 4 FKH-8 ChIP-seq peaks categories being used. Ranking 13th (p-value = 1.00E-29) among the 30 *de novo* motifs found for the set of proximal peaks centred within 50 base pairs upstream of a TSS, 195 of the sequences (9.54% of the total sequences in the category) were able to generate a motif matching the known binding site for the *C. elegans* TF *pha-4* (see **Figure R.28**). Additionally, ranking 14th (p-value = 1.00E-29), a small set of 48 peak sequences (2.35% of the total sequences in the category) were able to generate a motif matching the known binding site for the *S. cerevisiae* TF *FKH1* (see **Figure R.28**).
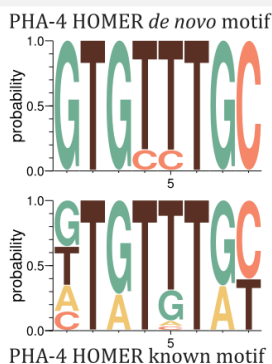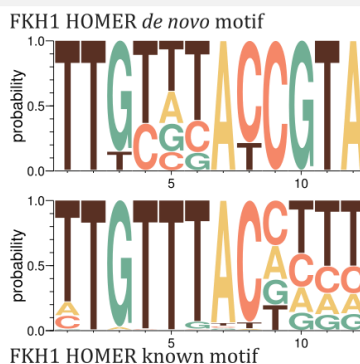
▲ Figure R.28. Sequence logos for motifs generated from HOMER *de novo* discovery analysis. A subset of sequences from proximal FKH-8 ChIP-seq peaks centred within 50 base pairs upstream of the closest TSS generate motifs that match with known binding sites for members of the FKH family. PHA-4 *de novo* motif was found in 195 out of the 2045 sequences in the category, placed within ± 78.5 base pairs from the centred of its target sequences. FKH-1 *de novo* motif was found in 48 out of the 2045 sequences in the category, placed within ± 56 base pairs from the centred of its target sequences.

Acknowledging that different *de novo* motif discovery algorithms may render different results, we decided to expand our analysis by employing the RSAT peak-motif tool (Thomas-Chollier et al. 2012). Accounting for the results of the aforementioned HOMER approach, in which only a small subset of proximal peaks was able to generate FKH-related sites, we decided to perform this new analysis using the whole set of FKH-8 ChIP-seq sequences. Once discovered, we asked for the comparison of the *de novo* motifs obtained against the CIS-BP 2.00 database. At the time of writing, the binding sites for 5 of the 17 FKH members encoded in the *C. elegans* genome have been experimentally determined (namely, those for *daf-16*, *fkh-9*, *let-381*, *lin-31* and *pha-4*). Additionally, 6 *C. elegans* FKH members has had their binding sites inferred from those established for other members of the FKH family, both in *C. elegans* or in other organisms (such are the cases of *fkh-*

2/6/7/8/10 and *unc-130*). Finally, no binding sites have been defined for the FKH TFs *fkh-3/4/5*, *pes-1*, *C34B4.2* and *T27A8.2*. All 4 *de novo* motif discovery algorithms available at the RSAT peak-motif tool were used in this analysis (namely, *positions*, *oligos*, *dyads* and *local_words*), each asked to provide up to 10 motifs. From the final 40 *de novo* motifs that we retrieved, 9 were able to match known CIS-BP 2.00 FKH-related binding sites (see **Table R.9** and **Figure R.29**).

Experimental evidence has shown that the binding sites of assayed TFs in successful TF ChIP-seq experiments cluster near the centres of declared ChIP-seq peaks (Bailey and MacHanick 2012). Accordingly, the distribution of a biologically relevant motif is expected to peak around the centre of such sequences whereas less relevant motifs are likely to produce a more flattened pattern (Mercier et al. 2011). Thus, to discern the

biological relevance of our *de novo* FKH-related motifs, we profiled their distribution within the sequences of their targeted peaks (see **Figure R.30**). Flattened distribution could be observed for Motifs 37, 39, 35 and 36, hence suggestive for artefactual or background motifs. However, Motifs 23 and 30 displayed a peak-like distribution clustered in regions of 100 base pairs centred around the peak, a common trend exhibited for TFs

when directly binding DNA (Bailey and MacHanick 2012). Additionally, bimodal distributions in which sites located at the edges of their targeted peaks could be observed for Motifs 40, 31 and 9. Much interestingly, such distribution has been reported for members of the FKH family when acting as pioneer factors in regions to open chromatin (Grossman et al. 2018).

▼ **Table R.9. A subset of FKH-8 ChIP-seq peaks contain motifs matching known binding sites for members of the FKH family.**
RSAT peak-motif tool was used to analyse the whole set of FKH-8 ChIP-seq peak sequences. For each enriched motif consensus retrieved, several matches against members from different TF families were detected. In this table, only motifs with matches against members of the FKH family are collected. For each FKH-related motif found, best *C. elegans* FKH candidate is indicated. Source of matched motif appears as CIS-BP 2.00 ID; name and organism from matched FKH TF are also shown. Similitude between found motifs and the matched known motifs is represented through a Pearson coefficient of correlation between the frequency matrices (C). A width-normalized correlation (N.C) accounting for the number of individual matching positions between frequency matrices is used to prevent spurious correlation arising from the partial matches between motifs. Number of individual occurrences (O) and individual peaks (P) in which a given motif appear are also indicated.

Name	Best *C. elegans* *fkh* candidate	Matched CIS-BP id	Matched *fkh* TF	C	N.c	O	P
Motif 40	*lin-31*	M00665_2.00	*lin-31* (*C. elegans*)	0.737	0.676	882	780
Motif 31	*fkh-7*	M03753_2.00	FoxP (*D. melanogaster*)	0.731	0.67	1110	983
Motif 37	*fkh-2, fkh-6, unc-130*	M00258_2.00	FoxC1 (*H. sapiens*)	0.744	0.546	653	577
Motif 39	*pha-4*	M08201_2.00	*pha-4* (*C. elegans*)	0.744	0.541	701	640
Motif 23	*fkh-2, fkh-8*	M03027_2.00	FoxI1 (*H. sapiens*)	0.776	0.502	525	400
Motif 30	*fkh-2, fkh-6, unc-130*	M00256_2.00	FoxC1 (*H. sapiens*)	0.826	0.496	1307	1130
Motif 35	*fkh-7*	M09086_2.00	FoxP1 (*H. sapiens*)	0.702	0.432	437	423
Motif 36	*fkh-2, unc-130*	M03045_2.00	FoxD3 (*H. sapiens*)	0.739	0.431	513	486
Motif 9	*fkh-7*	M06202_2.00	FoxP (*D. melanogaster*)	0.709	0.413	358	319
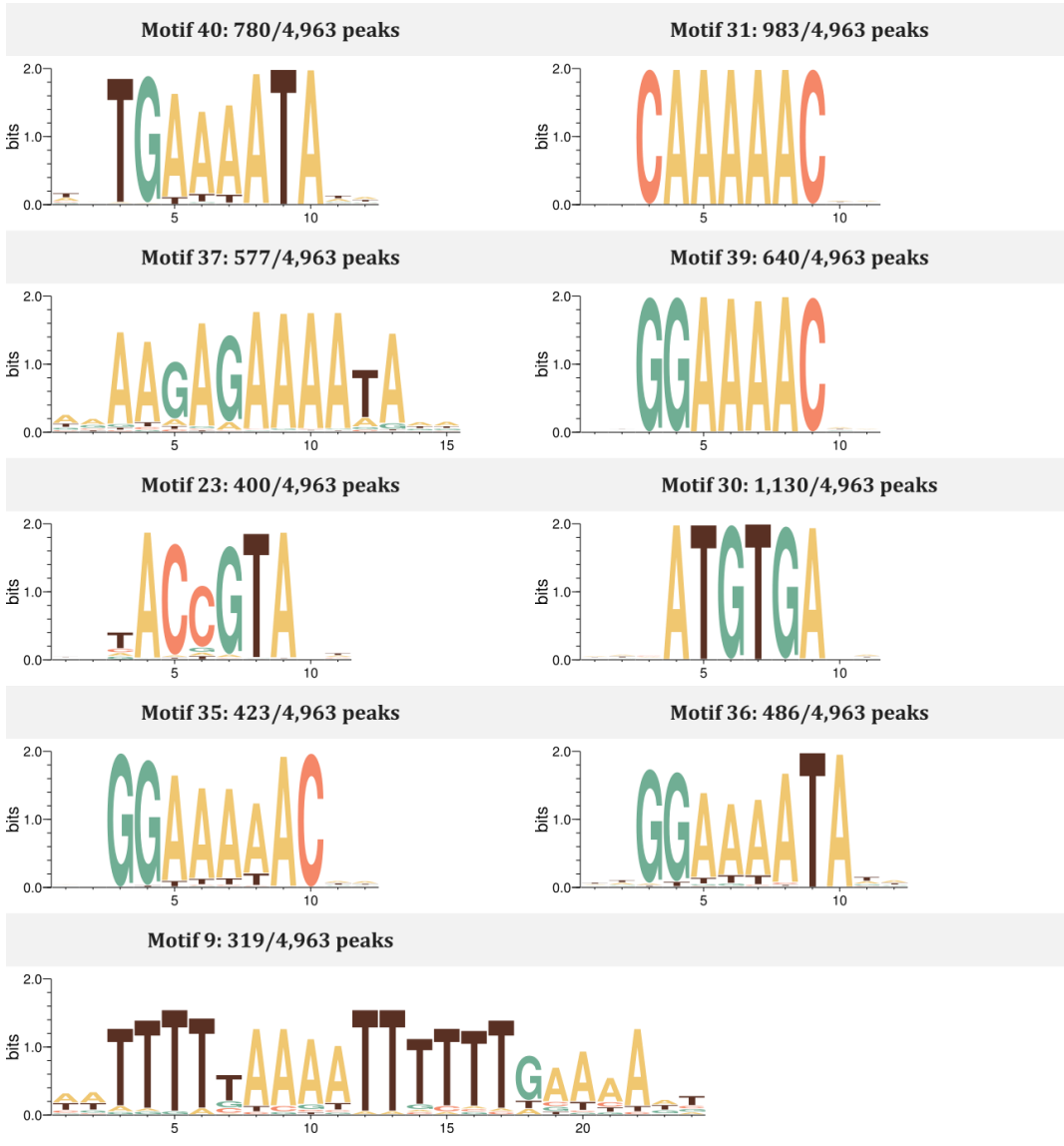
⬆ Figure R.29. Sequence logos for PWMs generated from RSAT peak-motif *de novo* discovery analysis that match FKH-related binding sites

Subsets of sequences from FKH-8 ChIP-seq peaks generate motifs that match with known binding sites for members of the FKH family. Motifs 40, 31, 37, 39, 35 and 36 were discovered through the *dyads* algorithm, Motifs 23 and 30 were discovered through the *oligos* algorithm whereas Motif 9 was found through the *positions* algorithm. Motif headers indicate the number of peaks harbouring at least one site matching such motif over the total number of FKH-8 ChIP-seq peaks. From top to bottom and left to right, sequence logos appear ordered by decreasing normalized correlation value (see **Table R.9**).
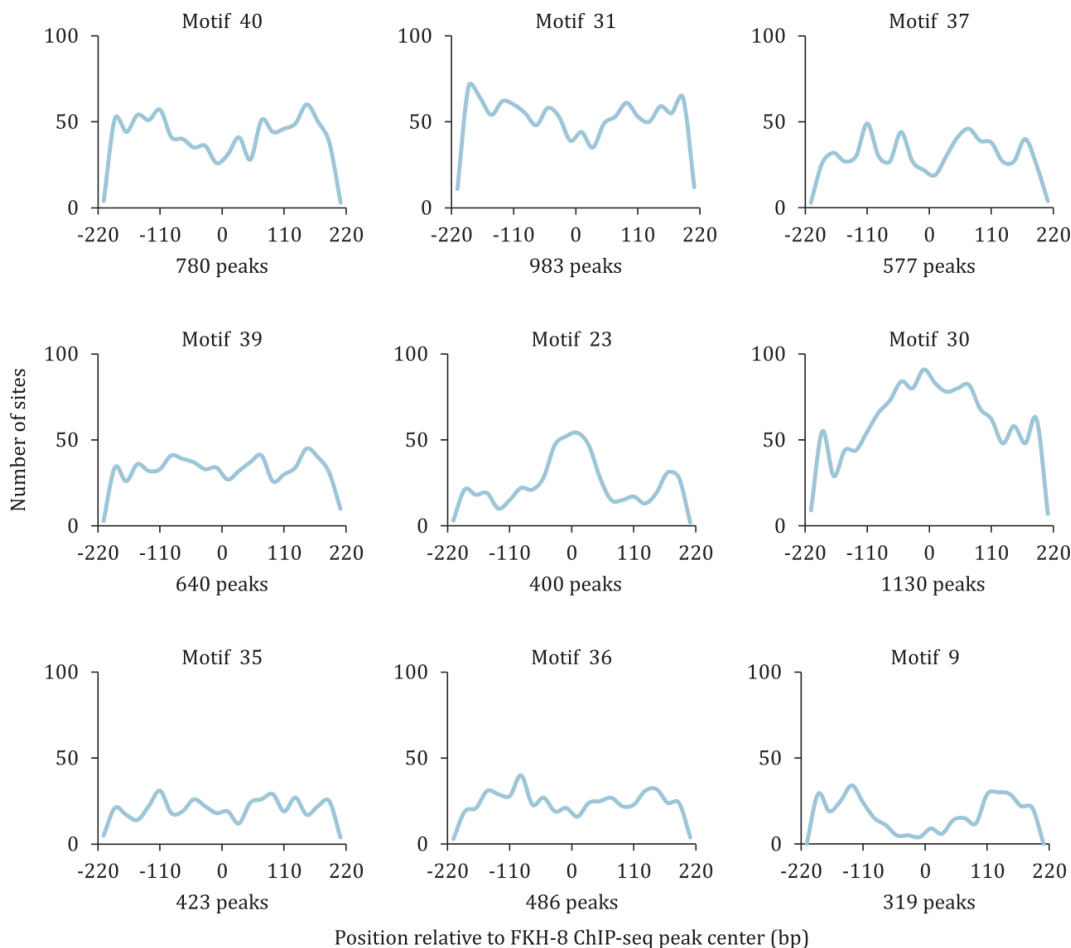
⌃ Figure R.30. Profile distribution of FKH-related motifs generated from RSAT peak-motif *de novo* discovery analysis.
Distribution for the position of FKH-related motifs along the sequences of their targeted peaks is shown in blue. For each motif, the number of targeted peaks is indicated. From top to bottom and left to right, distribution profiles for each motif appear ordered by decreasing normalized correlation value (see **Table R.9**).

Since structural ciliary genes contain *fkh-8* ChIP-seq peaks and, as already published and found in our motif analysis, these genes are also enriched for X-box sites, we aimed to identify if FKH-8 bound regions matched with those containing X-boxes. To this aim, we performed a local motif enrichment analysis through the CentriMo tool from the MEME Suite (Bailey and MacHanick 2012), employing the same motif databases previously used (*C. elegans* CIS-BP 1.02 and JASPAR CORE non-redundant 2018). Through this approach we discovered that 1.467 out of 4.963 individual sequences bound by FKH-8 were harbouring at least one matching site for an RFX transcription factor,

hence accounting for roughly 29% of the total sequences in the set. Much interestingly, a statistically significant enrichment for those motifs was found at the central region of the peaks (see **Figure R.31**), even when the algorithm was allowed to retrieve local motif enrichments at any point in the sequence and not only at its centre (which is the default mode for the CentriMo tool). Since motifs distributed peaking at central regions of ChIP-seq peak sequences often match with direct DNA binding, these results further suggest the co-binding of *fkh-8* and *daf-19* within regulatory regions of ciliome genes.
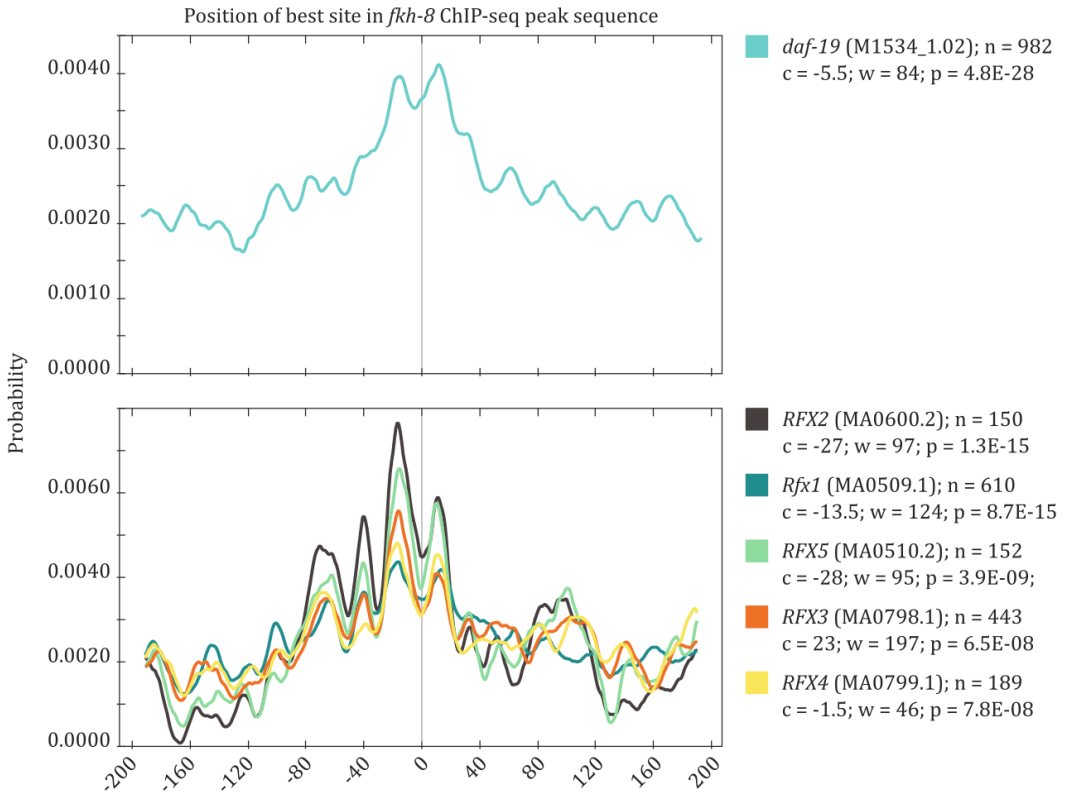
▲ **Figure R.31. 29% of *fkh-8* ChIP-seq peaks harbour RFX binding sites in their sequences.**
daf-19 (upper panel) and RFX (bottom panel) motif distribution within the sequences of *fkh-8* ChIP-seq peaks. Origin for representation of binding profiles was established at the centre of the sequences. Peaks were found to typically span for 416 base pairs. Legend shows here: name of the transcription factor, ID of the PWM the database (between brackets), number of peaks in which at least one binding site was found (n), location of the centre of the most enriched region (c), width of the most enriched central region (w) and statistical significance for the enrichment of the motif (p). Top panel results from the interrogation of the *C. elegans* CIS-BP 1.02 database, whereas bottom panel is retrieved when the JASPAR CORE non-redundant 2018 database is used. Notice that the CentriMo algorithm only computes the highest scored match for a given motif in a given sequence when creating the probability curve, that being denoted as the best site in sequence.

In summary, FKH-8 genomic binding analysis through ontology enrichments clearly expose that *fkh-8* preferentially binds ciliome genes. In addition, the presence of FKH-related motifs clustered at the centre of FKH-8 ChIP-seq peaks suggests direct binding to the regulatory regions of at least a subset of those genes. Moreover, the presence of X-boxes within the sequences of *fkh-8* bound genes suggests that *daf-19* and *fkh-8* may co-operate in the transcriptional regulation of the sensory ciliome in *C. elegans*.

4. RNA-seq profiling of *fkh-8* and *daf-19* single and double mutants.

Downregulated genes in young adult *fkh-8(tm292)* mutants do not enrich ciliary terms or functions.

To further expand our knowledge on the transcriptional regulatory role of *fkh-8*, we decided to perform whole RNA sequencing in the available allele *fkh-8(tm292)*, a deletion allele that removes 561 base pairs eliminating or affecting 3 out of the 8 exons of *fkh-8*

(see **Figure R.32**). From these RNA-seq experiments, performed on young adult animals, we retrieved expression data from 23,616 transcribed elements. Most of these elements belonged to the protein-coding genes category (roughly 80% of the total); however, other types of elements were also found, such as non-coding or piwi-associated RNAs (see **Figure R.33**).
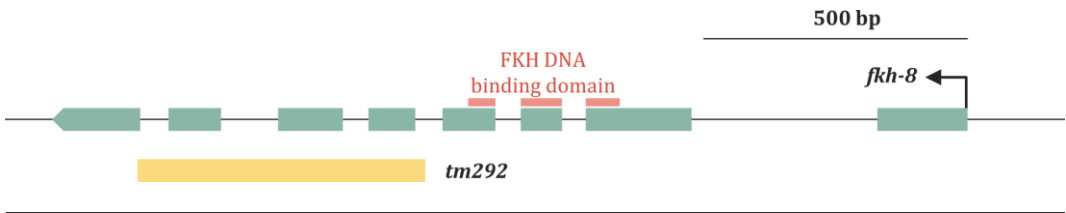
▲ **Figure R.32.** *fkh-8(tm292)* **is a partial deletion allele for** *fkh-8.*
Turquoise boxes represent here all 8 exons coding for FKH-8. Upper pink boxes denote the regions in which the FKH DNA binding domain is encoded. A yellow bar represents the width of the *tm292* deletion. Notice that, in this mutant allele, the FKH DNA binding domain remains unaltered.

Next, we performed differential expression analysis comparing the retrieved data from *fkh-8(tm292)* mutants and N2 isogenic wild type animals. Unexpectedly, when considering the threshold imposed by the false discovery rate (FDR) method correcting for multiple comparisons, we found that only 2 out of the 23,616 transcribed elements were significantly deregulated between both strains (see **Figure R.34.A**). Those were the cases of the astacin-like metalloprotease TOH-1, whose expression resulted downregulated in *tm292* mutants, and the ortholog of human NAA25 (N-alpha-acetyltransferase 25) CRA-1, in which a mild yet significant upregulation was found. We hypothesized that

transcriptional regulatory effects of *fkh-8* – a gene that we saw mainly expressed within the 60 ciliated neurons of the worm – could be diluted when analysing RNA-seq data retrieved from a bulk pool of cells obtained from whole animals. Accordingly, we analysed to which extend the raw comparison between expression values for all 23,616 elements from both datasets could be rendering false positives when not attending to the FDR threshold. Interestingly, for the 4,070 transcribed elements that were differentially expressed in the raw comparison between *fkh-8(tm292)* mutants and wild type animals, the probability of false positives was always kept below 29% (see **Figure R.34B**).

Acknowledging for those limitations, we proceeded to analyse the differences in the raw comparison between expression values for both strains.
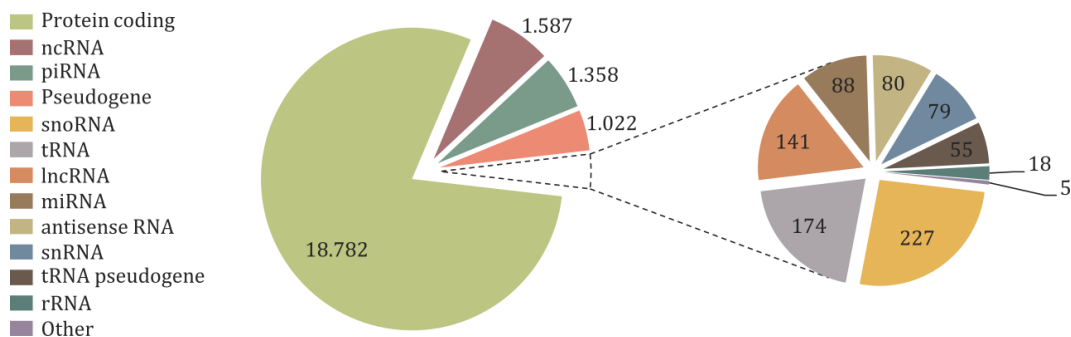
▲ **Figure R.33. Classification of the total transcribed elements retrieved from RNA-seq experiments.**

Pie charts show the number and type of transcribed elements detected from RNA-seq experiments performed on wild type, single *fkh-8(tm292)* mutants, doble *daf-12(sa204)* and *daf-19(m86)* mutants, and triple *fkh-8(tm292)*, *daf-12(sa204)* and *daf-19(m86)* mutants. Most elements correspond to protein coding genes (79.53% of the total). We also detected expression of other elements such as non-coding RNAs (ncRNA, 6.72%), piwi-interacting RNAs (piRNA, 5.75%), pseudogenes (4.33% %), small nucleolar RNAs (snoRNA, 0.96%), transfer RNAs (tRNAs, 0.74%), long non-coding RNAs (lncRNA, 0.60%), microRNAs (miRNA, 0.37%), antisense RNAs (0.34%), small nuclear RNAs (snRNA, 0.33%), transfer RNAs pseudogenes (0.23%) and ribosomal RNAs (rRNA, 0.08%, leftovers not removed in the RNA extraction procedure).

This way, we found 2,154 transcribed elements significantly downregulated in *fkh-8(tm292)* mutants when compared against wild type animals. To gain some insights into the nature of these downregulated elements we started by performing a Gene Ontology analysis using the Gene Ontology enRIchment anaLysis and visuaLizAtion tool GORILLA. GO Cellular Component analysis showed enrichment of extracellular space components as well as the peroxisome. GO enrichment for Molecular Function terms identified three major branches: the first one leading to binding of galactosidase; the second one related to transmembrane transport of amino acids and, finally, a huge enrichment of kinase and phosphatase activities.

However, the clearest gene enrichment was found for GO terms related to the Biological Process class (see **Figure R.35**), when a broad downregulation of metabolic processes related to de- and phosphorylation was found. Another highly enriched process was the immune response, as 89 of the 219 genes associated to this category were downregulated in *fkh-8(tm292)* mutants. In addition, GO terms for both defence responses to Gram-negative and Gram-positive bacteria were also enriched, reinforcing the role of *fkh-8* in the regulation of the innate immune response. Statistical significance and number of related genes for the selected GO terms mentioned in this paragraph can be seen in **Figure R.36**. Of note, no obvious enrichment

for GO terms related to ciliary function or sensory neurons was found for the downregulated genes of *fkh-8(tm292)* mutants.
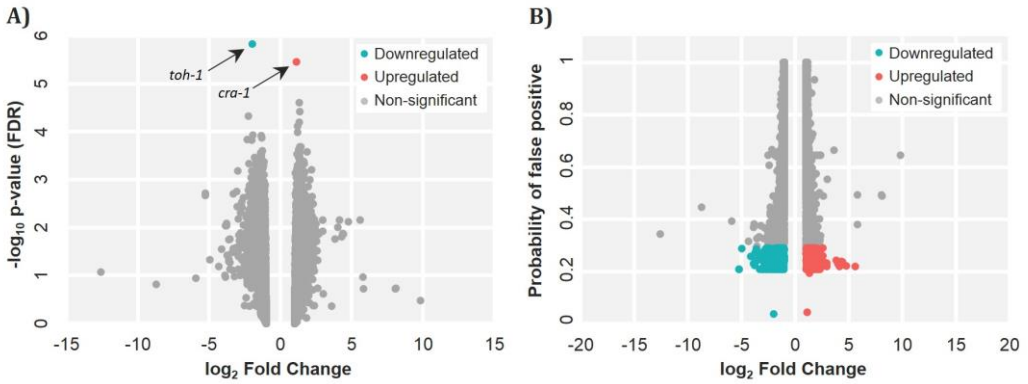
⬆ **Figure R.34. Differential expression analysis between *fkh-8(tm292)* mutants and wild type animals.**
A) Volcano plot for the differential expression analysis comparing the expression values of 23,616 transcribed elements between *fkh-8(tm292)* mutants and the isogenic wild type control strain. Those elements (represented by dots) that failed to pass the threshold imposed by the false discovery rate (FDR) method correcting for multiple comparisons appear in grey. Notice that only 2 genes (*toh-1* in turquoise and *cra-1* in pink) are located below the FDR threshold, hence being significantly deregulated in *fkh-8(tm292)* mutants. **B)** Probability of false positive for the differential expression values of 23,616 transcribed elements between *fkh-8(tm292)* mutants and the isogenic wild type control strain. For all 4,070 transcribed elements in which a differential expression was found in the raw comparison between the values of both strains, the probability of false positives remains below 29%.

An analogous analysis was performed with the 1,916 upregulated elements that we found in the raw comparison between *fkh-8(tm292)* mutants and wild type animals. This set of genes showed enrichment in up to 35 terms from the GO Cellular Component category, thus being located in several parts of the cell. Analysis of the GO Molecular Function class showed enrichment of 55 terms mostly related to catalytic functions. Finally, GO analysis of the Biological Process category enriched 144 different terms emerging from nine major branches ranging from developmental, cellular of metabolic processes. As a whole, upregulated genes in *fkh-8(tm292)* mutants do not seem to constitute a functionally coherent set of genes. A summary of the most significantly enriched GO terms referred in this paragraph, as well as statistical significance and number of genes per category appear gathered in **Figure R.37**.
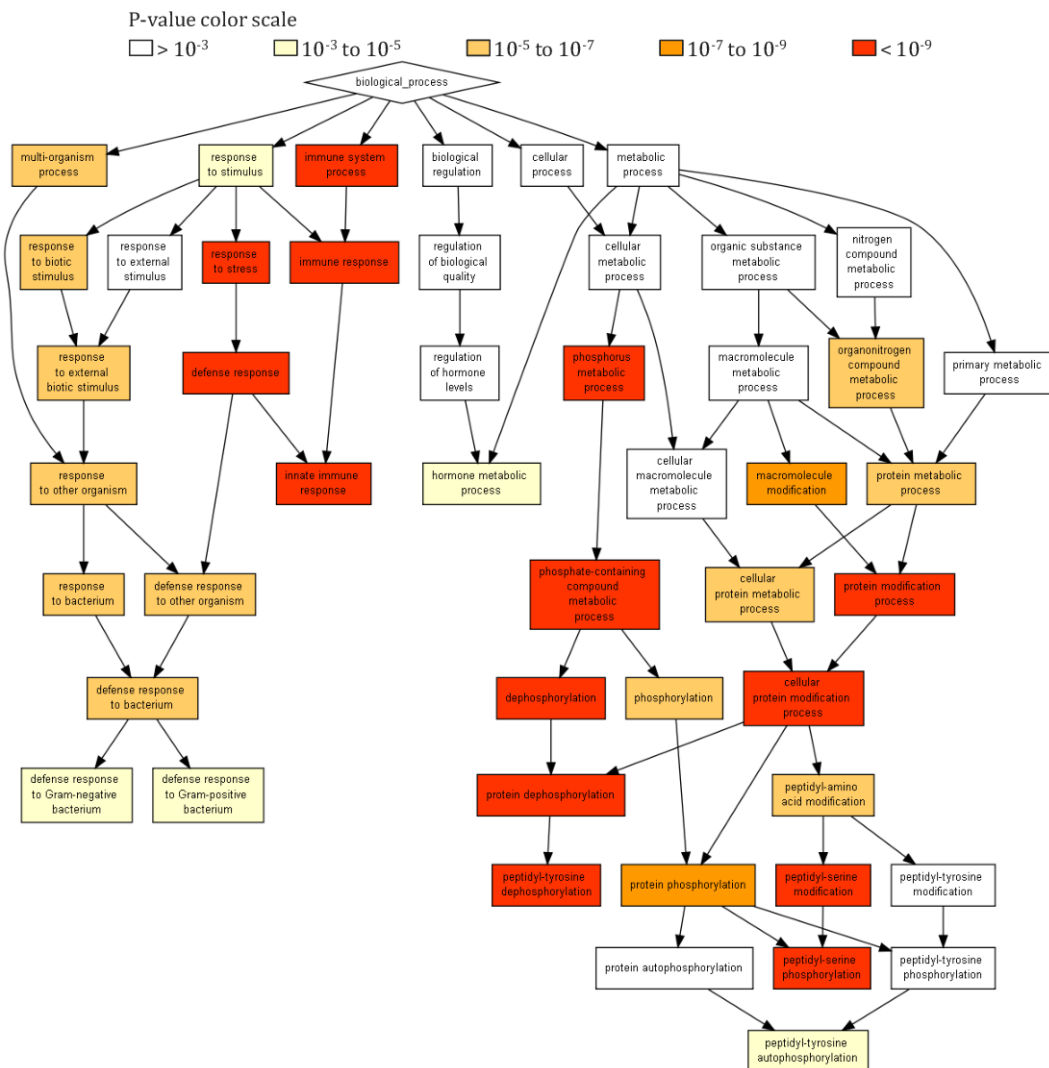
⌃ Figure R.35. Hierarchization for the GO biological process terms enriched by downregulated genes in *fkh-8(tm292)* mutants show enrichment for immune response but not for cilia related terms.

Coloured boxes represent here the statistical significance for the enrichment of terms retrieved through the GOrilla two unranked list of genes analysis. Arrows indicate the hierarchization for downstream processes. Results arise from the information provided by 1,320 of the 2,154 genes used in this analysis. Background control consists of 23,616 *C. elegans* genes, for which 12,282 were related to Gene Ontology terms. P-values meet the criterium for multiple comparisons imposed by the Benjamini-Hochberg method.
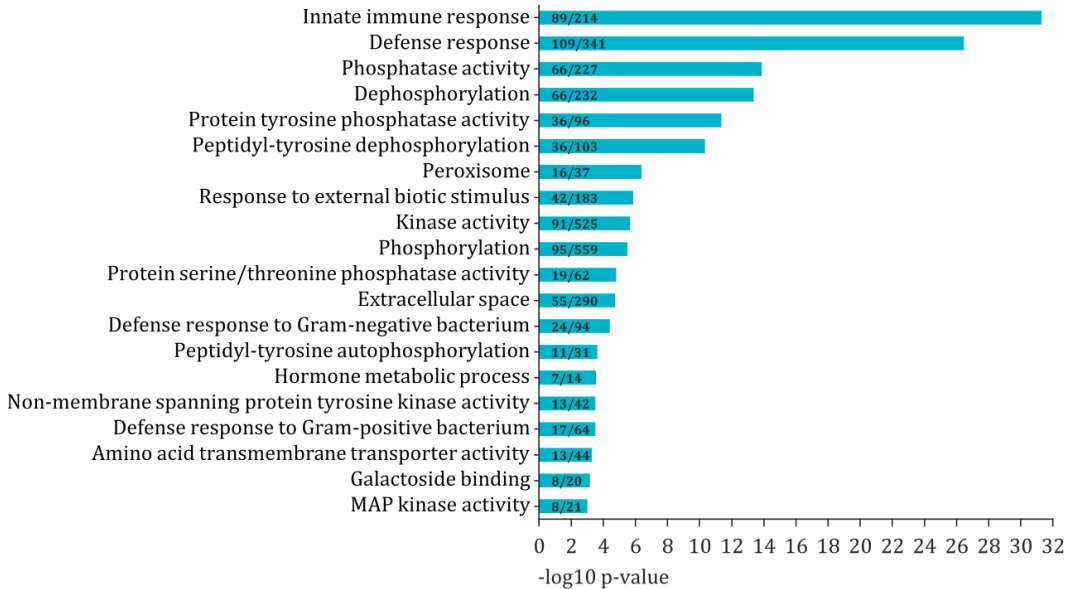
Figure R.36. Summary of representative GO terms (cellular component, molecular function or biological process) enriched by the downregulated genes in *fkh-8(tm292)* mutants.
Coloured bars indicate the statistical significance for the enrichment of selected GO terms retrieved through the GOrilla two unranked list of genes analysis. Results arise from the information provided by 1,320 of the 2,154 genes used in this analysis. Background control consists of 23,616 *C. elegans* genes, for which 12,282 were related to Gene Ontology terms. For each term, number of associated genes found in the sample over the total number of genes associated with the term is indicated. P-values meet the criterium for multiple comparisons imposed by the Benjamini-Hochberg method.

Figure R.37. Summary of representative GO terms (cellular component, molecular function or biological process) enriched by the upregulated genes in *fkh-8(tm292)* mutants.
fkh-8(tm292) upregulated genes enrich for a diverse and unrelated set of GO terms. Coloured bars indicate here the statistical significance for the enrichment of selected GO terms retrieved through the GOrilla. Results arise from the information provided by 1,246 of the 1,916 genes used in this analysis. Background control consists of 23,616 *C. elegans* genes, for which 12,282 were related to Gene Ontology terms. For each term, number of associated genes found in the sample over the total number of genes associated with the term is indicated. P-values meet the criterium for multiple comparisons imposed by the Benjamini-Hochberg method.
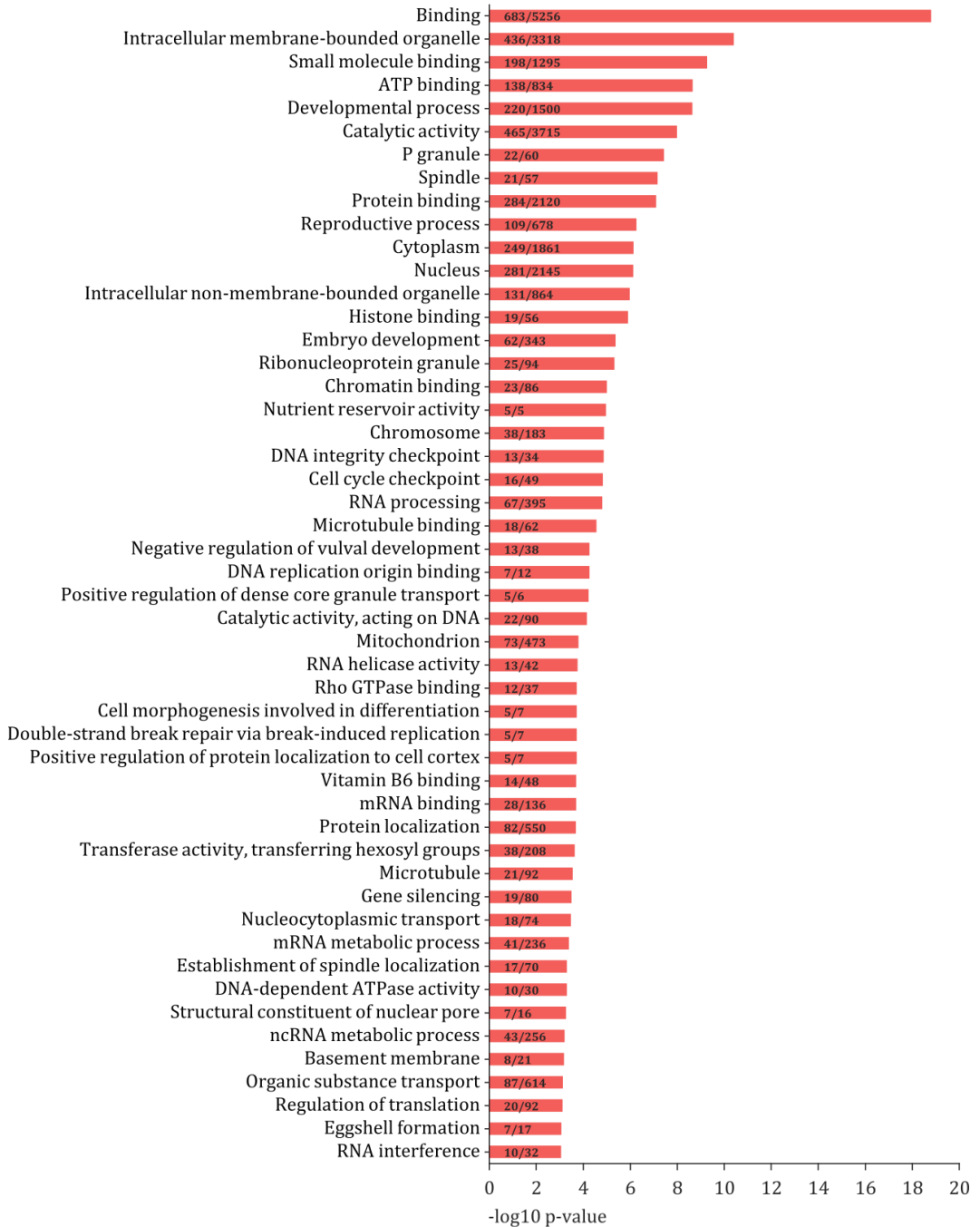
As differential expression analysis of *fkh-8(tm292)* mutants did not show enrichment in ciliary components or cilium-related functions, we decided to perform a more stringent GO analysis. To this aim, we selected only those genes that were most highly down or up-regulated by establishing a new cut-off in the absolute value of expression fold change equal or greater than 1.5 in a base-2 logarithmic scale. Following this criterion, 588 genes were downregulated and 434 were upregulated in *fkh-8(tm292)* mutants compared to wild type worms. For this new analysis, the WormBase Gene Set Enrichment Analysis tool was used. Gene enrichment of downregulated genes within the Anatomical category of the tool revealed terms such as intestine, head mesodermal cell, epithelial or muscular systems. Interestingly, sensory-related terms were also enriched, such as the ones referring to the ASE neuron, the thermosensory neurons or the amphid sensillum. Through this methodology, the GO Biological Process terms enriched showed association to immune response, similar to our previous enrichment analysis; however, other terms related to sensory neuron functions such as *dauer*-related terms or hypersensitivity to cadmium or neuropeptide signalling were also found. Thus, restricting our differentially expressed gene analysis to the most highly downregulated genes showed a slight enrichment on cilium-related functions or localization. Nevertheless, this link was still small. Analogous enrichment analysis for the most upregulated genes showed enrichment for non-neuronal tissues such as cephalic sheath cell, epithelial system and accessory cell as well as terms related to the reproductive system. Few genes were shown to enrich

phenotype-related categories. GO terms for this subset of upregulated genes unravelled enrichment for structural constituents of the cuticle, but also were shown to affect the defence response or the response to biotic stimulus. Statistical significance and number of related genes for the selected enrichment terms mentioned in this paragraph can be seen in **Figure R.38**.

Next, we wanted to test if genes directly bound by FKH-8 showed differential expression levels in *fkh-8(tm292)* mutants when compared to wild type worms. From our RNA-seq experiment, we retrieved expression data for 3,148 out of the 3,987 unique genes that we found associated to FKH-8 ChIP-seq peaks through the HOMER criterium (see preceding section). Most of these genes, roughly 83% of them, were not statistically deregulated between *fkh-8(tm292)* mutants and wild type worms (see **Figure R.39**). However, a small subset composed of 225 genes were downregulated in this *fkh-8* mutant background. Interestingly, this subset of downregulated genes was able to enrich the WormBase tissue category related to the ciliated ASE neuron (p-value = 9.90E-05). On the other hand, 309 FKH-8 ChIP-seq associated genes resulted upregulated in *fkh-8(tm292)* mutants when compared to wild type worms. A single WormBase tissue category was enriched by this subset of genes, namely that related to the germline (p-value = 6.00E-06). As a whole, from this analysis, we found that a small subset of putative direct targets of FKH-8 integrated by 534 genes (roughly a 17% of the total putative targets) were significantly deregulated in *fkh-8(tm292)* mutants. Additionally, downregulated genes could be related to the ciliated

ASE neurons.

A)

B)
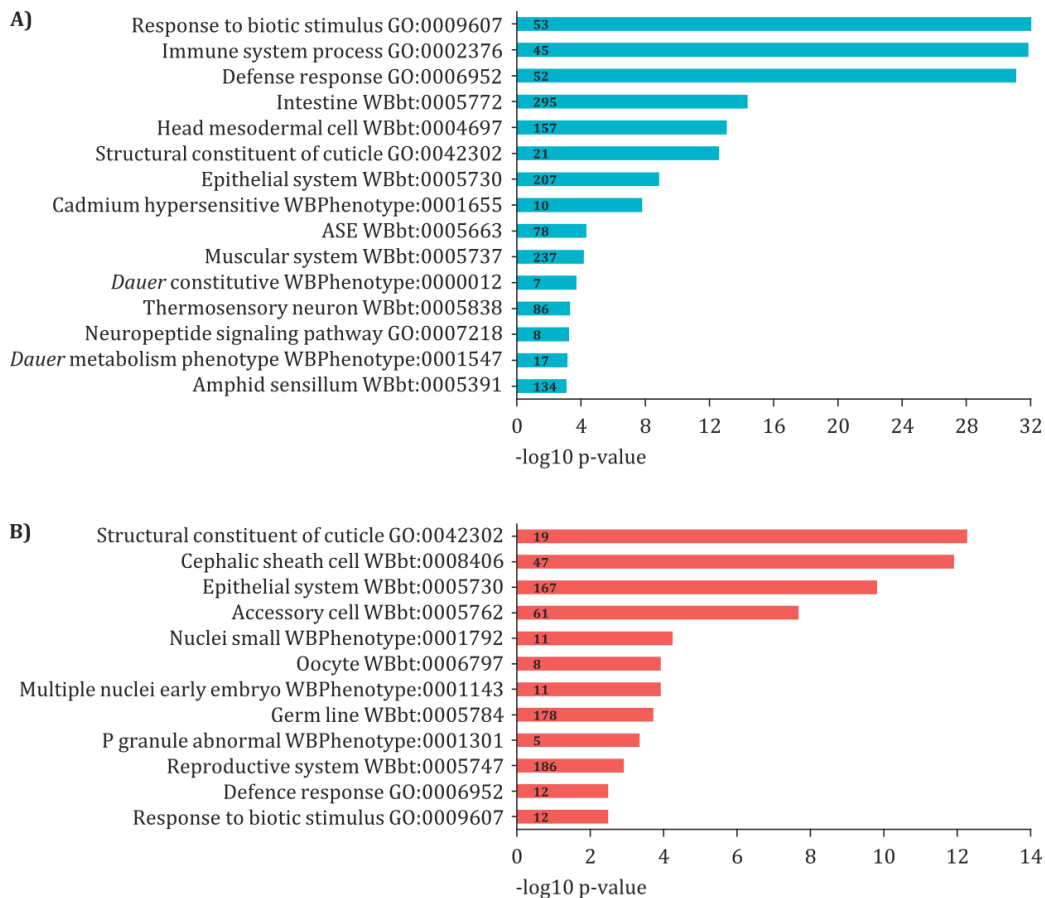
▲ **Figure R.38. WormBase Gene Set Enrichment Analysis of** *fkh-8(tm292)* **mutants uncovers additional GO terms related to sensory functions.**

In addition to Inmune response, downregulated *fkh-8(tm292)* genes enrich GO categories related to ciliated sensory neurons, such as Response to biotic stimulus, Cadmium response, ASE, *dauer*, thermosensory neuron, and Amphid sensillum. Coloured bars indicate here the statistical significance for the enrichment of selected WormBase and GO terms for both downregulated **(A)** and upregulated **(B)** genes retrieved through the WormBase Gene Set Enrichment Analysis tool. For each term, number of associated genes is indicated. P-values gathered here meet the criterium for multiple comparisons when applying the Benjamini-Hochberg method.

Consequently, we also wanted to test possible changes in expression levels for the genes composing our curated ciliome gene list in the *fkh-8(tm292)* mutant background. Of notice, and further exposing the limitations of RNA-seq experiments performed with a bulk pool of cells obtained from whole animals, no expression data were retrieved for 5 of the genes in that list. Those were the cases of the transmembrane protein

and receptors *tmem-17* (panciliary expressed), *srbc-64* (expressed in the ASK neurons), *srg-36* (expressed in ASI neurons), *str-2* (expressed in the ASI and AWC) and *str-112* (expressed in the AWA neurons). For the remaining 158 curated cilia genes for which expression data were retrieved, 151 were not statistically deregulated between *fkh-8(tm292)* mutants and wild type worms; 2 were significantly upregulated (*gcy-23* and *sulp-2*) and 5 were significantly downregulated (*dcar-1*, *mks-1*, *osm-5*, *pitp-1* and *tax-6*) (see **Figure R.40**). Remarkably, 79 out of the 158 genes for which expression data were retrieved were also putative direct targets of FKH-8, as assessed by the presence of FKH-8 ChIP-seq peaks within 2,000 base pairs upstream and 1,000 base pair downstream their translational start site. However, of those 79 genes, only 2 (*osm-5* and *tax-6*) were significantly downregulated.
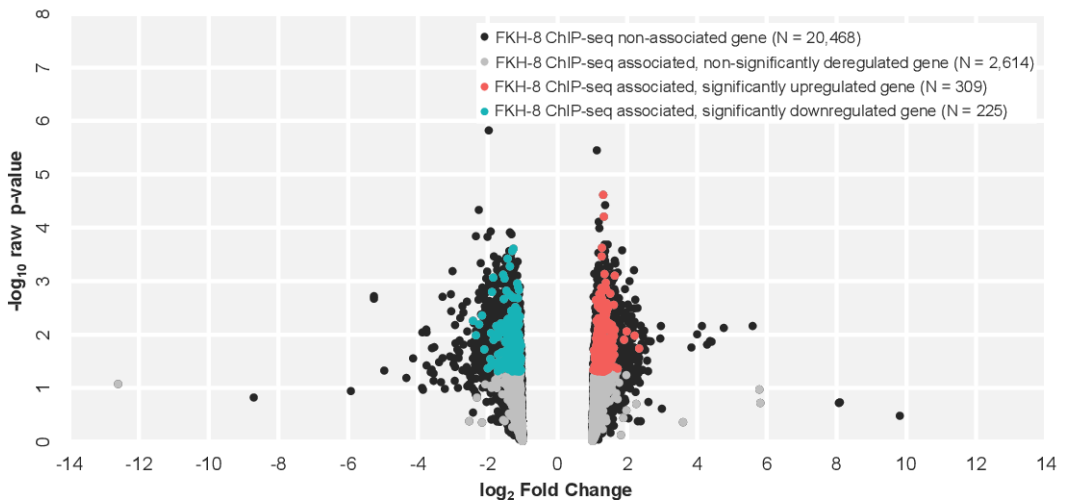
⌖ Figure R.39. Subsets of putative direct targets of FKH-8 are deregulated in young adult *fkh-8(tm292)* mutants.

Volcano plot for the differential expression analysis comparing the expression values of 23,616 transcribed elements between *fkh-8(tm292)* mutants and the isogenic wild type control strain. Genes not associated to FKH-8 ChIP-seq peaks appear in black. Genes with associated FKH-8 ChIP-seq peaks with no differential expression in *fkh-8(tm292)* mutants are represented in grey. Turquoise dots indicate genes targeted by FKH-8 significantly downregulated in *fkh-8(tm292)* mutants whereas pink dots portray analogous upregulated genes. Note that statistical significance has been established not accounting for the threshold imposed by the false discovery rate (FDR) method. However, probability of false positive for each deregulated element in this graph remains below 29%.

Finally, to further deepen on the transcriptional regulatory role of *fkh-8* over the ciliated features of *C. elegans*, we benefited from embryonic single-cell RNA-seq data by (Packer et al. 2019) to generate a yet more extensive, curated list of panciliary expressed genes, as determined by the authors. We could retrieve expression data for 348 out of

the 354 panciliary genes we selected. As expected by preceding results, most of those genes (namely, 311, roughly 89% of the total) were not statistically deregulated between *fkh-8(tm292)* mutants and wild type worms (see **Figure R.41**). On this occasion, a small subset composed of 22 genes resulted significantly downregulated in this *fkh-8* mutant background. On the other hand, 15 of these panciliary genes were found significantly upregulated in *fkh-8(tm292)* mutants when compared to wild type worms.

▲ Figure R.40. Most curated cilium genes are not deregulated in young adult *fkh-8(tm292)* mutants.

Volcano plot for the differential expression analysis comparing the expression values of 23,616 transcribed elements between *fkh-8(tm292)* mutants and the isogenic wild type control strain. Genes absent from our curated cilium gene list appear in black. Curated cilium genes for which no differential expression in *fkh-8(tm292)* mutants was found when compared to wild type worms are represented in grey. Turquoise dots indicate curated cilium genes significantly downregulated in *fkh-8(tm292)* mutants whereas pink dots portray analogous upregulated genes. Note that statistical significance has been established not accounting for the threshold imposed by the false discovery rate (FDR) method. However, probability of false positive for each deregulated element in this graph remains below 29%.

As a whole, differential expression analysis between *fkh-8(tm292)* mutants and wild type animals did not show clear differential gene expression directly related to ciliary genes, neuronal ciliary functions or FKH-8 bound associated genes. *fkh-8(tm292)* mutants seemed to have an impairment in the innate immune response, with de-regulated genes implicated in the detection and reaction to environmental stimuli as well as a noteworthy upregulation of genes related to reproductive structures or processes. As *fkh-8* is mainly expressed within the ciliated sensory neurons of *C. elegans*, we conclude this impact in gene expression is likely to reflect non-cell-autonomous effects induced by targets of *fkh-8* yet to be identified in ciliated neurons. However, our lack of cellular

resolution in this analysis precludes us from making strong statements regarding *fkh-8* gene control.
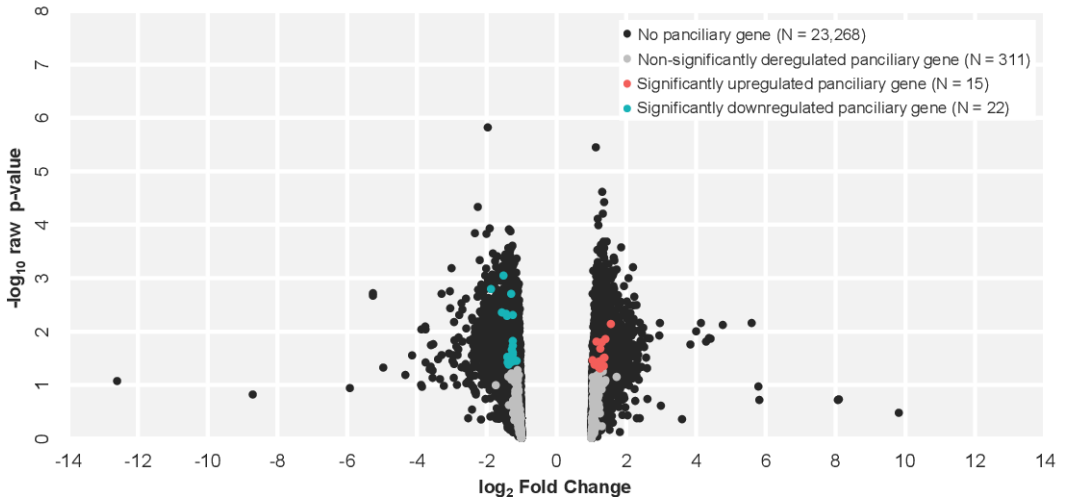
⬏ Figure R.41. Few panciliary genes are deregulated in young adult *fkh-8(tm292)* mutants.
Volcano plot for the differential expression analysis comparing the expression values of 23,616 transcribed elements between *fkh-8(tm292)* mutants and the isogenic wild type control strain. Genes not present in a curated gene list composed by panciliary expressed genes taken from (Packer et al. 2019) appear in black. Curated panciliary expressed genes for which no differential expression in *fkh-8(tm292)* mutants was found when compared to wild type worms are represented in grey. Turquoise dots indicate curated panciliary genes significantly downregulated in *fkh-8(tm292)* mutants whereas pink dots portray analogous upregulated genes. Note that statistical significance has been established not accounting for the threshold imposed by the false discovery rate (FDR) method. However, probability of false positive for each deregulated element in this graph remains below 29%.

Downregulated genes in double *daf-12* and *daf-19* young adult mutants relate to neuronal processes.

The lack of gene expression defects in *fkh-8(tm292)* mutants contrasted with FKH-8 genomic binding, which we found preferentially associated with genes expressed in cilia. We hypothesized that the lack of FKH-8 in ciliated neurons could be compensated by *daf-19*, thus precluding us to assess *fkh-8* effects in *fkh-8(tm292)* mutants. To test this hypothesis, we expanded our RNA-seq analysis comparing double *daf-19/daf-12* mutants against wild type worms.

The impact of the double null mutation of *daf-12* and *daf-19* in the regulation of the young adult transcriptome was evidenced when roughly 40% of the transcribed elements retrieved from our RNA-seq experiment were shown to be significantly deregulated when compared to their expression values in wild type animals (see **Figure R.42**). Enrichments for neuronal and sensory-related categories were retrieved from the WormBase tissue ontology when analysing the set of 4,088 downregulated genes.

Specifically, terms for the PVD interneurons and the outer labial sensillum were significantly highly enriched, gathering up to 1,085 and 1,090 genes, respectively. Other neuronal-related terms included dorsal and lateral nerve cords as well as the nerve ring. Regarding sensory-related categories, enrichments for terms such as cephalic sensillum or touch receptor neuron were also found. Impairment of the muscular system was represented through several terms, such as head mesodermal cell or the anal depressor and uterine muscles. Numerous phenotypes were associated with subsets of these downregulated genes within the WormBase phenotype ontology. Larval development was represented through the molt variant term. Ailment in environmental stimulus response manifested by enriching avoidance of bacterial lawn and nicotine hypersensitive terms. Most significant term within the Gene Ontology, which results are also retrieved directly from the WormBase Gene Set Enrichment Analysis tool, also referred to the structural constituents of the cuticle. Within the GO Molecular Function category, several terms related to the binding activity were enriched, such as nucleoside phosphate, purine nucleotide, ribonucleotide or small GTPase binding. Within the GO Biological Process category, the term for biological adhesion showed the biggest enriching significance. Several neuronal-related terms were also enriched in this ontology, with neuron development, neurogenesis, neuron projection guidance or regulation of neuron differentiation being among them. Finally, immune system process, defence response and response to biotic stimulus terms were also found enriched. Detailed information about the ontology terms mentioned in this paragraph appears collected in **Table R.10**.
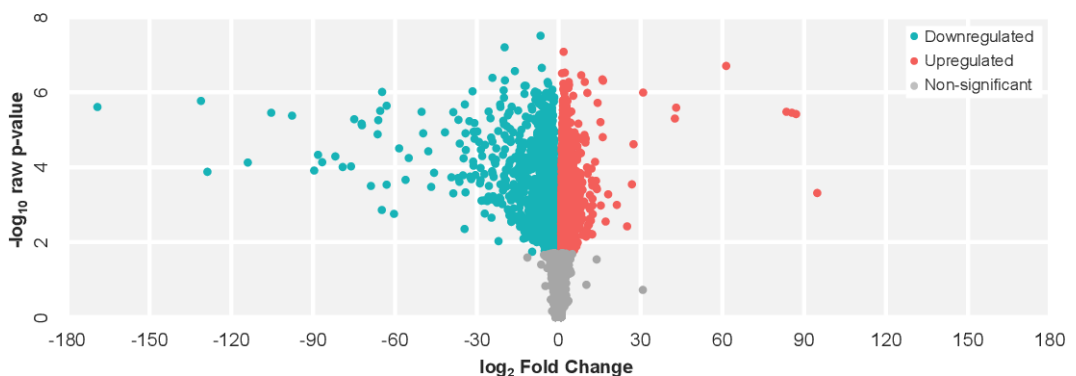
▲ **Figure R.42. Null *daf-19/daf-12* mutation greatly impacts gene expression in the young adult transcriptome of *C. elegans*.**
Volcano plot for the differential expression analysis comparing the expression values of 23,616 transcribed elements between double *daf-19(m86)II*; *daf-12(sa204)X* null mutants and the wild type control strain. Those elements (represented by dots) that failed to pass the threshold imposed by the false discovery rate (FDR) method correcting for multiple comparisons appear in grey (N = 14,086). Significantly downregulated genes (N = 4,088) appear in turquoise whereas significantly upregulated genes (N = 5,442) are depicted in pink.

