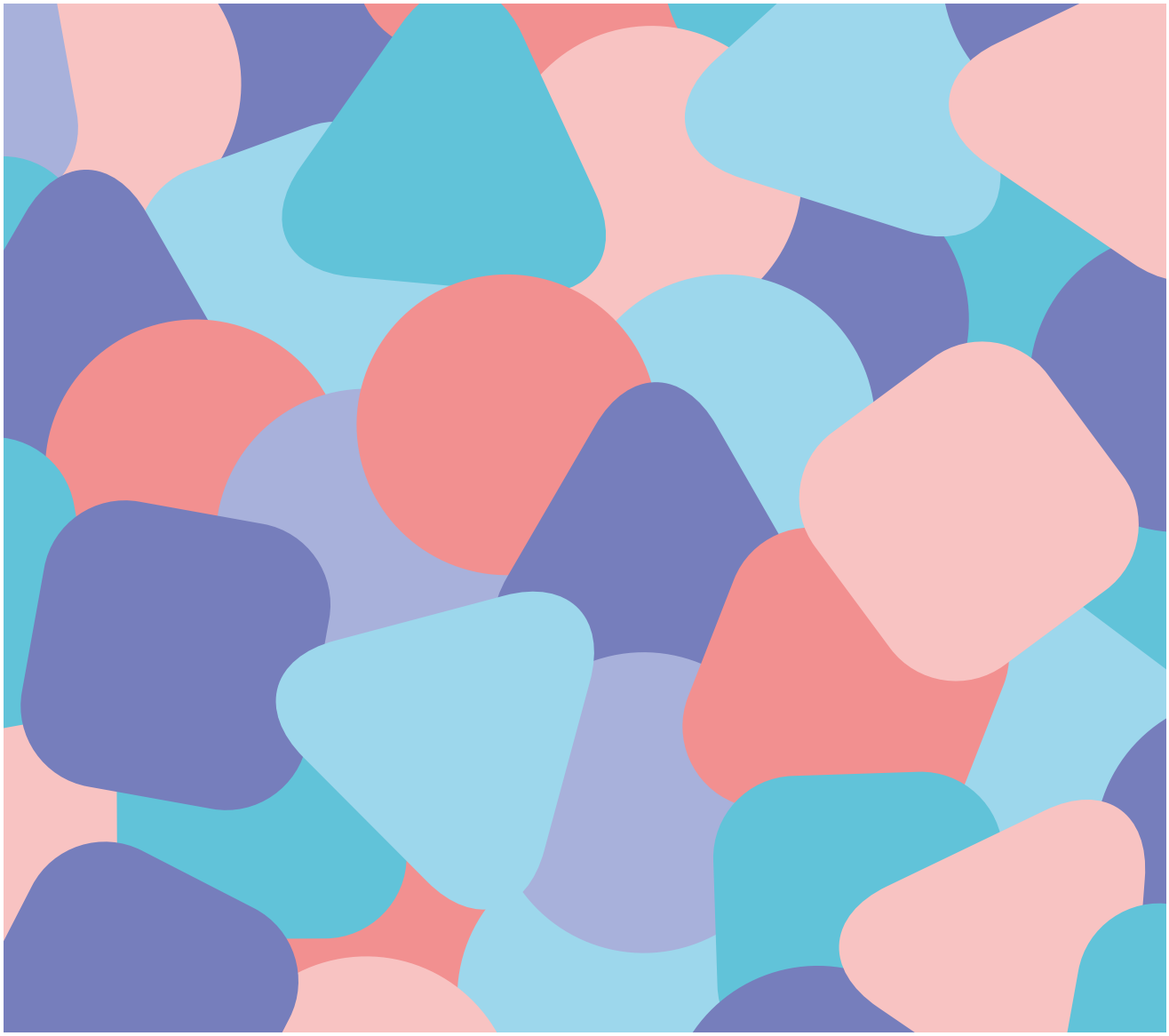
Serotonergic transcriptional regulatory logic in *Caenorhabditis elegans*

Doctoral Thesis by:
Carla Lloret Fernández
Thesis Supervisor:
Dr. Nuria Flames Bonilla

Doctoral Programme in Neurobiology, University of Valencia

Instituto de Biomedicina de Valencia (IBV) Consejo Superior de Investigaciones Científicas (CSIC)

Valencia, May 2017

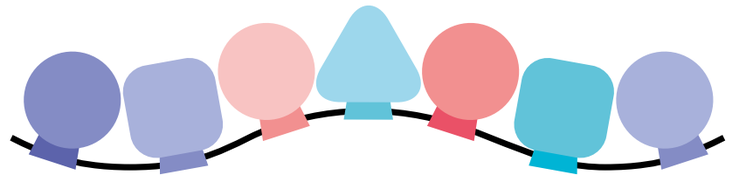


VNIVERSITAT DE VALÈNCIA

CSIC
Consejo Superior de Investigaciones Científicas

ibv INSTITUTO DE BIOMEDICINA DE VALENCIA CSIC

A mi familia



Nuria Flames Bonilla, Doctora en Ciencias Biológicas, Científico Titular y Directora de la Unidad de Neurobiología del Desarrollo del Instituto de Biomedicina de Valencia del CSIC,

CERTIFICA:

Que Carla Lloret Fernández, licenciada en Biotecnología por la Universidad Politécnica de Valencia, ha realizado bajo mi supervisión, la Tesis Doctoral titulada "Serotonergic transcriptional regulatory logic in *Caenorhabditis elegans*".

En Valencia, 29 de Mayo de 2017

Dra. Nuria Flames Bonilla

[EN]

This Thesis has been made possible thanks to a pre-doctoral fellowship from the VALi+d Programme, conferred by the Ministry of Education, Investigation, Culture and Sports of the Valencian Community (79/2013).

The investigation has been funded by the following investigation projects: 'Dissecting the gene regulatory mechanisms that generate serotonergic neurons and their link to mental disorders', European Research Council, Starting Grant; and 'Estudio de los mecanismos transcripcionales que regulan la diferenciación de las neurtonas monoaminérgicas y su conservación evolutiva', Ministry of Economy and Competitiveness, Spanish National Plan I+D, SAF2014-56877-R.

[ES]

Este trabajo de tesis ha sido posible gracias a una beca predoctoral del Programa VALi+d, otorgada por la Consellería de Educación, Investigación, Cultura y Deporte de la Comunidad Valenciana (Orden 79/2013).

La investigación ha sido financiada por los siguientes proyectos de investigación: "Dissecting the gene regulatory mechanisms that generate serotonergic neurons and their link to mental disorders", financiado por el European Research Council (Starting Grant); y "Estudio de los mecanismos transcripcionales que regulan la diferenciación de las neuronas monoaminérgicas y su conservación evolutiva", financiado por el Ministerio de Economia y Competitividad, Plan Nacional I+D, SAF2014-56877-R.

Serotonergic transcriptional regulatory logic in *Caenorhabditis elegans*

Abstract — English

Serotonergic transcriptional regulatory logic in *Caenorhabditis elegans*

Neuronal diversity in the nervous system is generated through the activation of multiple unique batteries of terminal differentiation genes, which determine the functional properties of the distinct mature neurons. It is generally accepted that transcription factors (TFs) bind in a combinatorial and cooperative manner to DNA sequences of the genome called enhancers, placing TFs as the main regulators of gene expression. However, how these combinations of TFs identify and activate their target sequences is poorly understood. In this work we use as a paradigm the serotonergic neurons to unravel the regulatory rules that select a cell type-specific transcriptome during terminal differentiation.

Serotonergic neurons are present in all eumetazoan groups and are universally defined by their ability to synthesise and release serotonin (5-HT), which is achieved by the expression of the '5-HT pathway genes'. Taking advantage of this phylogenetic conservation, we use the simple model organism *Caenorhabditis elegans* to dissect the transcriptional regulatory logic of serotonergic neurons. *C. elegans* hermaphrodites have three functionally different serotonergic subclasses: the HSN motorneuron, the ADF sensory neuron and the NSM neurosecretory motorneuron. All three neuron subtypes express the 5-HT pathway genes. Through an *in vivo cis*-regulatory analysis of these genes we have identified independent *cis*-regulatory modules (CRM) responsible for their expression in each serotonergic neuron subtype. This modular organisation suggests that different regulatory logics are employed in each neuron subclass to activate its terminal transcriptome. To deepen in our understanding of how cell type-specific transcriptional programmes are implemented we decided to focus the rest of our work on the best characterised serotonergic neuron subtype, the HSN neuron, and carried out an extensive dissection of HSN terminal differentiation transcriptional rules.

Loss of function mutant and *cis*-regulatory analyses reveal that direct activation of the HSN transcriptome is orchestrated by a code of six TFs, that we have termed HSN TF collective. This TF code is composed by AST-1 (ETS TF family), UNC-86 (POU TF family), SEM-4 (SPALT TF family), HLH-3 (bHLH TF family), EGL-46 (INSM TF family) and EGL-18 (GATA TF family). The expression of the HSN TF collective is sufficient to induce serotonergic fate in some specific contexts and is required throughout the life of the animal in order to maintain the identity of the HSN neuron.

Bioinformatically identified binding site clusters for the six TFs of the HSN TF collective are enriched in known HSN expressed genes compared to a random set of genes. Through *in vivo* reporter analysis, we demonstrate that this clustering constitutes a regulatory signature that is sufficient for *de novo* identification of HSN neuron functional enhancers. This regulatory signature contains certain syntactic constrains that further improve the prediction of enhancer expression in the cell.

Mouse orthologues of most members of the HSN TF collective are known regulators of the mammalian serotonergic differentiation programme. This homology in both serotonergic regulatory programmes allows for the identification of an additional candidate TF in the worm (PHA-4), orthologue to the mouse FOXA2, and a mouse TF (SALL2), orthologue of the worm SEM-4. Moreover, we prove that mouse orthologues can functionally substitute for their worm counterparts. Finally, Principal Coordinates Analysis suggests that, among *C. elegans* neurons, the HSN transcriptome most closely resembles that of mouse serotonergic neurons, which reveals deep homology.

Our results show that a regulatory signature based on a defined set of TFs is sufficient for enhancer identification using primary DNA sequence. Moreover, our results identify rules governing the transcriptional regulatory code of a critically important neuronal type in two species separated by over 700 million years.

Abstract — Spanish

Lógica de regulación transcripcional de las neuronas serotonérgicas en *Caenorhabditis elegans*

La diversidad del sistema nervioso se genera mediante la activación de múltiples baterías únicas de genes efectores, que definen las propiedades funcionales de los diferentes subtipos neuronales. Está bien establecido que los factores de transcripción (FT) se unen de una manera combinatoria y cooperativa a secuencias de ADN presentes en los elementos de regulación en *cis* del genoma, llamados potenciadores (*enhancers* en inglés). Esto otorga a los FT un papel central en la regulación de la expresión génica. Sin embargo, no se conocen los mecanismos por los que estas combinaciones de FT identifican y activan sus secuencias diana. En este trabajo se han utilizado las neuronas serotonérgicas como paradigma de investigación de las leyes que regulan la selección del transcriptoma de un tipo neuronal concreto durante la diferenciación terminal.

Las neuronas serotonérgicas se encuentran presentes en todos los grupos de eumetazoos y se definen por su habilidad de sintetizar y liberar serotonina (5-HT), lo cual es posible gracias a la expresión de los llamados 'genes de la vía de la 5-HT'. Aprovechando esta conservación filogenética, hemos utilizado el organismo modelo *Caenorhabditis elegans* para diseccionar la lógica de regulación transcripcional de las neuronas serotonérgicas. Los hermafroditas *C. elegans* contienen tres subclases de neuronas serotonérgicas con diferente función: la neurona motora HSN, la neurona secretora ADF y la neurona motora neurosecretora NSM. Mediante un análisis de regulación *in vivo* de los genes de la vía de la 5-HT, hemos identificado módulos de regulación en *cis* (MRC) independientes responsables de su expresión en cada uno de los tres subtipos serotonérgicos. Esta organización modular sugiere que cada subclase utiliza una lógica de regulación diferente. Para profundizar en los mecanismos de selección y activación del transcriptoma específico de un tipo neuronal decidimos enfocar el resto de nuestro trabajo en el estudio de la neurona HSN, por ser la mejor caracterizada hasta la fecha.

El análisis de mutantes de pérdida de función, junto con el estudio detallado de los MRC de la neurona HSN, revelan que un código de seis FT es capaz de activar directamente el transcriptoma de la neurona HSN. Este código, al que hemos llamado 'Colectivo de FT de HSN', está formado por AST-1 (de la familia de FT ETS), UNC-86 (POU), SEM-4 (SPALT), HLH-3 (bHLH), EGL-46 (INSM) y EGL-18 (GATA). Esta combinación, es suficiente, en algunos contextos celulares para la inducción del fenotipo serotonérgico y necesario durante toda la vida del animal para mantener la identidad de la neurona HSN.

Por otro lado, estudios bioinformáticos de predicción de sitios de unión para los seis FT del código, muestran que los genes expresados en la neurona HSN están enriquecidos en la presencia de agrupaciones de estos seis sitios de unión, en comparación a un conjunto de genes elegidos al azar. Mediante el análisis de reporteros *in vivo,* demostramos que esta agrupación constituye una huella reguladora que es suficiente para la identificación de nuevos potenciadores funcionales para la neurona HSN. Además, esta huella reguladora contiene normas sintácticas que mejoran la predicción de potenciadores expresados en la célula.

Curiosamente, el programa de diferenciación de las neuronas serotonérgicas en ratón está controlado por FT que son ortólogos a los del nematodo. Esta elevada homología en la regulación nos ha permitido identificar nuevos candidatos a regular las neuronas serotonérgicas del gusano (PHA-4, ortólogo a FOXA2) y del ratón (SALL2, ortólogo a SEM-4). Asimismo, los ortólogos de ratón son capaces de sustituir funcionalmente a los FT equivalentes en gusano. Finalmente, el Análisis de Coordenadas Principales sugiere que, de entre todas las neuronas del gusano, el transcriptoma de la neurona HSN es el que más se asemeja a aquel de las neuronas serotonérgicas de ratón, revelando relaciones de homología profunda.

En conclusión, hemos demostrado que la presencia de una huella reguladora basada en un conjunto definido de FT es suficiente para identificar potenciadores, utilizando únicamente la secuencia primaria de ADN. Además, hemos identificado las reglas que gobiernan el código de regulación transcripcional de un tipo neuronal relevante en dos especies separadas hace más de 700 millones de años.

Contents

Index of figures and tables

Introduction

Introduction

The nervous system is the most complex tissue, regarding cell diversity and structure. At the functional level it controls a wide repertoire of behaviours; from basic behaviours like appetite, breathing or arousal, to more sophisticated behaviours such as learning and memory. In order to understand how the brain functions, it is essential to learn the principles of how the nervous system develops: how neurons are made, how they differentiate into distinct and specialised neuron types and how they are assembled into circuits that produce behaviours. According to recent estimates there are 86-100 billion neurons in the human brain (Muotri & Gage 2006; Azevedo et al. 2009), with a total of 10^{15} synapses, an average of 5,000-20,000 synapses per neuron in the neocortex (Pakkenberg et al. 2003; Milo et al. 2010) and about 10,000 different neuronal types (Muotri & Gage 2006). The first descriptions of nerve cells were attributed to Ehrenberg and Purkinje during the mid-1830s and in the early 1870s Camillo Golgi discovered a revolutionary way to stain the nervous system (reviewed in (López-Muñoz et al. 2006)). Ramón y Cajal was the first to postulate that the functional units of the nervous system were discrete cells that were beautifully represented in his drawings. His work established the notion that brain function could be understood by studying its component cells (Ramón y Cajal 1909). This, together with the discovery that nerve cells communicate with each other through the release of chemical neurotransmitters at specialised sites (Dale 1954; Dale 1934; Loewi 1954), marked the beginning of modern neuroscience. An additional revolution in the neuroscience field has taken place recently, with the advent of deep sequencing technologies and single cell transcriptomics. In the light of molecular neurobiology and transcriptome analysis of different neuron types, we now know that the morphological and functional diversity of neuronal cell types is a reflection of the astounding degree of diversity in their molecular composition (Hawrylycz et al. 2012; Darmanis et al. 2015). On top of this, a higher level of complexity appears when neuronal connections organise into functional circuits constituting the brain connectome (Sporns 2011).

Although research in the past decades has shed light on the basic principles of nervous system development, less is known about the molecular mechanisms that control neuron subtype specification and how those identities are maintained throughout the life of the animal.

In this Thesis we are interested in understanding how particular neuronal types are specified at the terminal level, with a special focus on how the cell can decode the information of the regulatory genome to select the complement of genes that are required for its function. As a paradigm we use the serotonergic neurons to unravel the regulatory rules that select a cell type-specific transcriptome during terminal differentiation. The Introduction has been divided in three main parts: the first part deals with neuron terminal specification from a molecular and transcriptional point of view, the second one focuses on the serotonergic system, providing an update on serotonergic terminal differentiation in mammals, and the third proposes the nematode *Caenorhabditis elegans* as an optimal model to study serotonergic terminal differentiation.

Part I

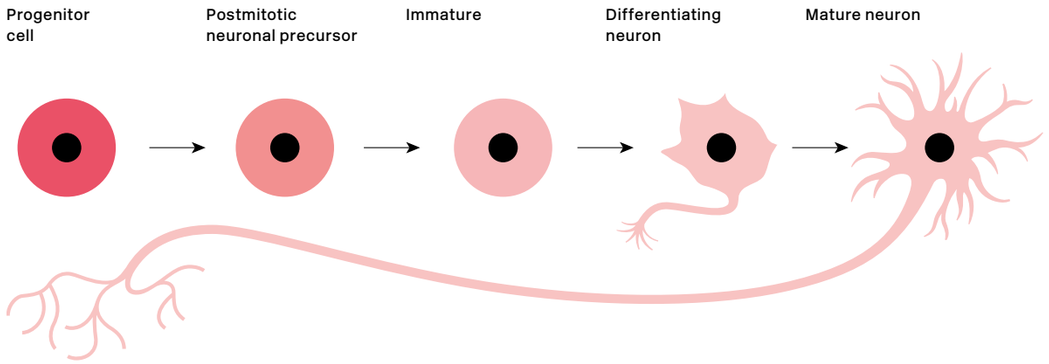
Neuronal cell fate specification

In order to generate a neuron, a cell or lineage generally will undergo the following steps: choose a neural versus non-neural fate, commit to a specific neuronal lineage, mature and terminally differentiate into a specific neuron class or subclass (Hobert 2005). Next, we will describe the developmental states needed until terminal differentiation is achieved.

In vertebrates, the earliest step in the generation of the nervous system is the acquisition of a neuronal fate by a specific group of ectodermal cells of the embryo that generate a structure termed neuronal plate (neuronal precursors), which will generate the entire nervous system (reviewed in (Stern 2006)) → **Figure 1.1**. Through temporal and concentration gradients of signalling molecules, known as morphogens, the expression of patterning transcription factors (TFs) is induced, which mainly belong to the homeodomain (HD) TF family. Through mutual cross-inhibition, transcriptional domains are established, which define the progenitor cells that will give birth to specific neuron types (Briscoe et al. 1999; Ericson et al. 1997; Wilson & Maden 2005). Proneural genes are activated, which promote the generation of progenitors that are committed to neuronal differentiation. Proneural genes mainly belong to the basic-helix-loop-helix (bHLH) TF family and, via Notch signalling, inhibit their own expression in adjacent cells in a process called lateral inhibition (Bertrand et al. 2002). In vertebrates, mutations in the Asc TF *Ascl1* (also known as *Mash1*) and Ato related *Neurogenin 2* (*Ngn2*) result in loss of neuronal progenitors and premature generation of astrocytic progenitors (Bertrand et al. 2002; Gómez-Skarmeta et al. 2003). This observation suggests that, consistent with a proneural role, *Ngn2* and *Ascl1* play a role in both the commitment of multipotent progenitors to a neuronal fate and in the inhibition of glial fate.

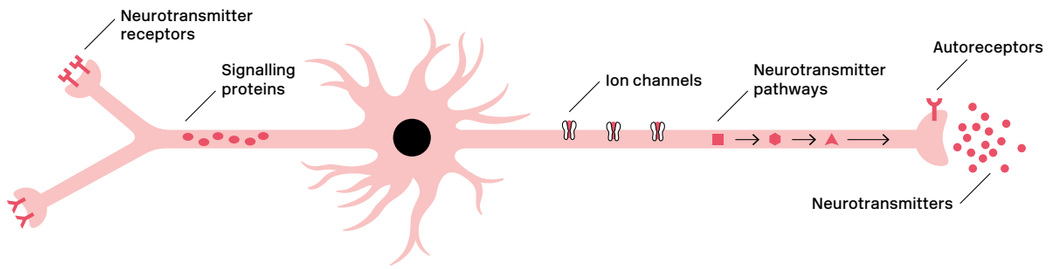
In order to acquire their highly specialised features, neurons must proceed through two final stages of neuronal development: maturation and differentiation. During maturation newly born neurons undergo axonal and dendritic morphogenesis, synaptogenesis and synapse elimination thereby assembling into functional circuits. This process occurs simultaneously with differentiation, in which neurons come to express the battery of genes required for their mature function in the circuit → **Figure 1.2**. This set of genes, known as effector genes, are expressed throughout the life of an adult differentiated neuron and confer the unique identity to a neuron (Hobert 2016); in other words, the 'molecular identity' of a neuron or its 'molecular signature'. The composition of the neuron type-specific gene batteries is combinatorial, meaning that individual neuron types do not uniquely express exclusive gene products, but it is rather the unique combination of genes that are more broadly expressed what defines a neuron type-specific gene battery (Wenick & Hobert 2004). This so called 'combinatorial coding' can support

Figure 1.1
Stages of neuronal development



Progenitor cells first acquire the neuronal fate through the action of morphogens and signalling molecules. Next, they exit the cell cycle to become postmitotic neuronal precursors that do not yet possess neuron type identity. Identity is acquired after neuronal differentiation through coordinate expression of neuron type gene batteries. The newborn neurons remain phenotypically immature until they accomplish the maturation process that includes cell body migration, dendritic growth, expression of neurotransmitter pathways, formation of axonal pathways and synaptogenesis. At this point, neurons become integrated into neural circuitry and acquire functional properties.

Figure 1.2
Terminal features of a mature neuron

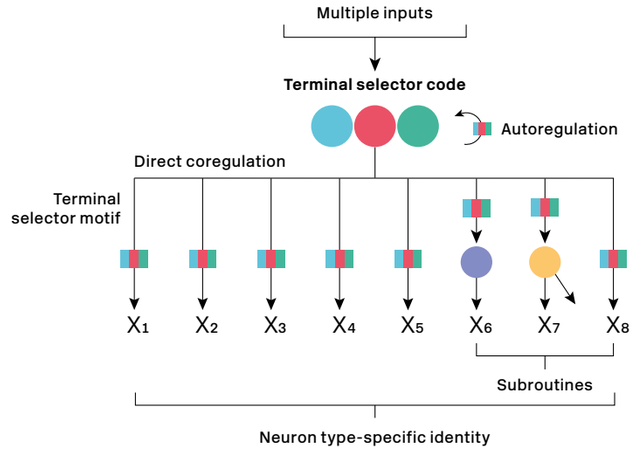


Terminal selectors directly control the expression of terminal differentiation genes (also known as effector genes), of which representative examples are depicted here. Adapted from (Hobert 2011).

Figure 1.3
Key features of gene regulatory logic in mature neuron types

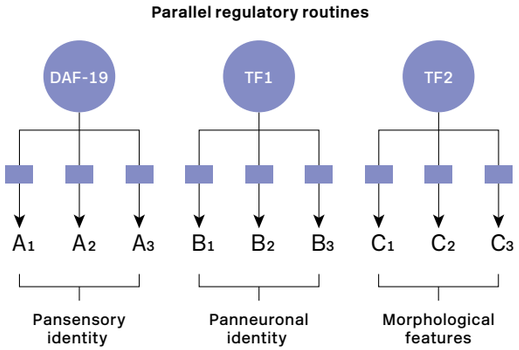
A) Unique neuronal identity

Terminal selector codes co-regulate a battery of terminal identity features (subtype-specific transcriptome) of a given neuron type. They do it through recognition of the terminal selector motif (blue/red/green rectangle). TFs appear as coloured circles. Sustained expression of terminal selectors is often, but not always, ensured by autoregulation. Xn are effector genes. Sometimes, terminal selector codes further activate additional TFs that control other subroutines (purple and yellow circles).



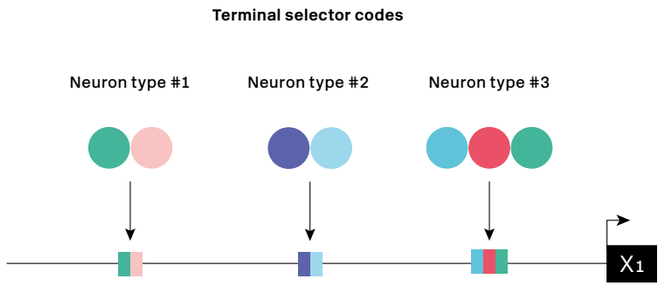
B) Identity shared with other neuron types

Parallel regulatory routines such as those that regulate pansensory features (via DAF-19) or panneuronal features (via TF1 in the scheme) are controlled by factors that could also be considered terminal selectors, with the difference that they do not assign unique identities. For example, morphology regulators (TF2) control generic aspects of the morphology of a neuron, such as placement of axons/dendrites into specific fascicles or axo/dendritic polarity.



C) Modular organisation of terminal effector genes

Modular organisation of the regulatory sequences of effector genes: X1 represents a hypothetical effector gene expressed in more than one neuron type. It contains different *cis*-regulatory modules activated by the different combination of terminal selectors that are active in each cell type. This schematic reflects a key principle of combinatorial 're-usage' of the same terminal selector in different neuron types (green circles). Adapted from (Hobert 2016).



the construction of an almost infinite number of different neuron type-specific expression patterns, and thus almost infinite number of neuron types. The problem of neuronal differentiation and neuron type specification is therefore to be reframed as how these neuron type-specific gene expression programmes are executed.

Terminal selector codes regulate terminal differentiation and neuron type-specific gene expression

A prevailing model for how neurons acquire a postmitotic identity is the 'terminal selector hypothesis', proposed by Dr. Oliver Hobert (Hobert 2008). A few decades ago, Dr. Garcia-Bellido coined the term 'selector gene' for genes that define the identity of specific domains of a developing organism and that act transiently during specific phases of development (Garcia-Bellido 1975). Building on this concept, the term 'terminal selector' was later proposed for TFs that are activated around the time of the final mitosis or in early postmitotic neurons and that directly control the terminal identity of individual cell types in the nervous system (Hobert 2008) → **Figure 1.3-A**. Terminal selectors act via recognition and direct binding to specific regulatory regions common to all effector genes of the cell, named 'terminal selector motifs' (Heinz et al. 2015; Hobert 2016). It is believed that the expression of these terminal selectors is initially activated by proneural genes, transient regulatory factors and signalling cues such as the Wnt pathway (Bertrand & Hobert 2009; Hobert 2016).

The terminal selector concept was first described in the nematode *Caenorhabditis elegans*. Examples exist in the worm where a unique terminal selector controls the terminal differentiation programme of a specific neuronal subtype. Such is the case of the COE-type Zn-finger TF UNC-3 and the cholinergic command interneurons of the worm (Kratsios et al. 2011; Pereira et al. 2015). However, much more common is the observation of combinatorial activity of several TFs that constitute what has been termed 'terminal selector codes'. For example, in the dopaminergic neurons of the worm, AST-1 (ETS TF) acts in combination with the CEH-43 (Dlx TF) and CEH-20/CEH-40 (Pbx TFs) to directly regulate the terminal fate of the four dopaminergic subtypes (Flames & Hobert 2011; Doitsidou et al. 2013). In the glutamatergic system, thirteen different TFs act in distinct combinations in the twenty-five different glutamatergic neuron classes to initiate and maintain the expression of *eat-4* (vesicular glutamate transporter *Vglut*), the key defining feature of glutamatergic neurons (Serrano-Saiz et al. 2013). Many more examples have been described in the worm and are listed in → **Table 1.1**.

Although terminal selectors act mostly through gene activation, repressive mechanisms have also been described to be important to achieve proper cell type-specific profiles. For example, in *C. elegans* ventral cord motorneurons there is a subset of these neurons that are directly activated by the common terminal selector UNC-3 (Kratsios et al. 2011; Kratsios et al. 2015). However, this motorneurons regulated by UNC-3 belong to different subclasses. Kerk et al. have recently shown that their diversification is controlled by distinct combinations of class-specific transcriptional repressors. Furthermore, these repressors are continuously required in postmitotic neurons to prevent UNC-3 from activating class-specific effector genes in specific motorneuron subsets and they do it via discrete *cis*-regulatory elements. This works proposes that antagonising the activity of broadly acting terminal selectors of neuron identity in a subtype-specific fashion may constitute a general principle of neuron subtype diversification (Kerk et al. 2017). In vertebrates, although not studied in such depth, repressor TFs are also relevant for neuron subtype specification (William et al. 2003; Muhr et al. 2001).

Table 1.1
Terminal selectors of *Caenorhabditis elegans* neurons

Known regulators of most *C. elegans* sensory, inter- and motorneurons have been included. These genes code for TFs that show key features of terminal selectors and are expressed in mature neurons throughout their lifetime, likely a reflection of their continuous role in maintaining the differentiated, terminal state. Early or transiently acting regulators are not included. 'NT': neurotransmitter. '(–)': unknown neurotransmitter used. '/': redundant factors. Adapted from (Hobert, 2016).

Neuron	NT	Type	Terminal Selectors	Neuron	NT	Type	Terminal Selectors
ADE	DA	sensory	*ast-1, ceh-43, ceh-20/ceh-40*	IL2	Ach	sensory	*unc-86, cfi-1*
ADL	Glu	sensory	*lin-11*	OLL	Glu	sensory	*sox-2, vab-3*
AFD	Glu	sensory	*ttx-1, ceh-14*	PDA	Ach	motor	*unc-3*
AIA	Ach	inter	*ttx-3*	PDB	Ach	motor	*unc-3*
AIB	Glu	inter	*unc-42*	PDE	DA	sensory	*ceh-43, ceh-20/ceh-40*
AIM	Glu, 5-HT	inter	*unc-86, ceh-14*	PHA	Glu	sensory	*ceh-14*
AIY	Ach	inter	*ttx-3, ceh-10*	PHB	Glu	sensory	*ceh-14*
AIZ	Glu	inter	*unc-86, ceh-14*	PHC	Glu	sensory	*ceh-14*
ALA	GABA	sensory	*ceh-14, ceh-17*	PLM	Glu	sensory	*unc-86, mec-3*
ALM	Glu	sensory	*unc-86, mec-3*	PQR	Glu	sensory	*unc-86, egl-13, ahr-1*
AQR	Glu	sensory	*unc-86, egl-13, ahr-1*	PVC	Ach	inter	*ceh-14, cfi-1*
AS	Ach	motor	*unc-3*	PVD	Glu	sensory	*unc-86, mec-3*
ASE	Glu	sensory	*che-1*	PVM	(–)	sensory	*unc-86, mec-3*
ASG	Glu	sensory	*lin-11, ceh-37*	PVN	Ach	inter	*unc-3, ceh-14*
ASH	Glu	sensory	*unc-42*	PVP	Ach	inter	*lin-11, unc-30*
ASJ	Ach	sensory	*sptf-1*	PVQ	Glu	inter	*pag-3, zag-1*
ASK	Glu	sensory	*ttx-3*	PVR	Glu	inter	*unc-86, ceh-14*
AUA	Glu	inter	*ceh-6*	RID	(–)	inter	*lim-4*
AVG	Ach	inter	*lin-11, ast-1*	RIH	Ach, 5-HT	inter	*unc-86*
AVK	(–)	inter	*unc-42, fax-1*	RIS	GABA	inter	*nhr-67, lim-6*
AVL	GABA	motor	*nhr-67, lim-6*	RIV	Ach	inter	*unc-42*
AVM	Glu	sensory	*unc-86, mec-3*	RMD	Ach	motor	*unc-42*
AWA	(–)	sensory	*odr-7*	RME	GABA	motor	*nhr-67, ceh-10, tab-1*
AWB	Ach	sensory	*lim-4, sox-2*	SAA	Ach	inter	*sox-3*
AWC	Glu	sensory	*ceh-36, sox-2*	SAB	Ach	motor	*unc-3*
BAG	Glu	sensory	*ets-5, ceh-37, egl-13, egl-46*	SMB	Ach	motor	*lim-4*
BDU	(–)	inter	*pag-3, zag-1*	SMD	Ach	motor	*unc-42*
CEP	DA	sensory	*ast-1, ceh-43, ceh-20/ceh-40*	URA	Ach	sensory	*unc-86, cfi-1*
DA	Ach	motor	*unc-3*	URB	Ach	sensory	*unc-86*
DB	Ach	motor	*unc-3*	URX	Ach	sensory	*unc-86, ahr-1, egl-13*
DD	GABA	motor	*unc-30*	URY	Glu	sensory	*vab-3*
DVC	Glu	inter	*ceh-14*	VA	Ach	motor	*unc-3*
FLP	Glu	sensory	*unc-86, mec-3*	VB	Ach	motor	*unc-3*
HSN	Ach, 5-HT	motor	*unc-86, sem-4*	VD	GABA	motor	*unc-30*
IL1	Glu	sensory	*sox-2, vab-3*				

Terminal selectors are not only required to induce specific differentiation programmes, but also in some cellular contexts are sufficient to do so (i.e. gain-of-function mutants for the terminal selectors result in the activation of the effector genes in other cells) (Gordon & Hobert 2015; Flames & Hobert 2009). This context dependency is likely dictated by the need of proper cofactors (Gordon & Hobert 2015), but also by a chromatin environment that may be refractory to terminal selector activity (Tursun et al. 2011; Patel & Hobert 2017).

The terminal selector concept entails that an individual neuron type may not require a large number of TFs to regulate different terminal features or regulatory routines. Rather, functionally unrelated effector genes of a cell type (cell-specific transcriptomes) appear to be co-regulated through one common terminal selector or, as described, a combination of terminal selectors (Xue et al. 1993; Wenick & Hobert 2004; Doitsidou et al. 2013). Parallel regulatory routines (partial cell transcriptomes), such as panneuronal (Stefanakis et al. 2015) or pansensory identity (Swoboda et al. 2000) may be regulated by parallel-acting terminal selector combinations (Hobert 2011) → **Figure 1.3-B**.

Additionally, another quality of terminal selectors is that, as they act in combinations, individual members can be re-used in distinct neuron types constituting different codes that select different terminal programmes → **Figure 1.3-C**. This combinatorial action reveals fundamental to TF function as it increases the number of roles that a given TF can play, it increases the sequence-specificity and diversity of DNA-binding, and enhances the signal to noise ratio of gene regulation.

Remarkably, terminal differentiation genes are often continuously expressed throughout the life of a neuron and it has been observed that they are required to actively maintain the active transcriptome of a mature neuron (Deneris & Hobert 2014).

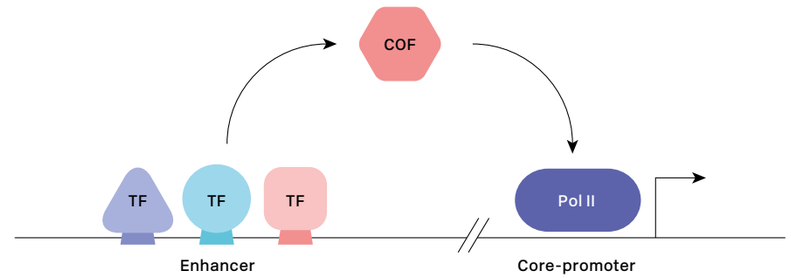
Although originally described in the invertebrate nematode *C. elegans* (Hobert 2008), the terminal selector concept is common to more complex organisms such as *Drosophila* and vertebrates. In rodents, PET1 (ETS TF) exhibits key features of a terminal selector for the serotonergic raphe neurons (Hendricks et al. 1999; Hendricks et al. 2003), whereas NURR1 (orphan nuclear hormone receptor TF) and PITX3 (homeodomain TF) have been described to regulate the expression of the dopaminergic gene battery in the mesencephalic dopaminergic neurons (Jacobs, van der Linden et al. 2009; Jacobs, van Erp et al. 2009). Moreover, the combination of NGN2, ISL1 and LHX3 are sufficient to reprogramme mouse embryonic stem cells into functional spinal motorneurons (Mazzoni et al. 2013). Furthermore, in some cases nematode and mammalian homologs factors have been shown to have similar roles on neuron type specification, as is the case for AST-1/ER81 regarding dopaminergic fate. *ast-1* codes for a *Pea3/Er81*-like ETS TF in the worm whose expression is maintained throughout the animal's life and that directly regulates the expression of the dopaminergic pathway genes, whereas its mouse homolog, *Er81/Etv1*, also appears to control at least some aspects of the dopaminergic fate in olfactory bulb neurons (Flames & Hobert 2009). Similarly, the COE (Collier/Olf/EBF)-type Zn-finger factor UNC-3 controls cholinergic identity of most motorneuron classes in the ventral nerve cord of *C. elegans* and a *Ciona intestinalis* COE factor is also required and sufficient for inducing cholinergic fate (Kratsios et al. 2011; Kratsios et al. 2015). These studies point to a possible phylogenetical conservation in the terminal selector regulation of some neuronal subtypes.

Cis-regulatory logic of gene expression

According to the terminal selector hypothesis, transcriptional regulation of neuron type-specific effector genes is mediated by particular combi-

Figure 1.4
Transcriptional regulation and its main players

Enhancers contain short sequence motifs that can be recognised by TFs. TFs, in turn, recruit transcriptional cofactors (COF) that recruit and activate RNA polymerase II (Pol II) at core-promoters (short sequences surrounding the transcriptional start site) to enable transcription. Adapted from (Reiter et al. 2017).



nations of TFs. In line with this idea, it is generally accepted that TFs bind in a combinatorial manner to DNA sequences of genomic *cis*-regulatory elements called enhancers (Banerji et al. 1981), placing TFs as the main regulators of gene expression, either activating or repressing it. TFs act as adaptor molecules that recognise the basic building blocks of regulatory sequences, the TF binding sites (TFBSs), thus essentially reading the regulatory information contained in the enhancer sequence. TF binding to enhancers prompts the recruitment of the transcription machinery to the core promoter (generally spanning ~40 bp upstream and downstream of the transcription start site), resulting in transcription initiation and the formation of robust expression patterns (Ptashne & Gann 1997) → **Figure 1.4**.

Despite the extensive research on transcriptional regulation, how TFs identify their target sequences and achieve combinatorial enhancer control remains a central question in biology. Here, we will present the current understanding on regulatory enhancer function, recently reviewed elsewhere (Levo & Segal 2014; Reiter et al. 2017; Spitz and Furlong 2012).

Enhancers are typically a few hundred base pairs in size and regulate the location, timing and levels of gene transcription. They can be located in non-coding sequences including introns and, in less frequency, in coding exons (Birnbaum et al. 2014) and can regulate their target gene or genes both in the same chromosome and in different chromosomes (Sanyal et al. 2012). Nucleotide (nt) variation in enhancers has been shown to lead to a multitude of phenotypes, including morphological differences between species (Carroll 2005) and human disease. In fact, most genetic associations to disease are located in non-coding sequences that are thought to be regulatory sequences (Mathelier et al. 2015).

In the past decade, development of several high-throughput methods have enabled the characterisation of genome wide TFBSs and active enhancers *in vitro* and *in vivo*. Methods such as chromatin immunoprecipitation followed by microarray (ChIP–chip) (Harbison et al. 2004; Venters et al. 2011) or high-throughput sequencing (ChIP–seq) (Arvey et al. 2012), its variation using exonuclease trimming (ChIP-exo) (Rhee & Pugh 2011), DNase I hypersensitive site sequencing (DNase-seq) or ATAC-seq to measure accessible chromatin, have measured the occupancy of sites along

the genome, and this has been used to delineate TF binding preferences. ChIP profiling of chromatin marks has also been used to trace the active regulatory landscape of specific cell types as histone H3K27ac correlates with active enhancers (Creyghton et al. 2010; Rada-Iglesias et al. 2011). Complementary to the ChIP experiments, *in vitro* affinity measurements of chosen TFs to many short sequences have been used to predict potential binding events genome wide. For example, methodologies like protein binding microarrays (PBMs) (Berger & Bulyk 2009), high-throughput systematic evolution of ligands by exponential enrichment (HT-SELEX) (Jolma et al. 2013), mechanically induced trapping of molecular interactions (MITOMI) (Maerkl & Quake 2007) and high-throughput sequencing–fluorescent ligand interaction profiling (HiTS-FLIP) (Nutiu et al. 2011) have examined the binding of hundreds of TFs from various organisms including yeast (Badis et al. 2008), *C. elegans* (Grove et al. 2009), mice (Badis et al. 2008) and humans (Jolma et al. 2013). However, functional characterisation of the enhancer activity of many genomic sequences (Whitfield et al. 2012; Kheradpour et al. 2013; White et al. 2013; Kwasnieski et al. 2014) have revealed that only a small fraction of the potential TFBSs in eukaryotic genomes are actually occupied by TFs in any given cell type, and that these sites vary substantially across cell types and conditions (Spitz & Furlong 2012; Whitfield et al. 2012; Biggin 2011). Moreover, only a subset (25-50%) of bound TFBSs correspond to active enhancers assessed by their ability to drive transcription in reporter assays (Kwasnieski et al. 2014; White et al. 2013; Fisher et al. 2012), stressing that it is unclear what distinguishes TFBSs actually bound by the TF from those that are unoccupied, as well as functional sites. Understanding the 'transcriptional code' involves being able to explain the sequence features and mechanisms underlying the ability of TFs to bind specific enhancers and to drive transcription in a given cellular context. It is likely that, going beyond the isolated TFBSs and considering combinatorial regulatory properties, regulatory architecture and sequence context effects will help in our better understanding of the regulatory genome.

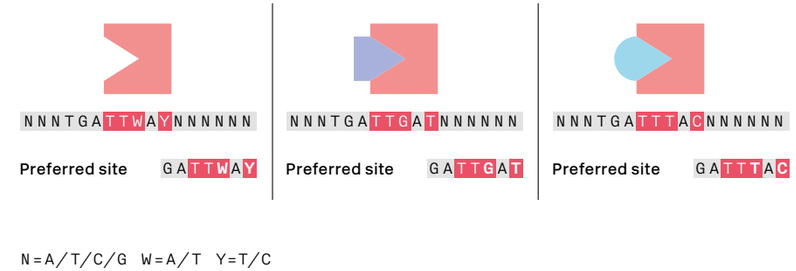
Transcription factor cooperativity at the level of DNA binding is common among enhancers

Multiple lines of evidence suggest that cooperativity (synergism) between TFs can be established at the level of DNA binding. For example, experimentally disrupting the recognition sequences of some TFs or depleting the corresponding TF proteins can cause loss of binding of other TFs (Heinz et al. 2010; Yanez-Cuna et al. 2012; Schulz et al. 2015). Combinatorial TF binding enables cell type- and time-specific enhancer expression. Such differential TF binding has indeed been seen in *D. melanogaster*. pMAD (the phosphorylated form of MAD) provides the competence for cells to adopt particular cell specific fates through combinatorial binding with Tinman in the dorsal mesoderm and Scalloped in the wing imaginal disc (Xu et al. 1998; Guss et al. 2001). Moreover, a similar scenario can be observed within the same cell, but at different developmental times. For instance, Twist binds to sites co-bound by Zelda during early *Drosophila* embryogenesis, but at later stages, when Zelda is not expressed, it binds to different sites co-bound by different partners (Schulz et al. 2015; Yanez-Cuna et al. 2012). Interactions between TFs or between a TF and a cofactor (protein-protein interactions; PPIs) can, directly → **Figure 1.5-A** or indirectly → **Figure 1.5-B**, result in modified DNA binding preferences (Slattery et al. 2011). Moreover, the enhancer architecture, meaning TFBSs arrangements, can also indirectly influence TF cooperativity. For example, the binding affinity of a TF to the DNA and its ability to promote expression can increase as a result of PPIs with a nearby binding TF → **Figure 1.5-C'**. Also DNA bend-

Figure 1.5
Mechanisms of transcription factor cooperativity

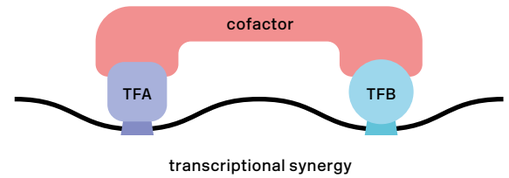
A) Latent specifity

Interactions between TFs, termed protein-protein interactions (PPIs), or between a TF and a cofactor, can result in modified DNA binding preferences. In the figure, TFA (red) preferentially binds different nucleotide sequences depending on the presence of TFB (purple) or TFC (blue).



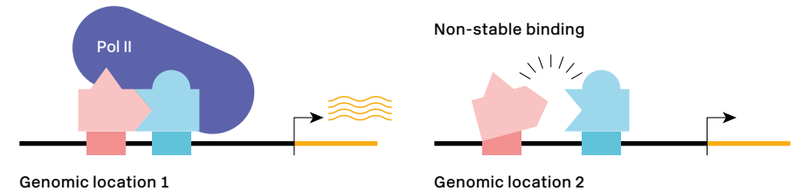
B) Co-binding

Two or more TFs co-bound to the same enhancer element may recruit a common cofactor (e.g. p300), or different components of a multiprotein complex (e.g. the Mediator or SAGA complexes), which may lead to a net increase in the affinity of each TF for their TFBS. Alternatively, it may increase the retention time of the TFs at the enhancer.



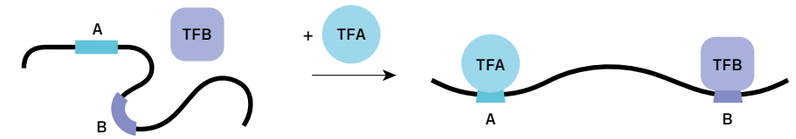
C') Enhancer architecture: Binding through PPI

The regulatory sequence architecture (location, orientation and distance of a TFBS relative to nearby TFBSs) can favour or inhibit specific PPIs, affecting enhancer expression. In the figure, TFA is only able to bind to the DNA, effectively recruit Pol II and drive messenger RNA transcription (orange curved lines) if its BS is close enough to TFB so that both proteins can interact.

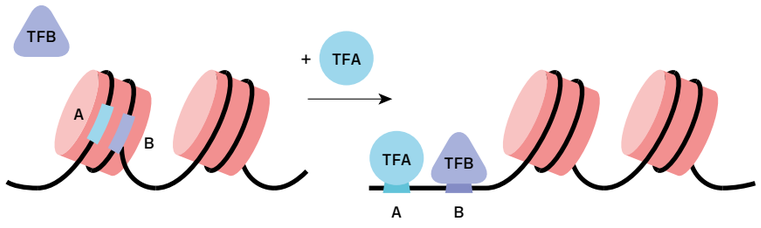


C'') Enhancer architecture: Binding through DNA bending

Some TFs can induce local DNA bending, making additional TFBSs more accessible and thus, increasing the affinity of other TFs for sites in the enhancer.

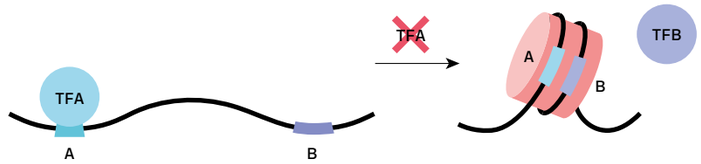


↳



D) Activating chromatin remodeling

Some TFs may act cooperatively by activating chromatin remodelling. In the schematic, the binding motifs for TFA and TFB are located in a nucleosome-bound region of the DNA. Following the binding of TFA to site A, the nucleosome is actively repositioned, which exposes BS B. This allows TFB to bind without the requirement for PPIs.



E) Blocking nucleosome repositioning

By remaining bound to a given site, a TF (TFA in the figure) can prevent nucleosome repositioning and may therefore serve as a place-holder to facilitate the binding of another factor (TFB in the figure) to a neighbouring site that otherwise could become inaccessible if the place-holder factor was removed. This may act as a passive method of enhancer priming. Adapted from (Spitz and Furlong, 2012; Levo and Segal, 2014).

Figure 1.5
Mechanisms of transcription factor cooperativity

ing can help reveal BSs and increase the affinity of other TFs for sites in the enhancer → **Figure 1.5-C''** (Falvo et al. 1995). Moreover, some TFs may act cooperatively by activating chromatin remodelling, without the need to show PPIs, and are called 'pioneer factors'. Pioneer factor binding can also passively reduce the number of additional factors that are needed to bind at a later time in order to create an active enhancer. Such 'priming' can increase the velocity of a transcriptional response and is seen during development and in hormonal regulation (Zaret & Carroll 2011). For example, PHA-4 (FOXA TF in *C. elegans*), has been shown to recruit a histone variant to promoters (Updike & Mango 2006) and to open chromatin *in vivo* (Fakhouri et al. 2010). Similar roles have been described for mammalian FOXA1 and GATA4 (Cirillo et al. 2002) → **Figure 1.5-D**. Finally, blocking nucleosome repositioning may also serve as a place-holder to facilitate the binding of another factor (assisted loading or collaborative competition) to a neighbouring site that otherwise could become inaccessible if the place-holder factor was removed (Miller & Widom 2003; Voss et al. 2011) → **Figure 1.5-E**.

Regulatory architecture has an effect on enhancer functionality

The regulatory architecture refers to the multiplicity, identity, affinity and position of TFBSs present in an enhancer or a regulatory sequence. The distribution of TFBSs has been studied in detail in only a few enhancers. For example, a study of the developmental enhancer *Sparkling*, a specific enhancer of the *Drosophila Pax2* gene, revealed an unexpected high density of essential TFBSs that required specific arrangements for its functionality (Swanson et al. 2010). Although informative, these one-by-one approaches are not able to reveal any general molecular logic underlying regulatory landscapes.

The recently introduced techique of massive parallel reporter assays (MPRAs) allows to test thousands of synthetic sequences for their enhancer activity *in vivo,* thereby obtaining a better understanding of the rules determining enhancer function ('the regulatory code') (Grossman et al. 2017). For example, using twelve liver-associated TFBSs, it was determined that gene expression levels increment monotonically with increasing numbers of TFBSs in a homotypic TFBS cluster and, more importantly, that the highest level of expression directed from an enhancer is achieved with clusters of binding sites (BSs) for different TFs (Smith et al. 2013). Further experiments support these results (Levo & Segal 2014), suggesting that the stronger activity of heterotypic clusters of TFBSs compared to homotypic ones may be a general *cis*-regulatory rule. Studies in yeast found that a ~10 bp periodicity of TFBS location was important for transcriptional activity (Sharon et al. 2012). Not only the number of TFBSs but also the affinity has an effect in enhancer functionality. Low-affinity TF–DNA interactions are abundant *in vivo* and quantifiable from current high-throughput ChIP experiments. They have been shown to contribute quantitatively and spatially to the formation of proper expression patterns and have implications in evolution (Segal et al. 2008; Evans et al. 2012; Tanay 2006).

MPRAs have also been used to carry out exhaustive mutational analysis of genomic regulatory elements to discover their functional architecture at single-nucleotide resolution (also called mutagenesis saturation) (recently reviewed in (Inoue & Ahituv 2015)). *In vivo* studies using the *RhoCRE3* enhancer of the *Rho* gene (mouse retina) found that 86% of single nucleotide substitutions showed significant effects on enhancer activity. Changes in activity were explained not only by mutations within putative TFBSs but also by complex phenomena, including TF competition and TFBS turnover during evolution (Mogno et al. 2013). Additionally, disruption of repressor BSs sometimes results in reporter activity in the wrong cell type, suggesting that the cell type specificity in many cases requires the presence of a repressor (Kheradpour et al. 2013).

Models for enhancer functionality

Specific enhancer architectures might contain constraints on properties such as the number, location, orientation and order of TFBSs, which are referred to as 'grammatical or syntactic rules'. It is still not well stablished if such rules have an important role in the regulation of gene expression, accordingly, three models have been proposed to explain enhancer function based on their syntactic constrains (Spitz & Furlong 2012).

In the first model, termed the 'enhanceosome model', TFBSs show a rigid distribution in order, spacing and orientation → **Figure 1.6-A.** Here the DNA serves as a scaffold for cooperative protein binding. With such an enhancer, the target gene

would be activated only upon the assembly of a complex, providing a precise on/off binary transcriptional switch in response to the appropriate stimulus. This also implies that motif composition is fixed regarding number, distance and orientation between TFBSs; in other words, all TFs must bind to generate an enhancer output. One of the best studied examples of an enhanceosome is the interferon ß-enhancer, where small sequence changes within the 55 bp element alter the binding potential of the eight factors that occupy the enhancer (Thanos & Maniatis 1995; Merika & Thanos 2001). The alternative 'billboard model' proposes a totally flexible distribution of TFBSs → **Figure 1.6-B**. In this second model, also termed 'information display', different enhancer configurations are possible; i.e. motif composition can vary. TFs can act additively or cooperatively to recruit the transcriptional machinery but with no constraints on the relative positioning of their BSs (Smith et al. 2013). In this model it is the transcription machinery that 'reads' or 'samples' discrete regions of the enhancer, giving different and graded enhancer outputs. This was proposed for the *Drosophila* enhancers containing repressors Giant or Knirps and the activators Twist and Dorsal (Kulkarni & Arnosti 2003). The billboard and the enhanceosome models, both shaped by evidence derived from a relatively small set of prototypic examples, are useful approaches to explain general characteristics of enhancers, but evidence available for many other enhancers suggests that they merely represent the extreme ends of a spectrum of architectural diversity (Borok et al. 2010). Supporting this, studies in *Drosophila* and human adipocytes indicate that enhancers often fall somewhere on a continuum between complete modularity, where the spatial relationship between domains is unimportant, and total spatial constraint (Swanson et al. 2010; Grossman et al. 2017). Thus, the final model called the 'transcription factor collective model' represents an intermediate situation → **Figure 1.6-C**. It shows flexibility in the spacing and order of TFBSs and in the motif composition; some TFs will bind to DNA and others may interact with already DNA-bound TFs. An elegant example came from the analysis of five TFs that are essential for cardiac development in *Drosophila* (Junion et al. 2012). The five TFs are found at a large set of enhancers even though each enhancer only harbours a variable subset of motifs required for DNA-binding of all factors. PPIs are though to facilitate the collective occurrence of all TFs, as loss of one member inhibits enhancer activation. In this example, no specific grammar rules were reported but the model predicts the presence of some rules that will or will not be required depending on the specific context of each regulatory sequence.

The importance of sequence context and regulatory landscapes

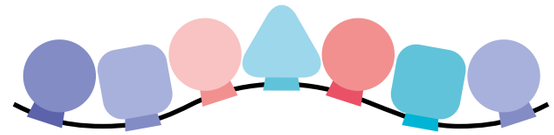
A description of regulatory sequences that only account for the regulatory building blocks and their arrangements views regulatory sequences as inert strings on which functional elements are threaded. However, accumulating evidence suggests that TF access to motif sites may be governed by sequence context, nucleosomes or the larger chromatin landscape.

TFBSs flanking base pairs are important to determine if a particular TFBS will be active or not, as they contribute to TF binding specificity (Maerkl & Quake 2007). Such effect may be also mediated by DNA shape, such as deviations from ideal β-DNA structure (Aow et al. 2013; Siggers & Gordan 2014). One specific type of flanking sequences that have a role in TF binding are A- or T-tracts (Jolma et al. 2013), yet other studies point to the importance of high local GC content for transcriptional activation (White et al. 2013). As described earlier, TF binding can also affect nucleosome occupancy and positioning → **Figure 1.5-E**. Two main se-

Figure 1.6
Current models of enhancer activity

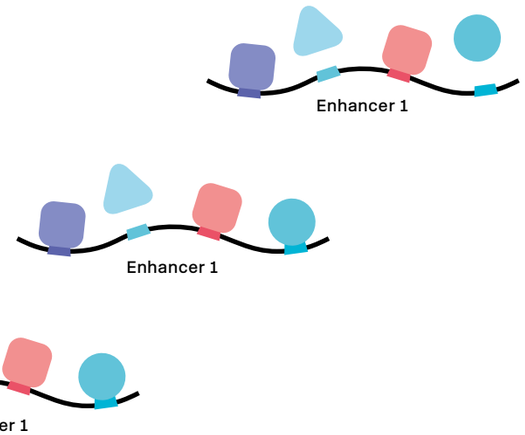
A) Enhanceosome

The enhanceosome model represents a situation in which all TFs that bind to an enhancer are essential for the cooperative occupancy and activation of the enhancer. The DNA motif composition and its relative positioning (motif grammar) act as a scaffold to cooperatively recruit all TFs, which form a higher-order protein interface to regulate transcription.



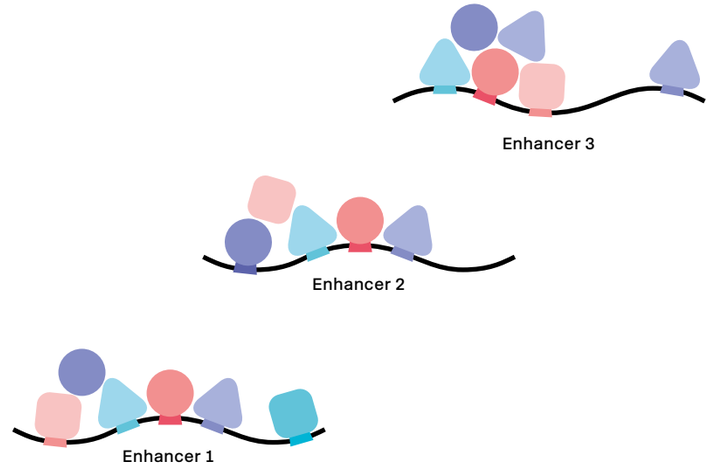
B) Billboard

The billboard model. For any given enhancer, the positioning of TFBSs is flexible and subject to loose distance or organisational constraints. Only a subset of sites in the enhancer may be active at a given time, reflected in the different binding profiles of the enhancer 1.



C) TF collective

The TF collective represents a situation in which the same set of TFs bind to many enhancers. They can occupy each one of these enhancers in a different manner, with all or a subset of TFs directly contacting the DNA. For cases in which the motifs for only a subset of TFs are present, the remaining TFs are still recruited to the enhancer through protein–protein interactions between the participating TFs. The collective binding can therefore occur using diverse motif composition and flexible motif positioning. Adapted from (Spitz and Furlong 2014).



quence features are associated with this: a ~10 bp periodic signal of dinucleotides (AA, TT, AT or TA) that are favoured when the DNA backbone faces inwards towards the histone core and vice versa, and poly(dA:dT) tracts, which facilitate the accessibility of the DNA to binding TFs, thereby influencing the resulting expression (Struhl & Segal 2013; Segal & Widom 2009).

Many works in the last decades have shown that regulatory DNA can act at long distances, often more than 1 Mb, by contacting the promoters of target genes through chromatin loops, forcing a change from a two-dimension to a three-dimension vision of the genome (reviewed in (Maeso et al. 2016)). Chromosome conformation capture techniques that identify DNA–DNA contacts throughout the genome (reviewed in (Denker & De Laat 2016)) have shed light on the *cis*-regulatory architecture of the genome, which is compartmentalised into structures known as topologically associating domains (TADs), and are elucidating the complexity intricate to gene regulatory landscapes, such as the existence of 'regulatory archipelagos' (Montavon et al. 2011), hierarchal relationships between different enhancer elements (Leddin et al. 2011) and even hierarchies within TADs that are specific of each genomic locus (architectural signature) that is thought to reflect the functional activity of that region (Phillips-Cremins et al. 2013). Finally, it has also been suggested that the active regulatory landscapes of a cell could be organised in what has been termed 'transcription factories', although the existence of this nuclear domains is still a matter of debate (Iborra et al. 1996; Jackson et al. 1993).

In summary, thanks to the work from may laboratories in this last decade, we are rapidly increasing our knowledge on transcriptional regulation, however, additional studies and data are required in order to be able to predict the functional regulatory sequences from the genome sequence of a given organism as well as when and where they are activated.

Part II

The next part of the Chapter consists of a detailed description of how the murine serotonergic system is specified, paying special attention to the key regulators of the terminal differentiation steps, and the clinical relevance of serotonergic neurons.

The mammalian serotonergic system

The mammalian serotonergic system is composed by a relative small number of neurons, between 300,000 in humans and 26,000 in mice (Baker et al. 1991; Ishimura et al. 1988; Hornung 2003), yet it innervates nearly all of the cytoarchitectonic regions of the brain and spinal cord and has been implicated in the modulation of seemingly every human behaviour and physiological process orchestrated by the nervous system. For example, it regulates body temperature, sleep, appetite, pain and motor activity and modulates higher brain functions, including cognition and emotional behaviour (Jacobs & Azmitia 1992).

The defining feature of all serotonergic neurons in any organisms is the ability to use serotonin (5-HT) as a neurotransmitter. The biosynthesis of 5-HT is regulated by the coordinated action of a battery of phylogenetically conserved enzymes and transporters known as the 5-HT pathway genes → **Figure 1.7**. Tryptophan hydroxylase (TPH2 in mouse and humans) catalyses the first and rate-limiting step of the pathway, transforming the amino acid tryptophan into 5-hydroxytryptophan (5-HTP). To do so, it requires the GTP cyclohydrolase I (GCH1). Next, the dual functional 5-HTP/L-DOPA decarboxylase (AADC) matures 5-HTP to 5-HT. The vesicular monoamine transporter SLC6A1/2 (also called VMAT2) pumps 5-HT from the cytoplasm into small synaptic vesicles or dense core vesicles (Liu & Edwards 1997) for their transport to the synaptic terminal, thereby controlling the releasable pool of 5-HT. The 5-HT reuptake transporter SLC6A4 (also called SERT) in the plasma membrane absorbs extracellular 5-HT into the cytoplasm (Ramamoorthy et al. 1993; Blakely et al. 1991), and 5-HT is degraded by the oxidase MAO, common to all monoamine biosynthesis pathways (Youdim et al. 2006). SERT is present not only in the presynaptic plasma membrane of 5-HT-producing neurons to reuptake 5-HT from the synaptic cleft, but also in a range of neurons that are capable of absorbing 5-HT from extrasynaptic space but do not synthesise it. Moreover, also present in the pre-synaptic terminal are the two 5-HT auto-receptors (HTR1A and HTR1B).

The 5-HT synthesis pathway is highly conserved in evolution as there are known homologues for all of its components in multiple organisms from humans and mouse, to fish, flies and worms (Flames & Hobert 2011).

Mouse serotonergic neuron specification

Brain serotonergic progenitors are located in the ventral hindbrain and form two main clusters – rostral (rombomeres 1 to 3) and caudal (rombomeres 4 to 6) → **Figure 1.8-B** (Alonso et al. 2013; Okaty et

al. 2015). The rostral cluster includes subgroups of dorsal and median raphe nuclei and sends mostly ascending projections to the forebrain and midbrain, modulating higher brain functions, whereas the caudal cluster has primarily descending connections to the spinal cord. At the molecular level, recent transcriptome analyses and single-cell sequencing have mapped the differences in gene expression between and within the rostral and caudal raphe nuclei (Wylie et al. 2010; Okaty et al. 2015). These studies propose the existence of specialised subtypes of serotonergic neurons that regulate specific biological functions and will have transcendental health implications, such as tailored therapies to particular serotonergic dysfunctions.

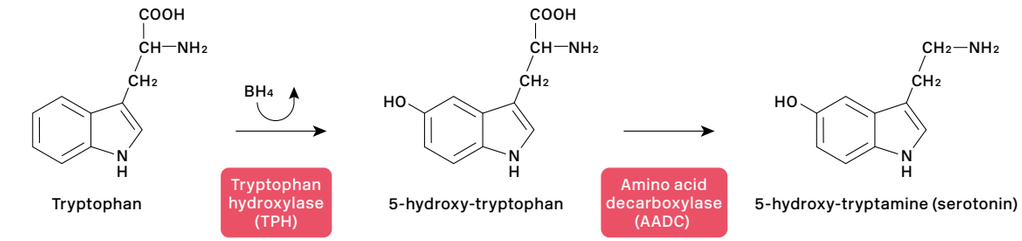
In the developing embryo, the secreted signalling molecules sonic hedgehog (SHH) and fibroblast growth factors 4 and 8 (FGF4, FGF8) regionally pattern the mid-hindbrain neuroepithelium, to specify serotonergic progenitors in the ventral hindbrain that will later give rise to the different serotonergic nuclei (Ye et al. 1998; Cordes 2005). Visceral motor neurons (VMN) and serotonergic neurons arise sequentially from the same progenitors → **Figure 1.8-A**. Two rhombomeres are the exception to this rule: r1, which produces only serotonergic neurons and will constitute more than half of the total serotonergic population, and r4, which never produces serotonergic neurons. At embryonic stage E10.5, rhombomeres r2, r3, and r5–7 shift from VMN generation to serotonergic neuron production by inhibiting the expression of the homeodomain TF *Phox2b* (paired-like homeodomain protein 2b). This temporal switch is controlled by the homeobox TF NKX2.2 (Pattyn et al. 2003).

Several TFs are known to be expressed in the progenitors of the serotonergic neurons and instruct serotonergic fate. The proneural bHLH factor ASCL1 (also known as MASH1 and recently shown to act as a pioneer factor (Raposo et al. 2015)), is expressed during the serotonergic progenitor stages at all rhombomeric levels of the hindbrain, and its induction is likely an early response to morphogens (Briscoe et al. 1999; Pattyn et al. 2003). Its expression is extinguished as progenitors exit the cell cycle to become postmitotic precursors → **Figure 1.9**. FOXA proteins (forkhead TFs) are well known for having multiple roles in single-cell lineages (Kaestner 2010). In the case of the serotonergic lineage, FOXA2 expression is highly induced as the progenitors switch to serotonergic neurogenesis. ASCL1 and FOXA2 act upstream and regulate the expression of the serotonergic postmitotic factors, as described next → **Figure 1.9**.

Two GATA TFs (GATA2 and GATA3) participate in serotonergic postmitotic specification. GATA2 is required for the activation of the serotonergic neuron-specific TF *Pet1* (Craven et al. 2004) and also *Gata3* (Haugas et al. 2016) → **Figure 1.9**. It seems to act high in the regulatory network as neuronal precursors adopt alternative fates in *Gata2* conditional mutants (Haugas et al. 2016). Contrary to GATA2, GATA3 seems to act late in the developmental pathway as it has been proposed to regulate *Tph2*, although GATA3 BSs in the *Tph2* regulatory region have not been identified → **Figure 1.9** (van Doorninck et al. 1999; Pattyn et al. 2004). Rostral serotonergic neurons are not affected in *Gata3* mutant embryos. This observation indicates the presence of slightly different regulatory mechanisms between anterior and posterior serotonergic populations. *Gata3* expression is maintained in adult serotonergic neurons (Zhao et al. 2008), but it has not been determined if *Gata3* conditional mutant adult animals have serotonergic neuron defects (van Doorninck et al. 1999). Another serotonergic postmitotic TF is INSM1, a Zn Finger TF that regulates LMX1B, GATA2 and PET1 expression. In null *Insm1* mutants, the expression of these TFs and the serotonergic terminal marker *Tph2* are significantly downregulated → **Figure 1.9** (Jacob et al. 2009). However, whether the regulation of the tryptophan hydroxylase gene

Figure 1.7
Serotonin biosynthetic pathway

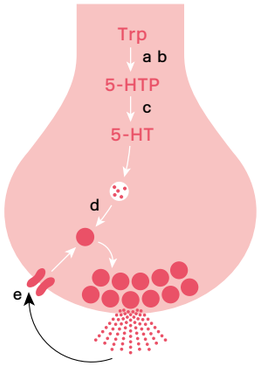
A) Biosynthesis of serotonin (5-HT)



B) Serotonin battery of genes that define serotonergic identity

Abbreviations:
5-HT: serotonin;
5-HTP: 5-hydroxytryptophan;
BAS-1: biogenic amine synthesis related 1; AADC: amino acid decarboxylase;
GCH1: GTP cyclohydrolase 1;
MOD-5: modulation of locomotion defective;
SERT: serotonin transporter;
TPH: tryptophan hydroxylase;
Trp: tryptophan;
SLC18A1/2: Solute Carrier Family 18 Member A1/A2 (also called VMAT (vesicular monoamine transporter));
SLC6A4: Solute Carrier Family 6 Member A4.

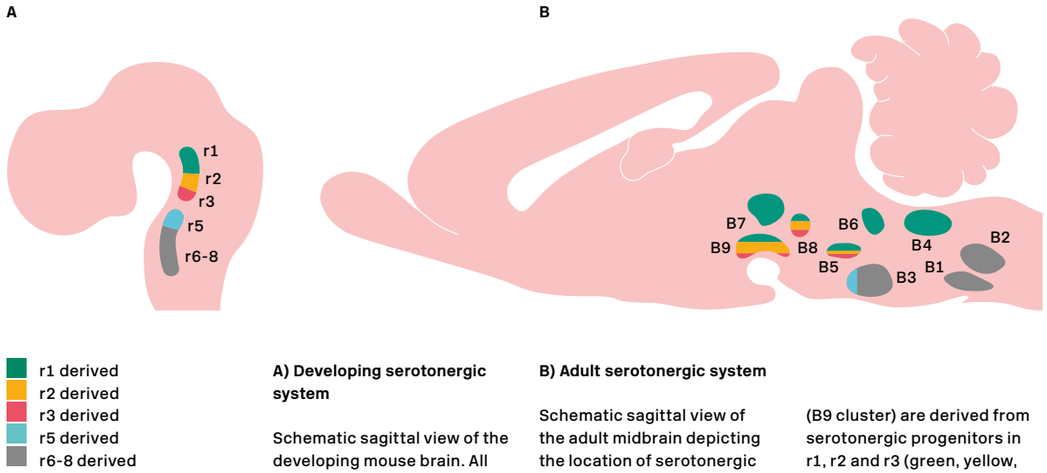
	Serotonin pathway genes:	*M. musculus*	*C. elegans*
a	TPH2	TPH-1
b	GCH1	CAT-4
c	AADC	BAS-1
d	SLC18A1/2	CAT-1
e	SLC6A4	MOD-5



is direct or is an indirect consequence of the loss of expression of the other post-mitotic serotonergic fate determinants remains unclear.

LMX1B and PET1 are expressed in postmitotic serotonergic neurons just before the onset of *Tph2* expression, and their expression is maintained throughout the life of the animal (Ding et al. 2003; Hendricks et al. 2003; Asbreuk et al. 2002). *Pet1* and *Lmx1b* null mutants show defects in expression of the 5-HT pathway genes (*Tph2*, *Sert*, *Vmat2*, *Aaad*) but show no defects in the expression of earlier markers → **Figure 1.9**. In both mutants, postmitotic cells are still generated, and at least in the case of *Pet1* mutants, they still express panneuronal features and they do not switch their fate to that of another neuron type (Cheng et al. 2003; Ding et al. 2003; Hendricks et al. 2003). This phenotype is very similar to the ones described for the terminal selectors NURR1 and PITX3 in dopaminergic specification (Jacobs, van Erp et al. 2009). In addition, PET1 and LMX1B have other characteristics typical of terminal selectors. PET1 directly binds to regulato-

Figure 1.8
Neuroanatomical features of serotonergic neuron development in the mouse



A) Developing serotonergic system

Schematic sagittal view of the developing mouse brain. All mouse serotonergic neurons are born in two longitudinal domains, rostral (r1–r3) and caudal (r5–r8), on either side of the floor plate in the ventral rhombencephalon or hindbrain.

B) Adult serotonergic system

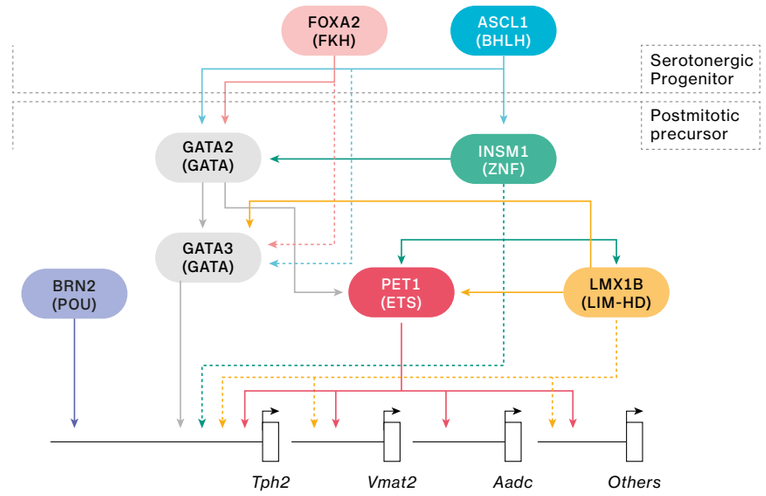
Schematic sagittal view of the adult midbrain depicting the location of serotonergic neurons clusters (raphe nuclei). All serotonergic neurons in the mature dorsal raphe nucleus (B4, B6 and B7 groups) are born in r1 (green). Serotonergic neurons in the median raphe nucleus (B5 and B8 groups) and laterally extending supraleminscal serotonergic neurons (B9 cluster) are derived from serotonergic progenitors in r1, r2 and r3 (green, yellow, red). Serotonergic neurons born in r5–r8 migrate to form the raphe pallidus (B1), raphe obscurus (B2), raphe magnus (B3) and cell bodies in the ventrolateral medulla (B3) (grey). Serotonergic neurons in the rostral portion of B3 are born in r5 (blue). Adapted from (Deneris & Wyler 2012).

ry elements in all the 5-HT pathway genes and also autoregulates its own maintenance of expression, and both PET1 and LMX1B are required to maintain the serotonergic phenotype (Hendricks et al. 1999; Liu et al. 2010; Scott 2005). More recently, it has been shown that continuously expressed PET1 acts as a postnatal maturation-promoting factor of serotonergic neuron excitability, controlling synaptic input, and it also directly binds to secondary maturation regulatory factors, such as *Engrailed* (Wyler et al. 2016; Fox & Deneris 2012). Despite the central role for PET1 in serotonergic specification, several observations indicate that it does not act alone to control serotonergic terminal fate: 1) although in the brain is specific for serotonergic neurons, PET1 is expressed in additional tissues that do not express the serotonergic gene battery (Fyodorov et al 1998), 2) 20% of serotonergic cells remain unaltered in *Pet1* null mutants (Hendricks et al. 2003), and 3) ectopic expression of LMX1B or PET1 individually is not sufficient to turn on expression of the 5-HT pathway genes when electroporated into E2 chick neural tube (Cheng et al. 2003).

One such additional factor could be BRN2. A hypomorphic allele of *Brn2* (also known as POU3F2), shows decreased levels of 5-HT production in the

Figure 1.9
Mouse serotonergic transcription regulatory logic

Arrows indicate activation of a downstream TFs. 5-HT pathway genes appear as white boxes. Short dotted lines indicate previously described relationships between TFs that more recent reports question. Long dotted lines indicate a described role of a TF in the regulation of the 5-HT pathway genes that could be due, however, to an indirect effect. Adapted from (Deneris and Wyler 2012; Haugas et al. 2016, Haugas et al. 2016, Scheuch et al. 2007 and Nasu et al. 2014).



brain (Nasu et al. 2014). Interestingly, a polymorphism in the regulatory region of *Tph2* caused reduced binding of BRN2 (Scheuch et al. 2007), suggesting that the serotonergic phenotype observed in *Brn2* mutants could be due to a direct effect of the TF over *Tph2* regulation → **Figure 1.9**. However, the precise role of *Brn2* on serotonergic specification should be further addressed.

In contrast to the mammalian serotonergic system that arises only from the hindbrain, other vertebrates contain additional serotonergic nuclei in other brain regions. In zebrafish, for example, apart from the raphe serotonergic neurons that, as in mammals, express and are dependent on *Pet1*, there are additional PET1-independent serotonergic cell groups in the forebrain (Lillesaar et al. 2007). Invertebrate serotonergic systems are also composed by nuclei from different origins. In *Drosophila*, two broad groups of serotonergic neurons are known (brain and ventral ganglion) that are regulated by distinct sets of TFs (Dittrich et al. 1997; Lundell & Hirsh 1998).

Serotonergic link to mental disorders

Evidence from decades of studies in humans, non-human primates and rodents strongly support an association of altered serotonergic function with behavioural and physiological pathogenesis such as depression, obsessive-compulsive disorder, anxiety or autism (Albert et al. 2011; Holmes 2008; Waider et al. 2011; López-Arvizu et al. 2011). Genetic manipulation of the levels of *Sert*, *Tph2* or *Htr1a* in mice have been linked to emotional and stress-related behaviours (Murphy et al. 2008; Richardson-Jones et al. 2011). These studies provide a compelling impetus to find gene variation that affects serotonergic signaling and confer risk for neuropsychiatric diseases. Perhaps the most convincing example of disease-associated variation of the serotonergic pathway is the multiple, rare, non-synonymous variants in SERT that create gain-of-function alleles that produce elevated levels of the transporter or enhanced trafficking-independent 5-HT transport activity (Prasad et al. 2005). Notably in a large sam-

ple of multiplex autism families, it was shown that these variants are significantly associated with autism and rigid compulsive behaviours (Sutcliffe et al. 2005). However, as described in the previous section, most of the genetic associations lie outside the coding genome. Whether variation in the serotonergic transcription regulatory network affects the functionality of serotonergic neuron and susceptibility to mental illness is still unknown.

Despite variation with such an effect has not been discovered in humans, different pieces of evidence support this idea. First, *Lmx1b*- and *Pet1*–deficient mice show several emotion- and stress-related behavioural abnormalities that mimic those observed in *Tph2*, *Sert* and *Htr1a* targeted mice, including increase aggression, anxiety-like behaviour and fear responses (Hendricks et al. 2003; Kiyasova et al. 2011). Notably, LMX1B and PET1–dependent transcription is required in adulthood to maintain normal anxiety-like behaviours (Liu et al. 2010), raising the possibility that behavioural pathogenesis might derive from adult onset disruption, genetically or environmentally, of serotonergic transcription. Finally, a recent Genome-wide association study (GWAS) detected 56 significant single nucleotide polimorphisms (SNPs) associated with bipolar disorder including a novel region between MIR2113 and BRN2 (Mühleisen et al. 2014), once again, pointing to BRN2 action in serotonergic differentiation and highlighting network variation as a potentially important mechanism in neuropsychiatric disease pathogenesis.

Part 3

Caenorhabditis elegans as a model system

Sydney Brenner defined *Caenorhabditis elegans* as '*an experimental organism which was suitable for genetical study and in which one could determine the complete structure of the nervous system*' (Brenner 1974), highlighting for the first time its great applicability in the field of neurobiology. Since Brenner's pioneering work more than four decades ago, many have joined in his quest to exploit the simplicity of *C. elegans* nervous system to answer fundamental questions that can be translated to far more complex nervous systems such as the human brain.
In nature, *C. elegans* can be found in actively growing stages in compost and rotting fruits or in arrested dauer stage in various locations including shells of the snail *Helix aspera*, in parasitic association with isopods or in soil (Chen et al. 2006). The canonical wild type strain was isolated in 1954 from a compost heap at Bristol (England) and was given the strain designation 'N2'. In the laboratory, *C. elegans* can be cultured on a diet of *Escherichia coli* on a nutrient agar surface. Its small size (approximately 1 mm in length), rapid generation time (three and a half days at 20 °C) and large brood size (~300 offspring) facilitates the culture of large populations amenable to genetic screens. *C. elegans* is androdioecious, which means it is primarily self-fertilising (hermaphrodite), facilitating strain maintenance and assuring an isogenic background (Brenner 1974). However, males (XO) occur due to rare meiotic non-disjunction of the X chromosome, with a frequency lower than 0.2 %. *C. elegans* nuclei contain five pairs of autosomal chromosomes (I-V) and one sexual chromosome (X). Hermaphrodites are XX and males XO. The existence of males allows for cross-fertilisation, a very useful trait for genetic manipulation. Additionall, for long-term storage, lines can be maintained as frozen stocks.

C. elegans was the first multicellular organism for which the complete genome sequence was determined and annotated (The C. elegans Sequencing Consortium 1998; Waterston & Sulston 1995). This was made feasible by the compact size of its genome: 100 Mb in comparison to the 3137 Mb in humans. Importantly, at least 83% of *C. elegans* proteome has human homologous genes and 70% of human genes contain a *C. elegans* orthologue (Lai et al. 2000). Thus, a very important feature of the *C. elegans* genome is the compactness of the non-coding genome, including the regulatory genome. This feature makes it unique to study the *cis*-regulatory logic of cell differentiation. Several related nematode species have also been isolated and fully sequenced (Stein et al. 2003; Ghedin et al. 2007; Abad et al. 2008; Opperman et al. 2008; Dieterich et al. 2008), being very useful for comparative genomics and phylogenetic filtering of conserved gene regulatory regions.
Another advantage of *C. elegans* is the resolved cell lineage of the worm → **Figure 1.10**. Animals develop from fertilised zygotes through an invariant cell lineage into adult hermaphrodites contain-

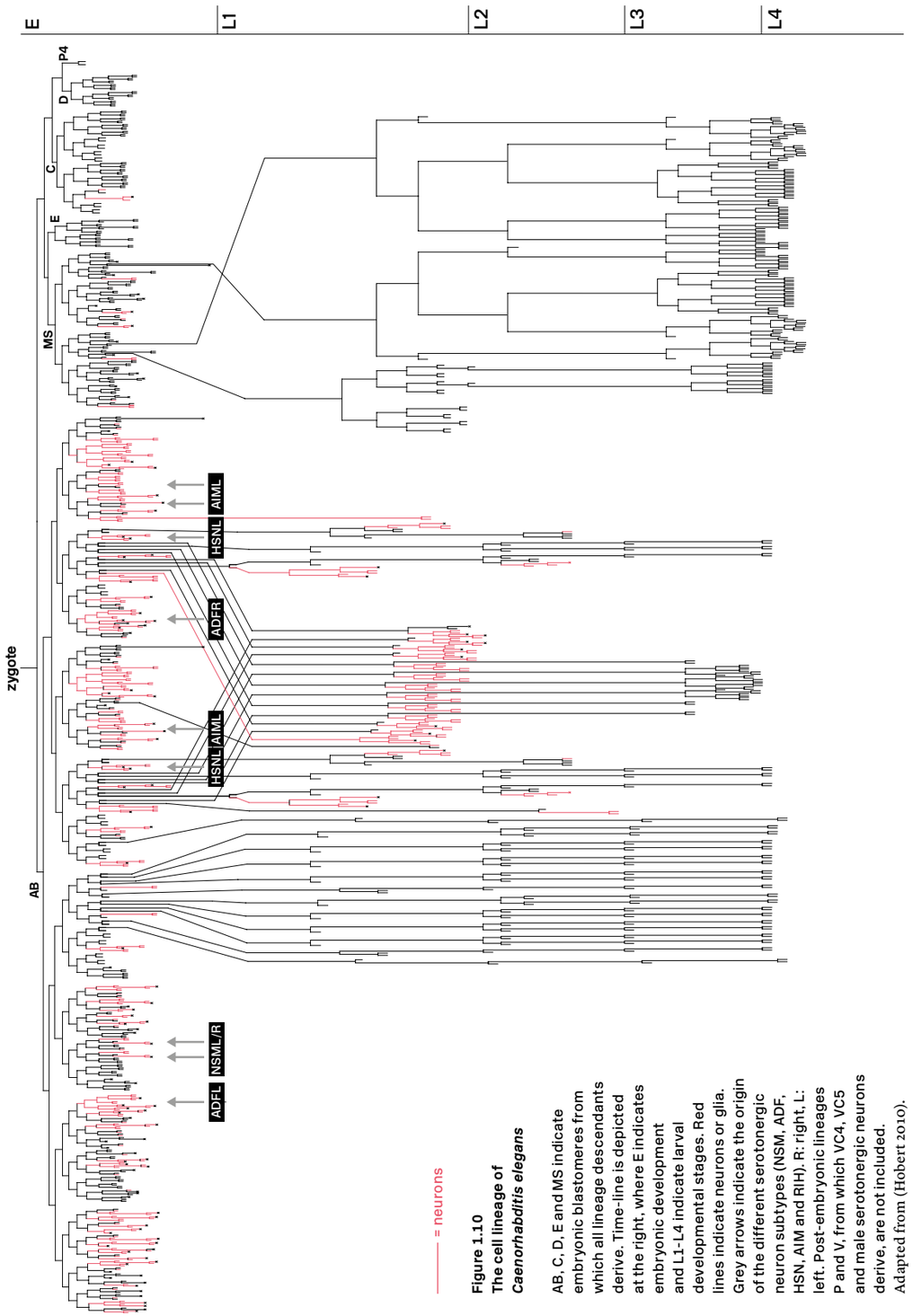


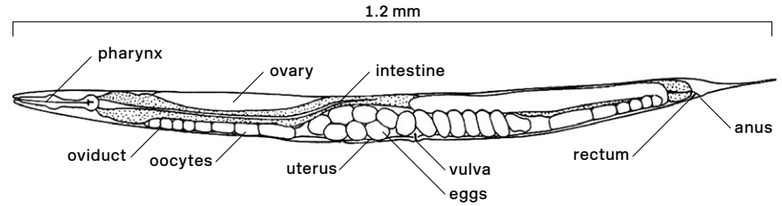
Figure 1.10
The cell lineage of *Caenorhabditis elegans*

AB, C, D, E and MS indicate embryonic blastomeres from which all lineage descendants derive. Time-line is depicted at the right, where E indicates embryonic development and L1-L4 indicate larval developmental stages. Red lines indicate neurons or glia. Grey arrows indicate the origin of the different serotonergic neuron subtypes (NSM, ADF, HSN, AIM and RIH). R: right, L: left. Post-embryonic lineages P and V, from which VC4, VC5 and male serotonergic neurons derive, are not included. Adapted from (Hobert 2010).

Figure 1. 11
Schematic diagrams showing anatomical features of *Caenorhabditis elegans*

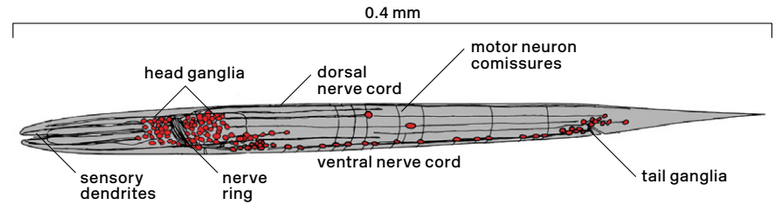
A) Different tissues in the adult hermaphrodite

From (Sulston and Horvitz 1977).



B) *Caenorhabditis elegans* nervous system

C. elegans nervous system. Neurons appear in red. In the first larval stage, shown here schematically, there are 222 neurons. During larval development a number of neurons are added resulting in a total number of 302 neurons. From (Hobert 2010).



ing 959 somatic cells (Sulston et al. 1983), organised in differentiated tissues as pharynx, gut, cuticle, muscle, reproductive system and nervous system → **Figure 1.11-A**. This knowledge allows us to study how a cell establishes its fate and to analyse the effect of mutations and environmental factors at single cell resolution.

In addition, *C. elegans* is highly amenable to genetic studies. Mutants are readily generated and are made available upon request to the research community by multiple consortia. Interferance RNA (RNAi) provides another powerful method for gene knockdown and it is easily performed just by feeding the animals with RNAi producing bacterial clones (Timmons & Fire 1998). Forward and reverse genetics have led to the molecular identification of many key genes in a plethora of developmental and cell biological processes. Moreover, germline transformation is easily achieved by microinjection of DNA into the gonad, yielding transgenic lines in a few days (Mello et al. 1991). Recent introduction of CRISPR-Cas9 technology to *C. elegans* has further improved the specificity in the creation of mutant and transgenic worms (Dickinson et al. 2013; Friedland et al. 2013; Shen et al. 2014). The transparency of the worm allows for *in vivo* imaging techniques. Expressing fluorescent proteins under the control of specific promoters allows *in vivo* visualization of specific neuron types (Chalfie et al. 1994). The possibility to use different fluorescent proteins (Fradkov et al. 2000) next to the knowledge of the genetic linage of every cell in the worm favours for neuron identification.

Caenorhabditis elegans nervous system

Even though *C. elegans* is one of the simplest organisms with a nervous system, its nervous system is the most complex tissue in the organism, both in terms of numbers and neural diversity. The adult hermaphrodite contains exactly 302 neurons and 56 glial cells, comprising more than one third of the somatic cells (White et al. 1986). A similar ratio is maintained in the males: 385 neurons in a total of 1031 somatic nuclei (Sammut et al. 2015; White et al. 1986). Neurons are organised in several ganglia in the head called the nerve ring (similar to a worm brain) and tail, and into a ventral nerve cord resembling the spinal cord → **Figure 1.11-B**. Perhaps the most striking difference between neuronal specification in *C. elegans* and other organisms is that neurons are largely non-clonally derived → **Figure 1.10** and arise from several different lineages (Sulston et al. 1983). Some lineage sub-branches give rise to neurons only, but most oftenly lineages give rise to neurons and non-neuronal cells such as muscle.

C. elegans neurons may look simple on a gross anatomical level, but their connectivity, behaviour and gene expression batteries are intricate and complex. The pattern of connectivity of the entire *C. elegans* nervous system has been successfully mapped for hermaphrodites (White et al. 1986) and males (Jarrell et al. 2012). To date there is no similar report for any other living organism. Based on this work, the 302 neurons of the hermaphrodite are categorised in 118 morphologically and anatomically distinct neuron classes (White et al. 1986). In many cases, contrary to more complex nervous systems, individual cells rather than groups of cells define neuronal subtypes. *C. elegans* contains interneurons, motorneuron and sensory neurons; and the diversity of neurotransmitters and neuropeptides used is comparable to vertebrate nervous systems (Rand & Nonet 1997; Hobert 2013). Functional assignments have been made to different neuronal classes based on genetic screens and laser ablation studies. Only two neuron classes are essential to the survival (Bargmann & Avery 1995), allowing for perturbation of many aspects of *C. elegans* neurobiology without affecting viability.

At a genomic level, the neuronal genome of *C. elegans* consists of approximately 2800 effector genes, excluding gene regulatory factors and structural and regulatory genes involved in cytoskeleton organisation or basic cellular processes (Hobert 2013), and represent most of the families present in vertebrates.

The extraordinary wealth of knowledge already available, combined with the genetic amenability of the worm and its straight-forward neuronal classification, confers the possibility to study neuron type and subtype specification at single cell resolution, making of *C. elegans* a peerless system with which to dissect the nuts and bolts of gene expression regulation in the context of terminal cell fate specification within the nervous system.

Caenorhabditis elegans serotonergic system

The serotonergic system of the hermaphrodite *C. elegans* is defined by a subset of five neuronal subtypes that are immunoreactive to 5-HT antibodies. They are the NSM, ADF, AIM and RIH neurons in the head, and the HSN neuron in the mid-body of the worm → **Figure 1.12-A** (Rand & Nonet 1997; Desai et al. 1988; Sze et al. 2000; McIntire et al. 1992; Horvitz et al. 1982). All except RIH appear as bilateral pairs, summing a total of nine serotonergic neurons. An additional pair of neurons, termed VC4 and VC5, have been reported to present weak and unreliable 5-HT immunoreactivity → **Figure 1.12-A** (Duerr et al. 2001; Duerr et al. 1999). Moreover, 5-HT staining has been reported in the I5 head neuron and in the PHB bilateral neurons in the tail, although it has not been further replicated (Sawin et al. 2000). The serotonergic system is sexually dimorphic, being HSN and VC4/5 hermaphrodite specific neurons; i.e. HSN undergoes programmed cell death in the males and VC4/5 are not generated. Males contain at least twelve additional 5-HT-immunoreactive neurons that are born postembryonically: six CPs (CP1-6) located along the ventral nerve cord (Loer & Kenyon 1993; Sze et al. 2000) and three pairs of Ray B neurons in the tail (Ray 1, 3 and 9) associated to the sensory ray structures at the tip of the tail of the male (Loer & Kenyon 1993) → **Figure 12-D**. An additional unilateral cell located at the right preanal ganglion (RPAG) has also been described to be 5-HT-immunoreactive (Loer & Kenyon 1993).

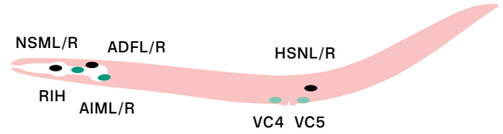
According to their source of 5-HT, serotonergic neurons can be classified in two distinct groups: 5-HT producing neurons are those that express the enzymes for the biosynthesis of 5-HT, including tryptophan hydroxylase enzyme (TPH-1), and thus are able to synthesise 5-HT cell-autonomously → **Figure 1.12-B**. These cells may or may not express the sole serotonin reuptake transporter in *C. elegans* (MOD-5). Within this subgroup lay the NSM and ADF neurons, and the sexually dimorphic HSN, CP1-6 and Ray 1, 3 and 9 neurons. All of them express the *tph-1, cat-1, bas-1* and *cat-4* genes (Loer & Kenyon 1993; Duerr et al. 1999; Sze et al. 2000; Sze et al. 2002; Hare & Loer 2004). Contrary, serotonin absorbing neurons are those that use 5-HT as a neurotransmitter but do not synthesise it. They do not express TPH-1 so, in order to achieve 5-HT neurotransmission, they uptake the molecule from the extrasynaptic space using MOD-5 transporter (Jafari et al. 2011). 5-HT absorbing neurons

Figure 1.12
***Caenorhabditis elegans* serotonergic system**

5-HT-producing neuron (NSM, ADF and HSN)
5-HT-absorbing neuron (AIM and RIH)
Unreliable 5-HT staining (VC4 and VC5)

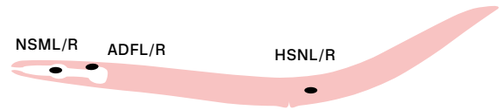
A) Complete serotonergic system

Complete hermaphrodite serotonergic system, including all 5-HT immunoreactive neurons.



