UNIVERSITY OF CALIFORNIA SAN DIEGO
SAN DIEGO STATE UNIVERSITY

Regulatory Interactions Governing Peripheral Nervous System Specification in the Ascidian *Ciona robusta*

A dissertation submitted in partial satisfaction of the
Requirements for the degree Doctor of Philosophy

in
Biology
by
Clifford Dennis Pickett

Committee in charge:

University of California San Diego

Professor Kim Cooper
Professor James Posakony

San Diego State University

Professor Robert Zeller, Chair
Professor Nick Shikuma
Professor Ricardo Zayas

2020

The Dissertation of Clifford Dennis Pickett is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

San Diego State University

2020

DEDICATION

I dedicate this dissertation to my family. Most especially to my mother, whose steadfast encouragement and never-ending support of all manner, since the beginning, has been instrumental toward this achievement. To my wife Danielle, by my side in everything always. And to Patrick- I may not have survived writing this during stay-at-home orders without you.

EPIGRAPH

Nothing in this world can take the place of persistence. Talent will not; nothing is more common than unsuccessful people with talent. Genius will not; unrewarded genius is almost a proverb. Education will not; the world is full of educated derelicts. Persistence and determination alone are omnipotent.

Calvin Coolidge

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

Supplementary Table 2.1: Full list of associated GO terms associated with CrPou4 and MmPou4F2 up-regulated genes.

Supplementary Table 2.2: DESEQ analysis of genes up- and down-regulated in response to any of the ectopically-expressed *Pou4* genes.

LIST OF ABBREVIATIONS

bHLH	basic helix-loop-helix
BMP	bone morphogenic protein
CFP	cyan fluorescent protein
CNS	central nervous system
DNA	deoxyribonucleic acid
EPI	epidermal promoter
ESN	epidermal sensory neuron
FGF	fibroblast growth factor
GFP	green fluorescent protein
GO	Gene Ontology
GRN	gene regulatory network
hpf	hours post fertilization
miRNA	microRNA
mRNA	messenger RNA
nt	nucleotide
PCR	polymerase chain reaction
PNS	peripheral nervous system
RNA	ribonucleic acid
RNA-SEQ	RNA-sequencing
TF	transcription factor
ISH	*in situ* hybridization
YFP	yellow fluorescent protein

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to first thank my PI Dr. Robert Zeller. Over the last 7 years you have provided me with the most important tools for the success I was looking for- space and freedom. Never once did you formally direct what I have done, while at the same time you consistently, at a moment's notice, offered generous advice and excellent ideas. It has been an honor to serve with you three times at the Marine Biological Laboratory's Embryology Course, and I also thank you for your help facilitating my participation in the course in 2016.

I would also like to thank my committee, Dr. Ricardo Zayas, Dr. Nick Shikuma, Dr. Kim Cooper, and Dr. James Posakony. I sincerely appreciate your time as well as the practical nature of your guidance; the ideas and conclusions we reached as a group at our meetings went far towards making this work as solid as it could be.

I would be remiss not to acknowledge the many hours spent by the near dozen undergraduates I have had the pleasure of mentoring. Sonya Choe, Lexy Cartwright, Samantha Silva, Dylan Castillo, Sharon Onggo, Martin Mueller, Ema Suarez, Karl Garcia, Tiffany Hoang, Alyna Lam and Savannah Ringgenberg, it was a pleasure training you and thank you for your hard work.

Finally I would like to thank Mentor #2, Dr. Thomas Meedel. You helped me develop, as an undergraduate, an excellent set of skills and lessons I used to launch into my graduate career, and then remained a valuable mentor. So much of what I have achieved I imagine unlikely without your help over the years.

The Introduction to the Dissertation, in full, is under review for publication. C. J. Pickett and Robert Zeller, 2020. The dissertation author was the primary investigator and author of this manuscript.

Chapter 1, in full, is a reprint of the material as it appears in the journal, *genesis*. C. J. Pickett and Robert Zeller, 2018. The dissertation author was the primary investigator and author of this manuscript.

VITA

Education

Doctor of Philosophy in Biology; *2020*; University of California San Diego and San Diego State University, San Diego, CA

B.S., Biology *with honors*; 2013; Rhode Island College; Providence, RI
Graduated *May 2013*, *Magna Cum Laude*

Embryology: Concepts & Techniques in Modern Developmental Biology
MBL Advanced Research Training Course completed *Summer 2016*

Publications and Papers

- "Efficient genome editing using CRISPR-Cas-mediated homology directed repair in the ascidian *Ciona robusta*" Pickett, C., Zeller, R. *genesis* 56.11-12 (2018): e23260.

- "The *Ciona* Myogenic Regulatory Factor Functions as a Typical MRF but Possesses a Novel N-terminus that is Essential for Activity" Lindsay E. Ratcliffe, Emmanuel K. Aseidu, C. J. Pickett, Megan A. Warburton, Stephanie A. Izzi, and Thomas H. Meedel *Developmental Biology* 448.2 (2019): 210-225.

- "A Cross-Scale Examination of Pou4 Genes During Development: insights from a phylogenetically-comprehensive analysis" (advanced review article) Pickett, C. J., Zeller, R. (in review; submitted April 2020)

- "Mouse Pou4 Proteins Regulate Dissimilar Sets of Genes in the Ascidian *Ciona robusta* and Paradoxically Can Substitute for the Endogenous Ascidian Pou4" Pickett, C., Hurless, V., Hoang, T., Zeller, R. (in prep)

- Regulatory Interactions Governing the Specification of Epidermal Sensory Neurons in the Ascidian *Ciona robusta* " Pickett, C., Zeller, R. (in prep)

Honors and Awards

Achievement Rewards for College Scientists Foundation Award *2018, 2019*
Elliott Family Fund Scholarship *2016*
Kurt and Rhoda Isselbacher Endowed Scholarship *2016*
Burroughs Wellcome Fund Scholarship *2016*
RI-INBRE Research Fellowship *2013*
W. Christina Carlson Award for Excellence in Biology *2013*
Mary M. Keefe Departmental Award for Excellence in Biology *2012*

Honors and Awards (*Continued*)

Rhode Island College Alumni Scholarship *2012, 2011*

Summer Undergraduate Research Fellow (SURF) *2012, 2011*

Professional Presentations

9th International Tunicate Meeting, New York University
Advancing CRISPR-Cas Technologies for Effective Gene Perturbations in the Ascidian Ciona robusta Pickett, C., Zeller, R. *July 2017* (Poster)

Graduate Student Symposium, San Diego State University
Developing CRISPR-Cas Technologies for Effective Gene Knockdown in the Marine Invertebrate Chordate Ciona intestinalis Pickett, C., Zeller, R.
April 2016 (poster); *April 2015* (poster); *April 2014* (poster)

Student Research Symposium, San Diego State University
Developing CRISPR-Cas Technologies for Effective Gene Knockdown in the Ascidian Ciona intestinalis
Pickett, C., Zeller, R. *March 2015* (poster); *March 2014* (talk)

Summer Undergraduate Research Fellow Conference, University of Rhode Island
The CiMRF N-terminus: A Functionally Important Evolutionary Novelty?
Pickett, C., Meedel, T. *August 2013* (poster)

Eastern Colleges Science Conference, Providence College
Stimulating Myogenesis in the Embryonic Endoderm of Ciona intestinalis
Pickett, C., Meedel, T. *April 2013* (talk)

Northeast Undergraduate R&D Symposium, University of New England
Stimulating Myogenesis in the Embryonic Endoderm of Ciona intestinalis
Pickett, C., Meedel, T. *February 2013* (talk)

Summer Undergraduate Research Fellow Conference, University of Rhode Island
Functional Evolution of the Myogenic Regulatory Factor Family
Pickett, C., O'Connor, K., Meedel, T. *August 2012* (poster)

Relevant Educational Activities

Student Mentoring; Zeller Lab, personal research mentor to undergraduates Sonya Choe, Lexy Cartwright, Samantha Silva, Dylan Castillo, Sharon Onggo, Martin Mueller, Ema Suarez, Karl Garcia, Tiffany Hoang, Alyna Lam and Savannah Ringgenberg *2014-2019*

Educational Volunteering

San Diego County Science and Engineering Fair, Judge *2014-2019*

Julian Charter School San Diego Academy Science Fair, Judge *2016-2019*

San Diego Ocean Discovery Institute: *Snorkel with a Scientist*, Volunteer *2014, 2019*

SDSU Undergraduate Research Symposium, Judge *2016;* Facilitator *2014*

Educational Volunteering (*Continued*)

San Diego Ocean Discovery Institute: *Research Day*, mentored literature review *2014*

The Fleet Science Center: *Science on the Rocks*, volunteer *2014: An Artist and a Scientist Walk into a Bar...*, volunteer *2017; Two Scientists Walk into a Bar...*, volunteer *2014 – present*

Coastal and Marine Institute Laboratory: Open House, Zeller Lab Display Lead *2014-2019*

High Tech Fair, volunteer with MEBSA *2014*

Rhode Island State Science and Engineering Fair, Judge *2013*

Rhode Island Science Olympiad, Volunteer *2011*

Aptos Junior High School, Aptos CA, *2009* (100+ hours) volunteering and tutoring Physics, Chemistry, and Biology

ABSTRACT OF THE DISSERATION

Regulatory Interactions Governing Peripheral Nervous System Specification in the Ascidian *Ciona robusta*

By

Clifford Dennis Pickett

Doctor of Philosophy in Biology

University of California San Diego 2020

San Diego State University 2020

Professor Robert Zeller, Chair

Ciliated mechanoreceptors, such as our inner ear hair cells, receive vibrational and touch input from an animal's surroundings and direct that information into the their brain via neuronal transmission. In ascidians, the vertebrates' closest extant relatives, ciliated mechanoreceptors called epidermal sensory neurons (ESNs) are found primarily dorsally and ventrally along the larval tails. Ascidian ESNs are ciliated neurons which project their own axon, and thus only analogous to vertebrate inner ear mechanosensory systems.

Initially governed by Nodal signaling, the dominant gene that sets up the territory from which ESNs emerge is *Msxb*. In conjunction with Fgf9/16/20 signaling dorsally and Admp signaling ventrally, *Msxb* is responsible for establishing the future dorsal and ventral larval midlines as neurogenic. Once the field is further established by expression of *Achaete-scute homolog* (among others), Notch signaling locks in the number and spacing of ESNs through lateral inhibition. *Pou4*, a gene necessary to produce an ESN, is also sufficient: *Pou4* has a special capacity for converting all epidermal cells into ESNs (or ESN-like cells). During normal development, *Pou4* is involved in activating other transcription factors such as *NeuroD* and *Atonal*, and, as studied herein, perhaps *MyT1*. *Pou4*, an ancient homeobox gene probably involved in neurogenesis of the metazoan last common ancestor, remains primarily neurogenic in all animals in which it has been studied.

The Introduction of this Dissertation is an extensive review of *Pou4* genes throughout the animal kingdom, examining their original roles, gene copy number history, and the patterns of systems they are deployed in. Chapter 1 is previously published work detailing techniques for using CRISPR-Cas technology to insert DNA sequences of interest into the *C. robusta* genome. Chapter 2 examines subsequent transcriptomes from expressing the three mouse *Pou4* genes in the *C. robusta* larval epidermis, and demonstrates the ability of those mouse *Pou4* genes to

rescue a *Pou4* knockdown loss of cilia phenotype. Chapter 3 concludes with my principal interest throughout my graduate tenure: an examination of genetic relationships that work to specify and differentiate ESNs; I use overexpression and knockdown data to build a working model for a GRN that involves gene relationships between ESNs and adjacent non-ESN midline epidermis.

INTRODUCTION TO THE DISSERTATION

A Cross-Scale Examination of *Pou4* Genes During Development: insights from a phylogenetically-comprehensive analysis

ABSTRACT

Analyzing a gene's expression data across the animal kingdom and reconstructing protein phylogenies can suggest roles and sequences for a protein as it existed in its last common ancestor. For the homeobox gene *Pou4,* in doing so we can propose its role the last common metazoan ancestor, study where utilization has been gained and lost, assume core sequences based on regions of 100% conservation, and even speculate on the sequence of the DNA binding site it bound nearly one billion years ago. This review, a cross-phyla analysis of *Pou4* gene expression and function, examines current knowledge of *Pou4* activity over a wide array of organisms, from sponges to humans. *Pou4* genes are found in all metazoans studied except for ctenophores, placing *Pou4* in the genome of the common metazoan ancestor. In all characterized cases, *Pou4* genes, although they have been coopted for several other uses, act to differentiate various neurons of the central nervous system, and in many cases the peripheral nervous system as well. Furthermore, *Pou4* genes are often critical for the normal differentiation and survival of the cell type in which they are deployed. The metazoan LCA had a *Pou4* gene that was proneural, with a DNA-binding domain that still exists today. Within the chordates, vertebrates appear to have lost the use of *Pou4* genes during olfactory development, while the invertebrate chordates never utilized their *Pou4* gene in photosensory development. With noted exceptions, the invertebrates possess a single *Pou4* gene, while the tetrapods −with the intriguing exception of mammals and marsupials− have four *Pou4* genes; fish and some amphibians have a larger copy number due to more genome duplication events. The four tetrapod *Pou4* genes are products of the two whole genome duplication events that occurred after the invertebrate-vertebrate chordate split, whereupon one of those four *Pou4* genes was summarily lost in the marsupial-mammal lineage.

INTRODUCTION

Multiple copies of similar genes within an organism's genome drew Susumu Ohno to develop the theory in 1970 that the vertebrates had experienced whole genome duplication events in their evolutionary history [1]; all evidence to date has supported this theory [2]. There are scores of genes that exist as ohnologs, paralogs derived from whole genome duplication (WGD) [3]. The best studied of which are arguably *Hox* genes, homeobox transcription factors that dominate early segmentation and axis formation events during development [4, 5]. Another large class of well-studied homeobox proteins are POU genes, which also often exist as ohnologs, and are involved with a sizable array of developmental processes likely due to several interesting physical characteristics, discussed below [6]. POU genes entered the biological lexicon after the mammalian *Pit1*, *Oct2*, and *Oct2* genes were discovered to share a highly conserved bipartite domain with the distantly related *C. elegans Unc-86* gene [7]. Since then, a large number of POU-domain proteins have been described [8]. A functionally diverse group of transcription factors, POU-domain proteins have been detected in every metazoan examined from ctenophores [9] to mammals [10]. In the nearly one billion years since the emergence of the POU genes [11], they have acquired a variety of transcriptional targets in a diverse array of organs and cell types [6]. Yet to our knowledge, all POU proteins have retained their primary function as critical mediators of cellular differentiation.

A sequence-level comparison of three mammalian POU proteins, *Brn3a*, *Brn3b*, and *Brn3c* (now referred to as *Pou4F1*, *Pou4F2*, and *Pou4F3*), to the *Drosophila Acj6* and *C. elegans Unc-86* [12], revealed these POU proteins share a unique property within the POU class: a conserved amino-terminal domain termed the POU-IV box [13] (Figure 1). Although the

function of this domain is not well studied or understood, the shared presence of this domain is one of the central reasons for placing these proteins into a functionally-defined category of their own, the POU IV class [14]. Pou IV class proteins appear to function almost exclusively in the determination and differentiation of neural tissues in the central nervous system (CNS) and peripheral nervous system (PNS) [8, 15]. *Pou4* gene utilization in non-neural structures such heart development [16] and gonadogenesis [17, 18], have been phylogenetically-limited inventions.

This review summarizes our current understanding of *Pou4* properties and begins to untangle the story of *Pou4* evolution. By analyzing *Pou4* gene function in a variety of cell types over a diverse range of phyla, the evolutionary trends of *Pou4* gene activity can be assessed. We can infer that the *Pou4* gene was expressed in the neurons of our early animal ancestors [11], and is now utilized in the differentiation of a wide variety of neuronal cell types in both the central and peripheral nervous systems of animals. In two prominent cases, heart development and gonad development, *Pou4* plays a role in major non-neural systems, while we find only one example of *Pou4* activity during early stages of development, germ layer formation [19]. Here we first review the Pou4 proteins and their original characterizations. Then we examine *Pou4* copy number and various protein features, and show that mammals and marsupials have lost a copy of their *Pou4* genes, yet seem to do the most with them. Lastly, we summarize the gene expression patterns and protein functions of a comprehensive set of invertebrate and vertebrate *Pou4* genes. Together, this information provides context for a fuller understanding the evolution of *Pou4* and its roles during animal development.

POU4 PROTEINS

Discussed in detail below, the common expression patterns of *Pou4* genes throughout the animal kingdom place *Pou4* in ciliated mechanoreceptive neurons, as it is expressed in all available cases from jellyfish to humans in that cell type [20, 21]. Classically acknowledged as a proneural transcription factor, it may come to be that, particularly if poriferan choanocytes express *Pou4*, these genes played a critical role in the evolution of that cell type. Today, *Pou4* genes are expressed in a wide variety of neuronal cells; some of their unique physical properties may be why so many neuronal types have come to utilize them in their development [22].

The POU domain of POU genes is bipartite, subdivided into the POU-Specific domain and the POU-Homeobox domain (Homeodomain) [15, 22]; members of the Pou IV class of proteins share an additional domain: an amino terminal region termed the POU-IV box (Figure 1A). However, the central feature and most analyzed region of POU-IV class proteins is their POU domains, approximately 150 amino acids located at the carboxy-terminal end of the protein. POU domains are categorically split into two sub domains: the POU-Specific domain and the POU-Homeobox domain [23, 24]. These domains are connected by a 13-19 amino acid variable linker region (Figure 1B, green text). This motif (and often the POU-IV box) is typically included in phylogenetic analyses of Pou4 proteins [25, 26], due to the fact that the POU-Specific and POU-Homeodomains are too highly conserved for comparisons producing phylogenetic inferences [27]. It is also clear that after the advent of multiple *Pou4* genes in vertebrates, their linker regions have been highly conserved; for example the human Pou4F1 linker and the chicken Pou4F1 linker are identical (Figure 1B). The sequence conservation of the linker region suggests that paralog-specific functional properties, such as allowing proper flexibility between POU-Specific and POU Homeodomain DNA binding domains [28], or protein-protein interactions[29], are likewise conserved.

Originally described by Gerrero, et al. [13], the POU-IV box is an unique motif of approximately 40 amino acids in the amino-terminal domain. Amino acid alignments of the various POU-IV boxes (Figure 1A) reveal conservation among the chordates with expected lesser conservation when compared to non-chordates. There is less than 45% conservation between the *C. elegans* Unc-86 and the mouse Pou4F1 POU-IV boxes, yet functional analyses suggest those particular amino acids of the *C. elegans Unc-86* constitute the domain [13, 30]. This motif is hypothesized to function in protein-protein interactions and as a transactivation domain: it has been shown that the POU-IV box does not participate in DNA binding, but it is required for normal Pou4 function [13, 31]. This was first determined by exploring the functional properties between two mammalian Pou4 variants. Pou4F1 and Pou4F2 have short isoforms producing truncated proteins which do not contain their respective POU-IV boxes [32]. Several reports performed in cell culture [33-35], or studied in adult tissues [36], reveal distinct properties between the short (s) and long forms (l) of Pou4F1 and Pou4F2. For example the short isoform of Pou4F2 may act to repress the function of Pou4F1 by forming a heterodimer complex and disabling DNA binding of Pou4F1 [37]. Additionally, when comparing the activity of Pou4F1(s) and Pou4F1(l), Pou4F1(s) was unable to confer to cells oncogenic activity while Pou4F1(l) was oncogenic, suggesting the N-terminal POU-IV box is critical for DNA binding and subsequent gene activation [38]. In *Drosophila*, the picture is more complicated; 13 different splice variants of the arthropod *Pou4* gene *Acj6* have been documented, but only 4 of the 13 contain the classic, conserved, 40-amino acid POU-IV box; the other 9 increase the size of their POU-IV boxes with inter-POU-IV box exons, maximally adding up to 28 additional amino acids between the Acj6 POU-IV box conserved regions [39]. It remains to be seen exactly what types

of protein-protein interactions might be involved with the POU-IV box and how this region adds to the flexibility within the class IV POU domain proteins.

POU4 COPY NUMBER AND GENOMIC ORGANIZATION

Mammalian genomes contain three *Pou4* genes, while another *Pou4* gene, *Brn3d* (*XtPou4F1.2*), has been characterized in *Xenopus tropicalis* [27] and in the Japanese killifish medaka (*Pou4F4*; *Oryzias latipes*) [40]. A protein BLAST search of Pou4F1.2 reveals proteins homologous to XtPou4F1.2 also present in numerous other vertebrate groups' proteomic databases, including chondrichthyes (whale shark), another teleost (zebrafish), reptiles (alligator), aves (chicken), and monotremes (platypus) (Figure 2; accession numbers for protein sequences referenced there and in Figure 3 located in Supplementary Table 2). The gene is absent from all therian (placentals and marsupials).

A PhyML phylogenetic tree of representative deuterostome Pou4 proteins suggests that four clades of Pou4 proteins exist in vertebrates (Figure 2). Based on this tree, *Pou4F1* and *PouF1.2* likely shared a common ancestral gene as did *Pou4F2* and *PouF3*. We were not able to find any representative *Pou4F1.2* genes from therians, suggesting that this gene was lost from marsupial and placental mammals. The PhyML tree was combined with a species tree derived from timetree.org [41] and Treerecs [42] was used to reconcile the PhyML tree to infer the number of gene losses and gains needed to explain the evolution of the vertebrate Pou4 genes from an invertebrate ancestral *Pou4* gene (Figure 4). It is hypothesized that vertebrate genomes have undergone two whole genome duplication compared to tunicates (their sister group) [2, 43] suggesting that four copies of *Pou4* should be present in vertebrates, but we were only able to find two *Pou4* genes in both lampreys and hagfish; four *Pou4* genes were identified in

vertebrates (*Pou4F1*, *Pou4F1.2*, *Pou4F2* and *Pou4F3*). As we are limited in the number agnathan genomes we can sample, it is unclear if these genomes we sampled lost two copies of the *Pou4* gene, or if additional *Pou4* gene duplications occurred prior to the evolution of gnathostomes to produce the compliment of four copies of Pou4. Based on the available sequence data we examined, the reconciled tree suggests that additional gene duplications occurred prior the evolution of gnathostomes. A simple amino acid sequence analysis comparing XtPou4F1.2 to XtPou4F1, Pou4F2, and XtPou4F3 reveals a higher percentage of similarity to XtPou4F3, however, various alignment programs such as PhyML [44] and MrBayes [45], group XtPou4F1.2 with XtPou4F1, which likely is a result of shared mutations in the highly conserved POU Specific and POU Homeodomains of the paralogs (Figure 1B), highlighting the importance of careful analysis when determining protein homologues.

In the majority of invertebrates analyzed, there is typically only one *Pou4* gene described. Known exceptions of those models with expression data include the platyhelminth flat worm *Schmidtea mediterranea* which has two *Pou4* genes, and the cnidarian freshwater hydrozoan *Craspedacusta sowerbyi* which has three *Pou4* genes [18].

Available genomic data suggests all vertebrate *Pou4* genes are organized into two exons, and this genomic arrangement is shared among all deuterostomes, with one exception – tunicates. The deuterostome first exon is shorter than the second, ranging from 96 to 288 base pairs (bp). Vertebrate and amphioxus introns range from 233 to 803 bp, while the sea urchin *Pou4* intron is nearly 8,000bp. Second exons range from 894 to 1137 bp in length. In the tunicate *Ciona robusta*, the *Pou4* gene is organized into eight exons with at least two splice forms [26], while in the genus *Ascidia*, the gene is organized into nine exons (unpublished Hess and Zeller).

All protostomes examined have their *Pou4* genes divided into a series of several exons and introns, the largest number of which is *Drosophila* with nine exons [46].

DIFFERENTIAL ACTIVITY FROM AN IDENTICAL BINDING MOTIF

The mammalian Pou4 proteins exhibit nearly complete homology in their POU-Specific domains (98.6%) and in their POU-Homeodomains (100%), and to a lesser extent in their POU-IV boxes (92.7%); unsurprisingly due to this fact Position Weight Matrices (PWM) or direct binding data for each protein has been shown to bind to the same DNA sequence [47-49] (5`-(TG(C/A)ATAATTAATGA-3`; Supplementary Figure 1) when activating, and in some cases repressing, their target genes [48]. Furthermore, this essential binding motif is shared between vertebrates and arthropods and to some degree nematodes [47, 49]. Thus, displaying almost no drift over time, the nucleotide code Pou4 proteins bind today is nearly the same nucleotide code bound by Pou4 proteins in the biletarian LCA.

Because the *Drosophila* Pou4 and mouse Pou4 proteins bind to the same DNA sequence [47, 48] and their DNA binding domains are nearly identical, the specific activities of Pou4F1, Pou4F2 and Pou4F3 are probably due to differences outside of the POU-Specific and POU-Homeodomains, for example the POU-IV box, as it has been shown to not participate in DNA binding yet confer differential activity [34, 50]. Additionally, between the four vertebrate Pou4 proteins, there is little similarity outside of the POU-IV box, Pou-Specific domains and Pou-Homeodomains, and because the homologs function differently, sequences outside those regions are additional candidates for each proteins particular functional properties. Within the vertebrate highly-conserved domains it has been shown however that there is one critical and functional difference between the POU-Homeodomain of Pou4F1 and Pou4F2: at Homeodomain position

22, Pou4F1 includes a valine, whereas Pou4F2 includes an isoleucine. If the isoleucine of Pou4F2 is swapped for a valine, the activities of the Pou4F2, which normally exhibits a repressive function to the Pou4F1 promoter, switches to being an activator [29]. As the binding sequences are the same for Pou4F1 and Pou4F2, this indicated that this section of the POU-Homeodomain is involved in protein-protein interactions such as dimerization [51], perhaps recruiting different partners and thus directing variable activity.

GENUS-SPECIFIC EXPRESSION AND FUNCTIONAL DATA

The current availability of *Pou4* data from various metazoan species ranges from the identification of the gene in genome or transcriptome reads, to well-characterized functional studies. Below we summarize the spatial and temporal expression data for *Pou4* beginning with our most distantly related metazoan relatives, and finishing with the well-studied vertebrates.

The Non-Bilaterians

Porifera: In an attempt to perform an inclusive phylogenetic analysis of all the *Pou4* genes, indeed all of the known genes from the entire POU-class, degenerate primer pairs were designed against POU genes and PCR amplifications were performed on a wide array of organisms that do not yet have available transcriptomes, such as the demosponges *Acarnus erithacus*, *Tethya aurantia*, and *Spongilla sp*. [11]. Although only short regions containing the highly conserved POU-Specific domains, linker regions, and POU-Homeodomains were obtained for analysis, good evidence for the existence of *Pou4* genes was obtained for each of the poriferan species. No other *Pou4* data is available for sponges. Nevertheless, as sponges emerged near the origin of

metazoans [52], this data places the *Pou4* gene in the last common ancestor (LCA) of all metazoan animals.

Placozoa: Sequence data was acquired through searching publicly available *Trichoplax adherans* genomes/transcriptomes [11]. Although the *Pou4* gene is present, no other data is available.

Cnidaria: *Pou4* is involved with neurogenesis in the anthozoan sea anemone *Nematostella vectensis* [53]. A recent study in *N. vectensis* provides an in depth analysis of the temporal and spatial expression as well as function of the *Nematostella Pou4* gene, *NvPou4*; so far unique to this organism, it is expressed early in development, prior to gastrulation. To date, *N. vectensis* is the only organism known to express the proneural transcription factor in cells other than post-mitotic, undifferentiated neuronal cells. Experiments performed on EdU-treated animals revealed that later in development *NvPou4* expression is restricted to non-proliferating pre-neural cells undergoing differentiation, cells that differentiate into neurons and cnidarian-specific cnidocytes [19]. This, combined with results showing that a *NvPou4* knockout line fails to differentiate cnidocysts, formally places *NvPou4* in the class of terminal selector genes within cnidarians, as is the case with mammals and *C. elegans* [54], and *Ciona* [55], where terminal differentiation of a neural cell population depends on *Pou4* for their proper development. Because *Pou4* acts as a terminal selector in a wide range of phyla, it may be revealed that the neuronal differentiation programs of other animals utilize *Pou4* as a terminal selector; thus far only the four cases mentioned above have cemented *Pou4* as a terminal selector in their respective scenarios.

The hydrozoan *Craspedacusta sowerbyi* has tripled its number of *Pou4* genes [18]; *Pou4* copy number in other cnidarians is either one, or unknown. The three *C. sowerbyi Pou4* genes were regrettably named *csPou4f1*, *csPou4f2*, and *csPou4f3* due to their percentages of sequence similarity to the corresponding mammalian *Pou4* genes. However, to have named them after vertebrate genes is misleading and potentially confusing. These three *Pou4* genes have distinct and/or overlapping expression patterns in adult tissues. *In situ* hybridization data shows that *csPou4f1* is expressed in the statocyst (a gravity detection organ). *csPou4f2* is expressed in the tissue of the statocyst and the gonad. *csPou4f3* was described as being expressed in the center of the bell quadrants, with gastric cavity expression reported as likely background staining [56]. It would be interesting to know whether any of these *Pou4* genes are expressed in the innervating neurons or the ciliated sensory cells.

Pou4 gene expression has also been examined during medusa development in the scyphozoan moon jelly *Aurelia sp. 1* [20]. Termed *AurBrn3*, this *Pou4* gene is expressed in the developing neuroectoderm of the rhopalia, sensory structures in scyphozoans and cubozoans involved in behavior coordination and perhaps regeneration [57]. *AurBrn3* is expressed specifically in ciliated sensory cells which are thought to be mechanosensory [58].

Sequence data is available for another cnidarian *Hydra vulgaris Pou4 gene*, *Pou4F2*, so named as it shares more sequence identity with the vertebrate *Pou4F2* than *Pou4F1*. Expression or functional data for the hydra *Pou4* is currently unknown [59].

The Bilaterians

Acoelomorpha: A *Pou4* gene has been described in a member of the order Acoela, the flatworm *Neochildia fusca* [60]. During embryogenesis, *Pou4* expression was detected by PCR

amplification from extracted mRNA, while *in situ* hybridization showed *Pou4* expression in neurons of the adult brain. This is especially meaningful as, due to the fact that Xenacoelomorpha, which contains Acoela, is regarded to be a sister group to the Nephrozoa (deuterostomes and protostomes) [61], it suggests *Pou4* functioned as a proneural gene in the bilaterian LCA.

Spiralians

Rotifera: The phylogenetic placement of rotifers is debatable, but they are typically characterized as Spiralians due in part to an articulated jaw structure. Most recently they have been placed in gnathifera, a sister group to living Spiralians [62]. Several rotiferan genomes have been sequenced and unsurprisingly they exhibit a *Pou4* gene [63], yet surprisingly -as Rotiferans have experienced a lineage specific WGD [64]- they appear to only have a single *Pou4* gene (GenBank BIOproject ID: PRJNA503132; UniProt ID: BpHYR1_004824). Their *Pou4* gene was electronically annotated from genome sequencing; no expression data is yet available.

Platyhelminthes: *Schmidtea mediterranea*, the freshwater flatworm, known for its remarkable capacity for full-body regeneration, has two copies of its *Pou4* gene, *Pou4-1* (*Pou4 like 1*; *Pou4l-1*) and *Pou4-2* (*Pou4f-2*). *Pou4-l1* is expressed in the cephalic ganglia, particularly the ChAT^{+} neurons of the brain, and found to be required for proper regeneration of the CNS: RNAi of *Pou4-l1* leads to a reduction in *synapsin* levels, *ChAT* levels, and *neuropeptide-like*$^{+}$ cells, important for generating a properly functioning CNS. *Pou4-2* is also expressed in the mechanosensory neurons of the dorsal and ventral ciliated stripes, as well as mechanosensory neurons of the periphery [65]. Reports show that, like vertebrates and *Ciona* [66, 67], *Pou4-2*$^{+}$

cells are modulated by Notch-Delta signaling, and that *Pou4-2* is required for the differentiation and maintenance of *Pou4*+ ciliated mechanosensory neurons [68].

In the parasitic tapeworm *Hymenolepis microstoma*, *Pou4* expression is detected in the anterior half of the embryo of which the developing juvenile emerges. As development proceeds, *Pou4* is expressed strongly in the region of the developing scolex, a sensory neuron-rich hook-like attachment structure [69].

Molluscs: During the embryogenesis of the pygmy squid *Idiosepius notoides*, *Ino-Pou4* expression is detected in the developing optic and palliovisceral ganglia beginning at Stage 19 (onset of mouth and eye invagination [70]). As development proceeds, *Ino-Pou4* expression extends into the pedal and interbranchial ganglia, while expression fades in the interbranchial ganglia as development reaches hatching (Stage 30) [71].

The abalone *Haliotis asinina* exhibits a similar pattern of *Ha-Pou4* expression during its development. Beginning after gastrulation at 9hpf (hatched trochophore larva), *Ha-Pou4* is detected in three cells; one of which migrates anteriorly to the developing mantle while the other pair contributes to the developing foot field, a region rich in chemo- and mechanosensory neurons. After torsion (where the foot rotates 180° relative to the mantle) at 21hpf, *Ha-Pou4* expression is maintained in the developing foot, and has expanded into the cephalic tentacles and the region of the developing eyes. At 26hpf, *Ha-Pou4* expression is detected in the developing statocyst, an organ analogous to the vertebrate balance-detecting organ, the vestibular system. *Ha-Pou4* is also expressed in the ctenidial and osphradial rudiment, regions that in the adult function primarily in chemosensation [72].

Annelida: In the marine annelid worn, *Platynereis dumerilii*, *Pou4* is involved with the formation of the noncephalic photoreceptors [73]. Although not coexpressed in the same cell as the *P. dumerilii* opsin, their tight proximity suggested that the cells are linked by lineage and that *Pou4* was likely expressed in the parental cell that gave rise to the *Opsin*+ and *Pou4*+ daughter cells. *Pou4* expression was also detected in the ventral nerve cord and parapodial ganglia.

Ecdysozoa

Nematodes: *C. elegans*, like most invertebrates, possesses only a single *Pou4* gene, *Unc-86*. *Unc-86* was quickly identified as sharing significant levels of homology with the mouse *Pit-1* and *Oct-2*; later it was shown an even closer relationship with the mouse *Brn3* factors, sharing homology between the POU-Specific and POU-Homeodomains [7]. Although little homology was revealed in the POU-IV homology box, this N-terminal region was found to be important in regulating protein-protein interactions [30] as it is in vertebrates [74].

During embryogenesis, *Unc-86* is expressed in an array of neuroblasts, almost entirely descended from founder cell AB, however one neuron is descended from founder cell C. Together, *Unc-86* is expressed in 6 neuroblasts leading to 29 classes of neurons [75]. These adult neurons vary widely in form (e.g. interneurons, chemoreceptors, and nociceptors), and in function (e.g. CO_2/O_2 detection, temperature sensation, and egg-laying behavior), and in transmitter type (e.g. Glutamate, Serotonin, Acetylcholine, as well as various neuroactive peptides) (Table 1) [76]. A highly comprehensive review of *Unc-86* has recently been published by Leyva-Diaz, *et al*., which focuses on *Unc-86* properties and how they relate to other *Pou4* genes [77]; thus, we will below only briefly mention several aspects of the well-studied Unc-86 gene.

In *Unc-86* null mutants, overall morphological development is normal, however entire neuronal cell lineages fail to differentiate properly or at all [75, 78]. All *Unc-86*+ cells divide asymmetrically [75]; *Unc-86* is responsible for establishing that asymmetric cell fate because in its absence daughter cells that would normally exhibit *Unc-86* expression maintain the profiles of the parent cell [78]. Differences in *Unc-86* activity are decided by distinct regulatory regions, whereby the resultant binding of various transcription factors confers to the *Unc-86* gene the property of producing asymmetric daughters [79]. How this relationship operates is still unclear. One upstream gene of *Unc-86* is *lin-32*, a vertebrate *Atonal* homologue (which is also a *Pou4F3* activator in inner ear hair cells [80]). Intriguingly, every neuron of the worm produced by the combination of *lin-32* and *Unc-86* differentiates into a sensory neuron [79]. Because of *Atonal*'s role in activating *Pou4F3* in vertebrate mechanosensory hair cells and RGCs [81], this may be an ancestral trait of this gene pair/cell type (for a deeper discussion on this matter, see [77]). *Pou4* genes are in many organisms involved in photosensory system development, so it is additionally interesting to note that although *Unc-86* is expressed in over 60 neuron types in *C. elegans*, from interneurons to CO_2 sensatory neurons, that it is not involved in developing any of the *C. elegans* photosensory neurons [82], i.e. neurons that express *LITE-1*, the transmembrane protein involved in *C. elegans* phototransduction [83].

Arthropods: The *Pou4* homologue in the fly *Acj6* was named for the behavior of the *Drosophila melanogaster* mutant: Abnormal chemosensory jump 6. *Acj6* was long ago identified as having an important role in specifying a suite of olfactory receptor neurons (ORNs) in the olfactory sensilla of the maxillary palp [84], but its complicated role in the regulatory programs of those different cells was only discovered recently. There are 120 ORNs: 2 neurons per 60 individual

sensilla. Of these 120 they are divided into six classes, each class involved in a different array of odor reception [85]. The *Drosophila* genome contains only one *Pou4* gene, and for years it appeared as if *Acj6* was alternatively spliced producing one variant that acted as an activator, and another that lacked two amino acids in the POU-Specific domain that appeared to act as a repressor, previously named *I-Pou* for inhibitory Pou [86, 87]. After further investigation this simplified scenario turned out not to be the case, however it remained true that some *Acj6* variants indeed act as repressors, as well as activators, in that particular splice forms bind and repress larval targets during adult stages [85]. *Acj6* undergoes a diversity of alternative splicing events, each particular to different cell types, producing an incredible array of 13 different transcripts, eight of which are expressed in the maxillary palp [39]. This is by far the most complex and variation-producing display of splicing events of any *Pou4* gene yet studied. Each of the 13 transcripts contains the exons 1, 4, 5, 6, 7, 8, and 9, the latter four encoding the amino acids of the POU Specific domain and the POU Homeodomain, while the inter-POU-IV box exons 2 and 3 are not always present. As exons 1 and 4 encode the POU-IV box, this fact strengthens formidably the suggestion that the POU-IV box is an important feature of this gene. The variety of exon combinations experienced by the Acj6 POU-IV box hints at the exciting possibility that different combinations of POU-IV box exons translate into a protein with different protein-protein interaction partners. Exons 5 and 8 have long and short forms, adding another layer of functional diversity and evolutionary plasticity to the *Drosophila Pou4* gene. Each of these transcripts is expressed in a spatially distinct manner in different ORNs of the maxillary palp, in the antennae, or in the whole larva. Overall, *Drosophila* has evolved a complicated system of splicing their *Pou4* gene and use the 13 described forms to differentiate different types of their many chemo- and mechanoreceptive neurons (Table 2). *Acj6* has been

implicated in proper synaptic connections [88], axon outgrowth [46], and choice of odorant detected by the neuron in which a particular splice form of *Acj6* is expressed [39]. *Acj6* is expressed in a visual processing section of the fly brain called the lobula [46], yet like *C. elegans*, *Drosophila* does not use any of its *Pou4* gene variants during photoreceptor development [89, 90].

Limulus polyphemus, the North Atlantic horseshoe crab, has two *Pou4* genes, probably as a result of their lineage specific (xiphosuran) WGD event [91]. Their *Pou4* genes were electronically annotated as, "inhibitory POU protein-like"; no expression data is available.

Deuterostomes

Echinodermata: Sea urchins have been a powerhouse of understanding gene regulatory networks in animals, yet little is known of *Pou4* gene expression and function in any of the echinoderm larval stages. This is likely due to the fact that *Pou4* genes are predominantly involved in post-mitotic neuronal differentiation and not typically found in the group of genes that establish germ layers [92], while the majority of investigation into sea urchin development is of early stages, such as germ layer formation and gastrulation. In *Strongylocentrotus purpuratus*, *Pou4* expression is detected via qPCR in the 48-hour pluteus larva [93], and it is electronically annotated in a number of other species [94, 95]. Sea urchin plutei larva, at 3 days post fertilization, develop a pair of photosensory structures that utilize photoreceptive opsins [96], but other molecular markers within them are currently unknown. Interestingly, the genomic organization of the *S. purpuratus Pou4* resembles that of the vertebrates and amphioxus, not like the rest of the invertebrates examined, having a short first exon, a single intron (yet far larger

than cephalochordates at over 7700bp), and a second final exon encoding most of the protein [97].

Although little information exists for embryonic/larval echinoderm *Pou4* expression, the tube feet of adult sea urchins do express *Pou4*. Tube feet are photoreceptive and express a set of opsins required or photoprocessing [98, 99] and *Pou4* is expressed and likely directly involved in the differentiation of these structures [100]. Interestingly, all of the neurons of the urchin tube feet are of the PNS [101], unlike *Pou4*-expressing retinal ganglion cells of the CNS. Thus, *Pou4* has a roles in photoreceptor differentiation regardless of whether it is PNS or CNS. Nevertheless, a series of experiments was designed to test the evolutionary fidelity of *Pou4* protein function whereby the sea urchin *Pou4* was knocked into (i.e. replaced) the *Pou4F2* mouse gene locus [100]. Remarkably, the sea urchin *Pou4* protein was able to compensate fully for the absence of mouse *Pou4F2*, restoring normal vision to the mice. Although the *S. purpuratus* Pou4 shows several amino acid differences in its POU-IV box and POU domain (vs. Pou4F2; Figure 1A, 1B), it is presumably able to interact with the same protein partners and DNA sequences that the mouse Pou4F2 protein normally interacts with [102].

Hemichordates: In the acorn worms *Saccoglossus kowalevskii* and *Ptychodera flava*, *Pou4* expression was detected via PCR amplification during embryonic and adult stages [60, 103]. No data yet exists for *Pou4* spatial expression in a hemichordate.

Cephalochordates: The spatial distribution of *Pou4*-expressing cells in the amphioxus *Branchistoma floridae* is detected primarily but not limited to ciliated epidermal sensory neurons (CESNs) and chemoreceptive neurons of the anterior-most region of the larva [25]. Additionally,

beginning at the 10hpf neurula stage, *Pou4* is also detected in the anterior neural plate, a CNS feature, expressed near the region of future eye development. However, *Pou4* expression is lost in this region long before eye formation; further investigation is required to determine whether *Pou4* is involved in amphioxus photoreception development. In the dorsolateral position of the neural tube, *Pou4* is expressed in a segmentally-spaced pattern, and it has been argued that it may be functioning to specify motorneurons, as *Unc-86* does in *C. elegans* [104, 105] or as in *Ciona* [106], and it remains to be seen if this expression is synonymous to ascidian motor ganglion expression (see below).

Ascidians: Ascidians have a single copy of the *Pou4* gene, typical of most invertebrates. In addition to having just one copy, the organizational structure of several exons and introns closely resembles the organization invertebrates exhibit, versus the short first intron and large second exon pattern of vertebrates, amphioxus, and urchins [25, 107]. In *Ciona*, *Pou4* identifies PNS neurons during larval development, first expressed during neurulation and marking ciliated epidermal sensory neurons (CESNs), the chemosensory adhesive palps, and bipolar tail neurons (BTNs) [26]. There is also evidence that *Pou4* is expressed in the developing motor ganglia of the neural tube in pre-hatched larva [106]. *Ciona* Pou4 has been implicated in gene regulatory networks of developing CESNs and BTNs in several reports [108-110], and is required for the formation of CESNs [55]. Little information is known about *Pou4* function or expression in the adult animal, but it is likely that *Pou4* is involved in the differentiation of mechanosensory hair cells within the adult siphon called coronal cells. Due to the fact that coronal cells share so many aspects of the vertebrate hair cell development, both genetically, structurally, and synaptically, they have been recently classified as homologous [111]. There is evidence that there are two

alternatively spliced forms of the *Ciona Pou4* [112], but there are no published reports of any temporal, spatial, or functional differences between the two.

Vertebrates: All vertebrates examined, with the exception of therians and some fish species (Figures 3 and 4), have four *Pou4* genes, *Pou4F1*, *Pou4F2*, *Pou4F3*, and *Pou4F4* (*Pou4F1.2*) likely due to two whole genome duplication events. Predominantly, *Pou4F1* and *Pou4F2* are involved in neuron specification and differentiation during brain development, eye development, and inner ear neuronal development, whereas *Pou4F3* is primarily involved in developing mechanosensory hair cells [113, 114]. With the exception of instances where more than one *Pou4* gene is expressed in a cell type, and whereby it has been shown that its partner can compensate for absence of the other [115], *Pou4* genes in all cases examined have been shown as indispensable for normal differentiation of the cell type in which they are expressed [116-118].

Due to an abundance of overlapping expression patterns between various vertebrate neural systems, we will examine vertebrate *Pou4* expression on a system by system basis, rather than a gene or organism type basis.

CNS (Eye): In all tetrapods examined, *Pou4F1*, *Pou4F2*, and *Pou4F3* are expressed in the retinal ganglion cells of the eye [119], but temporal variations exist. In the mouse *Pou4F1* is expressed first at E10.5 [120] vs. *Pou4F2* at E11.5 [121, 122] followed by *Pou4F3* at E12.5 [113]; whereas in the chick, *Pou4F3* is expressed first followed by a seeming co-initiation of *Pou4F1* and *Pou4F2* [102]. Expression of *Pou4F1*, *Pou4F2* and *Pou4F3* was detected in the amphibian eye (*Xenopus laevis*), but *in situ* data in the frog eye only exists for *Pou4F1* and *Pou4F2* [27, 123]. *Pou4F1.2* (X*Brn3d*), has been analyzed in *Xenopus* and also shown to be

involved in *Xenopus* eye development [27], and thus all four *Pou4* genes are expressed in the amphibian eye during its formation [124, 125]. Temporal sequence has not been examined thoroughly in the zebrafish (*Danio rerio*) eye; however *Pou4F1*, *Pou4F2*, and *Pou4F3* are expressed in that organ [66, 107]. In medaka, only *Pou4F2* and *Pou4F4* are expressed in the developing retina [40]. *Pou4F1* KO mice show disruption to the normal stratification of retinal ganglion cells (RGCs), but not a severe drop in RGC number experienced in *Pou4F2* KO mice [126]. *Pou4F2* KO mice show a startling loss of up to 70% of their RGCs [127], and delayed expression of *Pou4F1*, indicating partial reliance on *Pou4F2* for proper *Pou4F1* expression [126]. A subset of vertebrate RGCs are "intrinsically photosensitive" (ipRGCs, first described in [128]), and in the mouse this population is Pou4F1$_{+}$/Pou4F2$_{-}$ [129, 130].

CNS (Brain and Spinal Cord): Many regions of the mouse brain and elements of the spinal cord express *Pou4* genes (Table 3). The first detection of a *Pou4* gene is *Pou4F1* in the embryonic midbrain; interestingly this expression pattern is conserved in embryonic zebrafish [131]. *Pou4F1* is the most extensively expressed *Pou4* gene in the mouse CNS, ranging from the neural tube to the diencephalon to the dorsal grey horn of the spinal cord [132-134]. *Pou4F1* KO mice have severe disruptions to normal CNS and PNS development, displaying a malformed inferior olive and red nucleus, attributed to aberrant neuronal cell migration [126]. Xenopus *XBrn3d* (*Pou4F1.2*) is not surprisingly also involved with patterning the CNS during embryogenesis, expressed in a wide range of developing CNS structures [135]. Zebrafish and medaka express *Pou4F1* in the habenular nucleus and in the optic tectum, the major visual processing center in non-mammalian vertebrates [40, 123, 136] as well as sensory neurons of the developing spinal cord [137]. Mammalian brain regions expressing *Pou4F2* include (but are not

limited to) the dorsal root ganglia, acoustic ganglia, and brain stem [133, 138]; similarly the *Xenopus* neural tube, anterior neural fold, and trigeminal placode all express *Pou4F2* [27]. In zebrafish, *Pou4F2* labels the dorsal midbrain, cerebellum, and rostral forebrain, as well as rhombomeres 3-5 (features of the hindbrain) [139]; a similar but less detailed picture exists for Pou4F2 in medaka [40]. Available information suggests *Pou4F3* plays the most spatially- and temporally-limited role of the *Pou4* genes in vertebrate CNS development, where it is expressed in the developing mammalian diencephalon, the hind- mid- and forebrain, and the spinal cord [140]. Sequencing analysis of larval brain tissue in Xenopus identified *Pou4F3* expression [125], while there is no evidence that *Pou4F3* is expressed in developing brains of zebrafish or medaka.

PNS (auditory/mechanosensory): The ciliated sensory cells of the inner ear are probably the best-studied PNS cells expressing *Pou4F3* genes, and the organ where *Pou4F3* arguably plays its most important role [114]. In the mouse, beginning at E11.5, *Pou4F1* and *Pou4F2* are both expressed in the otic epithelium [141], while *Pou4F3* does not appear until E12.5 [113]. Fish do not have a middle or outer ear nor a cochlea, however the auditory hair cells of zebrafish saccule, utricle, and lagena (a cochlea homolog), are indeed marked by *Pou4F3* [142]. The lateral lines of zebrafish and amphibians are populated with neuromasts containing mechanosensory hair cells and various supporting cells; while no expression data exists for *Pou4* gene expression in the larval frog lateral line, in zebrafish *Pou4F2* is expressed in a subset of neuromast support cells [66], while *Pou4F3* is expressed in the ciliated sensory neurons of the neuromasts [143]. Both *Pou4F1* and *Pou4F2* KO mice have several PNS features malformed including sensory neurons and motorneurons; behaviorally, mice have a disturbed swallowing reflex and poor locomotion [132, 144]. *Pou4F3* KO mice are essentially normal, despite complete loss of hearing due to a

lack of hair cell development in the inner ear [127, 145]; *Pou4F3* in humans is known as a deafness gene, *DNF1* [21, 146]. This overall developmental normality suggests that *Pou4F3* is potentially dispensable in the non-hair-cell cell types it is expressed in. Loss of *Pou4F3* also demonstrates how critical this gene is to normal hair cell development. In these KO mice, hair cells begin to develop, however their normal differentiation, maturation, and survival is dependent on *Pou4F3* [113].