▼ Table R.10. Ontology terms enriched in the set of downregulated genes in *daf-12(sa204)*; *daf-19(m86)* mutants when compared to wild type worms.

WormBase Gene Set Enrichment Analysis was performed upon the set of 4,088 downregulated genes retrieved in the differential expression analysis between double null mutant animals for *daf-12* and *daf-19* against wild type worms. Columns gather the number of expected and observed genes per category as well as the corresponding fold change enrichment (F.C.) achieve for each term. P-values for term enrichment are shown, all meeting the criterium for multiple comparisons when applying the Benjamini-Hochberg method (Q-value ≤ 0.05).

Term	Expected	Observed	F.C.	P-value	Q-value
PVD	770	1,085	1.4	4.20E-39	1.20E-36
Outer labial sensillum	780	1,090	1.4	1.50E-36	2.10E-34
Head mesodermal cell	840	1,023	1.2	3.80E-13	2.70E-11
Anal depressor muscle	130	190	1.5	6.50E-11	2.30E-09
Nerve ring	240	324	1.4	1.30E-10	4.10E-09
Touch receptor neuron	520	645	1.2	4.30E-10	1.20E-08
Dorsal nerve cord	160	229	1.4	4.50E-10	1.20E-08
Lateral nerve cord	110	159	1.4	5.10E-08	1.20E-06
Uterine muscle	36	63	1.8	1.70E-07	3.80E-06
Cephalic sensillum	130	155	1.2	8.50E-03	4.30E-02
Molt variant	86	131	1.5	6.00E-09	1.40E-06
Avoids bacterial lawn	140	200	1.4	7.80E-09	1.40E-06
Nicotine hypersensitive	15	32	2.1	4.10E-07	3.20E-05
Structural constituent of cuticle	45	112	2.5	3.00E-28	3.60E-26
Biological adhesion	27	60	2.2	7.10E-13	2.80E-11
Neuron development	50	91	1.8	5.00E-11	1.20E-09
Neurogenesis	82	134	1.6	5.90E-11	1.20E-09
Neuron projection guidance	36	67	1.9	2.00E-09	2.40E-08
Immune system process	68	110	1.6	4.30E-09	4.30E-08
Regulation of neuron differentiation	33	62	1.9	1.10E-08	9.80E-08
Nucleoside phosphate binding	340	425	1.2	2.00E-07	1.50E-06
Purine nucleotide binding	310	377	1.2	1.00E-06	7.30E-06
Ribonucleotide binding	310	378	1.2	2.30E-06	1.50E-05
Small GTPase binding	27	48	1.8	2.50E-06	1.60E-05
Response to biotic stimulus	100	136	1.4	1.80E-05	8.50E-05
Defence response	100	132	1.3	9.60E-05	4.30E-04

Analogously, WormBase tissue enrichment analysis for the 5,442 upregulated elements showed an extraordinarily high enrichment for the terms germ line and reproductive system. Consequently, several phenotypes associated with reproduction or abnormal early development were also found. Interestingly, 761 genes enriched the term for thermosensory neuron. Within the Gene Ontology terms, the most enriched processes related to the metabolism of heterocycle, organic cyclic or cellular aromatic compounds were present. RNA splicing via transesterification reaction was also found affected. Detailed information about the ontology terms mentioned in this paragraph appear in **Table R.11**.

▼ **Table R.11. Ontology terms enriched in the set of upregulated genes in *daf-12(sa204)*; *daf-19(m86)* mutants compared to wild type worms.**
WormBase Gene Set Enrichment Analysis was performed upon the set of 5442 upregulated genes retrieved in the differential expression analysis between double null mutant animals for *daf-12* and *daf-19* against wild type worms. Columns gather the number of expected and observed genes per category as well as the corresponding enrichment fold change (F.C.) achieve for each term. P-values for term enrichment are shown, all meeting the criterium for multiple comparisons when applying the Benjamini-Hochberg method (Q-value ≤ 0.05).

Term	Expected	Observed	F.C.	P value	Q value
Germ line	2,000	3,284	1.6	2.40E-239	6.90E-237
Reproductive system	2,200	3,376	1.5	4.10E-190	5.90E-188
Thermosensory neuron	660	761	1.2	8.40E-07	2.40E-05
Aneuploidy	53	88	1.6	3.00E-08	1.90E-06
One cell arrest early embryo	17	35	2	7.00E-07	2.10E-05
Embryonic cell morphology variant	51	74	1.4	1.20E-04	1.40E-03
Heterocycle metabolic process	580	757	1.3	3.50E-18	2.10E-16
Organic cyclic compound metabolic process	590	769	1.3	1.30E-17	5.20E-16
Cellular aromatic compound metabolic process	580	753	1.3	3.00E-17	8.90E-16
RNA splicing via transesterification reactions	35	67	1.9	3.10E-10	6.10E-09

Altogether, lack of *daf-12* and *daf-19* showed a severe effect upon the transcriptional regulation of the young adult transcriptome, being traced, as excepted, to downregulation of neuronal and sensory-related features. In accordance with *fkh-8(tm292)* mutation, yet more exacerbated, a noteworthy upregulation of genes related to reproductive processes was also found.

Deregulated genes in double *daf-12* and *daf-19* young adult mutants with FKH-8 binding enrich ciliary features.

Despite the general impact that we could observe in neuronal and sensory processes

enriched by the downregulated genes of *daf-19/daf-12* null mutants, no clear differential expression could be directly related to ciliome genes or ciliary neuronal functions. This fact, which seemed to contradict the reported role of *daf-19* as the master regulator of the ciliogenesis in *C. elegans*, was actually in agreement with recently published data (De Stasio et al. 2018); hence further supporting the notion that cilia maintenance is less dependent on DAF-19 regulation in the adult stage. To further deepen this issue, we decided to analyse the differential expression values of the panciliary expressed genes selected from (Packer et al. 2019) between *daf-19/daf-12* null mutants and wild type worms. As expected by preceding results, the expression values for most of those genes (namely, 246 out of the 348 for which expression data were retrieved, hence 70% of the total genes) were unaffected between these two strains. A small subset composed of 42 genes were significantly downregulated whereas for 60 of these panciliary genes a significant upregulation was found (see **Figure R.43**).

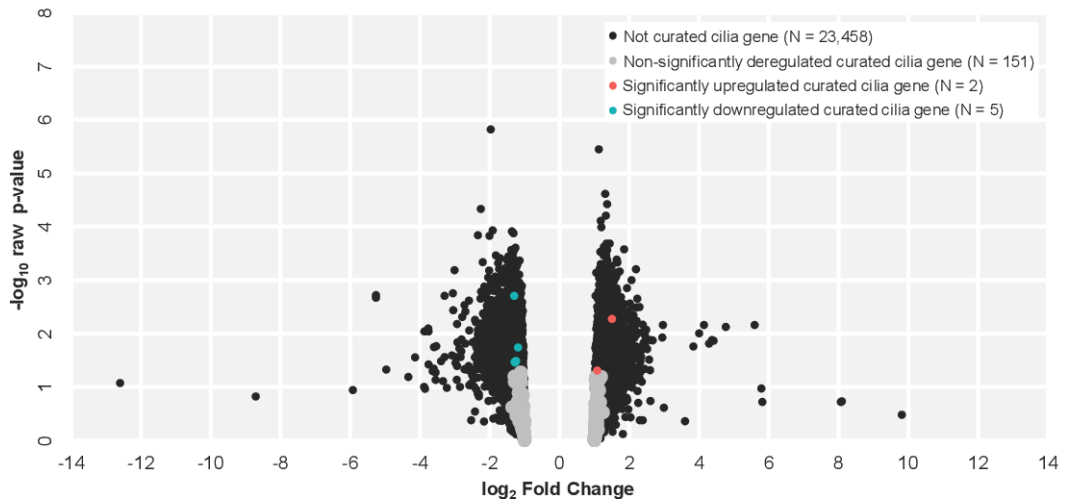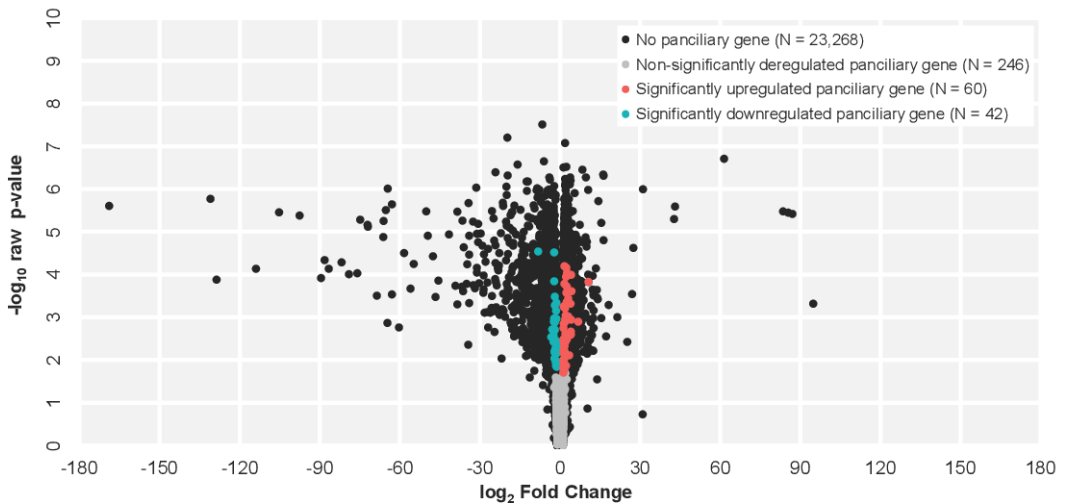▲ **Figure R.43. Few panciliary genes are deregulated in young adult *daf-19/daf-12* null mutants.** Volcano plot for the differential expression analysis comparing the expression values of 23,616 transcribed elements between *daf-19/daf-12* null mutants and the wild type control strain. Genes not present in a curated gene list composed of panciliary expressed genes taken from (Packer et al. 2019) appear in black. Curated panciliary expressed genes for which no differential expression in *daf-19/daf-12* null mutants was found when compared to wild type worms are represented in grey. Turquoise dots indicate curated panciliary genes significantly downregulated in *daf-19/daf-12* null mutants whereas pink dots portray analogous upregulated genes. Differential expression was assessed considering the threshold imposed by the false discovery rate method. Number of genes per category appears indicated in the legend.

Furthermore, the aforementioned results were also coincident with the analogous analysis we performed comparing differential gene expression between *fkh-8(tm292)*

mutants and wild type animals. However, as stated in previous sections, a clear relationship with the ciliome was established for putative direct targets of FKH-8 (see **Figures R.24 & R.25**). Moreover, known motif analysis of FKH-8 ChIP-seq peaks showed that 29% of the total peaks harbouring RFX-related binding sites (see **Figure R.31**). Accordingly, we wanted to analyse the differential expression between *daf-19/daf-12* null mutants and wild type animals for the 3,987 unique genes that we found associated with FKH-8 ChIP-seq peaks through the HOMER

criterium. For the 3,148 genes from which we retrieved expression data, 1,640 showed no differential expression between these two strains, whereas 698 resulted significantly downregulated and 810 resulted significantly upregulated (see **Figure R.44**). We wondered if the presence of a FKH-8 ChIP-seq peak was relating that subset of deregulated genes to particular ontology features. Thus, we performed a GOrilla enrichment analysis comparing those up- or downregulated subsets in *daf-19/daf-12* null mutants with FHK-8 binding.

▲ **Figure R.44. Subsets of putative direct targets of FKH-8 are deregulated in *daf-19/daf-12* null mutants.**
Volcano plot for the differential expression analysis comparing the expression values of 23,616 transcribed elements between *daf-19/daf-12* null mutants and the wild type control strain. Genes not associated with FKH-8 ChIP-seq peaks appear in black. Genes targeted by FKH-8 ChIP-seq peak in which no differential expression for *daf-19/daf-12* null mutants was found are represented in grey. Turquoise dots indicate genes targeted by FKH-8 significantly downregulated in *daf-19/daf-12* null mutants whereas pink dots portray the analogous upregulated genes. Differential expression was assessed considering the threshold imposed by the false discovery rate method. Number of genes per category appears indicated in the legend.

Surprisingly, up to 124 GO terms resulted enriched within the Biological Process

ontology by putative direct target of FKH-8 downregulated in *daf-19/daf-12* null

mutants. These terms could be tracked to 32 downstream categories emerging from several branches of regulation, localization or developmental processes. Of notice, cilium-related terms such as determination of adult lifespan or *dauer* larval development were among the more significantly enriched terms. Enriched Molecular Function terms were related to catalytic activity, finally leading to protein serine/threonine kinase activity, and

binding processed related to ATP, enzyme, and RNA binding. Finally, this subset of downregulated genes was shown to mainly enrich terms related to nuclear or axonal localization within the Cellular Component ontology. Statistical significance and number of related genes for the selected enrichment terms mentioned in this paragraph can be seen in **Figure R.45**.
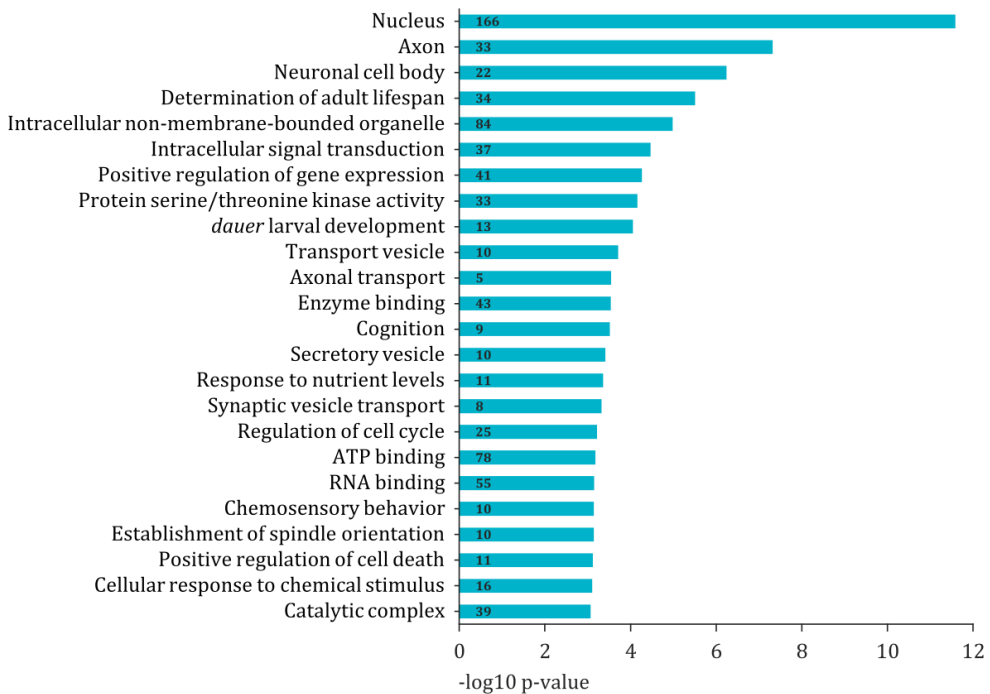
⬈ Figure R.45. Summary of representative GO terms (cellular component, molecular function or biological process) enriched by putative direct targets of FKH-8 downregulated in *daf-19/daf-12* null mutants.

Coloured bars indicate the statistical significance for the enrichment of selected GO terms retrieved through the GOrilla two unranked list of genes analysis. Results arise from the information provided by 554 of the 698 genes used in this analysis. Background control consists of 4,088 *C. elegans* genes significantly downregulated in *daf-19/daf-12* null mutants when compared to wild type worms, for which 2,833 were related to Gene Ontology terms. For each term, number of associated genes found in the analysed sample is indicated. P-values meet the criterium for multiple comparisons imposed by the Benjamini-Hochberg method.

Regarding the 810 putative direct targets of FKH-8 upregulated in *daf-19/daf-12* null mutants, several branches emerging from regulation of biological process within the Biological Process ontology were tracked down to enriched terms related to developmental growth, nervous system development and response to stimulus, among others. Of note, the hyperosmotic response term, a response mediated by ciliated neurons in *C. elegans*, was also enriched. A single Molecular Function category (binding) finally led to the enriched terms for protein and guanyl ribonucleotide binding. Finally, this set of upregulated genes could be traced to several cell compartments including the nucleus, dendrite or phagocytic vesicle. Interestingly, the Cell Component term for axoneme resulted enriched too. Statistical significance and number of related genes for the selected enrichment terms mentioned in this paragraph can be seen in **Figure R.46**.
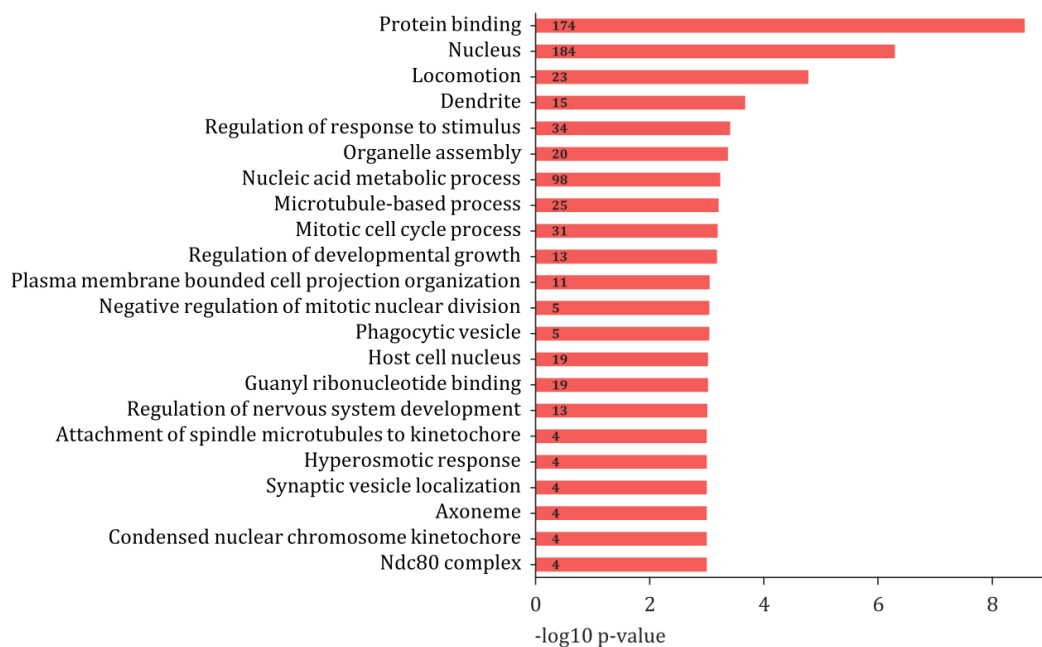
⬥ Figure R.46. Summary of representative GO terms (cellular component, molecular function or biological process) enriched by putative direct targets of FKH-8 upregulated in *daf-19/daf-12* null mutants.

Coloured bars indicate the statistical significance for the enrichment of selected GO terms retrieved through the GOrilla two unranked list of genes analysis. Results arise from the information provided by 602 of the 810 genes used in this analysis. Background control consists of 5,442 *C. elegans* genes significantly upregulated in *daf-19/daf-12* null mutants when compared to wild type worms, for which 3,381 were related to Gene Ontology terms. For each term, number of associated genes found in the analysed sample is indicated. P-values meet the criterium for multiple comparisons imposed by the Benjamini-Hochberg method.

In summary, the subset of putative direct targets of FKH-8 with deregulated expression in *daf-19/daf-12* null mutants was linked to neuronal, sensory-related and cilium-related categories that were absent when considering only the whole set of deregulated genes between *daf-19/daf-12* null mutants and wild type worms.

Finally, considering both shared expression patterns, presence of mutual binding motifs and relationship to ciliary features, we sought to analyse possible epistatic effects between *fkh-8* and *daf-19* in the regulation of the young adult transcriptome. To that aim, we generated a triple *daf-19/daf-12/fkh-8* mutant strain and performed a differential expression analysis between triple *daf-19/daf-12/fkh-8* and double *daf-19/daf-12* mutants. However, the huge impact elicited by the double null mutation of *daf-12* and *daf-19* along with the diluted effect that we could observe for the mutation of *fkh-8* ultimately lead to such differential analysis rendering results with a high probability of false positives. Thus, we decided to discard such data set for further analyses. To circumvent this issue and approximate a view on the epistatic effects of *fkh-8*, *daf-12* and *daf-19* we took advantage of the differential expression analysis of each of the mutant backgrounds against the wild type strain. We reasoned that if *fkh-8* had indeed no synergistic effect alongside *daf-12* or *daf-19* then the set of deregulated elements found in triple mutants would render similar results to those observed for double mutants. However, when comparing the 3 sets of downregulated elements that we retrieved from each of the aforementioned differential expression analyses, we saw that 275 genes were exclusively found in triple *daf-19/daf-12/fkh-8* mutants. Analogously, up to 658 genes were exclusively found in triple *daf-19/daf-12/fkh-8* mutants when comparing the 3 sets of upregulated elements. We then wanted to assess if these new unique sets of genes could be linked to a given biological process. However, when using the WormBase Gene Set Enrichment Analysis tool upon the set of downregulated genes, no single category was enriched. On the other hand, a few terms were enriched by the set of 658 upregulated elements (see **Table R.12**). Nevertheless, no clear ciliary relationship was obtained from that analysis. Still, the addition of the *fkh-8* mutation to the *daf-19/daf-12* mutant background rendered a significant difference in the expression levels of 933 genes that could be attributed to some form of genetic interaction among *fkh-8*, *daf-12* and/or *daf-19*.

Innate immune response might be impaired in young adult *fkh-8(tm292)* mutants.

When compared to wild type worms, the most significant downregulated process that we found within *fkh-8(tm292)* mutant animals was that of the innate immune response (see **Figure R.36**). Interestingly, in recent years, evidence for a link between the nervous and the immune systems have emerged (reviewed in Wani, Goswamy, and Irazoqui 2020). Moreover, this evidence has frequently pointed to subpopulations of ciliated neurons within the nervous system of the worm (X. Cao et al. 2017; Cao and Aballay 2016; Styer et al. 2008). Accordingly, we wanted to further explore the consequences of *fkh-8* mutation upon the functionality of *C. elegans* immune system.

▼ **Table R.12. Ontology terms enriched by the set of upregulated genes found in *triple daf-19/daf-12/fkh-8* mutants.**

WormBase Gene Set Enrichment Analysis was performed upon a set of 658 upregulated genes uniquely found in triple *daf-19/daf-12/fkh-8* mutant animals. Columns gather the number of expected and observed genes per category as well as the corresponding enrichment fold change (F.C.) achieve for each term. P-values for term enrichment are shown, all meeting the criterium for multiple comparisons when applying the Benjamini-Hochberg method (Q-value ≤ 0.05).

Term	Expected	Observed	F.C.	P value	Q value
Male	78	143	1.8	4.50E-12	1.30E-09
Amphid sheath cell	19	45	2.4	4.10E-08	6.20E-06
Coiling frequency decreased	1.8	8	4.5	7.40E-05	1.70E-02
Backing decreased	3	10	3.4	2.10E-04	2.50E-02
Fat content increased	6.1	15	2.5	4.80E-04	3.80E-02
Organic acid metabolic process	12	25	2.1	1.70E-04	2.00 E-02
Transmembrane transport	24	41	1.7	5.70E-04	3.30 E-02

Several neuronal pathways and circuits are known to regulate intestinal host defence in *C. elegans* (Ermolaeva and Schumacher 2014; Liu and Sun 2021; Wani et al. 2020). *tm292* mutants did not show any significant deregulation in the insulin signalling pathway, since the expression levels of the major components integrating this pathway (namely, the *daf-2*, *age-1*, *pdk-1*, *atk-1*, *atk-2* or *daf-16* genes) (see Murphy and Hu 2013) showed no differences when compared to those found in wild type worms. *daf-2* encodes the sole insulin/IGF-1-like receptor that *C. elegans* possesses; however, the products of up to 40 insulin-like peptides (ILPs) genes (from *ins-1* to *ins-39* plus *daf-28*) are known to bind and activate such receptor. Interestingly, we found a significant deregulation in 8 of the genes from the *ins* family (*ins-3, -6, -7, -11, -13, -29, -33* and *-37*) (see **Table R.13**) in *fkh-8(tm292)* mutants and, except for the *ins-11* and the *ins-37* genes, a significant upregulation in their expression levels was found when compared to wild type worms. No expression data was retrieved for the *ins-38* gene. Interestingly, most of the deregulated ILPs in *tm292* mutants (6 out of 8) were known agonists of *daf-2*, since agonistic activities have been reported for the *ins-3*, -6, -11, -13 and -29 genes, whereas an antagonist role is known for *ins-37*, no activity could be linked to *ins-33 in vivo* and a dual agonistic/antagonistic role has been reported for *ins-7* (Zheng et al. 2019).

The *C. elegans* genome contains at least 153 genes that give rise to over 300 predicted neuropeptides (De Fruyt et al. 2020). In addition to the *ins* family, a second major group of peptides is composed of the FMRFamide-related neuropeptides (or FLPs) whereas the remaining neuropeptides that are non-ILPs and non-FLPs define the so-called neuropeptide-like proteins (or NLPs) (Li and Kim

2008). Interestingly, expression of genes from the *nlp* family has been found to be induced upon fungal infection (Couillault et al. 2004). Accordingly, we wanted to assess NLPs' expression levels in *tm292* mutants. At the time of writing, WormBase lists 75 members within the *nlp* gene class and for the 55 NLPs from which we retrieved expression data, 18 of them were significantly deregulated in *tm292* mutants when compared to wild type worms (see **Table R.14**). To complement this analysis, we also checked FLPs' expression values; for all 31 members in the *flp* family currently listed in the WormBase we found deregulated expression for 6 of them in *tm292* mutants (see **Table R.14**). As a whole, we found roughly 21% of the *C. elegans* neuropeptide repertory was deregulated in *fkh-8(tm292)* when compared to wild type worms.

▼ **Table R.13.** *fkh-8(tm292)* **RNA-seq analysis shows expression defects in genes coding for insulin-like peptides.**
Differential gene expression analysis was performed between *fkh-8(tm292)* mutants and wild type worms. Fold change (F.C.) of gene expression between both genetic backgrounds is shown for genes coding for several insulin-like peptides genes. Raw p-values and probability of false positive (P.F.P) computed after the Benjamini–Hochberg correction procedure are indicated.

Gene	Description	log$_2$ FC	Raw p-value	P.F.P
ins-3	Probable insulin-like peptide beta-type 3	1,43	3,13E-02	0,26
ins-6	Probable insulin-like peptide beta-type 5	1,73	6,89E-04	0,21
ins-7	INSulin related	1,69	2,54E-02	0,25
ins-11	B-chain-like peptide	-1,77	4,67E-02	0,29
ins-13	INSulin related	1,43	2,11E-02	0,25
ins-29	INSulin related	1,26	3,77E-03	0,21
ins-33	INSulin related	1,52	1,15E-02	0,23
ins-37	INSulin related	-1,74	3,41E-03	0,21

Next, we analysed the expression data of genes involved in the p38 MAP Kinase pathway, which is sequentially composed of the genes *tir-1*, *nsy-1*, *sek-1*, *pmk-1* and *atf-7* (Couillault et al. 2004; Shivers et al. 2010). All 5 genes were represented in our data set and no differences between *tm292* mutants and wild type worms were observed for the first 3 genes in the pathway. However, we found that the PMK-1 kinase and the ATF-7 transcription factor were significantly downregulated in *fkh-8(tm292)* animals when compared to wild type worms (see **Table R.15**). Moreover, downregulation of ATF-7 could be direct, since three different FKH-8 ChIP-seq peaks were found located across the *atf-7* locus (namely, on both promoter and terminator regions plus an additional peak within the first intron). Of note, as assessed through the CeNGEN data, ATF-7 is expressed in all 25 classes of ciliated neurons in the worm. Consequently, these data strongly suggest that the final steps in the conserved p38 MAP kinase signalling pathway, a central

signalling module of antimicrobial response in the nematode (Kim et al. 2002), could be directly regulated by *fkh-8*.

▼ **Table R.14.** *fkh-8(tm292)* **RNA-seq analysis shows expression defects for several neuropeptides-coding genes.**
Differential gene expression analysis was performed between *fkh-8(tm292)* mutants and wild type worms. Fold change (F.C.) of gene expression between both genetic backgrounds is shown for genes encoding several neuropeptides. Raw p-values and probability of false positive (P.F.P) computed after the Benjamini–Hochberg correction procedure are indicated.

Gene	Description	log$_2$ FC	Raw p-value	P.F.P
nlp-1	Neuropeptide-like protein 1	-1,38	4,01E-02	0,28
nlp-10	Neuropeptide-Like Protein	-1,68	1,78E-02	0,24
nlp-14	Neuropeptide-Like Protein	1,36	4,46E-02	0,28
nlp-19	Neuropeptide-Like Protein	1,66	2,95E-02	0,26
nlp-20	Neuropeptide-Like Protein	-1,58	1,43E-02	0,24
nlp-23	Neuropeptide-Like Protein	-1,21	3,07E-02	0,26
nlp-26	Neuropeptide-Like Protein	1,55	1,02E-02	0,23
nlp-30	QWGYGGY-amide	-2,19	2,46E-02	0,25
nlp-33	Neuropeptide-like protein 33	-1,52	7,01E-04	0,21
nlp-34	Neuropeptide-Like Protein	-2,87	1,90E-02	0,24
nlp-35	Neuropeptide-Like Protein	1,40	2,24E-02	0,25
nlp-37	Neuropeptide-Like Protein	1,20	4,83E-02	0,29
nlp-40	Peptide P4	-1,25	1,82E-02	0,24
nlp-43	Neuropeptide-Like Protein	-1,43	3,31E-02	0,26
nlp-50	Neuropeptide-Like Protein	1,39	7,23E-03	0,22
nlp-54	Neuropeptide-Like Protein	1,45	1,46E-02	0,24
nlp-55	Neuropeptide-Like Protein	-1,51	1,76E-02	0,24
nlp-57	Neuropeptide-Like Protein	1,72	4,70E-02	0,29
flp-8	FMRF-Like Peptide	1,51	5,61E-03	0,21
flp-10	FMRF-Like Peptide	-1,24	2,64E-02	0,25
flp-18	SVPGVLRF-amide 3	1,36	2,82E-02	0,26
flp-19	WANQVRF-amide	1,23	3,96E-02	0,28
flp-32	FMRF-Like Peptide	1,59	2,51E-02	0,25
flp-34	FMRF-Like Peptide	-1,27	2,51E-02	0,25

▼ **Table R.15. Genes composing the conserved p38 MAP kinase signalling pathway are deregulated in *fkh-8(tm292)* mutants.**

Differential gene expression analysis was performed between *fkh-8(tm292)* mutants and wild type worms. Fold change (F.C.) of gene expression between both genetic backgrounds is shown for genes coding for p38 MAP kinase signalling pathway genes. Raw p-values and probability of false positive (P.F.P) computed after the Benjamini–Hochberg correction procedure are indicated.

Gene	Description	$\log_2$ FC	Raw p-value	P.F.P
tir-1	Sterile alpha and TIR motif-containing protein TIR-1	-1,10	3,23E-01	0,64
nsy-1	Mitogen-activated protein kinase kinase kinase NSY-1	1,11	6,37E-02	0,31
sek-1	Dual specificity mitogen-activated protein kinase kinase SEK-1	-1,05	3,55E-01	0,65
pmk-1	Mitogen-activated protein kinase PMK-1	-1,13	3,99E-02	0,28
atf-7	ATF (cAMP-dependent transcription factor) family	-1,18	3,88E-03	0,21

On the other hand, at least 25 different component are known to participate in Wnt signalling in *C. elegans* (Eisenmann 2005); namely, those encoded by the genes *lin-44*, *egl-20*, *mom-2*, *cwn-1*, *cwn-2*, *lin-17*, *mom-5*, *mig-1*, *cfz-2*, *mig-5*, *dsh-1*, *dsh-2*, *mom-1*, *kin-19*, *gsk-3*, *pry-1*, *bar-1*, *wrm-1*, *hmp-2*, *pop-1*, *apr-1*, *lit-1*, *mom-4*, *tap-1*, *lin-18.* Interestingly, 5 of those genes (namely, *egl-20*, *cwn-1*, *lin-17*, *cfz-2* and *mig-5*) were significantly deregulated in *fkh-8(tm292)* mutants when compared to wild type worms (see **Table R.16**). Of note, *fkh-8* actions over *mig-5* could be direct, since an FKH-8 ChIP-seq peak was located less than 90 base pairs upstream of the translational start site of *mig-5*. Additionally, activation of the Wnt signalling pathway in the intestinal epithelium has been shown to trigger the expression of antimicrobials genes such as those coding for lectins and/or lysozymes (Labed et al. 2018). Accordingly, we wanted to assess the expression levels of antimicrobial genes in *tm292* mutants. At the time of writing, WormBase lists 264 members within the C-type LECtin *clec* gene class and for the 219 *clec* genes from which we

retrieved expression data, 46 of them were significantly deregulated in *tm292* (see **Table R.17**). Interestingly, roughly 70% of those genes were significantly downregulated in the mutant background when compared to the wild type worms. Analogously, and much interestingly, 5 out of the 10 members from the LYSozyme *lys* gene class and 2 out of the 6 members from the Invertebrate LYSozyme *ilys* genes class were significantly downregulated in *tm292* mutant worms (see **Table R.18**), thus accounting for downregulation in roughly 44% of the lysozymes repertory encoded by the *C. elegans* genome.

Finally, we wanted to analyse the expression levels for the genes responsible for the antiviral response in *C. elegans*. Such response is mediated by the RNAi machinery and involves a cascade of genes that includes the RDE-1/4 dsRNA binding complex, the DExD box RNA helicase DRH-1 and the Dicer homolog DCR-1 (Tabara et al. 2002). None of these genes showed a deregulated level of expression in *tm292* mutants when compared to wild type worms. Thus, these results

suggest that the antiviral response elicited by this branch of the RNAi machinery of the worm is unaffected, at least, by the partial deletion of *fkh-8*. Interestingly, several components from the endogenous silencing machinery of *C. elegans* (Billi, Fischer, and Kim 2014) were indeed significantly upregulated in *tm292* mutants. Such were the cases of the *ergo-1*, *eri-7*, *eri-9*, *hrde-1*, *nrde-2*, and *nrde-3* genes (see **Table R.19**), although the biological meaning of this phenotype is unclear.

▼ **Table R.16. Components belonging to the Wnt signalling pathway of *C. elegans* are deregulated in *fkh-8(tm292)* mutants.**

Differential gene expression analysis was performed between *fkh-8(tm292)* mutants and wild type worms. Fold change (F.C.) of gene expression between both genetic backgrounds is shown for genes coding for components from the Wnt signalling pathway. Raw p-values and probability of false positive (P.F.P) computed after the Benjamini–Hochberg correction procedure are indicated.

Gene	Description	log$_2$ FC	Raw p-value	P.F.P
egl-20	Uncharacterized protein	-1,40	2,17E-02	0,25
cwn-1	Non-coding transcript of protein-coding gene *cwn-1*	-1,28	1,76E-02	0,24
lin-17	Non-coding transcript of protein-coding gene *lin-17*	-1,38	1,54E-02	0,24
cfz-2	Frizzled-2	1,24	2,24E-02	0,25
mig-5	Segment polarity protein dishevelled homolog *mig-5*	1,35	2,92E-02	0,26

▼ **Table R.17. *fkh-8(tm292)* RNA-seq analysis shows expression defects for several members of the antimicrobial C-type lectin gene class.**

Differential gene expression analysis was performed between *fkh-8(tm292)* mutants and wild type worms. Fold change (F.C.) of gene expression between both genetic backgrounds is shown for genes coding for antimicrobial *clec* genes. Raw p-values and probability of false positive (P.F.P) computed after the Benjamini–Hochberg correction procedure are indicated.

Gene	Description	log$_2$ FC	Raw p-value	P.F.P
clec-82	C-type LECtin	-2,34	4,07E-02	0,28
clec-230	C-type LECtin	-2,21	3,26E-02	0,26
clec-66	C-type LECtin	-2,21	1,17E-02	0,23
clec-60	C-type LECtin	-1,98	4,66E-02	0,29
clec-67	C-type LECtin	-1,85	1,62E-02	0,24
clec-229	C-type LECtin	-1,63	2,56E-02	0,25
clec-62	C-type LECtin	-1,57	3,30E-02	0,26
clec-83	C-type LECtin	-1,56	1,52E-02	0,24
clec-84	C-type LECtin	-1,55	2,72E-02	0,26
clec-152	C-type lectin pseudogene clec-152	-1,52	8,50E-03	0,23
clec-227	C-type LECtin	-1,52	2,88E-02	0,26

Gene	Description	log$_2$ FC	Raw p-value	P.F.P
clec-117	C-type LECtin	-1,51	1,23E-02	0,23
clec-260	C-type LECtin	-1,49	2,91E-02	0,26
clec-236	C-type lectin pseudogene clec-236	-1,44	1,53E-03	0,21
clec-123	C-type LECtin	-1,43	1,91E-02	0,24
clec-48	C-type LECtin	-1,43	4,87E-02	0,29
clec-119	C-type LECtin	-1,38	1,15E-02	0,23
clec-79	C-type LECtin	-1,38	4,95E-02	0,29
clec-154	C-type lectin pseudogene clec-154	-1,38	2,35E-02	0,25
clec-151	C-type LECtin	-1,36	2,04E-02	0,24
clec-73	C-type LECtin	-1,33	3,35E-02	0,27
clec-50	C-type LECtin	-1,32	2,85E-03	0,21
clec-65	C-type LECtin	-1,32	2,70E-02	0,26
clec-155	C-type lectin pseudogene clec-155	-1,31	3,76E-02	0,27
clec-85	C-type LECtin	-1,31	4,01E-02	0,28
clec-45	C-type LECtin	-1,31	4,71E-02	0,29
clec-153	C-type LECtin	-1,29	5,37E-04	0,21
clec-2	C-type LECtin	-1,25	2,10E-02	0,25
clec-4	C-type LECtin	-1,21	4,85E-02	0,29
clec-41	C-type LECtin	-1,20	4,14E-02	0,28
clec-47	C-type LECtin	-1,16	8,02E-03	0,22
clec-90	C-type LECtin	-1,14	4,46E-02	0,28
clec-140	C-type LECtin	1,15	2,77E-02	0,26
clec-71	C-type LECtin	1,17	2,54E-02	0,25
clec-170	C-type LECtin	1,23	1,28E-02	0,24
clec-204	C-type LECtin	1,34	4,54E-02	0,29
clec-10	C-type LECtin	1,35	4,44E-03	0,21
clec-166	C-type LECtin	1,37	5,00E-02	0,29
clec-245	C-type LECtin	1,39	3,07E-02	0,26
clec-31	C-type LECtin	1,42	4,06E-02	0,28
clec-146	C-type LECtin	1,43	2,74E-03	0,21
clec-70	C-type LECtin	1,61	9,89E-03	0,23
clec-147	C-type LECtin	1,67	4,85E-02	0,29
clec-223	C-type LECtin	1,84	7,48E-03	0,22
clec-218	C-type LECtin	1,95	1,28E-02	0,24
clec-97	C-type LECtin	2,67	1,36E-02	0,24

▼ **Table R.18. Several genes coding for *C. elegans* lysozymes are downregulated in *fkh-8(tm292)* mutant worms.**

Differential gene expression analysis was performed between *fkh-8(tm292)* mutants and wild type worms. Fold change (F.C.) of gene expression between both genetic backgrounds is shown for genes coding for members from the *ilys* and *lys* gene classes. Raw p-values and probability of false positive (P.F.P) computed after the Benjamini–Hochberg correction procedure are indicated.

Gene	Description	log$_2$ FC	Raw p-value	P.F.P
ilys-2	Invertebrate LYSozyme	-1,93	4,70E-02	0,29
ilys-5	Invertebrate LYSozyme	-1,25	7,32E-03	0,22
lys-1	LYSozyme	-1,14	4,33E-04	0,21
lys-2	LYSozyme	-1,99	2,58E-03	0,21
lys-3	Lysozyme-like protein 3	-2,34	8,55E-03	0,23
lys-4	LYSozyme	-1,29	3,35E-02	0,27
lys-5	LYSozyme	-1,54	2,33E-02	0,25

▼ **Table R.19. Members from the endogenous RNAi machinery are upregulated in *fkh-8(tm292)* mutants.**

Differential gene expression analysis was performed between *fkh-8(tm292)* mutants and wild type worms. Fold change (FC) of gene expression between both genetic backgrounds is shown for genes encoding for members of the endogenous RNAi machinery. Raw p-values and probability of false positive (P.F.P) computed after the Benjamini–Hochberg correction procedure are indicated.

Gene	Description	log$_2$ FC	Raw p-value	P.F.P
ergo-1	Endogenous-RNAi deficient arGOnaute	1,56	6,64E-03	0,22
eri-7	Enhanced RNAI (RNA interference)	1,49	3,14E-02	0,26
eri-9	Enhanced RNAI (RNA interference)	1,38	2,74E-02	0,26
hrde-1	Heritable RNAi Deficient	1,37	3,66E-03	0,21
nrde-2	Nuclear RNAi defective-2 protein	1,24	2,84E-04	0,21
nrde-3	Nuclear RNAi defective-3 protein	1,33	2,92E-02	0,26

In summary, we found integrating components from the p38 MAP kinase and Wnt signalling pathways significantly deregulated in *tm292* mutants when compared to wild type worms. Additionally, several different genes coding for neuropeptides and antimicrobial lysozymes and C-type lectins were also affected in their expression levels. As a whole, the detailed analysis of our *fkh-8(tm292)* RNA-seq experiment revealed the impairment of several major branches of the innate immune response in *C. elegans*.

5. Generation of a new null mutant allele of *fkh-8*.

fkh-8(tm292) is likely a hypomorphic allele.

Our RNA-seq analysis revealed the presence of *fkh-8* reads surrounding the *tm292* deletion. As previously stated, the *tm292* mutation consists of a partial deletion in which a region spanning from roughly the middle of the fourth intron to the first 10 nucleotides of the last exon are eliminated. Our analysis revealed the presence of a novel and stable mutant RNA messenger in which, due to the removal of exon 8 splice acceptor site, exon 4 splice donor site is not recognized. Thus exon 4, the remains from intron 4 and the remains of exon 8 are joined together in a new and big last exon (see **Figure R.47**). Due to this partial retention of intron 4, three new amino acids and a novel stop codon appear (see **Figure R.48.A**). Thus, it is likely that the *tm292* allele generates a truncated FKH-8 protein of 179 amino acids in which the whole FKH DNA binding domain remains unaffected (see **Figure R.48.B**).
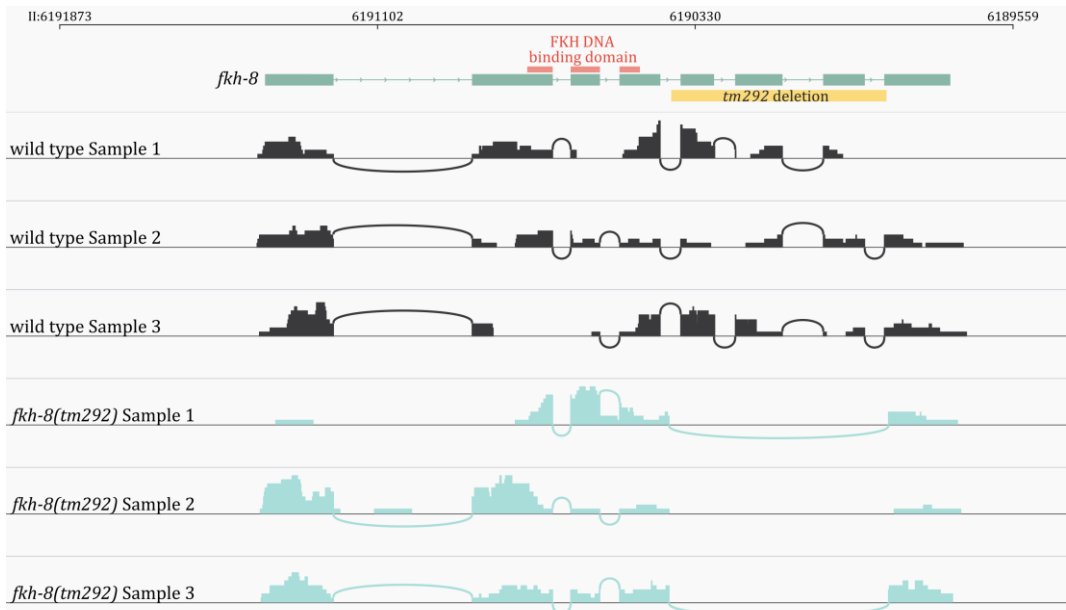
⤴ Figure R.47. *tm292* deletion generates a novel truncated messenger for *fkh-8*.
Sashimi plot portraying the read coverage retrieved for the *fkh-8* locus (on top, turquoise) in three independent samples of both wild type (dark grey) and *fkh-8(tm292)* mutants (light blue). Coloured connecting arcs represent the presence of reads overlapping adjacent exons when aligned against the genome of reference. Location of the FKH DNA binding domain is indicated in pink whereas a yellow box depicts the 561 base pairs deletion of the *tm292* allele. Chromosome coordinates appear on top of the figure.
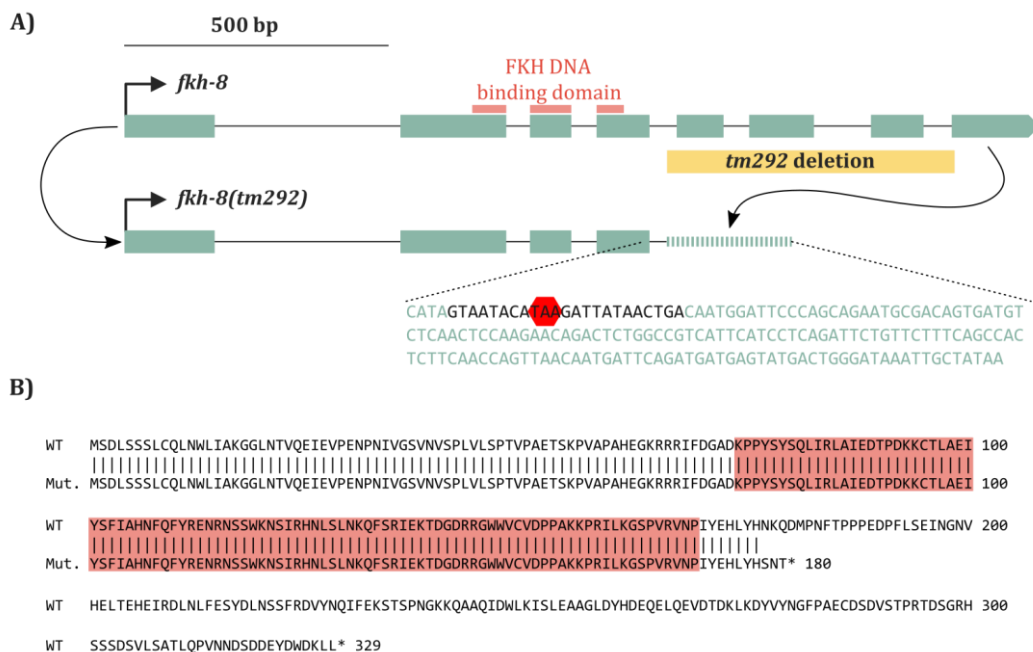
Figure R.48. *fkh-8(tm292)* **is likely a hypomorphic allele.**
A) Schematics representing the *fkh-8* locus in its wild type form (top) and within the context of the *tm292* mutation (bottom). A yellow box depicts the 561 base pairs deletion in which exons 5, 6 and 7 are fully removed along with the first 10 nucleotides of exon 8. RNA-seq data reveals the presence of a novel messenger in which the remains of intron 4 are joined together to those of exon 8. In this context, a TAA stop codon (marked in red) appears. **B)** Alignment of FKH-8 wilt type (WT) protein and predicted product generated by the *tm292* mutation (Mut.). The whole FKH DNA binding domain, highlighted in pink, is conserved in this predicted truncated FKH-8 protein. Notice that the 3 last amino acids of this mutant protein differ from those of the wild type product.

fkh-8(vlc43): an engineered full deletion of the *fkh-8* locus.

Considering our RNA sequencing results, we find it feasible that lack of expression defects for ciliary genes in *fkh-8(tm292)* mutants could be due to the remaining activity of a truncated protein. Accordingly, we aimed to generate our own null *fkh-8* mutant strain. Successful co-CRISPR strategy allowed us to obtain a 1,864 base pairs deletion strain in which the whole *fkh-8* locus was removed.

This new allele was confirmed through Sanger sequencing and mapped 79 base pairs upstream *fkh-8* translational start site and 122 base pairs downstream its TAA stop codon. As customary, this new mutant allele will be made available to the rest of the scientific community through the CGC. A schematic representation of both *fkh-8* mutant alleles used in this work appears in **Figure R.49**.
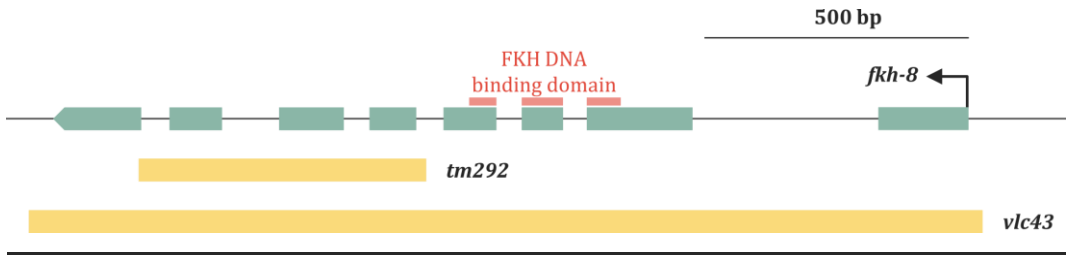
▲ **Figure R.49. Schematics for the two *fkh-8* mutant alleles used in this work.**
Turquoise boxes represent here all 8 exons coding for FKH-8. Red bars on top of the exons mark the FKH
DNA binding domain. Yellow bars represent the width for the deletions present in the two mutant alleles
of *fkh-8* used in this work. Notice that *tm292* mutation leave intact the FKH binding domain.

fkh-8 mutants do not exhibit dye-filling defects.

As previously stated, the subsets of ciliated neurons integrated by the ASK, ADL, ASI, AWB, ASH, ASJ, PHA and PHB neurons expose their endings directly to the external environment, hence allowing for their labelling when worms are transferred into fluorescent lipophilic dye solutions such as that of DiD. Accordingly, and due to the strong relationship that we could observe between *fkh-8* genomic binding and the ciliated fate, we decided to perform DiD staining as a first quick analysis of cilia integrity in *fkh-8* mutants. However, as seen in **Figure R.50**, dye-filling of both *fkh-8* mutant strains assayed is similar to that of wild type animals. Thus, these results indicate that general ciliary integrity is not compromised in *fkh-8* mutants at least in the subset of ciliated neurons assayed.
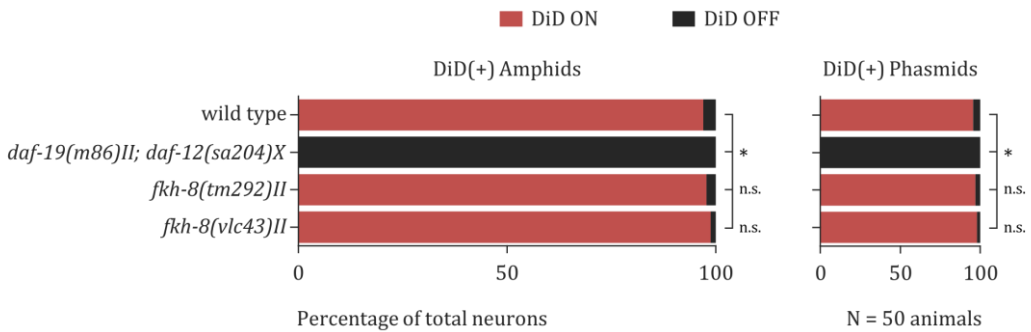
▲ **Figure R.50. *fkh-8* mutants do not exhibit dye-filling defects.**
For each sub-population of amphid and phasmid neurons, deep red bars represent the percentage of total neurons labelled by the fluorescent lipophilic dye DiD. Absence of labelling is depicted in black. Wild type worms are used as a control for DiD staining, whereas dye-filling defective *daf-19* mutants constitute the negative control strain. Worms used in these experiments were cultivated at 20 Celsius degrees to avoid dye-filling defects observed in wild type worms when cultivated at 25 Celsius degrees. Statistical significance was calculated through the Fisher exact test. A Bonferroni correction accounting for all 6 possible pair comparisons was applied. For ease of reference, only the comparisons against the wild type control strains are indicated. No statistical difference was found between both *fkh-8* mutant alleles and wild type.

6. *In vivo* reporter gene expression analysis uncovers the role of FKH-8 regulating structural ciliary genes.

Animals with whole deletion of the *fkh-8* locus show stronger defects in the expression of cilium-related reporters.

RNA-seq profiling of *fkh-8(tm292)* mutants failed to identify expression defects in genes expressed in ciliated neurons and for which FKH-8 binding had been found. As stated, this might be due to the hypomorphic nature of the *tm292* allele. An alternative explanation, not mutually exclusive, is that expression defects in *fkh-8* mutants might be too subtle to be detected by whole worm RNA sequencing. To overcome this issue, we decided to explore ciliated neuron gene expression in *tm292* and *vlc43* alleles by fluorescent reporter analysis, hence achieving cellular resolution.

To this aim, we took advantage of some of the reporter constructs that we generated for structural ciliary features. We found that 9 of such expand a genomic region associated to a FKH-8 ChIP-seq peak (see **Figure R.51**). Thus, we selected those reporters to analyse their expression in *fkh-8 tm292* and *vlc43* mutant backgrounds. Importantly, except for *osm-5*, none of these genes showed expression defects in our *fkh-8(tm292)* RNA-seq analysis (see **Table R.20**).

From this analysis (whose quantifications appear represented in **Figure R.52**), we noticed that 3 out of the 9 reporters showed no statistical differences when comparing the average number of reporter-expressing neurons between wild type animals and both *fkh-8* mutant backgrounds. Namely, those were the cases of the *peli-1*(long), *tub-1* and *xbx-1*(short) reporters. However, for the rest of the reporters, a significant decrease in the number of reporter-expressing neurons was found within a *fkh-8* mutant background, including alternative reporters for *peli-1* and *xbx-1* genes. Interestingly, for 3 of such reporters (those of *ift-20*, *osm-1* and *peli-1*(short)), a stronger phenotype was found within the null *vlc43* allele. Analogously, phenotypes associated with the *tmem-107* and *xbx-1*(long) reporters were only manifested in the full absence of *fkh-8*. These data support the hypothesis that *tm292* is not a null allele. Additionally, *fkh-8* requirement for *peli-1* reporter expression was found to be specific for the short reporter version, while *peli-1*(long) reporter was normally expressed in both *fkh-8* mutants. This outcome agreed with our RNA-seq results since *peli-1* expression value was not significantly different from that of the wild type strain (p-value = 6.12E-02). Such a phenomenon could be explained by the presence of redundant enhancers or compensatory TF bindind sites within the *peli-1*(long) construct.

In summary, this detailed fluorescent reporter analysis allowed us to detect significant expression defects in 6 out of the 7 structural ciliary genes selected that were not observed in our RNA-seq analysis. Moreover, affected genes harbour *fkh-8* ChIP-seq peaks within or nearby their reporter sequences, hence suggesting this regulation in gene expression might be direct.
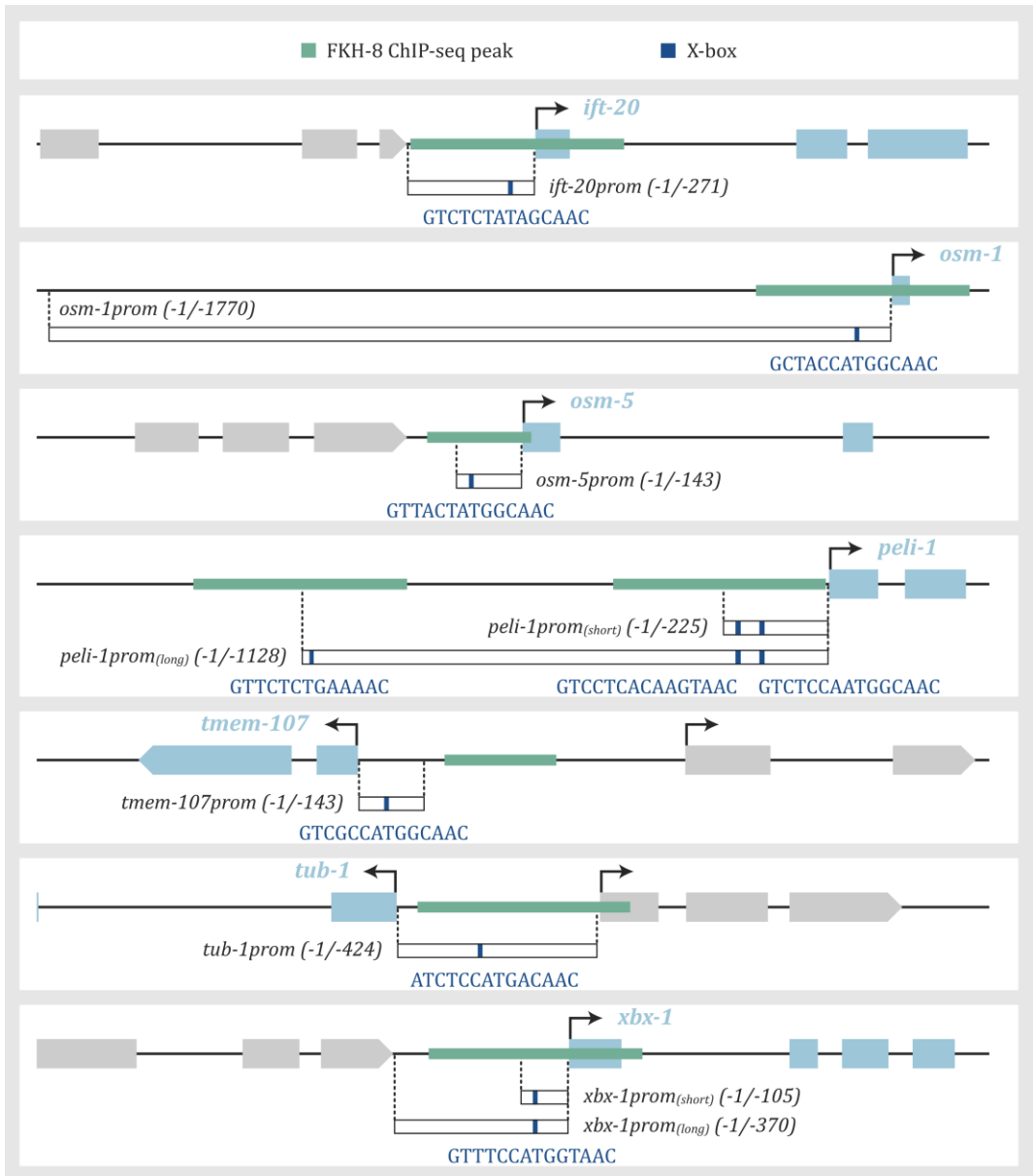
▲ Figure R.51. Genetic landscapes of genes selected for reporter analysis in *fkh-8* mutant backgrounds.