B) 5-HT-producing neurons

Specific serotonergic subpopulation of 5-HT producing neurons, expressing the tryptophan hydroxylase gene, in hermaphrodites.



C) 5-HT-absorbing neurons

Specific serotonergic subpopulation of 5-HT absorbing neurons, expressing the 5-HT re-uptake transporter, in hermaphrodites.

D) Male serotonergic system

Complete male serotonergic system. L/R: bilateral left/right neurons.

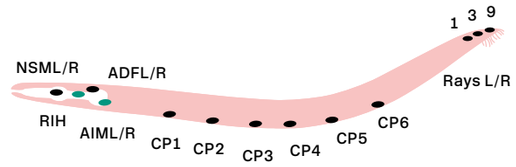
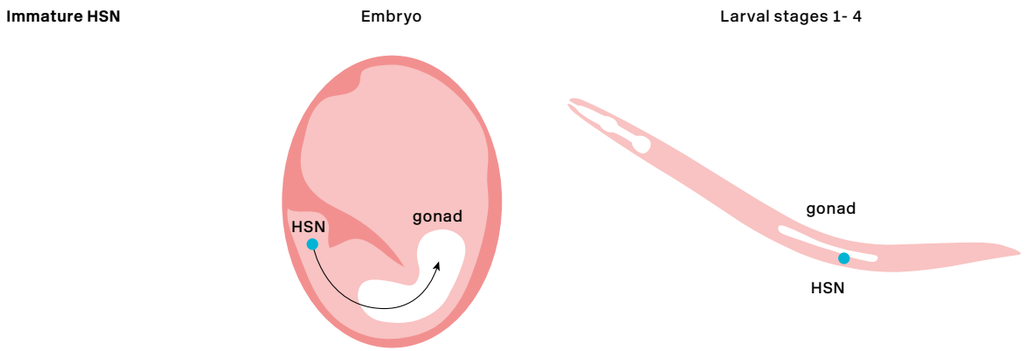


Figure 1.13
HSN anatomical description

A) Schematic representation of the localisation of the HSN neuron in the embryo (top-left) and the corresponding location in larvae (top-right)

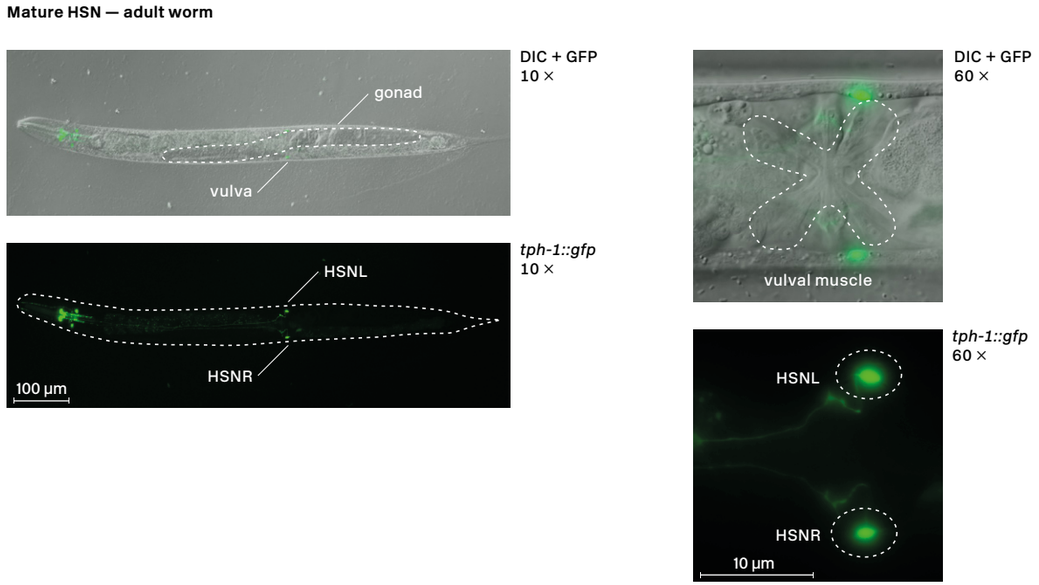
The HSNs are generated in the tail of the embryo and then migrate anteriorly to the gonad primordium, near the middle of the animal. This occurs in both sexes but the HSN undergoes programmed cell death in the males (Sulston & Horvitz 1977). At this point HSN are immature; they do not express the tryptophan hydroxylase enzyme TPH-1, nor they synthesise 5-HT.

Blue circles represent immature HSN neuron. The arrow indicates the migratory route. Note that only one of two bilaterally symmetric HSNs is illustrated.



B) Micrographs of bilateral HSNs expressing *tph-1::gfp* in a ventral plane at low magnification (left) and at higher magnification (right)

GFP signal in the head corresponds to the NSM and ADF serotonergic neurons. Gonad and vulva (arrow head) are indicated. R: right, L: left.



are AIM and RIH → **Figure 1.12-C**. Both neurons express the vesicular monoamine transporter (CAT-1) and AIM additionally expresses the decarboxylase BAS-1. Certain reporters for *mod-5* expression have been also reported to be expressed in the NSM and ADF neuron (Ranganathan et al. 2001).
Serotonergic neurons, like many other subclass specific neurons in the worm, arise from very different progenitors in development → **Figure 1.10**. However, in contrast to other aminergic neurons such as the dopaminergic that are all sensory neurons, serotonergic neurons belong to different functional classes. NSM is a neurosecretory motorneuron, ADF is a ciliated chemosensory neuron, AIM and RIH are interneurons and HSN and the special class VC4/5 are motorneurons. Male CP1-6 are motorneurons while the Rays 1,3 and 9 are ciliated sensory neurons. This heterogeneity is reflected in the diversity of behaviours that have been linked to serotonin signalling in the worm. For example, the ADF neurons play a role in chemosensation, aerotaxis, immunity and food detection (Pocock & Hobert 2010; Chang et al. 2006; Xie et al. 2013; Jafari et al. 2011), while the NSM neurons are involved in pharyngeal pumping and food memory (Albertson and Thomson, 1976; Sawin et al. 2000). AIM and RIH 5-HT absorbing neurons are temporal-spatial regulators of extrasynaptic 5-HT and modulate the response to food deprivation (Jafari et al. 2011). The hermaphrodite HSN neuron appears as a central regulator of the egg-laying motor programme (Waggoner et al. 1998; Shyn et al. 2003; Hardaker et al. 2001) and the male specific neurons have been mainly related with mating behaviour (Loer & Kenyon 1993). Four receptors have been identified that bind 5-HT, including three G protein coupled (metabotropic) receptors (SER-1, SER-4 and SER-7) and one serotonin-gated chloride channel (MOD-1) (Olde & McCombie 1997; Hamdan et al. 1999; Hobson et al. 2003; Ranganathan et al. 2001). Analysis of mutants for these receptors has shed light on the mechanisms by which 5-HT modulates locomotion, pharyngeal pumping and egg-laying.

The HSN serotonergic neuron

An important part of this Thesis focuses on dissecting HSN neuron regulatory programme, therefore, in this section we will describe in a little more detail some aspects of HSN biology.

Hermaphrodite specific motorneurons, HSNs, are generated in the tail of the embryo but migrate anteriorly before hatching. The HSN cell bodies are located lateral and slightly posterior to the vulva, and extend a long process ventrally, into the ventral nerve cord, and then anteriorly, into the nerve ring (Desai et al. 1988) → **Figure 1.13**. Contrary to the rest of the serotonergic neurons that mature prior hatching, synapse formation in the HSN occurs at L3 and L4 stages and, thus, HSN neuron fully matures and synthesises 5-HT at early young adult stage. The main function of the cell is to innervate the vulval muscles and stimulate egg-laying. Genetic ablation of HSN (via *egl-1* apoptotic semidominant mutants) or mutations in genes related with HSN function and development, result in animals that fail to lay eggs normally and retain them in the uterus, showing an egg laying defective (*egl*) phenotype (Trent et al. 1983; Desai et al. 1988). HSNs are particularly important for inducing the onset of egg-laying active phases, releasing 5-HT to the vulval muscles, and for inhibiting the process (Waggoner et al. 1998; Shyn et al. 2003). However, *tph-1* mutants show a mild *egl* phenotype (Sze et al. 2000), thus other neuropeptides and neurotransmitters as acetylcholine may be implicated in the process (Kim et al. 2001; Weinshenker et al. 1995). Moreover, 5-HT neurotransmission leads to a burst in velocity prior to egg-laying events (Hardaker et al. 2001), placing the HSN as a central regulator of the egg-laying motor programme.

Terminal selectors of *Caenorhabditis elegans* serotonergic neurons

As described earlier in the Introduction, neuronal specification and the terminal selector concept have been extensively studied in *C. elegans* (Hobert 2008). In the case of cholinergic neurons, the unique COE (Collier, Olf, EBF)-type TF UNC-3 co-regulates all members of the cholinergic gene battery in most cholinergic subtypes (VA, VB, DA, DB and AS) (Kratsios et al. 2011; Kratsios et al. 2015). In dopaminergic neurons, a combination of three TFs regulates all dopaminergic subtypes (Flames & Hobert 2009; Doitsidou et al. 2013). Contrary to these and other neuronal types in *C. elegans*, to date, there is no unique TF (or combination of TFs) known to be involved in the differentiation of all serotonergic neurons. This resembles, as aforementioned, the serotonergic systems of other organisms as mouse, fish and *Drosophila*.

Among the *C. elegans* serotonergic cells, the HSN is the best characterised. Using an extensive forward genetic screen, 38 distinct genetic loci affecting different aspects of HSN development and function were isolated → **Table 1.2** (Desai et al. 1988; Desai & Horvitz 1989). Several candidates to play a role in HSN terminal differentiation came out from this study. Especially interesting are those genes that code for TFs and whose mutants showed 5-HT staining defects and will be briefly described next.

egl-5 encodes a homeodomain TF, orthologous to *Drosophila* Abd-B and the vertebrate Hox9-13 proteins. EGL-5 is expressed in the HSN neuron (Baum et al. 1999) and this expression is maintained throughout adulthood (Ferreira et al. 1999). Two pieces of evidence suggest that *egl-5* acts higher in the cascade of HSN neural differentiation. One is that in *egl-5* mutant HSN cells change their fate into PHB neurons (HSN sister cell) (Baum et al. 1999). The second is that *egl-5* regulates the survival and fate of HSN/PHB precursor (Singhvi et al. 2008).

egl-43 encodes a zinc finger protein, related to the zinc fingers of the murine Evi-1 proto-oncoprotein (Morishita et al. 1988). Mutations in *egl-43* result in a modest loss of 5-HT stainig and the most severe HSN migration defect of all HSN migration mutants, and EGL-43 expression is restricted to embryonic stages (Baum et al. 1999, Garriga, Guenther et al. 1993). This observation, together with the fact that other mutations that cause severe HSN displacements also result in low penetrance defects in HSN 5-HT synthesis (Garriga, Desai et al. 2013), suggest that migration might be the only HSN trait that is affected directly by the absence of the *egl-43* gene.

ham-2 mutants (C2H2 zinc finger-containing TF), similarly to *egl-43*, show mild 5-HT staining defects that are probably due to their role in the control of HSN migration (Desai et al. 1988; Baum et al. 1999).

egl-44 encodes a member of the transcription enhancer factor family of the TEA domain (TEAD) class. Mutant animals show multiple HSN defects including abnormalities in cell migration, axonal outgrowth and 5-HT production (Desai et al. 1988; Desai & Horvitz 1989). The same phenotypes were observed in *egl-46* mutants (Wu et al. 2001). It has been determined that EGL-44 additionally regulates the expression of EGL-46 (Wu et al. 2001) but it remains unclear whether these TFs directly activate the 5-HT pathway genes.

sem-4 (SPALT type zinc finger TF) controls HSN cellular morphology, axon pathfinding and 5-HT synthesis (Basson & Horvitz 1996; Desai et al. 1988), and its expression is maintained in HSN during all the life of the animal (Grant et al. 2000). As before, it is unknown whether *sem-4* works at the terminal differentiation stage or if it is an upstream regulator of other factors.

The best candidate retrieved from Desai *et al.* work to be a terminal selector for HSN serotonergic differentiation (i.e., a TF that directly controls differentiated features of HSN) is the POU-homeodomain TF UNC-86 (Desai et al. 1988). HSN shows normal early differentiation in *unc-86* mutants but exhibits terminal differentiation defects, including lack of expression of some 5-HT pathway genes and consequently 5-HT synthesis (Sze et al. 2002). UNC-86 protein binds to the *tph-1* upstream regulatory region, arguing that the regulation is direct (Sze et al. 2002). UNC-86 also regulates terminal differentiation of the serotonergic NSM, AIM and RIH neurons (Sze et al. 2002). In fact, it is the only known regulator for the RIH neuron. In the AIM neuron, UNC-86 also regulates other features of the cell such as *mbr-1* expression; a TF involved in neurite pruning (Kage et al. 2005). Recently, CEH-14 (LIM homeodomain) has also been described to regulate AIM serotonergic fate (Pereira et al. 2015). However, UNC-86 is not exclusively expressed in serotonergic neurons. In fact, a classic terminal selector example in the worm is the combined action of UNC-86 and MEC-3 to determine the identity of a group of six mechanosensory neurons (Way & Chalfie 1988; Chalfie et al. 1981; Finney & Ruvkun 1990). Thus, it seems that, as it is the case for other terminal selectors, UNC-86 is used with different partners to select different cell fates.

HLH-3, a basic-Helix-Loop-Helix TF, also appears as a good candidate terminal selector of the HSN. *hlh-3* mutants show normal generation and migration of HSN but fail to express *tph-1* and to synthesise 5-HT, while other serotonergic neurons are not affected in these mutants. *unc-86* expression is not affected in *hlh-3* mutants either, which suggests that both factors act in parallel to specify HSN serotonergic terminal fate (Doonan et al. 2008).

Regarding NSM specification, as anticipated, UNC-86 is also a *bona fide* terminal selector for this neuron (Sze et al. 2002). Indeed, we have recently shown that it acts in combination with the LIM-homeodomain TF TTX-3 in order to regulate the terminal features of the NSM neuron (Zhang et al. 2014). TTX-3 action is specific of the NSM, as the

Gene name	Known phenotype / function	Reference
egl-5	5-HT staining defect. Acts earlier in HSN development; i.e. regulates the fate and survival of HSN/PHB precursor.	(Baum et al. 1999); (Guenther & Garriga 1996); (Singhvi et al. 2008)
egl-43	Small 5-HT staining defect. Exclusive role in migration.	(Baum et al. 1999)
egl-44	5-HT staining, migration, axon pathfinding defects. Acts upstream of other TFs.	(Wu et al. 2001)
egl-46	5-HT staining, migration, axon pathfinding.	(Wu et al. 2001)
ham-2	Small 5-HT staining defect. Exclusive role in migration.	(Baum et al. 1999)
hlh-3	5-HT staining, 5-HT pathway gene expression, axon pathfinding defects.	(Doonan et al. 2008)
sem-4	5-HT staining, morphology, axon pathfinding defects.	Basson & Horvitz 1996)
unc-86	5-HT staining, 5-HT pathway gene expression defects. Acts exclusively at the terminal step of differentiation.	(Sze et al. 2002)
ast-1	5-HT staining, 5-HT pathway gene expression defects.	Observation in the laboratory
egl-18	Migration and 5-HT staining defects. expression defects.	Observation in the laboratory

Table 1.2
Potential regulators of HSN development and function

The list includes genes that exclusively code for TFs and whose loss of function mutants show an egg-laying defective phenotype. Highlighted in pink are those genes that have been studied in this Thesis. Adapted from (Desai et al. 1988).

rest of serotonergic neurons remain unaffected in the single mutants, but TTX-3 acts as a terminal selector together with other TFs to select different terminal fates (Wenick & Hobert 2004).
Despite the ADF neuron is known to regulate multiple processes and behaviours, very little is known about the genes that regulate this serotonergic pair of neurons. ADF does not express *unc-86* and, as expected, is not affected in *unc-86* mutants (Sze et al. 2002). The LIM homeobox gene *lim-4* appears to control the serotonergic phenotype of the ADF neuron (Zheng et al. 2005). However, *lim-4* is only expressed in the progenitor cell of ADF, thus it must act upstream of an additional TF yet to be identified. DAF-19, orthologue of the regulatory factor X (RFX) TFs has been shown to regulate *tph-1* in the ADF and thus 5-HT staining. However, DAF-19 seems not to act directly to regulate *tph-1* expression.

To conclude the Introduction, we have reviewed how serotonergic neurons are specified both in nematodes and in mammals. Serotonergic neurons have been extensively studied during the past decades, probably due to their clinical relevance and the countless processes in which they are involved. Although several TFs are known to be involved in their differentiation (Flames & Hobert 2011; Deneris & Wyler 2012), very little is known about how they act together at the level of gene expression and enhancer regulation, activating the correct gene profile to confer their unique properties. Despite recent technical advances, addressing these questions *in vivo* in complex model organisms such as rodents is a challenging task, as their potential regulatory genome is large and genetic manipulations are time consuming. We believe that the use of simple model organisms like *C. elegans,* which is especially suitable for transcriptional regulatory studies, is key to unravel novel and fundamental aspects of cell type-specific transcriptional regulation.

Objectives

The global aim of this Thesis is to unravel the regulatory transcriptional logic that governs the selection and activation of the terminal features expressed in the phylogenetically conserved serotonergic neurons, using the nematode *Caenorhabditis elegans* as an animal model.

The specific objectives of this Thesis are the following:

1 To dissect *in vivo* the *cis*-regulatory logic of the serotonin pathway genes in the serotonergic neurons subtypes NSM, ADF and HSN.

2 To identify and characterise in detail the transcription factors that control the terminal differentiation programme of the HSN serotonergic neuron subtype (HSN terminal selectors).

3 To interrogate the transcriptome of the HSN neuron for the presence of a DNA-coded regulatory signature that allows for the identification of HSN active enhancers from the whole genome of *C. elegans.*

4 To determine if the serotonergic regulatory programme between nematodes and mammals is phylogenetically conserved, in molecular and functional terms.

Materials & methods

Experimental Procedures

This section describes how experiments were performed, following the same order as will be presented in the Results section. In this way, this part is divided into the corresponding Chapters I to IV.

Chapter I

C. elegans strains and genetics

C. elegans culture and genetics were performed as previously described (Brenner 1974). Briefly, worms were grown in NGM (Nematode Growth Medium, see composition in the Materials section) agar plates (55 mm × 16 mm, non-vented), over a lawn of OP50 bacteria (Caenorhabditis Genetic Center), a uracil-requiring mutant of *Escherichia coli*. The wild type strain used in this study was Bristol N2. All transgenic strains used in Chapter I were generated by microinjection into the N2 strain. Strain names and the transgenes that they express are listed in → **Table 2.21**.

Generation of *C. elegans* transgenic lines for *cis*-regulatory analysis

Gene constructs for *cis*-regulatory analyses were generated by standard cloning into the pPD95.75 expression vector, which contains the *gfp* coding sequence, the muscle myosin *unc-54* 3' UTR and seems to have a basal promoter in the synthetic intron ahead of the *gfp* (Addgene Plasmid #1494) → **Figure 2.1**. Plasmid DNA was purified using QIAprep Spin Miniprep Kit (QIAGEN #27106) and resuspended in miliQ water (Sigma, #W4502). DNA sequences were checked by Sanger sequencing using the ABI Prism 3100 platform (Applied Biosystems), at the Sequencing Unit of the Institute for Plant Molecular and Cellular Biology (IBMCP) — Polytechnic University of Valencia.

Transgenic strains were generated by intragonadal microinjection of the DNA as a simple array into the N2 strain. The injection mix consisted of 50 ng/µL of the plasmid and *rol-6(su1006)* (pRF4, (Mello et al. 1991)) at 100 ng/µL as a co-injection marker (final concentration: 150 ng/µL). *rol-6(su1006)* is a dominant negative mutation of a collagen gene that confers worms a 'roller' phenotype: they twist into a right-handed helix allowing for the easy identification of animals bearing the transgene under the dissecting scope (Kramer et al. 1990).

Prior to the injection, DNA was centrifuged at full speed for about 10 min in order to pellet impurities present in the tube, which could plug the needle. 1 µL of the mix was loaded into a 0.5 µm diameter-capillary tip or needle (Femtotip II, Eppendorf # 930000043). This needle was then adjusted to a micromanipulator that holds it firmly and brings it into the correct angle for gonad injection (15°-45°). Adult worms, with clearly visible gonads, were selected for injection and placed straight in 2 % agarose pads. Halocarbon oil 700 (Sigma, #H-8898) was used to prevent the worms from fast dehy-

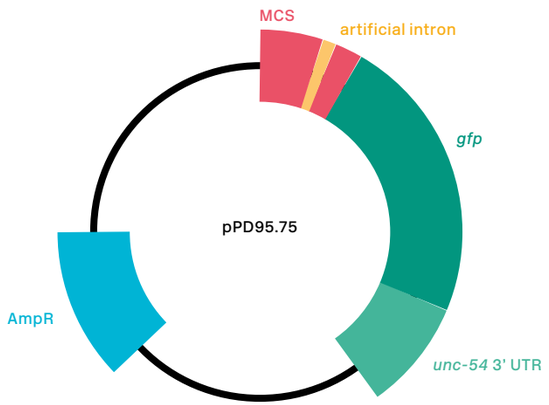


Figure 2.1
Schematic of the pPD95.75 plasmid

MCS: multiple cloning site, where *cis*-regulatory modules (CRM) are cloned.

dration while lying in the agarose bed. Finally, the pad was placed in a high-resolution inverted microscope (Axio Vert.A1 Zeiss) and worms were penetrated with the needle at the syncytial gonad arm, where DNA was liberated. Worms were allowed to recover for several minutes in M9 1X solution and then placed in individual plates (see composition in Materials section). All microinjections procedures were performed at 20 °C.

The progeny of the injected worms (F1) was followed up and worms expressing the *rol-6* co-marker were selected and singled in individual plates. Plates where worms were able to transmit the array to the second generation (F2) were considered stable independent lines. Several extrachromosal lines were normally generated in each injection event and 2-4 transgenic lines with high transmission efficiency were selected and maintained to score.

Scoring

Cis-regulatory reporter scoring was performed using young adult worms maintained at 25 °C. At least three independent lines per construct were scored to assess the variability between transgenic lines. A minimum of 30 animals (60 cells) per line was scored. For the scorings, worms were mounted in 4 % agarose pads (prepared with distilled water) placed over standard microscope slides (Rogo Sampaic #11854782) and sealed with standard coverslips (22 × 22 mm) (VWR #631-1570). Sodium azide (100 µM) (Sigma, #26628-22-8) was used to immobilise worms.

Scoring and images were performed using 60X objective in a Zeiss Axioplan 2 microscope. Percentages of GFP expression in the cell were calculated as the total number of GFP positive cells over the total number of cells scored. Results are shown for individual lines, where +: > 60% GFP positive cells; +/−: 20-60% GFP cells; − < 20% GFP cells, relative to mean wild type values.

Chapter II

Generation of mutant strains

Mutant animals for the six TFs of interest were crossed with several reporter lines. Many mutant and reporter strains used in this chapter were obtained from the Caenorhabditis Genetic Center (CGC). For the specific source of the strain see → **Table 2.21**. Newly generated mutant or double mutant strains were genotyped to confirm their mutant nature. The alleles used in this work have been previously curated (www.wormbase.org), unless indicated, so primers were designed to test the presence of the specific mutations by Polymerase Chain Reaction (PCR) → **Table 2.4**. Mutations were finally confirmed by sequencing.

Genomic DNA preparation for genotyping: worm lysis

Well-grown, non-starved plates were used to obtain genomic DNA to genotype strains. Worms were collected in M9 1X buffer and transferred to a 0.5 mL Eppendorf tube, on ice. Worms were allowed to deposit at the bottom of the tube for 10 min. Supernatant M9 1X was removed and replaced by lysis solution (see Materials section) containing 1 % Proteinase K (Roche Life Science, #3115879001). Tubes were vortexed and stored at −80 °C during ≥ 20 min. Then they were placed at 65 °C during ≥ 1 h, followed by 30 min at 95 °C, to inactivate Proteinase K. Genomic DNA was then ready to use as template for genotyping PCRs.

Mutant genotyping

PCRs to genotype mutant strains were carried out using 1 U of Go Taq® DNA polymerase (Promega, #M7806) and 1.5 mM $MgCl_2$ buffer in a final 25 µL volume. The DNA template used was genomic DNA obtained from the worm lysis protocol, without purification or quantification (approx. 200-300 ng). Primers were added to a final concentration of 0.5 µM each and dNTPs 0.2 mM each. Go Taq® DNA polymerase requires a 2 min initial denaturation step at 95 °C and a 5 min final extension step at 72 °C.

When genetic crosses where carried out using the *ast-1(hd92)* L1 lethal allele, we used a rescuing array containing a wild type copy of *ast-1*, represented as *ast-1(+)*, plus a *rol-6(su1006)* co-marker (*hdEx237*). In order to identify homozygous strains for the mutation, we selected plates were all animals showed a roller phenotype, indicating that the *hd92* lethality was rescued by the *hdEx237* array. Moreover, when genetic crosses where carried out using the *lin-11(n389)*, which has not been curated, the mutation was followed by chromosome repulsion using a fluorescent marker in chromosome I.

Mutant scoring

Mutant scoring was performed using young adult worms maintained at 25°C. At least 50 animals (100 HSN cells) were scored for each genotype and percentages of expression were calculated in the same way as described in Chapter I. Standard Error of the Proportion (SEP) was calculated and Fisher Exact Test, two tailed, was applied for statistical analysis. Calculations were performed using Graphpad QuickCalcs online software (www.graphpad.com/quickcalcs/). Each strain was usually scored on different days to assess the reproducibility of the results. The total number of cells scored over the different days were considered as the "n" in the final percentage of expression. Whenever the worms fell laterally on the slide, only one side was scored. This turned out to be particularly important with reporter strains whose intensity of fluorescence was low. Mutant strains were scored in parallel to wild type reporter strains. In → **Figure 3.2.5,** to globally represent how the terminal battery of features of the HSN neuron is affected in the different mutant backgrounds in comparison to wild type, we constructed a heatmap. Heatmaps.2 from R-gplots package was used.

Image processing

Images were acquired using the Zen System 2011 (Zeiss) and processed using the free software

For more details, see Figure 3.3.2.

Table 2.1
Description of the allelic nature of the mutations for the HSN transcription factor candidates

Gene (allele)	Mutation type	Chromosome	Reference
unc-86(n846)	G>A substitution disrupts splice acceptor site at intron 2, coinciding with the POU domain (potentially null).	III	(Finney 1987; Zhang et al. 2014)
unc-86(n848)	G>A substitution disrupts splice donor site in intron 4, coinciding with the homeodomain (hypomorph allele).	III	(Finney 1987; Zhang et al. 2014)
sem-4(n1971)	G>A substitution disrupts splice donor site in exon 2 of *sem-4* long isoform, before any of the seven ZnF domains (predicted null allele).	I	(Basson & Horvitz 1996)
sem-4(n2654)	C>T missense mutation in exon 6 of *sem-4* long isoform, coinciding with the second ZnF domain (hypomorph allele).	I	(Toker et al. 2003)
hlh-3(tm1688)	1244 bp deletion spanning all exon 1 (predicted null allele).	II	(Doonan et al. 2008)
egl-46(sy628)	G>A substitution generates early stop in exon 1. Although Yu et al. describe *sy628* as not necessarily null, in our hands it showed a stronger neuronal phenotype than *gk692* (predicted null allele), which instead showed a more severe *egl* phenotype.	V	(Yu et al. 2003; The C. elegans Deletion Mutant Consortium 2012)
ast-1(ot417)	G>A substitution in exon 6, coinciding with ETS domain (hypomorph allele).	II	(Flames & Hobert 2009)
ast-1(hd92)	Deletion spanning exon 6 and 7 that abolishes the ETS domain (predicted null allele).	II	(Flames & Hobert 2009)
egl-18(ok290)	698 bp deletion spanning from exon 2 to exon 4 that abolishes the zinc finger domain (predicted null allele).	V	(Koh et al. 2002)
end-1&ric-7(ok558)	879 bp deletion spanning part of intron 2 and exon 3. Includes last 246 bp the 3' UTR' of *ric-7* gene.	V	(Maduro et al. 2005; The C. elegans Deletion Mutant Consortium 2012)
ceh-14(ch3)	1277 bp deletion spanning from intron 2 to exon 4, including both LIM domains.	X	(Cassata et al. 2000)
lin-11(n389)	Not curated (putative null allele).	I	(Trent et al. 1983)

ImageJ 1.50i (Rasband, W.S., https://imagej.nih.gov/ij/), Adobe Photoshop CC and Adobe Illustrator CC.

Population synchronisation

Plates full of gravid adults were used for population synchronisation via bleaching (*Egg Prep*). M9 1X was poured onto the plates and they were gently swirled to dislodge worms. Using a glass pipette, worms were transferred to a 15 mL conical tube placed on ice. Worms were allowed to sediment for 10 min and/or centrifuged at 2500 rpm for 2 min. Most M9 1X buffer was aspired without disturbing the worm pellet. A 12% alkaline hypochlorite solution (*Egg Prep* Solution) (see Materials section) was added to the tube and vortexed during ≤ = 8 min. Tubes were centrifuged at 3000 rpm for 3 min. *Egg Prep* Solution was discarded and worms were resuspended in M9 1X solution. This 'bleaching' process was repeated a maximum of three times, until an obvious decrease in the number of intact adult worms and a consequential increase in free eggs was observed. Worms were then washed with M9 1X a minimum of 3 times to eliminate any possible persisting bleach. Finally, worms were resuspended in 5 mL fresh M9 1X and incubated

Table 2.2
PCR genotyping programme

Step	Temperature (°C)	Time (min)	# of cycles
Initial denaturation	95	2	1
Denaturation	95	0.5	20-35
Annealing	42-65	0.5	
Extension	72	1/kb	
Final Extension	72	5	1
Soak	10	∞	1

Table 2.3
PCR genotyping mix

Component	Final Volume (µL)	Final Concentration
5X GoTaq ® Reaction Buffer	5	1X (1,5 mM $MgCl_2$)
PCR Nucleotide Mix	0.5	0.2 mM
Fwd Primer	0.5	0.5 µM
Rev Primer	0.5	0.5 µM
GoTaq® DNA Polymerase (5U/ µL)	0.2	1 U/25 µL
Template DNA	X	0.2-0.3 µg/25 µL
Nuclease-Free Water up to	25	

Gene (allele)	Primer design	Sequence	Tm (°C)	Ext (s)	# of cycles	Product size (bp)
ast-1(ot417)	Flanking point mutation	Fwd ccaagcccaagcctaagtc Rev ggcgcacacctattttcatt	60	60	30	793
	Mismatch	Fwd (wt) aaagtcaaagccaaacatg Fwd (mut) aaagtcaaagccaaacata Rev cggcaatattcagagatcg	58	50	20	297
unc-86(n848)	Flanking point mutation	Fwd atagcctcttcagctttctccag Rev aatctacttaggcttctgccacc	58	60	30	551
unc-86(n846)	Flanking point mutation	Fwd ttatcccagtccacgatctgc Rev gtggacccattactgctgctg	56	60	30	555
sem-4(n2654)	Flanking point mutation	Fwd gaagagagagtgggcggagc Rev gccgctaaattatctgtgtaaatgg	59	60	30	500
sem-4(n1971)	Flanking point mutation	Fwd ttccaccgtttgcagcgtttc Rev cgtcgttggagttggcataacc	64	60	30	875
hlh-3(tm1688)	External to the deletion	Fwd cgacatgttctctccgtgtttctc Rev gctgattagaggacatcatttgtgg	65	120	30	1785(wt) 541(mut)
	Internal to the deletion	Fwd gccctcccttatttggttgcc Rev ctttgccttgtttccagcagc	68	60	30	505 (wt) 0 (mut)
egl-46(sy628)	Flanking point mutation	Fwd gctcactcgctccccttcttg Rev gctttgtctttttcgggtctatcgg	64	60	30	794
	Mismatch	Fwd (wt) tacacttccaatgttctgg Fwd (mut) tacacttccaatgttctga Rev gtcggttcttggaaaagc	65	50	20	344
egl-18(ok290)	External to the deletion	Fwd caacaatccgtgagcccacc Rev cttcaaggatcggcaggacc	64	120	30	1315(wt) 617(mut)
	Internal to the deletion	Fwd ccggaagctcccaaagttgc Rev cgatagtagagcccacacgg	65	60	30	498 (wt) 0 (mut)
end-1(ok558)	External to the deletion	Fwd cgaactctgtctgctccaatcc Rev cacctgttcgatcctgcaaccc	60	120	30	1354(wt) 475(mut)
ceh-14(ch3)	External to the deletion	Fwd tctctcttgttcctccaacctg Rev cgagtagctctttatggaggac	62	120	30	1818(wt) 541(mut)
	Internal to the deletion	Fwd gcttggtgcgacatgtttcc Rev gcaagtttacggtaacgcactg	58	60	30	525(wt) 0(mut)

← **Table 2.4**
Primers and PCR specifications for HSN transcription factor candidate genotyping

overnight with gentle rocking at the desired temperature, to let eggs hatch. Since there is no food in the media larvae are arrested at L1 stage allowing for population synchronisation.
The next day, tubes were centrifuged at 2500 rpm and M9 was discarded up to 3 mL. Worms were resuspended in the remaining buffer and the desired volume (50-100 uL) was aliquoted onto seeded plates.

Worm immunohistochemistry

Antibody staining was performed using a tube fixation protocol adapted from (McIntire et al. 1992). Briefly, synchronised young adult hermaphrodites were fixed with 4% paraformaldehyde (PFA) in phosphate buffer saline (PBS), for 18 h at 4 °C. The next day worms were washed with with a solution of 1% PBS – 0.5% Triton X-100 five times and incubated for 18 h at 37 °C in a nutator mixer with a solution of 5% β-mercaptoethanol – 1% Triton X-100 – 0.1M Tris (pH 7.5). The third day worms were rinsed five times with a solution of 1% PBS – 1% Triton X-100 – 0.1 M Tris (pH 7.5) and treated with 1 mg/mL collagenase type IV (Sigma, #C5138) in collagenase buffer (1% Triton X-100 – 0.1 M Tris – pH 7.5 – 1 mM $CaCl_2$) for 90 min at 37 °C, 700 rpm. Worms were washed with a solution of 1% PBS – 0.5% Triton X-100, and proceeded to stain. Blocking solution (PBS 1X – 0.2% Gelatine – 0.25% Triton X-100) was added to the worms for 30 min at room temperature and then they were incubated for 24 h at 4 °C in primary antibody (rabbit anti-5-HT antibody 1:5000 (Sigma, #S5545)) diluted in a solution of PBS 1X – 0.1% Gelatine – 0.25% Triton X-100. The last day worms were washed five times (PBS 1X – 0.25% Triton X-100) and incubated with the secondary antibody (Alexa 555-conjugated donkey anti-rabbit (Molecular Probes, #A31572)) for 3 h at room temperature.Finally, worms were rinsed two times more, incubated in 4,6-diamidino-2-phenylindole (DAPI) (Sigma, #D95425MG), and mounted on FluorSave (Merck Millipore, #34578920ML).

Silencing of the GATA family through RNAi feeding assays

RNAi plates were prepared using *E. coli* HT115 (CGC); an RNase III-deficient strain with *iso*-propyl-β-D-1-thiogalactopyranoside-inducible (IPTG) T7 polymerase activity. HT115 clones bearing the *C. elegans* genes of interest were obtained from Dr. Ahringer library (Kamath & Ahringer 2003) (generated at the Wellcome CRC Institute, Cambridge University; distributed by Source BioScience LifeSciences) → **Table 2.5**. These bacteria were cultured overnight (15-13 h) in LB media (Sigma, #L3522) with ampicillin (50 µg/mL) at 37 °C. 3 h before seeding, the cultures were inoculated with an IPTG solution (0.6 M), in order to induce the production of double strand (dsRNA). Clones were seeded on NGM plates containing ampicillin and IPTG, at the same concentration as mentioned above.
RNAi experiments were performed to induce gene silencing of all of the members of the GATA TF family, except for *elt-4* whose clone is not available. RNAi experiments were performed by the standard feeding protocol (Kamath et al. 2001). *rrf-3 (pk1426)* background was used to sensitise worms to the RNAi effects. This mutation turns out to be lethal at 25 °C, so all experiments were performed at 20 °C (Simmer et al. 2002). *rrf-3* adult worms were transferred to IPTG plates and deposited within a drop of alkaline hypochlorite solution (*Drop Bleach*, see Materials section). Larva that survived the treatment became the parental generation (P0) which experienced post embryonic effects of the RNAi. We analysed their progeny (F1), which developed under the embryonic effects of the RNAi (F1 scoring). Whenever the RNAi was lethal at F1, we performed P0 scoring. As a negative control the empty vector L4440 (Addgene, #1654) was used. The ex-

Gene name	TF family	Library
egl-18	GATA	Julie Ahringer
elt-1	GATA	Julie Ahringer
elt-2	GATA	Julie Ahringer
elt-3	GATA	Julie Ahringer
elt-6	GATA	Julie Ahringer
elt-7	GATA	Julie Ahringer
end-1	GATA	Julie Ahringer
end-3	GATA	Julie Ahringer
med-1/med-2	GATA	Julie Ahringer
ast-1	ETS	Julie Ahringer
egl-18	GATA	Julie Ahringer
egl-46	INSM-ZnF	Julie Ahringer
sem-4	SPALT-ZnF	Julie Ahringer
unc-86	POU-HD	Julie Ahringer
ceh-14	LIM-HD	Julie Ahringer
exc-9	LIM-HD	Julie Ahringer
lim-4	LIM-HD	Julie Ahringer
lim-6	LIM-HD	Julie Ahringer
lim-7	LIM-HD	Marc Vidal
lin-11	LIM-HD	Julie Ahringer
mec-3	LIM-HD	Julie Ahringer
ttx-3	LIM-HD	Julie Ahringer
unc-95	LIM-HD	Julie Ahringer
valv-1	LIM-HD	Julie Ahringer
attf-4	FKH	Julie Ahringer
C34B4.2	FKH	Our laboratory
daf-16	FKH	Julie Ahringer
fkh-10	FKH	Julie Ahringer
fkh-2	FKH	Marc Vidal
fkh-3/4	FKH	Marc Vidal
fkh-5	FKH	Julie Ahringer
fkh-6	FKH	Julie Ahringer
fkh-7	FKH	Julie Ahringer
fkh-8	FKH	Marc Vidal
fkh-9	FKH	Our laboratory
let-381	FKH	Julie Ahringer
lin-31	FKH	Julie Ahringer
pha-4	FKH	Marc Vidal
T27A8.2	FKH	Julie Ahringer
unc-130	FKH	Julie Ahringer

Table 2.5
RNAi clones used in this work

periment was done once. A minimum of 30 worms was scored. The statistic applied was Fisher exact test, *: pV< 0.05.

Identification of PHB neuron: DiI staining

The PHB neuron, generated in the same division as HSN, is a phasmid neuron that has its cilia exposed to the environment. It has been shown that amphid and phasmid neurons can take up lipophilic dyes, such as fluorescein isothiocyanate (FITC), DiI, DiO, and DiD, from their surroundings (Hedgecock et al. 1985; Collet et al. 1998). These dyes label all parts of the neuron, from cilia to soma. Thus, we used DiI to visualise PHB neurons in the different mutant backgrounds.
A young adult synchronised population was collected from 1-2 plates using 1mL M9 1X and transferred to an 1.5 mL Eppendorf tube. Worms were rinsed twice in M9 1X (2000 rpm, 2 min) to remove bacterial contamination. Worms were resuspended in 300 µL M9 1X, transferred to a 0.5 mL tube and centrifuged one last time. 1 µL of 2 mg/mL DiI (1,1'-Dioctadecyl-3, 3,3',3'-Tetramethylindocarbocyanine Perchlorate) (Molecular Probes, #D282) diluted in N,N-dimethyl formamide (Sigma, #D4551) was added to 200 µL of M9 1X, vortexed and then added to the worms. Tubes were covered with aluminium foil and incubated 3 h in a rocking nutator. After three washes with M9 1X, worms were immediately scored at the fluorescence microscope. A Texas Red filter (585 nm) was used to visualise the dye.

Mutagenesis analysis of *cis*-regulatory modules

To predict putative TFBSs, we searched for the corresponding position weight matrixes (PWM) in both orientations 5'-3' and 3'-5' → **Table 2.6**. The specific sequences of the TFBSs were obtained from published papers and online libraries such as Transcription Factor encyclopedia (TFe), CIS-BP, JASPAR and MatInspector.

Table 2.6
Position weight matrixes used in bioinformatics prediction analysis

Gene	TF Family	PWM	Mutation	Reference
ast-1	ETS	CGGA$^{A}/_{T}{}^{A}/_{G}$	CcGA$^{A}/_{T}{}^{A}/_{G}$, CaGA$^{A}/_{T}{}^{A}/_{G}$	TFe
unc-86	POU	$^{C}/_{T}{}^{G}/_{T}$CATN$^{A/T}/_{C}{}^{A}/_{T}$ /GCCATAATAAAA-CAAT	$^{C}/_{T}{}^{G}/_{T}$gggN$^{A/T}/_{C}{}^{A}/_{T}$ / GtGtATAccA-cAACAAT	(Sze et al. 2002; Verrijzer et al. 1992)
sem-4	SPALT / MYT	TTGT$^{C}/_{G}$T / AAATTT	CTag $^{C}/_{G}$T, TTag $^{C}/_{G}$T / AAgggg	(Toker et al. 2003)
hlh-3	bHLH	$^{C}/_{G}$CAGAA / TGACGTG	tttGAA , aaAaAA / TGcCGaa	MatInspector, TFe
egl-46	INSM	$^{G}/_{T}$NN$^{A}/_{T}$G$^{C}/_{G}$GG	$^{G}/_{T}$NNA/TGaaa, $^{G}/_{T}$NNA/TaaaG, $^{G}/_{T}$NNA/Taaaa	TFe
egl-18	GATA	$^{A/G}/_{T}$GATA$^{A/G}/_{T}$	DtATAD, DGAaAD	(Merika & Orkin 1993)

Construct	Sequence
t1p26	aggaggtgtctttgtttgtgtataccacaacaagcgatcaacacagcaaag
t1p31	ttctccggatattagattaggtggcaggcggctccattg
t1p44	gtatattacgtgccgaatttttgaagcaccacgccatcggat
t1p43	caatcaacacagcaaagatttctctcaacctcatttcatgattttc
t1p60mut1	cgtttttttttctccggaaattagattgtgtggcaggc
t1p60mut2	gaagcaccacgccatcgtatatttaaaagaggaggtg
t1p52	gaaaccatgacagcaaaaataggtagagtggcgccttattcg
t1p54	gaagcaccacgccatcgtatatttaaaagaggaggtg
t1p55	cgtttttttttctccggaaattagattgtgtggcaggc
t1p58	cgccttattcgactcatttcgtttttttttcttaatatctggtccggaaattagattgtgtggca
t1p59	cggctccattgtatattacgtgccgaagcttctggaaaccacgccatcggatatctaaaagagg
c1p63	gaatcattcatcattctggtttcggttgttacccattcc
c1p61	gtgttaagcattattctttactgaatcattgggcattctggtttccgttg
c1p60	cccaccatctactgttagaaaactagcttggatccccggga
c1p73mut1	ctttactgaatcattcatcatttttttttccgttgttacccattccgccc
c1p73mut2	ccaccaaatttttcaatgttttccctgccgaaagaaaatatgaaaatatcaacg
c1p71	gcccgttggtttctcttctcctgcttttaccatctactgttagaaaattg
c1p74	gttctttcaagtttatatcaacaaaatataaattccagtttttttttgatagcg
c1p75	caacaaaagataaattccagtttttttttttatagcgtgtcatacagtatg
c1p76mut1	caacaaaatataaattccagtttttttttttatagcgtgtcatacagtatg
c1p76mut2	gttctttcaagtttatatcaacaaaatataaattccagtttttttttgatagcg
c1p79mut1	ccgtgggaatctaaaacgtgtttttgctcgcttattcttgcgtatagac
c1p79mut2	gcccgttggtttctcttctcctgcttttaccatctactgttagaaaattg
c1p79mut3	gcgatgaacatagtgggtacttgttagaaatcgggccaaaatatcacc
c1p83	cgcagttttgttctttcaagtttatatcggcaaaagataaattccag
c1p85	cattattctttactgaatcattcataaccagaatgggtttccgttgttacccattccgcccgttgg
c1p86	cttctcctgcccccaccatctactgttagaacaagacaattggatccccgggattggccaaagg
c1p87	gcattattctttactgaatcattcatcattctggaacggaaaccgttacccattccgcccgttggtttctc
b1p73	ctcattctcaaaccagtttctatcggtttgtttgcattcaattaa
b1p71	ccagtttctatccgtttgtttggggtcaattaaatttttttttttcagcgta
b1p65	gaagaatacgctgaaaaaaacccccttaattgaatgcaaacaaacggatag
b1p78	catctcattctcaaaccagtttctttccgtttgtttgcattcaattaa
b1p77	cctatccccggctttctgtttgaattccagtaacacattgatattc
b1p76	ccagtaacacattgatattcttctttaacaccacattattcatgtatttcctcc
b1p83	cgcaaacgttttggagaatatagacaactttaggaagtcatc
b1p84	ccagaattccagtaacacatttatattcttccccaacaccac
b1p86mut1	ccagaattccagtaacacatttatattcttccccaacaccac
b1p86mut2	catctcattctcaaaccagtttctttccgtttgtttgcattcaattaa
b1p86mut3	cgcaaacgttttggagaatatagacaactttaggaagtcatc
b1p87	cgatcactatcctatccccggtggcagaaagtgccagaattccagtaacacattgatattcttcccc
b1p89	ccatctcattctcaaaccagtttctttccgttacagatagattgtttgcattcaattaaatttttt

Table 2.7
Primers for site-directed mutagenesis (forward sequences)

Electrophoretic Mobility Shift Assay (EMSA)

Full-length *unc-86* (kindly provided by Dr. Hobert) and *ast-1* cDNA were cloned into the pET-21b His tag expression vector (EMD Millipore, kindly provided by Dr. Hobert). They were transformed into *E. coli* Rosetta2(DE3) (Novagen, #71400) strain. Overexpression was done first by growing the cells at 37 °C in LB and PowerBroth medium (Molecular Dimensions, #MD121061) respectively, supplemented with 100 µg/mL ampicillin and 100 µg/mL chloramphenicol to OD_{600} = 0.5-0.6, and then inducing expression with 0.5 mM IPTG (Acros Organics, #BP1755100) at 37 °C for 3 h or 20 °C for 16 h, respectively.

UNC-86 protein was obtained as previously explained (Zhang et al. 2014) with minor changes. Briefly, cells were collected by centrifugation and resuspended in buffer A (100 mM NaH_2PO_4 – 10mM Tris (pH 7.5) – 10% glycerol) supplemented with 1 mM phenylmethanesulfonyl fluoride (PMSF). Cells were lysed by sonication and soluble and insoluble fractions were separated by centrifugation and analysed by SDS/PAGE. Protein was subtracted from insoluble fraction as follows: the insoluble fraction was resuspended in solubilisation buffer (buffer A supplemented with 8 M urea) and loaded on a pre-equilibrated His Trap HP column (GE Healthcare Life Sciences, #17524801). The resin was washed with solubilisation buffer supplemented with 10 mM imidazole, and protein was eluted with the same buffer supplemented with 500 mM imidazole. Elution buffer was exchanged by progressive dialysis to 20 mM HEPES (pH 7.5) – 100 mM NaCl – 10% glycerol – 2 mM $MgCl_2$, and the protein was concentrated by centrifugation up to 1.3 µg/µL and stored at -80°C.

For AST-1 protein, cells were collected by centrifugation and resuspended in buffer B (200 mM MES (pH 6.0) – 500 mM NaCl – 2 mM $MgCl_2$ – 10% glycerol) supplemented with 1 mM phenylmethanesulfonyl fluoride (PMSF). Cells were lysed by sonication and soluble proteins were loaded on a His Trap HP column pre-equilibrated with buffer B. The resin was washed with buffer B supplemented with 10 mM imidazole, and protein was eluted with buffer B supplemented with 300 mM imidazole. Eluted fraction was analysed by SDS/PAGE. Imidazole was removed and protein concentrated by centrifugation up to 0.3 µg/µL, and stored at –80°C.

egl-18 cDNA was cloned into pcDNA.3 vector followed by His tag sequence and transfected with Lipofectamine-2000 (Invitrogen, #11668019) in HEK293T cells (kindly provided by Dr. Hobert, #ATCC:CRL-3216). HEK293T cells were grown in DMEM – 10% FBS. After 24 h, cells were lysed with the following buffer: 1mM EDTA – 0.5% Triton – 20 mM β-glycerol – 0.2 mM PMSF – 100 µM Na_3VO_4 – protease inhibitor.

EMSAs were performed incubating UNC-86 and AST-1 proteins in a buffer containing the labelled probes and 10 mM Tris (pH 7.5) – 50 mM NaCl – 1 mM $MgCl_2$ – 4% glycerol – 0.5 mM DTT – 0.5 mM EDTA – 1µg of poly(dIdC) – 6 µg of bovine serum albumin (BSA) and labelled probes for 20 min at room temperature. For EGL-18, protein extracts were incubated in 20 mM HEPES – 50 mM NaCl – 5 mM $MgCl_2$ – 5% glycerol – 1 mM DTT – 0.1 mM EDTA – 1 µg of poly(dIdC) – 6 µg of BSA – 1 µg anti-6xhistag antibody (Abcam, #ab18184) at 4 °C for 30 min. As negative control anti-GFP antibody (Roche, #11814460001) was used. Then labelled probes were added and incubated for 20 min at room temperature. For AST-1 and UNC-86, 1 µl (30 ng/µl) labelled probe was added, and for EGL-18 4 µl (30 ng/µl) were added. Finally, samples were loaded onto a 6% (37.5:1 acrylamide: bisacrylamide) gel and run at 150 V for 4 h. Gels were then dried and visualised using Fujifilm FLA-500. Probe sequences are listed in → **Table 2.8**. Primers were annealed and end-labelled with ATP (γ-32P) (Perkin Elmer, #NEG502A250UC) using T4 PNK (Thermo

Protein	Probe	Size (bp)	Sequence
AST-1	*tph-1* wt	44	cgttttttttctccggatattagattgtgtggcaggcggctcc
	tph-1 mut	44	cgttttttttctcttagattgtgtggcaggcggctcc
	cat-1 wt	41	tttactgaatcattcatcattctggtttccgttgttaccca
	cat-1 mut	35	tttactgaatcattcatcattctggttgttaccca
	bas-1 wt	48	ctcattctcaaaccagtttctatccgtttgtttgcattcaattaaatt
	bas-1 mut	42	ctcattctcaaaccagtttctttgtttgcattcaattaaatt
UNC-86	*tph-1* wt	40	gtgtctttgtttgcgcataataaaacaatcaatcaacaca
	tph-1 mut	40	gtgtctttgtttgtgtataccacaacaagcgatcaacaca
	cat-1 wt	41	tttactgaatcattcatcattctggtttccgttgttaccca
	cat-1 mut	41	tttactgaatccccgggcattctggtttccgttgttaccca
	bas-1 wt	57	aaaccagtttctatccgtttgtttgcattcaattaaatttttttttttcagcgtattc
	bas-1 mut	56	aaccagtttctatccgtttgtttggggtcaattaaatttttttttttcagcgtattc
EGL-18	Control	31	catttatatcagccgtttttatcttttcctg
	tph-1.1 wt	97	ttttttctccggatattagattgtgtggcaggcggctccattgtatattacgtgccgaattccagaagcac-cacgccatcggatatctaaaagagga
	tph-1.1 mut	97	ttttttctccgtatattagattgtgtggcaggcggctccattgtatattacgtgccgaattccagaagcac-cacgccatcgtatatttaaaagagga
	cat-1 wt	54	gtttatatcaacaaaagataaattccagttttttttttgatagcgtgtcatacag
	cat-1 mut	54	gtttatatcaacaaaatataaattccagttttttttttatagcgtgtcatacag

Table 2.8
Probe sequences for EMSA analysis (forward sequences)

Scientific, #EK0031) according to the manufacturer's specifications.

In vivo transcription factor expression analysis

For TF developmental expression analysis different reporter strains were used → **Table 2.21**. To assess UNC-86, HLH-3, SEM-4 and EGL-18 expression we used fosmid reporters (kindly provided by Dr. Hobert and CGC). For SEM-4 we also analysed a truncated translational reporter that encodes approximately half of SEM-4 fused in frame to GFP, whose expression was significantly more intense than the fosmid strain. For *egl-46* we injected a transcriptional reporter that covered all the intergenic region (4,477 bp) (kindly provided by Dr. Pocock). The injection mix consisted of 50 ng/µL of the pNF303 plasmid (*egl-46prom::NLS::DsRed* in pPD96.04) and *rol-6 (su1006)* (pRF4) at 100 ng/µL as an co-injection marker (final concentration: 150 ng/µL). Finally, for AST-1 we generated a CRISPR/Cas9 mediated GFP knock-in (detailed in the next section).
Worms were scored at all developmental stages, from embryo to adult. Whenever reporter fluorescence intensity was low, identification of the HSN nucleus using DIC previous to fluorescence scoring was required. HLH-3 expression was only detected in embryonic stages. Embryos were selected and mounted at 1 to 4 cell-stage (0 hours post-fertilisation (hpf) to 1.25 hpf, respectively), incubated at 25 °C and analysed at 5 hpf. HSN neuroblast precursor cells were identified relative to nearby landmark cell deaths (Sulston et al. 1983). Scoring and images were performed using 60X objective in a Zeiss Axioplan 2 microscope. Two-tailed Fisher's exact test was used for statistical analysis.

Generation of fluorescent reporters via CRISPR/Cas9

ast-1 gene reporter strain was generated using CRISPR/Cas9-mediated GFP knock-in strategy, as described in (Dickinson et al. 2015; Dickinson et al. 2013). Plasmids used were kindly provided by Dr. Boxem. The protocol followed is described below and depicted in → **Figures 2.2-2.4**.
We first designed primers to amplify PCR products flanking *ast-1* stop codon, which would serve as homology arms. To increase the efficiency, we did nested PCRs → **Table 2.9**. As template we used N2 genomic DNA, purified using the DNeasy Blood & Tissue Kit (Qiagen, #69504). Expand Long template PCR system (Sigma, #11681834001) and Q5 Hot Start High-Fidelity 2X Master Mix (NEB, #M049S) polymerases were used. PCR products were purified using QIAquick PCR Purification Kit (QIAGEN, #28106). One PCR product (5' homology arm) included the last 715 bp of the *ast-1* gene except for the STOP codon, while the other (3' homology arm) included the *ast-1* stop codon and the following 653 bp. In this way, the desired fluorescent protein (FP) would be inserted in frame at *ast-1* C-terminus. Moreover, both homology arms contained a flexible linker (GGAGCATCGGGAG CCTCAGGAGCATCG) that will separate *ast-1* from FP::3XFlag in 9 amino acids, and sequence overlaps that will allow for Gibson assembly. For homology arm recombination we used the pJJR82 plasmid (Addgene, #75027) containing GFP, *ccdB* negative selection markers and a self-excising selection cassette. Homology arms were inserted in these *ccdB* sites (SEC) through Gibson assembly cloning method (Gibson et al. 2009). The mix consisted of 4 µL 2X AMM (see Materials section), 1 µL 5' homology arm, 1 µL 3' homology arm, 2 µL plasmid. TOP10 electrocompetent cells (Invitrogen, #C404010) were transformed with 1 µL of the mix and several clones were sequenced to confirm the presence of both homology arms in the plasmid.
In order to choose the Cas9 target site two CRISPR design tools were used: http://crispr.mit.edu and http://benchling.com, both based in the Zhang Laboratory's Method (Hsu et al. 2013). We selected as single guide RNA (sgRNA) the GGGGTGACTATCGATAAAGA sequence, which overlapped with the stop codon and showed 100 % specificity (only 1 off-target: B0001.2 gene, known to be involved in embryonic development). Through site-directed mutagenesis (Quickchange II XL site-directed mutagenesis kit, Agilent Technologies #200522) we introduced the Cas9-sgRNA into the pDD162 plasmid (Addgene, #4754) and confirmed its presence by sequencing.
60 N2 worms were injected with the following injection mix: 50 ng/µL Cas9–sgRNA plasmid, 10 ng/µL repair template, and 2.5 ng/µL pCFJ90 (*Pmyo-2::mCherry* pharyngeal co-injection marker, Addgene, #19328). Worms were grown in individual plates at 25 °C for 3 days, then filter-sterilised hygromycin B solution (final concentration of 250 µg/mL; Gibco, #10687010) was added to the plates to select initial knock-in worms. Plates with candidate worms were obvious: many animals survived the antibiotic, showed roller phenotype (due to the presence of the *sqt-1(e1350)* dominant allele in the plasmid) and expressed no red co-marker (loss of extrachromo-

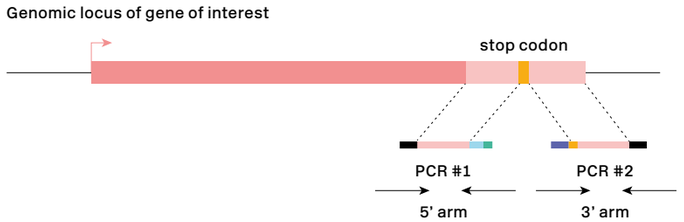
somal arrays). 20 worms from each candidate plate were picked and singled to new plates. Plates were all descendants were roller were selected as homozygous for the insertion. Only one line form each plate in the previous step was kept as independent lines, because it is impossible to tell whether two strains that originated from the same injection plate derive from independent insertion events or a single insertion event.

To remove the *rol* selectable marker, 8-10 L1-L2 larvae were picked to three new plates from several homozygous independent lines and heat shocked at 34 °C, for 4 h. This activated the expression of *hsp-16.41prom::Cre*, which recognised LoxP sites that flank the Cre, hygromycin and *sqt-1* sequences. 5-7 days later, F1 progeny L4 non-roller worms were selected as candidate knock-in worms that had lost the selectable cassette.

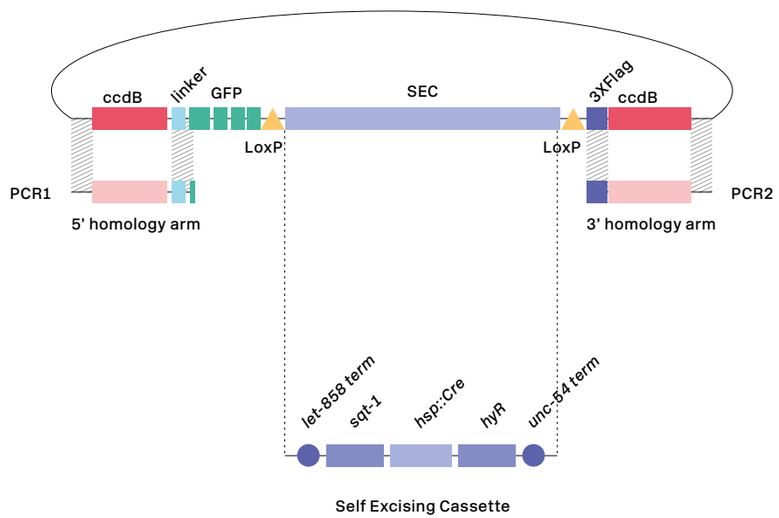
HSN fate maintenance assays. RNA interference (RNAi) by feeding at the adult stage

RNAi experiments were performed, as previously described, but to induce gene silencing of our candidate TFs after HSN maturation, i.e. adult stage → **Table 2.5**. A synchronised population of ani-

Figure 2.2
Genome engineering using a self-excising drug selection cassette. Adapted form (Dickinson et al. 2015).



A) In order to tag the gene of interest in the C-terminus, two PCRs are required; PCR 1 will amplify a region of 500-700 bp just before the gene stop codon, while PCR 2 will amplify a region of 500-700 bp starting with the stop codon. Primers (arrows) will add overhangs that overlap with the GFP-SEC vector.



B) Schematic of an expedited cloning procedure for insertion of homology arms into a GFP–SEC vector (pJJR82). The GFP–SEC vector is first digested with restriction enzymes to release the ccdB markers, and 500–700 bp homology arms (light pink) are inserted by Gibson assembly to generate the repair template plasmid. Grey angled lines indicate overlapping DNA. The self-excising cassette (SEC) for drug selection consists of a hygromycin resistance gene (hygR), a visible roller marker (*sqt-1(d)*), and an inducible *Cre* recombinase (*hsp::Cre*). SEC is flanked by LoxP sites and placed within a synthetic intron in an GFP::3xFlag tag, so that the LoxP site that remains after marker excision is within an intron.

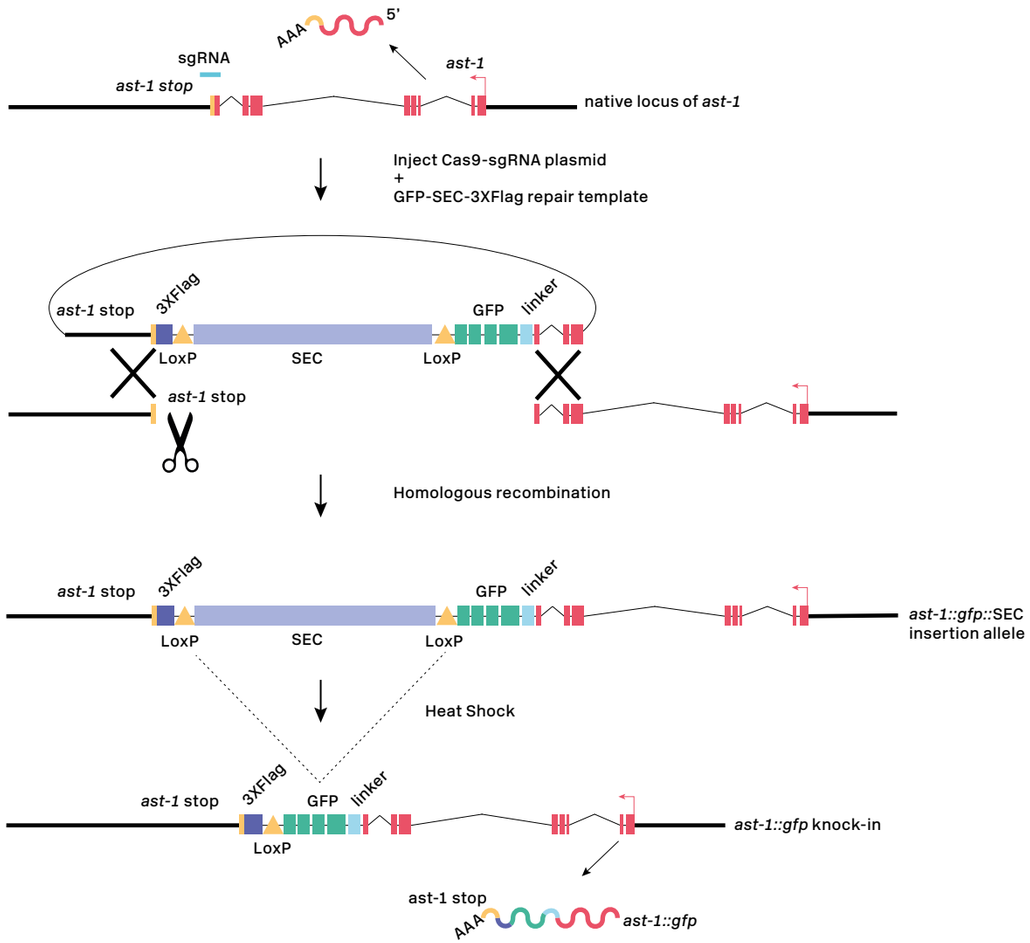
Figure 2.3
GFP *ast-1* gene tagging protocol using CRISPR/Cas9 and a self-excising drug selection cassette. Adapted form (Dickinson et al. 2015).

A) Endogenous locus of the *ast-1* gene; coding exons are in red and the stop codon appears in yellow. A small diagram shows the resulting mRNA of the gene. The single guide RNA (sgRNA, blue line) containing a PAM sequence that will be recognised by CRISPR/Cas9, is designed to overlap the stop codon region. Wild type worms are injected with a mixture of the GFP-SEC-3XFlag plasmid containing the homology arms previously introduced by Gibson assembly (repair template), and the Cas9-sgRNA plasmid (Figure 2.4).

B) Cas9 recognises the PAM sequence at the STOP codon of *ast-1* and cuts the DNA. Homology arms in the repair template allow for homologous recombination with the worm endogenous *ast-1* locus.

C) Homologous insertion strains are recognised by choosing animals that segregated 100% *rol* progeny, due to the presence of the *sqt-1(d)* marker in the SEC cassette.

D) Heat shock treatment activates the *hsp::Cre* present in the SEC cassette, which recognises the LoxP sites and excises the cassette. In this way, an *ast-1::gfp* knock-in is generated, that will be transcribed into an *ast-1::gfp* fused mRNA containing a flexible linker (blue) between the *ast-1* (red) and the *gfp* (green) cDNA, and a 3XFlag (purple) before the stop codon (yellow).



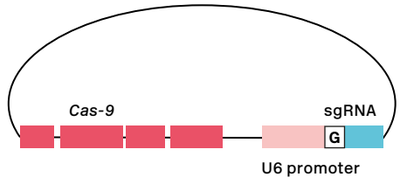


Figure 2.4
Cas9-sgRNA plasmid used in this Thesis

pDD162 plasmid contains the Cas9 cDNA, plus the sgRNA that will recognise the desired genomic target (i.e. C-terminus). sgRNA is expressed under the U6 small nuclear RNA promoter, which drives transcription by RNA polymerase II (Dickinson et al. 2013).

mals (alkaline hypochlorite treatment) were grown under normal food (OP50) until young adult stage, when the HSN has already matured. At this time point the worms were scored for the 5-HT pathway gene reporter *tph-1::yfp* (*otIs517*), and then they were transferred to RNAi plates. These plates contained dsRNA for *ast-1*, *unc-86*, *sem-4*, *egl-46* and *egl-18*. As a positive control dsRNA against GFP was used. As a negative control empty vector L4440 (Addgene, #1654) was used. Animals were incubated for 72 h at 15 °C before the final scoring. The experiment was performed in three independent replicates with similar results. A minimum of 50 worms was scored. The statistic applied was Fisher exact test, *: pV< 0.05.

Heat shock overexpression experiments

For overexpression of the six candidate TFs, cDNAs were obtained from total worm RNA extraction (Genelute Mammalian Total RNA kit (SIGMA, #RTN70-1KT)), followed by reverse transcription (Quantitec reverse Transcription Kit, (QIAGEN, #205311)) and PCR amplification. In the case of *unc-86*, the genomic locus was amplified from N2 genomic DNA (DNeasy Blood & Tissue Kit (Qiagen, #69504)). In all cases, cDNA or gene locus were cloned into the pPD49.78 (Addgene, #1447) expression vector, which includes the heat shock inducible promoter from *hsp-16.2* gene and the 3' UTR of muscle myosin *unc-54*. In the case of *egl-18*, (*hsp-16.2::egl-18 cDNA::3'UTR unc-54*) the backbone plasmid is pPD49.83 (kindly provided by Dr. Pocock). For *ast-1*, *otIs198 (hsp-16.2::ast-1*, *hsp-16.2::NLS::mCherry*, *ttx-3::DsRed*) integrated strain was used. Details of the primers and enzymes utilised are found in → **Table 2.10**. Plasmids were injected, alone or in combination, into N2 strain together with the co-markers *rol-6(su1006)* and *ttx-3::mCherry*. pUC19 was used as filler DNA when necessary. Different plasmid concentrations were used depending on the toxicity of the array → **Table 2.11**. Resulting extrachromosomal strains were crossed with *zdIs13(tph-1::gfp)*. This reporter is initially expressed in the NSM and ADF neurons in the embryo at comma stage. *ast-1* overexpression strain *(otIs198)* carries an extra internal control (*hsp-16.2::NLS::mCherry*) to validate the heat shock experiment. All strains generated are listed in → **Table 2.21**.

Analysis of the effect of transcription factor ectopic expression at embryonic stages

Gravid hermaphrodites were placed in an M9 1X drop on a glass slide and they were sectioned to release the eggs with the help of a 0.3 mm × 13 mm needle (BD Microlance 3, #304000). 1-2 cell embry-

Table 2.9
Primers for CRISPR-Cas9 mediated GFP knock-in

Use	Primer sequence
Amplify 5' homology arm. External.	Fwd tgctcctgatttctcatcgtgg Rev tcgataaagagggaatgctcg
Amplify 5' homology arm. Nested.	Fwd acgttgtaaaacgacggccagtcgccggcacgatctctgaatattgccggg Rev catcgatgctcctgaggctcccgatgctcctcgataaagagggaatgctcgtg
Amplify 3' homology arm. External.	Fwd tagtcacccccataattcct Rev gcgagacccaccaaattgattc
Amplify 3' homology arm. Nested.	Fwd cgtgattacaaggatgacgatgacaagagatagtcacccccataattcctcc Rev ggaaacagctatgaccatgttatcgatttcattgattccgtgcgcctttg
sgRNA	Fwd ctcctattgcgagatgtcttgggggtgactatcgataaagagttttagagctagaaatagcaag Rev cttgctatttctagctctaaaactctttatcgatagtcacccccaagacatctcgcaataggag

Table 2.10
Plasmids for the overexpression of HSN transcription factor candidates

Plasmid	Construct	Primers used to amplify cDNA	Enzymes	Backbone
pNF197	*hsp-16.2::unc-86* genomic	Kindly provided by Dr. Hobert	NheI / NcoI	pPD49.78
pNF204	*hsp-16.2::sem-4* cDNA	Fwd tagagagctagcatgaatgagctgctcgc Rev tagagaggtaccctaagagggtggtgg	NheI / KpnI	pPD49.78
pNF283	*hsp-16.2::hlh-3* cDNA	Fwd gagagagctagcatgaccgcatccacctc Rev gagagaggtaccttaataagtttctgtatgcg	NheI / KpnI	pPD49.78
pNF284	*hsp-16.2::egl-46* cDNA	Fwd gagagagctagcatggtgcctatgaatg Rev gagagaggtaccttacattgttggaataac	NheI / KpnI	pPD49.78
pNF314	*hsp-16.2::egl-18* cDNA	Kindly provided by Dr. Pocock	(–)	pPD49.83

Gene	TF cDNA (ng/µL)	*ttx-3::mCherry* (ng/µL)	*rol-6*(su1004) (ng/µL)	pUC19 (ng/µL)	Total concentration (ng/µL)
unc-86	50	25	25	50	150
sem-4	50	50	50	0	150
hlh-3	50	50	50	0	150
egl-46	50	50	50	0	150
ast-1	integrated	integrated	integrated	integrated	integrated
egl-18	50	50	50	0	150
Combo A+U+S	50	25	25	0	200
Combo 6	15	50	50	10	200

Table 2.11
Injection mix for the overexpression of HSN transcription factor candidates

Primer name	Primer sequence
C	agcttgcatgcctgcaggtcg
D	aagggcccgtacggccgacta
D*	ggaaacagttatgtttggtata

Table 2.12
Standard primers for Fusion PCR. From (Hobert 2002)

os were isolated by aspiration with a manually pulled micropipette (Blaubrand intraMARK, #6121414), and mounted in a 4% agarose slide sealed with Vaseline to avoid dehydration. Embryos were incubated at 20 °C for 4 h, transferred to a 37 °C incubator for 20 min (heat shock), and moved back to 20 °C. 22 h later embryos carrying the *ttx-3::mCherry* co-marker were scored for *tph-1::gfp* reporter expression at the fluorescent microscope. To note, animals bearing the different heat shock arrays and that received the heat shock treatment experienced developmental problems and did not overpass the embryonic stage. Control animals that received the heat shock but did not contain the heat shock construct developed normally. Mean, standard deviation (SD), standard error of the mean (SEM) and distribution of the population were calculated. To calculate normality in the populations the D'Agostino & Pearson omnibus normality test was performed. For statistics, the non-parametric Kruskal-Wallis test with Bonferroni correction was applied. A minimum of 50 embryos were scored. To calculate the percentage of embryos that respond to the heat shock treatment, Fisher exact test 2 tailed was performed, pV < 0.05.

Analysis of the effect of transcription factor ectopic expression at larval stages

A L1 synchronised population of worms (*Egg Prep*) was plated and incubated for 2 h at 20°C. Worms were transferred to a 37 °C incubator for 30 min and then moved back to 20°C. Heat shocks were replicated every 2 h, 3 times. The next L2 stage worms were scored for ectopic gfp expression in other neurons or tissues. Similar to embryonic overexpression, animals bearing the different heat shock arrays and that received the heat shock treatment at larval stage 1 experienced developmental problems and did not reach late larval stages. Control animals that received the heat shock but did not contain the heat shock construct developed normally. A minimum of 30 worms was scored at each age. For statistics, Fisher exact test 2 tailed was performed, pV < 0.05.

Chapter III

Bioinformatics analysis

Unless otherwise indicated, all analyses were performed using R (The R Team 2016) and Bioconductor (Huber et al. 2015).

HSN regulatory signature: 'sliding window' analysis

For HSN regulatory signature analysis, we built PWMs from the functional motifs found in the 5-HT pathway genes CRMs → **Figures 3.2.8-3.2.10**. Next, we downloaded upstream and intronic gene regions from WormBase version 220 and classified genes in three groups: genes expressed in HSN, genes expressed in neurons (according to WormBase annotations on gene expression and (Hobert et al. 2016)) and non-neuronal genes. PWMs were aligned to genomic sequences and we retrieved matches with a minimum score of 70%. To increase specificity, we removed all matches that did not bear an exact consensus sequence for the corresponding TF family and obtained the following PWMs: ETS: C/TA/TTCGG, GATA: A/G/TGATAA/G/T, HLH: C/GCAGAA, INSM: CCC/GCA/TNNA/C, SPALT: TTGTC/GT, POU: A/TTG/TCAT → **Figure 3.3.1**. Then, we performed a sliding window search to find regions that included at least one match for each TF type. Embryonic stem cell enhancers median size has been reported to be around 800 bp (Parker et al. 2013). Therefore, the initial search was performed with a maximum length restriction of either 600, 700 or 800 bp. Differences between HSN expressed genes and other gene groups was greater when the maximum length was set to 700bp, thus we kept this maximum window length for the rest of the analyses. In order to assess signature conservation, we performed similar analyses using other nematode genomes also available from WormBase *(C. briggsae, C. japonica, C. remanei, C. brenneri)*. We selected for the conservation analysis *C. elegans* genes with orthologues in at least two additional species and considered the signature as conserved if HSN regulatory windows were found in all orthologous genes.

Gene Ontology (GO) analysis

Gene Ontology analysis was performed using GOrilla software and *C. elegans* coding genome (19.276 genes) as control list (Eden et al. 2009).

Generation of reporters for *de novo* identification of HSN expressed genes

Reporters for HSN regulatory signature analysis were generated by fusion PCR (Hobert 2002) → **Figure 2.5**. Briefly, regulatory windows plus 50 bp flanks were PCR amplified from genomic N2 DNA and fused to GFP in the pPD95.75 expression vec-

tor. A list of tested windows and the primers used are listed in → **Table 2.12** and **2.13.** Expand Long Template Polymerase (Roche, #11681834001) was used. PCR products were injected at 50 ng/µL into wild type N2 strain together with 100 ng/µL *rol-6 (su1006)* (final concentration: 150 ng/µL). The resulting transgenic lines were analysed under the fluorescent dissecting scope and the 3 that showed strongest GFP were selected for scoring. A minimum of 30 worms was analysed per line.

HSN regulatory signature syntax analysis

Syntactic rule detection was performed with iTF software using the regulatory windows found in known HSN expressed genes as input and the short consensus sites (ETS: YWTCGG, GATA: DGATAD, HLH: SCAGAA, INSM: CCSCWNNM, SPALT: TTGTST, POU: YKCATNHW) as PWMs (Kazemian et al. 2013). To test syntax functionality, mutagenesis was performed over the *cis*-regulatory modules *tph-1prom2*, *cat-1prom14* and *bas-1prom13*, in order to invert BSs orientation. Specific nucleotidic changes are described in → **Annex 3.3.3**. When flipping one core site, two flanking nt were usually also considered. If two functional BSs overlapped (as in ETS-GATA), one was maintained in the original orientation and the other one was flipped and placed next to it, leaving a 2 nt spacing between them. Mutated plasmids were injected at 50 ng/µL into

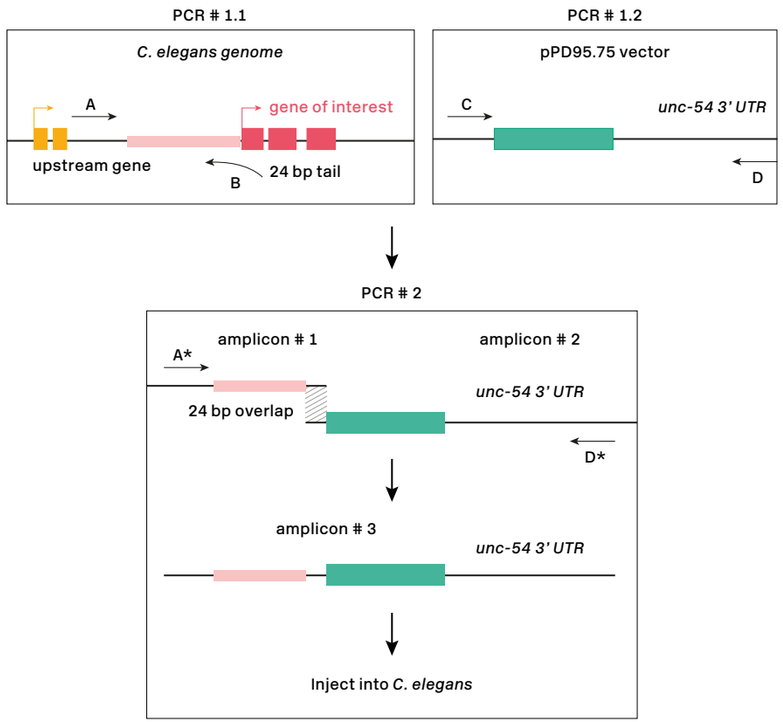


Figure 2.5
Fusion PCR protocol

Two templates are needed: worm genomic DNA and the pPD95.75 vector. Note that, in the schematic, the fusion is made to the enhancer or promoter of the gene (i.e. transcriptional fusion), but it could also be made to the coding sequence of the gene (i.e. translational fusion). Pink box represents the upstream regulatory region of interest to fuse to *gfp;* red boxes represent coding exons of the gene of interest; green box represents *gfp* coding DNA; small black arrows represent the primers used in this protocol. Primer A is 5' upstream to the DNA of interest; primer A* is nested to A; primer B spans the end of the enhancers/promoter of interest + 24 nucleotides of the *gfp* pPD95.75 vector. Primers C and D, flank the *gfp* cDNA plus *unc-54* 3' UTR of the pPD95.75 vector. PCR 1.1 renders amplicon 1, while PCR 1.2 renders amplicon 2. PCR 2 uses as template amplicons 1 and 2 and, taking advantage of the 24 overlapping nt conferred by primer B in PCR 1.1, renders amplicon 3 *(DNA of interest::gfp).* Purified PCR product can be directly injected into the worms. Adapted from (Hobert 2002).

Table 2.13
Primers used to amplify windows to test *de novo* expression in HSN

Forward primers correspond to primer A* and reverse primers corresponds to B in Figure 2.5. All primers B contain an additional 24 bp tail (agtcgacctgcaggcatgcaagctt) before the sequence included in the table.

Gene name	Primer sequence	
C16B8.4	Fwd	ggatgtggaacatgaatccgattg
	Rev	ggggtagaatgggtgaaaaaaaattgt
dgn-1	Fwd	gtgaaaaaagtctttcatcgcaaaacc
	Rev	gcaacacgcatacacacacaa
kel-8	Fwd	atcgtaaacataaacaatgcacccg
	Rev	gggtccaaaaaaatcgtatttttgcg
F37A8.5	Fwd	catgtgatgtgagaattcccaatgg
	Rev	tcgtcatgaatttatttatatattggtacaagcg
npr-3	Fwd	acaaatcaaaaccgcaaaaaacagg
	Rev	acaagcgatatggcatggacc
tkr-2	Fwd	ttaattgctacttaatcgagaagttcgaac
	Rev	agtgaaaaaattatttgaatggccacatc
twk-17	Fwd	ggtttacagctttgaagactagtaagc
	Rev	tcactgatgtctagacttaagcaagattc
tyra-3	Fwd	agtgcgtgtatgtctctaacatctaag
	Rev	gaaatctagttatcggttaggttaatttcgg
lgc-49	Fwd	atgaaccctctttcacttttggc
	Rev	cgaacaaggaacatgtctgtaatgtg
pde-3	Fwd	cgtacaatttttttttggaaaaatcaaaaaaaattaagc
	Rev	actgtgacttttttgaaatttttttcccg
daf-38	Fwd	catagaaatttctgatgatgagcctcg
	Rev	tgccttaaaaacatagtgaaacgttagg
fut-1	Fwd	tttttttttgaaaacatcgtcaccgc
	Rev	tttcctgtttttacagtgattcacataagc
kcc-1	Fwd	cagaaatccatccactgaaacagt
	Rev	caaaaacccagcagagcgtg
klp-7	Fwd	gtctaattgcctctgatatgttttgacc
	Rev	gatgttatgaaatggcgagtactgc
sprr-1	Fwd	cgtcattgtcttggtttgtctcg
	Rev	tcttgggaaaaaatatgttaagaaggtgc
unc-32	Fwd	ctcatgattcatttctctccttattggc
	Rev	cgcttctttcgagagaaaacatttaaga
C53B4.4	Fwd	gcttcaggaagcgaacaagc,
	Rev	ggagcacacgtttctaagatggg
ckr-2	Fwd	ctaagatttgctcctcttattttagccg
	Rev	catcttcttcttcttccccttcttcta
F32D8.10	Fwd	gaattttcatgattcatcttgaaaatcccc
	Rev	cctgggtagttgatcttctacttgaag
glb-20	Fwd	cggagagtgagaaagagagagagg
	Rev	agggtaaaaagtattcaaaactcaacagg
mgl-2	Fwd	agtgtaatgaaccaaaaaataggcgg
	Rev	acaaaaaagcgttccaattcctcg
snt-1	Fwd	aagaaaaggttatgcaacaaactggg
	Rev	gcgagaaccagcagaataaatacg
unc-7	Fwd	gactgagctatcctgcctgc
	Rev	tcaatgcaagaaagacacgcg
ast-1	Fwd	ggtaaattcccaaattttggccaaac
	Rev	tcaattgcatagaacaattaggttatgcc
acr-24	Fwd	aatctgatcaattctaaactatttttttcaacacc
	Rev	aggtttgaaacatttttttgaaacaaaatttgaag
gab-1	Fwd	ccgaggatcagttatgtgaagagtattg
	Rev	cagacacgaagaaattcattacgaaaatcg

↳

mec-10	Fwd	attgattgcactaataatccactggc
	Rev	tcctgttttagctcaaaatacgtgc
npr-1	Fwd	aatagctctgaatcattctaaaacgcc
	Rev	attgaatttggaacgagtaatgtgtagg
tol-1	Fwd	ccttgttaccttgactatcgggaa
	Rev	cctaattagtagtcacgagaagagcag
bam-2	Fwd	ggaacaagtcaaggtgttcatagaaatag
	Rev	acattcattggacgcgtacaatttc
shl-1	Fwd	gttcgaaaattttggaattgactaattttatccg
	Rev	gctagcacctattgagcaggaa
sto-5	Fwd	gtcactccgaggttctggc
	Rev	agaagaagaatgtacagatatagtgcgc
tiam-1	Fwd	actactcgagttgagtgtgttgc
	Rev	ctgttatggcaagttgaaactggag
pan-1	Fwd	cgaaagtacaacaatgattcctcatagg
	Rev	gatcgtaaaatcttaattcacagaagctcg
abts-4	Fwd	tcctcactctgtcatttgaatgttctc
	Rev	aactgttatcatatctcacacataatttacgc
cat-1	Fwd	gcatttagcagtccatgtttaggc
	Rev	gactatagctggagtccgcg
aak-2	Fwd	tcatcggaccaacttcccc
	Rev	cccactaaattttcctgtagtttcagc
kal-1	Fwd	tcgctaaaaaaatctcttgaagtctgc
	Rev	aaacatctgtactagtccgggttc
kcc-2	Fwd	atttaacaacattgcaaaacagaagaagtc
	Rev	cagaattggtattaataacgggatgaaagg
sem-4	Fwd	aactcttaatgttttgttgcgaccc
	Rev	ggtcgtaaaaatcgcaacaaaccg
F16G10.5	Fwd	cctaactggatgactcagtaccaaaaag
	Rev	atgaataccttcttcaatccaacagaaag
flp-27	Fwd	gcaaatcgacaatttgccgaaaatg
	Rev	acctctgtgaaagccacgc
gipc-2	Fwd	ggcgtcaactaacaatgacgtg
	Rev	atgtatagatttttgcctcaaaatccagg
irld-53	Fwd	gaaagagctaccactaaacgaaacatg
	Rev	agtttttttaaaaatgatttttgggaaattggaaaa
irld-62	Fwd	gtaaaaatcctaaacattaggatagtttttgatgtg
	Rev	caattgaacaaactataatatttttcatgaaaaaatactttaaaacc
lurp-2	Fwd	acatcgcagcgacaaagtttttg
	Rev	ctcgaacgttagagcctccttg
plep-1	Fwd	ccaatacatttccagttcaaaaaagtttttttttaatac
	Rev	tctgaatatttttttgtgaaatattgaaaaaactcttcg
slc-28.1	Fwd	ttcctagcggataaattcaaagttttttaatg
	Rev	agcttttggtaactgaaatcagatttttttc
stg-1	Fwd	gaaactttcaaattagctgaatcagttgattttc
	Rev	ttttaaaacaaatatgacggggcttttcg
tub-1	Fwd	gctaaaaattatacattcatttatgttg
	Rev	gattacctggaaactttgaatagtttttgaac

Table 2.14
Two-step fusion PCR programme

Step	Temperature (°C)	Time (s)	# of cycles
Initial denaturation	92	120	1
Denaturation	92	10	25
Annealing	58	10	
Extension	68	60/kb	
Final Extension	68	7	1
Soak	10	∞	1

Table 2.15
Two-step fusion PCR mix

Note that, in order to amplify the *gfp* coding DNA, primers and template indicated in the table must be changed by primers C and D, and vector pPD95.75.

PCR 1		
Component	Final Volume (µL)	Final Concentration
Expand Long Template Buffer 2	2.5	1X (2.75 mM $MgCl_2$)
PCR Nucleotide Mix	0.35	0.2 mM
Fwd Primer (A)	0.3	0.5 µM
Rev Primer (B)	0.3	0.5 µM
Expand Long Template DNA Polymerase (5U/ µL)	0.2	1 U/25 µL
Template DNA (genomic)	1	0.2 µg/25 µL
Nuclease-Free Water	20.35	(–)

PCR 2		
Component	Final Volume (µL)	Final Concentration
Expand Long Template Buffer 2	2.5	1X (2,75 mM $MgCl_2$)
PCR Nucleotide Mix	0.35	0.2 mM
Fwd Primer (A*)	0.3	0.5 µM
Rev Primer (D)	0.3	0.5 µM
Expand Long Template DNA Polymerase (5U/ µL)	0.2	1 U/25 µL
Template DNA (PCR1 + *gfp::3' utr unc-54*)	2	0.2 µg/25 µL
Nuclease-Free Water	19.35	(–)

wild type N2 strain next to 100 ng/µL *rol-6 (su1006)* (final concentration: 150 ng/µL). A minimum of 30 worms was scored per line. Three independent lines were scored.

Chapter IV

Mouse strains

Animals of C57Bl6/JRccHsd (ENVIGO, Harlan) genetic background were housed in our animal care facility with a 12 h dark/light cycle and had free access to food and water. Timed embryos were obtained from overnight mating and the morning of the vaginal plug was considered as embryonic day (E) 0.5. All experiments were performed according to the animal care guidelines of the European Community Council (86 609 EEC) and to Spanish regulations (RD1201 2005), following protocols approved by the ethics committees of the Consejo Superior Investigaciones Científicas (CSIC).

Mouse immunohistochemistry

Freshly isolated E11.5 embryos were fixed by immersion in 4% PFA for 3 h (for SALL2 detection) or overnight (for BRN2 detection), washed with PBS, cryoprotected overnight in 30% sucrose in PBS and sectioned coronally at 10 mm using a Leica CM1900 cryostat. Before SALL2 and BRN2 immunohistochemical staining, the antigen was unmasked by boiling samples in 10 mm sodium citrate, pH 6, for 5 min and allowing them to cool down slowly, or by incubation in HCl 2N at 37 °C for 20 min followed by washes in borate buffer, respectively. The sections were then incubated for 1 h at room temperature in blocking buffer (PBS containing 1% bovine serum albumin (BSA) and 0.2% Triton X-100), and incubated overnight at 4 °C with following primary antibodies: rabbit anti-SALL2 (1:100; Sigma, #HPA004162), goat anti-BRN2 (1:100; Santa Cruz Biotechnology, #SC6029), rabbit anti-5-HT (1:5000; Sigma, #S-5545), goat anti-5-HT (1:200; Abcam, #Ab66047). The next day, sections were washed several times with PBS and incubated at room temperature for 1 h with secondary antibodies: Alexa 555-conjugated donkey anti-rabbit (1: 600; Molecular Probes, #A31572), Alexa 555-conjugated donkey anti-goat (1: 600; Molecular Probes, #A21432), Alexa 488-conjugated donkey anti-goat (1: 600; Molecular Probes, #A11055) and Alexa 488-conjugated donkey anti-rabbit (1: 600; Molecular Probes, #A21206) diluted in blocking buffer. Cells were counterstained with DAPI, washed in PBS, and mounted with Fluorsave (Calbiochem). No labelling was observed when sections were incubated only with secondary antibody (negative control). Immunofluorescent samples were analysed and photographed using a confocal TCS-SP8 Leica microscope.

RNAi screen to test functionality of additional FKH and LIM-HD candidates to regulate HSN identity

RNAi plates were prepared as described earlier for experiments in Chapter II. We used RNAi clones obtained from J. Ahringer (Kamath & Ahringer 2003) and M. Vidal RNAi libraries (Rual et al. 2004) (generated at Dana-Farber Cancer Institute; distributed by Source BioScience LifeSciences) → **Table 2.5**. *rrf-3* adult worms were transferred to IPTG plates and deposited within a drop of alkaline hypochlorite solution (*Drop Bleach*). Plates were incubated and scored at 20 °C and we performed F1 scoring at the adult stage. As a negative control empty vector L4440 (Addgene, #1654) was used. The experiment was done once. A minimum of 30 worms was scored. The statistic applied was Fisher exact test, *: pV< 0.05.

Validating functionality of new HSN transcription factor binding site candidates

Putative BSs for the new candidate TFs *pha-4*, *ceh-14* and *lin-11* were searched in *tph-1*, *cat-1* and *bas-1 cis*-regulatory modules. Core binding motif used for PHA-4 is A/GC/TNAAC/TA (Wederell et al. 2008) and for CEH-14 and LIN-11 are TA/TA/TA/TA/TA (German et al. 1992). PHA-4 putative BSs were identified in *tph-1prom2* and *cat-1prom14* CRMs. Directed mutagenesis was carried out to disturb their BSs (RYNggYA) and constructs were injected at 50 ng/µL into wild type N2 strain next to 100 ng/µL *rol-6 (su1006)* (final concentration: 150 ng/µL). Everything was performed as previously described in the mutagenesis experiments in Chapter II.

Rescuing worm serotonergic defects using mouse orthologues

DNA corresponding to the entire coding sequence of *ast-1*, *unc-86* (entire genomic locus DNA), *sem-4*, *hlh-3*, *egl-46* and *egl-18* and their mouse orthologues *Pet1, Brn2, Sall2, Ascl1, Insm1* and *Gata3* were amplified by PCR. These DNAs were cloned in front of HSN-specific promoters: *bas-1prom1*, *cat-4prom2* or *kal-1promA*, in the pPD95.75 vector. All the information for the cloning is collected in → **Table 2.16**.