Viscera: Vertebrates have utilized both *Pou4F1* and *Pou4F2* in development of their hearts; experiments in mammals detected both genes expressed in cardiomyocytes and cardiac neurons [147, 148]. Expression patterns are largely overlapping, and gene function seems to be partially redundant [147]. In zebrafish, cardiac deformities are only seen when both *Pou4* genes are simultaneously knocked down; additionally, homozygous mutants develop normally whereas crossing heterozygous *Pou4F1* and *Pou4F2* mutant lines results in embryonic lethality [16]. In mammals, *Pou4F1* may be involved in various aspects of embryonic cardiac remodeling [147], whereas *Pou4F2* has been shown to be responsible for adaptive hypertrophic responses [16] and ischemic responses [149]. *Pou4F1* and *Pou4F2* are also expressed in the developing ovaries and testes in mammals [150, 151], whereas *Pou4* expression was not detected in two teleost fish species [152, 153]. Little information is known about cell-type expression in the mammalian ovary, while *Pou4F1* was shown to be expressed in spermatogonia vs. *Pou4F2* expression in spermatocytes and spermatids [154]. Information is unavailable regarding specific activities of *Pou4* genes in the reproductive organs.

DISCUSSION

Pou4: A Metazoan Gene Utilized in Various Systems

The organ systems in which we have knowledge of *Pou4* expression are summarized in Figure 4. However, an absence of data creates an opacity towards viewing a complete picture of the *Pou4* gene: in several cases presented, little information is known about *Pou4* function, indeed in three entire animal groups (hemichordates, placozoans, and sponges), we only have a single representative member, each having only EST or PCR data. As more data comes to light, it will be interesting to observe how the following suggestions are altered.

Except for the instances of *Pou4F1* and *Pou4F2* being expressed in reproductive organs [18, 37] and vertebrate hearts [16, 147], *Pou4* is predominantly a neuron-specific, proneural transcription factor critical to the differentiation or post-differentiation survival of the cell in which it is active. Figure 4 illuminates a feature perhaps shared by all metazoans carrying a Pou4 gene: the presence of *Pou4* expression in the development of ciliated mechanoreceptive cells. This cell type is ancient, dating back to the eumetazoan ancestor over 700 Mya [155, 156], so perhaps this gene/cell type combination is equally ancient; perhaps *Pou4* coevolved with this cell type. Additionally, one of the major sensory cells in sponges are the choanocytes, ciliated or flagellated cells that work together to establish water flow and eject flotsam [157]. It will be interesting to see if *Pou4* is involved in the development of this cell type. Sponges lost or never developed a bona fide nervous system [158] (even though they do have nervous system features [157] and express proneural NK homeobox genes [159]), and because sponges have very few cell types this supposition makes future reports of *Pou4* activity in sponges all the more interesting. *Pou4* activity in choanocytes could either bolster the idea of sponges having a nervous system-like anatomy and/or further push the functional *Pou4* origin squarely into coevolution with ciliated neuronal mechanoreceptors. To fully understand the origins of *Pou4*

function and the extent of its capacity throughout the animal kingdom will require considerably more study.

Further patterns emerge within this gene's evolutionary history when the cell types and organ systems *Pou4* is present in are compared. In addition to *Pou4* expression in all mechanoreceptive cells, in every animal for which we have the requisite data it is also expressed in all developing brains. Differences include how vertebrate chemosensation development requires *Pou4* genes but not photosensory development, whereas the inverse is true for the invertebrate chordates. While choanoflagellates do not contain a *Pou4* gene, nor any genes of the POU-domain class [160], the metazoan LCA has been inferred to contain the Pou4 gene [11]. Ctenophores stand out among the metazoans as not having a *Pou4* gene, it therefore must have been lost in this group, due to *Pou4* presence in several sponges [18]. Placozoans and sponges both have *Pou4* genes, however the only evidence of their presence is EST data. Figure 4 provides a framework for understanding the breadth and utility of the *Pou4* gene, there are certainly fillable gaps present: for example it has been shown that *Pou4* is expressed in the brain of the Acoela, and that *Pou4* is expressed in the photoreceptive neurons of the adult urchin. But in each of those examples other data, for example data regarding *Pou4* spatial expression in the larval urchin, is simply unknown. Urchin larva, in addition to containing photoreceptive structures, also contain ciliated mechanoreceptors within their ciliary band [161, 162]. It would be interesting to know if *Pou4* is involved with the generation of either. *Pou4* expression in the mechanoreceptive cilia is essentially expected, as all other animals seem to require it for theirs, whereas *Pou4* in the photosensory structures would certainly add to the discussion of whether or not there is a connection between *Pou4* and the generation of those structures within the

deuterostomes, while adding to the story about how the invertebrate chordates have perhaps abandoned *Pou4* use in developing theirs.

In the tetrapods, the available *Pou4* data is most detailed in the mammals; a wealth of expression and functional data exists for *Pou4* activity from mouse and rat studies and from mammalian cell culture [163]. The three mouse *Pou4* genes (*F1, F2* and *F3*) are expressed in a wide variety of neural tissues, with much expression overlap of the three in some cell types. The fact that there is much overlap and cases of functional redundancy, such as how all three *Pou4* genes can promote RGC differentiation [102], or how the urchin *Pou4* gene can functionally substitute for mammalian *Pou4* genes [100], present a case for little functional divergence at the protein-protein/protein-DNA interfaces of these proteins, despite noticeable differences throughout much of their sequences outside of highly conserved regions.

Paralogous Trends of Vertebrate Pou4 Proteins

Figure 2 and Figure 3 make clear that *Pou4F1*, *Pou4F1.2*, *Pou4F2*, and *Pou4F3* are paralogs, visualized by orthologs tightly grouping together. Aligning the highly conserved regions of their proteins reveals some interesting features, for example there are several amino acids that are specific to an ortholog and conserved since their evolutionary split. In Figure 1A, the alignment makes clear that all of the Pou4F2 proteins have a 2-amino acid spacing unique to this group (*yellow highlighting*). Additionally, within the Pou4F1 orthologs, there is a conserved serine at position 19, and, at position 36-37 where most have a valine-serine, Pou4F2 has a adenine-valine (*green highlighting*). There are also several examples of this in the POU-Specific and POU Homeodomains of Pou4F1.2, Pou4F2 and Pou4F3 (Figure 1B).

Figures 2 and 3 also demonstrate that one of the two ancestral *Pou4* genes, after the vertebrate's second WGD, diverged into *Pou4F1* and *Pou4F1.2*, while the other diverged into *Pou4F2* and *Pou4F3*. This concept is also visualized by several amino acids shared between either Pou4F1 and Pou4F1.2, or Pou4F2 and Pou4F3 (*bold magenta text between aligned rows*). Three examples of this exist in both the POU-IV Box (positions 4, 18, and 24), and two within the POU Homeodomain (positions 113 and 152). Additionally bolstering the concept of a *Pou4F1*/*Pou4F1.2* common ancestor and a Pou4F2/ Pou4F3 common ancestor, no amino acids are shared exclusively between non-paralogs.

Evolution of *Pou4* Copy Number

Protostomes: There are many documented cases of protostomes that have undergone WGD events, including bdelloid rotifers [64], the reef-building coral *Acrospora* [164], *Limulus* and the xiphosurans [91], Arachnopulmonata (i.e. spiders and scorpions) [165], the African Land Snail [166], free-living flatworms of the genus Macrostomum [167, 168], and possibly 18 WGD events in various hexapods [169, 170]. However, based on a survey of information regarding expression data and functional data, one might conclude that most protostomes have only one *Pou4* gene, e.g. *Acj6* and *Unc-86*. However, this is almost certainly not the case, e.g. *Limulus polyphemus* has two *Pou4* genes, likely due to the horseshoe crab's aforementioned WGD event, and it is expected that other organisms with duplicated *Pou4* genes will be discovered.

There are also examples of protostomes who have more than one copy of *Pou4*, yet without past WGD events. *S. mediterranea* has two *Pou4* genes, and they too are expressed in different territories and evidence suggests they perform similar proneural roles in the distinct cell types they function in, but clearly operate within distinguishably different types of neurons [65,

68]. How the differences in protein sequence act to confirm different properties in *S. mediterranea* are unknown. Also, the cnidarian *C. sowerbyi* has three *Pou4* genes. Studies show hydrozoans have duplicated several genes in their genomes [171, 172], yet this does not present a case for a proclivity towards gene duplications, and the significance of *C. sowerbyi* having three *Pou4* genes remains unknown; that they are expressed in different territories agues for neofunctionalization, however cell-*type* data thus hypotheses of their roles are not yet developed.

Deuterostomes: The chordate common ancestor and its single *Pou4* gene experienced two whole genome duplication events; mounting evidence demonstrates the second vertebrate WGD event preceding the agnathan/gnathostome split [173-175]. Our jawless vertebrate ancestors either A) underwent two WGD events whole cloth, or B) experienced a type of genome duplication recently classified as *2rjv allotetraploidy* [176], whereby the doubling of chromosomes occurred through interspecific hybridization after the first WGD. Curiously, the trees developed here for Pou4 depict the agnathan Pou4 genes as duplicated prior to the second duplication that led to the four extant vertebrate orthologous groups (Figures 3 and 4), and this is currently unresolved. Nevertheless, all scenarios allow for the generation of the extant vertebrate *Pou4* gene copies. Five *Pou4* genes are present in the sturgeon *A. ruthenus*, likely due to their lineage-specific third WGD around 350 million years ago [177]. Teleosts, including the basal teleost *P. kingsleyae*, have also undergone a third WGD [178]; the freshwater elephantfish *P. kingsleyae* has retained six *Pou4* genes while zebrafish and medaka have retained only four copies. Thus many of the WGD-doubled *Pou4* ohnologs have apparently been lost during rediploidization in teleosts, which curiously stands in contrast to the otherwise high rate of noted retention of proneural ohnologues [179]. In amphibians, even though *Xenopus tropicalis* in particular has not experienced another WGD [180], this phenomena is quite common [181], and

several species show up to dodecaploidy, i.e., twelve sets of chromosomes per somatic cell [182]. Within the tetrapods, it has been proposed that over the history of WGD events, particular chromosomal sections referred to as paralogons were conserved in non-therian vertebrates, regions that included the *Pou4F1.2* locus [183]. As such, evidence suggests during therian sex-chromosome evolution, the paralogon that included *Brn3d*/*Pou4F1.2*, was lost [184].

History of *Pou4* Systems Deployment

Within the Ecdysozoans and Spiralians, available information reveals that molluscs, arthropods, and nematodes utilize *Pou4* in their chemosensory system development. However, no data exists to suggest annelids use of *Pou4* genes to develop theirs. Absence of evidence is not evidence of absence, however, and the question should probably be best left as *unknown*, rather than at this time suggesting annelids lost the use of *Pou4* for chemoreception development.

Mammals and hydrozoans both express *Pou4* genes in their gonads (Figure 4, purple boxes); *Craspedacusta sowerbyi* expresses one of its three *Pou4* genes in the adult gonad, while the mouse *Pou4F1* and *Pou4F2* are both expressed in the reproductive organs of both sexes. Due to their wide phylogenetic separation and lack of other organisms expressing *Pou4* gonadally, this likely represents in both cases a redeployment of *Pou4* from previous functions in mammalian and hydrozoan life histories, rather than being ancestrally reproductive. Transcriptomes from the Bluefin Tuna (*Thunnus maccoyii*) [153], and Tilapia (*Oreochromis niloticus*) [152], as well as amphioxus [185] and two tunicate transcriptomes [112, 186], all revealed lack of *Pou4* gene expression in surveys of reproductive organ-specific genes. Thus *Pou4* involvement during gametogenesis may be tetrapod or mammalian-specific within the

chordates. Further investigation into protostome reproductive organ genes will unveil if *Pou4* has a genetic proclivity for reproductive organ development in addition to numerous neuronal cell types.

Within the chordates, what emerges most clearly -and most definitively, due to extensive data- is that *Pou4* is employed for chemosensory neurons in invertebrate chordates (Figure 4; blue box) and not for developing photoreception. This stands in contrast to vertebrates' utilization of all of their *Pou4* genes in the development of their visual systems (Figure 4; yellow box), while available evidence suggests *Pou4* genes are decidedly not involved in vertebrate chemosensory neuron development [17, 187].

The jury is still out whether or not the bilaterian common ancestor had a pair of anterior eyes, and whether or not they are ancestral to homologous structures, i.e. the paired eyes many animals exhibit today [188]. Although the visual systems vertebrates have evolved are more complex than those of the invertebrate chordates, the basic schema of photoreceptors integrating into processing centers of the brain is common to all, and it is unclear why the invertebrate chordates evolved their photoreception systems without *Pou4*. This problem is complicated though, and necessitates a brief discussion of chordate eye evolution. Amphioxus has four photoreceptive organs [189]; three of the four photoreceptive organs in amphioxus are likely related to vertebrate photoreceptive cells. The frontal eye shares morphological and genetic signatures with vertebrate eyes, while the two other photoreceptive organs, the Joseph Cells and the Organ of Hesse, are regarded as related to vertebrate ipRGCs [190] (which in mice do require a *Pou4* gene, *Pou4F1* [129, 130]). The fourth, termed the lamellar body, shares homology and genetic signature with the vertebrate pineal gland [190]. Ascidians on the other hand, developed photoreception exclusively from the pineal gland ancestor. Because their ocellus expresses an

opsin closely related to the vertebrate sensory opsin, the structure was initially thought to be homologous to vertebrate eyes [191]. However, more recent genetic and structural studies place the ocellus as a pineal derivative [192]. Thus, amphioxus, as ancestral to both tunicates and vertebrates, possesses both types of photoreception. There is no evidence, however, that pineal structures in either amphioxus, tunicates, or any vertebrates including mammals [193] express a *Pou4* gene. As the sea urchin expresses *Pou4* (in their photoreceptive tube feet), and the invertebrate chordates do not, a weakly supported suggestion could be made that invertebrate chordates may have lost the use of *Pou4* for photoreception development. However, better knowledge of larval sea urchin *Pou4* expression, as well as detailed *Pou4* data in hemichordates would be needed prior to advancing the idea.

A similar but inverse situation is presented for chemoreception, and the scenario is probably more straightforward compared to photoreception. In both ascidians and amphioxus, the best-studied chemoreceptive neurons (located in the anterior end of amphioxus juveniles/adults and ascidian tadpole larva) express *Pou4* [25, 26], whereas evidence suggests vertebrate chemosensory-related structure do not express *Pou4* [194]. Because of fairly widespread use of Pou4 in chemoreceptive systems, vertebrates have probably lost the use of *Pou4* for chemosensory development; again echinoderm or hemichordate data of *Pou4* in chemoreceptive neurons would significantly advance the hypothesis that vertebrates have lost this feature.

Pou4F1.2 **aka** ***Brn3d*** **aka** ***Pou4F4***

Due to whole genome duplication events and subsequent gene losses, vertebrates are typically regarded to have three *Pou4* genes, *Pou4F1* (*Brn3a*), *Pou4F2* (*Brn3b*), and *Pou4F3*

(*Brn3c*). However, as discussed above medaka and *Xenopus* have four *Pou4* genes, *Pou4F1.2* and *Pou4F4* respectively (both also referred to as *Brn3d*). A BLAST search of *Pou4F1.2* reveals homologues in representatives of all examined vertebrate groups including monotremes but with the exception of mammals and marsupials (Figures 3 and 4). Yet the story is more complicated, and an attempt to display the history of *Pou4* genes is presented in Figure 4. BLAST results from the Pou4F1.2 sequence gives an interesting result: scores of similar hits in avian, reptilian, and other amphibian Pou4 protein collections. No therians (placental mammals and marsupials) appear to have a *Pou4F1.2* gene; and its loss was suggested to have occurred during therian sex-chromosome evolution [184]. Currently, little is known about function or expression of this gene class, except for an expression pattern in medaka [40] and a handful of studies in frogs [27, 195, 196], where it has been shown to be important to amphibian neurogenesis. It thus stands to reason that other *Pou4F1.2* homologues will be revealed to have similar important roles.

Due to the importance of *Pou4* genes in neurogenesis, a hope is that enhanced recognition of non-mammalian vertebrate *Pou4F1.2* homologues may lead to stimulating developments in burgeoning yet already exciting models such as the little skate [197, 198] and several reptiles [199], as well as more established models such as the chick and zebrafish. Furthermore, reports sometimes mischaracterize *Pou4F1.2* as a *Pou4F2* homologue [200], and it may come to light that other proteins originally cast as Pou4F1, because they most closely aligned to the mammalian homologue, may turn out to be Pou4F1.2. For example the first chicken Pou4 protein identified was classified as Brn3a (Pou4F1) due to a high sequence similarity to the mammalian Pou4F1 [201], however, an assessment of the amino acid sequence of the Pou4 protein in that report makes clear that the protein they present is not Pou4F1, but in fact Pou4F1.2.

CONCLUSIONS

Since the advent of the *Pou4* gene class nearly 1 billion years ago, *Pou4* genes have stayed largely fidelitous to their initial roles as transcriptional mediators of neurogenesis. Comparing sister taxa has allowed the emergence of patterns of organ-system loss and gain of *Pou4* activity throughout evolutionary history. Only in the cases where gene duplication events have occurred do we see a departure from neural deployment as in vertebrate hearts and mammalian and hydrozoan gonads. This is most remarkable when combined with the fact that *Pou4* proteins have been shown to bind to the same DNA sequence. That point is even more interesting given the fact that mammals have apparently done more with less: non-therian vertebrates have four *Pou4* gene copies while mammals with their three copies have significantly widened the spatial deployment of their *Pou4* genes into the development of the heart and gonads.

Arguably the most interesting question that has emerged from this analysis is how Pou4 proteins, which are principally involved in neurogenesis throughout the animal kingdom, can diverge in activity and be part of the differentiation of structures like the heart and reproductive organs, even though they bind to the same DNA sequence? Which immediately leads to the question of regulatory circuit redeployment. Within Pou4+ heart cells and Pou4+ neurons, are there any upstream regulators e.g. *Atonal* [81] or downstream targets e.g. *BDNF* [202] conserved between the cell types? Or, have mammalian *Pou4* regulatory regions evolved to utilize other regulatory factors found in their hearts? There are sure to be additional questions brought forth as more *Pou4* data is generated for some of the understudied groups and gaps in our knowledge are filled. It is an intriguing and important problem involving various different types of research in

many types of organisms guaranteed to lead to a fuller and more interesting picture of the history of *Pou4* genes and the evolution of the genetics of development.

ACKNOWLEDGEMENTS

Clifford Pickett gratefully acknowledges ARCS Foundation support, San Diego Chapter. This Introduction, "A Cross-Scale Examination of *Pou4* Genes During Development: insights from a phylogenetically-comprehensive analysis**,** C. J. Pickett and Robert W. Zeller", in full, has been submitted for publication and at this time is in the process of formal review. The dissertation author was the primary investigator and author of this material.

A.

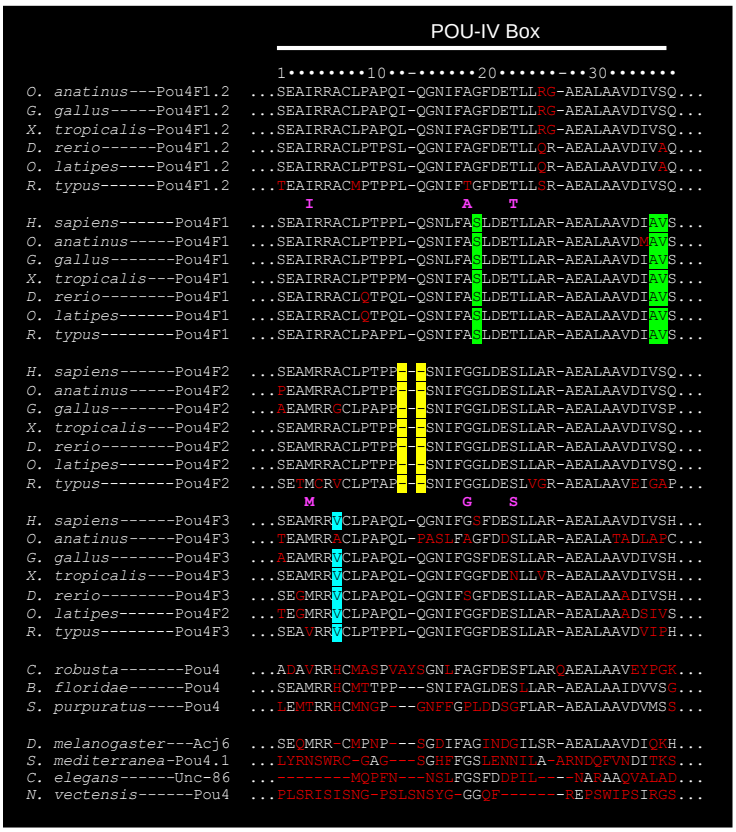


Figure I.1: Alignment of POU-IV boxes, POU-Specific Domains, and POU Homeodomains of various Pou4 proteins reveals a startling display of conservation across bilaterians. (A) POU-IV box of various protostomes and deuterostomes. (B) Green text in B highlights Pou4 variable linker regions. Red text in vertebrate Pou4 proteins identifies amino acid outliers compared to the other four paralogs; in the invertebrates red text identifies amino acids not shared with any of the vertebrate paralogs. Highlighting of various colors identifies amino acids unique in that position for a given vertebrate Pou4 gene. Bold magenta highlights amino acids shared between adjacent Pou4 paralogs. Publicly available sequences of each Pou4 protein listed were collected from NCBI and UniProt (Supplementary Table 1) and aligned using T-Coffee [203].

B.

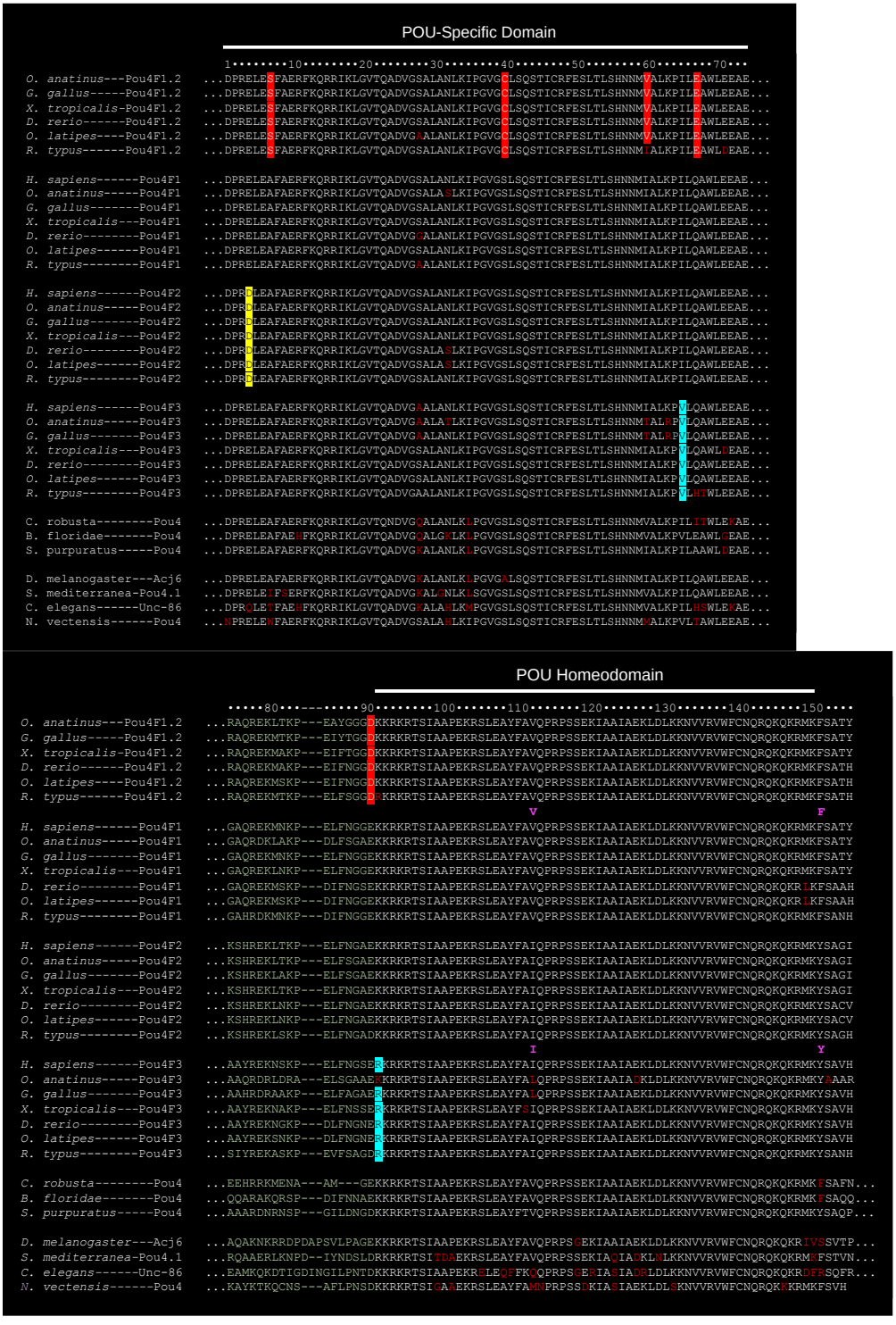


Figure I.1: Alignment of POU-IV boxes, POU-Specific Domains, and POU Homeodomains of various Pou4 proteins reveals a startling display of conservation across bilaterians, continued.

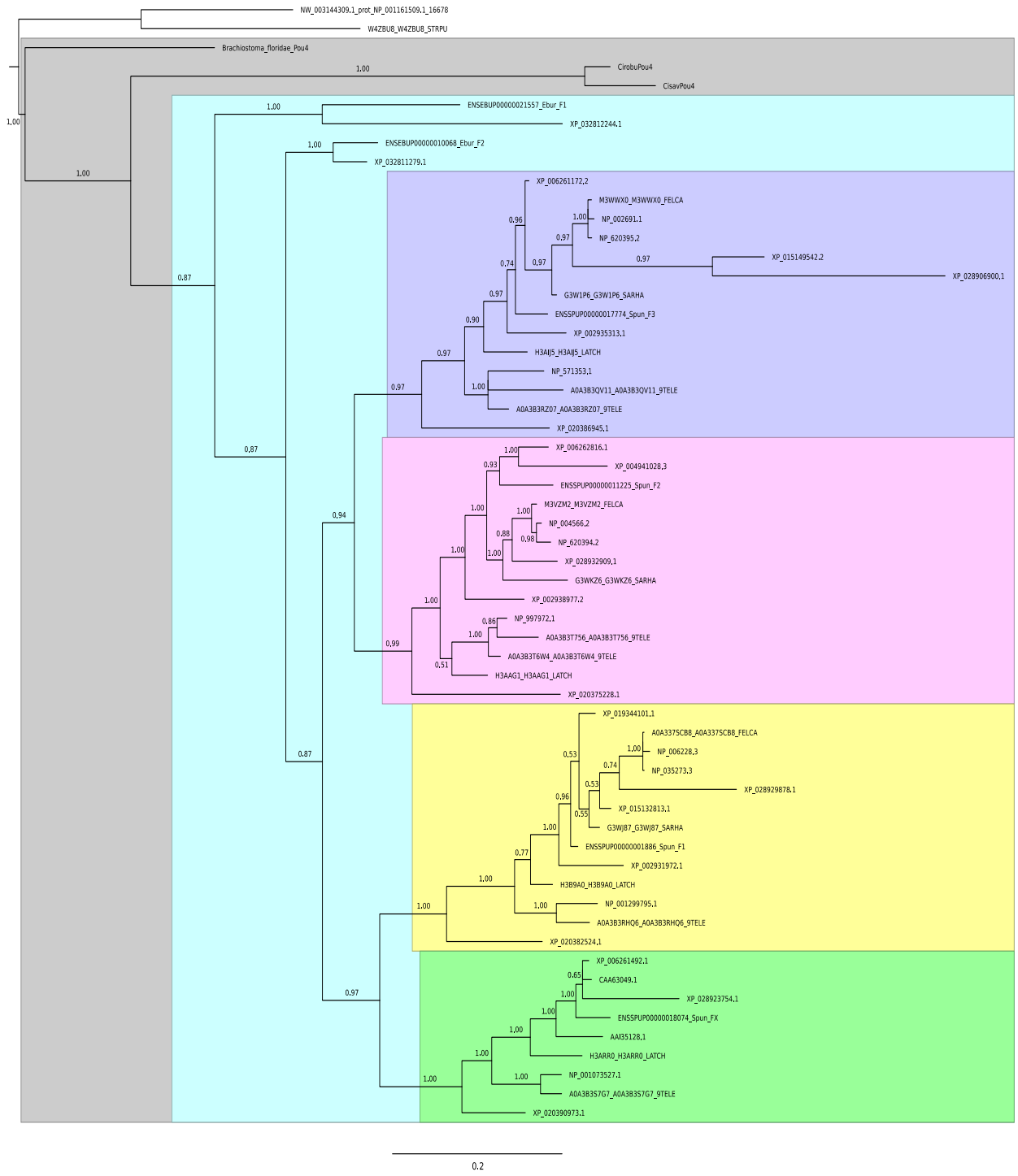


Figure I.2: Phylogenetic tree produced with PhyML [44] reveals four clades of vertebrate Pou4 proteins (Pou4F1, Pou4F1.2, Pou4F2 and Pou4F3). Pou4F1.2 is absent in the therians. Protein sequences were aligned with M-Coffee [204] with some manual editing and this alignment was used as input into PhyML. Field colors: Grey, invertebrate chordates; cyan, jawless vertebrates; Purple, Pou4F3; Pink, Pou4F2; Yellow, Pou4F1; Green, Pou4F1.2.

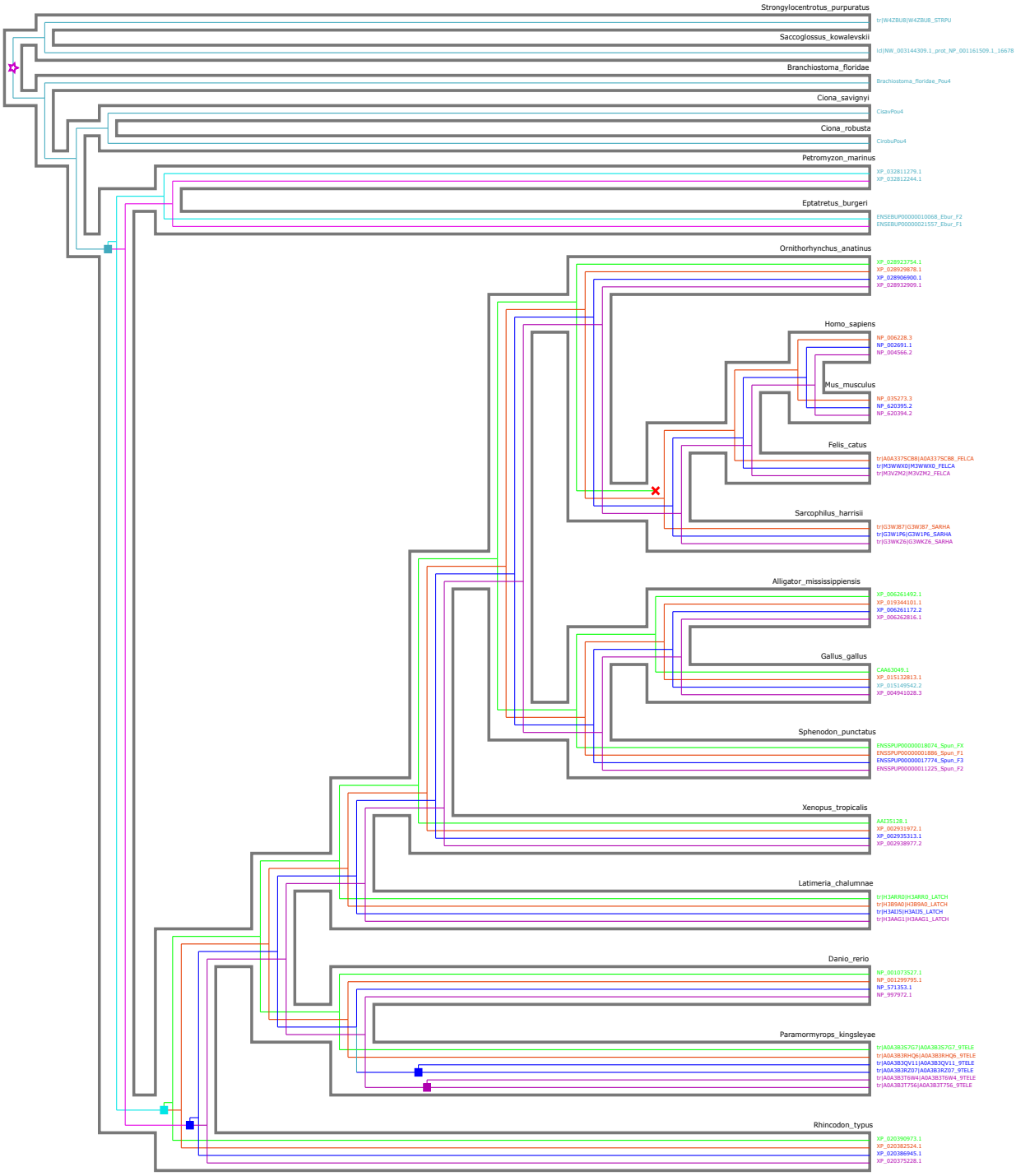


Figure I.3: Treerecs reconciled PhyML tree [41] depicts evolutionary history of deuterostome *Pou4* gene duplications and loss. Magenta star represents last common deuterostome *Pou4* gene; various colored squares represent Pou4 duplications, at least one of which could have preceded the one of the rounds of vertebrate WGD. We are unable to ascertain if the second round of Pou4 gene duplications was gene copy-specific or associated with a second round of WGD in vertebrates as this has not yet been resolved [173, 174]. Red X indicates Pou4F1.2 loss in therians.

Table I.1: *C. elegans* neurons expressing *Unc-86*. *Neuron name* abbreviations from WormBase [205]. Abbreviations: RL and DV indicate right and left paired neurons, and dorsal and ventral paired neurons, respectively; Glu, Glutamate; Ser, Serotonin; ACh, Acetylcholine. Data collected from [193] and from relevant citations within the text.

Neuron Name	Neuron Type	Neurotransmitter	Function
ADA RL	interneuron	Glu	chemosensation
AIM RL	interneuron	Ser, Glu, ACh	Ser regulation
AIZ RL	interneuron	Glu	locomotion
ALM RL	mechanosensory	Glu	touch
ALN RL	chemosensory	ACh	O_2 sensor
BDU RL	interneuron	peptide	AVM guidance
CEMD LR/DV	chemosensory	ACh	pheromone sensor
FLP LR	nociceptive	Glu	thermo/ mechanoreceptive
HSN LR	motorneuron	ACh, Ser	egg-laying
I2 LR	interneuron	Glu	unknown
IL2 LR/DV	mechanosensory	ACh	nictation (dauer dispersal)
NSM LR	proprioceptive/ mechanosensory	Ser	sense food/bacteria
PHC LR	thermoreceptor	Glu	temp avoidance
PLM LR	mechanosensory	Glu	gentle touch sense
PLN LR	chemosensory	ACh	O_2 sensor
PQR	chemosensory	Glu	O_2 CO_2 sensor
PVD LR	nociceptive	Glu	thermo/mechanoreceptive
PVM	mechanosensory	unknown	stretch sense
PVR	interneuron	Glu	pharyngeal sense
RIH	interneuron	ACh, Ser	unknown
RIP LR	interneuron	unknown	pharynx-soma connection
RIR	interneuron	ACh	unknown
RMG LR	interneuron	peptide	sensory neuron integration
SDQ LR	interneuron, sensory	ACh	O_2 sensor
URA LR/DV	motorneuron	ACh	unknown
URB LR	interneuron	ACh	unknown
URX LR	chemosensory	ACh	O_2 CO_2 sensor
URY LR/DV	chemosensory	Glu	mate/pathogen sense

Table I.2: *Drosophila* neurons and regions expressing *Acj6*. Drosophila neuron type abbreviations from FlyBase [206]. Data collected from [205] and relevant citations throughout text.

Embryonic *Acj6* Domains		
Neuron	**Type**	**Description**
lateral cord	NB7-4	glia progenitor
central brain	ND	neuron progenitor
Adult *Acj6* Domains		
brain	DL1	adult antennae lobe projection
brain	DVA3	adult antennae lobe projection
brain	VA1d	adult antennae lobe projection
brain	VM7	adult antennae lobe projection
brain	VM2	adult antennae lobe projection
brain	DM6	adult antennae lobe projection
brain	VA1lm	adult antennae lobe projection
olfactory receptor	Or92a	olfactory receptor neuron
neuron	Or47b	olfactory receptor neuron
optic lobe		neuron

Table I.3: Mouse Developmental Stages and Cells Expressing *Pou4* Genes. Abbreviations: DRG, Dorsal Root Ganglion; RGC, Retinal Ganglion Cell**.** Expression data was collected from [206], in addition to the specific citations in the corresponding text.

Pou4F1		***Pou4F2***		***Pou4F3***	
stage	**tissue/cell**	**stage**	**tissue/cell**	**stage**	**tissue/cell**
8.5	midbrain	9.5	facio-acoustic pre-ganglion complex	10.5	diencephalon
9	trigeminal nerve	E9.5	dorsal root ganglion	10.5	hindbrain
9.5	otic pit epithelium	10	trigeminal ganglion	10.5	midbrain
9.5	spinal cord precursors	10.5	hindbrain	12.5	inner ear
9.5	DRG	10.5	trunk somite	12.5	retina
10	neural tube	11	neural tube	12.5	spinal cord
10	neural fold	11	DRG	13.5	ventricular layer
10.5	hindbrain	11.5	inner ear epithelium	13.5	thalamus
10.5	otocyst epithelium	11.5	retina	13.5	midbrain
10.5	optic cup	11.5	acoustic ganglion	13.5	cochlear HCs
10.5	acoustic ganglion	12.5	RGCs	13.5	vestibular HCs
10.5	trigeminal ganglion	13.5	brainstem	14.5	inner ear vestibule
11.5	inner ear epithelium	15.5	cochlear ganglion	15.5	cochlea
11.5	retina	P adult	medulla oblongata	15.5	maculae
12.5	brainstem	P adult	pons	15.5	cristae
12.5	diencephalon	P adult	glomerular tuft		
12.5	vestibulo-cochlear ganglion				
12.5	cochlear ganglion				
13	spinal cord				
13.5	habenula				
13.5	neural retinal epithelium				
13.5	RGCs				
14.5	medulla oblongata				
14.5	pons				
14.5	facial ganglion				
14.5	glossopharyngeal ganglion				
14.5	dorsal grey horn				
18.5	inferior olivary nucleus				
P1	red nucleus				

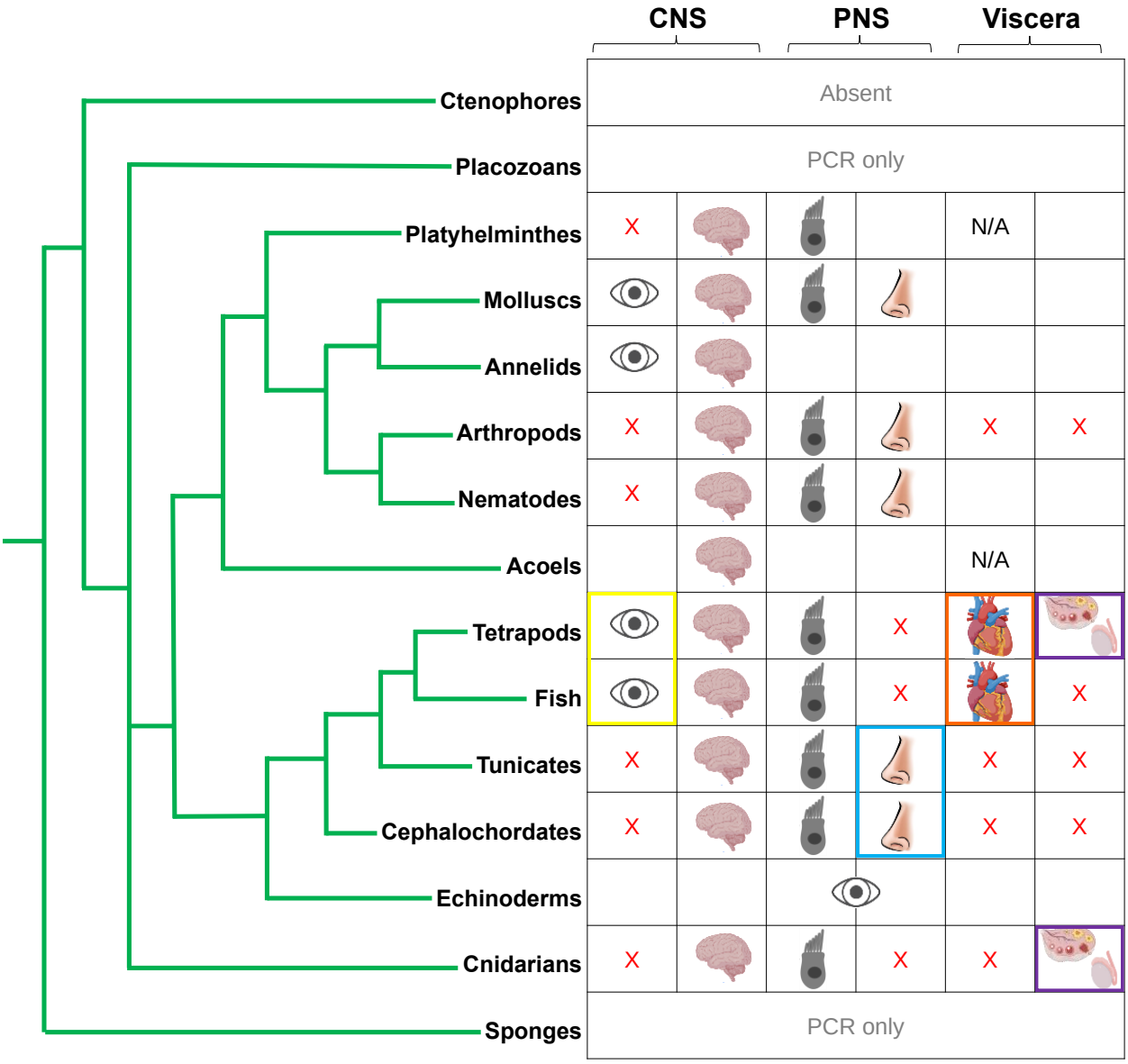
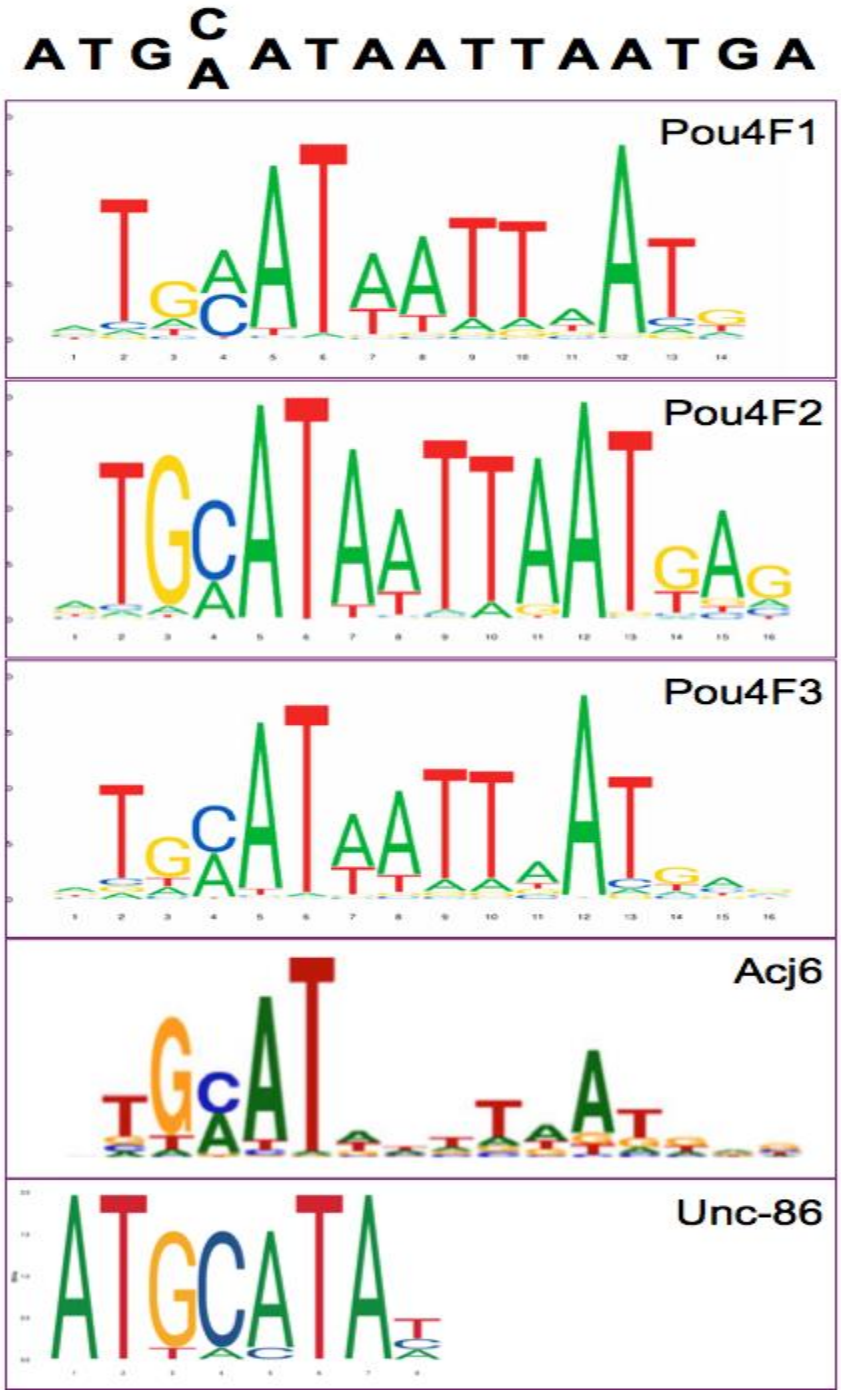


Figure I.4: Phylogeny (modeled after [11]) demonstrating categories of *Pou4* spatial expression in various systems reveals noteworthy patterns. "Eye" icon represents *Pou4* involved in photoreception; "Brain" icon represents *Pou4* activity in CNS (brain and spinal cord); "Cell with cilia" icon represents *Pou4* activity in ciliated mechanoreceptors; "Nose" icon represents chemoreception/olfactory *Pou4* activity; "Heart" icon represents *Pou4* activity in the heart; ovary and testis cartoons represent expression in the gonads. Yellow box highlights vertebrate properties invertebrate chordates do not share; Blue box groups invertebrate chordate properties vertebrates do not share; Orange box groups properties exclusive to vertebrates, while the purple box highlights a feature of Pou4 expression specific to mammals and a cnidarian. Red box highlights ctenophore absence, and X's denote demonstrated absence in a particular system.



Supplementary Figure I.1: An alignment of Pou4 binding site LOGOs reveals conserved bilaterian binding of Pou4 proteins. Top: Consensus sequence from the alignment of mammalian and arthropod Pou4 binding site LOGOs. Purple boxes below: Pou4F1, Pou4F2, and Pou4F3 binding site LOGOs were generated from ChIP-seq data PWMs [48], as were Acj6 [47] and Unc-86 [49], and loosely aligned to reveal the overall consensus sequence. Unc-86 fits the pattern, however, the "TTAAGTA" of the consensus is not present.

Supplementary Table I.1: Accession numbers and UniProt ID's (if available) for protein sequences analyzed or referenced in Figure 1.

Group *Species*	**Gene Name**	**NCBI ref seq**	**UniProt**
Therian	*H. sapiens Pou4F1*	NP_006228.3	Q01851
Homo sapiens	*H. sapiens Pou4F2*	NP_006228.3	Q12837
	H. sapiens Pou4F3	NP_002691.1	Q15319
Monotreme	*O. anatinus Pou4F1*	XP_028929878.1	N/A
Ornithorhynchus	*O. anatinus Pou4F1.2*	XP_028923754.1	N/A
anatinus	*O. anatinus Pou4F2*	XP_028932909.1	F7C2L5
	O. anatinus Pou4F3	XP_028906900.1	F7CAM9
Aves	*G. gallus Pou4F1*	XP_015132813.1	P55968
Gallus gallus	*G. gallus Pou4F1.2*	CAA63049.1	Q91998
	G. gallus Pou4F2	XP_004941028.3	A0A1D5PZI1
	G. gallus Pou4F3	XP_015149542.2	A0A1D5PDX7
Amphibian	*X. tropicalis Pou4F1*	XP_002931972.1	A0A5G3HSU0
Xenopus tropicalis	*X. tropicalis Pou4F1.2*	AAI35128.1	A4IGJ4
	X. tropicalis Pou4F2	XP_002938977.2	A0A5G3KJV7
	X. tropicalis Pou4F3	XP_002935313.1	A0A1L8GXL4
Teleost	*D. rerio Pou4F1*	NP_001299795.1	F1Q813
Danio rerio	*D. rerio Pou4F1.2*	AAI28836.1	A1A5W2
	D. rerio Pou4F2	NP_997972.1	Q6XZH1
	D. rerio Pou4F3	NP_571353.1	Q90435
Teleost	*O. latipes Pou4F1*	XP_024151581.1	N/A
Oryzias latipes	*O. latipes Pou4F1.2*	XP_004076857.1	N/A
	O. latipes Pou4F2	XP_024115081.1	N/A
	O. latipes Pou4F3	XP_004076179.2	N/A
Chondrichthyes	*R. typus Pou4F1*	XP_020382524.1	N/A
Rhincodon typus	*R. typus Pou4F1.2*	XP_020390973.1	N/A
	R. typus Pou4F2	XP_020375228.1	N/A
	R. typus Pou4F3	XP_020386945.1	N/A
Tunicate *Ciona intestinalis*	*C. intestinalis Pou*	XP_009857505.1	Q4H2X8
Cephalochordate *Branchistoma floridae*	*B. floridae Pou4*	ABC42926.1	Q1WFI8
Echinoderm *Strongylocentrotus purpuratus*	*S. purpuratus Sp-Pou4f2*	XP_786727.2	W4ZBU8

Supplementary Table I.1: Accession numbers and UniProt ID's (if available) for protein sequences analyzed or referenced in Figure 1, continued.