Schematics illustrating the genetic landscape within 2,000 base pairs regions (think black lines) surrounding the selected genes. Coloured boxes represent the exons; structural cilia features are depicted in blue whereas non-cilia related genes are shown in grey. Sense of coding sequences is depicted with black arrow heads; thus, genes encoded within the plus strand show arrows pointing to the right and genes encoded within the minus strand show arrows pointing to the left. Vertical dashed lines mark the genetic region used to build the reporters. Coordinates of the regions cloned, represented by white boxes, are indicated respect to the ATG. Location and sequence of X-box motifs appear as dark blue boxes and text. FKH-8 ChIP-seq peak sequences are represented by turquoise bars.

▼ Table R.20. *fkh-8(tm292)* RNA-seq analysis shows no expression defects for selected structural ciliary genes.

Differential gene expression analysis was performed between *fkh-8(tm292)* mutants and wild type worms. Fold change (FC) of gene expression between both genetical backgrounds is shown for selected genes coding for structural components of the cilium, also expanding a genomic region associated to a FKH-8 ChIP-seq peak. Raw p-values and probability of false positive (P.F.P) computed after the Benjamini–Hochberg correction procedure are indicated. Note that *osm-5* is the only gene for which expression level is significantly affected in the comparison between mutant and wild-type animals, with a probability of false positive approaching 21%.

Gene	WormBase ID	log$_2$ FC	Raw p-value	P.F.P
ift-20	WBGene00022465	-1.20	325E-02	0.64
osm-1	WBGene00003883	1.03	6.42E-02	0.84
osm-5	WBGene00003885	-1.29	1.98E-04	0.21
peli-1	WBGene00017766	1.00	8.47E-02	0.99
tmem-107	WBGene00043308	-1.36	2.42E-02	0.56
tub-1	WBGene00006655	1.04	6.12E-02	0.81
xbx-1	WBGene00006960	-1.01	9.56E-02	1
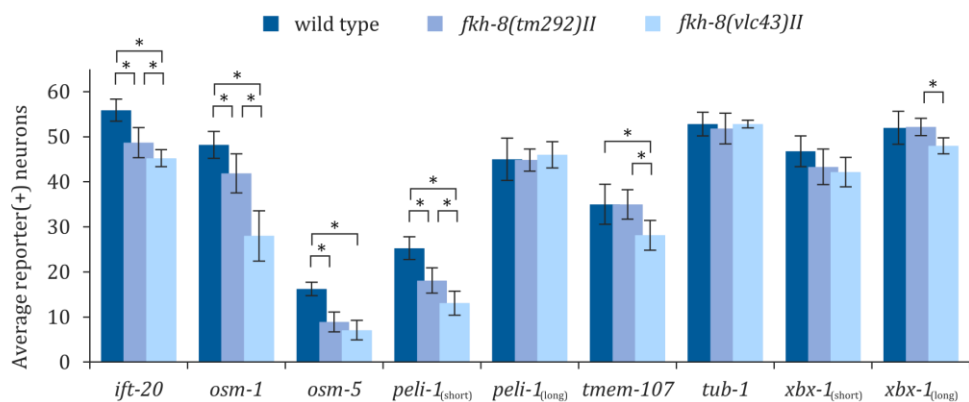
▲ Figure R.52. Null *fkh-8(vlc43)* mutant animals display stronger defects in the expression of cilia-related features than *fkh-8(tm292)*.

Average number of reporter-expressing neurons in wild type and two *fkh-8* alleles. For *peli-1* and *xbx-1*, short and long versions of the reporters were analysed. Error bars represent standard deviation of the sample. Statistical significance, depicted by asterisks, was established through a two-tailed Student *t* test once the variances of the two samples being compared had been analysed. A Bonferroni correction was applied considering all 3 simultaneous hypotheses being tested per reporter. N = 5 to 12 animals per reporter and genotype.

As a whole, these results demonstrate the role of *fkh-8* in the transcriptional regulation of structural ciliary features of *C. elegans*. Furthermore, differences observed between both *fkh-8* mutant strains used in this work unravel the hippomorphic nature of the *tm292* deletion.

fkh-8 mutation affects the expression of structural ciliary genes in a cell-type specific manner.

The small decrease in the total number of reporter-expressing cells that we could observe when comparing wild type animals and *fkh-8* mutants could be explained at least by two different phenotypes: first, *fkh-8* could be affecting gene expression in any of the cells expressing the reporter but with very low penetrance; alternatively, *fkh-8* could be affecting reporter expression only in specific subpopulations of neurons. To distinguish between these two possibilities, we analysed the phenotypes that we obtained in the context of the five distinct anatomical regions in which ciliated neurons are found (see **Figure R.53**).

Within Region 1, where amphid neurons locate, up to 5 different reporters showed significant expression defects in a *fkh-8* mutant background; namely, those for *oms-1*, *osm-5*, *peli-1*(short), *tmem-107* and *xbx-1*(long). Interestingly, except for the *peli-1*(short) case, expression defects for the rest of those reporters were observed only in null *fkh-8* mutants. In Region 2, where PQR and phasmid PHA and PHB neurons locate, expression defects were observed for the *ift-20*, *osm-1, osm-5* and *peli-1*(short) reporters in both *fkh-8* mutant backgrounds (except for the *osm-5* reporter that was affected only in the null *vlc43* allele). In Region 3, where cephalic and labial neurons are found, expression was affected for the *ift-20*, *osm-1*, *osm-5 peli-1*(short) and *tmem-107* reporters. *peli-1*(short) reporter expression within this region in N2 animals is limited to IL2 neurons, as assessed by co-localization with the cholinergic marker *cho-1/Acetylcholine transporter* (see **Figure R.54**); thus, *fkh-8* mutation is likely to affect *peli-1*(short) expression within the IL2 neurons. Expression defects in this region were generally stronger in the null *fkh-8* strain, with the extreme case of the *tmem-107* reporter that was only significantly affected in the *fkh-8(vlc43)* mutant background. Region 4, in which ADE, FLP and AQR neurons locate, showed defects in the expression of *ift-20*, *osm-1*, *osm-5* and *xbx-1*(short) reporters. Finally, Region 5, which corresponds to the unique neuron class PDE, revealed loss of expression for the *ift-20*, *osm-5*, *tmem-107* and *xbx-1*(short) reporters.

➤ **Figure R.53. *fkh-8* mutation affects expression of ciliary genes in a cell-type specific manner.** Top scheme shows a lateral left view of a *C. elegans* young adult hermaphrodite portraying the location of the ciliated system neurons. AQR unilateral neuron locates in the right side of the animal and hence appears depicted in dashes lines. For each of the five anatomical regions in which the system was divided, average number of reporter-expressing neurons for 7 different reporters is shown. Standard deviation of the mean is also indicated for each scoring. Regions with a significant decrease of reporter expression in *fkh-8* mutants are highlighted in pink. Statistical significance was established through a one-tailed *t* test once the variances of the two groups being compared were analysed. Last column gather the final average number of reporter-expressing neurons observed for all regions and animals (N) analysed.

	R3		R1		R4		R5		R2		TOTAL		N
ift-20													
Wild type	22.80 ± 1.32		23.50 ± 0.53		2.70 ± 1.25		2.00 ± 0.00		4.90 ± 0.32		55.90 ± 2.47		10
fkh-8(tm292)II	19.30 ± 2.41		23.80 ± 0.42		0.90 ± 0.74		0.50 ± 0.71		4.20 ± 0.42		48.70 ± 3.33		10
fkh-8(vlc-43)II	17.33 ± 1.30		23.42 ± 0.67		0.33 ± 0.49		0.00 ± 0.00		4.17 ± 0.39		45.25 ± 1.86		12
osm-1													
Wild type	19.78 ± 1.56		22.56 ± 2.65		1.22 ± 1.30		0.11 ± 0.33		4.56 ± 0.73		48.22 ± 2.99		9
fkh-8(tm292)II	15.89 ± 2.98		22.00 ± 2.00		0.22 ± 0.67		0.00 ± 0.00		3.78 ± 0.67		41.89 ± 4.34		9
fkh-8(vlc-43)II	9.67 ± 2.45		14.56 ± 3.43		0.00 ± 0.00		0.11 ± 0.33		3.67 ± 0.71		28.00 ± 5.57		9
osm-5													
Wild type	4.13 ± 1.96		6.25 ± 1.28		1.25 ± 0.89		0.88 ± 0.64		3.75 ± 0.71		16.25 ± 1.49		8
fkh-8(tm292)II	0.08 ± 0.29		5.42 ± 1.83		0.00 ± 0.00		0.00 ± 0.00		3.42 ± 0.67		8.92 ± 2.19		12
fkh-8(vlc-43)II	0.00 ± 0.00		4.00 ± 1.83		0.00 ± 0.00		0.00 ± 0.00		3.10 ± 0.74		7.10 ± 2.18		10
peli-1(short)													
Wild type	6.71 ± 2.14		14.14 ± 2.04		0.00 ± 0.00		0.00 ± 0.00		4.43 ± 0.79		25.29 ± 2.50		7
fkh-8(tm292)II	4.18 ± 0.40		10.27 ± 2.24		0.00 ± 0.00		0.00 ± 0.00		3.64 ± 0.92		18.09 ± 2.81		11
fkh-8(vlc-43)II	3.00 ± 1.04		6.42 ± 2.02		0.00 ± 0.00		0.00 ± 0.00		3.67 ± 0.65		13.08 ± 2.68		12
tmem-107													
Wild type	15.67 ± 1.75		13.67 ± 2.07		1.00 ± 0.63		0.50 ± 0.55		4.17 ± 0.75		35.00 ± 4.43		6
fkh-8(tm292)II	16.67 ± 0.82		13.00 ± 1.10		1.50 ± 1.38		0.00 ± 0.00		3.83 ± 1.94		35.00 ± 3.29		6
fkh-8(vlc-43)II	12.75 ± 2.49		10.50 ± 2.51		0.75 ± 0.71		0.13 ± 0.35		4.00 ± 0.93		28.13 ± 3.27		8
xbx-1(short)													
Wild type	19.40 ± 1.52		18.60 ± 2.70		2.60 ± 0.55		1.60 ± 0.55		4.60 ± 0.89		46.80 ± 3.42		5
fkh-8(tm292)II	19.33 ± 1.51		18.67 ± 1.51		1.17 ± 0.75		0.17 ± 0.41		4.00 ± 0.89		43.33 ± 3.93		6
fkh-8(vlc-43)II	19.17 ± 2.04		17.00 ± 2.04		1.83 ± 0.75		0.50 ± 0.84		3.67 ± 0.82		42.17 ± 3.25		6
xbx-1(long)													
Wild type	20.00 ± 1.22		24.20 ± 1.92		2.20 ± 0.45		1.20 ± 0.45		4.40 ± 0.89		52.00 ± 3.67		5
fkh-8(tm292)II	20.60 ± 0.89		23.20 ± 0.84		2.40 ± 0.55		1.80 ± 0.45		4.20 ± 0.84		52.20 ± 1.92		5
fkh-8(vlc-43)II	19.17 ± 1.72		19.67 ± 1.75		2.67 ± 0.52		1.67 ± 0.52		4.83 ± 0.41		48.00 ± 1.79		6

Legend: ○ Ciliated neuron   ▢ Expression defective region

DIC	*peli-1*(short)	*cho-1*	Merged

Head

30 μm

⬆ **Figure R.54. *peli-1* is expressed in the IL2 neurons in a wild type genetic background.**
Representative images of a lateral view of the head of a *C. elegans* young adult hermaphrodite expressing fluorescent reporters for the genes *peli-1* and *cho-1/Acetylcholine transporter*. First column displays the DIC image exposing the morphology of the worm, second and third columns show the expression for the fluorescent reporters of *peli-1(short)::gfp* and *cho-1::mCherry* (*otIs544*) whereas the final column reveals the colocalization between both reporters within the six IL2 neurons. For fluorescent images, body limits of the worm are represented as dashed lines. Final fluorescent images correspond to Z stacks projections of images at 1 micron distance taken under a vertical fluorescence microscope.

The pattern of gene expression loss for the *ift-20* and *osm-5* reporters was reminiscent of the anatomical configuration of the dopaminergic system of the worm, which is integrated by 4 CEP (located in Region 3), 2 ADE (located in Region 4) and 2 PDE neurons (located in Region 5). We hence analysed *ift-20* expression together with the dopaminergic reporter *dat-1*. Double reporter analysis showed *ift-20* expression being systematically lost in all dopaminergic neurons in both *fkh-8* mutant backgrounds (see **Figure R.55**).

In summary, our detailed anatomical analysis of reporter gene expression demonstrate that *fkh-8* is required for the expression of structural ciliary genes. Interestingly, the expression of each reporter is affected in specific and partially overlapping subpopulations of ciliated neurons. Altogether, our analysis identified expression defects in neuron classes distributed across the whole ciliated system of the worm; specifically including the IL2, CEP, ADE, PDE and AQR classes plus several unidentified amphid and phasmid neurons. This specificity in gene expression defects might reflect distinct compensatory effects present in each neuron for each reporter. Moreover, the severity of the phenotypes presented in this section is actually minimized, since all scorings correspond to the presence and not the intensity of the fluorescent signal. In this regard, decreases in signal intensity were observed for several (if not all) reporters and anatomical regions analysed, further exposing the broad requirement for *fkh-8* in the correct expression of cilium-related features.

fkh-8 acts cell-autonomously.

Next, we assessed cell-autonomous actions of *fkh-8* taking advantage of the penetrant phenotype we observed for the dopaminergic system, in which the expression of the panciliary gene *ift-20* was completely abolished, together with the fact that expression of the *dat-1* gene is not affected in *fkh-8* mutants. Thus, we generated transgenic lines in which expression of the *fkh-8* coding sequence was under the control of a *dat-1*/dopamine transporter promoter. To detect expression of the rescuing construct, we created a transcript coding for both FKH-8 and EGFP proteins joint together by a T2A autocatalytic sequence, hence ensuring *fkh-8* expression within the *gfp*-positive neurons.

We analysed three independent transgenic lines and, as expected, we found *dat-1*::*egfp* expression in all dopaminergic neuron types in *fkh-8(vlc43)* mutants (see **Figure R.56**). Importantly, *fkh-8* expression in the dopaminergic neurons was sufficient to rescue *ift-20* expression in those cells (see **Figure R.56**). Thus, these results demonstrate that *fkh-8* acts in a cell-autonomous manner to regulate the expression of this panciliary gene.

cis-mutation of putative FKH binding sites strongly affects expression of the panciliary gene *xbx-1*.

Finally, to complement our *fkh-8* mutant analysis, we decided to perform *cis*-regulatory analysis and site-directed mutagenesis of predicted *fkh-8* binding sites. In order to minimize redundancy among *fkh-8* binding sites or with shadow enhancers, we decided to analyse the shortest reporter affected by

fkh-8 mutation, corresponding to that of *xbx-1*(short). As explained above, the expression of the *xbx-1*(short) reporter is affected in the PDE and unidentified neurons located in Region 4 (likely the ADE or AQR neuron) in *fkh-8* mutants and this reporter overlaps with an FKH-8 ChIP-seq peak (see **Figure R.51**). The analysis of the 105 base pairs sequence of the *xbx-1*(short) reporter revealed the existence of an FKH-type binding motif that matches the FKH consensus RYMAAYA and is adjacent to that of an X-box motif (see **Figure R.57 and Table R.21**). All nine nucleotides of the putative FKH motif were mutated by changing the CTTGTTGAA wild type sequence into the AGGAACCCT motif (see Methods section for details on mutation criteria).

Surprisingly, and in contrast to the modest effect that the *fkh-8(tm292)* and *fkh-8(vlc-43) trans*-mutation had on the expression of the wild type *xbx-1*(short) reporter, this *cis*-mutation caused a dramatic reduction in both the number of neurons able to maintain reporter expression as well as in the intensity of the GFP signal (see **Figure R.58**). These results suggest that in the *fkh-8* mutant background other members of the FKH family could bind FKH-8 binding sites and compensate for its loss.

A)

B)

C)
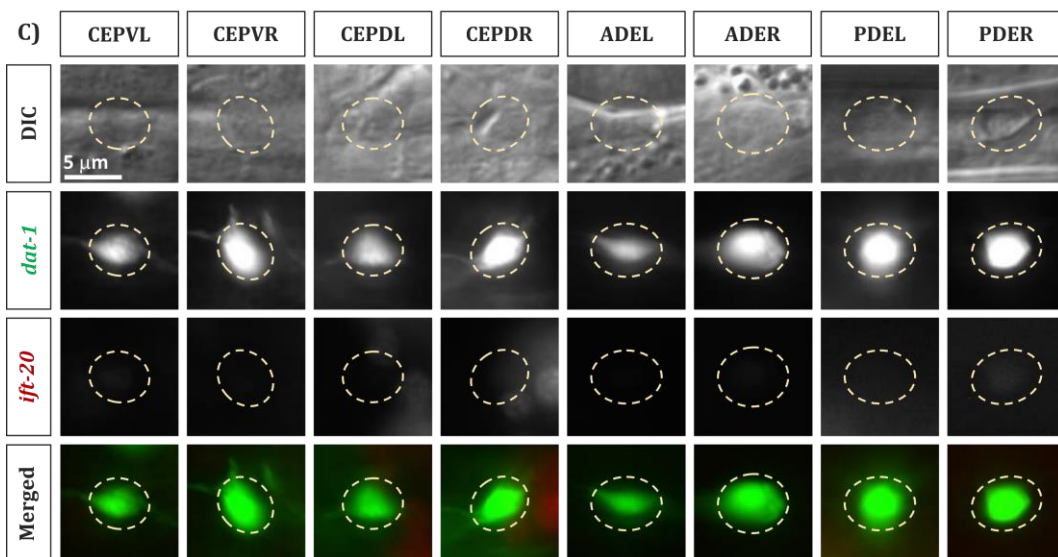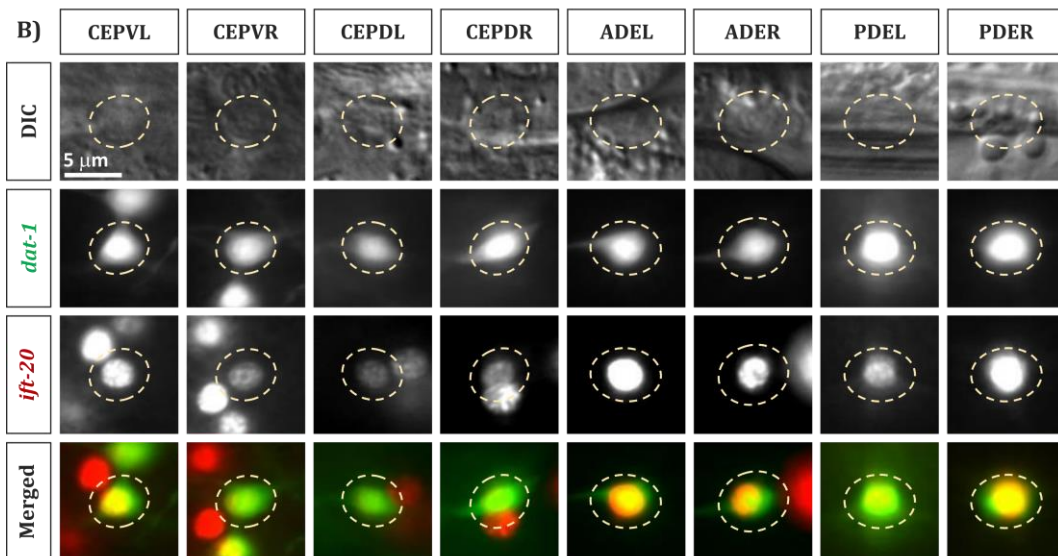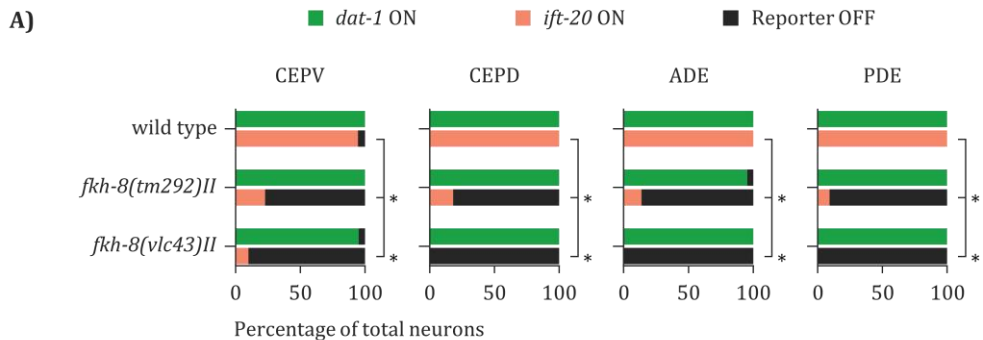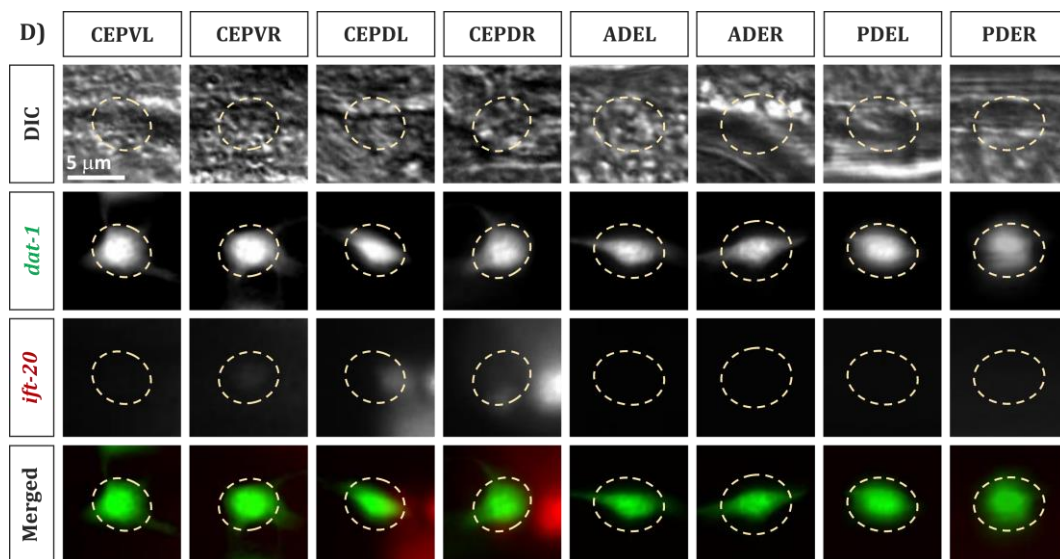
◀ ⬆ **Figure R.55.** ***fkh-8*** **mutation fully abolishes the expression of the panciliary** ***ift-20*** **reporter within the whole dopaminergic system of** ***C. elegans.***

A) Coloured bars represent here the percentage of total neurons expressing the dopaminergic marker *dat-1* (in green) and an integrated reporter for the panciliary gene *ift-20* (in red) in wild type and *fkh-8 tm292* and *vlc43* mutant alleles. Statistical significance was calculated through the Fisher exact test and appears depicted by asterisks. A Bonferroni correction accounting for all 4 possible pair comparisons was applied. For ease of reference, only the comparisons against the wild type control strains are indicated. No statistical difference was found between both *fkh-8* mutant alleles. N = 9, 11 and 10 animals for wild types, *fkh-8(tm292)* and *fkh-8(vlc43)* mutants, respectively. **B)**, **C)** and **D)** Representative images of all four dopaminergic neuron types in WT (**B**) *fkh-8(tm292)* (**C**) and *fkh-8(vlc43)* (**D**) animals. First row displays the DIC images exposing the morphology of the cells, second and third rows show the expression for the fluorescent reporters of *dat-1* and *ift-20* whereas the final row shows the colocalization between both reporters. For all images, cell nucleus appears surrounded by dashed lines. *ift-20* reporter expresses the ds-red protein fused to a nuclear localization signal and thus labels only the nucleus of the cell.
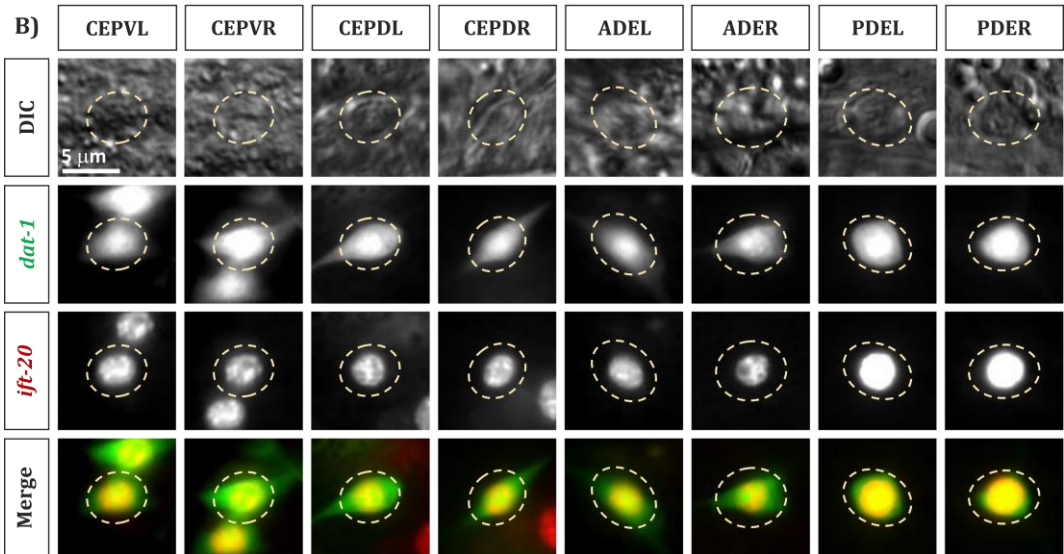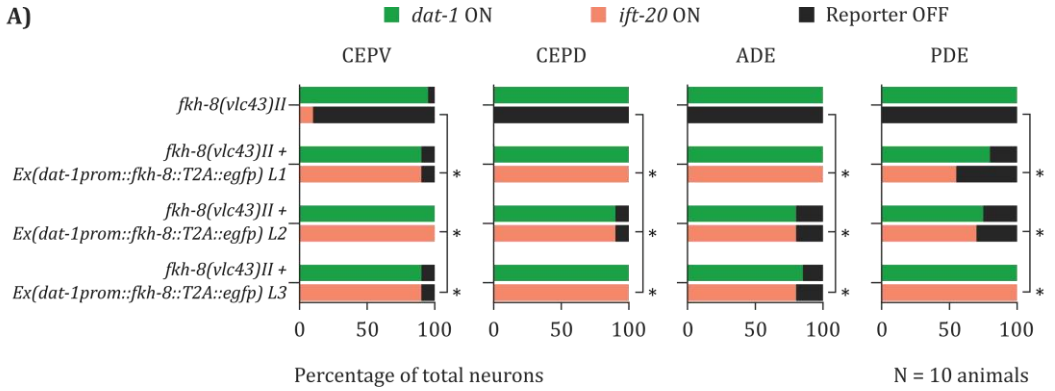
Figure R.56. The transcription factor *fkh-8* acts in a cell-autonomous manner.
A) For each row, coloured bars represent here the percentage of total neurons expressing the dopaminergic marker *dat-1* in green and an integrated reporter for the panciliary gene *ift-20* in red. First row shows expression of both reporters observed within the different subpopulations of the dopaminergic system in null *fkh-8* mutant animals. Same *dat-1* promoter was used to generate a rescuing extrachromosomal construct expressing both FKH-8 and EGFP proteins. Expression of the integrated reporter for *ift-20* was assessed for three different lines of null *fkh-8* mutants animals carrying this rescuing array. Statistical significance for the number of neurons expressing the reporter for *ift-20* between mutant and each rescued line was calculated through the Fisher exact test and appears depicted by asterisks. **B)** Representative images for all eight dopaminergic neurons in null *fkh-8(vlc43)II* rescued animals. First row displays the DIC images showing the morphology of the cells, second and third rows show the expression for the fluorescent reporters of *dat-1* and *ift-20* whereas the final row reveals the colocalization between both reporters with *dat-1* signal being represented in green and *ift-20* in red, *ift-20* reporter expresses the ds-red protein fused to a nuclear localization signal and thus labels only the nucleus of the cell. For all images, cell nucleus appears surrounded by dashed lines.
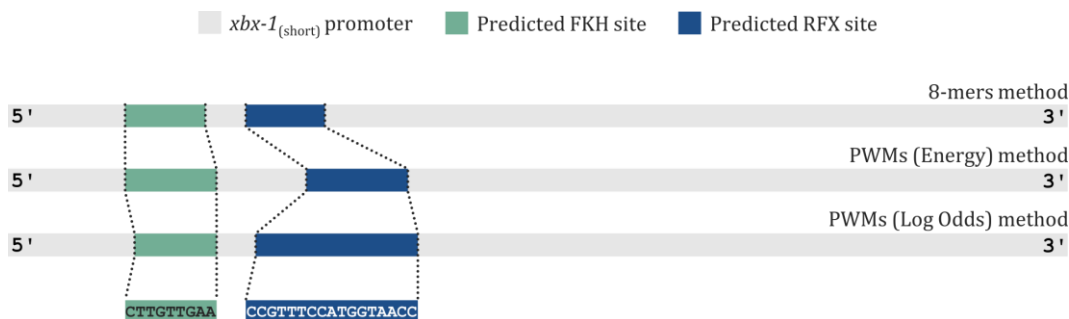
⋏ **Figure R.57. The structural panciliary gene *xbx-1* harbours putative FKH and RFX sites in its minimal promoter sequence.**

Grey bar represent the 105 base pairs of the *xbx-1*(short) promoter. Three methods (8-mer, PWM Energy and PWM Lag Odds) were used to scan the sequence for putative binding sites for members of the FKH and RFX families of transcription factors. Overlapping matches for both families were retrieved from all three methods, here depicted as green boxes for predicted FKH sites and blue boxes for putative X-boxes. The sequence integrating all nucleotides from overlapping matches appear at the bottom of the figure.

⋎ **Table R.21. Motifs within the *xbx-1* minimal promoter that match experimentally validated binding sites of FKH and RFX transcription factors.**

Three different methods from the CIS-BP 1.02 database toolset (8-mers, PWMs Energy and PWMs Log odds) were used to scan the 105 base pairs sequence of the *xbx-1*(short) promoter. Putative binding sites for members of the FKH and RFX families of transcription factors (TF) were found to pass the threshold for all three methods in overlapping regions.

Method	Motif found	Motif ID	Score	Related TF	CIS-BP TF ID	Family
8-mers	CTTGTTGA	M0730_1.02	0.467	*daf-16*	T082933_1.02	FKH
8-mers	CTTGTTGA	M0749_1.02	0.465	*unc-130*	T081000_1.02	FKH
8-mers	CCGTTTCC	M1526_1.02	0.458	*daf-19*	T138992_1.02	RFX
PWMs (Energy)	CTTGTTGAA	M0756_1.02	0.218	*fkh-6*	T080996_1.02	FKH
PWMs (Energy)	CTTGTTGAA	M0756_1.02	0.218	*fkh-10*	T080997_1.02	FKH
PWMs (Energy)	CTTGTTGAA	M0756_1.02	0.218	*unc-130*	T081000_1.02	FKH
PWMs (Energy)	CTTGTTGAA	M0756_1.02	0.218	*pha-4*	T082661_1.02	FKH
PWMs (Energy)	CTTGTTGAA	M0756_1.02	0.218	*lin-31*	T082896_1.02	FKH
PWMs (Energy)	CTTGTTGAA	M0756_1.02	0.218	*fkh-2*	T082938_1.02	FKH
PWMs (Energy)	CCATGGTAAC	M1528_1.02	0.368	*daf-19*	T138992_1.02	RFX
PWMs (Log Odds)	TTGTTGAA	M2122_1.02	8.808	*unc-130*	T081000_1.02	FKH
PWMs (Log Odds)	CGTTTCCATGGTAACC	M6077_1.02	22.514	*daf-19*	T138992_1.02	RFX

A)

Legend:
- ░ *xbx-1*(short) promoter
- ▨ FKH site
- ◼ RFX site (blue)
- **X** Mutation
- ○ Ciliated neuron
- ▮ GFP ON (green)
- ▮ GFP FAINT (light green)
- ◼ GFP OFF (black)

R3, R1, R4, R5, R2

Line 1; N = 5 — GFP
Line 2; N = 7 — GFP
↳ CTTGTTGAA

Line 1; N = 12 — GFP
Line 2; N = 12 — GFP
Line 3; N = 12 — GFP
↳ AGGAACCCT

Percentage of reporter(+) ciliated neurons

B)

xbx-1(short) WT FKH: Worm #1, Worm #2, Worm #3 (20 µm)

xbx-1(short) ΔFKH: Worm #1, Worm #2, Worm #3 (20 µm)
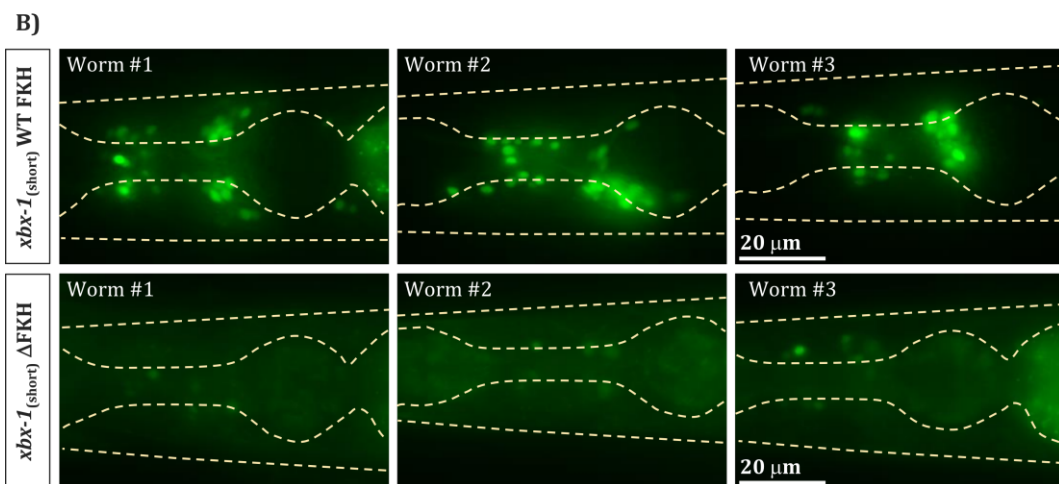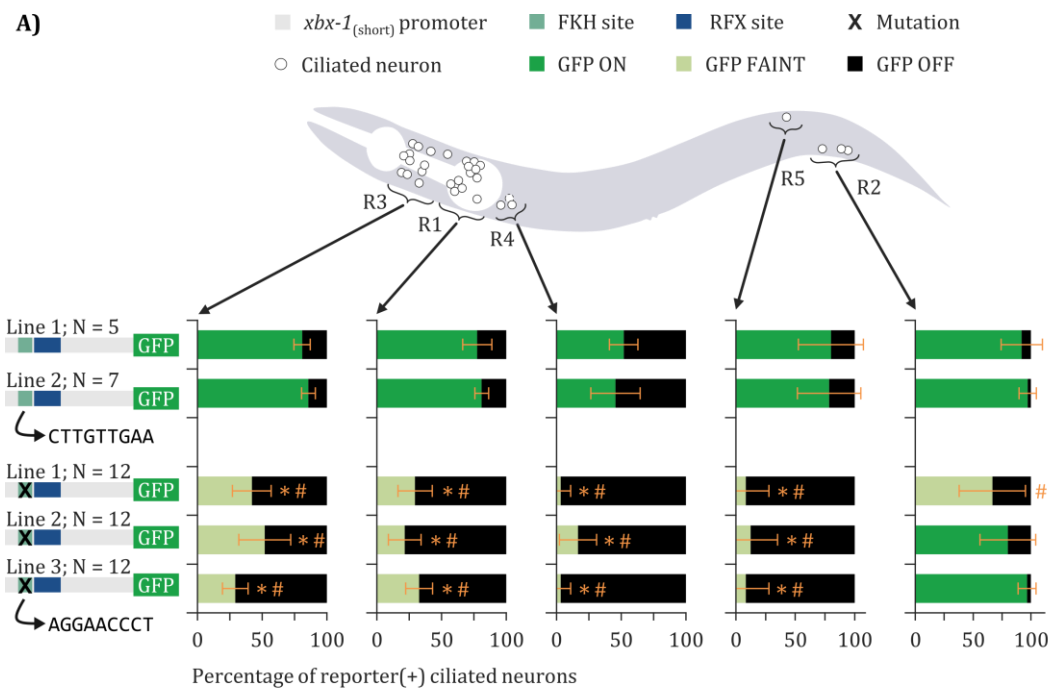
◄ **Figure R.58. Mutation of putative FKH site greatly reduces the expression of an *xbx-1* reporter.**
A) Top scheme shows a lateral left view of a *C. elegans* young adult hermaphrodite ciliated system. Note that AQR in the head locates in the right side of the animal and hence appears depicted in dashed lines. For each of the five anatomical regions in which the system was divided, corresponding graphs show the average percentage of ciliated neurons expressing a GFP-tagged reporter for the panciliary gene *xbx-1*. Top part of the graphs shows scorings for 2 independent lines of the wild type promoter. Bottom graphs represent results from 3 independent lines with a mutation in the putative FKH site. Error bars represent the standard deviation. Statistical significance was calculated through a two-tailed Student *t*-test once the variances between the two groups being compared were analysed. Significant differences are depicted with asterisks when comparing against Line 1 of the wild type reporter and with hashtags when the comparison is performed against the Line 2. Since each line carrying the mutated version of the reporter was compared against two control lines simultaneously, a Bonferroni correction was applied. **B)** Representative lateral views of young adult animals expressing both the wild type version (Line 1, top) and FKH site mutated version (Line 1, bottom) of the *xbx-1*(short) reporter. Final fluorescent images correspond to Z stacks projections of images at 1 micron distance taken under a vertical fluorescence microscope.

7. *fkh-8* mutants display defects in a wide range of sensory-mediated behaviours.

FKH-8 ChIP seq data and *in vivo* reporter analysis of structural ciliary genes suggested that *fkh-8* acts as a terminal selector of the sensory ciliome, together with *daf-19*. However, *fkh-8* mutant data so far only showed expression defects in a selected subpopulation of ciliated cells and for subsets of analysed reporters, whereas DiD staining of amphid and phasmid neurons was unaffected in *fkh-8* mutant backgrounds. Reporter gene expression analysis displays a very high cellular resolution; however, it lacks high throughput, being indeed very time-consuming and limiting the number of constructs that can be analysed. As an alternative to reporter gene expression, we decided to analyse sensory behaviours as a readout for the correct specification of sensory ciliated neurons. If, as we propose, *fkh-8* has a broad effect on cilia functionality, then sensory-mediated behaviours should be impaired in *fkh-8* mutant animals. All experimental procedures whose results appear in the following sections were performed by Ainara Esteve Serrano, a MSc student in our laboratory.

fkh-8 mutants show normal response to body touch.

First, to discard general motor or behavioural defects in *fkh-8* mutants, we decided to test behavioural paradigms not mediated by ciliated neurons. We hypothesized that if *fkh-8* was particularly required for the specification of the ciliated fate, sensory-mediated behaviours elicited through the action of non-ciliated neurons should not be affected in *fkh-8* mutant animals. To test this hypothesis, we

performed two types of mechanosensory assays: response to gentle and harsh touch.

When a wild type worm is gently touched with an eyebrow hair, the animal perceives the stimulus and exhibits an escape response. Neural circuitries eliciting this response depend on the anatomical location in which the stimulus is delivered. If the gentle touch is delivered to the anterior part of the animal's body, three non-ciliated mechanosensory neurons (namely, ALML, AMLR and AVM) elicit an escape response in which the young adult animal moves backward (Chalfie and Sulston 1981). On the other hand, if this gentle touch is delivered to the posterior part of the animal's body, two posterior non-ciliated mechanosensory neurons (namely, PLML and PLMR) induce a forward movement (Chalfie and Sulston 1981). When we performed the anterior gentle touch assay over a population of wild type worms, 92% of the animals exhibited a backward escape response (see **Figure R.59.A**). Both alleles of *fkh-8* showed a similar mean percentage of response (93% for *tm292* and 92% for *vlc43*), showing that *fkh-8* is dispensable for the correct functionality of ALM and AVM neurons. Nearly identical results were obtained when we performed the posterior gentle touch assay (see **Figure R.59.B**), hence proving that *fkh-8* is neither required for PLM functionality. As experimental control for defects in gentle touch, we used null *mec-3(e1338)* animals, in which all six touch-receptor neurons fail to properly differentiate (Chalfie and Sulston 1981). *mec-3(e1338)* animals have

been reported to be insensitive to both gentle and harsh touch assays (Li *et al.* 2011). As expected, we find that in our assays *mec-3(e1338)* mutants do not respond to gentle touch in the anterior nor the posterior part of the worm (**Figure R.59.A and B**).

Neural circuitry controlling behavioural responses to innocuous gentle touches differs from the circuits mediating noxious harsh touch. When a wild type worm is prodded with a platinum wire in the posterior part of its body, the animal perceives this aversive stimulus and exhibits an escape response, rapidly moving forward. This response is known to be mediated by the combined action of a ciliated (PDE) and a non-ciliated (PVD) mechanosensory neuron (Li *et al.* 2011). Virtually, the whole population of wild type worms reacted to the harsh touch, exhibiting the expected forward escape response (see **Figure R.59.C**). A similar response was obtained for the population of *fkh-8(tm292)* and *fkh-8(vlc43)* worms. Since we also found *fkh-8* expression in the PVD neurons and no behavioural defect was observed in a response mediated by their action, we conclude that *fkh-8* is not required for correct PVD or PDE functionality, at least mediating this response. However, we cannot discard that either PVD or PDE function is affected, as redundant roles have been reported for PVD and PDE neurons in the regulation of posterior harsh touch behavioural responses (Li *et al.* 2011). Indeed, our reporter analysis shows ciliary gene expression defects for the PDE neuron in *fkh-8* mutants. Similar to gentle touch, *mec-3(e1338)* animals failed to exhibit the appropriate harsh touch responses (see **Figure R.59.C**).

fkh-8 mutants display nose touch defects.

Once we determined the absence of defects in behaviours mediated by the action of non-ciliated mechanosensory neurons in *fkh-8* mutant animals, we aimed to specifically test the functionality of subpopulations of ciliated neurons.

When a free-wandering wild type worm encounters an obstacle in a nose-collision manner, the animal exhibits a mechanosensory reflex in which a backward movement is elicited. This reflex can be tested through the so-called nose touch assay, in which an eyelash attached to a pipette tip is placed in the path of an animal moving forward. Eight ciliated mechanosensory neurons have been identified to mediate the corresponding escape response, namely, ASH (2), FLP (2) and OLQ (4) neurons (Kaplan and Horvitz 1993).

As expected, virtually the whole population of wild type worms assayed elicited a backward movement when encountering the eyelash in this nose-collision manner (R.I. = 0.99) (see **Figure R.59.D**). Interestingly, a statistically significant fraction of the *fkh-8(tm292)* worms failed to exhibit this behaviour, hence decreasing the response index for the whole population to 0.81. This defect was yet increased in the population of *fkh-8(vlc43)* mutants, in which the response index dropped to an average value of 0.49. This difference, statistically significant when compared against both wild type and *fkh-8(tm292)* mutants, agrees with the stronger reporter expression defects found for *fkh-8(vlc43)* animals and supports either *fkh-8(tm292)* is a hypomorphic allele or it displays compensatory effects not present in

vlc43.

Laser-ablation experiments have shown that ASH neurons are in charge of roughly half of the normal response to the nose touch stimulus in wild type animals, followed by FLP and with OLQ playing only a minor role (Kaplan and Horvitz 1993). Since 63% of FLP LASER-ablated worms were found able to react to this assay (Kaplan and Horvitz 1993) and that percentage is higher than the one we observed for *fkh-8(vlc43)* null mutants, we conclude that ASH functionality is disrupted in *fkh-8* mutants; however, with this assay, we cannot discern if FLP and/or OLQ functionality is also affected.
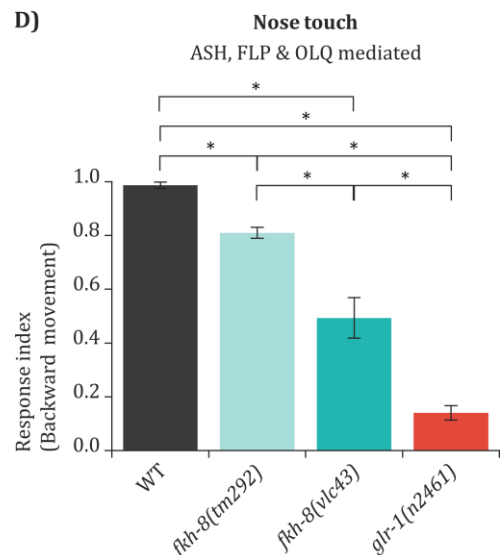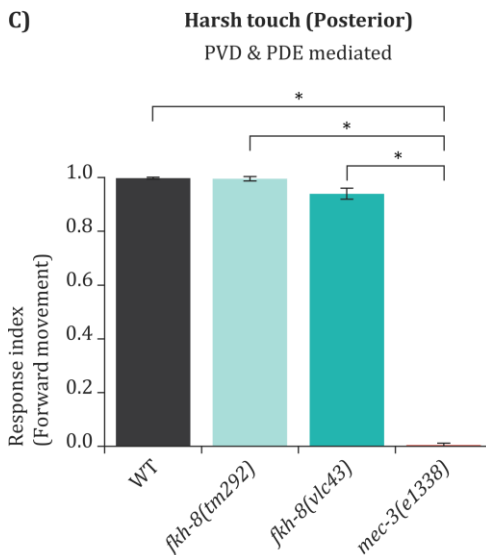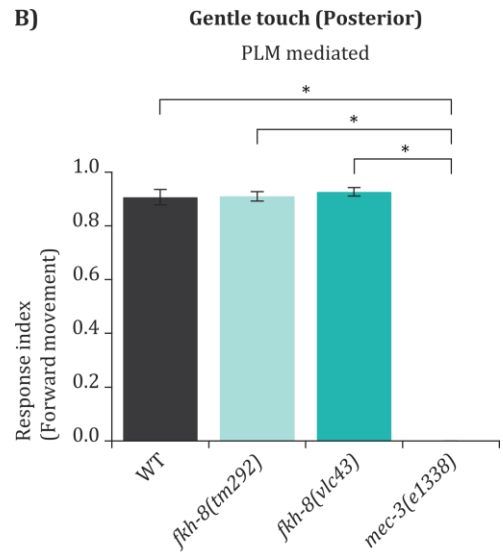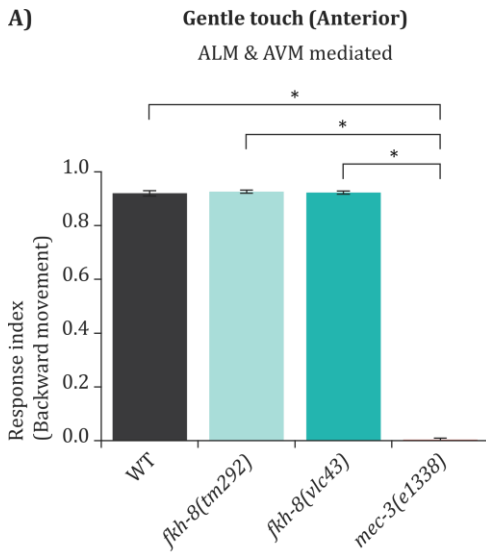
glr-1(n2461) mutant worms were used as the control strain, showing severe defects in this nose touch assay. These animals fail to express the glutamate receptor that the synaptic targets of the ASH need to elicit this nose touch reflex, hence disrupting the capability of those worms to react to nose touch but not affecting the responsiveness to high osmolarity that is also controlled by the polymodal ASH neurons (Hart et al. 1995). As expected, we found these worms fail to exhibit the corresponding backward movement, eliciting a response index with a mean value of 0.14 (see **Figure R.59.D**).

fkh-8(vlc43) mutants are defective in chemotaxis to a volatile attractant.

In *C. elegans*, the detection of environmental volatile cues has been found to be mainly mediated by the olfactory ciliated neurons AWA and AWC (Bargmann et al. 1993).

Diacetyl, a naturally occurring by-product of bacterial fermentation, has been long used as a volatile attractant for the worms and is mostly sensed by the AWA neurons (Bargmann *et al*. 1993). To assess if AWA functionality was compromised in *fkh-8*

mutant animals we analysed the response to diacetyl. Wild type worms exhibited the expected attraction response towards diacetyl, achieving a mean chemotaxis index of 0.86 for the whole population assayed (see **Figure R.60.A**). Although we could observe a drop in the response elicited by *fkh-8(tm292)* worms, the average value for their chemotaxis index (0.72) was not significantly different to wild type worms. This decrease in the response toward the chemoattractive diacetyl was further aggravated in null *fkh-8(vlc43)* mutant animals, in which the chemotaxis index dropped to an average value of 0.49. However, we found high variability among replicates of this assay; thus, comparison with wild type worms was not statistically significant. Although the trend was clearly visible, further experiments will be needed to clarify if there are diacetyl response defects in *fkh-8* mutant animals. Consequently, based on this experiment, we could not conclude a role for *fkh-8* in the correct functionality of the AWA neurons. *odr-10(ky32)* mutant animals, that fail to express the specific diacetyl receptor *odr-10* (Sengupta, Chou, and Bargmann 1996), were used as the control strain for behavioural defects. These animals scored a mean chemotaxis index of 0.07 (see **Figure R.60.A**).

◄ Figure R.59. *fkh8* mutants display cilia-related mechanosensory defects.
A) to C) Mechanosensory behaviours controlled by non-ciliated neurons (ALM, AVM, PLM and PVD), and corresponding to gentle and harsh touch, show no defects in neither allele of *fkh-8* mutant animals. **D)** Nose touch response, controlled by ASH, FLP and OLQ ciliated neurons, is defective in *fkh-8* mutant worms. Bars represent the mean response index from 3 independent replicates for each assayed genotype. Error bars correspond to the standard deviation among corresponding response indexes. Statistical significance was calculated through a *t*-test first considering the variances between the pairs being compared. Bonferroni correction was employed accounting for all 6 possible pair comparisons. For ease of reference, only statistically significant differences are depicted. N = 60 animals per genotype and 20 animals per replicate in A), B) and D). N = 90 animals per genotype and 30 animals per replicate in C). See **Table M1.12** for detailed quantification of each replicate.

Next, we tested *fkh-8* mutants' behaviour towards the olfactory cue 2-heptanone, a naturally occurring ketone, also released by bacteria, that is a known attractant for *C. elegans* (Bargmann *et al.* 1993).AWC neurons depend on the asymmetrical expression of *str-2,* either in the left or the right cell, for odour discrimination (Wes and Bargmann 2001). 2-heptanone has been found to be detected by the *str-2*-expressing AWC neuron (AWC[ON]) of the worm (Zhang *et al.* 2016). When we exposed wild type worms to the chemoattractive 2-heptanone, the population assayed reached, on average, a chemotaxis index of 0.73 (see **Figure M.60.B**). A slight decrease, not statistically significant, was observed for *fkh-8(tm292)* mutant animals (C.I = 0.65). Importantly, null *fkh-8(vlc43)* mutant worms failed to sense this volatile attractant since the mean chemotaxis index was 0.24. This difference, statistically significant when compared against both wild type and *fkh-8(tm292)* mutants was not different from the response exhibited by the population of *odr-1(n1936)* worms used as the negative control strain. These animals, which fail to express a receptor-class transmembrane guanylyl cyclase needed for olfaction (L'Etoile and Bargmann 2000), achieved a mean chemotaxis index of 0.11.

As a whole, these assays towards attractive volatile olfactory cues allowed us to demonstrate the implication of *fkh-8* in the correct functionality of odour-sensing ciliated neurons, namely AWC but, also probably, AWA. Moreover, the retained functionality exhibited by *fkh-8(tm292)* worms towards these stimuli further confirmed the hypomorphic nature of this allele.

fkh-8(vlc43) mutants show defects in the avoidance response to a volatile repellent.

C. elegans is also endorsed with the capability to sense and avoid volatile environmental cues that could be indicative of potential danger. Several compounds have been identified to elicit an aversive response from the worms (Bargmann *et al.* 1993), among them, 2-nonanone, a naturally occurring ketone. Repulsive 2-nonanone is mainly sensed by the AWB neurons (Troemel et al. 1997) and we decided to investigate if aversive behaviour toward volatile repellents were impaired in *fkh-8* mutant animals.

As expected, the wild type population exhibited a robust repulsion towards 2-nonanone, reaching a mean chemotaxis index of -0.74 (see **Figure M.60.C**). Although this response was less pronounced in *fkh-8(tm292)* mutant animals (C.I. = -0.57), no significant difference was found when comparing against the wild type worms. However, null *fkh-8(vlc43)* animals displayed more pronounced defects in the repulsive response towards volatile 2-nonanone, with a mean chemotaxis index of -0.46 that significantly differed from the value found for wild type animals. In this assay, a population of predicted null *che-2(e1033)* mutants were used as the negative control strain. These animals, defective in cilia formation and hence in sensory-mediated behaviours (Fujiwara, Ishihara, and Katsura 1999), performed with a mean chemotaxis index of -0.14.

These results proved AWB functionality to be compromised in *fkh-8* null animals and once more underscore the functional differences between the *tm292* and *vlc43* alleles.

A) Diacetyl Response
AWA mediated

B) 2-heptanone Response
AWC mediated

C) 2-nonanone Response
AWB mediated
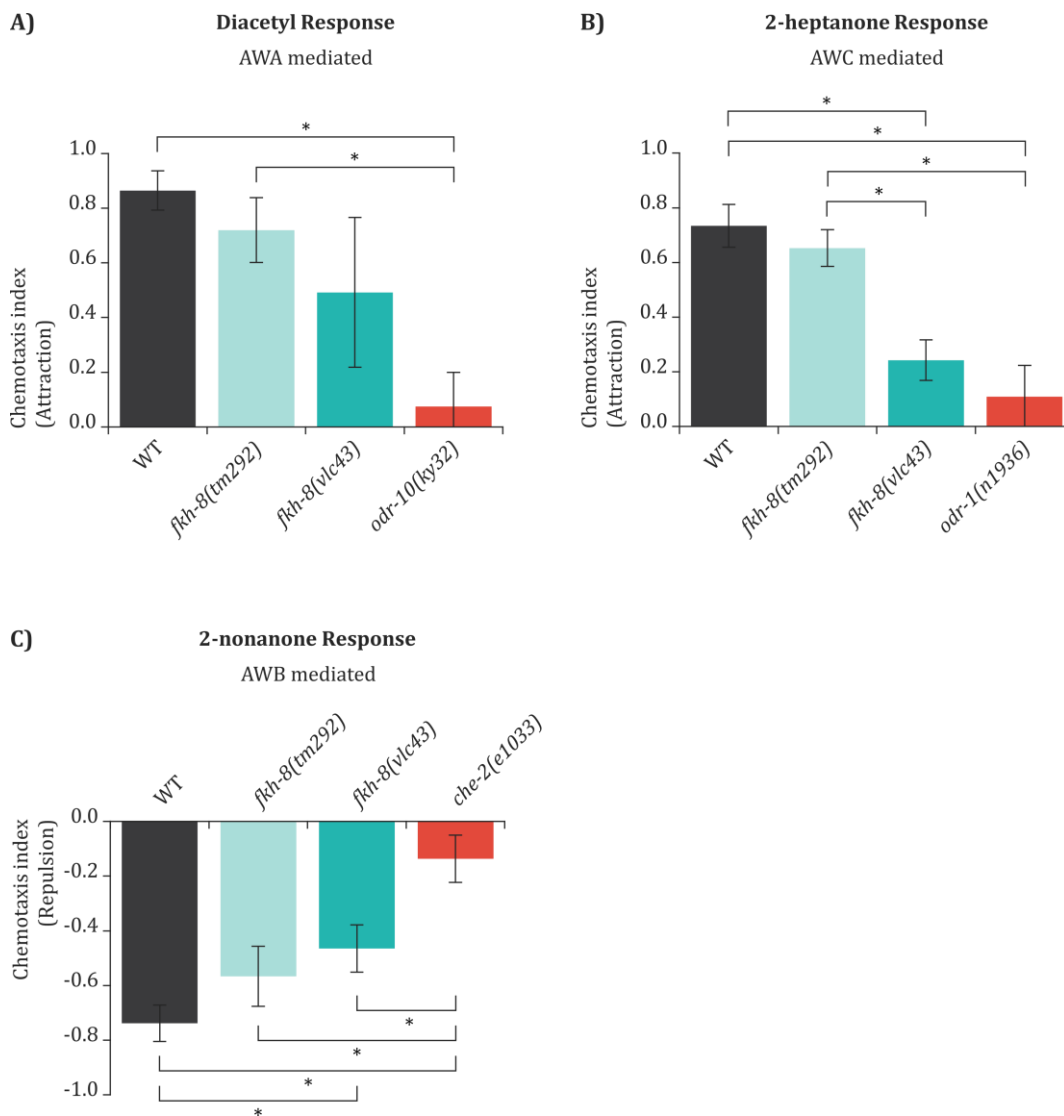
⮝ **Figure R.60. Absence of *fkh-8* impairs olfaction in *C. elegans*.**
A) Null *fkh-8* mutant animals display a decreasing trend in olfactory recognition of attractant diacetyl, suggesting functional defects in AWA neurons. **B)** Null *fkh-8* mutants fail to detect 2-heptanone, revealing an impairment of AWC functionality. **C)** Lack of *fkh-8* affects 2-nonanone aversive behaviour, indicative of deficiencies in AWB neurons. Bars represent here the mean response index from 3 independent replicates for each genotype assayed. Error bars correspond to the standard deviation among corresponding response indexes. Statistical significance was calculated through a *t*-test first considering the variances between the pairs being compared. Bonferroni correction was employed accounting for all 6 possible pair comparisons. For ease of reference, only statistically significant differences are depicted. N = 219 to 537 animals per genotype. See **Table M1.12** for detailed quantification of each replicate

fkh-8 mutants do not show chemotaxis defects to the water-soluble attractant NaCl.