DNA was injected into N2 animals at different concentrations depending on the toxicity of the array, being the total concentration of injected DNA 150 ng/µL → **Table 2.17**. Transgenic lines were then crossed with their respective mutant strain (*ast-1(ot417)*, *unc-86(n846)*, *sem-4(n1971)*, *hlh-3(tm1688)*, *egl-46(sy628)* or *egl-18(ok290)*) carrying the *zdIs13(tph-1::gfp)* transgene or the *yzIs71(tph-1::gfp*, *rol-6(su1006)* in the case of *egl-18*. Mutant scoring was performed using young adult worms maintained at 25 °C. As negative control mutant worms that did not harbour the extrachromosomal array were scored in parallel. At least 50 animals (100 HSN cells) were scored for each genotype and percentages of expression were calculated in the same way as described in Chapter II.

Principal Coordinate Analysis to compare expression profiles of worm neurons with expression profiles of mouse raphe serotonergic neurons (PCoA)

For Principal Coordinate Analysis and hierarchical clustering we used curated data from WormBase (Hobert et al. 2016) to generate a matrix with gene expression profiles for the 118 *C. elegans* hermaphrodite anatomical neuronal classes. Panneuronal genes and neurons in which less than 30 genes had been reported to be expressed were excluded. We built a similar matrix with mouse gene expression data from RNA-seq experiments, either from adult raphe nuclei divided into different rhombomeres (R1Dorsal, R1 Medial, R2, R3, R5, R6) (Okaty et al. 2015) or from cortical neurons, which were used as a control (Molyneaux et al. 2015). To transform the quantitative RNA-seq data into a presence-absence binary matrix, we considered values above 19 counts per million (CPM) as present and values below that threshold as absent. With this threshold, it appeared that one third of the genome is expressed in serotonergic cells, which is what it is estimated to be expressed in a cell.

To assign mouse orthologues to *C. elegans* genes we combined orthology relationships between mouse and worm genes annotated in the ENSEMBL database and worm-human orthology relationships reported in (Shaye & Greenwald 2011). In the last case, we used ENSEMBL data base again to assign mouse orthologues to human genes. In (Shaye & Greenwald 2011), ENSEMBL, OrthoMCL, InParanoid and Homologene methods are combined to identify orthologues. Thus, we combined

Plasmid	Construct	Primers used to amplify DNA	Template	Enzymes
pNF104	*bas-1prom1::ast-1 cDNA*	Fwd gagagaggtaccggtagaaaaaatgatgcaa-gtcgtctcgtcagcc Rev gagagagaattcctatcgataaagagggaatg	pNF17 (*ast-1* cDNA-pCDNA3.1)	KpnI / EcoRI
pNF391	*kal-1promA:: unc-86 genomic*	Fwd gagagaggtaccatgcaacctttcaac Rev gagagagaattcctaatcaaagaatccagg	pNF197 (Dr. Hobert)	EcoRI/KpnI
pNF395	*kal-1promA:: sem-4 cDNA*	Fwd tatatacccgggatgaatgagctgctcgccg Rev gagagagtatacctaagagggtggtggggt	pNF204	XmaI/Bstz17I
pNF370	*cat-4prom2:: hlh-3 cDNA*	Fwd gagagaggtaccatgaccgcatccacctc Rev gagagagaattcttaataagtttctgtatgcg	pNF283	KpnI / EcoRI
pNF371	*cat-4prom2:: egl-46 cDNA*	Fwd gagagaggtaccatggtgcctatgaatg Rev gagagagaattcttacattgttggaataac	pNF284	KpnI / EcoRI
pNF372	*cat-4prom2:: egl-18 cDNA*	Fwd gagagagctagcatgtcgatcagcataatg Rev gagagaggtaccgttagcagccggatctc	pNF314	KpnI / EcoRI
pNF185	*bas-1prom1:: Pet1 cDNA*	Fwd gagagaggtaccatgagacagagcggcacctc Rev gagagagaattcctagtgataat-gacccccaag	(Source BioScience, #8861455)	KpnI / EcoRI
pNF397	*kal-1promA:: Brn2 cDNA*	Fwd tatataggatccatggcgaccgcagcgtc Rev tatatagtatactcactggacgggcgtctgca	(Addgene, #27151)	BamHI/Bstz17I
pNF384	*kal-1promA:: Sall2 cDNA*	Fwd gagagacccgggatgtctcggcgaaag Rev gagagagtatactcatggcatggtgg	(OpenBiosystems, #5706710)	XmaI/Bstz17I
pNF380	*cat-4prom2:: Ascl1 cDNA*	Fwd gagagaggtaccatggagagctctggcaag Rev gagagagaattctcagaaccagttggtaaa-gtcc	pNF287 (p2Lox-ALP-V5-His) generated in the laboratory	KpnI / EcoRI
pNF385	*cat-4prom2:: Insm1 cDNA*	Fwd gagagaggtaccatgccacggggatttc Rev gagagagaattctcaacaagcgggc	Genescript	KpnI / EcoRI
pNF383	*cat-4prom2:: Gata3 cDNA*	Fwd gagagaaccggtatggaggtgactgcg Rev gagagagaattcctaacccatggcggt	(Source BioScience, #6826352)	Age/ EcoRI
Promoter	Construct	Primer	Template	Enzymes
	bas1prom1	Fwd aaaggatccggaaatggcaacatcttagac Rev tttggatccccgaactactactgaaagttc	N2 genomic DNA	PstI/BamHI
	cat-4prom2	Fwd gagagaaagcttcaatcagcccagaaatcgc Rev tttggatccgatattatgatgttgatagag	N2 genomic DNA	PstI/BamHI
	kal-1promA	Fwd gagagactgcagatttcgtatttggagc Rev gagagaggatcccatgtgctgtaagag	N2 genomic DNA	PstI/BamHI

both sources to have a wider coverage of orthology relationships than using ENSEMBL or (Shaye & Greenwald 2011) data alone. Worm genes without any mouse orthologue and genes that were not expressed in any worm neuron were removed. Whenever a worm gene had more than one mouse orthologue, it was duplicated in the worm data set. Simple matching coefficient (Sokal & Michener 1958) was calculated and Principal Coordinate Analysis was performed using the dudi.pco function from the ade4 R package (Dray et al. 2007). Finally, to assess which worm cell was closest to the mouse raphe nuclei, we calculated the euclidean distance between each of the worm cells and each of the raphe nuclei in the space defined by the three principal components. As a control, 100 random sets of 95 genes (the same number of genes that are expressed in the HSN) were generated from the worm gene pool, generated with the sample function of R (The R Team 2016). This data set

← **Table 2.16**
Rescuing constructs

Table 2.17
Injection mix for rescuing experiments using worm and mouse orthologue factors

Gene	TF cDNA (ng/µL)	*ttx-3::mCherry* (ng/µL)	*rol-6(su1006)* (ng/µL)	pUC19 (ng/µL)	Total concentration (ng/µL)
unc-86	50	50	50	0	150
sem-4	20	50	50	30	150
hlh-3	50	50	50	0	150
egl-46	50	50	50	0	150
ast-1	10	50	50	40	150
egl-18	50	50	50	0	150
Brn2	50	50	50	0	150
Sall2	20	50	50	30	150
Ascl1	50	50	50	0	150
Insm1	50	50	50	0	150
Pet1	10	50	50	40	150
Gata3	50	50	50	0	150

Table 2.18
Commercial kits used in this Thesis

Kit name	Source / Reference
QIAprep Spin Miniprep Kit	QIAGEN, # 27106
Genelute Mammalian Total RNA kit	Sigma, #RTN70-1KT
Quantitec Reverse Transcription Kit	QIAGEN, #205311
DNeasy Blood & Tissue Kit	QIAGEN, #69504
QIAquick PCR Purification Kit	QIAGEN, #28106
Quickchange II XL site-directed mutagenesis kit	Agilent Technologies, #200522

was merged with mouse raphe nuclei expression profile and Principal Coordinate Analysis was performed as before. For hierarchical clustering, the same binary matrix containing mouse and worm expression data was fed to the pvclust function in the pvclust R package (Suzuki & Shimodaira 2006), which uses a bootstrapping technique to calculate p-values for each cluster, the AU and BP values (Shimodaira 2002). Parameters were set as follows: method.hclust = 'average', method.dist= 'binary', nboot = 10000, r = seq(0.5, 1.4, by=0.1). The standard error of the PV and AU values was approximately 0.1% for most clusters, including the HSN-raphe cluster.

Materials

This section includes detailed information on relevant reagents, solutions, apparatus and software used in this Thesis.

Nematode Growth Media

The NGM agar contains NaCl (3 gL-1), agar (17 gL-1), peptone (2.5 gL-1), $CaCl_2$ (1M, 1 mL L-1), $MgSO_4$ (1M, 1 mL L-1), KH_2PO_4 buffer (1M, pH=6.0, 25 mL L-1), cholesterol (5 mg mL-1 in ethanol 95%, 1mL L-1), Milli-Q water (Merck Millipore) water and nystatin (Sigma), to prevent fungal and bacterial contamination.

M9 Buffer

The M9 buffer is used at 1X to collect worms from agar plates and to grow worms without food. M9 10X is prepared with the following components: $Na_2HPO_4 \times 12H_2O$ (146g L-1), KH_2PO_4 (30G L-1), NaCl (5g L-1) and NH_4Cl (10g L-1).

Worm Lysis Solution

This solution, after the addition of Proteinase K, is used to disaggregate the worms and obtain genomic DNA. It is stored at 4 °C and its components are: KCl (50 mM), Tris-HCl (10 mM, pH 8.3), $MgCl_2$ (2.5 mM), Triton X-100 (0.45% (v/v), Sigma), Tween 20 (0.45% (v/v), Sigma).

Gibson Assembly Reagents Gibson Assembly Master Mix (2X)

The Isothermal Start Mix is prepared with 1.5 g Poliethinelglycol 8000 (Promega, #V3011), 3 mL 1 M Tris-HCl, pH 8.0 (Sigma, # T3253) and 150 µl 2 M $MgCl_2$ (Sigma, #M8266) in a volume of 3150 µL.

The Gibson Assembly Master Mix (2X) consists of 405 µl Isothermal Start Mix (RT) (described above), 25 µl 1 M DTT (4°C) (Sigma, #GE17-1318-01), 50 µl 10 mM dNTPs (−20°C) (Labclinics, #GC-013-001), 50 µl NAD (−80°C) (NEB, # B9007S), 1 µl T5 exonuclease (−20°C) (10 U/µl) (NEB, #M0363S), 31.25 µl Phusion High Fidelity DNA Polymerse, (2 U/µl) (−20 °C) (NEB #M0530S), 250 µl Taq Ligase (40 U/µl) (NEB, #M0208L) and 437.75 µl H_2O, to a final volume of 1250 µl. This mix is aliquoted and stored at −20 °C.

Egg Preparation Solution (Egg Prep)

The *Egg Prep* solution is used to synchronise large worm populations. It consists of 1.2 mL NaClO (commercial bleach), 2.5 mL NaOH (1 M) and 6.3 mL ddH_20.

Drop bleach

The *Drop Bleach* solution is used to kill sensible bacteria or fungi that frequently contaminate worm plates. It is prepared with 500 µL NaClO (commercial bleach), 200 µL NaOH (5 M) and 300 µL ddH_20.

Table 2.19
Plasmids used in this Thesis

Plasmid name	Description	Source / Reference
pPD95.75	*gfp* vector used for promoter standard cloning and fusion PCR (Dr. Fire laboratory).	Addgene #1494
pRF4	*rol-6(su1006)* vector, used as co-marker in worm microinjection (Dr. Fire laboratory).	(Mello et al., 1991)
pET-21b	Expression vector used for cloning for EMSA experiments.	EMD Millipore, kindly provided by Dr. Hobert laboratory
pNF17	*ast-1* cDNA-pCDNA3.1, used in EMSA experiments.	This work
pET-21b-ast-1	Vector containing an *ast-1* probe for EMSA experiments.	This work
pET-21b-unc-86	Vector containing an *unc-86* probe for EMSA experiments.	(Zhang et al. 2014)
pCDNA3-egl-18	Vector containing an *egl-18* probe for EMSA experiments.	This work
pPD129.36	Also known as L4440, this vector is used as negative control in RNAi experiments.	Addgene, #1654
pNF303	*egl-46prom::NLS::DsRed* in pPD96.04. Transcriptional reporter for *egl-46*, used in cross-regulation analysis.	This work
pJJR82	GFP-Self-Excising-Cassette vector, used for CRISPR genome engineering strategies.	Addgene #75027, kindly provided by Dr. Boxem
pDD162	sgRNA/Cas9 containing vector, used for CRISPR genome engineering strategies.	Addgene #4754, kindly provided by Dr. Boxem
pCFJ90	*Pmyo-2::mCherry;* pharyngeal co-injection marker used in CRISPR/Cas9 GFP knock-in.	Addgene #19328, kindly provided by Dr. Boxem
pPD49.78	*hsp-16.2* (heat shock promoter) vector used in transcription factor overexpression experiments.	Addgene #1447
pPD49.83	*hsp-16.2* (heat shock promoter) vector used in transcription factor overexpression experiments.	Addgene #1448
pNF101	*ttx-3prom::mCherry,* used as co-marker in overexpression and rescuing experiments.	(Bertrand and Hobert, 2009)
pNF197	*hsp-16.2::unc-86 genomic*, used to ectopically express unc-86 in the worm.	This work
pNF204	*hsp-16.2::sem-4* cDNA, used to ectopically express *sem-4* in the worm.	This work
pNF283	*hsp-16.2::hlh-3 cDNA*, used to ectopically express *hlh-3* in the worm.	This work

↳

pNF284	*hsp-16.2::egl-46 cDNA*, used to ectopically express *egl-46* in the worm.	This work
pNF314	*hsp-16.2::egl-18 cDNA*, used to ectopically express *egl-18* in the worm.	This work
Pet1	Vector containing the cDNA of the mouse gene *Pet1*, used for generating new plasmids for rescuing experiments.	Source BioScience, #8861455
Brn2	Vector containing the cDNA of the mouse gene *Brn2*, used for generating new plasmids for rescuing experiments.	Addgene, #27151
Sall2	Vector containing the cDNA of the mouse gene *Sall2*, used for generating new plasmids for rescuing experiments.	OpenBiosystems, #5706710
Gata3	Vector containing the cDNA of the mouse gene *Gata3*, used for generating new plasmids for rescuing experiments.	Source BioScience, #6826352
pNF104	*bas-1prom1::ast-1* cDNA, used to rescue *tph-1* defects in HSN.	This work
pNF391	*kal-1promA:: unc-86 genomic*, used to rescue *tph-1* defects in HSN.	This work
pNF395	*kal-1promA:: sem-4 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF370	*cat-4prom2:: hlh-3 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF371	*cat-4prom2:: egl-46 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF372	*cat-4prom2:: egl-18 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF185	*bas-1prom1:: Pet1 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF397	*kal-1promA:: Brn2 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF384	*kal-1promA:: Sall2 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF380	*cat-4prom2:: Ascl1 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF385	*cat-4prom2:: Insm1 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pNF383	*cat-4prom2:: Gata3 cDNA*, used to rescue *tph-1* defects in HSN.	This work
pUC19	Empty vector used as filler DNA in worm microinjection mixes.	Addgene, #50005

Table 2.20
Living organisms used in this Thesis

Strain name	Description	Source / Reference
OP50	*Escherichia coli* strain used to feed worms.	CGC
HT115	*Escherichia coli* strain used to feed worms exclusively in RNAi experiments.	CGC
TOP10	*Escherichia coli* electrocompetent cells, used to transform and amplify plasmids.	Invitrogen, # C404010
Rosetta2(DE3)	*Escherichia coli* Rosetta2(DE3), used to transform pET-21b His tag expression vector.	Novagen, #71400
HEK293T human cells	Transfected with *egl-18*-pcDNA.3 vector in EMSA experiments.	ATCC: CRL-3216, kindly provided by Dr. Hobert
C57Bl/6JRccHsd mouse strain	Used for the expression analysis of the new mouse serotonergic candidates.	ATCC: CRL-3216, kindly provided by Dr. Hobert

Table 2.21
Worm strains used in this Thesis

Strains are listed in order of appearance in the Results section.

Chapter I — *Cis*-regulatory analysis of the 5-HT pathway genes (Figure 3.1.3)		
Strain name	Genotype	Source
N2	*Caenorhabditis elegans* wild type strain	CGC
NFB343	*vlcEx135[tph-1prom1::gfp, rol-6(su1006)]*	This work
NFB345	*vlcEx137[tph-1prom1::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom1* Line 3	This work
NFB133	*vlcEx32[tph-1prom8::gfp, rol-6(su1006)]*	This work
NFB134	*vlcEx33[tph-1prom8::gfp, rol-6(su1006)]*	This work
NFB135	*vlcEx34[tph-1prom8::gfp, rol-6(su1006)]*	This work
NFB70	*vlcEx1[tph-1prom2::gfp, rol-6(su1006)]*	This work
NFB71	*vlcEx2[tph-1prom2::gfp, rol-6(su1006)]*	This work
NFB72	*vlcEx3[tph-1prom2::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom2* Line 4	This work

No name	*tph-1prom2* Line 5	This work
No name	*tph-1prom2* Line 6	This work
No name	*tph-1prom6* Line 1	This work
No name	*tph-1prom6* Line 2	This work
NFB137	*vlcEx36[tph-1prom5::gfp, rol-6(su1006)]*	This work
NFB138	*vlcEx37[tph-1prom5::gfp, rol-6(su1006)]*	This work
NFB139	*vlcEx38[tph-1prom5::gfp, rol-6(su1006)]*	This work
NFB73	*vlcEx4[tph-1prom3::gfp, rol-6(su1006)]*	This work
NFB74	*vlcEx5[tph-1prom3::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom3* Line 3	This work
No name	*tph-1prom3* Line 4	This work
NFB75	*vlcEx6[tph-1prom17::gfp, rol-6(su1006)]*	This work
NFB76	*vlcEx7[tph-1prom17::gfp, rol-6(su1006)]*	This work
NFB77	*vlcEx8[tph-1prom17::gfp, rol-6(su1006)]*	This work
OH4194	*otEx2433[bas-1prom1::gfp, rol6(su1006)]*	This work
OH4196	*otEx2435[bas-1prom1::gfp, rol6(su1006)]*	This work
OH4198	*otEx2437[bas-1prom2::gfp, rol6(su1006)]*	This work
OH4200	*otEx2439[bas-1prom2::gfp, rol6(su1006)]*	This work
OH8681	*bas-1prom13* Line 1	This work
OH8682	*bas-1prom13* Line 2	This work
OH8684	*bas-1prom14* Line 1	This work
NFB149	*vlcEx48[bas-1prom15::gfp, rol6(su1006)]*	This work
NFB150	*vlcEx49[bas-1prom15::gfp, rol6(su1006)]*	This work
NFB151	*vlcEx50[bas-1prom15::gfp, rol6(su1006)]*	This work
NFB136	*vlcEx35[bas-1prom16::gfp, rol6(su1006)]*	This work
No name	*bas-1prom17* Line 1	This work
No name	*bas-1prom17* Line 2	This work
NFB117	*vlcEx16[bas-1prom18::gfp, rol6(su1006)]*	This work

NFB118	*vlcEx17[bas-1prom18::gfp, rol6(su1006)]*	This work
NFB307	*vlcEx161[bas-1prom18::gfp, rol6(su1006)]*	This work
NFB308	*vlcEx162[bas-1prom18::gfp, rol6(su1006)]*	This work
OH4261	*otEx2476[bas-1prom3::gfp, rol6(su1006)]*	This work
OH4263	*otEx2478[bas-1prom3::gfp, rol6(su1006)]*	This work
OH4812	*otEx2806[bas-1prom4::gfp, rol6(su1006)]*	This work
OH4814	*otEx2808[bas-1prom4::gfp, rol6(su1006)]*	This work
NFB206	*vlcEx84[bas-1prom5::gfp, rol6(su1006)]*	This work
NFB207	*vlcEx85[bas-1prom5::gfp, rol6(su1006)]*	This work
NFB208	*vlcEx86[bas-1prom5::gfp, rol6(su1006)]*	This work
No name	*bas-1prom6* Line 1	This work
No name	*bas-1prom6* Line 2	This work
NFB116	*vlcEx15[bas-1prom7::gfp, rol6(su1006)]*	This work
NFB120	*vlcEx19[bas-1prom7::gfp, rol6(su1006)]*	This work
NFB121	*vlcEx20[bas-1prom7::gfp, rol6(su1006)]*	This work
OH4209	*otEx2448[cat-1prom1::gfp, rol-6(su1006)]*	This work
OH4208	*otEx2447[cat-1prom1::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom1* Line 3	This work
OH4217	*otEx2455[cat-1prom2::gfp, rol-6(su1006)]*	This work
OH4218	*otEx2456[cat-1prom2::gfp, rol-6(su1006)]*	This work
OH4219	*otEx2457[cat-1prom3::gfp, rol-6(su1006)]*	This work
OH4228	*otEx2460[cat-1prom3::gfp, rol-6(su1006)]*	This work
OH7435	*otEx3249[cat-1prom12::gfp, rol-6(su1006)]*	This work
OH7436	*otEx3250[cat-1prom12::gfp, rol-6(su1006)]*	This work
NFB236	*vlcEx111[cat-1prom11::gfp, rol6(su1006)]*	This work
NFB237	*vlcEx112[cat-1prom11::gfp, rol6(su1006)]*	This work
NFB241	*vlcEx116[cat-1prom35::gfp, rol6(su1006)]*	This work
NFB242	*vlcEx117[cat-1prom35::gfp, rol6(su1006)]*	This work

NFB243	*vlcEx118[cat-1prom35::gfp, rol6(su1006)]*	This work
No name	*cat-1prom36* Line 1	This work
No name	*cat-1prom36* Line 2	This work
NFB244	*vlcEx119[cat-1prom37::gfp, rol6(su1006)]*	This work
NFB245	*vlcEx120[cat-1prom37::gfp, rol6(su1006)]*	This work
NFB246	*vlcEx121[cat-1prom37::gfp, rol6(su1006)]*	This work
OH6000	*otEx2991[cat-1prom13::gfp, rol-6(su1006)]*	This work
OH6001	*otEx2992[cat-1prom13::gfp, rol-6(su1006)]*	This work
OH6002	*otEx2993[cat-1prom13::gfp, rol-6(su1006)]*	This work
OH7443	*otEx3257[cat-1prom14::gfp, rol-6(su1006)]*	This work
OH7506	*otEx3303[cat-1prom14::gfp, rol-6(su1006)]*	This work
OH7508	*otEx3305[cat-1prom14::gfp, rol-6(su1006)]*	This work
NFB291	*vlcEx149[cat-1prom14::gfp, rol6(su1006)]*	This work
NFB292	*vlcEx150[cat-1prom14::gfp, rol6(su1006)]*	This work
NFB293	*vlcEx151[cat-1prom14::gfp, rol6(su1006)]*	This work
NFB180	*vlcEx66[cat-1prom26::gfp, rol6(su1006)]*	This work
NFB201	*vlcEx79[cat-1prom26::gfp, rol6(su1006)]*	This work
NFB202	*vlcEx80[cat-1prom26::gfp, rol6(su1006)]*	This work
NFB238	*vlcEx113[cat-1prom27::gfp, rol6(su1006)]*	This work
NFB239	*vlcEx114[cat-1prom27::gfp, rol6(su1006)]*	This work
NFB240	*vlcEx115[cat-1prom27::gfp, rol6(su1006)]*	This work
OH4753	*otEx2760[cat-4prom4::gfp, rol-6(su1006)]*	This work
OH4754	*otEx2761[cat-4prom4::gfp, rol-6(su1006)]*	This work
OH4755	*otEx2762[cat-4prom4::gfp, rol-6(su1006)]*	This work
OH4852	*otEx2829[cat-4prom5::gfp, rol-6(su1006)]*	This work
OH4853	*otEx2830[cat-4prom5::gfp, rol-6(su1006)]*	This work
OH7500	*otEx3297[cat-4prom6::gfp, rol-6(su1006)]*	This work
OH7501	*otEx3298[cat-4prom6::gfp, rol-6(su1006)]*	This work
OH7502	*otEx3299[cat-4prom6::gfp, rol-6(su1006)]*	This work
NFB280	*vlcEx143[cat-4prom8::gfp, rol-6(su1006)]*	This work
NFB281	*vlcEx144[cat-4prom8::gfp, rol-6(su1006)]*	This work
NFB332	*vlcEx177[cat-4prom58::gfp, rol-6(su1006)]*	This work
NFB333	*vlcEx178[cat-4prom58::gfp, rol-6(su1006)]*	This work
No name	*cat-4prom58* Line 3	This work
NFB340	*vlcEx182[cat-4prom59::gfp, rol-6(su1006)]*	This work
NFB341	*vlcEx183[cat-4prom59::gfp, rol-6(su1006)]*	This work
No name	*cat-4prom59* Line 3	This work
OH6005	*otEx2996[cat-4prom9::gfp, rol-6(su1006)]*	This work
OH6006	*otEx2997[cat-4prom9::gfp, rol-6(su1006)]*	This work
OH6007	*otEx2998[cat-4prom9::gfp, rol-6(su1006)]*	This work
NFB210	*vlcEx88[cat-4prom18::gfp, rol-6(su1006)]*	This work
NFB211	*vlcEx89[cat-4prom18::gfp, rol-6(su1006)]*	This work
NFB212	*vlcEx90[cat-4prom18::gfp, rol-6(su1006)]*	This work
NFB262	*lin-15B(n765)X; vxIs97[tph-1p::DsRed + lin15(+)]; vlcEx97[cat-4prom19::gfp, rol-6(su1006)]*	This work
NFB263	*lin-15B(n765)X; vxIs97[tph-1p::DsRed + lin15(+)]; vlcEx98[cat-4prom19::gfp, rol-6(su1006)]*	This work
NFB181	*vlcEx67[cat-4prom27::gfp, rol-6(su1006)]*	This work
NFB196	*vlcEx68[cat-4prom27::gfp, rol-6(su1006)]*	This work
NFB197	*vlcEx69[cat-4prom27::gfp, rol-6(su1006)]*	This work
NFB636	*vlcEx344[mod-5prom1::gfp, rol-6(su1006)]*	This work
NFB637	*vlcEx345[mod-5prom1::gfp, rol-6(su1006)]*	This work
No name	*mod-5prom1* Line 3	This work
NFB593	*vlcEx321[mod-5prom3::gfp, rol-6(su1006)]*	This work
NFB594	*vlcEx322[mod-5prom3::gfp, rol-6(su1006)]*	This work
NFB638	*vlcEx346[mod-5prom8::gfp, rol-6(su1006)]*	This work

NFB639	*vlcEx347[mod-5prom8::gfp, rol-6(su1006)]*	This work
No name	*mod-5prom8* Line 3	This work
NFB574	*vlcEx315[mod-5prom6::gfp, rol-6(su1006)]*	This work
NFB575	*vlcEx316[mod-5prom6::gfp, rol-6(su1006)]*	This work
No name	*mod-5prom6* Line 3	This work

Chapter II — Loss of function mutant analysis (Figures 3.2.2-3.2.6, 3.4.3)

Strain name	Genotype	Source
JR2370	*egl-18(ok290)IV*	CGC, (Koh et al. 2002)
VC271	*end-1&ric-7(ok558)V*	CGC, (Maduro et al. 2005)
TB528	*ceh-14(ch3)X*	CGC, (Cassata et al. 2000)
MT633	*lin-11(n389)I; him-5(e1467)V*	CGC, (Trent et al. 1983)
MK4013	*zdIs13(tph-1::gfp)IV*	(Clark & Chiu 2003)
GR1333	*yzIs71[tph-1::gfp, rol-6(su1006)]V*	(Sze et al. 2000)
OH8777	*ast-1(ot417)II; zdIs13(tph-1::gfp)IV*	Dr. Hobert Lab.
OH9423	*unc-86(n846)III; zdIs13(tph-1::gfp)IV*	Dr. Hobert Lab.
OH11963	*sem-4(n1971)I; zdIs13(tph-1::gfp)IV*	Dr. Hobert Lab.
NFB471	*hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV*	This work
NFB477	*egl-46(sy628)V; zdIs13(tph-1::gfp)IV*	This work
NFB683	*egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V*	This work
NFB730	*end-1&ric-7(ok558)V; zdIs13(tph-1::gfp)IV*	This work
OH8246	*otIs221(cat-1::gfp)III*	(Flames & Hobert 2009)
OH8249	*otIs224(cat-1::gfp)V*	(Flames & Hobert 2009)
OH8772	*ast-1(ot417)II; otIs221(cat-1::gfp)III*	(Flames & Hobert 2009)
OH10603	*unc-86(n846)III; otIs224(cat-1::gfp)V*	Dr. Hobert Lab.
NFB60	*ast-1(hd92)II; vlcEx845[ast-1(+), cat-1::DsRed]; otIs221(cat-1::gfp)III*	This work

OH11909	*sem-4(n1971)I; otIs221(cat-1::gfp)III*	Dr. Hobert Lab.
NFB489	*hlh-3(tm1688)II; otIs221(cat-1::gfp)III*	This work
NFB478	*egl-46(sy628)V; otIs221(cat-1::gfp)III*	This work
NFB687	*egl-18(ok290)IV; otIs221(cat-1::gfp)III*	This work
NFB733	*end-1&ric-7(ok558)V; otIs221[cat-1::gfp]III*	This work
NFB926	*lin-11 (n389)I; ceh-14 (ch3) X; otIs221(cat-1::gfp)III; him-5 (e1467)V*	This work
OH4255	*otEx2470[cat-4::gfp(50ng/ul), rol-6(su1006)]*	(Flames & Hobert 2009)
OH8250	*otIs225(cat-4::gfp)II*	(Flames & Hobert 2009)
NFB83	*ast-1(ot417)II; otEx2470[cat-4::gfp (50ng/ul), rol-6(su1006)]*	This work
OH10918	*unc-86(n846)III; otIs225(cat-4::gfp)II*	Dr. Hobert Lab.
OH11962	*sem-4(n1971)I; otIs225(cat-4::gfp)II*	Dr. Hobert Lab.
NFB472	*hlh-3(tm1688)II; otEx2470[cat-4::gfp (50ng/ul), rol-6(su1006)]*	This work
NFB479	*egl-46(sy628)V; otIs225(cat-4::gfp)II*	This work
NFB685	*egl-18(ok290)IV; otIs225(cat-4::gfp)II*	This work
NFB732	*end-1&ric-7(ok558)V; otIs225[cat-4::gfp]II*	This work
OH8251	*otIs226(bas-1::gfp)IV*	(Flames & Hobert 2009)
OH4196	*otEx2435[bas-1::gfp(50ng/ul), rol-6(su1006)]*	(Flames & Hobert 2009)
OH10562	*ast-1(ot417)II; otIs226(bas-1::gfp)IV*	(Flames & Hobert 2009)
NFB159	*ast-1(hd92); vlcEx844[ast-1(+), cat-1::DsRed]; otIs226(bas-1::gfp)IV*	This work
NFB160	*ast-1(hd92); vlcEx845[ast-1(+), cat-1::DsRed]; otIs226(bas-1::gfp)IV*	This work
OH10607	*unc-86(n846)III; otIs226(bas-1::gfp)IV*	Dr. Hobert Lab.
OH11910	*sem-4(n1971)I; otIs226(bas-1::gfp)IV*	Dr. Hobert Lab.
NFB455	*hlh-3(tm1688)II; otIs226(bas-1::gfp)IV*	This work
NFB536	*egl-46(sy628)V; otIs226(bas-1::gfp)IV*	This work
NFB715	*egl-18(ok290)IV; otEx2435[bas-1::gfp (50ng/ul), rol-6(su1006]*	This work
NFB731	*end-1&ric-7(ok558)V; otIs226(bas-1::gfp)IV*	This work
LX1376	*vsEx580[kcc-2c::gfp, myo-2::gfp]*	CGC, (Tanis et al. 2009)
NFB47	*ast-1(ot417)II; vsEx580[kcc-2c::gfp, myo-2::gfp]*	This work

NFB153	*unc-86(n846)III; vsEx580[kcc-2c::gfp, myo-2::gfp]*	This work
NFB152	*sem-4(n1971)I; vsEx580[kcc-2c::gfp, myo-2::gfp]*	This work
NFB510	*hlh-3(tm1688)II; vsEx580[kcc-2c::gfp, myo-2::gfp]*	This work
NFB511	*egl-46(sy628)V; vsEx580[kcc-2c::gfp, myo-2::gfp]*	This work
NFB1029	*egl-18(ok290)IV; vsEx580[kcc-2c::gfp, myo-2::gfp]*	This work
QW84	*zfIs4(lgc-55::mCherry)*	CGC, (Pirri et al. 2009)
QW122	*zfIs6(lgc-55::gfp)II*	CGC, (Pirri et al. 2009)
NFB187	*ast-1(ot417)II; zfIs4(lgc-55::mCherry)*	This work
NFB448	*unc-86(n846)III; zfIs6(lgc-55::gfp)II*	This work
NFB156	*sem-4(n1971)I; zfIs6(lgc-55::gfp)II*	This work
NFB473	*hlh-3(tm1688)II; zfIs4(lgc-55::mCherry)*	This work
NFB526	*egl-46(sy628)V; zfIs6(lgc-55::gfp)II*	This work
NFB996	*egl-18(ok290)IV; zfIs6(lgc-55::gfp)II*	This work
BL5752	*inIs181(ida-1::gfp); inIs182(ida-1::gfp)*	CGC, (Zahn et al. 2004)
BL5717	*inIs179(ida-1::gfp)II*	CGC, (Zahn et al. 2004)
NFB259	*ast-1(ot417)II; inIs181(ida-1::gfp); inIs182(ida-1::gfp)*	This work
NFB42	*unc-86(n846)III; inIs179(ida-1::gfp)II*	This work
NFB155	*sem-4(n1971)I; inIs179(ida-1::gfp)II*	This work
NFB539	*hlh-3(tm1688)II; inIs181(ida-1::gfp); inIs182(ida-1::gfp)*	This work
NFB538	*egl-46(sy628)V; inIs179(ida-1::gfp)II*	This work
NFB1405	*egl-18(ok290)IV; inIs179(ida-1::gfp)II*	This work
RJP255	*ynIs34(flp-19::gfp)IV*	CGC, (Kim & Li 2004)
NFB39	*ast-1(ot417)II; ynIs34(flp-19::gfp)IV; him-5(e1490)V*	This work
NFB38	*unc-86(n846)III; ynIs34(flp-19::gfp) IV; him-5(e1490)V*	This work
NFB157	*sem-4(n1971)I; ynIs34(flp-19::gfp) IV; him-5(e1490)V*	This work
NFB486	*hlh-3(tm1688)II; ynIs34(flp-19::gfp) IV*	This work
NFB518	*egl-46(sy628)V; ynIs34(flp-19::gfp) IV*	This work
OH13083	*otIs576(unc-17fosmid::GFP, lin-44::YFP)*	Dr. Hobert Lab.

NFB844	*ast-1(ot417)II; otIs576(unc-17fosmid::GFP, lin-44::YFP)*	This work
NFB855	*unc-86(n846)III; otIs576(unc-17fosmid::GFP, lin-44::YFP); him-5(e1490)V*	This work
NFB891	*sem-4(n1971)I; otIs576(unc-17fosmid::GFP, lin-44::YFP); him-5(e1490)V*	This work
NFB857	*hlh-3(tm1688)II; otIs576(unc-17fosmid::GFP, lin-44::YFP); him-5(e1490)V*	This work
NFB757	*egl-46(sy628)V; otIs576(unc-17fosmid::GFP, lin-44::YFP)*	This work
NFB858	*egl-18(ok290)IV; otIs576(unc-17fosmid::GFP, lin-44::YFP)*	This work
AL132	*icIs132(unc-40::gfp)*	CGC, (Chan et al. 1996)
NFB252	*ast-1(ot417)II; icIs132 (unc-40::gfp); him-8(e1489)IV*	This work
NFB178	*unc-86(n846)III; icIs132 (unc-40::gfp); him-8(e1489)IV*	This work
NFB179	*sem-4(n1971)I; icIs132 (unc-40::gfp); him-8(e1489)IV*	This work
NFB552	*hlh-3(tm1688)II; icIs132 (unc-40::gfp); him-8(e1489)IV*	This work
NFB453	*egl-46(sy628)V; icIs132 (unc-40::gfp); him-8(e1489)IV*	This work
NFB973	*egl-18(ok290)IV; icIs132 (unc-40::gfp); him-8(e1489)IV*	This work
OH9545	*otIs287[rab-3::yfp, rol-6(su1006)]IV*	(Stefanakis et al. 2015)
OH9609	*otIs291[rab-3::gfp, rol-6(su1006)]*	(Stefanakis et al. 2015)
NFB63	*ast-1(ot417)II; otIs287[rab-3::yfp, rol-6(su1006)]IV*	This work
OH9660	*unc-86(n846)III; otIs287[rab-3::yfp, rol-6(su1006)]IV*	Dr. Hobert Lab.
NFB154	*sem-4(n1971)I; otIs287[rab-3::yfp, rol-6(su1006)]IV*	This work
NFB474	*hlh-3(tm1688)II; otIs287[rab-3::yfp, rol-6(su1006)]IV*	This work
NFB537	*egl-46(sy628)V; otIs287[rab-3::yfp, rol-6(su1006)]IV*	This work
NFB1026	*egl-18(ok290)IV; otIs291[rab-3::gfp, rol-6(su1006)]*	This work
BC13535	*sIs13247(nlg-1::gfp)*	CGC, (McKay et al. 2003)
NFB251	*ast-1(ot417)II; sIs13247(nlg-1::gfp)*	This work
NFB158	*unc-86(n846)III; sIs13247(nlg-1::gfp)*	This work
NFB250	*sem-4(n1971)I; sIs13247(nlg-1::gfp)*	This work
NFB517	*hlh-3(tm1688)II; sIs13247(nlg-1::gfp)*	This work
NFB633	*egl-46(sy628)V; sIs13247(nlg-1::gfp)*	This work
OH904	*otIs33(kal-1::gfp)IV*	(Bülow et al. 2002)

NFB819	*vlcEx453[kal-1::gfp, ttx-3::mCherry, rol-6(su1006)]*	This work
NFB45	*ast-1(ot417)II; otIs33(kal-1::gfp)IV*	This work
NFB44	*unc-86(n846)III; otIs33(kal-1::gfp)IV*	This work
NFB186	*sem-4(n1971)I; otIs33(kal-1::gfp)IV*	This work
NFB475	*hlh-3(tm1688)II; otIs33(kal-1::gfp)IV*	This work
NFB480	*egl-46(sy628)V; otIs33 (kal-1::gfp)IV*	This work
NFB686	*egl-18(ok290)IV; vlcEx453[kal-1::gfp, ttx-3::mCherry, rol-6(su1006)]*	This work

Chapter II — Mutagenesis analysis (Figures 3.2.8-3.2.10, 3.4.3)

Strain name	Genotype	Source
NFB143	*vlcEx42[tph-1prom14::gfp, rol-6(su1006)]*	This work
NFB144	*vlcEx43[tph-1prom14::gfp, rol-6(su1006)]*	This work
NFB145	*vlcEx44[tph-1prom14::gfp, rol-6(su1006)]*	This work
NFB119	*vlcEx18[tph-1prom26::gfp, rol-6(su1006)]*	This work
NFB164	*vlcEx54[tph-1prom26::gfp, rol-6(su1006)]*	This work
NFB165	*vlcEx55[tph-1prom31::gfp, rol-6(su1006)]*	This work
NFB166	*vlcEx56[tph-1prom31::gfp, rol-6(su1006)]*	This work
NFB277	*vlcEx140[tph-1prom31::gfp, rol-6(su1006)]*	This work
NFB278	*vlcEx141[tph-1prom31::gfp, rol-6(su1006)]*	This work
NFB279	*vlcEx142[tph-1prom31::gfp, rol-6(su1006)]*	This work
NFB398	*vlcEx226[tph-1prom44::gfp, rol-6(su1006)]*	This work
NFB399	*vlcEx227[tph-1prom44::gfp, rol-6(su1006)]*	This work
NFB403	*vlcEx228[tph-1prom43::gfp, rol-6(su1006)]*	This work
NFB404	*vlcEx229[tph-1prom43::gfp, rol-6(su1006)]*	This work
NFB999	*vlcEx357[tph-1prom54::gfp, rol-6(su1006)]*	This work
NFB1000	*vlcEx358[tph-1prom54::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom54* Line 3	This work

NFB1014	*vlcEx546[tph-1prom55::gfp, rol-6(su1006)]*	This work
NFB1015	*vlcEx547[tph-1prom55::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom55* Line 3	This work
NFB1110	*vlcEx614[tph-1prom60::gfp, rol-6(su1006)]*	This work
NFB1113	*vlcEx615[tph-1prom60::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom60* Line 3	This work
NFB1025	*egl-18(ok290); vlcEx1[tph-1prom2::gfp, rol-6(su1006)]*	This work
NFB763	*vlcEx406[tph-1prom52::gfp, rol-6(su1006)]*	This work
NFB764	*vlcEx407[tph-1prom52::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom52* Line 3	This work
OH7443	*otEx3257[cat-1prom14::gfp, rol-6(su1006)]*	This work
OH7506	*otEx3303[cat-1prom14::gfp, rol-6(su1006)]*	This work
OH7508	*otEx3305[cat-1prom14::gfp, rol-6(su1006)]*	This work
NFB291	*vlcEx149[cat-1prom14::gfp, rol-6(su1006)]*	This work
NFB292	*vlcEx150[cat-1prom14::gfp, rol-6(su1006)]*	This work
NFB293	*vlcEx151[cat-1prom14::gfp, rol-6(su1006)]*	This work
NFB379	*vlcEx212[cat-1prom63::gfp, rol-6(su1006)]*	This work
NFB380	*vlcEx213[cat-1prom63::gfp, rol-6(su1006)]*	This work
NFB381	*vlcEx214[cat-1prom63::gfp, rol-6(su1006)]*	This work
NFB354	*vlcEx194[cat-1prom61::gfp, rol-6(su1006)]*	This work
NFB355	*vlcEx195[cat-1prom61::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom61* Line 3	This work
NFB330	*vlcEx175[cat-1prom60::gfp, rol-6(su1006)]*	This work
NFB331	*vlcEx176[cat-1prom60::gfp, rol-6(su1006)]*	This work
NFB457	*vlcEx259[cat-1prom73::gfp, rol-6(su1006)]*	This work
NFB460	*vlcEx262[cat-1prom73::gfp, rol-6(su1006)]*	This work
NFB411	*vlcEx236[cat-1prom71::gfp, rol-6(su1006)]*	This work
NFB412	*vlcEx237[cat-1prom71::gfp, rol-6(su1006)]*	This work

NFB501	*vlcEx277[cat-1prom71::gfp, rol-6(su1006)]*	This work
NFB502	*vlcEx278[cat-1prom71::gfp, rol-6(su1006)]*	This work
NFB528	*vlcEx292[cat-1prom74::gfp, rol-6(su1006)]*	This work
NFB529	*vlcEx293[cat-1prom74::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom74* Line 3	This work
NFB458	*vlcEx260[cat-1prom75::gfp, rol-6(su1006)]*	This work
NFB459	*vlcEx260[cat-1prom75::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom75* Line 3	This work
NFB557	*vlcEx302[cat-1prom76::gfp, rol-6(su1006)]*	This work
NFB558	*vlcEx303[cat-1prom76::gfp, rol-6(su1006)]*	This work
OH4219	*otEx2457[cat-1prom3::gfp, rol-6(su1006)]*	This work
OH4228	*otEx2460[cat-1prom3::gfp, rol-6(su1006)]*	This work
NFB773	*vlcEx410[cat-1prom83::gfp, rol-6(su1006)]*	This work
NFB774	*vlcEx411[cat-1prom83::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom83* Line 3	This work
NFB721	*vlcEx387[cat-1prom79::gfp, rol-6(su1006)]*	This work
NFB722	*vlcEx388[cat-1prom79::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom79 Line 3*	This work
NFB577	*egl-46(sy628)V; otEx3257[cat-1prom14::gfp, rol-6(su1006)]*	This work
NFB1027	*egl-18(ok290)IV; otEx3257[cat-1prom14::gfp, rol-6(su1006)]*	This work
NFB423	*vlcEx241[bas-1prom73::gfp, rol6(su1006)]*	This work
NFB424	*vlcEx242[bas-1prom73::gfp, rol6(su1006)]*	This work
No name	*bas-1prom73* Line 3	This work
No name	*bas-1prom71* Line 1	This work
NFB408	*vlcEx233[bas-1prom71::gfp, rol-6(su1006)]*	This work
No name	*bas-1prom71* Line 3	This work
NFB282	*vlcEx145[bas-1prom65::gfp, rol-6(su1006)]*	This work
NFB283	*vlcEx146[bas-1prom65::gfp, rol-6(su1006)]*	This work

NFB284	*vlcEx147[bas-1prom65::gfp, rol-6(su1006)]*	This work
NFB711	*vlcEx382[bas-1prom78::gfp, rol-6(su1006)]*	This work
NFB712	*vlcEx383[bas-1prom78::gfp, rol-6(su1006)]*	This work
No name	*bas-1prom78* Line 3	This work
NFB663	*vlcEx354[bas-1prom77::gfp, rol-6(su1006)]*	This work
NFB664	*vlcEx355[bas-1prom77::gfp, rol-6(su1006)]*	This work
NFB661	*vlcEx352[bas-1prom76::gfp, rol-6(su1006)]*	This work
NFB662	*vlcEx353[bas-1prom76::gfp, rol-6(su1006)]*	This work
NFB840	*vlcEx465[bas-1prom83::gfp, rol-6(su1006)]*	This work
NFB841	*vlcEx466[bas-1prom83::gfp, rol-6(su1006)]*	This work
No name	*bas-1prom83* Line 3	This work
NFB878	*vlcEx476[bas-1prom84::gfp, rol-6(su1006)]*	This work
NFB900	*vlcEx487[bas-1prom84::gfp, rol-6(su1006)]*	This work
NFB920	*vlcEx494[bas-1prom86::gfp, rol-6(su1006)]*	This work
NFB921	*vlcEx495[bas-1prom86::gfp, rol-6(su1006)]*	This work
No name	*bas-1prom86* Line 3	This work
NFB735	*ast-1(ot417)II; vlcEx16[bas-1prom18::gfp, rol-6(su1006)]*	This work
NFB837	*ast-1(ot417)II; vlcEx382[bas-1prom78::gfp, rol-6(su1006)]*	This work

Chapter II — Cross-regulation analysis (Figures 3.2.14-3.2.15)		
Strain name	Genotype	Source
OH10425	*otIs337(unc-86fosmid::NLS::YFP::H2B; ttx-3::mCherry)*	(Zhang et al. 2014)
VH1195	*hdIs42 [ast-1::YFP, rol-6(su1006)]*	(Schmid et al. 2006)
NFB1369	*ast-1(vlc19[ast-1::gfp])II*	This work
OP57	*unc-119(eds)III; wgIs57[sem-4TY1::EGFP::3XFLAG(92C12)+unc-119(+)]*	CGC, (Sarov et al. 2012)
MH1337	*kuIs34(sem-4::gfp)IV*	CGC, (Grant et al. 2000)
MH1346	*kuIs35(sem-4::gfp)*	CGC, (Grant et al. 2000)
RW11606	*unc-119(tm4063)III; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	CGC, Dr. Waterston Lab
NFB608	*vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]*	This work
OH9345	*otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]*	(Murgan et al. 2015)
OP650	*unc-119(tm4063)III; wgIs650[hlh-3::TY1::EGFP::3XFLAG + unc-119(+)]*	CGC, (Sarov et al. 2006)
NFB62	*ast-1(ot417)II; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)*	This work
NFB100	*ast-1(hd92)II; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)*	This work
NFB189	*sem-4(n1971)I; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)*	This work
NFB476	*hlh-3(tm1688)II; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)*	This work
NFB481	*egl-46(sy628)V; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)*	This work
NFB843	*egl-18(ok290)IV; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)*	This work
NFB1375	*unc-86(n846)III; ast-1(vlc19[ast-1::gfp])II*	This work
NFB1381	*sem-4(n1971)I; ast-1(vlc19[ast-1::gfp])II*	This work
NFB1377	*hlh-3(tm1688)II; ast-1(vlc19[ast-1::gfp])II*	This work
NFB1379	*egl-46(sy628)V; ast-1(vlc19[ast-1::gfp])II*	This work
NFB1383	*egl-18(ok290)IV; ast-1(vlc19[ast-1::gfp])II*	This work
NFB962	*ast-1(ot417)II; kuIs34(sem-4::gfp)IV*	This work
NFB173	*ast-1(hd92)II; kuIs34(sem-4::gfp)IV*	This work
NFB190	*unc-86(n846)III; kuIs34(sem-4::gfp)IV*	This work
NFB516	*hlh-3(tm1688)II; kuIs34(sem-4::gfp)IV*	This work
NFB525	*egl-46(sy628)V; kuIs34(sem-4::gfp)IV*	This work
NFB1300	*egl-18(ok290)IV; kuIs35(sem-4::gfp)*	This work
NFB754	*ast-1(ot417)II; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	This work
NFB997	*ast-1(hd92)II; hdEx237[ast-1(+), rol-6(su1006)]; unc-119(tm4063)III; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	This work
NFB688	*unc-86(n846)III; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	This work
NFB430	*sem-4(n1971)I; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	This work
NFB607	*hlh-3(tm1688)II; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	This work
NFB419	*egl-46(sy628)V; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	This work
NFB717	*ast-1(ot417)II; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB930	*ast-1(hd92)II; hdEx237[ast-1(+), rol-6(su1006)]; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB652	*unc-86(n846)III; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB720	*sem-4(n1971)I; zdIs13(tph-1::gfp)IV; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB651	*hlh-3(tm1688)II; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB871	*egl-18(ok290)IV; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB584	*ast-1(hd92)II; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]*	This work
NFB583	*unc-86(n846)III; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]*	This work
NFB586	*sem-4(n1971)I; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]*	This work
NFB969	*egl-46(sy628)V; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]*	This work
NFB1299	*egl-18(ok290)IV; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]*	This work
NFB188	*ast-1(hd92)II; zdIs13(tph-1::gfp)IV; hdIs42 [ast-1::YFP, rol-6(su1006)]0*	This work
LX960	*lin-15B(n765); vsIs97[tph-1p::DsRed + lin-15(+)]*	(Tanis et al. 2008)

Chapter II — HSN fate maintenance & GATA family RNAi (Figures 3.2.6, 3.2.16, 3.4.3)		
Strain name	Genotype	Source
NL2099	*rrf-3(pk1426)II*	CGC, Dr. Plasterk Lab.
NFB689	*rrf-3(pk1426)II; otIs517[(tph-1::SL2::YFP::H2B), ttx-3::mCherry, rol-6(su1006)]*	This work
NFB49	*rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV*	This work
NFB643	*rrf-3(pk1426)II; otIs221(cat-1::gfp)III*	This work

Chapter II — Overexpression of the HSN regulatory code (Figures 3.2.17-3.2.19)		
Strain name	Genotype	Source
NFB6	*zdIs13(tph-1::gfp); otIs198(hsp-16.2::ast-1; hsp-16.2::NLS::mCherry, ttx-3::DsRed)*	This work
NFB440	*zdIs13(tph-1::gfp)IV; vlcEx255[hsp-16.2:unc-86; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB257	*zdIs13(tph-1::gfp)IV; vlcEx96[hsp-16.2::sem-4; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB509	*zdIs13(tph-1::gfp)IV; vlcEx284 [hsp-16-2:hlh-3; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB624	*zdIs13(tph-1::gfp)IV; vlcEx334 [hsp-16.2:hlh-3; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB506	*zdIs13(tph-1::gfp)IV; vlcEx281 [hsp-16.2:egl-46; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB507	*zdIs13(tph-1::gfp)IV; vlcEx282 [hsp-16.2:egl-46; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB725	*zdIs13(tph-1::gfp)IV; vlcEx391(hsp-16.2::egl-18; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB726	*zdIs13(tph-1::gfp)IV; vlcEx392(hsp-16.2::egl-18; ttx-3::mCherry; rol-6(su1006)]*	This work
NFB267	*zdIs13(tph-1::gfp); vlcEx130[hsp-16.2::unc-86; hsp-16.2::ast-1; hsp-16.2::sem-4; ttx-3::mCherry; rol-6(su1006)]]*	This work
NFB268	*zdIs13(tph-1::gfp); vlcEx131[hsp-16.2::unc-86; hsp-16.2::ast-1; hsp-16.2::sem-4; ttx-3::mCherry; rol-6(su1006)]]*	This work
NFB1386	*zdIs13(tph-1::gfp)IV; vlcEx810[hsp::ast-1, hsp::unc-86, hsp::sem-4, hsp::hlh-3, hsp::egl-46, hsp::egl-18 (all at 15ng/ul), rol-6(su1006)(50ng/ul), ttx-3::mCherry(50ng/ul)]*	This work
NFB1387	*zdIs13(tph-1::gfp)IV; vlcEx811[hsp::ast-1, hsp::unc-86, hsp::sem-4, hsp::hlh-3, hsp::egl-46, hsp::egl-18 (all at 15ng/ul), rol-6(su1006)(50ng/ul), ttx-3::mCherry(50ng/ul)]*	This work

Chapter II — Genetic interaction within the HSN regulatory code (Figure 3.2.21)		
Strain name	Genotype	Source
MT1862	*unc-86(848)III*	CGC, Dr. Horvitz Lab.
MT6921	*sem-4(n2654)I*	CGC, (Basson & Horvitz 1996)
NFB1031	*hlh-3(tm1688)II; egl-46(sy628)V; otIs226(bas-1::gfp)IV*	This work
NFB695	*ast-1(ot417)II; egl-46(sy628)V; otIs226(bas-1::gfp)IV*	This work
NFB605	*sem-4(n2654)I; ast-1(ot417)II; otIs226(bas-1::gfp)IV*	This work
NFB253	*ast-1(ot417)II; unc-86(n848)III; otIs226[bas-1::gfp]IV*	This work
NFB958	*hlh-3(tm1688)II; egl-18(ok290)IV; otIs221(cat-1::gfp)III*	This work
NFB755	*egl-18(ok290)IV; sem-4(n2654)I; otIs221(cat-1::gp)III*	This work
NFB756	*egl-18(ok290)IV; sem-4(n2654)I; zdIs13(tph-1::gp)IV*	This work
NFB1033	*egl-18(ok290)IV; egl-46(sy628)V; otIs221(cat-1::gfp)III*	This work
NFB1032	*hlh-3(tm1688)II; egl-46(sy628)V; otIs221(cat-1::gfp)III*	This work
NFB395	*sem-4(n2654)I; ast-1(ot417)II; zdIs13(tph-1::gfp)IV*	This work
NFB648	*sem-4(n2654)I; unc-86(n848)III; otIs224(cat-1::gfp)V*	This work
NFB650	*hlh-3(tm1688)II; unc-86(n848)III; otIs224(cat-1::gfp)V*	This work
NFB931	*hlh-3(tm1688)II; egl-18(ok290)IV; otex2435 [bas1prom1 gfp (50ng/ul), rol-6]*	This work

Chapter III — Reporter fusion analysis for *de novo* expression in the HSN (Figures 3.3.3-3.3.6)		
Strain name	Genotype	Source
NFB1354	*vlcEx802 [abts-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1355	*vlcEx803 [abts-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1211	*vlcEx705 [acr-24::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1212	*vlcEx706 [acr-24::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1089	*vlcEx598 [ast-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1090	*vlcEx599 [ast-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1221	*vlcEx715 [bam-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work

↳

NFB1222	*vlcEx716 [bam-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1133	*vlcEx635 [C16B8.4::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1134	*vlcEx636 [C16B8.4::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1173	*vlcEx675 [C53B4.4::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1174	*vlcEx676 [C53B4.4::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1175	*vlcEx677 [ckr-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1176	*vlcEx678 [ckr-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1157	*vlcEx659 [daf-38::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1158	*vlcEx659 [daf-38::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1135	*vlcEx637 [dgn-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1136	*vlcEx638 [dgn-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1177	*vlcEx679 [F32D8.10::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1178	*vlcEx680 [F32D8.10::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1139	*vlcEx641 [nlp-10::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1140	*vlcEx642 [nlp-10::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1159	*vlcEx661 [fut-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1160	*vlcEx662 [fut-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1213	*vlcEx707 [gab-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1214	*vlcEx708 [gab-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1179	*vlcEx681 [glb-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1180	*vlcEx682 [glb-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1161	*vlcEx663 [kcc-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1162	*vlcEx664 [kcc-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1137	*vlcEx639 [kel-8::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1138	*vlcEx640 [kel-8::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1163	*vlcEx665 [klp-7::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1164	*vlcEx666 [klp-7::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1149	*vlcEx651 [lgc-49::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work

NFB1150	*vlcEx652 [lgc-49::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1215	*vlcEx709 [mec-10::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1216	*vlcEx710 [mec-10::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1181	*vlcEx683 [mgl-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1182	*vlcEx684 [mgl-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1217	*vlcEx711 [npr-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1218	*vlcEx712 [npr-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1141	*vlcEx643 [npr-3::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1142	*vlcEx644 [npr-3::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1352	*vlcEx800[pan-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1353	*vlcEx801 [pan-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1151	*vlcEx653 [pde-3::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1152	*vlcEx654 [pde-3::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1223	*vlcEx717 [shl-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1224	*vlcEx718 [shl-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1183	*vlcEx685 [snt-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1184	*vlcEx686 [snt-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1167	*vlcEx669 [sprr-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1168	*vlcEx670 [sprr-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1227	*vlcEx721 [sto-5::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1228	*vlcEx722 [sto-5::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1229	*vlcEx723 [tiam-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1230	*vlcEx724 [tiam-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1143	*vlcEx645 [tkr-2 w2-6::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1144	*vlcEx646 [tkr-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1219	*vlcEx713 [tol-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1220	*vlcEx714 [tol-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1145	*vlcEx647 [twk-17::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work

NFB1146	*vlcEx648 [twk-17::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1147	*vlcEx649 [tyra-3::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1148	*vlcEx650 [tyra-3::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1171	*vlcEx673 [unc-32::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1172	*vlcEx674 [unc-32::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1185	*vlcEx687 [unc-7::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1186	*vlcEx688 [unc-7::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1079	*vlcEx588 [aak-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1080	*vlcEx589[aak-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1077	*vlcEx586 [cat-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006)(100ng/ul)]*	This work
NFB1078	*vlcEx587 [cat-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1081	*vlcEx590 [kal-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1082	*vlcEx591 [kal-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1083	*vlcEx592 [kcc-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1084	*vlcEx593 [kcc-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1127	*vlcEx629 [sem-4 w3-13::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1128	*vlcEx630 [sem-4 w3-13::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1087	*vlcEx596 [sem-4 w15-16::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1088	*vlcEx597 [sem-4 w15-16::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1131	*vlcEx633 [sem-4 w18-23::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1132	*vlcEx634 [sem-4 w18-23::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1231	*vlcEx725 [f16g10.5::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1232	*vlcEx726 [f16g10.5::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1233	*vlcEx727 [flp-27::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1234	*vlcEx728 [flp-27::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1235	*vlcEx729 [gipc-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1236	*vlcEx730 [gipc-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1237	*vlcEx731 [irld-53::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work

NFB1238	*vlcEx732 [irld-53::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1239	*vlcEx733 [irld-62::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1240	*vlcEx734 [irld-62::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1241	*vlcEx735 [lurp-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1242	*vlcEx736 [lurp-2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1243	*vlcEx737 [plep-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1244	*vlcEx738 [plep-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1245	*vlcEx739 [slc-28.1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1246	*vlcEx740 [slc-28.1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1261	*vlcEx755 [stg-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1262	*vlcEx756 [stg-1::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)]*	This work
NFB1283	*vlcEx773[tub-1::MDM2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)*	This work
NFB1284	*vlcEx774[tub-1::MDM2::gfp::unc-54 3'UTR (50ng/ul), rol-6(su1006) (100ng/ul)*	This work

Chapter IV — Orientation bias in HSN *cis*-regulatory modules (Figure 3.3.7)		
Strain name	Genotype	Source
NFB1060	*vlcEx570[tph-1prom58::gfp, rol-6(su1006)]*	This work
NFB1061	*vlcEx571[tph-1prom58::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom58* Line 3	This work
NFB1062	*vlcEx572[tph-1prom59::gfp, rol-6(su1006)]*	This work
NFB1063	*vlcEx573[tph-1prom59::gfp, rol-6(su1006)]*	This work
No name	*tph-1prom59* Line 3	This work
NFB1064	*vlcEx574[cat-1prom85::gfp, rol-6(su1006)]*	This work
NFB1065	*vlcEx575[cat-1prom85::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom85 Line 3*	This work
NFB1066	*vlcEx576[cat-1prom86::gfp, rol-6(su1006)]*	This work
NFB1067	*vlcEx577[cat-1prom86::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom86* Line 3	This work
NFB1068	*vlcEx578[cat-1prom87::gfp, rol-6(su1006)]*	This work
NFB1069	*vlcEx579[cat-1prom87::gfp, rol-6(su1006)]*	This work
No name	*cat-1prom87* Line 3	This work
NFB1070	*vlcEx580[bas-1prom87::gfp, rol-6(su1006)]*	This work
NFB1071	*vlcEx581[bas-1prom87::gfp, rol-6(su1006)]*	This work
NFB1074	*vlcEx584[bas-1prom89::gfp, rol-6(su1006)]*	This work
NFB1075	*vlcEx585[bas-1prom89::gfp, rol-6(su1006)]*	This work
No name	*bas-1prom89* Line 3	This work

Chapter IV — Rescue of HSN mutant phenotype with mouse factors (Figure 3.4.4)		
Strain name	Genotype	Source
NFB290	*ast-1(ot417)II; zdIs13(tph-1::gfp)IV; vlcEx148[bas-1prom::ast-1, ttx-3::mCherry; rol-6(su1006)]*	This work
NFB336	*ast-1(ot417)II; zdIs13(tph-1::gfp)IV; vlcEx148[bas-1prom::Pet1, ttx-3::mCherry; rol-6(su1006)]*	This work
NFB499	*unc-86(n846)III; zdIs13(tph-1::gfp)IV; vlcEx503[kal-1prom::unc.86 genomic, ttx-3::mCherry, rol-6(su1006)]*	This work
NFB649	*sem-4(n1971)I; zdIs13(tph-1::gfp)IV; vlcEx511[kal-1prom::Sall2; ttx-3::mCherry, rol-6(su1006)]*	This work
NFB938	*hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV; vlcEx458[cat-4prom::hlh-3; ttx-3::mCherry, rol-6(su1006)]*	This work
NFB912	*hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV; vlcEx479[cat-4prom::Ascl1, ttx-3::mCherry, rol-6(su1006)]*	This work
NFB913	*hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV; vlcEx480[cat-4prom::Ascl1, ttx-3::mCherry, rol-6(su1006)]*	This work
NFB939	*egl-46(sy628)V; zdIs13(tph-1::gfp)IV; vlcEx471[cat-4prom::egl-46, ttx-3::mCherry, rol-6(su1006)]*	This work
NFB940	*egl-46(sy628)V; zdIs13(tph-1::gfp)IV; vlcEx472[cat-4prom::egl-46, ttx-3::mCherry, rol-6(su1006)]*	This work
NFB914	*egl-46(sy628)V; zdIs13(tph-1::gfp)IV; vlcEx481[cat-4prom::Inms1; ttx-3::mCherry, rol-6(su1006)]*	This work
NFB941	*egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V; vlcEx469[cat-4prom2::egl-18, ttx-3::mCherry, rol-6(su1006)]*	This work
NFB942	*egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V; vlcEx470[cat-4prom2::egl-18, ttx-3::mCherry, rol-6(su1006)]*	This work
NFB899	*egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V; vlcEx485[cat-4prom2::Gata3, ttx-3::mCherry, rol-6(su1006)]*	This work

Table 2.22
Other reagents and materials used in this Thesis

Reagent	Source / Reference
Worm mounting for visualisation under the microscope	
Sodium azide	Sigma, #26628-22-8
Microscope glass slides	Rogo Sampaic, #11854782
Coverslips (22*22mm)	VWR, #631-1570
Glass micropipette	Blaubrand intraMARK, #6121414
PCR	
Go Taq® DNA polymerase	Promega, #M7806
Expand Long template PCR system	Sigma, #11681834001
Q5 Hot Stat High-Fidelity 2X Master Mix	NEB, #M049S
dNTPs	Promega, #U1420
Nuclease-Free water	Sigma, #W4502
Proteinase K	Roche Life Science, #3115879001
Primers	Sigma and Metabion
Worm microinjection	
Halocarbon oil 700	Sigma, #H-8898
Microinjection capillary	Femtotip II, Eppendorf, #930000043
Needles	BD Microlance 3, #304000
Electrophoretic Mobility Shift Assay	
Anti-6xhistag antibody	Abcam, #ab18184
PowerBroth medium	Molecular Dimensions, #MD121061
His Trap HP column	GE Healthcare Life Sciences, #17-5248-01
Lipofectamine-2000	Invitrogen, #11668019
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Acros Organics, #BP1755-100
Anti-GFP antibody	Roche, #11814460001
ATP [γ-32P]	Perkin Elmer, #NEG002A250UC
T4 PNK	Thermo Scientific, #EK0031
Worm and mouse immunohistochemistry	
Collagenase type IV	Sigma, #5138
Rabbit anti-5-HT antibody	Sigma, #S-5545
Rabbit anti-Sall2	Sigma, #HPA004162
Goat anti-Brn2	Santa Cruz Biotechnology, #SC6029
Goat anti-5-HT	Abcam, #Ab66047
Alexa 555-conjugated donkey anti-rabbit	Molecular Probes, #A31572
Alexa 555-conjugated donkey anti-goat	Molecular Probes, #A21432
Alexa 488-conjugated donkey anti-goat	Molecular Probes, #A11055
Alexa 488-conjugated donkey anti-rabbit	Molecular Probes, #A21206
DAPI	Sigma, # D9542-5MG
FluorSave	Merck Millipore, #34578920ML
Phasmid neuron staining	
DiI	Molecular Probes, # D-282
N,N-dimethyl formamide	Sigma, #D4551
CRISPR/Cas9 GFP knock-in	
Hygromycin B solution	Gibco, #10687010
Isothermal Start Mix	See specific components above
Gibson Assembly Master Mix (2X)	See specific components above

Table 2.23
Apparatus used in this Thesis

Apparatus	Source / Reference
Dissecting scope	Zeiss Stemi 2000
Fluorescence scope	1FAxioZoom V16, Zeiss
Microinjection inverted microscope	Axio Vert.A1 Zeiss
Epifluorescence microscope	Zeiss Axioplan 2 microscope
Cryostat	Leica CM1900
Confocal microscope	TCS-SP8 Leica microscope
Image reader: detection of radioactive-marked membranes	Fujifilm FLA-500

Table 2.24
Software and Data Bases used in this Thesis

Software	Reference / Link to web
ImageJ	(-)
Adobe Photoshop CC	(-)
Adobe Illustrator CC	(-)
WormBase version 220	http://www.wormbase.org/#012-34-5
Genome Browser	https://genome.ucsc.edu
Transcription Factor Encyclopedia	http://www.cisreg.ca/cgi-bin/tfe/home.pl
CIS-BP	http://cisbp.ccbr.utoronto.ca
MatInspector (Genomatix)	https://www.genomatix.de/online_help/help_matinspector/matinspector_help.html
JASPAR	http://jaspar.genereg.net/cgi-bin/jaspar_db.pl
CRISPR Design	http://crispr.mit.edu
Benchling	http://benchling.com
iTF	(Kazemian et al. 2013); http://veda.cs.uiuc.edu/iTFs
GOrilla	(Eden et al. 2009); http://cbl-gorilla.cs.technion.ac.il
R	(The R Team 2016); https://www.r-project.org/
Bioconductor	(Huber et al. 2015); https://www.bioconductor.org/
pvclust (R package)	(Suzuki et al. 2006); www.sigmath.es.osaka-u.ac.jp/shimo-lab/prog/pvclust
ade4 (R package)	(Dray et al. 2007); pbil.univ-lyon1.fr/ade4/home.php?lang=eng
GraphPad QuickCalcs	http://www.graphpad.com/quickcalcs/

Results

Regulatory logic of serotonin pathway gene expression in the different serotonergic neuron classes of *Caenorhabditis elegans*

Chapter I

Serotonergic neurons share a battery of phylogenetically conserved enzymes and transporters, known as the 5-HT pathway genes, that allow the neurons to use 5-HT as a neurotransmitter → **Figure 1.7-B**. In this Chapter, we investigate how the expression of the 5-HT pathway genes is regulated in the different serotonergic neuron subtypes. First, we propose different models for serotonergic *cis*-regulatory logic and then we use GFP-based reporters to distinguish between them. This part of the project was performed in collaboration with Dr. Miren Maicas.

Establishment of possible models for the regulation of serotonin pathway gene expression in different serotonergic classes

C. elegans adult hermaphrodites contain three anatomically different 5-HT sinthesising neuron subclasses, which express the 5-HT pathway genes: the NSM neurosecretory motorneuron, the ADF chemosensory neuron and the HSN motorneuron → **Figure 1.12**. In our study we focused on the *tph-1* expressing, and thus 5-HT-producing neurons (NSM, ADF and HSN) and, from now on, will be referred to as serotonergic neurons, unless specified. These three serotonergic subclasses arise from different progenitors in development → **Figure 1.10**, fulfil different functions and, with the exception of the shared expression of 5-HT pathway genes, express different sets of terminal features → **Figure 3.1.1**.

As reviewed in Chapter I, distinct TFs are known to be required to control differentiation of NSM, ADF and HSN neurons (Desai et al. 1988; Sze et al. 2002; Xie et al. 2013; Zhang et al. 2014). However, how these TFs precisely regulate 5-HT pathway gene expression in each neuron subclass remains unknown.

One can envision two possible models to explain serotonergic regulatory logic. In model 1, subclass-specific TFs directly activate 5-HT pathway gene expression through independent *cis*-regulatory modules (CRMs) → **Figure 3.1.2, Model 1**. Alternatively, in model 2, subclass-specific TFs may activate a universal serotonergic subroutine of TFs that would then drive 5-HT pathway gene expression in all different subclasses of serotonergic neurons using the same CRM → **Figure 3.1.2, Model 2.**

To distinguish between these possibilities, we decided to perform *in vivo cis*-regulatory analyses of the *tph-1*(TPH2), *cat-1*(VMAT), *bas-1*(AADC), *cat-4*(GCH1) and *mod-5* (SLC6A4/SERT) 5-HT pathway genes. We systematically dissected the *cis*-regulatory regions of these five 5-HT pathway genes in the context of *gfp* reporters expressed in transgenic worms. We reasoned that the depth of this *cis*-regulatory analysis would provide evidence to identify the model or models that explain serotonergic regulatory logic. If serotonergic gene expression were controlled in a modular manner by distinct TFs in the different neuron subtypes (Model 1), we would observe that independent reporters would be specifically expressed in individual neuron subtypes. Alternatively, if 5-HT pathway

gene expression were defined by a master-regulator(s) that was activated by neuron-subtype TFs (Model 2) we would find CRMs globally expressed in all serotonergic subtypes.

Distinct *cis*-regulatory modules control serotonin pathway gene expression in the different subclasses of serotonergic neurons

To start, we analysed the expression pattern of the five 5-HT pathway genes using integrated reporters containing the complete upstream *cis*-regulatory region (full-length). The *tph-1* full-length reporters used in this project (*zdIs13* and *yzIs71*) (Clark & Chiu 2003; Sze et al. 2000) are exclusively expressed in the 5-HT producing neurons NSM, ADF and HSN. We found that *zdIs13* is additionally expressed in the pair of cholinergic neurons VC4 and VC5 that have been reported to contain weak and variable 5-HT immunoreactivity (Rand & Nonet 1997; Duerr et al. 1999) → **Table 3.1.1, → Figure 1.12**. Although VMAT antibody has been reported to be detected in all monoaminergic neurons (Duerr et al. 1999), *cat-1* full-length reporter (*otIs221*) is expressed in all monoaminergic except AIM. This includes the serotonergic NSM, ADF, RIH and HSN neurons, the serotonin-like VC4 and VC5 neurons, the dopaminergic CEPV, CEPD, ADE and PDE neurons, the tyraminergic neurons RIM and the octopaminergic neurons RIC → **Table 3.1.1**. *bas-1* full-length reporter (*otIs226*) is expressed in all serotonergic (NSM, ADF, AIM, HSN) neurons, except RIH, and dopaminergic neurons, as previously reported (Hare & Loer 2004) → **Table 3.1.1**. Finally, *cat-4* full-length reporter (*otIs225*) shows the same neuronal expression pattern as *bas-1* (Sze et al. 2002; Loer et al. 2015) → **Table 3.1.1**. In regard to *mod-5* expression, a 7.6 kb reporter containing all intergenic plus the first two exons and first intron of the *mod-5* gene (*mod-5B::gfp* in → **Figure 3.1.3-E, → Table 3.1.1**) has been described to be expressed in NSM, ADF and AIM neurons, but not in HSN (Jafari et al. 2011). A smaller version of the reporter (*mod-5A::gfp* in → **Figure 3.1.3-E**) containing only upstream *cis*-regulatory regions, was exclusively expressed in the ADF neuron (Jafari et al. 2011).

Our *cis*-regulatory study was carried out using extrachromosomal reporter lines. Prior to any analysis, we injected the full-length reporters containing upstream regulatory regions into the worms to see if they reproduced the expected expression. If this was the case, we kept these reporters for their posterior dissection. Conversely, if the upstream region did not recapitulate the expression of the gene, we also considered intronic regions as candidate CRMs, as is the case for *mod-5*. All transgenic strains generated are listed in → **Table 2.21**. We established three categories of GFP fluorescence according to the following criteria: expression values between 100-60% of scored cells are considered as '+' sign; values between 60-20% of expression would be considered 'partial expression' in the cells ('+/–'); values lower than 20% would be considered 'loss' of expression ('–'). Primary data for this part of the project is included in → **Annex 3.1.1.**

We began the analysis with the most specific serotonergic reporter, the tryptophan hydroxylase *tph-1* gene. We named *tph-1prom1* to the pPD95.75 plasmid containing all the intergenic regulatory sequence of the *tph-1* gene (1719 bp upstream plus the first 30 bp of exon 1, expressed as (–1719/+30)). As expected, we found expression in the NSM, ADF and HSN neurons. Additionally, we found expression in the VC4/5 neurons, similar to the integrated strain version (*zdIs13*) → **Figure 3.1.3-A.** We then cloned the first 377 bp preceding *tph-1* start codon into the pPD95.75 to generate *tph-1prom2*. This construct was expressed at comparable levels to the previous reporter in all the neuronal subtypes. The remaining 1341 bp upstream to *tph-1prom2* were cloned to generate a new construct named *tph-1prom8*, which was not expressed in any neuron, indicating that all the information about serotonergic regulatory logic must be contained in *tph-1prom2*. We then cloned the first 5' 146 bp from *tph-1prom2* to generate *tph-1prom6* but, again, showed no expression at all. The remaining 231 bp (*tph-1prom5*), however, was only expressed in the NSM neurons, indicating that it does not contain sufficient regulatory information to be activated in ADF and HSN neurons. Nonetheless, we wanted to know if an even smaller sequence would be still able to drive expression in the NSM neuron. We found out, indeed, that a 178 bp sequence right next

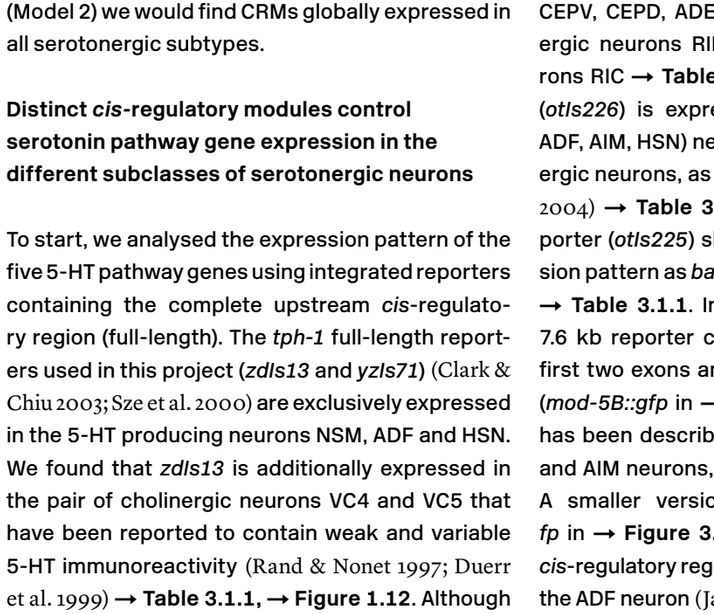


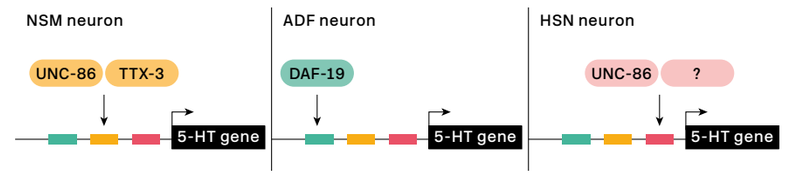
Figure 3.1.1
Heat map representation of HSN, NSM and ADF known expressed genes

Each serotonergic neuron subtype expresses different sets of genes. Only the 5-HT pathway genes and two additional genes *(aho-3 and nlp-3)* are expressed by the three serotonergic subtypes. Data obtained from www.wormbase.org and (Hobert 2016).

Figure 3.1.2
Models for serotonergic subclass specification

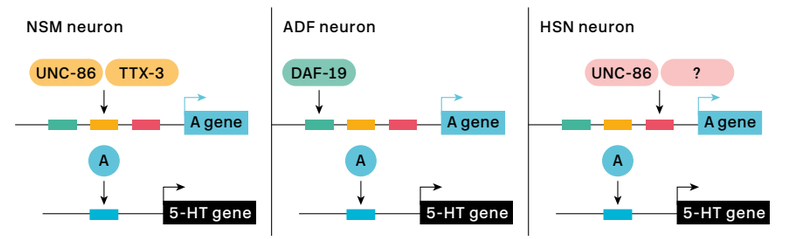
Model 1

Different TFs in each neuron subclass directly regulate 5-HT pathway gene expression through different *cis*-regulatory modules (CRMs), represented by different couloured boxes. 5-HT: serotonergic



Model 2

Different TFs in each neuron subclass are required to activate the expression of a common TF that directly regulates 5-HT pathway gene expression through the same CRM in all subclasses.



to the start codon (*tph-1prom3*) drove high levels of GFP in NSM. Therefore, we established this region as the CRM of the *tph-1* gene, in the NSM neuron. However, as GFP fluorescence was lost in ADF and HSN neurons in *tph-1prom6* and *tph-1prom5*, we hypothesised that maybe the regulatory elements controlling *tph-1* expression in these cells could be somewhere in the middle of both constructs. With this in mind, we cloned 99 bp containing the 3' end of *tph-1prom6* and the 5' start of *tph-1prom5* to generate *tph-1prom17* and assessed expression. ADF neuron regained robust GFP expression with this construct, so we considered this small DNA fragment the *tph-1* CRM in the ADF. As HSN GFP expression was only observed with the *tph-1prom2* reporter, we concluded that this is the *tph-1* CRM in the HSN. Our *cis*-regulatory analysis of the *tph-1* gene revealed that independent modules are required to achieve expression of this gene in the different serotonergic subtypes.

We next moved on to dissect the regulatory elements of the *cat-1* gene. *cat-1prom1* contained 2.5 kb upstream of the *cat-1* gene start codon; almost all the intergenic region. This construct is expressed in all the cells that are known to express CAT-1, as observed with antibody staining (Duerr et al. 1999), except for AIM → **Figure 3.1.3-B, → Annex 3.1.1**. We then cloned the furthest 752 bp from the ATG (*cat-1prom2*) and expression was lost in all serotonergic cells although maintained in dopaminergic cells, suggesting that this sequence is not required for serotonergic expression of *cat-1*. The remaining 1584 bp (*cat-1prom3*) were expressed in the same neurons as *cat-1prom1*, except for RIH and RIC neurons. We further divided this promoter in two new ones: *cat-1prom12* and *cat-1prom11*. The former was expressed in NSM, ADF and HSN serotonergic neurons and also in all dopaminergic neurons. The latter, however, was only expressed in the ADF serotonergic neuron. This construct was also expressed in VC4 and VC5 serotonergic-like neurons and in the octopaminergic RIC neuron. Next, we wanted to know if an even smaller DNA sequence could be enough to drive *cat-1* expression in the ADF neuron, similarly to its *tph-1* CRM. We further divided it into *cat-1prom35* and *cat-1prom36* and only the second one maintained GFP expression exclusively in the ADF cell. Finally *cat-1prom37*, an even shorter piece from the *cat-1prom36* that contains the first 185 bp from the start codon, was established as the minimal CRM for the *cat-1* gene in the ADF neuron. Next, we aimed to identify the CRM of NSM and HSN serotonergic neurons, so we moved back to *cat-1prom12*. We divided this promoter in two, generating *cat-1prom13* and *cat-1prom14*. The former lost expression in all the cells while the latter maintained GFP in the three serotonergic neurons, in similar levels. Expression in all dopaminergic neurons was still observed too. We divided this *cat-1prom14* into *cat-1prom26* and *cat-1prom27*. While the first one lost GFP expression, the second one was exclusively expressed in the NSM neuron. Therefore, we considered this sequence the minimal CRM of the *cat-1* gene in the NSM neuron. Interestingly, this reporter was ectopically expressed in several neurons located between the head and the vulva of the worm. This could be indicating that this DNA sequence is missing a repressor element that in wild type worms acts to suppress *cat-1* expression in these cells. Lastly, as happened with *tph-1*, we were unable to find a smaller region that was expressed in the HSN neuron. For this reason, we used *cat-1prom14* (522 bp) as the minimal CRM of the *cat-1* gene in the HSN neuron. This CRM, however, also contained regulatory information for the ADF neuron. Thus, we found two independent CRMs for the ADF neuron, suggesting that *cat-1* expression in this cell is redundantly regulated.

We then focused on the study of the *bas-1* gene. *bas-1prom1* reporter, which consists of 1.5 kb upstream of the start codon of the gene, recapitulates

Table 3.1.1
Expression of the serotonin pathway genes in the monoaminergic neurons of *Caenorhabditis elegans*

5-HT: serotonergic neurons, DA: dopaminergic neurons, Tyr: tyraminergic neurons, Oct: octopaminergic neurons. (+): Expression depends on the use of a reporter or an antibody; VC4/5 express *zdIs13(tph-1::gfp)* but not *yzIs71* or the fosmid reporter *otIs517;* AIM are immunoreactive to VMAT but do not express the *otIs221(cat-1::gfp)* reporter.

Gene	NSM	ADF	HSN	RIH	AIM	VC4/5	CEPV	CEPD	ADE	PDE	RIM	RIC	Reference
	5-HT						DA				Tyr	Oct	
tph-1	+	+	+	-	-	(+)	-	-	-	-	-	-	(Sze et al. 2000), VC4/5 laboratory observation
cat-1	+	+	+	+	(+)	+	+	+	+	+	+	+	(Duerr et al. 1999), (Sze et al. 2000)
bas-1	+	+	+	-	+	-	+	+	+	+	-	-	(Hare and Loer 2004)
cat-4	+	+	+	-	+	-	+	+	+	+	-	-	(Sze et al. 2002), (Loer et al. 2015)
mod-5	+	+	-	+	+	-	-	-	-	-	-	-	(Jafari et al. 2011)

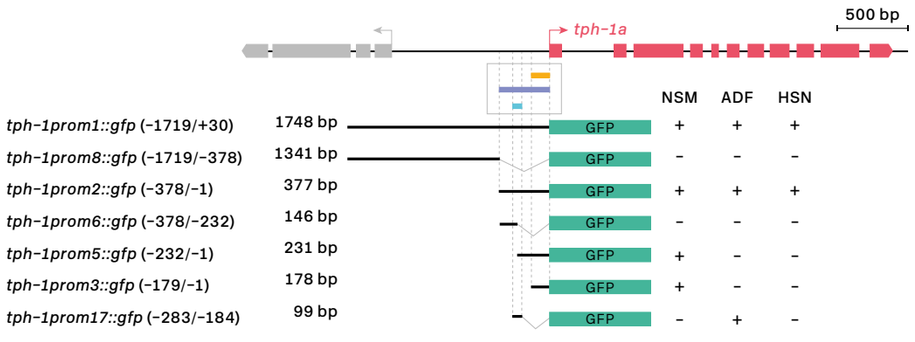
the previously described expression pattern (Hare & Loer 2004) → **Figure 3.1.3-C, → Annex 3.1.1**. We next isolated the first 647 bp of the reporter to create *bas-1prom2*, which was only expressed in NSM and in HSN. We concentrated in these pair of serotonergic neurons and divided the reporter into two parts (*bas-1prom13* and *bas-1prom14*). The first, but not the second, construct was also exclusively expressed in both neuronal subtypes. We repeated this procedure, dividing *bas-1prom13* into *bas-1prom15* and *bas-1prom16*, and only one reporter (*bas-1prom16*) maintained GFP expression in NSM and HSN. Once again, after dividing bas-1prom16 into *bas-1prom17* and *bas-1prom18*, only *bas-1prom18* kept the expression in both neuronal subtypes, although its penetrance in the HSN neuron was lower than that of *bas-1prom1* reporter. We concluded that NSM and HSN share the same minimal CRM of *bas-1* gene. Next, going back to ADF neuron, we analysed *bas-1prom3* that carries the 860 bp complementary to *bas-1prom2*. In this case, we did see GFP expression in the ADF neuron and also in all dopaminergic neurons. A further division of *bas-1prom3* into *bas-1prom4* and *bas-1prom5* revealed that the vital information for ADF neuron was found in the 285 bp contiguous to the *bas-1* start codon (*bas-1prom5*). We did an extra division of the reporter to generate *bas-1prom6* and *bas-1prom7*. The former only showed expression in the PDE dopaminergic neuron, while the latter was only expressed in the ADF neuron. Consequently, we assigned this 162 bp, adjacent to the start codon, as the minimal CRM of the *bas-1* gene in the ADF. As with the previous 5-HT pathway genes, we identified independent modules for different serotonergic neuron subtypes, with the difference that NSM and HSN seem to share a common regulatory module.

Next, we assessed *cat-4* regulatory logic. Previous studies in the laboratory had identified an 896 bp enhancer, just before the *cat-4* ATG codon, that mirrored *cat-4* full length integrated reporter strain (*otIs225*). Thus, we started our *cat-4* analysis using this shorter reporter (*cat-4prom4*)

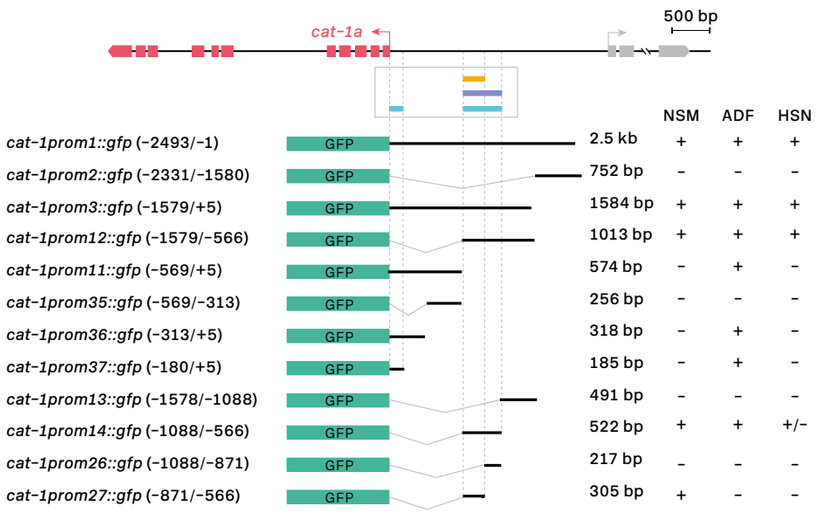
Figure 3.1.3
***Cis*-regulatory analysis of the serotonin pathway genes in the serotonergic neurons**

White boxes underneath each gene summarise the smallest CRM that drives expression in each serotonergic neuron subclass. Yellow line indicates NSM minimal, purple line indicates HSN minimal and blue line indicates ADF minimal. Thick black lines symbolise the genomic region placed in front of GFP (green box) and dashed lines are used to place each construct in the context of the locus. Numbers in brackets represent the coordinates of each construct referred to the ATG.
+ : > 60% GFP positive cells;
+/− : 20-60% GFP cells;
− < 20% GFP cells.
n > 30 worms per line.
See Annex 3.1.1 for primary data and for complete analysis in all monoaminergic neurons that express the 5-HT pathway genes.
Work performed next to Dr. Miren Maicas.