Group *Species*	**Gene Name**	**NCBI ref seq**	**UniProt**
Arthropod *Drosophila melanogaster*	*D. melanogaster Acj6*	NP_001245684.1	M9NDX7
Platyhelminthes *Schmidtea mediterranea*	*S. mediterranea Pou4l-1*	AIX99546.1	A0A0A1CRK9
Nematode *Caenorhabditis elegans*	*C. elegans Unc-86*	NP_001021191.2	T2MDR7
Hydra *Hydra vulgaris*	*H. vulgaris Pou4*	XP_004205460.2	T2MDR7
Nematostella *Nematostella vectensis*	*N. vectensis Nm-Pou4*	*XP_032219883.1*	N/A

Supplementary Table I.2: Protein IDs used to construct Figure 2 and Figure 3.

Name in figures	Sequence Identifier	Source	Species	Common Name
CirobuPou4	CirobuPou4	Internal sequence from clone, gene model KH.C2.42.v1.A.SL1-1 @www.aniseed.cnrs.fr	*Ciona robusta*	Ascidian
CisavPou4	H2ZNV6_CIOSA	ensembl.org	*Ciona savignyi*	Ascidian
tr|W4ZBU8|W4ZBU8_STRPU	W4ZBU8_STRPU	uniprot.org	*Strongylocentrotus purpuratus*	Purple sea urchin
Brachiostoma_floridae_Pou4	ABC42926.1	ncbi.nlm.nih.gov	*Branchiostoma floridae*	Lancelet
lcl|NW_003144309.1_prot_NP _001161509.1_16678	NP_001161509.1	ncbi.nlm.nih.gov	*Saccoglossus kowalevskii*	Acorn worm
XP_032811279.1	XP_032811279.1	ncbi.nlm.nih.gov	*Petromyzon marinus*	Lamprey
XP_032812244.1	XP_032812244.1	ncbi.nlm.nih.gov	*Petromyzon marinus*	Lamprey
ENSEBUP00000021557_Ebur_F1	ENSEBUP00000021557	ensembl.org	*Eptatretus burgeri*	Inshore hagfish
ENSEBUP00000010068_Ebur_F2	ENSEBUP00000010068	ensembl.org	*Eptatretus burgeri*	Inshore hagfish
XP_020386945.1	XP_020386945.1	ncbi.nlm.nih.gov	*Rhincodon typus*	Whale shark
XP_020390973.1	XP_020390973.1	ncbi.nlm.nih.gov	*Rhincodon typus*	Whale shark
XP_020382524.1	XP_020382524.1	ncbi.nlm.nih.gov	*Rhincodon typus*	Whale shark
XP_020375228.1	XP_020375228.1	ncbi.nlm.nih.gov	*Rhincodon typus*	Whale shark
XP_006261172.2	XP_006261172.2	ncbi.nlm.nih.gov	*Alligator mississippiensis*	American alligator
XP_006261492.1	XP_006261492.1	ncbi.nlm.nih.gov	*Alligator mississippiensis*	American alligator
XP_006262816.1	XP_006262816.1	ncbi.nlm.nih.gov	*Alligator mississippiensis*	American alligator
XP_019344101.1	XP_019344101.1	ncbi.nlm.nih.gov	*Alligator mississippiensis*	American alligator
CAA63049.1	CAA63049.1	ncbi.nlm.nih.gov	*Gallus gallus*	Chicken
XP_015149542.2	XP_015149542.2	ncbi.nlm.nih.gov	*Gallus gallus*	Chicken
XP_015132813.1	XP_015132813.1	ncbi.nlm.nih.gov	*Gallus gallus*	Chicken
XP_004941028.3	XP_004941028.3	ncbi.nlm.nih.gov	*Gallus gallus*	Chicken
NP_006228.3	NP_006228.3	ncbi.nlm.nih.gov	*Homo sapiens*	Human
NP_004566.2	NP_004566.2	ncbi.nlm.nih.gov	*Homo sapiens*	Human
NP_002691.1	NP_002691.1	ncbi.nlm.nih.gov	*Homo sapiens*	Human
tr|H3AAG1|H3AAG1_LATCH	H3AAG1_LATCH	Uniprot.org	*Latimeria chalumnae*	West Indian Ocean coelacanth
tr|H3AIJ5|H3AIJ5_LATCH	H3AIJ5_LATCH	Uniprot.org	*Latimeria chalumnae*	West Indian Ocean coelacanth
tr|H3ARR0|H3ARR0_LATCH	H3ARR0_LATCH	Uniprot.org	*Latimeria chalumnae*	West Indian Ocean coelacanth
tr|H3B9A0|H3B9A0_LATCH	H3B9A0_LATCH	Uniprot.org	*Latimeria chalumnae*	West Indian Ocean coelacanth
NP_035273.3	NP_035273.3	ncbi.nlm.nih.gov	*Mus musculus*	House Mouse
NP_620394.2	NP_620394.2	ncbi.nlm.nih.gov	*Mus musculus*	House Mouse
NP_620395.2	NP_620395.2	ncbi.nlm.nih.gov	*Mus musculus*	House Mouse
XP_028923754.1	XP_028923754.1	ncbi.nlm.nih.gov	*Ornithorhynchus anatinus*	Platypus
XP_028906900.1	XP_028906900.1	ncbi.nlm.nih.gov	*Ornithorhynchus anatinus*	Platypus
XP_028929878.1	XP_028929878.1	ncbi.nlm.nih.gov	*Ornithorhynchus anatinus*	Platypus
XP_028932909.1	XP_028932909.1	ncbi.nlm.nih.gov	*Ornithorhynchus anatinus*	Platypus
tr|A0A3B3QV11|A0A3B3QV11_9TELE	A0A3B3QV11_9TELE	Uniprot.org	*Paramormyrops kingsleyae*	Freshwater Elephantfish
tr|A0A3B3RHQ6|A0A3B3RHQ6_9TELE	A0A3B3RHQ6_9TELE	Uniprot.org	*Paramormyrops kingsleyae*	Freshwater Elephantfish
tr|A0A3B3S7G7|A0A3B3S7G7_9TELE	A0A3B3S7G7_9TELE	Uniprot.org	*Paramormyrops kingsleyae*	Freshwater Elephantfish

Supplementary Table I.2: Protein IDs used to construct Figure 2 and Figure 3, continued.

Name in figures	Sequence Identifier	Source	Species	Common Name
tr|A0A3B3RZ07|A0A3B3RZ07_9TELE	A0A3B3RZ07_9TELE	Uniprot.org	*Paramormyrops kingsleyae*	Freshwater Elephantfish
tr|A0A3B3T6W4|A0A3B3T6W4_9TELE	A0A3B3T6W4_9TELE	Uniprot.org	*Paramormyrops kingsleyae*	Freshwater Elephantfish
tr|A0A3B3T756|A0A3B3T756_9TELE	A0A3B3T756_9TELE	Uniprot.org	*Paramormyrops kingsleyae*	Freshwater Elephantfish
NP_571353.1	NP_571353.1	ncbi.nlm.nih.gov	*Danio rerio*	Zebrafish
NP_997972.1	NP_997972.1	ncbi.nlm.nih.gov	*Danio rerio*	Zebrafish
NP_001299795.1	NP_001299795.1	ncbi.nlm.nih.gov	*Danio rerio*	Zebrafish
NP_001073527.1	NP_001073527.1	ncbi.nlm.nih.gov	*Danio rerio*	Zebrafish
AAI35128.1	AAI35128.1	ncbi.nlm.nih.gov	*Xenopus tropicalis*	Western clawed frog
XP_002935313.1	XP_002935313.1	ncbi.nlm.nih.gov	*Xenopus tropicalis*	Western clawed frog
XP_002938977.2	XP_002938977.2	ncbi.nlm.nih.gov	*Xenopus tropicalis*	Western clawed frog
XP_002931972.1	XP_002931972.1	ncbi.nlm.nih.gov	*Xenopus tropicalis*	Western clawed frog
tr|A0A337SCB8|A0A337SCB8_FELCA	A0A337SCB8_FELCA	Uniprot.org	*Felis catus*	Domestic Cat
tr|M3VZM2|M3VZM2_FELCA	M3VZM2_FELCA	Uniprot.org	*Felis catus*	Domestic Cat
tr|M3WWX0|M3WWX0_FELCA	M3WWX0_FELCA	Uniprot.org	*Felis catus*	Domestic Cat
tr|G3W1P6|G3W1P6_SARHA	G3W1P6_SARHA	Uniprot.org	*Sarcophilus harrisii*	Tasmanian devil
tr|G3WJ87|G3WJ87_SARHA	G3WJ87_SARHA	Uniprot.org	*Sarcophilus harrisii*	Tasmanian devil
tr|G3WKZ6|G3WKZ6_SARHA	G3WKZ6_SARHA	Uniprot.org	*Sarcophilus harrisii*	Tasmanian devil
ENSSPUP00000018074_Spun_FX	ENSSPUP00000018074	ensembl.org	*Sphenodon punctatus*	Tuatara
ENSSPUP00000001886_Spun_F1	ENSSPUP00000001886	ensembl.org	*Sphenodon punctatus*	Tuatara
ENSSPUP00000011225_Spun_F2	ENSSPUP00000011225	ensembl.org	*Sphenodon punctatus*	Tuatara
ENSSPUP00000017774_Spun_F3	ENSSPUP00000017774	ensembl.org	*Sphenodon punctatus*	Tuatara

REFERENCES

1. Ohno, S., *Evolution by Gene Duplication.* 1970.

2. Kasahara, M., *The 2R hypothesis: an update.* Curr Opin Immunol, 2007. **19**(5): p. 547-52.

3. Singh, P.P., J. Arora, and H. Isambert, *Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes.* PLoS computational biology, 2015. **11**(7).

4. Krumlauf, R., *Hox genes, clusters and collinearity.* Int J Dev Biol, 2018. **62**(11-12): p. 659-663.

5. Morgan, B.A., *Hox genes and embryonic development.* Poult Sci, 1997. **76**(1): p. 96-104.

6. Ryan, A.K. and M.G. Rosenfeld, *POU domain family values: flexibility, partnerships, and developmental codes.* Genes Dev, 1997. **11**(10): p. 1207-25.

7. Finney, M., G. Ruvkun, and H.R. Horvitz, *The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors.* Cell, 1988. **55**(5): p. 757-69.

8. Malik, V., D. Zimmer, and R. Jauch, *Diversity among POU transcription factors in chromatin recognition and cell fate reprogramming.* Cell Mol Life Sci, 2018. **75**(9): p. 1587-1612.

9. Ryan, J.F., et al., *The homeodomain complement of the ctenophore Mnemiopsis leidyi suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa.* Evodevo, 2010. **1**(1): p. 9.

10. Scholer, H.R., *Octamania: the POU factors in murine development.* Trends Genet, 1991. **7**(10): p. 323-9.

11. Gold, D.A., R.D. Gates, and D.K. Jacobs, *The early expansion and evolutionary dynamics of POU class genes.* Mol Biol Evol, 2014. **31**(12): p. 3136-47.

12. Collum, R.G., et al., *A novel POU homeodomain gene specifically expressed in cells of the developing mammalian nervous system.* Nucleic Acids Res, 1992. **20**(18): p. 4919-25.

13. Gerrero, M.R., et al., *Brn-3.0: a POU-domain protein expressed in the sensory, immune, and endocrine systems that functions on elements distinct from known octamer motifs.* Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10841-5.

14. Deutsch, J.S., *Homeosis and beyond. What is the function of the Hox genes?* Adv Exp Med Biol, 2010. **689**: p. 155-65.

15. Benditt, J., *POU! Goes the homeobox.* Sci Am, 1989. **260**(2): p. 20, 22.

16. Maskell, L.J., et al., *Essential but partially redundant roles for POU4F1/Brn-3a and POU4F2/Brn-3b transcription factors in the developing heart.* Cell Death Dis, 2017. **8**(6): p. e2861.

17. Gray, P.A., et al., *Mouse brain organization revealed through direct genome-scale TF expression analysis.* Science, 2004. **306**(5705): p. 2255-7.

18. Hroudova, M., et al., *Diversity, phylogeny and expression patterns of Pou and Six homeodomain transcription factors in hydrozoan jellyfish Craspedacusta sowerbyi.* PLoS One, 2012. **7**(4): p. e36420.

19. Océane Tournière, D.D., Gemma Sian Richards, Kartik Sunagar, Yaara Y Columbus-Shenkar, Yehu Moran, Fabian Rentzsch, *NvPOU4/Brain3 functions as a terminal selector gene in the nervous system of the cnidarian Nematostella vectensis.* biorxiv, 2020.

20. Nakanishi, N., et al., *Evolutionary origin of rhopalia: insights from cellular-level analyses of Otx and POU expression patterns in the developing rhopalial nervous system.* Evolution & development, 2010. **12**(4): p. 404-415.

21. Weiss, S., et al., *The DFNA15 deafness mutation affects POU4F3 protein stability, localization, and transcriptional activity.* Mol Cell Biol, 2003. **23**(22): p. 7957-64.

22. Herr, W. and M.A. Cleary, *The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain.* Genes Dev, 1995. **9**(14): p. 1679-93.

23. Wegner, M., D.W. Drolet, and M.G. Rosenfeld, *POU-domain proteins: structure and function of developmental regulators.* Curr Opin Cell Biol, 1993. **5**(3): p. 488-98.

24. Herr, W., et al., *The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products.* Genes Dev, 1988. **2**(12A): p. 1513-6.

25. Candiani, S., et al., *Expression of AmphiPOU-IV in the developing neural tube and epidermal sensory neural precursors in amphioxus supports a conserved role of class IV POU genes in the sensory cells development.* Dev Genes Evol, 2006. **216**(10): p. 623-33.

26. Candiani, S., et al., *Ci-POU-IV expression identifies PNS neurons in embryos and larvae of the ascidian Ciona intestinalis.* Dev Genes Evol, 2005. **215**(1): p. 41-5.

27. Hutcheson, D.A. and M.L. Vetter, *The bHLH factors Xath5 and XNeuroD can upregulate the expression of XBrn3d, a POU-homeodomain transcription factor.* Dev Biol, 2001. **232**(2): p. 327-38.

28. Sturm, R.A. and W. Herr, *The POU domain is a bipartite DNA-binding structure.* Nature, 1988. **336**(6199): p. 601-4.

29. Dawson, S.J., et al., *Functional role of position 22 in the homeodomain of Brn-3 transcription factors.* Neuroreport, 1998. **9**(10): p. 2305-9.

30. Rockelein, I., et al., *Identification of amino acid residues in the Caenorhabditis elegans POU protein UNC-86 that mediate UNC-86-MEC-3-DNA ternary complex formation.* Mol Cell Biol, 2000. **20**(13): p. 4806-13.

31. Martin, S.E., X. Mu, and W.H. Klein, *Identification of an N-terminal transcriptional activation domain within Brn3b/POU4f2.* Differentiation, 2005. **73**(1): p. 18-27.

32. Smith, M.D., et al., *Bcl-2 transcription from the proximal P2 promoter is activated in neuronal cells by the Brn-3a POU family transcription factor.* Journal of Biological Chemistry, 1998. **273**(27): p. 16715-16722.

33. Begbie, J.L. and D.S. Latchman, *The two activation domains of the Brn-3a transcription factor have distinct functional properties.* The International Journal of Biochemistry & Cell Biology, 1997. **29**(12): p. 1493-1500.

34. Smith, M.D., et al., *The N-terminal domain unique to the long form of the Brn-3a transcription factor is essential to protect neuronal cells from apoptosis and for the activation of Bcl-2 gene expression.* Nucleic acids research, 1998. **26**(18): p. 4100-4107.

35. Thomas, G., et al., *EWS differentially activates transcription of the Brn-3a long and short isoform mRNAs from distinct promoters.* Biochemical and biophysical research communications, 2004. **318**(4): p. 1045-1051.

36. Budhram-Mahadeo, V., et al., *The closely related POU family transcription factors Brn-3a and Brn-3b are expressed in distinct cell types in the testis.* Int J Biochem Cell Biol, 2001. **33**(10): p. 1027-39.

37. Theil, T., et al., *Short isoform of POU factor Brn-3b can form a heterodimer with Brn-3a that is inactive for octamer motif binding.* Journal of Biological Chemistry, 1995. **270**(52): p. 30958-30964.

38. Theil, T., et al., *Mouse Brn-3 family of POU transcription factors: a new aminoterminal domain is crucial for the oncogenic activity of Brn-3a.* Nucleic Acids Res, 1993. **21**(25): p. 5921-9.

39. Bai, L. and J.R. Carlson, *Distinct functions of acj6 splice forms in odor receptor gene choice.* J Neurosci, 2010. **30**(14): p. 5028-36.

40. Brombin, A., et al., *Genome-wide analysis of the POU genes in medaka, focusing on expression in the optic tectum.* Dev Dyn, 2011. **240**(10): p. 2354-63.

41. Rambaut, A., *FigTree 1.4. 4 software.* Institute of Evolutionary Biology, Univ. Edinburgh, 2014.

42. Comte, N., et al., *Treerecs: an integrated phylogenetic tool, from sequences to reconciliations.* bioRxiv, 2019: p. 782946.

43. Dehal, P. and J.L. Boore, *Two rounds of whole genome duplication in the ancestral vertebrate.* PLoS Biol, 2005. **3**(10): p. e314.

44. Guindon, S. and O. Gascuel, *A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.* Syst Biol, 2003. **52**(5): p. 696-704.

45. Huelsenbeck, J.P. and F. Ronquist, *MRBAYES: Bayesian inference of phylogenetic trees.* Bioinformatics, 2001. **17**(8): p. 754-5.

46. Certel, S.J., et al., *Regulation of central neuron synaptic targeting by the Drosophila POU protein, Acj6.* Development, 2000. **127**(11): p. 2395-405.

47. Jiang, M., et al., *Evidence for a hierarchical transcriptional circuit in Drosophila male germline involving testis-specific TAF and two gene-specific transcription factors, Mod and Acj6.* FEBS Lett, 2018. **592**(1): p. 46-59.

48. Jolma, A., et al., *DNA-binding specificities of human transcription factors.* Cell, 2013. **152**(1-2): p. 327-39.

49. Weirauch, M.T., et al., *Determination and inference of eukaryotic transcription factor sequence specificity.* Cell, 2014. **158**(6): p. 1431-1443.

50. Latchman, D.S., *The Brn-3a transcription factor.* Int J Biochem Cell Biol, 1998. **30**(11): p. 1153-7.

51. Dawson, S.J., P.J. Morris, and D.S. Latchman, *A single amino acid change converts an inhibitory transcription factor into an activator.* J Biol Chem, 1996. **271**(20): p. 11631-3.

52. Feuda, R., et al., *Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals.* Curr Biol, 2017. **27**(24): p. 3864-3870 e4.

53. Sebe-Pedros, A., et al., *Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq.* Cell, 2018. **173**(6): p. 1520-1534 e20.

54. Serrano-Saiz, E., et al., *BRN3-type POU Homeobox Genes Maintain the Identity of Mature Postmitotic Neurons in Nematodes and Mice*. Curr Biol, 2018. **28**(17): p. 2813-2823 e2.

55. Pickett, C.J. and R.W. Zeller, *Efficient genome editing using CRISPR-Cas-mediated homology directed repair in the ascidian Ciona robusta.* Genesis, 2018. **56**(11-12): p. e23260.

56. Horridge, G.A., *Statocysts of medusae and evolution of stereocilia.* Tissue Cell, 1969. **1**(2): p. 341-53.

57. Cary, L.R., *The influence of the marginal sense organs on metabolic activity in Cassiopea xamachana bigelow.* Proceedings of the National Academy of Sciences of the United States of America, 1916. **2**(12): p. 709.

58. Spangenberg, D., et al., *Touch-plate and statolith formation in graviceptors of ephyrae which developed while weightless in space.* Scanning microscopy, 1996. **10**(3): p. 875-87; discussion 887-8.

59. Wenger, Y. and B. Galliot, *Punctuated emergences of genetic and phenotypic innovations in eumetazoan, bilaterian, euteleostome, and hominidae ancestors.* Genome Biol Evol, 2013. **5**(10): p. 1949-68.

60. Ramachandra, N.B., et al., *Embryonic development in the primitive bilaterian Neochildia fusca: normal morphogenesis and isolation of POU genes Brn-1 and Brn-3.* Dev Genes Evol, 2002. **212**(2): p. 55-69.

61. Philippe, H., et al., *Mitigating Anticipated Effects of Systematic Errors Supports Sister-Group Relationship between Xenacoelomorpha and Ambulacraria.* Curr Biol, 2019. **29**(11): p. 1818-1826 e6.

62. Fröbius, A.C. and P. Funch, *Rotiferan Hox genes give new insights into the evolution of metazoan bodyplans.* Nature communications, 2017. **8**(1): p. 1-10.

63. Gomez, A., et al., *Ecological genomics of adaptation to unpredictability in experimental rotifer populations.* 2019.

64. Flot, J.-F., et al., *Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga.* Nature, 2013. **500**(7463): p. 453-457.

65. Cowles, M.W., et al., *COE loss-of-function analysis reveals a genetic program underlying maintenance and regeneration of the nervous system in planarians.* PLoS Genet, 2014. **10**(10): p. e1004746.

66. DeCarvalho, A.C., S.L. Cappendijk, and J.M. Fadool, *Developmental expression of the POU domain transcription factor Brn-3b (Pou4f2) in the lateral line and visual system of zebrafish.* Dev Dyn, 2004. **229**(4): p. 869-76.

67. Joyce Tang, W., J.S. Chen, and R.W. Zeller, *Transcriptional regulation of the peripheral nervous system in Ciona intestinalis.* Dev Biol, 2013. **378**(2): p. 183-93.

68. Elliott, S.A., *Studies of conserved cell-cell signaling pathways in the planarian, Schmidtea mediterranea*. 2016, Department of Neurobiology and Anatomy, University of Utah. p. x, 478 pages.

69. Olson, P.D., et al., *Genome-wide transcriptome profiling and spatial expression analyses identify signals and switches of development in tapeworms.* Evodevo, 2018. **9**: p. 21.

70. Arnold, J.M., *Normal embryonic stages of the squid, Loligo pealii (Lesueur).* The Biological Bulletin, 1965. **128**(1): p. 24-32.

71. Wollesen, T., et al., *POU genes are expressed during the formation of individual ganglia of the cephalopod central nervous system.* Evodevo, 2014. **5**: p. 41.

72. O'Brien, E.K. and B.M. Degnan, *Developmental expression of a class IV POU gene in the gastropod Haliotis asinina supports a conserved role in sensory cell development in bilaterians.* Dev Genes Evol, 2002. **212**(8): p. 394-8.

73. Backfisch, B., et al., *Stable transgenesis in the marine annelid Platynereis dumerilii sheds new light on photoreceptor evolution.* Proc Natl Acad Sci U S A, 2013. **110**(1): p. 193-8.

74. Budhram-Mahadeo, V., et al., *Activation of the alpha-internexin promoter by the Brn-3a transcription factor is dependent on the N-terminal region of the protein.* J Biol Chem, 1995. **270**(6): p. 2853-8.

75. Finney, M. and G. Ruvkun, *The unc-86 gene product couples cell lineage and cell identity in C. elegans.* Cell, 1990. **63**(5): p. 895-905.

76. Pereira, L., et al., *A cellular and regulatory map of the cholinergic nervous system of C. elegans.* Elife, 2015. **4**.

77. Leyva-Diaz, E., et al., *Brn3/POU-IV-type POU homeobox genes-Paradigmatic regulators of neuronal identity across phylogeny.* Wiley Interdiscip Rev Dev Biol, 2020: p. e374.

78. Chalfie, M., H.R. Horvitz, and J.E. Sulston, *Mutations that lead to reiterations in the cell lineages of C. elegans.* Cell, 1981. **24**(1): p. 59-69.

79. Baumeister, R., Y. Liu, and G. Ruvkun, *Lineage-specific regulators couple cell lineage asymmetry to the transcription of the Caenorhabditis elegans POU gene unc-86 during neurogenesis.* Genes Dev, 1996. **10**(11): p. 1395-410.

80. Brown, N.L., et al., *Math5 is required for retinal ganglion cell and optic nerve formation.* Development, 2001. **128**(13): p. 2497-508.

81. Costa, A., et al., *Atoh1 in sensory hair cell development: constraints and cofactors.* Semin Cell Dev Biol, 2017. **65**: p. 60-68.

82. Edwards, S.L., et al., *A novel molecular solution for ultraviolet light detection in Caenorhabditis elegans.* PLoS Biol, 2008. **6**(8): p. e198.

83. Gong, J., et al., *The C. elegans Taste Receptor Homolog LITE-1 Is a Photoreceptor.* Cell, 2016. **167**(5): p. 1252-1263 e10.

84. Ayer, R.K., Jr. and J. Carlson, *acj6: a gene affecting olfactory physiology and behavior in Drosophila.* Proc Natl Acad Sci U S A, 1991. **88**(12): p. 5467-71.

85. Bai, L., A.L. Goldman, and J.R. Carlson, *Positive and negative regulation of odor receptor gene choice in Drosophila by acj6.* J Neurosci, 2009. **29**(41): p. 12940-7.

86. Treacy, M.N., et al., *Twin of I-POU: a two amino acid difference in the I-POU homeodomain distinguishes an activator from an inhibitor of transcription.* Cell, 1992. **68**(3): p. 491-505.

87. Treacy, M.N., X. He, and M.G. Rosenfeld, *I-POU: a POU-domain protein that inhibits neuron-specific gene activation.* Nature, 1991. **350**(6319): p. 577-84.

88. Komiyama, T., et al., *From lineage to wiring specificity. POU domain transcription factors control precise connections of Drosophila olfactory projection neurons.* Cell, 2003. **112**(2): p. 157-67.

89. Silver, S.J. and I. Rebay, *Signaling circuitries in development: insights from the retinal determination gene network.* Development, 2005. **132**(1): p. 3-13.

90. Kumar, J.P., *The molecular circuitry governing retinal determination.* Biochim Biophys Acta, 2009. **1789**(4): p. 306-14.

91. Kenny, N.J., et al., *Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs.* Heredity, 2016. **116**(2): p. 190-199.

92. Kiecker, C., T. Bates, and E. Bell, *Molecular specification of germ layers in vertebrate embryos.* Cellular and Molecular Life Sciences, 2016. **73**(5): p. 923-947.

93. Howard-Ashby, M., et al., *Identification and characterization of homeobox transcription factor genes in Strongylocentrotus purpuratus, and their expression in embryonic development.* Developmental biology, 2006. **300**(1): p. 74-89.

94. Howard-Ashby, M., et al., *High regulatory gene use in sea urchin embryogenesis: Implications for bilaterian development and evolution.* Dev Biol, 2006. **300**(1): p. 27-34.

95. Wei, Z., R.C. Angerer, and L.M. Angerer, *A database of mRNA expression patterns for the sea urchin embryo.* Dev Biol, 2006. **300**(1): p. 476-84.

96. Valero-Gracia, A., et al., *Non-directional photoreceptors in the pluteus of Strongylocentrotus purpuratus.* Frontiers in Ecology and Evolution, 2016. **4**: p. 127.

97. Kudtarkar, P. and R.A. Cameron, *Echinobase: an expanding resource for echinoderm genomic information.* Database (Oxford), 2017. **2017**.

98. Ullrich-Luter, E.M., S. D'Aniello, and M.I. Arnone, *C-opsin expressing photoreceptors in echinoderms.* Integr Comp Biol, 2013. **53**(1): p. 27-38.

99. Ullrich-Luter, E.M., et al., *Unique system of photoreceptors in sea urchin tube feet.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8367-72.

100. Mao, C.A., et al., *Substituting mouse transcription factor Pou4f2 with a sea urchin orthologue restores retinal ganglion cell development.* Proc Biol Sci, 2016. **283**(1826): p. 20152978.

101. Diaz-Balzac, C.A., et al., *Holothurian Nervous System Diversity Revealed by Neuroanatomical Analysis.* PLoS One, 2016. **11**(3): p. e0151129.

102. Liu, W., et al., *All Brn3 genes can promote retinal ganglion cell differentiation in the chick.* Development, 2000. **127**(15): p. 3237-47.

103. Combes, R.D., *New Measures on Animal Experimentation in the UK Will Improve Animal Welfare and Scientific Research.* Altern Lab Anim, 1999. **27**(3): p. 309-16.

104. Desai, C., et al., *A genetic pathway for the development of the Caenorhabditis elegans HSN motor neurons.* Nature, 1988. **336**(6200): p. 638-46.

105. Lloret-Fernandez, C., et al., *A transcription factor collective defines the HSN serotonergic neuron regulatory landscape.* Elife, 2018. **7**.

106. Stolfi, A., et al., *Neural tube patterning by Ephrin, FGF and Notch signaling relays.* Development, 2011. **138**(24): p. 5429-39.

107. Sampath, K. and G.W. Stuart, *Developmental expression of class III and IV POU domain genes in the zebrafish.* Biochem Biophys Res Commun, 1996. **219**(2): p. 565-71.

108. Horie, R., et al., *Shared evolutionary origin of vertebrate neural crest and cranial placodes.* Nature, 2018. **560**(7717): p. 228-232.

109. Stolfi, A., et al., *Migratory neuronal progenitors arise from the neural plate borders in tunicates.* Nature, 2015. **527**(7578): p. 371-4.

110. Sharma, S., W. Wang, and A. Stolfi, *Single-cell transcriptome profiling of the Ciona larval brain.* Dev Biol, 2019. **448**(2): p. 226-236.

111. Rigon, F., et al., *Developmental signature, synaptic connectivity and neurotransmission are conserved between vertebrate hair cells and tunicate coronal cells.* J Comp Neurol, 2018. **526**(6): p. 957-971.

112. Satou, Y., et al., *An integrated database of the ascidian, Ciona intestinalis: towards functional genomics.* Zoolog Sci, 2005. **22**(8): p. 837-43.

113. Xiang, M., et al., *Requirement for Brn-3c in maturation and survival, but not in fate determination of inner ear hair cells.* Development, 1998. **125**(20): p. 3935-46.

114. Xiang, M., et al., *Essential role of POU-domain factor Brn-3c in auditory and vestibular hair cell development.* Proc Natl Acad Sci U S A, 1997. **94**(17): p. 9445-50.

115. Trieu, M., et al., *Direct autoregulation and gene dosage compensation by POU-domain transcription factor Brn3a.* Development, 2003. **130**(1): p. 111-21.

116. Zou, M., et al., *Brn3a/Pou4f1 regulates dorsal root ganglion sensory neuron specification and axonal projection into the spinal cord.* Dev Biol, 2012. **364**(2): p. 114-27.

117. Xiang, M., et al., *Requirement for Brn-3c in maturation and survival, but not in fate determination of inner ear hair cells.* Development, 1998. **125**(20): p. 3935-3946.

118. Rasheed, V.A., et al., *Developmental wave of Brn3b expression leading to RGC fate specification is synergistically maintained by miR-23a and miR-374.* Dev Neurobiol, 2014. **74**(12): p. 1155-71.

119. Shi, M., et al., *Genetic interactions between Brn3 transcription factors in retinal ganglion cell type specification.* PLoS One, 2013. **8**(10): p. e76347.

120. Fedtsova, N.G. and E.E. Turner, *Brn-3.0 expression identifies early post-mitotic CNS neurons and sensory neural precursors.* Mech Dev, 1995. **53**(3): p. 291-304.

121. Philips, G.T., et al., *Precocious retinal neurons: Pax6 controls timing of differentiation and determination of cell type.* Dev Biol, 2005. **279**(2): p. 308-21.

122. Xiang, M., et al., *Brn-3b: a POU domain gene expressed in a subset of retinal ganglion cells.* Neuron, 1993. **11**(4): p. 689-701.

123. Hirsch, N. and W.A. Harris, *Xenopus Brn-3.0, a POU-domain gene expressed in the developing retina and tectum. Not regulated by innervation.* Invest Ophthalmol Vis Sci, 1997. **38**(5): p. 960-9.

124. Ding, Y., et al., *Spemann organizer transcriptome induction by early beta-catenin, Wnt, Nodal, and Siamois signals in Xenopus laevis.* Proc Natl Acad Sci U S A, 2017. **114**(15): p. E3081-E3090.

125. Session, A.M., et al., *Genome evolution in the allotetraploid frog Xenopus laevis.* Nature, 2016. **538**(7625): p. 336-343.

126. Badea, T.C., et al., *Distinct roles of transcription factors brn3a and brn3b in controlling the development, morphology, and function of retinal ganglion cells.* Neuron, 2009. **61**(6): p. 852-64.

127. Wang, S.W., et al., *Brn3b/Brn3c double knockout mice reveal an unsuspected role for Brn3c in retinal ganglion cell axon outgrowth.* Development, 2002. **129**(2): p. 467-77.

128. Keeler, C.E., *Iris movements in blind mice.* American Journal of Physiology-Legacy Content, 1927. **81**(1): p. 107-112.

129. Badea, T.C., et al., *Distinct roles of transcription factors brn3a and brn3b in controlling the development, morphology, and function of retinal ganglion cells.* Neuron, 2009. **61**(6): p. 852-864.

130. Badea, T.C. and J. Nathans, *Morphologies of mouse retinal ganglion cells expressing transcription factors Brn3a, Brn3b, and Brn3c: analysis of wild type and mutant cells using genetically-directed sparse labeling.* Vision research, 2011. **51**(2): p. 269-279.

131. Halluin, C., et al., *Habenular Neurogenesis in Zebrafish Is Regulated by a Hedgehog, Pax6 Proneural Gene Cascade.* PLoS One, 2016. **11**(7): p. e0158210.

132. Xiang, M., et al., *Targeted deletion of the mouse POU domain gene Brn-3a causes selective loss of neurons in the brainstem and trigeminal ganglion, uncoordinated limb movement, and impaired suckling.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11950-5.

133. Xiang, M., et al., *Role of the Brn-3 family of POU-domain genes in the development of the auditory/vestibular, somatosensory, and visual systems.* Cold Spring Harb Symp Quant Biol, 1997. **62**: p. 325-36.

134. Dykes, I.M., et al., *Brn3a and Islet1 act epistatically to regulate the gene expression program of sensory differentiation.* J Neurosci, 2011. **31**(27): p. 9789-99.

135. Riddiford, N. and G. Schlosser, *Dissecting the pre-placodal transcriptome to reveal presumptive direct targets of Six1 and Eya1 in cranial placodes.* Elife, 2016. **5**.

136. Sato, T., et al., *Genetic single-cell mosaic analysis implicates ephrinB2 reverse signaling in projections from the posterior tectum to the hindbrain in zebrafish.* J Neurosci, 2007. **27**(20): p. 5271-9.

137. Dyer, C., et al., *Specification of sensory neurons occurs through diverse developmental programs functioning in the brain and spinal cord.* Dev Dyn, 2014. **243**(11): p. 1429-39.

138. Sajgo, S., et al., *Dynamic expression of transcription factor Brn3b during mouse cranial nerve development.* J Comp Neurol, 2016. **524**(5): p. 1033-61.

139. Rohrschneider, M.R., G.E. Elsen, and V.E. Prince, *Zebrafish Hoxb1a regulates multiple downstream genes including prickle1b.* Dev Biol, 2007. **309**(2): p. 358-72.

140. Ninkina, N.N., et al., *A novel Brn3-like POU transcription factor expressed in subsets of rat sensory and spinal cord neurons.* Nucleic Acids Res, 1993. **21**(14): p. 3175-82.

141. Deng, M., et al., *Comparative expression analysis of POU4F1, POU4F2 and ISL1 in developing mouse cochleovestibular ganglion neurons.* Gene Expr Patterns, 2014. **15**(1): p. 31-7.

142. Barta, C.L., et al., *RNA-seq transcriptomic analysis of adult zebrafish inner ear hair cells.* Sci Data, 2018. **5**: p. 180005.

143. Xiao, T., et al., *A GFP-based genetic screen reveals mutations that disrupt the architecture of the zebrafish retinotectal projection.* Development, 2005. **132**(13): p. 2955-67.

144. McEvilly, R.J., et al., *Requirement for Brn-3.0 in differentiation and survival of sensory and motor neurons.* Nature, 1996. **384**(6609): p. 574-7.

145. Erkman, L., et al., *Role of transcription factors Brn-3.1 and Brn-3.2 in auditory and visual system development.* Nature, 1996. **381**(6583): p. 603-6.

146. Lin, Y.H., et al., *A novel missense variant in the nuclear localization signal of POU4F3 causes autosomal dominant non-syndromic hearing loss.* Sci Rep, 2017. **7**(1): p. 7551.

147. Farooqui-Kabir, S.R., et al., *Cardiac expression of Brn-3a and Brn-3b POU transcription factors and regulation of Hsp27 gene expression.* Cell Stress Chaperones, 2008. **13**(3): p. 297-312.

148. Simmons, A.B., et al., *Pou4f2 knock-in Cre mouse: A multifaceted genetic tool for vision researchers.* Molecular vision, 2016. **22**: p. 705.

149. Budhram-Mahadeo, V., et al., *Co-expression of POU4F2/Brn-3b with p53 may be important for controlling expression of pro-apoptotic genes in cardiomyocytes following ischaemic/hypoxic insults.* Cell Death Dis, 2014. **5**: p. e1503.

150. Cardoso-Moreira, M., et al., *Gene expression across mammalian organ development.* Nature, 2019. **571**(7766): p. 505-509.

151. Sajgo, S., et al., *Dre - Cre sequential recombination provides new tools for retinal ganglion cell labeling and manipulation in mice.* PLoS One, 2014. **9**(3): p. e91435.

152. Tao, W., et al., *Transcriptome display during tilapia sex determination and differentiation as revealed by RNA-Seq analysis.* BMC Genomics, 2018. **19**(1): p. 363.

153. Bar, I., S. Cummins, and A. Elizur, *Transcriptome analysis reveals differentially expressed genes associated with germ cell and gonad development in the Southern bluefin tuna (Thunnus maccoyii).* BMC Genomics, 2016. **17**: p. 217.

154. Budhram-Mahadeo, V., et al., *The closely related POU family transcription factors Brn-3a and Brn-3b are expressed in distinct cell types in the testis.* The international journal of biochemistry & cell biology, 2001. **33**(10): p. 1027-1039.

155. Schlosser, G., *A Short History of Nearly Every Sense-The Evolutionary History of Vertebrate Sensory Cell Types.* Integr Comp Biol, 2018. **58**(2): p. 301-316.

156. Peterson, K.J. and N.J. Butterfield, *Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record.* Proceedings of the National Academy of Sciences, 2005. **102**(27): p. 9547-9552.

157. Leys, S.P., *Elements of a 'nervous system' in sponges.* J Exp Biol, 2015. **218**(Pt 4): p. 581-91.

158. Ryan, J.F. and M. Chiodin, *Where is my mind? How sponges and placozoans may have lost neural cell types.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2015. **370**(1684): p. 20150059.

159. Gazave, E., et al., *NK homeobox genes with choanocyte-specific expression in homoscleromorph sponges.* Development genes and evolution, 2008. **218**(9): p. 479.

160. King, N., et al., *The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans.* Nature, 2008. **451**(7180): p. 783-788.

161. NAKAJIMA, Y., *Development of the Nervous System of Sea Urchin Embryos: Formation of Ciliary Bands and the Appearance of Two Types of Ectoneural Cells in the Pluteus: (ectoneural cell/sea urchin pluteus/ciliary band/axoneme).* Development, growth & differentiation, 1986. **28**(6): p. 531-542.

162. Wood, N.J., et al., *Neuropeptidergic systems in pluteus larvae of the sea urchin Strongylocentrotus purpuratus: neurochemical complexity in a "simple" nervous system.* Frontiers in endocrinology, 2018. **9**: p. 628.

163. Sippl, C. and E.R. Tamm, *What is the nature of the RGC-5 cell line?* Adv Exp Med Biol, 2014. **801**: p. 145-54.

164. Mao, Y. and N. Satoh, *A Likely Ancient Genome Duplication in the Speciose Reef-Building Coral Genus, Acropora.* iScience, 2019. **13**: p. 20-32.

165. Schwager, E.E., et al., *The house spider genome reveals an ancient whole-genome duplication during arachnid evolution.* BMC biology, 2017. **15**(1): p. 1-27.

166. Liu, C., et al., *Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition.* bioRxiv, 2020.

167. Zadesenets, K.S., et al., *Genome and Karyotype Reorganization after Whole Genome Duplication in Free-Living Flatworms of the Genus Macrostomum.* International journal of molecular sciences, 2020. **21**(2): p. 680.

168. Zadesenets, K.S., L. Schärer, and N.B. Rubtsov, *New insights into the karyotype evolution of the free-living flatworm Macrostomum lignano (Platyhelminthes, Turbellaria).* Scientific reports, 2017. **7**(1): p. 1-9.

169. Li, Z., et al., *Multiple large-scale gene and genome duplications during the evolution of hexapods.* Proceedings of the National Academy of Sciences, 2018. **115**(18): p. 4713-4718.

170. Julca, I., et al., *Phylogenomics identifies an ancestral burst of gene duplications predating the diversification of Aphidomorpha.* Molecular biology and evolution, 2020. **37**(3): p. 730-756.

171. Hislop, N.R., et al., *Tandem organization of independently duplicated homeobox genes in the basal cnidarian Acropora millepora.* Development genes and evolution, 2005. **215**(5): p. 268-273.

172. Torras, R., et al., *nanos expression at the embryonic posterior pole and the medusa phase in the hydrozoan Podocoryne carnea.* Evolution & development, 2004. **6**(5): p. 362-371.

173. Smith, J.J., et al., *Publisher Correction: The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution.* Nat Genet, 2018. **50**(11): p. 1617.

174. Holland, L.Z. and D. Ocampo Daza, *A new look at an old question: when did the second whole genome duplication occur in vertebrate evolution?* Genome Biol, 2018. **19**(1): p. 209.

175. Caputo Barucchi, V., et al., *Genome duplication in early vertebrates: insights from agnathan cytogenetics.* Cytogenet Genome Res, 2013. **141**(2-3): p. 80-9.

176. Simakov, O., et al., *Deeply conserved synteny resolves early events in vertebrate evolution.* Nature Ecology & Evolution, 2020: p. 1-11.

177. Ludwig, A., et al., *Genome duplication events and functional reduction of ploidy levels in sturgeon (Acipenser, Huso and Scaphirhynchus).* Genetics, 2001. **158**(3): p. 1203-1215.

178. Glasauer, S.M. and S.C. Neuhauss, *Whole-genome duplication in teleost fishes and its evolutionary consequences.* Mol Genet Genomics, 2014. **289**(6): p. 1045-60.

179. Roux, J., J. Liu, and M. Robinson-Rechavi, *Selective Constraints on Coding Sequences of Nervous System Genes Are a Major Determinant of Duplicate Gene Retention in Vertebrates.* Mol Biol Evol, 2017. **34**(11): p. 2773-2791.

180. Evans, B.J., et al., *Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (Xenopus, Pipidae) from West and Central Africa.* PLoS One, 2015. **10**(12).

181. Schmid, M., B.J. Evans, and J.P. Bogart, *Polyploidy in Amphibia.* Cytogenet Genome Res, 2015. **145**(3-4): p. 315-30.

182. Evans, B.J., *Ancestry influences the fate of duplicated genes millions of years after polyploidization of clawed frogs (Xenopus).* Genetics, 2007. **176**(2): p. 1119-1130.

183. Braasch, I. and M. Schartl, *Evolution of endothelin receptors in vertebrates.* Gen Comp Endocrinol, 2014. **209**: p. 21-34.

184. Braasch, I., J.N. Volff, and M. Schartl, *The endothelin system: evolution of vertebrate-specific ligand-receptor interactions by three rounds of genome duplication.* Mol Biol Evol, 2009. **26**(4): p. 783-99.

185. Yue, J.X., K.L. Li, and J.K. Yu, *Discovery of germline-related genes in Cephalochordate amphioxus: A genome wide survey using genome annotation and transcriptome data.* Mar Genomics, 2015. **24 Pt 2**: p. 147-57.

186. Danks, G., et al., *OikoBase: a genomics and developmental transcriptomics resource for the urochordate Oikopleura dioica.* Nucleic Acids Res, 2013. **41**(Database issue): p. D845-53.

187. Papatheodorou, I., et al., *Expression Atlas: gene and protein expression across multiple studies and organisms.* Nucleic acids research, 2018. **46**(D1): p. D246-D251.

188. Arendt, D. and J. Wittbrodt, *Reconstructing the eyes of Urbilateria.* Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 2001. **356**(1414): p. 1545-1563.

189. Lacalli, T.C., *Sensory systems in amphioxus: a window on the ancestral chordate condition.* Brain, behavior and evolution, 2004. **64**(3): p. 148-162.

190. Pergner, J. and Z. Kozmik, *Amphioxus photoreceptors - insights into the evolution of vertebrate opsins, vision and circadian rhythmicity.* Int J Dev Biol, 2017. **61**(10-11-12): p. 665-681.

191. Kusakabe, T., et al., *Ci-opsin1, a vertebrate-type opsin gene, expressed in the larval ocellus of the ascidian Ciona intestinalis.* FEBS Lett, 2001. **506**(1): p. 69-72.

192. Kusakabe, T. and M. Tsuda, *Photoreceptive systems in ascidians.* Photochemistry and photobiology, 2007. **83**(2): p. 248-252.

193. Bult, C.J., et al., *Mouse Genome Database (MGD) 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D801-D806.

194. Sokpor, G., et al., *Transcriptional and Epigenetic Control of Mammalian Olfactory Epithelium Development.* Mol Neurobiol, 2018. **55**(11): p. 8306-8327.

195. Hardwick, L.J. and A. Philpott, *xNgn2 induces expression of predominantly sensory neuron markers in Xenopus whole embryo ectoderm but induces mixed subtype expression in isolated ectoderm explants.* Wellcome Open Research, 2018. **3**(144): p. 144.

196. Schneider, M.L., D.L. Turner, and M.L. Vetter, *Notch signaling can inhibit Xath5 function in the neural plate and developing retina.* Molecular and Cellular Neuroscience, 2001. **18**(5): p. 458-472.

197. Jung, H., et al., *The Ancient Origins of Neural Substrates for Land Walking.* Cell, 2018. **172**(4): p. 667-682 e15.

198. Martik, M.L., et al., *Evolution of the new head by gradual acquisition of neural crest regulatory circuits.* Nature, 2019. **574**(7780): p. 675-678.

199. Nomura, T., et al., *Reptiles: a new model for brain evo-devo research.* Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 2013. **320**(2): p. 57-73.

200. Mu, X. and W.H. Klein. *A gene regulatory hierarchy for retinal ganglion cell specification and differentiation*. in *Seminars in cell & developmental biology*. 2004. Elsevier.

201. Lindeberg, J., et al., *Identification of a chicken homologue in the Brn-3 subfamily of POU-transcription factors.* Developmental brain research, 1997. **100**(2): p. 169-182.

202. McEvilly, R.J., et al., *Requirement for Brn-3.0 in differentiation and survival of sensory and motor neurons.* Nature, 1996. **384**(6609): p. 574-577.

203. Notredame, C., D.G. Higgins, and J. Heringa, *T-Coffee: A novel method for fast and accurate multiple sequence alignment.* J Mol Biol, 2000. **302**(1): p. 205-17.

204. Wallace, I.M., et al., *M-Coffee: combining multiple sequence alignment methods with T-Coffee.* Nucleic acids research, 2006. **34**(6): p. 1692-1699.

205. Harris, T.W., et al., *WormBase: a modern Model Organism Information Resource.* Nucleic Acids Res, 2019.

206. Thurmond, J., et al., *FlyBase 2.0: the next generation.* Nucleic Acids Res, 2019. **47**(D1): p. D759-D765.

CHAPTER 1

Efficient genome editing using CRISPR-Cas-mediated homology directed repair in the ascidian *Ciona robusta*

ABSTRACT

Eliminating or silencing a gene's level of activity, whether by antisense morpholinos, TALE nucleases or RNAi, is one of the classic approaches developmental biologists employ to determine a gene's function. A recently developed method of gene perturbation called CRISPR-Cas, which was derived from a prokaryotic adaptive immune system that protects against invading viral DNA, has been adapted for use in eukaryotic cells. This technology has been established in several model organisms as a powerful and efficient tool for knocking out or knocking down the function of a gene of interest. The system requires the coexpression of Cas9 nuclease and an artificial guide RNA (gRNA) that in combination targets and cleaves a genomic sequence of interest and introduces an indel-forming double-strand break. It has been recently shown that CRISPR-Cas functions with fidelity and efficiency in *Ciona robusta*.

Here, we show that in *C. robusta* CRISPR-Cas mediated genomic knock-ins can be efficiently generated. Electroporating a tissue-specific transgene driving Cas9 and a U6-driven gRNA transgene together with a fluorescent protein-containing homology directed repair (FP-HDR) template results in gene-specific patterns of fluorescence consistent with a targeted genomic insertion. Using the *Tyrosinase* locus to optimize reagents, we first examine homology arm-length efficiencies of FP-HDR templates. Next, we characterize a new Pol III promoter for expressing gRNAs from the *C. savignyi* H1 gene, and then adapt technology that flanks gRNAs by ribozymes allowing cell-specific expression from Pol II promoters. These novel methods of expressing gRNAs in ascidians are more effective than the U6 promoter. Reagents were then developed for targeting *Brachyury* and *Pou4* that resulted in expected patterns of fluorescence, and sequenced PCR products derived from single embryos validated predicted genomic

insertions. Finally, using two differentially colored FP-HDR templates, we show that biallelic FP-HDR template insertion can be detected in live embryos of the F0 generation.

Keywords: ascidian, Ciona, CRISPR-Cas9, homology-directed repair, fluorescent protein, genome editing

INTRODUCTION

In the few years since its development, CRISPR-Cas technology has proved to be an important biotechnological advancement used for a wide range of applications including genomic knock-outs [1], transcriptional repression [2], and genomic knock-ins [3, 4]. Adapted from a prokaryotic adaptive immune system, CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated) relies on a *Streptococcus pyogenes* Cas9 nuclease that pairs with an artificial single guide RNA molecule that contains a genomic target of 17-20 bp that provides sequence-specific nuclease activity to the Cas9-gRNA complex. CRISPR-Cas's baseline effect, double-strand breaks that result in insertion-deletion mutations, has been demonstrated in nearly every experimental model where the components can be delivered [5]. In the ascidian *Ciona robusta* (formerly *C. intestinalis* species A), CRISPR-Cas has been shown to efficiently cleave genes of interest via injection of Cas9 mRNA and gRNAs [6] and, through electroporation of plasmid transgene components [7, 8].