Gustatory capabilities of *C. elegans* towards soluble chemical cues have been found to be mainly mediated by the ASE neurons (Bargmann and Horvitz 1991). Several salts have been long known to induce chemotactic behaviours in the worm and NaCl is one of the most widely used since both chlorine and Na⁺ cations have been proved to be attractants for the animals (Ward 1973). Laser-ablation experiments have demonstrated that Cl⁻ and Na⁺ ions are mainly sensed by the gustatory ASE neurons, with ADF, ASG and ASI playing minor, redundant roles (Bargmann and Horvitz 1991). We thus aimed to explore if this behaviour was impaired in *fkh-8* mutant animals.

A) NaCl Response
ASE, ADF, ASG & ASI mediated

B) Drop test (M13)

C) Drop test (SDS)
ASH & ASK mediated

D) Drop test (Cu²⁺)
ADL, ASE & ASH mediated
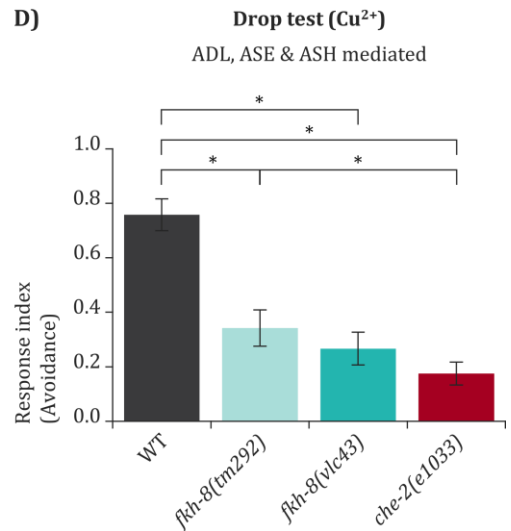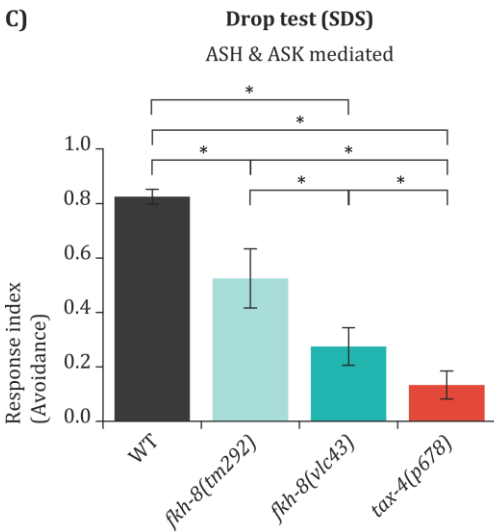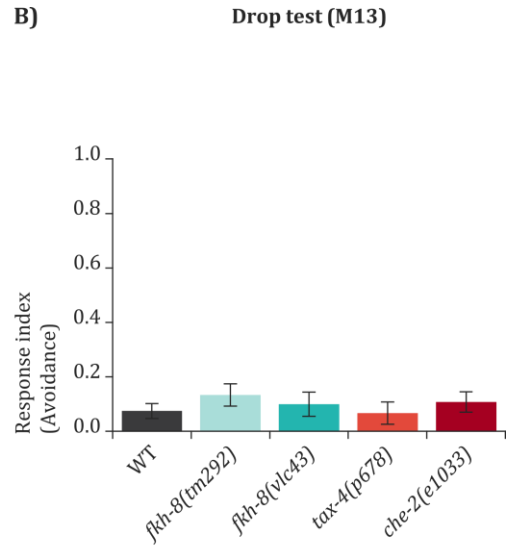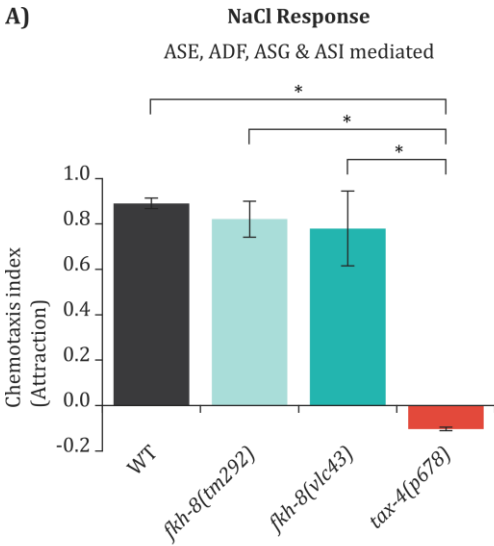
As expected, wild type worms exhibited high attraction (C.I. = 0.89) towards NaCl (see **Figure R.61.A**). Similar numbers were achieved by both *fkh-8* mutant genotypes, with mean chemotaxis indexes of 0.82 for *fkh-8(tm292)* and 0.78 *for fkh-8(vlc43)* mutants, respectively. This slight decrease is not statistically significant, even in the case of *vlc43*, due to high variability among assays. *tax-4(p678)* mutant animals, that fail to express an α subunit of a cyclic nucleotide-gated channel, were used as the control strain for behavioural defects. These worms, which have been found to be defective in thermotaxis, chemotaxis and *dauer* formation (Komatsu *et al*. 1996), were, as expected, defective in the response towards NaCl in this assay, scoring a mean population chemotaxis index of –0.10.

These results suggest that either there are no major defects in the functionality of the ASE neurons, at least to mediate NaCl attraction, or that those possible defects can be compensated by the combined, redundant actions from the rest of the ciliated neurons mediating the attracting response towards NaCl.

fkh-8 mutants display defects in the avoidance response to water-soluble toxic substances.

Several subsets of ciliated neurons have been implicated in the detection of toxic-water soluble substances by the drop test (Hilliard et al. 2002), in which the animal is exposed to the chemical by delivering a small drop near its tail. The drop assay is considered to be successfully performed if drops containing non-toxic substances are delivered without altering the ongoing movement of a worm. Consequently, we first assayed the response of the animals towards a neutral stimulus (M13 buffer) to discard a differential response due to mechanical sensory defects and not to toxic compound sensory defects. We found that none of the strains exhibited an escape response from this neutral compound, with response indexes ranging from 0.07 to 0.13 (**Figure R.61.B**). Thus, we proceeded to assess the responses towards water-soluble toxic substances.

◄ **Figure R.61. *fkh-8* mutants show defects in toxic gustatory-mediated behaviours.**
A) *fkh-8* mutant animals do not display gustatory defects towards attractive NaCl. **B)** Drop assay in which a neutral stimulus is used (M13 buffer) does not produce response differences among the different genotypes being assayed. **C)** *fkh-8* mutants fail to recognize aversive SDS, revealing an impairment in ASH and ASK functionality. **D)** *fkh-8* mutations severely affect worm's capability to recognize toxic copper ions, revealing deficiencies in ADL, ASE and ASH neuron functions. Bars represent the mean response index from 3 independent replicates for each assayed genotype. Error bars correspond to standard deviation among replicates. Statistical significance was calculated through a *t*-test first considering the variances between the pairs being compared. Bonferroni correction was employed accounting for all possible pair comparisons. For ease of reference, only statistically significant differences are depicted. N = 163 to 295 animals per genotype in A) and N=30 animals and genotype in the rest of the assays and 5 animals per replicate. See **Table M1.12** for detailed quantification of each replicate

The amphid ASH is the main neuron type mediating SDS avoidance followed by a secondary contribution of the ASK neurons (Hilliard et al. 2002). Knowing this, we wanted to assay if behaviours elicited by the action of those neurons could be affected in *fkh-8* mutant animals. As expected, wild type worms displayed a huge avoidance reaction towards toxic SDS (see **Figure R.61.C**), with a mean response index of 0.83 for the whole population assayed. Interestingly, this reaction was significantly diminished in *fkh-8(tm292)* mutants, in which a mean response index of 0.53 was observed. Null *fkh-8(vlc43)* worms proved to be even less reactive to the presence of SDS, with a response index of 0.28, a value significantly different from both wild type and *fkh-8(tm292)* animals. These results prove that ASH and ASK functionality is compromised in *fkh-8* mutant animals, and that these defects can be partially restored by the remaining activity of the *fkh-8(tm292)* allele. In this assay, *tax-4(p678)* animals, mutants for the α subunit of a cyclic nucleotide-gated channel, were again used as the control strain for behavioural defects, displaying on average a response index of 0.13.

Next, we tested avoidance to copper. This response is known to be mediated by the redundant action of ADL, ASE and ASH neurons since laser ablation of all six neurons is needed to abolish escape response towards toxic copper and cadmium ions (Sambongi *et al.* 1999). Hence, we decided to employ the drop test to assay the compromised functionality of the ADL, ASE and ASH neurons

towards copper ions. As expected, wild type worms strongly reacted against the presence of copper (see **Figure R.61.D**), displaying a mean response index of 0.76. Both *fkh-8* mutant strains significantly failed to avoid copper, since their response indexes dropped to 0.34 in the case of *fkh-8(tm292)* animals and to 0.27 for *fkh-8(vlc43)* worms. *che-2(e1033)* animals, used as the control strain for behavioural defects, failed to show an escape response toward copper ions, exhibiting a mean response index of 0.18.

Thus, these results proved *fkh-8* to be necessary for the correct functionality of all six ADL, ASE and ASH neurons regarding the aversive response against copper. Moreover, these results agreed with our RNA-seq data, since 4 genes that we found related to the copper response through the GO term GO:004668 (namely, the JUN kinase *kgb-1*, the phosphatase *vhp-1* plus the nematode-specific genes *kreg-1* and *lys-3*), all were significantly downregulated in *fkh-8(tm292)* compared to wild type worms.

As a whole, this set of behavioural analyses allowed us to expand the repertoire of sensory ciliated neurons affected in *fkh-8* mutant animals. In the preceding sections we have shown evidence of *fkh-8* mutants displaying defects in a wide range of sensory-mediated behaviours elicited by the ASE, ADL, ASH, AWC, AWB, FLP, OLQ, ASK, AWA neurons. Together with the results presented in previous sections, we conclude *fkh-8* is required for the correct specification of most, if not all, ciliated neurons of *C. elegans*.

8. *fkh-8* & *daf-19* exhibit cross-regulation and synergistic effects.

The results presented so far indicate that, similar to *daf-19*, *fkh-8* acts as a terminal selector of (at least) the structural ciliome. Consequently, we aimed to assess epistatic relationships between these two transcription factors.

First, we analysed the presence of reciprocal binding sites in each locus. Taking advantage of the RFX-related motifs we retrieved during the study of putative regulatory sequences of cilium-related genes (see **Figure M.5**), we scanned the *fkh-8* locus for the presence of *daf-19* binding sites. Three different sites could be identified (see **Figure R.62** and **Table R.22**). Due to its nearly palindromic nature, two matches for Motif 1 were located overlapping 202 base pairs upstream of the start codon of *fkh-8*. A second

site matching Motif 3 was found 526 base pairs upstream the ATG. Finally, a third site was located matching Motif 4 492 base pairs downstream the stop codon. We then proceed analogously over the locus of *daf-19* scanning its sequence for the presence of *fkh-8* binding sites profiting from the ChIP-seq data set. Interestingly, up to five distinct ChIP-seq peaks could be found overlapping different locations of the *daf-19* locus (see **Figure R.63**). In accordance with the terminal selector model, several putative binding sites for *daf-19* were also found within the *daf-19* locus, suggesting autoregulatory actions of *daf-19* (see **Figure R.63** & **Table 3.23**). Remarkably, these putative X-boxes were often located overlapping the regions in which FKH-8 ChIP-seq peaks were found.

II:6192064-6189064   3 kb
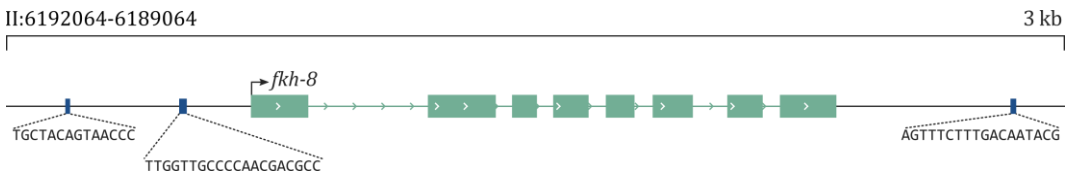
▲ Figure R.62. Putative binding sites for *daf-19* can be found in the *fkh-8* locus.
Turquoise boxes represent FKH-8 exons. Dark blue boxes illustrate putative X-box sites. Genomic coordinates are indicated in the top left corner. Note that *fkh-8* runs through the minus strand of the DNA.

Next, we used our RNA-seq data sets to check the expression levels of *fkh-8* and *daf-19* in the *daf-19*, *daf-12* double mutants and *fkh-8* mutants, respectively. We found that expression of *fkh-8* was increased by more than a 3.25 fold in *daf-19*, *daf-12* double mutants when compared against its value in wild type worms (p-value = 1.30E-02 and q-value

(correcting for multiple comparisons) = 3.59E-02). Conversely, *daf-19* expression was also found to be increased in *fkh-8(tm292)* mutant animals. Raw comparison of expression values showed a 2.33 fold increase between mutant and wild type animals (p-value = 3.66E-02); however, as it happened with the rest of the differentially

expressed genes in this experiment, the possibility of false positives overcame the threshold value of 5%, reaching for this particular case 27.44%. Nonetheless, on the basis of this evidence, and the evidence for direct binding, we decided to further investigate the possibility of a transcriptional cross-regulation between the two factors using *in vivo* reporter analysis.

▼ **Table R.22. Putative RFX sites are found in *fkh-8* regulatory sequences.**
PWMs matching experimentally validated binding sites for RFX transcription factors (see **Figure R.5**) were used to scan a 5,663 base pairs sequence containing the whole coding sequence of *fkh-8* plus 2,000 nucleotides of upstream and downstream regions. Position is referred to the ATG starting codon, where upstream locations are indicated through the sign minus. Binomial probability for site occurrence and binomial statistical significance obtained when considering all non-coding sequences of *C. elegans* are indicated. Note that within the binomial model an event occurrence is significant if significance is equal or greater than zero.

PWM	Target Sequence	Location	Position	P-value	Significance
Motif 3	TGCTACAGTAACCC	Promoter	-526	5.80E-05	4.237
Motif 4	CGTATTGTCAAAGAAACT	Terminator	2156	7.50E-05	4.127
Motif 1	GGCGTCGTTGGGGCAACCAA	Promoter	-202	9.10E-05	4.043
Motif 1	TTGGTTGCCCCAACGACGCC	Promoter	-202	9.10E-05	4.015
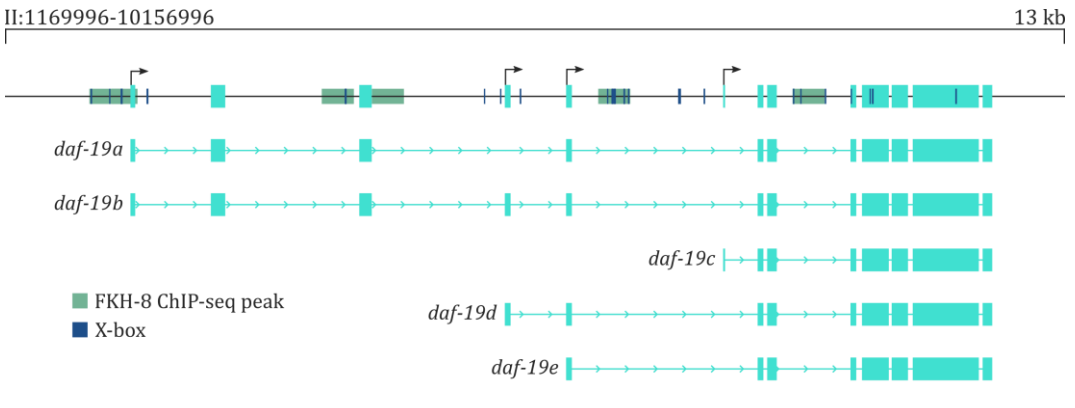
▲ **Figure R.63. *fkh-8* binds to putative regulatory sequences of the *daf-19 locus*.**
Four different promoters produce five different isoforms of the *C. elegans* transcription factor *daf-19*, whose exons appear here as aquamarine boxes. Turquoise boxes illustrate the regions in which FKH-8 ChIP-seq peaks are found. Genomic coordinates are indicated in the top left corner; note that *daf-19* runs through the minus strand of the DNA. Two peaks overlap in the putative promoter region of *daf-19* isoforms *a* and *b,* indicative of two different *fkh-8* binding events separated by roughly 230 base pairs. In accordance with the terminal selector model, several X-boxes (represented as vertical dark blue lines on top of the figure) are found within the *daf-19* locus, suggesting auto-regulation of *daf-19*. Isoform names vary in the bibliography; here, the WormBase nomenclature is shown. Due to its cilia-related functions, *daf-19d* is cited in several publications as *daf-19c* (for function in "<u>c</u>ilia"). Analogously, *daf-19c*, which control male-specific cilia features, is usually termed *daf-19m* (for function in "male <u>m</u>ating").

▼ Table R.23. Putative RFX sites are found in *daf-19* regulatory sequences.

PWMs matching experimentally validated binding sites for RFX transcription factors (see **Figure R.5**) were used to scan an 11,369 base pairs sequence containing the whole coding sequence of *daf-19* plus 2,000 nucleotides of upstream and downstream regions. Position is referred to the ATG starting codon, where upstream locations are indicated through the sign minus. Binomial probability for site occurrence and binomial statistical significance obtained when considering all non-coding sequences of *C. elegans* are indicated. Note that within the binomial model an event occurrence is significant if significance is equal or greater than zero.

PWM	Target Sequence	Location	Position	P-value	Significance
Motif 3	TTGCATAGGAACGG	Promoter	-507	8.4E-05	4.074
Motif 4	CTCGTTGCCTGGGGGAGT	Promoter	-266	7.5E-05	4.127
Motif 4	TTGGTTTCCATGGAAACT	Promoter	-111	5.3E-07	6.274
Motif 1	GTAGTTTCCATGGAAACCAA	Promoter	-111	5.6E-07	6.251
Motif 1	TTGGTTTCCATGGAAACTAC	Promoter	-111	6.1E-07	6.216
Motif 2	TTGGTTTCCATGGAAACT	Promoter	-111	2.6E-06	5.578
Motif 3	TTCCATGGAAACCA	Promoter	-110	3.0E-06	5.517
Motif 2	GTAGTTTCCATGGAAACC	Promoter	-109	9.9E-07	6.003
Motif 4	GTAGTTTCCATGGAAACC	Promoter	-109	1.1E-06	5.962
Motif 3	TTCCATGGAAACTA	Promoter	-106	4.5E-06	5.344
Motif 4	TGGGTGGCCATAGGAACA	Intron	229	1.1E-05	4.947
Motif 3	TCCTATGGCCACCC	Intron	230	2.6E-05	4.580
Motif 3	GGCCATAGGAACAT	Intron	234	1.4E-05	4.846
Motif 4	TGCGTTTCCATGGATACG	Intron	2837	1.1E-07	6.939
Motif 1	TGCGTTTCCATGGATACGGG	Intron	2837	4.7E-07	6.325
Motif 1	CCCGTATCCATGGAAACGCA	Intron	2837	5.2E-07	6.286
Motif 2	TGCGTTTCCATGGATACG	Intron	2837	2.1E-06	5.684
Motif 3	ATCCATGGAAACGC	Intron	2838	5.4E-06	5.269
Motif 4	CCCGTATCCATGGAAACG	Intron	2839	1.5E-07	6.824
Motif 2	CCCGTATCCATGGAAACG	Intron	2839	3.2E-07	6.490
Motif 3	TTCCATGGATACGG	Intron	2842	2.0E-05	4.696
Motif 2	CACAGTGCCCTAACGACT	Intron	4666	9.6E-05	4.019
Motif 2	ACGGCTTCCGTCACAACC	Intron	4878	7.9E-05	4.101
Motif 4	CCTGTAGCTATGTGAATT	Intron	5139	5.3E-05	4.280
Motif 3	TCACATAGCTACAG	Intron	5140	2.2E-05	4.664
Motif 2	ACCATCTCCAAGGTGACA	Intron	6285	1.0E-04	3.995
Motif 4	ACGGTTCCCAAGGAGATC	Intron	6347	9.3E-06	5.034
Motif 1	ACGGTTCCCAAGGAGATCCG	Intron	6347	3.2E-05	4.488

PWM	Target Sequence	Location	Position	P-value	Significance
Motif 1	CGGATCTCCTTGGGAACCGT	Intron	6347	3.4E-05	4.464
Motif 3	CTCCTTGGGAACCG	Intron	6348	8.0E-05	4.097
Motif 1	CGGGTGGACACGACGACGAC	Intron	6365	9.3E-06	5.031
Motif 1	GTCGTCGTCGTGTCCACCCG	Intron	6365	1.1E-05	4.969
Motif 1	GACGACCCATGGATACCGG	Intron	6379	8.1E-05	4.092
Motif 1	CCGGTATCCATGGGGTCGTC	Intron	6379	8.5E-05	4.069
Motif 4	CCGGTATCCATGGGGTCG	Intron	6381	1.5E-05	4.838
Motif 3	CCCCATGGATACCG	Intron	6384	1.1E-05	4.949
Motif 1	ACGGTGGCGATGGAAACCGC	Intron	6503	4.6E-06	5.334
Motif 1	GCGGTTTCCATCGCCACCGT	Intron	6503	5.0E-06	5.300
Motif 4	ACGGTGGCGATGGAAACC	Intron	6503	2.1E-05	4.671
Motif 2	ACGGTGGCGATGGAAACC	Intron	6503	5.8E-05	4.234
Motif 2	GCGGTTTCCATCGCCACC	Intron	6505	4.5E-05	4.344
Motif 4	GCGGTTTCCATCGCCACC	Intron	6505	1.0E-04	3.983
Motif 4	GAGGTTGTCATGCATATA	Intron	6561	7.0E-06	5.153
Motif 2	CATATATGCATGACAACC	Intron	6563	7.4E-05	4.128
Motif 1	ACGGTTACCAAGACAATTAA	Intron	7219	3.2E-05	4.488
Motif 1	TTAATTGTCTTGGTAACCGT	Intron	7219	3.4E-05	4.464
Motif 2	ACGGTTACCAAGACAATT	Intron	7221	2.1E-06	5.684
Motif 1	CCGATCGTCTTGGTAACAGC	Intron	7239	1.8E-05	4.741
Motif 1	GCTGTTACCAAGACGATCGG	Intron	7239	1.8E-05	4.741
Motif 2	GCTGTTACCAAGACGATC	Intron	7241	6.5E-06	5.185
Motif 2	CCTGTAACCATGACGGCG	Intron	7557	1.2E-05	4.931
Motif 1	GGCGCCGTCATGGTTACAGG	Intron	7557	8.5E-05	4.069
Motif 1	CCTGTAACCATGACGGCGCC	Intron	7557	9.7E-05	4.015
Motif 4	GGCGCCGTCATGGTTACA	Intron	7559	6.3E-05	4.201
Motif 1	CGGGTTGCCTAGGTGACAAG	Intron	8731	2.6E-06	5.582
Motif 1	CTTGTCACCTAGGCAACCCG	Intron	8731	3.0E-06	5.521
Motif 2	CTTGTCACCTAGGCAACC	Intron	8731	4.9E-06	5.314
Motif 4	CGGGTTGCCTAGGTGACA	Intron	8733	1.2E-06	5.932
Motif 2	CGGGTTGCCTAGGTGACA	Intron	8733	4.2E-05	4.375
Motif 4	GGTGGAGTTATGGTTACA	Intron	8828	7.1E-05	4.151
Motif 3	GCCAATAGCAACCA	Intron	9153	6.3E-06	5.202
Motif 2	CCAGTTCACCTGAAAACC	Intron/Exon	9492	9.6E-05	4.019
Motif 1	GCGGGTGCCCAGCCGTCGGG	Exon	9738	7.6E-05	4.120

PWM	Target Sequence	Location	Position	P-value	Significance
Motif 1	CCCGACGGCTGGGCACCCGC	Exon	9738	8.5E-05	4.069
Motif 2	ACCACTACTATGGCATCC	Exon	9771	4.5E-05	4.344
Motif 3	TACTATGGCATCCG	Exon	9776	3.0E-05	4.518
Motif 4	GTGGTTGCGATAGTGTCA	Exon	10871	9.9E-05	4.004
Motif 3	CACTATCGCAACCA	Exon	10872	4.8E-05	4.322
Motif 3	TTGCATAGGAACGG	Promoter	-507	8.4E-05	4.074

fkh-8(tm292) mutation has no major effect over the expression of *daf-19.*

To experimentally assess cross-regulation effects between *fkh-8* and *daf-19* we aimed to analyse reporter expression for each TF in their reciprocal mutant backgrounds. First, we analysed the effect of *fkh-8* mutation over the expression of *daf-19*. To accurately recapitulate *daf-19* endogenous expression, we generated a strain (NFB672) containing the fosmid WRM0622dH09, which includes the whole *daf-19* locus plus several up and downstream genes, spanning approximately 39 kilobases of chromosome II. In this strain, DAF-19 is tagged in its Carboxyl terminus (Sarov et al. 2012) with the TY1 epitope, enhanced GFP and three copies of the FLAG epitope (*fkh-8::TY1::EGFP::3x FLAG*) (see **Figure R.64**). As previously stated, *daf-19* encodes at least 5 different isoforms whose expression has been shown to be restricted to distinct subpopulations of cells. Whereas long *daf-19* isoforms (*a* and *b*) are known to be expressed in non-ciliated neurons, *daf-19d* has been shown to be expressed within the ciliated system of the worm (Senti and Swoboda 2008). Accordingly, the fosmid-based expression pattern of *daf-19* in wild type worms is virtually pan-neuronal, as expected from the combined expression of all its isoforms provided by a construct tagged in the Carboxyl terminal end that all isoforms share.
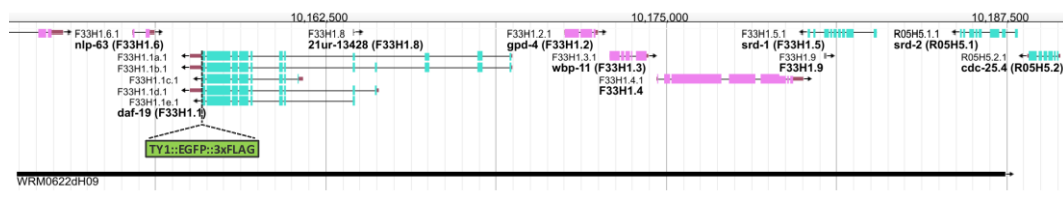
⬆ **Figure R.64. Structure of the fosmid reporter for *daf-19*.**
DAF-19 expression was assessed through a fosmid-based reporter strain carrying the fosmid WRM0622dH09. In the context of roughly 39 kilobases fosmid containing several genes upstream of and downstream from *daf-19*, the C-terminal end of the *daf-19* locus was engineered to contain the TY1 peptide, enhanced GFP and three copies of the FLAG epitope.

The pan-neuronal expression pattern of this *daf-19* fosmid reporter was similar in *fkh-8(tm292)* mutants. However, differences in intensity levels were not assessed. Additionally, it is possible that low penetrant effects in *daf-19* expression in only a few neurons in *fkh-8* mutants could be missed in the context of this broad pan-neuronal expression. Thus, we specifically tested the expression of *daf-19* in *fkh-8* mutants in two particular subpopulations of ciliated neurons: those integrating the dopaminergic system (CEP, ADE and PDE neurons) and those with the capability to uptake the fluorescent lipophilic dye DiD (ASK, ADL, ASI, AWB, ASH, ASJ, PHA and PHB neurons), hence accounting for roughly 46% of the whole ciliated system.

As expected, we found *daf-19* to be expressed in the whole dopaminergic system of the worm in a wild type genetic background and no significant differences were observed as a result of the *fkh-8(tm292)* deletion; thus, *fkh-8* does not seem to regulate *daf-19* expression at least within this subpopulation of ciliated neurons (see **Figure R.65.A**).

Surprisingly, fosmid-based expression of *daf-19* in wild type animals was not detected in half of the DiD-positive amphid neurons. Robust and bright expression was consistently seen in AWB neurons, with a less intense but reliable expression observed in the ASK and ASI neurons. *daf-19* expression was fairly uncommon and dim within the ASH and ASJ neurons, being found in less than 9 and 18% of the worms, respectively. Finally, no expression was spotted in ADL. *fkh-8(tm292)* mutants showed a similar expression pattern, hence suggesting that *fkh-8* does not regulate *daf-19* expression at least within the

ASK, ASI and AWB ciliated neurons. Dye-filling of PHA and PHB phasmid neurons was shown to be affected in both genetic backgrounds, a common defect that we have observed for animals cultured at 25 Celsius degrees. *daf-19* expression was generally dim in both phasmid neuron classes and no differences could be established between the two genotypes being compared (see **Figure R.65.B**).

As a whole, these experiments indicate that *fkh-8* does not regulate transcription of *daf-19,* at least in the subpopulation of ciliated neurons assayed; however, they are in contrast with the RNA-seq results that suggested a repressive action for *fkh-8* on *daf-19*. It is possible that slight increases in the GFP signal in the ciliated neurons could be missed with our approach. Alternatively, it is possible that *fkh-8* repression is specific for some isoforms and not others. As our reporter labels all *daf-19* isoforms, we cannot assess this possibility. Finally, this analysis was performed with the hypomorphic allele, which might display enough FKH-8 activity to mask possible regulation on *daf-19* transcription.

Long *daf-19* isoforms repress *fkh-8* expression in non-ciliated neurons.

As seen in the preceding sections, fosmid-based expression of *fkh-8* was mainly observed within the whole ciliated system of the wild type *C. elegans* hermaphrodite, with minor contributions from a few non-ciliated cells (see **Figures R.19** to **R.22**). On the other hand, our RNA-seq data showed a significant upregulation of *fkh-8* in double *daf-19/daf-12* null mutants. Thus, we decided to assess the

in vivo expression of *fkh-8* in such genetic background using its fosmid-based reporter.

A)

dat-1 ON   daf-19 ON   OFF

CEPV   CEPD   ADE   PDE

WT

fkh-8(tm292)II

Percentage of total neurons   N = 10 animals

B)

DiD ON   daf-19 ON   OFF

DiD(+) Amphids   DiD(+) Phasmids

WT

fkh-8(tm292)II

Percentage of total neurons   N = 10 animals
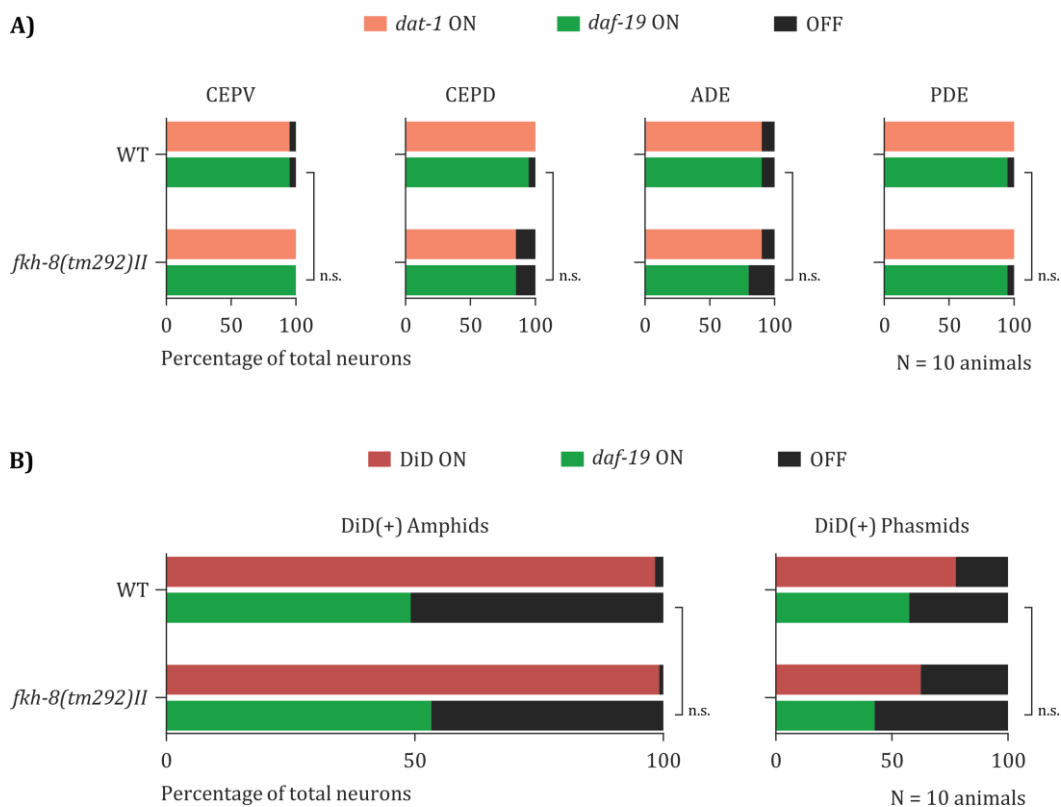
▲ **Figure R.65.** *fkh-8(tm292)* **mutation does not affect the expression of** *daf-19* **in subpopulations of ciliated neurons.**

A) Expression of a *daf-19* reporter was compared between wild type and *fkh-8(tm292)* mutant animals carrying an integrated reporter for the dopaminergic marker *dat-1* and a fosmid-based extrachromosomal array expressing *daf-19::GFP*. Performance of both reporters – expressed as the percentage of total neurons – is represented by coloured bars for each subpopulation of dopaminergic neurons, portraying *dat-1* in red and *daf-19* in green. Absence of reporter expression is depicted in black. Statistical significance for the performance of the *daf-19* reporter between both genetic backgrounds was calculated through the Fisher exact test and deemed non-significant (n.s.). **B)** Expression of a *daf-19* reporter was compared in the subpopulation of ciliated DiD-positive neurons of *C. elegans*, (corresponding to amphids ASK, ADL, ASI, AWB, ASH and ASJ and phasmids PHA and PHB) between wild type and *fkh-8(tm292)* mutant animals carrying a fosmid-based extrachromosomal array expressing *daf-19::GFP*. Co-expression of both markers – indicated as the percentage of total neurons – is represented by coloured bars for each sub-population of DiD-positive neurons, portraying the fluorescent lipophilic dye DiD in deep red and *daf-19* in green. Absence of reporter expression is depicted in black. Statistical significance for the performance of the *daf-19* reporter between both genetic backgrounds was calculated through the Fisher exact test and deemed non-significant (n.s.).

In agreement with our RNA-seq data, reporter expression of *fkh-8* was dramatically upregulated in double *daf-19/daf-12* null mutants, switching from the pan-ciliated expression pattern observed in wild type animals to a seemingly pan-neuronal expression pattern. Since *daf-19* isoforms are known to have a differential expression pattern in which long *a* and *b* isoforms are expressed in non-ciliated neurons, we wondered if the upregulation of *fkh-8* in the absence of all isoforms of *daf-19* was indeed the reflection of the repressing action of DAF-19A/B. To test this hypothesis, we analysed the expression pattern of the *fkh-8* fosmid-based reporter in the context of the specific absence of DAF-19A/B (see **Figure R.66.A**). As seen in **Figure R.66.B,** pan-neuronal deregulation of *fkh-8* was also observed in animals specifically lacking the long *daf-19a/b* isoforms, suggesting DAF-19A/B repress *fkh-8* expression.

Conversely, once the repressive role of DAF-19A/B upon *fkh-8* expression had been observed within the non-ciliated neurons, we wondered if DAF-19C could be exerting some activating actions upon *fkh-8* expression within the ciliated neurons themselves. To test this hypothesis, we decided to focus on an easily identifiable ciliated neuron such as the PDE. These neurons locate halfway between the vulva and the tail and are only surrounded by the bilateral PVD neurons anteriorly and the unilateral PVM neuron at the left side of the animal posteriorly. We had previously determined that *fkh-8* is expressed in both PDE and PVD neurons in a wild type genetic background (see **Figure 3.21**). As seen in **Figure R.66.C**, *fkh-8* expression in the PDE neurons resulted indistinguishable between wild type and double *daf-19/daf-12* null mutants. However, and as expected for the deregulation of *fkh-8* in non-ciliated neurons, the PVM neuron was shown to express *fkh-8* in the absence of all the isoforms of *daf-19* (see **Figure R.66.D**).

➤ **Figure R.66. Long *daf-19* isoforms repress *fkh-8* expression in non-ciliated neurons**
A) Schematics portraying both *daf-19* mutations used in this section. In the *of5* allele, a 12 base pairs deletion removes the original ATG. In this deletion, a putative new ATG appears in a different reading frame that hits a STOP codon in exon 2. Classical *m86* allele creates a stop codon in an exon present in all known isoforms of *daf-19*. **B)** Representative head lateral views from of young adult hermaphrodites showing *fkh-8* reporter expression in wild type (left), null *daf-19* mutants (middle) and mutants for the long a/b isoforms of *daf-19* (right). Final fluorescent images correspond to Z stacks projections of images at 1 micron distance taken under a vertical fluorescence microscope. **C)** Expression of a *fkh-8* reporter was compared in the subpopulation of ciliated dopaminergic neurons of *C. elegans* between wild type and double null mutant animals for *daf-12* and *daf-19* both carrying a fosmid-based extrachromosomal array expressing *fkh-8::EGFP*. Expression of both markers – depicted as the percentage of total neurons – is represented by coloured bars for each sub-population of *dat-1* positive neurons, portraying the dopaminergic neurons in deep red and *fkh-8* expression in green. Absence of reporter expression is depicted in black. **D)** Representative mid body lateral views of young adult hermaphrodites showing *fkh-8* reporter expression in wild type (top), and null *daf-19* mutants (bottom) and mutants. Lack of *daf-19* derepresses *fkh-8* expression in the PVM neuron.

As a whole, these experiments demonstrate that long *daf-19a/b* isoforms repress *fkh-8* expression in non-ciliated neurons hence ensuring *fkh-8* expression is limited to the ciliated system of *C. elegans*.

A)

daf-19

of5
12 bp deletion

m86
R > STOP

1 kb

B)

| wild type | *daf-19(m86); daf-12(sa204)* | *daf-19a/b(of5)* |

fkh-8

20 μm

C)

■ *dat-1* ON   ■ *fkh-8* ON   ■ Reporter OFF

PDE neurons

wild type

daf-19(m86)II; daf-12(sa204)X

n.s.

0   50   100

Percentage of total neurons

N = 53 animals

D)

| DIC | *fkh-8* | *dat-1* | Merged |

wild type

10 μm

PVDL   PDEL

daf-19(m86); daf-12(sa204)

10 μm

PDEL

PVDL   PVM

223
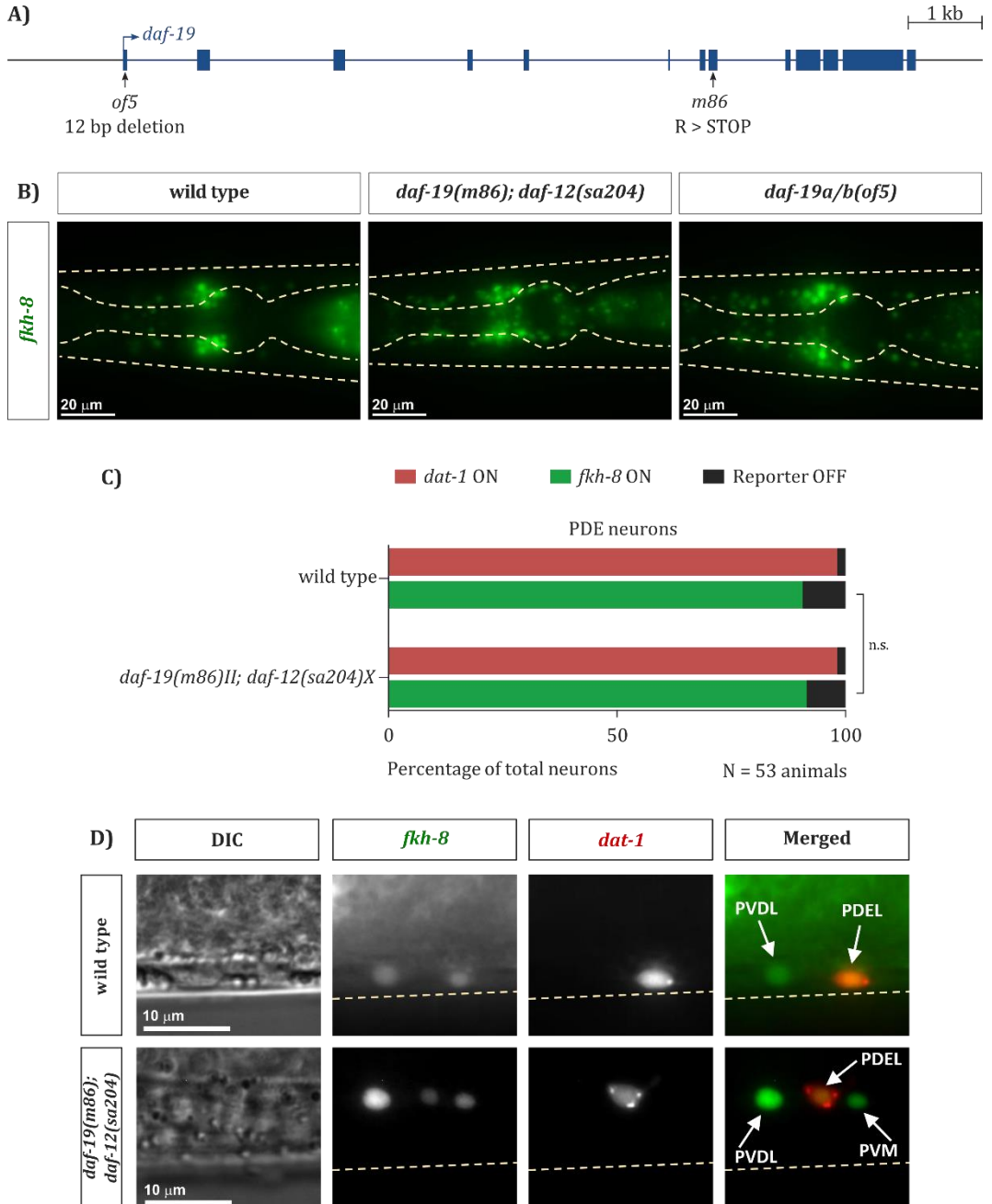
daf-19 and *fkh-8* act synergistically in the regulation of some structural ciliary genes.

Our results suggested that *fkh-8* and *daf-19* could genetically interact. First, in our analysis of the regulatory genome, we found that 29% of the *fkh-8* ChIP-seq peaks contain RFX binding sites (see **Figure R.31**) and, second, RNA-seq data for *daf-19/daf-12/fkh-8* triple mutants showed gene expression defects that were not present in either *fkh-8* or *daf-19/daf-12* mutants. Thus, to increase our sensitivity to measure *daf-19* and *fkh-8* genetic interactions, we turned again to the fluorescent reporter strategy.

We focused on four structural ciliary genes (namely, *ift-20*, *osm-1*, *peli-1* and *xbx-1*), whose expression we found partially regulated by *daf-19* and *fkh-8* (see **Figure R.2, R.3, R.52 and R.53**) and aimed to check their expressions in *daf-19/daf-12/fkh-8* triple mutants. For unknown reasons we were unable to obtain viable triple mutant animals using the *fkh-8(vlc43)* allele; thus, we performed this analysis using *fkh-8*(*tm292)* hypomorphic mutants. To include all corresponding controls, the expression of each reporter was assessed in 6 different genetic backgrounds (see **Figure R.67 and Table R.24**), also including a double *daf-12/fkh-8* mutant strain. Interestingly, for all the analysed reporters except for those of *peli-1*, this double mutant background further reduced the number of reporter-expressing neurons when comparing against each of the corresponding single mutant strains. This effect, that could be expected for those reporters for which we saw expression defects in *daf-12* single mutants (namely, those of *ift-20* and

osm-1), was novel for both *xbx-1* reporters and hence indicative of some kind of epistasis between *daf-12* and *fkh-8.* Finally, we analysed reporter's expression in *daf-19/daf-12/fkh-8* triple mutants. In accordance with previous results, the genetical interaction between *daf-19* and *fkh-8* was manifested, except for the *xbx-1*(short) and *peli-1*(long) reporters, by a further reduction in the number of reporter-expressing neurons when comparing against the corresponding double *daf-19/daf-12* mutant strain. Profiting from our detailed anatomical analysis, we found that, similar to previous results, the effect of each mutation or their combinations affected expression in different subpopulations of ciliated neurons depending on each reporter. For instance, expression of *ift-20* was mostly affected within the labial and cephalic neurons located in Region 3 in each of the single or double mutants analysed, whereas the presence of the *daf-19* mutation fully abolished expression within the amphid neurons in Region 1 (see **Figure R.67** and **R.68**). In contrast, residual expression of *xbx-1*(long) was still visible in Region 1 for double *daf-19/daf-12* mutants; however, this expression was fully abolished when adding the effect of the *fkh-8* mutation in the triple *daf-19/daf-12/fkh-8* mutant strain (see **Figure 3.67**).

To compute for the possible epistatic effects elicited by the combination of the different mutations we made use of a multiplicative model. Since the application of specific models demands specific type of data (Wagner 2015), average number of reporter-expressing neurons found for each genetic background were transformed into the corresponding fold change differences related to the mean value observed in the wild type

strain. Accordingly, expected fold change values for double *daf-12*/*fkh-8* mutants were calculated as the product of the mean observed values from the corresponding single mutants. Analogously, expected fold change values for the triple *daf-19*/*daf-12*/*fkh-8* mutants were calculated as the product of the mean observed values from the double *daf-19*/*daf-12* and single *fkh-8* mutant strains.
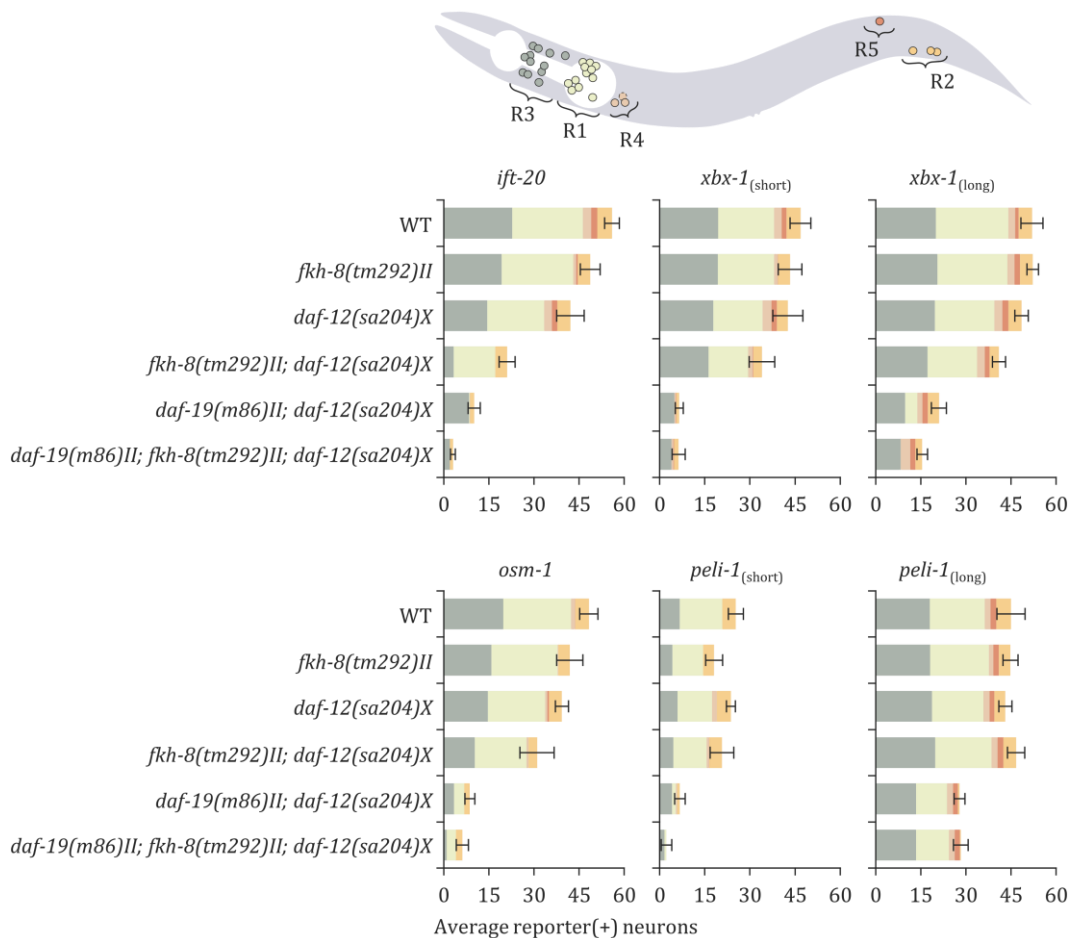
⬆ Figure R.67. Analysis for the expression of several fluorescent reporters from cilia-related features in *fkh-8*, *daf-12* and *daf-19* mutant backgrounds.
Top scheme shows here a lateral left view of a *C. elegans* young adult hermaphrodite ciliated system. Non-bilateral AQR neuron locates at the right side of the animal and hence appears depicted in dashed lines. Neurons gathered in the five different anatomical regions in which the ciliated system was divided are depicted in five different colours. In each graph, coloured bars show the average reporter-expressing neurons per region and genotype. Error bars represent the standard deviation for the final total mean. N = 5 to 12 animals per reporter and genotype.
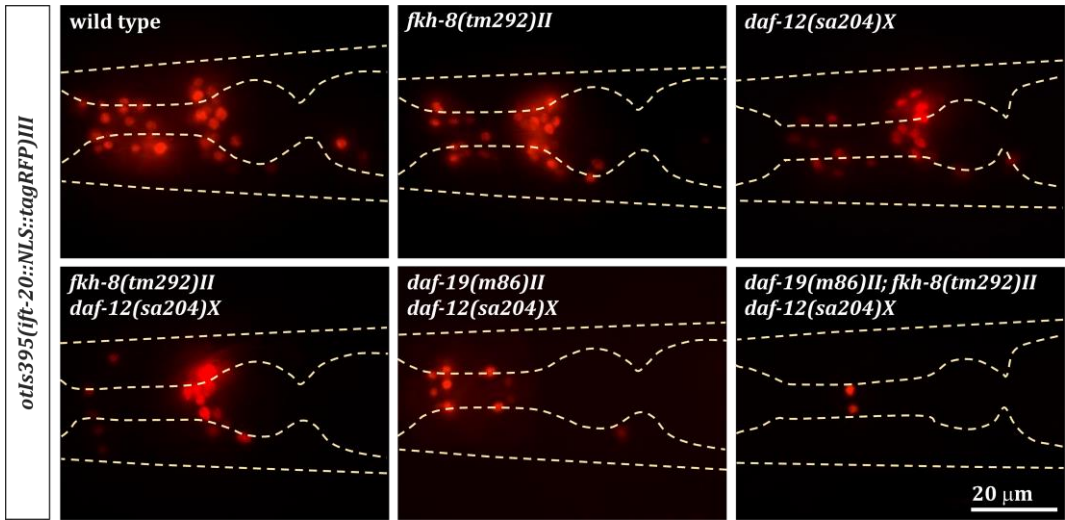
▲ **Figure R.68.** *daf-19* and *fkh-8* act synergistically in the transcriptional regulation of the panciliary gene *ift-20*.

Representative lateral view from heads of young adult hermaphrodites of *C. elegans* expressing a fluorescent reporter for the panciliary gene *ift-20* in different genetic backgrounds. Final fluorescent images correspond to Z stacks projections of images at 1 micron distance taken under a vertical fluorescence microscope.

As expected from the quantification of reporter-expressing neurons, statistically significant epistatic effects were found for the combination of *daf-12* and *fkh-8* mutations in all analysed reporters excepts for that of *osm-1* (see **Figure R.69**). Interestingly, a synergistic enhancement in the phenotype was found for the expression of *ift-20* and both *xbx-1* reporters, since the observed fold change values were lower than expected. However, and much curiously, both reporters for the *peli-1* gene showed significantly higher fold changes than expected, thus revealing a synergistic suppression between *fkh-8* and *daf-12* in the regulation of *peli-1*.

Finally, significant epistatic effects were also found for the combination of all 3 *daf-19*, *daf-12* and *fkh-8* mutations in 4 out of the 6 reporters that we analysed (see **Figure R.69**). In this case, a synergistic enhancement in the phenotype was found for the expression of all affected reporters, namely those for *ift-20*, *xbx-1*(long), *osm-1* and *peli-1*(short). Of note, this synergistic drop in the number of reporter-expressing neurons could not be exclusively attributed to the presence of synergy between *fkh-8* and *daf-12*, since the subpopulations of ciliated neurons that were affected were different in both double *daf-12*/*fkh-8* and triple *daf-19*/*daf-12*/*fkh-8* mutant strains (see **Figure R.68** as an example for the *ift-20* reporter). Indeed, synergy between *daf-19* and *fkh-8* was especially relevant in the context of the *peli-1*(short) reporter, since the drop in the number of reporter-expressing neurons overcame the suppressor effect that the *daf-12* mutation elicited over the mutation of *fkh-8*. Again, subpopulations

of ciliated neurons showing expression defects were different for each of the analysed reporters. Accordingly, whereas *ift-20* expression was fully abolished from the dopaminergic subpopulation of ciliated neurons individually in both *fkh-8* and *daf-19/daf-12* mutant strains, remaining EGFP-positive neurons of the *xbx-1*(long) reporter were located in anatomical regions compatible with the dopaminergic system of the worm, since at least the ADE and PDE neurons were found to keep expressing this reporter. Nevertheless, some residual expression was observed for all of the analysed reporters even in the context of the triple *daf-19/daf-12/fkh-8* mutation, hence indicating that additional and yet unidentified TFs have to be involved in the regulation of the ciliated fate in *C. elegans*.
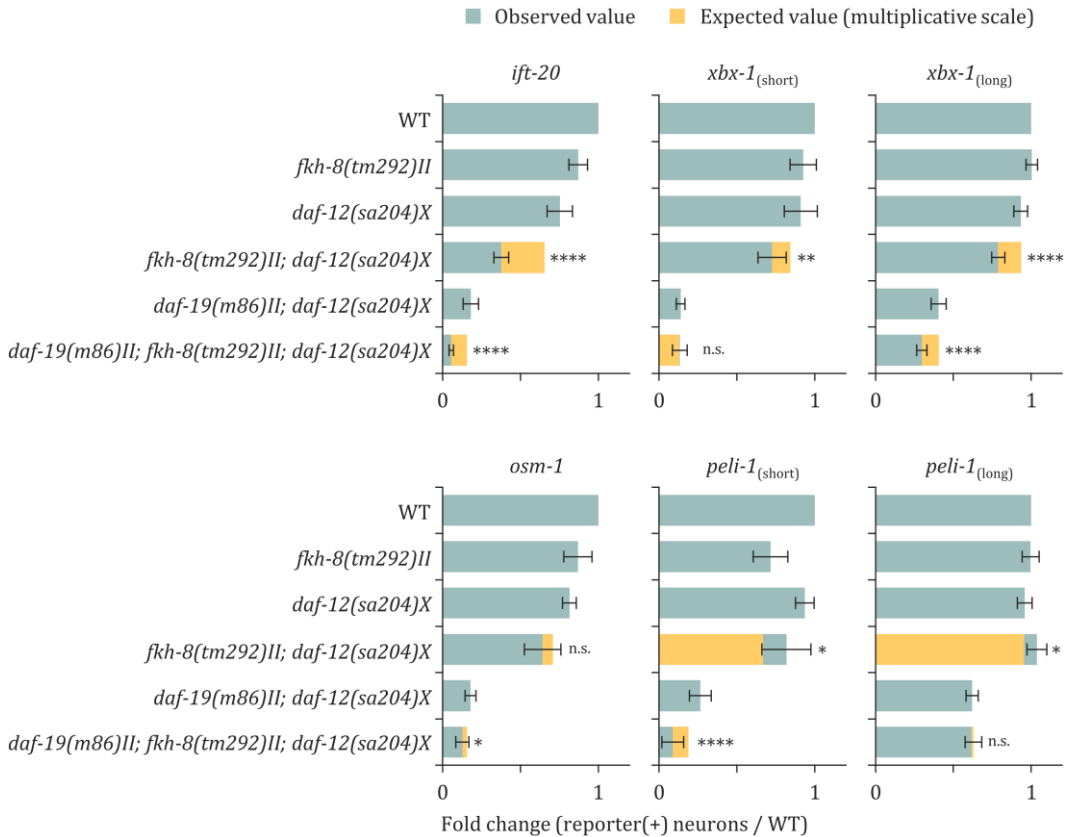
Figure R.69. Synergistic effects elicited by *fkh-8*, *daf-12* and *daf-19* mutant backgrounds in the regulation of several fluorescent reporters from cilia-structural features.

For each reporter, blue coloured bars represent here the mean fold change obtained when relating the average number of reporter-expressing neurons found in a given mutant background to the mean value observed for the wild type sample. The expected values for the combined mutations considering additivity are represented by orange bars overlapping the observed values. Error bars represent the standard deviation of the sample. Statistical significance between observed and expected values was calculated through a two-tailed one-sample *t*-test. n.s.: non-significant, * < 0.05, **< 0.01; *** < 0.001; **** < 0.0001. N = 5 to 12 animals per reporter and genotype.

As a whole, these results prove *daf-19* and *fkh-8* act synergistically in the transcriptional regulation of some structural ciliary genes. This detailed fluorescent reporter analysis allowed us to detect expression defects higher than expected in 4 out of the 6 reporters that we selected. In accordance with previous results, the combined mutation of both *daf-19* and *fkh-8* affected gene expression in a cell-type specific manner; however, neuronal subpopulations in which expression defects were assessed varied depending on the given reporter. Taken together, this body of results reveals that *fkh-8* acts as a terminal selector of the ciliated fate in *C. elegans*, working alongside *daf-19* to ensure the correct and robust expression of the ciliated features both in time and cell type.

▼ Table R.24. Statistics associated with Figure R.67.

Two-tailed Student *t*-tests were performed to compare the performance of six fluorescent reporters from cilium-related features in six different genetic backgrounds. In this table, p-values obtained for all possible pair comparisons between genetic backgrounds and reporter are gathered Tests were calculated attending at the variances of the two samples being compared. Final statistical significance (depicted in red) was established applying a Bonferroni correction considering all 15 hypotheses being tested for each reporter simultaneously, thus establishing an alpha value below 3,33E-03. Numbers in the first column represent the different genetic backgrounds: 1: wild type, 2: *fkh-8(tm292)II*, 3: *daf-12(sa204)X*, 4: *fkh-8(tm292)II; daf-12(sa204)X*, 5: *daf-19(m86)II; daf-12(sa204)X* and 6: *daf-19(m86)II; fkh-8(tm292)II; daf-12(sa204)X*.