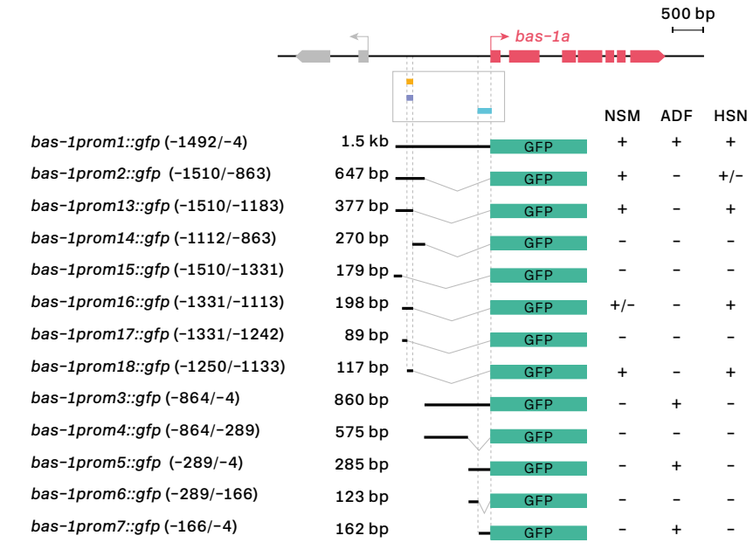
A



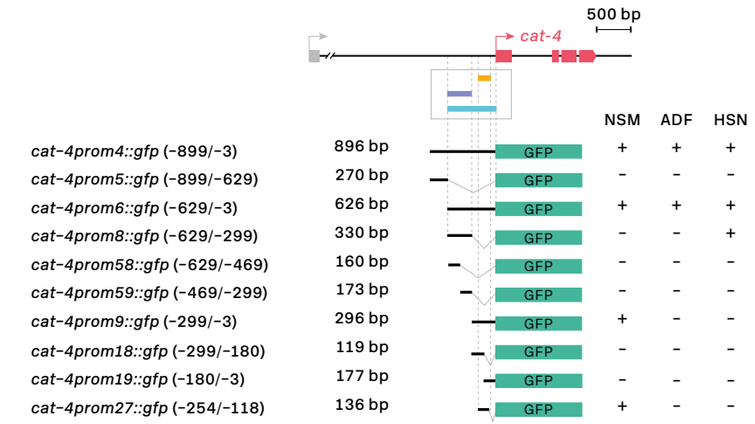
B



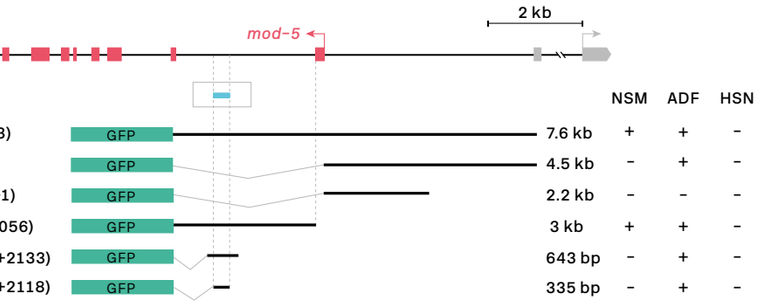
C



D



E



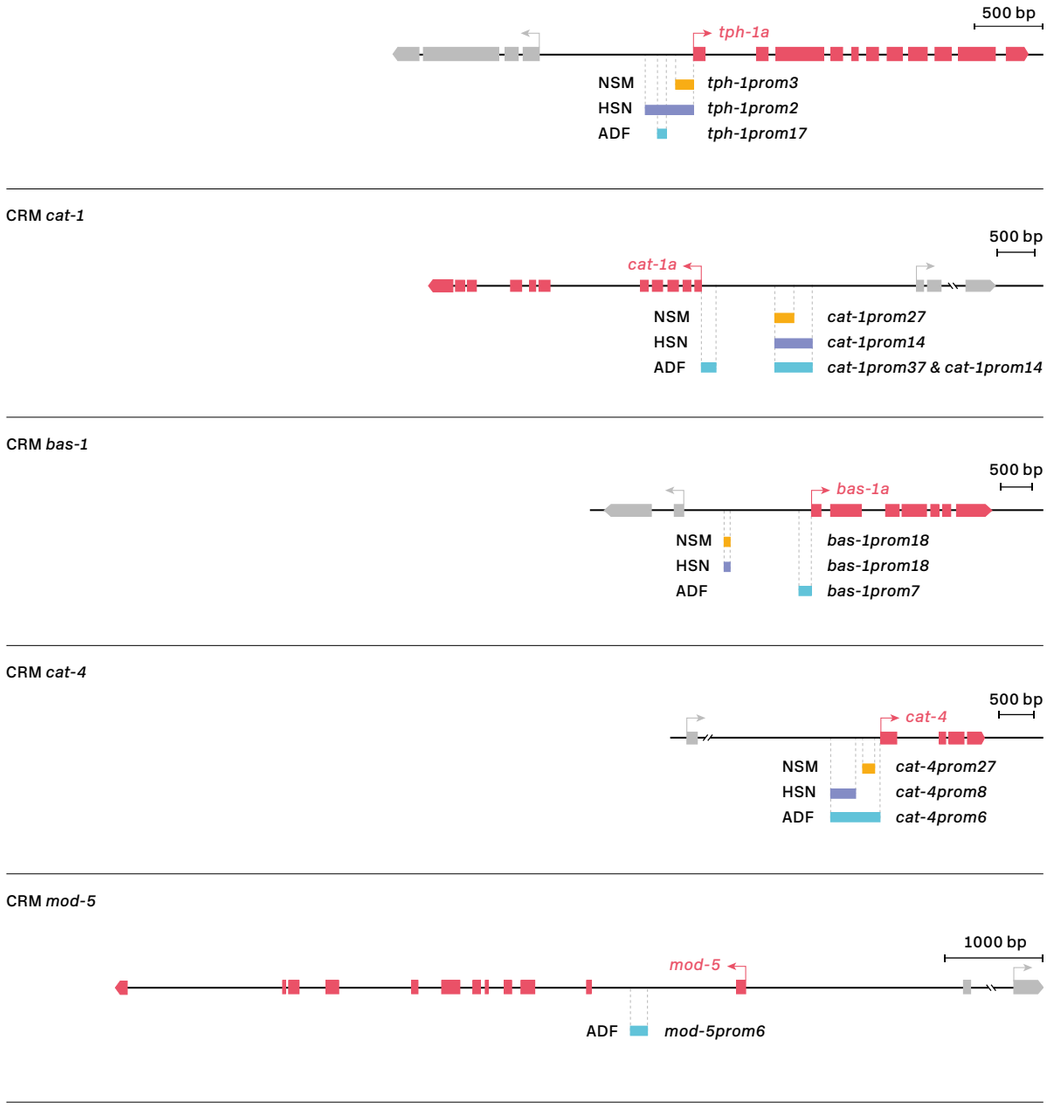
→ **Figure 3.1.3-D, → Annex 3.1.1**. The first 270 bp of the reporter (furthest from the start codon) did not show any GFP expression (*cat-4prom5*), while the complementary 626 bp (*cat-4prom6*) showed an identical expression pattern as *cat-4prom4*. We then divided *cat-4prom6* into two new reporter lines (*cat-4prom8* and *cat-4prom9*). *cat-4prom8* was exclusively expressed in the HSN neuron. To test if a smaller sequence could contain all the information required for *cat-4* expression in this cell, we further divided the promoter into two (*cat-4prom58* and *cat-4prom59*). However, none of them showed GFP expression in the HSN or in any other neuron. Thus, we concluded that *cat-4prom8* is the minimal CRM of *cat-4* in the HSN. To continue with ADF and NSM, we analysed *cat-4prom9*, the complementary sequence to *cat-4prom8*, with respect to *cat-4prom4*. This reporter showed GFP expression in the serotonergic NSM neuron, and also in all dopaminergic neurons. To delimit the NSM minimal, we further divided the sequence to generate two new reporters (*cat-4prom18* and *cat-4prom19*) but both lost expression in the cell, although some remained in the dopaminergic neurons. Nonetheless, we found that a 136 bp sequence that overlapped the 3' end of *cat-4prom18* and the 5' start of *cat-4prom19* was sufficient to drive GFP expression specifically in the NSM serotonergic neuron (*cat-4prom27*). Of note, some expression in the dopaminergic CEPV neurons persisted too. We considered this *cat-4prom27* as the minimal CRM of the *cat-4* gene in the NSM neuron. None of the reporters analysed but *cat-4prom6* was enough to drive GFP expression in the ADF. Hence, we considered that this should be the CRM of *cat-4* in the ADF neuron.

Finally, as *mod-5* is the only gene from the 5-HT pathway that is not expressed in all 5-HT producing neurons (it is absent in the HSN neuron) we did not analyse its regulatory logic in such detail. We designed a reporter with more than 2 kb of the upstream regulatory region (*mod-5prom1*) to see if it could recapitulate the previously described expression pattern (Jafari et al. 2011). However, this reporter showed no GFP expression in any serotonergic neuron, although it was expressed in a pair of unknown neurons in the tail → **Figure 3.1.3-E, → Annex 3.1.1**. As this *cis*-regulatory region did not contain the information required for expression in serotonergic neurons, we decided to analyse intronic regions. We generated a second reporter containing the whole intron 1 (*mod-5prom3*) and detected GFP expression in NSM and ADF neurons. A 335 bp long intronic region contained within this reporter (*mod-5prom6*) was considered the minimal CRM required for *mod-5* expression in the ADF neuron.

In summary, to decipher the regulatory logic of serotonergic gene expression we generated more than 100 transgenic lines containing 49 different reporter gene fusions, spanning from about 100 to 2500 base pairs. → **Figure 3.1.4** summarises what we have learnt from this promoter bashing regarding the 5-HT producing neurons NSM, ADF and HSN. Briefly, different CRMs are required for 5-HT pathway gene expression in the three neuron subtypes. These CRMs can be found in up to 2 kb upstream *cis*-regulatory regions of the genes, except for *mod-5* that exhibits intronic control of its expression. We detected ectopic expression in other neurons that do not normally express the 5-HT pathway genes in one of the short reporters for *cat-1*. Additionally, two independent and redundant *cat-1* CRMs were identified for the ADF neuron. The fact that different CRMs are active in specific subclasses of serotonergic neurons supports the idea that different TFs directly regulate 5-HT pathway gene expression in NSM, ADF, and HSN neurons (Model 1) and discards the possibility of having a common target gene that will in turn activate the same CRM for all of them (Model 2).

Figure 3.1.4
Summary of serotonergic *cis*-regulatory logic

Cis-regulatory analysis reveals that serotonergic regulatory logic follows model 1 (Figure 2.2), where different CRMs are required for the expression of the 5-HT pathway genes in a neuron subtype-specific manner and is predicted to be activated by different TFs.



A candidate approach to identify terminal selectors for HSN neuron serotonergic fate

Chapter II

In Chapter I we have shown that the *cis*-regulatory logic of *C. elegans* serotonergic system is neuron subtype-specific. We have identified independent and generally non-redundant CRMs that drive expression of the 5-HT pathway genes in NSM, ADF and HSN serotonergic neurons. Recently, our group has contributed to elucidate the NSM regulatory logic: a terminal selector code composed by TTX-3 and UNC-86 TFs directly regulates the terminal fate of the neuron, including direct activation of the 5-HT pathway genes (Zhang et al. 2014). Regarding the ADF neuron, only the TF DAF-19 is known to be required for tryptophan hydroxylase expression, although it is not clear if it is a direct or indirect action (Xie et al. 2013). The HSN neuron has been very well characterised in the past and several TFs are known to regulate its development (Desai et al. 1988; Doonan et al. 2008; Wu et al. 2001; Sze et al. 2002). To deepen our understanding of how cell type-specific transcriptional programmes are implemented we decided to focus the rest of this Thesis on the best characterised serotonergic neuron subtype, the HSN neuron, and carried out an extensive dissection of HSN terminal differentiation transcriptional rules. Dr. Miren Maicas and PhD student Ángela Jimeno, both coworkers at the laboratory of Dr. Nuria Flames, have collaborated in the elaboration of this Chapter, with the mutagenesis and electrophoretic mobility shift assay (EMSA) experiments and with the RNAi assays.

Transcription factors from six different families are required for HSN terminal differentiation

As described in the Introduction, Desai and colleagues carried out a mutant screen looking for genetic components of the HSN function (Desai et al. 1988). In this study, 38 genes were identified as HSN-defective mutants, mainly distinguished by an evident *egl* phenotype → **Table 1.2.** Additionally, as the *egl* phenotype is so easy to identify, several laboratories have reported mutants with HSN defects. In the present work, we decided to follow a candidate approach and select mutant alleles for TFs that showed reduced or absent amounts of HSN 5-HT staining and exhibited *egl* phenotype. We discarded genes that code for TFs known to act early in the developmental pathway of the HSN, such as *egl-5* that affects the HSN precursor cell (Baum et al. 1999; Guenther & Garriga 1996; Singhvi et al. 2008), or *egl-44* that is known to control other candidate regulators for the HSN as *egl-46* (Wu et al. 2001). We also did not consider those genes whose 5-HT staining defects in the mutant have been linked to severe migration defects, as *egl-43* and *ham-2* (Baum et al. 1999). Following these criteria we ended up with a list of four TFs that appear as potential regulators of the HSN terminal fate: EGL-46 (INSM ZnF TF) (Wu et al. 2001; Desai et al. 1988), HLH-3 (bHLH TF) (Doonan et al. 2008), SEM-4 (SPALT ZnF TF) (Basson & Horvitz 1996; Grant et al. 2000) and UNC-86 (POU TF) (Sze et al. 2002; Finney & Ruvkun 1990). All of them are known to

be expressed in the HSN and, additionally, UNC-86 and HLH-3 are known to regulate some of the 5-HT pathway genes (Doonan et al. 2008; Sze et al. 2002). In the specific case of UNC-86, this regulation is known to be direct upon *tph-1* and it is also lineage independent. For these reasons, UNC-86 appears as a good terminal selector candidate to regulate HSN serotonergic fate.

Furthermore, we included two more candidates in the analysis whose mutant alleles are defective for 5-HT staining although this phenotype has not been described in the literature to date: the ETS TF AST-1 and the GATA factor EGL-18. Previous work in the laboratory identified the *ast-1* gene as an important inducer of dopaminergic fate; i.e. AST-1 directly binds and activates the transcriptome of dopaminergic cells (Flames & Hobert 2009). Dopaminergic and serotonergic neurons are both monoaminergic and share some components of the monoamine synthetic pathway. During the analysis of *ast-1* mutants *cat-1* (vesicular monoamine transporter VMAT) gene expression defects were incidentally detected in the HSN neuron, while other serotonergic cell types remained unaffected. With regard to EGL-18, mutants were originally reported as HSN migration defective (in terms of cell position along HSN migrating path, cell position relative to the ventral nerve cord and branching) and to display an *egl* phenotype (Garriga, Desai et al. 1993; Desai et al. 1988). These studies used *n474*, *n475* and *n162* alleles that, although predicted to encode polypeptides that are truncated before the DNA-binding domain, showed normal levels of 5-HT in the HSN. However, there are reports describing how nonsense mutations can lead to exon skipping or alternative start site usage (Ginjaar et al. 2000; Davuluri et al. 2008), thus we decided to use the *ok290* allele, an 816 bp deletion spanning intron 2 and exon 3 that removes the zinc finger region (Koh et al. 2002) → **Figure 3.2.1**. We noticed a 5-HT staining phenotype with this allele that had not previously been described, and thus decided to include it in the analysis.

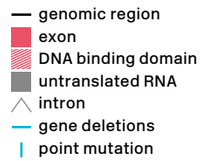
First, we obtained null loss-of-function mutants for the six genes from the Caenorhabditis Genetic Center (CGC). For the analysis of *ast-1*, we used the *ot417* hypomorphic allele (G>A substitution affecting the DNA binding domain) because null alleles show an L1 larval arrest phenotype, whilst HSN matures at late L4 larval stage. Details on the specific allelic nature of the mutants used in this work are summarised in → **Figure 3.2.1** and → **Table 2.1**.

Next, we wondered if the loss of these six TFs could induce an incapability of the cell to synthesise 5-HT. Anti-5-HT staining in mutant worms reveals significant defects in neurotransmitter production that ranges in severity; mutants for *ast-1*, *unc-86* and *sem-4* show practically no detectable levels of 5-HT, while in *hlh-3*, *egl-46* and *egl-18* mutants there is a partial loss of 5-HT staining → **Figure 3.2.2-A** and **B**, → **Annex 3.2.1**. In agreement to our *cis*-regulatory analysis, these 5-HT defects seem exclusive of the HSN neuron, as we did not observe any significant defect in the NSM or ADF neurons → **Annex 3.2.1**.

We then investigated to what extent was the 5-HT biosynthetic pathway gene expression affected in these mutants. We crossed mutant animals with the four 5-HT pathway gene transcriptional reporters that are expressed in HSN: *tph-1* (*zdIs13*, *yzIs71*), *cat-1* (*otIs221*, *otIs224*), *bas-1* (*otIs226*, *otEx2435*) and *cat-4* (*otIs225*, *otEx2470*) and scored the resulting fluorescent protein expression in the HSN. We found that gene expression is affected at different levels → **Figure 3.2.2-A** and **B**, → **Annex 3.2.1**. *unc-86* and *sem-4* showed the strongest phenotypes with a complete loss of expression of the four 5-HT pathway genes analysed, except for *tph-1* (TPH2) whose expression was reduced by half in *sem-4* mutant background. *ast-1* mutants also exhibited complete loss of *tph-1* and *cat-1* (VMAT) expression, but *bas-1* (AADC) and *cat-4* (GCH1)

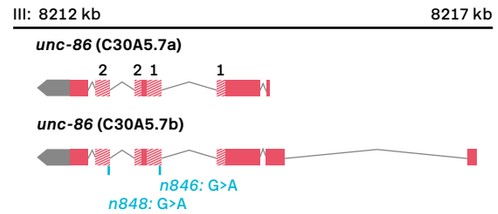
Figure 3.2.1
Genetic locus of HSN candidate regulators and mutant alleles used in this work

Schematic representations are to scale. Thick black lines symbolise the genomic region and chromosomal location is indicated. Red boxes symbolise exons. Exons, grey lines symbolise introns and grey boxes symbolise untranslated RNA of the gene. DNA binding domains are indicated with red/white stripes: alternative gene isoforms are included. Allelic mutations used in this work are indicated in blue: vertical lines correspond to point mutations and horizontal lines indicate gene deletions. Specific nucleotide changes are also included in blue.



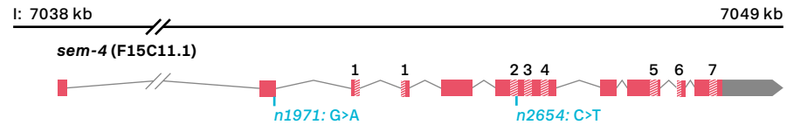
A) *unc-86* locus

1: POU DNA binding domain,
2: Homeodomain DNA binding domain.



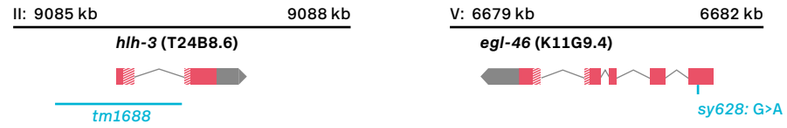
B) *sem-4* locus

1-7: Zinc finger domains. Intron 1 has been shortened in the image as indicated by –//–.



C) *hlh-3* locus

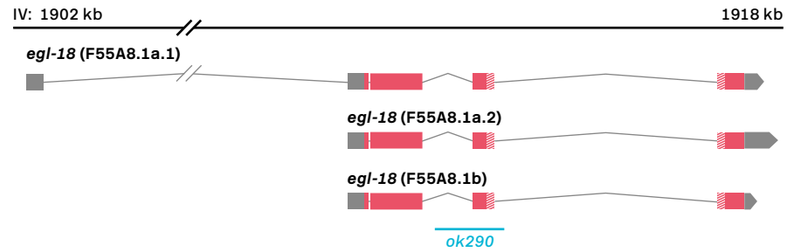
D) *egl-46* locus



E) *ast-1* locus

F) *egl-18* locus

egl-18 isoform a1 contains a long intron that has been shortened in the image, as indicated by –//–. For specific description of the alleles see Table 2.1.



remained unaffected. *hlh-3* and *egl-46* mutants showed severe defects in *tph-1* and partial defects in *cat-1* and *bas-1* expression, while *cat-4* levels remained comparable to wild type. Finally, *egl-18* showed the weakest phenotype, with only partial *tph-1* expression defects.

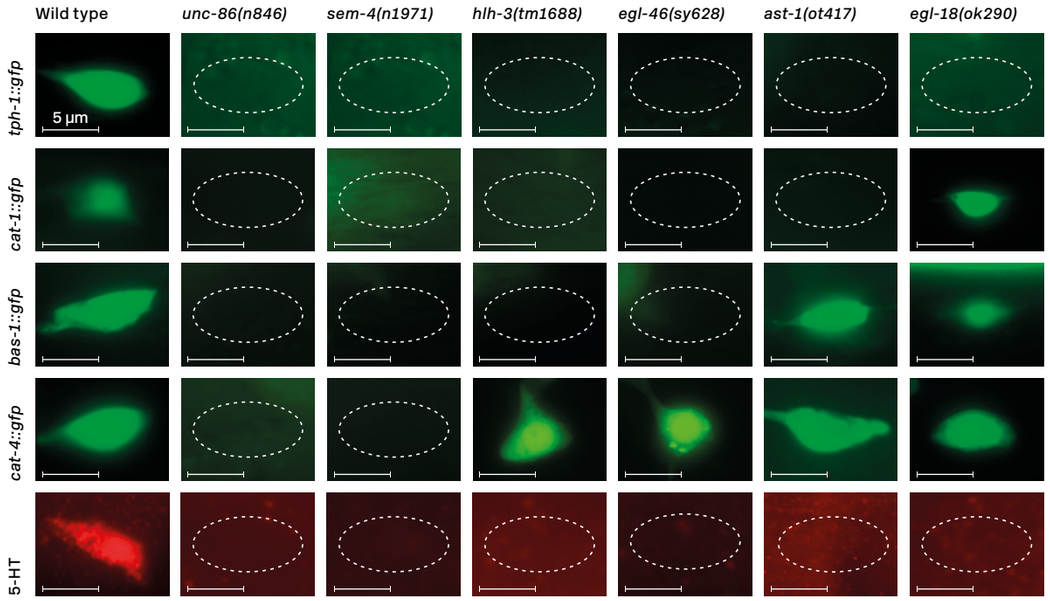
Regarding *ast-1* phenotype, to discard the possibility that the lack of *bas-1* and *cat-4* expression defects was due to the nature of the *ot417* hypomorph allele, we decided to perform mosaic analysis using the *hd92* null allele → **Figure 3.2.3,** → **Table 2.1**. We rescued null *ast-1(hd92)* lethality with an extrachromosomal array containing *ast-1* cDNA and a *cat-1::mCherry* red marker that is expressed in all monoaminergic neurons, including the HSN → **Figure 3.2.3-A,** → **Table 2.1**. We scored for *bas-1* expression in viable young adult worms, in which the array had rescued *ast-1* lethality, but that had lost the rescuing array in the HSN neuron, assessed by the lack of red marker. 87 out of 87 analysed mutant cells showed normal *bas-1* reporter expression. In this way, we confirmed that *bas-1::gfp* expression does not require AST-1 → **Figure 3.2.3-B**. As a control we analysed *cat-1::gfp* expression in these mosaic animals and saw similar defects in expression as in *ot417* animals (26 out of 29 mutant cells lost expression in the HSN).

Again, expression defects are generally specific for the HSN subclass while ADF and NSM remain mainly unaffected. An exception is *unc-86* mutant

Figure 3.2.2
Analysis of serotonin pathway gene expression in mutant animals for the six candidate regulators of the HSN

A) Serotonin pathway analysis in mutant animals

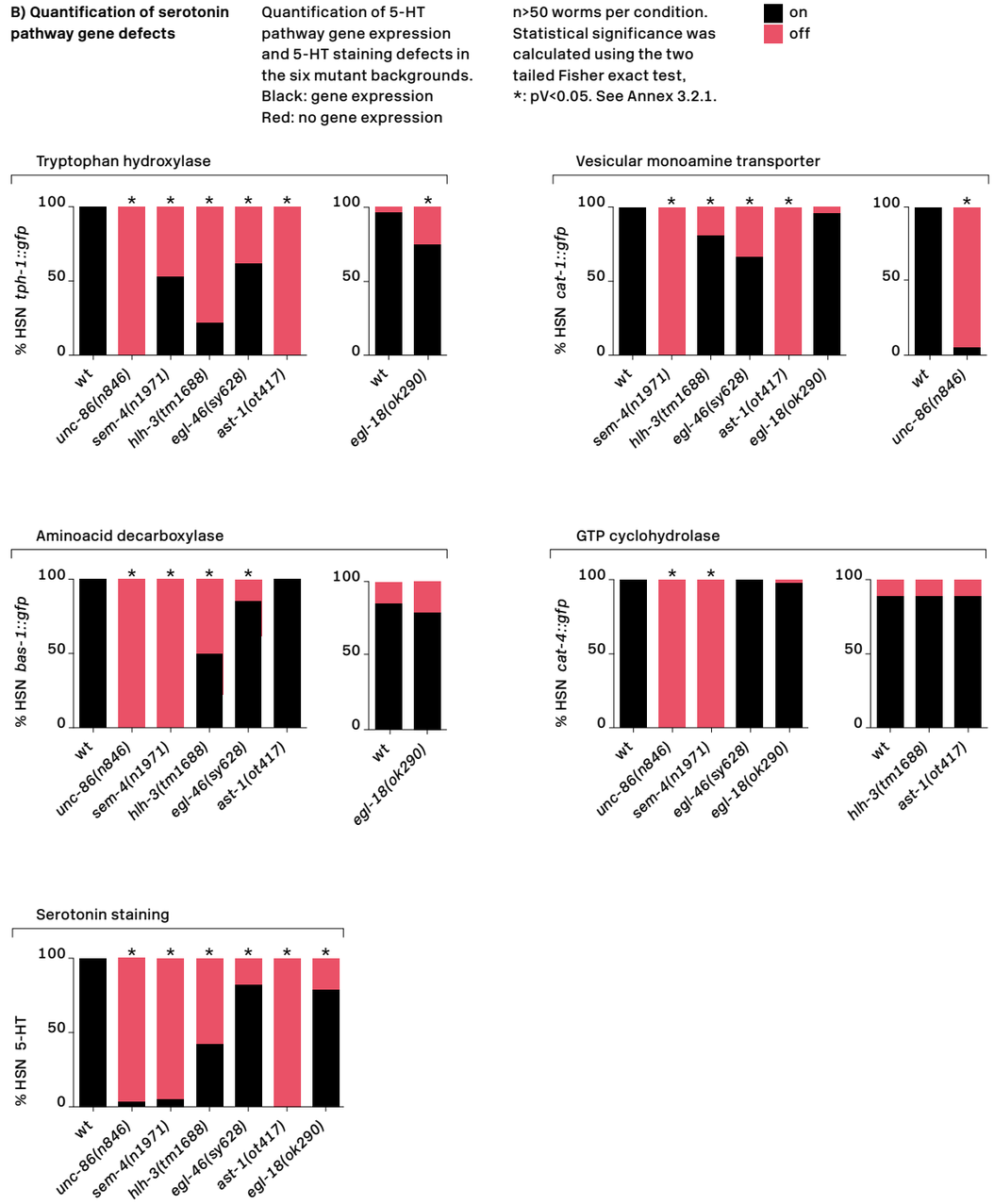
Micrographs showing 5-HT pathway gene expression and 5-HT staining defects in *unc-86(n846), sem-4(n1971), hlh-3(tm1688), egl-46(sy628), ast-1(ot417)* and *egl-18(ok290)* mutants. Different reporters were used for the same 5-HT pathway gene whenever the reporter transgene was integrated in the same chromosome as the mutation *(tph-1::gfp: zdIs13* and *yzIs71; cat-1::gfp: otIs221* and *otIs224, bas-1::gfp: otIs226* and *otEx2435, cat-4::gfp: otIs225* and *otEx2470).*



B) Quantification of serotonin pathway gene defects

Quantification of 5-HT pathway gene expression and 5-HT staining defects in the six mutant backgrounds.
Black: gene expression
Red: no gene expression

n>50 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test, *: pV<0.05. See Annex 3.2.1.

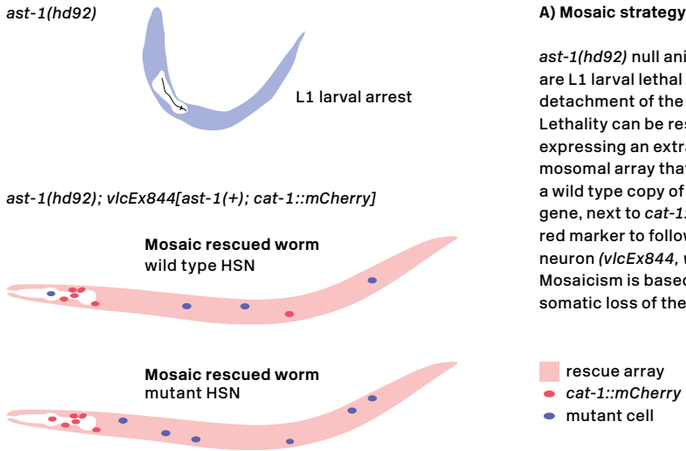


that showed expression defects for the four 5-HT pathways genes in the NSM neuron, as previously described (Sze et al. 2002; Zhang et al. 2014). In addition, we also observed very mild phenotypes in the ADF for *tph-1* in *egl-46* mutant background (90±3%) and for *bas-1* in *sem-4* (90±3%) and *hlh-3* (86±3%) mutants → **Annex 3.2.1**.

As explained in the introduction, terminal selectors do not only control a specific feature of the cell, as can be neurotransmitter type. Instead, they tend to broadly regulate expression of the terminal transcriptome of the neuron (Hobert 2008). Therefore, we wanted to test if these six TFs were also required for a more extensive regulation of the HSN transcriptome and analysed nine additional transcriptional reporters of HSN expressed genes not related to 5-HT biosynthesis: *kcc-2c* (potassium chloride co-transporter), *lgc-55* (amine-gated chloride channel), *ida-1* (tyrosine phosphatase-like receptor), *flp-19* (FMRF-like peptide), *unc-17* (vesicular acetylcholine transporter), *unc-40* (netrin receptor), *rab-3* (ras GTPase), *nlg-1* (neuroligin) and *kal-1* (human Kallmann syndrome homologue). We observed expression defects in all the alleles. *sem-4*, *hlh-3* and *egl-46* showed the broadest, although partially penetrant, defects affecting 9/9, 8/9 and 8/9 genes, respectively, while *unc-86* and *sem-4* showed the strongest phenotypes (exhibiting the greatest loss of reporter expression) in the genes that they regulate (4/9 in the case of *unc-86*) → **Figure 3.2.4-A** and **B**, → **Annex 3.2.2**. As with the 5-HT pathway genes, *egl-18* and *ast-1* showed the weakest phenotypes, regulating 3/7 genes analysed → **Figure 3.2.4-A** and **B**, → **Annex 3.2.2**. *flp-19* and *nlg-1* could not be analysed in *egl-18(ok290)* animals because we did not manage to achieve recombination between the reporter and the mutant allele in the same chromosome. Of note, the potassium chloride co-transporter *kcc-2* was affected in all mutant backgrounds, while the extracellular matrix gene *kal-1* was mainly unaffected → **Figure 3.2.4-A** and **B**, → **Annex 3.2.2**. The expression of all reporters in the HSN started at L4-young adult stage, when the HSN projects its axon and differentiates, except for *rab-3* (L3 stage) and *kal-1* (L1 stage). We reason that *kal-1* expression remains practically unaltered in all mutant backgrounds because its transcription might be regulated by an earlier-acting programme such as the factors we discarded (*egl-5*, *egl-44*, etc.). Quantification of gene expression in every mutant background is summarised in the heat map present in → **Figure 3.2.5**.

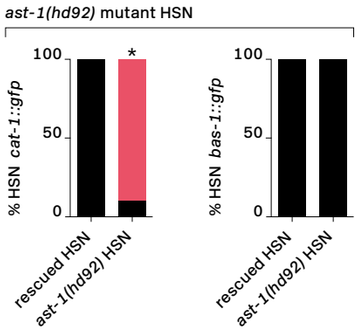
Our results demonstrate that the six TFs selected in our candidate approach are required for the expression of the 5-HT pathway genes as well as more globally for the acquisition of the HSN neuron specific identity. Although most terminal features are affected, the fact that the expression of some effector genes is maintained in every mu-

Figure 3.2.3
***ast-1* null mutant analysis in the HSN using a mosaic strategy**



A) Mosaic strategy

ast-1(hd92) null animals are L1 larval lethal due to a detachment of the pharynx. Lethality can be rescued expressing an extrachromosomal array that carries a wild type copy of the *ast-1* gene, next to *cat-1::mCherry* red marker to follow the HSN neuron *(vlcEx844, vlcEx845)*. Mosaicism is based on the somatic loss of the extrachromosomal DNA in some cells o lineages. Mutant HSNs (purple circle) in the context of an *ast-1* rescued viable worm (light pink) can be identified via loss of the red *cat-1* marker (red circles). Of note, many other cells may have lost the rescuing array (purple circles along the body) but only *cat-1* expressing cells can be detected.

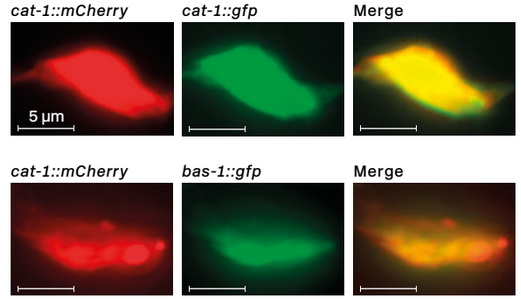


B) Quantification of *bas-1::gfp* and *cat-1::gfp* expression in mutant HSNs

For *bas-1::gfp* expression lines *vlcEx844* and *vlcEx845* were used. n= 87 mutant cells. *cat-1::gfp (otIs221)* was used as a control of the technique. 1 line was used. n=29 mutant cells. Statistical significance was calculated using the two tailed Fisher exact test, *: pV< 0.05.

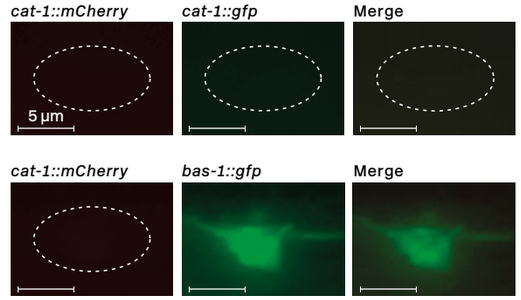
C) Mosaic rescued worm (wild type HSN)

Representative images of an *ast-1(hd92)* rescued HSN showing wild type *cat-1::gfp* and *bas-1::gfp* expression.



D) Mosaic rescued worm (mutant HSN)

Representative images of *ast-1(hd92)* mutant HSNs, in the context of an *ast-1(+)* rescued worm, lacking *cat-1::gfp* expression but showing normal *bas::gfp* expression.



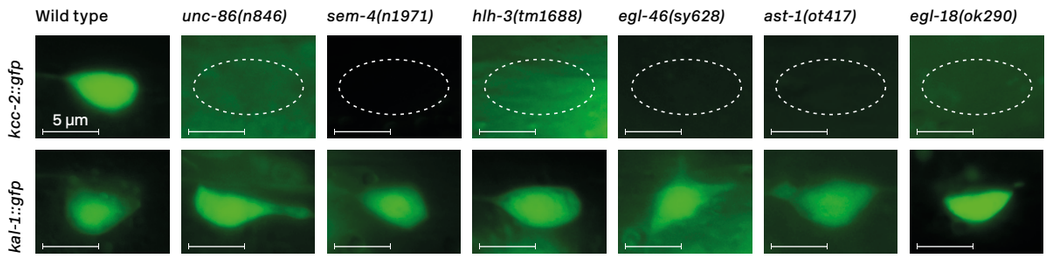
tant background indicates that HSN is generated and partially differentiates in each case, but fails to activate the expression of the complete HSN transcriptome. Furthermore, *sem-4*, *egl-46*, *egl-18* and, more severely, *hlh-3* exhibit a significant loss of the panneuronal marker *rab-3*, indicating that these mutants may not only be acting to specify the particular transcriptome of the HSN, but also more globally to regulate its neuronal features. Worth commenting is also the fact that the phenotypic profile of each mutant is slightly different from each other, which suggests that these TFs will probably not function in a cascade-like linear pathway. Importantly, although with some exceptions, the six TFs do not tend to act upon NSM or ADF serotonergic neurons. This matches our promoter bashing results presented in Chapter I, in which the independent CRMs found in the regulatory regions of the 5-HT pathway genes were predicted to be regulated by different subsets of TFs in a neuron specific manner.

Study of the possible redundant role of GATA transcription factor members in HSN serotonergic differentiation

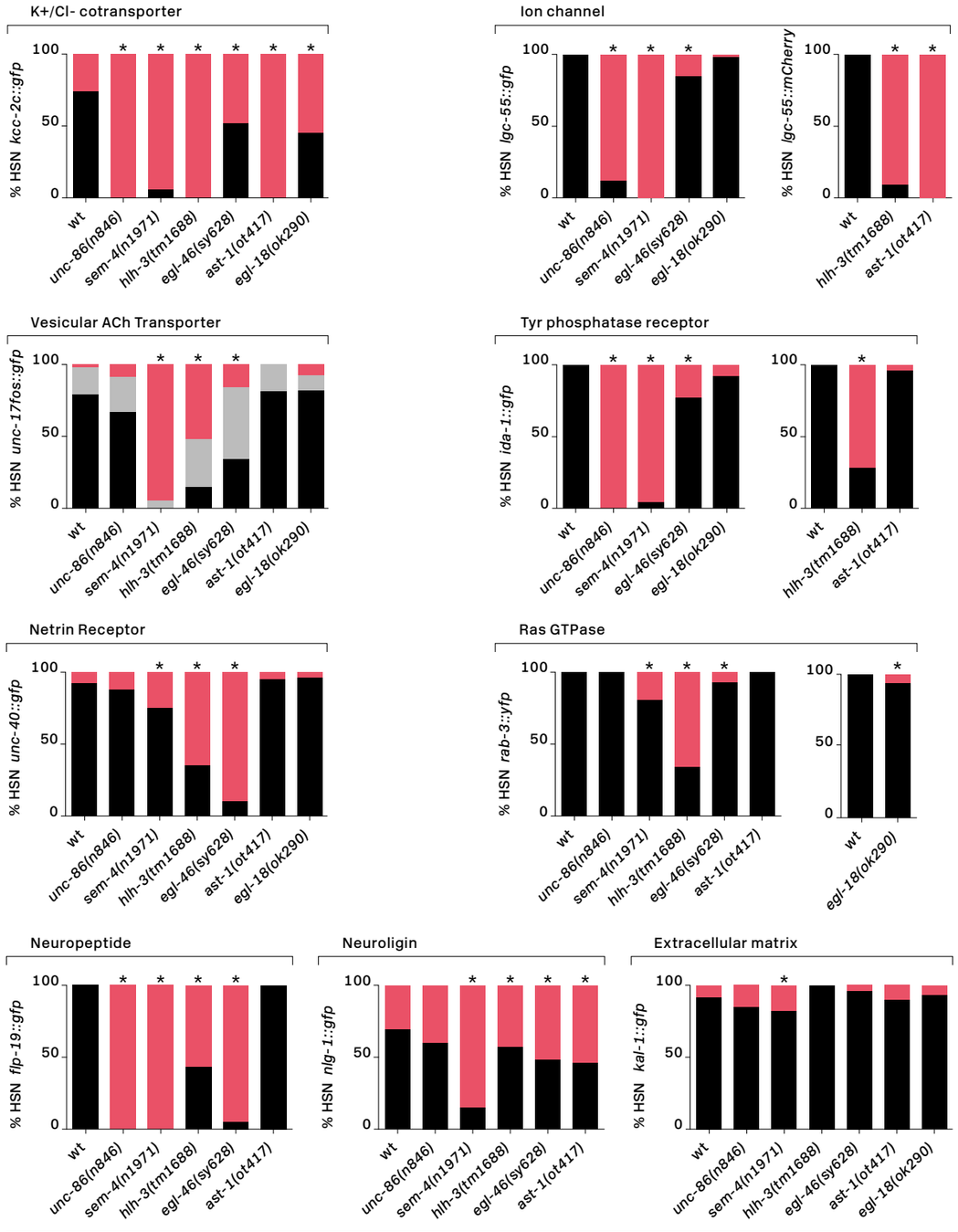
Inclusion of the GATA factor *egl-18* in our HSN analysis is particularly interesting because, to date, no GATA TF has been implicated in neuronal specification in nematodes. In mouse, however, GATA TFs have been shown to have neuronal functions. For example, GATA2 and GATA3 are required for the correct differentiation of certain serotonergic and glutamatergic neurons of the raphe nuclei and they act in a redundant manner (Haugas et al. 2016). Moreover, the same GATA pair acts redundantly as postmitotic selector genes to promote GABAergic and suppress glutamatergic identity in certain rhombencephalic regions (Lahti et al. 2016). Interestingly, more examples have been reported of GATA factors acting redundantly during the development of other tissues: in mouse, GATA1 and GATA2 redundantly regulate primitive hematopoie-

Figure 3.2.4
Analysis of non-serotonergic terminal features of the HSN neuron in mutant animals for the six candidate HSN regulators

A) Micrographs showing expression defects in the K+/Cl- cotransporter *kcc-2 (vsEx580)* and normal expression of the extracellular matrix *kal-1 (otIs33)*, in *unc-86(n846)*, *sem-4(n1971)*, *hlh-3(tm1688)*, *egl-46(sy628)*, *ast-1(ot417)* and *egl-18(ok290)* mutants.



B) Quantification of *kcc-2::gfp (vsEx580)*, *lgc-55::gfp (zfIs6)*, *lgc-55::mCherry (zfIs4)*, *flp-19::gfp (ynIs34)*, *ida-1::gfp (inIs179)*, *ida-1::gfp (inIs181, inIs182)*, *unc-17::gfp (otIs576)*, *unc-40::gfp (inIs132)*, *nlg-1::gfp (sIs13247)*, *rab-3::yfp (otIs287)*, *rab-3::gfp (otIs289)* and *kal-1 (otIs33)* expression defects in A). Black: gene expression, red: no gene expression, grey: dim expression. n>50 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test, *: pV< 0.05. See Annex 3.2.2.



tic development (Fujiwara et al. 2004), in *Xenopus*, GATA4, 5 and 6 act redundantly in the specification of the myocardium (Peterkin et al. 2005). Perhaps the most exacerbated example of redundancy in the GATA family is observed in *Arabidopsis*. There are 29 GATA factors encoding-genes, which is in contrast to the relatively low number of these TFs found in other eukaryotes: six in humans, eight in *D. melanogaster* and eleven in *C. elegans*. The explanation for the expansion of this gene family in plants remains obscure but it suggests a high functional redundancy and may explain the low success of classical genetic strategies in the elucidation of the function of GATA factors in plants (Reyes et al. 2004).

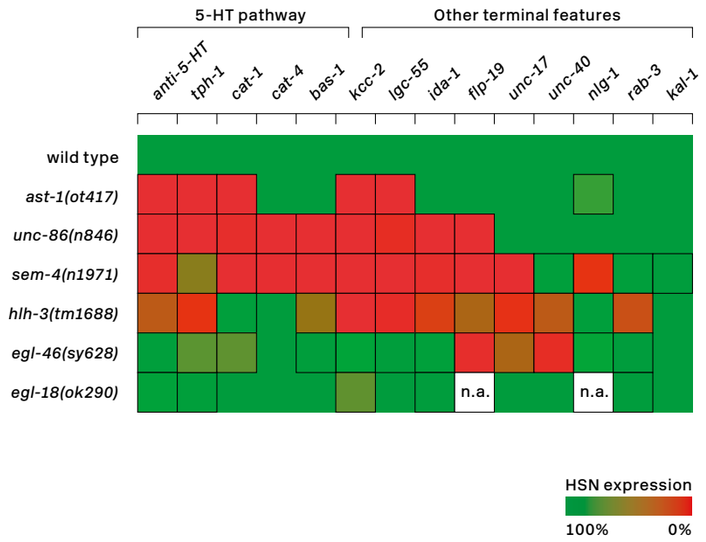
In *C. elegans* there are also examples of redundancy in the GATA family. For example, EGL-18 (a.k.a ELT-5) and its paralogue ELT-6, are redundantly required to regulate cell fates and fusion in the vulval primordium and are essential to form the vulva (Koh et al. 2002). ELT-6 and EGL-18 also function redundantly during larval seam cell development (Gorrepati et al. 2013). Although *C. elegans* gut specification was first explained through a sequential cascade of redundant GATA TFs (MED-1, MED-2; END-1, END3; ELT-2, ELT-7), non-redundant functions have been later assigned to these GATA factors (reviewed in (Maduro 2017)). What is common in all cases is that several GATA members tend to act together in the same pathway. For this reason, we hypothesised that the EGL-18 GATA factor could be working together with other GATA factors, in a redundant or in a non-redundant manner, to regulate HSN terminal differentiation. This redundancy might explain the relative subtile phenotype of *egl-18* mutants.

C. elegans GATA family consists of eleven members: ELT-1, ELT-2, ELT-3, ELT-4, EGL-18, ELT-6, ELT-7, END-1, END-3, MED-1 and MED-2. We decided to perform RNAi experiments upon all of them to see if some other GATA factors could have a role in HSN specification. As HSN differentiation markers we chose *cat-1::gfp* (*otIs221*) and *tph-1::gfp* (*zdIs13*) transcriptional reporters and we analysed F1 progeny to identify TFs that have a role during development. RNAi against *egl-18*, *end-1*, *elt-3* and *elt-6* showed a significant decrease in the number of *cat-1::gfp* positive HSN cells but only RNAi against *egl-18* and *end-1* showed a significant decrease in *tph-1::gfp* expression → **Figure 3.2.6-A,** → **Annex 3.2.3**. Moreover, *egl-18*, *elt-3* and *elt-6* show a migration defect in the HSN, which was normally posteriorly displaced (data not quantified). *egl-18* RNAi was used as a positive control, as we had already demonstrated its requirement for *tph-1* expression. However, interestingly, the effect on *cat-1::gfp* had not been observed in the null mutant what could be indicating off target effects for *egl-18* RNAi clone. As *elt-1* and *elt-2* were embryonic lethal, we performed RNAi at P0 stage with the same reporters. No defects were observed in terms of GFP expression → **Figure 3.2.6-B,** → **Annex 3.2.3**, yet *elt-1* treated animals showed *egl* phenotype and migration defects in the HSN.

Our results point to a possible role of *end-1* in the regulation of HSN that could act redundantly with *egl-18*. Therefore, we ordered the loss of function mutant *ok558* for this gene. This null allele consists of an 879 bp deletion that removes the zinc finger DNA binding domain. Not-conveniently, it also affects 246 bp of the 3' UTR of *ric-7* gene → **Figure 3.2.6-C** → **Table 2.1**. We crossed this allele with the four 5-HT pathway genes reporters (*tph-1*, *cat-1*, *cat-4* and *bas-1*) and stained for 5-HT, but only observed a very subtle 5-HT staining defect in the HSN neuron → **Figure 3.2.6-D,** → **Annex 3.2.1**. This, together with the fact that *end-1* reporter expression is only detected transiently in L3 and L4 stage and not in adult worms (data not shown) suggests that *end-1* could have a minor role in the induction of HSN serotonergic specification. Alternatively, it could act redundantly with *egl-18* in this process. As the contribution of *end-1* to HSN serotonergic differentiation seems to be small, from now we focuse on EGL-18 as the GATA member that is participating in the process.

Figure 3.2.5
Heat map summary of single mutant characterisation

Statistically significant expression defects, compared to wild type, are indicated with a black frame. n.a.: not analysed. *kcc-2:* potassium chloride co-transporter, *lgc-55:* amine-gated Cl- channel, *ida-1:* Tyr phosphatase-like receptor, *flp-19:* FMRF-like peptide, *unc-17:* vesicular acetylcholine transporter, *unc-40:* netrin receptor, *nlg-1:* neuroligin, *rab-3:* ras GTPase, *kal-1:* human Kallmann syndrome homologue. n>50 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test, *: pV< 0.05. See Annexes 3.2.1. and 3.2.2



HSN candidate regulatory factors do not affect HSN lineage

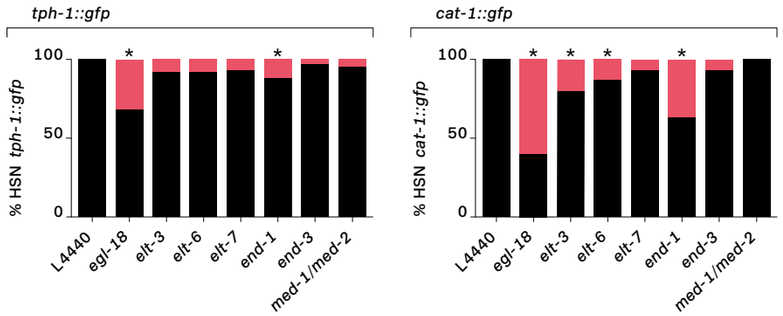
The described HSN defects could be specific of the HSN or an indirect consequence of a more anterior lineage defect. In fact, most of our TF candidates are known to affect cellular lineages in other contexts. For instance, it has been described that, in some lineages, *unc-86* mutants affect the dividing neuroblast giving rise to reiteration lineage defects where one neuronal subtype is not generated and another one appears repeatedly. This occurs, for example, in the dopaminergic deirid and postdeirid neurons ADE and PDE (Chalfie et al. 1981). In addition, *sem-4* is known to affect the M lineage in a way that cells that normally become sex myoblasts are generated but fail to exhibit the appropriate characteristics of sex myoblasts and also fail to undergo cell division. Also in *sem-4* mutants, the cells that normally become coelomocytes are generated but undergo an extra round of cell division (Basson & Horvitz 1996). Importantly, bHLH TFs are commonly referred to as proneural factors, meaning that they are usually both necessary and sufficient for the specification of neural precursors or neural lineages. In *Drosophila*, expression of Ac/Sc or Atonal genes (bHLH) within uncommitted ectodermal cells results in competence to adopt a neural cell fate (Bertrand et al. 2002). In addition, the lineages of the Q neuroblasts (the precursors of AVM and PVM neurons) have extra terminal divisions in *egl-46* mutants (Desai & Horvitz 1989). Finally, it has been reported that *egl-18* expression in selected embryonic lineages and larval seam cells is responsible for normal seam cell development and viability (Koh & Rothman 2001). Contrary, *ast-1* has never been reported to act early in cell lin-

eages. Instead, it is known to act late in dopaminergic and pharyngeal specification (Schmid et al. 2006; Flames & Hobert 2009). However, nothing is known about the action of these six TFs in the HSN lineage (AB pl/rapppappa).

Thus, in order to study the possible impact of the different mutants in the HSN lineage, we scored the presence of the PHB neuron (AB pl/rapppappp), the sister of the HSN neuron → **Figure 3.2.7-A**. The PHB is a phasmid neuron that can uptake lipophilic dyes from the environment and, thus can be detected using fluorescein isothiocyanate (FITC), DiI, DiO, or DiD (Collet et al. 1998; Hedgecock et al. 1985). These dyes diffuse and label all the plasma membrane of the neuron, from cilia to soma. In the tail of the worm, two bilateral neurons uptake this dye, the PHB and the PHA → **Figure 3.2.7-C**. Thus, we used DiI to visualise PHB neurons in the different mutant backgrounds. Although DiI staining is variable, this technique still allowed us to determine that the PHB neuron appears in the different mutant backgrounds in similar numbers as in the wild type N2 strain, except for *egl-18*, were we saw a significant loss of staining (87±4% in the wild type, vs 61±6% pV= 0.039 in the mutant) → **Figure 3.2.7-B** and → **3.2.7-D**, → **Annex 3.2.4**. Hence, *ast-1*, *unc-86*, *sem-4*, *hlh-3* and *egl-46* mutant defects observed in the HSN seem to be cell specific, while *egl-18* could have a dual role in HSN and PHB specification.

Figure 3.2.6
Analysis of the GATA transcription factor family as possible regulator of the HSN serotonergic fate

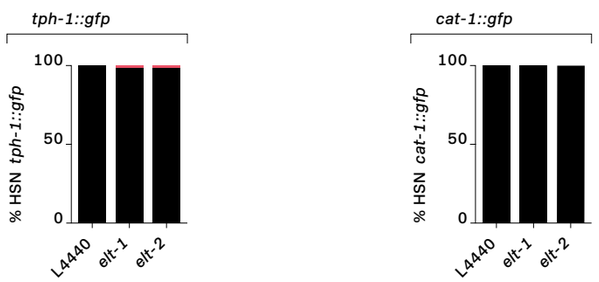
F1 scoring



A) Loss of function (RNAi) experiments against eight members of the GATA family

Quantification of *tph-1::gfp (zdIs13)* and *cat-1::gfp (otIs221)* in the HSN of adult worms. F1 scoring. The RNAi clone against *med-1* also has *med-2* as a predicted common target. n=30 worms per clone. Statistical significance was calculated using the two tailed Fisher exact test, *: pV< 0.05.

P0 scoring

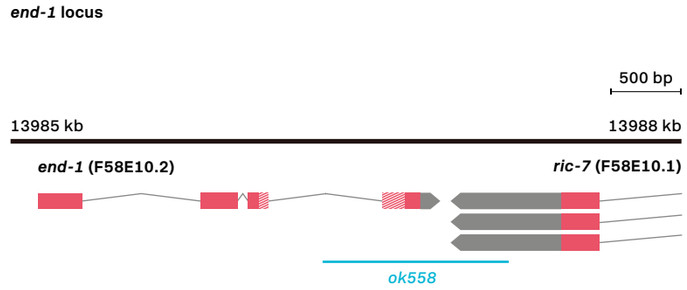


B) Loss of function (RNAi) experiments against *elt-1* and *elt-2* GATA members that showed RNAi lethality with F1 scoring

P0 scoring using the same *tph-1* and *cat-1* reporters. See Annex 3.2.3. RNAi experiments were performed by Ángela Jimeno.

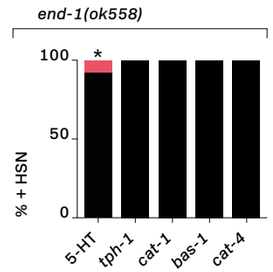
C) Genomic locus of *end-1*

Thick black line symbolises the genomic region, red boxes symbolise exons, grey lines symbolise introns and grey boxes symbolise untranslated RNA of the gene. DNA binding domain is indicated with red/white stripes. Alternative gene isoforms are included. The *ok558* deletion allele (blue) used in this work affects the 3' UTR of the *ric-7* neighbouring gene (Maduro et al. 2005).



D) Quantification of 5-HT staining and 5-HT pathway gene expression in the HSN of *end-1* null mutants

n>50 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test, *: pV< 0.05. See Annex 3.2.1.



The six transcription factor candidates act directly on their target genes to regulate their expression in the HSN neuron

We next aimed to assess if the TFs regulating HSN terminal differentiation act directly on the genes they regulate. To this end, we performed a comprehensive *in vivo cis*-regulatory analysis, using the CRMs previously determined in Chapter I.

We studied in depth the three 5-HT pathway genes that showed stronger phenotypes in the mutant analysis: *tph-1* (TPH2), *cat-1* (VMAT) and *bas-1* (AADC). We wanted to assess the existence of BS matches for the six TFs in the minimal CRMs that drive expression in the HSN. To this purpose we looked for the consensus BSs of each family according to several sources (published references, TF Data Bases (TFe) and JASPAR). The specific searched sequences were: CGGAA/TA/G (for the ETS BS), C/TG/TCATNA/T/CA/T (for the POU BS), TTGTC/GT (For SPALT BS), C/GCAGAA (for bHLH consensus), G/TNNA/TGC/GGG (for INSM BS) and A/G/TGATAA/G/T (for GATA BS). In the cases where several hits were found for a specific TFBS we prioritised mutations of phylogenetically conserved sites (present in different *Caenorhabditis* species) → **Figure 3.2.11**. If a phenotype was not detected upon mutation or if none of the sites were phyloge-

netically conserved, then we combined mutations in several sites.

The HSN minimal CRM for *tph-1*, *tph-1prom2*, is 377bp long and contains predicted BSs for the six TF members → **Figure 3.2.8-A**. This enhancer is highly expressed in the HSN neuron in 4 out of 6 lines tested. We established three levels of GFP expression according to the following criteria: if the mutated construct shows 100-60% expression of mean wild type construct values would be represented with a '+' sign; expression values 60-20% lower than mean wild type expression values would be considered 'partial loss' of expression in the HSN and would be represented with a '+/−' sign; values less than 20% of mean wild type values would be considered 'total loss' of expression in the cell and would be represented with a '−' sign.

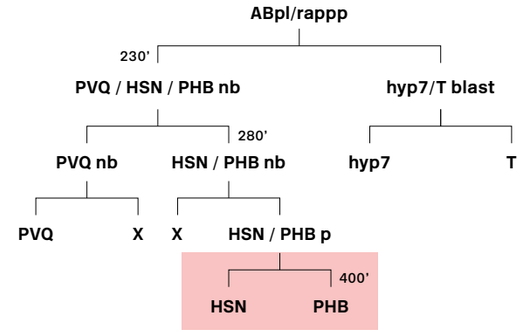
Starting with the ETS TF family, we found three putative ETS sites that matched the consensus. We decided to check if any of them could be conserved in six additional *Caenorhabditis* species (*brenneri*, *briggssae*, *japonica*, *remanei*, *sp. 5 ju800* and *tropicalis*) using Genome Browser database. Two of them were conserved in 6/6 species → **Figure 3.2.11**. We decided to mutate a conserved ETS BS first, through site directed mutagenesis, generating the *tph-1prom14* construct (*ETS MUT* in → **Figure 3.2.8-A**. We found GFP expression was specifically lost from HSN in 3/3 lines analysed. Searching for POU family BSs we found five matches, but only one of them conserved in 6/6 species that, importantly, coincided with a previously characterised functional BS (Sze et al. 2002) → **Figure 3.2.11**. We mutated this site (*tph-1prom26*, *POU MUT)* and, as expected, we saw a total loss of GFP in the cell in 2/2 lines → **Figure 3.2.8-A**. Regarding the SPALT family, only one BS was found that was conserved in 5/6 species considered → **Figure 3.2.11**. After truncating the motif (*tph-1prom31*, *SPALT MUT*) expression was partially lost in the HSN in 4/5 lines analysed → **Figure 3.2.8-A**. Only one HLH and one INSM motif were found in the *tph-1* CRM and they were not conserved in any other *Caenorhabditis* species → **Figure 3.2.11**. However, for HLH BS, 6/6 species aligned to a CAGAA motif instead of the SCAGAA motif that we had chosen for our analysis. Both HLH and INSM BSs turned out to be functional → **Figure 3.2.8-A**. Finally, we found three highly conserved GATA sites (6/6 species), one independent and two overlapping → **Figure 3.2.11**. Nonetheless, neither single (*tph-1prom54*, *tph-1prom55*) nor combinatorial (*tph-1prom60*) mutation affected *tph-1* expression → **Figure 3.2.8-A**. In summary, *in vivo* reporter analyses revealed that all except GATA BSs are required for proper *tph-1* expression in HSN. Paradoxically, *egl-18* (GATA) mutants show defects in *tph-1prom2* expression → **Figure 3.2.2,** → **Figure 3.2.8-B**. Therefore, EGL-18 may act upstream of another TF to regulate *tph-1prom2* expression. Alternatively, EGL-18 may be recruited the *tph-1* promoter even in the absence of functional GATA sites, perhaps through PPIs with other TFs of the regulatory code. Similar BS-independent recruitment of TFs has been previously reported for the LIM homeodomain TF *mec-3* (Xue et al. 1993).

The HSN minimal CRM for *cat-1*, *cat-1prom14*, is a bit larger than the one for *tph-1*, spanning 523 bp and it also contains predicted BSs for the six TF members → **Figure 3.2.9-A**. We found only one consensus BS for ETS that was conserved in 6/6 species considered → **Figure 3.2.11**, so we mutated that single site (*cat-1prom63*) and assessed the resulting expression in the HSN. 3/3 lines lost GFP expression specifically in the HSN, indicating that it is a functional site *in vivo* → **Figure 3.2.9-A**. With regard to the POU family, there were three predicted sites. We mutated the only conserved site, generating the *cat-1prom61* reporter, and saw a clear expression defect in HSN → **Figure 3.2.11,** → **Figure 3.2.9-A**. Similarly, only one predicted SPALT site was found which was con-

Figure 3.2.7
DiI staining analysis

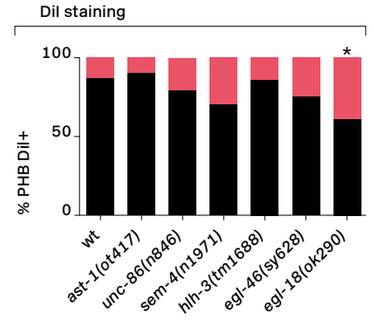
A) Partial lineage of the HSN

The HSN neuron is born at approximately 400 min post-fertilisation, next to its sister the PHB phasmid neuron. nb: neuroblast, p: precursor, X: death event.

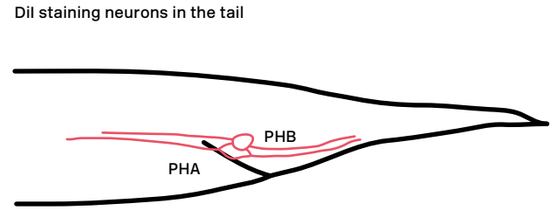


B) Quantification of DiI staining defects in the PHB neuron of adult worms

n>50 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test, *: pV< 0.05. See Annex 3.2.4.



C) Schematic representation of the tail of a worm with the two DiI filling neurons in red, PHA (anterior) and the PHB (posterior)



D) Representative micrographs of DiI staining in the PHB neuron in the different mutant backgrounds

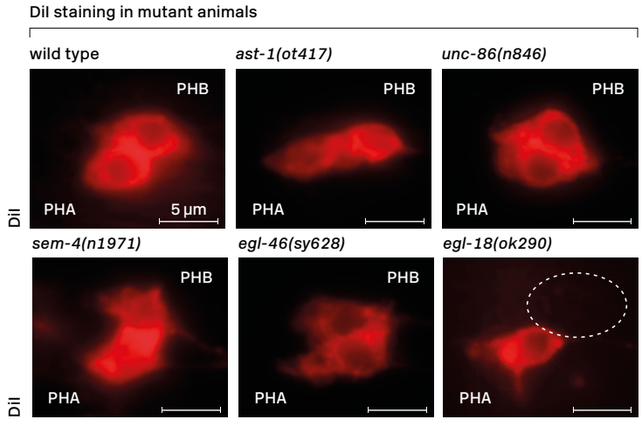
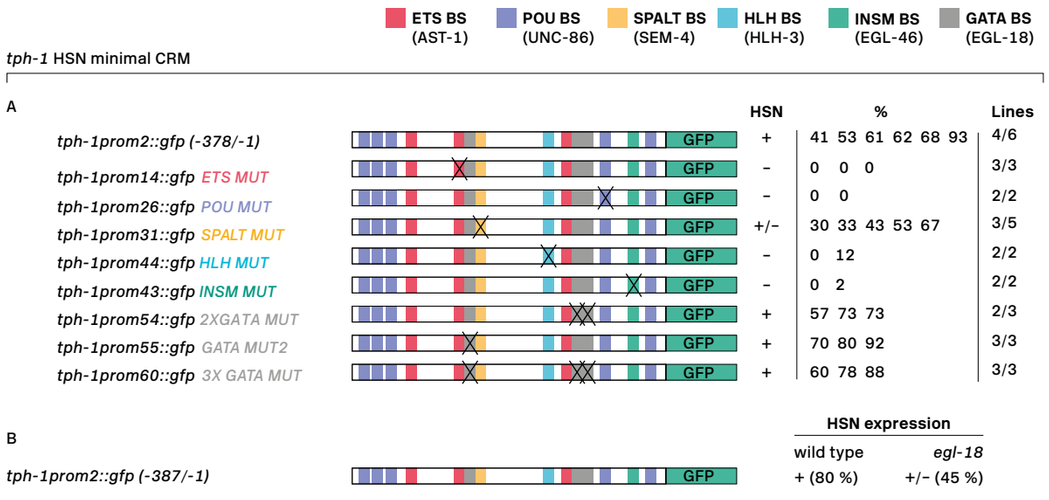


Figure 3.2.8
***tph-1* minimal HSN CRM mutational analysis**

A) Black crosses represent point mutations to disrupt the corresponding BS. TFBSs belonging to different families are represented with different coloured boxes. +: 100 to 60% of mean wildt type construct values, +/−: expression values 60-20% lower than mean wild type expression values; −: values are less than 20% of mean wildt type values. n>30 animals per line. See Annex 3.2.5 for mutated sequences.

B) *tph-1prom2::gfp* expression is partially affected in *egl-18(ok290)* mutants. n>50 each genotype. These experiments were performed by Dr. Miren Maicas.



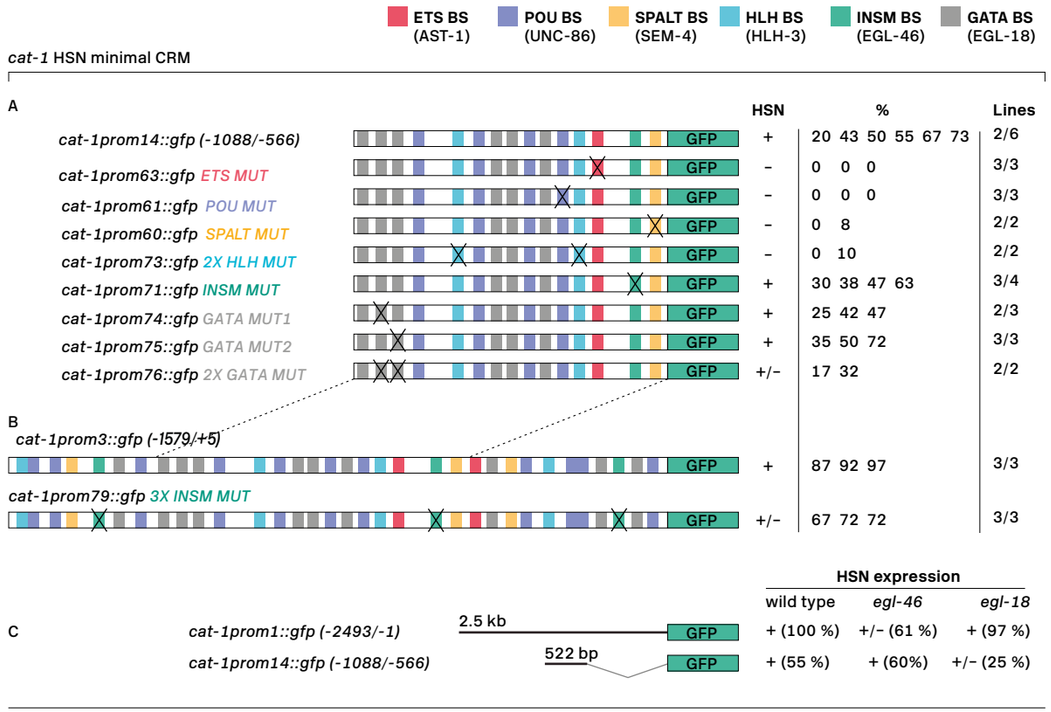
served in 6/6 species → **Figure 3.2.11**. Our point mutation analysis reveals that the site is functional (*cat-1prom60*) → **Figure 3.2.9-A**. Regarding HLH family, we found two putative BSs that, when mutated (*cat-1prom73*), exhibited a total loss of GFP in the HSN → **Figure 3.2.9-A, → Figure 3.2.11** and also in the ADF neuron (data not shown). Regarding the INSM family, as with *tph-1*, only one predicted site was found that was not conserved in any other *Caenorhabditis* species → **Figure 3.2.11**. After point mutation analysis (*cat-1prom71*) no defect was observed in the cell → **Figure 3.2.9-A**, indicating that *cat-1* CRM probably does not contain a functional BS for EGL-46. In agreement with this lack of functional INSM BSs in *cat-1prom14* we found that the expression of this reporter is not affected in *egl-46* mutants → **Figure 3.2.9-C**. However, we noticed that the penetrance of HSN expression for *cat-1prom14* is much lower than the full-length reporter *cat-1prom1* (55% compared to 100% expression respectively, → **Figure 3.2.9-C**), which indicates that additional information exists outside the minimal CRM to promote robust HSN expression. Interestingly, *cat-1prom1* expression is affected in *egl-46* mutants, suggesting that EGL-46 dependent *cis*-regulatory elements must exist and will be found outside *cat-1prom14* → **Figure 3.2.9-C**. In fact, we found three predicted INSM BSs in the larger CRM *cat-1prom3* (1584 bp) that is highly expressed in the HSN → **Figure 3.2.9-B**. After mutating all of the BSs, we detected a partial loss of expression in the HSN → **Figure 3.2.9-B**. These results suggest that, although partial *cat-1* expression can be achieved without EGL-46, this TF is required for robust expression in the context of the full *cat-1* promoter. To finalise *cat-1* analysis, we searched for GATA putative BSs. Six sites were identified but none of them was conserved → **Figure 3.2.11**. We individually mutated two GATA sites (*cat-1prom74* and *cat-1prom75*) without seeing an effect, but the double mutation (*cat-1prom76*) showed partial loss of expression in the HSN, indicating that GATA factors can bind *in vivo*, at least, to these sites in order to regulate *cat-1* expression → **Figure 3.2.9-A**. This surprisingly contrasts with the lack of expression defects of a full-length *cat-1* reporter in *egl-18* (GATA) mutants → **Figure 3.2.2**. We analysed minimal *cat-1* CRM (*cat-1prom14*) expression in *egl-18* mutants and found its expression is affected in this mutant background → **Figure 3.2.9-C**. These results reveal that *egl-18* has a direct role in regulating *cat-1* expression, but also that *egl-18* loss can be compensated in the context of big regulatory regions but not in the context of the minimal CRM *cat-1prom14*. As we will explain next, electrophoretic mobility shift assay (EMSA) experiments, con-

Figure 3.2.9
***cat-1* minimal HSN CRM mutational analysis**

A) *cat-1* minimal HSN CRM *(cat-1prom14)* mutational analysis. n>30 animals per line. These experiments were performed by Dr. Miren Maicas.

B) *cat-1* HSN CRM *(cat-1prom3)* mutational analysis. The shorter version of the CRM *(cat-1prom14)* does not contain functional INSM BSs, but the longer version *(cat-1prom3)* does.

C) *cat-1prom14::gfp* expression is unaffected in *egl-46(sy628)* mutants, which coincides with the lack of phenotype when INSM BS are mutated in this construct. *cat-1prom14::gfp* contains functional GATA sites and, as expected, its expression is affected in *egl-18(ok290)* mutants. Expression of a longer reporter *(cat-1prom1::gfp)* is independent of *egl-18*, revealing compensatory effects in the context of big regulatory sequences. n>50 animals per genotype. See Annex 3.2.5 for mutated sequences.



firmed direct binding of EGL-18 to *cat-1* CRM.

The last HSN minimal CRM that we studied was *bas-1* (AADC) (*bas-1prom18*) that is 117 bp long and shows no alignment at all with the rest of *Caenorhabditis* species → **Figure 3.2.11**. It contains predicted BSs for ETS, POU, GATA and SPALT TFs, but lacks any predicted INSM or HLH BSs → **Figure 3.2.10-A**. Interestingly, the penetrance of expression of this construct is lower than the full reporter: only 2/4 lines show expression levels above 60%, reinforcing the idea that although all TFBSs might not be strictly required, they might be required for robust expression. First, we focused on the minimal CRM (*bas-1prom18*) and our reporter analyses revealed that, indeed, ETS, POU and SPALT but not GATA BSs are required for reporter expression in HSN → **Figure 3.2.10-A**. Similar to *cat-1*, *bas-1* functional BSs in the context of small CRMs do not always correlate with the phenotypes observed for the full-length reporter in the mutant background. For example, we found functional ETS BSs in *bas-1prom18* while expression of the full-length *bas-1* reporter is unaffected in *ast-1(ot417)* hypomorph and *ast-1(hd92)* null mutants → **Figure 3.2.10-A, → Figure 3.2.2-A, → Figure 3.2.3-B**. We analysed *bas-1prom18* expression in *ast-1(ot417)* mutants and found a small but significant reduction in the percentage of GFP positive HSNs (66% in mutants vs 83% in N2 animals → **Figure 3.2.10-C**). Moreover, EMSA experiments indicate direct binding of AST-1 to *bas-1* CRM. Altogether, these results suggest that AST-1 can bind and activate transcription from the *bas-1* minimal CRM. This resembles the just described relationship between EGL-18 and *cat-1*. In both cases, however, other factors can compensate for their loss by activating transcription from regulatory sequences outside the minimal CRMs. This genetic redundancy for some members of the HSN TF collective at specific 5-HT pathway genes possibly acts as a mechanism to ensure that differentiation is robust. Although HLH-3 (HLH) and EGL-46 (INSM) are required for full-length *bas-1* expression → **Figure 3.2.2**, no functional HLH or INSM BSs were found in the minimal *bas-1* CRM (*bas-1prom18*). Similar to the minimal *cat-1* CRM, GFP expression of the short *bas-1prom18* is partially penetrant, while a longer construct (*bas-1prom13*) is more robustly expressed → **Figure 3.2.10-B**. We checked for presence of predicted HLH, INSM and other GATA BSs in *bas-1prom13*. We found two overlapping HLH sites that, when mutated simultaneously (*bas-1prom77*), slightly reduced the expression in the HSN. The single INSM site found was potentially functional, as mutations (*bas-1prom76*) clearly disrupt GFP expression in the HSN. This suggests a direct role for *hlh-3* and *egl-46* in robust *bas-1* expression. Finally, we mutated several GATA BSs, alone or in combination (*bas-1prom83*, *bas-1prom84*, *bas-1prom86*), but did not find any defect in the HSN → **Figure 3.2.10-B**. The fact that we did not find functional GATA BSs in *bas-1* CRMs, next to the observation that *egl-18* (GATA) mutants do not show *bas-1* expression defects, could suggest that GATA factors are dispensable for the regulation of this gene. However, as we had already observed genetic redundancy in other CRMs, we wondered whether this could also be the case for *bas-1* reg-ulation. To address this question, we analysed the expression of a *bas-1* minimal CRM carrying GATA BS mutations (*bas-1prom78*) in the *ast-1(ot417)* sensitised genetic background. Interestingly, while GATA BS mutations have no significant effects in wild type worms, we found a complete loss of expression of this construct in *ast-1(ot417)* mutants → **Figure 3.2.10-D**. These results revealed both a direct role for GATA factors in *bas-1* expression and redundancy and or compensatory effects between *egl-18* and *ast-1*.

Altogether, our exhaustive *in vivo cis*-regulatory analyses revealed that all six TFs (AST-1, UNC-86, SEM-4, HLH-3, EGL-46 and EGL-18) act directly on

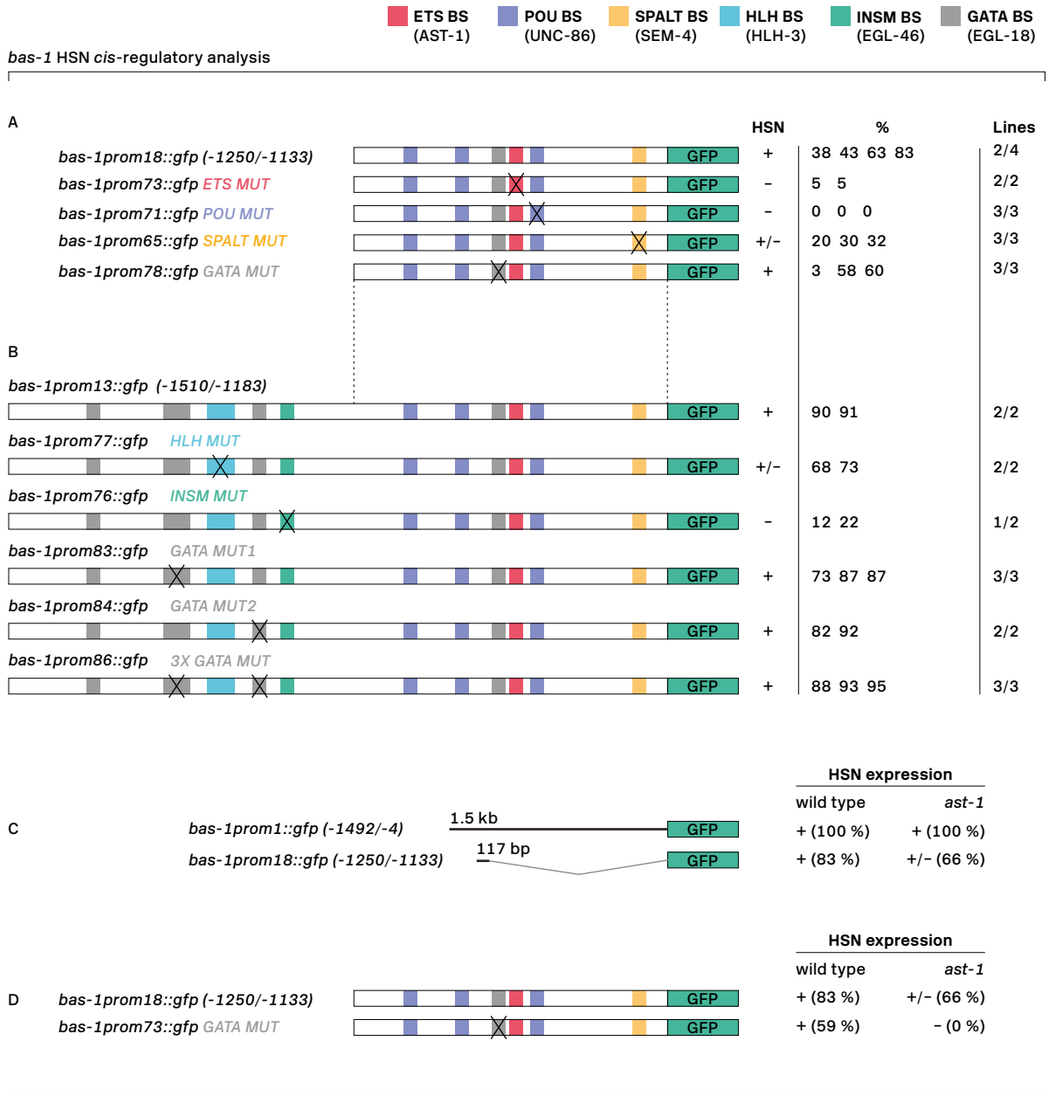
Figure 3.2.10
***bas-1* minimal HSN CRM mutational analysis**

A) *bas-1* minimal HSN CRM *(bas-1prom18)* mutational analysis. n>30 animals per line. See Annex 3.2.5 for mutated sequences. These experiments were performed by Dr. Miren Maicas.

B) A longer *bas-1* construct *(bas-1prom13)* is more robustly expressed in HSN (90% expression compared to 48% expression of *bas-1prom18)*. This construct contains functional HLH and INSM BSs, unlike the shorter version *bas-prom18*.

C) *bas-1prom18* contains functional ETS sites and its expression is affected in *ast-1(ot417)* mutants. Expression of a longer reporter *(bas-1::prom1)* is independent of *ast-1*, revealing compensatory effects in the context of big regulatory sequences. n>50 each genotype.

D) GATA BS point mutation does not significantly affect *bas-1::gfp* expression in the wild type background (in any CRM context). However, it synergises with *ast-1* mutant background leading to a complete loss of GFP expression. These results unravel a direct role for GATA sites in *bas-1* gene expression and synergy between *egl-18* and *ast-1*.



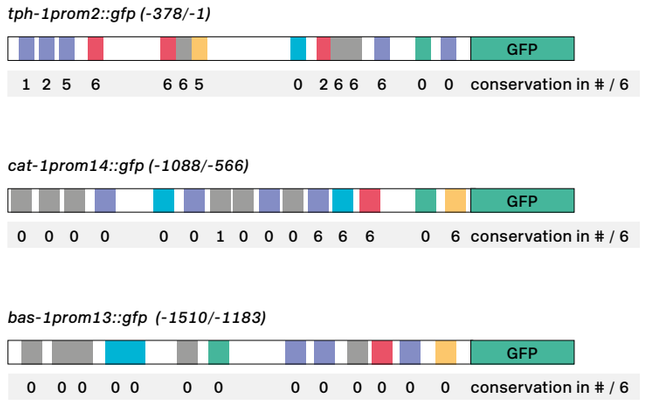


Figure 3.2.11
Conservation of the putative transcription factor binding sites of the six candidate regulators of the HSN in *tph-1*, *cat-1* and *bas-1* CRMs

Conservation between six *Caenorhabditis* species *(C. brenneri, C. brigssae, C. japonica, C. remanei, C. sp. 5 ju800, C. tropicalis)* was assessed generating multiple alignments using the Multiz and PhyloP tools from UCSC Genome Browser. TFBSs belonging to different families are represented with different coloured boxes. Numbers below TFBS indicate the number of species in which is conserved.

the regulatory regions of the 5-HT pathway genes. Further supporting these results, in the next section we will describe how UNC-86, AST-1 and EGL-18 bind to serotonergic CRMs *in vitro*. Our extensive analysis provides us with additional information about how the individual roles of each TF depend on specific DNA contexts. For instance, we detected several examples of genetic redundancy that provide robustness of expression to the system and that can be unravelled in the context of smaller CRMs or mutant backgrounds. Notably, redundancy is specific to the CRM architecture as two TFs can act redundantly in one CRM but not in others. In addition, we have observed clear examples of genetic enhancement between TFs, suggesting that they act as a regulatory code (HSN regulatory code). Moreover, each CRM has a different disposition of TFBS arrangements supporting a flexible function of these TFs in the HSN. Finally, we also found that short HSN CRMs that lack TFBSs for some HSN TF collective members can drive partially penetrant HSN expression, while longer CRMs with functional BSs for all HSN TF collective members drive more robust expression.

UNC-86, AST-1 and EGL-18 directly bind to the regulatory regions of the serotonin pathway genes *in vitro*

In order to validate the previously inferred direct binding of the HSN regulatory code to the 5-HT pathway genes, Electrophoretic Mobility Shift Assays (EMSA) were performed. DNA probes for *tph-1*, *cat-1* and *bas-1* genes were labelled with the radioactive isotope phosphorous-32 (32^P) and incubated with the purified proteins of some members of the HSN code. One or two probes targeting previously identified functional BSs for every member in the regulatory regions of the genes were designed.
EMSA experiments reveal that UNC-86 is able to bind to the *tph-1*, *cat-1* and *bas-1* probes *in vitro* → **Figure 3.2.12-A**. We had previously shown UNC-86 binding to *tph-1* and *bas-1* (Zhang et al. 2014) but not to *cat-1*. The binding is also dose dependent because we observe a stronger band in the gel as we add increasing concentrations of the probe. To test for POU BS specificity, the EMSA experiments were repeated using probes with mutations in the functional POU sites determined from our *cis*-regulatory analysis. UNC-86 binding is absent in these conditions → **Figure 3.2.12-A**, indicating that UNC-86 specifically and directly binds to the DNA at this specific location.
In addition, AST-1 is able to bind *in vitro* to *cat-1* and *bas-1* in a specific manner but we did not observe binding to *tph-1* regulatory regions → **Figure 3.2.12-B**. In the same way as with UNC-86, binding is dose dependent and deletion of the ETS site that is required for reporter gene expression *in vivo* resulted in the loss of AST-1 binding *in vitro* → **Figure 3.2.12-B**.
We also detected *in vitro* binding of EGL-18 to the *cat-1* probe, but no interaction was seen with two different *tph-1* probes (*tph-1.1* and *tph-1.2*; → **Figure 3.2.12-C**). EGL-18 binding to *cat-1* was specific as we only observed a supershift in the EGL-18 band when EGL-18-His tagged protein was incubated with the anti-6xhistag antibody and not when it was incubated with an anti-GFP antibody → **Figure 3.2.12-C**. As a positive control the 3' enhancer region of the Wilms Tumour 1 gene (*WT1*) was used. GATA2 (closest human orthologue to EGL-18) has been shown to bind to this region in several solid tumour cell lines and to be critical in the expression of *WT1* (Furuhata et al. 2009). Moreover, the *cat-1* DNA probe was modified in order to truncate both GATA BSs that had been shown to reduce expression in the HSN *in vivo* when simultaneously deleted → **Figure 3.2.9-A**. As with UNC-86 and AST-1, the band was lost, indicating that EGL-18 binds to *cat-1* regulatory regions *in vitro*, through GATA BSs.

Unfortunately, no binding corresponding to SEM-4, HLH-3 or EGL-46 was detected under these experimental conditions. Our results show that, at least, AST-1, UNC-86 and EGL-18 are able to bind *in vitro* to the *in vivo* determined CRMs. A summary of positive EMSA assays is shown in → **Figure 3.2.13**.

The HSN regulatory code is expressed in the HSN

So far, we have described that our six candidate TFs are required for proper HSN terminal differentiation. In order to confirm the cell autonomous actions of the proteins, and to start studying their inter-relationships, we analysed their expression in the HSN, as it had only been partially assessed. All strains and the corresponding genotypes used are listed in → **Table 2.21**.
When possible, integrated fosmid reporter strains were used. Unlike transcriptional reporters, which consist on a DNA fragment of a few kilobases immediately 5' upstream of the start codon of the gene of interest, fosmids cover large genomic regions (around 40 kb) and in *C. elegans* are considered a good approach for endogenous gene expression assessment, as they usually contain all regulatory information (Tursun et al. 2009). Fortunately, in *C. elegans* a fosmid library that covers 80% of the genome and 90% of the worm genes is available (http://www.sourcebioscience.com). As it shows 5.74 X clone coverage of the genome, one can usually find a fosmid where the gene of interest is close to the centre, with at least 2–3 additional genetic loci on either side. Engineering fluorescently labelled genes of interest in a genomic clone context allows for evaluation of the expression pattern and functionality of the tagged gene. Previous to the appearance of CRISPR-Cas9 technology, inserting tags at the target gene locus contained within these fosmids by homologous recombination (also called 'recombineering') represented the most accurate method to generate reporters that recapitulate full endogenous expression of a given gene.
Starting with UNC-86, it is well-known when and where this TF is expressed. Using rabbit antisera against UNC-86 protein, expression is detected in the embryonic Q lineage and is maintained in the adult, including in the HSN neuron. UNC-86 is first detected in the postmitotic HSN in the em-

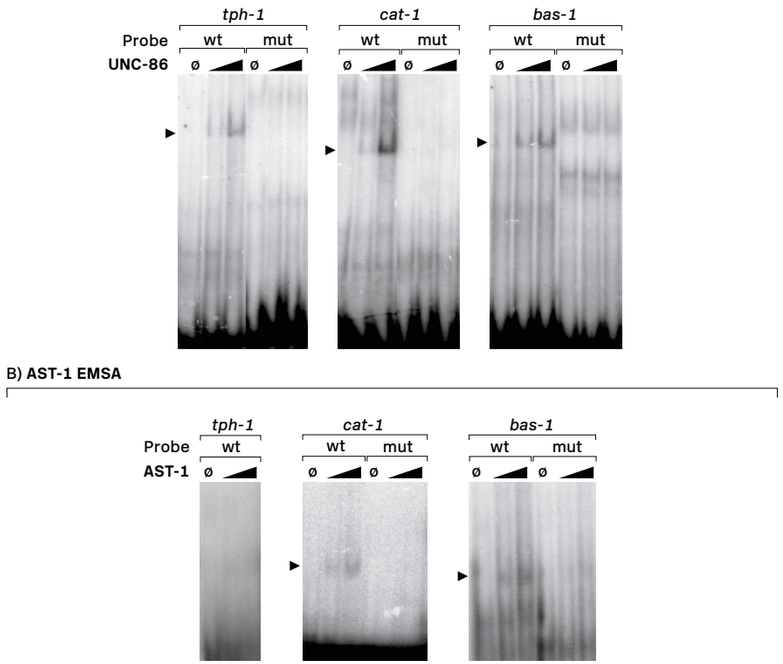


Figure 3.2.12 Electrophoretic mobility shift assays to assess direct binding of the six candidate regulators of the HSN

A) Purified UNC-86 binds *tph-1*, *cat-1* and *bas-1* CRMs in a concentration dependent manner (depicted by arrowheads). UNC-86 binding is abolished by point mutation in the POU BS.

B) Purified AST-1 binds to *cat-1* and *bas-1* CRMs in a concentration dependent manner (arrowheads). AST-1 binding is lost upon ETS BS mutation.

C) Cellular extracts from HEK293T cells overexpressing EGL-18:HIS bind to the *cat-1* CRM. Supershift of the band with HIS antibody, but not with GFP antibody, indicates that the binding involves EGL-18 protein. Moreover, point mutation of GATA site abolishes *cat-1* sequence binding by the cellular extract. These experiments were fully performed by Dr. Miren Maicas.

bryo (Finney & Ruvkun 1990). We wanted to check if we could reproduce these results using the integrated fosmid *otIs337* strain (Zhang et al. 2014), so we could next proceed to analyse its expression in different mutant backgrounds. With this configuration, the onset of expression and tissue specificity of *unc-86* gene product can be easily assessed. Expression was first detected in the HSN at embryonic comma stage (approximately 7 hours post-fertilisation), and was maintained in adult worms → **Figure 3.2.14-A** and **B**. We next crossed the *otIs337* reporter with a *tph-1* transcriptional reporter strain that carried a red fluorescent protein (*vsIs97*) and confirmed that *unc-86* is also expressed in the NSM but not in the ADF neuron, although no quantification was carried out.

Basson and Horvitz first reported *sem-4::lacZ* fusion transgene expression in the HSN (Basson & Horvitz 1996). In the present work, in order to evaluate *sem-4* expression, we first used the integrated fosmid *wgIs57* (Sarov et al. 2012). Expression was hard to detect in the HSN during embryonic stages but GFP was localised to the HSN in L1 larva and it was maintained throughout the life of the animals → **Figure 3.2.14-A**. However, as fluorescence intensity in the HSN was very faint and worms showed a relatively high level of background GFP signal, we next analysed a translational fusion reporter strain, *kuIs34*, that had been previously described to be expressed in the adult HSN (Grant et al. 2000). A smaller percentage of L1 worms showed GFP expression in the HSN but, already at L2 stage, GFP penetrance was comparable to those of the fosmid reporter and expression was also maintained during the rest of the life of the worms → **Figure 3.2.14-B**. We concluded that both reporters could be indistinctly used to study the HSN neuron. Similarly to *unc-86*, *sem-4* is expressed in many other neurons and cells: hypodermal 8, 9 and 10 cells, all rectal cells, DVC, VPCs, ventral nerve cord neurons, cells in the head and in the preanal ganglion (Grant et al. 2000; Jarriault et al. 2008). After crossing this *kuIs34* reporter with the *tph-1::DsRed* reporter strain *vsIs79*, we saw that *sem-4* is not expressed in any other serotonergic neuron (not in NSM nor in ADF).

According to some reports, *hlh-3* appears to be expressed in all neuronal precursors during embryogenesis (Krause et al. 1997). A different report describes that, using either a full-length translational reporter or transcriptional fusion reporters, this expression is maintained in most neurons of the nerve ring ganglia upon hatching and that already at larval stage 1, expression is almost undetectable except for the endodermal-like P cells (Doonan et al. 2008). Expression is maintained in the 53 resulting postmitotic motor neurons, including HSN, throughout larval development. In our work, we chose two fosmid reporter strains to assess HLH-3 expression in the HSN. One was the *otEx4140* extrachromosomal transgene (Murgan et al. 2015) and the other one was *wgIs650* (Sarov et al. 2006), both constructs corresponding to recombineered fosmids. In our hands, expression was only detected in the HSN precursor cell, the HSN/PHB neuroblast, at 5 hpf. Both reporters show comparable expression pattern and timing. → **Figure 3.2.14-A** and **B** show expression of *otEx4140*. Expression was rapidly lost as we did not manage to see *hlh-3::yfp* expression in the HSN at the time *unc-86::yfp* (*otIs337*) initiates its expression in the cell (430 min after the first cleavage, approximately 7 hpf). We mounted in parallel one cell stage embryos of the two reporter strains, *otIs337* and *otEx4140*, and 4 hours later we scored the total number and the position of fluorescent cells in the embryonic tail. We scored at different time points and up to 6 hours after mounting the embryos (7 hpf). We never saw YFP expressing cells at the same location and at the same time, so we concluded that *hlh-3* expression in HSN/PHB precursor cell must be turned off before the last postmitotic division takes place, coinciding with *unc-86* onset. This would be in agreement with

hlh-3 acting as a proneural gene during neurogenesis. We did not assess expression of *hlh-3* in NSM or ADF precursor neurons.

egl-18 reporters and EGL-18 anti-sera had been previously shown to be strongly expressed in seam cells, neurons in the head, VPCs and hyp7 cell, amongst others (Koh & Rothman 2001; Koh et al. 2002), but never before in the HSN cell. To study *egl-18* expression we used the fosmid strain *stIs11606* (Dr. Waterston laboratory). We observed fluorescent expression in the HSN in all larval stages and in adult worms → **Figure 3.2.14-A** and **B**. *egl-18* expression in embryos was broad, so we assessed expression in the embryonic HSN crossing *unc-86::yfp* reporter into the *egl-18::mCherry* background. Both reporters co-localised in the HSN postmitotic cell at 7 hpf. Using the *tph-1* reporter *zdIs13*, we determined that *egl-18* is also expressed in the NSM neuron, but not in the ADF in adult worms.

For the two other TF members, however, there was no available fosmid. To study *egl-46* expression we injected N2 worms with a transcriptional reporter that covered all the intergenic region (4,477 bp) (kindly provided by Dr. Pocock) to create the *vlcEx324* extrachromosomic reporter line. Although *DsRed* expression was rather variable from scoring to scoring, we detected expression in the HSN in all larval stages and in adult worms in 50-60% of the cases → **Figure 3.2.14-A** and **B**. We did not assess expression in the embryo as it had been already described that *egl-46* expression in the HSN starts at 1.5 fold stage (Wu et al. 2001). As with the previous TFs, *egl-46* is expressed in additional neurons in the head, ventral nerve cord and tail (Wu et al. 2001). We also determined that it is expressed in the NSM neuron, but not in the ADF neuron.

Finally, to examine *ast-1* expression we first analysed the fusion reporter line *hdIs42* that contains the entire coding sequence of *ast-1* (Schmid et al. 2006). YFP was observed in the nuclei of approximately 40 neurons in the worm, but not in the HSN neuron. As a way to test if the gene product of the *hdIs42* array was functional and rescued the *ast-1* HSN phenotype, we crossed it with *ast-1(hd92)* lethal mutants. The array was not able to rescue *ast-1* lethality, indicating that it missed relevant information for its proper expression. As we believed that *ast-1* must be expressed in the HSN, in the same way as the rest of its partners, we decided to generate CRISPR-Cas9 mediated GFP protein knock-in, tagging AST-1 protein at the C-terminus. More than ten integrated lines were recovered and all were undoubtedly expressed in the adult HSN → **Figure 3.2.14-A** and **B**. We chose one line to study the expression and temporal pattern of the gene (*ast-1(vlc19[ast-1::gfp])*). Interestingly, AST-1::GFP signal was first detected in the HSN at late L3 larval stage and was increasingly up-regulated until the adult stage, when almost a 100% of the worms showed GFP in the cell. Moreover, it was exclusively detected in neuronal nuclei, in contrast to what was previously described (Schmid et al. 2006). Apart from HSN, we scored 30-32 neurons in the head, two HSNs, two PDEs and two ALN neurons in the tail plus another four neurons in the tail. We did not observe expression in the NSM or ADF serotonergic neurons.

In summary, the six members of the HSN regulatory code are expressed in the HSN but the onset expression varies among the different TFs. *unc-86*, *sem-4*, *egl-18* and *egl-46* are expressed from early embryo or L1 stage and during the whole life of the worm, while *ast-1* specifically activates its expression in the cell in the transition between L3 and L4 stages. The five of them would be co-expressed at the larval stage 4, when HSN differentiates and starts expressing most terminal features. In contrast, *hlh-3* is only detected in the HSN at embryonic stages, prior to the onset of HSN terminal differentiation. As aforementioned, bHLH TFs have a proneural role during neurogenesis and contribute to the specification

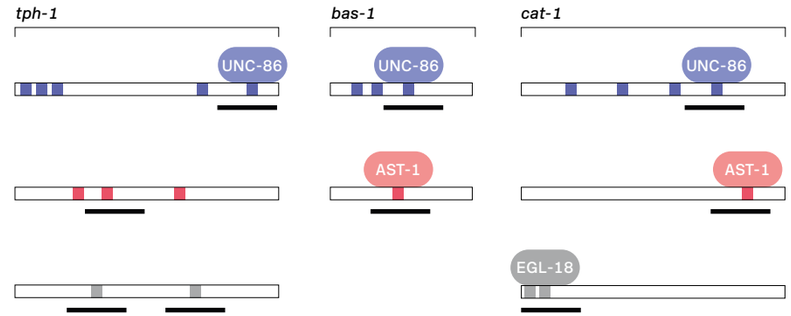


Figure 3.2.13
Summary of EMSAs: UNC-86, AST-1 and EGL-18 bind to serotonin pathway gene CRMs in electrophoretic mobility assays

UNC-86 binds to *tph-1*, *cat-1* and *bas-1*. AST-1 binds to *bas-1* and *cat-1*, but not to *tph-1*, at least with the probe and under the conditions tested. EGL-18 binds to *cat-1*, but not to either *tph-1* probe tested.

of progenitor-cell identity (Bertrand et al. 2002). This could suggest that *hlh-3* acts earlier in the developmental history of the HSN neuron. However, HSN sister neuron is not affected in *hlh-3* mutant background → **Figure 3.2.7.** There are examples in *C. elegans* where bHLH proneural genes, such as *lin-32* and *hlh-2*, have also later roles in terminal neuron differentiation (Portman & Emmons 2000).