When CRISPR-Cas is combined with a homology directed repair (HDR) template, it is possible to generate precise genomic edits and sequence insertions by taking advantage of endogenous cellular HDR mechanisms [9]. Introduced sequences of DNA that include stretches of genomic homology flanking a designated mutation or insertion are used by the cell as a template to repair the double-strand break created by the nuclease. Single-strand oligodeoxynucleotides (ssODNs) between 100-200 bp with homology arms of ~50 bp have been shown to effectively introduce single bp changes, or insert a FLAG-tag sequence [10]. For larger modifications, e.g. the insertion of a fluorescent protein (FP) sequence (~750 nt), homology arms of the insert size or larger are more effective [11].

Here, we take advantage of our ability to introduce DNA constructs into thousands of fertilized *Ciona* eggs via electroporation to develop a transgene-based approach to perform precise genomic modifications using CRISPR-Cas transgenes and transgenes containing FP-HDR templates. Using the melanocyte gene *Tyrosinase* (*Tyr*), in which introduced phenotypes are easily scorable, we describe the generation of FP-HDR reagents and discuss ways to optimize template-based HDR editing. Embryos and larvae that expressed fluorescent proteins in the predicted cells, a result of correct HDR-mediated insertions, were identified with fluorescent microscopy *in vivo.* Importantly, no cells were detected that expressed fluorescent proteins in control experiments in which gRNA transgenes were omitted suggesting that the editing process is precise and requires all editing reagents to be present. After demonstrating that HDR-mediated editing was possible, we analyzed how altering homology arm length affects HDR efficiency. Next, we generated an alternative RNA polymerase (pol) III promoter from the *Ciona savignyi* H1 (CsH1) gene that effectively facilitated HDR in the melanocytes. Although the H1 promoter is routinely used in vertebrates, its use in ascidians has not yet been reported. Furthermore, we adapted a technique that incorporates ribozymes to enable the expression of gRNAs with tissue-specific RNA polymerase II promoters and show that this technique is efficient at eliciting FP-HDR genomic insertions. To show that our CRISPR-Cas mediated FP-HDR genomic insertions are broadly applicable in *Ciona* and not specific to the *Tyr* locus, we demonstrate FP insertions at two additional loci: *Brachyury* (*Bra*), which is expressed exclusively in the mesoderm, and *Pou4*, which is specific to the peripheral nervous system (PNS). Confirmation of fluorescent protein insertions at the desired genomic locus was obtained by sequencing PCR amplicons, either from pools of a small number of embryos or from individual embryos. Finally, we used two

differently colored FP-HDR templates to show that both alleles in the same cell can be targeted for HDR-mediated genomic insertions simultaneously.

MATERIALS AND METHODS

Adult *C. robusta* were collected from marinas in Mission Bay, San Diego, and kept in a recirculating refrigerated tank under constant illumination to allow the accumulation of gametes. Varying amounts of transgene DNAs used for electroporations were pooled together, ethanol precipitated and the pellets dissolved in 0.77M Mannitol. Electroporations were carried out essentially as in Zeller, Virata [12] with no relevant modifications, and electroporated embryos were left to develop at 16-18°C in 0.45μm-filtered seawater supplemented with penicillin/streptomycin (10 U/mL and 10μg/mL, respectively) for the desired amount of time of each experiment. Embryos required to develop past mid tailbud stage were anesthetized with MS-222 (Sigma) at 12-15hpf. A biological replicate refers to an independent zap on a different day.

Immunohistochemistry was performed as in Joyce Tang, Chen [13]. Briefly, embryos were fixed in 2% paraformaldehyde, washed in PBST and briefly permeabilized in 100% methanol. Embryos were incubated at 4°C overnight with primary antibody (Rabbit-anti-GFP, Thermo Fisher Scientific, A11122; Mouse anti-acetylated tubulin antibody, Sigma, T7451) at 1:2000, washed thoroughly and incubated at 4°C overnight with secondary antibody (Alexa Fluor 546, A11003; Alexa Fluor 488, A-11001: Thermo Fisher Scientific,) at 1:500. Embryos were washed thoroughly, incubated for 10min in 50% glycerol +300nM DAPI, and then transferred to 70% glycerol for imaging.

Molecular cloning

All transgenes used in this study were cloned into our custom vector, pZapANX (pZANX), which is a pSP72 plasmid with a cryptic mesenchymal promoter sequence deletion (pSP72-1.27; Zeller et al 2006). Cas9 and the gRNA were obtained from pX330-U6-Chimeric_BB-CBh-hSpCas9 [14], Addgene (#42230). F+E gRNA modifications were performed to the gRNA backbone as described in Chen, Gilbert [15]. All *C. robusta* and *C. savignyi* fragments were amplified using genomic DNA isolated from sperm. PCR products were generated with Q5 DNA polymerase from New England Biolabs using standard PCR conditions.

Cloning Cas9 transgenes

Briefly, *cis*-regulatory DNA was amplified from genomic DNA, digested and purified, and ligated with the Cas9 coding region and pZANX to create the Cas9 transgenes. *Tyr* (KH.C12.469) *cis*-regulatory DNA was amplified with 5`-CGCTCTAGAACCATTTTTCATGCCTGACAAATGGCGG -3` and 5`-CGCGGTACCATGGTTGGAAGTGCTAGACAAATGTGTG -3`. *Fkh* (KH.C8.396) *cis*-regulatory DNA was generated with 5`-GCGGCTAAGTCGGCAGTTTCCAAG-3` and 5`-GACAACATCATTTTTGTACTTGCAAAAGTAATCC-3`. *EpiB* (KH.C7.154) *cis*-regulatory DNA was amplified with 5`-GTTTATACATTGCAGTCAAACGAAGCACC-3` and 5`-CGTTTTCCGTAATTAAAATAGTACCATG-3`. *SoxB1* (KH.C1.99) *cis*-regulatory DNA was amplified with 5`-AGCGTTGCGTGGCCTGCCTCCAGTCG-3`and 5`-CTGTAAAGTAGAAGTTTCAGTTAAAAATC-3`.

Cloning U6 and H1 gRNA transgenes

The putative Ciona H1 RNA was first identified from a computational screen for non-coding RNAs [16] and has also been annotated in the Ensembl Ciona genome browser as gene ENSCINT00000030147. U6::gRNA and H1::gRNA transgenes were constructed by first amplifying the Cr*U6* (Fwd: 5`-TGTACCACCACGTCGCAGACG-3`; Rev 5`-GATGGGTCTTCGAGAAGACCTGTTTAAGAGCTATGCTGGAAACAGCATAG-3`) and Cs*H1* (Fwd: 5`-CTAAAATAAATGTACAGACCTTGGTCATG-3`; Rev: 5`-GCCTGTACTAAATACAAAAGATAGTGTTTTTTAACCGGGTCTTCGAGAAGACCTG-3`) promoters from *C. robusta* and *C. savignyi* genomic DNA, respectively. The gRNA sequence was amplified from pX330 with forward primers (U6: 5`-CTGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCG-3`; H1: 5`-CCGGGTCTTCGAGAAGACCTGTTTAAGAGCTATGC-3`) and a common reverse primer (5`-GGCACCGAGTCGGTGCTTTTTTGTTTTAGAGGATTCGCG-3`). Overlap PCR was performed to create the empty U6::gRNA and H1::gRNA vectors into which desired targets could be cloned.

gRNA design began by looking for suitable targets that were closest to the desired point of YFP insertion. We performed a BLAST search of each potential target sequence+NGG against the *C. robusta* genome; targets with potential off-target binding were not considered for use. Targets were added to the U6/H1::gRNA transgenes by BbsI digestion followed by insertion of a phosphotased, complementary pair of annealed oligos that included BbsI overhangs as in Gandhi, Haeussler [8]. Using the CRISPOR website, http://crispor.tefor.net [17], we assessed specificity and efficacy scores and are reported in the following table.

gene targeted	gRNA sequence used in this study	specificity score	Doench '16 efficiency	Mor.-Mateos efficiency
Tyr	TGTGGGAGAAAGTGAAATCC	91	51	32

Bra-Cod	CGTCACCGGGGGTCATGCAA	not found	65	62
Bra-UTR	CATGATAGATCGAGTGTTTA	99	20	33
Pou4	GTTTCTTCAATTTCGGTGCG	100	59	34

To generate Tyr::RZ-gRNA$_{(Tyr)}$-RZ first we used a primer set (Fwd: 5`-GTGAGGACGAAACGAGTAAGCTCGTCGGGTCTTCGAGAAGACCTGTTTAAG-3`; Rev: 5`-GGCACCGAGTCGGTGCTTTTTTGGCCGGCATGGTCCCAGCCTCCTCGCTGG-3`) to amplify the gRNA sequence from our empty U6::gRNA adding ribozyme sequence to either end. We then performed a second PCR using a second primer set (Fwd: 5`-GTGAGGACGAAACGAGTAAGCTCGTCGGGTCTTCGAGAAGACCTGTTTAAG-3`; Rev: 5`- GGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGG-3`) using the previous PCR product as template to add additional ribozyme sequence. This fragment was cloned into pZANX. We then digested with Bbs I and inserted the annealed *Tyr* target oligos as described above. This incomplete transgene was used as a template for the final PCR reaction, which amplified the final fragment using a third primer set (Fwd: 5`-CCCACACTGATGAGTCCGTGAGGACGAAACG-3`; Rev: 5`-GGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC-3`). This final PCR added the remaining ribozyme sequence. We then combined this fragment with the Tyr cis-reg element and cloned them into pZANX.

Cloning FP-HDR transgenes

The Tyr FP-HDR$_{(2.0\ kB)}$ 5` arm was amplified with the following primers: 5`-CGTCAGCAGGTTATTCCCTGACGCAG-3` and 5`-GCAGCCGTTGACGTCATTACGTTACC-3`, and the 3` arm was amplified with the following

primers: 5`- CCTGGTTCGTGAACTTGGCCCG -3` and 5`-GTGTCATGACGTCACTTCGTTCTTCTCC -3`. These fragments were cloned into pZANX upstream and downstream of GFP.

To generate Tyr FP-HDR$_{(1.0\ kB)}$ and Tyr FP-HDR$_{(500\ bp)}$, we used Tyr FP-HDR$_{(2.0\ kB)}$ as a template for PCR. A fragment that included 1.0 kB homology arms 5` and 3` of GFP was generated with 5`- GGTGTCGCAGGTGAGCACAATAAC -3` and 5`-GCAATGCTACGTGTATCTAGCCG -3, and a fragment that included 500 bp homology arms 5` and 3` of GFP was generated with 5`- CCTATATCTGCACTGTAATGTATGCG -3` and 5`-CAATGAAGTATAAATGTGTTCTAACATAC -3. The fragments were each cloned into pZANX. Removing GFP from Tyr FP-HDR$_{(1.0\ kB)}$ and inserting either the CFP or mCherry sequences generated Tyr FP-HDR$_{(CFP)}$ and Tyr FP-HDR$_{(mCherry)}$.

To generate Bra FP-HDR, the 5` arm was amplified with a primer set (5`-CAGGAGACAAGGTTCATTGCTGTTACTGC-3` and 5`-CGCCGCTTACGCCACCTTCTTTG-3) that amplified 1.3kB upstream of the last codon. PAM mutations to the last *Bra* coding exon were introduced to the fragment with an additional primer set (5`-GGGCGTCATGCAACGTCCACAAGAAG-3` and 5`-GGGCGTCATGCAACGTCCACAAGAAG-3). The 3` arm was amplified (Fwd: 5`-TGACGTCACAATGCGAATATAATTATCG-3` and Rev: 5`-CCCAACACCCTCCATCATACATTCAA-3`) and UTR PAM mutations (27nt downstream of the stop codon) introduced with primers (Fwd: 5`-CAAAGGTTGATTGTCATGATAGATCGAGTGTTTAACCCATGTTCAAGG-3` and Rev: 5`-CGTCAACTTACAAACGTGACCAAACCTTGAACATGGGTTAAACACTCG-3`). The 5`

fragment was cloned upstream of YFP and the 3` arm was cloned downstream of YFP into pZANX.

To generate Pou4 FP-HDR, the 5` arm was amplified with a primer set (5`-CACCCACCCCGCCCCAACC -3` and 5`-ACATAATCACGTCCTGAAATATTAAATAAATAAAC -3) amplifying 1.3kB upstream of the last codon. The 3` arm was amplified (5`- TAGAATATGCTGTGACCTCTTCAATGC -3` and 5`- ACCGAGTCGTTAATATTTCATTTTTTAATG -3) and UTR mutations were introduced with primers (Fwd: 5`-GCACTTTTCACTTCTTCTTGTTTCTTCAATTTCGGTGCGACCTCG-3` and Rev: 5`-GGAACTGAATAGAAACTTGAAGTCGAGGTCGCACCGAAATTGAAGAAAC-3). The 5` fragment was cloned upstream of YFP and the 3` arm was cloned downstream of YFP into pZANX.

Sequencing genomic insertions

The Brachyury-YFP PCR fragment was amplified from pools of 10 YFP-positive embryos using a primer that annealed outside the Bra FP-HDR template (5'-CAGGAGACAAGGTTCATTGCTGTTACTGC-3') and a GFP AS primer (5'-TGACGTCACAATGCGAATATAATTATCG -3'). This fragment was used as a template for a second PCR with the GFP AS primer and a primer that was nested inside of the first amplicon yet still annealed outside of the Bra FP-HDR 5` arm of homology. The Pou4-YFP fragment was generated from single YFP-positive embryos with a primer that annealed outside of the Pou4 FP-HDR template (5'-CCGCCAACATTTAACTCGTCGGTGATGATG-3') and the GFP AS

primer. The Pou4-YFP and Bra-YFP fragments were cloned into pBS2KS+ and sequenced from both ends with T3 Fwd and T7 Rev primers.

RESULTS

To detect HDR in ascidian larvae, we designed our FP-HDR templates to have the fluorescent protein coding region in-frame with the coding region of *Tyr*, *Bra*, and *Pou4*. We hypothesized that if HDR occurred, then we would be able to detect fluorescence produced from the resulting fusion protein. FP-HDR templates were designed to fuse a fluorescent protein either in-frame with a truncated protein, *Tyr*, in which we predicted the function of the *Tyr* protein would be impaired or eliminated; or in-frame with a full-length protein to maintain its function, as with *Bra* and *Pou4*. Because we used fluorescence as our detection scheme, it was important to design HDR templates that lacked any functional *cis*-regulatory elements to reduce the possibility of false-positives. In all experiments, electroporation of a supercoiled FP-HDR plasmid template and a Cas9 transgene alone did not produce cells that expressed fluorescent protein.

The *Tyrosinase* locus can be efficiently targeted for FP-HDR

The *Ciona robusta tyrosinase* gene encodes the rate-limiting enzyme of the melanin synthesis pathway in the two larval melanocytes of the central nervous system (CNS), the ocellus and otolith. Disruption of the *Tyr* gene eliminates melanin synthesis and produces albino larvae [18, 19]. We first generated a transgene that drove ubiquitous expression of gRNAs with the *Ciona robusta* U6 promoter (CrU6) [18] that targeted the coding region of the *Tyr* locus (U6::gRNA$_{(Tyr)}$). We also generated a transgene that drove Cas9 expression exclusively in the

melanocytes by using 1.6 kB of *Tyr cis*-regulatory DNA (Tyr::Cas9). Electroporation of these constructs resulted in albino embryos, indicating that the *Tyr* locus was targeted and suggesting that the CRISPR-Cas9 system worked as expected. To test if we could precisely edit the genome by inserting a specific sequence into the *Tyr* locus, we constructed an HDR template that inserted a green fluorescent protein (GFP) coding sequence plus a stop codon amino-terminal to the Tyr transmembrane domain (Fig. 1A). This FP-HDR template (Tyr FP-HDR$_{(2.0\ kB)}$) contained 5` and 3` homology arms that were each 2.0 kB in length and upon successful recombination the resulting fusion was expected to produce a truncated, non-functioning Tyr protein that would generate GFP fluorescence in unpigmented melanocytes. We electroporated equal masses of Tyr::Cas9, U6::gRNA$_{(Tyr)}$, and Tyr FP-HDR$_{(2.0\ kB)}$ into fertilized eggs and cultured embryos to larval stage. We detected GFP expression in fixed larvae using an anti-GFP antibody due to the presence in the trunk of an autofluorescent signal in the GFP channel at relevant time points. The larval melanocytes provide a straightforward way to score the resulting FP insertions because in any electroporation, larvae had 0/2, 1/2, or 2/2 melanocytes expressing GFP (Fig. 1B). Of 308 larvae examined (4 biological replicates) 55.6% of the embryos (n=308) exhibited one fluorescing melanocyte while 16% had fluorescence in both melanocytes (28.4% of embryos counted displayed no fluorescing melanocytes). In control groups that lacked the gRNA plasmid, we scored one embryo having fluorescence in a pigment cell. These results demonstrated that the CRISPR-Cas system was effective and efficient at inserting a fluorescent protein sequence at the *Tyr* locus when expressed with an FP-HDR template.

Reducing the homology arm lengths reduces FP-HDR efficiency

It has been reported in other systems that reducing homology arm length below a certain limit can reduce the efficiency of HDR [11, 20]. To determine if homology arm length affected HDR efficiency in our system, we constructed two additional *Tyr* FP-HDR templates with homology arms of approximately 1.0 kB (Tyr FP-HDR$_{(1.0\ kB)}$), or 0.5 kB (Tyr FP-HDR$_{(500\ bp)}$) that targeted GFP to the same location in the *Tyr* locus (Fig. 1A). Electroporations were performed using an equal mass ratio of Tyr::Cas9, U6::gRNA$_{(Tyr)}$ and either Tyr FP-HDR$_{(2.0\ kB)}$, Tyr FP-HDR$_{(1.0\ kB)}$, or Tyr FP-HDR$_{(500\ bp)}$. Compared to the 2.0 kB control group, Tyr FP-HDR$_{(1.0\ kB)}$ produced approximately the same percentage (n=181; 2 biological replicates) of embryos expressing YFP in one or two pigment cells (58.3% and 17.2%, respectively). However, Tyr FP-HDR$_{(500\ bp)}$ (n=177; 2 biological replicates), only 50.3% expressed YFP in one pigment cell (8.4% decrease) and only 9.0% of embryos expressed YFP in both pigment cells (7.5% decrease). Represented as percentage of total melanocytes, Tyr FP-HDR$_{(2.0\ kB)}$ resulted in 42.2% of melanocytes expressing GFP, Tyr FP-HDR$_{(1.0\ kB)}$, resulted in 41.5% melanocytes expressing GFP, and Tyr FP-HDR$_{(500\ bp)}$ resulted in 32.1% of melanocytes expressing GFP (Fig. 1D, Supplementary Fig. 4A). These results were consistent with studies from other model systems that showed a dependence of HDR efficiency on homology arm length [11] and led us to construct all future FP-HDR templates with homology arm lengths of at least 1.0 kB.

The ascidian H1 promoter is effective at expressing gRNAs and increases rates of FP-HDR

In mammalian systems, the U6 RNA pol III and the H1 RNA pol III promoters are used for expressing short RNAs that do not require post-transcriptional processing, including gRNAs [21]. The H1 promoter generates the H1 RNA which is a component of the RNAse P complex [22]. We identified putative H1 promoters in *Ciona* (see materials and methods) and tested their

expression in embryos. Like their mammalian counterparts, the H1 promoters in the two *Ciona* species are short, only ~100bp, compared to the *Ciona* U6 promoter which is ~1.0 kB long. Because the H1 promoter is one-tenth the size of the U6 promoter, more copies of H1 transgenes are present compared to U6 transgenes for the same mass of DNA, which may result in higher levels of gRNA expression. The CsH1 promoter only required a single point mutation to introduce the appropriate restriction enzyme site for cloning gRNAs, compared to the *C. robusta* promoter; therefore we cloned the CsH1 promoter by amplifying 97 bp upstream of the *H1* sequence (Fig. 2A). We designed the H1 transgene to express the same *Tyr*-targeting gRNA sequence (H1::gRNA$_{(Tyr)}$) as the U6 transgene. We first compared the temporal and spatial activity of the CrU6 and CsH1 promoters by *in situ* hybridization and observed that while both promoters expressed gRNAs ubiquitously, transcripts from the CsH1 promoter were expressed a full cell cycle earlier than those from the CrU6 promoter (16-cells vs. 32-cells, respectively; data not shown). Like the U6 driver, electroporating H1::gRNA and Tyr::Cas9 resulted in albino embryos (Supplementary Fig. 1A). To examine efficacy of the CsH1 promoter in facilitating HDR, we electroporated equal masses of H1::gRNA$_{(Tyr)}$, Tyr::Cas9, and Tyr FP-HDR$_{(2.0\ kB)}$ and antibody labeled GFP at larval stage. Expressing gRNAs with this promoter increased the percentage of melanocytes expressing GFP (*n*=186 embryos; 3 biological replicates) from 42.0% to 46.3% (Fig. 2C, Supplementary Fig. 4B). In control electroporations that excluded H1::gRNA, we detected no fluorescent melanocytes.

Ribozyme-flanked gRNAs can be expressed tissue-specifically with RNA pol II promoters

In *Ciona*, RNA pol III promoters are not as well characterized as RNA pol II promoters, which often have detailed temporal and spatial expression data [23, 24]. To avoid concerns that

the pol III promoters might not express at high levels or at all in particular cell types of interest, and to prevent expressing guide RNAs in non-targeted cells, we adapted an approach that flanks the gRNAs by ribozymes (RZ) [25] and allows them to be expressed by tissue-specific RNA pol II type promoters. By flanking the gRNA with RZs that self-cleave at the exact 5` and 3` end of the gRNA sequence, the gRNA is consequently "cleaved-out" from the full RZ-gRNA-RZ transcript, thus freeing a precisely formed gRNA molecule to pair with Cas9. We generated a transgene in which the *Tyr* promoter was used to drive expression of a transcript that included a Hammerhead RZ [26] 5` of the gRNA as well as an HDV RZ [27] 3` of the gRNA sequence (Tyr::RZ-gRNA$_{(Tyr)}$-RZ; Fig. 2B). We examined the RZ construct function by electroporating an equal mass of the Tyr::Cas9 and Tyr::RZ-gRNA$_{(Tyr)}$-RZ and then assaying for loss of melanization in the melanocytes. This resulted in albino embryos, demonstrating that like the U6 and H1 gRNA drivers, the RZ-flanked gRNA transgene also facilitated CRISPR-Cas targeting of the *Tyr* locus. To examine if the activity of the RZ-flanked gRNA transgene translated to a change in FP-HDR insertion efficiency, we electroporated equal amounts of the Tyr::Cas9, Tyr::RZ-gRNA$_{(Tyr)}$-RZ and the *Tyr* FP-HDR construct. Driving ribozyme-flanked gRNAs with the *Tyr* cis-reg element resulted in 58.7% of embryos expressing YFP in a single melanocyte and 18.5% of embryos expressing YFP in both melanocytes (Fig. 2C, Supplementary Fig. 4B). Omitting the RZ construct produced no embryos with fluorescing melanocytes. These results demonstrate the effectiveness of using RNA pol II promoters to express gRNAs in the developing *Ciona* embryo.

The *Brachyury* and *Pou4* loci can be efficiently targeted for FP-HDR

To demonstrate the broad applicability of our optimized reagents, we chose two genes expressed in divergent tissues: *Bra*, which is expressed exclusively in the developing notochord; and *Pou4*, which is expressed exclusively in the PNS. We first developed reagents to target and disrupt *Bra* and *Pou4* to examine functionality of our CRISPR-Cas reagents and then we designed FP-HDR templates to *Bra* and *Pou4* to determine if our CRISPR-Cas FP-HDR technology broadly effective in *Ciona*.

Brachyury encodes a T-box transcription factor required for the differentiation and patterning of the embryonic notochord [28], and disruption of this gene causes major notochordal defects [29]. In initial experiments, we drove gRNAs with the CsH1 promoter targeted against two positions within the *Bra* coding region and expressed Cas9 with a 1.6 kB *Forkhead* (Fkh) promoter (Fkh::Cas9) that drives expression in the notochord precursor cells one cell cycle prior to the separation of the notochord from the endoderm [30]. The resulting embryos displayed a stubby-tail phenotype, consistent with the phenotype of a null mutant, suggesting that our CRISPR-Cas reagents targeted *Bra* (Supplementary Fig. 1B). In those experiments we titrated the amount of the Fkh::Cas9 driver and determined that 4 μg was the optimal amount to electroporate, because greater amounts of this driver (6 μg or 8 μg), when combined with a negative control gRNA targeting a GFP sequence resulted in embryos displaying stubby tail phenotype (Supplementary Fig. 1C).

The *Bra* FP-HDR template was constructed with a yellow fluorescent protein (YFP) sequence inserted in-frame at the 3` end of the *Bra* coding sequence designed to produce a Brachyury-YFP fusion protein (Fig. 5A) and the gRNA targeted the last coding exon of *Bra*. The protospacer-adjacent motif (PAM) was mutated in the FP-HDR template without altering the *Bra* coding sequence to reduce or eliminate Cas9-gRNA ribonucleoprotein binding. We

electroporated 4 μg of Fkh::Cas9 and 10 μg each of H1::gRNA$_{(Bra\text{-}Exon)}$ and Bra FP-HDR and let the embryos develop to tailbud stage. To score HDR efficiency, we binned anti-GFP immunostained embryos into three groups: no YFP expression, 1-30% expression, and >30% expression. Combined averages (3 biological replicates of 75-100 embryos scored for each replicate) revealed 48.6% of embryos had fluorescent nuclei in up to 30% of their notochord cells, while 19.3% of embryos had fluorescent nuclei in >30% of their notochord cells when the exon-targeting gRNA constructs were included (Fig. 3B, 3D). We saw one embryo in control electroporations without the gRNA construct that had fluorescence in the notochord. When we expressed gRNAs that targeted the 3`UTR of *Bra (*H1::gRNA$_{(Bra\text{-}UTR)}$), HDR efficiency was reduced and fluorescence was seen in non-notochordal cells (Fig. 3C). We also generated a transgene driving the same exon-targeting gRNAs as H1::gRNA$_{(Bra\text{-}exon)}$ but from the CrU6 promoter, however electroporating this construct with Fkh::Cas9 and Bra FP-HDR failed to produce cells expressing YFP. These observations suggest that combinations of gRNAs and FP-HDR templates should be tested and optimized to minimize off-target effects and maximize FP-HDR of targeted loci.

Subsequently, we designed CRISPR-Cas transgenes for FP-HDR to the *Pou4* gene. *Pou4* is a transcription factor that in *Ciona* is involved in the differentiation of ciliated epidermal sensory neurons (ESN) of the larval epidermis, including caudal, trunk and sensory neurons of the larval palps [13, 31]. In initial experiments, we expressed Cas9 in the epidermis with an *EpiB* promoter (EpiB::Cas9) [32] together with U6-driven gRNAs targeting the *Pou4* coding region. We hypothesized that disrupting *Pou4* would prevent cilia formation because of the role it plays in the regulatory network governing the formation of ESNs [13]. Embryos immunostained for acetylated tubulin had the number of ESNs markedly reduced, demonstrating that our *Pou4*-

targeting constructs were functional and suggesting that *Pou4* is indispensable for ESN differentiation (Supplementary Fig. 1D).

We then designed a Pou4 FP-HDR template with approximately 1.3 kB arms of homology and inserted *YFP* in-frame at the 3` end of the *Pou4* coding sequence. We could not identify an optimal targeting sequence near the end of the *Pou4* coding region so we selected a target in the *Pou4* 3`UTR ~200 bp from the stop codon and mutated the target's PAM sequence in the Pou4 FP-HDR template (Fig. 4A). We electroporated equal masses of EpiB::Cas9, U6::gRNA$_{(Pou4\text{-}UTR)}$ and Pou4 FP-HDR, let the embryos develop to larval stage, and detected YFP expressing cells using immunohistochemistry. For simplification of scoring efficiencies, we examined only caudal ESNs and did not assess abundance of YFP expressing cells in the palps or trunk. We binned the embryos into three groups: no caudal ESNs displaying YFP expression, 1-5 pair caudal ESNs displaying YFP expression (1-30% expressing), and 6-15 pair caudal ESNs displaying YFP expression (>30% expressing). In a typical electroporation, most (~85%) embryos had at least one ESN expressing YFP; 55% of the embryos expressed YFP in 1-5 pairs of CESNs and 30% had 6-15 CESNs expressing YFP (Fig. 4B, 4E). To examine whether fusing YFP to the C-terminal end of Pou4 would disrupt its function and thus disrupt cilia projections, we immunostained projecting cilia with an acetylated tubulin antibody. Localization of fluorescent nuclei coincided with the presence of cilia at each ESN pair, indicating that recombination likely generated a functional Pou4-YFP fusion protein (Fig. 4D) as disrupting *Pou4* expression disrupts cilia formation (Supplementary Fig. 1D). It is worth highlighting the observation that although Cas9 was expressed throughout the epidermis with the EpiB promoter, YFP fluorescence occurred only in ESNs, suggesting that the *Pou4* locus was specifically targeted for HDR.

In developing and optimizing the reagents for Pou4 FP-HDR, we also generated a transgene driving Cas9 with another epidermal promoter from the SoxB1 gene (SoxB1::Cas9) [7]. We found that electroporating SoxB1::Cas9 and U6::gRNA$_{(Pou4\text{-}UTR)}$ with the Pou4 FP-HDR template produced YFP expression in non-ESN epidermal cells (Fig. 4C). In addition, we constructed a transgene that drove *Pou4*-targeting gRNAs from the H1 promoter (H1::gRNA$_{(Pou4\text{-}UTR)}$). We found that electroporating H1::gRNA$_{(Pou4\text{-}UTR)}$ with EpiB::Cas9 and the Pou4 FP-HDR template was ineffective at generating FP-HDR (data not shown). These observations indicate that care must be taken to optimize not only the choice of targeting sequence but also the drivers used to express Cas9 and the gRNAs.

Sequencing PCR amplicons from single embryos demonstrates genomic FP-HDR integration

In each of the experiments examining FP-HDR mediated genome insertions, control groups never expressed FP, demonstrating our FP-HDR templates are transcriptionally silent and that targeting-gRNAs are necessary to produce predicted patterns of fluorescence. To more conclusively demonstrate that FP-HDR had occurred, we generated PCR amplicons from the region surrounding the site of genomic editing and sequenced the resulting fusion sequence. We generated primer sets designed to amplify fragments that should exist only if FP-HDR template recombination occurred: one primer was designed to anneal to YFP and the other primer was designed to anneal to an area of the genome upstream of the sequence of homology in the FP-HDR template (Supplementary Fig. 2). To generate amplicons revealing YFP integration at the *Bra* locus, we performed the same electroporation as above with Fkh::Cas9, H1::gRNA$_{(Bra\text{-}Cod)}$, and Bra FP-HDR. We live-sorted tailbud-stage embryos, pooled ten embryos that had 30% or

more of their notochord cells fluorescing, and subjected those pools to PCR. Sequencing analysis of cloned PCR amplicons demonstrated a correct YFP insertion into the genome from the *Bra* HDR template at the correct genomic location (Supplementary Fig. 3).

To generate amplicons revealing YFP integration at the *Pou4* locus, we electroporated equal masses of EpiB::Cas9, U6::gRNA$_{(Pou4\text{-}UTR)}$, and Pou4 FP-HDR and collected larval stage embryos that had ~50% of their ESNs fluorescing. Only ESN nuclei displayed fluorescence, but because Cas9 was expressed throughout the epidermis, it was likely that Cas9-directed FP-HDR occurred throughout the entire epidermis. Because the epidermis is a tissue that constitutes approximately 30% of the larval cell count, we reasoned there might be enough copies of modified genomes in a single embryo to detect the FP-HDR insertion and as such subjected individual embryos to PCR. Sequence analysis of cloned PCR amplicons from single YFP-positive Pou4 FP-HDR embryos demonstrated correct FP-HDR insertion into the genome (Supplementary Fig. 3). We were unable to produce this PCR product when embryos from control electroporations were used as template. Additionally, we performed control PCRs that attempted to amplify the same fragment using the Pou4 FP-HDR transgene and purified genomic DNA as template but were unsuccessful at generating amplicons. That we were successful at generating these amplicons only from embryos with predicted patterns of fluorescence suggests that the sequenced PCR product was produced from a genomic locus that had undergone recombination with the FP-HDR template.

Transgene-mediated biallelic knock-in

In other model systems, CRISPR-Cas9 has been shown to mediate HDR of both alleles [33, 34]. To examine if we could detect biallelic FP-HDR in *Ciona*, we constructed two

additional FP-HDR templates for *Tyr* that contained either cyan fluorescent protein (CFP) (Tyr FP-HDR$_{(CFP)}$) or mCherry (Tyr FP-HDR$_{(mCherry)}$) in place of GFP. We reasoned construction of two differently colored FP-HDR templates would allow us to detect instances of biallelic incorporation: mCherry expression alone (one or both alleles targeted), CFP expression (one or both alleles targeted), or, both mCherry and CFP expressed together in a single cell (one allele incorporated the CFP template and one allele incorporated the mCherry template). We electroporated Tyr::Cas9 and Tyr::RZ-gRNA$_{(Tyr)}$-RZ along with equal amounts of Tyr FP-HDR$_{(CFP)}$ and Tyr FP-HDR$_{(mCherry)}$. For the purpose of determining events of biallelic FP-HDR, we modified our scoring system and scored embryos as having no fluorophore expression, mCherry expression, CFP expression, or both mCherry and CFP expression. Electroporating these DNAs (2 biological replicates, 179 total embryos scored) resulted in an average of 27.5% of embryos expressed only mCherry in at least one of the two melanocytes, 32.5% of embryos expressed only CFP in at least one melanocyte, and 4.5% of embryos expressed both fluorescent proteins in a single cell (Fig. 5A, 5B). In some instances, we saw both melanocytes of a single embryo with CFP or mCherry expression, however we did not observe instances of CFP and mCherry expressed together in both melanocytes of a single embryo. Because this assay was unable to detect instances where both alleles in a single cell had incorporated either only CFP or only mCherry, these results suggest that our CRISPR-Cas FP-HDR templates mediated biallelic fluorophore insertion at the *Tyr* locus in *at least* 4.5% of the cells targeted. Accounting for instances where CFP or mCherry was targeted to both alleles suggests that the overall rate of biallelic targeting may be closer to 12%, i.e. 3X the rate measured above.

DISCUSSION

The ability to precisely edit the genomes of a variety of organisms has greatly expanded the usefulness of these systems as models for human development and disease [35-37]. Ascidians are gaining prominence as excellent models for examining a wide variety of biological problems including regeneration [38], Alzheimer's studies [39], and climate change research [40]. The ascidian embryo also provides a rapid and useful system for the analysis of complex regulatory elements [41-43]. In this study, we demonstrated CRISPR-Cas-mediated HDR is efficient at inserting a desired sequence into a specific genomic location in the ascidian *C. robusta*. To demonstrate FP-HDR is widely applicable in *Ciona*, we targeted three genes expressed in divergent tissues: *Tyrosinase*, a membrane-tethered enzyme that is the rate limiting step in the melanization of the two larval CNS melanocytes; *Brachyury*, a transcription factor expressed in the mesoderm; and *Pou4*, a transcription factor expressed in the PNS. We determined three aspects of the system need to be optimized: the type of promoter for expressing gRNAs, the promoter used for expressing Cas9, and the gRNA sequence selected to target the gene of interest and facilitate HDR. Our results demonstrate that each of these components must be tested to produce correct FP-HDR insertion and FP expression. We confirmed genomic recombination of the FP-HDR transgenes by sequencing PCR amplicons obtained from pools of embryos, and also from single embryos. Most notably, we demonstrate biallelic genome editing mediated by FP-HDR templates and this was detected in living F0 larvae.

We successfully modified three genes expressed in different cell types and at different developmental stages. *Tyrosinase* is expressed in only two melanocytes, the otolith and ocellus, which are likely involved in photo- and geotactic responses during larval swimming [44]. Disrupting *Tyrosinase* function caused either loss of pigmentation and/or pigment replacement by fluorescence in one or both CNS melanocytes of the larva. This easily scorable phenotype

made it straightforward to optimize the CRISPR-Cas FP-HDR components. *Brachyury* (mesoderm) and *Pou4* (ectoderm) encode well-characterized transcription factors, allowing easily identifiable FP-HDR phenotypes against established gene expression data. For each of these genes, we were able to demonstrate successful FP-HDR suggesting that CRISPR-Cas-mediated recombination is generally applicable throughout the genome.

While optimizing the delivery of the editing reagents, we evaluated three different means of expressing gRNAs for FP-HDR: expression from the CrU6 RNA pol III promoter, expression from a newly characterized CsH1 RNA pol III promoter, and expression from an RNA pol II promoter that drove ribozyme-flanked gRNAs. We found that all three gRNA delivery systems efficiently mediated HDR at the *Tyr* locus. However, HDR efficiency was much more varied at the other two loci. At the *Bra* locus, only H1 driven gRNAs, and not U6 driven gRNAs, were effective at mediating HDR. Conversely, the opposite pattern was observed at the *Pou4* locus: H1 was effective, while U6 was not. In human and mouse studies, the U6 and H1 promoters are both used to drive gRNAs as well as other small RNA molecules such as shRNAs. In the case of expressing shRNAs, the U6 promoter is thought to be more effective than the H1 promoter [45]. However, to express gRNAs for eliciting CRISPR-Cas activity in mammalian cell culture, the U6 and H1 resulted in similar efficiencies [21]. It is not clear why an apparent difference in activity exists in *Ciona*. When generating transgenic ascidians via electroporation, the amount of DNA that can be electroporated while still maintaining a high percentage of normally developing embryos is limited to less than about 125 ng/μL [12]. The H1 promoter is about one-tenth the length of the U6 promoter, so more copies of an H1 transgene can be delivered per mass of DNA compared to U6 transgenes, likely affording higher amounts of gRNA expression. There is also likely to be cell-type specific differences in the abilities of these promoters to drive gRNAs and

this could be examined in future experiments. Based on the optimizations reported here, we suggest driving ribozyme-flanked gRNAs from tissue-specific RNA pol II promoters. Although these constructs are more difficult to generate, the ability to control where gRNAs are expressed, or when they are expressed if using an inducible promoter, makes them a more attractive choice, in our opinion, for gRNA delivery. In addition, the ability to control which cells express gRNAs can help to minimize any adverse effects resulting from expressing reagents in cells that are not desired targets.

Titrating the amount of Cas9 protein and identifying appropriate gRNAs for each specific targeted locus is also essential. If null-phenotypes for a given gene are known, this is easily accomplished by performing a series of electroporations in which the Cas9 delivery transgene amount is varied and the resulting embryos/larvae carefully examined for non-specific phenotypic changes. To interpret these effects, it is important to have previously characterized the expression pattern of the gene being targeted. For example, in all three genes targeted in this report, we and others had previously characterized the spatiotemporal expression of each gene and thus could expect a known pattern of fluorescence that would indicate a successful editing event after FP insertion, In our opinion, it is important to determine the maximal amounts of deliverable Cas9 protein, both alone or co-expressed with a control gRNA, that will not introduce altered phenotypes in embryos/larvae. We recommend when possible to design gRNAs that target the coding regions of genomic loci. Ascidian genomes are highly polymorphic [46] but there should be a reduced frequency of polymorphisms in coding regions compared to non-coding regions because of functional constraints. Because we are using the expression of fluorescent protein fusions to indicate successful genome editing, it is important to ensure that the FP-HDR template alone has no inherent transcriptional activity, as this would confound the

experimental outcome. Therefore, it is important to run controls in which only the FP-HDR template is electroplated into fertilized eggs to ensure that is it transcriptionally silent.

One of the more noteworthy findings from this study is the ability to use CRISPR-Cas mediated FP-HDR to simultaneously target both alleles of a specific genetic locus in the F0 generation. By using two different FP-HDR templates containing either CFP or mCherry, we could show that both alleles of the *Tyrosinase* gene could be targeted together within the same cell of a living *Ciona* embryo. As we reported, about 4% of embryos in which we targeted both alleles had melanocytes that fluoresced both cyan and red, suggesting that each allele had been edited. Given that we are unable to detect biallelic knock-in events in melanocytes that express CFP/CFP or mCherry/mCherry, it is likely that 4% is an underestimate and that 10-12% of the resulting embryos/larvae have had both alleles edited in one of their melanocytes. In some cases, it may be more advantageous to generate knock-in lines of animals that carry genomic edits at specific loci. Here it should be possible to combine our knock-in approach with recently reported methods to induce germ cell mutations via forced regeneration following larval tail ablation [47] to generate a line of animals containing the desired genomic edit/insertion. It would still be possible to easily determine single versus double allele edits by generating lines that have either a CFP or mCherry knock-in and then crossing these lines together; single allelic knock-ins would glow with a single color, while those cells in which a double knock-in was present would glow with both inserted FP colors. By performing the appropriate crosses in this manner, it would be straightforward to separate out the resulting allelic modifications and easily determine whether one or both alleles were targeted in the experiment.

We see several implications based on the results presented in this study. The FP-HDR technology, as others have reported [9], can be used to insert a fluorophore sequence in-frame

with a gene of interest, effectively tagging the protein allowing researchers to follow not only the advent but also the cellular location of that particular protein. The tagged proteins provided by the FP-HDR approach should be expressed at endogenous levels and reflect a more accurate of expression and cellular location than transgenically expressed FP-tagged genes. This is often difficult to achieve when transgene drivers are used, primarily because all cis-regulatory elements should be present; transgenes likely lack some of these elements, particularly if the transgenes used contain minimal elements required for expression.

CRISPR-Cas-mediated knock-ins to transcription factors would also benefit ChIP-Seq studies in ascidians, as the use of tagged proteins obviates the current limitation of there being few antibodies available that recognize ascidian transcription factors and allows IP to be performed against the motif of a fusion protein. Previous ChIP-Seq studies in ascidians have been performed on embryos in which transgenes express tagged TF proteins [48, 49]. However, when these TF fusion proteins are expressed transgenically, the level of TF-fusion expression is unknown and in some cases outside of the native cell lineage [48]. Transgenic levels of a TF can cause non-native target binding if overexpressed, while the binding of a TF expressed outside of its native cells may also confound results [50]. The use of endogenously expressed GFP-tagged TFs has been reported in other systems [51, 52], and the use of the FP-HDR approach described here should facilitate the widespread adoption of ChIP-Seq in ascidians.

Ascidians are excellent models for studying gene regulation during development, and as simple chordates offer insight into more complicated vertebrate development. A significant potential application of this technology is to edit the germline. A recent report that demonstrated mutations introduced into embryonic somatic tissues can be inherited by regenerated germ cells [47] introduced the mutations using a TALEN-based approach [53, 54]; these components are far

more difficult to generate and test than CRISPR-based systems. Our CRISPR-Cas9 mediated approach for precisely modifying the ascidian genome, coupled with forced germ cell regeneration, is likely to accelerate the ability to generate stable knock-in lines in ascidians and will be of great benefit to the ascidian community.

ACKNOWLEDGEMENTS

We thank Christina Niemeyer for useful comments on the manuscript and thank Karl Garcia and Tiffany Hoang for assistance with molecular cloning. This work was supported by NSF grants IOS-0951347, IOS-1557448 and NIH grant 1R21DC01318001A1 (NIDCD) to RWZ.

Table 1.1: Descriptions of transgenes referenced in the figure legends.

Transgene Name	Description	Expression Pattern
Tyr::Cas9	*Tyr* cis-reg DNA drives expression of *Cas9*	melanocytes
Fkh::Cas9	*FoxD* cis-reg DNA drives expression of *Cas9*	notochord precursors
EpiB::Cas9	*EpiB* cis-reg DNA drives expression of *Cas9*	epidermis
SoxB1::Cas9	*SoxB1* cis-reg DNA drives expression of *Cas9*	epidermis
U6::gRNA$_{(Tyr)}$	*CrU6* cis-reg DNA drives expression of *Tyr*-targeting gRNAs	ubiquitous
H1::gRNA$_{(Tyr)}$	*CsH1* cis-reg DNA drives expression of *Tyr*-targeting gRNAs	ubiquitous
Tyr::RZ-gRNA$_{(Tyr)}$-RZ	*Tyr* cis-reg DNA drives expression of the ribozyme-flanked gRNA$_{(Tyr)}$ sequence	melanocytes
H1::gRNA$_{(Bra\text{-}Cod)}$	*CsH1* cis-reg drives expression of *Bra* coding region-targeting gRNAs	ubiquitous
H1::gRNA$_{(Bra\text{-}UTR)}$	*CsH1* cis-reg drives expression of *Bra* UTR-targeting gRNAs	ubiquitous
U6::gRNA$_{(Pou4\text{-}UTR)}$	*CrU6* cis-reg drives expression of *Pou4* UTR targeting gRNAs	ubiquitous
U6::gRNA$_{(GFP)}$	*CrU6* cis-reg drives expression of *GFP* targeting gRNAs	ubiquitous
Tyr FP-HDR$_{(2.0\ kB)}$	*YFP* flanked by 2.0 kB arms of *Tyr* DNA	N/A
Tyr FP-HDR$_{(1.0\ kB)}$	*YFP* flanked by 1.0 kB arms of *Tyr* DNA	N/A
Tyr FP-HDR$_{(500bp)}$	*YFP* flanked by 500 bp arms of *Tyr* DNA	N/A
Tyr FP-HDR$_{(CFP)}$	*CFP* flanked by 1.0 kB arms of *Tyr* DNA	N/A
Tyr FP-HDR$_{(mCherry)}$	*mCherry* flanked by 1.0 kB arms of *Tyr* DNA	N/A
Bra FP-HDR	*YFP* flanked by ~1.3 kB arms of *Bra* DNA	N/A
Pou4 FP-HDR	*YFP* flanked by ~1.2 kB arms of *Pou4* DNA	N/A

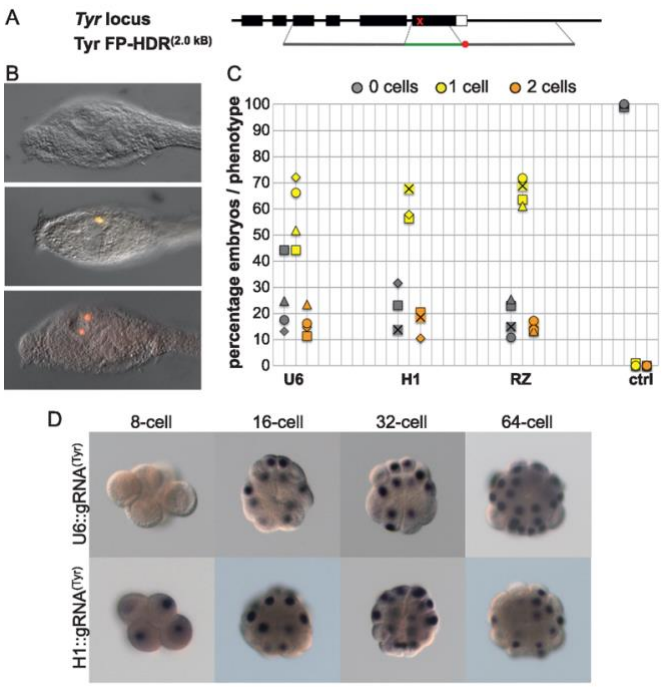


Figure 1.1: GFP is efficiently inserted into the *C. robusta Tyrosinase* locus using CRISPR-Cas mediated HDR. (**A**) Schematics of HDR constructs with varying homology arm lengths. The red X indicates the genomic location where gRNAs were targeted; this target site does not exist within the recombination template. Tyr FP-HDR templates have a stop codon at the 3` end of the GFP sequence (red octagon) producing a truncated Tyr protein fused in-frame with GFP. The same gRNA target was used in each experiment. (**B**) Examples of embryos with 2 (top), 1 (middle), or 0 (bottom) fluorescing melanocytes; a result of recombination with the HDR template. 8 µg of each construct (Tyr::Cas9, U6::gRNA$_{(Tyr)}$, and Tyr FP-HDR) was electroporated into fertilized, dechorionated eggs. GFP fluorescence is detected anti-GFP immunohistochemistry at larval stage. (**C**) The three CRISPR-Cas FP-HDR reagents are effective and necessary to generate melanocytes expressing GFP. Electroporating 8 µg each of Tyr::Cas9, U6::gRNA$_{(Tyr)}$ and Tyr FP-HDR$_{(2.0\ kB)}$ resulted in an average of 55.6% of the embryos (n=308; 3 biological replicates) exhibiting one fluorescing pigment cell while 16% had fluorescence in both pigment cells (43.9% of total melanocytes). In control electroporations that did not include U6::gRNA$_{(Tyr)}$, one embryo displayed GFP expression. (**D**) Shortening the arms of homology to 500 bp reduces the percentage of cells expressing GFP. Electroporations were performed (4 biological replicates) with 8 µg of each construct: Tyr::Cas9 and U6::gRNA$_{(Tyr)}$ and either Tyr FP-HDR$_{(2.0\ kB)}$, Tyr FP-HDR$_{(1.0\ kB)}$, or Tyr FP-HDR$_{(500bp)}$. Compared to the 2.0 kB control group (16.6% with two YFP+ melanocytes, 61.7% with one YFP+ melanocyte, 21.7% with no YFP+ melanocytes; n=575), the 1.0 kB transgene resulted in approximately the same percentage of embryos expressing YFP in two, one, or zero pigment cells (15.9%, 59.5%, and 24.7% respectively, n=496; Fig2B). However, in the 500 bp HDR template group (n= 491), the percentage of embryos with two YFP+ melanocytes was reduced to 7.6%, and percentage of embryos with one YFP+ melanocyte was reduced to 47% while the percentage of embryos with no YFP+ melanocytes increased to 43% (n=491). MANOVA followed by a pairwise ANOVA detected differences with statistical significance ($p<0.04801$) for the 500 bp construct vs. the 1.0 kB construct at eliciting 0 cells expressing YFP.