Comparison	*ift-20*	*xbx-1*(short)	*xbx-1*(long)	*osm-1*	*peli-1*(short)	*peli-1*(long)
1 vs. 2	3,28E-05	1,58E-01	9,17E-01	2,38E-03	4,67E-05	9,40E-01
1 vs. 3	3,77E-08	1,59E-01	8,37E-02	1,06E-05	1,29E-01	3,34E-01
1 vs. 4	1,01E-17	9,53E-05	6,39E-05	3,19E-06	1,68E-02	4,75E-01
1 vs. 5	6,40E-18	4,78E-06	2,56E-12	1,57E-19	1,39E-11	1,47E-04
1 vs. 6	4,52E-15	9,91E-15	6,69E-06	1,79E-19	1,99E-14	4,80E-06
2 vs. 3	1,07E-03	7,91E-01	1,80E-02	1,70E-01	4,41E-05	1,95E-01
2 vs. 4	1,24E-14	9,08E-04	3,25E-06	9,93E-04	9,61E-02	2,65E-01
2 vs. 5	1,77E-15	1,16E-06	1,42E-13	1,29E-09	1,13E-09	9,00E-10
2 vs. 6	1,82E-12	2,40E-07	1,97E-16	2,15E-10	2,02E-13	1,05E-07
3 vs. 4	9,89E-12	5,16E-03	9,46E-05	1,39E-02	4,92E-02	2,17E-02
3 vs. 5	1,15E-13	6,04E-05	1,47E-13	3,01E-17	4,58E-14	7,93E-11
3 vs. 6	2,72E-12	1,11E-12	1,88E-16	5,53E-17	2,83E-17	1,27E-08
4 vs. 5	4,21E-08	1,37E-08	2,60E-11	1,39E-04	2,18E-07	7,34E-10
4 vs. 6	9,79E-11	1,58E-09	5,23E-11	6,64E-05	1,25E-08	7,60E-08
5 vs. 6	4,04E-05	7,37E-01	2,15E-06	3,36E-03	9,31E-06	7,10E-01

Discussion

daf-19 does not act alone in the regulation of ciliary features in *C. elegans*.

The terminal selector model proposed by Dr Hobert stands that specific combinations of TFs, termed terminal selectors, are responsible for the direct transcriptional regulation of the effector genes that ultimately define the terminal state and functionality of a neuron (Hobert 2008). The acquisition of the ciliated fate constitutes a terminally differentiated state for a subpopulation of neurons in *C. elegans*, which is defined by the expression of hundreds of effector genes encoding for ciliary components. As stated in the Introduction, the role of *daf-19*/RFX as a terminal selector of the ciliome has been well established. To date, most of the described terminal selectors have been found to act in combinations (Hobert 2016a); however, at the time of writing, no other TF has been described as a terminal selector for the sensory ciliome in *C. elegans* – or in any other organism as well. In this Thesis project we aimed to identify such factor/s.

First, through bibliographic research, we selected a battery of genes coding both for panciliary features (constituting the structural building blocks of the cilium) as well as subtype-specific ciliary features (present only in sub-populations of ciliated neurons) in order to analyse their transcriptional dependency towards DAF-19. Interestingly, we found that the expression of subtype-specific ciliary genes of the TRPV family was largely independent of *daf-19*, as assessed when comparing reporter expression between wild type and double null *daf-19* and *daf-12* mutants. This is in agreement with the original

model proposed by Dr Swoboda, in which transcription factors other than *daf-19* were suggested to be responsible for the transcriptional regulation of cilia-related genes such as "cell-specific cilium structure elements, signal receptors, and signal transducers" (Swoboda et al. 2000b). However, later research has found examples of a role for *daf-19* in the regulation of subtype-specific cilia genes, such as *odr-4* (Efimenko et al. 2005), a gene that in turn has been shown to be important for ciliary localization of some odorant receptors (Dwyer et al. 1998).

In contrast, and as expected, reporter analysis of core structural ciliary features showed a dramatic decrease in the number of reporter-expressing neurons in *daf-19* mutants. Importantly, for 11 of the 12 reporters assayed, some neurons were still able to activate transcription of those reporters even in the null *daf-19* mutant background, demonstrating that DAF-19 is not absolutely required for ciliary gene expression and strongly suggesting the existence of DAF-19 co-factors. Interestingly, short reporters corresponding to less than 150 base pairs in length flanking the binding site of DAF-19/RFX showed stronger transcriptional dependency toward DAF-19. Comparison of long and short versions of the same reporter allowed us to drive two main conclusions. First, shorter sequences containing a DAF-19/RFX site were expressed in fewer neurons when compared to their longer counterparts and; second, shorter sequences were much more sensitive to the loss of *daf-19*. In conclusion, both observations strongly suggest for the existence of additional TFBSs for TFs other than DAF-19 that are able to compensate *daf-19* loss in longer *versus* shorter

regulatory sequences.

In-silico approaches help decipher transcriptional regulatory signatures.

According to the terminal selector model, regulatory sequences of terminal effector genes share a molecular signature of similar TFBSs. As we determined that *daf-19* does not act alone in the transcriptional regulation of ciliary features, we reasoned that putative regulatory sequences of cilium-related genes should contain additional TFBSs other than those for *daf-19*.

TFBS over-representation algorithms often constitute statistical black boxes in which the modification, addition or removal of a single in-put sequence can cause substantial changes in the corresponding out-put results. This is especially relevant when the number of in-put sequences is low. Other parameters, such as nature, composition or length of the sequences, have also a major impact on the final results (see Bailey 2008 for an introductory theoretical approach to motif discovery). In the last 20 years, the *in-silico* analysis of *cis*-regulatory elements of co-expressed genes has been extensively used and the availability of literature, tools and pipelines is rather overwhelming (see Tran and Huang 2014, 2018 for a recent survey of current methods). Another major issue derived from the high activity in the field is the advent of new tools and pipelines that last as online resources a limited amount of time just to be discontinued not long after the corresponding publication has been published. Within that ever-changing scenario, two resources have remained "up and running" for

more than 10 years: RSAT and the MEME Suite, which were used in this work.

An unbiased bottom-up approach performing *de novo* motif discovery in putative regulatory sequences of 163 curated cilium-related genes, both structural and subtype-specific, allowed us to identify 9 enriched motifs. As expected from ciliary genes, motif comparison performed against TFBSs databases showed that 4 out of the 9 motifs corresponded to RFX-related binding sites and roughly 63% of all the genes in our list harboured at least one match for such type of sites. DAF-19 binding sites, the so-called X-boxes, are imperfect palindromic sequences consisting of two half-sites of 6 base pairs in which a greater degeneration is allowed for one of the half-sites. Based on *in vivo* expression analysis accounting for nucleotide composition of the X-box, Swoboda and collaborators proposed a model for the different expression patterns exhibited by different ciliary features (Efimenko et al. 2005). According to this model, in palindromic or nearly palindromic X-boxes, DAF-19 binds as a homodimer and that correlates with a broad panciliary expression pattern. Consistently, we found that roughly 85% of the structural ciliary genes in our list harboured a predicted RFX motif and, moreover, more than 90% of them were able to match palindromic or nearly palindromic PWMs of known RFX binding motifs. In contrast, following the above-mentioned model, X-boxes with a more degenerated half-site could allow *daf-19* to interact as a heterodimer, binding to that site along with unidentified cell specific interactors that could direct gene expression to subsets of ciliated neurons. Interestingly, the RFX-related binding sites that we

retrieved from our *de novo* motif discovery analysis show increasing degrees of degeneration. And much interestingly, we found that only 37% of the function-specific genes in our list harboured a predicted RFX motif and, when present, that motif tended to be more degenerated since only 24% of those genes were able to match palindromic or nearly palindromic PWMs of known RFX binding sites. Thus, our motif enrichment analysis for RFX is in agreement with previous models on DAF-19 actions.

RFX binding sites are particularly long and complex compared to the binding sites of other TFs. Accordingly, the presence of an X-box motif has been previously used for the bioinformatical identification of candidate ciliary genes in several studies (Li et al. 2004, Blacque et al. 2005, Efimenko et al. 2005, Chen et al. 2006, Avidor-Reiss et al. 2004, Laurençon et al. 2007). However, as already illustrated in pioneering works by Swoboda and collaborators, the use of different types of X-box motifs may render different results (Efimenko et al. 2005), since different consensus had to be employed to maximize the discovery of ciliome genes. In agreement with these published results, it is interesting that our unbiased bottom-up approach has allowed us to retrieve 4 different PWMs that could be further used in the screening and categorization of putative ciliary genes genome-wide. The relevance of this fact can be exemplified with our analysis of the *fkh-8* locus. When we analysed the regulatory sequences of *fkh-8* for the presence of X-boxes, we found matches for our Motifs 1, 3 and 4, but not for the specific nucleotide composition of Motif 2. Hence, a search performed exclusively with the PMW of Motif 2 would have

rendered a result in which *fkh-8*, a gene that we have experimentally assessed to be regulated by *daf-19*, would have been labelled as a non-RFX-target gene. Interestingly, we find DAF-19 acts as a repressor of *fkh-8* expression in non-ciliated neurons; thus, it is tempting to speculate that activating or repressing actions of DAF-19 could be mediated through different isoforms acting on different types of motifs. Further experiments would be needed to test this hypothesis.

However, the presence of RFX sites in ciliome genes was already well-known and although it allows us to validate our approach, we were interested in the identification of new motifs that could help us identify new co-factors for DAF-19. Other than RFX-related motifs, our *de novo* motif discovery analysis showed that the regulatory sequences of roughly 28% of the genes in our list contain putative bHLH binding sites. This could actually be a reflection of the neuronal nature of those genes since some members of the bHLH, such as the subfamily *Achaete-Scute* in which *lin-32* and *hlh-14* are included, are known proneuronal genes with conserved functions in different metazoan such as Drosophila, mouse, *C. elegans* and also other organisms with very simple and ancestral nervous systems like *Nematostella vectensis* (Portman and Emmons 2000). In addition to RFX and bHLH, 4 non-degenerated motifs were also retrieved that did not match any known motif either from the *C. elegans* CIS-BP 1.02 or the JASPAR CORE non-redundant 2018 databases. Since the percentage of appearance among our ciliome genes was rather low, we decided not to further study those motifs.

Over-representation algorithms are known to surfer from noise induced from the analysed sequences (Bailey 2008). Accordingly, we decided to perform a second *de novo* motif discovery analysis employing only the sequenced from the 102 genes harbouring any of the RFX-related motifs that we retrieved in our first analysis and that mainly corresponded to structural ciliary genes. In this alternative approach, we retrieved 10 motifs; 6 of them matching known RFXs motifs (as expected), 2 non-degenerated motifs that did not match any known motif from the aforementioned databases (and present in so low frequencies that we discarded them) and 2 degenerated motifs that, being able to match PWMs of known motifs, were unable to pass the statistical thresholds imposed by multiple comparisons methods. Interestingly, each of these two motifs were found in roughly half of the genes also harbouring an X-box motif. More specifically, a molecular signature composed by an X-box plus Motif 9 was found in 25% of these genes; a second signature including the X-box plus Motif 10 was found in 21% of them, and finally, a third signature including all 3 motifs was present in 26% of these genes. Thus, as a whole, around 72% of the 102 genes considered in our second *de novo* motif discovery analysis contain an X-box motif combine with Motif 9, Motif 10 or both.

Trying to identify the TF that recognizes a given *de novo* motif by comparison with other known motifs impose an inherent drawback simply derived from the number of known motifs present in a database. At the time of writing, MEME suite makes use of the *C. elegans* CIS-BP 1.02 database from which less than 30% of the total TFs of the worm are represented. Taking this into consideration and also considering that paralogous and orthologous TFs usually recognize similar PWMs, we took a closer look to identify TF families that could bind the two degenerated non-RFX-related motifs that we retrieved in our second *de novo* motif discovery analysis. The suboptimal alignments between such motifs and those present in the database seemed to predict putative binding for members of the Forkhead, Homeodomain and Nuclear Hormone Receptor families of TFs. Interestingly, 9 out of the 10 TF candidates that we later found enriched in ciliated sensory neurons belonged to these families.

Our *in-silico* analysis of *cis*-regulatory elements of co-expressed cilium-related genes in *C. elegans* finds a precedent in (Burghoorn et al. 2012). Authors also used a bottom-up approach to identify TFBSs in sequences of known RFX target genes. Instead of RSAT, they used MEME. Instead of 700 base pairs, their analysed sequences were 400 base pairs in length. Through that strategy a CT rich motif in close proximity to the X-box was identified that was also responsible for the transcriptional regulation of the analysed genes. However, the TF binding to such motif was not determined. When we replicated their experiment, we also found such CT rich motif both through the RSAT and MEME tools. Moreover, we found that the PWM generated by that motif matched that of *eor-1* with high statistical significance. However, *eor-1* expression pattern is rather broad, being found in all 27 tissue categories established in the single cell data published by (J. Cao et al. 2017).

In summary, the implementation of this

in-silico approach in which putative regulatory sequences of cilium-related genes were used to perform a *de novo* motif discovery analysis allowed us to identify specific families of TFs candidates to be involved in the regulation of the sensory ciliome in *C. elegans*. Besides the expected RFX-related motifs and the general pro-neural bHLH factors, binding motifs for members of the Forkhead, Homeodomain and Nuclear Hormone Receptor families could be part of the molecular signature present in putative regulatory sequences of structural ciliary genes in *C. elegans*.

Expression analyses allow for the identification of candidate regulators of the ciliated fate.

In the past 3 years several transcriptomic efforts have been aimed to define the differential gene expression status of *C. elegans* at the single-cell resolution. This thesis project has greatly benefited from them. The first resource to be publicly available was that of (J. Cao et al. 2017), where authors performed single-cell RNA-seq from bulk L2 animals. Using that dataset with the most astringent conditions offered by the GExplore$_{1.4}$ platform we were able to identify 10 TFs highly and specifically enriched within the ciliated sensory neurons of *C. elegans*. These ten candidates belonged to only four different families of TFs: one Zinc Finger (ZF - C2H2) (*Y22D7AL.16*), one Forkhead (FKH) (*fkh-8*), seven nuclear hormone receptors (NHR) (*nhr-158*, *nhr-188*, *nhr-124*, *nhr-30*, *nhr-38*, *nhr-216*, *nhr-277*) and one homeodomain (HD) (*ceh-57*). Curiously, those four families bind to consensus sites similar to the motifs

that we retrieved in our *de novo* discovery analysis for putative regulatory sequences of RFX-related genes (namely, those for Motif 9 and Motif 10).

Of note, 7 out of the 10 candidates belonged to the NHR family of TFs. This is not surprising as in *C. elegans* the NHR family has undergone massive expansion and it accounts for 284 members in contrast to the 48 members found in mouse and human (Arda et al. 2010), hence corresponding to roughly one third of all TFs in the *C. elegans* genome. Several NHRs function as metabolic sensors; accordingly, this expansion has been proposed to enable rapid and adaptive responses to different cues rather than participate in neuronal specification (Arda et al. 2010).

From the 10 candidates we retrieved, most were (and still remain) poorly described in the literature. To gain some insight about their expression patterns, we applied a t-SNE visualization technique over the original data from (J. Cao et al. 2017) and determined that 4 of the candidates exhibited the broadest expression pattern within the ciliated system of the worm. Those TFs were, ranked in decreasing broadness of expression pattern: *nhr-277*, *fkh-8*, *nhr-158* and *ceh-57*. Within the terminal selector model, many terminal selectors are required for both initiation and maintenance of effector gene expression (Deneris and Hobert 2014); thus, terminal selectors are expressed throughout the whole life of a neuron. To check candidates' expression at later stages but also to verify our analysis from (J. Cao et al. 2017) data, we benefited from the CeNGEN project (Hammarlund et al. 2018). There authors

followed two approaches on L4 animals: 1) single-cell RNA-sequencing, allowing for the identification of each neuronal class, and 2) deep bulk RNA sequencing of sorted neurons from each class enabling full characterization of their transcriptome. These data are currently freely available through the cengen.org web page and shows that *fkh-8* expression is found in most ciliated neurons of *C. elegans*, ranking first among the rest of the candidates.

Thus, accounting for expression data in both datasets, *fkh-8* stood as our best candidate as a TF broadly expressed in all ciliated neurons of *C. elegans* (similar to that of DAF-19 isoform d), and thus a good candidate as a DAF-19 co-factor in the regulation of the ciliated fate. Also considering previous reports identifying FKH TFs as terminal selector of the motile ciliome in different animal groups, we decided to focus the rest of this Thesis on the characterization of *fkh-8*.

A multi-angle approach reveals *fkh-8* is a terminal selector of the ciliated fate in *C. elegans*.

Our characterization of *fkh-8* function combining different strategies and techniques has allowed us to uncover the cell-autonomous role of *fkh-8* as a terminal selector of the ciliated fate in *C. elegans*.

fkh-8 expression, using a fosmid-based EGFP reporter, starts at the early bean stage around 300 minutes post-fertilization; coincident with neurogenesis and early differentiation of ciliated neurons. At the two-fold embryonic stage, *fkh-8* reaches an apparent panciliary expression pattern that is maintained throughout the life of the animal.

Although the expression pattern that we observe for *fkh-8* in the young adult animal is mostly restricted to the ciliated system, some consistent expression is also detected in the VC4, VC5 and PVD neurons. Indeed, FKH-8 expression in those non ciliated neurons can also be confirmed through the CeNGEN data set and the intensity of their signals equal those of the ciliated neurons. These data set also reveals *fkh-8* expression in two additional neuron types: URX and URY. None of those neurons (VC4, VC5, PVD, URX and URY) are ciliated but, remarkably, they are all related to sensory functions. VC4 and VC5 are actually motor neurons but, in contrast to the rest of the VC neurons, they do not innervate body wall muscles. VC4/5 neurons are known to extend processes dorsally along the vulval hypodermis and were originally suggested to be mechanosensory (White, J.G., Southgate, E., Thomson, J.N. and Brenner 1986). Several authors have hypothesized that the function of such processes is to sense the stretch produced by individual egg-laying events (Zhang, Schafer, and Breitling 2010). Since mechanical deformation of VCs presynaptic termini have been observed during strong twitching and egg-laying contractions, the VC neurons have been suggested to act as baroreceptors that are mechanically activated by vulval muscle contraction which, in cooperation with several other neurons, coordinate locomotion and egg-laying (Collins et al. 2016). The PVD neurons are polymodal nociceptors, sensing both harsh touch and cold temperatures (Chatzigeorgiou et al. 2010). On the other hand, URX neurons functions at aerotaxis, mainly sensing oxygen (Gray et al. 2004) with a minor contribution to CO_2 sensation (Bretscher et al. 2011).

Finally, URY neurons are involved in mate-searching behaviour in males (Barrios et al. 2012) and have also been postulated to act in pathogen sensing (Pradel et al. 2007). Moreover and much interestingly, PVD, URX and URY neurons exert their actions in collaboration with ciliated neurons and are developmental sister cells of the ciliated PDE, CEP and OLQ neurons, respectively (Sulston et al. 1983; Sulston and Horvitz 1977).

Both, onset of *fkh-8* expression as well as the anatomical expression pattern that we could observe by means of the fosmid-based EGFP reporter, can be further confirmed through a yet another transcriptomic effort: the *C. elegans* embryogenesis single cell data by (Packer et al. 2019). Data from this study, in which single-cell RNA-seq was performed over *C. elegans* embryos of different developmental stages ranging from gastrulation to terminal cell differentiation, are freely and easily available through the VisCello webtool (the *C. elegans* embryogenesis visualizer). VisCello data shows *fkh-8* expression in branches descending exclusively from the AB lineage and starting around 300 minutes after the first cleavage of the fertilized egg. Consistent with our own observations, the highest levels of FKH-8 expression are consistently found within the ciliated neurons of the worm. Moreover, FKH-8 expression can be further confirmed in the URX sensory neurons, as well as the excretory and rectal gland. Additionally, a few interneurons related to sensory processes such as the RIM, AUA, AIZ, AIN or AIY, as well as the putative sensory motor neurons URA and URB, show FKH-8 expression at levels equal to those of some ciliated neurons. However, this expression pattern should be corroborated through further fosmid expression analyses.

Considering its pan-ciliary expression pattern, and to decipher the role of *fkh-8* in the specification of the ciliated fate, we made use of a yet another resource of genomic information: the one provided by both the modENCODE and the modERN consortia. Through the analysis of 446 ChIP-seq data sets from different developmental stages of *C. elegans* we found that *fkh-8* ranked first binding ciliome genes in our curated list when compared against 258 different TFs. It is worth noting that *daf-19* data is, to date, not present in any of such consortia. Additionally, we found that TFs such as *C34F6.9, dpl-1, dsc-1, lin-35, lsy-2, nfya-1, pha-4* and *efl-1* also show an enrichment in the binding of ciliome genes and it might be interesting to further explore the role of these TFs in ciliogenesis. Extending the analysis of FKH-8 binding profile genome-wide clearly exposed (through ontology enrichment analysis) the preferential binding of FKH-8 to ciliome genes. Moreover, applying both known and *de novo* motif discovery analysis, we could assess the presence of FKH-related motifs at the centre of the FKH-8 ChIP-seq peaks. Since the distribution of a biologically relevant motif is expected to peak around the centre of such sequences (Bailey and MacHanick 2012; Mercier et al. 2011) and two of the motifs found (namely, Motif 23 and Motif 30) behaved that way, we suggest that a direct binding of FKH-8 to the regulatory regions of at least a subset of FKH-8-targeted genes might be occurring. In this regard, it is interesting to note that peaks targeted by Motif 23 could be related to 390 unique genes. When applying an ontology enrichment analysis, those genes are mainly related to germ line terms,

suggesting that *fkh-8* might repress its expression in somatic cells. However, those peaks with predicted matches for Motif 30 identify 1,038 unique genes (of which only 90 were shared with Motif 23) that clearly enriched cilium-related categories in the ontology enrichment analysis. Surprisingly, bimodal distributions could be observed for FKH-related Motifs 40, 31 and 9. Such distribution, in which sites locate at the edges of their targeted peaks, has been observed for members of the FKH family when they act as pioneer factors in regions of open chromatin (Grossman et al. 2018).

After expression and genome-wide binding analysis, we aimed to test the functionality of *fkh-8* by performing RNA-seq experiments. Unexpectedly, despite specific expression of *fkh-8* in ciliated neurons and genomic binding to ciliome genes, we could not find much evidence for a role of *fkh-8* in ciliogenesis when comparing *fkh-8(tm292)* mutants against wild type worms. The first warning for our generalist approach lacking in sensitivity arose when the low differences between both strains failed to pass a multiple comparisons test. We postulate that transcriptional regulatory effects of *fkh-8* – a gene that we observed is mainly expressed within 60 ciliated neurons – could be diluted when analysing RNA-seq data retrieved from a bulk pool of cells obtained from whole animals. However, we assessed that probability of false positive within the whole data set comparing those strains remained below 29% for each of the 23,616 transcribed elements retrieved in the experiment. Accounting for those limitations, we proceeded to analyse the differences in the raw comparison between expression values of *fkh-8(tm292)*

mutants and wild type worms.

When compared to wild type worms, *fkh-8(tm292)* mutants show a noteworthy upregulation of genes related to reproductive structures or processes. Much interestingly, when considering genes differentially expressed in *fkh-8* mutants that are also targeted by FKH-8 ChIP-seq peaks, a single WormBase tissue category is enriched: that of the germline (p-value = 6.00E-06). Repression of somatic-promoting genes in the germline is essential for germ cells to maintain their unique pluripotent identity. The opposite program, in which somatic cells repress germline genes' expression, runs in parallel in the developing and developed organism, ensuring somatic development, differentiation, and function. Both repressive programs have been linked to chromatin remodelling through histone modifications (Rechtsteiner et al. 2019; Robert, Garvis, and Palladino 2015), but some TFs have also been found to participate in these programs. That is the case for Blimp1, a critical determinant of the germ cell lineage in mice that repress somatic genes and directly binds to its targets (Ohinata et al. 2005). Thus, one possibility is that, just as some TFs have been shown to repress somatic genes within the germline, *fkh-8* is able to block germline genes' expression within the ciliated system of the worm.

In addition, our RNA-seq results suggest that the *tm292* deletion likely constitutes a hypomorphic allele in which a putative truncated protein of 180 amino acids that is also harbouring the whole FKH binding domain is generated. Accordingly, we decided to create a null mutant strain for *fkh-8*. Successful co-CRISPR strategy allowed us to obtain the

whole deletion of the *fkh-8* locus in the form of the novel *vlc43* allele. This deletion has been sequenced to expand from 79 base pairs upstream of the *fkh-8* translational site to 122 base pairs downstream its TAA stop codon. This new resource will be made available to the rest of the "worm community" through the CGC.

The outstanding contribution of OMIC technologies to the understanding of transcriptional regulation needs no further defence nowadays. However, in the case of subtle, yet crucial changes in the expression level of a gene (or even more dramatic changes that affect only a few cells), a generalist approach such as whole animal bulk RNA-seq experiment may suffer from a lack in sensitivity. That was indeed the case when our RNA-seq profiling of *fkh-8(tm292)* mutants failed to identify gene expression defects in cilium-related genes harbouring FKH-8 binding sites in their sequences (as assessed through the presence of ChIP-seq peaks). Moreover, limitations on the statistical methods of analysis for RNA-seq approaches were further manifested when low differences between wild type and mutant animals failed to pass a multiple comparisons test.

Much interestingly, this lack of sensibility was also present in the RNA-seq profiling of double *daf-19* and *daf-12* mutants. In contrast with the modest effect retrieved for *fkh-8(tm292)*, the influence of the former double null mutations exhibited a huge impact in the regulation of the young adult transcriptome. Accordingly, expression levels for roughly 40% of the genes recovered in our experiment were significantly deregulated when compared against those of wild type worms.

On this occasion, the criterium for multiple comparisons was also met for the whole data set. However, disagreement between reporter expression analysis and RNA-seq performance was also observed. For all 14 genes that we tested in the double *daf-19* and *daf-12* null mutant strain (namely, *che-11*, *che-13*, *ift-20*, *mks-1*, *osm-1*, *osm-5*, *peli-1*, *tmem-107*, *tub-1*, *xbx-1*, *ocr-1*, *ocr-2*, *ocr-4* and *osm-9*), agreement between techniques was met only for *osm-5* , (fold change = 0.41, p-value = 4.63E-4), similar to what we found in the *fkh-8* mutant analysis.

Thus, we decided to move onto a cellular resolution strategy employing fluorescent reporter analysis. By those means we were able to assess differential expression defects in four out of the nine structural cilium-related reporters that we analysed in *fkh-8(tm292)* mutant animals. When assessing the expression levels of those genes (namely: *ift-20*, *osm-1*, *osm-5* and *peli-1*) in our RNA-seq profiling, only *osm-5* exhibited a significant difference between wild type and mutant animals (fold change = 0.41, p-value = 1.98E-03), with a 21.66% probability to score as a false positive. Through this fluorescent reporter methodology we found expression defects in several different classes of ciliated neurons. Such defects, that were cell-specific depending on each given reporter, included missing expression in the IL2, CEP, ADE and PDE neurons in both *fkh-8* mutant backgrounds, with a smaller impact affecting the phasmid neurons as well. In addition, many yet unidentified labial and amphid neurons saw reporter gene expression greatly compromised in the null *fkh-8* mutant strain. These results are noteworthy considering that, in contrast to our RNA-seq experiment, they allowed us to

unravel a role for *fkh-8* in the transcriptional regulation of ciliome genes.

Nevertheless, fluorescent reporter analysis test the activity of specific genetic sequences (enhancers). In consequence, we have demostrated the impact of either *fkh-8* or *daf-19* in the transcriptional regulation of such sequences, not over the average transcriptional rate of the target gene. It is well known that transcriptional regulation is robust and makes use of redundant enhancers (also known as shadow enhancers) (see Hobert 2010); thus, it is still possible that in *fkh-8* and *daf-19/daf-12* mutants, transcription of ciliary components in all ciliated neurons is mostly unaffected through the use of alternative enhancers. In this regard, it will be important to further assess transcription of ciliome genes in sensory neurons at single cell resolution using, for example, fluorescent in sity hybridization (FISH). In any case, it should be noted that *fkh-8* mutants show broad defects in cilia-mediated behaviours and this fact strongly suggests that *fkh-8* function upon ciliome transcriptional regulation cannot be fully compensated.

However, despite its very high cellular resolution, the fluorescent reporter strategy proved to be very time-consuming and with a very limited throughput. Therefore, to complement this approach, we decided to analyse sensory-mediated behaviours elicited by the action of specific ciliated neurons. We reasoned that, if *fkh-8* had a broad effect on expression of ciliary components, then sensory-mediated behaviours would be impaired in *fkh-8* mutant animals. Moreover, behavioural defects would demonstrate that endogenous gene expression, and not only reporter genes, was affected in those mutants. Thus, we decided to assess mechanosensation, olfaction and gustatory behaviours elicited by different classes of ciliated and non-ciliated neurons.

Regarding mechanosensation, anterior gentle touch, mediated by the non-ciliated neurons ALML, AMLR and AVM, and posterior gentle touch, mediated by the non-ciliated neurons PLML and PLMR, allowed us to discard general motor defects in both *fkh-8* mutant strains. Harsh touch, mediated by the combined action of both ciliated PDE and non-ciliated PVD, was not affected either. This was rather surprising considering that 1) we found compromised ciliome reporter gene expression in the PDE neurons in *fkh-8* mutant animals and 2) that both neurons (PDE and PVD) express *fkh-8*. Finally, nose touch response, that is controlled by the ciliated ASH, FLP and OLQ neurons, was significantly decreased for both *fkh-8* mutant strains, displaying a stronger phenotype for the null mutant.

Next, we assayed behaviours mediated by the detection of olfactory cues. Attractive response to diacetyl, mediated by the AWA neurons, was affected in *fkh-8* mutants. However, these defects were not statistically significant due to large variations among replicates. In contrast, attractive response to 2-heptanone, mediated by the AWC neurons, was dramatically impaired in the null *fkh-8(vlc43)* mutant strain but not in the animals carrying the partial deletion allele *tm292*. This experimental evidence further supports the hypomorphic nature of this allele. Analogous behavioural defects were observed for

the aversive response to 2-nonanone mediated by AWB. As a whole, these results demonstrate the requirement of *fkh-8* for appropriate olfaction in *C. elegans*.

Finally, several behaviours mediated by the detection of gustatory cues were also assayed. Attractive response towards NaCl, a cue mainly sensed by the gustatory ASE neurons, with the ADF, ASG and ASI neurons playing minor, redundant roles, showed no defects in any of the *fkh-8* mutant alleles. Drop test assaying the avoidance response mediated by ASH and ASK towards SDS showed a significant decrease in both *fkh-8* mutant alleles, again displaying a stronger phenotype in the null mutant strain. In the case of the avoidance response to copper, mediated by the ADL, ASE and ASH ciliated neurons, both *fkh-8* mutant alleles showed defects in the detection of the toxic. Interestingly, susceptibility to copper was first suggested through our RNA-seq profiling of *fkh-8(tm292)* mutants, since the downregulated genes *vhp-1*, *kgb-1*, *kreg-1* and *lys-3* were found to enrich the GO term related to the copper response (GO:004668). *vhp-1* encodes a MAP kinase phosphatase that is required for the regulation of the KGB-1/JNK-like MAPK-mediated stress response pathway (Mizuno et al. 2004). In turn, *kgb-1* regulates the expression of both *kreg-1* and *lys-3*, which are required for the defence response against heavy metal stress (Hattori et al. 2013). Of note, the *vhp-1* locus contains 8 different FKH-8 ChIP-seq peaks, suggestive of a direct action of *fkh-8* upon this regulatory cascade.

As a whole, our behavioural analyses demonstrate the requirement of *fkh-8* for the correct functionality of the ciliated ASH, FLP, OLQ, AWC, AWB, ASK, ADL, and ASE neurons. These behavioural studies are a good example of how complementary approaches can be combined to better understand the role of TFs in specific transcriptional regulatory programs. Other examples exist that remark the power of behavioural analysis. Notably, *daf-19* was first isolated for its abnormal behaviour producing *dauer* (Perkins et al. 1986) and later cloned and identified as an RFX TFs acting as a terminal selector of the ciliated fate (Swoboda et al. 2000a). Analogously, *che-1* was first identified as a chemosensory mutant displaying behavioural defects (Dusenbery, Sheridan, and Russell 1975), that was later identified as a Zinc Finger TF which is a terminal selector for the ASE neuron (Uchida et al. 2003). Similarly, *unc-3* was first identified as a locomotor defective mutant (S Brenner 1974) and later described as a terminal selector for cholinergic motoneurons (Kratsios et al. 2012).

Ultimately, combined results from our reporter expression analysis and our behavioural studies show compromised functionality for the IL2, CEP, ADE, PDE, ASH, FLP, OLQ, AWC, AWB, ASK, ADL, and ASE neurons, in addition to several yet unidentified labial, amphid and phasmid neurons in which reporter expression defects were also assessed. These results thus reveal broad requirements of *fkh-8* in the specification of the ciliated fate. Much importantly, we show that *fkh-8* elicits its action in a cell-autonomous manner, since expression of *fkh-8* in specific subpopulations of ciliated neurons was sufficient to rescue reporter expression defects.

As already stated, in many of our

behavioural assays, null *fkh-8(vlc43)* mutant animals showed stronger defects than those carrying the partial deletion allele *tm292*. The increase severity of the phenotypes for *fkh-8(vlc43)* mutants was also observed in the expression of many of the gene reporters analysed. This body of evidence portrays *fkh-8(tm292)* as an hypomorphic allele capable to translate a truncated protein with retained functionality. A second possibility, not mutually exclusive, is that compensatory effects take place in *tm292* that are not present in the null allele. For example, the phenomenon of transcriptional adaptation could be triggered if the messenger generated by the *tm292* allele undergoes decay. Transcriptional adaptation, a process that has been recently observed also in *C. elegans* (Serobyan et al. 2020), relate to the capability of particular mutations in a given gene to modulate the expression levels of other related genes (the so-called adapting genes). This mechanism, for which mRNA decay as well as small RNA biogenesis and transport are critical, explain the phenotypic differences observed among different mutant alleles of the same gene. However, the nature of the *tm292* mutation, in which no intronic region appears downstream the deletion, could theoretically avoid the cellular checkpoint of the nonsense mediated decay response against aberrant messengers. Moreover, the presence of reads in our *fkh-8(tm292)* RNA-seq profiling allowed us to assess the retention of a small intronic region upstream the *tm292* deletion. As a results, both the remains from *fkh-8* intron 4 and exon 8 were joined together, generating a novel stop codon that we consider likely able to generate a truncated FKH-8 protein of 179 amino acids in which the whole FKH DNA binding domain remains unaffected. Accounting for the amino acid changes induced by the *tm292* allele, one could wonder if protein-protein interactions of FKH-8 are also affected in that mutant strain.

cis-mutation of FKH sites suggests compensatory effects upon *fkh-8* loss.

In our hands, analysis of fluorescent reporters of ciliary features in *fkh-8* mutant animals resulted in generally mild phenotypes. Surprisingly, and in contrast to this *trans* analysis, *cis* mutation of a putative FKH site in the promoter sequence of the core panciliary gen *xbx-1* caused a dramatic reduction in both the number of neurons expressing the reporter as well as in the intensity of the fluorescent signal. These results suggest compensatory effects might be elicited by other members of the FKH family binding to the FKH-8 site in its absence or impairment.

C. elegans genome encodes for 18 transcription factors from the FKH family, namely: *daf-16*, *C34B4.2*, *K04C1.3*, *fkh-2* to *fkh-10*, *let-381*, *lin-31*, *pes-1*, *pha-4*, *T27A8.2* and *unc-130*. We checked their expression pattern as assessed by single cell RNA-seq from bulk L2 animals in (J. Cao et al. 2017). We found that 11 of these factors were not expressed in ciliated sensory neurons or expressed at very low levels. Those were the cases of *let-381* (expressed in coelomocytes), *fkh-6* (in excretory cells), *fkh-4* (in body wall muscle), *lin-31* (in excretory cells and vulval precursors), *unc-130* (mainly found in rectum and sex myoblasts), *fkh-2* (in body wall muscle) and *pha-4* (mainly found in pharyngeal muscle, epithelia and gland). On the other hand, *fkh-3*, *T27A8.2*, *fkh-5* and *pes-1*

were reported to be expressed at very low levels within the ciliated system. However, this reported lack of *fkh-2* expression in sensory neurons contrasts with previously described postembryonic expression of a functional *gfp*-tagged reporter being restricted to the nuclei of the AWB neurons, with a less consistent expression in a few additional chemosensory neurons (Mukhopadhyay et al. 2007). Indeed, in AWB, *fkh-2* acts downstream *daf-19*, and it is responsible for the acquisition of the winged-shaped morphology of the AWB cilium through the AWB-exclusive regulation of *kap-1*, a broadly expressed subunit of the heterotrimeric kinesin-II motor protein. Thus, considering both expression pattern and known cilium-related function, *fkh-2* might be seen as a candidate to compensate for *fkh-8* defects in this small subset of ciliated neurons.

On the other hand, 4 FKH transcription factors exhibit broad expression patters across the whole 27 tissue categories from (J. Cao et al. 2017). Namely, *C34B4.2*, *daf-16*, *fkh-7* and *fkh-9*. Interestingly, two of these factors, *daf-16* and *fkh-7* show their most enriched category in that of the ciliated sensory neurons. Finally, *K04C1.3* and *fkh-10* appear mainly expressed in neuronal categories, both ciliated and not ciliated (such as the cholinergic neurons). Profiting for the CeNGEN data (choosing their single-cell RNA-seq from FACS L4 neurons dataset) is possible to further refine the expression pattern of these 6 TFs within the whole ciliated system of the worm. Indeed (see **Table D.1.**), overlapping expression of at least two of these FKH members is found for each of the ciliated neurons, hence further supporting the idea of potential compensatory contributions to the

expression of ciliary features in the absence or impairment of *fkh-8*.

Although other FKH TFs displaying expression within the ciliated system seem the most likely candidates to elicit compensatory effects in *fkh-8* mutants, it is also conceivable that *fkh-8* could act as a repressor of specific FKH members that would be de-repressed in *fkh-8* mutant backgrounds, being then able to compensate for the lack of FKH-8. Interestingly, among the 18 additional FKH TFs, *T27A8.2*, a gene the CeNGEN data address expressed only in the ciliated ADF neuron and the ventral motor neuron VB01, is the only FKH member to be significantly upregulated within the *tm292* mutant background (fold change = 3.61, p-value = 3.12E-03 and a 26.12% probability to score as a false positive). In the future, it will be interesting to use reporter expression analysis for this TF to assess where it is upregulated in *fkh*-8 mutants.

In addition, as an alternative explanation to the loss of reporter expression following the *cis*-mutation of the FKH site present in the promoter of the panciliary gene *xbx-1*, one could wonder if such mutation might be affecting the binding of DAF-19 itself – or any other (yet) unidentified TFs. The putative FKH site found in the promoter sequence of *xbx-1* is in close proximity to that of DAF-19. It is known that specificity in the binding of TFs is compromised by the nucleotides flanking a given TFBSs (Maerkl and Quake 2007) and, additionally, changes in the nucleotide sequence may also induce deviations in the structure of the DNA in which TFs need to allocate (Siggers and Gordân 2014). Since the site-directed mutagenesis that we performed was rather big (changing 9 consecutive

nucleotides) and this sequence is 4 nucleotides apart from the DAF-19 binding site, we cannot rule out this second possibility. Accordingly, we are currently performing additional *cis*-regulatory analysis on other cilium-related enhancers in which FKH TFBSs are more distant to the X-box to assess if similar expression defects to those observed for the *xbx-1* reporter are also observed.

▼ **Table D.1. Several members of the FKH family are expressed in the ciliated sensory neurons of *C. elegans*.**

Expression (•) at the L4 stage, as assessed by (Hammarlund et al. 2018), is indicated for 7 members of the FKH family in all 25 neuronal classes integrating the ciliated system of *C. elegans.*

	IL2	OLL	IL1	OLQ	BAG	CEP	ASK	ADL	ASI	AWB	ASG	AWA	ASE	AFD	ADF	ASH	AWC	ASJ	FLP	ADE	AQR	PDE	PHA	PHB	PQR
fkh-8	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•
C34B4.2	•						•	•			•	•	•	•	•	•	•	•		•			•	•	
daf-16	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
fkh-7	•			•			•	•	•		•		•	•	•	•				•				•	
fkh-9	•		•	•	•	•	•	•			•	•	•		•	•	•	•	•	•	•		•		•
fkh-10	•				•	•	•	•			•	•	•	•	•	•									
K04C1.3	•				•	•	•	•			•	•	•		•	•	•	•							

Interplay between RFX and FKH members is also seen in *C. elegans.*

In vertebrates, members of the RFX family plus the FKH TF FoxJ1 work together in the transcriptional regulation of components in the motile cilium (Didon et al. 2013; Quigley and Kintner 2017). In addition, motile cilia components in chordotonal neurons of *Drosophila* have been shown to be transcriptionally regulated by both *Rfx* (RFX) and *fd3F* (FKH) (Newton et al. 2012). Examples are found in the literature illustrating mutual positive cross-regulation between both families of TFs, with RFX members activating FKH members and *vice versa*. However, no example for FKH members has been described to date to broadly participate in the transcriptional regulation of the sensory ciliome.

As analogously seen in motile cilia systems (reviewed in Thomas et al. 2010 and Choksi et al. 2014), we have also found cooperation between RFX and FKH members regulating the expression of sensory ciliary features in *C. elegans*. Reporter gene expression analyses have revealed that the combined mutation of both *daf-19* and *fkh-8* elicits more severe and synergistic defects in the expression of given structural cilia-related features beyond those expected by the combined effects of single mutants. The possibility of cooperation between both factors was initially supported by the sequence analysis of *fkh-8* genomic binding, since roughly 29% of the *fkh-8* ChIP-seq peaks contain a predicted RFX binding site. Moreover, RNA-seq profiling of *daf-19*, *daf-12*, and *fkh-8* triple mutants show specific gene expression defects not shown in either single *fkh-8* or double *daf-19*, *daf-12* mutants.

As already described in promoters of motile cilia components in *Drosophila* (Newton et al. 2012), our own promoter analysis of structural ciliary features of *C. elegans* showed that these regulatory sequences contain predicted motifs matching both RFX (*daf-19*) and FKH (*fkh-8*) binding sites. A third example of RFX and FKH TFs co-binding is found in the multiciliated cells of *Xenopus* (Quigley and Kintner 2017). However, in the *Xenopus* context, FoxJ1 (FKH) was found to bind at distal sites in the genome and a loop created by the crosslinking of both distal and proximal Rfx2 (RFX) binding sites was able to recruit FoxJ1 to proximal promoter regions to activate the expression of target genes. One could wonder if such docking activity of RFX is also present in *C. elegans*. Consequently, the proximal position of the Rfx2 binding sites in the promoters of motile cilia components of *Xenopus* agrees with the known positional bias observed for the X-boxes motifs bound by *daf-19* in the promoters of structural ciliary features of *C. elegans*. However, in contrast to *Xenopus*, and similar to *Drosophila*, putative FKH binding sites are also found in those promoters (as assessed through ChIP-seq profiling) and located in close proximity to *daf-19* X-boxes. Moreover, our own promoter analysis of ciliome genes often showed the presence of a single X-box motif located within 300 base pairs upstream of their translational site. Since those sequences were 700 base pairs in length, the whole intergenic region was often represented or even surpassed. And although the analysis we performed cannot exclude the participation of FKH members bound at distal regions, both the presence of a single X-box motif plus a nearby putative FKH binding site as well as the compact nature of the *C. elegans* genome suggest that the transcriptional regulatory module for ciliary features is mostly located within the proximal regions of the target genes.

In addition, in this work, through the analysis of fosmid-based reporters, we have unravelled the repressive role that specific isoforms of the RFX member *daf-19* not expressed in ciliated neurons, elicit over the expression of *fkh-8*. In the presence of *daf-19*, *fkh-8* expression remains restricted mainly to the ciliated system of the worm. However, in the full absence of *daf-19*, expression of *fkh-8* is de-repressed showing a pan-neuronal expression pattern. Interestingly, we have discovered that the repressive role of *daf-19* is elicited by the long DAF19A/B isoforms that are mainly express in the non-ciliated neurons of *C. elegans*. In contrast, no major effect is observed when a fosmid-based reporter of *daf-19* is analysed in a *fkh-8* mutant background. Thus, in the interplay between *daf-19* and *fkh-8*, the RFX member seems to modulate the correct expression of the FKH member within the ciliated system of the animal. However, since this interplay was analysed in the *fkh-8(tm292)* allele, further analyses in the *fkh-8* null mutant strain will be necessary to fully elucidate the relationship between the two factors. In addition, our fosmid based *daf-19* reporter labels all *daf-19* isoforms; thus, we cannot assess if *fkh-8* mutation affects the balance of *daf-19* isoform expression (e.g., loss of *daf-19d* isoform and gain of expression of *daf-19a/b* in ciliated neurons will be undetectable with our reporter expression analysis).

RNA-seq uncovers a role for *fkh-8*

in the regulation of the innate immune response in *C. elegans.*

In recent years, several publications (reviewed in Wani, Goswamy, and Irazoqui 2020) have illustrated the role of the nervous system in the regulation of the innate immunity against pathogen infection in *C. elegans.* Intriguingly, as assessed through our RNA-seq experiment, *fkh-8(tm292)* mutants seem to have an impairment in the innate immune response as well as in the detection and reaction to environmental stimuli when compared to wild type worms. Over the past 20 years, accumulated evidence has shown that pathogen-specific programs of gene expression are elicited in response to specific infection with fungi, viruses or bacteria (Ermolaeva and Schumacher 2014; Liu and Sun 2021; Wani et al. 2020).

One of the first evidences demonstrating the key role of the nervous system in the *C. elegans* host defence response was found when mutants defective in neurotransmitter secretion were shown to be resistant to infection by *P. aeruginosa.* (Kawli and Tan 2008). The modulation of the innate immune response by neuroendocrine signals was found to happen through the insulin signalling pathway, since the molecular mechanism behind that response involved the secretion of the insulin-like peptide INS-7, a known DAF-2 agonist. Indeed, loss-of-function mutants lacking the entire coding region of *ins-7* were more resistant to *P. aeruginosa* infection. *ins-7* is expressed in many sensory ciliated neurons and other tissues. In agreement with an altered response to pathogen infection, when compared to wild type worms, *tm292* mutants show a significant upregulation of the *ins-7* gene (fold change = 3.24, p-value = 2.54E-02 with a 25.25% of false positive probability).

Years before this finding, a genetic screen for *C. elegans* mutants displaying enhanced susceptibility to killing by *P. aeruginosa* allowed to identify the conserved p38 MAP kinase signalling pathway as the central signalling module of antimicrobial response in the nematode (Kim et al. 2002). This module is sequentially compose by the genes *tir-1*, *nsy-1*, *sek-1*, *pmk-1* and *atf-7* (Couillault et al. 2004; Shivers et al. 2010). Whereas expression level of *tir-1*, *nsy-1* and *sek-1* remains unaffected in *tm292* mutants, the PMK-1 kinase and the ATF-7 transcription factor are significantly downregulated in these mutants (*pmk-1*: fold change = 0.46, p-value = 3.99E-02 with a 27.57% of false positive probability; *atf-7*: fold change = 0.44, p-value = 3.88E-03 with a 20.81% of false positive probability). Downregulation of ATF-7 could be direct since three different FKH-8 ChIP-seq peaks are located across the *atf-7* locus and ATF-7 is expressed in all 25 classes of ciliated neurons of *C. elegans* (as assessed through the CeNGEN data). Thus, these data strongly suggest a role for *fkh-8* in the modulation of the immune response through the regulation of the final steps in the conserved p38 MAP kinase signalling pathway.

Downregulation of intestinal *pmk-1* has also been reported to occur in a cell-non-autonomous manner through dopamine secreted by the ciliated CEP neurons (Cao and Aballay 2016). In this mechanism, dopamine signalling negatively regulates the innate immune response against *P. aeruginosa* infection through the DOP-4 receptor expressed

by the ciliated ASG neurons. Expression levels of all 5 genes integrating the dopamine pathway in *C. elegans* (namely, *cat-2*, *cat-4*, *bas-1*, *cat-1* and *dat-1*) as well as the *dop-4* gene remain unaffected in *fkh-8(tm292)* when compared to those of wild type worms, and we confirmed no expression defects for *dat-1* by reporter analysis. In addition, down-regulation of *pmk-1* had also been described to occur through the putative octopamine G protein-coupled catecholamine receptor OCTR-1 (Sun et al. 2011). Interestingly, activation of OCTR-1 in the ciliated ASH neurons inhibit the p38 MAK kinase signalling pathway, downregulating the immune response. Accordingly, *octr-1* mutants show enhanced resistance to infection (X. Cao et al. 2017). Intriguingly, these mutants also show an upregulation of the *nlp-20* gene, a neuropeptide-like protein (*nlp*) which is expressed in a small subset of neurons located in head and tail and is required for the enhanced resistance *P. aeruginosa* infection. Consistent with these results, *tm292* mutants show a significant downregulation of the *nlp-20* gene (fold change = 0.33, p-value = 1.43E-02 with a 23.73% of false positive probability) when compared to wild type worms. Of note, *ocrt-1* expression remains unaffected in *tm292* mutants.

In *C. elegans*, following infection by *S. aureus*, the release of neuronal acetylcholine triggers intestinal induction of host defence genes via conserved muscarinic and Wnt pathways (Labed et al. 2018). No gene encoding for muscarinic receptors in the worm (*gar* class), appears deregulated in *fkh-8(tm292)* mutants. However, from the 25 different component that are known to participate in the Wnt signalling in *C. elegans*

(Eisenmann 2005), 5 appear significantly deregulated in *fkh-8(tm292)* mutants. This muscarinic-Wnt pathway constitute a brain-gut axis that has been shown to be linked by the action of several TFs, including the ETS TF *lin-1* (not affected in *tm292* mutants). In turn, activation of the canonical Wnt pathway in the intestinal epithelium triggers the expression of antimicrobials genes such as the C-type lectin *clec-60* and as well as lysozymes (Labed et al. 2018). From the 219 members of the *clec* class from which we retrieved expression data in our RNA-seq experiment, 32 resulted significantly downregulated in *tm292* mutants. Among them we found *clec-60* (fold change = 0.25, p-value = 4.66E-02 with a 28.68% of false positive probability). Additionally, 14 *clec* genes resulted significantly upregulated in the comparison between *fkh-8(tm292)* mutants and wildtype worms. Moreover, we also found half of the genes (5 out of 10) encoding for lysozymes in *C. elegans* significantly downregulated (namely, *lys-1* to *lys-5*) as well as 2 out of the 6 members of the invertebrate lysozyme gene class (namely, *ilys-2* and *ilys-5*) the worm possess. To date, the mechanism by which the sensory nervous system of *C. elegans* detects *S. aureus* remains unknown.

Concluding remarks.

Cilia are evolutionary conserved eukaryotic organelles that play major roles in the control of key features for animal survival such as the perception, cell locomotion or the capability to induce the circulation of extracellular fluids. A reflection of their importance is seen in the plethora of human diseases, globally termed as ciliopathies, that arise when genes related to or coding for cilia

components are mutated. These inherited human disorders constitute complex conditions that are often difficult to diagnose, being both multisystemic and highly pleiotropic. To worsen this scenario, and since a marked overlap exists among different ciliopathies, patients often encounter a situation in which they are diagnosed with different diseases depending on which pathologies their clinicians have expertise on.

Since cilia were uncovered as a hub for cell signalling, additional layers of complexity were revealed regarding the control of ciliogenesis. Today, cilia have been found to be regulated by a net of interconnected pathways in which reciprocal antagonistic or synergistic influences often act in parallel involving cell cycle, structural influences of the cytoskeleton, cellular proteostasis and cilia-mediated signalling itself (see Malicki and Johnson 2017).

Twenty years have passed since the transcriptional regulation of ciliome genes was found to be mediated by transcription factors of the RFX family. Originally uncovered in *C. elegans* (Swoboda, Adler, and Thomas 2000), the RFX-mediated control of ciliogenesis was soon established as conserved among ciliated organisms (see Choksi et al. 2014). A second milestone in deciphering the regulatory logic behind ciliogenesis was found in the FKH family of TFs. Today, the key role of FOXJ1 in the specification of the motile ciliated fate is well known. Moreover, solid evidence has been accumulated showing the cooperative action between RFX and FOXJ1 acting as terminal selectors of motile cilia. In addition, several examples gathered from different model organisms (reviewed in Thomas et al.

2010 and Choksi et al. 2014) have linked cell type-specific transcription factors such as Atonal (ATO), MCIDAS, MYB or the notochord homeobox (NOTO) to the regulation of the acquisition of motile cilia diversity.

In contrast, in these last 20 years, no additional TF other than members from the RFX family have been found to directly regulate the activation of the genes integrating the sensory ciliome. In this work we show, for the first time, experimental evidence linking the correct expression of core structural ciliary features of *C. elegans* to a member of the FKH family: *fkh-8*. Illustrating its broad requirements for the proper functionality of ciliated neurons, we show that *fkh-8* mutant animals are defective in mechanosensory, olfactory and gustatory responses.

Several questions arise as a result of this Thesis work. Maybe the most obvious one stands for the mechanism by which *fkh-8* exerts its action. Since gene expression analyses revealed generally mild phenotypes even for the genes harbouring *fkh-8* ChIP-seq peaks, what is then responsible for the mechanosensory, olfactory and gustatory defects that *fkh-8* mutants display? If those features are being correctly expressed (or at least, adequately compensated), is *fkh-8* related to the correct trafficking of ciliary features? Moreover, is the ultrastructure of the cilium affected in *fkh-8* mutant animals?

Similar to the interplay seen between the RFX family and FOXJ1 in the transcriptional regulation of the motile ciliome, we show evidence for the cooperation between *daf-19* (RFX) and *fkh-8* (FKH) in the transcriptional regulation of the sensory ciliome in *C. elegans*. Considering this, are TFs of the FKH

family also involved in transcriptional regulation of the sensory ciliome in other organisms? Additionally, our promoter analysis of structural ciliary features has shown how these regulatory sequences harbours both *daf-19* and *fkh-8* binding sites in close proximity to each other. This evidence allows for the envisioning of a physical interaction between both TFs. Is *daf-19* regulating *fkh-8* accessibility to the DNA or is it the opposite? Since we observed specific *daf-19* isoforms repress *fkh-8* expression outside the ciliated system, how is that repression elicited if all *daf-19* isoforms possess the same DNA binding domain? And moving away from the ciliated system, what links the neuronal defects induced by the mutation of *fkh-8* with the dramatic deregulation of the innate immune response in the worm?

Much more research will be needed to decipher all of these questions. Meanwhile, we honestly wish that the insights this work has provided regarding the understanding of the transcriptional regulatory logic behind primary cilia formation will ultimately benefit those people whose lives have been impacted by ciliopathies.

Conclussions

In this Thesis project we aimed to identify *daf-19* co-operators and new transcriptional regulators of the ciliogenesis in *C. elegans*. The results obtained in this work lead to the following conclusions:

1 ➤ *daf-19* is not required for the expression of some ciliome genes in specific sensory neurons, suggesting the existence of other transcriptional regulators of the ciliogenesis in *C. elegans*.

2 ➤ X-box binding sites embedded in small regulatory sequences are more dependent on DAF-19 than the same X-box analysed in the context of a larger regulatory sequence, hence unravelling compensatory actions by additional TFs on ciliome gene expression.

3 ➤ Features coding for structural panciliary components show a high dependency towards DAF-19 loss whereas expression of members from the TRPV family, coding for subtype-specific features, seems largely independent of DAF-19.

4 ➤ Lack of *daf-12*, a nuclear hormone receptor TF, has a small but significant effect on the expression of both structural and subtype-specific ciliary features.

5 ➤ Putative regulatory sequences of genes encoding for structural panciliary features are enriched in RFX-related motifs and pro-neuronal bHLH-matching sites. Suboptimal alignments for additional motifs identify enrichment for Forkhead, Homeodomain and Nuclear Hormone Receptor binding sites.

6 ➤ Ten transcription factors are highly and specifically expressed within the ciliated sensory neurons of *C. elegans*. In agreement with our *in silico* findings, these TFs belong to only four different TF families, namely: Zinc Finger (ZF - C2H2) (*Y22D7AL.16*), Forkhead (FKH) (*fkh-8*), nuclear hormone receptors (NHR) (*nhr-158*, *nhr-188*, *nhr-124*, *nhr-30*, *nhr-38*, *nhr-216*, *nhr-277*) and homeodomain (HD) (*ceh-57*).

7 ➤ *In vivo* reporter analysis demonstrates that *fkh-8* is selectively expressed in the whole ciliated system of *C. elegans*. FKH-8 expression onset in development overlaps that of the neurogenic period and early differentiation of ciliated neurons. FKH-8 expression is maintained throughout the whole life of the animal.

8 ➤ FKH-8 preferentially binds regulatory regions of genes encoding for structural components of the cilium. The presence of FKH-related motifs clustered at the centre of FKH-8 ChIP-seq peaks suggests direct binding at least for a subset of those genes.