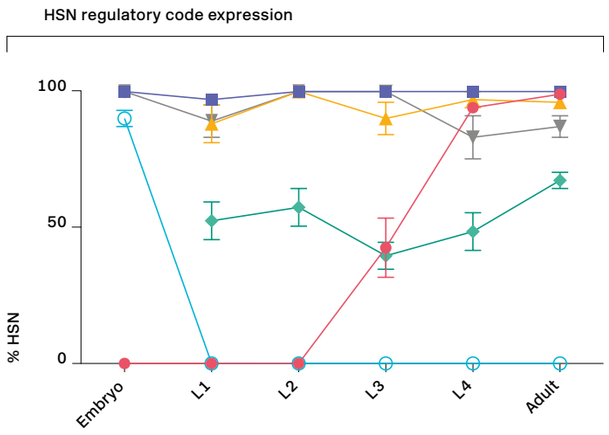
HSN terminal differentiation involves parallel pathways

To further explore how the HSN TF code regulates HSN terminal fate we aimed to assess if they show cross-regulation. We analysed TF expression of the previously analysed reporters in the different mutant backgrounds of young adult worms. We selected this age for analysis because the HSN neuron already should have become mature. For *hlh-3* reporter analysis, as it is only expressed in the embryo, expression was assessed in 5 hours old embryos. → **Figure 3.2.15** shows how individual factors affect the expression of the different TF reporters. Although statistics between wild type and mutant values were calculated using raw data, we have represented in the graphs mutant values relative to wild type values for an easier interpretation.

Moving on to the analysis, the *ast-1(ot417)* mutation does not affect the expression of any other HSN code member. The same happens with *egl-18(ok290)* and *egl-46(sy628)* mutants, with the exception of *egl-46* mutants that show a small but significant decrease in *ast-1* expression → **Figure 3.2.15-A,** → **Annex 3.2.6**. This implies that AST-1, EGL-18 and EGL-46 TFs are not required for the expression of the rest of the code. Note that for *egl-46* expression analysis in *egl-18(ok290)* mutants, *kuIs34* strain could not be used due to an incompatibility in chromosome location. Instead, the *kuIs35* transgene was used, which corresponds to an independent integration event of the same construct as *kuIs34* (Grant et al. 2000). Although we did not characterise *kuIs35* in the same depth as *kuIs34*, both transgenes seemed to show identical expression patterns.

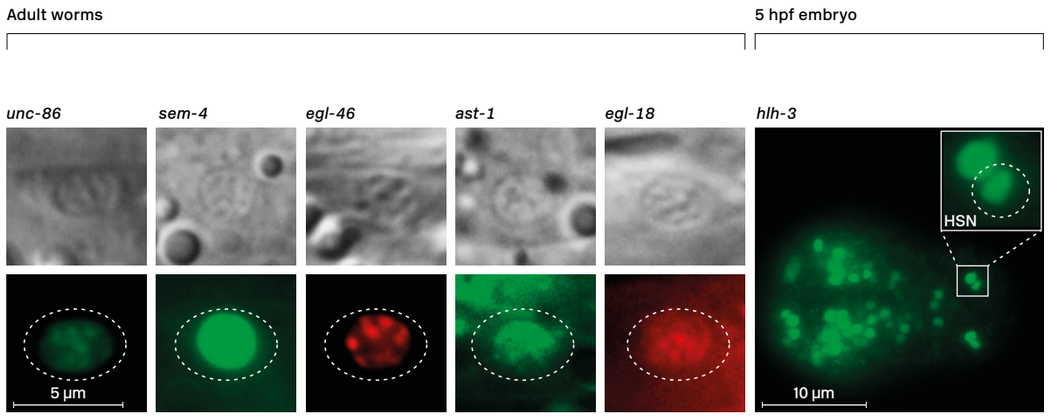
In contrast to the previous cases, UNC-86 seems to be required for the expression of several members of the HSN TF code. We observed a severe phenotype in *unc-86(n846)* mutants upon *ast-1* expression, *sem-4* expression and *egl-46* expression. *sem-4(n1971)* and *hlh-3(tm1688)* mutants, in turn, also seem to affect *ast-1* expression. *hlh-3(tm1688)* mutants show an additional mild phenotype over *egl-46* expression → **Figure 3.2.15-A**. Our results

Figure 3.2.14
Expression of the HSN regulatory code in the HSN neuron



A) Developmental expression of the six TFs in the HSN neuron, using fluorescent reporters. Expression of all, except *hlh-3*, was maintained in adult animals. Strains used: otIs337*(unc-86::yfp), wgIs75(sem-4::gfp)* and *stIs11606(egl-18::mCherry)* fosmid reporters; *vlcEx324(egl-46::DsRed)* transcriptional reporter; *ast-1[vlc19(ast-1::gfp)]* CRISPR-Cas9-mediated *gfp* knock-in. Error bars indicate standard error of the proportion (SEP). n>50 worms at every developmental stage. See Annex 3.2.6.

- *unc-86::yfp*
- *sem-4::gfp*
- *hlh-3::yfp*
- *egl-46::DsRed*
- *ast-1::gfp*
- *egl-18::mCherry*



B) Micrographs showing expression of the HSN regulatory code in adult and embryonic HSN neurons. Pictures correspond to the same reporters as in A), except for *sem-4*, where the brighter *kuIs34* transcriptional reporter is shown.

DIC images show the HSN nucleus and soma (top panel), while fluorescence images show expression in the cell (bottom panel). *hlh-3* was only detected in the HSN precursor (PHB/HSN), prior to the last mitotic division. hpf: hours post-fertilisation.

seem to indicate that UNC-86, SEM-4 and HLH-3 TFs are required for proper *ast-1* expression in the HSN. As UNC-86 regulates *sem-4* expression, loss of *ast-1* expression in *unc-86* mutants could be indirectly due to *sem-4* loss. Alternatively, *ast-1* expression may require both SEM-4 and UNC-86.

To discard the possibility that no phenotype was observed in *ast-1(ot417)* animals due to the hypomorphic nature of the mutation, we analysed the expression of the same reporters in *ast-1(hd92)* null mutants at larval stage 1. Again, no significant effect was observed, supporting the idea that AST-1 is not required for the expression of the rest of the HSN regulatory code → **Figure 3.2.15-B**.

Regarding *hlh-3* expression analysis at embryonic stages, only loss of EGL-18 protein seems to slightly affect its expression → **Figure 3.2.15-C**. We have previously shown that *egl-18* is expressed as early as *unc-86* (7 hpf). However, this observation suggests that *egl-18* must be already expressed at this earlier embryonic stage (5 hpf) in order to regulate *hlh-3* expression in the HSN precursor.

Next, in the few cases where we observed cross-regulation between factors, we wanted to check if this effect could already be observed at early developmental stages, what would suggest a role in initiating the expression, or, on the contrary, it could be specific to late stages, indicating a role in maintenance of expression → **Figure 3.2.15-B**. We assessed *sem-4* and *egl-46* expression in the HSNs of *unc-86(n846)* mutants, at L1 larval stage. While *sem-4* still showed significant reduced expression in the cell supporting the requirement of UNC-86 to start its expression, *egl-46* levels were comparable to wild type. The same occurred with *egl-46* expression in *hlh-3(tm1688)* mutant animals. This indicates that UNC-86 and HLH-3 are required for the maintenance but not for the initiation of *egl-46* expression in the HSN. Regarding AST-1, this TF is not expressed till late third larval stage so L1 scoring could not be addressed.

Our genetic interaction analysis, summarised in → **Figure 3.2.15-D**, suggests that the expression of the six TFs is mainly independent of each other, which strongly suggest that the TFs regulating HSN terminal differentiation act through parallel pathways. However, we do see certain degree of cross-regulation between TFs, where some players like UNC-86 seem to act higher in the pathway to assure the expression of other members of the HSN TF code, while the rest of the players seem to exclusively act at the terminal level. Moreover, AST-1 expression seems to be tightly regulated by other members. Importantly, regardless of the presence or absence of cross-regulation, our previous mutagenesis experiments and EMSA analyses, show that all of them act at the terminal level of HSN differentiation regulating in parallel the expression of the 5-HT pathway genes.

AST-1, UNC-86, SEM-4, EGL-46 and EGL-18 are continuously required to maintain the serotonergic identity in the HSN

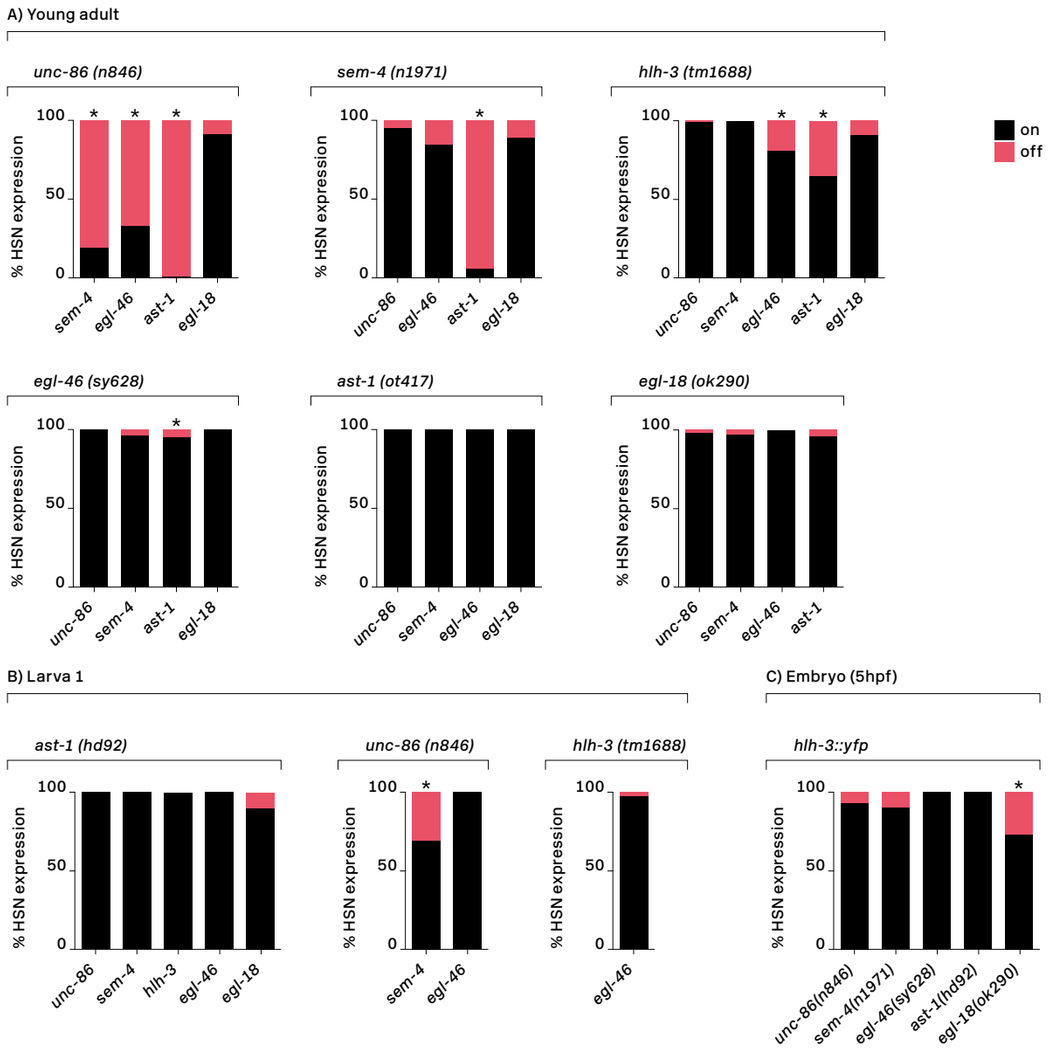
Although specific features of individual neurons are remarkably plastic, most neuron identity features remain stable throughout the life of a terminally differentiated postmitotic neuron. Neuron types and their elaborate connectivity patterns are maintained for up to many decades of life in mitotically quiescent cells. It is believed that gene regulatory programmes launched early in fetal life to specify neuronal type identities continue to function later in life to maintain the correct differentiated state and that interference with these maintenance mechanisms can lead to loss of neuron identity, function and circuitry (Deneris & Hobert 2014). We just have shown that AST-1, UNC-86, SEM-4, EGL-46 and EGL-18 are expressed in the HSN of adult worms. Therefore, we explored if the stable serotonergic identity of the HSN requires the continuous action of the HSN TF code in the cell.

Figure 3.2.15
Cross-regulation between the six members of the HSN regulatory code

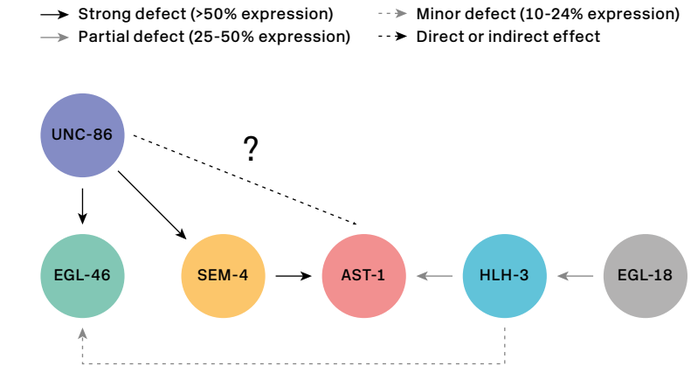
A) Quantification of reporter expression at adult stage in the *unc-86(n846)*, *sem-4(n1971)*, *egl-46(sy628)*, *ast-1(ot417)* and *egl-18(ok290)* mutant backgrounds. Mutant expression relative to wild type expression is represented. Strains used: *otIs337(unc-86::yfp)* and *stIs11606(egl-18::mCherry)* fosmid reporters; *kuIs34(sem-4::gfp)*, *kuIs35(sem-4::gfp)* and *vlcEx324(egl-46::DsRed)* transcriptional reporters; *ast-1[vlc19(ast-1::gfp)]* CRISPR-Cas9-mediated *gfp* knock-in. n>50 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test, *: pV< 0.05. See Annex 3.2.6.

B) Quantification of reporter expression at L1 larval stage in the *ast-1(hd92)* L1 lethal mutants and in the *unc-86(n846)* and *hlh-3(tm1688)* mutants that showed expression defects at adult stages.

C) Quantification of *otEx4140(hlh-3::yfp)* fosmid reporter expression at 5 hours post-fertilisation (hpf) in the five null mutant backgrounds *(unc-86(n846), sem-4(n19719), egl-46(sy628), ast-1(hd92)* and *egl-18(ok290))*. n>35 embryos per condition. Statistical significance was calculated using the two tailed Fisher exact test, *: pV< 0.05.



D) Summary schematic of the cross-regulation relationships between TFs. UNC-86 is epistatic to SEM-4, EGL-46 and AST-1 or, alternatively, *ast-1* reduced expression in *unc-86* mutants is due to an indirect effect of SEM-4 control over AST-1. EGL-18 is also epistatic to HLH-3 and HLH-3 is, in turn, epistatic to EGL-46.



To address if these TFs are required for a maintained expression of the 5-HT pathway genes in the neuron, we took advantage of the readily available RNAi technology in *C. elegans*. → **Figure 3.2.16-A** describes the experimental set up to knock down the expression of our candidates after HSN maturation. Of note, as HLH-3 was not expressed in the adult HSN, it was not considered for the maintenance analysis. We used the *rrf-3 (pk1426)* neuronal RNAi sensitised strain carrying the *tph-1* fosmid reporter (*otIs517*). Worms were grown under normal bacteria (OP50) until young adult stage, when HSN has already matured. At this time point *tph-1* expression in the HSN was assessed, ranging between 93-98% in all of the cases, and then worms were transferred to RNAi treated plates. These plates contained dsRNA for *ast-1*, *unc-86*, *sem-4*, *egl-46* and *egl-18*. As a positive control, dsRNA against *gfp* was used. As a negative control the empty vector L4440 was used. Animals were allowed to grow at 15 °C for three days before the final scoring. This was considered enough time to allow GFP signal to degrade and see any possible defects in *tph-1* expression in the cell. The temperature was set at 15 °C to diminish the metabolic rate of the worms and allow for late scorings → **Figure 3.2.16-A**. Under these experimental conditions, if any of our candidates is required for the maintenance of the *tph-1* gene in the adult and aging HSN cell, after RNAi knock down of the gene, fluorescent protein expression should be lost. On the contrary, if they are not required for this process, expression in the cell should remain invariant. In all cases, we observed a decrease in the *tph-1::yfp* signal → **Figure 3.2.16-B,** → **Annex 3.2.3.** These results reveal that all members of the HSN TF collective that are expressed in adult HSN are required for identity maintenance.

Overexpression of the HSN regulatory code is sufficient to induce serotonergic fate in some cellular contexts

We have shown that a specific combination of six TFs is co-ordinately and continuously required for proper terminal differentiation of the HSN neuron. Next we wanted to explore if, in the same way as depletion of any of these candidates causes general loss of identity in the HSN, ectopic expression of any of these candidates in other cells could force the serotonergic fate. This is not unconceivable as there are plenty of examples in *C. elegans* and in vertebrates where ectopic expression, or miss-expression, of specific factors leads to the production of extra cells with that particular fate (Duggan

et al. 1998; Flames & Hobert 2009; Lodato et al. 2014)

To address whether the HSN regulatory code is not only necessary for proper differentiation of the neuron but also sufficient, we used the heat shock inducible promoter *hsp-16.2* to ectopically induce expression of all the members of the code, either individually or in combination, at two developmental times: embryogenesis and L1 larval stage.

The heat shock response, a temperature-dependent stress defence mechanism, offers a straightforward strategy for temporal control of transgene expression. The heat shock response is common to bacteria, plants and mammals. In *C. elegans*, it is mediated by heat shock factor 1 (HSF1), a TF that is synthesised constitutively but remains latent during unstressed conditions (Lis & Wu 1993). In response to heat stress, HSF1 trimerises and binds with high affinity to promoters containing specific binding elements, leading to the transcription of different families of heat shock proteins (HSPs) (Pelham 1982; Lis & Wu 1993). Thus, transgenes containing HSF1-binding elements (*hsp* promoters) can be induced, albeit with little cellular specificity, following a shift to stressful temperatures. Such *hsp* promoters are frequently used in *C. elegans* as drivers of gene expression in a time-controlled manner. The *hsp-16.2* promoter drives expression most strongly in hypodermal cells and neurons, while the *hsp-16.41* promoter is more efficient in directing expression in the intestine and pharyngeal tissue (Fire et al. 1990). We chose the former for our analysis, reasoning that the HSN regulatory code, if able, would induce the serotonergic fate in neurons more easily than in non-neuronal tissues.

The strains used in the analysis are N2 worms expressing the cDNA of the six TFs under the control of the *hsp-16.2* promoter (what we call the 'heat shock array'). Due to the extrachromosomal nature of the strains, the red *ttx-3::mCherry* co-marker allowed us to identify animals bearing the heat shock array. As a serotonergic identity marker we chose *tph-1*, which is specifically expressed in serotonergic neurons and is not shared by other monoaminergic neurons.

We carried out a lineal study to know which was the optimum time to induce expression of two factors, *ast-1* and *unc-86*, in terms of average number of *tph-1::gfp* expressing cells in the whole population of heat shocked embryos. Starting with *ast-1*, we analysed embryos that the day before had received a heat shock (20 min at 37 °C) at different developmental times: 180 min after first cleavage (3 hpf), 300 min after first cleavage (5 hpf), 420 min after first cleavage (7 hpf) and 540 min after first cleavage (9 hpf) → **Figure 3.2.17**. Optimum time for *ast-1* expression was 7 hpf, coinciding with the start of embryonic elongation (mean ± SEM: 7 ± 0.4 cells), followed by 5 hpf (6 ± 0.8 cells), corresponding to the start of embryonic neurogenesis → **Figure 3.2.17-A**. Embryos that received the heat shock at 3 hpf, at the start of gastrulation, showed no more than 4 *tph-1::gfp* expressing cells or no GFP at all, suggesting embryo arrest or embryo death. For this reason, we did not consider this temperature in the study of *unc-86*. The optimum temperature for *unc-86* over expression was 5 hpf (9 ± 1.4 cells), followed by 7 hpf (6 ± 0.8 cells). The same results were observed for the last temperature analysed (9 hpf, 6.0 ± 0.5 cells) → **Figure 3.2.17-B**. Considering that the difference in number of *tph-1::gfp* expressing cells between the two treatments (5 hpf and 7 hpf) is greater for *unc-86* than for *ast-1*, we decided to select 5 hpf as the standard time point for heat shock treatment in the analysis of the HSN regulatory code. Moreover, it has also been reported that *ast-1* over expression at 5 hpf conferred the maximum response in *dat-1::gfp* (dopaminergic marker) expressing embryos (Flames & Hobert 2009).

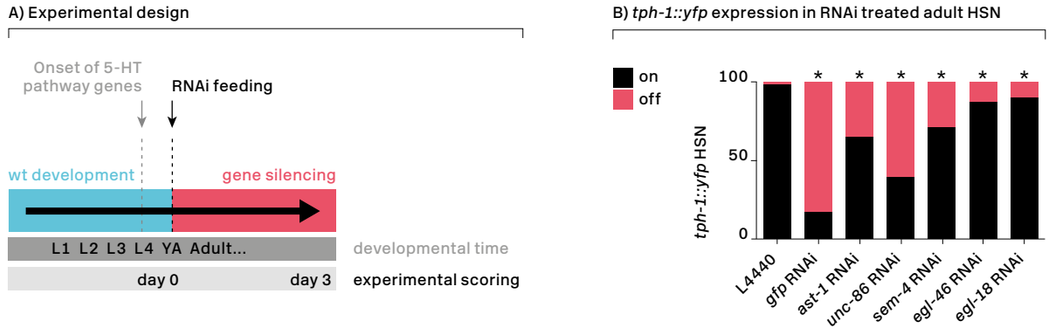
Following the same heat shock protocol we ectopically expressed the six members of the HSN regulatory code individually and scored the resulting

Figure 3.2.16
Requirement of the HSN regulatory code for the maintenance of HSN identity

A) Experimental set up for RNA interference assay. *rrf-3(pk1426)* animals are grown at 20 °C on normal food (*E. coli* OP50 strain) until late L4 stage, when the 5-HT pathway genes activate their expression in the HSN neuron. Animals were scored for *tph-1::yfp* expression in the HSN neuron (day 0). Next, animals were moved to plates containing RNAi treated bacteria (HT115); RNase deficient *E. coli* strain transformed with individual clones for the *unc-86*, *sem-4*, *egl-46*, *ast-1* and *egl-18*. Animals were allowed to feed upon this food during 72 hours, at 15 °C, and then *tph-1::yfp* expression in the HSN was scored (day 3).

B) Loss of function (RNAi) experiments after HSN differentiation show that AST-1, UNC-86, SEM-4, EGL-46 and EGL-18 are required to maintain proper *tph-1* expression.

L4440 is the .empty vector usead as negative control. n > 50 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test; *pV < 0.05. See Annex 3.2.3.



number of *tph-1::gfp* expressing cells. As control we used heat shocked reporter animals (*zdIs13*), without the heat shock array (termed 'wild type'). We analysed different independent lines with the exception of the *ast-1* strain that was integrated.

→ **Figure 3.2.18-A** shows the mean and standard error of the mean of *tph-1::gfp* expressing cells in the whole population of embryos bearing the heat shock array (*ttx-3::mCherry* positive), pooling together the different lines for the same TF. → **Figure 3.2.18-B** shows this information broken down into the different lines → **Table 3.2.1 and** → **Table 3.2.2**. Our results indicate that over expression at embryonic stages of *ast-1*, *unc-86* and *sem-4*, but not *hlh-3*, *egl-46* and *egl-18* provokes an increase in the number of *tph-1::gfp* expressing cells, in comparison to wild type.

Next, we decided to analyse the effect of the combined expression of the 6 members of the HSN regulatory code (termed 'combo 6'), and also of the three members that significantly increased the number of *tph-1::gfp* cells in the population: *ast-1*, *unc-86* and *sem-4* (termed 'combo A+U+S'), to see if *tph-1::gfp* ectopic expression could be enhanced. Indeed, both combinations achieved a higher number of *tph-1::gfp* expressing cells (combo A+U+S: mean = 10.7 ± 0.5; combo 6: mean = 10.2 ± 0.6) that

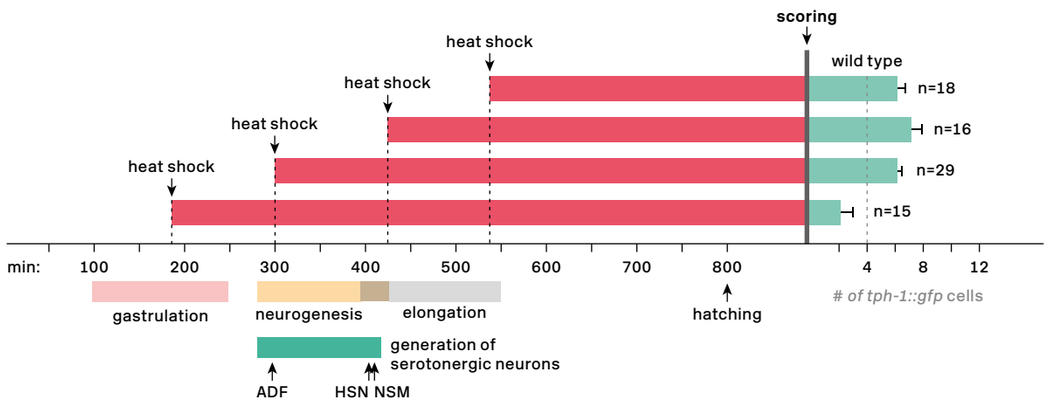
Figure 3.2.17
Lineal study of *ast-1* and *unc-86* ectopic expression in the embryo

A) Schematic representation of the heat shock-induced *ast-1* and *unc-86* factors

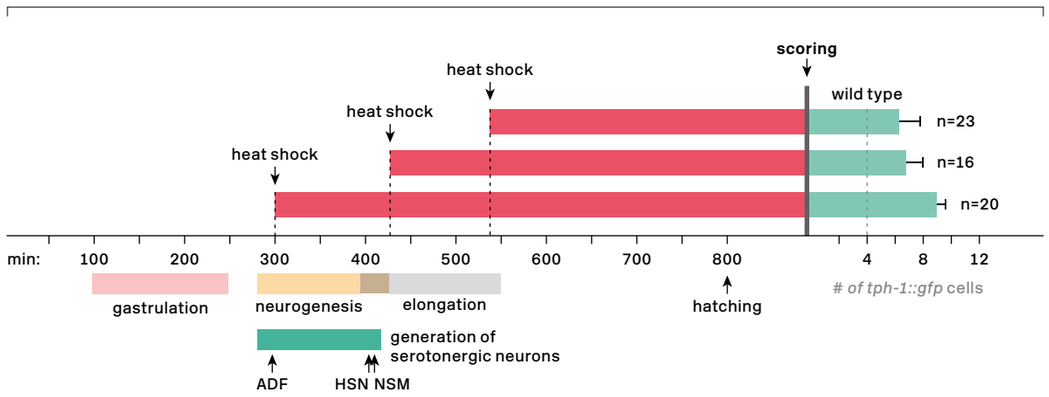
B) Schematic representation of the heat shock-induced *ast-1* overexpression experiments

Embryos carrying the heat shock array received the treatment (20' at 37 °C) at different developmental time points, and were scored the next day (23 hours post-fertilisation). Red lines indicate misexpression of the gene. Green lines represent the mean number of GFP cells after the treatment and error bars represent the standard error of the mean (SEM). *ast-1* and *unc-86* overexpression leads to ectopic *tph-1::gfp* expression in 7 and 8 cells, respectively, compared to 4 *gfp* cells in wild type.

A) *ast-1* ectopic expression



B) *unc-86* ectopic expression



was statistically different from wild type animals and from all single members except for *unc-86*. However, there was no difference between ectopic expression of the complete HSN regulatory code (combo 6) and combo A+U+S. Again, jointly or independent analysis of the lines showed the same results, except for the difference between *sem-4* and combo A+U+S, which is no longer statistically significant → **Figure 3.2.18-A, B** and **E,** → **Table 3.2.1** and → **Annex 3.2.7.**

We realised that in → **Figure 3.2.18-A** and **B** there are two components contributing to the ectopic number of *tph-1::gfp* positive cells: the total number of GFP positive cells that every individual embryo expresses, and the total number of embryos that respond to the heat shock. For this reason we next treated our data separately and represented, on the one hand, the average of *tph-1::gfp* positive cells per 'positive embryos' (>4 *tph-1::gfp* cells) in comparison to all wild type animals → **Figure 3.2.18-C** for combined lines, and → **Figure 3.2.18-D** for individual lines). In this way we see how embryos, in which over expression of a factor (or factors) has an effect, differ from wild type animals. Notably, the maximum number of *tph-1::gfp* positive cells scored in wild type animals is six → **Table 3.2.1 and 3.2.2**, two more than expected. This additional pair cells are ASG neurons. Other reports have also described that 5% of ASG neurons express *tph-1::gfp* in normoxic conditions (Pocock & Hobert 2010).

On the other hand, we represented the percentage of 'positive embryos' in relation to the total number of embryos analysed, to see the ability that a particular TF, or a combination of them, has to elicit an ectopic response → **Figure 3.2.18-F.** → **Figure 3.2.18-C** shows that embryos that respond to over expression of *ast-1*, *unc-86*, *sem-4*, *hlh-3*, *egl-18*, combo A+U+S and combo 6, but not to *egl-46*, show higher numbers of serotonergic-like cells that wild type animals, being *unc-86* the single factor that produces the most *tph-1::gfp* + cells (mean = 11.1 ± 0.5; maximum = 18) → **Table 3.2.3.** → **Figure 3.2.18-G** shows representative images of 'positive embryos' and non-responding embryos. Although combo A+U+S and combo 6 show embryos with up to 19 and 20 cells, respectively, their means (combo A+U+S: 11.6 ± 0.4; combo 6: 11,2 ± 0.6) are not statistically different from *unc-86* single over expression and neither are their means different from each other. However, *sem-4* single is statistically different from combo 'A + U + S' but not to combo 6, and *ast-1*, *hlh-3* and *eg-18* singles are statistically different to both combos. Similar results are obtained when considering the different lines individually → **Figure 3.18-D,** → **Table 3.2.4.**

Regarding the percentage of embryos that respond to the heat shock treatment ('positive embryos'), → **Figure 3.2.19-F** clearly shows that over expression of the members of the HSN regulatory code (either individually or in combination) dramatically increases the number of embryos that show more than 4 GFP positive cells in all treatments. Although overexpression of *egl-46* did not lead to a higher mean number of *tph-1::gfp* positive cells compared to wild type animals, the fraction of embryos that are able to express the serotonergic marker in more than four cells is higher than wild type animals (pV=0.0026). Remarkably, 'singles' render lower percentages of positive embryos than 'combos', with the exception of *sem-4* that shows no statistical difference only with combo 6 (pV= 0.0967). Moreover, there are two clearly differentiated groups within single factors: high responding (*ast-1*, *unc-86* and *sem-4*) and low responding (*hlh-3*, *egl-46* and *egl-18*). This correlates with the fact that combo 6 does not further enhance the percentage of embryos that are able to respond to the ectopic expression of the complete HSN regulatory code, in comparison to combo A + U + S. For specific data see → **Annex 3.2.7.**

Overexpression of the HSN regulatory code, either of single members or in combination, inhibited pos-

terior development of the worms. Embryos rarely reached comma stage or, if they did, their morphology was so atrophied that the embryonic stage was difficult to determine. This did not occur, however, in the majority of the wild type worms. We argued that if we over expressed the HSN TF collective at posterior developmental times, we would allow worms to grow and we would be able to identify the ectopic GFP cells, if any, as neuronal or non-neuronal. For this reason, we decided to perform similar analyses at the first larval stage.

Taking into consideration the results obtained in the previous heat shock over expression experiments, we decided to analyse ectopic expression at L1 stage of the 'high responding' strains (*ast-1*, *unc-86* and *sem-4*). As we observed no differences between the two combos used, we selected combo 6 for L1 overexpression. The heat shock regime followed was different: we heat shocked a synchronised population of L1 larva three times (30 min, 37 °C) with 2 hours resting intervals at 20 °C. Worms were scored the next morning, when most of them were already L2 larva although some remained at L1 stage. In this case, ectopic expression of *tph-1::gfp* was not as broad as during embryonic development. In animals where we induced ectopic expression of *ast-1*, we repeatedly observed *tph-1::gfp* expression in two pairs of neurons in the tail. 27% of the worms analysed (20/74) showed GFP in at least one and up to four cells in the tail → **Figure 3.2.19-B**. Due to its morphology, position and co-localisation with other reporters (*lgc-55*, *unc-17*, *ast-1*, *unc-86* but not *eat-4*) we identified the pair that stains fainter as ALN → **Figure 3.2.19-A.** The identity of the other pair of neurons remains uncertain. Moreover, 7% of worms (7/74) showed GFP expression in the PVT neuron → **Figure 3.2.19-A** and **B**. This neuron is unilateral and has a big soma, making it really easy to identify. However, this value was not statistically different from wild type, as a small percentage of them (3%) also showed GFP expression in this cell (pV=0.2975). Nonetheless, overexpression of *unc-86* induces *tph-1* expression in the PVT cell in 69% of the cases (pV<0.0001) and *sem-4* over expression in 17% of the worms (pV=0.04). Also in *hsp::unc-86* and *hsp:.sem-4* heat shocked animals, *tph-1::gfp* is observed in a single unidentified neuron in the head (91% in the first case, 33% in the second) → **Figure 3.2.19-A** and **B**. None of these strains show *tph-1::gfp* expression in the tail neurons. When we ectopically express the six members of the HSN regulatory code, we do not observe an enhanced phenotype. Instead, it seems to reflect the previously reported phenotype of the additive single factor over expression → **Figure 3.2.19-B**. Thus, the ectopic induction of *tph-1* expression seems mostly restricted to embryonic stages. This is probably due to the more compact state of the chromatin in mature postmitotic neurons. On the other hand, the limited effect observed with the embryonic induction of the six factors could suggest that there are additional factors playing a role on the selection of the terminal genes expressed by the HSN neuron.

HSN regulatory code shows both synergic and additive genetic interactions

Our phenotypic analysis of HSN regulatory code single mutants shows that all members of the HSN regulatory code are required for proper HSN terminal differentiation. However, each null mutant often shows only partial defects in the expression of most analysed reporters (only *unc-86* and to a less extent *sem-4* generally show 100% off phenotypes) → **Figure 3.2.5**. Moreover, our *cis*-regulatory analysis indicates that all members of the TF regulatory code act on the CRMs to direct their expression → **Figure 3.2.8-3.2.10**. Cooperativity is common among TFs acting on the same enhancer, *via* protein-DNA interaction and/or protein-pro-

Table 3.2.1
Variability and distribution analysis of *tph-1::gfp* expression in the whole population of embryos, considering the average of all lines

Statistic description of embryonic over expression of the HSN regulatory code collective, considering the whole population of animals. Wild type values refer to *tph-1::gfp* expressing animals without the heat shock array. Combo A+U+S refers to the combination of *ast-1*, *unc-86*, *sem-4*. Combo 6 refers to the combination of all the members of the HSN regulatory code. Independent lines analysed for individual factors are considered jointly. Is: integrated line, Ex: extrachromosomal array.

Measure	wt	*ast-1*	*unc-86*	*sem-4*	*hlh-3*	*egl-46*	Combo A+U+S	Combo 6
N	196	150	68	57	56	52	75	54
Mean	3.3	5.1	8.2	6.5	3.6	3.4	10.8	10.2
SD	1.2	2.2	4.7	2.5	1.8	1.3	4.5	4.6
SEM	0.1	0.2	0.3	0.3	0.2	0.2	0.5	0.6
Median	4	5	8	7	3	4	11	10
25-50% percentiles	2-4	4-7	4-12	5-8	2-5	2-4	8-15	6.75-12.5
Max #	6	10	18	14	9	6	19	20
Ex	1 (Is)	1 (Is)	1 (Ex)	1 (Ex)	2 (Ex)	2 (Ex)	2 (Ex)	2 (Ex)

Measure	wt	*ast-1*	*unc-86*	*sem-4*	*hlh-3*		*egl-18*		*egl-46*		Combo A+U+S	Combo 6	
Lines					1	2	1	2	1	2		1	2
N	196	154	68	57	26	30	19	33	52	15	75	46	8
Mean	3.3	5.1	8.2	6.5	3.0	4.1	3.5	3.4	3.3	4.5	10.8	10.1	10.8
SD	1.2	2.2	4.7	2.5	1.1	2.1	1.2	1.3	2.1	2.2	4.5	4.8	3.4
SEM	0.1	0.2	0.7	0.3	0.2	0.4	0.3	0.2	0.3	0.6	0.5	0.7	1.2
Median	4	5	8	7	3	4	3	4	3	4	11	10	10.5
25-50% percentiles	2-4	4-7	4-12	5-8	2-4	2-6	3-4	2-4	2-5	2-6	8-15	6-14	9.25-12
Max #	6	10	18	14	5	9	6	6	5-7	6-6	19	20	17
Lines (Is, ex)	Is	Is	Ex	Ex	Ex	Ex	Ex	Ex	Ex	Ex	Ex	Ex	Ex

Table 3.2.2
Variability and distribution analysis of *tph-1::gfp* expression in the whole population embryos, considering individual lines independently

See Table 3.2.1 for explanation.

Measure	wt	*ast-1*	*unc-86*	*sem-4*	*hlh-3*	*egl-46*	*egl-18*	Combo A+U+S	Combo 6
N	196	82	44	44	16	9	22	68	47
Mean	3.3	6.7	11	7.4	5.8	5.3	6.1	11.7	11.2
SD	1.2	1.3	3.1	2.1	1.1	0.5	1.1	3.6	4.0
SEM	0.1	0.1	0.5	0.3	0.3	0.2	0.2	0.4	0.6
Median	4	6	11	7	5.5	5	6	11	11
25-50% percentiles	2-4	6-7.25	8.25-13.75	6-8.75	5-6	5-6	5-7	9-15	8-14
Max #	6	10	18	14	9	6	10	19	20
positive embryos (%)	3.1	54.7	62.9	77.2	28.6	17.3	32.8	90.7	87.0
lines (Is. ex)	1 (Is)	1 (Is)	1 (Ex)	1 (Ex)	2 (Ex)	2 (Ex)	2 (Ex)	2 (Ex)	2 (Ex)

Table 3.2.3
Variability and distribution analysis of *tph-1::gfp* expression in heat shock responding embryos, considering the average of all lines

Statistic description of embryonic overexpression of the HSN regulatory code. Wild type values refer to the total embryos analysed. Single and combo values exclusively refer to heat shock-responding embryos; 'positive embryos' (carriers of the heat shock array and positive for the treatment > 4GFP cells). Combo A+U+S refers to the combination of *ast-1*, *unc-86*, *sem-4*. Combo 6 refers to the combination of all the members of the HSN regulatory code collective. % of positive embryos refers to the proportion of positive embryos regarding the total number of embryos analysed (positive and negative). Independent lines analysed for individual factors are considered jointly. Is: integrated line, Ex: extra-chromosomal array.

tein interaction (Levo & Segal 2014). A readout of cooperativity is observing genetic interaction between TFs and a way to assess genetic interactions is doing double mutant analysis. Before presenting the results, we will explain the concepts of genetic interaction and cooperativity.

Frequently, genes interact with one another, distorting simple Mendelian ratios or even leading to novel phenotypes. In general terms, a genetic interaction occurs when two alleles affecting different genes combine within an organism to yield a phenotype not simply explained by adding together the phenotypes associated with each of the two alleles. Hence, when analysing a double mutant animal, the null hypothesis is that the two genes act in independent pathways and the expected outcome is that double mutant phenotype is the product of the sum of its component single-locus effects (additive phenotype). Whatever deviates from this result will be considered synergy or a genetic interaction. Synergy and genetic interaction are terms indistinctly used by geneticists. Genetic interactions indicate a functional connection between two genes, which is distinct from physical interactions. Cooperative binding of TFs to the enhancer DNA is one mode of synergy. TFs can bind cooperatively to DNA interacting physically with another TF or cofactor, or without the need

Measure	wt	*ast-1*	*unc-86*	*sem-4*	*hlh-3*		*egl-46*		*egl-18*		Combo A+U+S	Combo 6	
Lines					1	2	1	2	1	2		1	2
N	196	82	44	44	2	14	4	5	1	2	68	39	8
Mean	3.3	6.7	11	7.4	5	5.9	5.25	5.4	15	7	11.7	11.3	10.8
SD	1.2	1.3	3.1	2.1	0	1.1	0.5	0.5	6.0	6.4	3.6	4.1	3.4
SEM	0.1	0.1	0.5	0.3	0	0.3	0.3	0.2	0.8	1.6	0.4	0.7	1.2
Median	4	6	11	7	5	6	5	5	0.2	0.6	11	11	10.5
25-50% percentiles	2-4	6- 7.25	8.25-13.75	6 - 8.75	5-5	5- 6.25	5.75-6	5-6	6	6	9-15	8-14	9.25-12
Max #	6	10	18	14	5	9	6	6	5-7	6-6	19	20	17
positive embryos (%)	3.1	54.7	62.9	77.2	7.8	46.7	21.1	15.5	28.8	46.7	90.7	84.8	100
lines (Is. ex)	Is	Is	Ex	Ex	Ex	Ex	Ex	Ex	Ex	Ex	Ex	Ex	Ex

Table 3.2.4
Variability and distribution analysis of *tph-1::gfp* expression in heat shock responding embryos, considering individual lines independently.

See Table 3.2.3 for explanation.

to physically interact, via changes in DNA accessibility or DNA conformation (Levo & Segal 2014; Spitz & Furlong 2012) → **Figure 1.5**. It has been proposed that the RNA polymerase II transcriptional machinery is designed to respond in a synergistic (greater-than-additive) manner upon binding of multiple activators to achieve specificity to signal response. This is possible if an RNA pol II enhancer is organised in unique combinations of activators, closely packed, that promote their interaction and cooperative binding to DNA. In this way, the binding of one TF will facilitate o hamper the binding of another TF and the recruitment of the RNA pol II complex, hence affecting gene expression, in a greater-than-additive fashion (Carey 1998).

Although genetic analysis by itself will not allow for elucidation of the detailed molecular mechanism by which the members of HSN regulatory code interact, it will allow us to distinguish between two situations: TFs bind independently to the DNA to regulate the expression of the 5-HT pathway genes, or they genetically interact in some way to achieve gene expression. A third situation can be envisioned, where one factor regulates the expression of the other (epistasis). W. Bateson coined the term epistasis to describe those cases where a mutation in one gene masks the effect of a mutation in a second gene (Bateson 1911). We have de-

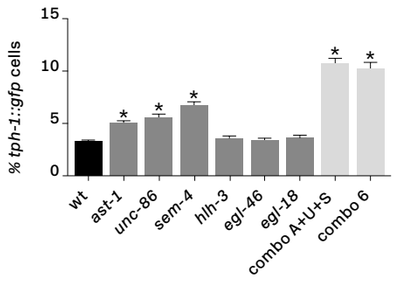
Figure 3.2.18
Induction of serotonergic fate through over expression of the HSN regulatory code at embryonic stages

Embryos carrying the heat shock array received the treatment (20' at 37 °C) during neurogenesis (5 hours post-fertilisation) and were scored the next day. Wild type animals (without the transgene) were compared to animals that over expressed *ast-1, unc-86, sem-4, hlh-3, egl-46* and *egl-18* single factors, and the combinations of *ast-1, unc-86* and *sem-4* (combo 'A+U+S') and the six factors (combo 6).

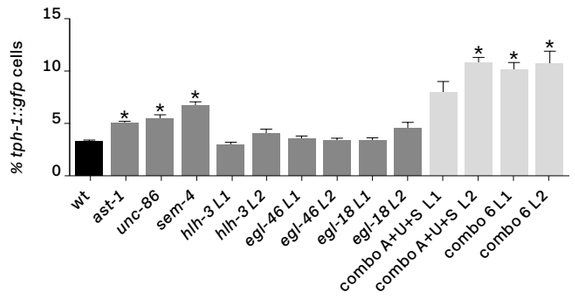
A), B) Mean number of *tph-1::gfp* expressing cells in the whole population of embryos analysed, considering independent lines together (A) or separately (B). Error bars represent the SEM. n > 50 animals per condition. D'Agostino & Pearson omnibus normality test indicate that two groups, wild type and *unc-86*, do not follow a normal distribution. L: independent line number. *: statistical significant difference between wild type and the rest of conditions using the non-parametric Kruskal-Wallis test with Dunn's correction for multiple comparisons. See Figure 3.2.18-E for all comparisons.

Analysis of the whole population of embryos

A)# GFP cells (average of lines)



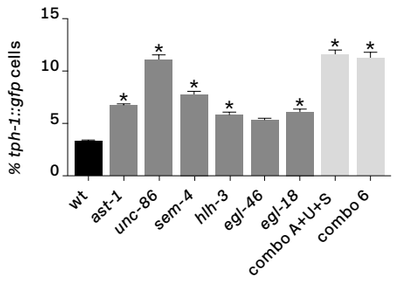
B) # GFP cells (individual lines)



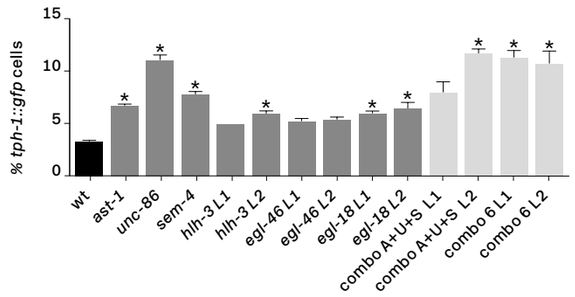
C), D) Mean number of *tph-1::gfp* expressing cells exclusively in the population of heat shock responding embryos with more than 4 *gfp* positive cells (positive embryos), considering independent lines together (C) or separately (D). D'Agostino & Pearson omnibus normality test indicate that none except *unc-86*, *egl-46* and combo 6 follow a normal distribution. L: independent line number. *: statistical significant difference between wild type and the rest of conditions using the non-parametric Kruskal-Wallis test with Dunn's correction for multiple comparisons. See Figure 3.2.18-E for all comparisons.

Analysis of the heat shock-responding population of embryos (positive)

C) # GFP cells in positive embryos (average of lines)



D) # GFP cells in positive embryos (individual lines)



E) Statistical significance in (A)-(D) was calculated using Kruskal-Wallis non-parametric test, with Dunn's multiple test correction. The schematic includes relationships between wild type, single factors and combinations of factors, considering all the independent lines for each condition together. See Annex 3.2.7 for complete statistical analysis.

E) Comparison between wt, single factors and combos

Whole population of embryos

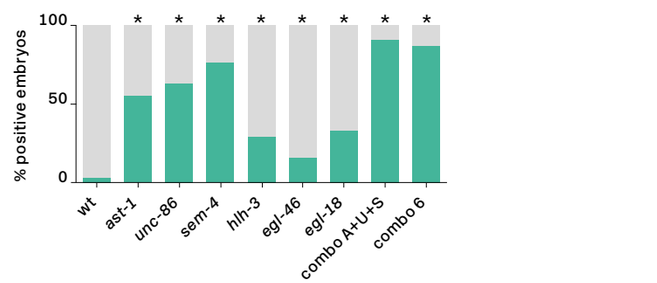
	hsp::ast-1	hsp::unc-86	hsp::sem-4	hsp::hlh-3	hsp::egl-46	hsp::egl-18	combo 'A+U+S'	combo 6
wt	****	****	****	ns	ns	ns	****	****
hsp::ast-1		*	ns	**	**	**	****	****
hsp::unc-86			ns	****	****	****	ns	ns
hsp::sem-4				****	****	****	*	ns
hsp::hlh-3					ns	ns	****	****
hsp::egl-46						ns	****	****
hsp::egl-18							****	****
combo 'A+U+S'								ns

'Positive embryos'

	hsp::ast-1	hsp::unc-86	hsp::sem-4	hsp::hlh-3	hsp::egl-46	hsp::egl-18	combo 'A+U+S'	combo 6
wt	****	****	****	**	ns	****	****	****
hsp::ast-1		***	ns	ns	ns	ns	****	***
hsp::unc-86			ns	**	*	**	ns	ns
hsp::sem-4				ns	ns	ns	**	*
hsp::hlh-3					ns	ns	***	**
hsp::egl-46						ns	**	*
hsp::egl-18							***	**
combo 'A+U+S'								ns

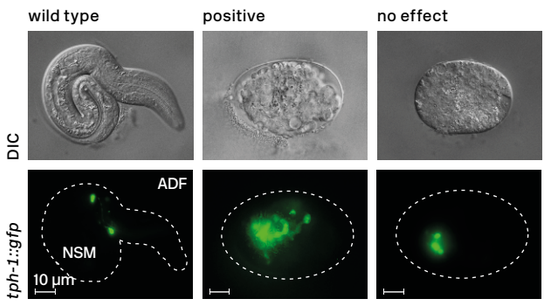
F) Percentage of positive embryos relative to the whole population of embryos analysed. All factors, alone or in combination, are able to enhance *tph-1::gfp* expression in more than 4 cells, when overexpressed during neurogenesis. Statistical significance was calculated using Fisher exact test; *pV<0.05. See Annex 3.2.7.

Percentage of positive embryos



G) Representative images of a wild type (control animal that receive the heat shock but does not carry the heat shock array), a positive embryo (that carries the heat shock array and responds to the treatment) and a negative embryo (that carries the array but does not respond to the treatment).

Overexpression of HSN regulatory code (5 hpf)



termined that *unc-86* is epistatic of *sem-4*, and *egl-46*, *sem-4* and to a less extent *hlh-3* are epistatic of *ast-1* → **Figure 3.2.15**. In these cases, double mutant analysis cannot be used to assess cooperativity. Alternatively, we can use hypomorphic alleles, although interpretation of the results is more difficult.

In order to better understand these genetic interaction concepts and to interpret correctly the double mutant results presented next, we have included schematic representations of all possible outcomes in → **Figure 3.2.20**. → **Figures 3.2.20-A-E** show the case of two genes, *a* and *b*, that code for two TFs A and B that regulate the expression of gene *x*. In → **Figure 3.2.20-A** two recessive null loss-of-function alleles, a_1 and b_1, show an incomplete penetrant phenotype over the expression of gene *x*. As an example, mutation in gene *a* shows 60% expression of the gene (a_1 = 40% loss of gene expression), while mutation in gene *b*, leads to 80% of expression in gene *x* (b_1 = 20%). When both mutant alleles are present in homozygosis in the same animal gene *x* expression is 40% (a_1b_1 = 60%). So in this case, assuming the null hypothesis of additivity $a_1b_1 = a_1 + b_1$ = 40% + 20% = 60%. The phenotype observed matches a simple additive relationship and thus, we infer that A and B act independently → **Figure 3.2.20-B**. By contrast, if the double mutant phenotype also conferred 60% of gene *x* expression (40% off), we would infer that A and B act together in the same linear pathway. As double mutant phenotype mimics a_1 phenotype, this could be interpreted as 'epistasis' where gene *a* is epistatic to gene *b* (or gene *b* is hypostatic to gene *a*) → **Figure 3.2.20-E**.

Alternatively, double mutant analysis can unravel synergistic enhancement. Considering, for example, that single mutant a_1 shows a 40% reduction in gene *x* expression (60% of the worms will still express the gene) and single mutant b_1 shows a 20% decrease (80% on), if the double mutant a_1b_1 phenotype is that 90% of the animals lose the expression of gene *x* (and only 10% maintain it), we talk about genetic or synergic enhancement → **Figure 3.2.20-C**. In this case, the addition of the two single mutant phenotypes cannot explain the total loss observed in the double mutant (a_1b_1 (90%) > $a_1 + b_1$ (40% + 20% = 60%)). Instead, if the double mutant animals showed less phenotype than any of the single mutants, then we would talk about genetic or synergic suppression → **Figure 3.2.20-D**. Synergic interactions typically result from apparent functional redundancy. One common reason for redundancy involves genetic pathways that act in parallel to elicit a similar outcome, where either pathway can functionally compensate for the other. In these two cases, both genes will be acting in parallel pathways over the same substrate (gene *x*). In our previous results we have seen some examples of redundancy when analysing *cis*-regulatory motifs, thus we predict similar synergistic effects will be found in our double mutant analysis.

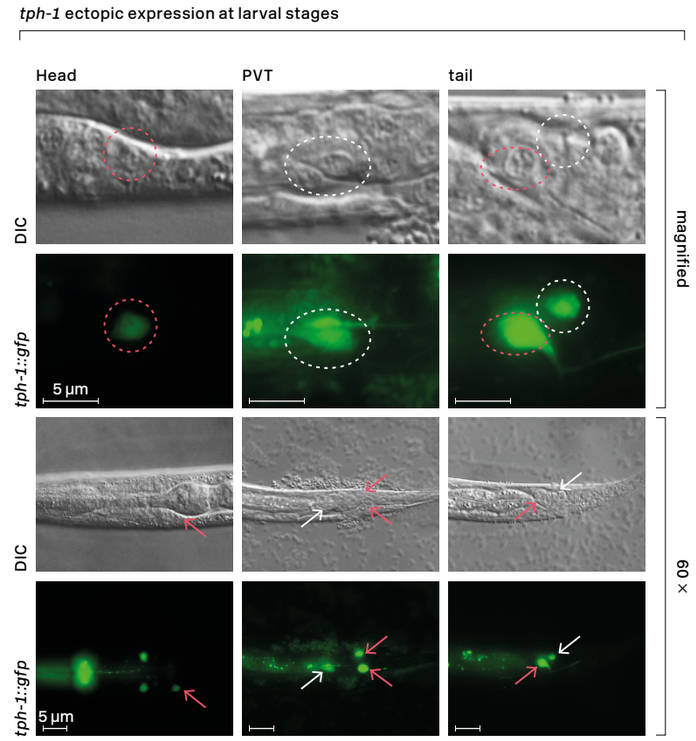
Finally, another type of genetic interaction, which differs from synergic enhancement, is synthetic lethality. In this case, two independent mutations are viable on their own, but in combination they lead to dead worms. In this case, both genes must genetically interact because neither allele alone provokes lethality. The lethal phenotype appears when both alleles are combined; i.e. the lethal phenotype is synthetic, meaning created *de novo*. These interactions tell us that normally in the organism, both genes work together to keep worms alive.

To be able to detect synergy between the members of the HSN regulatory code, we selected reporters for genes whose expression is only partially affected in null mutants for the HSN TF collective → **Figure 3.2.5**. We focused our analysis on *tph-1*, *cat-1* and *bas-1* gene expression because their HSN CRMs contain functional BSs for all six factors, thus we know all of them have a termi-

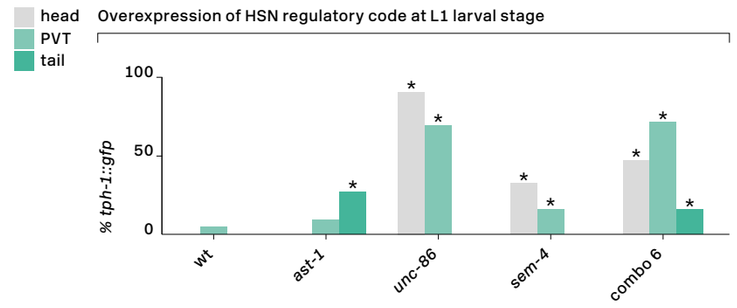
Figure 3.2.19
Induction of serotonergic fate in specific neurons through overexpression of the HSN regulatory code at larval stages

Synchronised population of animals bearing the heat shock array for *ast-1*, *unc-86*, *sem-4*, combo A+U+S and combo 6, received three heat shock treatments (30', 37 °C) every two hours and were assessed for *tph-1::gfp (zdIs13)* expression the next day.

A) Micrographs showing ectopic *tph-1::gfp* expression in the head and tail of L1-L2 larvae. Top panels show highly magnified DIC and fluorescence images of the neurons. White dotted circles indicate neurons whose identity has been determined (PVT and ALN), while red dotted circles indicate unknown neurons. The bottom panel shows a low magnification picture in order to visualise the anatomy of the worm. White arrows indicate PVT and ALN neurons, while red arrows indicate unknown neurons.



B) Percentage of larvae expressing *tph-1::gfp* in the head, PVT and at least one neuron in the tail, after the heat shock treatment. n > 30 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test; *pV < 0.05. See Annex 3.2.7.



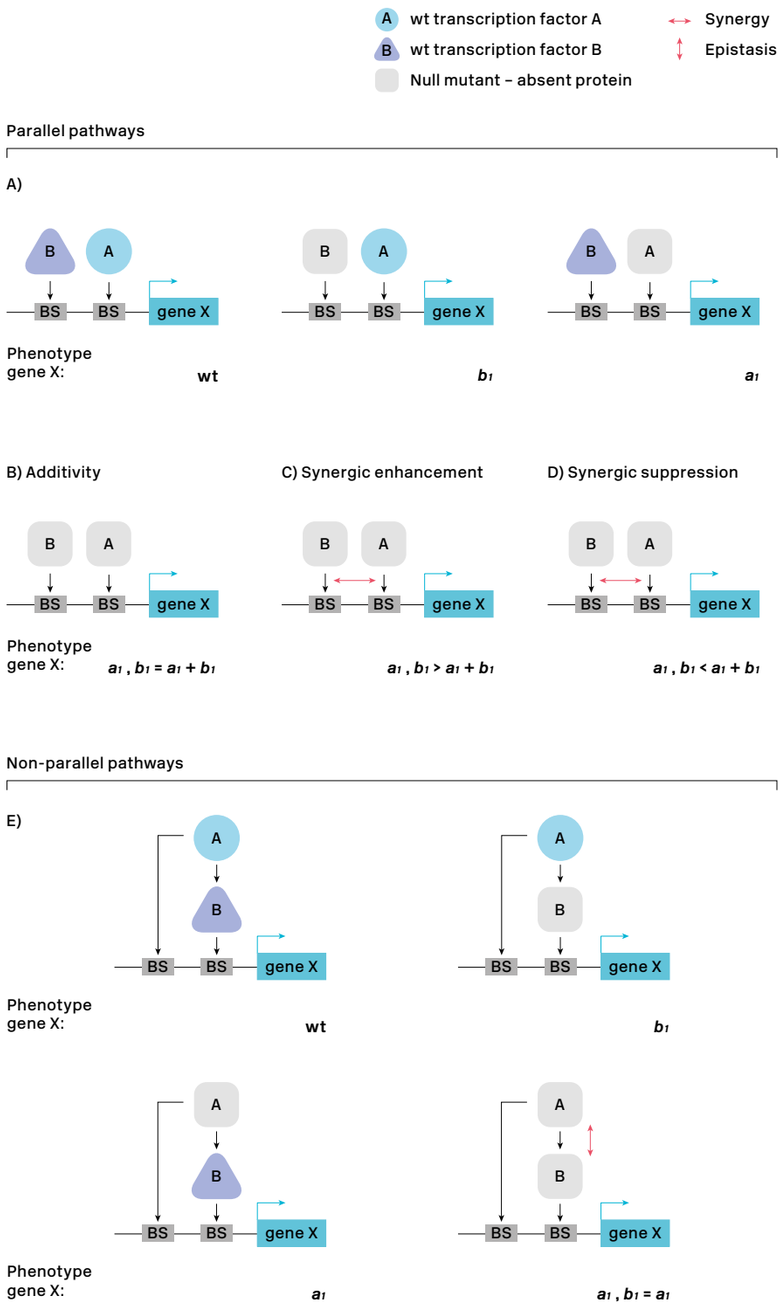


Figure 3.2.20
Models of genetic interaction between transcription factors to regulate the expression of a gene.

A) A and B are two TFs that regulate the expression of gene *x* in a particular cell. a_1 and b_1 represent mutant phenotypes for gene *a* and *b* null loss of function alleles, in terms of loss of expression of gene x. a_1b_1 indicates double mutant phenotype. wt reflects normal levels of expression of gene *x* in that particular cell. Figures (A)-(D) represent hypothetical examples in which A and B act in parallel to regulate the expression of gene *x*, whereas in Figure E, A and B act in the same linear pathway.

B) TFs act independently to regulate the expression of gene *x;* example of additivity. Double mutant for genes *a* and *b* exhibit a loss of expression in gene *x* comparable to the sum of the single mutant phenotypes.

C) TFs genetically interact to regulate the expression of gene *x;* example of synergic enhancement. Double mutant phenotype is more severe than the sum of single mutant phenotypes.

D) TFs genetically interact to regulate the expression of gene *x;* example of synergic suppression. Double mutant phenotype is less severe than the sum of single mutant phenotypes.

E) TFs act in the same linear pathway to regulate the expression of gene *x*. In this example, *a* has an additional direct role on gene *x*. Double mutant phenotype mimics a_1 single mutant phenotype; A acts upstream to B and *a* is epistatic to *b*.

nal role → **Figure 3.2.8-3.2.10**. For example, in → **Figure 3.2.21-A** we show that *bas-1* is slightly affected in *egl-46(gk692)* (85±3% expression) and in *hlh-3(tm1688)* (49±5% expression) null mutants but in the double mutants, *bas-1* expression is completely abolished (0±0%, pV<0.0001). The double mutant phenotype is greater that the addition of both single mutant phenotypes (100% OFF compared to 15% + 51% = 66% OFF), thus our results indicate synergy between these two factors. Of note, from our cross-regulation studies, we know that HLH-3 down-regulates the expression of a transcriptional reporter of *egl-46* in 20% → **Figure 3.2.15**. Hence, perhaps *hlh-3* mutant phenotype is due to its direct action on *bas-1* expression together with the slight effect on *egl-46* expression. However, if *egl-46* did not have an additional effect in the regulation of *bas-1*, the double mutant would reveal that *hlh-3* is epistatic to *egl-46* (as in the example → **Figure 3.2.20-E**). As this is not the case, we can conclude that HLH-3 and EGL-46 act in parallel and show synergic enhancement over *bas-1* regulation.

As *ast-1(hd92)* allele shows larval lethality and we had previously shown that a mosaic *ast-1* strain showed the same lack of phenotype for *bas-1* expression as the *ast-1(ot417)* allele, we used this hypomorph allele in the analyses. Double mutant *ast-1*; *egl-46* abolishes completely the expression of the *bas-1* gene, which largely exceeds the sum of single mutant phenotypes (pV<0.0001) → **Figure 3.2.21-A**. Although cross-regulation analysis shows that EGL-46 downregulates in 5% the expression of endogenous *ast-1* expression → **Figure 3.2.15**, the double mutant has a much greater phenotype than *egl-46* mutant alone, which indicates that both factors act synergistically. Combining *ast-1(ot417)* with hypomorph alleles of *unc-86(n848)* or *sem-4(n2654)* also abolishes *bas-1* expression from the HSN (pV<0.0001) → **Figures 3.2.21-A**. However, as *ast-1* expression is dependent on *unc-86* and *sem-4*, and this analysis is done with hypomorphic alleles we cannot discard that the synergism observed is due to partial epistatic effects on AST-1 → **Figure 3.2.15**.

ast-1; *hlh-3* and *ast-1*; *egl-18* double mutant animals showed synthetic lethality. We know that absence of the 5-HT neurotransmitter is not lethal as *tph-1(mg280)* null mutants are perfectly viable (Sze et al. 2000). Therefore, these results tell us that these TFs are acting together to regulate a different vital process and preclude us from their study in the 5-HT pathway genes.

Finally, we analysed the relationship between *egl-18* and *hlh-3*. Null mutants for the former gene show 79±3% of *bas-1* expression, while mutants for the latter 41±5%. Double mutants show 33±5%, which is slightly higher than expected by simple addition of phenotype (pV = 0.0355) → **Figure 3.2.21-D**. As *hlh-3* single mutant phenotype is not statistically different form the double mutant phenotype (pV = 0.2885), this is an example of epistasis, where *hlh-3* is epistatic to *egl-18*.

cat-1 is another 5-HT pathway gene whose expression in mutant animals is not completely deleted in some cases. Once again, we observed examples of synergic enhancement between TFs in the regulation this gene. For example, → **Figure 3.2.21-A** shows that *egl-18* null mutants show wild type levels of *cat-1* expression (97±2%) and *hlh-3* null mutants only see its expression partially reduced (81±3%). Double mutants show 29±4% expression in the HSN, which exceeds the sum of phenotypes (pV<0.0001). A similar phenotype is observed in the double mutant *egl-18* null, *sem-4* hypomorph → **Figure 3.2.21-A**. As there is no cross-regulation between them, this phenotype likely reflects synergic enhancement between two parallel pathways. Interestingly, when we analysed the double null mutants between *egl-46*; *egl-18* and *egl-46*; *hlh-3*, synergistic suppression was observed → **Figure 3.2.21-B.** Focusing on the first case, *egl-*

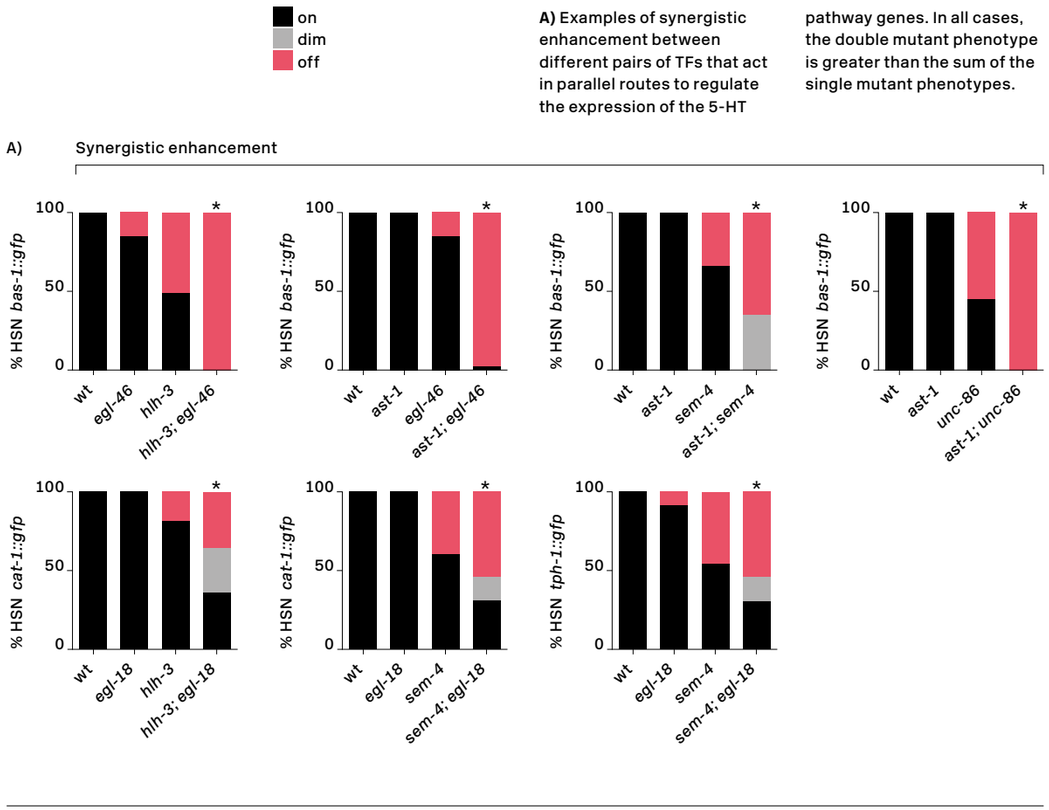
18 mutants show normal levels of *cat-1* (97±2%) and *egl-46* shows a partial phenotype (61±5%) but double mutants rescue *cat-1* expression defects of *egl-46* (94±2% expression in double mutants). Once more, this suggest that in the absence of *egl-18* additional factors could be recruited to the promoters to induce expression.

Lastly, we observed a couple of examples where TFs acted in an independent manner to control *cat-1* expression. This is the case of the double mutants *hlh-3*; *unc-86* and *unc-86*; *sem-4* **→ Figures 3.2.21-C.**

Regarding *tph-1* expression *sem-4(n2654)*; *egl-18(ok290)* double mutants show synergic enhancement **→ Figure 3.2.21.A**. Considering that reporter analysis does not show cross-regulation between them, these results suggest both factors act cooperatively to regulate expression. Interestingly, we find that *sem-4* mutants in combination with *ast-1* mutants, shows antagonistic effects (synergic suppression) **→ Figure 3.2.21-B** suggesting, once again, that in the absence of both factors additional factors are recruited to the regulatory regions of those genes.

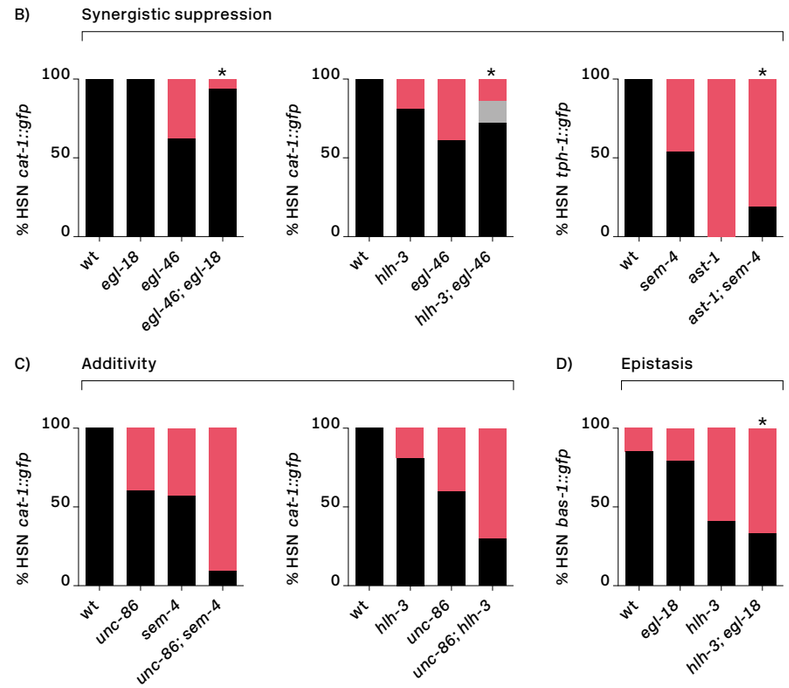
→ Figure 3.2.21.E summarises the different types of genetic interactions observed between TFs of the HSN regulatory code, regardless of the reporter gene analysed. For specific data **→ Annex 3.2.8.**

Figure 3.2.21
Double mutant analysis to assess cooperativity between members of the HSN regulatory code in the regulation of the serotonin pathway genes



A) Examples of synergistic enhancement between different pairs of TFs that act in parallel routes to regulate the expression of the 5-HT pathway genes. In all cases, the double mutant phenotype is greater than the sum of the single mutant phenotypes.

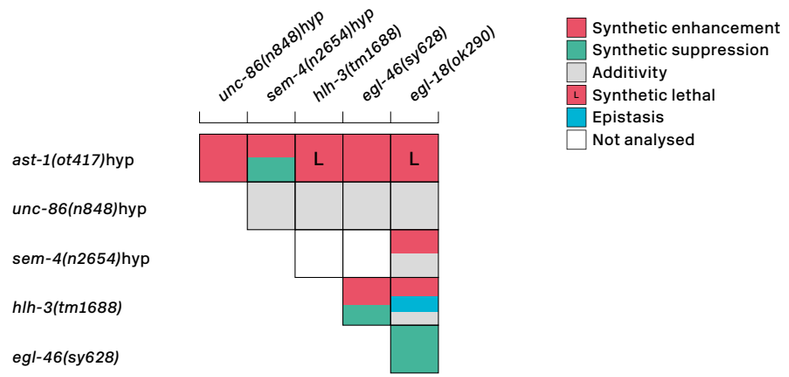
Our double mutant analysis mainly detects synergistic interactions between TFs, suggesting extensive cooperative relationships among them. This cooperativity can explain partial phenotypes observed in the single mutant analysis. Moreover, it gives us a hint of the great complexity of the system, as different combinations of TFs synergise to regulate the expression of some genes but not in others.



B) Examples of synergistic suppression between different pairs of TFs. The double mutant phenotype is smaller than the sum of the single mutant phenotypes.

C) Examples of no cooperativity. The double mutant phenotype can be explained by the sum of the single mutant phenotypes.

D) Example of epistasis, where both TFs act in the same linear pathway. The double mutant *hlh-3, egl-18* phenotype is equivalent to *hlh-3* single mutant.



E) Summary of all the genetic relationships observed in this work, considering those that have not been included in A-D. hyp: hypomorph alelle. n>50 worms in each condition. Fisher exact test, *: pV< 0.05. For specific data and analysis see Annex 3.2.8.

The HSN regulatory signature selects the HSN transcriptome

Chapter III

It is largely unknown why specific regions of the DNA function as active regulatory modules in some cellular contexts. TFs are the main regulators of enhancer function. Each enhancer is bound by specific combinations of TFs that will either activate or repress transcription (Reiter et al. 2017). TFBSs are small and degenerate, thus predicted matches for TFBSs are widely distributed throughout the entire non-coding genome. So far, it is impossible to predict which ones are actually bound by the corresponding TF.

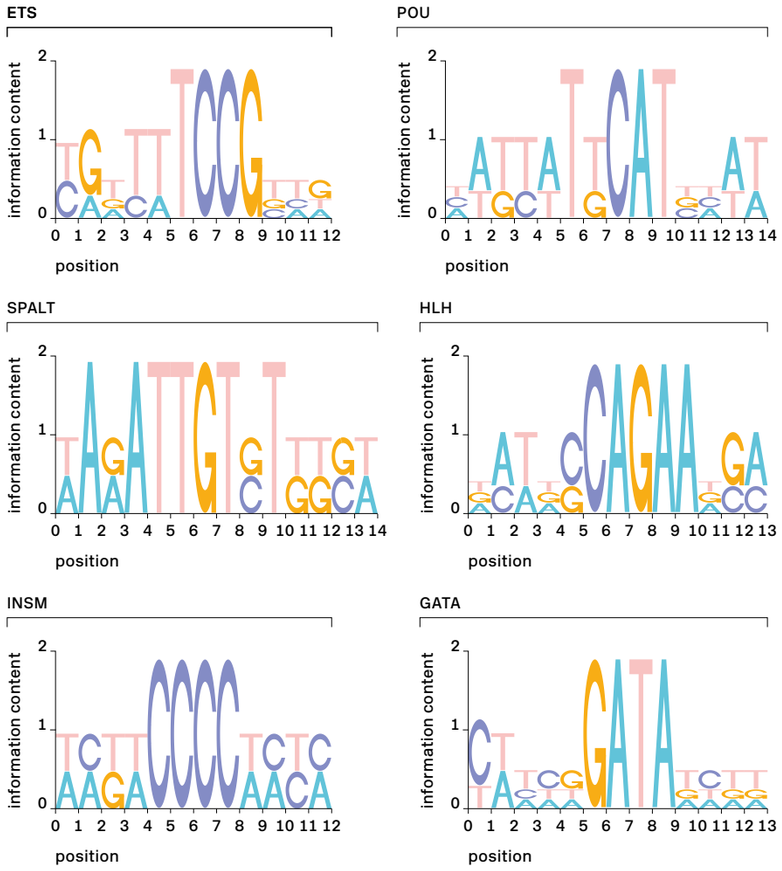
Our results suggest that the HSN regulatory code is required for broad HSN specification (and not only for 5-HT gene expression) and it acts directly on the regulatory regions of their target genes. Since the members of the code belong to six different TF families that recognise very different BSs → **Figures 3.3.1-A**, we wondered whether the clustering of BSs for the HSN regulatory code in putative regulatory regions of HSN expressed genes might confer sufficient specificity to impose a defining regulatory signature.

This part of the project was done in collaboration with Dr. Alejandro Artacho, informatician at the Department of Genomics and Health, in the Centre for Public Health Research (CSISP). Carlos Mora, PhD student, and Dr. Miren Maicas also participated in this part of the project.

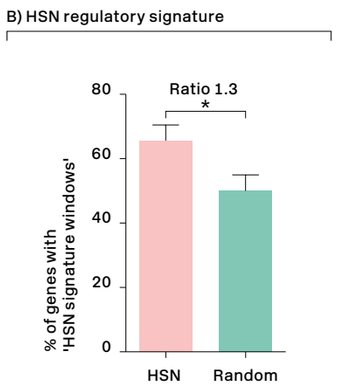
The HSN signature is enriched in regulatory regions of HSN expressed genes.

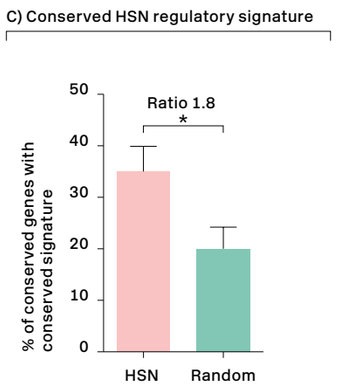
First, based on the functional BSs that we had previously identified → **Figures 3.2.8-3.2.10**, we built PWMs for each of the six TFBSs of the HSN regulatory code → **Figure 3.3.1-A**. There are 96 genes known to be expressed in the HSN (Hobert et al. 2016), excluding panneuronal features, which are regulated in a very redundant manner (Stefanakis et al. 2015) → **Annex 3.3.1**. We analysed upstream and intronic sequences of HSN expressed genes in search of DNA windows (up to 700 bp length) containing at least one PWM match for each of the six members of the code, termed from now on the 'HSN signature' → **Figure 3.3.1-A**. We compared the number of windows that contained the HSN signature in HSN expressed genes with a random set of 100 genes → **Annex 3.3.2**. We realised that known HSN expressed genes contain large upstream and intronic sequences, thus, for comparison purposes, we selected random genes with similar upstream and intronic distribution → **Figure 3.3.1-D** and **E**. CRMs and regulatory enhancers comprise defined DNA regions usually ranging from 50 to 1500 bp. We tried different window sizes and obtained best results, in terms of largest difference between the two sets of genes analysed, using a maximum window size of 700 bp. This value is consistent with our regulatory analysis results in which the longest CRM is 522 bp (*cat-1prom14*), coinciding with the mean size of enhancers described for mouse

Figure 3.3.1
HSN regulatory signature characterisation



A) Position Weight Matrix logos of the HSN transcription factor code calculated from the functional BSs in Figures 3.2.8-3.2.10.

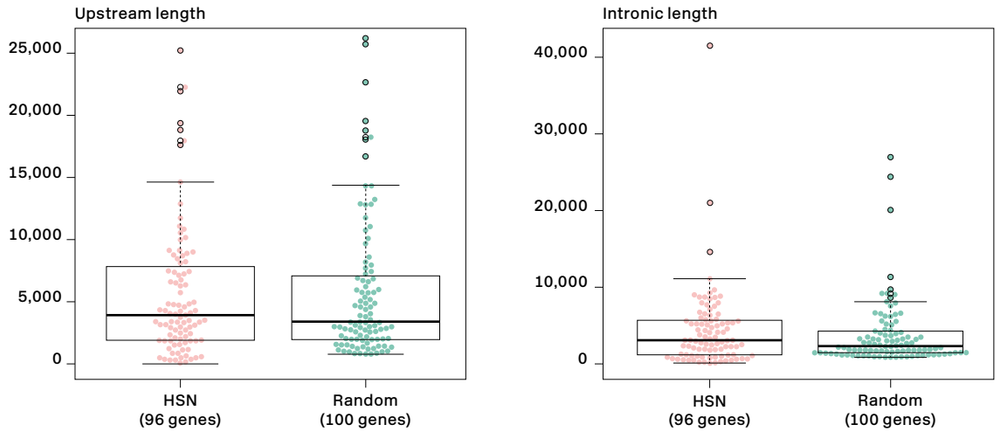




B) Analysis of the number of genes with positive windows for the HSN regulatory signature. 66% of HSN expressed genes contain the HSN regulatory signature compared to only 50% of a comparable random gene set. Statistical significance was calculated using Fisher exact test; *: pV<0.05.

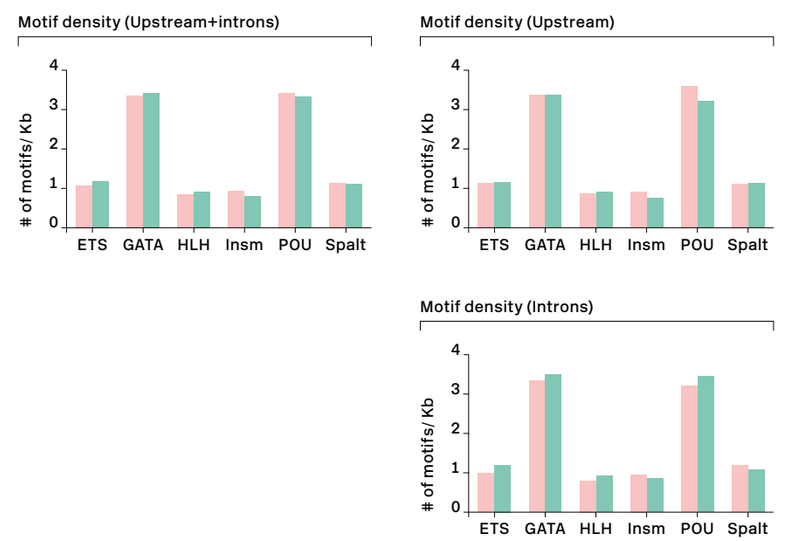
C) Inclusion of the conservation criteria in the HSN signature analysis strongly increases the difference between HSN and random genes. These result indicate that the HSN regulatory signature is enriched in the regulatory regions of HSN expressed genes.

D) Comparison of upstream sequence length between HSN expressed and selected random genes shows no significant differences. Statistical significance was calculated using t test (pV=0.89) and wilcox test (pV=0.9). Similarly, comparison of intronic sequence length between HSN expressed and selected random genes shows not significant differences. Statistical significance was calculated using t test (pV=0.36) and wilcox test (pV=0.36).



E) Comparison of the number of motifs per kilo base, considering upstream regulatory regions, introns, and both together, in both gene lists. Density of TFBSs for the HSN regulatory code (ETS, GATA, HLH, INSM, POU and SPALT BSs) is similar between HSN (pink) and random (green) genes.

These experiments were performed in collaboration with Dr. Alejandro Artacho.



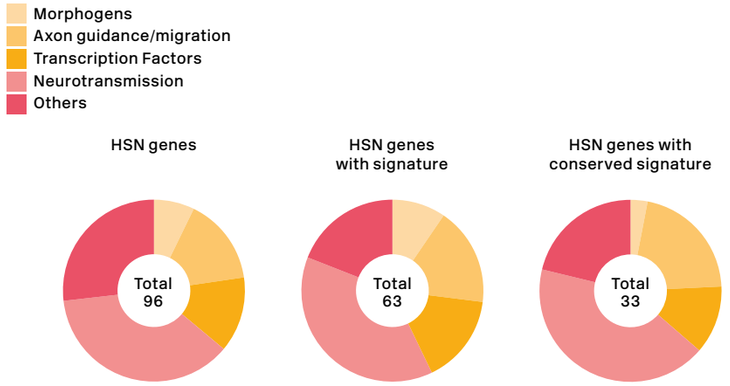


Figure 3.3.2
Functional distribution of the HSN regulatory signature

HSN expressed genes are distributed in five functional categories. No significant difference was found in the distribution of these categories in all HSN expressed genes compared to genes with HSN regulatory signature or HSN genes with conserved signature. Statistical significance was calculated using Chi square test computing p-values by Monte Carlo simulation; pV = 0.59.

embryonic stem cells (Whyte et al. 2013; Parker et al. 2013). We found that a higher percentage of HSN expressed genes (66%) contained the HSN signature compared to the random gene list (50%), being the ratio between them 1.3 (pV<0.05) → **Figure 3.3.1-B**. We wondered if this difference could be due to a higher frequency of TFBSs for some or all members of the HSN regulatory code in the HSN expressed genes in comparison to random genes, or if it is due to specific clustering of the six classes of sites. For this purpose, we compared the number of motifs per kilo base, considering upstream regulatory regions, introns, and both together, in both gene lists. → **Figure 3.3.1-E** clearly shows that there is no difference in the global number of BS matches for each TF found in HSN genes compared to random genes. Thus, we conclude that this difference must be specifically due to the clustering of the different TFBS classes.

Non-coding regulatory regions evolve rapidly, which limits the use of direct multispecies alignment to identify regulatory regions (Villar et al. 2014). In spite of the fast turnover of specific TFBSs, the regulatory logic itself is often conserved among species (Doitsidou et al. 2013; Flames & Hobert 2009; Villar et al. 2014). To analyse whether our HSN signature was also conserved, we performed similar bioinformatics analyses of the regulatory regions in additional *Caenorhabditis* species. Specifically, we selected HSN expressed genes that had orthologues in at least two additional species of the *Caenorhabditis* genus (from the *C. brenneri*, *C. remanei*, *C. briggsae* and *C. japonica* genomes). We considered the HSN signature as phylogenetically conserved when orthologous genes in all species displayed the signature within their upstream or intronic regions. We found that the inclusion of the conservation criteria in this analysis strongly increased the difference between HSN and random genes; i.e. HSN expressed genes contain almost twice the number of windows with HSN conserved signature than random genes (ratio 1.8; pV<0.05) → **Figure 3.3.1-C**. This suggests that the HSN signature is used by the HSN regulatory code to select the genes expressed in the neuron, and thus is strongly preserved in evolution.

Our results indicate that the HSN signature is only found in a subset of HSN expressed genes (64% in total HSN genes and 33% in HSN conserved genes). We reasoned that maybe only certain genes that develop a specific function or participate in a specific biological process were the ones containing the HSN signature. Hence, we explored signature distribution across gene categories. We divided the 96 genes expressed in the HSN into five groups: morphogen signalling (components of the Wnt, Notch pathway, etc.), axon guidance and migration, TFs, neurotransmission and others (terminal features, synaptogenesis and extracellular matrix components) → **Figure 3.3.2**, → **Annex 3.3.1**. The category that accounts for the highest percentage of genes known to be expressed in HSN is neurotransmission (37%), followed by 'others' (28%), axon guidance and migration (15%), TFs (13%) and morphogen signalling (7%). Next we did the same only with the 63 HSN genes that contain the HSN signature and with the 33 genes that contain conserved HSN signature. In both cases we saw similar distribution of categories. Therefore, it seems that HSN genes with the HSN regulatory signature were equally distributed across functional categories compared to all HSN expressed genes → **Figure 3.3.2**. This suggests that the HSN regulatory code acts broadly upon the HSN transcriptome and does not select specific functional subsets of genes.

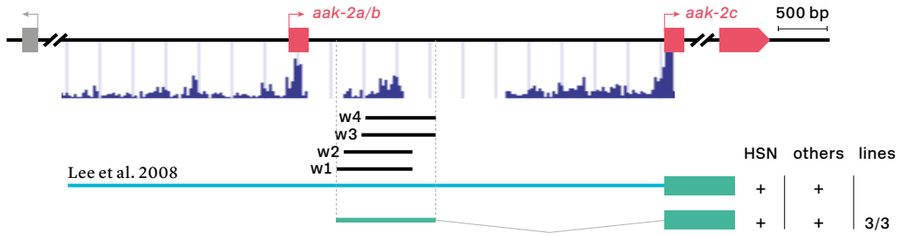
Our hypothesis is that the HSN signature selects HSN expressed genes. If this is true, then the identified windows should correspond to functional enhancers. We tested HSN signature windows from five genes by fusing PCR amplified HSN windows to *gfp* and injecting them into N2 worms. *In vivo* reporter analysis confirmed that they drive expression in the HSN neuron (4 out of 5 genes tested, → **Figure 3.3.3, → Table 3.3.1**.) Of note, GFP was not exclusively expressed in the HSN. Similar to our *cis*-regulatory analysis of the 5-HT pathway genes, we noticed that *C. elegans* functional HSN signature windows do not show a high level of sequence conservation, which is in agreement with rapid evolution of regulatory sequences.

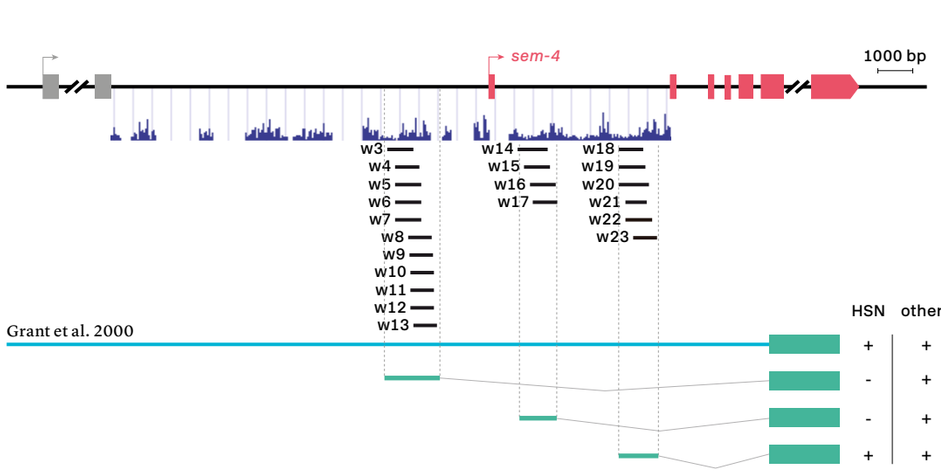
The HSN signature allows *de novo* identification of HSN expressed genes

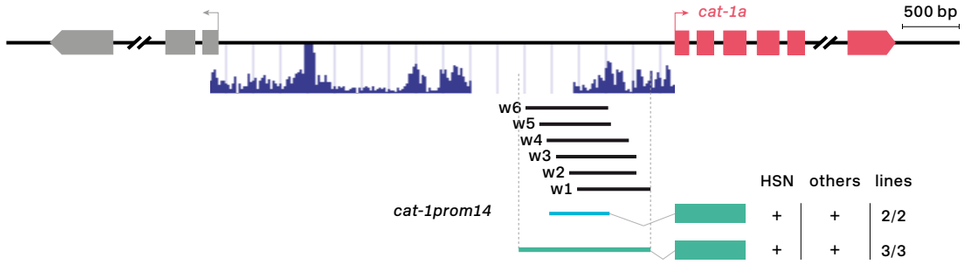
Next, we aimed to identify new genes expressed in HSN based solely on the presence of the HSN signature. First, using the same strategy as before, we examined the distribution of the HSN signature windows across the entire *C. elegans* genome. We classified the genome in two groups: neuronal genes and non-neuronal genes. The first group, to which we subtracted the previously mentioned 96 HSN expressed genes → **Annex 3.3.1**, consists of 1.839 genes. The second group corresponds to the remaining 18.786 protein coding genes from *C. elegans* genome (Hobert 2013, www.wormbase.org). Remarkably, the HSN signature is preferentially found in the putative regulatory sequences of genes known to be expressed in neurons or that have a neuronal function, compared to the rest of the genome (ratio 1.7), as would be expected from putative HSN expressed genes → **Figure 3.3.4-A**. As before, filtering of conserved signatures strongly increased the difference between 'neuronal' and 'non-neuronal' genomes, which adds support to its functionality (ratio 2.5) → **Figure 3.3.4-B**. Moreover, Gene Ontology analysis of all genes in the *C. elegans* genome with the HSN signature revealed enrichment of processes characteristic of HSN differentiation and function → **Figure 3.3.4-C** and **D**. For example, we found that regulation of locomotion, positive regulation of transcription and regulation of cell differentiation are the most significantly enriched processes → **Figure 3.3.4-C**, whereas more than 200 and more than 100 genes are associated to G-protein coupled receptor signalling and oviposition, respectively → **Figure 3.3.4-D**. The main function of the HSN neuron is to regulate the egg-laying behaviour (Desai et al. 1988), also HSN regulates muscle contraction and thus is considered a motorneuron, which will correlate with locomotion as GO term, (Collins et al. 2016).