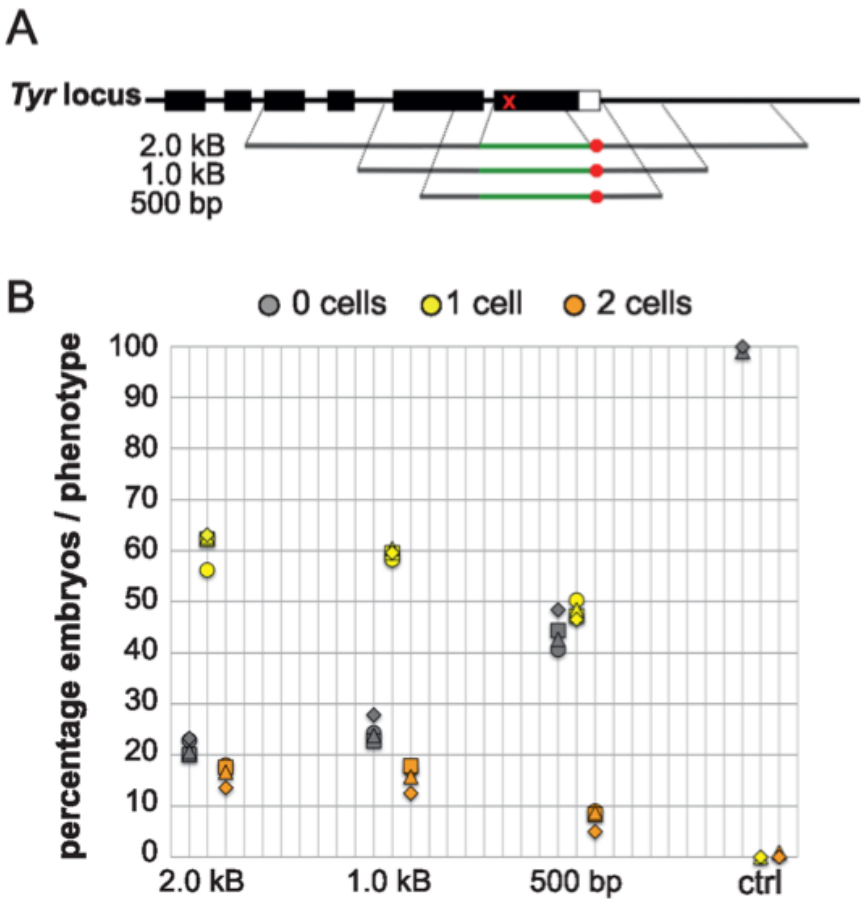


Figure 1.2: Expressing gRNAs from the CsH1 promoter or gRNAs flanked by ribozymes from an RNA polymerase II promoter is effective at eliciting FP-HDR than gRNAs expressed from the CrU6 promoter. (**A**) Sequence of the *Ciona savignyi* H1 promoter and *Tyr*-targeting gRNA (entire gRNA underlined; *Tyr* target sequence in blue). (**B**) Sequence of the hammerhead ribozyme fused to the *Tyr*-targeting gRNA (entire gRNA sequence underlined; *Tyr* target sequence in blue) followed by the HDV ribozyme sequence. (**C**) Guide RNAs driven from the CrU6 or the CsH1 promoters or from an RNA Polymerase II promoter (ribozyme-flanked gRNA) facilitate FP-HDR. Electroporating 8 µg each of Tyr::Cas9, Tyr FP-HDR$_{(2.0\ kB)}$ and either the CrU6::gRNA$_{(Tyr)}$, CsH1::gRNA$_{(Tyr)}$ or Tyr::RZ-gRNA$_{(Tyr)}$-RZ resulted in varying percentages of melanocytes expressing GFP. The control group, CrU6::gRNA$_{(Tyr)}$, resulted in an average of (n=370; 3 biological replicates) 16.5% of embryos displaying fluorescence in both pigment cells, 55.6% of the embryos with one pigment cell fluorescing, and 27.9% of embryos displayed no fluorescence. H1::gRNA$_{(Tyr)}$ resulted in an average of (n=186; 3 biological replicates) 16% of embryos displaying fluorescence in both pigment cells, 55.6% of the embryos with one pigment cell fluorescing, and 28.4% of embryos displayed no fluorescence. Electroporating 8 µg each of the Tyr::Cas9, Tyr FP-HDR$_{(2.0\ kB)}$, and Tyr::RZ-gRNA$_{(Tyr)}$-RZ resulted in an average of (n=312; 3 biological replicates) 19.1% of embryos had fluorescence in both pigment cells, 62.3% of the embryos in one pigment cell, and 18.6% of embryos displayed no fluorescence. A MANOVA followed by pairwise ANOVA identified Tyr::RZ-gRNA$_{(Tyr)}$-RZ as having an increase in efficacy over the U6 control group and H1::gRNA$_{(Tyr)}$ for generating one YFP+ melanocyte.

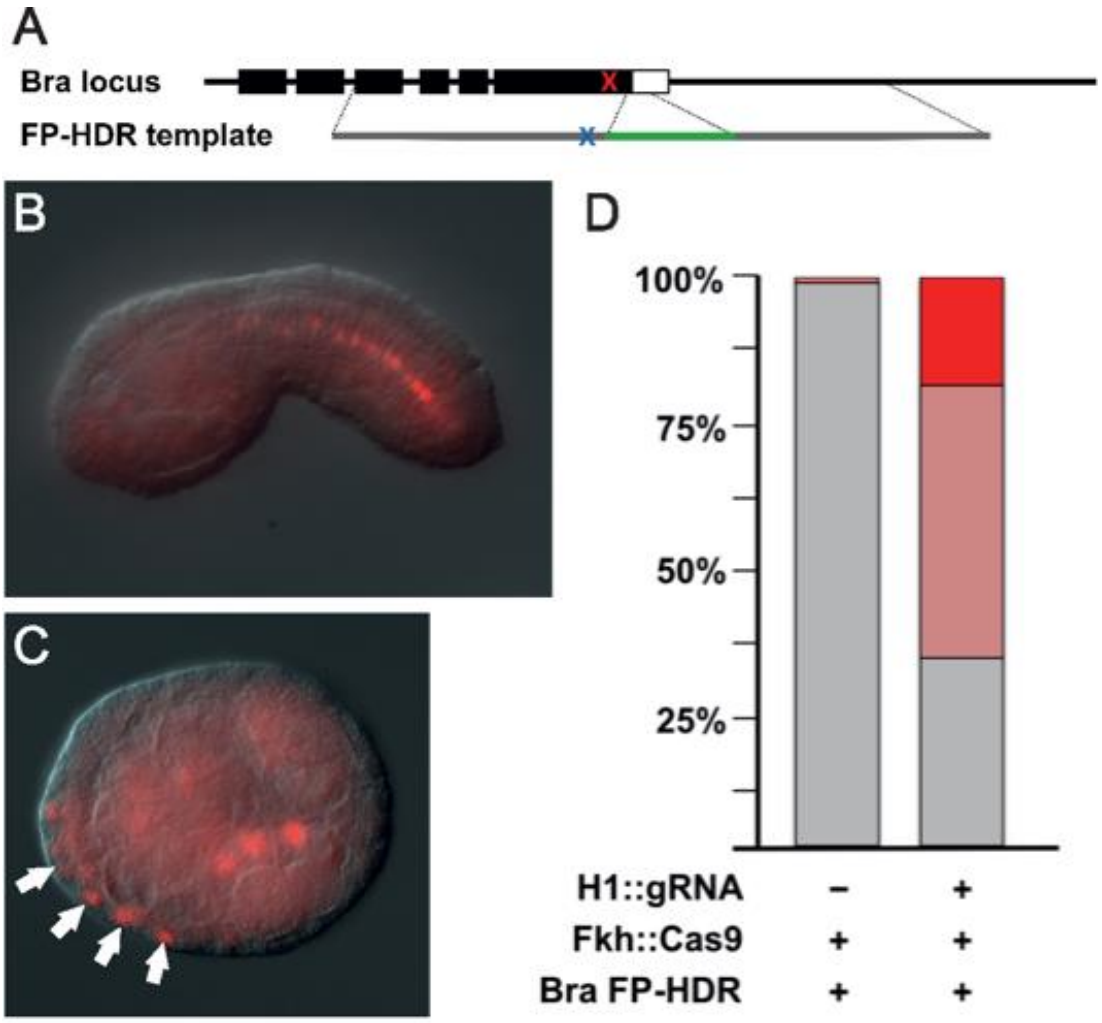


Figure 1.3: YFP is efficiently inserted into the *C. robusta Brachyury* locus using CRISPR-Cas mediated HDR resulting in YFP expression in the notochord. (**A**) Schematic of *Brachyury* HDR template. The red X indicates genomic site in the Bra coding region where gRNAs were targeted; the blue X indicates conservatively mutated PAM in the HDR template upstream of the YFP sequence (green). (**B**) An example of a tailbud embryo with fluorescing notochord nuclei. Fertilized eggs were electroporated with 4 µg of Fkh::Cas9, 10 µg of CsH1::gRNA$_{(Bra\text{-}Cod)}$ targeting the 3` end of the Bra coding region, and 10 µg of the Bra FP-HDR template. YFP was detected by immunohistochemistry with an anti-GFP antibody. (**C**) Correct combinations of Cas9 drivers and gRNA drivers are required for proper genome editing. Electroporating 4 µg of Fkh::Cas9 and 8 µg each of Bra FP-HDR and H1::gRNA$_{(Bra\text{-}UTR)}$ resulted in YFP being expressed in a non-notochordal lineage. Image shows an early gastrula embryo with YFP expression in several Fkh-expressing cells (arrows) that are not of the notochord lineage. (**D**) Quantitation of embryos with YFP fluorescing notochord cells. Percentage (3 biological replicates of 75-100 embryos scored for each replicate) of embryos with no fluorescing notochord cells (32.1%, gray); percentage of embryos with 1-30% of notochord cells fluorescing (48.6%, pink); percentage of embryos with >30% of notochord cells fluorescing (19.3%, red). Excluding H1::gRNA$_{(Bra\text{-}Cod)}$ resulted in scoring one embryo expressing YFP.

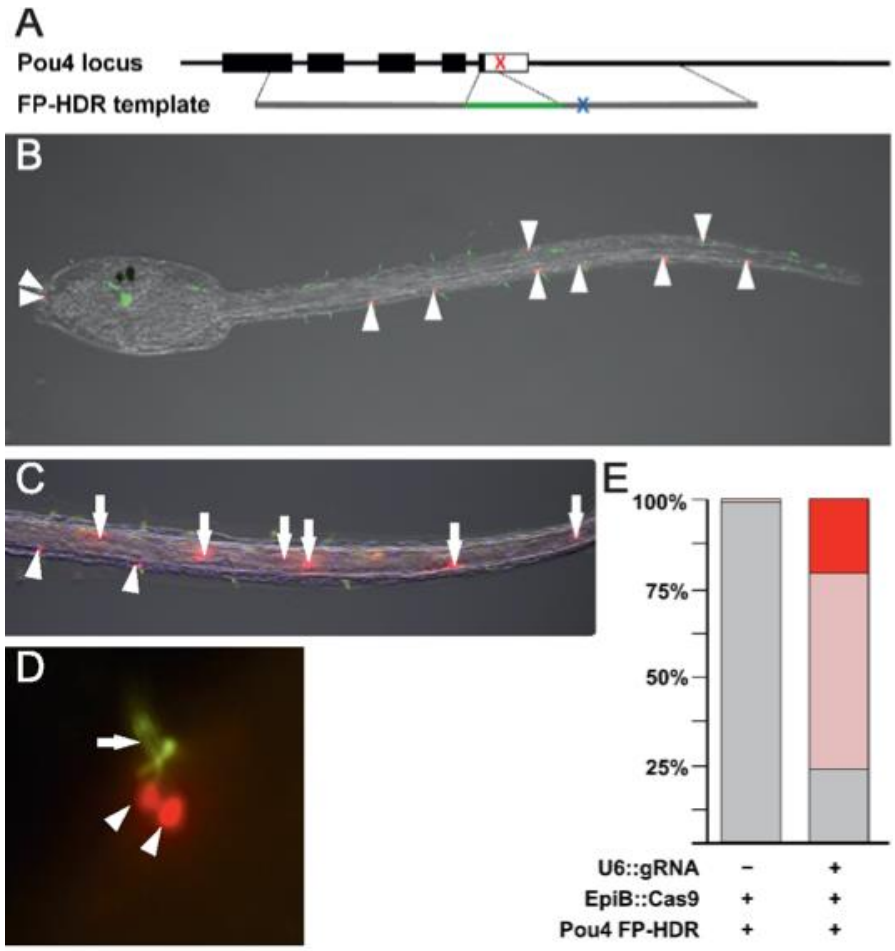


Figure 1.4: YFP is efficiently inserted into the *C. robusta Pou4* locus using CRISPR-Cas mediated HDR resulting in YFP expression in the ESNs. (**A**) Schematic of *Pou4* HDR template. A red X indicates genomic site in the Pou4 3` UTR where gRNAs were targeted; a blue X indicates mutated target sequence in template upstream of the YFP sequence (green). *Pou4* was targeted for FP-HDR by electroporating 8 µg each of EpiB::Cas9, U6::gRNA$_{(Pou4\text{-}UTR)}$, and Pou4 FP-HDR. (**B**) Example of larval stage embryo with 8/18 caudal ESN pairs expressing YFP (red, arrowheads). Palp ESNs also express YFP. Embryos were fixed and antibody stained for GFP (red) and anti-acetylated tubulin to label ESN cilia (green). (**C**) Expressing Cas9 with the *SoxB1* promoter resulted in YFP expression in ESNs and ectopic expression of YFP. 8 µg each of SoxB1::Cas9, U6::gRNA$_{(Pou4\text{-}UTR)}$, and Pou4 FP-HDR was electroporated and embryos were fixed and stained for GFP and acetylated tubulin. Arrows point to non-ESN epidermal cells expressing YFP; arrowheads point to ESN nuclei expressing YFP. (**D**) ESN cilia (green) projecting from fluorescing ESN nuclei pair (red) indicate functional Pou4-YFP fusion protein. (**E**) Quantitation of embryos with YFP fluorescing ESNs. Combined percentages (n= 373; 3 biological replicates) of embryos with fluorescing ESNs. Percentage of embryos with no ESNs expressing YFP (15%, gray); percentage of embryos with 1-30% ESNs fluorescing (55%, pink); percentage of embryos with >30% of ESNs fluorescing (30%, red). We scored 1/100 embryos as having 1-30% fluorescent ESNs.

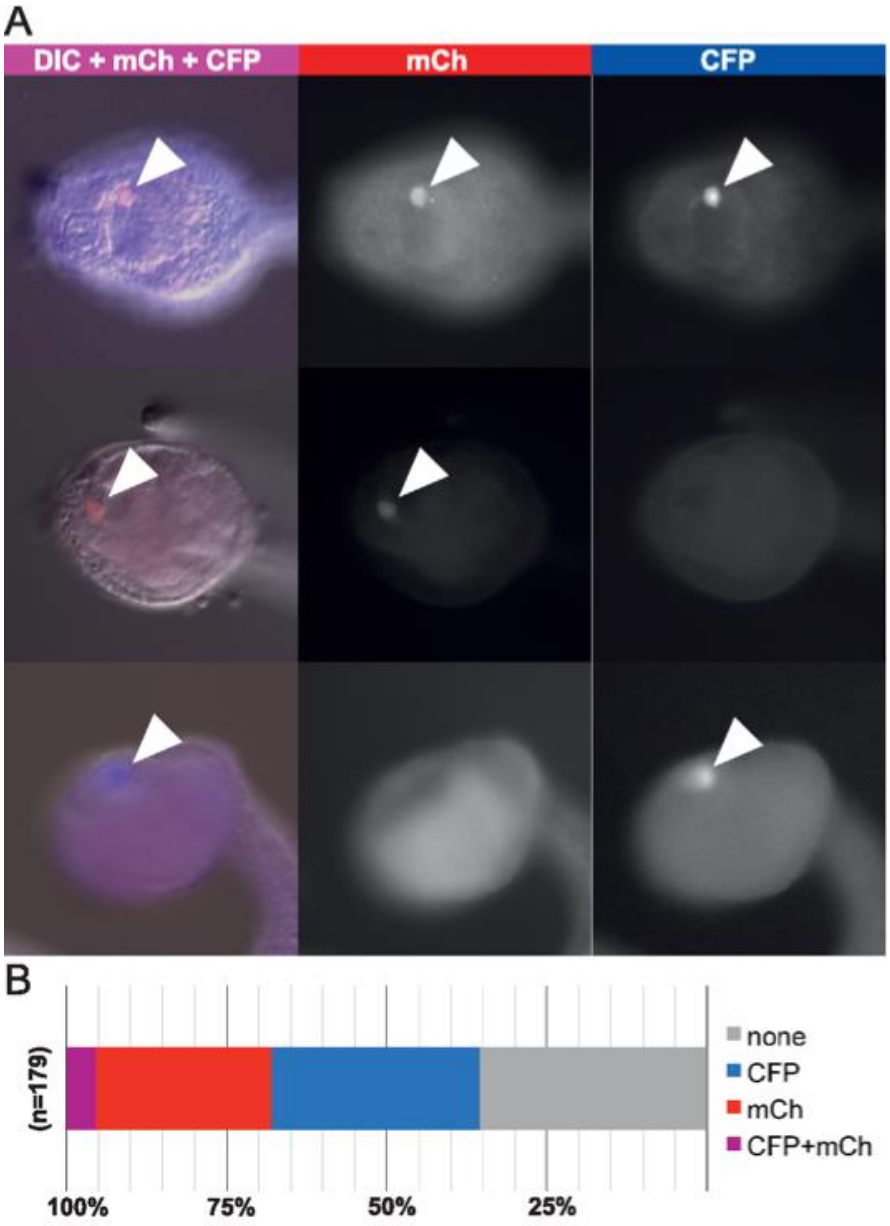
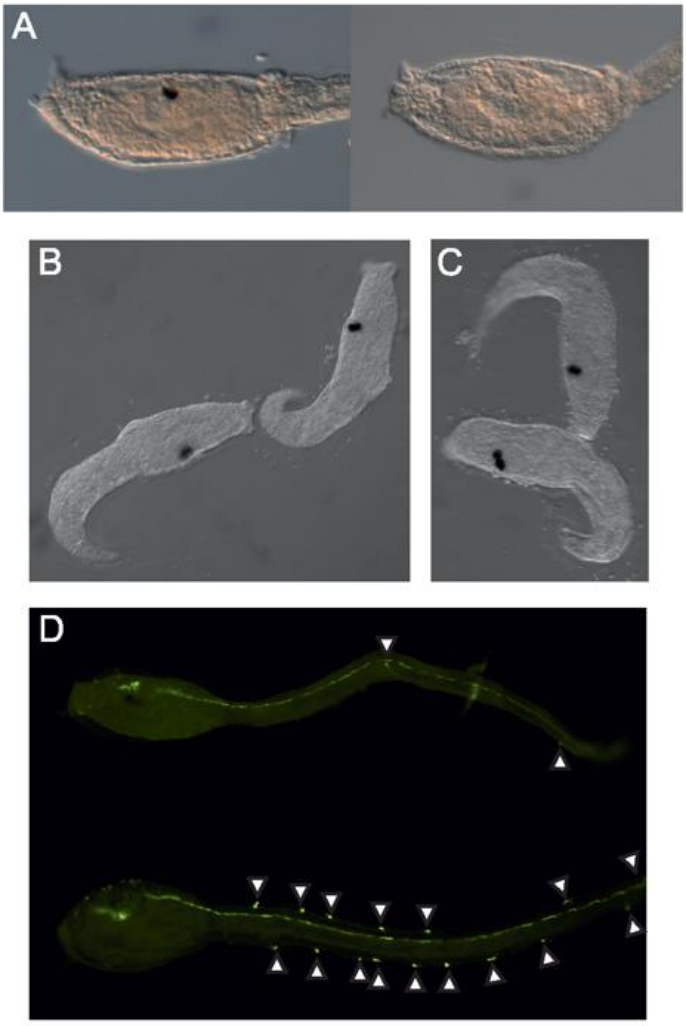
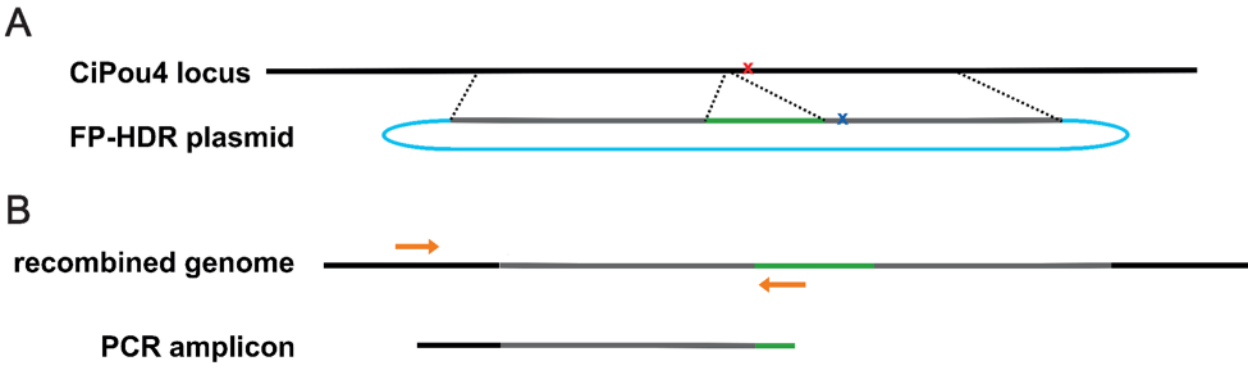


Figure 1.5: A FP can be inserted into both alleles of the Tyr locus simultaneously using CRISPR-Cas 9 mediated HDR. Fertilized eggs were electroporated with 8 µg each of Tyr::Cas9 and Tyr::RZ-gRNA$_{(Tyr)}$-RZ and with 4 µg each of Tyr FP-HDR$_{(CFP)}$ and Tyr FP-HDR$_{(mCherry)}$. (**A**) Representative live embryos (13 hpf) showing CFP and/or mCherry expression in melanocytes. Top row: embryo with melanocyte displaying both CFP and mCherry in a single melanocyte; middle row: embryo with melanocyte displaying mCherry only; bottom row: embryo with melanocyte displaying CFP only. CFP channel (right column), mCherry channel (middle column), composite channels (left column). (**B**) Percentages of embryos expressing both fluorophores (purple, 4.5%), only mCherry expressing cells (red, 27.5%), only CFP expressing cells (blue, 32.5%), and percentage of embryos that expressed no fluorescence (gray, 35.5%). Two biological replicates, n=179.



Supplemental Figure 1.1: CRISPR-Cas reagents targeting *Tyr*, *Bra* and *Pou4* are effective at eliciting null-phenotypes. (**A**) The CsH1 promoter is effective at expressing gRNAs and disrupting the *Tyr* gene resulting in unpigmented melanocytes. 8 µg of Tyr::Cas9 and H1::gRNA$_{(Tyr)}$ were electroporated and embryos developed to larval stage. Examples of embryos with 1 (left) or both (right) unmelanized melanocytes. (**B**) Targeting the *Bra* locus results in embryos displaying a phenotype consistent with a described null phenotype (Chiba et al., 2009). Electroporating 4 µg Fkh::Cas9 and 8 µg H1::gRNA$_{(Bra)}$ results in embryos with stubby tails. (**C**) Negative control electroporations indicate 8 µg of Fkh::Cas9 is an excessive amount. Electroporating 8 µg of Fkh::Cas9 and 8 µg of a negative control H1::gRNA$_{(GFP)}$ results in embryos displaying a phenotype consistent with a described null phenotype, whereas electroporating 4 µg Fkh::Cas9 and 8 µg of H1::gRNA$_{(GFP)}$ resulted in normal development (image not shown). (**D**) Targeting the *Pou4* locus results in a reduction of ESN cilia. Embryos electroporated with 8 µg EpiB::Cas9 and 4 µg each of four different U6-driven gRNA transgenes targeting the Pou4 coding region had reduced numbers of ESN cilia (2 total ESNs, top, green). The average number of cilia (dorsal/ventral) was reduced to 3.9/2.3, vs. 8.5/7.2 in control group (n>100 per group). Embryos electroporated with only the EpiB::Cas9 transgene had normal numbers of ESNs (bottom). White arrowheads indicate ESNs.



Supplemental Figure 1.2: Schematic of *Pou4* HDR template and amplification strategy as an example of how genomic insertions were verified. (**A**) Red X indicates genomic site where gRNAs were targeted, dark blue X indicates mutated PAM site in the Pou4 FP-HDR plasmid preventing CRISPR-Cas binding. Light blue represents pZapANX vector. (**B**) Diagram of hypothesized recombined genome with YFP insert (green). Orange arrows indicate locations of FWD and REV primers used to generate a PCR amplicon that should be produced only if the genome has incorporated the FP-HDR template. Amplicons were cloned into pBS2KS+ and sequenced with T7 and T3 primers to examine both ends of the fragment.

Supplementary Figure 1.3: Sequencing Brachyury FP-HDR and Pou4 FP-HDR PCR amplicons demonstrates genomic FP-HDR integration. A total of 10 amplicons from 10 different Bra-YFP+ batches of embryos and 8 amplicons from 8 different Pou4-YFP+ embryos were cloned into pBSK; 4 Bra clones were sequenced and 4 Pou4 clones were sequenced. Each of the examined sequences was nearly identical to the sequences reported above. The primers used (orange) generated sequences that spans from a genomic region outside of the homology arm (black) through the template arm (grey; primers used to generate homology arm in blue) and into the YFP sequence (green). The polylinker fusing Bra (top) or Pou4 (bottom) to YFP is shown in lowercase. The sequences shown above are one of four sequences derived from eight cloned Bra-YFP or Pou4-YFP amplicons. SNPs found within the sequenced clones are indicated as bold grey.

REFERENCES

1. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.

2. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.

3. Gokcezade, J., G. Sienski, and P. Duchek, *Efficient CRISPR/Cas9 plasmids for rapid and versatile genome editing in Drosophila.* G3 (Bethesda), 2014. **4**(11): p. 2279-82.

4. Kimura, Y., et al., *Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering.* Sci Rep, 2014. **4**: p. 6545.

5. Ma, D. and F. Liu, *Genome Editing and Its Applications in Model Organisms.* Genomics Proteomics Bioinformatics, 2015. **13**(6): p. 336-44.

6. Sasaki, H., et al., *CRISPR/Cas9-mediated gene knockout in the ascidian Ciona intestinalis.* Dev Growth Differ, 2014. **56**(7): p. 499-510.

7. Stolfi, A., et al., *Tissue-specific genome editing in Ciona embryos by CRISPR/Cas9.* Development, 2014. **141**(21): p. 4115-20.

8. Gandhi, S., et al., *Evaluation and rational design of guide RNAs for efficient CRISPR/Cas9-mediated mutagenesis in Ciona.* Dev Biol, 2017. **425**(1): p. 8-20.

9. Gaj, T., C.A. Gersbach, and C.F. Barbas, 3rd, *ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering.* Trends Biotechnol, 2013. **31**(7): p. 397-405.

10. Singh, P., J.C. Schimenti, and E. Bolcun-Filas, *A mouse geneticist's practical guide to CRISPR applications.* Genetics, 2015. **199**(1): p. 1-15.

11. Li, K., et al., *Optimization of genome engineering approaches with the CRISPR/Cas9 system.* PLoS One, 2014. **9**(8): p. e105779.

12. Zeller, R.W., M.J. Virata, and A.C. Cone, *Predictable mosaic transgene expression in ascidian embryos produced with a simple electroporation device.* Dev Dyn, 2006. **235**(7): p. 1921-32.

13. Joyce Tang, W., J.S. Chen, and R.W. Zeller, *Transcriptional regulation of the peripheral nervous system in Ciona intestinalis.* Dev Biol, 2013. **378**(2): p. 183-93.

14. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

15. Chen, B., et al., *Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system.* Cell, 2013. **155**(7): p. 1479-91.

16. Piccinelli, P., M.A. Rosenblad, and T. Samuelsson, *Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes.* Nucleic Acids Res, 2005. **33**(14): p. 4485-95.

17. Haeussler, M., et al., *Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR.* Genome Biol, 2016. **17**(1): p. 148.

18. Nishiyama, A. and S. Fujiwara, *RNA interference by expressing short hairpin RNA in the Ciona intestinalis embryo.* Dev Growth Differ, 2008. **50**(6): p. 521-9.

19. Jiang, D., et al., *Pigmentation in the sensory organs of the ascidian larva is essential for normal behavior.* J Exp Biol, 2005. **208**(Pt 3): p. 433-8.

20. Song, F. and K. Stieger, *Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks.* Mol Ther Nucleic Acids, 2017. **7**: p. 53-60.

21. Ranganathan, V., et al., *Expansion of the CRISPR-Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs.* Nat Commun, 2014. **5**: p. 4516.

22. Bartkiewicz, M., H. Gold, and S. Altman, *Identification and characterization of an RNA molecule that copurifies with RNase P activity from HeLa cells.* Genes Dev, 1989. **3**(4): p. 488-99.

23. Wang, W. and L. Christiaen, *Transcriptional enhancers in ascidian development.* Curr Top Dev Biol, 2012. **98**: p. 147-72.

24. Sierro, N., et al., *DBTGR: a database of tunicate promoters and their regulatory elements.* Nucleic Acids Res, 2006. **34**(Database issue): p. D552-5.

25. Gao, Y. and Y. Zhao, *Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing.* J Integr Plant Biol, 2014. **56**(4): p. 343-9.

26. Prody, G.A., et al., *Autolytic processing of dimeric plant virus satellite RNA.* Science, 1986. **231**(4745): p. 1577-80.

27. Ferre-D'Amare, A.R., K. Zhou, and J.A. Doudna, *Crystal structure of a hepatitis delta virus ribozyme.* Nature, 1998. **395**(6702): p. 567-74.

28. Di Gregorio, A., *T-Box Genes and Developmental Gene Regulatory Networks in Ascidians.* Curr Top Dev Biol, 2017. **122**: p. 55-91.

29. Chiba, S., et al., *Brachyury null mutant-induced defects in juvenile ascidian endodermal organs.* Development, 2009. **136**(1): p. 35-9.

30. Yagi, K., Y. Satou, and N. Satoh, *A zinc finger transcription factor, ZicL, is a direct activator of Brachyury in the notochord specification of Ciona intestinalis.* Development, 2004. **131**(6): p. 1279-88.

31. Candiani, S., et al., *Ci-POU-IV expression identifies PNS neurons in embryos and larvae of the ascidian Ciona intestinalis.* Dev Genes Evol, 2005. **215**(1): p. 41-5.

32. Zeller, R.W., et al., *Optimized green fluorescent protein variants provide improved single cell resolution of transgene expression in ascidian embryos.* Dev Dyn, 2006. **235**(2): p. 456-67.

33. Chang, N., et al., *Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos.* Cell Res, 2013. **23**(4): p. 465-72.

34. Gratz, S.J., et al., *Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila.* Genetics, 2014. **196**(4): p. 961-71.

35. Singh, V., D. Braddick, and P.K. Dhar, *Exploring the potential of genome editing CRISPR-Cas9 technology.* Gene, 2017. **599**: p. 1-18.

36. Cardi, T., N. D'Agostino, and P. Tripodi, *Genetic Transformation and Genomic Resources for Next-Generation Precise Genome Engineering in Vegetable Crops.* Front Plant Sci, 2017. **8**: p. 241.

37. Chen, L., et al., *Advances in genome editing technology and its promising application in evolutionary and ecological studies.* Gigascience, 2014. **3**: p. 24.

38. Spina, E.J., et al., *A microRNA-mRNA expression network during oral siphon regeneration in Ciona.* Development, 2017. **144**(10): p. 1787-1797.

39. Virata, M.J. and R.W. Zeller, *Ascidians: an invertebrate chordate model to study Alzheimer's disease pathogenesis.* Dis Model Mech, 2010. **3**(5-6): p. 377-85.

40. Lopez, C.E., et al., *Proteomic responses to elevated ocean temperature in ovaries of the ascidian Ciona intestinalis.* Biol Open, 2017. **6**(7): p. 943-955.

41. Kusakabe, T., *Decoding cis-regulatory systems in ascidians.* Zoolog Sci, 2005. **22**(2): p. 129-46.

42. Farley, E.K., et al., *Suboptimization of developmental enhancers.* Science, 2015. **350**(6258): p. 325-8.

43. Satou, Y. and K.S. Imai, *Gene regulatory systems that control gene expression in the Ciona embryo.* Proc Jpn Acad Ser B Phys Biol Sci, 2015. **91**(2): p. 33-51.

44. Nicol, D. and I.A. Meinertzhagen, *Cell counts and maps in the larval central nervous system of the ascidian Ciona intestinalis (L.).* J Comp Neurol, 1991. **309**(4): p. 415-29.

45. Makinen, P.I., et al., *Stable RNA interference: comparison of U6 and H1 promoters in endothelial cells and in mouse brain.* J Gene Med, 2006. **8**(4): p. 433-41.

46. Dehal, P., et al., *The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins.* Science, 2002. **298**(5601): p. 2157-67.

47. Yoshida, K., et al., *Germ cell regeneration-mediated, enhanced mutagenesis in the ascidian Ciona intestinalis reveals flexible germ cell formation from different somatic cells.* Dev Biol, 2017. **423**(2): p. 111-125.

48. Kubo, A., et al., *Genomic cis-regulatory networks in the early Ciona intestinalis embryo.* Development, 2010. **137**(10): p. 1613-23.

49. Oda-Ishii, I., et al., *A Maternal System Initiating the Zygotic Developmental Program through Combinatorial Repression in the Ascidian Embryo.* PLoS Genet, 2016. **12**(5): p. e1006045.

50. Mahen, R., et al., *Comparative assessment of fluorescent transgene methods for quantitative imaging in human cells.* Mol Biol Cell, 2014. **25**(22): p. 3610-8.

51. Niu, W., et al., *Diverse transcription factor binding features revealed by genome-wide ChIP-seq in C. elegans.* Genome Res, 2011. **21**(2): p. 245-54.

52. Samuel, A., et al., *Otx2 ChIP-seq reveals unique and redundant functions in the mature mouse retina.* PLoS One, 2014. **9**(2): p. e89110.

53. Wright, D.A., et al., *TALEN-mediated genome editing: prospects and perspectives.* Biochem J, 2014. **462**(1): p. 15-24.

54. Sasakura, Y., K. Yoshida, and N. Treen, *Genome Editing of the Ascidian Ciona intestinalis with TALE Nuclease.* Methods Mol Biol, 2017. **1630**: p. 235-245.

CHAPTER 2

Mouse Pou4 Proteins Regulate Dissimilar Sets of Genes in the Ascidian *Ciona robusta* and Paradoxically Can Substitute for the Endogenous Ascidian Pou4

INTRODUCTION

Class IV POU genes have been involved with the differentiation of ciliated mechanoreceptive cells since at least the last common bilaterian ancestor [1]. *Pou4* genes are expressed in a wide variety of neuronal cell types in all organisms examined; principally Pou4 genes encode transcription factors (TFs) and are expressed in neural tissues during embryonic development [2-6]. They also play roles in germ layer formation [7], cardiogenesis [8], and gonadogenesis in several organisms [9], but much less is known about *Pou4* function in these cells types and tissues. Most invertebrates have a single copy of their *Pou4* gene; in the invertebrate chordates *Branchistoma floridae* and *Ciona robusta*, their single Pou4 genes are expressed in differentiating ciliated sensory neurons, caudal bipolar neurons, chemosensory neurons, as well as several features of the central nervous system (CNS) [10-12]. In organisms such as the elephantfish *P. kingsleyae* and the hydrozoan *C. sowerbyi* where gene copy has expanded to between three and six known Pou4 genes, respectively (due to gene duplications and genome duplications plus gene loss [13, 14]) their roles have expanded spatially into the heart and gonads, representing examples of neofunctionalization [15].

Mammalian Pou4 genes, of which there are three (*Pou4F1* [*Brn3a*], *Pou4F2* [*Brn3b*] and *Pou4F3* [*Brn3c*]) display both distinct and sometimes overlapping roles in the CNS [16-18]. The proteins share highly conserved POU Domains (DNA-binding) and POU-IV Boxes (putatively protein-protein interaction) [19], bind to almost identical DNA sequences [20], and can functionally substitute for one another [21-23]. All three mammalian *Pou4* genes are predominately expressed in sensory neurons [24-26]. One example of a territory utilizing all three *Pou4* genes is in the mammalian inner ear, where ciliated mechanosensory cells (inner hair

cells (IHCs) and outer hair cells (OHCs)) transmit externally-received impulses to adjacent bipolar neurons of the spiral ganglia in the organ of Corti [27, 28]. During inner ear development, the mouse MmPou4F1 and MmPou4F2 proteins are both present in neurons of the cochlear ganglia [25, 29], whereas MmPou4F3 is found exclusively in IHCs and OHCs [30]. In humans, *Pou4F3* is also known as the deafness gene *DFNA15*, as mutations disrupt normal IHCs/OHCs damaging the sense of hearing [31].

The swimming larva of the solitary ascidian *Ciona robusta* (*Ciona intestinalis Species B* [32]) has a comparatively simple peripheral nervous system composed of various epidermal neurons, including palp neurons and ciliated neurons of the larval trunk and tail [33-36]. *CrPou4* is expressed in all ciliated sensory neurons of the larval epidermis (CESNs) [10], and is required for caudal ESNs formation [37]. In addition to normally specifying ESNs in *Ciona*, this gene has the remarkable ability to convert presumptive epidermal cells into ESNs by forced expression in that territory [38, 39], placing it into the category of terminal selector gene [40]. The *C. elegans Pou4* gene *Unc-86* and the mammalian *Pou4F1* (*Brn3a*) genes have also been shown to act as terminal selector genes [41]. *CrMyT1* and *CrPou4* act at the top of the regulatory cascade in differentiating ESNs; downstream TFs include *Atonal* and *NeuroD* [42]. The genes directly responsible for the activation of *CrPou4* have not yet been identified [42, 43], but, much is known about earlier events such as how the midline territory from which ESNs arise is formed [44-46], how Notch signaling dynamics affect ESN patterning [38, 47], as well as post-specification aspects of differentiation such as the previously mentioned downstream proneural gene activation [39, 48]. Caudal ESNs (CESNs) are found along the dorsal and ventral midlines of the larval tail [34, 49] and contain properties of both inner ear ganglia as well as IHCs/OHCs [33]. Unlike mechanosensory hair cells, the mechanotransduction reception and impulse delivery

into the CNS occur within each CESN [33-35]. Despite the differences between ascidian and mammalian systems of mechanotransduction, ascidian ESNs share many aspects of development with mammalian IHCs and OHCs, such as patterning by Notch-signaling [34, 38], bHLH proneural gene expression [39], and putative micro-RNA interaction with Notch target genes [47, 48]. Differentiated CESNs also express structural gene homologs of those in differentiated mammalian hair cells [34, 50].

Direct *CrPou4* activators and downstream targets are currently unknown, yet expressing ectopic *CrPou4* in the ascidian larval epidermis will override the normal transcriptional program of those epidermal precursors and reprogram those cells into becoming ciliated neurons [38, 39]. In contrast to ascidians, however, the *Pou4* gene active in mammalian hair cells, *Pou4F3*, is not required for differentiation but only survival [51], while ectopic expression of *Pou4F3* in the region of non-sensory supporting cells does not induce formation of ectopic hair cells [52, 53]. Interestingly, the bHLH gene *Atoh1* can produce ectopic hair cells when ectopically expressed, and this gene appears to be at the top of the regulatory cascade specifying this cell type, directly activating *Pou4F3* [54, 55]. The atonal homolog *lin*-32 regulates the nematode *Pou4* gene *Unc-86* [56], and a similar situation exists in *Drosophila* [3], which interestingly is not how a previous study demonstrated *Pou4* and an *Atonal* ortholog interact in *Ciona* [39]. In that study *Pou4* was shown to regulate *Atonal*.

Here we report on the identification of putative *Pou4*-regulated genes in *Ciona* and investigate the ability of the *Ciona* and mouse *Pou4* genes to rescue CRISPR morphants of targeting *CrPou4*. We first identified putative CrPou4 target genes in RNA-SEQ experiments by comparing gene expression profiles between wild type and transgenic embryos expressing *CrPou4*. As CESNS represent only about 1% of the total cell number of the larvae, we reasoned

that ectopic expression of *CrPou4* to produce greater numbers of ESNs would help to identify putative target genes; a similar approach was previously used to identify putative target of the *Brachyury* gene in *Ciona* [57]. We next ectopically expressed each mouse *Pou4* gene independently and used RNA-SEQ and comparative transcriptomics to identify putative targets for each of the mouse genes. As we detail below, this approach allowed us to identify similarities and differences in gene expression that may help to explain the differences we observed in the ectopic expression phenotypes. Lastly, we investigated the ability of each of the four *Pou4* genes to rescue CRISPR-Cas9 *CrPou4* morphants. Surprisingly, each of the genes rescued the morphants to a similar extent, even though we observed differences in the ability of these genes to produce ectopic ESNs. Our studies suggest that the context in which these genes are expressed can dramatically alter their function and provides insight to the role of *Pou4* in ESN development.

MATERIALS AND METHODS

Animals, Immunostaining, and Imaging

Adult *C. robusta* were collected from marinas in Mission Bay, San Diego, and kept in a recirculating refrigerated tank under constant illumination allowing for the accumulation of gametes. Immunohistochemistry was performed as in Pickett and Zeller (2018) [37]. Whole mount *in situ* hybridization (ISH) was performed essentially as in Satou et al., 2001 [58], with one relevant modification. To prevent colorimetric reactions from taking place in the larval tunic, probes were "recycled": after the hybridization step, each probe in hybridization buffer was aspirated from the hybridized embryos and placed at -80°C to be used again. Recycling of probe/hyb was repeated 4-5 times until the tunic no longer stained. After colorimetric

development, embryos were fixed for 5 minutes in 4% paraformaldehyde to stop the reaction, washed 2X in PBST, rinsed in 100% EtOH to clear background staining (if needed), washed 1X in PBST, then transferred to 70% glycerol/PBST for imaging. All imaging was performed on a Zeiss AxioPlan 2E Imaging epifluorescence microscope. Brightfield images were captured with an AxioCam ICc1 color camera and fluorescent images were captured with a Hamamatsu ORCA Flash 4.0 monochromatic camera.

Molecular Cloning

Generation of ISH probes: cDNA from mixed stage embryos was used in conjunction with the particular oligo primers listed below in Table 1 to generate PCR products that were subsequently subcloned into pBS2K+. All plasmids were sequenced prior to use; each was linearized with Acc65I preceding riboprobe synthesis.

Ectopic expression of *Pou4* genes : EpiB::Pou4-YFP, which expresses *Ciona Pou4* cDNA fused in-frame with YFP throughout the embryonic epidermis using the EpiB promoter, was previously described in Chen et al 2011 [38]. Mouse *Pou4* cDNA sequences were kindly provided by Bill Klein and Xiuqian Mu. EpiB::MmPou4F1-YFP, EpiB::MmPou4F2-YFP, and EpiB::MmPou4F3-YFP were cloned in the same manner as Epi::CrPou4-YFP; mouse *Pou4* cDNAs were PCR amplified, restriction enzyme-digested, and subcloned into an empty Epi::__YFP vector.

Transgenes for *CrPou4* KD and rescue experiments: Generation of Epi::Cas9 and U6::gRNA transgenes described previously in Pickett & Zeller (2018) [37]; see Table 1 for gRNA target

sequences used to knockout *CrPou4*. SCP::H2YFP::AB was generated by cloning a Super Core Promoter (AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGTCGAGCCGAGTGGTTGTGCCTCCATAGAA [59]) into our pZapANX-H2YFP vector upstream of H2YFP to create pZapANX-SCP::H2YFP. 2.0kB of cis-regulatory DNA from the *Pou4* locus (henceforth referred to as "AB") was PCR-amplified from *C. robusta* sperm-extracted genomic DNA and cloned downstream of H2YFP to create (pZapANX) SCP::H2YFP::AB. SCP::CrPou4::AB, SCP::MmPou4F1::AB, SCP:: MmPou4F2::AB, and SCP:: MmPou4F3::AB were generated by digesting SCP::H2YFP::AB with Acc65I and BamHI (New England BioLabs) and removing the H2 fragment, whereupon each of the coding sequences of each *Pou4* gene was inserted upstream of the YFP sequence. All transgenes were sequenced prior to use.

Electroporations and RNA-SEQ

Electroporations were carried out essentially as in Zeller et al. 2006 [60], and electroporated embryos were left to develop at 16-18°C in 0.22µm-filtered seawater supplemented with penicillin/streptomycin (10 U/mL and 10µg/mL, respectively) for the required amount of time for each experiment. Developing embryos were immobilized with MS-222 (Sigma) at 12-15hpf. A biological replicate refers to an independent electroporation on a different day using different parental animals.

To perform RNA-SEQ experiments, zygotes were electroporated with 40µg of transgene DNA in a total volume of 400 µL and developed at 18°C for 22hrs. Embryos were collected and spun down at 5,000xG for 1 minute, seawater was removed, and 300 µL of ice-cold RNA-extraction buffer (200mM Tris pH7.5; 200mM NaCl; 100mM EDTA; 0.2% SDS) was added.

Tadpoles in buffer were transferred to a glass homogenizer and homogenized on ice. RNA was extracted from homogenate by phenol-chloroform extraction followed by chloroform extraction. RNA precipitation performed by adding 1/10 volume 3M NaOAc (pH 7.3) and 2.5 volumes 100% EtOH, placing at -80°C for 20 min., followed by centrifugation at 17,000xG for 10min at 4°C. Pelleted RNA was dried briefly and suspended in DEPC-treated DI water and 1X DNase I buffer and 10U DNase I for DNA removal (New England Biolabs). Following DNase I treatment, RNA was re-extracted beginning at the phenol/chloroform step. Each experiment containing 5 samples (Epi::H2YFP control, Epi::CrPou4, Epi::MmPou4F3, Epi::MmPou4F2, Epi::MmPou4F1) was performed in triplicate. Samples were kept at -80°C until sequencing. Samples were sequenced at the Next Generation Sequencing Core, Scripps Research.

To perform *CrPou4*-knockdown rescue experiments, transgene DNA was pooled together, precipitated, and dissolved in 0.77M Mannitol. Each experimental group was electroporated with 8µg of Epi::Cas9 and 4µg each of the two CrPou4-targeting U6::gRNA transgenes, plus 8µg of either SCP::CrPou4::AB, SCP::MmPou4F1::AB, SCP::MmPou4F2::AB, or SCP::MmPou4F3::AB for a total of 24µg in a final volume of 200 µL. The positive control group included 8µg of Epi::Cas9 and 8µg of SCP::H2YFP::AB, while the negative control group included 8µg of Epi::Cas9, 4µg each of the two *Pou4*-targeting U6::gRNA transgenes, and 8µg of SCP::H2YFP::AB.

Bioinformatics

Sequences were pre-processed with Trimmomatic [61] to remove contaminating adapter sequences and for quality trimming. Sequences were then checked with FastQC [62]. Reads were aligned against gene models using Salmon [63]. The gene models used were

from the Aniseed genome site [64] and correspond to the KH2012 model set that incorporates additional genes from the NCBI genome annotation that were missing in the Kyoto KH models ([65]; http://ghost.zool.kyoto-u.ac.jp/cgi-bin/gb2/gbrowse/kh/) from 2012. This model set was supplemented with several gene models that our own lab corrected from internal sequencing data.

RNA-SEQ reads were quantitated with Salmon [66] and DESEQ2 [67] was used to identify differentially expressed genes. EggNog Mapper [68] was used to assign GO terms to the gene models and GO enrichment analyses were performed using WebGestaultR package [69]. To make comparisons between *Ciona* proteins and published mouse and zebrafish datasets, we used reciprocal Blast/Diamond searches [70, 71] which reported 1:1 relationships as well as 1:2/2:1 and 1:3/3:1 and 2:3/3:2 relationships.

Statistics

For data presented in Table 2, Table 3, and Table S2, statistical analysis was performed using the R programming language (version 4.0.2). For the rescue experiments, a Shapiro-Wilk test [72] indicated that data was not distributed normally when we compared total counts of ESNs, so instead of an ANOVA analysis we performed a Kruskal-Wallis analysis [73]. We also performed a nonparametric multivariate analysis using the R package npmv (Nonparametric Comparison of Multivariate Samples) [74].

RESULTS

In *Ciona*, *CrPou4* is required for ESN development [37], and has also been shown to be sufficient for eliciting ESN specification and differentiation [38]. We used ectopic expression

experiments followed by RNA-SEQ to identify putative transcriptional targets of *CrPou4*. As previous reports demonstrated that TFs from distantly related bilaterians are able to functionally substitute for one another [75], we wondered if the same was true for the mouse *Pou4* genes and repeated the ectopic expression experiments with the mouse genes. When ectopically expressed in the epidermis of the *Ciona* embryo, the mouse *Pou4* genes could convert epidermal cells to putative ESNs, but not to the same extent as *CrPou4*. Surprisingly, all of the mouse *Pou4* genes were able to rescue a CRISPR *Pou4* morphant to a similar level of the *Ciona Pou4* gene. Our analysis suggests that when ectopically expressed, each *Pou4* gene regulated unique combinations of genes, with little overlap. The observation that each protein can rescue a *CrPou4* morphant, although producing different phenotypes when ectopically expressed, suggests that proper cell-type context is important for the ability of *Pou4* to function.

Pou4 proteins generate dissimilar phenotypes when ectopically expressed throughout the ascidian epidermis

We have previously shown that ectopic expression of *CrPou4* produces supernumerary ESNs [38]. Because the DNA binding domains are well conserved, we hypothesized that ectopic expression of the mouse genes in *Ciona* would also produce similar supernumerary ESNs phenotypes. To test this, we cloned the mouse *Pou4* genes downstream of the epidermally-active EpiB promoter and expressed the mouse Pou4 genes in embryos transgenically. A small portion of embryos were removed from the pool and subjected to immunohistochemistry. Images of representative embryos from each experimental group shown in Figure 1. After staining for cilia, we observed that each Pou4 protein generated a slightly different phenotype. EpiB::CrPou4 as previously shown produced embryos that have their entire epidermises converted to ciliated

neurons (Figure1A-1D). Both EpiB::MmPou4F1 and EpiB::MmPou4F2 produced supernumerary cilia but only within the neurogenic territory of the larval midlines (Figure1E-1H and 1I-1L, respectively). EpiB::MmPou4F3 produced a similarly-shaped embryos compared to EpiB::CrPou4 (Figure1M-1P), yet there appeared to be fewer fully-developed cilia (compare Fig1D with Figure1P).