9 ➤ *fkh-8(tm292)* young adult mutants do not show downregulation of genes directly related to ciliary genes, cilium-related neuronal functions, or FKH-8-bound loci.

fkh-8, as assessed through RNA-seq data, regulates several branches of the innate immune response in *C. elegans*. In the case of the conserved p38 MAP Kinase pathway, this action could be mediated through direct activation of the ATF-7 TF in ciliated neurons. ◄ **10**

Double null mutation of *daf-12* and *daf-19* produces severe deregulation of the transcriptome, including downregulation of neuronal and sensory-related features. However, ciliary genes or functions are not affected. In addition, a notable upregulation of genes related to germline and reproductive processes is also observed. ◄ **11**

fkh-8(tm292) is likely a hypomorphic allele in which a novel, truncated messenger is effectively transcribed. This messenger could generate a truncated protein of 180 amino acids in which the whole FKH DNA domain remains intact. ◄ **12**

Expression of structural ciliary features is affected in *fkh-8* mutants in a cell-type-specific and gene-specific manner. Expression defects were observed in IL2, CEP, ADE, PDE and several unidentified labial, amphid and phasmid neurons. ◄ **13**

fkh-8 is required for the correct display of a wide range of sensory-mediated behaviours, including olfaction, gustation and cilium-related mechanosensation. Assayed behavioural paradigms demonstrate FKH-8 requirements for the correct functionality of the ciliated ASH, FLP, OLQ, AWC, AWB, ASK, ADL, and ASE neurons. ◄ **14**

The partial deletion of the *fkh-8(tm292)* allele has no effect over the *in vivo* fosmid-based expression of *daf-19*. However, RNA-seq results suggest a repressive action for *fkh-8* over *daf-19*. ◄ **15**

Long DAF-19 isoforms, not expressed in ciliated neurons, repress the expression of *fkh-8* in non-ciliated neurons. ◄ **16**

fkh-8 and *daf-19* act synergistically in the transcriptional regulation of structural ciliary features. ◄ **17**

Accounting for the aforementioned statements, we conclude that:

fkh-8 acts as a terminal selector of the sensory ciliome in *C. elegans*

Annexes

➤ **Annex 1. List of *C. elegans* strains used in this work.**

Strain	Genotype	Source
N2	*wild type*	CGC
JT204	*daf-12(sa204) X*	CGC
JT6924	*daf-12(sa204) X; daf-19(m86) II*	CGC
NFB1426	*vlcEx827[che-11p::NLS::gfp, rol-6(su1006)]*	This work
NFB2113	*vlcEx827[che-11p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2098	*vlcEx827[che-11p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB1411	*vlcEx812[che-13p::NLS::gfp, rol-6(su1006)]*	This work
NFB2114	*vlcEx812[che-13p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2099	*vlcEx812[che-13p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB2084	*vlcEx1133[osm-5p::NLS::gfp, rol-6(su1006)]*	This work
NFB2120	*vlcEx1133[osm-5p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2105	*vlcEx1133[osm-5p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
-	*otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; him-5(e1490) V*	Dr. Hobert Lab.
NFB2128	*otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; daf-12(sa204)*	This work
NFB2019	*otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; daf-12(sa204); daf-19(m86)*	This work
NFB1462	*vlcEx835[peli-1p::NLS::gfp, rol-6(su1006)]*	This work
NFB2122	*vlcEx835[peli-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2107	*vlcEx835[peli-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB1423	*vlcEx824[peli-1p::NLS::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]*	This work
NFB2123	*vlcEx824[peli-1p::NLS::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; daf-12(sa204)*	This work
NFB2108	*vlcEx824[peli-1p::NLS::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; daf-12(sa204); daf-19(m86)*	This work
NFB2091	*vlcEx1137[xbx-1p::NLS::gfp, rol-6(su1006)]*	This work
NFB2126	*vlcEx1137[xbx-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2111	*vlcEx1137[xbx-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB1421	*vlcEx822[xbx-1p::NLS::gfp, rol-6(su1006)]*	This work
NFB2127	*vlcEx822[xbx-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2112	*vlcEx822[xbx-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB2088	*vlcEx1135[tmem-107p::NLS::gfp, rol-6(su1006)]*	This work
NFB2124	*vlcEx1135[tmem-107p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2109	*vlcEx1135[tmem-107p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work

Strain	Genotype	Source
NFB1420	*vlcEx821[tub-1p::NLS::gfp, rol-6(su1006)]*	This work
NFB2125	*vlcEx821[tub-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2110	*vlcEx821[tub-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB2081	*vlcEx1131[mks-1p::NLS::gfp, rol-6(su1006)]*	This work
NFB2115	*vlcEx1131[mks-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2100	*vlcEx1131[mks-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB1414	*vlcEx815[osm-1p::NLS::gfp, rol-6(su1006)]*	This work
NFB2119	*vlcEx815[osm-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2104	*vlcEx815[osm-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB1424	*vlcEx825[ocr-1p::NLS::gfp, rol-6(su1006)]*	This work
NFB2116	*vlcEx825[ocr-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2101	*vlcEx825[ocr-1p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB1413	*vlcEx832[ocr-2p::NLS::gfp, rol-6(su1006)]*	This work
NFB2117	*vlcEx832[ocr-2p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2102	*vlcEx832[ocr-2p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB1433	*vlcEx833[ocr-4p::NLS::gfp, rol-6(su1006)]*	This work
NFB2118	*vlcEx833[ocr-4p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2103	*vlcEx833[ocr-4p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
NFB1417	*vlcEx818[osm-9p::NLS::gfp, rol-6(su1006)]*	This work
NFB2121	*vlcEx818[osm-9p::NLS::gfp, rol-6(su1006)]; daf-12(sa204)*	This work
NFB2106	*vlcEx818[osm-9p::NLS::gfp, rol-6(su1006)]; daf-12(sa204); daf-19(m86)*	This work
OP652	*wgIs652 [fkh-8::TY1::EGFP::3xFLAG + unc-119(+)]; unc-119(tm4063) III*	CGC
NFB2055	*wgIs652 [fkh-8::TY1::EGFP::3xFLAG + unc-119(+)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]*	This work
BC13336	*sIs13113 [rCesY110A7A.20::GFP + pCeh361]; dpy-5(e907) I*	CGC
-	*fkh-8(tm292) II*	NBRP
NFB1776	*otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(tm292) II*	This work
NFB2057	*otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(vlc43) II*	This work
NFB2132	*vlcEx815[osm-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II*	This work
NFB2235	*vlcEx815[osm-1p::NLS::gfp, rol-6(su1006)]; fkh-8(vlc43) II*	This work
NFB2138	*vlcEx821[tub-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II*	This work
NFB2145	*vlcEx821[tub-1p::NLS::gfp, rol-6(su1006)]; fkh-8(vlc43) II*	This work
NFB2140	*vlcEx822[xbx-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II*	This work

Strain	Genotype	Source
NFB2147	vlcEx822[xbx-1p::NLS::gfp, rol-6(su1006)]; fkh-8(vlc43) II	This work
NFB2139	vlcEx1137[xbx-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II	This work
NFB2146	vlcEx1137[xbx-1p::NLS::gfp, rol-6(su1006)]; fkh-8(vlc43) II	This work
NFB2136	vlcEx824[peli-1p::NLS::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(tm292) II	This work
NFB2143	vlcEx824[peli-1p::NLS::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(vlc43) II	This work
NFB2135	vlcEx835[peli-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II	This work
NFB2142	vlcEx835[peli-1p::NLS::gfp, rol-6(su1006)]; fkh-8(vlc43) II	This work
NFB2133	vlcEx1133[osm-5p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II	This work
NFB2155	vlcEx1133[osm-5p::NLS::gfp, rol-6(su1006)]; fkh-8(vlc43) II	This work
NFB2137	vlcEx1135[tmem-107p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II	This work
NFB2144	vlcEx1135[tmem-107p::NLS::gfp, rol-6(su1006)]; fkh-8(vlc43) II	This work
OH13646	otIs544 [cho-1(fosmid)::SL2::mCherry::H2B + pha-1(+)]; pha-1(e2123) III; him-5(e1490) V	Dr. Hobert Lab.
NFB2520	vlcEx835[peli-1p::NLS::gfp, rol-6(su1006)]; otIs544 [cho-1(fos-mid)::SL2::mCherry::H2B + pha-1(+)]	This work
NFB177	otIs259(dat-1p::gfp, rol-6(su1006)]	This work
NFB1756	otIs259(dat-1p::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]	This work
NFB1776	otIs259(dat-1p::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(tm292) II	This work
NFB2447	otIs259(dat-1p::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(vlc43) II	This work
NFB2363	vlcEx1241[dat-1p::fkh-8::T2A::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(vlc43) II	This work
NFB2364	vlcEx1242[dat-1p::fkh-8::T2A::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(vlc43) II	This work
NFB2365	vlcEx1243[dat-1p::fkh-8::T2A::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(vlc43) II	This work
NFB2092	vlcEx1138[xbx-1p::NLS::gfp, rol-6(su1006)]	This work
NFB2095	vlcEx1140[xbx-1p::NLS::gfp, rol-6(su1006)]	This work
NFB2096	vlcEx1141[xbx-1p::NLS::gfp, rol-6(su1006)]	This work
NFB2097	vlcEx1142[xbx-1p::NLS::gfp, rol-6(su1006)]	This work
CB1033	che-2(e1033) X	CGC
CX2065	odr-1(n1936) X	CGC
CX32	odr-10(ky32) X	CGC
PR678	tax-4(p678) III	CGC

Strain	Genotype	Source
KP4	*glr-1(n2461) III*	CGC
CB1338	*mec-3(e1338) IV*	CGC
OH7193	*otIs181[dat-1::mCherry + ttx-3::mCherry]III; him-8(e1489)IV*	Dr. Hobert Lab.
NFB2267	*vlcEx361[daf-19fos::GFP, rol-6(su1006)]; otIs181[dat-1::mCherry + ttx-3::mCherry] III*	This work
NFB2266	*vlcEx361[daf-19fos::GFP, rol-6(su1006)]; otIs181[dat-1::mCherry + ttx-3::mCherry] III; fkh-8(tm292) II*	This work
OH13645	*otIs518 [eat-4(fosmid)::SL2::mCherry::H2B + pha-1(+)]; pha-1(e2123) III; him-5(e1490) V*	Dr. Hobert Lab.
EO5	*daf-19(of5) II*	Dr. Swoboda Lab.
NFB2237	*wgIs652 [fkh-8::TY1::EGFP::3xFLAG + unc-119(+)]; otIs518 [eat-4(fosmid)::SL2::mCherry::H2B + pha-1(+)]; daf-12(sa204); daf-19(m86)*	This work
NFB2456	*wgIs652 [fkh-8::TY1::EGFP::3xFLAG + unc-119(+)]; daf-19(of5) II*	This work
NFB2519	*wgIs652 [fkh-8::TY1::EGFP::3xFLAG + unc-119(+)]; otIs181[dat-1::mCherry + ttx-3::mCherry]III*	This work
NFB2476	*wgIs652 [fkh-8::TY1::EGFP::3xFLAG + unc-119(+)]; otIs181[dat-1::mCherry + ttx-3::mCherry]III; daf-12(sa204); daf-19(m86)*	This work
NFB2150	*otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(tm292) II; daf-12(sa204) X*	This work
NFB2153	*otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(tm292) II; daf-12(sa204) X; daf-19(m86) II*	This work
NFB2239	*vlcEx1137[xbx-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II; daf-12(sa204) X*	This work
NFB2428	*vlcEx1137[xbx-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II; daf-12(sa204) X; daf-19(m86) II*	This work
NFB2149	*vlcEx822[xbx-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II; daf-12(sa204) X*	This work
NFB2152	*vlcEx822[xbx-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II; daf-12(sa204) X; daf-19(m86) II*	This work
NFB2238	*vlcEx815[osm-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II; daf-12(sa204) X*	This work
NFB2350	*vlcEx815[osm-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II; daf-12(sa204) X; daf-19(m86) II*	This work
NFB2236	*vlcEx835[peli-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II; daf-12(sa204) X*	This work
NFB2349	*vlcEx835[peli-1p::NLS::gfp, rol-6(su1006)]; fkh-8(tm292) II; daf-12(sa204) X; daf-19(m86) II*	This work
NFB2192	*vlcEx824[peli-1p::NLS::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(tm292) II; daf-12(sa204) X*	This work
NFB2191	*vlcEx824[peli-1p::NLS::gfp, rol-6(su1006)]; otIs395[ift-20p::NLS::tagRFP + pha-1(+)]; fkh-8(tm292) II; daf-12(sa204) X; daf-19(m86) II*	This work

➤ **Annex 2. On-line resources used in this work.**

This table gather the URLs of websites and on-line resources named within the different sections of the main text.

Resource	URL
"A Cell Atlas of Worm"	http://atlas.gs.washington.edu/worm-rna/
ChIPseek	http://chipseek.cgu.edu.tw.
CIS-BP	http://cisbp.ccbr.utoronto.ca/
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/
ENCODE	https://www.encodeproject.org/
Ensembl BioMart	https://www.ensembl.org/biomart/martview/
***fkh-8* ChIP-seq bed nar-rowPeak File**	https://www.encodeproject.org/files/ENCFF653QKE/
Gene Ontology	http://geneontology.org/
GExplore$_{1.4}$	http://genome.sfu.ca/gexplore/gexplore_about.html
IGV	https://software.broadinstitute.org/software/igv/
MEME Suite	http://meme-suite.org/ (to avoid massification issues, the use of the Alternate Servers is advisable)
modERN *fkh-8* ChIP-seq experiment summary	https://www.encodeproject.org/experiments/ENCSR435RXY/
RSAT	http://rsat.sb-roscoff.fr/
RStudio	https://rstudio.com/
SCeNGEA	https://cengen.shinyapps.io/SCeNGEA/
SCope	http://scope.aertslab.org/
VassarStats	http://vassarstats.net/
WormBase Gene Set En-richment Analysis	https://wormbase.org/tools/enrichment/tea/tea.cgi
WormEnrichr	https://amp.pharm.mssm.edu/WormEnrichr/

➤ **Annex 3. Detailed scorings from fluorescent reporters of cilium-related features of *C. elegans*.**
The following tables gather the values obtained in the detailed region-by-region analysis of fluorescent reporters for ciliary features of *C. elegans*. Values show the mean and standard deviation for the number of reporter-positive cells observed in each region. N: total number of worms analysed per reporter. Standard deviation is also shown for the final total mean.

Subtype-specific ciliary features:

*ocr-1*p::NLS::GFP Line 1	Region 3			Region 1			Region 4			Region 5			Region 2			N	Total Mean		
N2	0,75	±	0,87	2,00	±	0,43	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	12	2,75	±	1,06
daf-12(sa204)	0,33	±	0,62	2,60	±	1,06	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	15	2,93	±	1,28
daf-19(m86); *daf-12*(sa204)	0,50	±	1,08	1,60	±	0,70	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	10	2,10	±	1,10

*ocr-2*p::NLS::GFP Line 1																			
N2	0,00	±	0,00	12,83	±	1,03	1,67	±	0,65	1,58	±	0,51	5,92	±	0,51	12	22,00	±	2,00
daf-12(sa204)	0,00	±	0,00	11,00	±	1,00	0,22	±	0,44	1,78	±	0,44	5,33	±	1,00	9	18,33	±	1,12
daf-19(m86); *daf-12*(sa204)	0,00	±	0,00	9,33	±	1,00	0,22	±	0,67	2,00	±	0,00	5,33	±	0,71	9	16,89	±	1,45

*ocr-4*p::NLS::GFP Line 1																			
N2	3,70	±	0,48	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	10	3,70	±	0,48
daf-12(sa204)	3,83	±	0,39	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	12	3,83	±	0,39
daf-19(m86); *daf-12*(sa204)	3,90	±	0,32	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	10	3,90	±	0,32

*osm9*p::NLS::GFP Line 1																			
N2	4,10	±	0,99	5,10	±	1,10	0,00	±	0,00	0,00	±	0,00	0,10	±	0,32	10	9,30	±	1,77
daf-12(sa204)	4,00	±	0,00	4,25	±	0,75	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	12	8,25	±	0,75
daf-19(m86); *daf-12*(sa204)	4,00	±	0,50	3,67	±	1,00	0,00	±	0,00	0,00	±	0,00	0,00	±	0,00	9	7,67	±	1,22

Structural ciliary features:

*che-11*p::NLS::GFP Line 1	Region 3	Region 1	Region 4	Region 5	Region 2	N	Total Mean
N2	13,38 ± 2,97	19,88 ± 1,13	1,13 ± 0,64	0,75 ± 0,71	4,13 ± 0,35	8	39,25 ± 4,20
daf-12(sa204)	15,25 ± 1,39	17,75 ± 1,98	1,25 ± 1,04	0,50 ± 0,76	4,75 ± 0,46	8	39,50 ± 2,88
daf-19(m86); *daf-12*(sa204)	5,25 ± 1,28	0,88 ± 0,99	0,38 ± 1,06	0,00 ± 0,00	0,88 ± 0,35	8	7,38 ± 1,60

*che-13*p::NLS::GFP Line 1							
N2	18,60 ± 1,52	23,60 ± 2,70	2,80 ± 0,45	1,60 ± 0,55	5,00 ± 0,00	5	51,60 ± 1,82
daf-12(sa204)	17,71 ± 2,50	20,71 ± 1,89	2,00 ± 1,15	0,71 ± 0,76	4,71 ± 0,49	7	45,86 ± 3,85
daf-19(m86); *daf-12*(sa204)	1,79 ± 0,58	1,86 ± 0,53	0,00 ± 0,00	0,00 ± 0,00	0,00 ± 0,00	14	3,64 ± 0,93

ift-20::NLS::RFP Int.							
N2	22,80 ± 1,32	23,50 ± 0,53	2,70 ± 1,25	2,00 ± 0,00	4,90 ± 0,32	10	55,90 ± 2,47
daf-12(sa204)	14,50 ± 2,54	18,92 ± 2,71	2,50 ± 0,67	1,83 ± 0,39	4,33 ± 0,49	12	42,08 ± 4,54
daf-19(m86); *daf-12*(sa204)	8,44 ± 2,35	0,22 ± 0,67	0,11 ± 0,33	0,00 ± 0,00	1,33 ± 1,00	9	10,11 ± 2,76

*mks-1*p::NLS::GFP Line 1							
N2	5,31 ± 0,63	6,08 ± 1,38	0,00 ± 0,00	0,00 ± 0,00	3,77 ± 0,60	13	15,15 ± 1,21
daf-12(sa204)	4,90 ± 0,99	4,30 ± 1,16	0,00 ± 0,00	0,00 ± 0,00	2,50 ± 1,27	10	11,70 ± 2,50
daf-19(m86); *daf-12*(sa204)	0,00 ± 0,00	0,00 ± 0,00	0,00 ± 0,00	0,00 ± 0,00	0,00 ± 0,00	12	0,00 ± 0,00

*osm-1*p::NLS::GFP Line 1							
N2	19,78 ± 1,56	22,56 ± 2,65	1,22 ± 1,30	0,11 ± 0,33	4,56 ± 0,73	9	48,22 ± 2,99
daf-12(sa204)	14,71 ± 0,95	19,00 ± 1,83	0,71 ± 0,49	0,56 ± 0,79	4,29 ± 0,76	7	39,29 ± 2,14
daf-19(m86); *daf-12*(sa204)	3,42 ± 1,08	3,42 ± 2,02	0,00 ± 0,00	0,00 ± 0,00	1,83 ± 0,39	12	8,67 ± 1,67

*osm-5*p::NLS::GFP Line 1							
N2	4,13 ± 1,96	6,25 ± 1,28	1,25 ± 0,89	0,88 ± 0,64	3,75 ± 0,71	8	16,25 ± 1,49
daf-12(sa204)	6,60 ± 2,37	6,90 ± 2,18	1,80 ± 1,14	0,40 ± 0,70	3,60 ± 1,26	10	19,30 ± 4,24
daf-19(m86); *daf-12*(sa204)	0,42 ± 0,79	0,00 ± 0,00	0,00 ± 0,00	0,00 ± 0,00	0,08 ± 0,29	12	0,50 ± 0,80

*peli-1*p::NLS::GFP (S) Line 1							
N2	6,71 ± 2,14	14,14 ± 2,04	0,00 ± 0,00	0,00 ± 0,00	4,43 ± 0,79	7	25,29 ± 2,50
daf-12(sa204)	5,89 ± 1,17	11,56 ± 0,53	1,53 ± 0,73	0,00 ± 0,00	4,67 ± 0,50	9	23,67 ± 1,50
daf-19(m86); *daf-12*(sa204)	4,10 ± 0,99	1,30 ± 0,82	0,40 ± 0,52	0,00 ± 0,00	0,90 ± 0,57	10	6,70 ± 1,77

*peli-1*p::NLS::GFP (L) Line 1							
N2	18,00 ± 2,61	18,33 ± 1,97	1,83 ± 0,98	2,00 ± 0,00	4,83 ± 1,17	6	45,00 ± 4,69
daf-12(sa204)	18,75 ± 0,71	17,13 ± 1,36	2,00 ± 0,00	1,63 ± 0,74	3,63 ± 0,92	8	43,13 ± 2,17
daf-19(m86); *daf-12*(sa204)	13,44 ± 1,51	10,33 ± 2,12	2,00 ± 0,50	1,56 ± 0,53	0,56 ± 0,88	9	27,89 ± 1,76

*tmem-107*p::NLS::GFP Line 1							
N2	15,67 ± 1,75	13,67 ± 2,07	1,00 ± 0,63	0,50 ± 0,55	4,17 ± 0,75	6	35,00 ± 4,43
daf-12(sa204)	8,90 ± 2,02	8,00 ± 2,31	0,90 ± 0,88	0,00 ± 0,00	4,40 ± 1,07	10	22,20 ± 4,54
daf-19(m86); *daf-12*(sa204)	0,00 ± 0,00	0,00 ± 0,00	0,00 ± 0,00	0,00 ± 0,00	0,00 ± 0,00	12	0,00 ± 0,00

*tub-1*p::NLS::GFP Line 2							
N2	20,90 ± 1,37	23,30 ± 1,70	2,30 ± 0,82	1,90 ± 0,32	4,40 ± 0,84	10	52,80 ± 2,62
daf-12(sa204)	20,60 ± 1,65	21,90 ± 1,73	4,00 ± 0,94	2,00 ± 0,00	4,60 ± 0,97	10	53,10 ± 2,92
daf-19(m86); *daf-12*(sa204)	7,55 ± 2,11	6,73 ± 0,90	3,73 ± 1,35	1,36 ± 0,50	6,00 ± 1,10	11	25,36 ± 2,29

*xbx-1*p::NLS::GFP (S) Line 1							
N2	19,40 ± 1,52	18,60 ± 2,70	2,60 ± 0,55	1,60 ± 0,55	4,60 ± 0,89	5	46,80 ± 3,42
daf-12(sa204)	17,80 ± 1,30	16,40 ± 2,07	3,00 ± 2,12	1,80 ± 0,45	3,60 ± 0,55	5	42,60 ± 4,98
daf-19(m86); *daf-12*(sa204)	4,92 ± 1,08	0,00 ± 0,00	0,33 ± 0,65	0,25 ± 0,45	1,00 ± 0,74	12	6,50 ± 1,31

*xbx-1*p::NLS::GFP (L) Line 1							
N2	20,00 ± 1,22	24,20 ± 1,92	2,20 ± 0,45	1,20 ± 0,45	4,40 ± 0,89	5	52,00 ± 3,67
daf-12(sa204)	19,67 ± 1,37	19,83 ± 1,17	2,67 ± 0,52	2,00 ± 0,00	4,33 ± 0,52	6	48,50 ± 2,26
daf-19(m86); *daf-12*(sa204)	9,83 ± 1,27	4,00 ± 1,04	1,75 ± 0,62	1,75 ± 0,45	3,75 ± 1,36	12	21,08 ± 2,50

> ➤ **Annex 4. Ciliome gene list built for this work.**

Three primary sources were used to select well-known genes coding for cilia components: GO cellular component (C.C.) and biological process (B.P.) ontologies together with genes used in (Burghoorn *et al.* 2012). A manual curation procedure (C) using Wormbase information was also performed to further refine this list. Functional categorization was employed to divide the genes in four categories. The references appear as numbers whose key is as follows:

| 1: (Aoki et al. 2011) | 2: (Bae et al. 2009) | 3: (Barr and Sternberg 1999) | 4: (Birnby et al. 2000) | 5: (Blacque et al. 2005b) | 6: (Brear et al. 2014) | 7: (Burghoorn et al. 2012) | 8: (Chang | Johnston | and Hobert 2003) | 9: (Chatzigeorgiou et al. 2013) | 10: (C. Chen et al. 2014) | 11: (Dammermann et al. 2009) | 12: (Schafer et al. 2003) | 13: (Hallem and Sternberg 2008) | 14: (Hao et al. 2011) | 15: (Hu et al. 2006) | 16: (Hu and Barr 2005) | 17: (Huang et al. 2007) | 18: (Hunt-Newbury et al. 2007) | 19: (Inada et al. 2006) | 20: (Inglis et al. 2007) | 21: (Iwata et al. 2011) | 22: (Jauregui et al. 2008) | 23: (Jensen et al. 2016) | 24: (Kagoshima and Kohara 2015) | 25: (Kim and Li 2004) | 26: (Kunitomo et al. 2005) | 27: (Kypri et al. 2014) | 28: (Lambacher et al. 2016) | 29: (Lee et al. 2000) | 30: (Li et al. 2016) | 31: (Li | Yi | and Ou 2015) | 32: (Maguire et al. 2015) | 33: (McGrath et al. 2011) | 34: (Miller and Portman 2010) | 35: (Moussaif and Sze 2009) | 36: (Mukhopadhyay et al. 2005) | 37: (Nguyen et al. 2014) | 38: (Olivier-Mason et al. 2013) | 39: (Ortiz et al. 2006) | 40: (Ou et al. 2007) | 41: (Park et al. 2012) | 42: (Phirke et al. 2011) | 43: (Ramulu and Nathans 2001) | 44: (Russell et al. 2014) | 45: (Sherman et al. 2005) | 46: (Swoboda et al. 2000a) | 47: (Tanis et al. 2013) | 48: (Tobin et al. 2002) | 49: (VanHoven et al. 2006) | 50: (Venkatachalam, Luo, and Montell 2014) | 51: (Wei et al. 2013) | 52: (Williams et al. 2011) | 53: (Wojtyniak et al. 2013) | 54: (Xu et al. 2015) | 55: (Yu et al. 1997) |

Gene	WormBase ID	C.C.	B.P.	B	C	Category	References
aqp-6	WBGene00000174	X				Subtype-specific	17
arl-13	WBGene00021349	X				Structural	5, 6, 18, 28, 30, 37
arl-3	WBGene00000188				X	Subtype-specific	3, 4, 8, 13, 19, 41, 46
arl-6	WBGene00000193	X				Subtype-specific	13, 22
asb-1	WBGene00000206	X				Male	50
asg-2	WBGene00000210	X				Male	50
asic-1	WBGene00022815				X	Subtype-specific	44
asic-2	WBGene00012137			X		Subtype-specific	7
atp-2	WBGene00000229	X				Male	18, 50
bbs-1	WBGene00000241	X		X		Structural	5, 6, 7, 12, 18, 20, 26, 40, 42
bbs-2	WBGene00000242	X		X		Structural	5, 7, 12, 18, 20, 26, 51
bbs-4	WBGene00043992	X				Structural	20, 54
bbs-5	WBGene00010974	X		X		Structural	2, 5, 7, 18, 20, 54
bbs-8	WBGene00000244	X		X		Structural	6, 7, 12, 14, 18, 20, 23, 26, 37, 38, 40, 55
bbs-9	WBGene00016744	X				Structural	5, 20, 51
ccep-290	WBGene00021643	X				Structural	30
cdkl-1	WBGene00012779	X				Expression	30
cfap-36	WBGene00022435				X	Structural	23
che-10	WBGene00018051	X	X			Structural	8, 13, 20, 40
che-11	WBGene00000490	X		X		Structural	5, 7, 8, 11, 12, 18, 20, 22, 23, 31, 37, 39, 40, 42, 51, 55
che-12	WBGene00000491		X			Structural	3, 8, 20,
che-13	WBGene00000492	X		X		Structural	5, 7, 8, 12, 20, 22, 30, 40, 42, 59, 55
che-2	WBGene00000484	X		X		Structural	5, 7, 8, 12, 13, 20, 22, 26, 40, 42, 46, 51
che-3	WBGene00000485	X				Structural	3, 8, 13, 20, 22, 31, 46, 59
cil-1	WBGene00086546	X				Male	2
cil-7	WBGene00021017	X				Male	32
clhm-1	WBGene00016626	X				Subtype-specific	47
clic-1	WBGene00020246	X				Structural	18
cng-3	WBGene00000563	X				Subtype-specific	24, 53

Gene	Wormbase ID	C.C.	B.P.	B	C	Category	References
cwp-5	WBGene00009844	X				Male	34
daf-10	WBGene00000906	X				Structural	3, 6, 20, 41, 46
daf-11	WBGene00000907	X				Subtype-specific	6, 13
daf-25	WBGene00000917	X				Subtype-specific	43
daf-37	WBGene00016246	X				Subtype-specific	41
dcar-1	WBGene00007395	X				Subtype-specific	1
dli-1	WBGene00001007				X	Structural	5, 31,
dyci-1	WBGene00015927				X	Structural	31
dyf-1	WBGene00001117	X				Structural	5, 14, 20, 22, 31, 40, 55
dyf-11	WBGene00001127	X		X		Structural	5, 18, 20, 26, 30, 39, 40
dyf-13	WBGene00001129	X		X		Structural	5, 7, 20, 55
dyf-17	WBGene00012673	X				Expression	42
dyf-19	WBGene00012826	X	X			Structural	51
dyf-2	WBGene00001118	X		X		Structural	5, 7, 18, 20, 40, 51
dyf-3	WBGene00001119	X		X		Structural	7, 18, 20, 40
dyf-5	WBGene00001121	X		X		Structural	7, 18, 20, 40
dyf-6	WBGene00001122	X				Structural	14, 20, 40
dyf-7	WBGene00001123				X	Structural	18, 20, 40
dylt-2	WBGene00017014	X				Structural	5, 18,20
flp-17	WBGene00001460				X	Expression	25
gar-2	WBGene00001518				X	Subtype-specific	29
gasr-8	WBGene00008144			X		Expression	5, 7
gcy-1	WBGene00001528				X	Subtype-specific	4, 19, 39 ,52
gcy-12	WBGene00001538	X				Subtype-specific	4, 19, 52
gcy-14	WBGene00001540	X				Subtype-specific	4, 19, 39 ,52
gcy-15	WBGene00001541				X	Subtype-specific	19, 39
gcy-18	WBGene00001543	X				Subtype-specific	18, 19, 24, 37, 39
gcy-19	WBGene00001544				X	Subtype-specific	19, 39
gcy-2	WBGene00001529				X	Subtype-specific	4, 19, 39 ,52
gcy-20	WBGene00001545				X	Subtype-specific	19, 39
gcy-21	WBGene00001546				X	Subtype-specific	4, 19,39
gcy-22	WBGene00001547				X	Subtype-specific	4, 19
gcy-23	WBGene00001548	X				Subtype-specific	13, 19, 37, 39

Gene	Wormbase ID	C.C.	B.P.	B	C	Category	References
gcy-27	WBGene00001550				X	Subtype-specific	19, 39
gcy-29	WBGene00007314				X	Subtype-specific	19, 39
gcy-3	WBGene00001530				X	Subtype-specific	4, 19, 39 ,52
gcy-31	WBGene00001551				X	Subtype-specific	13, 52
gcy-33	WBGene00001553				X	Subtype-specific	13, 52
gcy-4	WBGene00001531				X	Subtype-specific	4, 19, 39 ,52
gcy-5	WBGene00001532				X	Subtype-specific	4, 8, 13, 18, 19, 21, 39, 46, 52
gcy-6	WBGene00001533				X	Subtype-specific	4, 8, 19, 26, 52
gcy-7	WBGene00001534				X	Subtype-specific	4, 8, 19, 52
gcy-8	WBGene00001535	X				Subtype-specific	19, 24, 33, 37, 46, 52
gcy-9	WBGene00001536	X				Subtype-specific	4, 19, 39 ,52
gpa-13	WBGene00001675	X				Structural	10, 18
gpa-3	WBGene00001665	X				Subtype-specific	21
hyls-1	WBGene00015466	X				Structural	11
ift-139	WBGene00022696	X		X		Structural	5, 7
ift-20	WBGene00022465	X		X		Structural	7, 18, 20, 40, 51,
ift-74	WBGene00016005	X				Structural	14, 20
ift-81	WBGene00017973	X				Structural	5, 14, 18, 20
ifta-1	WBGene00016935	X				Structural	5, 18, 20
ifta-2	WBGene00012132	X				Structural	20
jbts-14	WBGene00018727	X				Structural	26, 28
K07C11.10	WBGene00019482	X				Expression	26
kap-1	WBGene00002182	X				Structural	6, 20, 22, 38, 42, 51, 55
kin-10	WBGene00002196	X				Male	15
klp-11	WBGene00002222				X	Structural	14, 20
klp-20	WBGene00002230		X			Structural	18, 20
klp-6	WBGene00002218	X			X	Male	18, 20, 32
lov-1	WBGene00003058	X				Male	2, 3, 15, 20, 32, 34
mks-1	WBGene00020100	X	X	X		Structural	6, 7, 12, 20, 28, 39
mks-2	WBGene00194710	X	X			Structural	28, 30
mks-3	WBGene00018042	X		X		Structural	5, 7, 20, 28, 30, 39, 42, 51
mks-5	WBGene00007490	X				Structural	6, 28, 30, 37, 39, 51

Gene	Wormbase ID	C.C.	B.P.	B	C	Category	References
mks-6	WBGene00010642	X	X			Structural	5, 18, 28, 37, 39,51
mksr-1	WBGene00019364	X	X	X		Structural	5, 6, 7, 18, 28, 30, 39
mksr-2	WBGene00021416	X	X	X		Structural	5, 6, 7, 28, 30, 38, 39
nphp-1	WBGene00010898	X		X		Structural	7, 18, 26, 30, 37, 39, 51, 55
nphp-2	WBGene00021304	X	X			Structural	5, 6, 37
nphp-4	WBGene00011261	X		X		Structural	6, 7, 11, 26, 28, 30, 38, 39, 55
nsy-4	WBGene00021415	X				Subtype-specific	59
nubp-1	WBGene00008664	X				Structural	27
nud-1	WBGene00003829			X		Expression	7, 12
ocr-1	WBGene00003838				X	Subtype-specific	35, 48
ocr-2	WBGene00003839	X				Subtype-specific	1, 13, 20, 22, 48
ocr-4	WBGene00003841				X	Subtype-specific	35, 48
odr-1	WBGene00003848	X				Subtype-specific	3, 4, 8, 19, 24, 59, 52
odr-10	WBGene00003856	X				Subtype-specific	2, 3, 26, 42, 46
odr-3	WBGene00003850	X				Subtype-specific	1, 3, 8, 13, 18, 46, 59
odr-4	WBGene00003851		X	X		Structural	36
odr-8	WBGene00018160		X			Expression	8, 13, 15, 18, 22, 37
osm-1	WBGene00003883	X		X		Structural	3, 5, 7, 12, 18, 20, 46
osm-12	WBGene00003892	X		X		Structural	3, 5, 7, 12, 18, 20, 22, 37, 40, 55
osm-3	WBGene00003884	X				Structural	3, 5, 6, 8, 13, 14, 20, 22, 31, 38, 42, 46, 51, 55
osm-5	WBGene00003885	X		X		Structural	3, 5, 7, 10, 11, 12, 13, 20, 37, 39, 40, 42, 51, 55
osm-6	WBGene00003886	X		X		Structural	2, 3, 5, 7, 8, 11, 12, 13, 14, 20, 31, 38, 46, 59, 51, 55
osm-9	WBGene00003889	X				Subtype-specific	1, 2 ,9, 13, 14, 20, 22 42, 46, 48
osta-1	WBGene00015287	X				Structural	5, 6, 12
pef-1	WBGene00003969	X				Structural	18, 20
peli-1	WBGene00017766				X	Expression	6, 37
pitp-1	WBGene00010813	X				Subtype-specific	3, 6, 7, 12, 13, 46
pkd-2	WBGene00004035	X				Subtype-specific	2, 3, 15, 20, 32, 34, 42, 55

Gene	Wormbase ID	C.C.	B.P.	B	C	Category	References
rab-8	WBGene00004272				X	Structural	6, 38
rbg-3	WBGene00007167				X	Subtype-specific	16, 18
rpi-2	WBGene00019536	X				Structural	5, 28, 30, 39
srab-16	WBGene00020060	X				Subtype-specific	6, 18
srbc-64	WBGene00021477	X				Subtype-specific	6, 38
srbc-66	WBGene00020746	X				Subtype-specific	38
srd-23	WBGene00005101	X				Subtype-specific	6, 38
srg-2	WBGene00005160				X	Subtype-specific	12, 42
srg-36	WBGene00005193	X				Subtype-specific	33, 38
srg-37	WBGene00005194				X	Subtype-specific	33
srsx-3	WBGene00022408	X				Subtype-specific	6
sru-38	WBGene00005701	X				Subtype-specific	6, 18
str-1	WBGene00006069	X				Subtype-specific	6, 12, 38, 46
str-112	WBGene00006163				X	Subtype-specific	16
str-113	WBGene00006164				X	Subtype-specific	16
str-163	WBGene00006208	X				Subtype-specific	6, 18
str-2	WBGene00006070	X				Subtype-specific	8, 59
str-44	WBGene00006109	X				Subtype-specific	6, 12
sulp-2	WBGene00017464	X				Subtype-specific	45
tax-2	WBGene00006525	X				Subtype-specific	8, 13, 18, 22, 24, 38, 53
tax-4	WBGene00006526	X				Subtype-specific	4, 8, 13, 18, 19, 37, 53
tax-6	WBGene00006527	X				Structural	38
tba-5	WBGene00006531	X				Structural	14
tba-6	WBGene00006532				X	Structural	5
tbb-4	WBGene00006538	X				Structural	14
tmc-1	WBGene00020490	X				Subtype-specific	9
tmem-107	WBGene00043308	X	X			Structural	28
tmem-138	WBGene00008643	X				Structural	30
tmem-17	WBGene00022733	X				Structural	28, 30,42
tmem-218	WBGene00044771	X	X			Structural	30
tmem-231	WBGene00020825	X	X			Structural	28, 30, 42
tram-1	WBGene00007696	X				Subtype-specific	28, 30, 39
trp-4	WBGene00006616				X	Subtype-specific	35

Gene	Wormbase ID	C.C.	B.P.	B	C	Category	References
trpa-1	WBGene00007801	X				Subtype-specific	18, 35
tub-1	WBGene00006655	X		X		Structural	6, 7, 12, 20, 36, 46
xbx-1	WBGene00006960	X		X		Structural	7, 12, 20, 28, 30, 31, 37, 39, 51
xbx-3	WBGene00010864			X		Expression	7, 12
xbx-4	WBGene00016025			X		Expression	7, 12
xbx-5	WBGene00011974			X		Expression	7, 12
xbx-6	WBGene00009580			X		Expression	7, 12, 18
xbx-9	WBGene00007604			X		Expression	7

➤ **Annex 5. Schematic representing the location of sites matching PWMs generated from enriched oligonucleotides found in putative regulatory sequences of *bona fide* cilia genes of *C. elegans*.**

Black lines represent here putative promoter sequences of 700 base pairs in length from all coding isoforms of the 163 cilium-related genes used in this work. For every sequence, WormBase identifier, gene name and transcript name are shown. Coloured boxes represent the locations within the sequences in which a matching site for a PWM was found. Thickness of the boxes symbolized here the score of the matches, so sites that align comparatively better are represented thicker. Matches located in the positive strand of a sequence appear over the black lines whereas matches found in the negative strand appear below the lines. As established by protocol, the set of sequences was purged prior to the feeding of the overrepresentation algorithm, so the repeated sequences from different isoforms of the same gene do not bias the final result.

* For ease of reference, RFX-related motifs are depicted in purplish or bluish colours.
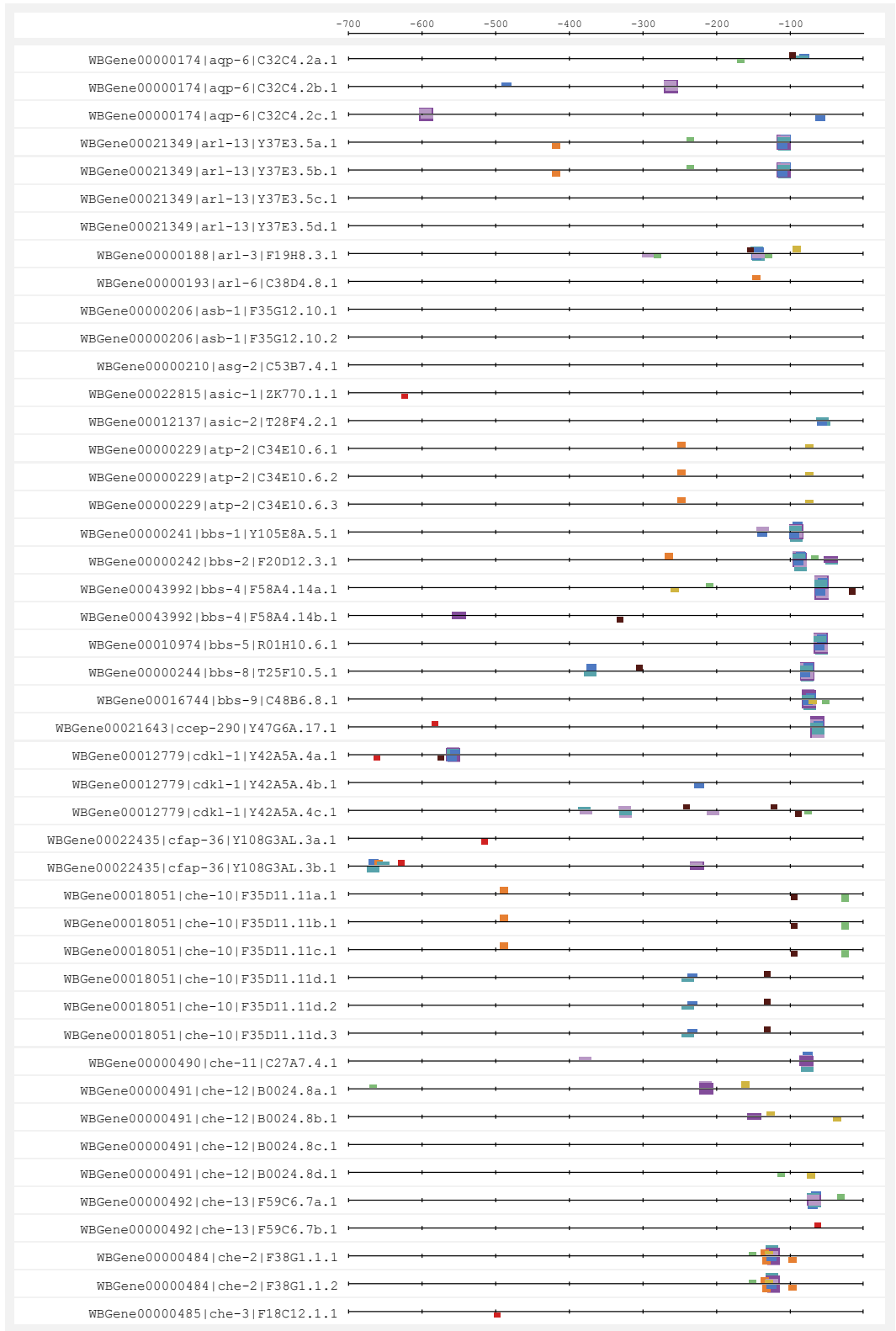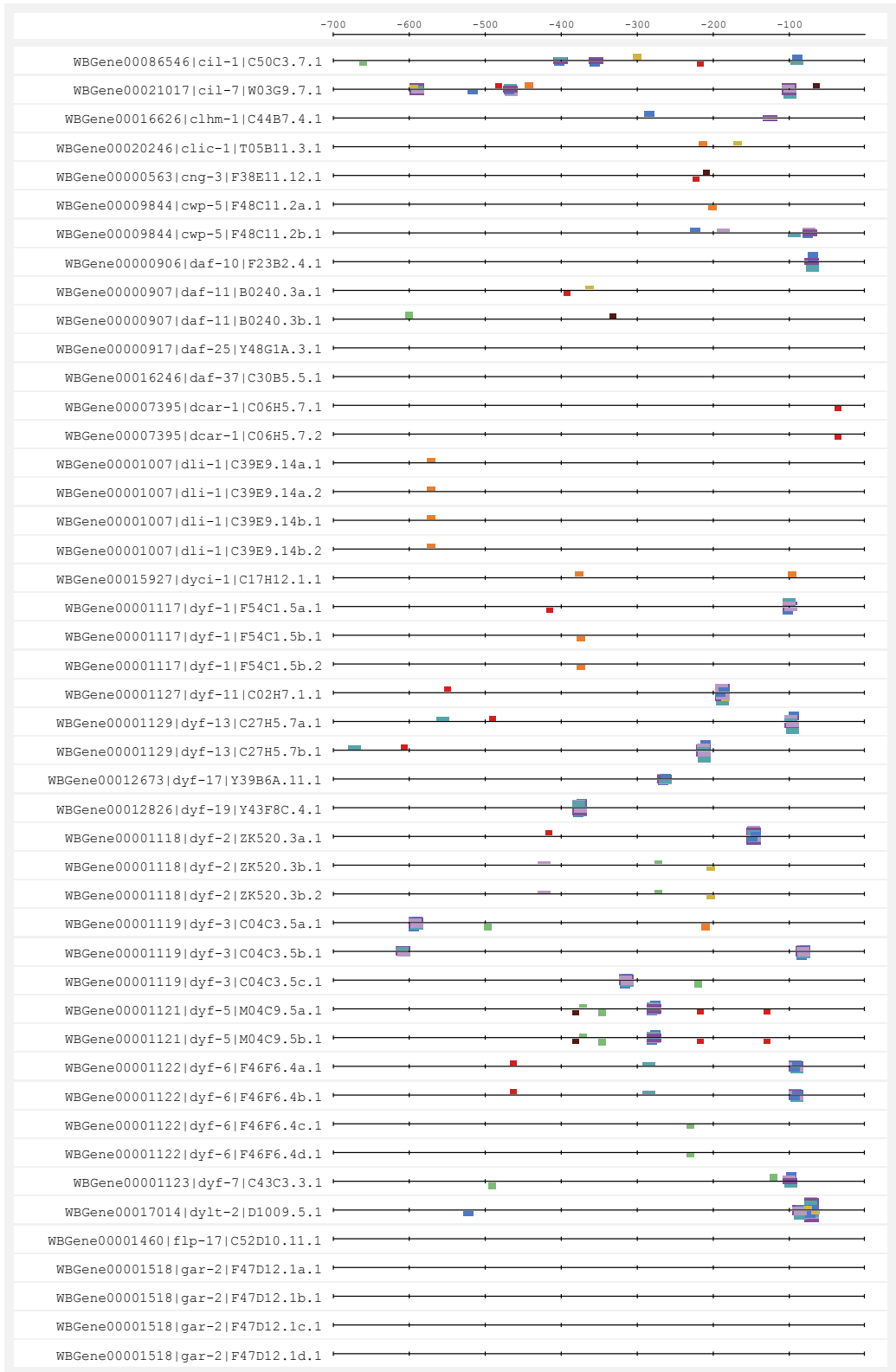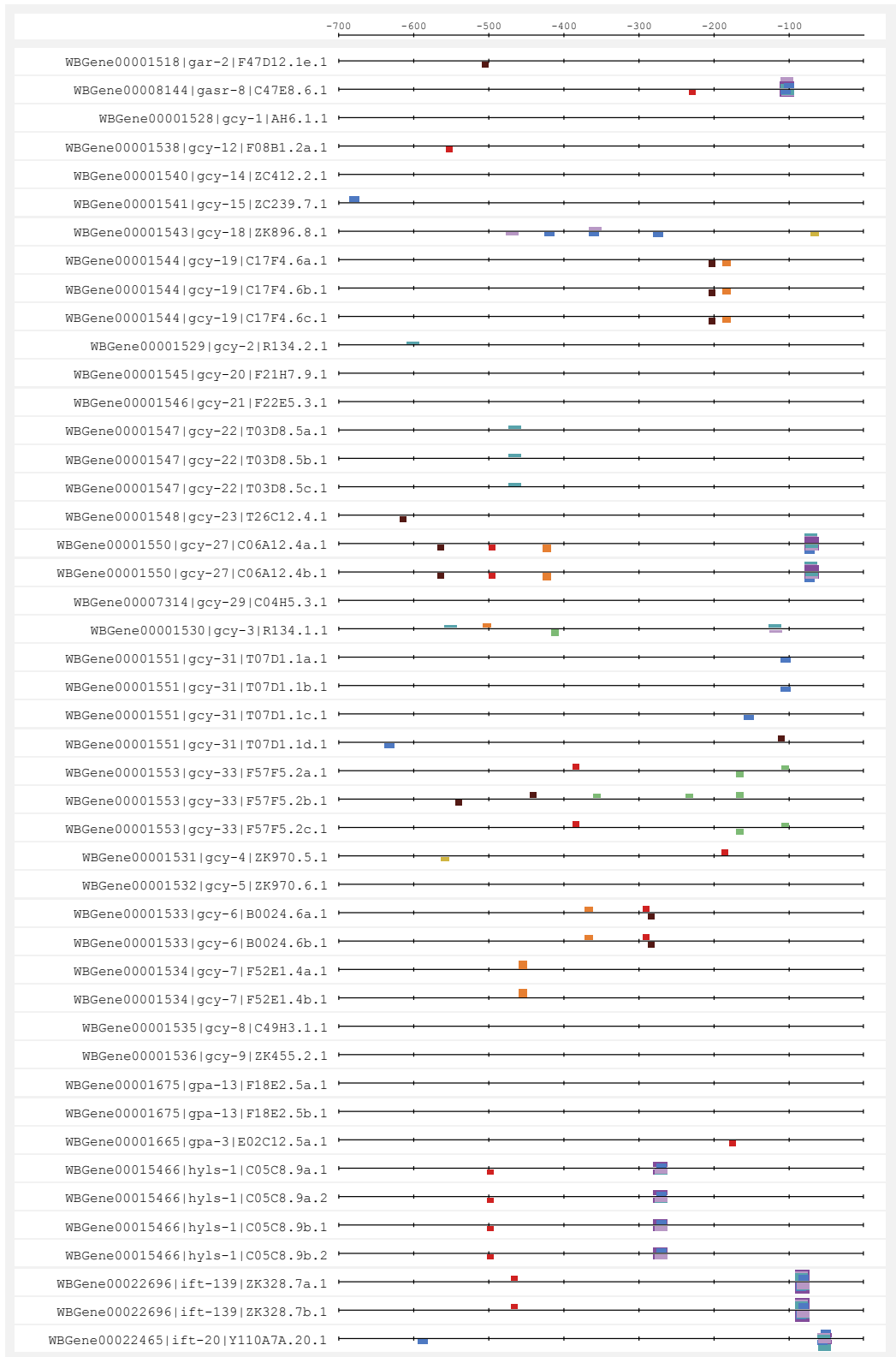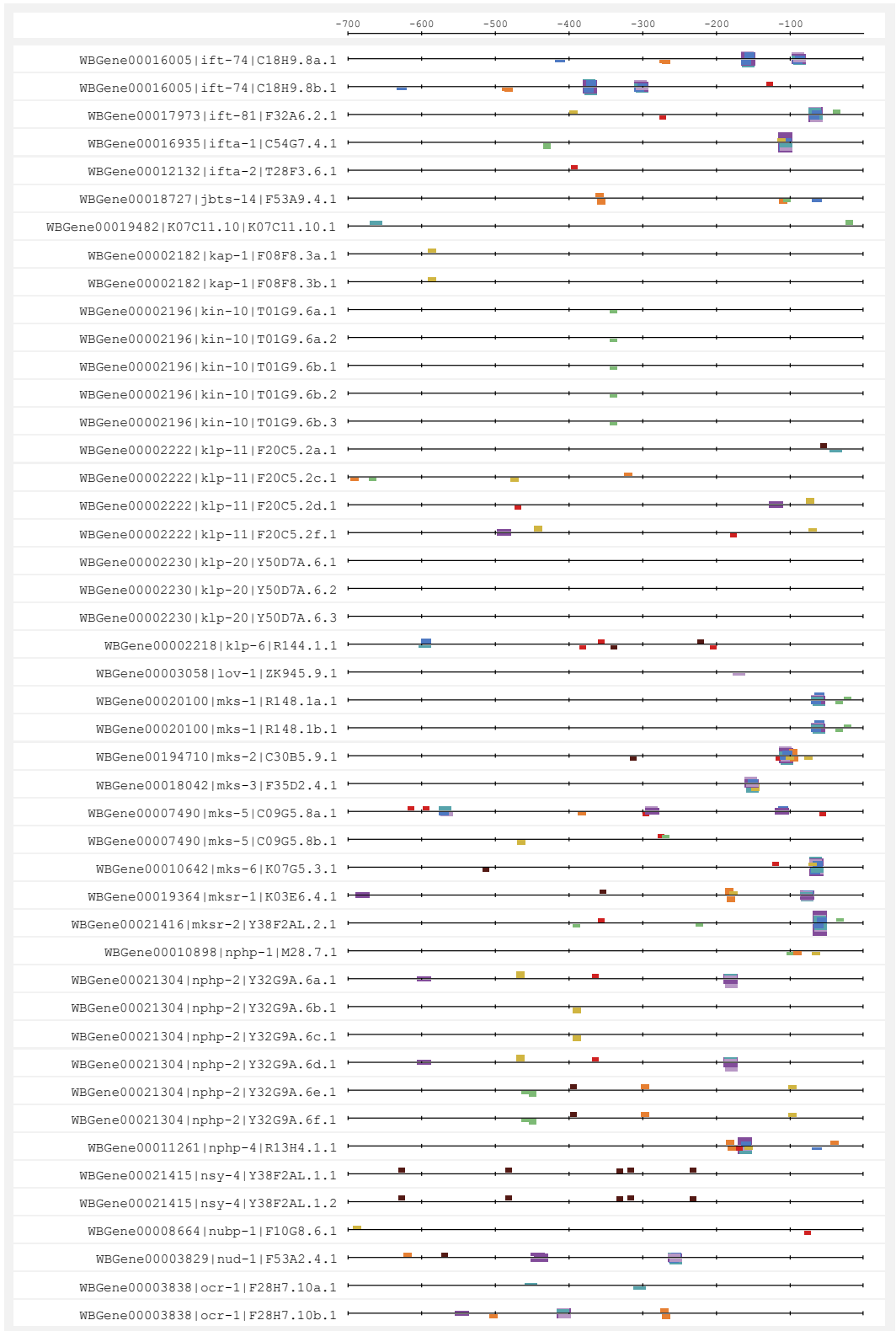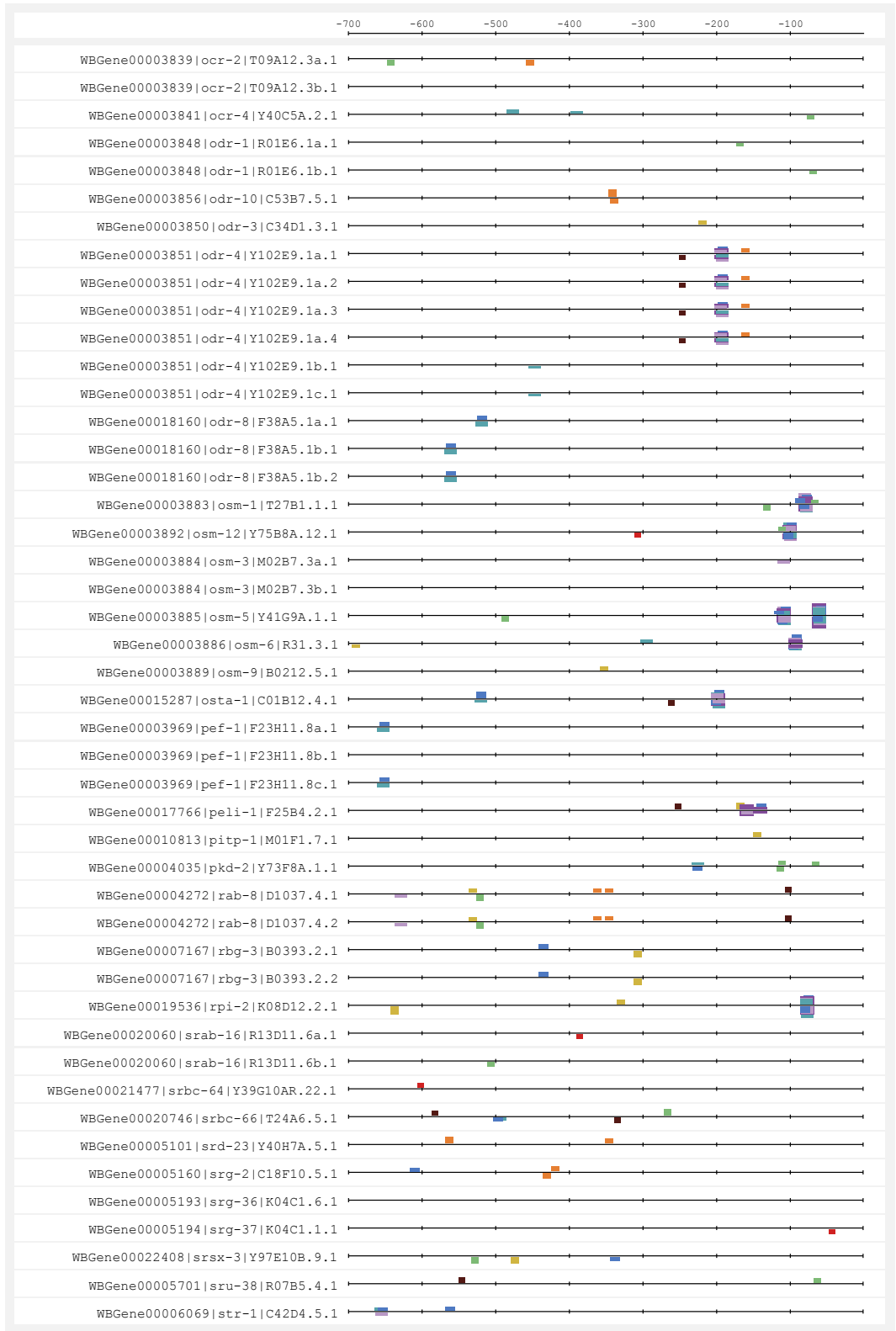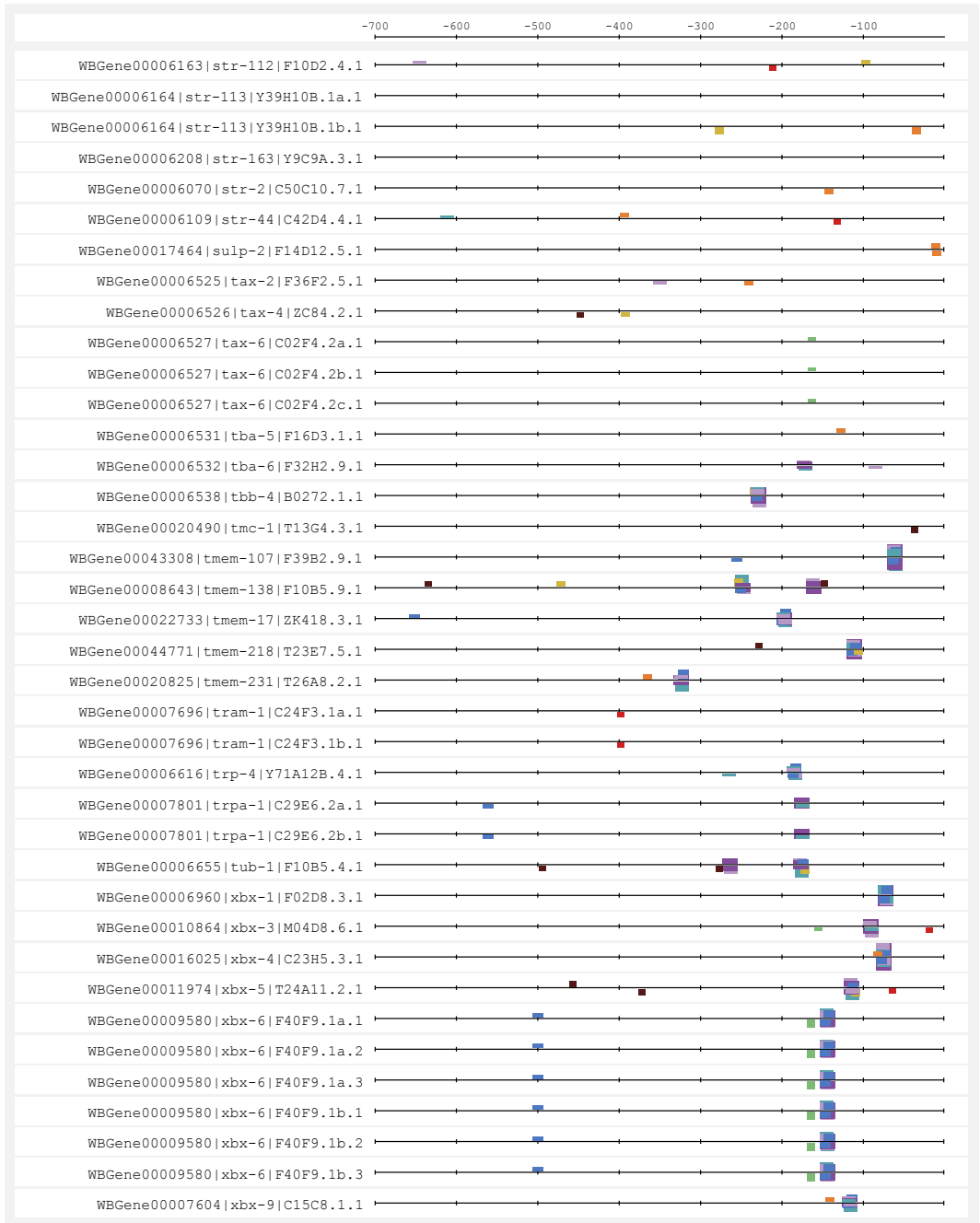
5 out of the 9 motifs generated in this analysis matched with experimentally validated binding sites of worm, mouse and human transcription factors. To reduce the number of false positives, three different statistical thresholds were used. p-value indicates the probability for a random motif of equivalent length to have a match score as good or better than the query motif. q-value implements a false discovery rate method correcting for the number of all the simultaneous significant matches for each query motif. E-value further corrects for multiple comparisons now accounting for all the motifs present in the target database being used. This progressively restrictive method for statistical significance reduced the total matches from the initial 73 that passed the first threshold for p-value to only 4 when comparing against the CIS-BP, and from 99 to 15 in the case of JASPAR CORE non-redundant 2018.

| Query	Target DB	Target ID	Gene	Organism	p-value	q-value	E-value
Motif 1	JASPAR	MA0600.2	RFX2	*H. sapiens*	3.63E-12	5.09E-09	5.10E-09
Motif 1	JASPAR	MA0510.2	RFX5	*H. sapiens*	2.05E-11	1.43E-08	2.87E-08
Motif 1	JASPAR	MA0798.1	RFX3	*H. sapiens*	3.99E-11	1.86E-08	5.60E-08
Motif 1	JASPAR	MA0799.1	RFX4	*H. sapiens*	1.31E-10	4.60E-08	1.84E-07
Motif 1	JASPAR	MA0509.1	Rfx1	*M. musculus*	5.08E-10	1.42E-07	7.13E-07
Motif 2	JASPAR	MA0600.2	RFX2	*H. sapiens*	1.01E-08	7.13E-06	1.42E-05
Motif 2	JASPAR	MA0798.1	RFX3	*H. sapiens*	1.26E-08	7.13E-06	1.77E-05
Motif 2	JASPAR	MA0509.1	Rfx1	*M. musculus*	1.27E-08	7.13E-06	1.78E-05
Motif 2	JASPAR	MA0799.1	RFX4	*H. sapiens*	3.25E-08	1.12E-05	4.56E-05
Motif 2	JASPAR	MA0510.2	RFX5	*H. sapiens*	3.59E-08	1.12E-05	5.04E-05
Motif 3	JASPAR	MA0509.1	Rfx1	*M. musculus*	1.75E-05	2.78E-02	2.45E-02
Motif 3	CIS-BP	M1534_1.02	*daf-19*	*C. elegans*	5.45E-06	2.93E-03	1.47E-03
Motif 4	JASPAR	MA0509.1	Rfx1	*M. musculus*	5.55E-08	1.56E-04	7.79E-05
Motif 4	JASPAR	MA0600.2	RFX2	*H. sapiens*	1.27E-05	5.88E-03	1.78E-02
Motif 4	JASPAR	MA0798.1	RFX3	*H. sapiens*	1.27E-05	5.88E-03	1.78E-02
Motif 4	JASPAR	MA0799.1	RFX4	*H. sapiens*	1.47E-05	5.88E-03	2.06E-02
Motif 4	JASPAR	MA0510.2	RFX5	*H. sapiens*	1.94E-05	6.06E-03	2.73E-02
Motif 4	CIS-BP	M1534_1.02	*daf-19*	*C. elegans*	7.11E-05	3.83E-02	1.91E-02
Motif 7	CIS-BP	M0228_1.02	*hlh-11*	*C. elegans*	3.29E-05	1.66E-02	8.86E-03
Motif 7	CIS-BP	M2264_1.02	*lin-32*	*C. elegans*	6.18E-05	1.66E-02	1.66E-02

➤ **Annex 7. Schematic representing the location of sites matching PWMs generated from enriched oligonucleotides found in putative regulatory sequences of *bona fide* cilia genes of *C. elegans* harbouring RFX-related sites.**

Black lines represent here putative promoter sequences of 700 base pairs in length from all coding isoforms of the 102 cilia-related genes in which RFX-related sites were found. For every sequence, WormBase identifier, gene name and transcript name are shown. Coloured boxes represent the locations within the sequences in which a matching site for a PWM was found. Thickness of the boxes symbolized here the score of the matches, so sites that align comparatively better are represented thicker. Matches located in the positive strand of a sequence appear over the black lines whereas matches found in the negative strand appear below the lines. As established by protocol, the set of sequences was purged prior to the feeding of the overrepresentation algorithm, so the repeated sequences from different isoforms of the same gene do not bias the final result.