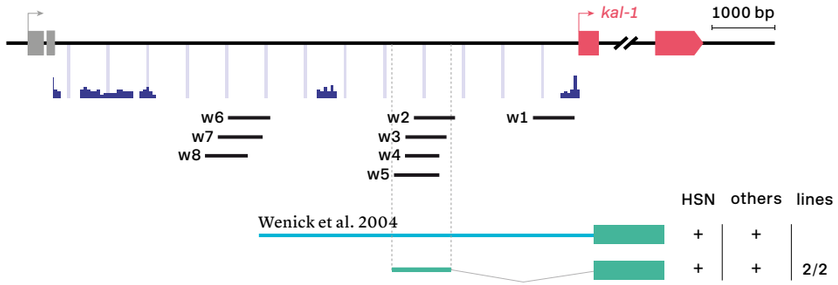
Figure 3.3.3
Validation of the functionality of HSN regulatory signature windows

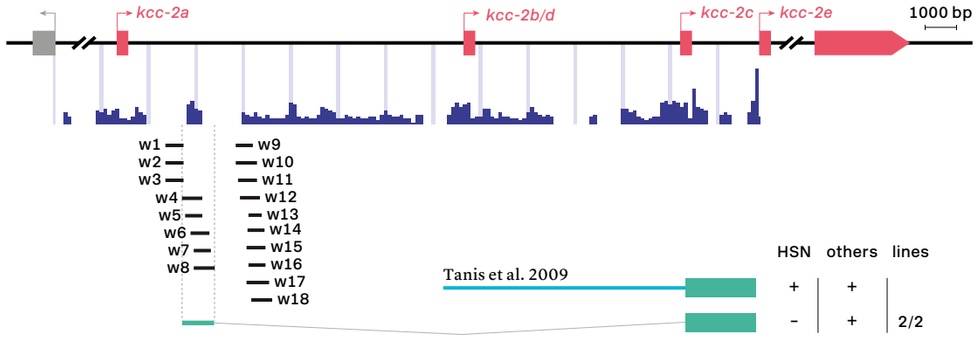
In vivo reporter fusion analysis to test functionality of HSN signature windows in HSN expressed genes. Black lines represent the coordinates covered by bioinformatically predicted HSN signature windows (indicated by 'w' and a number). Light blue lines indicate already published reporter constructs. Green lines indicate the region used in our analysis. Dark blue bar profiles represent sequence conservation in *C. briggsae*, *C. brenneri*, *C. remanei and C. japonica*. See Table 3.3.1 for a list of all reporters tested.











Having characterised the HSN signature in the whole *C. elegans* genome, we attempted to identify *de novo* genes expressed in the HSN. To this end, we randomly selected 35 neuronal genes with a conserved HSN regulatory signature and generated transgenic reporter lines containing the predicted HSN signature → **Table 3.3.1,** → **Table 2.21**. As a control, we randomly picked 10 similar-sized intergenic regions of neuronal genes lacking the HSN signature → **Table 3.3.1**. We found that 13 out of the 35 constructs (37%) showed GFP expression in HSN, while none of the controls led to reporter expression in this cell → **Figure 3.3.5-A-D**, → **Table 3.3.1**. We considered positive reporters those that fulfilled any of these criteria: 1) expression in at least two independent lines in >10% of HSN, or 2) expression in one independent line in >20% of HSN cells. Importantly, all reporter constructs, including the negative controls, did drive GFP expression in other neurons → **Figure 3.3.6,** → **Table 3.3.1**. Our results reveal that the presence of the HSN signature can be successfully used to *de novo* identify HSN expressed genes in more than one third of the cases.

HSN functional enhancers exhibit a distance bias in relation to the start codon

Next, we tried to identify any defining characteristic of the HSN expressed windows (or functional enhancers). Multiple BSs for the same TF, also known as homotypic clusters of TFBSs, are statistically enriched in proximal promoters and distal enhancers and have been shown to enhance gene expression (Markstein et al. 2002; Lifanov 2003). Conservation of such site clusters between vertebrate and invertebrates suggests that homotypic clustering could be a general organisation principle of *cis*-regulatory regions (Gotea et al. 2010). Thus, we calculated the number of motifs per kilo base of the different members of the HSN regulatory code in the HSN signature windows. We did not observe differences in motif frequency between the 13 HSN signature windows that are expressed in the cell and the 22 that are not expressed in the cell → **Figure 3.3.5-E**. We also did not observe any difference in the mean size of the HSN signature windows that were expressed in HSN (756bp ± 47.22) from those that were not expressed in the neuron (817 bp ± 28.62) → **Figure 3.3.5-F**. In addition, HSN expressed windows showed a mean GC content of 38 ± 1%, which was not significantly different from the 36 ± 1% GC content of the non-expressed windows → **Figure 3.3.5-G**. It is known that enhancers can be found at distances ranging from hundreds of bases to megabases from the transcription starting site (TSS) (Bulger & Groudine 2011). We wanted to check if there could be any distance bias in the location of the HSN signature expressed windows relative to the start codon of the gene assigned to the window. Distance was calculated independently of the sign; in other words, with independence of being upstream or in intronic regions downstream of the ATG of the gene. We saw that functional HSN signature windows are located significantly closer to the ATG of the gene compared to windows that are not expressed in the HSN. HSN expressed signature windows are found at a mean distance of 3.3 kb, while those that are not expressed are found at a mean distance of 6.7 kb → **Figure 3.3.5-H**. 11/13 HSN functional enhancers are found 3.3 kb away or closer to the start codon of the gene.

The HSN signature contains syntactic rules

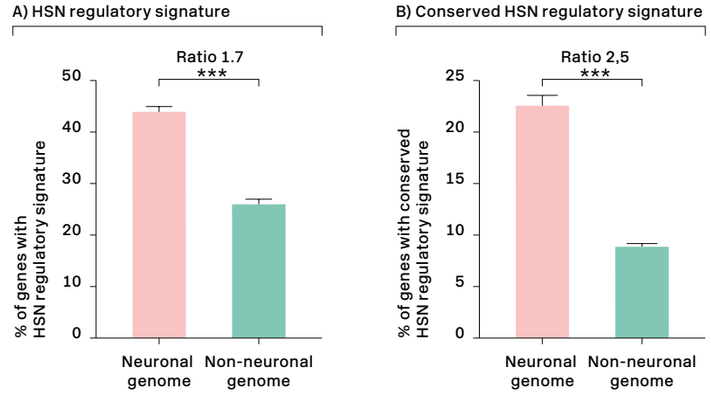
Motif positioning, often referred to as 'syntax' or 'grammar', is the relative order, orientation, spacing and helical phasing of TFBSs within an enhancer. Motif positioning typically ensures that TFs are arranged appropriately to facilitate PPIs and thereby promote cooperative binding, as well as the recruitment of cofactors and the transcriptional

Figure 3.3.4
Distribution of the HSN regulatory signature in *Caenorhabditis elegans* genome

A) HSN regulatory signature is enriched in neuronal genes compared to the non-neuronal genome. Statistical significance was calculated using Chi square with Yates correction. *: pV<0.0001.

B) Inclusion of the conservation criteria in the HSN signature analysis strongly increases the difference between neuronal and non-neuronal genome. *: pV<0.0001.

These experiments were performed in collaboration with Dr. Alejandro Artacho.



C)-D) Gene ontology analysis of genes with HSN regulatory signature. p values and number of genes corresponding to the biological processes enriched in the genes with HSN regulatory signature are represented in (C) and (D), respectively.

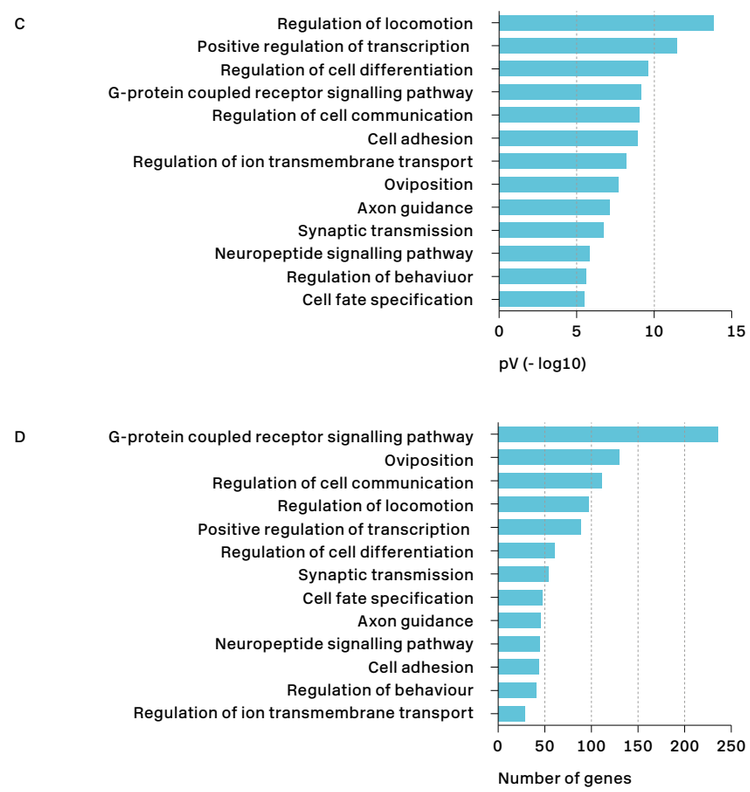
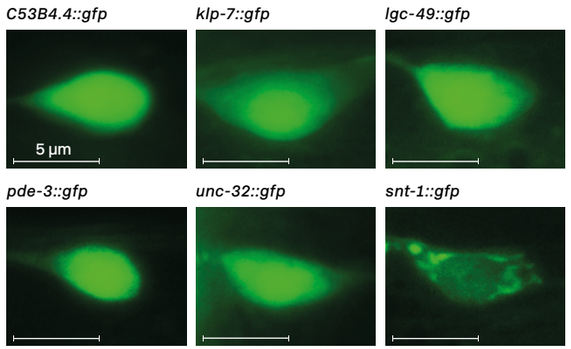


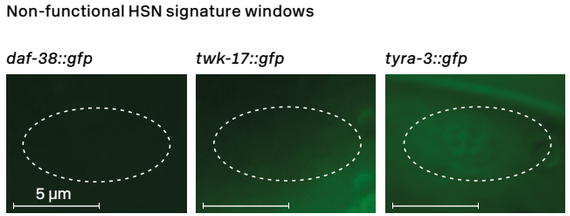
Figure 3.3.5
HSN regulatory signature can be used to identify *de novo* HSN expressed genes

A



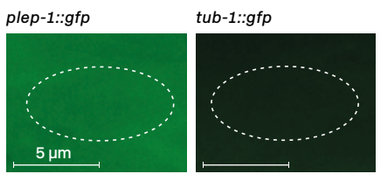
A) Micrographs showing representative examples of *de novo* identified HSN active enhancers; i.e. HSN signature positive windows fused to *gfp* that drive expression in the HSN neuron.

B



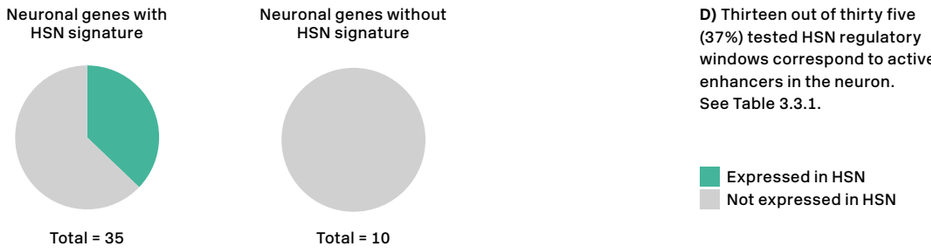
B) Micrographs showing representative examples of false positive enhancers of the HSN; i.e. HSN signature positive windows fused to *gfp* that do not drive expression in the HSN neuron.

C



C) Micrographs showing representative examples of negative controls; i.e. reporter fusions of windows without HSN regulatory signature that do not drive *gfp* expression in the HSN neuron.

D



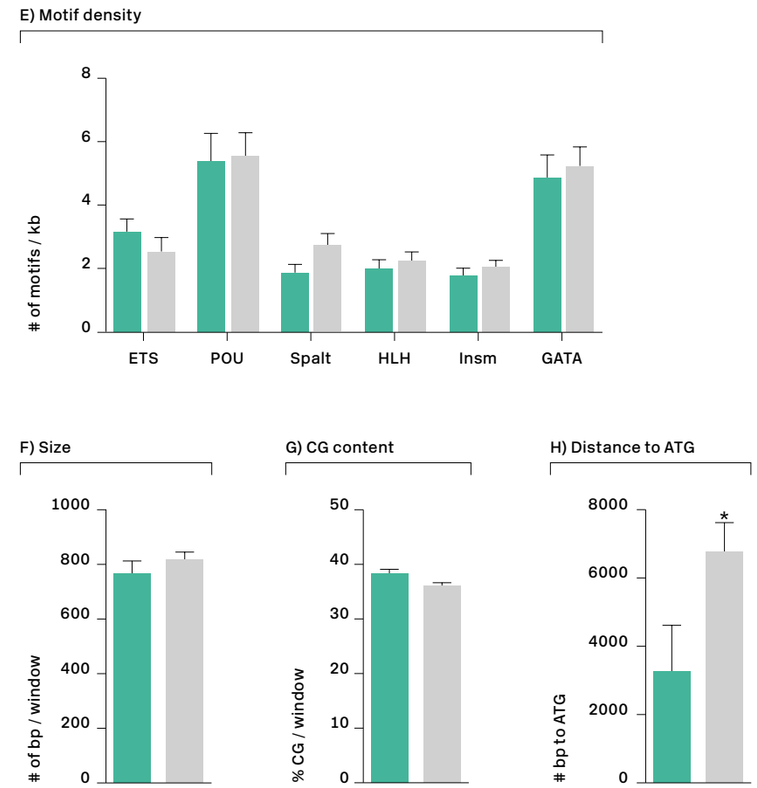
D) Thirteen out of thirty five (37%) tested HSN regulatory windows correspond to active enhancers in the neuron. See Table 3.3.1.

machinery. However, it is still a matter of debate if syntax rules do play a role in enhancer function as few examples have been reported.

From our *cis*-regulatory analysis we know that the HSN regulatory code acts in a flexible manner, as it can activate enhancers with a variable distribution and order of TFBSs. Moreover, we observed that, in some genomic contexts, the absence of BSs for certain TFs can be compensated by the rest of the TF code → **Figure 3.2.10**. This would suggest that the HSN regulatory code follows the Billboard model for enhancer function (Kulkarni & Arnosti 2003). However, we decided to use our HSN regulatory window analysis to try to identify syntax rules governing HSN enhancer functionality.

Thus, we explored if a particular grammar could be present in functional HSN signature windows compared to non-functional ones. We failed to find any preferential TFBS arrangement, similar to our 5-HT pathway gene *cis*-regulatory analysis and in agreement with the Billboard model. However, bioinformatic analysis (iTF software (Kazemian et al. 2013)) revealed particular biases in specific orientations between TF pairs. For instance, we found that ETS BSs show a statistically significant bias for 3' to 3' orientations with GATA BSs, and a 5' to 5' disposi-

Figure 3.3.5
HSN regulatory signature can be used to identify *de novo* HSN expressed genes

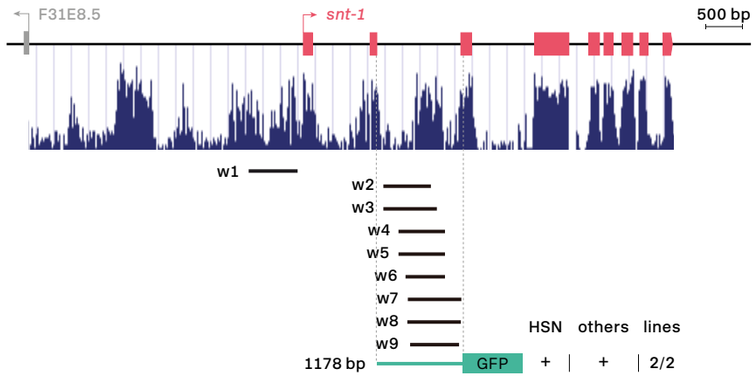


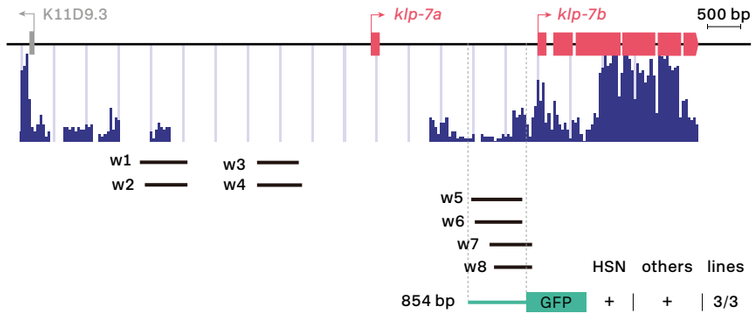
E) Comparison of the number of motifs per kilobase and per gene, between expressed and non-expressed HSN signature positive windows. Statistical significance was calculated using the Unpaired t test (ETS: pV=0.3499; POU: pV=0.8825; SPALT: pV= 0.1123; HLH: pV=0.5763; INSM pV=0.4265; GATA: pV=0.7044).

F) Comparison of the mean size of expressed and non-expressed HSN signature positive windows. Statistical significance was calculated using the Unpaired t test (pV=0.3333).

G) Comparison of the mean GC content in expressed and non-expressed HSN signature positive windows. Statistical significance was calculated using the Unpaired t test (pV=0,2421).

H) Comparison of the mean distance to the start codon between expressed and non-expressed HSN signature positive windows. Statistical significance was calculated using the Unpaired t test (pV=0,0275).





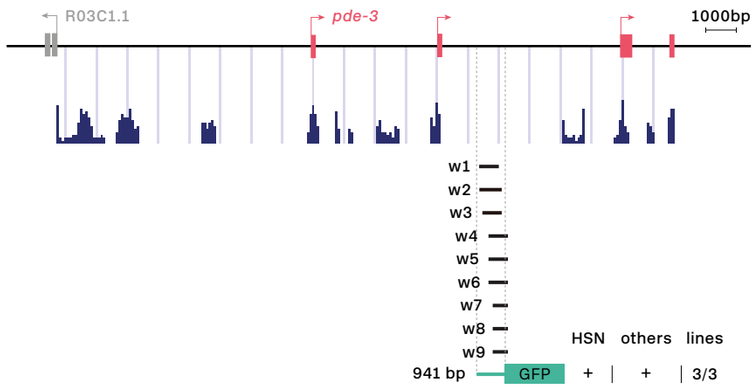
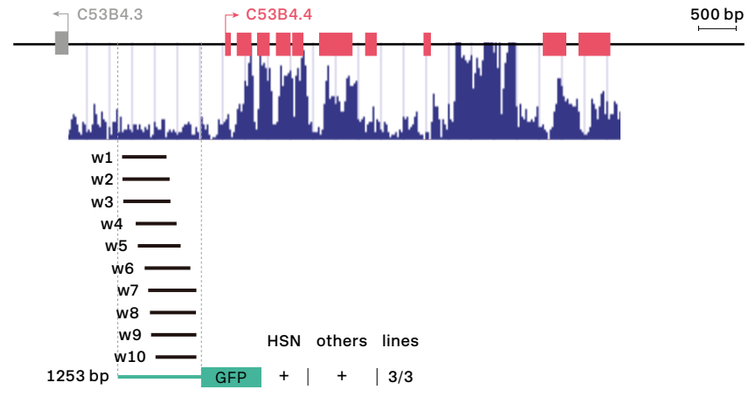
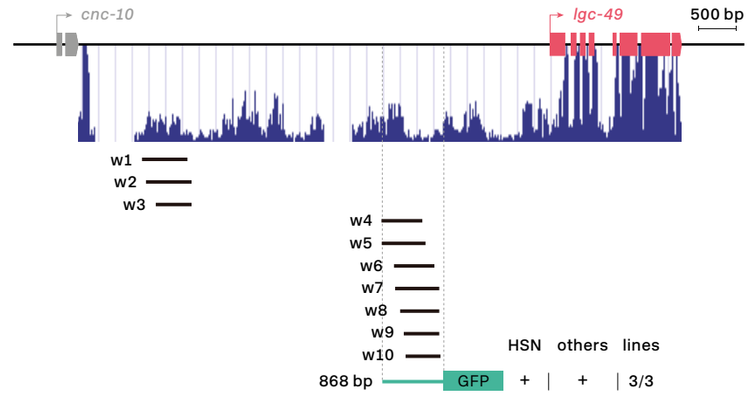


Figure 3.3.6
Representative examples of *de novo* identified HSN active enhancers

Black lines represent the coordinates covered by bioinformatically predicted HSN signature windows (indicated by 'w' and a number). Green lines indicate the region used in our analysis. Dark blue bar profiles represent sequence conservation in *C. briggsae, C. brenneri, C. remanei and C. japonica*. See Table 3.3.1 for a list of all reporters tested.





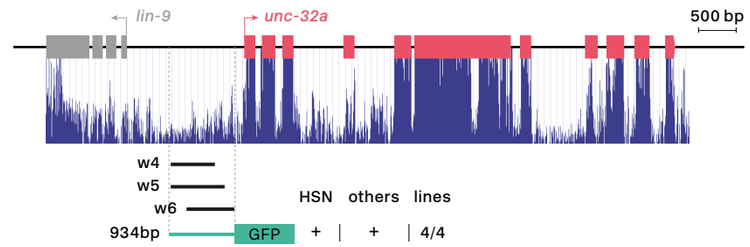


Figure 3.3.7
HSN regulatory signature contains syntactic rules

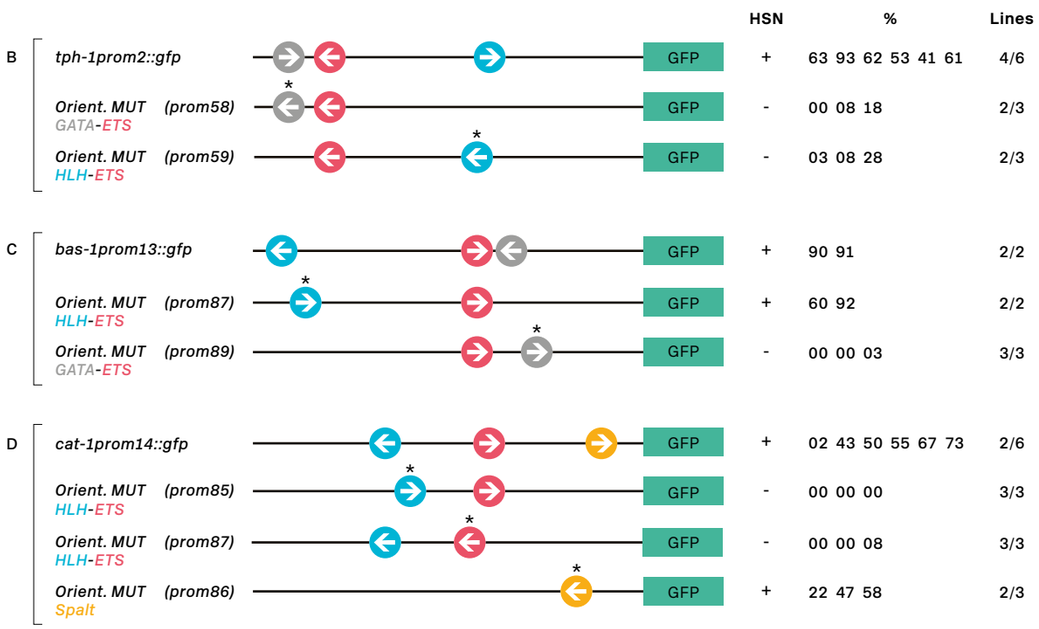
A

A) Transcription factor binding site orientation bias

HSN regulatory windows of HSN expressed genes show statistically significant biases in the orientations of ETS BSs with GATA and HLH BSs.

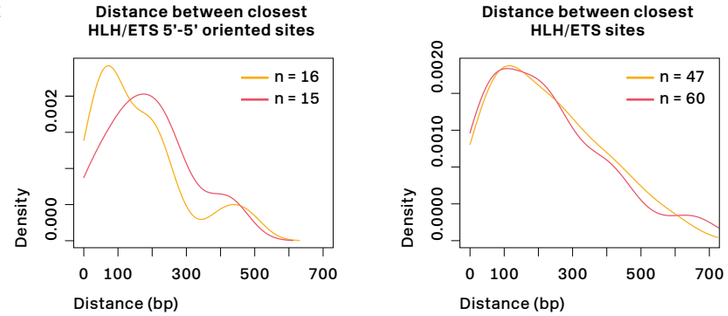
B)-D) Functional transcription factor binding site orientation bias in HSN CRMs

Experimentally identified minimal HSN CRM for *tph-1*, *cat-1* and *bas-1* show examples of TFBS orientation biases between GATA/ETS and HLH/ETS. Each number represents the % of GFP cells in a particular transgenic line. +: values rank between 100 to 60% of mean wild type construct expression; +/−: values indicate 20-60 % lower penetrance than mean wild type expression; −: values are less than 20% of mean wild type values. n > 30 worms per line. Disruption of the original TFBS orientation without affecting TFBS sequence ***per se*** produces defects in *gfp* expression. As a negative control, change in orientation of a SPALT BS in *cat-1prom14* does not affect expression. Arrows indicate the orientation of the BS. In each construct, the TFBS for which the orientation has been changed is marked with an asterisk. See Annex 3.3.3 for specific nucleotide changes.



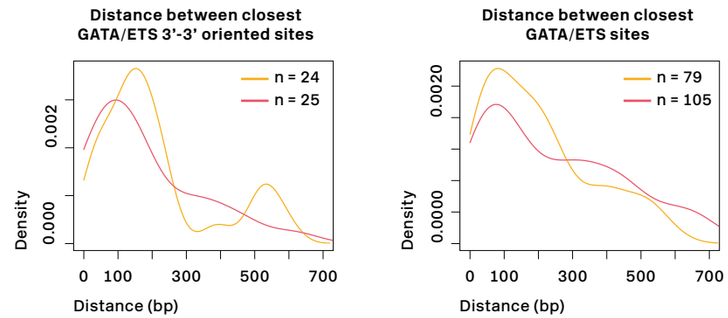
E)-F) Transcription factor binding site distance bias

Kernel density plots representing the distance between closest HLH-ETS (E) and GATA-ETS (F) BSs. Left graphs consider only ETS-HLH 5'-5' and GATA-ETS 3'-3' oriented sites and right graphs show distances between closest TFBS pairs in all orientations. 'n' indicates total number of motif pairs.



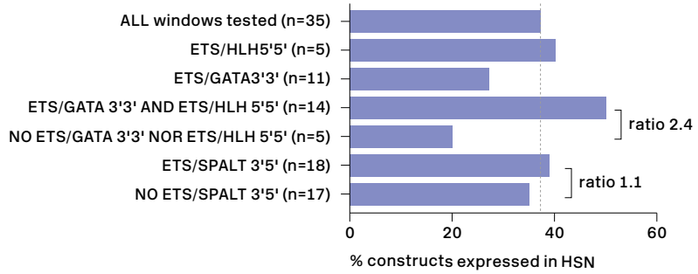
signature windows with HSN expression

signature windows without HSN expression



G) Orientation bias in HSN signature windows

37% tested HSN regulatory windows show expression in the HSN. Windows with both GATA/ETS 3'3' and HLH/ETS 5'5' syntax show a higher HSN expression rate compared to windows without these syntactic rules (50% and 20% respectively). This differential expression does not occur when considering non-functional syntactic rules such as specific SPALT/ETS orientations.



tion bias in relation to HLH BSs → **Figure 3.3.7-A**. These orientation biases were not found in HSN signature windows from the non-neuronal genome. Interestingly, these specific TF-TF dispositions were also found in our experimentally identified *tph-1*, *cat-1* and *bas-1* CRMs. Starting with *tph-1* CRM (*tph-1prom2*), we found an example of both orientation biases → **Figure 3.3.7-B**. Experimental rearrangement of the GATA BS from 3' to 5' orientation (*tph-1prom58*), led to a complete loss of GFP expression in the cell. The same happened when we altered the 5' orientation of the HLH site to 3' (*tph-1prom59*). Next, we also found both overrepresented motif arrangements in the *bas-1* CRM (*bas-1prom13*) → **Figure 3.3.7-C**. Rearrangement of the HLH site from 5' to 3' orientation had no effect in GFP expression (*bas-1prom87*). Flipping the GATA BS from 3' to 5' orientation *(bas-1prom89),* however, did provoke complete loss of GFP expression in the cell. Finally, we found one more example of the HLH-ETS 5'-5' overrepresented pair in the *cat-1* CRM (*cat-1prom14*) → **Figure 3.3.7-D**. Altering the orientation to 3'-5' (*cat-1prom85*) or to 5'3' (*cat-1prom87*) leads to a loss of GFP expression in the HSN.

Our results indicate that these syntactic rules are required in some contexts, although they are not absolutely necessary for enhancer functionality. Moreover, syntactic restrictions seem to be specific to some TFBS pairs, as we did not observe a statistical enrichment in other TF pairs, nor a phenotype when the SPALT BS in the *cat-1* CRM was flipped (*cat-1prom86*) → **Figure 3.3.7-D**. Specific BSs rearrangements are listed in → **Annex 3.3.3**.

We also explored the possibility that these TF pairs (HLH-ETS and GATA-ETS), in addition to showing an orientation bias, could also exhibit a distance bias. To this end, we compared the distance distribution frequencies between the closest HLH-ETS and ETS-GATA motif pairs, in HSN functional enhancers and in signature windows that were not expressed in the cell. → **Figure 3.3.7-E** shows that the HLH-ETS pair shows a different distance relationship depending on the group (HSN expressed windows vs not expressed in HSN). This difference is not appreciable or less pronounced when orientation relationships are not considered. The same is true for the ETS-GATA pair → **Figure 3.3.7-F**.

Having proven the functionality of the orientation syntactic rules in our CRMs, we explored the possibility that the presence of these rules could allow discrimination between functional and non-functional HSN signature windows. We found that, from our 35 tested HSN signatures, constructs in which both ETS/GATA and ETS/HLH syntactic rules were obeyed were more likely to be expressed in HSN compared to constructs that do not show neither these TFBS dispositions (50% compared to 20% HSN expression respectively; ratio 2.4). This difference was not observed when the ETS-SPALT motif pair was considered (ratio 1.1) → **Figure 3.3.7-G**.

In conclusion, we have shown that the HSN signature is flexible but obeys specific syntactic rules, which supports the TF collective model. Syntactic rules improve the probability to *de novo* identify functional enhancers. Enhancer distance from the starting codon and distance between TF motifs with the overrepresented orientation can also be used as a guide to distinguish between functional and non-functional enhancers. However, the presence of the HSN signature, even with correct syntactic rules, is not sufficient in all cases to induce HSN expression. This observation suggests that additional factors (either activating or repressing TFs or chromatin remodelers) might be also involved.

Table 3.3.1
***In vivo* reporter fusion analysis to identify *de novo* HSN active enhancers**
→

Gene name	HSN signature	Reported HSN expression	HSN expression	% HSN expression	Other neurons	Other cells	Lines
abts-4	Yes	No	-	0, 0, 0	+	-	0/3
acr-24	Yes	No	-	0, 0, 0	+	+	0/3
ast-1	Yes	No	+	38, 47, 48	+	-	3/3
bam-2	Yes	No	-	0, 0, 0	+	+	0/3
C16B8.4	Yes	No	-	0, 0	+	+	0/2
C53B4.4	Yes	No	+	85, 89, 90	+	+	3/3
ckr-2	Yes	No	-	0, 0, 0	+	+	0/3
daf-38	Yes	No	-	0, 0, 0	+	+	0/3
dgn-1	Yes	No	-	0, 0, 0	+	-	0/3
F32D8.10	Yes	No	-	0, 0, 0	+	+	0/3
F37A8.5	Yes	No	+	37, 63, 65	+	+	3/3
fut-1	Yes	No	-	0, 0, 0	+	+	0/3
gab-1	Yes	No	-	0, 0, 0	+	+	0/3
glb-20	Yes	No	-	0, 0, 0	+	+	0/3
kcc-1	Yes	No	-	0, 0, 0	+	+	0/3
kel-8	Yes	No	-	0, 0, 0	+	-	0/3
klp-7	Yes	No	+	82, 85, 86, 92	+	+	4/4
lgc-49	Yes	No	+	36, 52, 60	+	+	3/3
mec-10	Yes	No	+	56, 68, 74	+	+	3/3
mgl-2	Yes	No	+	93, 93, 95	+	+	3/3
npr-1	Yes	No	+	43, 51, 70	+	+	3/3
npr-3	Yes	No	-	0, 0, 12	+	+	2/3
pan-1	Yes	No	+	78, 98	+	+	2/2
pde-3	Yes	No	+	2, 13, 50	+	+	2/3
shl-1	Yes	No	-	0, 0, 0	+	+	0/3
snt-1	Yes	No	+	0, 14, 48	+	-	2/3
sprr-1	Yes	No	+	5, 8, 46	+	+	1/3
sto-5	Yes	No	-	0, 0, 0	+	+	0/3
tiam-1	Yes	No	-	0, 0, 0	+	+	0/3
tkr-2	Yes	No	-	0, 0, 0	+	+	0/3
tol-1	Yes	No	-	0, 0, 0	+	+	0/3
twk-17	Yes	No	-	0, 0, 0	+	+	0/3
tyra-3	Yes	No	-	0, 0, 0	+	+	0/3
unc-32	Yes	No	+	41, 53, 71, 87	+	+	4/4
unc-7	Yes	No	-	0, 0, 0	+	+	0/3
Controls							
aak-2	Yes	yes	+	72, 83, 87	+	+	3/3
cat-1	Yes	yes	+	83, 86, 93	+	-	3/3
kal-1	Yes	yes	+	56, 72, 81	+	+	3/3
kcc-2	Yes	yes	-	0, 0, 0	+	-	0/3
sem-4	Yes	yes	+	72, 81	+	+	2/2
F16G10.5	No	No	-	0, 0, 0	+	+	0/3
flp-27	No	No	-	0, 0, 0	+	+	0/3
gipc-2	No	No	-	0, 0, 0	+	+	0/3
irld-53	No	No	-	0, 0, 0	+	+	0/3
irld-62	No	No	-	0, 0, 0	+	+	0/3
lurp-2	No	No	-	0, 0, 0	+	+	0/3
plep-1	No	No	-	0, 0, 0	+	+	0/3
slc-28.1	No	No	-	0, 0, 0	+	+	0/3
stg-1	No	No	-	0, 0, 0	+	+	0/3
tub-1	No	No	-	0, 0, 0	+	+	0/3

Deep homology in the genetic programme regulating serotonergic differentiation

Chapter IV

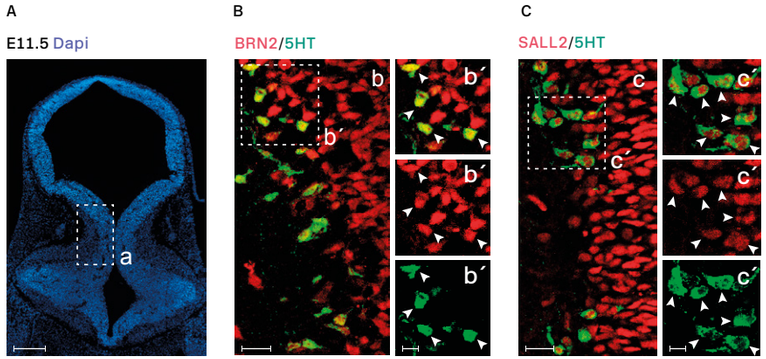
Mouse serotonergic differentiation has been extensively studied, as described in the Introduction. In this work we have identified six members of the HSN TF collective that regulate serotonergic specification in the HSN neuron. Mouse orthologues for several of these members are known regulators of mammalian serotonergic differentiation, arising the question of whether mice and nematodes could share a phylogenetically conserved serotonergic regulatory programme. This type of phylogenetic conservation between *C. elegans* neurons subtypes and more complex organisms has been previously shown for the dopaminergic system in the mouse olfactory bulb (Flames & Hobert 2009; Doitsidou et al. 2013), for a subpopulation of glutamatergic neurons in the mouse hippocampus and inferior olive (Serrano-Saiz et al. 2013), for cholinergic neurons of *Ciona intestinalis* (Kratsios et al. 2011) and midbrain GABAergic neurons (Gendrel et al. 2016; Kala et al. 2009; Lahti et al. 2016). In this final Chapter participated Dr. Laura Chirivella and Dr. Isabel Reillo, who performed expression pattern analysis in mouse tissue; Ángela Jimeno and Miren Maicas, who helped to characterise new regulators of the HSN; and Dr. Alejandro Artacho and Carlos Mora, who carried out all the bioinformatics.

HSN and mouse serotonergic neuron differentiation are controlled by homologous regulatory programmes

Mouse orthologues for four out of the six TFs of the HSN TF collective are known regulators of mammalian serotonergic neuron specification that act at different stages of the pathway: ASCL1 (bHLH TF orthologue to HLH-3), GATA2 and GATA3 (orthologue factors to EGL-18), INSM1 (Zn Finger TF orthologue to EGL-46) and PET1 (ETS TF orthologue of *ast-1*) (see Introduction for a detailed explanation of the role of these factors) → **Figure 1.9.** Additionally, BRN2 (also known as POU3F2, a POU TF from the same family than UNC-86) has been recently associated with serotonergic neuron specification (Nasu et al. 2014). However, this paper focused on the role of BRN2 in maternal behaviour during pup retrieval and did not assess its expression in serotonergic neurons. The effect observed could be due to a very early event in the serotonergic lineage or could even be non-cell autonomous. Thus, we analysed BRN2 protein expression in mouse hindbrain at E11.5, when mouse serotonergic differentiation occurs (Pattyn et al. 2003) → **Figure 3.4.1**. Double fluorescence immunohistochemistry against BRN2 and 5-HT reveals that this TF is expressed in serotonergic neurons → **Figure 3.4.1-B**. BRN2 expression is observed both in progenitors (closer to the ventricle) and in differentiating serotonergic neurons, although not in posterior developmental stages. This could be indicating that the TPH2 staining defect observed in mice with a truncated version of BRN2 could be a cell-autonomous phenotype.

No SPALT TF (TF family of SEM-4) is known to play a role in serotonergic specification. → **Figure 3.4.2** shows the phylogenetic relationship between mouse and worm TFs (EMBL-EBI TreeFam soft-

Figure 3.4.1
BRN2 and SALL2 expression in mouse raphe serotonergic neurons



A) Micrograph of mouse embryonic day 11.5 hindbrain coronal section with DAPI staining. Square box indicates the region in a, b' and c' panels. Scale bar represents: 100 µm.

B) BRN2 and serotonin co-staining. BRN2 is expressed in progenitors and differentiating serotonergic neurons. Arrowheads indicate double labelled cells. Scale bar represents: 20 µm.

C) SALL2 and serotonin co-staining. SALL2 is expressed in progenitors and differentiating serotonergic neurons. Arrowheads indicate double labelled cells. Scale bar represents: 20 µm.

These experiments were performed by Dr. Laura Chirivella and Dr. Isabel Reillo.

ware; (Ruan et al. 2008)). Phylogenetic analysis reveals that SALL2 is more closely related to SEM-4 that any other member of the mouse TF family → **Figure 3.4.2-A**. Hence, we analysed first if SALL2 could have a homologous role in serotonergic specification in the mouse. As with BRN2, we assessed SALL2 expression in mouse serotonergic neurons at embryonic stage E11.5. Similarly, we found that it is expressed in progenitors and differentiating serotonergic neurons → **Figure 3.4.1-C**, suggesting it could also be involved in mouse serotonergic specification.

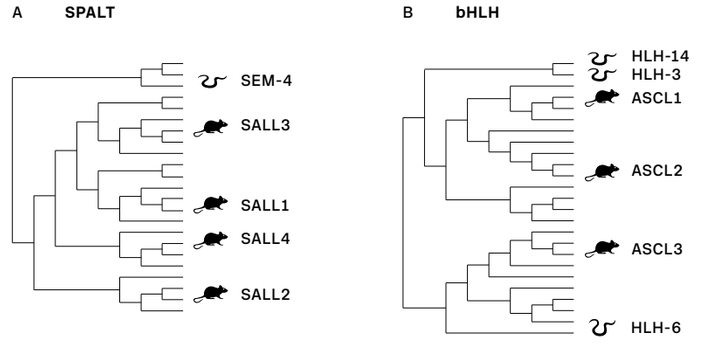
Looking at the mouse serotonergic regulatory programme from the opposite perspective, we realised that two transcription factors FOXA2 and LMX1B, belonging to the forkhead (FKH) and LIM-homeodomain (LIM-HD) TF families respectively, had no orthologous member in the HSN TF collective. As RNAi in *C. elegans* is a quick strategy to get insights in gene functions we decided to carry out an RNAi screen against all of the members of the FKH and LIM-HD TF family. *C. elegans* FKH family is composed by 18 members: ATTF-4, C34B4.2, DAF-16, FKH-2, FKH-3, FKH-4, FKH-5, FKH-6, FKH-7, FKH-8, FKH-9, FKH-10, LET-381, LIN-31, PES-1, PHA-4, T27A8.2 and UNC-130. We did RNAi against all these factors and found that only RNAi targeting *pha-4* showed reduced levels of *tph-1::yfp* (*otIs517*) and *cat-1::gfp* (*otIs221*) expression in the HSN, suggesting this member of the regulatory network could also be conserved between nematodes and mammals → **Figure 3.4.3-A**. *C. elegans* LIM-HD

family has 7 members: CEH-14, LIM-4, LIM-6, LIM-7, LIN-11, MEC-3 and TTX-3. RNAi against *ceh-14* and *lin-11* showed a downregulation of *cat-1::gfp* expression, that was only maintained for *tph-1::yfp* expression in the case of *ceh-14* → **Figure 3.4.3-B**. In this way, we identified one FKH member, *pha-4* and two LIM-HD members, *ceh-14* and *lin-11*, as potential regulators of HSN serotonergic terminal differentiation.

Going back to the phylogenetic tree, we found that, in most cases, the worm members of the HSN TF collective were closely related to their mouse orthologues. For example, HLH-3 appears as the phylogenetically closest worm TF to the mouse ASCL1 → **Figure 3.4.2-B**. EGL-46 is the only member of the INSM TF family in *C. elegans* and equally phylogenetically distant to INSM1 and INSM2 → **Figure 3.4.2-C**. AST-1 is the worm second closest TF to PET1 → **Figure 3.4.2-D**. Unpublished results from the laboratory demonstrated that *ets-5* mutants (the phylogenetically closest TF to PET1) show no 5-HT pathway defects in any serotonergic neuron of the worm. Therefore, AST-1 appears as the only functional homologue of PET1. BRN2 appears closer in the phylogenetic tree to the worm POU TF CEH-6, known to be involved in the regulation of several processes as locomotion, molting and ectodermal and excretory function, but not in serotonin regulation → **Figure 3.4.2-E**. UNC-86, in turn, is closer to BRN3.1 (*Pou4f3*), that controls the development of the auditory system (Lee et al. 2010). However we did not detect *Pou4f3* expression in the serotonergic neurons by *in situ* hybridisation (data not shown). The second closest common ancestor of BRN2 is UNC-86. In the case of GATA2 and GATA3, however, they share the closest common ancestor with ELT-1, then ELT-2 and finally EGL-18 and ELT-6 → **Figure 3.4.2-F**. However, *elt-6* RNAi treated worms showed no obvious phenotype at F1 scoring → **Figure 3.2.7**, whereas *elt-1* and *elt-2* RNAi were lethal during development but showed no phenotype at P0 scorings → **Figure 3.2.7**. This exemplifies that, although a tendency in serotonergic regulation, not always the closest orthologues are the ones that share a specific function. The newly identified PHA-4 candi-

Figure 3.4.2
Phylogenetic relationship between mouse and worm transcription factors

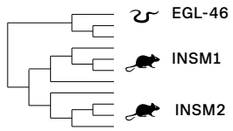
Cladograms show phylogenetic relationships between mouse and *C. elegans* TFs known to regulate serotonergic identity in one or both organisms. Cladograms were calculated using animal model data from TreeFam software (Ruan et al., 2008), although only worm and mouse are highlighted.



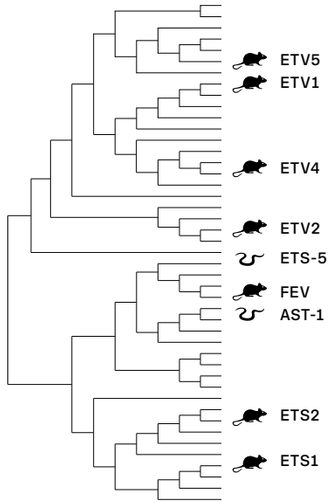
↳

↳ **Figure 3.4.2**
Phylogenetic relationship between mouse and worm transcription factors

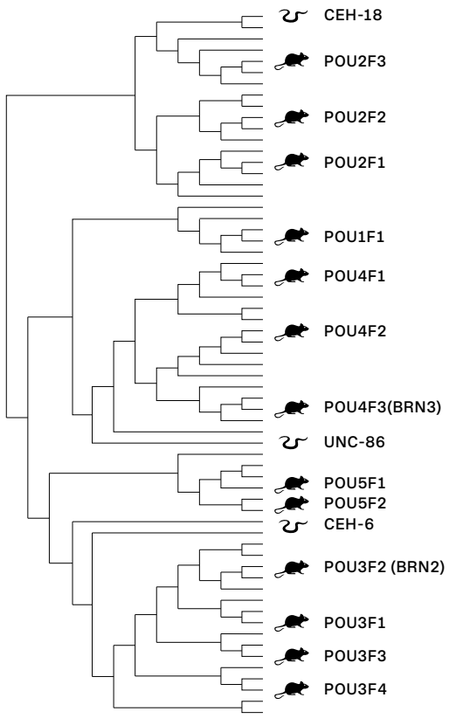
C **INSM**



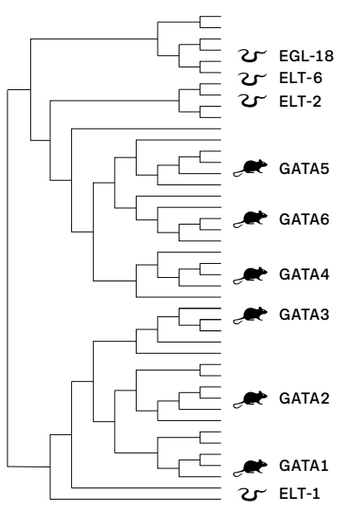
D **ETS**



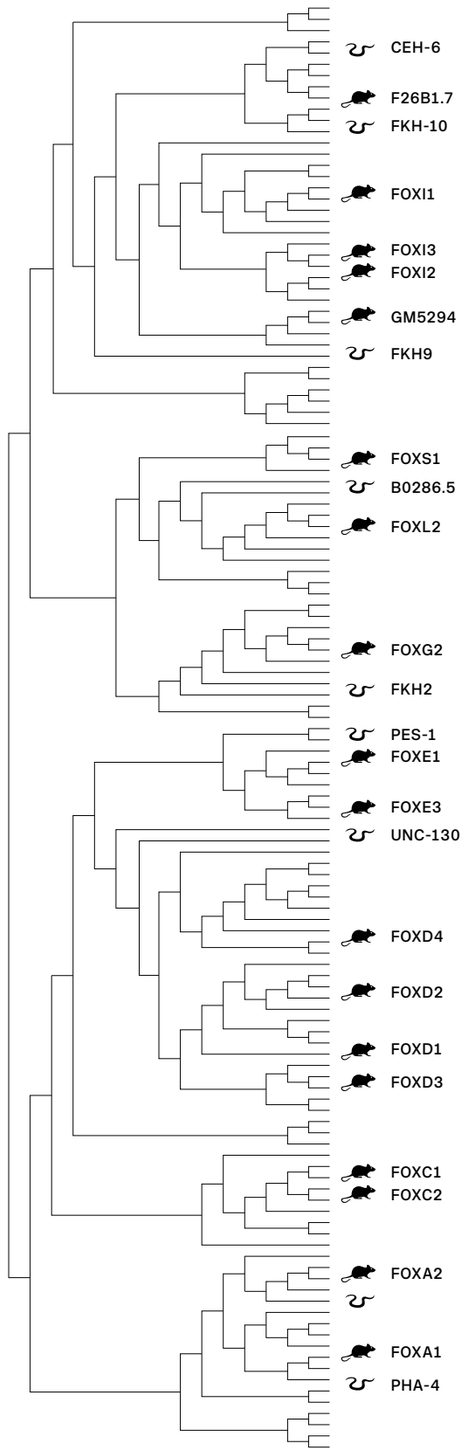
E **POU**



F **GATA**



G **FKH**



H **LIM**

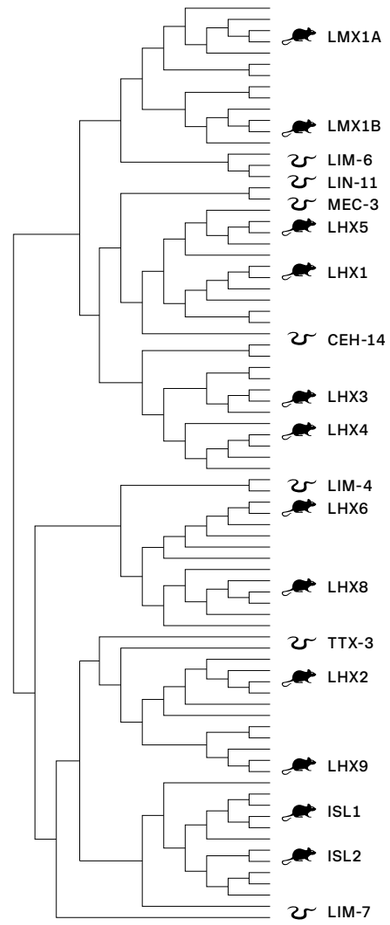
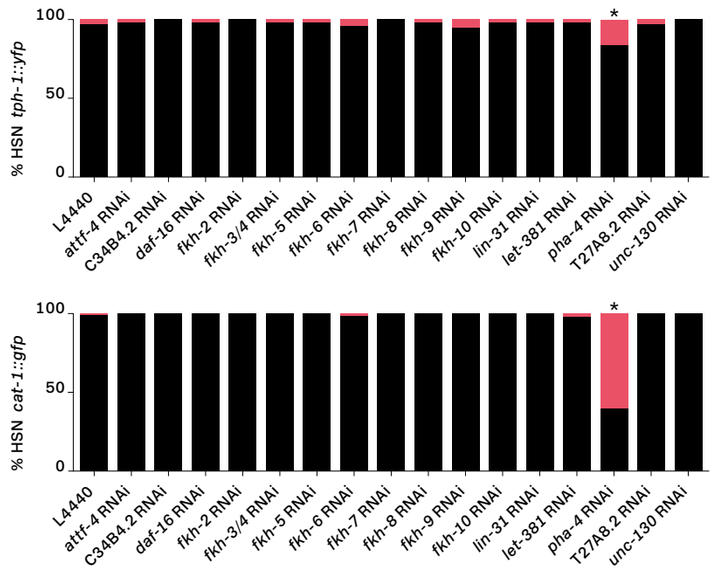


Figure 3.4.3
Characterisation of forkhead and LIM-homeodomain transcription factor candidates for HSN serotonergic regulation

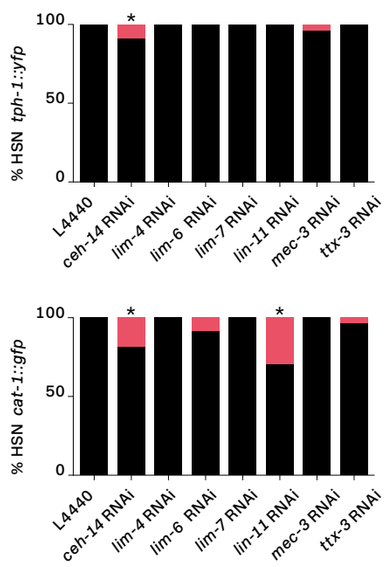
A



A) RNA interference screen against 16 of the 17 members of the forkhead (FKH) family. L4440 is the empty vector negative control. *tph-1* and *cat-1* reporter expression are significantly downregulated after *pha-4* RNAi mediated knock-down. > 30 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test; *pV < 0.05.
See Annex 3.2.3.

on
off

B



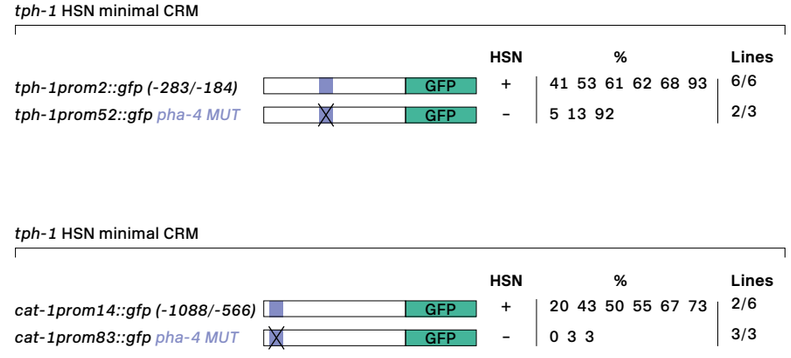
B) RNA interference screen against the seven members of the LIM-homeodomain (LIM-HD) family. *tph-1* and *cat-1* reporter expression are significantly downregulated after *ceh-14* and *lin-11* RNAi mediated knock-down.
See Annex 3.2.3.

RNAi experiments were performed by Ángela Jimeno.

C) *tph-1* and *cat-1* minimal *cis*-regulatory module analysis of *pha-4* BSs. Black crosses represent point mutations to disrupt the FKH BS (purple box). +: 100 to 60% of mean wild type construct values; +/−: expression values 60-20% lower than mean wild type expression values; −: values are less than 20% of mean wild type values. n>30 animals per line. See Annex 3.2.3 for mutated sequences.

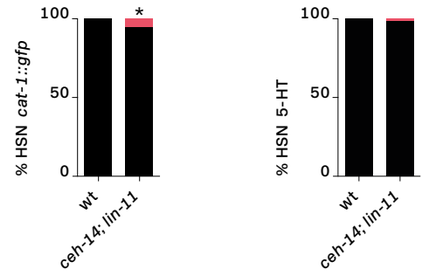
These experiments were performed by Dr. Miren Maicas.

C



D) Double mutant animals for the two LIM-HD candidates, *ceh-14* and *lin-11*, show normal levels of 5-HT staining and a mild phenotype for *cat-1::gfp* reporter expression. n > 50 worms per condition. Statistical significance was calculated using the two tailed Fisher exact test; *pV < 0.05.

D



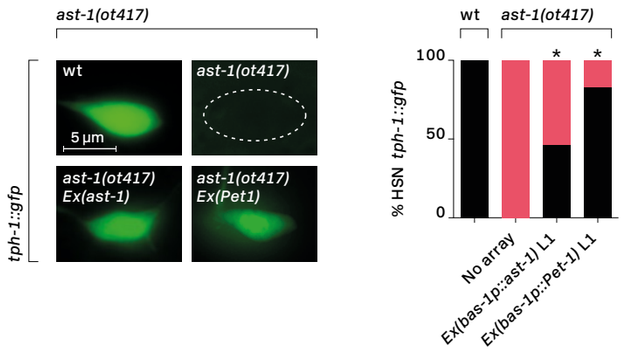
date is the closest worm FKH to the mouse FOXA2 → **Figure 3.4.2-G.** However, LMX1B closest orthologue in *C. elegans* is LIM-6, which showed no phenotype in RNAi experiments → **Figure 3.4.3-B.** The second closest worm TFs are CEH-14, LIN-11 and MEC-3 → **Figure 3.4.2-H.**

Forkhead, but not LIM-homeodomain, transcription factors have a role in HSN serotonergic terminal differentiation

In order to further characterise the new candidates *pha-4*, *ceh-14* and *lin-11*, we took two complementary approaches. We first looked within our 5-HT pathway gene CRMs for putative FKH and LIM-HD BSs (TF encyclopedia) (Wederell et al. 2008). We did find FHK sites in the *tph-1* (*tph-1prom2*) and *cat-1* (*cat-1prom14*) CRMs, but no LIM-HD were retrieved form the bioinformatics analysis → **Figure 3.4.3-C.** Directed mutagenesis upon these FKH sites led to loss of *tph-1* expression (*tph-1prom52*) and *cat-1* expression (*cat-1prom83*) in the HSN. Next, we aimed to analyse null loss of function mutants for our candidates. Unfortunately, *pha-4* null mutants are embryonic lethal, precluding us from studying its role in HSN terminal differentiation. We generated a double mutant strain using the null alleles *ceh-14*(*ch3*) and *lin-11*(*n389*), and the *cat-1::gfp* re-

Figure 3.4.4
Rescue of HSN transcription factor collective mutant phenotype with orthologous mouse factors

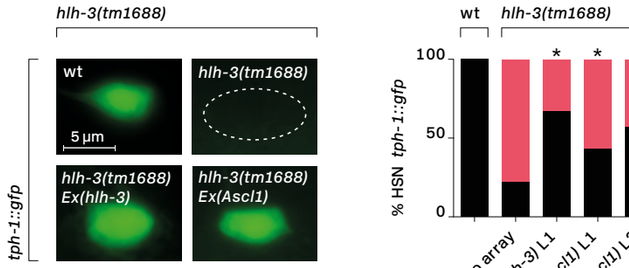
A



A) Micrographs showing *tph-1::gfp* expression in wild type animals, *ast-1(ot417)* mutant, and *ast-1(ot417)* mutants rescued with *ast-1* cDNA or mouse *Pet1* cDNA expressed under *bas-1* promoter. To the right, the quantification. n > 50 worms per condition. *: pV<0.05 Fisher's exact test. 'L#' indicates the transgenic line number. See Annex 3.4.1.

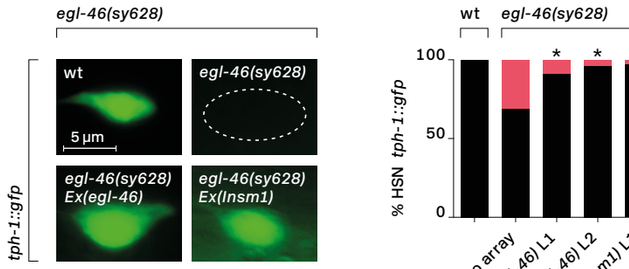
on
off

B

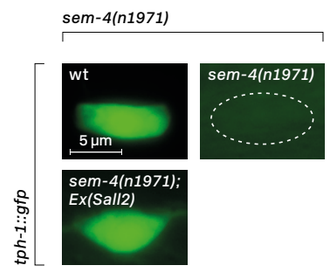


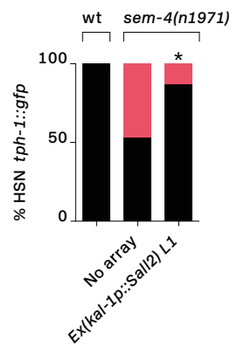
B)-D) Same rescue experiments and quantification as in (A), using *hlh-3/Ascl1*, *egl-46/Insm1* and *Sall2* cDNAs. A *cat-4* promoter was used to drive expression of all the factors except for *Sall2*, where a *kal-1* promoter was used.

C



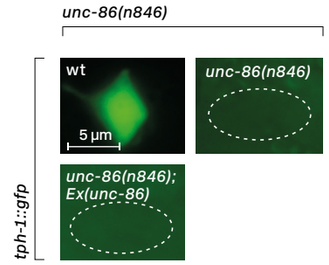
D

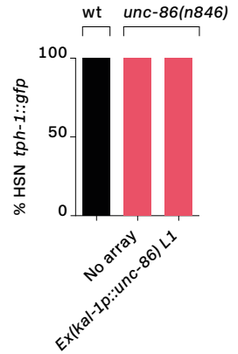




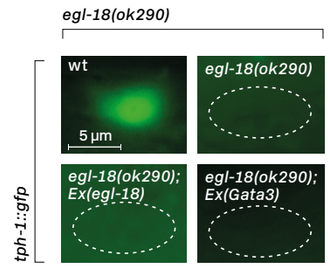
E)-F) Neither *unc-86* nor *egl-18/Gata3* are able to rescue HSN specific mutant defects.

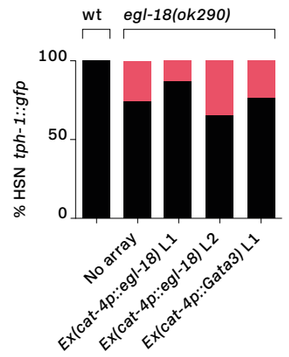
E





F





porter (*otIs221*) that showed the greatest phenotype in the RNAi screen. Double mutants showed very mild defects in *cat-1::gfp* expression in the HSN (7%±2 off phenotype) and wild type levels of 5-HT staining → **Figure 3.4.3-D,** → **Annexes 3.2.1.** and → **3.2.3.** Our findings suggest that *pha-4,* but not *ceh-14* or *lin-11*, could also be involved in HSN terminal differentiation.

The serotonergic transcription factor collective is functionally conserved between worms and mammals

The striking degree of homology of HSN and mouse serotonergic regulatory programmes made us wonder if the TF regulatory code could be functionally conserved between these two species. To answer this question, we performed cell specific rescue experiments of *C. elegans* mutants with the corresponding mouse homologue. First of all, we performed cell specific rescue of *C. elegans* mutants with the *C. elegans* gene and then performed similar experiments using the mouse orthologue gene. We expressed the cDNA of the TF, under the control of a promoter that satisfied these criteria: 1) drive GFP expression in the HSN, 2) be expressed in the minimum cells possible, in addition to HSN, and 3) not be regulated by the gene that the animal in which the construct will be injected is mutant for. For *ast-1* we used the *bas-1prom1* promoter, that drives 90% GFP expression in the HSN and whose expression is not affected in *ast-1* mutant background → **Figure 3.2.5.** For *hlh-3*, *egl-46* and *egl-18*, we used instead *cat-4prom4* that is also not affected in these mutant backgrounds and drives 87% GFP expression in the HSN → **Figure 3.2.5.** Finally, as almost every terminal feature tested is affected in *unc-86* and *sem-4* mutant backgrounds, we chose a *kal-1* promoter that is only slightly affected in *sem-4* mutant background and drives expression in the HSN in 84% of the cases → **Figure 3.2.5.**

In the case of *unc-86*, we used genomic DNA instead of cDNA. Mutant animals bearing the *tph-1* reporters (*zdIs13* or *yzIs71*) were injected with these constructs (termed the 'rescue array'), together with *ttx-3::mCherry* co-marker, and GFP expression in the cell was assessed. Once we confirmed that the worm constructs were able to rescue *tph-1* defects, we moved on to test the mouse *Pet1*, *Brn2*, *Sall2*, *Ascl1*, *Insm1* and *Gata3* genes. *Gata2* was not analysed because we were unable to obtain the cDNA. Plasmids and strains are listed in → **Table 2.16** and → **Table 2.21**.
We successfully rescued *ast-1*, *hlh-3* and *egl-46* phenotypes with the expression of the worm cDNA → **Figures 3.4.4-A, B** and **C.** However, we did not achieve *unc-86* or *egl-18* rescue → **Figures 3.4.4-E** and **F.** In the case of *sem-4*, no single line was retrieved form the microinjection, even when it was injected as low as 10 ng/µl. We hypothesise that extra doses of *sem-4* are lethal to the worms → **Figure 3.4.4-D.**

When we moved on to analyse the mouse rescues, we found that *Pet1*, *Ascl1*, *Inms1* and *Sall2* can functionally substitute *ast*-1, *hlh-3*, *egl-46* and *sem-4*, respectively → **Figures 3.3.4-A, B, C** and **D.** In the case of the SPALT family, *Sall2* does not seem as toxic to the worms as we were able to retrieve at least one line from the microinjection. These results suggest that both regulatory programmes are functionally conserved. Additionally, they confirm that the HSN TF collective, as expected, acts cell-autonomously, as specific expression of its components in the HSN is sufficient to restore the wild type function in a mutant background.

HSN and mouse raphe serotonergic neurons are molecularly similar

In evolutionary biology, deep homology refers to two structures that share the genetic mechanisms

Figure 3.4.5
Molecular homology between HSN and mouse serotonergic raphe neurons

HSN profile
Mouse raphe serotonergic profile
NSM and ADF serotonergic profile

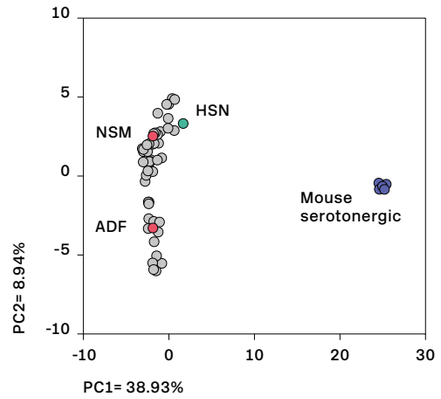
A) Worm-to-mouse vs. raphe serotonergic neurons

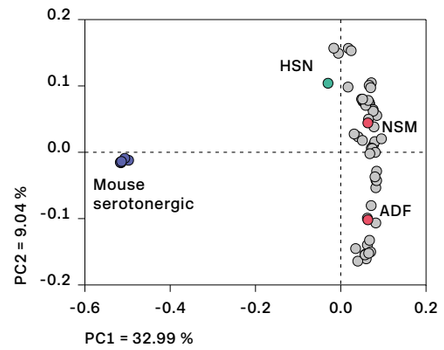
Principal Coordinate Analysis comparing expression profiles of worm neurons (grey dots, built by assigning mouse orthologues to *C. elegans* expressed genes (Hobert et al. 2016)) with expression profile of mouse raphe serotonergic neurons (blue dots, built from RNAseq data (Okaty et al. 2015)). HSN profile (green dot) is molecularly the closest to mouse raphe. See Annex 3.4.2 for the list of worm neurons considered in the analysis.

These experiments were performed in collaboration with Dr. Alejandro Artacho.

B) HSN profile without 5-HT pathway genes

Principal component analysis *C. elegans* neurons and mouse raphe neurons in which four 5-HT pathway genes *(tph-1, cat-1, bas-1* and *cat-4)* have been eliminated from HSN expression profile. HSN neuron is molecularly closest to mouse raphe even without considering 5-HT pathway gene expression. 5-HT pathway gene expression in other 5-HT neuron subtypes (red dots) is not sufficient to provide similarity to mouse raphe.



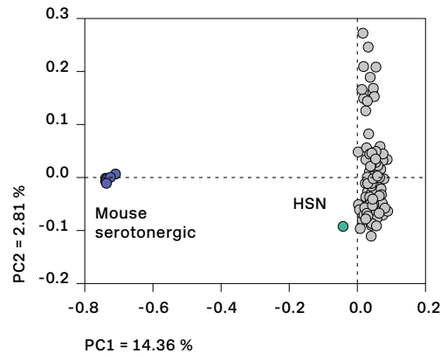


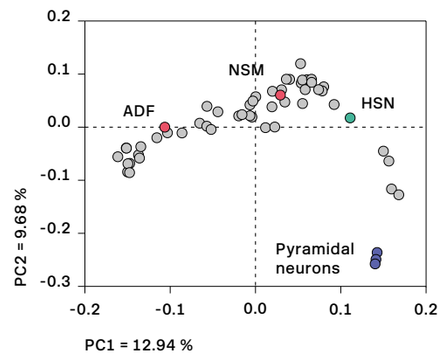
C) Randomised HSN-like profiles

HSN expression profile is composed by the four 5-HT pathway genes plus additional 92 genes. Analysis of 100 artificial HSN profiles composed by the same four 5-HT pathway genes plus 92 genes randomly selected from neuronal expressed genes shows that real HSN is still closest to mouse raphe.

D) Cortical neurons comparison

Principal Coordinates Analysis of *C .elegans* neurons compared to three different populations of cortical neurons at postnatal day one (corticothalamic neurons, cortico-callosal neurons and subcerebral cortical neurons) shows HSN is not molecularly the closest neuron to any of them (RNAseq data obtained from (Molyneaux et al. 2015)).





governing their differentiation (Shubin et al. 1997). The first example described that showed deep homology was the distribution of the *Distal-less* gene along the proximo-distal axis of all sorts of appendages and body outgrowths. Specifically, *Distal-less* is expressed along the proximo-distal axis of six developing coelomate phyla. To explain this, the authors could imagine two situations. One, that *Dll/Dlx* expression in such diverse (analogous) appendages could be convergent, although this would have required the independent co-option of *Dll/Dlx* several times in evolution. The other situation, which the authors interpreted as more likely, is that ectodermal *Dll/Dlx* expression along proximo-distal axis originated once in a common ancestor (homology) and has been used subsequently to pattern analogous body wall outgrowths in a variety of organisms. In this sense, there is a deep homology of genetic mechanism in relation to disparate analogous organs across a wide range of taxa.

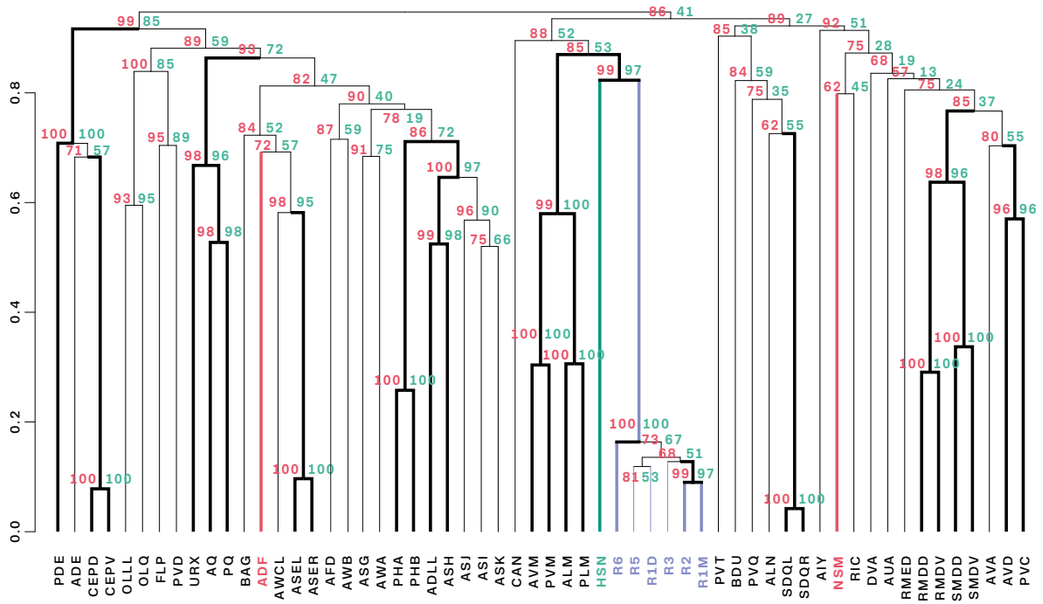
Our observation that both mouse and *C. elegans* serotonergic genetic programmes are homologous suggests that these two neuronal types share deep homology and, thus, correspond to homologous structures. If this were the case, then HSN neurons and mouse serotonergic raphe neurons should not merely share the expression of 5-HT pathway genes, which are also present in the other *C. elegans* serotonergic neurons NSM and ADF, but also should be broadly similar in molecular terms. To address this question, we used available gene expression information from Wormbase to generate partial expression profiles for the 118 neuronal classes of the *C. elegans* hermaphrodite (Hobert et al. 2016). Due to the incompleteness of worm neuronal expression profiles, we selected neuron classes defined by the expression of at least 30 different genes (49 different classes of neurons match this criteria, listed in → **Annex 3.4.2.**) Next, we assigned mouse orthologues to *C. elegans* neuronal genes to create a new dataset of expression profiles termed 'worm to mouse neuron profiles'. A detailed analysis of adult mouse serotonergic neuron transcriptome has been recently published (Okaty et al. 2015). We thus used this data to compare mouse serotonergic neuron molecular profile to all 'murine-like' worm neuron profiles. Principal Coordinates Analysis (PCoA) revealed that, out of the 49 analysed *C. elegans* neuronal classes, HSN is molecularly the closest to mouse serotonergic neurons → **Figure 3.4.5-A**. Using the same data as for the PCoA analysis, we performed hierarchical clustering analysis (HCA). Raphe serotonergic neurons are closer to HSN in the tree and, indeed, form a very robust cluster (AU = 99 ± 0.1 and BP = 97 ± 0.1) → **Figure 3.4.6**. Moreover, 65% of HSN expressed genes have at least one orthologous gene expressed in mouse serotonergic neurons, which correspond to different functional categories including axon guidance and migration, neurotransmission and synaptogenesis, transcriptional regulation, morphogenetic pathways and, of course, 5-HT biosynthetic pathway → **Table 3.4.1**. Interestingly, most of the *C. elegans* genes that have mouse orthologues expressed in the raphe contain the HSN regulatory signature → **Table 3.4.1**. Finally, we noticed that several of the mouse genes with *C. elegans* orthologues expressed in HSN have been associated to serotonin related disorders in genome wide association studies → **Table 3.4.2**.

Several controls were carried out to verify the robustness of this analysis. Firstly, we know that the similarity observed between the HSN and the raphe serotonergic neurons is not merely due to the expression of the 5-HT pathway genes as the NSM and ADF neurons, which also express these genes, are molecularly more distant to the mouse raphe neurons than HSN → **Figure 3.4.5-A**. In this line, we removed the 5-HT pathway genes from the HSN expression profile and saw that HSN remains the closest neuron to mouse serotonergic raphe neurons → **Figure 3.4.5-B**. Next, to discard that the close

Figure 3.4.6
Hierarchical clustering analysis between the HSN and the mouse raphe serotonergic neuron profiles

A) Same data used for Principal Coordinates Analysis (Figure 3.4.5) was used to perform hierarchical clustering. AU (red numbers) and BP (green numbers) represent the bootstrap probability with which each cluster forms. Thicker lines highlight clusters with AU > 95 (AU was preferred over BP because it systematically varies sample size and thus it is less biased). Note that HSN and raphe nuclei form a very robust cluster (AU = 99 ± 0.1 and BP = 97 ± 0.1, which means that the chances that this cluster does not represent a real cluster and that it is due to sampling error is, at most, 3%). Moreover, the other worm serotonergic neurons NSM and ADF do not cluster together with HSN or mouse raphe serotonergic neurons.

This analysis was performed by Carlos Mora.



* indicates *C. elegans* gene
with HSN regulatory signature

Table 3.4.1
Caenorhabditis elegans
HSN neurons and mouse raphe serotonergic neurons homology

C. elegans gene name	Mammalian gene name	Description
*bas-1**	Dopamine decarboxylase	*Ddc*
*cat-1**	Vesicular monoamine transporter	*Slc18a2*
cat-4	GTP cyclohydrolase 1	*Gch1*
*tph-1**	Tryptophan hydroxylase	*Tph2*
	Axon guidance and Migration	
*ebax-1**	Elongin-B/C E3 ligase	*Zswim5/6/8*
*egl-43**	PR domain containing	*Prdm16*
fmi-1	Flamingo homologue	*Celsr2/3, Fat1/3, Dchs1*
*madd-2**	Trim protein	*Trim9/36/46, Fsd1/1l, Mid2*
*mau-2**	Chromatid cohesion factor	*Mau2*
*mig-10**	Protein with an RA-like, PH domains and proline-rich motif	*Raph1, Grb10*
nck-1	SH2/SH3 domain-containing protein	*Nck1*
*rig-6**	Neuronal IgCAM	*Cntn1, 2, 3, 4, 5, 6*
*tbb-4**	Tubulin	*Tubb2a/2b/4a/4b/5*
*unc-40**	Netrin receptor	*Dcc, Neo1*
*unc-51**	Serine/threonine protein kinase	*Ulk1/2*
*unc-53**	Neuron navigator	*Nav1/2/3*
	Neurotransmission/Synaptogenesis	
*abts-1**	Anion/Bicarbonate Transporter family	*Slc4a7/8/10*
clh-3	Voltage sensitive chloride channel	*Clcn2*
eat-16	Regulator of G protein signalling	*Rgs11/19*
*gar-2**	G protein-coupled acetylcholine receptor	*Hrh3*
*ggr-2**	GABA/Glycine Receptor	*Glra1/2, Glrb*
*glr-5**	Glu Receptor	*Grid1/2, Grik1*
*gsa-1**	G protein, Subunit Alpha	*Gnal, Gnas*
*ida-1**	Protein tyrosine phosphatase-like receptor	*Ptprn, Ptpm2*
*irk-1**	Inward Rectifying K (potassium) channel family	*Kcnj3/5/6/9/11/16/*
*kcc-2**	K/Cl cotransporter	*Slc12a5/6*
*mpz-1**	Multiple PDZ domain protein	*Mpdz, Pdzd2, Inadl, Lnx1*
nhx-5	Na/H exchanger	*Slc9a6/7/9*
*nid-1**	Nidogen (basement membrane protein)	*Lrp1/1b*
nra-4	Nicotinic Receptor Associated	*Nomo1*
rsy-1	Regulator of synapse formation	*Pnisr*
*syg-1**	Ig transmembrane protein	*Kirrel, Kirrel3*
*nlg-1**	Neuroligin family	*Nlg1/2/3*
*unc-2**	Calcium channel alpha subunit	*Cacna1a/1b/1e*
unc-77	Voltage-insensitive cation leak channel	*Nalcn*
*unc-103**	K+ channel	*Kcnh2/7*
	Transcriptional regulation	
*ceh-20**	PBX TF	*Pbx1/2/3*
*egl-44**	TEA domain TF	*Tead1*
gei-8	Nuclear receptor co-repressor	*Ncor1*
*hlh-3**	bHLH TF	*Ascl1*
*ife-4**	Initiation factor 4E	*Eif4e2*
*sem-4**	Spalt TF	*Sall2, Zfp236/Znf236*
	Morphogenetic pathways	
*dsh-1**	Homologue of disheveled	*Dvl1/3*
*plr-1**	Ring finger protein	*Rnf215*
*prkl-1**	Drosophila Prickle homologue	*Prickle1/2*
*sel-10**	Suppressor/Enhancer of Lin-12(Notch)	*Fbxw7*
	Others	
*aak-2**	AMP-activated protein kinases	*Prkaa1/2*
ags-3	G protein signalling modulator	*Gpsm1*
*aho-3**	Hydrolase	*Abhd17a/17b*
ari-1	Ubiquitin-protein transferase	*Arih1*
arr-1	G protein signalling adaptor	*Arrb1/2*
*arrd-17**	Arrestin domain protein	*Arrdc3*
baz-2	Bromodomain adjacent to zinc finger domain	*Baz2a/2b*
elpc-1	Elongator complex protein component	*Ikbkap*
elpc-3	Elongator complex protein component	*Elp3*
*goa-1**	G protein,O, Alpha subunit	*Gnao1*
kin-20	Protein kinase	*Csnk1d/1e*
puf-9	Pumilio/FBF domain- containing	*Pum1/2*
*pxf-1**	Rap guanine nucleotide exchange factor	*Rapgef2/6*
rep-1	Rab escort protein	*Chm, Chml*
*ten-1**	Type II transmembrane EGF-like repeats	*Tenm1/3/4*
*top-1**	Topoisomerase	*Top1/1mt*

proximity between mouse serotonergic neurons and HSN is due to a random combination of genes with high degree of homology to raphe neurons, we built 100 random HSN profiles composed by the four 5-HT pathway genes (*tph-1/Tph*, *bas-1/Ddc*, *cat-1/Vmat*, *cat-4/Gch*) plus 92 additional genes from the pool of genes known to be expressed in all of the neurons of the worm. Comparison analysis of HSN profile and the 100 random HSN profiles shows that the real HSN is much closer to mouse serotonergic neurons than any of the random profiles → **Figure 3.4.5-C**. Additionally, we tested if HSN similarity to serotonergic neurons is specific to this neuron subtype. To test this, we performed similar analysis using RNAseq data obtained for cortical populations (corticothalamic neurons, cortico-callosal neurons and subcerebral cortical neurons) (Molyneaux et al. 2015). HSN proximity was not maintained with these cortical populations, suggesting that it is specific of the serotonergic fate → **Figure 3.4.5-D**. Finally, although serotonergic raphe neurons can be subdivided in different nuclei with slightly different transcriptome profiles (Okaty et al. 2015), we found that HSN does not show any obvious proximity to any specific subtype of mouse raphe serotonergic neurons (data not shown).

In sum, these results reveal an unexpected level of molecular proximity between *C. elegans* HSN and mouse serotonergic raphe neurons, deep homology in the genetic programme that regulates their terminal differentiation, and the presence of similar regulatory signatures in genes expressed in both cell populations. These results demonstrate that the serotonergic transcriptional regulatory code is highly conserved in evolution.

Table 3.4.2
Molecular homology between HSN and raphe neurons include genes associated to serotonin related disorders

Data obtained from the Genome-wide associated study catalogue.

→

C. elegans gene name	Description	Mammalian gene name	GWAS
Axon guidance and Migration			
egl-43	PR domain containing	*Prdm16*	FAntipsychotic Agents (HGVST461)
fmi-1	Flamingo homologue	*Fat3*	Narcolepsy (HGVST115)
mig-10	Protein with an RA-like, PH domains and and a proline-rich motif	*Grb10*	Narcolepsy (HGVST115), Schizophrenia (HGVST320)
unc-53	Neuron navigator	*Nav3*	Antipsychotic Agents (HGVST461)
Neurotransmission/Synaptogenesis			
abts-1	Anion/Bicarbonate Transporter family	*Slc4a10*	Bipolar Disorder (HGVST889), (HGVST472)
glr-5	Glu Receptor	*Grik1*	Attention Deficit Disorder with Hyperactivity (HGVST429)
mpz-1	Multiple PDZ domain protein	*Inadl*	Schizophrenia (HGVST903 and HGVST320)
		Pdzd2	Narcolepsy (HGVST115)
nhx-5	Na/H exchanger	*Slc9a9*	Tobacco Use Disorder (HGVST89)
syg-1	Ig transmembrane protein	*Kirrel3*	Schizophrenia (HGVST903)
nlg-1	Neuroligin family	*Nlg1/Nlgn1*	Narcolepsy (HGVST115)
unc-77	Voltage-insensitive cation leak channel	Nalcn	Bipolar Disorder (HGVST889 and HGVST472), Schizophrenia (HGVST320)
Transcriptional regulation			
ceh-20	PBX TF	*Pbx3*	Narcolepsy (HGVST115)
Others/Undetermined			
elpc-1	Elongator complex protein component	*Ikbkap*	Bipolar Disorder (HGVST316)
elpc-3	Elongator complex protein component	*Elp3*	Schizophrenia (HGVST903, HGVST320)
		Cntn4	Antipsychotic Agents (HGVST461),
		Cntn6	Bipolar Disorder (HGVST889 and HGVST472)
ten-1	Type II transmembrane protein containing EGF-like repeats	*Tenm4*	Bipolar Disorder (HGVST163), Narcolepsy (HGVST115), Schizophrenia (HGVST903)

Discussion

In this work we have revealed insights into how serotonergic neuron identity is globally controlled and, focusing on the regulatory logic of the HSN serotonergic subtype, we have increased our understanding of how the complement of cell type-specific enhancers is selected. We found that numerous TFs (at least six analysed here) act in conjunction to directly activate HSN expressed enhancers. This high number of TFs helps to provide specificity and robustness to the HSN regulatory signature, which is preferentially found in genes of the neuronal genome that are important for HSN function.

Serotonergic neuron subtypes are regulated by independent *cis*-regulatory modules

We first analysed *C. elegans* serotonergic subtype terminal regulation through a *cis*-regulatory analysis of the 5-HT pathway genes. We found that expression of the 5-HT pathway genes in the different serotonergic neuron subtypes, NSM, ADF and HSN, is controlled through different *cis*-regulatory modules.
This modular and independent subtype regulation is in agreement with the terminal selector model in which, for each cell type, a combinatorial code of TFs directly regulates the expression of most terminal features (Hobert et al. 2016). Taking into consideration that NSM, ADF and HSN, despite being all serotonergic, contain very different transcriptomes, it was expected that they are regulated by different TF codes and, thus, different CRMs are required for subtype-specific expression of each 5-HT pathway gene. We have previously reported that the POU TF UNC-86 next to the LIM TF TTX-3 terminally control NSM differentiation programme (Zhang et al. 2014) and, here, we have shown that a different combination of TFs regulates HSN terminal differentiation. Little is known about the TFs required for ADF terminal differentiation but, apparently, a different set of TF

families, including the RFX family (Xie et al. 2013), controls this process. In this sense, it seems that NSM, ADF and HSN could be more properly considered different classes of neurons that share a particular 'group identity' (the serotonergic identity), rather than different neuron subclasses. Similar regulatory logic has been found in the specification of *C. elegans* glutamatergic, cholinergic and GABAergic neuron subtypes (Serrano-Saiz et al. 2013; Pereira et al. 2015; Kratsios et al. 2011; Kerk et al. 2017; Gendrel et al. 2016). Within these cell types, despite they all share the battery of genes responsible for glutamate, acetylcholine or GABA metabolism, cell subtypes are regulated by different TFs. The exception would be *C. elegans* dopaminergic neuron subtype specification. Although they are also born from different cell lineages, the eight dopaminergic neurons are classified in three different anatomical subtypes (CEP, ADE and PDE) that are functionally and molecularly equivalent (i.e. they are all mechanosensory neurons). In this case, a unique code of TFs controls dopaminergic terminal differentiation of all dopaminergic neuron subtypes (Flames & Hobert 2009; Doitsidou et al. 2013). Contrary, serotonergic, glutamatergic, cholinergic and GABAergic neurons exhibit very disparate functions and are regulated by different combinations of TFs. Moreover, all serotonergic neurons except NSM, have been described to signal via additional neurotransmitters, while dopaminergic neurons exclusively use dopamine (Rand & Nonet 1997; Pereira et al. 2015; Loer & Rand 2016). This diversification of function plus signalling promiscuity inevitable adds layers of complexity to the regulation of neuron subtype terminal fate that must be reflected in the transcriptomes of the neuron subtypes and, hence, in their regulation by TFs. The higher complexity cannot be only associated to broad neuron types like GABAergic (26 neurons), glutamatergic (78 neurons) or cholinergic neurons (159 neurons), as the serotonergic system, like the dopaminergic system, is

rather small (Gendrel et al. 2016; Serrano-Saiz et al. 2013; Pereira et al. 2015; Chase & Koelle 2007).

Our study on serotonergic regulatory logic also revealed partial overlap between the CRMs of the 5-HT pathway genes. This suggests that some of the TFBSs are commonly used by the same or different TFs in the different serotonergic populations of the worm. In fact, UNC-86 is required for both NSM and HSN terminal differentiation (Sze et al. 2002) and we have identified POU TFBSs in the *tph-1* and *bas-1* CRMs that are functional both in HSN and NSM neurons. In addition, we also identified functional TFBSs (putative bHLH) that are shared between the HSN and the ADF neurons. In this case, we found partial ADF differentiation defects in *hlh-3* null mutants, suggesting a possible role for this TF in the ADF neuron. In other cases, we found effects in the *cis* mutation but no ADF phenotype in the corresponding HSN TF mutant. Hence, a different member of the same TF family may be binding to the motif in the ADF. Redeployment of a *cis*-regulatory motif has already been shown for POU sites in distinct glutamatergic neurons; the same POU site is apparently recognised by UNC-86 in light touch receptor neurons and by CEH-6 in the AUA neurons (Duggan et al. 1998; Serrano-Saiz et al. 2013). Alternatively, the mutations introduced in the CRMs could be affecting alternative TFBSs that have not been considered in this work.

Interestingly, we identified a case of two redundant CRMs for the ADF neuron. This redundancy in CRMs has been previously described in other systems as shadow enhancers and are usually associated to robustness in gene expression (Hong et al. 2008). Furthermore, although seldom, we also observed in our *cis*-regulatory analysis events of ectopic GFP expression in neurons other than those that normally express the 5-HT pathway genes. This brings together the classical view of terminal selectors mainly acting as activators of specific gene batteries in distinct

neuronal types (Hobert 2008), with the increasing evidence that neuronal subtype diversity can be achieved, or at least modulated, through repressor elements (Esmaeili et al. 2002; Chang et al. 2003; Kerk et al. 2017; Miller et al. 1992). Following this idea, although not addressed in this work, we cannot discard the possibility that neuron subtype specificity in the serotonergic system is conferred by the presence of repressor elements in the set of differentially expressed genes between the three serotonergic neuron subtypes.

A complex code of six transcription factors is required and sufficient to induce serotonergic fate specifically in the serotonergic HSN subtype

Neuronal terminal differentiation programmes have been best characterised in *C. elegans*. So far, relatively simple TF codes, composed of two or three members, have been shown to be required, and in some contexts sufficient, to select specific neuronal types (van Buskirk & Sternberg 2010; Serrano-Saiz et al. 2013; Doitsidou et al. 2008; Zhang et al. 2014). Thus, previous work suggested a rather simple organisation of CRMs controlling expression of neuronal terminal features in *C. elegans*.
Our results, however, demonstrate a more complex scenario in the regulation of the HSN transcriptome. In light of our findings, nematode neuronal terminal differentiation programmes are not necessarily significantly simpler than those found in vertebrates, as previously proposed (Holmberg & Perlmann 2012). Considering the technical advantages of *C. elegans* as a simple model system, our work is an example on how its study may help us to identify the general rules of terminal differentiation in eumetazoa.
In our characterisation of the regulatory mechanisms that control HSN terminal differentiation, we found a complex code of at least six terminal selectors belonging to different TF families, termed

the HSN TF collective. This TF collective does not only control the serotonergic fate of the HSN, but also many additional effector genes of the neuron, supporting the principle of co-regulation of many distinct terminal identity features by terminal selector-type TFs (Hobert 2011) that has been shown for many other neuron types in *C. elegans* (Wenick & Hobert 2004; Flames & Hobert 2009; Doitsidou et al. 2013; Serrano-Saiz et al. 2013) → **Table 1.1**. In the absence of these six terminal selectors, HSN neurons appear to remain in an undifferentiated neuronal ground state, as the expression of certain neuronal genes remain. It would be interesting to assess if they additionally show a switch in identity, as occurs in the mouse dorsal spinal cord, where *Tlx3* and *Tlx1* determine excitatory over inhibitory cell fates (Cheng et al. 2004). The mechanistic basis for this scenario could be that HSN terminal selector genes not only activate terminal differentiation genes, but also inhibit alternative fates by repressing the expression of other terminal selector genes. This has been elegantly demonstrated in *C. elegans* ALM and BDU neurons (Gordon & Hobert 2015). Although not included in the results, we observed ectopic expression of 5-HT pathway reporters in the PVT and an unidentified neuron in the posterior body of the worm, in *sem-4* and *unc-86* loss of function mutants, respectively. This suggests a direct or indirect role of these factors in the repression of the serotonergic phenotype in these unrelated neurons.
Moreover, the action of the six members of the HSN regulatory code seem to be exclusively required at the latest steps of HSN neuron differentiation and it must be maintained throughout the rest of the life of the animal in order to preserve the terminal differentiation programme unaltered, as described for many other terminal selectors (Deneris & Hobert 2014). Albeit meeting both conditions, *egl-18* seems to have a dual role in the regulation of the HSN neuron: late, as demonstrated with the

CRM mutational analysis and specific *in vitro* binding to GATA sites, but also early, as revealed by a significant lineage defect observed in *egl-18* mutants. Indeed, *egl-18* expression in the HSN matches this idea, as our data suggests that its expression begins before the postmitotic HSN neuron is generated and is maintained throughout the life of the animal. Regarding *hlh-3*, however, our CRM mutagenesis analysis indicates direct binding to the 5-HT pathway genes, yet we have been unable to detect *hlh-3::yfp* signal in the HSN after the precursor HSN/PHB stage, before the 5-HT pathway genes are even expressed. This could be explained in two ways. On the one hand, it is well known that bHLH TFs act as proneural factors that activate target genes in proliferating and differentiating progenitors during neurogenesis (Bertrand et al. 2002). However, to explain our results, we could further envision HLH-3 as a pioneer TF, as has been recently shown for its mouse homologue *Ascl1* (Raposo et al. 2015). These TFs are able to bind closed and open chromatin in proliferating cells, promoting accessibility and activation of differentiation specific genes (Zaret & Mango 2016). On the other hand, maybe we are not using the right tools to analyse *hlh-3* expression in the cell. If *hlh-3* expression in HSN were very low, then using standard fluorescent microscopy would not be enough to detect it. Alternatively, we have used a fosmid reporter strain and it is known that fosmids form episome-like structures that alter their accessibility to TF regulation and chromatin environment (Kelly et al. 1997). The use of CRISPR-Cas9 technology would help settle these doubts (Dickinson et al. 2015).

The HSN TF collective is not only required to establish the serotonergic fate in the HSN neuron, but also sufficient, at least in some cellular contexts. Ectopic expression of all the members of the code, except *egl-46*, increases the number of serotonergic-like cells in the embryo and the effect seems to be stronger with some members of the code (*ast-1*, *sem-4* and

mainly *unc-86*). Ectopic expression of a combination of the six members of the HSN TF collective, a combination of the three 'highly responsive' factors or *unc-86* alone, induces the maximum response observed. Importantly, both combinations are statistically higher in terms of percentage of embryos that respond to the heat shock. Similar results were obtained with the dopaminergic TF code (Flames & Hobert 2009; Doitsidou et al. 2013). When we analysed later developmental stages, the over-expression response is much more modest and restricted to neuronal cells. Similarly to what happens in the embryo study, the combination of the six members of the HSN TF collective does not increase the number of ectopic cells and this lack of plasticity has been shown to be mediated by repressive chromatin marks deposited at the end of development (Patel & Hobert 2017). Nonetheless, as *unc-86* regulates both HSN and NSM serotonergic fate (Sze et al. 2002; Zhang et al. 2014), we cannot distinguish between ectopic neurons generated by creating a 'HSN-like' or an 'NSM-like' environment. The fact that the six TF combo further enhances the ectopic response of a larger number of embryos, suggests that the phenotype observed could be due to this 'HSN-like' environment. Furthermore, we have shown that the HSN TF collective is not only required and sufficient for the establishment of the HSN fate, but also for the maintenance of the serotonergic phenotype throughout the life of the worm, as has been claimed for many other regulators of terminal differentiation (Deneris & Hobert 2014).