Ectopic *CrPou4* expression produces supernumerary ESNs and Up-regulates Mammalian Hair Cell and Proneural Gene Homologs

In the wild-type embryo, CESNs represent about 1% of the total cell population. Because our previous experiments had demonstrated that ectopic expression of *Pou4* in the epidermis could convert those cells into ESN-like cells, we reasoned that this approach could allow us to identify putative downstream targets of *CrPou4*; a similar strategy was used to identify targets of the *Brachyury* transcription factor that is expressed in the larval notochord [57]. We ectopically expressed *CrPou4* in the larval epidermis using a transgene driven by the epidermal promoter from the *EpiB* gene [38]. We used DESEQ2 to identify differentially expressed genes (DEGs) between embryos ectopically expressing *CrPou4* and embryos expressing a GFP transgene control. We considered up- or down-regulated genes (relative to controls) to be expressed at log 2-fold changes of +/- 0.58 respectively with adjusted P values of < 0.05. Using these cutoffs, we identified 1568 up-regulated and 1217 down-regulated genes, respectively. We first determined if any of the genes known to be expressed in the larval PNS were up-regulated by *CrPou4* expression and indeed regulatory genes we and others had previously identified were up-regulated including *MyT1*, *Atonal*, *NeuroD* and *miR-124* (Table 2 summarizes these observations). In addition, *CrPou4* expression slightly up-

regulated the expression of *Notch* and the Notch ligand *Jagged*. Lastly, *CrPou4* expression up-regulated a number of marker genes that are expressed in various PNS cell types including the pan-PNS genes *Gelsolin* [38, 76] and *beta-thymosin* [34], the palps marker *Islet* [77], and the BTN markers *Asic* and *Neurogenin* [12, 78]. Thus genes normally expressed in at least three different PNS cell types could be detected. Although a recent single cell sequencing study identified genes expressed in other PNS cells types such as the RTENs and ATENs [79] we could not identify genes that specifically identified those cell types.

We used EggNog Mapper [68] to assign Gene Ontology (GO) terms to each gene model and the WebGestaultR package [69] to identify over-represented GO terms from embryos ectopically expressing *CrPou4*. Figure 2 depicts a number of GO terms that were enriched in the embryos and that relate to possible roles in the PNS; all of which showed enrichment ratios greater than two and FDR <=0.05. Importantly, a number of categories relevant to PNS develop were identified including otic vesicle morphogenesis (GO:0071600), otic vesicle development (GO:0071599), inner ear receptor cell differentiation (GO:0060113) and a number of other GO terms related to neuronal function and morphogenesis. Additionally, a number of GO terms associated with other known expression domains of the vertebrate *Pou4* genes were also enriched including several cardiac-related GO terms (GO:0099622, GO:0086065, GO:0086003) [8, 80], and eye-associated GO terms (GO:0045494, GO:0050962, GO:0042675) [81]. The full list of identified GO terms can be found in Table S1.

To gain additional insight into the genes regulated by *CrPou4*, we compared our DEG list to several published datasets for both mouse and zebrafish hair cell transcriptomes (Barta et al., 2018; Y. Li et al., 2018; Scheffer, Shen, Corey, & Chen, 2015) and from a microarray

study that identified putative target genes of mouse *Pou4F3* (Hertzano et al., 2004). Table 2 provides a summary of the findings; the full analyzed dataset can be found in Table S2. We used Diamond [71] to compare the *Ciona* and mouse or zebrafish proteomes and reciprocal best hits to establish orthologies. Depending on the data set analyzed, we were able to map between 16% and 40% of the mouse/zebrafish genes to *Ciona*. Of these mapped genes, between 17% and 30% were either up or down regulated in embryos ectopically expressing *CrPou4* with the majority of those genes (50%-80%) being up-regulated (Table 3). We identified a number of genes such as *Atonal*, *MyT1*, and *Neurexin* that are present in the vertebrate datasets and up-regulated by *CrPou4*; some of these genes have been previously implicated in ESN development in *Ciona* [39]. Unvalidated genes such as *Cfap52* (*cilia and flagella associated protein 52*), *Phox2b*, and *Whirlin*, as well as dozens of others that were present in both a vertebrate dataset and the *CrPou4* up-regulated list will serve as a source of putative ESN genes for future study (Table 2 and Table S2).

To independently validate the identification of up-regulated genes in the RNA-SEQ analysis, we performed ISH against several candidates (Figure 3). In this examination of putative ESN genes, we detected expression of nine homologs in various neuronal cell populations of 22hpf larva. Five of these genes (*Fgf11/12/13/14*, *Myosin VII-a*, *Neurexin*, *Cadherin-23*, and *Usherin*) were expressed in ESNs (Figure 2A). Four up-regulated putative *CrPou4* targets (*FoxD*, *Lhx3*, *Synaptotagmin XI*, and *Retinal homeobox*) produced neuronal expression patterns, such as expression in regions of the CNS or palp sensory neurons (PNS), but were not visibly detected in the CESNs (Figure 2B). Several putative targets did not produce interpretable ISH patterns (such as *Foxc2* and *Olfr*; data not shown). We are currently determining the reasons for lack of signal, e.g. perhaps those genes are natively expressed at a

different time point yet were still activated by *CrPou4*.

Highly Variable Genes Up- and Down-regulated by Individual Pou4 Proteins

To gain insight into why the mouse *Pou4* genes produced different phenotypes when ectopically expressed in the epidermis, we performed RNA-SEQ analysis on the mouse Pou4 genes to examine their transcriptomes and performed comparisons with the *Ciona* Pou4 gene expression data including a GO term analysis. *Pou4* genes in mammals can functionally substitute for one another [21], so although phenotypes from the ectopic expression of the four *Pou4* genes varied, they each appeared to produce supernumerary cilia, and we expected there to be a substantial set of genes commonly up- and down-regulated by the four proteins. However, this common set of genes was actually quite small. Table 2 summarizes how ectopic expression of each of the mouse *Pou4* gene effects the genes known to be important for PNS specification. Surprisingly, the expression of the mouse *Pou4F2* gene reduced expression of the PNS markers *Gelsolin* and *beta-thymosin* and all three mouse genes reduced expression of the palp marker *crystallin-β*. *MmPou4F2* also reduced expression of *Atonal* and endogenous *CrPou4* which may help to explain why it did not efficiently produce ectopic ESNs outside of the embryonic midlines. *MmPou4F3* did not have as significant reduction in *Gelsolin* expression as compared to *MmPou4F2*. *MmPou4F2* was the only mouse gene that produced a reduction in *EpiB* expression (like *CrPou4*) indicating that the other mouse *Pou4* genes were not as effective in converting epidermal cells to another cell type. Given the initial threshold values of 1.5X up- or down-regulated compared to controls with an adjusted P value of less than 0.05, there was a cumulative 2369 *C. robusta* genes up-regulated and 3435 genes down-regulated by the *Pou4* genes (Table 3). From the set of 2369 up-regulated genes, only 34 were

up-regulated by all four Pou4 proteins, and out of the 3435 down-regulated genes only 79 were commonly down-regulated (Table S1). Many genes were up- or down-regulated by a particular *Pou4* gene, for example of the 1568 genes up-regulated by *CrPou4*, 830 of those were not up- or down-regulated by any of the mouse *Pou4* genes (Table 3). Compared to the responses of *CrPou4*, *MmPou4F2*, and *MmPou4F3*, ectopic expression of *MmPou4F1* resulted in far fewer up- and down-regulated genes. However, although *MmPou4F1* elicited the down-regulation of only 119 genes compared to controls, for example, 89 of those genes (74.8%) were also down-regulated by *CrPou4*. In all of our comparisons of gene expression between *CrPou4* and the mouse *Pou4* genes, it was not immediately obvious what changes in gene expression between these genes were most critical for the differences in phenotypes produced by the various proteins. Some possible candidate genes include *Nkx-C* and an *Orphan bHLH* transcription factor which have been shown to be important for midline specification in *Ciona* [44], both are down-regulated in *MmPou4F2* expressing embryos. *MmPou4F2* expressing embryos also had reduced expression of *Hes-C* (which could affect Notch signaling), *FGF11/12/13/14* (which is expressed in ESNs) and *CrPhox* (which has no known roles in ESNs). Unlike MmPou4F2, MmPou4F3 expressing embryos had increased levels of several putative transcriptional repressors (*Emc*, *Emc2*, *Ripply*) and the transcription factors *RX* and *HesB* which were also up-regulated in embryos ectopically expressing *CrPou4*.

Because the mouse *Pou4F1* and *Pou4F3* genes regulated far fewer genes than the mouse *Pou4F2* gene, our GO term analysis only identified enriched terms for the embryos ectopically expressing mouse *Pou4F2* (Table S1) and none of these GO terms included inner ear or mechanoreceptor-related terms. We next mapped genes with GO terms from each dataset (*CrPou4* and the three mouse *Pou4* genes) and counted the number of genes that were

up or down-regulated for the GO terms enriched in the embryos ectopically expressing *CrPou4* (Table S2). For all GO terms, there were more genes up-regulated by *CrPou4* then for any of the mouse genes; generally, the number for up-regulated genes followed the patterns *MmPou4* F2 > *MmPou4F3* > *MmPou4F1*. When examining the number of down-regulated genes within the enriched GO terms, in almost all cases *MmPou4F2* down-regulated more genes than *Ciona Pou4*. These results support the hypothesis that the mouse *Pou4* genes produce different phenotypes when ectopically expressed as a result of the differing numbers of genes regulated compared to *Ciona Pou4*.

Generation of a Minimal Transgene Designed to Express *Pou4* Genes ESN-Specifically

Ectopic expression of *Pou4* induces forced conversion of epidermal cells into ESNs [38], thus one of the obstacles we faced when examining *Pou4* rescue abilities was how to deliver the *Ciona* and mouse Pou4 proteins in an ESN-specific manner. Our lab has previously generated an ESN-specific transgene reporter generated from *Gelsolin* cis-regulatory DNA (KH.C9.512) [38]. However *gelsolin* is expressed too late in larval development to be relevant. Therefore we developed a streamlined set of CRMs that faithfully recapitulated the endogenous *Pou4* expression pattern.

We began by generating a VISTA plot [82] comparing the *CrPou4* locus to the *C. savignyi Pou4* locus (Figure3) whereupon we observed several conserved non-coding elements (CNEs), primarily the 1.8kB of DNA upstream of the transcription start site (TSS), the first intron, and approximately 3kB of non-coding DNA downstream of the coding region. Our first transgenes contained the 1.8kb CNE driving YFP. This transgene directed YFP expression in the mesenchyme, a known hotspot of ectopic expression of transgenes (unpublished

observations). Summarily we extended the 5` CNE to include 3.0kb of upstream CNE, which produced indistinguishable results from the first transgene. Next we extended the 1.8kb fragment 3` to include the first intron. This transgene drove YFP expression in the dorsal midline, suggesting this intronic CNE likely acts as an enhancer element to the *CrPou4* gene. We then tested sections of the 3` CNE with a transgene composed of a supercore promoter [59], followed by the YFP coding sequence, then by 3' CNE sections positioned in their native 3' position relative to the coding sequence. Our VISTA analysis demonstrated that the 3` CNE is divided into several peaks (Figure 3), and we divided the 3kb of CNE into three 1kb sections. Briefly, the first 1kb CNE and the third 1kb CNE did not on their own produce YFP expression, while incorporating the middle 1kb CNE fragment directed YFP expression in the ESNs, but weakly (data not shown). Finally, we tried the first two 1kb CNEs together, which drove robust YFP expression in the larval ESNs (Figure 3). This was likely due to the inclusion of the *CrPou4* 3` UTR which makes up the first 700bp of the 3` CNE [83, 84]. Having derived a robust ESN-specific reporter, we could now clone each mouse *Pou4* gene into the location of H2 and have the four *Pou4* genes expressed only in the cells of interest.

Mouse Pou4 Proteins Rescue a CrPou4 KD Phenotype

During ascidian larval development, the *Pou4* gene is required to properly differentiate ESNs; the phenotype of eliminating *CrPou4* is loss of ESN cilia [37]. We reasoned it was unlikely that the mouse *Pou4* genes would be able to rescue the null phenotype and restore ESN number, due to the observation that 1) the mouse Pou4 proteins up- or down-regulated quite different sets of genes compared to CrPou4, and 2) that when overexpressed the resultant mouse Pou4 phenotypes did not closely resemble the Epi::CrPou4 phenotype. We were surprised to

observe, however, that each MmPou4 protein functionally substituted for the absence of CrPou4, restoring the quantity of cilia nearly as well as the native CrPou4 did.

Prior to performing the rescue experiments, we first assessed if the rescue constructs themselves altered the number of ESNs by comparing the responses of SCP::H2YFP::AB to SCP::CrPou4::AB, delivering the same quantity (8µg/200uL) of those constructs as we planned to in the rescue experiments. We found that SCP::CrPou4::AB, which expresses *CrPou4* in an ESN-specific manner, did not alter ESN number vs. SCP::H2YFP::AB (data not shown).

We tested each of the mouse *Pou4* genes ability to rescue the null phenotype by using CRISPR-Cas reagents targeting *CrPou4* while coexpressing the mouse *Pou4* genes. To target *CrPou4*, we expressed Cas9 throughout the epidermis and expressed gRNAs ubiquitously with the *CrU6* promoter. gRNAs targeted the *CrPou4* TSS, as well as a location within the first intron, both locations of which do not exist in SCP::CrPou4::AB. To perform the rescue experiments, it was critical that we express the genes in the proper spatiotemporal manner, thus *CrPou4*, *MmPou4F1*, *MmPou4F2*, and *MmPou4F3* were expressed exclusively in the ESNs using specific cis-regulatory DNA from the *CrPou4* gene (see methods and Figure 3). In our negative control group electroporated with 8µg of SCP::H2YFP::AB and 8µg of EpiB::Cas9, the average number of cilia per embryo was 16.2. In the positive control group which included SCP::H2YFP::AB, 8µg of EpiB::Cas9, and *CrPou4*-targeting gRNA transgenes, average number of cilia decreased to 10.4.

Knocking down Pou4 and including SCP::CrPou4::AB in electroporations almost fully eliminated the decrease from the CRISPR reagents and restored cilia numbers to an average of 15.5 per larva. Including SCP::MmPou4F3::AB with *CrPou4* targeting regents resulted in an average of 14.7 cilia per larva, including SCP::MmPou4F2::AB resulted in an average of 14.5

cilia per larva, and including SCP::MmPou4F1::AB resulted in an average of 14.0 cilia per larva. In summary, delivering CRISPR reagents reduced average cilia per tadpole from approximately 16 to 10, while including any of the *Pou4* genes generated cilia numbers of around 14-15 per larva (four biological replicates).

To examine whether there were any statistically significant differences between the activities of the four Pou4 proteins we performed a Kruskal-Wallis statistical test of cilia generated by each experimental group (Figure 4). Data for cilia per embryo was collected for dorsal and ventral quantities, and using the Kruskal-Wallis statistical test we were able to determine whether there were any differences between conditions for numbers of either dorsal, ventral, or total cilia. In all three scenarios, numbers of cilia in the control group ("YFP" in Figure 4) was different than all other conditions ($p<0.001$) and that the KD group was also different than the other conditions ($p<1X10^{-6}$). When comparing all four conditions, F1, F2, F3, and P4 were all not significantly different from one another, while in pairwise comparisons F1 was different than P4 to a statistically significant degree ($p<0.05$). For ventral cilia only, F2 was different than P4 to a statistically significant degree ($p<0.05$).

DISCUSSION

We have shown through an electroporation based approach that the sets of genes up- or down-regulated by the four *Pou4* genes varies considerably. Results show that only a handful of genes are similarly regulated between the four *Pou4* genes examined in this study. Out of the 2369 unique genes up-regulated by any of the four *Pou4* genes, only 34 were commonly up-regulated. Additionally, from the 3435 unique genes down-regulated by any of the four *Pou4* genes, only 79 were commonly down-regulated. These findings stand in contrast to the results

showing that each of the three mouse Pou4 proteins can rescue a *CrPou4* KD to the same degree that CrPou4 can, restoring cilia number comparable to control group levels.

These approaches allowed us to generate a set of genes we could examine the ISH patterns of and determine if they are expressed in mature ESNs. The list of genes up-regulated by *CrPou4* totaled 2369. Cross-referencing that list with previously documented genes involved in vertebrate hair cell development and regulation allowed us to narrow down the set from which we could examine ISH data. We demonstrated expression in ESNs of five previously unreported expression patterns of genes, and four more genes involved in non-ESN neuronal cell types. These findings advance this line of research towards generating a comprehensive landscape of genes involved in developing and maintaining ascidian ESNs.

Due to the mouse Pou4 proteins binding the same DNA sequence [20], we expected there to be a large set of the same genes up- and down-regulated by the four proteins. However, this common set of genes was actually quite small, while the expression of genes differentially affected by CrPou4 and each of the mouse Pou4 proteins dominated the lists. While there were many genes up- or down-regulated between one, two, or three of the *Pou4* genes, in contrast, genes exclusively up- or down-regulated by each of the *Pou4* genes dominated. For example, 65% of the genes up-regulated by *CrPou4* were not up-regulated by either of the three mouse *Pou4* genes.

In *Pou4* KD experiments we have performed previously, our CRISPR-Cas reagents were able to reduce the number of cilia to an average of ~7 per larva (data not shown). However, in this study the *Pou4* rescue experiments required the electroporation of four separate constructs. Due to the empirically-determined maximum of 25µg plasmid DNA/200ul electroporation volume [60], we were forced to include less DNA per construct. We believe this played a part in

only reducing cilia averages to 10.4 when targeting *CrPou4*. Additionally, when designing gRNAs against *CrPou4*, we were restricted to sequences that are not included in the SCP::CrPou4::AB transgene. Nevertheless, including a rescue construct while knocking down *CrPou4* clearly rescued the knockdown phenotype.

The discrepancy between the highly variable sets of genes each Pou4 protein up- and down-regulated and the apparent ease with which mouse Pou4 proteins could rescue the *CrPou4* null phenotype was initially perplexing. After analyzing the RNA-SEQ data we assumed that the mouse Pou4 proteins would vary in their capacity for rescue, and that perhaps we would see a correlation between the number of genes commonly up- and down-regulated with *CrPou4*, and each protein's ability to rescue the null phenotype. We found this was not the case, as although the RNA-SEQ results varied widely, there was little difference between the degrees to which either of the four Pou4 proteins could rescue the decrease in cilia produced by targeting the *CrPou4* locus.

Several previously published experiments may help to explain the discrepancy between the genes up- and down-regulated by each Pou4 protein and their respective abilities to rescue the CRISPR morphants. Firstly, an experiment was performed by Shi, et al. (2013) whereby the cDNA sequence of the mouse *Pou4F1* gene was knocked into the *Pou4F2* locus, creating a mouse that expressed no Pou4F2 protein but instead expressed Pou4F1 in a spatial and temporal manner as Pou4F2 normally is in the wildtype mouse [22]. Interestingly, the mouse was entirely normal, displaying no detectable developmental or behavioral abnormalities. This demonstrated that the mouse Pou4 paralogs are functionally equivalent in that context, even though they vary considerably outside of their DNA binding domains (Figure S1A). Second, in a separate study, it was shown that a distantly-related *Pou4* gene from the echinoderm *Strongylocentrotus*

purpuratus could, if knocked-in to the mouse *Pou4F2* locus thereby replacing that gene, completely rescued the null phenotype. The *S. purpuratus* Pou4 protein completely restored RGC quantity and gauges of vision to normal levels [75], compared to the *Pou4F2* KO mouse and wildtype mouse. Therefore, although mouse Pou4 proteins function to regulate different genes and are active in different cell types, one Pou4 protein may functionally substitute in the absence of another.

Two quite opposite descriptions of the Pou4 proteins used in this study may also help to reconcile the difference between gene up- and down-regulated and ability to rescue. When examining conservation between Pou4 proteins, what stands out most clearly is the nearly 100% conservation between all four Pou4 proteins in their POU Domains, and, several smaller sections of lesser conservation including the functionally important POU-IV boxes [19] (Figure S1A). However, it can also be noted that there are large stretches of each protein that are conserved between orthologs but not paralogs, i.e., amino acids not conserved between MmPou4F2 and MmPou4F1 that are conserved between MmPou4F2 and the Xenopus Pou4F2 (Figure S1B). These regions vary widely between each mouse Pou4 protein. That the areas outside the conserved regions between orthologs are conserved within paralogs argues for functional importance, and perhaps functional differences. The mouse Pou4 proteins share a highly conserved POU domain yet each have their own specific and probably functionally important amino acid sequences outside of the highly conserved regions. Therefore, in conclusion, we suggest attributing each protein's ability to rescue the null phenotype to their highly conserved regions, while attributing the variations in the genes up- or down-regulated to the regions of each protein outside of the highly conserved domains, as a way to reconcile our disparate findings. Assays involving chimeric Pou4 proteins could begin to address this hypothesis. For example, if

MmPou4F1 and MmPou4F2 had their inter-conserved areas exchanged, one could test the hypothesis that the MmPou4F1 protein with MmPou4F2 inter-conserved sequences should generate sets of genes that correspond more closely to the sets of genes produced in this study by MmPou4F2 than MmPou4F1.

Tunicates diverged from the lineage that led to the mammals approximately 550 million years ago [85, 86]. Since then, they have evolved dramatically different life histories and body plans, both derived from divergent genetic instructions and divergent cell types. Cis-regulatory changes are regarded as largely what accounts for the diversity of animal life [87, 88], but it is nevertheless interesting to reflect on the fact that the proteins involved in producing such differences, e.g. the Pou4 proteins examined here, were very similar to the proteins present in their common ancestors, likely performing similar activities as they do today. Furthermore we can partially demonstrate that fact through studies of functional conservation such as those we have performed in this study. While our results show mammalian Pou4 proteins activate disparate genes in the context of the ascidian embryo, they have nevertheless retained the ability to substitute for a protein they haven't shared a genomic landscape with in over 500 million years.

ACKNOWLEDGEMENTS

Clifford Pickett gratefully acknowledges ARCS Foundation support, San Diego Chapter. This work, Chapter 2, "*Mouse Pou4 Proteins Regulate Dissimilar Sets of Genes in the Ascidian* Ciona robusta *and Paradoxically Can Substitute for the Endogenous Ascidian Pou4*; C.J. Pickett, Victoria Hurless, Tiffany Hoang, and Robert W. Zeller", in full, is being prepared for publication. The dissertation author was the primary investigator and author of this material.

Table 2.1: Gene model identifiers and oligo sequences used for genes examined by ISH and gRNA target sequences used in CRISPR-Cas targeting of *CrPou4*.

Gene	KH Model	Forward Oligo	Reverse Oligo
Pou4	KH.C2.42	ATGTTTACTAACATGCTTGCTCCACAC	CAGCATTTAATGGGGACGTGATTATG
Myosin-VIIa	KH.C5.147	ATGTCTCGCCAAATCTTTGTGAGACA	TGGTTTTAAGCCTCTTACAATCATACGGTA
cadherin-like 23	KH.L62.2	GGCACCGAACTACGATTTGTGTCCCA	CATCACAATTATGTCAATTATGACATCATCACTTATG
neurexin-2 isoform x14	KH.C3.273	ATGCTCTGCTGGACTACTGCTATGCC	GATTGCGGCAGTAGGGTCCAATAT
Usherin	KH.L170.57	TTTGTGACGGATTTTGGTGATGCTGATAATG	TAAATGTTTGTAGCTCCCCATTGACTTGAG
Fibroblast Growth Factor 11/12/13/14	KH.L28.8	ATGGATAAACTAAAAGTCGGCGGGC	TAAATTCTTTCTCTTCTTATTATCCGATG
Forkhead box D3	KH.C8.890	ATGATGACAGTGCAGTGTTGTGTGCA	CTAATGTTGGAACTGTGGGGAAGAAACTGT
Retinal homeobox factor	KH.C12.152	ATGAGTACAGACACATCTAAAGGTGAAG	CTGCTGACTCAAAACTGGAACTTTTTC
Lim homeobox Lhx3-like 2	KH.S215.4	ATGATTCTCGATACTAAAGCGCTCGATG	TTGGAAATGTGTCACGTGGTCAAG
Synaptotagmin XI	KH.C4.304	ATGTGGTCCCAACTGAAAGAAAGCG	AAAAGGTAAACTCAATAAGTGCCAAAATGTGAC

gRNA Name	Target Location	Target Sequence
CrPou4 gRNA1	TSS	ACGTTCATATTATTATCTAT
CrPou4 gRNA2	intron 1	ATTACATGTTGCGTCGGACA

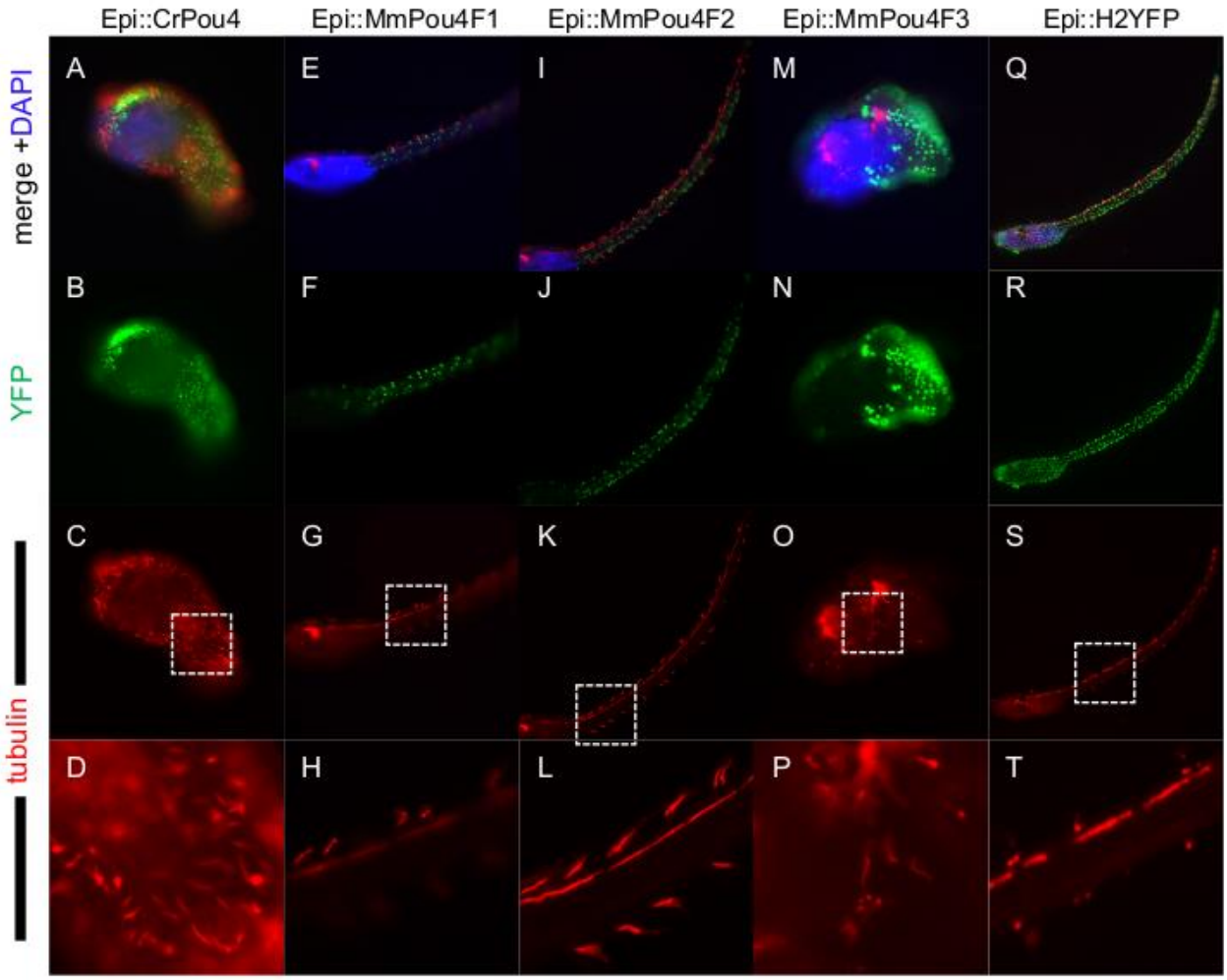


Figure 2.1: Ectopic expression of *C. robusta* and mouse *Pou4* genes converts presumptive epidermal cells into ESNs. Each image represents the dominant phenotype of expressing either of the four Pou4 genes throughout the embryonic epidermis. Epi::CrPou4 resulted in a conversion of the entire epidermis into ESNs (A-D) while Epi::MmPou4F1 (E-H) and Epi::MmPou4F2 (I-L) resulted in an increase of ESNs exclusively within the dorsal and ventral larval midlines. The overall morphology of Epi::MmPou4F3 larva appeared similar to Epi::CrPou4 larva, but Epi::MmPou4F3 larva exhibited fewer acetylated tubulin-positive epidermal cells (P). B, F, J, N, and R display YFP fluorescence demonstrating that each TF-FP was present throughout the epidermis. C, G, K, O, and S display fluorescently labeled acetylated tubulin. D, H, L, P, and T are magnified boxed areas from each corresponding 10X-magnified images in the row above. Blue, nuclear DAPI; Green, epidermally-expressed *Ciona* and mouse Pou4-YFP fusion proteins; Red, acetylated tubulin (cilia).

Table 2.2: CrPou4 up-regulates genes with functions related to neural development. Many genes that are involved with PNS development (left column) are up-regulated in response to *CrPou4* ectopic expression. Rightmost eight columns display fold-change (P4/F1/F2/F3_log2FC) vs. controls and adjusted P values (P4/F1/F2/F3_pAdj) for the corresponding change resulting from ectopic expression of *CrPou4* and the mouse *Pou4* genes.

Gene_Name	GeneBase	P4_log2FC	P4_pAdj	F1_log2FC	F1_pAdj	F2_log2FC	F2_pAdj	F3_log2FC	F3_pAdj
Early Acting Genes									
Bmp2/4 (bone morphogenetic protein 2/4)	KH.C4.125.	-5.80E-01	6.02E-06	1.97E-02	9.71E-01	-3.05E-01	3.46E-02	-2.37E-01	1.93E-01
Jag (jagged)	KH.C11.186.	6.41E-01	2.22E-02	-1.57E-01	8.23E-01	-1.08E-01	7.70E-01	1.89E-01	6.99E-01
Nodal (nodal growth differentiation factor)	KH.L106.16.	4.78E-01	3.51E-02	1.84E-01	7.12E-01	5.24E-01	1.95E-02	9.41E-01	7.82E-06
ADMP	KH.C2.421.								
Fgf9/16/20 (fibroblast growth factorf 9/16/20)	KH.C2.125.								
msxb	KH.C2.957.	3.53E-02	9.40E-01	2.26E-01	7.05E-01	-8.18E-01	2.67E-03	-5.79E-01	8.04E-02
Ascl.a (achaete-scute family bHLH transcription factor.a)	KH.L9.13.	8.46E-01	7.09E-02	2.07E+00	5.75E-06	1.41E+00	9.11E-04	2.46E+00	4.68E-10
Zinc Finger (C2H2)-24 // Ci-ZF265 (KLF1/2/4)	KH.C5.154.	2.28E-01	5.25E-01	1.72E+00	7.50E-12	4.88E-01	9.16E-02	1.18E+00	3.57E-06
Tox (thymocyte selection associated HMG box)	KH.C3.330.	3.14E-01	2.55E-02	1.64E-01	5.44E-01	3.96E-01	3.72E-03	6.68E-01	2.51E-07
Nkx-C	KH.C1.922.	-2.75E-01	5.43E-01	8.82E-01	6.26E-02	-7.85E-01	2.79E-02	7.59E-03	9.94E-01
Dlx.c (distal-less homeobox.c)	KH.C7.770.	-1.55E+00	3.12E-10	-6.79E-02	9.38E-01	-7.72E-01	4.80E-03	-9.75E-01	4.78E-04
Bhlhtun1 (Tunicate bhlh 1) (Orphan-1)	KH.C7.269.	-1.58E-01	5.99E-01	2.80E-01	4.97E-01	-7.09E-01	8.87E-04	-1.88E-02	9.73E-01
Dll (delta like canonical Notch ligand)	KH.L155.7.	6.98E-01	1.94E-02	9.94E-01	5.98E-03	1.35E+00	6.36E-07	1.78E+00	2.05E-11
myelin transcription factor // Ci-ZF223	KH.C1.274.	1.64E+00	9.75E-09	9.80E-01	1.90E-02	1.09E+00	4.38E-04	1.10E+00	7.58E-04
Pou4	KH.C2.42					-7.81E-01	9.83E-03		
Net (NeuroD)	KH.C9.872.	8.20E-01	2.88E-04	3.19E-01	4.80E-01	1.82E-01	5.39E-01	3.48E-01	2.94E-01
ATOH1,ATOH7,NEUROD6	CiAmos_C8.27_Trinity	3.50E+00	1.56E-21	5.19E-01	5.31E-01	-1.04E+00	2.04E-02	4.39E-01	5.16E-01
miR-124	KH.C7.140.	8.25E-01	4.85E-04	2.73E-01	5.79E-01	-4.59E-01	8.50E-02	-8.59E-02	8.76E-01
Markers (Pan PNS)									
b-thymosin	KH.C2.140.	1.01E+00	8.57E-06	-6.40E-02	9.35E-01	-1.52E+00	7.96E-12	-3.57E-01	2.96E-01
Gelsolin	KH.C9.512	1.74E+00	5.55E-32	-2.86E-01	3.45E-01	-1.23E+00	2.75E-15	-5.08E-01	6.27E-03
Markers (Palps)									
islet	KH.L152.2	8.06E-01	4.28E-09	1.62E-01	6.06E-01	-2.92E-01	8.05E-02	-6.97E-02	8.27E-01
Crytsallin beta (palps)	KH.S605.3	-2.72E-01	6.19E-01	-1.10E+00	4.53E-02	-3.90E+00	1.14E-29	-1.88E+00	4.60E-07
Markers (BTNs)									
Glutamate decarboxylase (aBTN)	KH.S761.6	6.34E-01	3.74E-02	3.50E-01	5.36E-01	3.93E-01	2.35E-01	4.22E-01	2.98E-01
asics	KH.C1.215	1.04E+00	6.48E-04	4.61E-01	4.28E-01	5.22E-01	1.36E-01	6.29E-01	1.12E-01
NeuroG	KH.C6.129	1.48E+00	1.08E-06	1.37E+00	2.53E-04	1.02E+00	2.07E-03	9.00E-01	1.49E-02
EpiB	KH.C7.154.	-1.63E+00	1.33E-07	-5.55E-02	9.65E-01	-7.48E-01	3.59E-02	-1.93E-01	7.67E-01
Others:									
Dlx.a (distal-less homeobox.a)	KH.C7.346.	1.35E+00	4.12E-06	-2.23E-01	7.73E-01	-2.42E+00	1.91E-14	-1.31E+00	9.71E-05
EVX1,EVX2,NKX3-1	CiEVXB_C3.836	5.96E-01	1.22E-01	-7.48E-03	9.95E-01	-1.82E+00	5.70E-08	-1.02E+00	8.66E-03
Fgf11/12/13/14 (fibroblast growth factorf 11/12/13/14)	KH.L28.8.	2.23E+00	3.19E-28	-4.40E-01	2.70E-01	-1.02E+00	5.18E-06	1.21E-02	9.84E-01
Gata.b (GATA binding protein.b) // Ci-ZF261	KH.S696.1.	2.46E+00	4.65E-13	6.68E-01	3.42E-01	8.44E-01	4.27E-02	4.31E-01	4.98E-01
Hes.b	KH.C3.312.	1.14E+00	8.19E-07	4.70E-01	2.73E-01	-4.75E-01	8.47E-02	6.33E-01	2.93E-02
Hes.c	KH.L34.9.	-4.69E-01	1.25E-01	1.92E-01	7.68E-01	-1.08E+00	4.64E-05	3.41E-01	3.97E-01
FGFR	KH.S742.2.	6.95E-01	2.92E-04	1.31E-01	7.94E-01	3.62E-01	9.86E-02	4.80E-01	4.12E-02
PHOX2B; PRRX1; PRRX2	KH.C1.414.	4.64E+00	5.28E-12	5.40E-01		7.80E-01	4.27E-01	8.14E-01	5.45E-01
APC	KH.C8.457.	1.38E+00	4.11E-06	3.79E-01	5.54E-01	5.79E-01	1.02E-01	2.93E-01	5.91E-01
ARX,PHOX2A,PHOX2B	CrPhox_C14.100.119	-3.87E-01	4.53E-01	-6.42E-01		-1.00E+00	2.05E-02	-6.22E-01	2.79E-01
ATBF // Ci-ZF113 // CiZF-313 // Ci-ZF314	KH.C3.443.	1.58E+00	2.32E-06	4.28E-01	5.50E-01	6.32E-01	1.12E-01	5.11E-01	3.21E-01
C/EBPa	KH.C3.644.	6.79E-01	1.56E-02	6.24E-01	1.48E-01	4.11E-01	1.83E-01	1.84E-01	7.16E-01
Chox10	KH.C11.689.	9.74E-01	5.38E-03	1.04E-01	9.30E-01	-1.37E-02	9.82E-01	3.51E-01	5.62E-01
Emc	KH.C7.692.	1.31E+00	9.34E-03	5.26E-01	6.15E-01	8.68E-01	1.14E-01	1.68E+00	1.18E-03
Emc2	KH.C7.157.	1.38E+00	4.92E-03	6.84E-01	4.62E-01	8.21E-01	1.33E-01	1.42E+00	8.69E-03
Foxb (forkhead box b)	KH.C4.341.	-1.05E+00	2.03E-03	3.97E-01	5.59E-01	-5.76E-01	1.30E-01	-3.33E-01	5.55E-01
Foxc (forkhead box c)	KH.L57.25.	1.54E+00	2.43E-05	5.54E-02	9.73E-01	-5.46E-01	2.55E-01	-6.41E-01	2.79E-01
Foxd.a (forkhead box d.a)	KH.C8.890.	2.14E+00	5.77E-04	1.51E+00		1.18E+00	1.00E-01	1.18E+00	1.72E-01
Foxd.b (forkhead box d.b)	KH.C8.396.	1.00E+00	9.85E-05	3.90E-01	4.48E-01	-1.15E-01	7.62E-01	-1.11E-01	8.55E-01
Fz3/6	KH.L9.43.	1.24E+00	5.70E-23	-9.36E-03	9.88E-01	-2.07E-01	2.15E-01	1.17E-01	6.45E-01
Groucho1	KH.L96.50.	1.66E+00	1.15E-12	2.21E-01	7.09E-01	3.59E-01	2.35E-01	2.23E-01	6.18E-01
HMX1,HMX3	CiHMX_S563.4	2.62E+00	7.41E-16	-5.67E-02	9.69E-01	-5.00E-02	9.32E-01	5.97E-02	9.49E-01
Lhx3/4 (LIM homeobox 3/4)	KH.S215.4.	1.64E+00	5.70E-23	7.40E-01	1.14E-03	3.73E-01	7.58E-02	8.58E-01	5.16E-06
Nkx-A	KH.C12.577.	-1.03E+00	1.33E-02	5.24E-01	4.90E-01	-5.20E-02	9.32E-01	-3.47E-01	6.10E-01
Nkx-B	KH.C8.709.	1.53E+00	6.19E-01	3.57E+00		3.59E+00	1.08E-01	3.18E+00	
Pax2/5/8-A	KH.S1363.2.	6.34E-01	6.13E-03	-3.57E-02	9.68E-01	-6.01E-01	1.29E-02	-6.14E-01	2.17E-02
Prx	KH.C1.414.	4.64E+00	5.28E-12	5.40E-01		7.80E-01	4.27E-01	8.14E-01	5.45E-01
ripply transcriptional repressor 1/2/3	KH.108949545	2.33E+00	3.16E-14	7.96E-01	1.65E-01	5.65E-01	1.59E-01	1.11E+00	3.54E-03
Rx	KH.C12.152.	3.18E+00	1.07E-22	9.38E-01	9.70E-02	7.64E-01	5.70E-02	1.20E+00	2.54E-03
Shox	KH.L112.34.	4.64E-01	6.37E-01	1.12E+00		1.02E+00	1.73E-01	3.19E+00	7.38E-08
Smad2/3.b (Smad family member 2/3.b)	KH.C12.26.	-6.22E-01	7.18E-04	-5.37E-02	9.28E-01	-5.54E-01	3.86E-03	-3.73E-01	1.19E-01
Tfap2-r.b (transcription factor AP-2-related.b)	KH.C7.43.	-1.03E+00	5.54E-04	4.08E-02	9.74E-01	-5.74E-01	8.72E-02	-2.36E-02	9.75E-01
Wnt3 (Wnt family member 3/3A)	KH.C9.27.	2.05E+00	1.75E-03	5.97E-01		3.05E-01	7.44E-01	-2.16E-01	8.93E-01

Table 2.3: Quantities of *C. robusta* genes up- or down-regulated in response to each *Pou4* gene expressed. Values represent genes with an adjusted P value (FDR-corrected) of 0.05 or less, and 1.5-fold over (up-regulated) or under (down-regulated) expression levels of corresponding genes in the control group. "Total Genes" is the sum of all genes up- or down-regulated by a particular Pou4 protein. "Genes Commonly Up- and Down-Regulated" are numbers of genes commonly regulated by any two of the Pou4 proteins. "Exclusive Genes" lists numbers of genes up- or down-regulated that were not up- or down-regulated by any other Pou4 protein than the one listed. "Total Unique" are the cumulative totals of different *C. robusta* genes affected by the four Pou4 proteins.

	Total Genes		Genes Commonly Up- and Down-Regulated						Exclusive Genes	
			MmPou4F3		MmPou4F2		MmPou4F1			
	up	down	up	down	up	down	up	down	up	down
CrPou4	1568	1995	311	743	434	1235	35	89	830	643
MmPou4F3	579	992			242	848	51	91	45	153
MmPou4F2	1026	2597					44	106	492	1040
MmPou4F1	76	119							7	7
Total Unique	2369	3435								

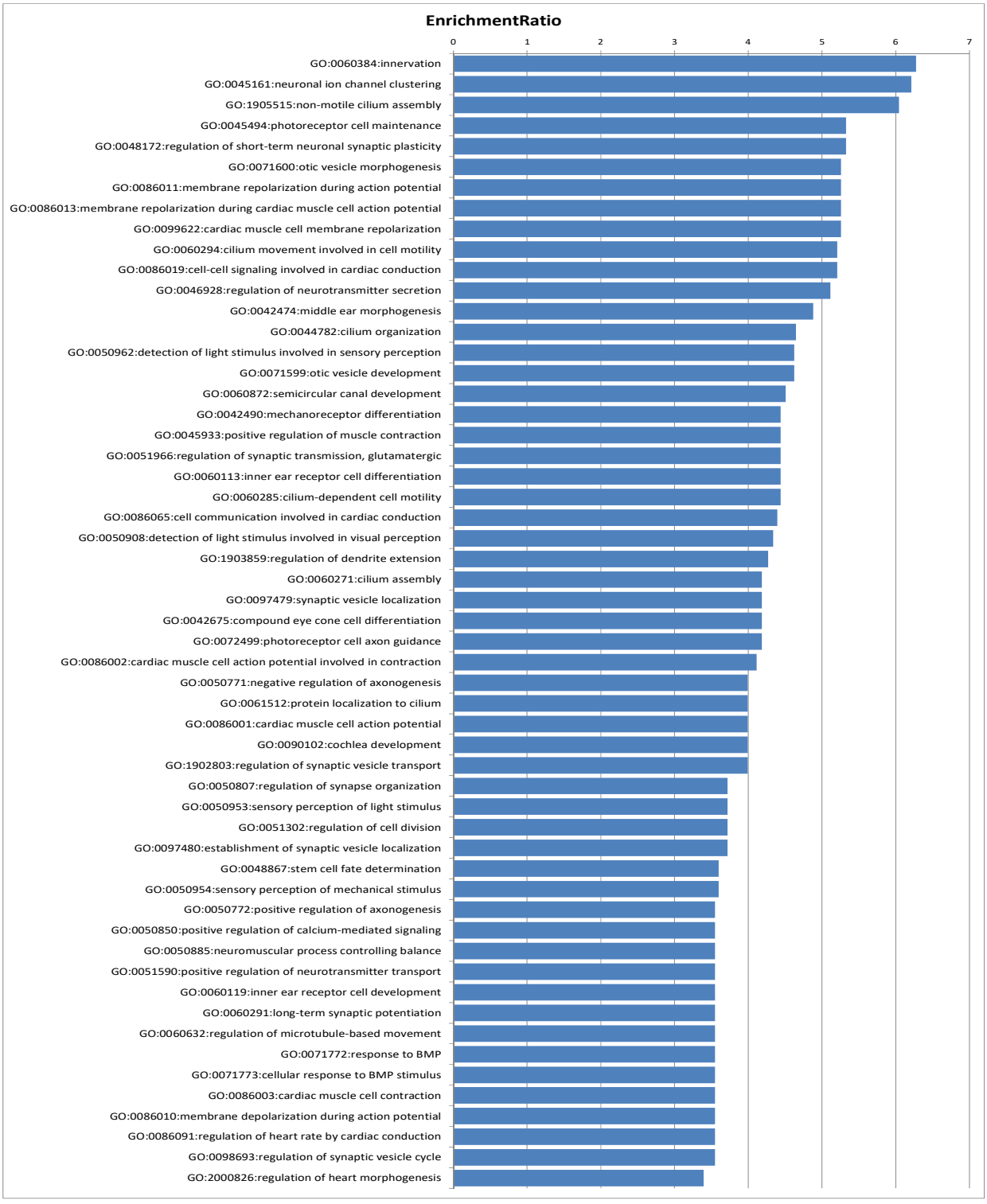


Figure 2.2: GO Term analysis of CrPou4 up-regulated genes displays enrichment of terms related to neurogenesis and neuronal function. GO terms were assigned using EggNog Mapper [68], and WebGestaultR package [69] was used to perform the GO enrichment analysis.

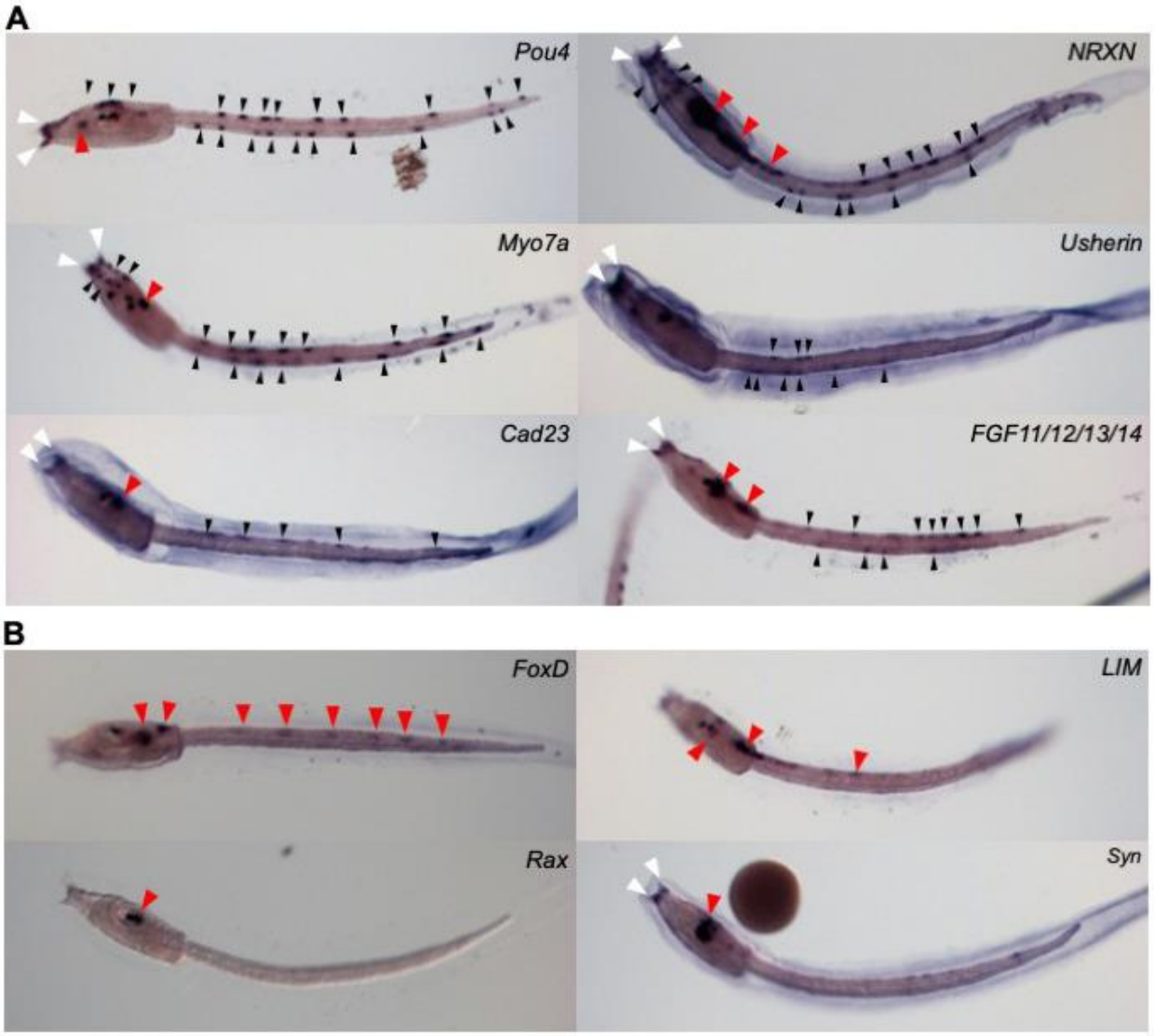


Figure 2.3: In situ hybridization patterns for neuronal genes up-regulated by ectopically expressing *CrPou4*. Top left: Representative image of 22hpf larva probed for *Pou4* expression. *Pou4* is expressed in the sensory palps (white arrowheads throughout figure), trunk epidermal ciliated neurons and ESNs (black arrowheads throughout figure), and also in the CNS (red arrowheads throughout figure). All other images are of representative larva displaying relevant putative *Pou4* gene target ISH results. All larva are displayed anterior to the left and dorsal to the top. All images are brightfield and were taken with a 10X objective.