■ **Assembly 1***
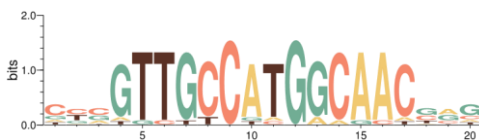
■ **Assembly 2***
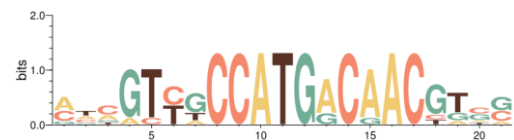
■ **Assembly 3***

■ **Assembly 4***
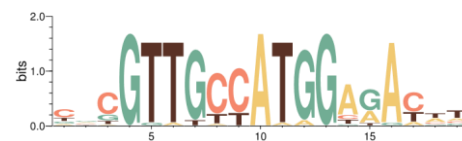
■ **Assembly 5***

■ **Assembly 6***
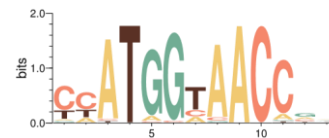
■ **Assembly 7**
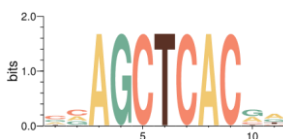
■ **Assembly 8**
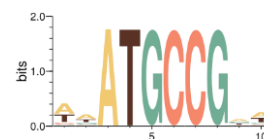
■ **Assembly 9**
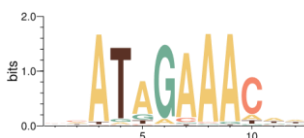
■ **Assembly 10**

* For ease of reference, RFX-related motifs are depicted in purplish or bluish colours.

	-700	-600	-500	-400	-300	-200	-100

WBGene00002222|klp-11|F20C5.2a.1
WBGene00002222|klp-11|F20C5.2c.1
WBGene00002222|klp-11|F20C5.2d.1
WBGene00002222|klp-11|F20C5.2f.1
WBGene00002218|klp-6|R144.1.1
WBGene00003058|lov-1|ZK945.9.1
WBGene00020100|mks-1|R148.1a.1
WBGene00020100|mks-1|R148.1b.1
WBGene00194710|mks-2|C30B5.9.1
WBGene00018042|mks-3|F35D2.4.1
WBGene00007490|mks-5|C09G5.8a.1
WBGene00007490|mks-5|C09G5.8b.1
WBGene00010642|mks-6|K07G5.3.1
WBGene00019364|mksr-1|K03E6.4.1
WBGene00021416|mksr-2|Y38F2AL.2.1
WBGene00021304|nphp-2|Y32G9A.6a.1
WBGene00021304|nphp-2|Y32G9A.6b.1
WBGene00021304|nphp-2|Y32G9A.6c.1
WBGene00021304|nphp-2|Y32G9A.6d.1
WBGene00021304|nphp-2|Y32G9A.6e.1
WBGene00021304|nphp-2|Y32G9A.6f.1
WBGene00011261|nphp-4|R13H4.1.1
WBGene00003829|nud-1|F53A2.4.1
WBGene00003838|ocr-1|F28H7.10a.1
WBGene00003838|ocr-1|F28H7.10b.1
WBGene00003841|ocr-4|Y40C5A.2.1
WBGene00003851|odr-4|Y102E9.1a.1
WBGene00003851|odr-4|Y102E9.1a.2
WBGene00003851|odr-4|Y102E9.1a.3
WBGene00003851|odr-4|Y102E9.1a.4
WBGene00003851|odr-4|Y102E9.1b.1
WBGene00003851|odr-4|Y102E9.1c.1
WBGene00018160|odr-8|F38A5.1a.1
WBGene00018160|odr-8|F38A5.1b.1
WBGene00018160|odr-8|F38A5.1b.2
WBGene00003883|osm-1|T27B1.1.1
WBGene00003892|osm-12|Y75B8A.12.1
WBGene00003884|osm-3|M02B7.3a.1
WBGene00003884|osm-3|M02B7.3b.1
WBGene00003885|osm-5|Y41G9A.1.1
WBGene00003886|osm-6|R31.3.1
WBGene00015287|osta-1|C01B12.4.1
WBGene00003969|pef-1|F23H11.8a.1
WBGene00003969|pef-1|F23H11.8b.1
WBGene00003969|pef-1|F23H11.8c.1
WBGene00017766|peli-1|F25B4.2.1

References

Aftab, Syed, Lucie Semenec, Jeffrey Shih Chieh Chu, and Nansheng Chen. 2008. "Identification and Characterization of Novel Human Tissue-Specific RFX Transcription Factors." *BMC Evolutionary Biology*.

Ahier, Arnaud and Sophie Jarriault. 2014. "Simultaneous Expression of Multiple Proteins under a Single Promoter in Caenorhabditis Elegans via a Versatile 2A-Based Toolkit." *Genetics* 196(3):605–13.

Akalin, Altuna, David Fredman, Erik Arner, Xianjun Dong, Jan Christian Bryne, Harukazu Suzuki, Carsten O. Daub, Yoshihide Hayashizaki, and Boris Lenhard. 2009. "Transcriptional Features of Genomic Regulatory Blocks." *Genome Biology*.

Allen, E. N., J. Ren, Y. Zhang, and J. Alcedo. 2015. "Sensory Systems: Their Impact on C. Elegans Survival." *Neuroscience* 296:15–25.

Alten, Leonie, Karin Schuster-Gossler, Anja Beckers, Stephanie Groos, Bärbel Ulmer, Jan Hegermann, Matthias Ochs, and Achim Gossler. 2012. "Differential Regulation of Node Formation, Nodal Ciliogenesis and Cilia Positioning by Noto and Foxj1." *Development*.

Altun, Z. F. and D. H. Hall. 2009. "Introduction." *WormAtlas*.

Altun, Z. F. and D. H. Hall. 2010. "Nervous System, Neuronal Support Cells." in *WormAtlas*.

Altun, ZF and DH Hall. 2011. "Handbook - Nervous System General Description." *Wormatlas.Org*.

Andersson, Robin and Albin Sandelin. 2020. "Determinants of Enhancer and Promoter Activities of Regulatory Elements." *Nature Reviews Genetics*.

Angeles-Albores, David, Raymond Lee, Juancarlos Chan, and Paul Sternberg. 2018. "Two New Functions in the WormBase Enrichment Suite." *MicroPublication Biology*.

Angeles-Albores, David, Raymond Y. Raymond, Juancarlos Chan, and Paul W. Sternberg. 2016. "Tissue Enrichment Analysis for C. Elegans Genomics." *BMC Bioinformatics* 17(1).

Aoki, Reina, Tatsurou Yagami, Hiroyuki Sasakura, Ken Ichi Ogura, Yasuhiro Kajihara, Masakazu Ibi, Takeaki Miyamae, Fumio Nakamura, Taro Asakura, Yoshikatsu Kanai, Yoshimi Misu, Yuichi Iino, Marina Ezcurra, William R. Schafer, Ikue Mori, and Yoshio Goshima. 2011. "A Seven-Transmembrane Receptor That Mediates Avoidance Response to Dihydrocaffeic Acid, a Water-Soluble Repellent in Caenorhabditis Elegans." *Journal of Neuroscience*.

Apfeld, Javier and Cynthia Kenyon. 1999. "Regulation of Lifespan by Sensory Perception in Caenorhabditis Elegans." *Nature* 402(6763):804–9.

Arda, H. Efsun, Stefan Taubert, Lesley T. MacNeil, Colin C. Conine, Ben Tsuda, Marc Van Gilst, Reynaldo Sequerra, Lynn Doucette-Stamm, Keith R. Yamamoto, and Albertha J. M. Walhout. 2010. "Functional Modularity of Nuclear Hormone Receptors in a Caenorhabditis Elegans Metabolic Gene Regulatory Network." *Molecular Systems Biology*.

Arnaiz, O., J. Cohen, A. M. Tassin, and F. Koll. 2015. "Remodeling Cildb, a Popular Database for Cilia and Links for Ciliopathies." *Cilia*.

Arnaiz, Olivier, Agata Malinowska, Catherine Klotz, Linda Sperling, Michal Dadlez, France Koll, and Jean Cohen. 2009. "Cildb: A Knowledgebase for Centrosomes and Cilia." *Database*.

Arribere, Joshua A., Ryan T. Bell, Becky X. H. Fu, Karen L. Artiles, Phil S. Hartman, and Andrew Z. Fire. 2014. "Efficient Marker-Free Recovery of Custom Genetic Modifications with CRISPR/Cas9 in Caenorhabditis Elegans." *Genetics* 198(3):837–46.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25(1):25–29.

Ashique, Amir M., Youngshik Choe, Mattias Karlen, Scott R. May, Khanhky Phamluong, Mark J. Solloway, Johan Ericson, and Andrew S. Peterson. 2009. "The Rfx4 Transcription Factor Modulates Shh Signaling by Regional Control of Ciliogenesis." *Science Signaling*.

Ashrafi, Kaveh, Francesca Y. Chang, Jennifer L. Watts, Andrew G. Fraser, Ravi S. Kamath, Julie Ahringer, and Gary Ruvkun. 2003. "Genome-Wide RNAi Analysis of Caenorhabditis Elegans Fat Regulatory Genes." *Nature* 421(6920):268–72.

Avidor-Reiss, Tomer, Andreia M. Maer, Edmund Koundakjian, Andrey Polyanovsky, Thomas Keil, Shankar Subramaniam, and Charles S. Zuker. 2004. "Decoding Cilia Function: Defining Specialized Genes Required for Compartmentalized Cilia Biogenesis." *Cell*.

Azimzadeh, Juliette and Michel Bornens. 2007. "Structure and Duplication of the Centrosome." *Journal of Cell Science*.

Bae, Young Kyung, Eunsoo Kim, Steven W. L'Hernault,

and Maureen M. Barr. 2009. "The CIL-1 PI 5-Phosphatase Localizes TRP Polycystins to Cilia and Activates Sperm in C. Elegans." *Current Biology*.

Bailey, Timothy L. 2008. "Discovering Sequence Motifs." *Methods in Molecular Biology*.

Bailey, Timothy L., James Johnson, Charles E. Grant, and William S. Noble. 2015. "The MEME Suite." *Nucleic Acids Research* 43(W1):W39–49.

Bailey, Timothy L. and Philip MacHanick. 2012. "Inferring Direct DNA Binding from ChIP-Seq." *Nucleic Acids Research*.

Bainbridge, Chance, Anjelica Rodriguez, Andrew Schuler, Michael Cisneros, and Andrés G. Vidal-Gadea. 2016. "Magnetic Orientation in C. Elegans Relies on the Integrity of the Villi of the AFD Magnetosensory Neurons." *Journal of Physiology Paris* 110(3):76–82.

Balleza, Enrique, J. Mark Kim, and Philippe Cluzel. 2018. "Systematic Characterization of Maturation Time of Fluorescent Proteins in Living Cells." *Nature Methods*.

Bargmann, C. I., J. H. Thomas, and H. R. Horvitz. 1990. "Chemosensory Cell Function in the Behavior and Development of Caenorhabditis Elegans." in *Cold Spring Harbor Symposia on Quantitative Biology*.

Bargmann, Cornelia I. 1998. "Neurobiology of the Caenorhabditis Elegans Genome." *Science*.

Bargmann, Cornelia I. 2006. "Chemosensation in C. Elegans. BT - Wormbook." in *Wormbook*.

Bargmann, Cornelia I., Erika Hartwieg, and H. Robert Horvitz. 1993. "Odorant-Selective Genes and Neurons Mediate Olfaction in C. Elegans." *Cell* 74(3):515–27.

Bargmann, Cornelia I. and H. Robert Horvitz. 1991. "Chemosensory Neurons with Overlapping Functions Direct Chemotaxis to Multiple Chemicals in C. Elegans." *Neuron*.

Barr, Maureen M. and L. Rene Garcia. 2006. "Male Mating Behavior." *WormBook : The Online Review of C. Elegans Biology* 1–11.

Barr, Maureen M. and Paul W. Sternberg. 1999. "A Polycystic Kidney-Disease Gene Homologue Required for Male Mating Behaviour in C. Elegans." *Nature*.

Barrios, Arantza, Rajarshi Ghosh, Chunhui Fang, Scott W. Emmons, and Maureen M. Barr. 2012. "PDF-1 Neuropeptide Signaling Modulates a Neural Circuit for Mate-Searching Behavior in C. Elegans." *Nature Neuroscience*.

Barstead, Robert, Gary Moulder, Beth Cobb, Stephen

Frazee, Diane Henthorn, Jeff Holmes, Daniela Jerebie, Martin Landsdale, Jamie Osborn, Cherilyn Pritchett, James Robertson, John Rummage, Ed Stokes, Malani Vishwanathan, Shohei Mitani, Keiko Gengyo-Ando, Osamu Funatsu, Sayaka Hori, Rieko Imae, Eriko Kage-Nakadai, Hiroyuki Kobuna, Etsuko Machiyama, Tomoko Motohashi, Muneyoshi Otori, Yuji Suehiro, Sawako Yoshina, Donald G. Moerman, Mark Edgley, Ryan Adair, BJ Allan, Vinci Au, Iasha Chaudhry, Rene Cheung, Owen Dadivas, Simon Eng, Lisa Fernando, Angela Fisher, Stephane Flibotte, Erin Gilchrist, Allison Hay, Peter Huang, Rebecca Worsley Hunt, Christine Kwitkowski, Joanne Lau, Norris Lee, Lucy Liu, Adam Lorch, Candy Luck, Jason Maydan, Sheldon McKay, Angela Miller, Greg Mullen, Candice Navaroli, Sarah Neil, Rebecca Hunt-Newbury, Mikhaela Partridge, Jaryn Perkins, Anna Rankin, Greta Raymant, Nadereh Rezania, Alexandra Rogula, Bin Shen, Greg Stegeman, Angela Tardif, Jon Taylor, Mariana Veiga, Tina Wang, and Rick Zapf. 2012. "Large-Scale Screening for Targeted Knockouts in the Caenorhabditis Elegans Genome." *G3: Genes, Genomes, Genetics* 2(11):1415–25.

Bell, Leslie R., Steven Stone, John Yochem, Jocelyn E. Shaw, and Robert K. Herman. 2006. "The Molecular Identities of the Caenorhabditis Elegans Intraflagellar Transport Genes Dyf-6, Daf-10 and Osm-1." *Genetics* 173(3):1275–86.

Bell, Oliver, Vijay K. Tiwari, Nicolas H. Thomä, and Dirk Schübeler. 2011. "Determinants and Dynamics of Genome Accessibility." *Nature Reviews Genetics*.

Benzinou, Michael, Andrew Walley, Stephan Lobbens, Marie Aline Charles, Béatrice Jouret, Frédéric Fumeron, Beverley Balkau, David Meyre, and Philippe Froguel. 2006. "Bardet-Biedl Syndrome Gene Variants Are Associated with Both Childhood and Adult Common Obesity in French Caucasians." *Diabetes*.

Berg, Otto G. 1988. "Selection of Dna Binding Sites by Regulatory Proteins. Functional Specificity and Pseudosite Competition." *Journal of Biomolecular Structure and Dynamics*.

Bertrand, Vincent and Oliver Hobert. 2009. "Linking Asymmetric Cell Division to the Terminal Differentiation Program of Postmitotic Neurons in C. Elegans." *Developmental Cell* 16(4):563–75.

Billi, Allison C., Sylvia E. J. Fischer, and John K. Kim. 2014. "Endogenous RNAi Pathways in C. Elegans." *WormBook : The Online Review of C. Elegans Biology*.

Birnbaum, Ramon Y., Rupali P. Patwardhan, Mee J. Kim, Gregory M. Findlay, Beth Martin, Jingjing Zhao, Robert J. A. Bell, Robin P. Smith, Angel A. Ku, Jay Shendure, and Nadav Ahituv. 2014. "Systematic Dissection of Coding Exons at Single Nucleotide

Resolution Supports an Additional Role in Cell-Specific Transcriptional Regulation." *PLoS Genetics*.

Birnby, Deborah A., Elizabeth Malone Link, Jennifer J. Vowels, Hong Tian, Patrick L. Colacurcio, and James H. Thomas. 2000. "A Transmembrane Guanylyl Cyclase (DAF-11) and Hsp90 (DAF-21) Regulate a Common Set of Chemosensory Behaviors in Caenorhabditis Elegans." *Genetics*.

Bisgrove, Brent W., Svetlana Makova, H. Joseph Yost, and Martina Brueckner. 2012. "RFX2 Is Essential in the Ciliated Organ of Asymmetry and an RFX2 Transgene Identifies a Population of Ciliated Cells Sufficient for Fluid Flow." *Developmental Biology*.

Blacque, Oliver E., Elliot A. Perens, Keith A. Boroevich, Peter N. Inglis, Chunmei Li, Adam Warner, Jaswinder Khattra, Rob A. Holt, Guangshuo Ou, Allan K. Mah, Sheldon J. McKay, Peter Huang, Peter Swoboda, Steve J. M. Jones, Marco A. Marra, David L. Baillie, Donald G. Moerman, Shai Shaham, and Michel R. Leroux. 2005a. "Functional Genomics of the Cilium, a Sensory Organelle." *Current Biology*.

Blacque, Oliver E., Elliot A. Perens, Keith A. Boroevich, Peter N. Inglis, Chunmei Li, Adam Warner, Jaswinder Khattra, Rob A. Holt, Guangshuo Ou, Allan K. Mah, Sheldon J. McKay, Peter Huang, Peter Swoboda, Steve J. M. Jones, Marco A. Marra, David L. Baillie, Donald G. Moerman, Shai Shaham, and Michel R. Leroux. 2005b. "Functional Genomics of the Cilium, a Sensory Organelle." *Current Biology* 15(10):935–41.

Bloodgood, Robert A. 1977. "Motility Occurring in Association with the Surface of the Chlamydomonas Flagellum." *Journal of Cell Biology*.

Bloodgood, Robert A. 2009. "From Central to Rudimentary to Primary: The History of an Underappreciated Organelle Whose Time Has Come. The Primary Cilium." *Methods in Cell Biology*.

Boehlke, Christopher, Fruzsina Kotsis, Vishal Patel, Simone Braeg, Henriette Voelker, Saskia Bredt, Theresa Beyer, Heike Janusch, Christoph Hamann, Markus Gödel, Klaus Müller, Martin Herbst, Miriam Hornung, Mara Doerken, Michael Köttgen, Roland Nitschke, Peter Igarashi, Gerd Walz, and E. Wolfgang Kuehn. 2010. "Primary Cilia Regulate MTORC1 Activity and Cell Size through Lkb1." *Nature Cell Biology* 12(11):1115–22.

Bonnafe, E., M. Touka, A. AitLounis, D. Baas, E. Barras, C. Ucla, A. Moreau, F. Flamant, R. Dubruille, P. Couble, J. Collignon, B. Durand, and W. Reith. 2004. "The Transcription Factor RFX3 Directs Nodal Cilium Development and Left-Right

Asymmetry Specification." *Molecular and Cellular Biology*.

Boucher C. 2011. "Normal Approximation to a Poisson Random Variable." *Wolfram Demonstrations Project.*

Bowler, Mathew, Dong Kong, Shufeng Sun, Rashmi Nanjundappa, Lauren Evans, Veronica Farmer, Andrew Holland, Moe R. Mahjoub, Haixin Sui, and Jadranka Loncarek. 2019. "High-Resolution Characterization of Centriole Distal Appendage Morphology and Dynamics by Correlative STORM and Electron Microscopy." *Nature Communications*.

Brear, Andrea G., Jason Yoon, Martin Wojtyniak, and Piali Sengupta. 2014. "Diverse Cell Type-Specific Mechanisms Localize G Protein-Coupled Receptors to Caenorhabditis Elegans Sensory Cilia." *Genetics* 197(2):667–84.

Brenner, S. 1974. "The Genetics of *Caenorabditis Elegans*." *Genetics* 77(May):71–94.

Brenner, S. 1974. "The Genetics of Caenorhabditis Elegans." *Genetics*.

Breslow, David K. and Andrew J. Holland. 2019. "Mechanism and Regulation of Centriole and Cilium Biogenesis." *Annual Review of Biochemistry*.

Bretscher, Andrew Jonathan, Karl Emanuel Busch, and Mario De Bono. 2008. "A Carbon Dioxide Avoidance Behavior Is Integrated with Responses to Ambient Oxygen and Food in Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences of the United States of America* 105(23):8044–49.

Bretscher, Andrew Jonathan, Eiji Kodama-Namba, Karl Emanuel Busch, Robin Joseph Murphy, Zoltan Soltesz, Patrick Laurent, and Mario de Bono. 2011. "Temperature, Oxygen, and Salt-Sensing Neurons in C. Elegans Are Carbon Dioxide Sensors That Control Avoidance Behavior." *Neuron*.

Brockie, Penelope J., Jerry E. Mellem, Thomas Hills, David M. Madsen, and Andres V. Maricq. 2001. "The C. Elegans Glutamate Receptor Subunit NMR-1 Is Required for Slow NMDA-Activated Currents That Regulate Reversal Frequency during Locomotion." *Neuron* 31(4):617–30.

Brody, Steven L., Xiu Hua Yan, Mary K. Wuerffel, Sheng Kwei Song, and Steven D. Shapiro. 2000. "Ciliogenesis and Left-Right Axis Defects in Forkhead Factor HFH-4-Null Mice." *American Journal of Respiratory Cell and Molecular Biology*.

Burghoorn, Jan, Brian P. Piasecki, Filip Crona, Prasad Phirke, Kristian E. Jeppsson, and Peter Swoboda. 2012. "The in Vivo Dissection of Direct RFX-

Target Gene Promoters in C. Elegans Reveals a Novel Cis-Regulatory Element, the C-Box." *Developmental Biology* 368(2):415–26.

Byerly, L., R. C. Cassada, and R. L. Russell. 1976. "The Life Cycle of the Nematode Caenorhabditis Elegans." *Developmental Biology*.

C. elegans sequencing consortium. 1998. "Genome Sequence of the Nematode C. Elegans: A Platform for Investigating Biology. The C. Elegans Sequencing Consortium [Published Erratum Appears in Science 1999 Jan 1;283(5398):35]." *Science*.

Cachero, Sebastián, T. Ian Simpson, Petra I. zur Lage, Lina Ma, Fay G. Newton, Eimear E. Holohan, J. Douglas Armstrong, and Andrew P. Jarman. 2011. "The Gene Regulatory Cascade Linking Proneural Specification with Differentiation in Drosophila Sensory Neurons." *PLoS Biology*.

Cao, Junyue, Jonathan S. Packer, Vijay Ramani, Darren A. Cusanovich, Chau Huynh, Riza Daza, Xiaojie Qiu, Choli Lee, Scott N. Furlan, Frank J. Steemers, Andrew Adey, Robert H. Waterston, Cole Trapnell, and Jay Shendure. 2017. "Comprehensive Single-Cell Transcriptional Profiling of a Multicellular Organism." *Science* 357(6352):661–67.

Cao, Xiou and Alejandro Aballay. 2016. "Neural Inhibition of Dopaminergic Signaling Enhances Immunity in a Cell-Non-Autonomous Manner." *Current Biology*.

Cao, Xiou, Rie Kajino-Sakamoto, Argenia Doss, and Alejandro Aballay. 2017. "Distinct Roles of Sensory Neurons in Mediating Pathogen Avoidance and Neuropeptide-Dependent Immune Regulation." *Cell Reports*.

Carbon, S., E. Douglass, N. Dunn, B. Good, N. L. Harris, S. E. Lewis, C. J. Mungall, S. Basu, R. L. Chisholm, R. J. Dodson, E. Hartline, P. Fey, P. D. Thomas, L. P. Albou, D. Ebert, M. J. Kesling, H. Mi, A. Muruganujan, X. Huang, S. Poudel, T. Mushayahama, J. C. Hu, S. A. LaBonte, D. A. Siegele, G. Antonazzo, H. Attrill, N. H. Brown, S. Fexova, P. Garapati, T. E. M. Jones, S. J. Marygold, G. H. Millburn, A. J. Rey, V. Trovisco, G. Dos Santos, D. B. Emmert, K. Falls, P. Zhou, J. L. Goodman, V. B. Strelets, J. Thurmond, M. Courtot, D. S. Osumi, H. Parkinson, P. Roncaglia, M. L. Acencio, M. Kuiper, A. Lreid, C. Logie, R. C. Lovering, R. P. Huntley, P. Denny, N. H. Campbell, B. Kramarz, V. Acquaah, S. H. Ahmad, H. Chen, J. H. Rawson, M. C. Chibucos, M. Giglio, S. Nadendla, R. Tauber, M. J. Duesbury, N. T. Del, B. H. M. Meldal, L. Perfetto, P. Porras, S. Orchard, A. Shrivastava, Z. Xie, H. Y. Chang, R. D. Finn, A. L. Mitchell, N. D. Rawlings, L. Richardson, A. Sangrador-Vegas, J. A. Blake, K. R. Christie, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, M. A. Harris, S. G. Oliver, K. Rutherford, V. Wood, J. Hayles, J. Bahler, A. Lock, E. R. Bolton, J. De Pons, M. Dwinell, G. T. Hayman, S. J. F. Lauderkind, M. Shimoyama, M. Tutaj, S. J. Wang, P. D'Eustachio, L. Matthews, J. P. Balhoff, S. A. Aleksander, G. Binkley, B. L. Dunn, J. M. Cherry, S. R. Engel, F. Gondwe, K. Karra, K. A. MacPherson, S. R. Miyasato, R. S. Nash, P. C. Ng, T. K. Sheppard, A. Shrivatsav Vp, M. Simison, M. S. Skrzypek, S. Weng, E. D. Wong, M. Feuermann, P. Gaudet, E. Bakker, T. Z. Berardini, L. Reiser, S. Subramaniam, E. Huala, C. Arighi, A. Auchincloss, K. Axelsen, G. P. Argoud, A. Bateman, B. Bely, M. C. Blatter, E. Boutet, L. Breuza, A. Bridge, R. Britto, H. Bye-A-Jee, C. Casals-Casas, E. Coudert, A. Estreicher, L. Famiglietti, P. Garmiri, G. Georghiou, A. Gos, N. Gruaz-Gumowski, E. Hatton-Ellis, U. Hinz, C. Hulo, A. Ignatchenko, F. Jungo, G. Keller, K. Laiho, P. Lemercier, D. Lieberherr, Y. Lussi, A. Mac-Dougall, M. Magrane, M. J. Martin, P. Masson, D. A. Natale, N. N. Hyka, I. Pedruzzi, K. Pichler, S. Poux, C. Rivoire, M. Rodriguez-Lopez, T. Sawford, E. Speretta, A. Shypitsyna, A. Stutz, S. Sundaram, M. Tognolli, N. Tyagi, K. Warner, R. Zaru, C. Wu, J. Chan, J. Cho, S. Gao, C. Grove, M. C. Harrison, K. Howe, R. Lee, J. Mendel, H. M. Muller, D. Raciti, K. Van Auken, M. Berriman, L. Stein, P. W. Sternberg, D. Howe, S. Toro, and M. Westerfield. 2019. "The Gene Ontology Resource: 20 Years and Still GOing Strong." *Nucleic Acids Research* 47(D1):D330–38.

Caron, Barthélémy, Yufei Luo, and Antonio Rausell. 2019. "NCBoost Classifies Pathogenic Non-Coding Variants in Mendelian Diseases through Supervised Learning on Purifying Selection Signals in Humans." *Genome Biology*.

Chalfie, M., J. E. Sulston, J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. 1985. "The Neural Circuit for Touch Sensitivity in Caenorhabditis Elegans." *Journal of Neuroscience* 5(4):956–64.

Chalfie, Martin, Anne C. Hart, Catharine H. Rankin, and Miriam B. Goodman. 2014. "Assaying Mechanosensation." *WormBook : The Online Review of C. Elegans Biology*.

Chalfie, Martin and John Sulston. 1981. "Developmental Genetics of the Mechanosensory Neurons of Caenorhabditis Elegans." *Developmental Biology*.

Chang, Andy J., Nikolas Chronis, David S. Karow, Michael A. Marletta, and Cornelia I. Bargmann. 2006. "A Distributed Chemosensory Circuit for Oxygen Preference in C. Elegans." *PLoS Biology*.

Chang, Sarah, Robert J. Johnston, and Oliver Hobert. 2003. "A Transcriptional Regulatory Cascade That Controls Left/Right Asymmetry in Chemosensory Neurons of c. Elegans." *Genes and Development* 17(17):2123–37.

Chatzigeorgiou, Marios, Sangsu Bang, Sun Wook Hwang, and William R. Schafer. 2013. "Tmc-1 Encodes a Sodium-Sensitive Channel Required for Salt

Chemosensation in C. Elegans." *Nature*.

Chatzigeorgiou, Marios, Sungjae Yoo, Joseph D. Watson, Wei Hsiang Lee, W. Clay Spencer, Katie S. Kindt, Sun Wook Hwang, David M. Miller, Millet Treinin, Monica Driscoll, and William R. Schafer. 2010. "Specific Roles for DEG/ENaC and TRP Channels in Touch and Thermosensation in C. Elegans Nociceptors." *Nature Neuroscience* 13(7):861–68.

Chen, Changchun, Eisuke Itakura, Katherine P. Weber, Ramanujan S. Hegde, and Mario de Bono. 2014. "An ER Complex of ODR-4 and ODR-8/Ufm1 Specific Protease 2 Promotes GPCR Maturation by a Ufm1-Independent Mechanism." *PLoS Genetics*.

Chen, Edward Y., Christopher M. Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela V. Meirelles, Neil R. Clark, and Avi Ma'ayan. 2013. "Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool." *BMC Bioinformatics* 14.

Chen, Jianchun, Heather J. Knowles, Jennifer L. Hebert, and Brian P. Hackett. 1998. "Mutation of the Mouse Hepatocyte Nuclear Factor/Forkhead Homologue 4 Gene Results in an Absence of Cilia and Random Left-Right Asymmetry." *Journal of Clinical Investigation*.

Chen, Nansheng, Allan Mah, Oliver E. Blacque, Jeffrey Chu, Kiran Phgora, Mathieu W. Bakhoum, C. Rebecca Hunt Newburry, Jaswinder Khattra, Susanna Chan, Anne Go, Evgeni Efimenko, Robert Johnsen, Prasad Phirke, Peter Swoboda, Marco Marra, Donald G. Moerman, Michel R. Leroux, David L. Baillie, and Lincoln D. Stein. 2006. "Identification of Ciliary and Ciliopathy Genes in Caenorhabditis Elegans through Comparative Genomics." *Genome Biology.*

Chen, Ting Wen, Hsin Pai Li, Chi Ching Lee, Ruei Chi Gan, Po Jung Huang, Timothy H. Wu, Cheng Yang Lee, Yi Feng Chang, and Petrus Tang. 2014. "ChIPseek, a Web-Based Analysis Tool for ChIP Data." *BMC Genomics* 15(1).

Cheung, Benny H. H., Merav Cohen, Candida Rogers, Onder Albayram, and Mario De Bono. 2005. "Experience-Dependent Modulation of C. Elegans Behavior by Ambient Oxygen." *Current Biology*.

Cho, Yongmin, Daniel A. Porto, Hyundoo Hwang, Laura J. Grundy, William R. Schafer, and Hang Lu. 2017. "Automated and Controlled Mechanical Stimulation and Functional Imaging: In Vivo in C. Elegans." *Lab on a Chip*.

Choksi, Semil P., Deepak Babu, Doreen Lau, Xianwen Yu, and Sudipto Roy. 2014. "Systematic Discovery of Novel Ciliary Genes through Functional Genomics in the Zebrafish." *Development (Cambridge)*.

Choksi, Semil P., Gilbert Lauter, Peter Swoboda, and Sudipto Roy. 2014. "Switching on Cilia: Transcriptional Networks Regulating Ciliogenesis." *Development (Cambridge)*.

Chrisman, Steven D., Christopher B. Waite, Alison G. Scoville, and Lucinda Carnell. 2016. "C. Elegans Demonstrates Distinct Behaviors within a Fixed and Uniform Electric Field." *PLoS ONE* 11(3).

Chu, Jeffery S. C., Maja Tarailo-Graovac, Di Zhang, Jun Wang, Bora Uyar, Domena Tu, Joanne Trinh, David L. Baillie, and Nansheng Chen. 2012. "Fine Tuning of RFX/DAF-19-Regulated Target Gene Expression through Binding to Multiple Sites in Caenorhabditis Elegans." *Nucleic Acids Research*.

Chu, Jeffrey Sc, David L. Baillie, and Nansheng Chen. 2010. "Convergent Evolution of RFX Transcription Factors and Ciliary Genes Predated the Origin of Metazoans." *BMC Evolutionary Biology*.

Chung, Mei I., Sara M. Peyrot, Sarah LeBoeuf, Tae Joo Park, Kriston L. McGary, Edward M. Marcotte, and John B. Wallingford. 2012. "RFX2 Is Broadly Required for Ciliogenesis during Vertebrate Development." *Developmental Biology*.

Clark, Kirk L., Elaine D. Halay, Eseng Lai, and Stephen K. Burley. 1993. "Co-Crystal Structure of the HNF-3/Fork Head DNA-Recognition Motif Resembles Histone H5." *Nature*.

Coates, Juliet C. and Mario De Bono. 2002. "Antagonistic Pathways in Neurons Exposed to Body Fluid Regulate Social Feeding in Caenorhabditis Elegans." *Nature*.

Collet, Joan, Caroline A. Spike, Erik A. Lundquist, Jocelyn E. Shaw, and Robert K. Herman. 1998. "Analysis of Osm-6, a Gene That Affects Sensory Cilium Structure and Sensory Neuron Function in Caenorhabditis Elegans." *Genetics*.

Collins, Kevin M., Addys Bode, Robert W. Fernandez, Jessica E. Tanis, Jacob C. Brewer, Matthew S. Creamer, and Michael R. Koelle. 2016. "Activity of the C. Elegans Egg-Laying Behavior Circuit Is Controlled by Competing Activation and Feedback Inhibition." *ELife*.

Cornaglia, Matteo, Laurent Mouchiroud, Alexis Marette, Shreya Narasimhan, Thomas Lehnert, Virginija Jovaisaite, Johan Auwerx, and Martin A. M. Gijs. 2015. "An Automated Microfluidic Platform for C. Elegans Embryo Arraying, Phenotyping, and Long-Term Live Imaging." *Scientific Reports*.

Couillault, Carole, Nathalie Pujol, Jérôme Reboul, Laurence Sabatier, Jean François Guichou, Yuji Kohara, and Jonathan J. Ewbank. 2004. "TLR-Independent Control of Innate Immunity in Caenorhabditis Elegans by the TIR Domain

Adaptor Protein TIR-1, an Ortholog of Human SARM." *Nature Immunology*.

Cruz, Catarina, Vanessa Ribes, Eva Kutejova, Jordi Cayuso, Victoria Lawson, Dominic Norris, Jonathan Stevens, Megan Davey, Ken Blight, Fiona Bangs, Anita Mynett, Elizabeth Hirst, Rachel Chung, Nikolaos Balaskas, Steven L. Brody, Elisa Marti, and James Briscoe. 2010. "Foxj1 Regulates Floor Plate Cilia Architecture and Modifies the Response of Cells to Sonic Hedgehog Signalling." *Development*.

Van Dam, Teunis J. P., Julie Kennedy, Robin van der Lee, Erik de Vrieze, Kirsten A. Wunderlich, Suzanne Rix, Gerard W. Dougherty, Nils J. Lambacher, Chunmei Li, Victor L. Jensen, Michel R. Leroux, Rim Hjeij, Nicola Horn, Yves Texier, Yasmin Wissinger, Jeroen Van Reeuwijk, Gabrielle Wheway, Barbara Knapp, Jan F. Scheel, Brunella Franco, Dorus A. Mans, Erwin Van Wijk, François Képès, Gisela G. Slaats, Grischa Toedt, Hannie Kremer, Heymut Omran, Katarzyna Szymanska, Konstantinos Koutroumpas, Marius Ueffing, Thanh Minh T. Nguyen, Stef J. F. Letteboer, Machteld M. Oud, Sylvia E. C. Van Beersum, Miriam Schmidts, Philip L. Beales, Qianhao Lu, Rachel H. Giles, Radek Szklarczyk, Robert B. Russell, Toby J. Gibson, Colin A. Johnson, Oliver E. Blacque, Uwe Wolfrum, Karsten Boldt, Ronald Roepman, Victor Hernandez-Hernandez, and Martijn A. Huynen. 2019. "Ciliacarta: An Integrated and Validated Compendium of Ciliary Genes." *PLoS ONE*.

Van Dam, Teunis J. P., Gabrielle Wheway, Gisela G. Slaats, Martijn A. Huynen, and Rachel H. Giles. 2013. "The SYSCILIA Gold Standard (SCGSv1) of Known Ciliary Components and Its Applications within a Systems Biology Consortium." *Cilia*.

Dammermann, Alexander, Hayley Pemble, Brian J. Mitchell, Ian McLeod, John R. Yates, Chris Kintner, Arshad B. Desai, and Karen Oegema. 2009. "The Hydrolethalus Syndrome Protein HYLS-1 Links Core Centriole Structure to Cilia Formation." *Genes and Development*.

Davidson, Eric H. 2010. "Emerging Properties of Animal Gene Regulatory Networks." *Nature*.

Defrance, Matthieu, Rekin's Janky, Olivier Sand, and Jacques van Helden. 2008. "Using RSAT Oligo-Analysis and Dyad-Analysis Tools to Discover Regulatory Signals in Nucleic Sequences." *Nature Protocols*.

Deneris, Evan S. and Oliver Hobert. 2014. "Maintenance of Postmitotic Neuronal Cell Identity." *Nature Neuroscience*.

Didon, Lukas, Rachel K. Zwick, Ion Wa Chao, Matthew S. Walters, Rui Wang, Neil R. Hackett, and Ronald G. Crystal. 2013. "RFX3 Modulation of FOXJ1 Regulation of Cilia Genes in the Human Airway Epithelium." *Respiratory Research*.

Doroquez, David B., Cristina Berciu, James R. Anderson, Piali Sengupta, and Daniela Nicastro. 2014. "A High-Resolution Morphological and Ultrastructural Map of Anterior Sensory Cilia and Glia in Caenorhabditis Elegans." *ELife*.

Dubruille, Raphaelle, Anne Laurençon, Camille Vandaele, Emiko Shishido, Madeleine Coulon-Bublex, Peter Swoboda, Pierre Couble, Maurice Kernan, and Bénédicte Durand. 2002. "Drosophila Regulatory Factor X Is Necessary for Ciliated Sensory Neuron Differentiation." *Development*.

Dusenbery, David B., Robert E. Sheridan, and Richard L. Russell. 1975. "CHEMOTAXIS-DEFECTIVE MUTANTS OF THE NEMATODE <Em>CAENORHABDITIS ELEGANS</Em>" *Genetics* 80(2):297 LP – 309.

Dwyer, Noelle D., Emily R. Troemel, Piali Sengupta, and Cornelia I. Bargmann. 1998. "Odorant Receptor Localization to Olfactory Cilia Is Mediated by ODR-4, a Novel Membrane-Associated Protein." *Cell*.

Edwards, Stacey L., Nicole K. Charlie, Marie C. Milfort, Brandon S. Brown, Christen N. Gravlin, Jamie E. Knecht, and Kenneth G. Miller. 2008. "A Novel Molecular Solution for Ultraviolet Light Detection in Caenorhabditis Elegans." *PLoS Biology*.

Efimenko, Evgeni, Kerry Bubb, Ho Yi Mark, Ted Holzman, Michel R. Leroux, Gary Ruvkun, James H. Thomas, and Peter Swoboda. 2005. "Analysis of Xbx Genes in C. Elegans." *Development* 132(8):1923–34.

Eisenmann, David M. 2005. "Wnt Signaling." *WormBook : The Online Review of C. Elegans Biology*.

Ellingford, Jamie, Glenda Beaman, Kevin Webb, Christopher O'Callaghan, Robert Hirst, Graeme CM Black, and William Newman. 2018. "Whole Genome Sequencing Enables Definitive Diagnosis of Cystic Fibrosis and Primary Ciliary Dyskinesia." *BioRxiv*.

Emery, P, M. Strubin, K. Hofmann, P. Bucher, B. Mach, and W. Reith. 1996. "A Consensus Motif in the RFX DNA Binding Domain and Binding Domain Mutants with Altered Specificity." *Molecular and Cellular Biology*.

Emery, Patrick, Bénédicte Durand, Bernard Mach, and Walter Reith. 1996. "RFX Proteins, a Novel Family of DNA Binding Proteins Conserved in the Eukaryotic Kingdom." *Nucleic Acids Research*.

Ermolaeva, Maria A. and Björn Schumacher. 2014. "Insights from the Worm: The C. Elegans Model for Innate Immunity." *Seminars in Immunology*.

Ernst, Jason, Pouya Kheradpour, Tarjei S. Mikkelsen, Noam Shoresh, Lucas D. Ward, Charles B. Epstein, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael Coyne, Manching Ku, Timothy Durham, Manolis Kellis, and Bradley E. Bernstein. 2011. "Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types." *Nature*.

Falk, Nathalie, Marlene Lösl, Nadja Schröder, and Andreas Gießl. 2015. "Specialized Cilia in Mammalian Sensory Systems." *Cells*.

FitzGerald, Peter C., David Sturgill, Andrey Shyakhtenko, Brian Oliver, and Charles Vinson. 2006. "Comparative Genomics of Drosophila and Human Core Promoters." *Genome Biology*.

Fornes, Oriol, Jaime A. Castro-Mondragon, Aziz Khan, Robin Van Der Lee, Xi Zhang, Phillip A. Richmond, Bhavi P. Modi, Solenne Correard, Marius Gheorghe, Damir Baranašić, Walter Santana-Garcia, Ge Tan, Jeanne Chèneby, Benoit Ballester, François Parcy, Albin Sandelin, Boris Lenhard, Wyeth W. Wasserman, and Anthony Mathelier. 2020. "JASPAR 2020: Update of the Open-Access Database of Transcription Factor Binding Profiles." *Nucleic Acids Research*.

Foster, Kyle J., Hilary K. Cheesman, Pengpeng Liu, Nicholas D. Peterson, Sarah M. Anderson, and Read Pukkila-Worley. 2020. "Innate Immunity in the C. Elegans Intestine Is Programmed by a Neuronal Regulator of AWC Olfactory Neuron Development." *Cell Reports* 31(1):107478.

Frézal, Lise and Marie Anne Félix. 2015. "C. Elegans Outside the Petri Dish." *ELife*.

Frøkjær-Jensen, Christian, Michael Ailion, and Shawn R. Lockery. 2008. "Ammonium-Acetate Is Sensed by Gustatory and Olfactory Neurons in Caenorhabditis Elegans." *PLoS ONE* 3(6).

De Fruyt, Nathan, Alex J. Yu, Catharine H. Rankin, Isabel Beets, and Yee Lian Chew. 2020. "The Role of Neuropeptides in Learning: Insights from C. Elegans." *International Journal of Biochemistry and Cell Biology*.

Fujiwara, Manabi, Takeshi Ishihara, and Isao Katsura. 1999. "A Novel WD40 Protein, CHE-2, Acts Cell-Autonomously in the Formation of C. Elegans Sensory Cilia." *Development*.

Gabel, Christopher V., Harrison Gabel, Dmitri Pavlichin, Albert Kao, Damon A. Clark, and Aravinthan D. T. Samuel. 2007. "Neural Circuits Mediate Electrosensory Behavior in Caenorhabditis Elegans." *Journal of Neuroscience* 27(28):7586–96.

Gajiwala, Ketan S., Hua Chen, Fabrice Cornille, Bernard P. Roques, Walter Reith, Bernard Mach, and Stephen K. Burley. 2000. "Structure of the Winged-Helix Protein HRFX1 Reveals a New Mode of DNA Binding." *Nature*.

Gao, Zhen, Lu Liu, and Jianhua Ruan. 2017. "Logo2PWM: A Tool to Convert Sequence Logo to Position Weight Matrix." *BMC Genomics*.

Garcia-Gonzalo, Francesc R., Siew Cheng Phua, Elle C. Roberson, Galo Garcia, Monika Abedin, Stéphane Schurmans, Takanari Inoue, and Jeremy F. Reiter. 2015. "Phosphoinositides Regulate Ciliary Protein Trafficking to Modulate Hedgehog Signaling." *Developmental Cell* 34(4):400–409.

Garcia Bellido, A. 1975. "Genetic Control of Wing Disc Development in Drosophila." *CIBA FOUND.SYMP.*

Garcia, Galo, David R. Raleigh, and Jeremy F. Reiter. 2018. "How the Ciliary Membrane Is Organized Inside-Out to Communicate Outside-In." *Current Biology*.

Golson, Maria L. and Klaus H. Kaestner. 2016. "Fox Transcription Factors: From Development to Disease." *Development (Cambridge)*.

Goodman, Miriam B. and Piali Sengupta. 2019. "How Caenorhabditis Elegans Senses Mechanical Stress, Temperature, and Other Physical Stimuli." *Genetics*.

Gray, Jesse M., David S. Karow, Hang Lu, Andy J. Chang, Jennifer S. Chang, Ronald E. Ellis, Michael A. Marietta, and Cornelia I. Bargmann. 2004. "Oxygen Sensation and Social Feeding Mediated by a C. Elegans Guanylate Cyclase Homologue." *Nature* 430(6997):317–22.

Gresh, Lionel, Evelyne Fischer, Andreas Reimann, Myriam Tanguy, Serge Garbay, Xinli Shao, Thomas Hiesberger, Laurence Fiette, Peter Igarashi, Moshe Yaniv, and Marco Pontoglio. 2004. "A Transcriptional Network in Polycystic Kidney Disease." *EMBO Journal*.

Grossman, Sharon R., Jesse Engreitz, John P. Ray, Tung H. Nguyen, Nir Hacohen, and Eric S. Lander. 2018. "Positional Specificity of Different Transcription Factor Classes within Enhancers." *Proceedings of the National Academy of Sciences of the United States of America*.

Guenni L.B.. 2011. "Poisson Distribution and Its Application in Statistics." *In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg*

Gupta, Shobhit, John A. Stamatoyannopoulos, Timothy L. Bailey, and William Stafford Noble. 2007. "Quantifying Similarity between Motifs." *Genome Biology* 8(2).

Hagenlocher, Cathrin, Peter Walentek, Christina Müller, Thomas Thumberger, and Kerstin Feistel. 2013. "Ciliogenesis and Cerebrospinal Fluid Flow in the Developing Xenopus Brain Are Regulated by

Foxj1." *Cilia*.

Hall, D. H. and R. L. Russell. 1991. "The Posterior Nervous System of the Nematode Caenorhabditis Elegans: Serial Reconstruction of Identified Neurons and Complete Pattern of Synaptic Interactions." *Journal of Neuroscience*.

Hallem, Elissa A. and Paul W. Sternberg. 2008. "Acute Carbon Dioxide Avoidance in Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences of the United States of America* 105(23):8038–43.

Hammarlund, Marc, Oliver Hobert, David M. Miller, and Nenad Sestan. 2018. "The CeNGEN Project: The Complete Gene Expression Map of an Entire Nervous System." *Neuron*.

Hao, Limin and Jonathan M. Scholey. 2009. "Intraflagellar Transport at a Glance." *Journal of Cell Science*.

Hao, Limin, Melanie Thein, Ingrid Brust-Mascher, Gul Civelekoglu-Scholey, Yun Lu, Seyda Acar, Bram Prevo, Shai Shaham, and Jonathan M. Scholey. 2011. "Intraflagellar Transport Delivers Tubulin Isotypes to Sensory Cilium Middle and Distal Segments." *Nature Cell Biology*.

Harlow, P. and M. Nemer. 1987. "Coordinate and Selective Beta-Tubulin Gene Expression Associated with Cilium Formation in Sea Urchin Embryos." *Genes & Development*.

Hart, Anne C., Shannon Sims, and Joshua M. Kaplan. 1995. "Synaptic Code for Sensory Modalities Revealed by C. Elegans GLR-1 Glutamate Receptor." *Nature* 378(6552):82–85.

Hattori, Ayuna, Tomoaki Mizuno, Mayuko Akamatsu, Naoki Hisamoto, and Kunihiro Matsumoto. 2013. "The Caenorhabditis Elegans JNK Signaling Pathway Activates Expression of Stress Response Genes by Derepressing the Fos/HDAC Repressor Complex." *PLoS Genetics*.

Hedgecock, E. M. and R. L. Russell. 1975. "Normal and Mutant Thermotaxis in the Nematode Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences of the United States of America*.

Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. 2010. "Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities." *Molecular Cell.*

Hellman, Nathan E., Yan Liu, Erin Merkel, Christina Austin, Stephanie Le Corre, David R. Beier, Zhaoxia Sun, Neeraj Sharma, Bradley K. Yoder, and Iain A. Drummond. 2010. "The Zebrafish Foxj1a Transcription Factor Regulates Cilia Function in Response to Injury and Epithelial Stretch." *Proceedings of the National Academy of Sciences of the United States of America*.

Hiesberger, Thomas, Xinli Shao, Eric Gourley, Andreas Reimann, Marco Pontoglio, and Peter Igarashi. 2005. "Role of the Hepatocyte Nuclear Factor-1β (HNF-1β) C-Terminal Domain in Pkhd1 (ARPKD) Gene Transcription and Renal Cystogenesis." *Journal of Biological Chemistry*.

Hilbe J.M. 2011. "Modeling Count Data". *In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg*

Hilliard, Massimo A., Cornelia I. Bargmann, and Paolo Bazzicalupo. 2002. "C. Elegans Responds to Chemical Repellents by Integrating Sensory Inputs from the Head and the Tail." *Current Biology* 12(9):730–34.

Hilliard, Massimo A., Carmela Bergamasco, Salvatore Arbucci, Ronald H. A. Plasterk, and Paolo Bazzicalupo. 2004. "Worms Taste Bitter: ASH Neurons, QUI-1, GPA-3 and ODR-3 Mediate Quinine Avoidance in Caenorhabditis Elegans." *EMBO Journal* 23(5):1101–11.

Hobert, Oliver. 2002. "PCR Fusion-Based Approach to Create Reporter Gene Constructs for Expression Analysis in Transgenic C. Elegans." *BioTechniques* 32(4):728–30.

Hobert, Oliver. 2005. "Specification of the Nervous System." *WormBook : The Online Review of C. Elegans Biology*.

Hobert, Oliver. 2008. "Regulatory Logic of Neuronal Diversity: Terminal Selector Genes and Selector Motifs." *Proceedings of the National Academy of Sciences of the United States of America*.

Hobert, Oliver. 2010. "Gene Regulation: Enhancers Stepping out of the Shadow." *Current Biology*.

Hobert, Oliver. 2011. "Regulation of Terminal Differentiation Programs in the Nervous System." *Annual Review of Cell and Developmental Biology*.

Hobert, Oliver. 2016a. "A Map of Terminal Regulators of Neuronal Identity in Caenorhabditis Elegans." *Wiley Interdisciplinary Reviews: Developmental Biology*.

Hobert, Oliver. 2016b. "Terminal Selectors of Neuronal Identity." in *Current Topics in Developmental Biology*.

Hobert, Oliver, Donald G. Moerman, Kathleen A. Clark, Mary C. Beckerle, and Gary Ruvkun. 1999. "A Conserved LIM Protein That Affects Muscular Adherens Junction Integrity and Mechanosensory Function in Caenorhabditis Elegans." *Journal of Cell Biology* 144(1):45–57.

Hope, Ian A., Andrew Mounsey, Petra Bauer, and Sobia Aslam. 2003. "The Forkhead Gene Family of Caenorhabditis Elegans." *Gene*.

Horani, Amjad, Thomas W. Ferkol, Susan K. Dutcher, and Steven L. Brody. 2016. "Genetics and Biology of Primary Ciliary Dyskinesia." *Paediatric Respiratory Reviews*.

Howell, Kelly and Oliver Hobert. 2017. "Morphological Diversity of C. Elegans Sensory Cilia Instructed by the Differential Expression of an Immunoglobulin Domain Protein." *Current Biology*.

Hu, Jinghua, Young Kyung Bae, Karla M. Knobel, and Maureen M. Barr. 2006. "Casein Kinase II and Calcineurin Modulate TRPP Function and Ciliary Localization." *Molecular Biology of the Cell*.

Hu, Jinghua and Maureen M. Barr. 2005. "ATP-2 Interacts with the PLAT Domain of LOV-1 and Is Involved in Caenorhabditis Elegans Polycystin Signaling." *Molecular Biology of the Cell*.

Huang, Chunyi George, Todd Lamitina, Peter Agre, and Kevin Strange. 2007. "Functional Analysis of the Aquaporin Gene Family in Caenorhabditis Elegans." *American Journal of Physiology - Cell Physiology*.

Huang, Mingxia, Zheng Zhou, and Stephen J. Elledge. 1998. "The DNA Replication and Damage Checkpoint Pathways Induce Transcription by Inhibition of the Crt1 Repressor." *Cell*.

Huck, S.W. 2008. "Statistical misconceptions". *Routledge Academic, New York.*

Hume, Maxwell A., Luis A. Barrera, Stephen S. Gisselbrecht, and Martha L. Bulyk. 2015. "UniPROBE, Update 2015: New Tools and Content for the Online Database of Protein-Binding Microarray Data on Protein-DNA Interactions." *Nucleic Acids Research*.

Hunt-Newbury, Rebecca, Ryan Viveiros, Robert Johnsen, Allan Mah, Dina Anastas, Lily Fang, Erin Halfnight, David Lee, John Lin, Adam Lorch, Sheldon McKay, H. Mark Okada, Jie Pan, Ana K. Schulz, Domena Tu, Kim Wong, Z. Zhao, Andrey Alexeyenko, Thomas Burglin, Eric Sonnhammer, Ralf Schnabel, Steven J. Jones, Marco A. Marra, David L. Baillie, and Donald G. Moerman. 2007. "High-Throughput in Vivo Analysis of Gene Expression in Caenorhabditis Elegans." *PLoS Biology*.

Hutter, Harald and Jinkyo Suh. 2016. " GExplore 1.4: An Expanded Web Interface for Queries on Caenorhabditis Elegans Protein and Gene Function ." *Worm* 5(4):e1234659.