The HSN transcription factor collective acts through parallel pathways and shows synergistic relationships to regulate the terminal features of the HSN neuron

Our epistatic analysis indicates that the different members of the HSN TF collective tend to act mostly in an independent manner, reinforcing the results obtained from mutagenesis and EMSA analyses. Together, this evidence actively indicates that the six TFs act through parallel pathways to bind and directly regulate the terminal features of the HSN neuron. However, there are a few examples of cross-regulation between certain members of the HSN TF collective, suggesting the existence of a more complex serotonergic transcriptional regulatory network.
While some TFs have no effect on the expression of the rest of the HSN TF collective, others seem to have key roles in the HSN transcriptional programme. Importantly, UNC-86 appears as master regulator of the HSN terminal differentiation as it is epistatic to *ast-1*, *sem-4* and, to less extent, *egl-46*. Moreover, it is the single factor that, when ectopically expressed, induces the highest number of serotonergic cells in the embryo and, in combination with two of its probably direct targets *sem-4* and/or *ast-1*, further increases the penetrance. Its relevance in serotonergic specification can be expanded to additional subtypes of the serotonergic system, as its requirement for NSM, AIM and RIH acquisition of the serotonergic fate is already known (Sze et al. 2002). UNC-86 is also expressed in many non-serotonergic neurons, where it also acts as a terminal selector in combination with other TFs (Topalidou & Chalfie 2011; Gordon & Hobert 2015; Duggan et al. 1998). Another interesting observation is that several TFs (UNC-86, SEM-4, HLH-3 and, very subtly, EGL-46) control the expression of *ast-1* and all of them, in turn, are required for the regulation of HSN terminal features. This type of regulation is known as feedforward loops and has been described to attain, stabilise and maintain the complete signature of a cell-specific programme of gene expression, increasing the robustness of the system (Davidson 2006; Altun-Gultekin et al. 2001; Alon 2007). Remarkably, similar regulatory mechanisms are observed with *ast-1* mouse homologue *Pet1* (Wyler et al. 2016; Deneris & Wyler 2012b).
Combinatorial TF binding enables cell type-specific enhancer expression and usually implies cooperativity between TFs.
Our detailed phenotypic analysis of single and double TF mutants points to a synergistic and redundant control of the expression of the terminal features of the HSN. Firstly, co-regulation of terminal features by the HSN TF collective is not exactly equivalent in each target gene. AST-1, for example, is absolutely required for *tph-1* and *cat-1* expression, however it is dispensable for *bas-1* expression. Considering the single mutant analysis in isolation, one would come to the erroneous conclusion that AST-1 does not play a role in *bas-1* expression. However, we find functional ETS BSs in *bas-1* minimal CRM and the role of AST-1 in *bas-1* expression is revealed in double mutant analysis with other members of the HSN regulatory code (*egl-46*, *sem-4* or *unc-86*), showing an exacerbated phenotype or synergy between TFs. This is known to occur in over specified pathways, where organisms are buffered when redundant genes suffer a mutation that disrupts, but does not eliminate, the function of their respective proteins, increasing the robustness of the system. However, when both genes are deleted, there is not enough recruitment to the enhancer and no, or less, gene expression is achieved. A similar de-coupling of regulation of terminal identity features has been observed in the specification of the serotonergic neuron type NSM (Zhang et al. 2014) and in cholinergic command interneurons (Pereira et al. 2015). We also find other examples of synthetic enhancement and suppression between several pairs of HSN TF collective members. Similar to

ast-1, for any given TF pair, additive or synergic regulation varies between target genes. Although our analysis does not inform about the possible protein-DNA or protein-protein relationships between the members of the HSN TF collective, it is likely that the subtle differences in the regulation of each terminal feature are determined by the specific number and disposition of functional TFBSs found in the CRM of each gene as will be discussed next.

The HSN regulatory signature identifies HSN expressed genes

Co-binding of specific combinations of TFs to the same genomic region, assessed by chromatin immunoprecipitation sequencing (ChIP-seq), has been successfully used to identify, *de novo*, cell type-specific enhancers in *Drosophila* embryos (Busser et al. 2015; Zinzen et al. 2009). However, this approach is based on experimental data and fails to address why some predicted TFBSs are actually bound by the TF while others are not. Following the terminal selector model, we hypothesised that the specific co-expression of the six members of the HSN TF collective in the HSN neuron could directly regulate its terminal transcriptome.
In agreement with this idea, we find that known HSN expressed genes contain DNA elements enriched in clusters of bioinformatically-predicted TFBS for the six members of the HSN regulatory code, termed HSN signature. Taking advantage of this, we take a step further and demonstrate that the HSN regulatory signature is preferentially found in neuronal genes that show HSN-related functions, and can be used for the *de novo* identification of HSN active enhancers. Conceptually, this means that if the combinatorial code of TFs is complex enough (in our case six TFs), it is sufficient to impose a defining signature to the enhancers they regulate allowing for the discrimination of these

HSN functional enhancers from the whole genome. The *C. elegans* genome is particularly compact compared to fruit flies and vertebrates, which could, in part, explain the success of our approach. Additionally, the number of TFs included in our analysis is higher than previous reports, which could help to confer sequence specificity to the HSN regulatory signature. Unfortunately, the use of *C. elegans cis*-regulatory bioinformatics analyses has been anecdotic to date (Beer & Tavazoie 2004). Our results suggest that this approach might be transformative to decipher the rules underlying the regulatory genome. Of note, our analysis still shows a high rate of false positives, which suggests that additional features are present in HSN functional enhancers. It would be interesting to determine if this combination of six TFs is exclusively expressed in the HSN neuron. In fact, this seems to be the case based on our work using fluorescent reporters; we have determined that the HSN is the only serotonergic neuron where *unc-86*, *sem-4*, *egl-46*, *ast-1* and *egl-18* are simultaneously expressed at the L4 and adult stage. *hlh-3* expression in the worm, however, is restricted to embryonic stages, including the HSN. Future analyses based on more complex paradigms should facilitate the identification of HSN functional enhancers.

HSN regulatory signature contains syntactic rules

Three models have been proposed to explain enhancer function (Spitz & Furlong 2012). In the enhanceosome model, TFBSs show a rigid distribution in order, spacing and orientation. Conversely, the billboard model proposes a totally flexible distribution of TFBSs. Finally, the TF collective model shows flexibility in the arrangement of TFBSs or the requirement of all of them, but also considers some possible constraints in their disposition. These flexible constraints, which are important only in specific genomic

contexts, represent a challenge for the identification of general syntactic rules. In this work we show that the HSN regulatory programme matches best the TF collective model: the six TFs belonging to the HSN regulatory code act in a flexible manner, activating enhancers with variable TFBS order and distribution, as seen in the 5-HT pathway genes CRMs and functional enhancers of the HSN. Moreover, in agreement with the TF collective model, we observed that, in some genomic contexts, the presence of BSs for certain TFs is dispensable and can be compensated for by the rest of the TF code. Additionally, we determined that some TF pair orientations (ETS-GATA and ETS-HLH) and distance bias between them are required for the activation of some HSN enhancers. This finding is further supported by the observation that ETS TFs physically interact with GATA or bHLH factors in other systems (Li et al. 2000; Shi et al. 2014).

Limited examples of TF pair orientation requirements have been reported so far. Synthetic enhancers have been used to show that a specific pMad-Tin orientation drives stronger expression in *Drosophila* mesoderm, although it is unclear whether similar constraints are also important in the context of endogenous enhancers (Erceg et al. 2014). More recently, Farley et al. showed that, in *Ciona intestinalis*, specific orientation constraints between an ETS site and a Zinc Finger site is required for notochord expression of a newly identified developmental enhancer of the *Brachyury* gene (Farley et al. 2016). Our results not only provide an additional example of endogenous enhancers with TF pair orientation restrictions, but also extend the importance of syntactic rules not only for developmental enhancers but also in the transcriptional regulation of terminal features. This suggests that orientation constraints might be widely present in regulatory modules.

Deep homology, molecular homology and functional homology between *Caenorhabditis elegans* HSN and mouse serotonergic neurons

The diversity of *C. elegans* serotonergic neuronal classes contrasts with that of tetrapod vertebrates, in which serotonergic neurons are genetically and molecularly rather uniform and limited to the raphe system. In contrast, other chordates contain additional serotonergic populations. Serotonergic subclass diversity is also prevalent in other phyla such as arthropoda and mollusca, which suggests a loss of serotonergic diversity in the tetrapod branch. As in nematodes, serotonergic subclass specification in other organisms is likely to be independently regulated: in *Drosophila*, the TFs Islet, Hunchback and Engrailed are required for serotonergic specification of the ventral ganglion while dispensable for brain serotonergic specification, whereas in zebrafish, *Pet1* regulates raphe serotonergic specification but not other serotonergic subclasses (reviewed in (Flames & Hobert 2011)).
Our results reveal that the regulatory programme of the HSN neuron, but not that of the NSM or ADF, strikingly resembles the serotonergic regulatory programme in mouse (Deneris & Wyler 2012; Haugas et al. 2016; Scheuch et al. 2007; Nasu et al. 2014). This high similarity allowed for the identification of PHA-4, a new regulator of HSN serotonergic terminal identity in the worm, and the identification of SALL2 as a candidate regulator of serotonergic specification in mouse. In the light of these results, we propose that the HSN neuron shares deep homology with mouse raphe neurons and that this deep homology might be the result of a common ancestor cell type. However, as we do not have enough information about the serotonergic regulatory programmes in other animal groups, an alternative scenario is that they might have arisen independently in nematodes and

vertebrates. If HSN and mouse raphe serotonergic neurons were homologous cell types, we would predict that they are also functionally homologous. Serotonergic systems in all animal groups function as facilitators of motor output, often of repetitive nature, with 5-HT promoting a switch between states (Gillette 2006). In mammals, serotonergic projections to spinal cord, which is the most ancient component of the serotonergic system, produce a long-lasting facilitation of spinal reflexes. In molluscs, for example, serotonergic activity accompanies motor activity and frequently precedes motor onset. Interestingly, *C. elegans* 5-HT signalling in HSN neurons also facilitates motor output. Egg-laying behaviour transitions from inactive to active states of egg-laying, and 5-HT signalling in HSN mediates the onset of the active phase (Waggoner et al. 1998). Thus, HSN and mouse serotonergic neurons share deep homology, as well as molecular and functional homology. Our data, together with previous reports of deep homology of other neuronal types (Nomaksteinsky et al. 2013; Strausfeld & Hirth 2013; Tomer et al. 2010), suggest that deep homology might underlie the specification of a wide variety of neuron subtypes. The identification of homologous regulatory programmes using similar approaches to those described here will help identify homologous neuronal types in distant species.

In summary, we believe that our careful dissection of HSN regulatory enhancers, in the context of global serotonergic regulatory logic, has helped improve our understanding on the general laws of transcriptional regulation and supports that phylogenetically conserved mechanisms underlying these rules exist. Beyond fundamental rules, our results also show, for the first time, that a regulatory signature based on a defined set of TFs is sufficient for enhancer identification merely based on primary DNA sequence without exclusively using experimental

data. By defining the molecular logic underlying the function of DNA enhancer elements, it is starting to be possible to identify the regulatory genome based only on DNA sequence. This opens up the possibility of predicting the biological consequences of disease-associated mutations, which are generally located in non-coding regions of the genome. However, we are aware that our predictive model for the HSN still misses additional important elements (activator and/or repressor TFs) and possibly grammar rules (motif positioning and chromatin states) in order to build a proper enhancer language. Long-standing open questions and challenges still remain, mainly: how is regulatory information encoded in the four-letter ‘alphabet’ of enhancer sequences, how is the cross-talk between enhancer–target gene specified in a three dimension genome and if there are additional functions for the key player TFs. Hopefully these questions will be answered in the following years with the use of innovative approaches and state-of-the-art technology. Thus, these are exciting times to continue studying transcriptional regulation.

Conclusions

In this Thesis we have dissected the *cis*-regulatory logic underlying the specification of *C. elegans* serotonergic system and, focusing on the HSN neuron subtype, we have studied how the complement of cell type specific enhancers is selected.

The results obtained in this Thesis lead to the following conclusions:

1 Distinct *cis*-regulatory modules control serotonin pathway gene expression in the different subclasses of serotonergic neurons. This modular and independent subtype regulation is in agreement with the terminal selector model in which, for each cell type, a different combinatorial code of transcription factors directly activates the expression of its terminal features.

2 The serotonin pathway genes *cis*-regulatory modules active in different subclasses are sometimes partially overlapping suggesting that they can be regulated by common transcription factors, or members of the same transcription factor family.

3 Small *cis*-regulatory modules of the serotonin pathway genes occasionally show ectopic expression suggesting that repressor elements may contribute to restricted expression of these genes in the serotonergic neurons.

4 A complex code of six transcription factors is required to induce the serotonergic fate specifically in the HSN neuron subtype. This code, that we have

termed the HSN transcription factor collective, is composed by UNC-86, SEM-4, HLH-3, EGL-46, AST-1 and EGL-18 that belong to the POU, SPALT, HLH, INSM, ETS and GATA transcription factor families, respectively. These six transcription factors directly bind to the regulatory regions of the serotonin pathway genes in order to activate their expression in the HSN neuron.

5 The HSN transcription factor collective is expressed in the HSN neuron at larval stage L4 and acts at the terminal steps of differentiation to establish the terminal fate of the neuron, except for HLH-3, whose expression is restricted to the HSN neuroblast. We propose that, similar to its mouse homologue ASCL1, HLH-3 has a dual role as a proneural and a pioneer factor, sequentially promoting neuronal specification and serotonergic differentiation of the HSN precursor cell, binding to the serotonin pathway genes in closed chromatin states. The expression of the HSN transcription factor collective, except for HLH-3, is required throughout the life of the animal in order to maintain the serotonergic identity of the HSN neuron.

6 The HSN transcription factor collective acts through parallel pathways to activate the expression of the serotonin pathway genes in the HSN. Importantly, UNC-86 appears as a master-regulator, whereas the activity of AST-1 is highly regulated by other members of the code.

7 Ectopic expression of the HSN transcription factor collective is sufficient to induce serotonergic fate, in some cellular contexts.

8 The transcription factors that belong to the HSN collective act cooperatively and redundantly to regulate the expression of the serotonin pathway genes. The individual roles of each member of the HSN transcription factor collective and the synergistic relationships among them depend on the specific DNA regulatory context where they bind.

9 The HSN transcriptome contains a specific signature composed by the clustering of transcription factor binding sites for the six members of the HSN transcription factor collective, that is enriched in neuronal genes and allows for *de novo*

identification of HSN expressed genes. We show, for the first time, that a regulatory signature merely based on primary DNA sequence is sufficient for enhancer identification.

10 The HSN regulatory programme matches best the transcription factor collective model: the six transcription factors that belong to the HSN regulatory code act in a flexible manner, activating enhancers with variable transcription factor binding site order and distribution. Moreover, in agreement with this model, the presence of binding sites for certain transcription factors is dispensable in certain genomic contexts and can be compensated for by the rest of the transcription factor collective. Furthermore, syntactic rules such as transcription factor pair orientation are required for the activation of some HSN enhancers.

11 The regulatory programme of the HSN neuron, but not that of the NSM or ADF neurons, strikingly resembles the serotonergic regulatory programme in mouse: AST-1/PET1, HLH-3/ASCL1, EGL-18/GATA2/3, EGL-46/INSM1 and UNC-86/BRN2 appear

as orthologue transcription factors. This homology allows for the identification of the new regulatory candidates in the worm PHA-4 (FOXA2), and in the mouse SALL2 (SEM-4).

12 *C. elegans* HSN neuron and mouse serotonergic raphe neurons share deep homology, as well as molecular and functional homology. We propose that this deep homology may have arisen from a common ancestor cell type.

Indexes, Annexes and Summary

Bibliography

Abad, P. et al., 2008. Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. *Nature biotechnology*, 26(8), pp.909–915.

Albert, P.R., Le Francois, B. & Millar, A.M., 2011. Transcriptional dysregulation of 5-HT1A autoreceptors in mental illness. *Molecular Brain*, 4(1), p.21.

Albertson, D.G. & Thomson, J. N., 1976. The pharynx of Caenorhabditis elegans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 275, pp.299-325

Allan, D.W. & Thor, S., 2015. Transcriptional selectors, masters, and combinatorial codes: Regulatory principles of neural subtype specification. *Wiley Interdisciplinary Reviews: Developmental Biology*, 4(5), pp.505–528.

Alon, U. 2007. Network motifs: theory and experimental approaches. *Nature Reviews. Genetics,* 8(6), pp.450–61.

Alonso, A. et al., 2013. Development of the serotonergic cells in murine raphe nuclei and their relations with rhombomeric domains. *Brain Structure and Function*, 218(5), pp.1229–1277.

Altun-Gultekin, Z., et al. 2001. A regulatory cascade of three homeobox genes, ceh-10, ttx-3 and ceh-23, controls cell fate specification of a defined interneuron class in C. elegans. *Development*, 128(11), pp.1951–69.

Aow, J.S.Z. et al., 2013. Differential binding of the related transcription factors Pho4 and Cbf1 can tune the sensitivity of promoters to different levels of an induction signal. *Nucleic Acids Research*, 41(9), pp.4877–4887.

Arvey, A. et al., 2012. Sequence and chromatin determinants of cell-type – specific transcription factor binding. *Genome Research*, 22(9), pp.1723–1734.

Asbreuk, C.H.J. et al., 2002. CNS Expression Pattern of Lmx1b and Coexpression with Ptx Genes Suggest Functional Cooperativity in the Development of Forebrain Motor Control Systems. *Molecular and Cellular Neuroscience*, 21(3), pp.410–420.

Azevedo, F.A.C. et al., 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5), pp.532–541.

Badis, G. et al., 2008. A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters. *Molecular Cell*, 32(6), pp.878–887.

Baker, K.G. et al. 1991. Cytoarchitecture of serotonin-synthesizing neurons in the pontine tegmentum of the human brain. *Synapse,* 7, pp-301–320.

Banerji, J., Rusconi, S. & Schaffner, W., 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27 (2 PART 1), pp.299–308.

Bargmann, C.I. & Avery, L., 1995. Laser Killing of Cells in Caenorhabditis elegans. *Methods Cell Biol.*, 48, pp.225–250.

Basson, M. & Horvitz, H.R., 1996. The Caenorhabditis elegans gene sem-4 controls neuronal and mesodermal cell development and encodes a zinc finger protein. *Genes and Development*, 10(15), pp.1953–1965.

Bateson, W., 1911. The Principles of Heredity. *The Edinburgh Review*, 213(435), pp.77–105.

Baum, P.D. et al., 1999. The Caenorhabditis elegans gene ham-2 links Hox patterning to migration of the HSN motor neuron. *Genes and Development*, 13(4), pp.472–483.

Beer, M.A. & Tavazoie, S., 2004. Predicting gene expression from sequence. TL - 117. *Cell*, 117 (2), pp.185–198.

Berger, M.F. & Bulyk, M.L., 2009. Universal protein binding microarrays for the comprehensive characterization of the DNA binding specificities of transcription factors. *Nature Protocol*, 4(3), pp.393–411.

Bertrand, N., Castro, D.S. & Guillemot, F., 2002. Proneural genes and the specification of neural cell types. *Nature Reviews. Neuroscience,* 3 pp.517-530.

Bertrand, V. & Hobert, O., 2009. Linking asymmetric cell division to the terminal differentiation program of postmitotic neurons in C. elegans. *Developmental Cell*, 20(11), pp.563–575.

Biggin, M.D., 2011. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Developmental Cell*, 21(4), pp.611–626.

Birnbaum, R.Y. et al., 2014. Systematic Dissection of Coding Exons at Single Nucleotide Resolution Supports an Additional Role in Cell-Specific Transcriptional Regulation. *PLoS Genetics*, 10(10).

Blakely, R.D. et al. 1991. Cloning and expression of a functional serotonin transporter from rat brain. *Nature,* 354, pp.66–70

Borok, M.J. et al., 2010. Dissecting the regulatory switches of development: lessons from enhancer evolution in Drosophila. *Development*, 137(1), pp.5–13.

Brenner, S., 1974. The genetics of Caenorhabditis elegans. *Genetics*, 77(1), pp.71–94.

Briscoe, J. et al., 1999. Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. *Nature*, 398(6728), pp.622–627.

Bulger, M. & Groudine, M., 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell*, 144(3), pp.327–339.

Bülow, H.E. et al., 2002. Heparan sulfate proteoglycan-dependent induction of axon branching and axon misrouting by the Kallmann syndrome gene kal-1. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), pp.6346–51.

Van Buskirk, C. & Sternberg, P.W., 2010. Paired and LIM class homeodomain proteins coordinate differentiation of the C. elegans ALA neuron. *Development*, 137(12), pp.2065–2074.

Busser, B.W. et al., 2015. Enhancer modeling uncovers transcriptional signatures of individual cardiac cell states in Drosophila. *Nucleic Acids Research*, 43(3), pp.1726–7139.

Carey, M., 1998. The enhanceosome and transcriptional synergy. *Cell*, 92(1), pp.5–8.

Carroll, S.B., 2005. Evolution at two levels: On genes and form. *PLoS Biology*, 3(7), pp.1159–1166.

Cassata, G. et al., 2000. The LIM Homeobox Gene ceh-14 Confers Thermosensory Function to the AFD Neurons in Caenorhabditis elegans. *Neuron*, 25(3), pp.587–597.

Chalfie, M. et al., 1994. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148), pp.802–5.

Chalfie, M., Horvitz, H.R. & Sulston, J.E., 1981. Mutations That Lead to Reiterations in the Cell Lineages of C. elegans. *Genetics*, 153, pp.59–69.

Chalfie, M. & Sulston, J., 1981. Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. *Developmental Biology*, 82(2), pp.358-370.

Chan, S.S.Y. et al., 1996. UNC-40, a C. elegans homolog of DCC (Deleted in Colorectal Cancer), is required in motile cells responding to UNC-6 netrin cues. *Cell*, 87(2), pp.187–195.

Chang, A.J. et al., 2006. A distributed chemosensory circuit for oxygen preference in C. elegans. *PLoS Biology*, 4(9), pp.1588–1602.

Chang, S., Johnston, R.J. & Hobert, O., 2003. A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of c. elegans. *Genes and Development*, 17(17), pp.2123–2137.

Chase, D.L. & Koelle, M.R., 2007. Biogenic amine neurotransmitters in C. elegans. *WormBook : the online review of C. elegans biology*, pp.1–15.

Chen, J. et al., 2006. The ecology and biodemography of Caenorhabditis elegans. *Experimental Gerontology*, 41(10), pp.1059–1065.

Cheng, L. et al., 2003. Lmx1b, Pet-1, and Nkx2.2 coordinately specify serotonergic neurotransmitter phenotype. *The Journal of neuroscience*, 23(31), pp.9961–7.

Cheng, L. et al., 2004. Tlx3 and Tlx1 are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nature Neuroscience*, 7(5), pp.510–517.

Cirillo, L.A. et al., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, 9(2), pp.279–289.

Clark, S.G. & Chiu, C., 2003. C. elegans ZAG-1, a Zn-finger-homeodomain protein, regulates axonal development and neuronal differentiation. *Development* , 130(16), pp.3781–3794.

Collet, J. et al., 1998. Analysis of osm-6, a gene that affects sensory cilium structure and sensory neuron function in Caenorhabditis elegans. *Genetics*, 148(1), pp.187–200.

Collins, K.M. et al., 2016. Activity of the *C. elegans* egg-laying behavior circuit is controlled by competing activation and feedback inhibition. *eLife*, 5, pp.13819–13840.

Cordes, S.P., 2005. Molecular genetics of the early development of hindbrain serotonergic neurons. *Clinical Genetics*, 68(6), pp.487–494.

Craven, S.E. et al., 2004. Gata2 specifies serotonergic neurons downstream of sonic hedgehog. *Development*, 131(5), pp.1165–1173.

Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp.21931–21936.

Dale, H., 1934. Chemical Transmission of the Effects of Nerve Impulses. *British medical journal*, 1(3827), pp.835–41.

Dale, H.H., 1954. The beginnings and the prospects of neurohumoral transmission. *Pharmacological Reviews*, 6(1), pp.7–13.

Darmanis, S. et al., 2015. A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences*, 112(23), p.201507125.

Davidson, E., 2006. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. *Academic Press. pp125-185.*

Davuluri, R. V. et al., 2008. The functional consequences of alternative promoter use in mammalian genomes. *Trends in Genetics*, 24(4), pp.167–177.

Deneris, E.S. & Hobert, O., 2014. Maintenance of postmitotic neuronal cell identity. *Nature neuroscience*, 17(7), pp.899–907.

Deneris, E.S. & Wyler, S.C., 2012a. Serotonergic transcriptional networks and potential importance to mental health. *Nature neuroscience*, 15(4), pp.519–27.

Denker, A. & De Laat, W., 2016. The second decade of 3C technologies: Detailed insights into nuclear organization. *Genes and Development*, 30(12), pp.1357–1382.

Desai, C. et al., 1988. A genetic pathway for the development of the Caenorhabditis elegans HSN motor neurons. *Nature*, 336(6200), pp.638–646.

Desai, C. & Horvitz, H.R., 1989. Caenorhabditis elegans mutants defective in the functioning of the motor neurons responsible for egg laying. *Genetics*, 121(4), pp.703–721.

Dickinson, D.J. et al., 2013. Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. *Nature methods*, 10(10), pp.1028–34.

Dickinson, D.J. et al., 2015. Streamlined genome engineering with a self-excising drug selection cassette. *Genetics*, 200(4), pp.1035–1049.

Dieterich, C. et al., 2008. The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism. *Nature genetics*, 40(10), pp.1193–1198.

Ding, Y.Q. et al., 2003. Lmx1b is essential for the development of serotonergic neurons. *Nature Neuroscience*, 6(9), pp.933–938.

Dittrich, R. et al., 1997. The differentiation of the serotonergic neurons in the Drosophila ventral nerve cord depends on the combined function of the zinc finger proteins Eagle and Huckebein. *Development*, 124(13), pp.2515–2525.

Doitsidou, M. et al., 2013. A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in C. elegans. *Genes and Development*, 27(12), pp.1391–1405.

Doitsidou, M. et al., 2008. Automated screening for mutants affecting dopaminergic-neuron specification in C. elegans. *Nature methods*, 5(10), pp.869–872.

Doonan, R. et al., 2008. HLH-3 is a C. elegans Achaete/Scute protein required for differentiation of the hermaphrodite-specific motor neurons. *Mechanisms of Development*, 125(9–10), pp.883–893.

van Doorninck, J.H. et al., 1999. GATA-3 is involved in the development of serotonergic neurons in the caudal raphe nuclei. *The Journal of neuroscience*, 19(12), p.RC12.

Dray, S., Dufour, A.B. & Chessel, D., 2007. The ade4 Package — II : Two-table and K -table methods. *R News*, 7, pp.47–52.

Duerr, J.S., Gaskin, J. & Rand, J. 2001. Identified neurons in C. elegans coexpress vesicular transporters for acetylcholine and monoamines. *The American Journal of Physiology — Cell Physiology*, 280, pp.1616-1622.

Duerr, J.S. et al., 1999. The *cat-1* gene of *Caenorhabditis elegans* encodes a vesicular monoamine transporter required for specific monoamine-dependent behaviors. *The Journal of neuroscience*, 19(1), pp.72–84.

Duggan, A., Ma, C. & Chalfie, M., 1998. Regulation of touch receptor differentiation by the Caenorhabditis elegans mec-3 and unc-86 genes. *Development*, 125(20), pp.4107–4119.

Eden, E. et al., 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10(1), p.48.

Erceg, J. et al., 2014. Subtle Changes in Motif Positioning Cause Tissue-Specific Effects on Robustness of an Enhancer's Activity. *PLoS Genetics*, 10(1).

Ericson, J. et al., 1997. Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell*, 90(1), pp.169–180.

Esmaeili, B. et al., 2002. The C. elegans even-skipped homologue, vab-7, specifies DB motoneurone identity and axon trajectory. *Development*, 129(4), pp.853–62.

Evans, N.C., Swanson, C.I. & Barolo, S., 2012. Sparkling Insights into Enhancer Structure, Function, and Evolution. *Current Topics in Developmental Biology*, 98, pp.97-121.

Fakhouri, T.H.I. et al., 2010. Dynamic chromatin organization during foregut development mediated by the organ selector gene PHA-4/FoxA. *PLoS Genetics*, 6(8).

Falvo, J. V., Thanos, D. & Maniatis, T., 1995. Reversal of intrinsic DNA bends in the IFNβ gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell*, 83(7), pp.1101–1111.

Farley, E.K. et al., 2016. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National Academy of Sciences of the United States of America*, 113(23), pp.6508–13.

Ferreira, H.B. et al., 1999. Patterning of Caenorhabditis elegans posterior structures by the Abdominal-B homolog, egl-5. *Developmental biology*, 207(1), pp.215–228.

Finney, M., 1987. *The genetics and molecular biology of unc-86, a C. elegans cell lineage gene.* Massachusetts Institute of Technology, Cambridge, Massachusetts. Thesis Disseretation.

Finney, M. & Ruvkun, G., 1990. The unc-86 gene product couples cell lineage and cell identity in Caenorhabditis elegans. *Cell*, 63, pp.895–905.

Fire, A., Harrison, S.W. & Dixon, D., 1990. A modular set of lacZ fusion vectors for studying gene expression in Caenorhabditis elegans. *Gene*, 93(2), pp.189–190.

Fisher, W.W. et al., 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 109(52), pp.21330–21335.

Flames, N. & Hobert, O., 2009. Gene regulatory logic of dopaminergic neuron differentiation. *Nature*, 458(7240), pp.885–889.

Flames, N. & Hobert, O., 2011. Transcriptional control of the terminal fate of monoaminergic neurons. *Annual review of neuroscience*, 34, pp.153–184.

Fox, S.R. & Deneris, E.S., 2012. Engrailed is required in maturing serotonin neurons to regulate the cytoarchitecture and survival of the dorsal raphe nucleus. *The Journal of neuroscience*, 32(23), pp.7832–42.

Fradkov, A.F. et al., 2000. Novel fluorescent protein from Discosoma coral and its mutants possesses a unique far-red fluorescence. *FEBS letters*, 479(3), pp.127–130.

Friedland, A.E. et al., 2013. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nature methods*, 10(8), pp.741–3.

Fujiwara, Y. et al., 2004. Functional overlap of GATA-1 and GATA-2 in primitive hematopoietic development. *Blood*, 103(2), pp.583–585.

Furuhata, a et al., 2009. GATA-1 and GATA-2 binding to 3' enhancer of WT1 gene is essential for its transcription in acute leukemia and solid tumor cell lines. *Leukemia*, 23(7), pp.1270–1277.

Fyodorov, D. Nelson, T. & Deneris, E., 1998. Pet-1, a novel ETS domain factor that can activate neuronal nAchR gene transcription. *Journal of neurobiology*, 34(2), pp.151-63.

Garcia-Bellido, A., 1975. Genetic control of wing disc development in Drosophila. *Ciba Foundation Symposium*, 29, pp.161–82

Garriga, G., Desai, C. & Horvitz, H.R., 1993. Cell interactions control the direction of outgrowth, branching and fasciculation of the HSN axons of Caenorhabditis elegans. *Development (Cambridge, England)*, 117, pp.1071–1087.

Garriga, G., Guenther, C. & Horvitz, H.R., 1993. Migrations of the Caenorhabditis elegans HSNs are regulated by egl-43, a gene encoding two zinc finger proteins. *Genes and development*, 7(11), pp.2097–2109.

Gendrel, M., Atlas, E.G. & Hobert, O., 2016. A cellular and regulatory map of the GABAergic nervous system of *C. elegans*. *eLife*, 5, pp.1–38.

German, M.S. et al., 1992. Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: Building a functional insulin minienhancer complex. *Genes and Development*, 6(11), pp.2165–2176.

Ghedin, E. et al., 2007. Supporting Online Material for Draft Genome of the Filarial Nematode Parasite Brugia malayi. *Science*, 317.

Gibson, D.G. et al., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, 6(5), pp.343–5.

Gillette, R., 2006. Evolution and function in serotonergic systems. *Integrative and Comparative Biology*, 46(6), pp.838–846.

Ginjaar, I.B. et al., 2000. Dystrophin nonsense mutation induces different levels of exon 29 skipping and leads to variable phenotypes within one BMD family. *European journal of human genetics : EJHG*, 8(10), pp.793–6.

Gómez-Skarmeta, J.L., Campuzano, S. & Modolell, J., 2003. Half a century of neural prepatterning: the story of a few bristles and many genes. *Nature reviews. Neuroscience*, 4(7), pp.587–598.

Gordon, P.M. & Hobert, O., 2015. A competition mechanism for a homeotic neuron identity transformation in C. elegans. *Developmental cell*, 34(2), pp.206–219.

Gorrepati, L., Thompson, K.W. & Eisenmann, D.M., 2013. C. elegans GATA factors EGL-18 and ELT-6 function downstream of Wnt signaling to maintain the progenitor fate during larval asymmetric divisions of the seam cells. *Development*, 140(10), pp.2093–2102.

Gotea, V. et al., 2010. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Research*, 20, pp.565–577.

Grant, K., Hanna-Rose, W. & Han, M., 2000. sem-4 promotes vulval cell-fate determination in Caenorhabditis elegans through regulation of lin-39 Hox. *Developmental biology*, 224(2), pp.496–506.

Grossman, S.R. et al., 2017. Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proceedings of the National Academy of Sciences*, E1291-E1300.

Grove, C.A. et al., 2009. A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors. *Cell*, 138(2), pp.314–327.

Guenther, C. & Garriga, G., 1996. Asymmetric distribution of the C. elegans HAM-1 protein in neuroblasts enables daughter cells to adopt distinct fates. *Development*, 3518(11), pp.3509–3518.

Guss, K.A. et al., 2001. Control of a Genetic Regulatory Network by a Selector Gene. *Science*, 292(5519), pp.1164-1167.

Hamdan, F.F. et al., 1999. Characterization of a novel serotonin receptor from Caenorhabditis elegans: cloning and expression of two splice variants. *Journal of neurochemistry*, 72(4), pp.1372–1383.

Harbison, C.T. et al., 2004. Scoring and Clustering and Conservation Testing Assignment of Single Motif to Each Regulator. *Nature*, 431(7004), pp.1–5.

Hardaker, L.A. et al., 2001. Serotonin modulates locomotory behavior and coordinates egg-laying and movement in Caenorhabditis elegans. *Journal of Neurobiology*, 49(4), pp.303–313.

Hare, E. & Loer, C., 2004. Function and evolution of the serotonin-synthetic bas-1 gene and other aromatic amino acid decarboxylase genes in Caenorhabditis. *BMC Evolutionary Biology*, 4(1), p.24.

Haugas, M. et al., 2016. Gata2 and Gata3 regulate the differentiation of serotonergic and glutamatergic neuron subtypes of the dorsal raphe. *Development*, 1, pp.4495–4508.

Hawrylycz, M.J. et al., 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489(7416), pp.391–399.

Hedgecock, E.M. et al., 1985. Axonal guidance mutants of Caenorhabditis elegans identified by filling sensory neurons with fluorescein dyes. *Developmental Biology*, 111(1), pp.158–170.

Heinz, S. et al., 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, 38(4), pp.576–589.

Heinz, S. et al., 2015. The selection and function of cell type-specific enhancers. *Nature reviews. Molecular cell biology*, 16(3), pp.144–54.

Hendricks, T. et al., 1999. The ETS domain factor Pet-1 is an early and precise marker of central serotonin neurons and interacts with a conserved element in serotonergic genes. *The Journal of neuroscience*, 19(23), pp.10348–10356.

Hendricks, T.J. et al., 2003. Pet-1 ETS gene plays a critical role in 5-HT neuron development and is required for normal anxiety-like and aggressive behavior. *Neuron*, 37(2), pp.233–247.

Hobert, O., 2002. PCR fusion-based approach to create reporter Gene constructs for expression analysis in transgenic C. elegans. *BioTechniques*, 32(4), pp.728–730.

Hobert O. 2010. Neurogenesis in the nematode Caenorhabditis elegans. *WormBook: The Online Review of C. elegans Biology*.

Hobert, O., 2011. Regulation of terminal differentiation programs in the nervous system. *Annual review of cell and developmental biology*, 27(1), pp.681–96.

Hobert, O., 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*, 105(51), pp.20067–71.

Hobert, O., 2005. Specification of the nervous system. *WormBook : the online review of C. elegans biology*, pp.1–19.

Hobert, O., 2016. Terminal Selectors of Neuronal Identity. *Current Topics in Developmental Biology*, 116, pp.445-475.

Hobert, O., 2013. The neuronal genome of Caenorhabditis elegans. *WormBook: The Online Review of C. Elegans Biology*, pp.1–22.

Hobert, O., Glenwinkel, L. & White, J., 2016. Revisiting Neuronal Cell Type Classification in Caenorhabditis elegans. *Current Biology*, 26(22), pp.1197–1203.

Hobson, R.J. et al., 2003. SER-7b, a constitutively active Gαs coupled 5-HT 7-like receptor expressed in the Caenorhabditis elegans M4 pharyngeal motorneuron. *Journal of Neurochemistry*, 87(1), pp.22–29.

Holmberg, J. & Perlmann, T., 2012. Maintaining differentiated cellular identity. *Nature Reviews. Genetics*, 13(6), pp.429–439.

Holmes, A., 2008. Genetic variation in cortico-amygdala serotonin function and risk for stress-related disease. *Neuroscience and Biobehavioral Reviews*, 32(7), pp.1293–1314.

Hong, J.-W., Hendrix, D.A. & Levine, M.S., 2008. Shadow Enhancers as a Source of Evolutionary Novelty. *Science*, 321(5894), p.1314.

Hornung, J.P. 2003. The human raphe nuclei and the serotonergic system. *Journal of Chemical Neuroanatomy*, 26, pp.331–343.

Horvitz, H.R. et al., 1982. Serotonin and Octopamine in the Nematode Caenorhabditis elegans. *Science*, 216(4549), pp.1012-1014.

Hsu, P.D. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), pp.827–32.

Huber, W. et al., 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, 12(2), pp.115–121.

Iborra, F.J. et al., 1996. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *Journal of cell science*, 109, pp.1427–1436.

Inoue, F. & Ahituv, N., 2015. Decoding enhancers using massively parallel reporter assays. *Genomics*, 106(3), pp.159–164.

Ishimura, K. et al. 1988. Quantitative analysis of the distribution of serotonin-immunoreactive cell bodies in the mouse brain. *Neuroscience Letters*, 91, pp.265–270.

Jackson, D.A. et al., 1993. Visualization of focal sites of transcription within human nuclei. *The EMBO journal*, 12(3), pp.1059–65.

Jacob, J. et al., 2009. Insm1 (IA-1) is an essential component of the regulatory network that specifies monoaminergic neuronal phenotypes in the vertebrate hindbrain. *Development*, 136(14), pp.2477–85.

Jacobs, B. & Azmitia, E., 1992. Structure and function of the brain serotonin system. *Physiological Reviews*, 72(1), pp.165–229.

Jacobs, F.M.J., van der Linden, A.J. a, et al., 2009. Identification of Dlk1, Ptpru and Klhl1 as novel Nurr1 target genes in meso-diencephalic dopamine neurons. *Development*, 136(14), pp.2363–2373.

Jacobs, F.M.J., van Erp, S., et al., 2009. Pitx3 potentiates Nurr1 in dopamine neuron terminal differentiation through release of SMRT-mediated repression. *Development (Cambridge, England)*, 136(4), pp.531–540.

Jafari, G. et al., 2011. Regulation of extrasynaptic 5-HT by serotonin reuptake transporter function in 5-HT-absorbing neurons underscores adaptation behavior in Caenorhabditis elegans. *The Journal of neuroscience*, 31(24), pp.8948–57.

Jarrell, T.A. et al., 2012. The Connectome of a Decision-Making Neural Network. *Science*, 337, pp.437–444.

Jarriault, S., Schwab, Y. & Greenwald, I., 2008. A Caenorhabditis elegans model for epithelial-neuronal transdifferentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(10), pp.3790–3795.

Jolma, A. et al., 2013. DNA-binding specificities of human transcription factors. *Cell*, 152(1–2), pp.327–339.

Junion, G. et al., 2012. A transcription factor collective defines cardiac cell

fate and reflects lineage history. *Cell*, 148(3), pp.473–486.

Kaestner, K.H., 2010. The FoxA factors in organogenesis and differentiation. *Current Opinion in Genetics and Development*, 20(5), pp.527–532.

Kage, E. et al., 2005. MBR-1, a novel helix-turn-helix transcription factor, is required for pruning excessive neurites in Caenorhabditis elegans. *Current Biology*, 15(17), pp.1554–1559.

Kala, K. et al., 2009. Gata2 is a tissue-specific post-mitotic selector gene for midbrain GABAergic neurons. *Development*, 136(2), pp.253–262.

Kamath, R.S. et al., 2001. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans. *Genome biology*, 2(1).

Kamath, R.S. & Ahringer, J., 2003. Genome-wide RNAi screening in Caenorhabditis elegans. *Methods*, 30(4), pp.313–321.

Kazemian, M. et al., 2013. Widespread evidence of cooperative DNA binding by transcription factors in Drosophila development. *Nucleic Acids Research*, 41(17), pp.8237–8252.

Kelly, W.G. et al., 1997. Distinct requirements for somatic and germline expression of a generally expressed Caenorhabditis elegans gene. *Genetics*, 146(1), pp.227–238.

Kerk, S.Y. et al., 2017. Diversification of C . elegans Motor Neuron Identity Article Diversification of C . elegans Motor Neuron Identity via Selective Effector Gene Repression. *Neuron*, 93(1), pp.80–98.

Kheradpour, P. et al., 2013. Systematic dissection of motif instances using a massively parallel reporter assay. *Genome Research*, 23, pp.800–811.

Kim, J. et al., 2001. Genes affecting the activity of nicotinic receptors involved in Caenorhabditis elegans egg-laying behavior. *Genetics*, 157(4), pp.1599–1610.

Kim, K. & Li, C., 2004. Expression and regulation of an FMRFamide-related neuropeptide gene family in Caenorhabditis elegans. *Journal of Comparative Neurology*, 475(4), pp.540–550.

Kiyasova, V. et al., 2011. A genetically defined morphologically and functionally unique subset of 5-HT neurons in the mouse raphe nuclei. *The Journal of neuroscience*, 31(8), pp.2756–2768.

Koh, K. et al., 2002. Cell fates and fusion in the C. elegans vulval primordium are regulated by the EGL-18 and ELT-6 GATA factors — apparent direct targets of the LIN-39 Hox protein. *Development*, 129, pp.5171–5180.

Koh, K. & Rothman, J.H., 2001. ELT-5 and ELT-6 are required continuously to regulate epidermal seam cell differentiation and cell fusion in C. elegans. *Development*, 128, pp.2867–2880.

Kramer, J.M. et al., 1990. The Caenorhabditis elegans rol-6 Gene, Which Interacts with the sqt-1 Collagen Gene To Determine Organismal Morphology, Encodes a Collagen. *Molecular and Cellular Biology*, 10(5), pp.2081–2089.

Kratsios, P. et al., 2011. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nature neuroscience*, 15(2), pp.205–214.

Kratsios, P. et al., 2015. Transcriptional coordination of synaptogenesis and neurotransmitter signaling. *Current Biology*, 25(10), pp.1282–1295.

Krause, M. et al., 1997. A C. elegans E/Daughterless bHLH protein marks neuronal but not striated muscle development. *Development*, 124(11), pp.2179–2189.

Kulkarni, M.M. & Arnosti, D.N., 2003. Information display by transcriptional enhancers. *Development*, 130(26), pp.6569–6575.

Kullyev, A. et al., 2010. A genetic survey of fluoxetine action on synaptic transmission in Caenorhabditis elegans. *Genetics*, 186(3), pp.929–941.

Kwasnieski, J.C. et al., 2014. High-throughput functional testing of ENCODE segmentation predictions. *Genome Research*, 24(10), pp.1595–1602.

Lahti, L. et al., 2016. Differentiation and molecular heterogeneity of inhibitory and excitatory neurons associated with midbrain dopaminergic nuclei. *Development*, 143(3), pp.516–529.

Lai, C.H. et al., 2000. Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics. *Genome research*, 10(5), pp.703–13.

Leddin, M. et al., 2011. Two distinct auto-regulatory loops operate at the PU . 1 locus in B cells and myeloid cells. *Hematopoiesis and Stem Cells*, 117(10), pp.2827–2839.

Lee, H.K. et al., 2010. A novel frameshift mutation of POU4F3 gene associated with autosomal dominant non-syndromic hearing loss. *Biochemical and Biophysical Research Communications*, 396(3), pp.626–630.

Lerch-Haner, J.K. et al., 2008. Serotonergic transcriptional programming determines maternal behavior and offspring survival. *Nature neuroscience*, 11(9), pp.1001–1003.

Levo, M. & Segal, E., 2014. In pursuit of design principles of regulatory sequences. *Nature reviews. Genetics*, 15(7), pp.453–68.

Li, R., Pei, H. & Watson, D.K., 2000. Regulation of Ets function by protein - protein interactions. *Oncogene*, 19, pp.6514–6523.

Lifanov, A.P., 2003. Homotypic Regulatory Clusters in Drosophila. *Genome Research*, 13(4), pp.579–588.

Lillesaar, C. et al., 2007. The serotonergic phenotype is acquired by converging genetic mechanisms within the zebrafish central nervous system. *Developmental Dynamics*, 236(4), pp.1072–1084.

Lis, J. & Wu, C., 1993. Protein traffic on the heat shock promoter: Parking, stalling, and trucking along. *Cell*, 74(1), pp.1–4.

Liu, C. et al., 2010. Pet-1 is required across different stages of life to regulate serotonergic function. *Nature neuroscience*, 13(10), pp.1190–1198.

Liu, Y. & Edwards, R.H., 1997. The Role of Vesicular Transport Proteins in Synaptic Transmission and Neural Degeneration. *Annual Review of Neuroscience*, 20(1), pp.125–156.

Lodato, S. et al., 2014. Gene co-regulation by Fezf2 selects neurotransmitter identity and connectivity of corticospinal neurons. *Nature neuroscience*, 17(8), pp.1046–1054.

Loer, C.M. et al., 2015. Cuticle integrity and biogenic amine synthesis in Caenorhabditis elegans require the cofactor tetrahydrobiopterin (BH4). *Genetics*, 200(1), pp.237–253.

Loer, C.M. & Kenyon, C.J., 1993. Serotonin-deficient mutants and male mating behavior in the nematode Caenorhabditis elegans. *The Journal of neuroscience*, 13(12), pp.5407–5417.

Loer, C.M. & Rand, J.B., 2016. The Evidence for Classical Neurotransmitters in *Caenorhabditis elegans*. *WormAtlas*.

Loewi, O., 1954. Introduction. *Pharmacological Reviews*, 6(1), pp.1–6.

López-Arvizu, C. et al., 2011. Increased symptoms of attention deficit hyperactivity disorder and major depressive disorder symptoms in nail-patella syndrome: Potential association with LMX1B loss-of-function. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 156(1), pp.59–66.

López-Muñoz, F., Boya, J. & Alamo, C., 2006. Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal. *Brain Research Bulletin*, 70(4–6), pp.391–405.

Lundell, M.J. & Hirsh, J., 1998. eagle is required for the specification of serotonin neurons and other neuroblast 7-3 progeny in the Drosophila CNS. *Development*, 125, pp.463–472.

Maduro, M.F. et al., 2005. Genetic redundancy in endoderm specification within the genus Caenorhabditis. *Developmental Biology*, 284(2), pp.509–522.

Maduro, M.F., 2017. Gut development in C. elegans. *Seminars in Cell and Developmental Biology*, (2016).

Maerkl, S.J. & Quake, S.R., 2007. A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science*, 315(5809), pp.233–237.

Maeso, I., Acemel, R.D. & Gómez-Skarmeta, J.L., 2016. Cis-regulatory landscapes in development and evolution. *Current opinion in genetics and development*, 43, pp.17–22.

Markstein, M. et al., 2002. Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proceedings of the National Academy of Sciences of the United States of America*, 99(2), pp.763–768.

Mathelier, A., Shi, W. & Wasserman, W.W., 2015. Identification of altered cis-regulatory elements in human disease. *Trends in Genetics*, 31(2), pp.67–76.

Mazzoni, E.O. et al., 2013. Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nature neuroscience*, 16(9), pp.1219–27.

McKay, R. J, 2003. Gene Expression Profiling of Cells, Tissues, and Developmental Stages of the Nematode *C. elegans*. *Cold Spring Harbour Symposium Quanitativet Biology*, 68, pp159-170.

McIntire, S.L. et al., 1992. Genes necessary for directed axonal elongation or fasciculation in C. elegans. *Neuron*, 8(2), pp.307–322.

Mello, C.C. et al., 1991. Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO Journal*, 10(12), pp.3959–3970.

Merika, M. & Orkin, S.H., 1993. DNA-binding specificity of GATA family transcription factors. *Molecular and cellular biology*, 13(7), pp.3999–4010.

Merika, M. & Thanos, D., 2001. Enhanceosomes. *Current Opinion in Genetics and Development*, 11(2), pp.205–208.

Miller, J. a & Widom, J., 2003. Collaborative Competition Mechanism for Gene Activation In Vivo. *Molecular and Cellular Biology*, 23(5), pp.1623–1632.

Miller, D.M., 1992. *C. elegans unc-4* gene encodes a homeodomain protein that determines the pattern of synaptic input to specific motor neurons. *Nature*, 355(6363), pp 841-845.

Milo, R. et al., 2010. BioNumbers The database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38, pp.750–753.

Mogno, I., Kwasnieski, J.C. & Cohen, B.A., 2013. Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome research*,23, pp.1908–1915.

Molyneaux, B.J. et al., 2015. DeCoN: Genome-wide analysis of invivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron*, 85(2), pp.275–288.

Montavon, T. et al., 2011. A regulatory archipelago controls hox genes transcription in digits. *Cell*, 147(5), pp.1132–1145.

Morishita, K. et al., 1988. Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell*, 54(6), pp.831–840.

Mühleisen, T.W. et al., 2014. Genome-wide association study reveals two new risk loci for bipolar disorder. *Nature communications*, 5, p.3339.

Muhr, J. et al., 2001. Groucho-mediated transcriptional repression establishes progenitor cell pattern and neuronal fate in the ventral neural tube. *Cell*, 104(6), pp.861–873.

Muotri, A.R. & Gage, F.H., 2006. Generation of neuronal variability

and complexity. *Nature*, 441(7097), pp.1087–1093.

Murgan, S. et al., 2015. Atypical Transcriptional Activation by TCF via a Zic Transcription Factor in C. elegans Neuronal Precursors. *Developmental Cell*, 33(6), pp.737–745.

Murphy, D.L. et al., 2008. How the serotonin story is being rewritten by new gene-based discoveries principally related to SLC6A4, the serotonin transporter gene, which functions to influence all cellular serotonin systems. *Neuropharmacology*, 55(6), pp.932–960.

Nasu, M. et al., 2014. Mammalian-specific sequences in Pou3f2 contribute to maternal behavior. *Genome Biology and Evolution*, 6(5), pp.1145–1156.

Nomaksteinsky, M. et al., 2013. Ancient origin of somatic and visceral neurons. *BMC biology*, 11, p.53.

Nutiu, R. et al., 2011. Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature Biotechnology*, 29(7), pp.659–664.

Okaty, B.W. et al., 2015. Multi-Scale Molecular Deconstruction of the Serotonin Neuron System. *Neuron*, 88(4), pp.774–791.

Olde, B. & McCombie, W.R., 1997. Molecular cloning and functional expression of a serotonin receptor from Caenorhabditis elegans. *Journal of molecular neuroscience*, 8(1), pp.53–62.

Opperman, C.H. et al., 2008. Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39), pp.14802–14807.

Pakkenberg, B. et al., 2003. Aging and the human neocortex. *Experimental Gerontology*, 38(1–2), pp.95–99.

Parker, S.C.J. et al., 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences of the United States of America*, 110(44), pp.17921–17926.

Patel, T. & Hobert, O., 2017. Coordinated control of terminal differentiation and restriction of cellular plasticity. *eLife*, 6, pp.1–26.

Pattyn, A. et al., 2004. Ascl1/Mash1 is required for the development of central serotonergic neurons. *Nature Neuroscience*, 7(6), pp.589–595.

Pattyn, A. et al., 2003. Coordinated temporal and spatial control of motor neuron and serotonergic neuron generation from a common pool of CNS progenitors. *Genes and Development*, 17(6), pp.729–737.

Pelham, H.R.B., 1982. A regulatory upstream promoter element in Drosphila hsp 70 heat shock gene. *Cell*, 30, pp.517–528.

Pereira, L. et al., 2015. A cellular and regulatory map of the cholinergic nervous system of C. Elegans. *eLife*, 4, pp.1–46.

Peterkin, T. et al., 2005. The roles of GATA-4, -5 and -6 in vertebrate heart development. *Cell and Developmental Biology*, 16(1), pp.83–94.

Phillips-Cremins, J.E. et al., 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153(6), pp.1281–1295.

Pirri, J.K. et al., 2009. A Tyramine-Gated Chloride Channel Coordinates Distinct Motor Programs of a Caenorhabditis elegans Escape Response. *Neuron*, 62(4), pp.526–538.

Pocock, R. & Hobert, O., 2010. Hypoxia activates a latent circuit for processing gustatory information in C. elegans. *Nature neuroscience*, 13(5), pp.610–614.

Portman, D.S. & Emmons, S.W., 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*, 127(24), pp.5415–5426.

Prasad, H.C. et al., 2005. Human serotonin transporter variants display altered sensitivity to protein kinase G and p38 mitogen-activated protein kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32), pp.11545–11550.

Ptashne, M. & Gann, A., 1997. Transcriptional activation by recruitment. *Nature*, 386, pp.569–577.

Rada-Iglesias, A. et al., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), pp.279–283.

Ramamoorthy, S. et al., 1993. Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. *Proceedings of the National Academy of Sciences of the United States of America*, 90(6), pp.2542–2546.

Ramón y Cajal, S., 1909. Histologie du Systeme Nerveux de l'Homme et des Vertebres. Maloine, Paris: 1911. chap. II. *Demography*, v.1(90), pp.3–43.

Rand, J.B. & Nonet, M.L., 1997. Neurotransmitter assignments for specific neurons. *C. elegans*, II, pp.1049–1052.

Ranganathan, R. et al., 2001. Mutations in the Caenorhabditis elegans serotonin reuptake transporter MOD-5 reveal serotonin-dependent and -independent activities of fluoxetine. *The Journal of neuroscience*, 21(16), pp.5871–84.

Raposo, A.A.S.F. et al., 2015. Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis. *Cell Reports*, 10(9), pp.1544–1556.

Reiter, F., Wienerroither, S. & Stark, A., 2017. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics and Development*, 43, pp.73–81.

Reyes, J.C., Muro-Pastor, M.I. & Florencio, F.J., 2004. The GATA family of transcription factors in Arabidopsis and rice. *Plant physiology*, 134, pp.1718–1732.

Rhee, H.S. & Pugh, B.F., 2011. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, 147(6), pp.1408–1419.

Richardson-Jones, J.W. et al., 2011. Serotonin-1A autoreceptors are necessary and sufficient for the normal formation of circuits underlying innate anxiety. *The Journal of neuroscience*, 31(16), pp.6008–6018.

Rual, J.-F. et al., 2004. Toward Improving Caenorhabditis elegans Phenome Mapping With an ORFeome-Based RNAi Library. *Genome Research*, 14, pp.2162–2168.

Ruan, J. et al., 2008. TreeFam: 2008 Update. *Nucleic Acids Research*, 36, pp.735–740.

Sammut, M. et al., 2015. Glia-derived neurons are required for sex-specific learning in C. elegans. *Nature*, 526(7573), pp.385–390.

Sanyal, A. et al., 2012. The long-range interaction landscape of gene promoters. *Nature*, 489(7414), pp.109–113.

Sarov, M. et al., 2012. A genome-scale resource for in vivo tag-based protein function exploration in C. elegans. *Cell*, 150(4), pp.855–866.

Sarov, M. et al., 2006. A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. *Nature methods*, 3(10), pp.839–844.

Sawin, E.R., Ranganathan, R. & Horvitz, H.R., 2000. C. elegans Locomotory Rate Is Modulated by the Environment through a Dopaminergic Pathway and by Experience through a Serotonergic Pathway. *Neuron*, 26(3), pp.619–631.

Scheuch, K. et al., 2007. Characterization of a Functional Promoter Polymorphism of the Human Tryptophan Hydroxylase 2 Gene in Serotonergic Raphe Neurons. *Biological Psychiatry*, 62(11), pp.1288–1294.

Schmid, C., Schwarz, V. & Hutter, H., 2006. AST-1, a novel ETS-box transcription factor, controls axon guidance and pharynx development in C. elegans. *Developmental Biology*, 293(2), pp.403–413.

Schulz, K.N. et al., 2015. Zelda is differentially required for chromatin accessibility, transcription-factor binding and gene expression in the early Drosophila embryo. *Genome research*, 25, pp.1715–1726.

Scott, M.M., 2005. A Differentially Autoregulated Pet-1 Enhancer Region Is a Critical Target of the Transcriptional Cascade That Governs Serotonin Neuron Development. *Journal of Neuroscience*, 25(10), pp.2628–2636.

Segal, E. et al., 2008. Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature*, 451(7178), pp.535–540.

Segal, E. & Widom, J., 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Current Opinion in Structural Biology*, 19(1), pp.65–71.

Serrano-Saiz, E. et al., 2013. Modular control of glutamatergic neuronal identity in C. elegans by distinct homeodomain proteins. *Cell*, 155(3), pp.659–673.

Sharon, E. et al., 2012. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotechnology*, 30(6), pp.521–530.

Shaye, D.D. & Greenwald, I., 2011. Ortholist: A compendium of C. elegans genes with human orthologs. *PLoS ONE*, 6(5).

Shen, Z. et al., 2014. Conditional knockouts generated by engineered CRISPR-Cas9 endonuclease reveal the roles of coronin in c.elegans neural development. *Developmental Cell*, 30(5), pp.625–636.

Shi, X. et al., 2014. Cooperative interaction of Etv2 and Gata2 regulates the development of endothelial and hematopoietic lineages. *Developmental Biology*, 389(2), pp.208–218.

Shimodaira, H., 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, 51(3), pp.492–508.

Shubin, N., Tabin, C. & Carroll, S., 1997. Fossils, genes and the evolution of animal limbs. *Nature*, 388(6643), pp.639–48.

Shyn, S.I., Kerr, R. & Schafer, W.R., 2003. Serotonin and Go Modulate Functional States of Neurons and Muscles Controlling C. elegans Egg-Laying Behavior. *Current Biology*, 13(21), pp.1910–1915.

Siggers, T. & Gordan, R., 2014. Protein-DNA binding: Complexities and multi-protein codes. *Nucleic Acids Research*, 42(4), pp.2099–2111.

Simmer, F. et al., 2002. Loss of the Putative RNA-Directed RNA Polymerase RRF-3 Makes C. elegans Hypersensitive to RNAi. *Current Biology*, 12(15), pp.1317–1319.

Singhvi, A., Frank, C.A. & Garriga, G., 2008. The T-box gene tbx-2, the homeobox gene egl-5 and the asymmetric cell division gene ham-1 specify neural fate in the HSN/PHB lineage. *Genetics*, 179(2), pp.887–898.

Slattery, M. et al., 2011. Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. *Cell*, 147(6), pp.1270–1282.

Smith, R.P. et al., 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nature genetics*, 45(9), pp.1021–1028.

Sokal, R.R. & Michener, C.D., 1958. A Statistical Method for Evaluating Systematic Relationships. *University of Kansas Science Bulletin*, 38, pp.1409-1438

Song, N.N. et al., 2011. Adult raphe-specific deletion of Lmx1B leads to central serotonin deficiency. *PLoS ONE*, 6(1).

Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: From enhancer binding to developmental control. *Nature reviews. Genetics*, 13(9), pp.613–626.

Sporns, O., 2011. The human connectome: A complex network. *Annals of the New York Academy of Sciences*, 1224(1), pp.109–125.

Stefanakis, N., Carrera, I. & Hobert, O., 2015. Regulatory Logic of Pan-Neuronal Gene Expression in C. elegans. *Neuron*, 87(4), pp.733–750.

Stein, L.D. et al., 2003. The genome sequence of Caenorhabditis briggsae: A platform for comparative genomics. *PLoS Biology*, 1(2).

Stern, C.D., 2006. Neural induction: 10 years on since the 'default model.' *Current Opinion in Cell Biology*, 18(6), pp.692–697.

Strausfeld, N.J. & Hirth, F., 2013. Deep Homology of Arthropod Central Complex and Vertebrate Basal Ganglia. *Science*, 340(6129), pp.157–161.

Struhl, K. & Segal, E., 2013. Determinants of nucleosome positioning. *Nature structural and molecular biology*, 20(3), pp.267–273.

Sulston, J.E. et al., 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Developmental Biology*, 100(1), pp.64–119.

Sutcliffe, J.S. et al., 2005. Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors. *The American Journal of Human Genetics*, 77(2), pp.265–279.

Suzuki, R. & Shimodaira, H., 2006. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), pp.1540–1542.

Swanson, C.I., Evans, N.C. & Barolo, S., 2010. Structural Rules and Complex Regulatory Circuitry Constrain Expression of a Notch- and EGFR-Regulated Eye Enhancer. *Developmental Cell*, 18, pp.359–370.

Swoboda, P., Adler, H.T. & Thomas, J.H., 2000. The RFX-Type Transcription Factor DAF-19 Regulates Sensory Neuron Cilium Formation in C. elegans. *Molecular Cell*, 5(3), pp.411–421.

Sze, J.Y. et al., 2000. Food and metabolic signalling defects in a *Caenorhabditis elegans* serotonin-synthesis mutant. *Nature*, 403(6769), pp.560–564.

Sze, J.Y. et al., 2002. The *C. elegans* POU-domain transcription factor UNC-86 regulates the *tph-1* tryptophan hydroxylase gene and neurite outgrowth in specific serotonergic neurons. *Development*, 129(16), pp.3901–3911.

Tanay, A., 2006. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Research*, 16(8), pp.962–972.

Tanis, J.E. et al., 2008. Regulation of serotonin biosynthesis by the G proteins Gαo and Gαq controls serotonin signaling in Caenorhabditis elegans. *Genetics*, 178(1), pp.157–169.

Tanis, J.E. et al., 2009. The potassium chloride cotransporter KCC-2 coordinates development of inhibitory neurotransmission and synapse structure in C. elegans. *Journal of Neuroscience*, 29(32), pp.9943–9954.

Thanos, D. & Maniatis, T., 1995. Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell*, 83(7), pp.1091–1100.

The C. elegans Deletion Mutant Consortium, 2012. Large-Scale Screening for Targeted Knockouts in the Caenorhabditis elegans Genome. *G3 Genes|Genomes|Genetics*, 2(11), pp.1415–1425.

The C. elegans Sequencing Consortium, 1998. Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science*, 282(5396), pp.2012–2018.

The R Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.

Timmons, L. & Fire, a, 1998. Specific interference by ingested dsRNA. *Nature*, 395(6705), p.854.

Toker, A.S. et al., 2003. The Caenorhabditis elegans spalt-like gene sem-4 restricts touch cell fate by repressing the selector Hox gene egl-5 and the effector gene mec-3. *Development*, 130(16), pp.3831–3840.

Tomer, R. et al., 2010. Profiling by Image Registration Reveals Common Origin of Annelid Mushroom Bodies and Vertebrate Pallium. *Cell*, 142(5), pp.800–809.

Topalidou, I. & Chalfie, M., 2011. Shared gene expression in distinct neurons expressing common selector genes. *Proceedings of the National Academy of Sciences*, 108(48), pp.19258–19263.

Trent, C., Tsuing, N. & Horvitz, H.R., 1983. Egg-laying defective mutants of the nematode Caenorhabditis elegans. *Genetics*, 104(4), pp.619–647.

Tursun, B. et al., 2009. A toolkit and robust pipeline for the generation of fosmid-based reporter genes in C. elegans. *PLoS ONE*, 4(3).

Tursun, B. et al., 2011. Direct conversion of C. elegans germ cells into specific neuron types. *Science*, 331(6015), pp.304–308.

Updike, D.L. & Mango, S.E., 2006. Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. *PLoS Genetics*, 2(9), pp.1500–1510.

Venters, B.J. et al., 2011. A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in Saccharomyces. *Molecular Cell*, 41(4), pp.480–492.

Verrijzer, C.P. et al., 1992. The DNA binding specificity of the bipartite POU domain and its subdomains. *The EMBO journal*, 11(13), pp.4993–5003.

Villar, D., Flicek, P. & Odom, D.T., 2014. Evolution of transcription factor binding in metazoans-mechanisms and functional implications. *Nature reviews. Genetics*, 15(4), pp.221–233.

Voss, T.C. et al., 2011. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell*, 146(4), pp.544–554.

Waggoner, L.E. et al., 1998. Control of alternative behavioral states by serotonin in Caenorhabditis elegans. *Neuron*, 21(1), pp.203–214.

Waider, J. et al., 2011. Tryptophan hydroxylase-2 (TPH2) in disorders of cognitive control and emotion regulation: A perspective. *Psychoneuroendocrinology*, 36(3), pp.393–405.

Waterston, R. & Sulston, J., 1995. The genome of Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, 92, pp.10836–10840.

Way, J.C. & Chalfie, M., 1988. mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C. elegans. *Cell*, 54(1), pp.5–16.

Wederell, E.D. et al., 2008. Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Research*, 36(14), pp.4549–4564.

Weinshenker, D., Garriga, G. & Thomas, J.H., 1995. Genetic and pharmacological analysis of neurotransmitters controlling egg laying in C. elegans. *The Journal of neuroscience*, 15(10), pp.6975–6985.

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

White, J.G., et al., 1986. The structure of the of the nervous system of the nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 314, pp.1–340.

White, M. a et al., 2013. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proceedings of the National Academy of Sciences of the United States of America*, 109(29), pp.11952–11957.

Whitfield, T.W. et al., 2012. Functional analysis of transcription factor binding sites in human promoters. *Genome Biology*, 13(9), p.50.

Whyte, W.A. et al., 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2), pp.307–319.

William, C.M., Tanabe, Y. & Jessell, T.M., 2003. Regulation of motor neuron subtype identity by repressor activity of Mnx class homeodomain proteins. *Development*, 130(8), pp.1523–1536.

Wilson, L. & Maden, M., 2005. The mechanisms of dorsoventral patterning in the vertebrate neural tube. *Developmental Biology*, 282(1), pp.1–13.

Wu, J., Duggan, A. & Chalfie, M., 2001. Inhibition of touch cell fate by egl-44 and egl-46 in C. elegans. *Genes and Development*, 15(6), pp.789–802.

Wyler, S.C. et al., 2016. Pet-1 Switches Transcriptional Targets Postnatally to Regulate Maturation of Serotonin Neuron Excitability. *Journal of Neuroscience*, 36(5), pp.1758–1774.

Wylie, C.J. et al., 2010. Distinct Transcriptomes Define Rostral and Caudal Serotonin Neurons. *Journal of Neuroscience*, 30(2), pp.670–684.

Xie, Y. et al., 2013. RFX Transcription Factor DAF-19 Regulates 5-HT and Innate Immune Responses to Pathogenic Bacteria in Caenorhabditis elegans. *PLoS Genetics*, 9(3), pp.1-17.

Xu, X. et al., 1998. Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the Drosophila mesoderm. *Genes and Development*, 12(15), pp.2354–2370.

Xue, D., Tu, Y. & Chalfie, M., 1993. Cooperative Interactions Between the Caenorhabditis elegans Homeoproteins UNC-86 and MEC-3. *Science*, 261(5126), pp.1324–1328.

Yanez-Cuna, J.O. et al., 2012. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. *Genome Research*, 22, pp.2018–2030.

Ye, W. et al., 1998. FGF and Shh signals control dopaminergic and serotonergic cell fate in the anterior neural plate. *Cell*, 93(5), pp.755–766.

Youdim, M.B.H., Edmondson, D. & Tipton, K.F., 2006. The therapeutic potential of monoamine oxidase inhibitors. *Nature reviews. Neuroscience*, 7(4), pp.295–309.

Yu, H. et al., 2003. Distinct roles of transcription factors EGL-46 and DAF-19 in specifying the functionality of a polycystin-expressing sensory neuron necessary for C. elegans male vulva location behavior. *Development*, 130(21), pp.5217–5227.

Zahn, T.R. et al., 2004. Dense core vesicle dynamics in Caenorhabditis elegans neurons and the role of kinesin UNC-104. *Traffic*, 5(7), pp.544–559.

Zaret, K.S. & Carroll, J.S., 2011. Pioneer transcription factors: Establishing competence for gene expression. *Genes and Development*, 25(21), pp.2227–2241.

Zaret, K.S. & Mango, S.E., 2016. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current opinion in genetics and development*, 37, pp.76–81.

Zhang, F. et al., 2014. The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. *Development*, 141(2), pp.422–435.

Zhao, G.Y. et al., 2008. Expression of the transcription factor GATA3 in the postnatal mouse central nervous system. *Neurosci Res*, 61(4), pp.420–428.

Zhao, Z. et al., 2006. Lmx1b Is Required for Maintenance of Central Serotonergic Neurons and Mice Lacking Central Serotonergic System Exhibit Normal Locomotor Activity. *The Journal of neuroscience*, 26(49), pp.12781–12788.

Zheng, X. et al., 2005. Cell-type specific regulation of serotonergic identity by the C. elegans LIM-homeodomain factor LIM-4. *Developmental Biology*, 286(2), pp.618–628.

Zinzen, R.P. et al., 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature*, 462(7269), pp.65–70.

Abbreviations and achronyms

5-HT	serotonin	**HSN**	Hermaphrodite Specific Motorneuron
5-HTP	5 hydroxytryptophan	**IPTG**	Isopropyl β-D-1-thiogalactopyranoside
AADC	amino acid decarboxylase	**kb**	kilo base
ADF	Amphid neuron Dual F	**MOD-5**	modulation of locomotion defective
BAS-1	biogenic amine synthesis related 1	**NSM**	Neuro Secretory Motorneuron
bHLH	basic Helix-Loop-Helix	**nt**	nucleotide
bp	base pair	**PAM**	Protospacer Adjacent Motif
BS	binding site	**PBS**	Phosphate Buffered Saline
C. briggsae	*Caenorhabditis briggsae*	**PCR**	Polymerase Chain Reaction
C. brenneri	*Caenorhabditis brenneri*	**PHB**	Phasmid neuron B
C. elegans	*Caenorhabditis elegans*	**PFA**	paraformaldehyde
CGC	Caenorhabditis Genetic Center	**POU**	Pit-Oct-Unc
C. japonica	*Caenorhabditis japonica*	**PPI**	protein-protein interaction
C. remanei	*Caenorhabditis japonica*	**PWM**	position weight matrix
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats	**pV**	p value
		Rev	reverse
CRM	*cis*-regulatory module	**r_n**	rhombomere
DiD	1,1'-Dioctadecyl-3,3,3',3'-tetramethylindodicarbocyanine, 4-chlorobenzenesulfonate salt	**RNA**	Ribonucleic acid
		RNAi	RNA interference
DiI	1,1'-Dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate	**RT**	room temperature
		SD	standard deviation
DiO	3,3'-Dioctadecyloxacarbocyanine perchlorate	**SEC**	Self-Excising Cassette
dpy	dumpy	**SEM**	standard error of the mean
egl	egg-laying defective	**SERT**	serotonin transporter
DAPI	4',6-diamidino-2-phenylindole	**sgRNA**	single guide RNA
DNA	Deoxyribonucleic acid	**SLC18A1/2**	Solute Carrier Family 18 Member A1/A2
E. coli	*Escherichia coli*	**SLC6A4**	Solute Carrier Family 6 Member A4.
EMSA	Electrophoretic Mobility Shift Assay	**SNP**	Single Nucleotide Polimorphism
ETS	E26 transformation-specific or E-twenty-six	**TF**	transcription factor
FKH	forkhead	**TFBS**	transcription factor binding site
Fwd	forward	**TPH-1**	tryptophan hydroxylase
GCH1	GTP cyclohydrolase 1	**Trp**	tryptophan
GFP	Green Fluorescent Protein	**TSS**	transcription starting site
GWAS	Genome-wide association study	**UTR**	untranslated region
HD	Homeo Domain	**ZnF**	Zinc Finger
hs	heat shock	**VMAT**	vesicular monoamine transporter
hsp	heat shock promoter	**wt**	wild type
hpf	hours post-fertilisation		

Annexes

Annex 3.1.1
Primary data of serotonin pathway gene *cis*-regulatory analysis

Apart from the serotonergic neurons (NSM, ADF, AIM, RIH and VC4/5), we included in the analysis all monoaminergic neurons that share the expression of some 5-HT pathway genes, including dopaminergic (CEPD, CEPV, ADE, PDE), octopaminergic (RIC) and tyraminergic (RIM) neurons. (–): not expected to be expressed. See Figure 3.1.3.

5-HT								
Promoter	% HSN	% NSM	%ADF	% VC4/5	%AIM	%RIH	Other cells	Lines
tph-1prom1	81,84,91	31, 76, 86	69,90,93	31, 69, 76	(–)	(–)	no	3
tph-1prom8	0,0,0	0,0,0	0,0,0	0,0,0	(–)	(–)	yes	3
tph-1prom2	41,53,61, 62,68,93	89,92,95, 98,98,98	90,03,04, 95,95,98	43,48,55, 58,81,90	(–)	(–)	yes	6
tph-1prom6	0,0	0,0	0,0	0,0	(–)	(–)	no	2
tph-1prom5	0,0,0	0,09,98	0,0,0	0,0,0	(–)	(–)	no	3
tph-1prom3	0,0,0,0	85,86,89,94	0,0,0,0	0,0,0,0	(–)	(–)	no	4
tph-1prom17	0,0,0	0,0,0	90,93,95	53,53,59	(–)	(–)	no	3
bas-1prom1	88,92	92,100	100,100	(–)	67,88	(–)	yes	2
bas-1prom2	42,50	100,93	0,0	(–)	0,0	(–)	yes	2
bas-1prom13	90,91	31,78	0,0	(–)	0,0	(–)	yes	2
bas-1prom14	0	0	0	(–)	0	(–)	no	1
bas-1prom15	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	yes	3
bas-1prom16	75	45	12	(–)	0	(–)	no	1
bas-1prom17	0,0	0,0	0,0	(–)	0,0	(–)	no	2
bas1prom18	38,43,43,63	82,92,93,95	0,0,0,0	(–)	0,0,0,0	(–)	yes	4
bas-1prom3	8,17	0,0	95,95	(–)	0,0	(–)	no	2
bas-1prom4	0,0	0,0	0,0	(–)	0,0	(–)	yes	2
bas-1prom5	0,0,0	0,0,0	67,85,87	(–)	0,0,0	(–)	yes	3
bas-1prom6	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	no	3
bas-1prom7	0,0,0	0,0,0	88,97,97	(–)	0,0,0	(–)	no	3
cat-1prom1	0,70,87	93,93,100	2,83,93	3,82,97	(–)	0,70,83	no	3
cat-1prom2	0,0	0,0	0,0	0,0	(–)	0,0	yes	2
cat-1prom3	87,92,97	93,93,98	78,88,98	83,92,97	(–)	0,8,12	no	3
cat-1prom12	43,68	98,98	70,73	0,0	(–)	0,0	no	2

↳

cat-1prom11	0,0	0,0	92,97	92,95	(–)	0,13	no	2
cat-1prom35	0,0,0	0,0,	0,0,0	0,0,0	(–)	0,0,0	no	3
cat-1prom36	0,0	0,0	77,80	0,0	(–)	0,0	no	2
cat-1prom37	0,0,0	0,0,0	92,95,98	0,0,0	(–)	0,0,0	yes	3
cat-1prom13	0,0,0	0,0,0	0,0,0	0,0,0	(–)	0,0,0	no	3
cat-1prom14	20,43,50, 55,67,73	82,82,88, 92,93,95	10,10,72, 80,82,87	0,0,0, 0,0,0	(–)	0,0,0, 0,0,0	no	6
cat-1prom26	0,0,0	0,0,0	0,0,0	0,0,0	(–)	0,0,0	no	3
cat-1prom27	7,8,30	90,95,98	0,2,12	0,0,0	(–)	0,0,0	yes	3
cat-4prom4	85,87,95	90,97,98	87,88,93	(–)	80,82,92	(–)	no	3
cat-4prom5	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	no	3
cat-4prom6	52,78,80	93,95,97	38,55,72	(–)	72,80,87	(–)	yes	3
cat-4prom8	48,75	0,0	0,0	(–)	0,0	(–)	no	2
cat-4prom58	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	no	3
cat-4prom59	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	no	3
cat-4prom9	0,0,0	87,88,93	0,0,0	(–)	0,0,0	(–)	no	3
cat-4prom18	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	yes	3
cat-4prom19	0,0	0,0	0,0	(–)	0,0	(–)	yes	2
cat-4prom27	0,0,0	20,72,72	0,0,0	(–)	0,0,0	(–)	no	3
mod-5prom1	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	no	3
mod-5prom3	0,0,0	97,98	95,97	(–)	0,0,0	(–)	yes	3
mod-5prom8	0,0,0	0,0,0	53,58,92	(–)	0,0,0	(–)	yes	3
mod-5prom6	0,0,0	0,0,0	61,83,87	(–)	0,0,0	(–)	yes	3

	DA				Tyr	Oct		
Promoter	% CEPV	% CEPD	%ADE	% PDE	%RIM	% RIC	Other cells	Lines
bas-1prom1	92,95	87,97	90,97	100,100	(–)	(–)	yes	2
bas-1prom2	0,0	0,0	0,0	0,0	(–)	(–)	yes	2
bas-1prom13	0,0	0,0	0,0	0,0	(–)	(–)	yes	2
bas-1prom14	0	0	0	0	(–)	(–)	no	1
bas-1prom15	0,0,0	0,0,0	0,0,0	0,0,0	(–)	(–)	yes	3
bas-1prom16	0	0	0	0	(–)	(–)	no	1
bas-1prom17	0,0	0,0	0,0	0,0	(–)	(–)	no	2
bas-1prom18	0,0,0,0	0,0,0,0	0,0,0,0	0,0,0,0	(–)	(–)	yes	4
bas-1prom3	82,93	92,93	82,90	90,90	(–)	(–)	no	2
bas-1prom4	0,2	7,8	0,3	0,0	(–)	(–)	yes	2
bas-1prom5	87,88,90	87,92,93	83,88,97	68,80,83	(–)	(–)	yes	3
bas-1prom6	0,0,0	0,0,3	8,10,18	32,40,73	(–)	(–)	no	3
bas-1prom7	0,0,0	0,0,0	0,0,0	0,0,0	(–)	(–)	no	3
cat-1prom1	0,70,87	93,93,100	2,83,93	42,93,95	0,93,97	0,92,97	no	3
cat-1prom2	0,0	0,0	0,0	87,88	0,0	0,0	yes	2
cat-1prom3	87,92,97	93,93,98	78,88,98	85,90,97	0,0,0	0,0,98	no	3
cat-1prom12	43,68	98,98	70,73	90,90	0,0	0,0	no	2
cat-1prom11	0,0	0,0	92,97	0,0	9,97	0,0	no	2
cat-1prom35	0,0,0	0,0,	0,0,0	0,0,0	0,0,0	0,0,0	no	3
cat-1prom36	0,0	0,0	77,80	0,0	0,0	0,0	no	2
cat-1prom37	0,0,0	0,0,0	92,95,98	0,0,0	0,0,0	0,0,0	yes	3
cat-1prom13	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	no	3
cat-1prom14	20,43,50, 55,67,73	82,82,88, 92,93,95	10,10,72, 80,82,87	20,30,55, 78,85,95	0,0,0, 0,0,0	0,0,0, 0,0,0	no	6
cat-1prom26	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	no	3
cat-1prom27	7,8,30	90,95,98	0,2,12	2,13,23	0,0,0	0,0,0	yes	3
cat-4prom4	85,87,97	87,88,98	63,72,77	(–)	(–)	(–)	no	3
cat-4prom5	0,0,0	0,0,0	0,0,0	(–)	(–)	(–)	no	3

cat-4prom6	97,97,100	87,95,97	70,80,90	(–)	(–)	(–)	yes	3
cat-4prom8	0,0	0,0	0,0	(–)	(–)	(–)	no	2
cat-4prom58	0,0,0	0,0,0	0,0,0	(–)	(–)	(–)	no	3
cat-4prom59	0,0,0	0,0,0	0,0,0	(–)	(–)	(–)	no	3
cat-4prom9	100,100,100	100,100,100	45,52,72	(–)	(–)	(–)	no	3
cat-4prom18	0,0,0	0,0,0	0,0,0	(–)	(–)	(–)	yes	3
cat-4prom19	7,63	0,0	0,27	(–)	(–)	(–)	yes	2
cat-4prom27	7,90,93	0,0,0	0,0,0	(–)	(–)	(–)	no	3
cat-4prom8	48,75	0,0	0,0	(–)	0,0	(–)	no	2
cat-4prom58	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	no	3
cat-4prom59	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	no	3
cat-4prom9	0,0,0	87,88,93	0,0,0	(–)	0,0,0	(–)	no	3
cat-4prom18	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	yes	3
cat-4prom19	0,0	0,0	0,0	(–)	0,0	(–)	yes	2
cat-4prom27	0,0,0	20,72,72	0,0,0	(–)	0,0,0	(–)	no	3
mod-5prom1	0,0,0	0,0,0	0,0,0	(–)	0,0,0	(–)	no	3
mod-5prom3	0,0,0	97,98	95,97	(–)	0,0,0	(–)	yes	3
mod-5prom8	0,0,0	0,0,0	53,58,92	(–)	0,0,0	(–)	yes	3
mod-5prom6	0,0,0	0,0,0	61,83,87	(–)	0,0,0	(–)	yes	3

Annex 3.2.1
Primary data of serotonin pathway gene expression in loss of function mutants

Analysis of *tph-1*, *bas-1*, *cat-1* and *cat-4* expression in the different mutant backgrounds for the six candidate regulators of the HSN neuron. Expression in the NSM and ADF neurons was also considered.

See Figures 3.2.2, 3.2.3 and 3.2.5.

Gene	Genotype	% HSN	SEP	pV	% NSM	SEP	pV	% ADF	SEP	pV
5-HT staining	N2	100	0	(–)	100	0	(–)	97	0	(–)
	ast-1(ot417)II	0	0	0.0001	100	0	1.000	100	0	1.000
	unc-86(n846)III	1	1	0.0001	95	2	0.059	98	1	1.000
	sem-4(n1971)I	4	2	0.0001	100	0	1.000	100	0	1.000
	hlh-3(tm1688)II	38	3	0.0001	100	0	1.000	97	2	1.000
	egl-46(sy628)V	82	2	0.0001	96	2	0.121	89	3	0.0489
	egl-18(ok290)IV	74	3	0.0001	97	2	0.246	90	3	0.0818
	end-1&ric-7(ok558)V	92	2	0.0025	100	0	1.000	100	0	1.000
	lin-11 (n389)I; ceh-14 (ch3) X	95	2	0.0594	99	1	1.000	99	1	1.000
tph-1	*zdIs13(tph-1::gfp)IV*	100	0	(–)	100	0	(–)	100	0	(–)
	yzIs71[tph-1::gfp, rol-6(su1006)]V	96	2	(–)	99	1	(–)	99	1	(–)
	ast-1(ot417)II; zdIs13(tph-1::gfp)IV	0	0	0.0001	100	0	1.000	100	0	1.000
	unc-86(n846)III; zdIs13(tph-1::gfp)IV	0	0	0.0001	0	0	0.0001	100	0	1.000
	sem-4(n1971)I; zdIs13(tph-1::gfp)IV	52	3	0.0001	99	1	0.560	99	1	1.000
	hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV	22	4	0.0001	100	0	1.000	99	1	1.000
	egl-46(sy628)V; zdIs13(tph-1::gfp)IV	62	5	0.0001	98	1	1.000	90	3	0.001
	egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V	75	4	0.0001	99	1	1.000	100	0	1.000
	end-1&ric-7(ok558)V; zdIs13(tph-1::gfp)IV	100	0	1.000	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
cat-1	*otIs221(cat-1::gfp)III*	100	0	(–)	100	0	(–)	97	2	(–)
	otIs224(cat-1::gfp)V	100	0	(–)	100	0	(–)	100	0	(–)
	ast-1(ot417)II; otIs221(cat-1::gfp)III	0	0	0.0001	100	0	1.000	100	0	1.000
	unc-86(n846)III;otIs224(cat-1::gfp)V	5	2	0.0001	90	3	0.002	100	0	1.000
	sem-4(n1971)I; otIs221(cat-1::gfp)III	0	0	0.0001	100	0	1.000	100	0	1.000
	hlh-3(tm1688)II; otIs221(cat-1::gfp)III	81	3	0.0001	99	1	1.000	100	0	1.000
	egl-46(sy628)V; otIs221(cat-1::gfp)III	61	5	0.0001	100	0	1.000	100	0	1.000
	egl-18(ok290)IV; otIs221(cat-1::gfp)III	97	2	0.130	99	1	1.000	96	2	0.687

	end-1&ric-7(ok558)V; otIs221[cat-1::gfp]III	100	0	1.000	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	lin-11 (n389)I; ceh-14 (ch3) X; otIs221(cat-1::gfp)III; him-5 (e1467)V	93	2	0.006	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
cat-4	*otEx2470[cat-4::gfp(50ng/ul), rol-6(su1006)]*	89	3	(–)	95	3	(–)	91	4	(–)
	otIs225(cat-4::gfp)II	100	0	(–)	100	0	(–)	100	0	(–)
	ast-1(ot417)II; otEx2470[cat-4::gfp (50ng/ul), rol-6(su1006)]	89	3	1.000	90	3	0.359	90	3	1.000
	unc-86(n846)III; otIs225(cat-4::gfp)II	0	0	0.0001	86	3	0.0001	100	0	1.000
	sem-4(n1971)I; otIs225(cat-4::gfp)II	1	1	0.0001	100	0	1.000	100	0	1.000
	hlh-3(tm1688)II; otEx2470[cat-4::gfp (50ng/ul), rol-6(su1006)]	89	3	1.000	85	8	0.171	90	7	1.000
	egl-46(sy628)V; otIs225(cat-4::gfp)II	100	0	1.000	100	0	1.000	98	1	0.498
	egl-18(ok290)IV; otIs225(cat-4::gfp)II	99	3	1.000	100	0	1.000	100	0	0.567
	end-1&ric-7(ok558)V; otIs225[cat-4::gfp]II	100	0	1.000	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
bas-1	*otIs226(bas-1::gfp)IV*	100	0	(–)	100	0	(–)	100	0	(–)
	otEx2435[bas-1::gfp(50ng/ul), rol-6(su1006)]	85	4	(–)	85	3	(–)	91	2	(–)
	ast-1(ot417)II; otIs226(bas-1::gfp)IV	100	0	1.000	100	0	1.000	100	0	1.000
	ast-1(hd92); vlcEx844[ast-1(+), cat-1::DsRed]; otIs226(bas-1::gfp)IV	100	0	1.000	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	ast-1(hd92); vlcEx845[ast-1(+), cat-1::DsRed]; otIs226(bas-1::gfp)IV	100	0	1.000	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	unc-86(n846)III; cat-1::DsRed]; otIs226(bas-1::gfp)IV	0	0	0.0001	83	4	0.0001	99	1	1.000
	sem-4(n1971)I; otIs226(bas-1::gfp)IV	0	0	0.0001	100	0	1.000	90	3	0.001
	hlh-3(tm1688)II; otIs226(bas-1::gfp)IV	49	5	0.0001	100	0	1.000	86	3	0.0001
	egl-46(sy628)V; otIs226(bas-1::gfp)IV	85	3	0.0001	98	1	0.500	97	1	0.251
	egl-18(ok290)IV; otEx2435[bas-1::gfp (50ng/ul), rol-6(su1006]	79	1	0.162	97	1	0.001	90	3	0.836
	end-1&ric-7(ok558)V; otIs226(bas-1::gfp)IV	100	0	1.000	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Annex 3.2.2
Primary data of non-serotonin related gene expression in loss of function mutants

Analysis of a battery of terminal features of the HSN that are independent of the 5-HT biosynthetic pathway, in the different mutant backgrounds for the six candidate regulators of the HSN neuron. See Figures 3.2.4 and 3.2.5.

Gene	Genotype	% HSN	SEP	pV
kcc-2c	*vsEx580[kcc-2c::gfp, myo-2::gfp]*	74	4	(–)
	ast-1(ot417)II; vsEx580[kcc-2c::gfp, myo-2::gfp]	0	0	0 .0001
	unc-86(n846)III; vsEx580[kcc-2c::gfp, myo-2::gfp]	0	0	0 .0001
	sem-4(n1971)I; vsEx580[kcc-2c::gfp, myo-2::gfp]	6	2	0 .0001
	hlh-3(tm1688)II; vsEx580[kcc-2c::gfp, myo-2::gfp]	0	0	0 .0001
	egl-46(sy628)V; vsEx580[kcc-2c::gfp, myo-2::gfp]	52	5	0 .0016
	egl-18(ok290)IV; vsEx580[kcc-2c::gfp, myo-2::gfp]	45	5	0 .0001
lgc-55	*zfIs4(lgc-55::mCherry)*	100	0	(–)
	zfIs6(lgc-55::gfp)II	100	0	(–)
	ast-1(ot417)II; zfIs4(lgc-55::mCherry)	0	0	0 .0001
	unc-86(n846)III; zfIs6(lgc-55::gfp)II	12	3	0 .0001
	sem-4(n1971)I; zfIs6(lgc-55::gfp)II	0	0	0 .0001
	hlh-3(tm1688)II; zfIs4(lgc-55::mCherry)	8	3	0 .0001
	egl-46(sy628)V; zfIs6(lgc-55::gfp)II	85	3	0 .0001
	egl-18(ok290)IV; zfIs6(lgc-55::gfp)II	98	1	1
ida-1	*inIs181(ida-1::gfp); inIs182(ida-1::gfp)*	100	0	(–)
	inIs179(ida-1::gfp)II	100	0	(–)
	ast-1(ot417)II; inIs181(ida-1::gfp); inIs182(ida-1::gfp)	96	2	0 .449
	unc-86(n846)III; inIs179(ida-1::gfp)II	0	0	0 .0001
	sem-4(n1971)I; inIs179(ida-1::gfp)II	4	2	0 .0001
	hlh-3(tm1688)II; inIs181(ida-1::gfp); inIs182(ida-1::gfp)	28	4	0 .0001
	egl-46(sy628)V; inIs179(ida-1::gfp)II	77	4	0 .0001
	egl-18(ok290)IV; inIs179(ida-1::gfp)II	92	2	0 .0019
flp-19	*ynIs34(flp-19::gfp)IV*	100	0	(–)
	ast-1(ot417)II; ynIs34(flp-19::gfp)IV; him-5(e1490)V	100	0	1
	unc-86(n846)III; ynIs34(flp-19::gfp) IV; him-5(e1490)V	0	0	0 .0001
	sem-4(n1971)I; ynIs34(flp-19::gfp) IV; him-5(e1490)V	0	0	0 .0001
	hlh-3(tm1688)II; ynIs34(flp-19::gfp) IV	43	5	0 .0001
	egl-46(sy628)V; ynIs34(flp-19::gfp) IV	5	2	0 .0001

unc-17	*otIs576(unc-17fosmid::GFP, lin-44::YFP)*	79	3	(–)
	ast-1(ot417)II; otIs576(unc-17fosmid::GFP, lin-44::YFP)	81	3	0 .86
	unc-86(n846)III; otIs576(unc-17fosmid::GFP, lin-44::YFP); him-5(e1490)V	67	3	0 .0587
	sem-4(n1971)I; otIs576(unc-17fosmid::GFP, lin-44::YFP); him-5(e1490)V	1	0	0 .0001
	hlh-3(tm1688)II; otIs576(unc-17fosmid::GFP, lin-44::YFP); him-5(e1490)V	15	3	0 .0001
	egl-46(sy628)V; otIs576(unc-17fosmid::GFP, lin-44::YFP)	34	3	0 .0001
	egl-18(ok290)IV; otIs576(unc-17fosmid::GFP, lin-44::YFP)	82	3	0 .1922
unc-40	*icIs132(unc-40::gfp)*	92	2	(–)
	ast-1(ot417)II; icIs132 (unc-40::gfp); him-8(e1489)IV	95	2	0 .3421
	unc-86(n846)III; icIs132 (unc-40::gfp); him-8(e1489)IV	88	3	0 .24
	sem-4(n1971)I; icIs132 (unc-40::gfp); him-8(e1489)IV	75	4	0 .0001
	hlh-3(tm1688)II; icIs132 (unc-40::gfp); him-8(e1489)IV	35	5	0 .0001
	egl-46(sy628)V; icIs132 (unc-40::gfp); him-8(e1489)IV	10	3	0 .0001
	egl-18(ok290)IV; icIs132 (unc-40::gfp); him-8(e1489)IV	96	2	0 .2208
rab-3	*otIs287[rab-3::yfp, rol-6(su1006)]IV*	100	0	(–)
	otIs291[rab-3::gfp, rol-6(su1006)]	100	0	(–)
	ast-1(ot417)II; otIs287[rab-3::yfp, rol-6(su1006)]IV	100	0	1
	unc-86(n846)III; otIs287[rab-3::yfp, rol-6(su1006)]IV	100	0	1
	sem-4(n1971)I; otIs287[rab-3::yfp, rol-6(su1006)]IV	81	4	0 .0001
	hlh-3(tm1688)II; otIs287[rab-3::yfp, rol-6(su1006)]IV	34	4	0 .0001
	egl-46(sy628)V; otIs287[rab-3::yfp, rol-6(su1006)]IV	93	2	0 .0142
	egl-18(ok290)IV; otIs291[rab-3::gfp, rol-6(su1006)]	94	2	0 .029
nlg-1	*sIs13247(nlg-1::gfp)*	69	3	(–)
	ast-1(ot417)II; sIs13247(nlg-1::gfp)	46	5	0 .0001
	unc-86(n846)III; sIs13247(nlg-1::gfp)	60	4	0 .0783
	sem-4(n1971)I; sIs13247(nlg-1::gfp)	15	3	0 .0001
	hlh-3(tm1688)II; sIs13247(nlg-1::gfp)	57	5	0 .0364
	egl-46(sy628)V; sIs13247(nlg-1::gfp)	48	6	0 .0007

kal-1	*otIs33(kal-1::gfp)IV*	92	2	(–)
	vlcEx453[kal-1::gfp, ttx-3::mCherry, rol-6(su1006)]	84	4	(–)
	ast-1(ot417)II; otIs33(kal-1::gfp)IV	90	3	0 .8181
	unc-86(n846)III; otIs33(kal-1::gfp)IV	85	4	0 .1432
	sem-4(n1971)I; otIs33(kal-1::gfp)IV	82	3	0 .0161
	hlh-3(tm1688)II; otIs33(kal-1::gfp)IV	100	0	0 .0014
	egl-46(sy628)V; otIs33 (kal-1::gfp)IV	96	3	0 .4117
	egl-18(ok290)IV; vlcEx453[kal-1::gfp, ttx-3::mCherry, rol-6(su1006)]	93	2	0 .8059

Annex 3.2.3
Primary data of RNA interference assays

Data from RNAi screen of the GATA (Figure 3.2.6), FKH and LIM-HD (Figure 3.4.3) TF families, and from RNAi maintenance assays of the HSN regulatory code members (Figure 3.2.16).