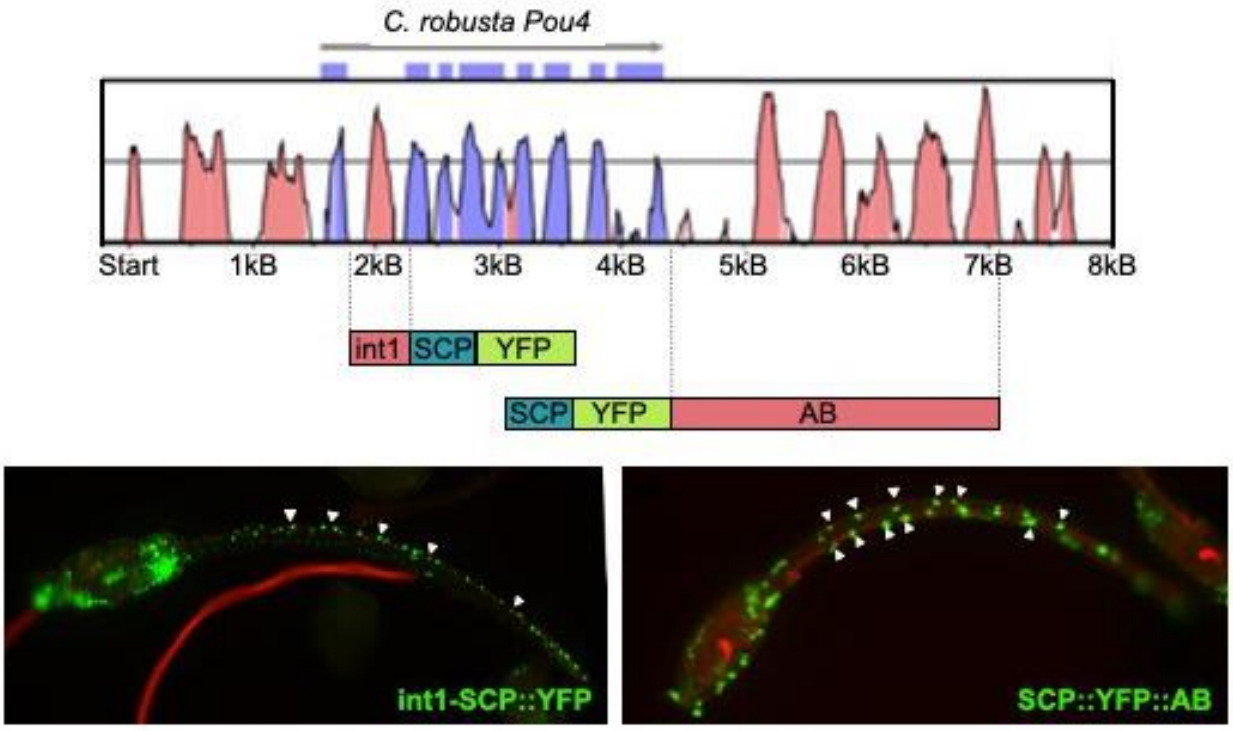


Figure 2.4: Deconstructing *CrPou4* regulatory DNA determines essential elements for *CrPou4* embryonic regulation. (A) VISTA plot comparing the *C. robusta Pou4* locus with that of its sister species *C. savignyi*. Pink peaks indicate conserved non-coding DNA and blue peaks indicate exons. Transgenes were constructed using this data to include a variety of *Pou4* CNEs. Diagrammed beneath the Pou4 VISTA plot is "intron1-SCP::YFP" which includes the intronic sequence upstream of a "Super Core Promoter" (to foster RNA pol and TFIID binding [59]) which is upstream of H2YFP, Also shown is SCP::YFP::AB, with has the SCP upstream of H2YFP and places the first 2.0kB of *Pou4* 3` CNE downstream of YFP. Bottom left: 22hpf transgenic larva electroporated with intron1-SCP::YFP results in strong expression in the dorsal midline (and some non-midline ectopic expression). Bottom right: Electroporating SCP::YFP::AB results in YFP expression in the ESNs with some ectopic expression in unidentified mesenchymal cells. This pattern is reminiscent of *Pou4* expression shown in Figure3A. Visible cilia projecting from YFP+ cells highlighted with white arrowheads.

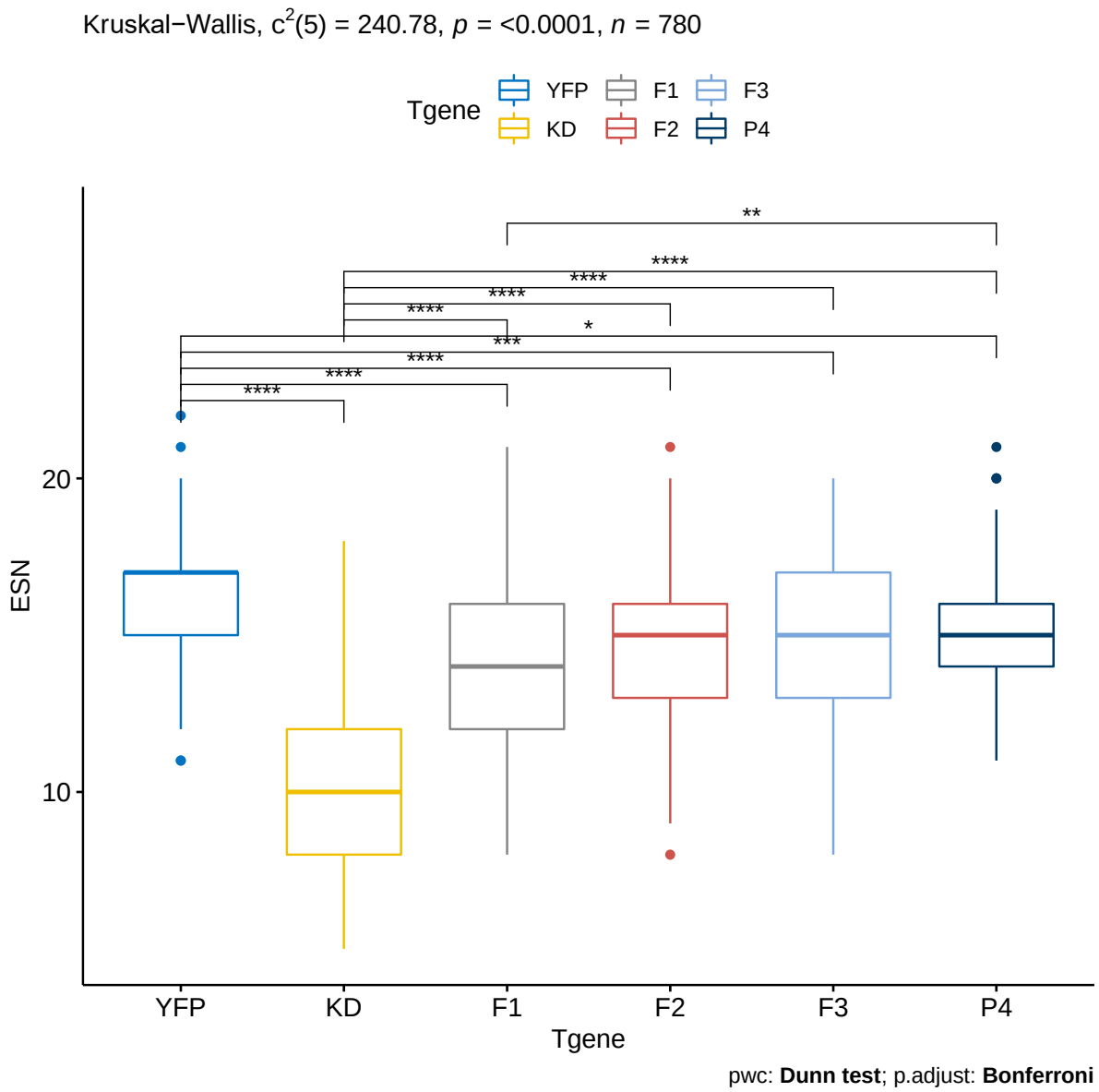
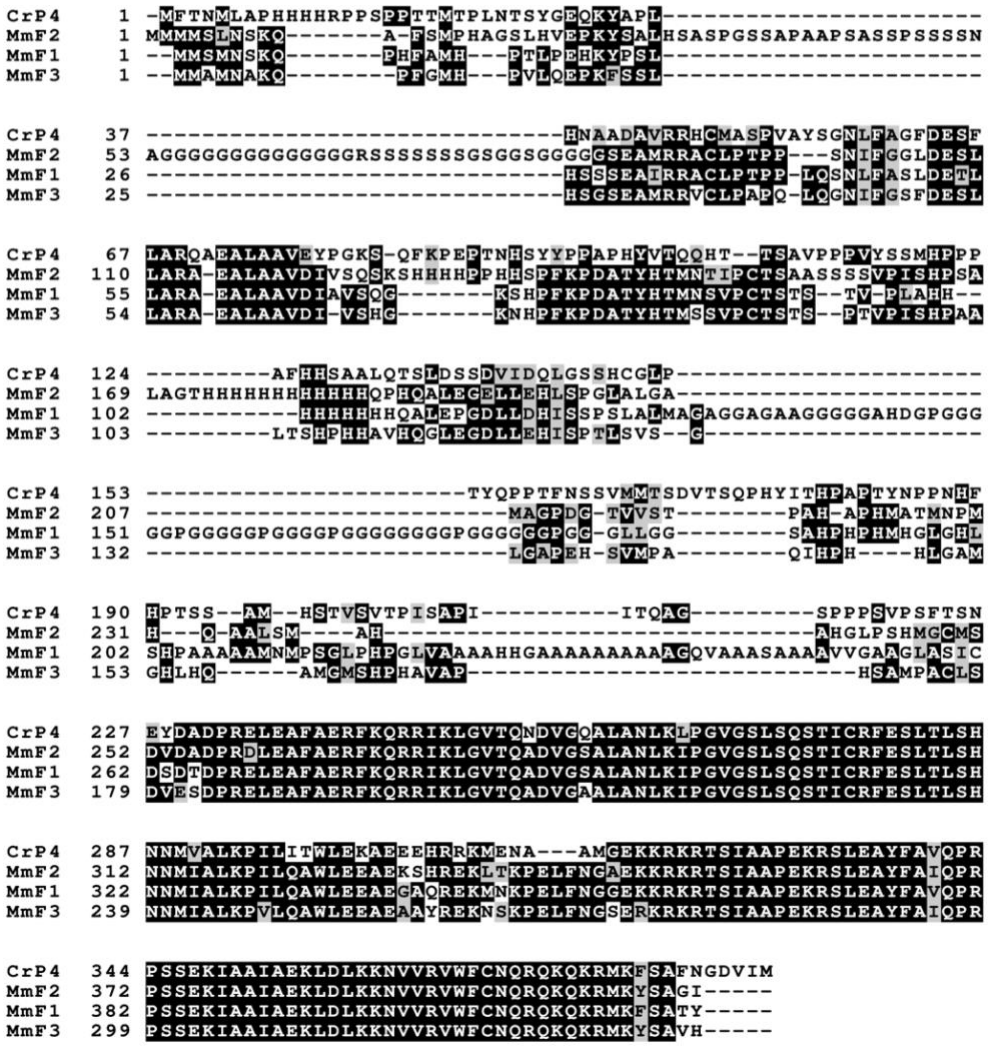


Figure 2.5: Mouse *Pou4* genes can rescue the *CrPou4* knock-down phenotype. The control group, "YFP", included SCP::H2YFP::AB and Epi::Cas9. The *CrPou4*-targeted control group, "KD", included SCP::H2YFP::AB, Epi::Cas9, and two different *CrPou4*-targeting U6::gRNA transgenes. The four experimental groups, "F1", "F2", "F3", and "P4", were electroporated with EpiB::Cas9, the two Pou4-targeting U6::gRNA transgenes, and the corresponding rescue constructs expressing any of the four *Pou4* genes ESN-specifically. In the YFP control group, larva presented an average of 16.2 cilia, whereas in the KD control group cilia number averaged 10.2 per larva. Adding ESN-specific MmPou4F1 resulted in an average of 14.0 cilia per larva, MmPou4F2 resulted in an average of 14.5 cilia per larva, MmPou4F3 resulted in an average of 15.2 cilia per larva, and CrPou4 resulted in an average of 15.5 cilia per larva. "Tgene" refers to the particular experimental transgene tested in conjunction with CRISPR-Cas reagents. All six experimental groups were included per experiment; each experiment was repeated four times (four biological replicates) and a total of n=130 larva were used to assess statistical differences.

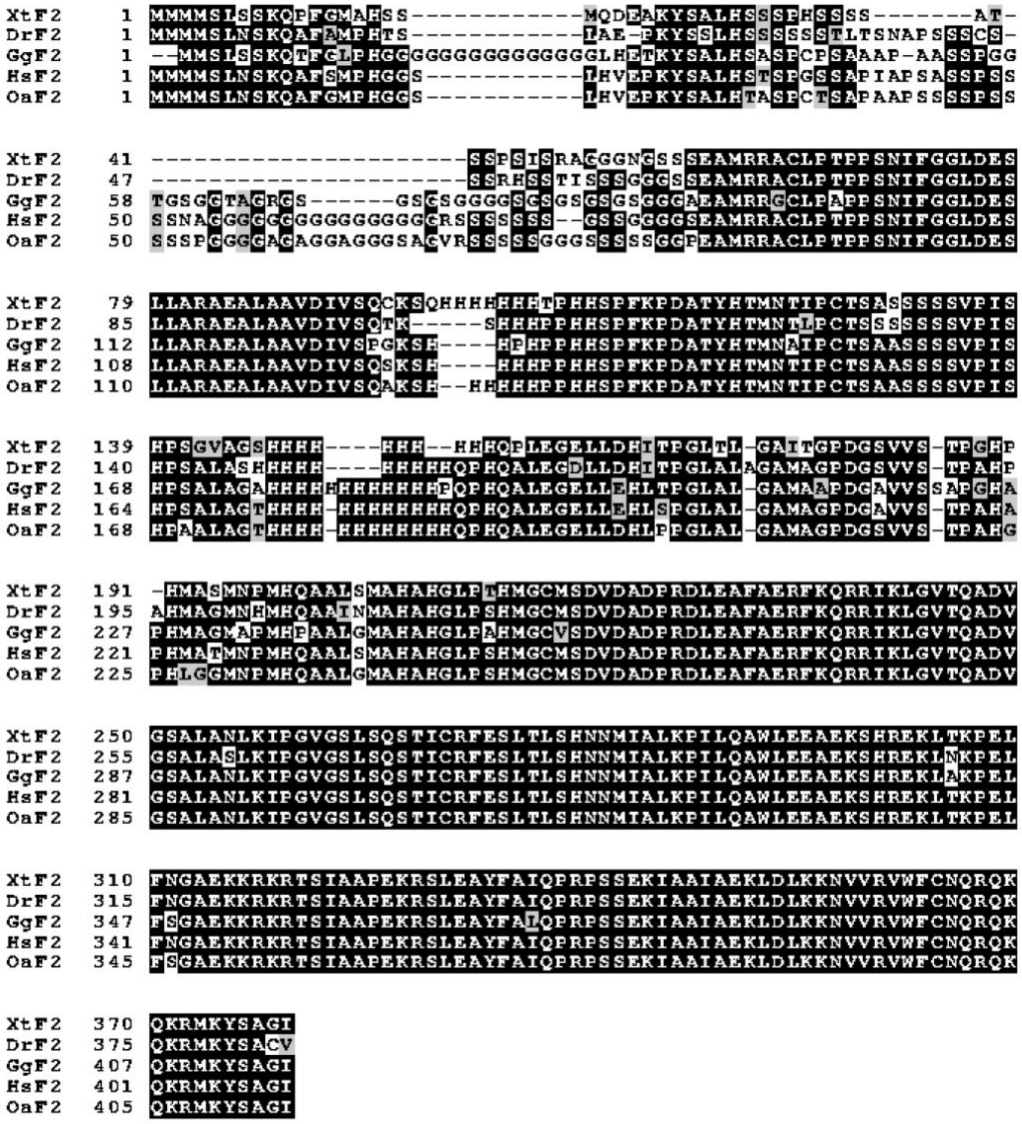
SUPPLEMENTAL INFORMATION

A



Supplementary Figure 2.1: Pou4 protein alignments highlight areas of conservation as well as areas of divergence. A) alignment of Ciona and mouse Pou4 protein sequences used in this study. Blue shading highlights the POU-IV boxes and POU domains. B) Alignment of Xenopus, Zebrafish, Chicken, Human, and Platypus Pou4F2 proteins. Pou4 proteins were aligned with MUSCLE [89] and the resulting alignments were secondarily parsed through BOXSHADE [90]. Green boxes highlight regions of conservation between orthologs that are not conserved between paralogs.

B



Supplementary Figure 2.1: Pou4 protein alignments highlight areas of conservation as well as areas of divergence, continued.

REFERENCES

1. Gold, D.A., R.D. Gates, and D.K. Jacobs, *The early expansion and evolutionary dynamics of POU class genes.* Mol Biol Evol, 2014. **31**(12): p. 3136-47.

2. Latchman, D.S., *The Brn-3a transcription factor.* Int J Biochem Cell Biol, 1998. **30**(11): p. 1153-7.

3. Leyva-Diaz, E., et al., *Brn3/POU-IV-type POU homeobox genes-Paradigmatic regulators of neuronal identity across phylogeny.* Wiley Interdiscip Rev Dev Biol, 2020: p. e374.

4. Xiang, M., et al., *Role of the Brn-3 family of POU-domain genes in the development of the auditory/vestibular, somatosensory, and visual systems.* Cold Spring Harb Symp Quant Biol, 1997. **62**: p. 325-36.

5. Latchman, D.S., *POU family transcription factors in the nervous system.* J Cell Physiol, 1999. **179**(2): p. 126-33.

6. Leyva-Díaz, E., et al., *Brn3/POU-IV-type POU homeobox genes—Paradigmatic regulators of neuronal identity across phylogeny.* Wiley Interdisciplinary Reviews: Developmental Biology, 2020: p. e374.

7. Océane Tournière, D.D., Gemma Sian Richards, Kartik Sunagar, Yaara Y Columbus-Shenkar, Yehu Moran, Fabian Rentzsch, *NvPOU4/Brain3 functions as a terminal selector gene in the nervous system of the cnidarian Nematostella vectensis.* biorxiv, 2020.

8. Maskell, L.J., et al., *Essential but partially redundant roles for POU4F1/Brn-3a and POU4F2/Brn-3b transcription factors in the developing heart.* Cell Death Dis, 2017. **8**(6): p. e2861.

9. Hroudova, M., et al., *Diversity, phylogeny and expression patterns of Pou and Six homeodomain transcription factors in hydrozoan jellyfish Craspedacusta sowerbyi.* PLoS One, 2012. **7**(4): p. e36420.

10. Candiani, S., et al., *Ci-POU-IV expression identifies PNS neurons in embryos and larvae of the ascidian Ciona intestinalis.* Dev Genes Evol, 2005. **215**(1): p. 41-5.

11. Candiani, S., et al., *Expression of AmphiPOU-IV in the developing neural tube and epidermal sensory neural precursors in amphioxus supports a conserved role of class IV POU genes in the sensory cells development.* Dev Genes Evol, 2006. **216**(10): p. 623-33.

12. Stolfi, A., et al., *Migratory neuronal progenitors arise from the neural plate borders in tunicates.* Nature, 2015. **527**(7578): p. 371-4.

13. Holland, L.Z. and D. Ocampo Daza, *A new look at an old question: when did the second whole genome duplication occur in vertebrate evolution?* Genome Biol, 2018. **19**(1): p. 209.

14. Simakov, O., et al., *Deeply conserved synteny resolves early events in vertebrate evolution.* Nature Ecology & Evolution, 2020: p. 1-11.

15. Li, W.H., J. Yang, and X. Gu, *Expression divergence between duplicate genes.* Trends Genet, 2005. **21**(11): p. 602-7.

16. Badea, T.C., et al., *Distinct roles of transcription factors brn3a and brn3b in controlling the development, morphology, and function of retinal ganglion cells.* Neuron, 2009. **61**(6): p. 852-864.

17. Badea, T.C. and J. Nathans, *Morphologies of mouse retinal ganglion cells expressing transcription factors Brn3a, Brn3b, and Brn3c: analysis of wild type and mutant cells using genetically-directed sparse labeling.* Vision research, 2011. **51**(2): p. 269-279.

18. Ninkina, N.N., et al., *A novel Brn3-like POU transcription factor expressed in subsets of rat sensory and spinal cord neurons.* Nucleic Acids Res, 1993. **21**(14): p. 3175-82.

19. Theil, T., et al., *Mouse Brn-3 family of POU transcription factors: a new aminoterminal domain is crucial for the oncogenic activity of Brn-3a.* Nucleic Acids Res, 1993. **21**(25): p. 5921-9.

20. Ryan, A.K. and M.G. Rosenfeld, *POU domain family values: flexibility, partnerships, and developmental codes.* Genes Dev, 1997. **11**(10): p. 1207-25.

21. Pan, L., et al., *Functional equivalence of Brn3 POU-domain transcription factors in mouse retinal neurogenesis.* Development, 2005. **132**(4): p. 703-712.

22. Shi, M., et al., *Genetic interactions between Brn3 transcription factors in retinal ganglion cell type specification.* PLoS One, 2013. **8**(10): p. e76347.

23. Wang, S.W., et al., *Brn3b/Brn3c double knockout mice reveal an unsuspected role for Brn3c in retinal ganglion cell axon outgrowth.* Development, 2002. **129**(2): p. 467-477.

24. Budhram-Mahadeo, V., et al., *The closely related POU family transcription factors Brn-3a and Brn-3b are expressed in distinct cell types in the testis.* Int J Biochem Cell Biol, 2001. **33**(10): p. 1027-39.

25. Huang, E.J., et al., *Brn3a is a transcriptional regulator of soma size, target field innervation and axon pathfinding of inner ear sensory neurons.* Development, 2001. **128**(13): p. 2421-32.

26. Badea, T.C., et al., *Distinct roles of transcription factors brn3a and brn3b in controlling the development, morphology, and function of retinal ganglion cells.* Neuron, 2009. **61**(6): p. 852-64.

27. Goutman, J.D., A.B. Elgoyhen, and M.E. Gómez-Casati, *Cochlear hair cells: the sound-sensing machines.* FEBS letters, 2015. **589**(22): p. 3354-3361.

28. Carricondo, F. and B. Romero-Gómez, *The cochlear spiral ganglion neurons: the auditory portion of the VIII nerve.* The Anatomical Record, 2019. **302**(3): p. 463-471.

29. Goodrich, L.V., *Early development of the spiral ganglion*, in *The primary auditory neurons of the mammalian cochlea*. 2016, Springer. p. 11-48.

30. Xiang, M., et al., *Essential role of POU-domain factor Brn-3c in auditory and vestibular hair cell development.* Proc Natl Acad Sci U S A, 1997. **94**(17): p. 9445-50.

31. Pauw, R.J., et al., *Audiometric characteristics of a Dutch family linked to DFNA15 with a novel mutation (p. L289F) in POU4F3.* Archives of Otolaryngology–Head & Neck Surgery, 2008. **134**(3): p. 294-300.

32. Brunetti, R., et al., *Morphological evidence that the molecularly determined Ciona intestinalis type A and type B are different species: Ciona robusta and Ciona intestinalis.* Journal of Zoological Systematics and Evolutionary Research, 2015. **53**(3): p. 186-193.

33. Imai, J.H. and I.A. Meinertzhagen, *Neurons of the ascidian larval nervous system in Ciona intestinalis: II. Peripheral nervous system.* Journal of Comparative Neurology, 2007. **501**(3): p. 335-352.

34. Pasini, A., et al., *Formation of the ascidian epidermal sensory neurons: insights into the origin of the chordate peripheral nervous system.* PLoS biology, 2006. **4**(7).

35. Takamura, K., *Nervous network in larvae of the ascidian Ciona intestinalis.* Development genes and evolution, 1998. **208**(1): p. 1-8.

36. Ryan, K., Z. Lu, and I.A. Meinertzhagen, *The peripheral nervous system of the ascidian tadpole larva: Types of neurons and their synaptic networks.* Journal of Comparative Neurology, 2018. **526**(4): p. 583-608.

37. Pickett, C.J. and R.W. Zeller, *Efficient genome editing using CRISPR-Cas-mediated homology directed repair in the ascidian Ciona robusta.* Genesis, 2018. **56**(11-12): p. e23260.

38. Chen, J.S., M. San Pedro, and R.W. Zeller, *miR-124 function during Ciona intestinalis neuronal development includes extensive interaction with the Notch signaling pathway.* Development, 2011. **138**(22): p. 4943-4953.

39. Joyce Tang, W., J.S. Chen, and R.W. Zeller, *Transcriptional regulation of the peripheral nervous system in Ciona intestinalis.* Dev Biol, 2013. **378**(2): p. 183-93.

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

41. Serrano-Saiz, E., et al., *BRN3-type POU Homeobox Genes Maintain the Identity of Mature Postmitotic Neurons in Nematodes and Mice*. Curr Biol, 2018. **28**(17): p. 2813-2823 e2.

42. Tang, W.J., J.S. Chen, and R.W. Zeller, *Transcriptional regulation of the peripheral nervous system in Ciona intestinalis.* Developmental biology, 2013. **378**(2): p. 183-193.

43. Waki, K., K. Imai, and Y. Satou, *Genetic pathways for differentiation of the peripheral nervous system in ascidians. Nat Commun 6: 8719*. 2015.

44. Roure, A. and S. Darras, *Msxb is a core component of the genetic circuitry specifying the dorsal and ventral neurogenic midlines in the ascidian embryo.* Developmental biology, 2016. **409**(1): p. 277-287.

45. Katsuyama, Y. and H. Saiga, *Retinoic acid affects patterning along the anterior–posterior axis of the ascidian embryo.* Development, growth & differentiation, 1998. **40**(4): p. 413-422.

46. Roure, A., P. Lemaire, and S. Darras, *An otx/nodal regulatory signature for posterior neural development in ascidians.* PLoS genetics, 2014. **10**(8).

47. Chen, J.S., et al., *An expanded Notch-Delta model exhibiting long-range patterning and incorporating MicroRNA regulation*. PLoS computational biology, 2014. **10**(6).

48. Chen, J.S., *Role of the microRNA miR-124 in the regulatory network governingPNS development in Ciona intestinalis*. 2013, The Claremont Graduate University.

49. Crowther, R.J. and J. Whittaker, *Serial repetition of cilia pairs along the tail surface of an ascidian larva.* Journal of Experimental Zoology, 1994. **268**(1): p. 9-16.

50. Thompson, H., et al., *The formation and positioning of cilia in Ciona intestinalis embryos in relation to the generation and evolution of chordate left–right asymmetry.* Developmental biology, 2012. **364**(2): p. 214-223.

51. Xiang, M., et al., *Requirement for Brn-3c in maturation and survival, but not in fate determination of inner ear hair cells.* Development, 1998. **125**(20): p. 3935-3946.

52. Liu, Z., et al., *In vivo generation of immature inner hair cells in neonatal mouse cochleae by ectopic Atoh1 expression.* PloS one, 2014. **9**(2).

53. Walters, B.J., et al., *In vivo interplay between p27Kip1, GATA3, ATOH1, and POU4F3 converts non-sensory cells to hair cells in adult mice*. Cell reports, 2017. **19**(2): p. 307-320.

54. Costa, A., et al., *Atoh1 in sensory hair cell development: constraints and cofactors.* Semin Cell Dev Biol, 2017. **65**: p. 60-68.

55. Costa, A., et al., *Generation of sensory hair cells by genetic programming with a combination of transcription factors.* Development, 2015. **142**(11): p. 1948-1959.

56. Baumeister, R., Y. Liu, and G. Ruvkun, *Lineage-specific regulators couple cell lineage asymmetry to the transcription of the Caenorhabditis elegans POU gene unc-86 during neurogenesis.* Genes Dev, 1996. **10**(11): p. 1395-410.

57. Takahashi, H., et al., *Brachyury downstream notochord differentiation in the ascidian embryo.* Genes & Development, 1999. **13**(12): p. 1519-1523.

58. Satou, Y., et al., *Gene expression profiles in Ciona intestinalis tailbud embryos.* Development, 2001. **128**(15): p. 2893-2904.

59. Juven-Gershon, T., S. Cheng, and J.T. Kadonaga, *Rational design of a super core promoter that enhances gene expression.* Nature Methods, 2006. **3**(11): p. 917-922.

60. Zeller, R.W., M.J. Virata, and A.C. Cone, *Predictable mosaic transgene expression in ascidian embryos produced with a simple electroporation device.* Developmental dynamics: an official publication of the American Association of Anatomists, 2006. **235**(7): p. 1921-1932.

61. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.

62. Andrews, S. and A. FastQC, *A quality control tool for high throughput sequence data. 2010*. 2015.

63. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nature methods, 2017. **14**(4): p. 417-419.

64. Brozovic, M., et al., *ANISEED 2015: a digital framework for the comparative developmental biology of ascidians.* Nucleic acids research, 2016. **44**(D1): p. D808-D818.

65. Satou, Y., et al., *An integrated database of the ascidian, Ciona intestinalis: towards functional genomics.* Zoolog Sci, 2005. **22**(8): p. 837-43.

66. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nat Methods, 2017. **14**(4): p. 417-419.

67. Love, M., S. Anders, and W. Huber, *Differential analysis of count data–the DESeq2 package.* Genome Biol, 2014. **15**(550): p. 10.1186.

68. Huerta-Cepas, J., et al., *Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper.* Mol Biol Evol, 2017. **34**(8): p. 2115-2122.

69. Liao, Y., et al., *WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs.* Nucleic Acids Res, 2019. **47**(W1): p. W199-W205.

70. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.

71. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND.* Nature methods, 2015. **12**(1): p. 59-60.

72. Shapiro, S.S. and M.B. Wilk, *An analysis of variance test for normality (complete samples).* Biometrika, 1965. **52**(3/4): p. 591-611.

73. Theodorsson-Norheim, E., *Kruskal-Wallis test: BASIC computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples.* Computer methods and programs in biomedicine, 1986. **23**(1): p. 57-62.

74. Ellis, A.R., et al., *Nonparametric inference for multivariate data: the R package npmv.* Journal of Statistical Software, 2017. **76**(4): p. 1-18.

75. Mao, C.A., et al., *Substituting mouse transcription factor Pou4f2 with a sea urchin orthologue restores retinal ganglion cell development.* Proc Biol Sci, 2016. **283**(1826): p. 20152978.

76. Ohtsuka, Y., T. Obinata, and Y. Okamura, *Induction of ascidian peripheral neuron by vegetal blastomeres.* Developmental biology, 2001. **239**(1): p. 107-117.

77. Wagner, E., et al., *Islet is a key determinant of ascidian palp morphogenesis.* Development, 2014. **141**(15): p. 3084-3092.

78. Imai, K.S., et al., *Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system.* Development, 2009. **136**(2): p. 285-93.

79. Cao, C., et al., *Comprehensive single-cell transcriptome lineages of a proto-vertebrate.* Nature, 2019. **571**(7765): p. 349-354.

80. Farooqui-Kabir, S.R., et al., *Cardiac expression of Brn-3a and Brn-3b POU transcription factors and regulation of Hsp27 gene expression.* Cell Stress Chaperones, 2008. **13**(3): p. 297-312.

81. Veenstra, G.J., P.C. van der Vliet, and O.H. Destree, *POU domain transcription factors in embryonic development.* Mol Biol Rep, 1997. **24**(3): p. 139-55.

82. Frazer, K.A., et al., *VISTA: computational tools for comparative genomics.* Nucleic acids research, 2004. **32**(suppl_2): p. W273-W279.

83. Satou, Y., et al., *Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations.* Genome biology, 2008. **9**(10): p. R152.

84. Satou, Y., et al., *An integrated database of the ascidian, Ciona intestinalis: towards functional genomics.* Zoological science, 2005. **22**(8): p. 837-843.

85. Dehal, P. and J.L. Boore, *Two rounds of whole genome duplication in the ancestral vertebrate.* PLoS Biol, 2005. **3**(10): p. e314.

86. Erwin, D.H. and E.H. Davidson, *The last common bilaterian ancestor.* Development, 2002. **129**(13): p. 3021-32.

87. Vavouri, T. and B. Lehner, *Conserved noncoding elements and the evolution of animal body plans.* Bioessays, 2009. **31**(7): p. 727-735.

88. Swalla, B., *Building divergent body plans with similar genetic pathways.* Heredity, 2006. **97**(3): p. 235-243.

89. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.

90. Hofmann, K. and M. Baron, *Boxshade 3.21.* Pretty printing and shading of multiple-alignment files Kay Hofmann ISREC Bioinformatics Group, Lausanne, Switzerland, 1996.

Chapter 3

Regulatory Interactions Governing the Specification of Epidermal Sensory Neurons in the Ascidian *Ciona robusta*

INTRODUCTION

Ciliated mechanosensory neurons are an ancient cell type, some form of which likely existed in the common metazoan ancestor [1]. A variety exist today, and they display a wide diversity in the arrangement of their cilia [2]. These cells receive input from their surroundings and transmit that signal, in many cases, to a processing center in the central nervous system (CNS); thus mechanoreceptive neurons are components of the peripheral nervous system (PNS) [3]. In some organisms, there has been an uncoupling of this cell into a ciliated mechanoreceptor synapsed with a nerve cell, e.g. in the arrangement of the mammalian inner ear, hair cells transmit impulses to the cochlear nerve [4]. Within the diversity of mechanosensory neurons found in the animal kingdom, common regulatory network components during development include homeobox genes, POU-domain genes, proneural bHLH genes, and Notch signaling that has been conserved from cnidarians [5, 6] to the hair cells of the mammalian ear [7, 8]; some of these genes are potentially expressed within the mechanoreceptive choanocytes of sponges as well [9, 10]. During the specification and differentiation of mechanoreceptors, a conserved set of regulatory genes are associated with these processes and include basic helix-loop-helix (bHLH) *Achaete-scute* genes, *Nkx* homeobox genes, Notch signaling, and POU domain genes [2, 11, 12]. In mammals, ascidians, and cnidarians, members of the class IV POU domain genes act as terminal selector genes- transcription factors (TFs) that are necessary and sufficient to drive the differentiation of cell type [5, 13, 14].

Tunicates (which include the ascidians) occupy an important position on the tree of life; among extant invertebrate chordates they are the closest related group to the vertebrates [15]. Ascidians are an experimentally tractable animals that have small genomes, undergo rapid development to produce swimming tadpole-like larvae, and have a well-defined cell lineage

(refs). Gene expression studies and gene manipulation experiments benefit from the ability to generate thousands of transgenic embryos using a simple electroporation system [16]. The ascidian larva has simplified and well defined central and peripheral nervous systems [17, 18], the latter consisting primarily of ciliated epidermal sensory neurons (ESNs) found in the palps, along the dorsal trunk, and 15-20 pairs projecting into the cellulose-based tunic of the tail fin along the dorsal and ventral midlines [18-20]. ESNs share structural similarities [21], genetic regulatory network components [20], and express extensive numbers of homologous genes in common with vertebrate hair cells and cochlea of the inner ear [22]. Caudal ESNs (CESNs) are presumed to be mechanoreceptors involved in the coordination of larval swimming behavior [23, 24].

The earliest characterized events leading to the formation of CESNs begin at the fifth cell division [25]. At the 32-cell stage, zygotic FGF signaling activates the TGFb ligand Nodal, and is one of the earliest zygotic events that establishes the field of cells from which the future dorsal CESNs are derived [26]. Zygotic BMP signaling establishes the field of cells from which the ventral CESNs arise [27]. At the 64-cell stage, the TF *Msxb* is expressed in a bilateral pair of cells which will divide to produce the dorsal midline of the early tailbud embryo [28]. *Msxb* is later expressed at late gastrula in the cells of the presumptive ventral midline [27]. *Msxb* has been shown to activate the expression of additional midline TFs such as *Tox* (*CAGF9*), *Nkx-C*, *Dll-C*, and *Achaete-scute homolog* (*Ash*; *Ash a-like 2*) [29]. These TFs are expressed both dorsally and ventrally and are thought to operate within a common regulatory network for CESN development [20, 29]. In the lateral- and mediolateral tail cells, putative transcriptional repressors such as *Hesb* and *SoxB2* [20, 27, 28, 30] are expressed, but it is not yet known if they play a role in CESNs specification. Notch-Delta signaling has been shown to mediate the number

and spacing of CESNs [20, 31-33]. As development proceeds and the tail is extended out from the trunk by a combination of cell divisions and cell rearrangements [34], the neurogenic homeobox TF *Pou4* is expressed in a punctate manner; rarely are there adjacent *Pou4*+ cells during the tailbud stages [35]. As development proceeds, the inter-ESN spacing increases to an average of ~five midline epidermal cells due to both intercalation of the midline cells from the right and left sides of the epidermis and midline cell divisions [31]. After patterning by Notch-signaling, a gene regulatory network (GRN) has been proposed for CESN development [30]. Following the regulatory network that establishes midline territory [20, 26], the model describes a cascade of proneural and neurogenic TFs which also activate the expression of the microRNA *miR-124*. A separate publication suggested that *miR-124* can repress Notch pathway genes in the CESNs, binding to the 3` UTR of five Notch pathway genes such as *Notch* and *Hesb* [30, 32]. Previous studies showed *MyT1* and *Pou4* are the first TFs expressed in CESNs post-specification and that their activation triggers a cascade of other transcription factors, in turn working to differentiate the precursor into a mature ciliated epidermal sensory neuron [30]. In that CESN-specific GRN, *MyT1* is at the top of the cascade, activating the downstream gene *Pou4*. However, in other organisms evidence suggests that *MyT1* negatively regulates Notch genes during neuronal development [36-38]; here we examine that regulatory relationship in more detail.

There is a significant gap in our knowledge of CESN development, for it is currently unknown how cells of established midline identity are selected to become either CESNs or remain non-ESN epidermal midline. To gain insight into this problem, we first perform an ISH-based screen for putative CESN-regulator candidate genes that may operate in the neurogenic territory of the dorsal and ventral caudal midlines. Then we perform electroporation-based

functional studies and examine how these genes affect cilia formation and the expression of established CESN regulators. Finally, we propose an updated GRN that provides a bridge between previously established midline formation networks [29] and the post-Notch signaling network of CESN differentiation [30]. This GRN is then placed into the context of wave-like expression dynamics and Notch signaling in a proposal for how initial inter-CESN spacing is generated and CESNs are specified.

MATERIALS AND METHODS

Gravid *C. robusta* were collected from Mission Bay, San Diego, and kept under constant illumination in a recirculating refrigerated tank to force the accumulation of gametes. Transgene DNAs used for electroporations were pooled together, ethanol precipitated and pellets were dissolved in 0.77M Mannitol. Electroporations were carried out essentially as in Zeller et al, 2006 [16] , and electroporated embryos were left to develop at ~18°C in 0.22µm-filtered seawater. If embryos were required to develop past 12-15 hours of development, sea water was supplemented with penicillin/streptomycin (10 U/mL and 10µg/mL, respectively) and embryos were paralyzed with MS-222 (Sigma) to prohibit tail movement. A biological replicate refers to an independent electroporation on a different day using different parent animals. All imaging was performed on a Zeiss AxioPlan 2E Imaging epifluorescence microscope. Fluorescent images were captured with a Hamamatsu ORCA 4.0 monochromatic camera and brightfield images were captured with an AxioCam ICc1 color camera.

Immunohistochemistry was performed as in Joyce Tang, Chen [30]. Briefly, embryos were fixed in 2% paraformaldehyde, washed in Phosphate-Buffered Saline plus 0.2% Tween 20

(PBST) and briefly permeabilized in ice-cold 100% methanol. Embryos were incubated at 4°C overnight with primary antibody (Rabbit-anti-GFP, Thermo Fisher Scientific, A11122; Mouse anti-acetylated tubulin antibody, Sigma, T7451) at 1:2000, washed thoroughly and incubated at 4°C overnight with secondary antibody (Alexa Fluor 546, A11003; Alexa Fluor 488, A-11001: Thermo Fisher Scientific,) at 1:500. Embryos were washed thoroughly in PBST, incubated for 10min in 50% glycerol/PBST +300nM DAPI, and then transferred to 70% glycerol/PBST for imaging.

Whole mount *in situ* hybridization was performed essentially as in [39] without modifications. After colorimetric development, embryos were fixed briefly in 4% paraformaldehyde, washed 2X in PBST, transferred into 100% EtOH for approximately 5 minutes, washed once in PBST and transferred to 70% glycerol for imaging.

Molecular cloning

All *C. robusta* gene fragments were amplified from cDNA generated using the ProtoScript First Strand cDNA Synthesis Kit (New England BioLabs) with RNA extracted from pooled embryos spanning the 64-cell to swimming larva stages. Primers used to generate gene products listed in Table S1. PCR products were produced with Q5 DNA polymerase (New England Biolabs) using standard PCR conditions. All transgenes used in this study were cloned into our custom vector, pZapANX, described in [16]. All CRISPR-Cas reagents used were previously described in [14]; see Table S1 for gRNA target sequences used. Overexpression constructs were generated by cloning cDNA sequences of each gene analyzed into our EpiB::H2-YFP transgene [32]. This construct uses an epidermal-specific promoter to drive histone-

localized (H2) YFP throughout the epidermis. We cloned epidermally-expressing TF constructs by restriction digest-deletion of H2 then replacing it with TF cDNAs of interest. These transgenes express TF-YFP fusion proteins throughout the epidermis. All TF-WRPW constructs were produced by PCR-amplifying the sequence encoding the C-terminal domain of *CrHairy-b*, deleting YFP and replacing it with the WRPW fragment. All transgenes were sequenced prior to use.

RESULTS

Candidate genes involved in the developmental regulation of CESNs

Candidate genes were identified through three main approaches: 1) literature analysis, 2) *in situ* hybridization patterns and 3) TFs predicted to bind to cis-regulatory regions of the *CrPou4* gene. We identified putative CESN genes by conducting a keyword search on the expression pattern database found at the *Ciona* Kyoto genome browser [40], This search yielded genes such as *Prx*, *Smad6/7*, *SoxC*, and *Mindbomb*. In a separate line of inquiry, our lab generated a minimal transgene utilizing approximately 1.0kB of *cis*-regulatory DNA from the *CrPou4* gene that directs expression in the CESNs (manuscript in preparation). This DNA sequence was used as input into binding site prediction programs such as CIS-BP database [41] and LASAGNA-search [42], and we additionally performed a HOMER analysis [43]. TFs predicted to bind sites present in the *Pou4 cis*-regulatory DNA were subjected to a BLAST search [44] to identify *Ciona* homologs. Genes that were not expressed at the correct time, using EST expression data at the Kyoto browser, were excluded. Lastly, sensory neuron differentiation is conserved [1, 21, 45] thus an examination of the literature also procured a series of candidates for further analysis (Table 1). Many genes from the aforementioned sources overlapped, e.g. we

found homologs such as *Hmx3* and *Barx-C* that were predicted to bind to sequences within the *Pou4* cis-reg DNA while also known to play a role in neuronal development [46-48].

Candidate CESN developmental regulators produce diverse ISH patterns of expression

We generated antisense RNA probes for twenty three genes listed in table one and performed ISH against each. We utilized a pool of embryos from multiple timepoints that spanned from early neurula to late tailbud stage. This span of stages allowed us to examine genes that may play a role early in midline formation up until final epidermal cell divisions have occurred [20]. Below, we mention briefly some roles of homologous genes in other organisms related to epithelial and proneural function, then describe the expression pattern of each *Ciona* gene.

Figure 1 displays expression patterns for 13/29 genes. Eight additional genes did not produce interpretable ISH patterns during the timepoints we assessed and six were unable to be successfully cloned. Additionally, we found that *EvxB* was expressed post-ESN specification (confirming [49]). *EMC2* expression is shown as an electroporation control embryo (see Fig. 4Fii), and is expressed throughout most of the embryonic midline with the presumable exception of ESNs. *Lhx1*, a gene involved in the differentiation of many CNS cell types [50, 51] as well as early axis formation [52, 53], was expressed in *Ciona* tailbud embryos in the anterior vesicle as well as the developing motor ganglia (Fig. 1A, [25]). *AP-2* TFs in mammals are expressed during differentiation of many tissues such as trigeminal nerves [54], the optic up and stalk [55] and neural crest/neural tube [54]. In *Ciona*, *Ap-2* (*Tfap2-r.b*) plays an important role in ectodermal specification and downstream neural differentiation events [56]. Our analysis of *Tfap2-r.b* during tailbud stages showed the gene is expressed strongly throughout the anterior and posterior dorsal

and ventral midlines with the exception of the posterior trunk midline cells (Fig. 1B, [25]). One of the four *C/EBP* TFs in mammals, *C/EBP-delta*, is a transcriptional activator that plays a role in inner ear development [57]. We detected expression of the single *Ciona C/EBP* (*C/EBP-beta/gamma/delta/epsilon*) in all embryonic tissues, most strongly in the developing sensory vesicle and the dorsal and ventral midlines (Fig. 1D). *Emx* is expressed in a variety of developing neural tissues in mammals [58], while similar activity is performed in *Drosophila* by the homolog *empty spiracles* [59]. We did not detect *Emx* in caudal neurogenic territories, only the lateral cells, however, we did detect expression in the region of the developing palps (Fig. 1E, [25]). *Sprouty* is an FGF antagonist in *Drosophila* important for eye morphogenesis and photoreceptor development [60]. In mammals it performs similar roles [61] and is also a deafness gene expressed in the developing inner ear [62]. In *Ciona*, we found that *Sprouty* is expressed in the tail tip region and the trunk epidermis of the early tailbud embryo (Fig. 1F). In mammals, a *ZF-248* homolog is expressed in numerous epithelial tissues [63]. During early tailbud stage in *Ciona*, *ZF-248* (previously named *Orphan Fox-4*) is expressed in the caudal midlines and the b-lineage endoderm (Fig. 1G; young adult expression in the endostyle [64]). *Bicaudal* is highly expressed and required for spinal cord development in mammals [65], while in *Drosophila* acts primarily during oocyte development [66]. We detected *Bicaudal-C* during early tailbud stage in the developing nerve cord and trunk lateral mesenchyme (TLM) (Fig. 1Hi, Hii). *FoxC* in zebrafish and mice is expressed around the region of the developing eye [67, 68]. Interestingly, the expression pattern we observed early was anterior trunk which resolved into the region of the developing ocellus and otolith (Fig. 1Ii, Iii, and [25]). *Hmx3* (*NK5*) is widely known to be involved in neuronal differentiation and specification of the inner ear [48, 69]. During the *Ciona* early tailbud stage it is expressed in BTN precursors (Fig. 1 Ji and Jii). We also

detected expression in the anterior developing brain and developing motor ganglion. *Lhx3* is required for differentiation of various neuronal cells [70] and inner ear development [71]. The *Ciona* expression pattern of *Lhx3* mirrored that of *Hmx3* (anterior brain, motor ganglion, and BTNs; Fig. 1 Ki, 1Kii, and [25]); *Lhx3* is also expressed similarly to *Hmx3* in mice. *Dlx-1* (*Dlx.b* in *Ciona*) is required predominantly during forebrain development in mammals [72], while its *Drosophila* homolog *Distal-less* patterns appendages and various sensory organs [73]. At mid-neurula stage we detected *Dlx.b* in the anterior caudal lateral cells, as well as the anterior trunk (Fig. 1 Li). This pattern weakened caudally, while anteriorly it formed a ring of expression in the trunk that excluded developing palps (Fig. 1 Lii, confirmed form [25]). *SoxC* genes in mammals (*Sox4*, *Sox11*, *Sox12*) are required for normal neural tube development. We confirmed [25] that *Sox4/11/12* was detected in the developing brain vesicle and region of the palps during neurula stage (Fig. 1Mii) which expanded into the CESNs during tailbud stage (Fig. 1Mii).

We also confirmed ISH patterns of several genes previously implicated in neuronal development in *Ciona*, and reanalyzed them at particularly relevant timepoints during ESN specification such as mid-late neurula stages. *Pax3/7* homologs are required for normal dorsal neural tube development in vertebrates [74]; in *Ciona Pax3/7* is expressed in the CNS and developing motor ganglion ([25, 75], and Fig. 1Oi, 1Oii), but we detected no evidence of expression in the territory of developing ESNs. *Neurogenin* (*Ngn*)is expressed in the developing brain, sensory vesicle, motor ganglia, nerve cord and a pair of posterior cells which are likely the BTN precursors (Fig. 1Pi, 1Pii; [76-78]). *Pou4* genes are predominantly neuronal TFs expressed in the CNS and PNS of vertebrates [79, 80] and various invertebrates [5, 81, 82]. Mentioned above, in *Ciona Pou4* is expressed in the BTNs, the palps, and all ESNs (Fig. 1Qi, 1Qii; [30, 35]). The Notch ligand *Delta2* is required for proper patterning of many neural structures

including the inner ear [83]; its encoded protein provokes Notch signaling in adjacent cells which leads to lateral inhibition inhibiting neurogenesis [84]. The Notch ligand *Delta2* during *Ciona* neurulation is expressed in the developing nerve cord (Fig. 1Ri), while expression in CESNs is present during tailbud stages (Fig. 1Rii; [30, 76]). *Orphan-bHLH* (*Orphan*) was identified as a proneural repressor in the embryonic midline, having the capacity to prohibit neural marker genes if ectopically expressed [29]. *Orphan* is expressed in the intercalating notochord, and the embryonic midlines at the start of neurulation (Fig. 1 Si; [29]). During tailbud formation *Orphan* expression is down-regulated in the notochord and lack of expression (in the presumptive CESNs) can be detected interspersed throughout the caudal midlines (Fig. 1 Sii). *MyT1* in vertebrates is expressed in a variety of neural cells [37, 38] and has been shown to promote proneural gene (PNG) expression by blocking repressive Notch target genes [36]. In *Ciona*, *MyT1* is strongly expressed in the developing CNS and perhaps early BTN precursors during neurula stages (Fig. 1 Ti); this territory expands to include the caudal CESNs (Fig. 1 Tii; [30, 85]). During neurulation, *Ash a-like 2* (*Ash*, *Achaete-scute homolog*) is expressed in a posteroanterior wave-like manner in the developing midlines (Fig. 1U, St. 15 - St. 19), and the developing palps beginning at St. 17 ([29, 30], Fig. 1U). *Ash a-like 2* is then is rapidly downregulated beginning at mid-tailbud stage (Fig. 1U, St. 21.5; [20, 29]).