Inada, Hitoshi, Hiroko Ito, John Satterlee, Piali Sengupta, Kunihiro Matsumoto, and Ikue Mori. 2006. "Identification of Guanylyl Cyclases That Function in Thermosensory Neurons of Caenorhabditis Elegans." *Genetics* 172(4):2239–52.

Inglis, Peter N., Guangshuo Ou, Michel R. Leroux, and Jonathan M. Scholey. 2007. "The Sensory Cilia of Caenorhabditis Elegans." *WormBook : The Online Review of C. Elegans Biology* 1–22.

Inukai, Sachi, Kian Hong Kock, and Martha L. Bulyk. 2017. "Transcription Factor-DNA Binding: Beyond Binding Site Motifs." *Current Opinion in Genetics & Development* 43:110–19.

Ishikawa, Hiroaki and Wallace F. Marshall. 2011. "Ciliogenesis: Building the Cell's Antenna." *Nature Reviews Molecular Cell Biology*.

Iwata, Ryo, Shigekazu Oda, Hirofumi Kunitomo, and Yuichi Iino. 2011. "Roles for Class IIA Phosphatidylinositol Transfer Protein in Neurotransmission and Behavioral Plasticity at the Sensory Neuron Synapses of Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences of the United States of America*.

Jackson, Brian C., Christopher Carpenter, Daniel W. Nebert, and Vasilis Vasiliou. 2010. "Update of Human and Mouse Forkhead Box (FOX) Gene Families." *Human Genomics*.

Jauregui, Andrew R., Ken C. Q. Nguyen, David H. Hall, and Maureen M. Barr. 2008. "The Caenorhabditis Elegans Nephrocystins Act as Global Modifiers of Cilium Structure." *Journal of Cell Biology*.

Jensen, Victor L., Stephen Carter, Anna A. W. M. Sanders, Chunmei Li, Julie Kennedy, Tiffany A. Timbers, Jerry Cai, Noemie Scheidel, Breandán N. Kennedy, Ryan D. Morin, Michel R. Leroux, and Oliver E. Blacque. 2016. "Whole-Organism Developmental Expression Profiling Identifies RAB-28 as a Novel Ciliary GTPase Associated with the BBSome and Intraflagellar Transport." *PLoS Genetics*.

Julius, David and Jeremy Nathans. 2012. "Signaling by Sensory Receptors." *Cold Spring Harbor Perspectives in Biology*.

Kaestner, Klaus H., Walter Knöchel, and Daniel E. Martínez. 2000. "Unified Nomenclature for the Winged Helix/Forkhead Transcription Factors." *Genes and Development*.

Kagoshima, Hiroshi and Yuji Kohara. 2015. "Co-Expression of the Transcription Factors CEH-14 and TTX-1 Regulates AFD Neuron-Specific Genes Gcy-8 and Gcy-18 in C. Elegans." *Developmental Biology*.

Kalderon, Daniel, Bruce L. Roberts, William D. Richardson, and Alan E. Smith. 1984. "A Short Amino Acid Sequence Able to Specify Nuclear Location." *Cell* 39(3 PART 2):499–509.

Kaplan, Joshua M. and H. Robert Horvitz. 1993. "A Dual

Mechanosensory and Chemosensory Neuron in Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences of the United States of America* 90(6):2227–31.

Kaplan, Oktay I., David B. Doroquez, Sebiha Cevik, Rachel V. Bowie, Lara Clarke, Anna A. W. M. Sanders, Katarzyna Kida, Joshua Z. Rappoport, Piali Sengupta, and Oliver E. Blacque. 2012. "Endocytosis Genes Facilitate Protein and Membrane Transport in C. Elegans Sensory Cilia." *Current Biology*.

Katan-Khaykovich, Yael and Yosef Shaul. 1998. "RFX1, a Single DNA-Binding Protein with a Split Dimerization Domain, Generates Alternative Complexes." *Journal of Biological Chemistry*.

Kaufmann, Eckhard and Walter Knöchel. 1996. "Five Years on the Wings of Fork Head." *Mechanisms of Development*.

Kawli, Trupti and Man Wah Tan. 2008. "Neuroendocrine Signals Modulate the Innate Immunity of Caenorhabditis Elegans through Insulin Signaling." *Nature Immunology*.

Keppel, G. 1991. "Design and analysis: A researcher's handbook." *(3rd ed.). Prentice-Hall, Inc.*

Kim, Dennis H., Rhonda Feinbaum, Geneviève Alloing, Fred E. Emerson, Danielle A. Garsin, Hideki Inoue, Miho Tanaka-Hino, Naoki Hisamoto, Kunihiro Hatsumoto, Man Wah Tan, and Frederick H. Ausubel. 2002. "A Conserved P38 MAP Kinase Pathway in Caenorhabditis Elegans Innate Immunity." *Science*.

Kim, Heesun, Takao Ishidate, Krishna S. Ghanta, Meetu Seth, Darryl Conte, Masaki Shirayama, and Craig C. Mello. 2014. "A Co-CRISPR Strategy for Efficient Genome Editing in Caenorhabditis Elegans." *Genetics* 197(4):1069–80.

Kim, Kyuhyung and Chris Li. 2004. "Expression and Regulation of an FMRFamide-Related Neuropeptide Gene Family in Caenorhabditis Elegans." *Journal of Comparative Neurology*.

Kiselak, Elizabeth Anne, Xuening Shen, Jingmei Song, David Roberto Gude, Jiannan Wang, Steven L. Brody, Jerome F. Strauss, and Zhibing Zhang. 2010. "Transcriptional Regulation of an Axonemal Central Apparatus Gene, Sperm-Associated Antigen 6, by a SRY-Related High Mobility Group Transcription Factor, S-SOX5." *Journal of Biological Chemistry*.

Komatsu, Hidetoshi, Ikue Mori, Jeong Seop Rhee, Norio Akaike, and Yasumi Ohshima. 1996. "Mutations in a Cyclic Nucleotide-Gated Channel Lead to Abnormal Thermosensation and Chemosensation in C. Elegans." *Neuron*.

Kratsios, Paschalis, Alberto Stolfi, Michael Levine, and Oliver Hobert. 2012. "Coordinated Regulation of Cholinergic Motor Neuron Traits through a Conserved Terminal Selector Gene." *Nature Neuroscience*.

Kuleshov, Maxim V., Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L. Jenkins, Kathleen M. Jagodnik, Alexander Lachmann, Michael G. McDermott, Caroline D. Monteiro, Gregory W. Gundersen, and Avi Ma'ayan. 2016. "Enrichr: A Comprehensive Gene Set Enrichment Analysis Web Server 2016 Update." *Nucleic Acids Research* 44(W1):W90–97.

Kunitomo, Hirofumi, Hiroko Uesugi, Yuji Kohara, and Yuichi Iino. 2005. "Identification of Ciliated Sensory Neuron-Expressed Genes in Caenorhabditis Elegans Using Targeted Pull-down of Poly(A) Tails." *Genome Biology*.

Kypri, Elena, Andri Christodoulou, Giannis Maimaris, Mette Lethan, Maria Markaki, Costas Lysandrou, Carsten W. Lederer, Nektarios Tavernarakis, Stefan Geimer, Lotte B. Pedersen, and Niovi Santama. 2014. "The Nucleotide-Binding Proteins Nubp1 and Nubp2 Are Negative Regulators of Ciliogenesis." *Cellular and Molecular Life Sciences*.

L'Etoile, Noelle D. and Cornelia I. Bargmann. 2000. "Olfaction and Odor Discrimination Are Mediated by the C. Elegans Guanylyl Cyclase ODR-1." *Neuron*.

Labed, Sid Ahmed, Khursheed A. Wani, Sakthimala Jagadeesan, Abdul Hakkim, Mehran Najibi, and Javier Elbio Irazoqui. 2018. "Intestinal Epithelial Wnt Signaling Mediates Acetylcholine-Triggered Host Defense against Infection." *Immunity*.

Lalmansingh, Avin S., Sudipan Karmakar, Yetao Jin, and Akhilesh K. Nagaich. 2012. "Multiple Modes of Chromatin Remodeling by Forkhead Box Proteins." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*.

Lambacher, Nils J., Ange Line Bruel, Teunis J. P. Van Dam, Katarzyna Szymaska, Gisela G. Slaats, Stefanie Kuhns, Gavin J. McManus, Julie E. Kennedy, Karl Gaff, Ka Man Wu, Robin Van Der Lee, Lydie Burglen, Diane Doummar, Jean Baptiste Rivière, Laurence Faivre, Tania Attié-Bitach, Sophie Saunier, Alistair Curd, Michelle Peckham, Rachel H. Giles, Colin A. Johnson, Martijn A. Huynen, Christel Thauvin-Robinet, and Oliver E. Blacque. 2016. "TMEM107 Recruits Ciliopathy Proteins to Subdomains of the Ciliary Transition Zone and Causes Joubert Syndrome." *Nature Cell Biology*.

Landler, Lukas, Simon Nimpf, Tobias Hochstoeger, Gregory C. Nordmann, Artemis Papadaki-Anastasopoulou, and David A. Keays. 2018. "Comment on 'Magnetosensitive Neurons Mediate Geomagnetic Orientation in

Caenorhabditis Elegans.'" *ELife*.

Laurençon, Anne, Raphaëlle Dubruille, Evgeni Efimenko, Guillaume Grenier, Ryan Bissett, Elisabeth Cortier, Vivien Rolland, Peter Swoboda, and Bénédicte Durand. 2007. "Identification of Novel Regulatory Factor X (RFX) Target Genes by Comparative Genomics in Drosophila Species." *Genome Biology*.

Lee, Hsiu Hsiang and Manfred Frasch. 2004. "Survey of Forkhead Domain Encoding Genes in the Drosophila Genome: Classification and Embryonic Expression Patterns." *Developmental Dynamics*.

Lee, Sangseok and Dong Kyu Lee. 2018. "What Is the Proper Way to Apply the Multiple Comparison Test?" *Korean Journal of Anesthesiology* 71(5):353–60.

Lee, Yong Seok, Yang Seo Park, Seunghee Nam, Su Jeong Suh, Junho Lee, Bong Kiun Kaang, and Nam Jeong Cho. 2000. "Characterization of GAR-2, a Novel G Protein-Linked Acetylcholine Receptor from Caenorhabditis Elegans." *Journal of Neurochemistry*.

Leewenhoeck, Antony van. 1676. "Concerning Little Animals by Him Observed in Rain- Well- Sea- and Snow-Water; as Also in Water Wherein Pepper Had Lain Infused." *Philosophical Transactions*.

Lefebvre, P. A., S. A. Nordstrom, J. E. Moulder, and J. L. Rosenbaum. 1978. "Flagellar Elongation and Shortening in Chlamydomonas. IV. Effects of Flagellar Detachment, Regeneration, and Resorption on the Induction of Flagellar Protein Synthesis." *Journal of Cell Biology*.

Lenhard, Boris, Albin Sandelin, and Piero Carninci. 2012. "Metazoan Promoters: Emerging Characteristics and Insights into Transcriptional Regulation." *Nature Reviews Genetics*.

Levo, Michal and Eran Segal. 2014. "In Pursuit of Design Principles of Regulatory Sequences." *Nature Reviews Genetics*.

Li, Chris and Kyuhyung Kim. 2008. "Neuropeptides." *WormBook: The Online Review of C. Elegans Biology*.

Li, Chunmei, Victor L. Jensen, Kwangjin Park, Julie Kennedy, Francesc R. Garcia-Gonzalo, Marta Romani, Roberta De Mori, Ange Line Bruel, Dominique Gaillard, Bérénice Doray, Estelle Lopez, Jean Baptiste Rivière, Laurence Faivre, Christel Thauvin-Robinet, Jeremy F. Reiter, Oliver E. Blacque, Enza Maria Valente, and Michel R. Leroux. 2016. "MKS5 and CEP290 Dependent Assembly Pathway of the Ciliary Transition Zone." *PLoS Biology* 14(3).

Li, Jin Billy, Jantje M. Gerdes, Courtney J. Haycraft, Yanli Fan, Tanya M. Teslovich, Helen May-Simera, Haitao Li, Oliver E. Blacque, Linya Li, Carmen C. Leitch, Richard Allan Lewis, Jane S. Green, Patrick S. Parfrey, Michel R. Leroux, William S. Davidson, Philip L. Beales, Lisa M. Guay-Woodford, Bradley K. Yoder, Gary D. Stormo, Nicholas Katsanis, and Susan K. Dutcher. 2004. "Comparative Genomics Identifies a Flagellar and Basal Body Proteome That Includes the BBS5 Human Disease Gene." *Cell*.

Li, Lili, Guangmei Tian, Hai Peng, Dan Meng, Liang Wang, Xiao Hu, Cheng Tian, Miao He, Junfei Zhou, Lihong Chen, Cheng Fu, Weixiong Zhang, and Zhangfeng Hu. 2018. "New Class of Transcription Factors Controls Flagellar Assembly by Recruiting RNA Polymerase II in Chlamydomonas." *Proceedings of the National Academy of Sciences of the United States of America*.

Li, Wei, Lijun Kang, Beverly J. Piggott, Zhaoyang Feng, and X. Z. Shaw. Xu. 2011. "The Neural Circuits and Sensory Channels Mediating Harsh Touch Sensation in Caenorhabditis Elegans." *Nature Communications* 2(1).

Li, Wenjing, Peishan Yi, and Guangshuo Ou. 2015. "Somatic CRISPR-Cas9-Induced Mutations Reveal Roles of Embryonically Essential Dynein Chains in Caenorhabditis Elegans Cilia." *Journal of Cell Biology*.

Lieberman-Aiden, Erez, Nynke L. Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner, Richard Sandstrom, Bradley Bernstein, M. A. Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander, and Job Dekker. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science*.

Liu, Jie, Alex Ward, Jingwei Gao, Yongming Dong, Nana Nishio, Hitoshi Inada, Lijun Kang, Yong Yu, Di Ma, Tao Xu, Ikue Mori, Zhixiong Xie, and X. Z. Shaw. Xu. 2010. "C. Elegans Phototransduction Requires a G Protein-Dependent CGMP Pathway and a Taste Receptor Homolog." *Nature Neuroscience*.

Liu, Yan, Narendra Pathak, Albrecht Kramer-Zucker, and Iain A. Drummond. 2007. "Notch Signaling Controls the Differentiation of Transporting Epithelia and Multiciliated Cells in the Zebrafish Pronephros." *Development*.

Liu, Yiyong and Jingru Sun. 2021. "Detection of Pathogens and Regulation of Immunity by the Caenorhabditis Elegans Nervous System." *MBio*.

Lubelsky, Yoav, Nina Reuven, and Yosef Shaul. 2005. "Autorepression of Rfx1 Gene Expression: Functional Conservation from Yeast to Humans in Response to DNA Replication Arrest." *Molecular*

and Cellular Biology.

Maddison, David R., Katja Sabine Schulz, and Wayne P. Maddison. 2007. "The Tree of Life Web Project." *Zootaxa*.

Madeira, Fábio, Young Mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, Adrian R. N. Tivey, Simon C. Potter, Robert D. Finn, and Rodrigo Lopez. 2019. "The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019." *Nucleic Acids Research*.

Maerkl, Sebastian J. and Stephen R. Quake. 2007. "A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors." *Science*.

Maguire, Julie E., Malan Silva, Ken C. Q. Nguyen, Elizabeth Hellen, Andrew D. Kern, David H. Hall, and Maureen M. Barr. 2015. "Myristoylated CIL-7 Regulates Ciliary Extracellular Vesicle Biogenesis." *Molecular Biology of the Cell*.

Malicki, Jarema J. and Colin A. Johnson. 2017. "The Cilium: Cellular Antenna and Central Processing Unit." *Trends in Cell Biology*.

Manojlovic, Zarko, Ryan Earwood, Akiko Kato, Branko Stefanovic, and Yoichi Kato. 2014. "RFX7 Is Required for the Formation of Cilia in the Neural Tube." *Mechanisms of Development*.

Margie, Olivia, Chris Palmer, and Ian Chin-Sang. 2013. "C. Elegans Chemotaxis Assay." *Journal of Visualized Experiments* (74).

Mathelier, Anthony, Wenqiang Shi, and Wyeth W. Wasserman. 2015. "Identification of Altered Cis-Regulatory Elements in Human Disease." *Trends in Genetics*.

Matys, V., O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A. E. Kel, and E. Wingender. 2006. "TRANSFAC and Its Module TRANSCompel: Transcriptional Gene Regulation in Eukaryotes." *Nucleic Acids Research*.

Mazo, Gregory, Nadine Soplop, Won Jing Wang, Kunihiro Uryu, and Meng Fu Bryan Tsou. 2016. "Spatial Control of Primary Ciliogenesis by Subdistal Appendages Alters Sensation-Associated Properties of Cilia." *Developmental Cell*.

McGrath, Patrick T., Yifan Xu, Michael Ailion, Jennifer L. Garrison, Rebecca A. Butcher, and Cornelia I. Bargmann. 2011. "Parallel Evolution of Domesticated Caenorhabditis Species Targets Pheromone Receptor Genes." *Nature*.

Mello, C. C., J. M. Kramer, D. Stinchcomb, and V. Ambros. 1991. "Efficient Gene Transfer in C.Elegans: Extrachromosomal Maintenance and Integration

of Transforming Sequences." *The EMBO Journal* 10(12):3959–70.

Mello, Craig and Andrew Fire. 1995. "DNA Transformation." *Methods in Cell Biology* 48(C):451–82.

de Mendoza, Alex and Arnau Sebé-Pedrós. 2019. "Origin and Evolution of Eukaryotic Transcription Factors." *Current Opinion in Genetics and Development*.

Mercier, Eloi, Arnaud Droit, Leping Li, Gordon Robertson, Xuekui Zhang, and Raphael Gottardo. 2011. "An Integrated Pipeline for the Genome-Wide Analysis of Transcription Factor Binding Sites from ChiP-Seq." *PLoS ONE*.

Meyer, Clifford A. and X. Shirley Liu. 2014. "Identifying and Mitigating Bias in Next-Generation Sequencing Methods for Chromatin Biology." *Nature Reviews Genetics*.

Milanetti, Edoardo, Giorgio Gosti, Luca De Flaviis, Pier Paolo Olimpieri, Silvia Schwartz, Davide Caprini, Giancarlo Ruocco, and Viola Folli. 2019. "Investigation of the Binding between Olfactory Receptors and Odorant Molecules in C. Elegans Organism." *Biophysical Chemistry*.

Miller, Renee M. and Douglas S. Portman. 2010. "A Latent Capacity of the C. Elegans Polycystins to Disrupt Sensory Transduction Is Repressed by the Single-Pass Ciliary Membrane Protein CWP-5." *DMM Disease Models and Mechanisms*.

Mitchell, David R. 2017. "Evolution of Cilia." *Cold Spring Harbor Perspectives in Biology*.

Mizuno, Tomoaki, Naoki Hisamoto, Takashi Terada, Tea Kondo, Makoto Adachi, Eisuke Nishida, Dennis H. Kim, Frederick M. Ausubel, and Kunihiro Matsumoto. 2004. "The Caenorhabditis Elegans MAPK Phosphatase VHP-1 Mediates a Novel JNK-like Signaling Pathway in Stress Response." *EMBO Journal*.

MM, Barr and Sternberg PW. 1999. "A Polycystic Kidney-Disease Gene Homologue Required for Male Mating Behavior in C. Elegans." *Nature*.

Molla-Herman, Anahi, Rania Ghossoub, Thierry Blisnick, Alice Meunier, Catherine Serres, Flora Silbermann, Chris Emmerson, Kelly Romeo, Pierre Bourdoncle, Alain Schmitt, Sophie Saunier, Nathalie Spassky, Philippe Bastin, and Alexandre Benmerah. 2010. "The Ciliary Pocket: An Endocytic Membrane Domain at the Base of Primary and Motile Cilia." *Journal of Cell Science*.

Moussaif, Mustapha and Ji Ying Sze. 2009. "Intraflagellar Transport/Hedgehog-Related Signaling Components Couple Sensory Cilium Morphology and Serotonin Biosynthesis in Caenorhabditis Elegans." *Journal of Neuroscience*.

Mukherjee, Ishita, Sudipto Roy, and Saikat Chakrabarti. 2019. "Identification of Important Effector Proteins in the FOXJ1 Transcriptional Network Associated with Ciliogenesis and Ciliary Function." *Frontiers in Genetics*.

Mukhopadhyay, Arnab, Bart Deplancke, Albertha J. M. Walhout, and Heidi A. Tissenbaum. 2005. "C. Elegans Tubby Regulates Life Span and Fat Storage by Two Independent Mechanisms." *Cell Metabolism*.

Mukhopadhyay, Saikat, Yun Lu, Hongmin Qin, Anne Lanjuin, Shai Shaham, and Piali Sengupta. 2007. "Distinct IFT Mechanisms Contribute to the Generation of Ciliary Structural Diversity in C. Elegans." *EMBO Journal*.

Murphy, Coleen T. and Patrick J. Hu. 2013. "Insulin/Insulin-like Growth Factor Signaling in C. Elegans." *WormBook : The Online Review of C. Elegans Biology*.

Nachury, Maxence V., Alexander V. Loktev, Qihong Zhang, Christopher J. Westlake, Johan Peränen, Andreas Merdes, Diane C. Slusarski, Richard H. Scheller, J. Fernando Bazan, Val C. Sheffield, and Peter K. Jackson. 2007. "A Core Complex of BBS Proteins Cooperates with the GTPase Rab8 to Promote Ciliary Membrane Biogenesis." *Cell*.

Nachury, Maxence V. and David U. Mick. 2019. "Establishing and Regulating the Composition of Cilia for Signal Transduction." *Nature Reviews Molecular Cell Biology*.

Nachury, Maxence V., E. Scott Seeley, and Hua Jin. 2010. "Trafficking to the Ciliary Membrane: How to Get across the Periciliary Diffusion Barrier?" *Annual Review of Cell and Developmental Biology*.

Nakagawa, So, Stephen S. Gisselbrecht, Julia M. Rogers, Daniel L. Hartl, and Martha L. Bulyk. 2013. "DNA-Binding Specificity Changes in the Evolution of Forkhead Transcription Factors." *Proceedings of the National Academy of Sciences of the United States of America*.

Nam, Deokhwa and Erin L. Reineke. 2017. "Timing and Targeting of Treatment in Left Ventricular Hypertrophy." *Methodist DeBakey Cardiovascular Journal*.

Narasimhan, Kamesh, Samuel A. Lambert, Ally W. H. Yang, Jeremy Riddell, Sanie Mnaimneh, Hong Zheng, Mihai Albu, Hamed S. Najafabadi, John S. Reece-Hoyes, Juan I. Fuxman Bass, Albertha J. M. Walhout, Matthew T. Weirauch, and Timothy R. Hughes. 2015. "Mapping and Analysis of Caenorhabditis Elegans Transcription Factor Sequence Specificities." *ELife*.

Nechipurenko, Inna V. and Piali Sengupta. 2017. "The Rise and Fall of Basal Bodies in the Nematode Caenorhabditis Elegans." *Cilia*.

Newton, Fay G., Petra I. zur Lage, Somdatta Karak, Daniel J. Moore, Martin C. Göpfert, and Andrew P. Jarman. 2012. "Forkhead Transcription Factor Fd3F Cooperates with Rfx to Regulate a Gene Expression Program for Mechanosensory Cilia Specialization." *Developmental Cell*.

Nguyen, Phuong Anh T., Willisa Liou, David H. Hall, and Michel R. Leroux. 2014. "Ciliopathy Proteins Establish a Bipartite Signaling Compartment in a C. Elegans Thermosensory Neuron." *Journal of Cell Science* 127(24):5317–30.

Nigon V.M., and Félix M. A. 2017. "History of Research on C. Elegans and Other Free-Living Nematodes as Model Organisms." *WormBook*.

Niimura, Yoshihito and Masatoshi Nei. 2003. "Evolution of Olfactory Receptor Genes in the Human Genome." *Proceedings of the National Academy of Sciences of the United States of America*.

Nonaka, Shigenori, Yosuke Tanaka, Yasushi Okada, Sen Takeda, Akihiro Harada, Yoshimitsu Kanai, Mizuho Kido, and Nobutaka Hirokawa. 1998. "Randomization of Left-Right Asymmetry Due to Loss of Nodal Cilia Generating Leftward Flow of Extraembryonic Fluid in Mice Lacking KIF3B Motor Protein." *Cell*.

Ohinata, Yasuhide, Bernhard Payer, Dónal O'Carroll, Katia Ancelin, Yukiko Ono, Mitsue Sano, Sheila C. Barton, Tetyana Obukhanych, Michel Nussenzweig, Alexander Tarakhovsky, Mitinori Saitou, and M. Azim Surani. 2005. "Blimp1 Is a Critical Determinant of the Germ Cell Lineage in Mice." *Nature*.

Olivier-Mason, Anique, Martin Wojtyniak, Rachel V. Bowie, Inna V. Nechipurenko, Oliver E. Blacque, and Piali Sengupta. 2013. "Transmembrane Protein OSTA-1 Shapes Sensory Cilia Morphology via Regulation of Intracellular Membrane Trafficking in C. Elegans." *Development (Cambridge)*.

Omori, Yoshihiro, Chengtian Zhao, Arunesh Saras, Saikat Mukhopadhyay, Woong Kim, Takahisa Furukawa, Piali Sengupta, Alexey Veraksa, and Jarema Malicki. 2008. "Elipsa Is an Early Determinant of Ciliogenesis That Links the IFT Particle to Membrane-Associated Small GTPase Rab8." *Nature Cell Biology*.

Ortiz, Christopher O., John F. Etchberger, Shoshana L. Posy, Christian Frøkjær-Jensen, Shawn Lockery, Barry Honig, and Oliver Hobert. 2006. "Searching for Neuronal Left/Right Asymmetry: Genomewide Analysis of Nematode Receptor-Type Guanylyl Cyclases." *Genetics*.

Ou, Guangshuo, Makato Koga, Oliver E. Blacque, Takashi

Murayama, Yasumi Ohshima, Jenny C. Schafer, Chunmei Li, Bradley K. Yoder, Michel R. Leroux, and Jonathan M. Scholey. 2007. "Sensory Ciliogenesis in Caenorhabditis Elegans: Assignment of IFT Components into Distinct Modules Based on Transport and Phenotypic Profiles." *Molecular Biology of the Cell*.

Overdier, D. G., A. Porcella, and R. H. Costa. 1994. "The DNA-Binding Specificity of the Hepatocyte Nuclear Factor 3/Forkhead Domain Is Influenced by Amino-Acid Residues Adjacent to the Recognition Helix." *Molecular and Cellular Biology*.

Packer, Jonathan S., Qin Zhu, Chau Huynh, Priya Sivaramakrishnan, Elicia Preston, Hannah Dueck, Derek Stefanik, Kai Tan, Cole Trapnell, Junhyong Kim, Robert H. Waterston, and John I. Murray. 2019. "A Lineage-Resolved Molecular Atlas of C. Elegans Embryogenesis at Single-Cell Resolution." *Science*.

Park, Donha, Inish O'Doherty, Rishi K. Somvanshi, Axel Bethke, Frank C. Schroeder, Ujendra Kumar, and Donald L. Riddle. 2012. "Interaction of Structure-Specific and Promiscuous G-Protein-Coupled Receptors Mediates Small-Molecule Signaling in Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences of the United States of America*.

Patir, Anirudh, Amy M. Fraser, Mark W. Barnett, Lynn McTeir, Joe Rainger, Megan G. Davey, and Tom C. Freeman. 2020. "The Transcriptional Signature Associated with Human Motile Cilia." *Scientific Reports*.

Pedersen, Lotte B., Jacob M. Schrøder, Peter Satir, and Søren T. Christensen. 2012. "The Ciliary Cytoskeleton." *Comprehensive Physiology*.

Pedersen, Lotte B., Iben R. Veland, Jacob M. Schrøder, and Søren T. Christensen. 2008. "Assembly of Primary Cilia." *Developmental Dynamics*.

Pereira, Laura, Paschalis Kratsios, Esther Serrano-Saiz, Hila Sheftel, Avi E. Mayo, David H. Hall, John G. White, Brigitte LeBoeuf, L. Rene Garcia, Uri Alon, and Oliver Hobert. 2015. "A Cellular and Regulatory Map of the Cholinergic Nervous System of C. Elegans." *ELife*.

Perens, Elliot A. and Shai Shaham. 2005. "C. Elegans Daf-6 Encodes a Patched-Related Protein Required for Lumen Formation." *Developmental Cell* 8(6):893–906.

Perkins, Lizabeth A., Edward M. Hedgecock, J. Nichol Thomson, and Joseph G. Culotti. 1986. "Mutant Sensory Cilia in the Nematode Caenorhabditis Elegans." *Developmental Biology*.

Phirke, Prasad, Evgeni Efimenko, Swetha Mohan, Jan Burghoorn, Filip Crona, Mathieu W. Bakhoum, Maria Trieb, Kim Schuske, Erik M. Jorgensen, Brian P. Piasecki, Michel R. Leroux, and Peter Swoboda. 2011. "Transcriptional Profiling of C. Elegans DAF-19 Uncovers a Ciliary Base-Associated Protein and a CDK/CCRK/LF2p-Related Kinase Required for Intraflagellar Transport." *Developmental Biology*.

Piasecki, Brian P., Jan Burghoorn, and Peter Swoboda. 2010. "Regulatory Factor X (RFX)-Mediated Transcriptional Rewiring of Ciliary Genes in Animals." *Proceedings of the National Academy of Sciences of the United States of America*.

Pierce-Shimomura, J. T., S. Faumont, M. R. Gaston, B. J. Pearson, and S. R. Lockery. 2001. "The Homeobox Gene Lim-6 Is Required for Distinct Chemosensory Representations in C. Elegans." *Nature*.

Pierreux, Christophe E., Aurélie V. Poll, Caroline R. Kemp, Frédéric Clotman, Miguel A. Maestro, Sabine Cordi, Jorge Ferrer, Luc Leyns, Guy G. Rousseau, and Frédéric P. Lemaigre. 2006. "The Transcription Factor Hepatocyte Nuclear Factor-6 Controls the Development of Pancreatic Ducts in the Mouse." *Gastroenterology*.

Pohl, Barbara S. and Walter Knöchel. 2005. "Of Fox and Frogs: Fox (Fork Head/Winged Helix) Transcription Factors in Xenopus Development." *Gene*.

Portman, D. S. and S. W. Emmons. 2000. "The Basic Helix-Loop-Helix Transcription Factors LIN-32 and HLH-2 Function Together in Multiple Steps of a C. Elegans Neuronal Sublineage." *Development*.

Postema, Merel C., Amaia Carrion-Castillo, Simon E. Fisher, Guy Vingerhoets, and Clyde Francks. 2020. "The Genetics of Situs Inversus without Primary Ciliary Dyskinesia." *Scientific Reports* 10(1):3677.

Pradel, Elizabeth, Yun Zhang, Nathalie Pujol, Tohey Matsuyama, Cornelia I. Bargmann, and Jonathan J. Ewbank. 2007. "Detection and Avoidance of a Natural Product from the Pathogenic Bacterium Serratia Marcescens by Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences of the United States of America*.

Prevo, Bram, Jonathan M. Scholey, and Erwin J. G. Peterman. 2017. "Intraflagellar Transport: Mechanisms of Motor Action, Cooperation, and Cargo Delivery." *FEBS Journal*.

Quigley, Ian K. and Chris Kintner. 2017. "Rfx2 Stabilizes Foxj1 Binding at Chromatin Loops to Enable Multiciliated Cell Gene Expression." *PLoS Genetics*.

Ramulu, Pradeep and Jeremy Nathans. 2001. "Cellular

and Subcellular Localization, N-Terminal Acylation, and Calcium Binding of Caenorhabditis Elegans Protein Phosphatase with EF-Hands." *Journal of Biological Chemistry*.

Rankin, Catharine H. 2002. "From Gene to Identified Neuron to Behaviour in Caenorhabditis Elegans." *Nature Reviews Genetics*.

Rayamajhi, Dheeraj and Sudipto Roy. 2020. "Multiciliated Cells: Rise and Fall of the Deuterosomes." *Trends in Cell Biology*.

Rechtsteiner, Andreas, Meghan E. Costello, Thea A. Egelhofer, Jacob M. Garrigues, Susan Strome, and Lisa N. Petrella. 2019. "Repression of Germline Genes in Caenorhabditis Elegans Somatic Tissues by H3K9 Dimethylation of Their Promoters." *Genetics*.

Reiter, Jeremy F., Oliver E. Blacque, and Michel R. Leroux. 2012. "The Base of the Cilium: Roles for Transition Fibres and the Transition Zone in Ciliary Formation, Maintenance and Compartmentalization." *EMBO Reports*.

Reiter, Jeremy F. and Michel R. Leroux. 2017. "Genes and Molecular Pathways Underpinning Ciliopathies." *Nature Reviews Molecular Cell Biology*.

Reith, Walter, Carlos Herrero-Sanchez, Michel Kobr, Paolo Silacci, Christine Berte, Emanuelle Barras, Sylvia Fey, and Bernard Mach. 1990. "MHC Class II Regulatory Factor RFX Has a Novel DNA-Binding Domain and a Functionally Independent Dimerization Domain." *Genes and Development*.

Richards, Thomas A. and Thomas Cavalier-Smith. 2005. "Myosin Domain Evolution and the Primary Divergence of Eukaryotes." *Nature*.

Robert, Valérie J., Steve Garvis, and Francesca Palladino. 2015. "Repression of Somatic Cell Fate in the Germline." *Cellular and Molecular Life Sciences* 72(19):3599–3620.

Roeder, Robert G. 1996. "The Role of General Initiation Factors in Transcription by RNA Polymerase II." *Trends in Biochemical Sciences*.

Rogers, Julia M., Colin T. Waters, Tom C. M. Seegar, Sanchez M. Jarrett, Amelia N. Hallworth, Stephen C. Blacklow, and Martha L. Bulyk. 2019. "Bispecific Forkhead Transcription Factor FoxN3 Recognizes Two Distinct Motifs with Different DNA Shapes." *Molecular Cell*.

Russell, Joshua, Andrés G. Vidal-Gadea, Alex Makay, Carolyn Lanam, and Jonathan T. Pierce-Shimomura. 2014. "Humidity Sensation Requires Both Mechanosensory and Thermosensory Pathways in Caenorhabditis Elegans." *Proceedings of the National Academy of Sciences of the United States of America*.

Sambongi, Yoshihiro, Takashi Nagae, Yanna Liu, Takao Yoshimizu, Kenji Takeda, Yoh Wada, and Masamitsu Futai. 1999. "Sensing of Cadmium and Copper Ions by Externally Exposed ADL, ASE, and ASH Neurons Elicits Avoidance Response in Caenorhabditis Elegans." *NeuroReport*.

Sanyal, Amartya, Bryan R. Lajoie, Gaurav Jain, and Job Dekker. 2012. "The Long-Range Interaction Landscape of Gene Promoters." *Nature*.

Sarov, Mihail, John I. Murray, Kristin Schanze, Andrei Pozniakovski, Wei Niu, Karolin Angermann, Susanne Hasse, Michaela Rupprecht, Elisabeth Vinis, Matthew Tinney, Elicia Preston, Andrea Zinke, Susanne Enst, Tina Teichgraber, Judith Janette, Kadri Reis, Stephan Janosch, Siegfried Schloissnig, Radoslaw K. Ejsmont, Cindie Slightam, Xiao Xu, Stuart K. Kim, Valerie Reinke, A. Francis Stewart, Michael Snyder, Robert H. Waterston, and Anthony A. Hyman. 2012. "A Genome-Scale Resource for in Vivo Tag-Based Protein Function Exploration in C. Elegans." *Cell*.

Satir, Peter and Birgit H. Satir. 2019. "The Conserved Ancestral Signaling Pathway from Cilium to Nucleus." *Journal of Cell Science*.

Sawin, ER, R. Ranganathan, and HR Horvitz. 2000. "C. Elegans Locomotory Rate Is Modulated by the Environment through a Dopaminergic Pathway and by Experience through a Serotonergic Pathway." *Neuron*.

Schafer, Jenny C., Courtney J. Haycraft, James H. Thomas, Bradley K. Yoder, and Peter Swoboda. 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*.

Schafer, Jenny C., Marlene E. Winkelbauer, Corey L. Williams, Courtney J. Haycraft, Renee A. Desmond, and Bradley K. Yoder. 2006. "IFTA-2 Is a Conserved Cilia Protein Involved in Pathways Regulating Longevity Dauer Formation in Caenorhabditis Elegans." *Journal of Cell Science* 119(19):4088–4100.

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. 2012. "Fiji: An Open-Source Platform for Biological-Image Analysis." *Nature Methods* 9(7):676–82.

Schneider, Thomas D. and R. Michael Stephens. 1990. "Sequence Logos: A New Way to Display Consensus Sequences." *Nucleic Acids Research*.

Sengupta, P., J. H. Chou, and C. I. Bargmann. 1996. "Odr-

10 Encodes a Seven Transmembrane Domain Olfactory Receptor Required for Responses to the Odorant Diacetyl." *Cell*.

Senti, Gabriele, Marina Ezcurra, Jana Löbner, William R. Schafer, and Peter Swoboda. 2009. "Worms with a Single Functional Sensory Cilium Generate Proper Neuron-Specific Behavioral Output." *Genetics* 183(2):595–605.

Senti, Gabriele and Peter Swoboda. 2008. "Distinct Isoforms of the RFX Transcription Factor DAF-19 Regulate Ciliogenesis and Maintenance of Synaptic Activity." *Molecular Biology of the Cell*.

Serobyan, Vahan, Zacharias Kontarakis, Mohamed A. El-Brolosy, Jordan M. Welker, Oleg Tolstenkov, Amr M. Saadeldein, Nicholas Retzer, Alexander Gottschalk, Ann M. Wehman, and Didier Y. R. Stainier. 2020. "Transcriptional Adaptation in Caenorhabditis Elegans." *ELife*.

Serrano-Saiz, Esther, Laura Pereira, Marie Gendrel, Ulkar Aghayeva, Abhishek Battacharya, Kelly Howell, L. Rene Garcia, and Oliver Hobert. 2017. "A Neurotransmitter Atlas of the Caenorhabditis Elegans Male Nervous System Reveals Sexually Dimorphic Neurotransmitter Usage." *Genetics*.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal*.

Shapiro, Samuel Sanford and B. Wilk Martin. 1965. "An Analysis of Variance Test for Normality." *Biometrika*.

Sherman, Teresa, Marina N. Chernova, Jeffrey S. Clark, Lianwei Jiang, Seth L. Alper, and Keith Nehrke. 2005. "The Abts and Sulp Families of Anion Transporters from Caenorhabditis Elegans." *American Journal of Physiology - Cell Physiology*.

Shivers, Robert P., Daniel J. Pagano, Tristan Kooistra, Claire E. Richardson, Kirthi C. Reddy, Janelle K. Whitney, Odile Kamanzi, Kunihiro Matsumoto, Naoki Hisamoto, and Dennis H. Kim. 2010. "Phosphorylation of the Conserved Transcription Factor ATF-7 by PMK-1 P38 MAPK Regulates Innate Immunity in Caenorhabditis Elegans." *PLoS Genetics*.

Shlyueva, Daria, Gerald Stampfel, and Alexander Stark. 2014. "Transcriptional Enhancers: From Properties to Genome-Wide Predictions." *Nature Reviews Genetics*.

Siggers, Trevor and Raluca Gordân. 2014. "Protein-DNA Binding: Complexities and Multi-Protein Codes." *Nucleic Acids Research*.

Singer J.M. 2011. "Central Limit Theorems." *In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg*

Silverman, Michael A. and Michel R. Leroux. 2009.

"Intraflagellar Transport and the Generation of Dynamic, Structurally and Functionally Diverse Cilia." *Trends in Cell Biology*.

Snouffer, Ashley, Desmond Brown, Hankyu Lee, Jonathon Walsh, Floria Lupu, Ryan Norman, Karl Lechtreck, Hyuk Wan Ko, and Jonathan Eggenschwiler. 2017. "Cell Cycle-Related Kinase (CCRK) Regulates Ciliogenesis and Hedgehog Signaling in Mice." *PLoS Genetics*.

Sorokin, S. P. 1968. "Reconstructions of Centriole Formation and Ciliogenesis in Mammalian Lungs." *Journal of Cell Science*.

Spassky, Nathalie and Alice Meunier. 2017. "The Development and Functions of Multiciliated Epithelia." *Nature Reviews Molecular Cell Biology*.

Spitz, François and Eileen E. M. Furlong. 2012. "Transcription Factors: From Enhancer Binding to Developmental Control." *Nature Reviews Genetics*.

Stadhouders, Ralph, Guillaume J. Filion, and Thomas Graf. 2019. "Transcription Factors and 3D Genome Conformation in Cell-Fate Decisions." *Nature*.

De Stasio, Elizabeth A., Katherine P. Mueller, Rosemary J. Bauer, Alexander J. Hurlburt, Sophie A. Bice, Sophie L. Scholtz, Prasad Phirke, Debora Sugiaman-Trapman, Loraina A. Stinson, Haili B. Olson, Savannah L. Vogel, Zabdiel Ek-Vazquez, Yagmur Esemen, Jessica Korzynski, Kelsey Wolfe, Bonnie N. Arbuckle, He Zhang, Gaelen Lombard-Knapp, Brian P. Piasecki, and Peter Swoboda. 2018. "An Expanded Role for the RFX Transcription Factor DAF-19, with Dual Functions in Ciliated and Nonciliated Neurons." *Genetics*.

Stefanakis, Nikolaos, Ines Carrera, and Oliver Hobert. 2015. "Regulatory Logic of Pan-Neuronal Gene Expression in C. Elegans." *Neuron*.

Steimle, Viktor, Bénédicte Durand, Emmanuèle Barras, Madeleine Zufferey, Martin R. Hadam, Bernard Mach, and Walter Reith. 1995. "A Novel DNA-Binding Regulatory Factor Is Mutated in Primary MHC Class II Deficiency (Bare Lymphocyte Syndrome)." *Genes and Development*.

Stormo, Gary D. and Yue Zhao. 2010. "Determining the Specificity of Protein-DNA Interactions." *Nature Reviews Genetics*.

Stubbs, Jennifer L., Isao Oishi, Juan Carlos Izpisúa Belmonte, and Chris Kintner. 2008. "The Forkhead Protein Foxj1 Specifies Node-like Cilia in Xenopus and Zebrafish Embryos." *Nature Genetics*.

Styer, Katie L., Varsha Singh, Evan Macosko, Sarah E. Steele, Cornelia I. Bargmann, and Alejandro

Aballay. 2008. "Innate Immunity in Caenorhabditis Elegans Is Regulated by Neurons Expressing NPR-1/GPCR." *Science*.

Sukul, N. C. and N. A. Croll. 1978. "Influence of Potential Difference and Current on the Electrotaxis of Caenorhaditis Elegans." *Journal of Nematology*.

Sulston, J., M. Dew, and S. Brenner. 1975. "Dopaminergic Neurons in the Nematode Caenorhabditis Elegans." *Journal of Comparative Neurology* 163(2):215–26.

Sulston, J. E., D. G. Albertson, and J. N. Thomson. 1980. "The Caenorhabditis Elegans Male: Postembryonic Development of Nongonadal Structures." *Developmental Biology*.

Sulston, J. E. and H. R. Horvitz. 1977. "Post-Embryonic Cell Lineages of the Nematode, Caenorhabditis Elegans." *Developmental Biology*.

Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thomson. 1983. "The Embryonic Cell Lineage of the Nematode Caenorhabditis Elegans." *Developmental Biology*.

Sun, Jingru, Varsha Singh, Rie Kajino-Sakamoto, and Alejandro Aballay. 2011. "Neuronal GPCR Controls Innate Immunity by Regulating Noncanonical Unfolded Protein Response Genes." *Science.*

Swoboda, Peter, Haskell T. Adler, and James H. Thomas. 2000a. "The RFX-Type Transcription Factor DAF-19 Regulates Sensory Neuron Cilium Formation in C. Elegans." *Molecular Cell* 5(3):411–21.

Swoboda, Peter, Haskell T. Adler, and James H. Thomas. 2000b. "The RFX-Type Transcription Factor DAF-19 Regulates Sensory Neuron Cilium Formation in C. Elegans." *Molecular Cell*.

Tabara, Hiroaki, Erbay Yigit, Haruhiko Siomi, and Craig C. Mello. 2002. "The DsRNA Binding Protein RDE-4 Interacts with RDE-1, DCR-1, and a DExH-Box Helicase to Direct RNAi in C. Elegans." *Cell*.

Tam, Lai Wa, Nedra F. Wilson, and Paul A. Lefebvre. 2007. "A CDK-Related Kinase Regulates the Length and Assembly of Flagella in Chlamydomonas." *Journal of Cell Biology*.

Tanis, Jessica E., Zhongming Ma, Predrag Krajacic, Liping He, Kevin J. Foskett, and Todd Lamitina. 2013. "CLHM-1 Is a Functionally Conserved and Conditionally Toxic Ca2+-Permeable Ion Channel in Caenorhabditis Elegans." *Journal of Neuroscience.*

Taschner, Michael and Esben Lorentzen. 2016. "The Intraflagellar Transport Machinery." *Cold Spring Harbor Perspectives in Biology*.

Taylor, Seth R., Gabriel Santpere, Molly Reilly, Lori

Glenwinkel, Abigail Poff, Rebecca McWhirter, Chuan Xu, Alexis Weinreb, Manasa Basavaraju, Steven J. Cook, Alec Barrett, Alexander Abrams, Berta Vidal, Cyril Cros, Ibnul Rafi, Nenad Sestan, Marc Hammarlund, Oliver Hobert, and David M. Miller. 2019. "Expression Profiling of the Mature <Em>C. Elegans</Em> Nervous System by Single-Cell RNA-Sequencing." *BioRxiv* 737577.

Thomas-Chollier, Morgane, Carl Herrmann, Matthieu Defrance, Olivier Sand, Denis Thieffry, and Jacques Van Helden. 2012. "RSAT Peak-Motifs: Motif Analysis in Full-Size ChIP-Seq Datasets." *Nucleic Acids Research*.

Thomas, Joëlle, Laurette Morlé, Fabien Soulavie, Anne Laurençon, Sébastien Sagnol, and Bénédicte Durand. 2010. "Transcriptional Control of Genes Involved in Ciliogenesis: A First Step in Making Cilia." *Biology of the Cell*.

Tobin, David M., David M. Madsen, Amanda Kahn-Kirby, Erin L. Peckol, Gary Moulder, Robert Barstead, Andres V. Maricq, and Cornelia I. Bargmann. 2002. "Combinatorial Expression of TRPV Channel Proteins Defines Their Sensory Functions and Subcellular Localization in C. Elegans Neurons." *Neuron*.

Tobin, Jonathan L. and Philip L. Beales. 2009. "The Nonmotile Ciliopathies." *Genetics in Medicine*.

Tran, Ngoc Tam L. and Chun Hsi Huang. 2014. "A Survey of Motif Finding Web Tools for Detecting Binding Site Motifs in ChIP-Seq Data." *Biology Direct.*

Tran, Ngoc Tam L. and Chun Hsi Huang. 2018. "MODSIDE: A Motif Discovery Pipeline and Similarity Detector." *BMC Genomics*.

Troemel, Emily R., Bruce E. Kimmel, and Cornelia I. Bargmann. 1997. "Reprogramming Chemotaxis Responses: Sensory Neurons Define Olfactory Preferences in C. Elegans." *Cell* 91(2):161–69.

Troemel, Emily R., Alvaro Sagasti, and Cornelia I. Bargmann. 1999. "Lateral Signaling Mediated by Axon Contact and Calcium Entry Regulates Asymmetric Odorant Receptor Expression in C. Elegans." *Cell*.

Turatsinze, Jean Valery, Morgane Thomas-Chollier, Matthieu Defrance, and Jacques van Helden. 2008. "Using RSAT to Scan Genome Sequences for Transcription Factor Binding Sites and Cis-Regulatory Modules." *Nature Protocols*.

Uchida, Okiko, Hiroyuki Nakano, Makoto Koga, and Yasumi Ohshima. 2003. "The C. Elegans Che-1 Gene Encodes a Zinc Finger Transcription Factor Required for Specification of the ASE Chemosensory Neurons." *Development.*

Uzbekov, Rustem and Irina Alieva. 2018. "Who Are You,

Subdistal Appendages of Centriole?" *Royal Society Open Science*.

VanHoven, Miri K., Sarah L. Bauer Huang, Stephanie D. Albin, and Cornelia I. Bargmann. 2006. "The Claudin Superfamily Protein NSY-4 Biases Lateral Signaling to Generate Left-Right Asymmetry in C. Elegans Olfactory Neurons." *Neuron*.

Venkatachalam, Kartik, Junjie Luo, and Craig Montell. 2014. "Evolutionarily Conserved, Multitasking TRP Channels: Lessons from Worms and Flies." *Handbook of Experimental Pharmacology*.

Vidal-Gadea, Andres, Chance Bainbridge, Ben Clites, Bridgitte E. Palacios, Layla Bakhtiari, Vernita Gordon, and Jonathan Pierce-Shimomura. 2018. "Response to Comment on 'Magnetosensitive Neurons Mediate Geomagnetic Orientation in Caenorhabditis Elegans.'" *ELife*.

Vidal-Gadea, Andrés, Kristi Ward, Celia Beron, Navid Ghorashian, Sertan Gokce, Joshua Russell, Nicholas Truong, Adhishri Parikh, Otilia Gadea, Adela Ben-Yakar, and Jonathan Pierce-Shimomura. 2015. "Magnetosensitive Neurons Mediate Geomagnetic Orientation in Caenorhabditis Elegans." *ELife*.

Van Voorhies, W. A. and S. Ward. 2000. "Broad Oxygen Tolerance in the Nematode Caenorhabditis Elegans." *Journal of Experimental Biology*.

Wagner, Günter P. 2015. "Two Rules for the Detection and Quantification of Epistasis and Other Interaction Effects." *Methods in Molecular Biology*.

Wang, Daniel Y. C., Sudhir Kumar, and S. Blair Hedges. 1999. "Divergence Time Estimates for the Early History of Animal Phyla and the Origin of Plants, Animals and Fungi." *Proceedings of the Royal Society B: Biological Sciences*.

Wang, Jinke, Jie Lu, Guangming Gu, and Yingxun Liu. 2011. "In Vitro DNA-Binding Profile of Transcription Factors: Methods and New Insights." *Journal of Endocrinology*.

Wang, Juan, Hillel T. Schwartz, and Maureen M. Barr. 2010. "Functional Specialization of Sensory Cilia by an RFX Transcription Factor Isoform." *Genetics*.

Wang, Lei and Brian D. Dynlacht. 2018. "The Regulation of Cilium Assembly and Disassembly in Development and Disease." *Development (Cambridge)*.

Wani, Khursheed A., Debanjan Goswamy, and Javier E. Irazoqui. 2020. "Nervous System Control of Intestinal Host Defense in C. Elegans." *Current Opinion in Neurobiology*.

Ward, S. 1973. "Chemotaxis by the Nematode Caenorhabditis Elegans: Identification of Attractants and Analysis of the Response by Use of Mutants." *Proceedings of the National Academy of Sciences of the United States of America* 70(3):817–21.

Ward, Samuel, Nichol Thomson, John G. White, and Sydney Brenner. 1975. "Electron Microscopical Reconstruction of the Anterior Sensory Anatomy of the Nematode Caenorhabditis Elegans." *Journal of Comparative Neurology*.

Ware, Randle W., David Clark, Kathryn Crossland, and Richard L. Russell. 1975. "The Nerve Ring of the Nematode Caenorhabditis Elegans: Sensory Input and Motor Output." *Journal of Comparative Neurology*.

Wei, Qing, Qingwen Xu, Yuxia Zhang, Yujie Li, Qing Zhang, Zeng Hu, Peter C. Harris, Vicente E. Torres, Kun Ling, and Jinghua Hu. 2013. "Transition Fibre Protein FBF1 Is Required for the Ciliary Entry of Assembled Intraflagellar Transport Complexes." *Nature Communications*.

Wei, Qing, Yuxia Zhang, Yujie Li, Qing Zhang, Kun Ling, and Jinghua Hu. 2012. "The BBSome Controls IFT Assembly and Turnaround in Cilia." *Nature Cell Biology*.

Weigel, Detlef, Gerd Jürgens, Frank Küttner, Eveline Seifert, and Herbert Jäckle. 1989. "The Homeotic Gene Fork Head Encodes a Nuclear Protein and Is Expressed in the Terminal Regions of the Drosophila Embryo." *Cell*.

Weirauch, Matthew T., Ally Yang, Mihai Albu, Atina G. Cote, Alejandro Montenegro-Montero, Philipp Drewe, Hamed S. Najafabadi, Samuel A. Lambert, Ishminder Mann, Kate Cook, Hong Zheng, Alejandra Goity, Harm van Bakel, Jean Claude Lozano, Mary Galli, Mathew G. Lewsey, Eryong Huang, Tuhin Mukherjee, Xiaoting Chen, John S. Reece-Hoyes, Sridhar Govindarajan, Gad Shaulsky, Albertha J. M. Walhout, François Yves Bouget, Gunnar Ratsch, Luis F. Larrondo, Joseph R. Ecker, and Timothy R. Hughes. 2014. "Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity." *Cell* 158(6):1431–43.

Wells, Kristen L., Mazhgan Rowneki, and Darrell J. Killian. 2015. "A Splice Acceptor Mutation in C. Elegans Daf-19/Rfx Disrupts Functional Specialization of Male-Specific Ciliated Neurons but Does Not Affect Ciliogenesis." *Gene*.

Wenick, Adam S. and Oliver Hobert. 2004. "Genomic Cis-Regulatory Architecture and Trans-Acting Regulators of a Single Interneuron-Specific Gene Battery in C. Elegans." *Developmental Cell*.

Wenzel, Dirk, Francesca Palladino, and Monika Jedrusik-Bode. 2011. "Epigenetics in C. Elegans: Facts and Challenges." *Genesis*.

Wes, P. D. and C. I. Bargmann. 2001. "C. Elegans Odour Discrimination Requires Asymmetric Diversity in Olfactory Neurons." *Nature*.

White, J.G., Southgate, E., Thomson, J.N. and Brenner, S. 1986. "The Structure of the Nervous System of the Nematode Caenorhabditis Elegans." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*.

Williams, Corey L., Chunmei Li, Katarzyna Kida, Peter N. Inglis, Swetha Mohan, Lucie Semenec, Nathan J. Bialas, Rachel M. Stupay, Nansheng Chen, Oliver E. Blacque, Bradley K. Yoder, and Michel R. Leroux. 2011. "MKS and NPHP Modules Cooperate to Establish Basal Body/Transition Zone Membrane Associations and Ciliary Gate Function during Ciliogenesis." *Journal of Cell Biology*.

Williams, Corey L., Marlene E. Winkelbauer, Jenny C. Schafer, Edward J. Michaud, and Bradley K. Yoder. 2008. "Functional Redundancy of the B9 Proteins and Nephrocystins in Caenorhabditis Elegans Ciliogenesis." *Molecular Biology of the Cell*.

Wingender, Edgar, Torsten Schoeps, and Jürgen Dönitz. 2013. "TFClass: An Expandable Hierarchical Classification of Human Transcription Factors." *Nucleic Acids Research*.

Winkelbauer, Marlene E., Jenny C. Schafer, Courtney J. Haycraft, Peter Swoboda, and Bradley K. Yoder. 2005. "The C. Elegans Homologs of Nephrocystin-1 and Nephrocystin-4 Are Cilia Transition Zone Proteins Involved in Chemosensory Perception." *Journal of Cell Science*.

Wojtyniak, Martin, Andrea G. Brear, Damien M. O'Halloran, and Piali Sengupta. 2013. "Cell- and Subunit-Specific Mechanisms of CNG Channel Ciliary Trafficking and Localization in C. Elega." *Journal of Cell Science*.

Xu, Qingwen, Yuxia Zhang, Qing Wei, Yan Huang, Yan Li, Kun Ling, and Jinghua Hu. 2015. "BBS4 and BBS5 Show Functional Redundancy in the BBSome to Regulate the Degradative Sorting of Ciliary Sensory Receptors." *Scientific Reports*.

Yang, Jichen and Stephen A. Ramsey. 2015. "A DNA Shape-Based Regulatory Score Improves Position-Weight Matrix-Based Recognition of Transcription Factor Binding Sites." *Bioinformatics*.

Yu, Hui, René F. Prétôt, Thomas R. Bürglin, and Paul W. Sternberg. 2003. "Distinct Roles of Transcription Factors EGL-46 and DAF-19 in Specifying the Functionally of a Polycystin-Expressing Sensory Neuron Necessary for C. Elegans Male Vulva Location Behavior." *Development*.

Yu, Sidney, Leon Avery, Eric Baude, and David L. Garbers. 1997. "Guanylyl Cyclase Expression in Specific Sensory Neurons: A New Family of Chemosensory Receptors." *Proceedings of the National Academy of Sciences of the United States of America*.

Yu, Xianwen, Doreen Lau, Chee Peng Ng, and Sudipto Roy. 2011. "Cilia-Driven Fluid Flow as an Epigenetic Cue for Otolith Biomineralization on Sensory Hair Cells of the Inner Ear." *Development*.

Yu, Xianwen, Chee Peng Ng, Hermann Habacher, and Sudipto Roy. 2008. "Foxj1 Transcription Factors Are Master Regulators of the Motile Ciliogenic Program." *Nature Genetics*.

Zaret, Kenneth S. and Jason S. Carroll. 2011. "Pioneer Transcription Factors: Establishing Competence for Gene Expression." *Genes and Development*.

Zarkower, David. 2006. "Somatic Sex Determination." *WormBook : The Online Review of C. Elegans Biology*.

El Zein, Loubna, Aouatef Ait-Lounis, Laurette Morlé, Joëlle Thomas, Brigitte Chhin, Nathalie Spassky, Walter Reith, and Bénédicte Durand. 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*.

Zhang, Chunmei, Ninghui Zhao, Yao Chen, Donghua Zhang, Jinyuan Yan, Wei Zou, Keqin Zhang, and Xiaowei Huang. 2016. "The Signaling Pathway of Caenorhabditis Elegans Mediates Chemotaxis Response to the Attractant 2-Heptanone in a Trojan Horse-like Pathogenesis." *Journal of Biological Chemistry* 291(45):23618–27.

Zhang, Feifan, Abhishek Bhattacharya, Jessica C. Nelson, Namiko Abe, Patricia Gordon, Carla Lloret-Fernandez, Miren Maicas, Nuria Flames, Richard S. Mann, Daniel A. Colón-Ramos, and Oliver Hobert. 2014. "The LIM and POU Homeobox Genes Ttx-3 and Unc-86 Act as Terminal Selectors in Distinct Cholinergic and Serotonergic Neuron Types." *Development (Cambridge)*.

Zhang, Mi, William R. Schafer, and Rainer Breitling. 2010. "A Circuit Model of the Temporal Pattern Generator of Caenorhabditis Egg-Laying Behavior." *BMC Systems Biology*.

Zheng, Shanqing, Hilton Chiu, Jeffrey Boudreau, Tony Papanicolaou, William Bendena, and Ian Chin-Sang. 2019. "A Functional Study of All 40 Caenorhabditis Elegans Insulin-like Peptides." *Journal of Biological Chemistry*.