TF Family	*Genotype*	Scoring	% HSN	SEP	pV
GATA	*rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + L4440*	F1	100	0	(-)
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + egl-18 RNAi	F1	68	6	0.0001
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + elt-3 RNAi	F1	92	4	0.0573
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + elt-6 RNAi	F1	92	4	0.0573
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + elt-7 RNAi	F1	93	3	0.1187
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + end-1 RNAi	F1	88	4	0.013
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + end-3 RNAi	F1	96	3	0.4958
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + med-1/med-2 RNAi	F1	95	3	0.2437
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + L4440	P0	100	0	(-)
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + elt-1 RNAi	P0	100	0	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + elt-2 RNAi	P0	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ L4440	F1	100	0	(-)
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ egl-18 RNAi	F1	40	6	0.0001
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ elt-3 RNAi	F1	80	5	0.0003
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ elt-6 RNAi	F1	87	4	0.0061
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ elt-7 RNAi	F1	93	3	0.1187
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ end-1 RNAi	F1	63	6	0.0001
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ end-3 RNAi	F1	93	3	0.1187
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ med-1/med-2 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ L4440	P0	100	0	(-)
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ elt-1 RNAi	P0	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III+ elt-2 RNAi	P0	100	0	1
FKH	*rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + L4440*	F1	97	2	0.4958
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + attf-4 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + C34B4.2 RNAi	F1	100	0	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + daf-16 RNAi	F1	98	2	1
FKH	*rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + fkh-2 RNAi*	F1	100	0	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + fkh-3/4 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + fkh-5 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + fkh-6 RNAi	F1	96	3	0.4958
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + fkh-7 RNAi	F1	100	0	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + fkh-8 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + fkh-9 RNAi	F1	95	3	0.2437
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + fkh-10 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + lin-31 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + let-381 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + pha-4 RNAi	F1	82	5	0.013
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + T27A8.2 RNAi	F1	97	2	0.4958
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + unc-130 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + L4440	F1	99	1	0.4958
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + attf-4 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + C34B4.2 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + daf-16 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + fkh-2 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + fkh-3/4 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + fkh-5 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + fkh-6 RNAi	F1	98	2	0.4958
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + fkh-7 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + fkh-8 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + fkh-9 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + fkh-10 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + lin-31 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + let-381 RNAi	F1	97	2	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + pha-4 RNAi	F1	40	6	0.0001
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + T27A8.2 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + unc-130 RNAi	F1	100	0	1

↳

LIM-HD	*rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + L4440*	F1	100	0	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + ceh-14 RNAi	F1	91	4	0.0573
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + lim-4 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + lim-6 RNAi	F1	98	2	1
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + lim-7 RNAi	F1	97	2	0.4958
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + lin-11 RNAi	F1	92	4	0.0573
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + mec-3 RNAi	F1	96	3	0.4958
	rrf-3(pk1426)II; zdIs13(tph-1::gfp)IV + ttx-3 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + L4440	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + ceh-14 RNAi	F1	80	5	0.0003
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + lim-4 RNAi	F1	100	0	1
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + lim-6 RNAi	F1	91	4	0.0573
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + lim-7 RNAi	F1	100	0	0.4958
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + lin-11 RNAi	F1	70	6	0.0001
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + mec-3 RNAi	F1	100	0	0.4958
	rrf-3(pk1426)II; otIs221(cat-1::gfp)III + ttx-3 RNAi	F1	95	3	1

Genotype	% PHB	SEP	pV
N2	87	4	(–)
ast-1(ot417)II; otIs221(cat-1::gfp)III	90	4	1
unc-86(n846)III; otIs224(cat-1::gfp)V	79	5	0.7306
sem-4(n1971)I; otIs221(cat-1::gfp)III	70	6	0.2092
hlh-3(tm1688)II; otIs221(cat-1::gfp)III	86	4	1
egl-46(sy628)V; otIs221(cat-1::gfp)III	75	6	0.3354
egl-18(ok290)IV; otIs221(cat-1::gfp)III	61	6	0.0391

Annex 3.2.4
Primary data of PHB neuron DiI staining analysis.

Scoring of the phasmid PHB neuron (HSN sister) that is located in the tail.
See Figure 3.2.7.

Annex 3.2.5
Specific DNA modifications for mutagenesis *cis*-regulatory analysis

Uppercase letters indicate wild type nucleotides, while lowercase letters indicate mutated nucleotides that alter a specific BS motif.
See Figures 3.2.8-3.2.10.

Promoter	Genotype	Target Binding Site	Mutation (wt BS > mutant BS)
tph-1prom2	N2	(–)	(–)
tph-1prom14	N2	ETS	CGGATA > CaGATA
tph-1prom26	N2	POU	GCGCATAATAAAACAATCA > GtGtATAccAcAACAAgCg
tph-1prom31	N2	SPALT	TTGTGT > TTagGT
tph-1prom44	N2	HLH	CCAGAA > tttGAA
tph-1prom43	N2	INSM	CCCCTCTC > tttCTCTC
tph-1prom54	N2	GATA	GGATATCT > GtATATtT
tph-1prom55	N2	GATA	GGATAT > GGAaAT
tph-1prom60	N2	GATA	GGATATCT > GtATATtT; GGATAT > GGAaAT
tph-1prom52	N2	FKH	ATAAATA > ATAggTA
cat-1prom14	N2	(–)	(–)
cat-1prom63	N2	ETS	TTTCCG > TTTCgG
cat-1prom61	N2	POU	TTCATCAT > TTgggCAT
cat-1prom60	N2	SPALT	TTGTCT > cTagCT
cat-1prom73	N2	HLH	TTCTGG > TTtTtt
cat-1prom71	N2	INSM	CCCCACCA > ttttACCA
cat-1prom74	N2	GATA	AGATAA > AtATAA
cat-1prom75	N2	GATA	TGATAG > TtATAG
cat-1prom76	N2	GATA	AGATAA > AtATAA ; TGATAG > TtATAG
cat-1prom83	N2	FHK	ATCAACA > ATCggCA
cat-1prom3	N2	(–)	(–)
cat-1prom79	N2	INSM	CCGCTAGA > ttGtTAGA; CCCCACCA > tttACCA[1]; CCCCTTGG > ttttTTGG
bas-1prom18	N2	(–)	(–)
bas-1prom73	N2	ETS	TATCCG > TATCgG
bas-1prom71	N2	POU	TGCATTCA > TGgggTCA
bas-1prom65	N2	SPALT/MYT	AAATTT > AAgggg
bas-1prom78	N2	GATA	CTATCC > CTtTCC
bas-1prom13	N2	(–)	(–)
bas-1prom77	N2	HLH	CCAGAA > tttGAA
bas-1prom76	N2	INSM	CCCCAACA > CtttAACA
bas-1prom83	N2	GATA	ATATC > ATATa
bas-1prom84	N2	GATA	TGATAT > TtATAT
bas-1prom86	N2	GATA	ATATC > ATATa; TGATAT > TtATAT; TGATAT > TtATAT
bas-1prom78	*ast-1(ot417)*	GATA	CTATCC > CTtTCC

Gene		Embryo	L1	L2	L3	L4	Young adult
unc-86	%	100.00	97.00	100.00	100.00	100.00	100.00
	SEP	0.00	2.00	0.00	0.00	0.00	0.00
	N	40.00	45.00	50.00	40.00	50.00	50.00
sem-4	%	n.a.	45.00	100.00	100.00	100.00	92.00
	SEP	n.a.	5.00	0.00	0.00	0.00	4.00
	N	n.a.	45.00	13.00	15.00	14.00	24.00
egl-46	%	n.a.	53.00	57.00	39.00	48.00	67.00
	SEP	n.a.	7.00	5.00	5.00	7.00	3.00
	N	n.a.	27.00	29.00	46.00	24.00	95.00
ast-1	%	0.00	0.00	0.00	40.00	94.00	99.00
	SEP	0.00	0.00	0.00	11.00	3.00	1.00
	N	10.00	20.00	20.00	10.00	44.00	71.00
egl-18	%	n.a.	91.00	100.00	100.00	83.00	87.00
	SEP	n.a.	4.00	9.00	0.00	8.00	4.00
	N	n.a.	22.00	12.00	20.00	12.00	50.00
hlh-3	%	90.00	0.00	0.00	0.00	0.00	0.00
	SEP	3.00	0.00	0.00	0.00	0.00	0.00
	N	50.00	20.00	20.00	20.00	20.00	20.00

Annex 3.2.6
Primary data of cross-regulation between the six members of the HSN regulatory code

In the first table we include the expression pattern date for the six members of the HSN regulatory code over time. In the second table we resume the expression of reporters of the individual members of the HSN regulatory code in the six mutant backgrounds. See Figures 3.2.14 and 3.2.15.

Gene	Genotype	Age	% HSN	SEP	pV
unc-86	*otIs337(unc-86fosmid::NLS::YFP::H2B; ttx-3::mCherry)*	Young adult	100	0	(–)
	otIs337(unc-86fosmid::NLS::YFP::H2B; ttx-3::mCherry)	L1	97	2	(–)
	sem-4(n1971)I; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)	Young adult	95	2	0.0595
	hlh-3(tm1688)II; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)	Young adult	99	1	0.4845
	egl-46(sy628)V; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)	Young adult	100	0	1
	ast-1(ot417)II; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)	Young adult	100	0	1
	ast-1(hd92)II; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)	L1	100	0	1
	egl-18(ok290)IV; otIs337(unc-86fosmid::NLS::YFP::H2B, ttx-3::mCherry)	Young adult	98	1	0.254
sem-4	*kuIs34(sem-4::gfp)IV*	Young adult	100	0	(–)
	kuIs34(sem-4::gfp)IV	L1	44	3	(–)
sem-4	*kuIs35(sem-4::gfp)*	Young adult	100	0	(–)
	unc-86(n846)III; kuIs34(sem-4::gfp)IV	Young adult	19	3	0.0001
	unc-86(n846)III; kuIs34(sem-4::gfp)IV	L1	30	4	0.0087
	hlh-3(tm1688)II; kuIs34(sem-4::gfp)IV	Young adult	100	0	1
	egl-46(sy628)V; kuIs34(sem-4::gfp)IV	Young adult	96	2	0.0648
	ast-1(ot417)II; kuIs34(sem-4::gfp)IV	Young adult	100	0	1
	ast-1(hd92)II; kuIs34(sem-4::gfp)IV	L1	57	5	0.0392
	egl-18(ok290)IV; kuIs35(sem-4::gfp)	Young adult	95	2	0.0594
egl-46	*vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]*	Young adult	67	3	(–)
	vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	L1	53	7	(–)
	unc-86(n846)III; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	Young adult	22	4	0.0001
	unc-86(n846)III; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	L1	4	2	0.0001
	sem-4(n1971)I; zdIs13(tph-1::gfp)IV; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	Young adult	58	6	0.472
	hlh-3(tm1688)II; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	Young adult	28	4	0.0001
	hlh-3(tm1688)II; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	L1	52	5	1
	ast-1(ot417)II; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	Young adult	67	5	1
	ast-1(hd92)II; hdEx237[ast-1(+), rol-6(su1006)]; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	L1	96	2	0.449
	ast-1(hd92)II; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	L1	31	4	0.752
	egl-18(ok290)IV; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	Young adult	71	5	0.5985

↳

ast-1	*ast-1(vlc19[ast-1::gfp])II*	Young adult	99	1	(–)
	unc-86(n846)III; ast-1(vlc19[ast-1::gfp])II	Young adult	1	1	0.0001
	sem-4(n1971)I; ast-1(vlc19[ast-1::gfp])II	Young adult	6	2	0.0001
	hlh-3(tm1688)II; ast-1(vlc19[ast-1::gfp])II	Young adult	64	4	0.0001
	egl-46(sy628)V; ast-1(vlc19[ast-1::gfp])II	Young adult	94	2	0.0361
	egl-18(ok290)IV; ast-1(vlc19[ast-1::gfp])II	Young adult	95	2	0.085
egl-18	*unc-119(tm4063)III; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	Young adult	87	3	(–)
	unc-86(n846)III; stIs11606[egl-18a::H1-mCherry + unc-119(+)]	Young adult	79	4	0.1807
	sem-4(n1971)I; stIs11606[egl-18a::H1-mCherry + unc-119(+)]	Young adult	77	4	0.0694
	hlh-3(tm1688)II; stIs11606[egl-18a::H1-mCherry + unc-119(+)]	Young adult	79	4	0.1743
egl-18	*egl-46(sy628)V; stIs11606[egl-18a::H1-mCherry + unc-119(+)]*	Young adult	88	3	1
	ast-1(ot417)II; vlcEx324[egl-46::DsRed; ttx-3::mCherry; rol-6(su1006)]	Young adult	87	3	0.0002
	ast-1(hd92)II; hdEx237[ast-1(+), rol-6(su1006)]; unc-119(tm4063)III; stIs11606[egl-18a::H1-mCherry + unc-119(+)]	L1	72	4	(–)
	ast-1(hd92)II; vlcEx845[cat-1:.mCherry, ast-1(+)]; stIs11606[egl-18a::H1- mCherry + unc-119(+)]	L1	65	5	0.752
hlh-3	*otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]*	Embryo	90	3	(–)
	unc-86(n846)III; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]	Embryo	84	4	0.29
	sem-4(n1971)I; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]	Embryo	81	4	0.0884
	egl-46(sy628)V; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]	Embryo	92	3	0.7934
	ast-1(hd92)II; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]	Embryo	93	3	0.5895
	egl-18(ok290)IV; otEx4140[hlh-3fosmid::YFP, rol-6(su1006)]	Embryo	67	6	0.0002

Annex 3.2.7
Primary data and statistics for overexpression of the HSN regulatory code experiments, at different developmental stages (embryonic and larval)

Comparison analysis of the number of *tph-1::gfp* positive cells in the different experimental conditions of overexpression of factors (wt, single or combinations of factors). Three independent experiments were performed and all data has been pooled together. Statistical relationships between all of the conditions tested are showed. Non-parametric Kruskal-Wallis analysis with Dunn's correction for multiple comparisons was performed. ns: non-significant difference. See Figure 3.2.18.

Figure 3.2.18-A		Figure 3.2.18-B		Figure 3.2.18-C		Figure 3.2.18-D	
Dunn's multiple comparison	Significancy	Dunn's multiple comparison	Significancy	Dunn's multiple comparison	Significancy	Dunn's multiple comparison	Significancy
wt vs. *ast-1*	****	wt vs. *ast-1*	****	wt vs. *ast-1*	****	wt vs. *ast-1*	****
wt vs. *unc-86*	****	wt vs. *unc-86*	****	wt vs. *unc-86*	****	wt vs. *unc-86*	****
wt vs. *sem-4*	****	wt vs. *sem-4*	****	wt vs. *sem-4*	****	wt vs. *sem-4*	****
wt vs. *hlh-3*	ns	wt vs. *hlh-3* L1	ns	wt vs. *hlh-3*	**	wt vs. *hlh-3* L1	ns
wt vs. *egl-46*	ns	wt vs. *hlh-3* L2	ns	wt vs. *egl-46*	ns	wt vs. *hlh-3* L2	**
wt vs. *egl-18*	ns	wt vs. *egl-46* L1	ns	wt vs. *egl-18*	****	wt vs. *egl-46* L1	ns
wt vs. combo 'A+U+S'	****	wt vs. *egl-46* L2	ns	wt vs. combo 'A+U+S'	****	wt vs. *egl-46* L2	ns
wt vs. combo 6	****	wt vs. *egl-18* L1	ns	wt vs. combo 6	****	wt vs. *egl-18* L1	**
ast-1 vs. *unc-86*	*	wt vs. *egl-18* L2	ns	*ast-1* vs. *unc-86*	***	wt vs. *egl-18* L2	ns
ast-1 vs. *sem-4*	ns	wt vs. combo 'A+U+S'	****	*ast-1* vs. *sem-4*	ns	wt vs. combo 'A+U+S'	****
ast-1 vs. *hlh-3*	**	wt vs. combo 6 L1	****	*ast-1* vs. *hlh-3*	ns	wt vs. combo 6 L1	****
ast-1 vs. *egl-46*	**	wt vs. combo 6 L2	****	*ast-1* vs. *egl-46*	ns	wt vs. combo 6 L2	****
ast-1 vs. *egl-18*	**	*ast-1* vs. *unc-86*	ns	*ast-1* vs. *egl-18*	ns	*ast-1* vs. *unc-86*	**
ast-1 vs. combo 'A+U+S'	****	*ast-1* vs. *sem-4*	ns	*ast-1* vs. combo 'A+U+S'	****	*ast-1* vs. *sem-4*	ns
ast-1 vs. combo 6	****	*ast-1* vs. *hlh-3* L1	**	*ast-1* vs. combo 6	***	*ast-1* vs. *hlh-3* L1	ns
unc-86 vs. sem-4	ns	*ast-1* vs. *hlh-3* L2	ns	*unc-86* vs. *sem-4*	ns	*ast-1* vs. *hlh-3* L2	ns

unc-86 vs. *hlh-3*	****	*ast-1* vs. *egl-46* L1	ns	*unc-86* vs. *hlh-3*	**	*ast-1* vs. *egl-46* L1	ns
unc-86 vs. *egl-46*	****	*ast-1* vs. *egl-46* L2	*	*unc-86* vs. *egl-46*	*	*ast-1* vs. *egl-46* L2	ns
unc-86 vs. *egl-18*	****	*ast-1* vs. *egl-18* L1	**	*unc-86* vs. *egl-18*	**	ast-1 vs. egl-18 L1	ns
unc-86 vs. combo 'A+U+S'	ns	*ast-1* vs. *egl-18* L2	ns	*unc-86* vs. combo 'A+U+S'	ns	*ast-1* vs. *egl-18* L2	ns
unc-86 vs. combo 6	ns	*ast-1* vs. combo 'A+U+S'	****	*unc-86* vs. combo 6	ns	*ast-1* vs. combo 'A+U+S'	****
sem-4 vs. *hlh-3*	****	*ast-1* vs. combo 6 L1	****	*sem-4* vs. *hlh-3*	ns	*ast-1* vs. combo 6 L1	**
sem-4 vs. *egl-46*	****	*ast-1* vs. combo 6 L2	ns	*sem-4* vs. *egl-46*	ns	*ast-1* vs. combo 6 L2	ns
sem-4 vs. *egl-18*	****	*unc-86* vs. *sem-4*	ns	*sem-4* vs. *egl-18*	ns	*unc-86* vs. *sem-4*	ns
sem-4 vs. combo 'A+U+S'	*	*unc-86* vs. *hlh-3* L1	****	*sem-4* vs. combo 'A+U+S'	**	*unc-86* vs. *hlh-3* L1	ns
sem-4 vs. combo 6	ns	*unc-86* vs. *hlh-3* L2	**	*sem-4* vs. combo 6	ns	*unc-86* vs. *hlh-3* L2	*
hlh-3 vs. *egl-46*	ns	*unc-86* vs. *egl-46* L1	**	*hlh-3* vs. *egl-46*	ns	*unc-86* vs. *egl-46* L1	ns
hlh-3 vs. *egl-18*	ns	*unc-86* vs. *egl-46* L2	****	*hlh-3* vs. *egl-18*	ns	*unc-86* vs. *egl-46* L2	ns
hlh-3 vs. combo 'A+U+S'	****	*unc-86* vs. *egl-18* L1	****	*hlh-3* vs. combo 'A+U+S'	***	*unc-86* vs. *egl-18* L1	*
hlh-3 vs. combo 6	****	*unc-86* vs. *egl-18* L2	ns	*hlh-3* vs. combo 6	**	*unc-86* vs. *egl-18* L2	ns
egl-46 vs. *egl-18*	ns	*unc-86* vs. combo 'A+U+S'	ns	*egl-46* vs. *egl-18*	ns	*unc-86* vs. combo 'A+U+S'	ns
egl-46 vs. combo 'A+U+S'	****	*unc-86* vs. combo 6 L1	ns	*egl-46* vs. combo 'A+U+S'	**	*unc-86* vs. combo 6 L1	ns
egl-46 vs. combo 6	****	*unc-86* vs. combo 6 L2	ns	*egl-46* vs. combo 6	*	*unc-86* vs. combo 6 L2	ns
egl-18 vs. combo 'A+U+S'	****	*sem-4* vs. hlh-3 L1	****	*egl-18* vs. combo 'A+U+S'	***	*sem-4* vs. *hlh-3* L1	ns
egl-18 vs. combo 6	****	*sem-4* vs. hlh-3 L2	*	*egl-18* vs. combo 6	**	*sem-4* vs. *hlh-3* L2	ns
combo 'A+U+S' vs. combo 6	ns	*sem-4* vs. egl-46 L1	**	combo 'A+U+S' vs. combo 6	ns	*sem-4* vs. *egl-46* L1	ns
		sem-4 vs. egl-46 L2	****			*sem-4* vs. *egl-46* L2	ns

		sem-4 vs. egl-18 L1	****			*sem-4* vs. *egl-18* L1	ns
		sem-4 vs. egl-18 L2	ns			*sem-4* vs. *egl-18* L2	ns
		sem-4 vs. combo 'A+U+S'	ns			*sem-4* vs. combo 'A+U+S'	*
		sem-4 vs. combo 6 L1	ns			*sem-4* vs. combo 6 L1	ns
		sem-4 vs. combo 6 L2	ns			*sem-4* vs. combo 6 L2	ns
		hlh-3 L1 vs. *hlh-3* L2	ns			*hlh-3* L1 vs. *hlh-3* L2	ns
		hlh-3 L1 vs. *egl-46* L1	ns			*hlh-3* L1 vs. *egl-46* L1	ns
		hlh-3 L1 vs. *egl-46* L2	ns			*hlh-3* L1 vs. *egl-46* L2	ns
		hlh-3 L1 vs. *egl-18* L1	ns			*hlh-3* L1 vs. *egl-18* L1	ns
		hlh-3 L1 vs. egl-18 L2	ns			*hlh-3* L1 vs. *egl-18* L2	ns
		hlh-3 L1 vs. combo 'A+U+S'	****			*hlh-3* L1 vs. combo 'A+U+S'	ns
		hlh-3 L1 vs. combo 6 L1	****			*hlh-3* L1 vs. combo 6 L1	ns
		hlh-3 L1 vs. combo 6 L2	****			*hlh-3* L1 vs. combo 6 L2	ns
		hlh-3 L2 vs. *egl-46* L1	ns			*hlh-3* L2 vs. *egl-46* L1	ns
		hlh-3 L2 vs. *egl-46* L2	ns			*hlh-3* L2 vs. *egl-46* L2	ns
		hlh-3 L2 vs. *egl-18* L1	ns			*hlh-3* L2 vs. *egl-18* L1	ns
		hlh-3 L2 vs. *egl-18* L2	ns			*hlh-3* L2 vs. *egl-18* L2	ns
		hlh-3 L2 vs. combo 'A+U+S'	****			*hlh-3* L2 vs. combo 'A+U+S'	**
		hlh-3 L2 vs. combo 6 L1	****			*hlh-3* L2 vs. combo 6 L1	ns
		hlh-3 L2 vs. combo 6 L2	**			*hlh-3* L2 vs. combo 6 L2	ns
		egl-46 L1 vs. *egl-46* L2	ns			*egl-46* L1 vs. *egl-46* L2	ns

		egl-46 L1 vs. *egl-18* L1	ns			*egl-46* L1 vs. *egl-18* L1	ns
		egl-46 L1 vs. *egl-18* L2	ns			*egl-46* L1 vs. *egl-18* L2	ns
		egl-46 L1 vs. combo 'A+U+S'	****			*egl-46* L1 vs. combo 'A+U+S'	ns
		egl-46 L1 vs. combo 6 L1	****			*egl-46* L1 vs. combo 6 L1	ns
		egl-46 L1 vs. combo 6 L2	**			*egl-46* L1 vs. combo 6 L2	ns
		egl-46 L2 vs. *egl-18* L1	ns			*egl-46* L2 vs. *egl-18* L1	ns
		egl-46 L2 vs. *egl-18* L2	ns			*egl-46* L2 vs. *egl-18* L2	ns
		egl-46 L2 vs. combo 'A+U+S'	****			*egl-46* L2 vs. combo 'A+U+S'	ns
		egl-46 L2 vs. combo 6 L1	****			*egl-46* L2 vs. combo 6 L1	ns
		egl-46 L2 vs. combo 6 L2	***			*egl-46* L2 vs. combo 6 L2	ns
		egl-18 L1 vs. *egl 18* L2	ns			*egl-18* L1 vs. *egl-18* L2	ns
		egl-18 L1 vs. combo 'A+U+S'	****			*egl-18* L1 vs. combo 'A+U+S'	*
		egl-18 L1 vs. combo 6 L1	****			*egl-18* L1 vs. combo 6 L1	ns
		egl-18 L1 vs. combo 6 L2	****			*egl-18* L1 vs. combo 6 L2	ns
		egl-18 L2 vs. combo 'A+U+S'	***			*egl-18* L2 vs. combo 'A+U+S'	ns
		egl-18 L2 vs. combo 6 L1	*			*egl-18* L2 vs. combo 6 L1	ns
		egl-18 L2 vs. combo 6 L2	ns			*egl-18* L2 vs. combo 6 L2	ns
		combo 'A+U+S' vs. combo 6 L1	ns			combo 'A+U+S' vs. combo 6 L1	ns
		combo 'A+U+S' vs. combo 6 L2	ns			combo 'A+U+S' vs. combo 6 L2	ns
		combo 6 L1 vs. combo 6 L2	ns			combo 6 L1 vs. combo 6 L2	ns

Percentage of 'positive embryos'

'Positive embryos' are those that express *tph-1::gfp* in more than four cells after ectopic expression of single factors, or combinations of them, via heat shock treatment.

Three independent experiments were performed and all data has been pooled together. Statistical significance was calculated using two-tailed Fisher test; *:pV<0,05.

See Figure 3.2.18.

Overexpressed factor (hsp)	% *tph(–)1::gfp* +	SEP	N embryos	pV (vs wt)	pV (specified)
wt	3	1	196	<0.0001	(–)
ast-1	55	4	150	<0.0001	(–)
unc-86	63	6	70	<0.0001	(–)
sem-4	77	6	57	<0.0001	(–)
hlh-3	29	6	56	<0.0001	(–)
egl-46	21	6	43	0.0002	(–)
egl-18	33	6	67	<0.0001	(–)
combo 3	9 1	3	77	<0.0001	(–)
combo 6	87	5	54	<0.0001	(–)
unc-86 vs combo 'A+U+S'	(–)	(–)	(–)	(–)	<0.0001
unc-86 vs combo 6	(–)	(–)	(–)	(–)	0.0037
sem-4 vs combo 'A+U+S'	(–)	(–)	(–)	(–)	0.0477
sem-4 vs combo 6	(–)	(–)	(–)	(–)	0.2204

Percentage of ectopic cells at larval stages

Scoring of ectopic *tph-1::gfp* expressing cells at larval stages L1 and L2, after overexpression of single factors, or combinations of them, via heat shock treatment at L1 stage. Three independent experiments were performed and all data has been pooled together.

Statistical significance was calculated using two-tailed Fisher test; *:pV<0,05.
See Figure 3.2.18.

	Head neuron				PVT				Tail neurons (ALN + unknown)			
Overexpressed factor	%	SEP	N	pV	%	SEP	N	pV	%	SEP	N	pV
wt	0	0	58	(–)	3	2	58	(–)	0	0	58	(–)
hsp::ast-1	0	0	74	1	9	3	74	0.2975	27	5	74	0.0001
hsp::unc-86	91	3	108	0.0001	69	4	108	0.0001	0	0	108	1
hsp::sem-4	33	9	30	0.0001	17	7	30	0.0429	0	0	30	1
hsp::combo 6	47	6	70	0.001	71	5	70	0.0001	15	4	70	0.0001

Annex 3.2.8
Primary data of serotonin pathway gene expression in double mutants

Analysis of *tph-1*, *bas-1* and *cat-1* in the different mutant backgrounds for the six candidate regulators of the HSN neuron. See Figure 3.2.21.

Genetic relationship	Genotype	% HSN	SEP	pV
Synergistic enhancement	*egl-46(sy628)V; otIs226(bas-1::gfp)IV*	84.5	3.4	(–)
	hlh-3(tm1688)II; otIs226(bas-1::gfp)IV	49.0	5.0	(–)
	hlh-3(tm1688)II; egl-46(sy628)V; otIs226(bas-1::gfp)IV	0.0	0.0	0.0001
	ast-1(ot417)II; otIs226(bas-1::gfp)IV	100.0	0.0	(–)
	egl-46(sy628)V; otIs226(bas-1::gfp)IV	84.5	3.4	(–)
	ast-1(ot417)II; egl-46(sy628)V; otIs226(bas-1::gfp)IV	1.9	1.3	0.0001
	ast-1(ot417)II; otIs226(bas-1::gfp)IV	100.0	0.0	(–)
	sem-4(n2654)I; otIs226(bas-1::gfp)IV	66.0	4.7	(–)
	sem-4(n2654)I; ast-1(ot417)II; otIs226(bas-1::gfp)IV	0.0	0.0	0.0001
	ast-1(ot417)II; otIs226(bas-1::gfp)IV	100.0	0.0	(–)
	ast-1(ot417)II; otIs226(bas-1::gfp)IV	100.0	0.0	(–)
	unc-86(n848)III; otIs226[bas-1::gfp]IV	45.0	5.0	(–)
	ast-1(ot417)II; unc-86(n848)III; otIs226[bas-1::gfp]IV	0.0	0.0	0.0001
	egl-18(ok290)IV; otIs221(cat-1::gfp)III	96.8	1.6	(–)
	hlh-3(tm1688)II; otIs221(cat-1::gfp)III	81.0	3.5	(–)
	hlh-3(tm1688)II; egl-18(ok290)IV; otIs221(cat-1::gfp)III	28.7	4.4	0.0001
	egl-18(ok290)IV; otIs221(cat-1::gfp)III	96.8	1.6	(–)
	sem-4(n2654)I; otIs221(cat-1::gp)III	48.0	4.9	(–)
	egl-18(ok290)IV; sem-4(n2654)I; otIs221(cat-1::gp)III	30.0	4.6	0.0405
	egl-18(ok290)IV; zdIs13(tph-1::gp)IV	91.0	2.9	(–)
	sem-4(n2654)I; zdIs13(tph-1::gp)IV	54.0	5.0	(–)
	egl-18(ok290)IV; sem-4(n2654)I; zdIs13(tph-1::gp)IV	30.0	4.6	0.0405
Synergistic suppression	*egl-18(ok290)IV; otIs221(cat-1::gfp)III*	96.8	1.6	(–)
	egl-46(sy628)V; otIs221(cat-1::gfp)III	61.3	4.7	(–)
	egl-18(ok290)IV; egl-46(sy628)V; otIs221(cat-1::gfp)III	94.3	2.2	0.0001
	hlh-3(tm1688)II; otIs221(cat-1::gfp)III	81.0	3.5	(–)
	egl-46(sy628)V; otIs221(cat-1::gfp)III	61.3	4.7	(–)
	hlh-3(tm1688)II; egl-46(sy628)V; otIs221(cat-1::gfp)III	72.0	4.5	0.0001
	sem-4(n2654)I; zdIs13(tph-1::gfp)IV	54.0	5.0	(–)
	ast-1(ot417)II; zdIs13(tph-1::gfp)IV	0.0	0.0	(–)
	sem-4(n2654)I; ast-1(ot417)II; zdIs13(tph-1::gfp)IV	19.0	4.0	0.0001
Additivity	*unc-86(n848)III; otIs224(cat-1::gfp)V*	60.0	4.5	(–)
	sem-4(n2654)I; otIs224(cat-1::gfp)V	57.0	5.0	(–)
	sem-4(n2654)I; unc-86(n848)III; otIs224(cat-1::gfp)V	9.0	2.9	1
	hlh-3(tm1688)II; otIs224(cat-1::gfp)V	82.0	3.6	(–)
	unc-86(n848)III; otIs224(cat-1::gfp)V	60.0	4.5	(–)
	hlh-3(tm1688)II; unc-86(n848)III; otIs224(cat-1::gfp)V	30.0	4.5	0.1048
Epistasis	*egl-18(ok290)IV; otex2435 [bas1prom1 gfp (50ng/ul), rol6]*	78.6	3.5	(–)
	hlh-3(tm1688)II; otex2435 [bas1prom1 gfp (50ng/ul), rol6]	41.0	5.2	(–)
	hlh-3(tm1688)II; egl-18(ok290)IV; otex2435 [bas1prom1 gfp (50ng/ul), rol6]	33.0	4.6	0.0355

Gene name	Functional category
aak-2	Terminal feature
abts-1	Neurotransmission
ags-3	Terminal feature
aho-3	Terminal feature
ari-1	Terminal feature
arr-1	Terminal feature
arrd-17	Terminal feature
bas-1	Neurotransmission
baz-2	Other
cam-1	Migration
cat-1	Neurotransmission
cat-4	Neurotransmission
cdh-3	Sinaptogenesis
ceh-20	TF
che-7	Terminal feature
clh-3	Neurotransmission
dhc-3	Terminal feature
dsh-1	Wnt pathway
eat-16	Terminal feature
ebax-1	Migration
egl-18	TF
egl-43	TF
egl-44	TF
egl-46	TF
egl-47	Terminal feature
egl-5	TF
egl-6	Neurotransmission
elpc-1	Terminal feature
elpc-3	Terminal feature
eor-2	Wnt pathway
flp-19	Neurotransmission
fmi-1	Axon guidance
gar-2	Neurotransmission
gei-8	TF
ggr-2	Neurotransmission
glr-5	Neurotransmission
goa-1	Terminal feature
grd-6	Terminal feature
gsa-1	Terminal feature
ham-2	Migration
hlh-14	TF
hlh-3	TF
ida-1	Neurotransmission
ife-4	TF
ins-18	Neurotransmission
inx-3	Neurotransmission
inx-7	Neurotransmission
irk-1	Neurotransmission

Gene name	Functional category
kal-1	Migration
kcc-2	Neurotransmission
kin-20	Terminal feature
lgc-55	Neurotransmission
mab-23	TF
madd-2	Migration
mau-2	Migration
mec-6	Neurotransmission
mig-10	Migration
mig-2	Migration
mom-2	Wnt pathway
mpz-1	Neurotransmission
nck-1	Axon guidance
nhx-5	Neurotransmission
nid-1	Migration / NT
nlg-1	Neurotransmission
nlp-1	Neurotransmission
nlp-15	Neurotransmission
nlp-3	Neurotransmission
nra-4	Neurotransmission
plr-1	Wnt pathway
prkl-1	Wnt pathway
ptps-1	Neurotransmission
puf-9	Terminal feature
pxf-1	Terminal feature
rep-1	Terminal feature
rig-6	Terminal feature
rsbp-1	Terminal feature
rsy-1	Neurotransmission
sel-10	Notch pathway
sem-4	TF
syg-1	Neurotransmission
tba-6	Terminal feature
tbb-4	Terminal feature
ten-1	Terminal feature
top-1	Terminal feature
tph-1	Neurotransmission
unc-103	Neurotransmission
unc-14	Migration
unc-17	Neurotransmission
unc-2	Neurotransmission
unc-40	Migration
unc-51	Migration
unc-53	Migration
unc-77	Neurotransmission
unc-86	TF
unc-9	Neurotransmission
vang-1	Wnt pathway

Annex 3.3.1
Gene expression profile of the HSN neuron

96 genes are known to be expressed in the HSN neuron, excluding panneuronal features ((Hobert et al. 2016), (www.wormbase.org)). See Figure 3.3.1 and 3.3.2.

Annex 3.3.2
List of random genes used in the 'sliding window analysis'

The selected genes have a similar upstream and intronic distribution to HSN expressed genes. See Figure 3.3.1.

Gene name	Gene name
aat-8	*elo-5*
abu-12	*ent-4*
acly-1	*eri-12*
alh-13	*F09C8.2*
B0403.6	*F10E7.9*
B0511.6	*F26F12.8*
bus-19	*F26G1.1*
C01B10.6	*F28C1.3*
C05G5.2	*F39G3.2*
C07A9.2	*F52G3.5*
C27F2.8	*F53F8.7*
C33D3.5	*F59E12.8*
C34B2.9	*fut-6*
C39D10.7	*gcy-23*
C43H6.6	*gon-1*
C44F1.1	*H11E01.3*
C49H3.16	*H12D21.9*
C54G6.2	*hbl-1*
cand-1	*her-1*
ceh-10	*hlh-13*
cutl-3	*hum-1*
D1086.17	*inx-11*
daf-16	*ist-1*
eef-1A.2	*K08D12.8*
egl-13	*K09E4.1*

Gene name	Gene name
kvs-4	*T21F4.1*
lgc-25	*T24B8.3*
lgc-52	*T25B9.3*
M03D4.4	*T28F3.5*
M04C3.5	*tag-273*
mct-2	*tbc-7*
mua-6	*tbx-33*
mvk-1	*tiam-1*
nep-18	*toca-1*
nhr-45	*tsp-3*
perm-5	*unc-44*
prk-1	*xol-1*
R01B10.2	*Y37A1A.4*
R05G6.1	*Y38H8A.1*
rgs-4	*Y39G10AR.11*
set-23	*Y47D3A.32*
shl-1	*Y48G1BL.7*
sqv-5	*Y55F3AM.21*
srg-51	*Y62H9A.13*
srh-48	*Y66C5A.2*
sru-47	*Y71H2B.1*
T02G6.5	*Y73B3A.16*
T07G12.5	*Y7A5A.7*
T13H5.1	*Y87G2A.2*
T14G8.3	*ZC84.1*

Promoter	TF pair	Modified binding site	Mutation (wt BS> mutant BS)
tph-1prom2	(–)	(–)	(–)
tph-1prom58	GATA-ETS	GATA	TCCGGATATTA > taatatcticcTCCGGAaATTA
tph-1prom59	HLH-ETS	HLH	TTCCAGAAGC > ttccagaag
bas-1prom13	(–)	(–)	(–)
bas-1prom87	HLH-ETS	HLH	CTTTCTGCCAGAATT > tggcagaaagTGCCAGAATT
bas-1prom89	GATA-ETS	GATA	TCTATCCGTT > tctatccgttacagataga
cat-1prom14	(–)	(–)	(–)
cat-1prom85	HLH-ETS	HLH	CATTCTGGTTTTCCG > aaccagaatggGTTTCCG
cat-1prom86	SPALT-ETS	SPALT	AATTGTCTTG > cagacaatt
cat-1prom87	HLH-ETS	ETS	GGTTTCCGTT > aacggaaacc

Annex 3.3.3
Specific DNA modifications for motif orientation analysis of the HSN regulatory signature

Specific BSs were flipped in order to test functionality of the overrepresented TFBS pairs (ETS-HLH and ETS-GATA). The two flanking nucleotides to the motif were also considered when flipping. Uppercase letters indicate wild type nucleotides, while lowercase letters indicate mutated nucleotides that alter a specific BS motif.

Whenever two motifs overlapped, as in the case of ETS-GATA pair, or were directly next to each other, as in one particular HLH-ETS pair, additional point mutations were introduced in the non-flipped binding site, or immediately next to it, in order to maintain a unique motif in each orientation, and are indicated in light grey. See Figure 3.3.7.

Annex 3.4.1
Primary data of HSN rescue experiments using mouse factors

Analysis of *tph-1::gfp* expression in mutant animals for the HSN TF collective, carrying an extrachromosomal 'rescue array' that contains the worm or its orthologue mouse factor, under an HSN specific promoter. (See Figure 3.4.4).

Genotype	% HSN	SEP	pV
zdIs13(tph-1::gfp)IV	100	0	(–)
ast-1(ot417)II; zdIs13(tph-1::gfp)IV	0	0	(–)
ast-1(ot417)II; zdIs13(tph-1::gfp)IV; vlcEx148[bas-1prom::ast-1, ttx-3::mCherry; rol-6(su1006)]	46	5	0.0001
ast-1(ot417)II; zdIs13(tph-1::gfp)IV; vlcEx148[bas-1prom::Pet1, ttx-3::mCherry; rol-6(su1006)]	83	4	0.0001
zdIs13(tph-1::gfp)IV	100	0	(–)
unc-86(n846)III; zdIs13(tph-1::gfp)IV	0	0	(–)
unc-86(n846)III; zdIs13(tph-1::gfp)IV;vlcEx503[kal-1prom::unc.86 genomic, ttx-3::mCherry, rol-6(su1006)]	0	0	1
zdIs13(tph-1::gfp)IV	100	0	(–)
sem-4(n1971)I; zdIs13(tph-1::gfp)IV	52	3	(–)
sem-4(n1971)I; zdIs13(tph-1::gfp)IV; vlcEx511[kal-1prom::Sall2; ttx-3::mCherry, rol-6(su1006)]	87	4	0.0001
zdIs13(tph-1::gfp)IV	100	0	(–)
hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV	22	4	(–)
hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV; vlcEx458[cat-4prom::hlh-3; ttx-3::mCherry, rol-6(su1006)]	66	4	0.0001
hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV; vlcEx479[cat-4prom::Ascl1, ttx-3::mCherry, rol-6(su1006)]	43	5	0.0024
hlh-3(tm1688)II; zdIs13(tph-1::gfp)IV; vlcEx480[cat-4prom::Ascl1, ttx-3::mCherry, rol-6(su1006)]	57	5	0.0001
zdIs13(tph-1::gfp)IV	100	0	(–)
egl-46(sy628)V; zdIs13(tph-1::gfp)IV	62	5	(–)
egl-46(sy628)V; zdIs13(tph-1::gfp)IV; vlcEx471[cat-4prom::egl-46, ttx-3::mCherry, rol-6(su1006)]	90	2	0.0001
egl-46(sy628)V; zdIs13(tph-1::gfp)IV; vlcEx472[cat-4prom::egl-46, ttx-3::mCherry, rol-6(su1006)]	96	2	0.0001
egl-46(sy628)V; zdIs13(tph-1::gfp)IV; vlcEx481[cat-4prom::Inms1; ttx-3::mCherry, rol-6(su1006)]	97	2	0.0001
yzIs71[tph-1::gfp, rol-6(su1006)]V	96	1	(–)
egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V	75	4	(–)
egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V; vlcEx469[cat-4prom2::egl-18, ttx-3::mCherry, rol-6(su1006)]	86	5	0.1858
egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V; vlcEx470[cat-4prom2::egl-18, ttx-3::mCherry, rol-6(su1006)]	69	4	0.3594
egl-18(ok290)IV; yzIs71[tph-1::gfp, rol-6(su1006)]V; vlcEx485[cat-4prom2::Gata3, ttx-3::mCherry, rol-6(su1006)]	79	4	0.5042

Neuron	Number of genes
ADE	32
ADF	77
ADL	101
AFD	56
AIY	71
ALM	83
ALN	31
AQR	45
ASE	132
ASG	58
ASH	98
ASI	133
ASJ	79
ASK	99
AUA	30
AVA	53
AVM	64
AWA	45
AWB	74
AWC	102
BAG	41
BDU	32
CAN	65
CEPD	36

Neuron	Number of genes
CEPV	32
DVA	36
DVC	30
FLP	37
HSN	98
NSM	44
OLL	38
OLQD	36
OLQV	36
PDE	39
PHA	90
PHB	89
PLM	84
PQR	52
PVC	43
PVD	49
PVM	54
PVQ	45
PVT	85
RIC	32
RMED	30
RMEV	30
SDQ	35
SMDD	32
URX	42

Annex 3.4.2
Caenorhabditis elegans **neuronal profiles**

Worm neurons used in the Principal Coordinate Analysis, which are known to express at least 30 genes. See Figures 3.4.5 and 3.4.6.

Summary — Spanish

Lógica de la regulación transcripcional de las neuronas serotonérgicas en *Caenorhabditis elegans*

Introducción

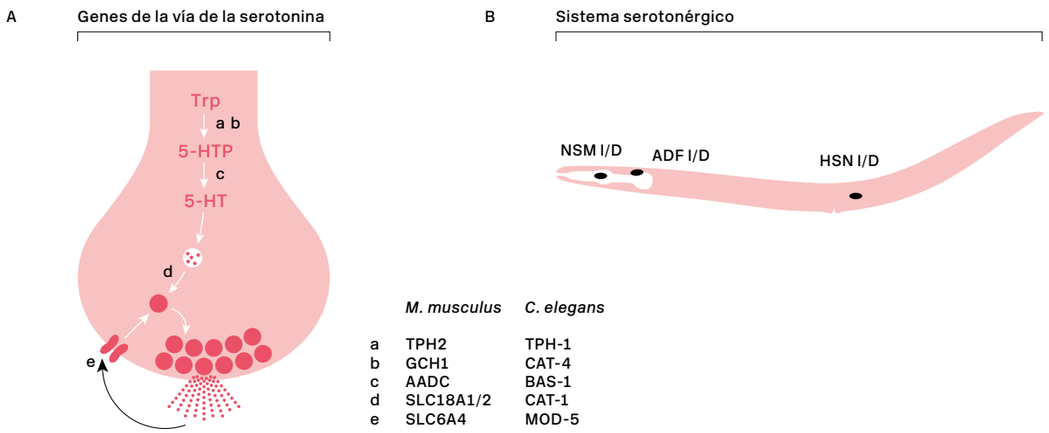
La generación de una clase neuronal concreta del sistema nervioso es un proceso que consta de varias etapas de elección y compromiso de destino celular neuronal y que culmina con la activación de una batería de genes específica, que define las propiedades únicas de la neurona madura y funcional (Hobert 2005). Este conjunto de genes específicos, llamados genes efectores, se expresan a lo largo de la vida de la neurona adulta (Hobert 2016b) y le confiere su identidad única; es decir, su identidad o huella molecular. La composición de estas baterías de genes es combinatoria, en el sentido de que cada subtipo neuronal no expresa unos genes exclusivos, sino una combinación exclusiva de genes que, a su vez, se pueden expresar de una manera más amplia (Wenick & Hobert 2004). Este mecanismo puede dar lugar a la construcción de un número casi infinito de diferentes patrones de expresión específicos de tipo neuronal y, por tanto, de tipos neuronales. Por tanto, la cuestión de la diferenciación de los diferentes tipos neuronales se podría reformular a cómo se ejecutan los programas de expresión génica específicos de tipo neuronal.

Un modelo prevalente para explicar la adquisición de la identidad postmitótica de una neurona es el de los 'selectores terminales' (Hobert 2008). Selectores terminales son aquellos factores de transcripción (FT) que se activan entre el final de la mitosis y el estadio postmitótico, y que controlan directamente la identidad terminal de los tipos neuronales en el sistema nervioso. Estos FT reconocen y se unen a regiones reguladoras del ADN que son comunes a todos los genes de diferenciación terminal de la neurona, generalmente activando la expresión génica, aunque cada vez hay más evidencias del papel de genes represores en la generación de diversidad neuronal (Kerk et al. 2017). El concepto de selector terminal implica que un tipo neuronal específico no necesitará, *a priori*, un gran número de FT para regular cada una de sus características terminales (transcriptoma celular), sino que todas estos genes efectores, aunque no estén relacionados entre sí, serán co-regulados por uno o, más habitualmente, una combinación de selectores terminales (Xue et al. 1993; Wenick & Hobert 2004; Doitsidou et al. 2013). El nematodo *Caenorhabditis elegans* ha sido ampliamente utilizado en el estudio de la lógica de la regulación transcripcional de muchos tipos neuronales. Por ejemplo, el FT UNC-3, de la familia de los dedos de zinc, regula la diferenciación terminal de varios tipos de neuronas colinérgicas (Kratsios et al. 2011; Pereira et al. 2015), mientras que una combinación específica de tres FT pertenecientes a diferentes familias (AST-1 (FT ETS), CEH-43 (FT DLX) y CEH-20/CEH-40 (FT PBX)) regula de manera directa la identidad terminal de todas las neuronas dopaminérgicas del gusano (Flames & Hobert 2011; Doitsidou et al. 2013). Aunque descrito por primera vez en *C. elegans*, este modelo se extiende a otros

Figura 1
Neuronas serotonérgicas y sistema serotonérgico en *C. elegans*

A) Ruta de biosíntesis de serotonina (5-HT). Abreviaturas: 5-HT: serotonina; 5-HTP: 5 hidroxitriptófano; BAS-1: síntesis de aminas biógenas 1; AADC: aminoácido descarboxilasa; GCH1: GTP ciclohidrolasa 1; MOD-5: modulación de locomoción defectuosa; SERT: transportador de serotonina; TPH: triptófano hidroxilasa; Trp: triptófano; SLC18A1/2: familia de transportador de solutos 18 miembro A1/A2 (también llamado VMAT: transportador vesicular de monoaminas); SLC6A4: familia de transportador de solutos 6 miembro 4.

B) Sistema serotonérgico del hermafrodita *C. elegans*. I/D: neuronas bilaterales (izquierda/derecha).



	M. musculus	*C. elegans*
a	TPH2	TPH-1
b	GCH1	CAT-4
c	AADC	BAS-1
d	SLC18A1/2	CAT-1
e	SLC6A4	MOD-5

animales más complejos como los vertebrados. Por ejemplo, PET1 muestra características clave de selector terminal para las neuronas serotonérgicas del raphe en ratones (Hendricks et al. 1999; Hendricks et al. 2003), mientras que una combinación de dos FT, NURR1 y PITX3, regula la diferenciación terminal de las neuronas dopaminérgicas del mesencéfalo de ratón (Jacobs, van der Linden et al. 2009; Jacobs, van Erp, et al. 2009). Además, el conjunto de NGN2, ISL1 y LHX3 es suficiente para reprogramar células embrionarias de ratón a neuronas motoras espinales (Mazzoni et al. 2013).

De acuerdo con la hipótesis de los selectores terminales, combinaciones específicas de FT se encargan de la regulación total o parcial del transcriptoma, de un tipo neuronal concreto. Está bien establecido que los FT se unen de una manera combinatoria y cooperativa a secuencias de ADN presentes en los elementos de regulación en *cis* del genoma, llamados potenciadores *(enhancers)* (Reiter et al. 2017). Esto otorga a los FT un papel central en la regulación de la expresión génica. A lo largo de la última década, el desarrollo de tecnología de alto rendimiento ha permitido la caracterización, tanto *in vitro* como *in vivo,* de la presencia de sitios de unión de FT y potenciadores funcionales a nivel genómico. Métodos tales como la inmunoprecipitación de cromatina seguida de secuenciación han permitido identificar potenciadores activos asociados a marcas específicas de cromatina, medir el

grado de ocupación de los sitios de unión a nivel genómico, trazar las preferencias de unión de numerosos FT y asociar zonas reguladoras activas en el genoma (*regulatory landscape*) con su correspondiente tipo celular (revisado en (Levo & Segal 2014)). Por otro lado, se ha descrito que la llamada arquitectura de un potenciador puede contener ciertas propiedades restrictivas en cuanto al número, localización, orientación y orden de los sitios de unión a FT, las cuales se conocen como 'normas sintácticas o gramaticales' de una secuencia reguladora (Spitz & Furlong 2012). Por ejemplo, a mayor número de sitios de unión para un mismo FT (agrupaciones homotípicas) y, sobro todo, a mayor número de diferentes sitios de unión para diferentes FT (agrupaciones heterotípicas) mayor es la expresión predicha para un mismo potenciador (Smith et al. 2013).

Sin embargo, experimentos masivos de caracterización funcional de la actividad de los potenciadores en el genoma han desvelado que sólo una pequeña fracción de los potenciales sitios de unión de FT del genoma eucariota se encuentran realmente ocupados por FTs en cualquier tipo celular dado (Whitfield et al. 2012; Kheradpour et al. 2013; White et al. 2013; Kwasnieski et al. 2014). Además sólo una fracción de los FT que se unen a sitios de unión se corresponden con potenciadores activos (Kwasnieski et al. 2014; White et al. 2013; Fisher et al. 2012). Todo ello pone de manifiesto que todavía no somos capaces de distinguir sitios de unión y potenciadores funcionales de los no funcionales, así como que desconocemos los mecanismos por los que estas combinaciones de FT identifican y activan sus secuencias diana. En este trabajo hemos utilizado las neuronas serotonérgicas como paradigma de investigación de las leyes que regulan la selección y activación del transcriptoma de un tipo neuronal en concreto, las neuronas serotonérgicas, durante la diferenciación terminal.

Las neuronas serotonérgicas se encuentran presentes en todos los grupos de eumetazoos y se definen por su habilidad de sintetizar y liberar serotonina (5-HT), lo cual es posible gracias a la expresión de los llamados 'genes de la vía de la 5-HT' → **Figura 1-A.** Dada su relevancia clínica y el gran número de procesos en los que están implicadas (Deneris & Wyler 2012), estas neuronas han sido ampliamente estudiadas en los últimos años, tanto en mamíferos como en nematodos.

El sistema serotonérgico de *C. elegans* consta de tres pares de neuronas con diferente función: la neurona motora neurosecretora NSM, la neurona secretora ADF y la neurona motora HSN → **Figura 1-B.** En cuanto la regulación de su diferenciación terminal, se sabe que la identidad celular de la neurona NSM viene determinada por la pareja de selectores terminales TTX-3 (FT LIM homeodominio) y UNC-86 (FT POU homeodominio) (Zhang et al. 2014), mientras que en la regulación de la neurona ADF participa DAF-19 (FT RFX) (Xie et al. 2013). Múltiples FT se han asociado al desarrollo de la neurona HSN (Desai et al. 1988; Basson & H Robert Horvitz 1996; Doonan et al. 2008; Sze et al. 2002), destacando UNC-86 como el mejor candidato a selector terminal.

Aprovechando la elevada conservación filogenética de las neuronas serotonérgicas, hemos utilizado el organismo modelo *C. elegans* para diseccionar su lógica de regulación transcripcional.

Objetivos

Los objetivos específicos de esta tesis son los siguientes:

1 Diseccionar *in vivo* la lógica de regulación en *cis* de los genes de la vía de la 5-HT en los diferentes subtipos serotonérgicos de *C. elegans,* NSM, ADF y HSN.

2 Identificar y caracterizar en profundidad los FT que controlan el programa de diferenciación ter-

minal del subtipo serotonérgico HSN (selectores terminales de la neurona HSN), siguiendo una aproximación por genes candidatos.

3 Interrogar el transcriptoma de la neurona HSN para la presencia de una huella de identidad reguladora, codificada en la secuencia primaria de ADN, que permita la identificación de potenciadores funcionales en la neurona HSN a nivel genómico.

4 Determinar si el programa de regulación de las neuronas serotonérgicas está conservado filogenéticamente, en términos moleculares y funcionales, entre nematodos y mamíferos.

Resultados y metodología

1 Diferentes módulos de regulación en *cis* controlan la expresión de los genes de la vía de la serotonina en las diferentes subclases de neuronas serotonérgicas

Llevamos a cabo un análisis de las regiones reguladoras de los genes de la vía de la 5-HT (*tph-1*, *cat-1*, *bas-1*, *cat-4* y *mod-5*) mediante la creación de reporteros transgénicos que expresan la proteína fluorescente GFP bajo el control de estas regiones reguladoras, en diferentes longitudes. De este modo, aislamos la región reguladora mínima de cada gen que es capaz de activar la expresión de GFP en cada uno de los subtipos neuronales, a las que denominamos módulos de regulación en *cis* (MRC). Los resultados de este análisis revelan que MRC independientes son necesarios para activar la expresión de cada gen en cada subtipo serotonérgico, como se ha esquematizado en la **→ Figura 2.** Esta organización modular concuerda con un modelo de regulación donde diferentes selectores terminales regulan la expresión de los genes de la vía de la 5-HT en NSM, ADF y HSN. Teniendo en cuenta esta lógica de regulación serotonérgica dependiente de subtipo celular, decidimos enfocar el resto de nuestro trabajo en el estudio en la neurona HSN, por ser la mejor caracterizada hasta la fecha.

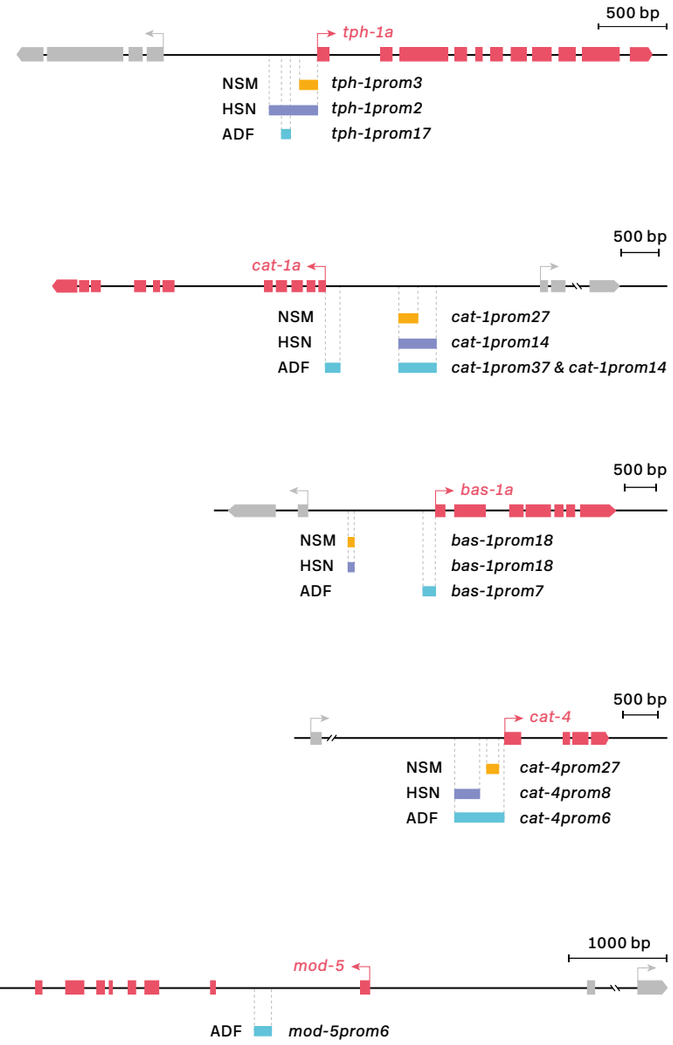
2.1 FT pertenecientes a seis familias diferentes son necesarios para la diferenciación terminal de la neurona HSN

La neurona HSN regula los músculos de la vulva del gusano y, por tanto, su disfunción provoca un fenotipo muy evidente de defecto en la puesta de huevos. Siguiendo una aproximación por gen candidato, elegimos seis genes que codifican para FT y cuyos mutantes presentan este fenotipo y defectos de tinción de 5-HT en la neurona HSN. De este modo seleccionamos como posibles selectores terminales de la neurona HSN al FT UNC-86 de la familia POU, al FT SEM-4 de la familia SPALT, al FT HLH-3 de la familia bHLH, al FT EGL-46 de la familia INSM (Desai et al. 1988; Basson & Horvitz 1996; Doonan et al. 2008; Wu et al. 2001), al FT AST-1 de la familia ETS y al FT EGL-18 de la familia GATA (ambos por observaciones en nuestro laboratorio).

Cruzamos alelos de pérdida de función para los seis genes candidatos con reporteros transcripcionales fluorescentes de los cuatro genes de la vía de la 5-HT que se expresan en la neurona HSN (*tph-1*, *cat-1*, *bas-1* y *cat-4*) y de nueve genes del transcriptoma de HSN que no están relacionados con la síntesis de 5-HT (*kcc-2*, *lgc-55*, *ida-1*, *flp-19*, *unc-17*, *unc-40*, *nlg-1*, *rab-3* y *kal-1*). Observamos que la expresión de estos reporteros se ve afectada a distintos niveles en los distintos fondos mutantes, como se representa en el *heatmap* de la **→ Figura 3.** Por ejemplo, en mutantes *unc-86* y *sem-4* la expresión de casi todos los genes analizados se vio altamente afectada, mientras que en mutantes para *egl-18* observamos una pérdida de expresión génica más modesta y sólo en algunos genes. En cualquier caso, la expresión de los seis FT es necesaria para una correcta diferenciación terminal de HSN y, por motivos de brevedad, los llamamos conjunta-

Figura 2
Resumen de la lógica de regulación de las neuronas serotonérgicas de *C. elegans*

El análisis de las regiones reguladoras revela que módulos de regulación en *cis* (MRC) independientes controlan la expresión de los genes de la vía de la 5-HT en los distintos subtipos neuronales, y predice que estos MRC serán activas por diferentes factores de transcripción.



mente 'código regulador de HSN'. De acuerdo con los resultados del análisis de MRC, los defectos de expresión son principalmente específicos de la neurona HSN. Algunos genes como *kal-1* no vieron su expresión prácticamente afectada, lo que nos permite saber que, en estos mutantes, la neurona HSN sí que se genera y adquiere un fenotipo neuronal, pero no alcanza a completar su programa de diferenciación terminal y ve su transcriptoma alterado. También es interesante la apreciación de que el patrón de expresión de cada mutante es ligera o radicalmente diferente, sugiriendo que cada FT actúa en rutas independientes.

2.2 Los seis FT del código regulador de la HSN actúan directamente sobre sus genes diana

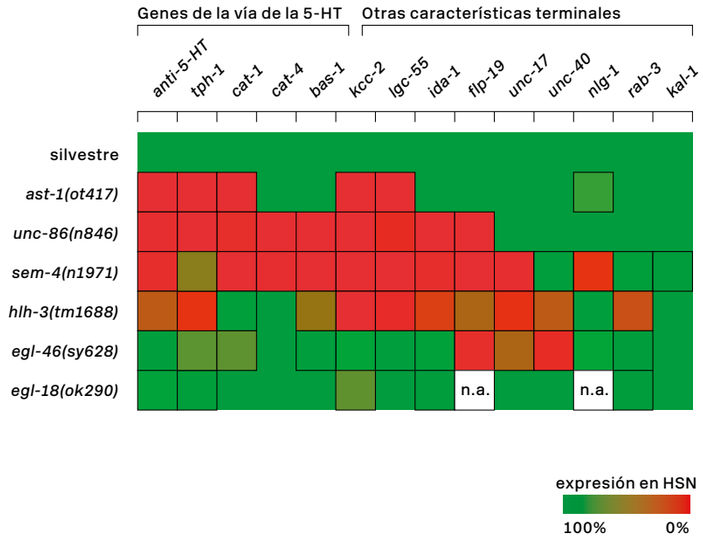
Quisimos determinar si el fenotipo observado en los animales mutantes sobre el transcriptoma de la neurona HSN era debido a una activación directa por parte de los FT que componen el código regulador de HSN o, por el contario, podría ser un efecto indirecto en el que alguno o todos los miembros del código actúan aguas arriba en la cascada transcripcional. Con esta finalidad, llevamos a cabo un análisis de los MRC previamente aislados de los genes *tph-1, cat-1* y *bas-1*, los que mayor fenotipo mutante tenían, donde buscamos sitios de unión predichos bioinformáticamente para los seis miembros del código. En todos los MRC encontramos al menos un sitio para cada una de las seis familias de FT, los cuales truncamos mediante mutagénesis dirigida y analizamos su expresión resultante *in vivo* en la HSN. En los casos donde la señal de GFP disminuyó o se perdió, inferimos que ese sitio de unión en concreto es funcional y el FT complementario se une de manera directa. De este modo encontramos sitios funcionales para los seis miembros del código y concluimos que regulan de manera directa la expresión de los genes de la vía de la 5-HT. En la → **Figura 4** se ha incluido, a modo de ejemplo, el MRC para *tph-1*. Curiosamente, observamos casos de redundancia entre sitios de unión de la misma familia de FT, como con los GATA, o incluso entre diferentes familias, entre los GATA y ETS.

2.3 El 'código regulador de HSN' actúa mediante rutas paralelas

Los resultados obtenidos hasta el momento sugieren que los seis miembros del código regulador de HSN actúan de manera directa e independiente sobre los genes de la vía de la serotonina. Nos planteamos si, además, podrían mostrar relaciones de regulación cruzada. Para comprobarlo, primero obtuvimos cepas reporteras para los seis miembros del código, tanto de otros laboratorios y de la CGC (Caenorhabditis Genetics Center), como por métodos de clonación tradicional o CRISPR en nuestro laboratorio con aquellos que no estaban disponibles. Confirmamos que los seis FT se expresan en la HSN en el gusano adulto, a excepción de *hlh-3,* cuya expresión sólo se observa en la célula precursora de la HSN. A continuación, cruzamos estos reporteros con los diferentes mutantes y analizamos su expresión en la neurona HSN. Descubrimos que la expresión de cada miembro es principalmente independiente del código en sí mismo, con algunas excepciones, indicando que los seis FT actúan de manera independiente para regular la identidad de la neurona HSN. A destacar, UNC-86 aparece como regulador principal del código (controlando la expresión de EGL-46, SEM-4 y/o AST-1), mientras que la actividad de AST-1 está regulada por otros miembros del código (HLH-3, SEM-4 y/o UNC-86) → **Figura 5.**

Figura 3
Heatmap resumen de la caracterización de los seis mutantes candidatos para la HSN

kcc-2: co-transportador de cloruro de potasio; *lgc-55:* canal de cloro dependiente de aminas, *ida-1:* receptor tirosina fosfatasa; *flp-19:* péptido FMRF; *unc-17:* transportador vesicular de acetilcolina; *unc-40:* receptor de netrina; *nlg-1:* neuroligina; *rab-3:* ras GTPasa; *kal-1:* síndrome de Kallmann. Aquellos defectos de expresión en animales mutantes que son estadísticamente significativos con respecto a los animales silvestres, se indican con un cuadrado negro. n.a.: no analizado. La significación estadística fue calculada con el test exacto de Fisher, pV< 0,05. n > 50 animales por condición.



2.4 El código regulador de HSN es necesario durante toda la vida del animal para el mantenimiento de la identidad celular de la neurona HSN

También exploramos si el código regulador de HSN es prescindible una vez establecida la identidad serotonérgica de la neurona o si, en cambio, es necesario durante toda la vida del animal para mantener el estado correcto de diferenciación de la neurona HSN (Deneris & Wyler 2012). Para ello llevamos a cabo experimentos de silenciamiento de la expresión génica mediante ARN de interferencia, dónde alimentamos a los gusanos con clones complementarios a cinco de nuestros genes candidatos. Estos clones se administraron en la edad adulta, una vez los genes de la vía de la 5-HT se han expresado en la neurona HSN y ésta es funcional. Estos experimentos no se realizaron para *hlh-3,* puesto que no parece ser expresado en la neurona HSN en la edad adulta. Nuestros resultados indican que el silenciamiento de los cinco FT en la edad adulta provoca una pérdida de expresión de *tph-1::yfp,* uno de los genes de la vía de la 5-HT y principal marcador de las neuronas serotonérgicas y , por tanto, que el código regulador de HSN no es sólo necesario para establecer el fenotipo serotonérgico en la neurona, sino también para mantenerlo → **Figura 6.**

3.1 La huella de identidad reguladora de la neurona HSN permite la identificación *de novo* de genes expresados en la neurona HSN

Existe un gran desconocimiento sobre por qué ciertas regiones del ADN tienen actividad reguladora y otras no. En la actualidad, no es posible predecir qué potenciadores se encuentran unidos a FT y si son activos o no. En este trabajo, debido a que los miembros del código regulador de HSN pertenecen a seis familias de FT diferentes que reco-

Figura 4
Análisis de mutagénesis en el MRC para *tph-1*

Los sitios de unión para las diferentes familias de factores de transcripción se han representado con diferentes cajas de colores. Las cruces negras indican mutaciones puntuales para truncar el sitio de unión correspondiente. +: valores de expresión 100-60% de la media de los valores silvestres; +/−: valores de expresión más bajos que el 60-20% de la media de los valore silvestres; −: valores de expresión más bajos que el 20% de la media de los valores silvestres. n>30 animales por línea.

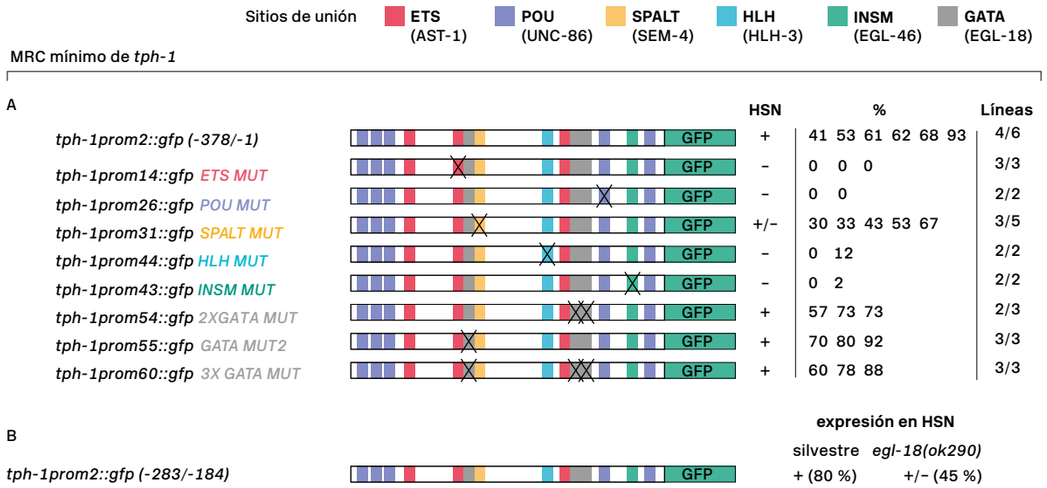


Figura 5
Análisis de regulación cruzada

El esquema resume las relaciones entre los seis miembros del código regulador de la neurona HSN, determinadas por el análisis de expresión de reporteros para los seis miembros en los diferentes fondos mutantes. Las flechas negras indican defectos fuertes de expresión, las flechas grises indican defectos parciales de expresión, mientras que las flechas grises discontinuas indican defectos débiles. Las flechas negras discontinuas indican que el efecto que ejerce UNC-86 sobre AST-1 puede ser directo o mediado por SEM-4. La significación estadística fue calculada con el test exacto de Fisher, pV< 0,05. n > 50 animales por condición.

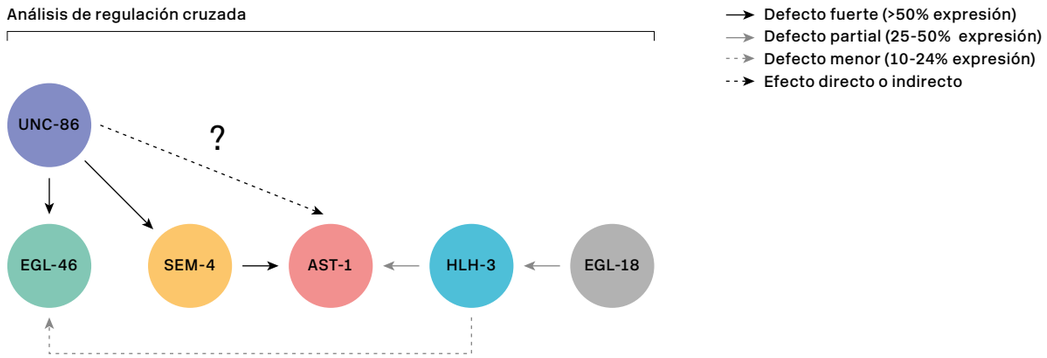
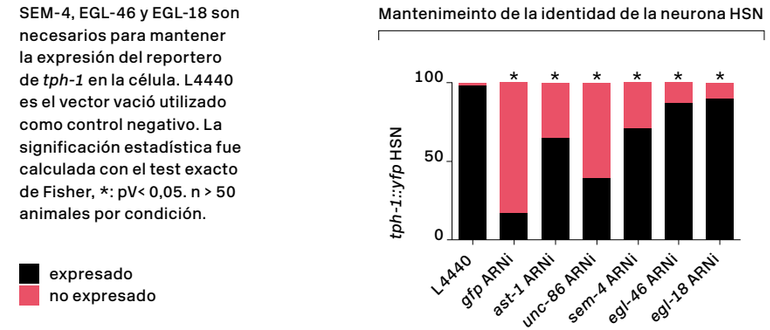


Figura 6
Análisis del mantenimiento de la identidad serotonérgica en la neurona HSN

Los experimentos de silenciamiento génico mediante ARN de interferencia sobre la neurona HSN madura indican que AST-1, UNC-86, SEM-4, EGL-46 y EGL-18 son necesarios para mantener la expresión del reportero de *tph-1* en la célula. L4440 es el vector vació utilizado como control negativo. La significación estadística fue calculada con el test exacto de Fisher, *: pV< 0,05. n > 50 animales por condición.



nocen secuencias de unión al ADN muy diferentes entre sí, nos planteamos si la agrupación de estos seis sitios de unión en los genes expresados en la neurona HSN podría conferir suficiente especificidad como para imponer una huella única de la neurona HSN.

Llevamos a cabo análisis bioinformáticos utilizando R (The R Team 2016) y Bioconductor (Huber et al. 2015) para analizar las secuencias reguladoras (aguas arriba e intrónicas) de los 96 genes que se expresan en la neurona HSN (Hobert et al. 2016) en busca de ventanas de ADN de hasta 700 pares de bases que contuviesen, por lo menos, un sitio de unión para cada FT del código regulador de HSN. A esto lo llamamos huella de identidad reguladora de HSN. Encontramos que los genes que se expresan en la HSN contienen esta huella de identidad reguladora en mayor medida que un conjunto de genes control (ratio 1,3), y que la ratio aumenta si sólo consideramos genes conservados en diferentes especies de *Caenorhabditis* (ratio 1,8) → **Figura 7-A,** lo que apoyaría la idea de que esta huella se haya seleccionado en la evolución para definir los potenciadores de este tipo neuronal concreto. De ser así, pensamos que esta huella se debería corresponder con potenciadores funcionales de la neurona HSN y lo comprobamos fusionando 5 de estas ventanas con *gfp* y corroborando su expresión en HSN en 4/5 líneas.

A continuación, examinamos la distribución de la huella de identidad reguladora de HSN en todo el genoma de *C. elegans* y la encontramos preferentemente en genes que se expresan en neuronas o tienen alguna función neuronal (genoma neuronal) → **Figura 7-B.** Por tanto, decidimos utilizarla para identificar genes nuevos que se expresen en la neurona HSN. Escogimos al azar 35 de estos genes neuronales que contenían la huella de identidad reguladora de HSN conservada en varias especies del nematodo y generamos reporteros transgénicos. 13 de estos genes se expresan en la HSN, mientras que ninguno de los controles negativos mostró GFP en la neurona. Nuestros resultados demuestran, por primera vez, que la la presencia de una secuencia primaria de ADN (la huella de identidad reguladora de HSN) puede ser utilizada para identificar genes *de novo* → **Figura 7-C.** Sin embargo, el alto número de falsos positivos indica que la huella, por si misma, no es suficiente en todos los casos para inducir expresión en la neurona HSN, y que nos falta información relevante sobre el mecanismo de activación del transcriptoma de HSN.

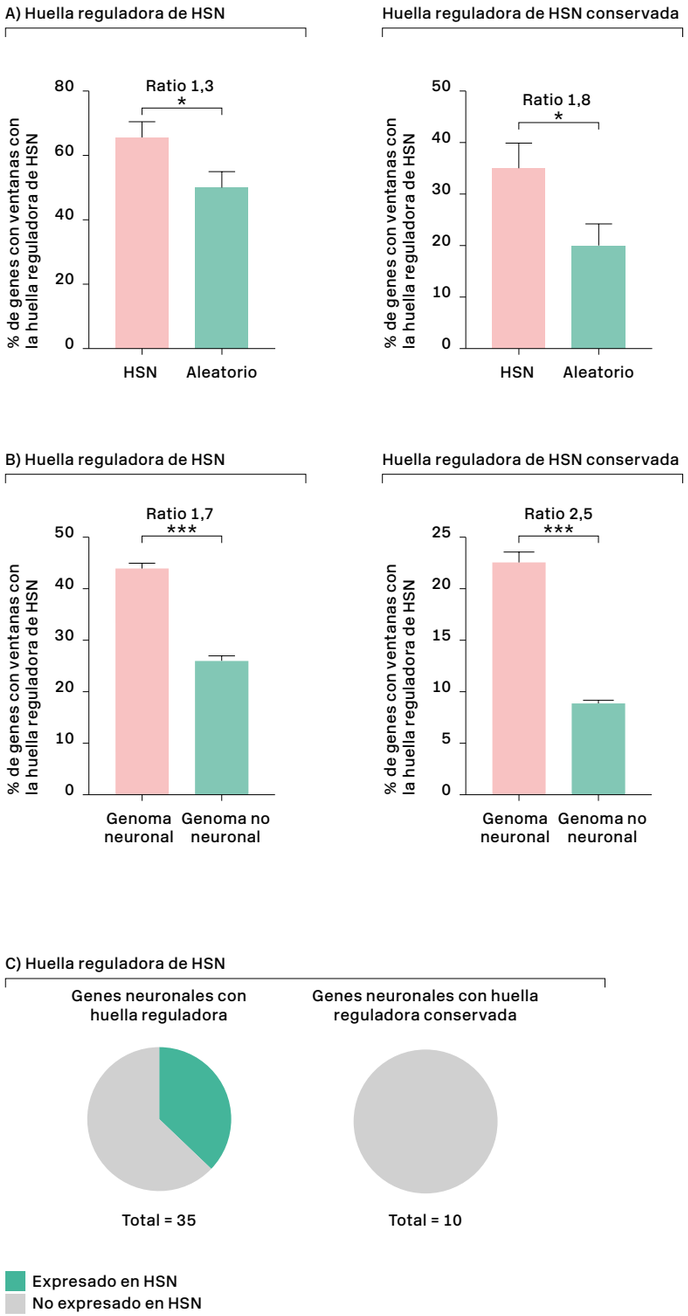


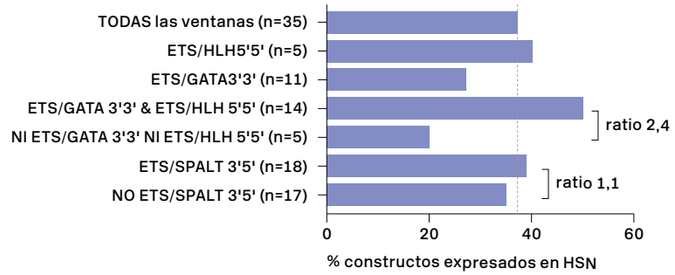
Figura 7
Análisis de la huella reguladora de la HSN y su capacidad predictiva para identificar potenciadores funcionales de la HSN

A) Un 66% de los genes expresados en la neurona HSN contienen la huella reguladora de HSN, en comparación con un 50% en los genes control (100 genes con un tamaño similar a los presentes en la HSN, elegidos al azar). La inclusión de los criterios de conservación de la huella reguladora de HSN aumenta las diferencias entre los genes expresados en la HSN y los genes control. *: pV<0,05. Test exacto de Fisher.

B) El análisis de la distribución de la huella reguladora de la HSN en todo el genoma de *C. elegans* indica la existencia de un enriquecimiento de ésta en genes neuronales o con una función neuronal (genoma neuronal) en comparación con genes sin función neuronal asignada. La inclusión de los criterios de conservación de la huella reguladora de HSN aumenta las diferencias entre el genoma neuronal y no-neuronal. Chi cuadrado con corrección de Yates. ***: pV<0,0001.

C) La presencia de la huella reguladora de HSN permite identificar potenciadores asociados a genes neuronales expresados en la neurona HSN en un 37% de los casos (13/35 ventanas analizadas), mientras que ninguno de los controles (ventanas correspondientes a genes neuronales y de longitud comparable, pero sin huella reguladora de HSN) se expresa en la neurona.

D) Ventanas con la huella reguladora de HSN muestran una preferencia de orientación estadísticamente significativa entre sitios de unión de la familia ETS-GATA (3'3') y ETS-HLH (5'5'). Aquellas ventanas que presentan ambas configuraciones gramaticales presentan mayor probabilidad de expresión en la HSN (50% en comparación con 20% en ventanas sin ninguna de las dos sintaxis).



3.2 Los potenciadores de la HSN siguen el modelo del 'colectivo de FT' y contienen reglas sintácticas

La arquitectura reguladora hace referencia a la multiplicidad, identidad, afinidad y posición de los sitios de unión de los FT presentes en una secuencia reguladora o potenciador. Arquitecturas específicas pueden contener restricciones en relación a estas variables, lo que recibe el nombre de 'reglas gramaticales o sintácticas'. Se han propuesto tres modelos para explicar la función de los potenciadores basados en estas normas gramaticales (Spitz & Furlong 2012). En el primer modelo (*enhanceosome*), los sitios de unión a FT muestran una distribución muy rígida en cuanto a orden, espaciado y orientación, lo que implica elevada cooperatividad entre los FT para unirse al ADN. En cambio, el segundo modelo (*billboard* o cartelera) propone una distribución y orientación mucho más laxa de los sitios de unión de los FT, donde no es necesario que todos los FT se unan al ADN ni que haya cooperatividad de unión al ADN. Finalmente, el tercer modelo (colectivo de FT) representa una situación intermedia entre los dos anteriores. Como el segundo, muestra flexibilidad en el espaciado y orden de los sitios de unión e incluso en su composición; es decir, aunque se requiere la presencia de todo el colectivo de FT, no todos se unirán al ADN necesariamente, sino que algunos podrán ejercer su función mediante la interacción con otros FT del grupo. Además, este modelo predice la existencia de reglas gramaticales que podrán ser requeridas o no dependiendo del contexto de cada potenciador.

Los datos obtenidos del análisis de los MRC sostienen que el código regulador de HSN podría seguir el segundo modelo. Sin embargo, decidimos investigar la posibilidad de que la huella reguladora de la HSN contuviese reglas gramaticales que ayudasen a explicar la funcionalidad de los potenciadores de la neurona HSN. Descubrimos que cuando dos parejas de FT (ETS-GATA y ETS-HLH) se encontraban en una orientación particular en una secuencia de ADN con la huella reguladora, la probabilidad de que se expresara en la HSN era mayor → **Figura 7-D,** confirmando la existencia de normas sintácticas y, por tanto, concluyendo que los potenciadores de la HSN siguen el modelo del colectivo de FT. A partir de entonces, decidimos llamar más correctamente a los seis FT que regulan a la neurona HSN 'colectivo de FT de la HSN'.

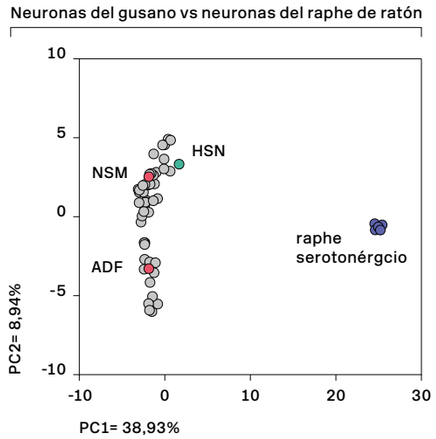


Figura 8
Análisis de homología molecular entre la neurona HSN y las neuronas serotonérgicas del raphe de ratón

Comparación entre los perfiles de expresión de las neuronas del gusanos (construidos mediante la asignación de ortólogos de ratón a los genes de *C. elegans*) y los perfiles de expresión de las neuronas serotonérgicas del raphe de ratón (construidos a partir de datos de RNAseq), utilizando un análisis de coordinadas principales. El perfil molecular de la neurona HSN (verde) es el más cercano a las neuronas del raphe de ratón (azul). Este resultado no es debido a su naturaleza serotonérgica, puesto que los perfiles de NSM y ADF (rojo), también serotonérgicas del gusano, están más alejadas.

4.1 El programa serotonérgico de la neurona HSN y el de las neuronas del raphe de ratón muestran una elevada homología

El programa de diferenciación de las neuronas serotonérgicas de ratón ha sido ampliamente estudiado, por lo que se sabe que está controlado por varios FT, entre los cuales se encuentran ortólogos a los del nematodo: HLH-3/ASCL1, EGL-18/GATA2/3, EGL-46/INSM1 y AST-1/PET1. Esta elevada homología entre los FT del ratón y el colectivo de FT de la HSN nos hizo plantearnos si sería posible que el resto de factores que intervienen en *C. elegans* también pudieran tener un papel en el ratón, y viceversa. Ensayos de silenciamiento génico y mutagénesis dirigida en gusano indican que PHA-4 (FT de la familia FKH y ortólogo del factor de ratón FOXA2) podría tener un papel en la diferenciación de la neurona HSN. Además, exploramos si los FT murinos BRN2 (ortólogo de UNC-86) y SALL2 (ortólogo de SEM-4) podrían tener un rol en la diferenciación serotonérgica. Descubrimos que ambos factores se expresan en las neuronas serotonérgicas recién nacidas del raphe en ratón. Asimismo, demostramos mediante experimentos de rescate en gusanos mutantes que los ortólogos de ratón son capaces de sustituir funcionalmente a los FT equivalentes en gusano.

4.2 La neurona HSN y las neuronas serotonérgicas del raphe de ratón son molecularmente similares y comparten homología profunda

En biología evolutiva el término 'homología profunda' hace referencia a dos estructuras que comparten los mecanismos genéticos que regulan su diferenciación (Shubin et al. 1997). La gran homología que comparten los programas genéticos de diferenciación serotonérgica en *C. elegans* y en ratón sugiere que estos dos tipos neuronales podrían compartir una homología profunda y, por tanto, serían estructuras homólogas. Para que esto fuese cierto, la neurona HSN y las neuronas del raphe de ratón no sólo deberían compartir la expresión de los genes de la vía de la 5-HT, que además también los expresan las neuronas serotonérgicas del gusano NSM y ADF, sino que deberían ser ampliamente parecidas en términos moleculares. Para abordar esta cuestión, utilizamos la información disponible en www.wormbase.org y en (Hobert 2016a) sobre el perfil de expresión génica en HSN, para compararla con los datos disponibles del transcriptoma de las neuronas serotonérgicas del raphe de ratón (Okaty et al. 2015). Realizamos un análisis de coordinadas principales que sugiere que, de entre todas las neuronas del gusano, el transcriptoma de la neurona HSN es el que más se asemeja a aquel de las neuronas serotonérgicas de ratón, apoyando la teoría de que comparten homología profunda → **Figura 8.**

Conclusiones

En esta Tesis hemos diseccionado la lógica de regulación transcripcional que subyace la especificación del sistema serotonérgico de *C. elegans* y, centrándonos en el subtipo neuronal HSN, hemos estudiado los mecanismos que seleccionan el complemento de potenciadores específicos de un tipo celular.
Los resultados obtenidos en esta Tesis han conducido a las siguientes conclusiones:

1 Módulos de regulación en *cis* diferentes controlan la expresión de los genes de la vía de la serotonina en los distintos tipos de neuronas serotonérgicas de *C. elegans*. Esta regulación modular e independiente está de acuerdo con el modelo de los selectores terminales donde, para cada subtipo neuronal, una combinación de factores de transcripción diferente activa la expresión de sus características terminales.

2 Un complejo código de factores de transcripción es necesario para inducir el fenotipo serotonérgico en la neurona HSN. Este código, al que hemos denominado colectivo de factores de transcripción de HSN, está compuesto por UNC-86, SEM-4, HLH-3, EGL-46, AST-1 y EGL-18, que pertenecen a las familias POU, SPALT, HLH, INSM, ETS y GATA, respectivamente. Estos seis factores de transcripción se unen directamente a las regiones reguladoras de los genes de la vía de la serotonina con la finalidad de activar su expresión en la neurona HSN.

3 El colectivo de factores de transcripción de HSN se expresa en la neurona a estadio larvario L4 y actúa específicamente en el paso de diferenciación terminal de la célula, a excepción de HLH-3, cuya expresión se limita al neuroblasto de HSN. Proponemos que, al igual que su homólogo en ratón ASCL1, HLH-3 muestra un rol dual como factor proneural y pionero, promocionando secuencialmente la especificación neuronal y la diferenciación terminal serotonérgica de la célula precursora de HSN, siendo capaz de unirse a los genes de la vía de la serotonina aún en estados de cromatina cerrada. El colectivo de factores de transcripción de HSN, a excepción de HLH-3, es necesario que se exprese durante toda la vida del animal para mantener la identidad serotonérgica de la neurona HSN.

4 El colectivo de factores de transcripción de HSN activa la expresión de los genes de la vía de la 5-HT a través de rutas paralelas en la neurona HSN.

5 El colectivo de factores de transcripción de HSN regula la expresión de los genes de la vía de la serotonina de manera cooperativa y redundante. El rol individual de cada miembro del colectivo y las relaciones sinergísticas entre ellos depende del gen en particular y del contexto donde se unen.

6 El colectivo de factores de transcripción de HSN contiene una huella de identidad reguladora compuesta por la agrupación de sitios de unión de los seis factores de transcripción del colectivo que está enriquecida en los genes neuronales y permi-

te la identificación *de novo* de genes expresados en la neurona HSN. En este trabajo demostramos, por primera vez, que una huella de identidad reguladora meramente basada en secuencia primaria de ADN es suficiente para la identificación de potenciadores.

7 El programa de regulación de la neurona HSN encaja con el modelo del colectivo de factores de transcripción: los seis factores de transcripción actúan de manera flexible, consiguiendo la activación de potenciadores con un orden y una distribución variable de sitios de unión. Además, de acuerdo con este modelo, la presencia de sitios de unión para ciertos factores de transcripción puede ser dispensable en determinados contextos genómicos al ser compensada por el resto de miembros del colectivo.

8 El programa de regulación serotonérgico de la neurona HSN, pero no el de las otras neuronas serotonérgicas NSM y ADF, se asemeja al de las neuronas serotonérgicas del raphe del ratón: AST-1/PET1, HLH-3/ASCL1, EGL-18/GATA2/3, EGL-46/INSM1 y UNC-86/BRN2 aparecen como factores de transcripción ortólogos. Esta homología permite la identificación de nuevos candidatos para la neurona HSN del gusano PHA-4 (FOXA2), and para el ratón SALL2 (SEM-4).

9 La neurona HSN de *C. elegans* y las neuronas serotonérgicas del raphe de ratón comparten homología profunda, además de homología funcional. Proponemos que esta homología profunda podría haber aparecido a partir de un tipo celular ancestral común.