Ectopic epidermal expression of TFs results in altered cilia distribution and number

We generated transgenic embryos expressing various TFs using a previously characterized epidermal driver from the *EpiB* gene (KH.C7.154, [32]) and examined the resulting embryos for alterations in PNS cilia patterning and gene expression. Following electroporation, transgenic embryos were subjected to immunohistochemistry to assay for cilia

number and position or to whole mount in situ hybridization (ISH). The control transgene Epi::H2YFP produces a histone-localized fluorescent protein (FP), and in Fig. 2A caudal cilia projecting from CESNs are detected scattered throughout the dorsal and ventral midlines. Fig. 2B shows the dominant phenotype produced by expressing *C/EBP* throughout the epidermis. Epi::C/EBP larval cilia were disrupted, and for reasons not yet understood, the cellular location of the C/EBP-YFP fusion protein was not primarily the nucleus as TF-FP but rather appeared to be imbedded in the borders of unidentified cellular compartments. Epi::Ngn (which did not contain a FP tag) radically transformed the overall shape of the larva into a twisted ball (Fig. 2C) reminiscent of the Epi::Pou4 phenotype [32]. There did appear to be numerous ectopic cilia, however the majority of those detected appeared not fully developed. *EMC2* homologs have known repressive roles [86, 87], and unsurprisingly expressing *EMC2* throughout the epidermis completely abolished formation of cilia in the tail; trunk epidermal ciliated neurons were still present. Epi::Smad6/7 (which also did not contain an FP tag) resulted in a near total reduction in cilia, but with the consistent exception of four cilia pairs in the anterior dorsal midline. Due to the mosaic nature of transgene expression, some larva did have cilia in other locations within the midlines, however there was never a *Smad6/7* transgenic larva that did not contain those four stereotypical cilia pairs. Our candidate gene search yielded another gene, "*Prx*", a *Prrx* homologue. We were unable to produce an ISH pattern for this gene, however ISH results in the *Ciona* Kyoto Database showed expression in the same manner as *Orphan*, i.e., dorsal/ventral midline expression with the exception of presumptive CESN gaps. Due to the repressive activity of its vertebrate homolog, we predicted *Prx* may act as a repressor in *Ciona*. Epi::Prx-YFP transgenic larva had severely reduced midline cilia (Fig. 2F), hinting at a conserved function. Epi::SoxC transgenic larva displayed supernumerary cilia, but only within the larval midlines

(Fig. 2G). Epi::Tfap2-r.b resulted in uninterpretable phenotypes, whereby malformed cilia, loss of cilia, and disfigurement of the trunk region were present within individual larva (Fig. 2H). Ectopic expression of *Ash a-like 2* produced disfigured tails with cilia that were often either clustered together, absent from, or malformed over relatively large sections of midline territory (Fig. 2I).

After examining cilia phenotypes resulting from ectopic expression of various genes, we examined their capacity to alter expression of known CESN regulators e.g. *Dlk* (*Delta2*), *Pou4*, and *MyT1*. Expressing *Ngn* throughout the epidermis resulted in no change to *EMC2* expression, and an expansion of *Orphan*, *Pou4*, and *Delta2* expression in much of the caudal and trunk epidermis (Fig. 3Ai-3Aiv). Epi::Smad6/7 which resulted in the elimination of all but four cilia pairs in the anterior dorsal midline (Fig. 2E) resulted in consistent elimination of *Pou4* from the dorsal midline while ventral midline patterns were maintained (compare to Fig. 1Rii). When we probed *Smad6/7* embryos for *Delta2* expression, we found it was absent ventrally and dorsally (Fig. 3Bii; the intense *Delta2* staining dorsally is nerve cord). *Prx*, which abolished CESN formation, led to an increase in the expression domain of *Orphan* while almost completely eliminating *Pou4* and *Delta2* expression. *EMC2*, which produced a similar absent cilia phenotype, eliminated *Pou4* expression (Fig. 3D1 vs. Fig. 4Fv) and *MyT1* expression Fig. 3Diii vs. Fig. 4Fiv) but interestingly had no discernable effect on *Delta2* expression (Fig. 3Dii vs. Fig. 4Fvi).

We also examined TFs previously shown to be involved in midline formation and CESN patterning and differentiation, but 1) focused our attention at timepoints more relevant to ESN specification events and 2) assessed their roles using novel combinations of overexpressed TF and genes probed for ISH. For example, *MyT1* had previously been positioned in a post-

specification CESN GRN [30], however here we examined its relationship with several different genes such as *Ash a-like 2* and *Orphan*. Epi::MyT1 transgenic embryos probed for *Neurogenin* demonstrated a slight expansion of that gene at the tail tip at early neurulation (Fig. 4Ai), while exhibiting a reduction in the expression of *EMC2* and *Orphan* (Fig. 4Aii and 4Aiii). *Ash a-like 2* expression was expanded into the mediolateral cells, *Pou4*+ cells increased (as previously shown [30]), and *Delta2* expression expanded into patches within the mediolateral and lateral cells (Fig. 4Ai-4Avi). Additionally, coelectroporating EpiB::Ash with Delta::YFP, resulted in YFP expression throughout the epidermis (Fig. S1). *Msxb* has been extensively studied [26, 29], and positioned as an initiator of midline territory formation. Epi::Msxb embryos resulted in no change to *Ngn* expression or *EMC2* expression (Fig. 4Bi and 4Bii), while as expected *Pou4* and *Delta2* expanded their expression domains to include patches of cells throughout the epidermis (Fig. 4Biii and 4Biv). We also wanted to test if ectopic *Pou4* expression would alter the pattern of a putative repressor such as *Orphan*. *Pou4* transgenic embryos exhibited a loss of epidermal *Orphan* expression (Fig. 4Ci), and EpiB::Pou4 embryos probed for *Delta2* demonstrated expansion throughout the epidermis (Fig. Cii). Epi::Ash a-like 2 transgenic embryos probed for *Pou4* expression displayed a near total absence of caudal expression (Fig. 4Di), consistent with a previous publication [29]. *Ash a-like 2* was shown to be predominantly repressive in that report, so we were surprised to see that Epi::Ash a-like 2 resulted in a dramatic increase in the expression domains of *Delta2* and *MyT1* (Fig. 4Dii and 4Diii). *MyT1* expression expanded into all of the midline cells and some mediolateral cells, whereas *Delta2* expression included all of the caudal epidermis and some cells in the trunk epidermis. In Roure and Darras 2016, ectopic *Orphan* expression led to a reduction of *Ash a-like 2* as well as the nervous system marker gene *ETR*. Epi::Orphan-bHLH in our hands did not affect *Ash a-like 2* expression (Fig. 4Ei; see

discussion), while it completely abolished epidermal *Pou4* and *Delta2* expression (Fig. 4Eii and 4Eiii).

Forced conversion of various TFs into putative repressors reveals roles of various TFs as either repressors or activators

After examining ISH results from ectopically expressing various TFs, we decided to test the effect of adding the WPRW motif (which converts transcriptional activators into repressors [88]) to the C-terminal end of Ash, MyT1, Pou4, and Prx. By comparing the resulting phenotypes between expression of the WRPW-fusion gene and the wild-type gene, one can infer if the endogenous gene normally functions as a repressor. We electroporated EpiB::Ash-WRPW and in response, *Delta2* and *MyT1* expression were both reduced (Fig. 5Ai and 5Aii), while *Pou4* expression expanded into clusters of stained cells (Fig. 5Aiii). Immunostaining of Ash-WRPW transgenic larva revealed an increase in cilia number (compare Fig. 5Av to Fig. 2A). Next, we electroporated EpiB::MyT1-WRPW. Similarly to ISH results from electroporating EpiB::MyT1, *Delta2* expression expanded throughout the midlines (Fig. 5Bi), while *Pou4* expression increased slightly (Fig. 5Bii). Immunostaining of cilia demonstrated an increase within the larval midlines (Fig. 5Biii), resembling results with Epi::MyT1 by Joyce Tang et al (2012) [30]. Ectopically expressing Pou4-WRPW resulted in a dramatic decrease of both *Delta2* (Fig. 5Ci) and *MyT1* (Fig. 5Cii), while immunostaining showed a total loss of CESN cilia (Fig. 5Ciii). Finally, we electroporated Epi::Prx-WRPW which expanded *Delta2*, *MyT1*, and *Pou4* expression throughout the epidermis (Fig. 5Di-5Diii). Embryo morphology was disturbed by the expression of this fusion protein so exact patterns could not be assessed, however it was clear that *Orphan* staining was reduced in regions of the dorsal and ventral midlines (Fig. 5Div).

CRISPR-Cas targeting of CESN regulators demonstrates genetic dependencies and complements previous assays

Overexpressed genes can lead to off target TF binding and can generate artifacts, and the amount of ectopically expressed genes via transgenesis is generally unknown [89, 90]. Therefore, we decided to complement our ectopic expression experiments by knocking down key genes using CRISPR-Cas technology [91, 92]. We tested whether *Delta2* and *Pou4* expression were dependent on the presence of various CESN regulators, so we designed reagents to transgenically knockdown *Pou4*, *Delta2*, *Ash a-like 2* and *MyT1*. Knocking down *Pou4* reduces cilia formation [14], and here decreased caudal midline *Delta2* expression (Fig. 6A); conversely, knocking down *Delta2* led to an increase in *Pou4* expression (Fig. 5C). Knocking down *Ash a-like 2* resulted in a decrease in *Delta2* expression (Fig. 6Bi) and a moderate expansion of *Pou4* expression (Fig. 6Bii). Finally, knocking down *MyT1* led to a severe decrease in *Delta2* and *Pou4* (Fig. 6Di, 6Dii).

DISCUSSION

Fields of neurogenic epidermis, such as in *Drosophila*, often generate neurons through complicated protein-level-dependent Notch-Delta signaling mechanisms [83, 84, 93, 94]. In those scenarios, Delta protein is present at low levels throughout the tissue, and *Delta* expression only becomes punctate after Notch signaling locks in a pattern. However in ascidians, dorsal CESNs emerge from a single row of midline cells that result from the fusion of the left and right halves of the dorsal epidermis. Prior to fusion, the right and left midline precursor cells are bordered by mediolateral cells expressing transcriptional repressors [20]. Upon fusion, these

mediolateral cells continue to express transcriptional repressors. ESN specification within the dorsal midline thus presents an opportunity to examine patterning by Notch signaling in a remarkably discrete set of cells. Unlike the field-dynamics of Notch signaling in other animals (see above references), gene expression analysis suggests that in the *Ciona* dorsal midline, *Delta* is expressed in binary manner within midline cells; either on or off, with no evidence that it is initially broadly expressed throughout the midline. The initiation of this patterning process is still unclear. Midline territory formation is well characterized [26, 29], and all midline cells can become ESNs [20, 27, 32, 95], yet it is not known clear what establishes the initial pattern prior to lateral inhibition refinement. Below, we summarize the experimental data herein, and assemble working models that propose new genetic relationships and how ESN patterning and specification may be operating.

In our search to identify additional factors that regulate ESN development, we presented ISH data for over a dozen candidate genes that may be involved in ESN development, and functional data for the TFs *EMC2*, *Smad6/7*, *Neurogenin*, *Prx*, and *CEBP*. *EMC2*, known in vertebrates and arthropods alike as a repressor of proneural genes, has a functionally conserved role in *Ciona*. Its expression pattern (Fig. 4Fii) combined with results of overexpression suggests *EMC2* is a non-ESN midline repressor of CESN genes (Fig. 2D, 3Di, and 3Diii). *Smad6/7* in other organisms is an inhibitory *Smad* [96, 97], and indeed when we ectopically expressed *Smad6/7* we saw a reduction of cilia, dorsal *Pou4*, as well as dorsal and ventral *Delta2* reduction (Figure 3Bi and 3Bii, respectively). However, *Smad6/7* is expressed in the CESNs (KH.C3.230.; [40]), suggesting that the ESNs are being prevented from responding to BMP/TGFb signaling. Because ectopic *Smad6/7* could abolish all cilia except for the anterior-most four dorsal pairs (Fig 2E), this hints intriguingly that there are specific populations of CESNs that are not

specified in the same way as others. For example, we showed here and previously [14] that CESN gene expression and cilia formation requires *Pou4*. However, it appears that if *Smad6/7* is overexpressed ectopically, one small population of stereotypically-positioned cilia may nevertheless develop in the absence of *Pou4* (compare Fig. 2E with 3Ei).

We showed that ectopic epidermal expression of *Neurogenin* induces the expression of ectopic *Pou4* and the BTN marker *Asic* [78], but it is unclear if it also regulates ESN specification (images and further description in Fig. S2). Ectopic epidermal expression of *Neurogenin* resulted in the activation of *Pou4* and *Delta2* (Fig. 3Aiii and 3Aiv), however Notch signaling is thought not to be required for normal BTN development [78]. *Neurogenin* is not expressed in CESNs but it is expressed early in the lineage that leads to both BTNs and CESNs (Fig. 1Qi and 1Qii, and [78]); *Pou4* is also expressed in the BTN lineage [35, 98]. While ectopic expression of *Pou4* throughout the epidermis does not induce ectopic *Ngn* expression (Fig. S3), it does ectopically activate *Hmx3* and *Lhx3* [99], both of which are expressed in BTNs [25, 78]. Our data suggests that in addition to playing a key role in the production of CESNs, *Pou4* also plays a role in the regulatory network responsible for producing BTNs. *Pou4* thus appears to have the ability to convert epidermal cells into at least two different neuronal cell types: CESNs and BTNs. Future experiments will be needed to determine whether CESN development and patterning requires early *Ngn* activity (see below).

Figure 7 is a model of the genetic interactions inferred from the experiments performed in this study. The *Msxb* gene [29] probably plays an indirect role in activating *Delta2* (dotted arrow, Fig. 7A). Ectopic *Msxb* expression did induce *Delta2* expression (Fig. 4Biv), but it also produced ectopic *Pou4* expression (Fig. 4Biii). In the wild-type embryo, *Msxb* is expressed hours earlier in the midline lineage prior to midline *Delta2* expression [29] and it is unclear if *Msxb* must

activate *Pou4* prior to activating *Delta2*. Future experiments in which embryos ectopically express *Msxb* while also knocking down *Pou4* could address this. To date, we have been unable to determine if *Pou4* is expressed before *Delta2*, or vice-versa in CESNs. Ectopic expression of *Pou4* produces ectopic *Delta2* (Fig. 4Cii), and both Pou4-WRPW and CRISPR-Cas targeting of *Pou4* leads to reduced *Delta2* expression (Fig. 5Ci and 6A, respectively) suggesting that *Pou4* is needed to activate *Delta2* in CESNs.

Ash a-like 2 expression begins at the tail tip at the start of neurulation [29], and spreads within the dorsal and ventral midlines anteriorly throughout the process of neural tube closure (Fig. 1U). A previous report showed *Ash a-like 2* is activated by *Msxb*, *Klf1/2/4*, and *Dll-C*, and represses several genes including the pan-neuronal marker *ETR* [29]. Our results provide evidence that *Ash a-like 2* is a transcriptional activator, and can induce *Delta2* expression (Fig. 7C). Ectopic epidermal expression of *Ash a-like 2* led to *Delta2* expression throughout the epidermis (Fig. 4Dii), and converting *Ash a-like 2* into a repressive form switched its activity including perturbed (but not expanded) *Delta2* and reduced *Pou4* expression (Fig. 5Ai and 5Aii, respectively), including the number of cilia in the midlines (Fig. 5Ai, 5Av). In embryos where *Ash a-like 2* was knocked-down, *Delta2* expression was reduced (Fig. 6Bii). The discrepancy between *Ash a-like 2* acting as a repressor [29] or as an activator may be explained by the fact that *Ash a-like 2* activates the expression of *Delta2* and the subsequent activation of Notch signaling throughout the epidermis which has been shown to prevent CESN development [20, 32].

Our data is consistent with *MyT1* regulating *Delta2*, at least indirectly. Expressing ectopic *MyT1* in the epidermis led to an increase in *Delta* expression (Fig. 4Avi) while CRISPR-Cas targeting of *MyT1* led to a decrease in *Delta2* expression (Fig. 6Di). A recent paper on the *MyT1*

homologue in *C. elegans* [100] provides insight into the role *MyT1* plays in the ascidian CESNs. In *C. elegans*, *MyT1* is expressed in neuronal cells and represses non-neuronal genes in those cells; ectopic expression of *MyT1* produces ectopic neurons. Our results are consistent with a similar role for *MyT1* in *Ciona*: 1) ectopic epidermal expression of *MyT1* produces ectopic ESNs and ectopic *Delta2* expression, 2) Expressing the forced repressive form of *MyT1* (MyT1-WRPW) in the epidermis produces the same result suggesting that *MyT1* acts as a transcriptional repressor, 3) *MyT1* function is consistent with it repressing at least two midline, but non-ESN, transcriptional repressors: *EMC2* and *Orphan-bHLH*. *Orphan-bHLH* has been shown to act as a repressor [29] and we show that it also represses *Delta2* and *Pou4* (Fig. 4Eii, 4Eiii, 7F, 7I). *EMC2* has a similar role; ectopic expression of either of these genes prevents CESNs from forming (Fig. 2D); ectopic *MyT1* expression prevents these repressors from being expressed, thus permitting midline cells to become CESNs (Fig. 4Aii, 4Aiii).

Ectopic epidermal expression of *Pou4* resulted in a decrease in *Orphan* expression, and likely only indirectly inhibits that gene. *Pou4* genes in vertebrates have been shown to in some cases act as repressors [101], however the reduction in *Orphan* expression by ectopic *Pou4* is likely indirect. Ectopic *Pou4* activation of *MyT1* (Fig. 7G) would in turn repress *Orphan*. Figure 7G is also supported by the result that Pou4-WRPW inhibited *MyT1* expression (Fig. 5Cii). In a recent RNA-SEQ analysis of ectopically-expressed *Ciona Pou4* and all three mouse *Pou4* genes, *MyT1* was upregulated in all four datasets, indicating that *Pou4* regulates *MyT1* and the conserved mouse *Pou4* genes have the same property (manuscript in preparation). Furthermore, vertebrate *Pou4* likely activates *MyT1* demonstrated by the down-regulation of *MyT1* in the absence of *Pou4* [102, 103]. That *CrPou4* appeared unable to elicit ectopic *CrMyT1* expression in Joyce Tang et al. (2012) may simply have to do with the quantity of *Pou4* delivered during

electroporation. EpiB::Pou4 quantity was reduced in those experiments so as to not alter embryo morphology so strongly it prohibited accurate counting of affected midline cells [30]. The amount of transgene DNA we used is twice the amount delivered by Joyce Tang et al, 2012. Based on the results here, we conclude that *MyT1* functions in CESNs as a repressor of at least two repressors (*EMC2* and *Orphan-bHLH*) that otherwise would inhibit *Pou4* expression and CESN development.

Prx is expressed in a similar pattern to *Orphan* and *Dll-C* [29, 40]. Expression of these three genes is found throughout almost all of the dorsal and ventral midlines, except for single-cell gaps that probably account for punctate expression of CESN genes such as *Pou4* and *MyT1*. Different *Prx* homologs are considered transcriptional activators and repressors in mammals [104], and the results from experiments we performed indicate that the *Ciona Prx* is a transcriptional activator of repressors and indirect repressor of CESN genes. Ectopically expressing *Prx* resulted in loss of cilia (Fig. 2F) and loss of CESN gene expression (Fig. 3Cii, 3Ciii), but also resulted in an increase in *Orphan* expression (Fig. 3Ci). Conversely, Prx-WRPW resulted in the ectopic activation of *Delta*, *MyT1*, and *Pou4*, but reduced *Orphan* expression (Fig. 5Di – 5Div). Because adding the repressive domain switched the function of the protein as indicated by a reversal of gene expression outcomes, we conclude that *Prx* activates (at least) Orphan and thus indirectly represses CESN genes. We recognize that generating a working ISH assay for this gene is critical to further understanding its roles and activity.

In summary, Figure 7 describes a core network of genes involved in the specification of CESNs, and adjacent non-CESN midline epidermis. This CESN-specification network partially integrates the midline specification network from Roure and Darras [29] with the post-specification CESN GRN from Joyce Tang et al [30], with two distinct caveats. The first is that

our CESN differentiation network places *MyT1* upstream of *Pou4* while we propose an inverse relationship. However, concluding *MyT1* as a repressor of repressors in the ESNs fits experimental evidence here and elsewhere [30]. That role is critical to the GRN in Figure 7 and our overall understanding of ESN development. The second caveat is that the network described in Figure 7 does not account for how an undetermined cell specified by the midline network selects between a CESN fate and a non-CESN midline fate. *Msxb* which can or in part activate pro-CESN genes (e.g. *Pou4*) *and* anti-CESN genes (e.g. *Orphan*), argues that there is a missing regulatory component. We present a hypothetical model of this problem in Figure 8.

Figure 1U shows *Ash a-like 2* expanding its expression domain over time throughout the dorsal and ventral midlines. Expansion of gene expression within a particular territory has previously been documented in *Ciona*, for example ventral midline *Msxb* expression was shown to expand posteriorly between stages 15-18 [28]. The mechanism by which *Msxb* expands posteriorly is unknown, but we may speculate about mechanisms driving posteroanterior *Ash a-like 2* expression. In ascidian embryogenesis, cell division is well synchronized, particularly animal lineages prior to neurulation [105]. At the onset of neurulation, the cells of the epidermis have undergone 10 divisions since fertilization, all synchronous and without a G-phase [106]. At this point in development, epidermal cells introduce a G-phase to their cell cycles speculated to provide the time needed by the process of neural tube closure [107]. Both of those reports document an posteroanterior wave of epidermal mitoses occurring roughly in step with neurulation [106, 107]. In *Drosophila*, both aforementioned events (proneural gene wave-like expansion [108], and mitotic wave activity [109]) have roles in developing neural tissues. Waves of proneural bHLH genes are also documented in vertebrates [110, 111]. The mitotic wave in the *Ciona* epidermis is regulated by *Tfap2-r.b* and *GATA3* expression [107], genes known to be

involved in other aspects of ascidian larval development [27, 56, 89]. Therefore it may be that these genes, while regulating mitotic control of the epidermis, are also underlying contributors to the anteriorly-expanding expression of *Ash a-like 2* and the subsequent pattern of CESNs.

As *Ash a-like 2* expression spreads anteriorly in the two sides of the dorsal midline cells, *Delta2*, *MyT1* and *Pou4* are activated in the first presumptive CESNs; this may require cooperation with *Msxb* activity (Fig. 8). *Delta2* activity in the presumptive CESNs would then signal the anterior-adjacent cell activating Notch signaling and preventing that cell from adopting a CESN fate. As *Ash a-like 2* expression continues to spread anteriorly in the dorsal midline cells, it will then be expressed in cells that are expressing *Msxb* but without Notch activity, and can thus activate *Delta2*/*Pou4*/*MyT1* and begin the initial patterning cycle again. This necessarily leads to the every-other pattern documented during initial *Pou4* expression (Fig. 8). Later in development, the increase number of inter-ESN midline cells would be a result of cell division and cell intercalation from the right and left sides of the embryo, spacing may be provided and maintained by long-distance Notch signaling [31]. This data-driven model is remarkably similar to the model proposed by Quan and Hassan 2005 [112] of Drosophila neurogenesis. Their summarized network of neural specification portrays the activities of *Ash a-like 2*, *Delta2*, *MyT1*, and other homologs in nearly identical ways as we propose here. In that model, achaete-scute bHLH proteins activate Delta which in turn induces Notch activity in the neighboring cell, culminating in the prohibition of PNG expression. Any Notch activity in the neuronal precursor is blocked by a *MyT1* homolog allowing PNGs to be expressed and neuronal development to proceed. In *Ciona*, the expression of *miR-124* in CESNs also contributes to silencing Notch activity in those cells [32].

The number of genes required for proper differentiation of any given cell type is probably in the many hundreds; genes for function likely push this number into the thousands. We have here added several candidates to the list of known genes that are involved with the specification and differentiation of the ascidian CESN, while also strengthening and refining previously studied relationships. Foremost, the model we present in Figure 8 is the first to suggest how these cells are selected and specified as either ESN or non-ESN midline cell. Numerous follow-up experiments have been proposed above that will clarify various remaining ambiguities. A fuller understanding of this cell type provides insights into the evolutionary history of ciliated mechanoreceptors, and also advances the field toward the increasingly tangible goal of a complete understanding of gene networks that drive ascidian embryogenesis.

ACKNOWLEDGEMENTS

Clifford Pickett gratefully acknowledges ARCS Foundation support, San Diego Chapter. This work, Chapter 3, "*Regulatory Interactions Governing the Specification of Epidermal Sensory Neurons in the Ascidian Ciona robusta***;** C.J. Pickett and Robert W. Zeller", in full, is being prepared for publication. The dissertation author was the primary investigator and author of this material.

Table 3.1: Candidate CESN-regulatory genes and gene models. The first column lists the gene name of the twenty nine candidates; second column lists Ghost Database gene model ID [40]; third column lists reasons for each candidates presence. Highlighting: Blue identifies genes with ISH data presented below, Red identifies candidates we were unable to generate interpretable ISH patterns for (or unable to clone), Green identifies genes that have previously published expression patterns that we subsequently reanalyzed.

Gene Name	Model	Motivations for Analysis
Pax1/9	KH.C11.488	predicted CNE binding (Pax1)
ATFc	KH.C1.26.	predicted CNE binding (MmTF homolog)
C/EBP	KH.C3.176.	predicted CNE binding (MmTF homolog)
ZF163	KH.L153.63.	predicted CNE binding (MmTF homolog)
FoxC	KH.L57.25.	predicted CNE binding (MmTF homolog)
AP-4	KH.C14.448.	predicted CNE binding (MmTF homolog)
Jun	KH.C5.610.	predicted CNE binding (MmTF homolog)
Lhx3	KH.S215.4.	predicted CNE binding (MmTF homolog)
Fox7	KH.C14.230.	predicted CNE binding (MmTF homolog)
Tlx	KH.L84.30.	predicted CNE binding (MmTF homolog)
NFIL3	KH.C5.520.	predicted CNE binding (HsTF homolog)
NF-YB	KH.S1818.1.	predicted CNE binding (HsTF homolog)
Lhx1 (LIM)	KH.L107.7.	predicted CNE binding (HsTF homolog)
Bicaudal-C	KH.C1.971.	up-reg in RNAseq dataset [21], predicted CNE binding
S1313	KH.S1313.1.	up-reg in RNAseq dataset [21]
nas-15	KH.L96.44.	up-reg in RNAseq dataset [21]
Sox4/11/12	KH.C7.523.	sensory neuron development [55]; ISH pattern
Smad6/7	KH.C3.230.	repressive Smad [59, 60] expressed in ESNs
Neurogenin	KH.C6.129.	pre Notch-signaling TF [52, 53]; expression pattern
AP-2-like	KH.C7.43.	pre Notch-signaling TF [51]; predicted CNE binding; MmTF homolog
Emx	KH.L142.14.	Notch-signaling TF; sensory neuron development [56]
EMC2	KH.C7.157.	repressive TF: Mm Id-3 homolog [61]; Dm extra-macrochaete homolog [62]
Prx	KH.C1.414.	ISH pattern and repressor evidence [49, 50]
EvxB	KH.C3.836.	ISH pattern [54]
Barx-C	KH.L41.33.	inner ear TF [47], predicted CNE binding
Hmx3	KH.S563.4.	inner ear TF [46, 64], predicted CNE binding
Dlx.b	KH.C7.243.	inner ear development [57, 58]
Pax3/7	KH.C10.150.	implicated in sensory neuron development [63]
Sprouty	KH.C2.825	FGF antagonist, deafness gene [65]
ZF-248	KH.C5.74	up-reg in RNAseq dataset [21]
MyT1	KH.C1.274.	[29, 67]
Pou4	KH.C2.42.	[29, 34, 66]
Orphan-bHLH	KH.C7.269.	[28, 40]
Delta2	KH.L50.6.	[19, 29, 31, 32]
Ash a-like 2	KH.C9.13	[19, 28, 29]

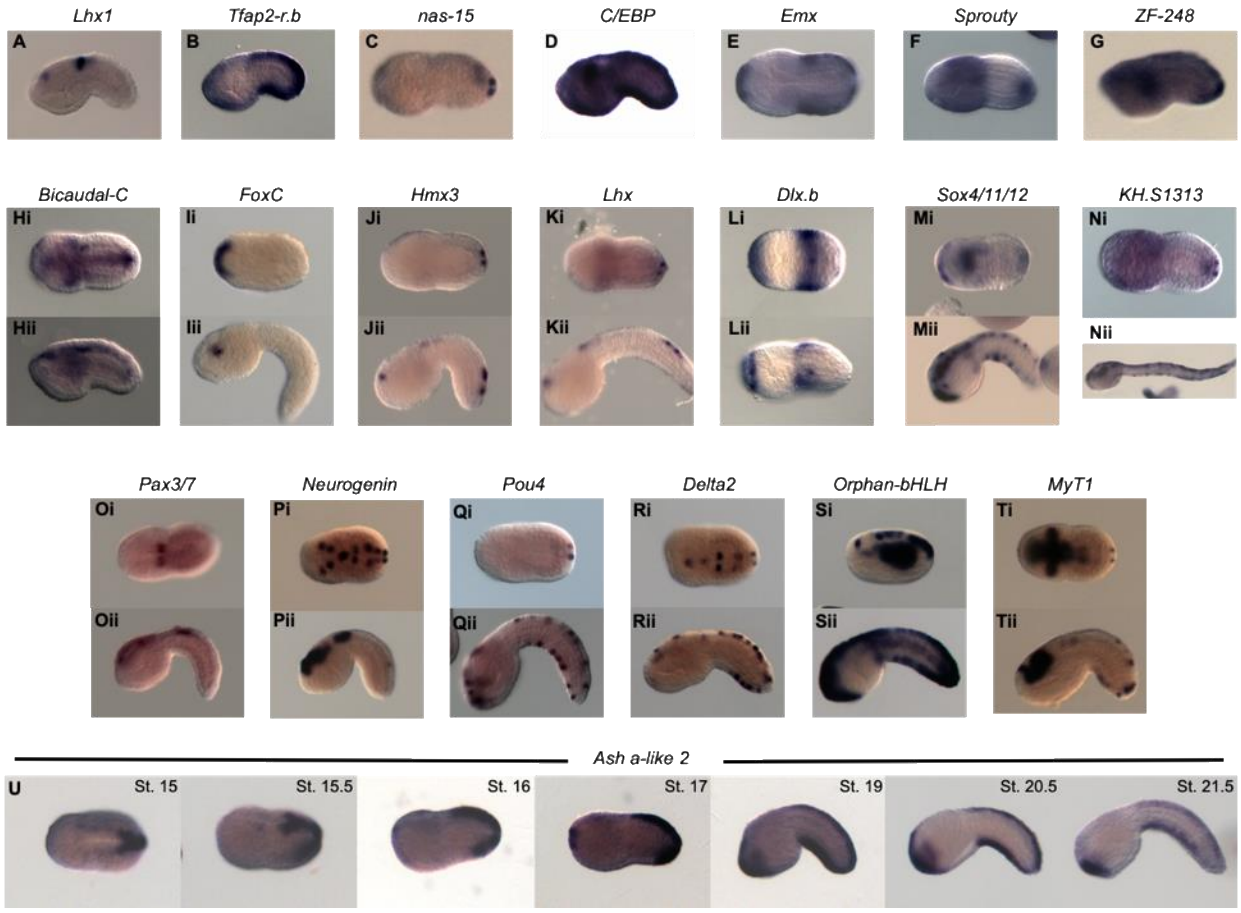


Figure 3.1: Wild type gene expression patterns of genes listed in Table 1. All embryos displayed anterior to the left. Embryos in E, F, Hi-Ni, Oi-Ti and U (St. 15) are dorsal views. All other embryos positioned laterally with dorsal side to the top. In Hi-Tii, embryos in upper row for a particular expression pattern shown at an earlier stage than the embryos in the corresponding row below. U displays time course of *Ash a-like 2* expression, spanning left to right from St. 15 to St. 21.5.

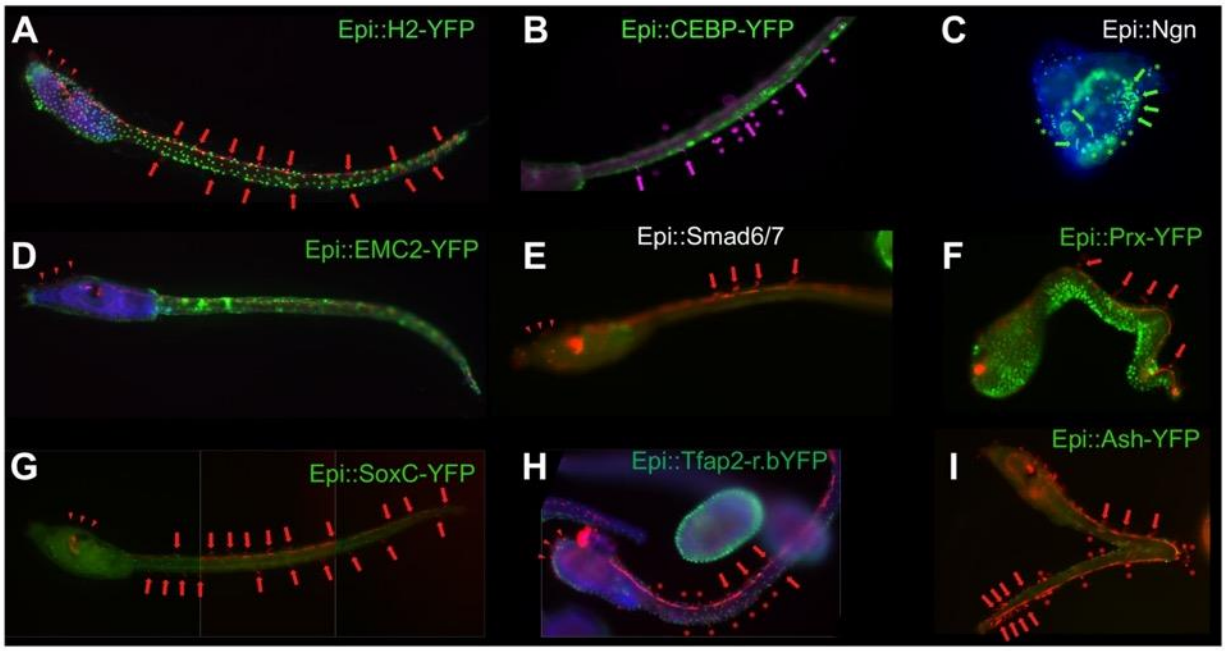


Figure 3.2: Ectopic epidermal expression of select candidate genes disrupts cilia patterning and formation. Fluorescently immunostained acetylated tubulin; Red in A and C-I, Magenta in B. Green: fluorescence from H2-YFP (control), TF-FP fusion protein (experimental, B-I), or slight background autofluorescence in mesodermal tissues (e.g. trunk regions of E and G). Red or magenta arrows indicate presence of a cilia; red or magenta asterisks indicate globular acetylated tubulin staining, likely detecting malformed cilia. Red arrowheads indicate trunk ESNs. Blue: DAPI staining of nuclei. Vertical grey lines in "G" to demarcate different focal planes. All larva positioned anterior to the left and dorsal to the top. Each experiment was repeated at least 2X; at least 50 embryos were examined per experiment. Images represent dominant phenotypes.

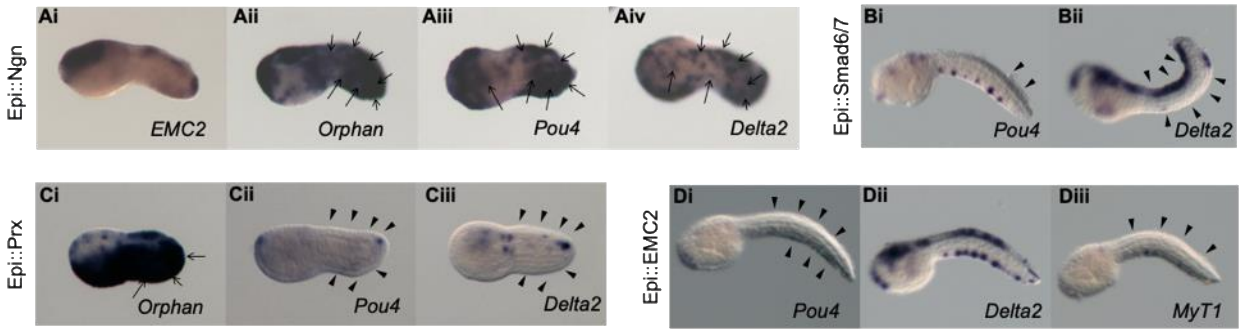


Figure 3.3: Ectopic expression of candidate genes alters downstream wild type expression patterns. Ai-Aiv, Epi::Ngn probed for *EMC2*, *Pou4*, *Delta2*, and *MyT1*; Bi and Bii, Epi::Smad6/7 probed for *Pou4* and *Delta2* (both normal and curled tail morphologies present in batch probed); Ci-Ciii, Epi::Prx probed for *Pou4*, *Delta2*, and *MyT1*; Di-Diii, Epi::EMC2 probed for *Pou4*, *Delta2*, and *MyT1*. All embryos are positioned anterior left and dorsal to the top. All results captured are from the same electroporations. Arrows indicate expanded expression of assayed gene vs. controls; Arrowheads indicate reduced expression vs. controls. Epi::Smad6/7 and Epi::EMC2 experiments repeated 2X while Epi::Ngn and Epi::Prx experiments were repeated 3X; at least 50 embryos were examined per experiment. Images represent dominant phenotypes. Control images displayed in Fig. 4 Fi-4Fvi.

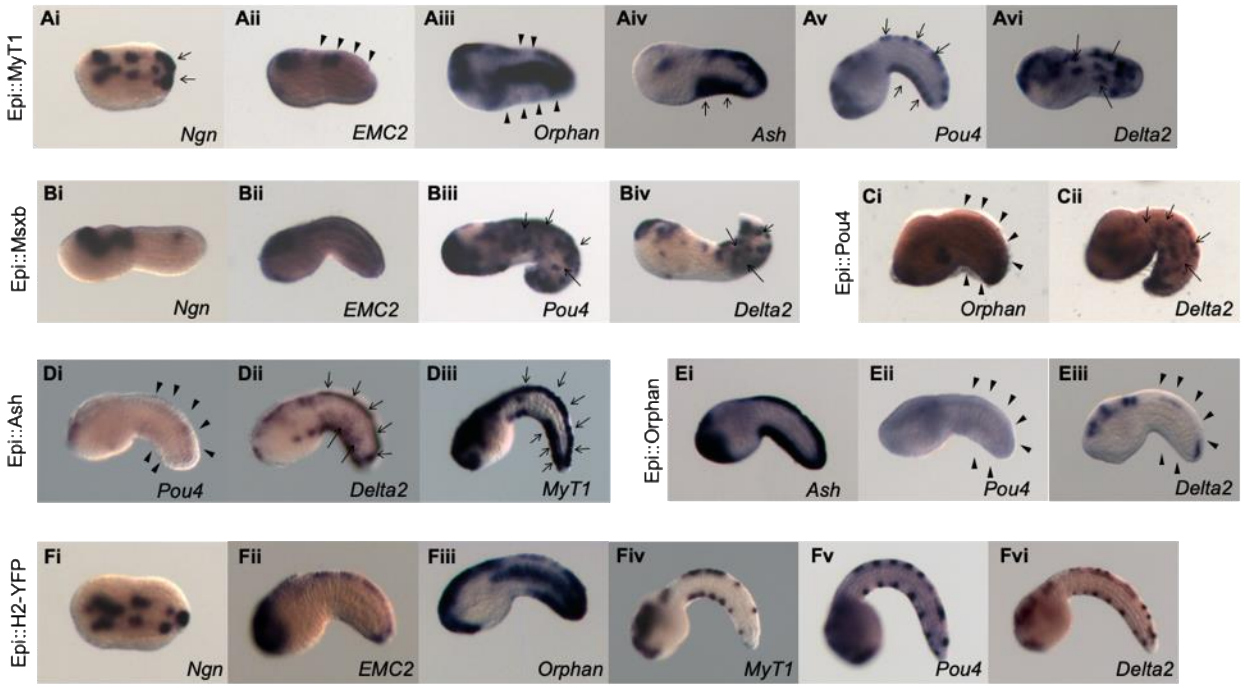


Figure 3.4: Ectopic expression of previously examined TFs alters expression patterns of various genes. Ai-Avi, Epi::MyT1; Bi-Biv, Epi::Msxb; Ci and Cii, Epi::Pou4; Di-Diii, Epi::Ash; Ei-Eiii, Epi::Orphan-bHLH. Genes probed per electroporation indicated in bottom right of each image. All embryos are positioned anterior to the left and dorsal to the top, with the exception of Ai which was rolled to present a more dorsal view. Fii: note the presumptive ESN "holes" in dorsal *EMC2* expression. Arrows indicate expanded expression of indicated gene vs. controls; Arrowheads indicate reduced expression vs. controls (EpiB::H2YFP, Fi-Fvi). Each experiment was repeated 3X; at least 50 embryos were examined per experiment. Images represent dominant ISH expression patterns.

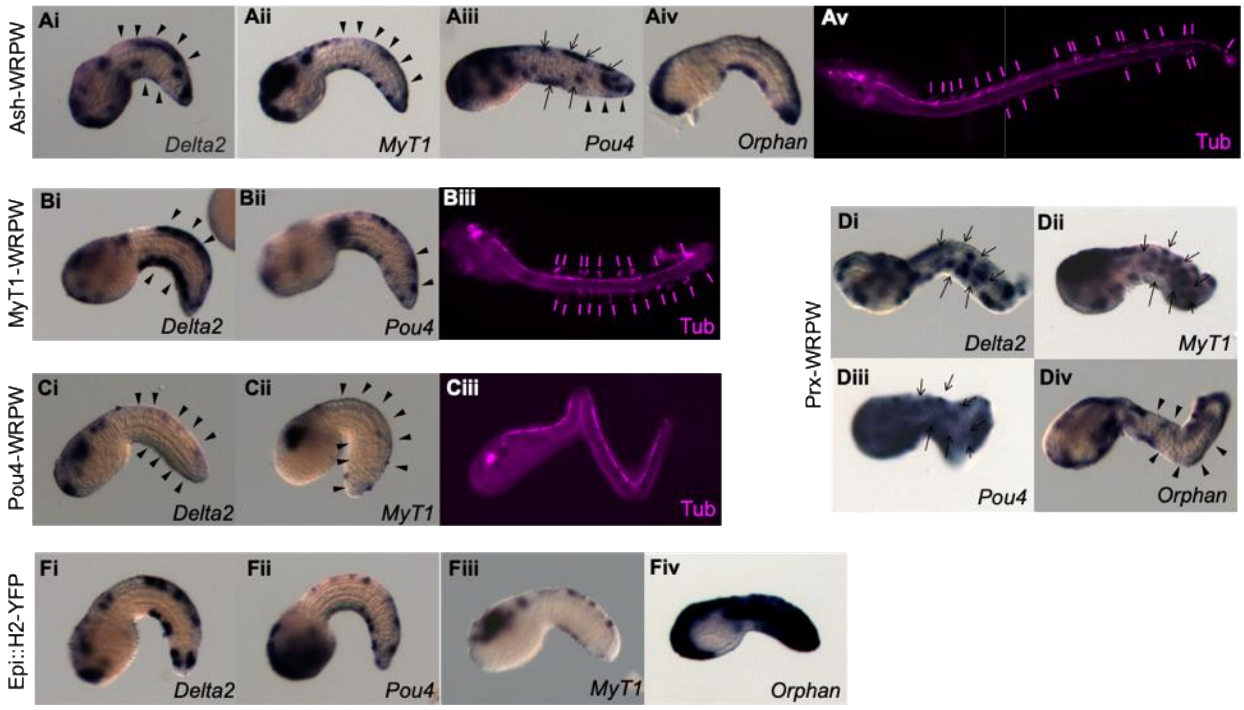


Figure 3.5: Ectopic expression of TF-WRPW fusion proteins forces repressor activity and alters ISH and cilia patterning/formation. Ai-Aiv, Epi::Ash-WRPW transgenic embryos; Av, Epi::Ash-WRPW transgenic larva. Bi and Bii, Epi::MyT1-WRPW transgenic embryos; Biii, Epi::MyT1-WRPW transgenic larva. Ci and Cii, Epi::Pou4-WRPW transgenic embryos; Ciii, Epi::Pou4-WRPW transgenic larva. Di-Div, Epi::Prx-WRPW transgenic embryos. Black arrows indicate expanded expression of indicated gene vs. controls (Fi-Fiv); Black arrowheads indicate reduced expression vs. controls. Magenta arrows indicate immunostained acetylated tubulin which marks the ESNs and neural tube. All embryos and larva anterior to the left and dorsal to the top. All WRPW-TF experiments repeated twice.

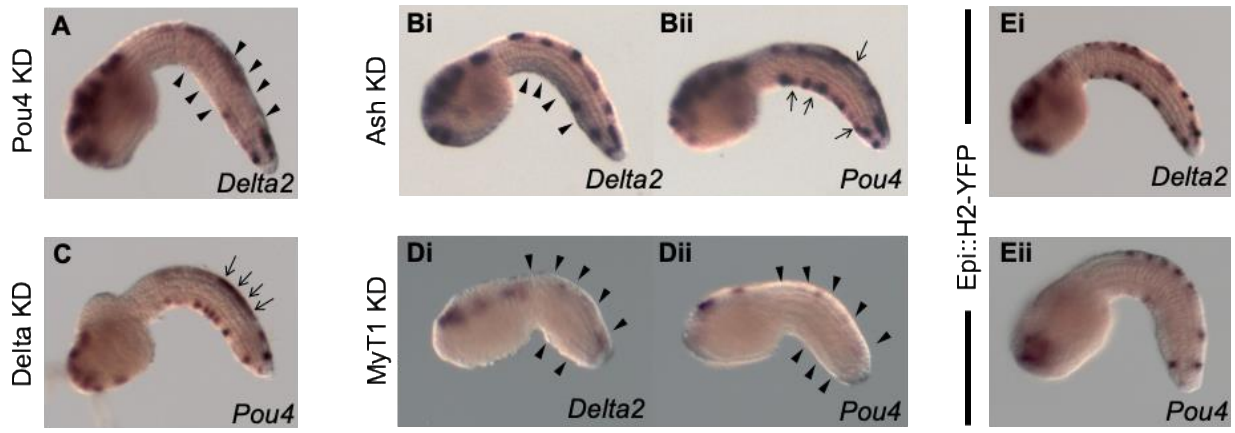


Figure 3.6: Knocking down genes with CRISPR-Cas reagents results in disrupted downstream gene expression patterns. CRISPR-Cas reagents were designed against *Pou4* (A; ISH for *Delta2*), *Ash a-like 2* (Bi and Bii; ISH for *Delta2* and *Pou4*, respectively), *Delta2* (C; ISH for *Pou4*), and *MyT1* (Di and Dii; *Delta2* and *Pou4*, respectively). All embryos positioned anterior to the left and dorsal to the top. Black arrows indicate expanded expression of indicated gene vs. controls (Ei and Eii); Black arrowheads indicate reduced expression vs. controls. Experiments knocking down *Pou4* and *Delta2* repeated three times, *MyT1* and *Ash a-like 2* knockdowns were repeated twice.

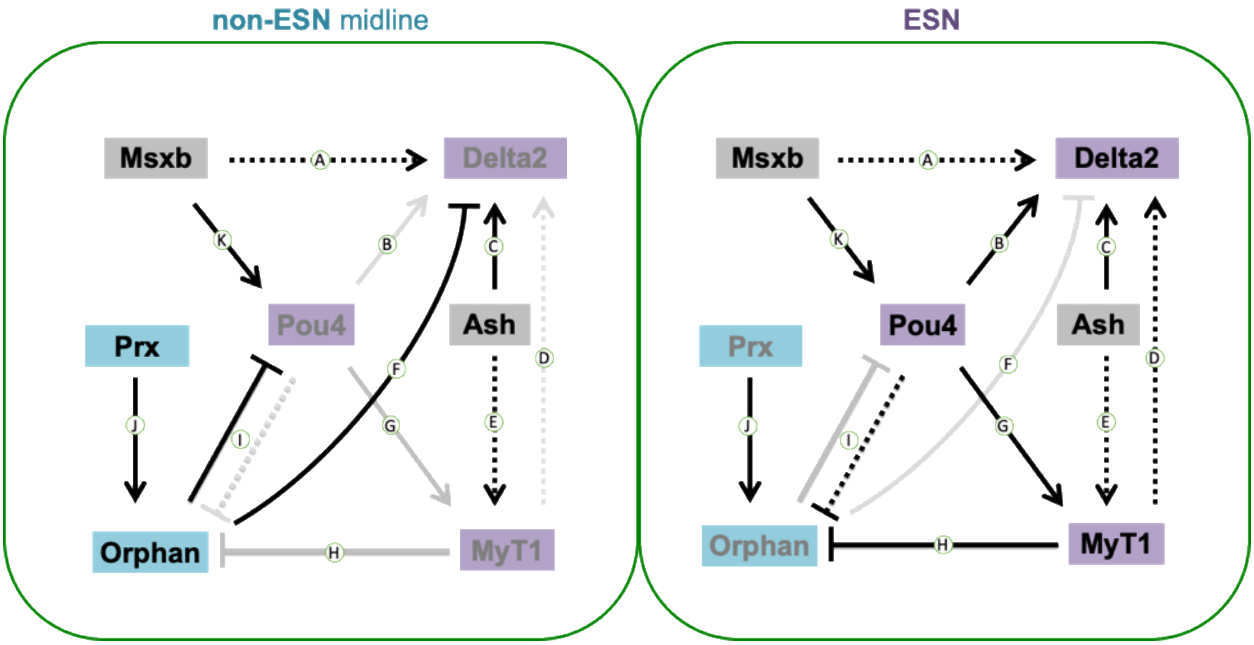


Figure 3.7: Hypothesized regulatory relationships based on experiments performed in this study. A-J mark particular examined relationships discussed in the text. Briefly: "A" supported by Fig. 4Biv (also [29]). "B" supported by Fig. 4Cii, 5Ci, and 6A. "C" supported by Fig. 4Dii, 5Ai, and 6Bi (also [30]). "D" supported by Fig. 4Aiv, 5Bi, and 6Di (also [30]). "E" supported by Fig. 4Diii and 5Aii. "F" supported by Fig. 4Eiii. "G" supported by Fig. 5Cii. "H" supported by Fig. 4Aiii. "I" supported by Fig. 4Ci and 4Eii. "J" supported by Fig. 3Ci and 5Div. "K" supported by Fig. 4Biii. Grey boxes highlight genes expressed throughout the midline; Purple boxes highlight CESN genes; Blue boxes highlight non-ESN midline genes. Solid arrows indicate hypothesized direct activation; Straight "T"'s indicate direct repression. Dotted arrows indicate hypothesized indirect activation; Elongated dotted "T"'s indicate indirect repression. Black solid or dotted arrows and "T"'s indicate relationships active in either the CESN or non-ESN midline cell; Lt. Grey solid or dotted arrows and "T"'s indicate relationships not active in the particular cell type.

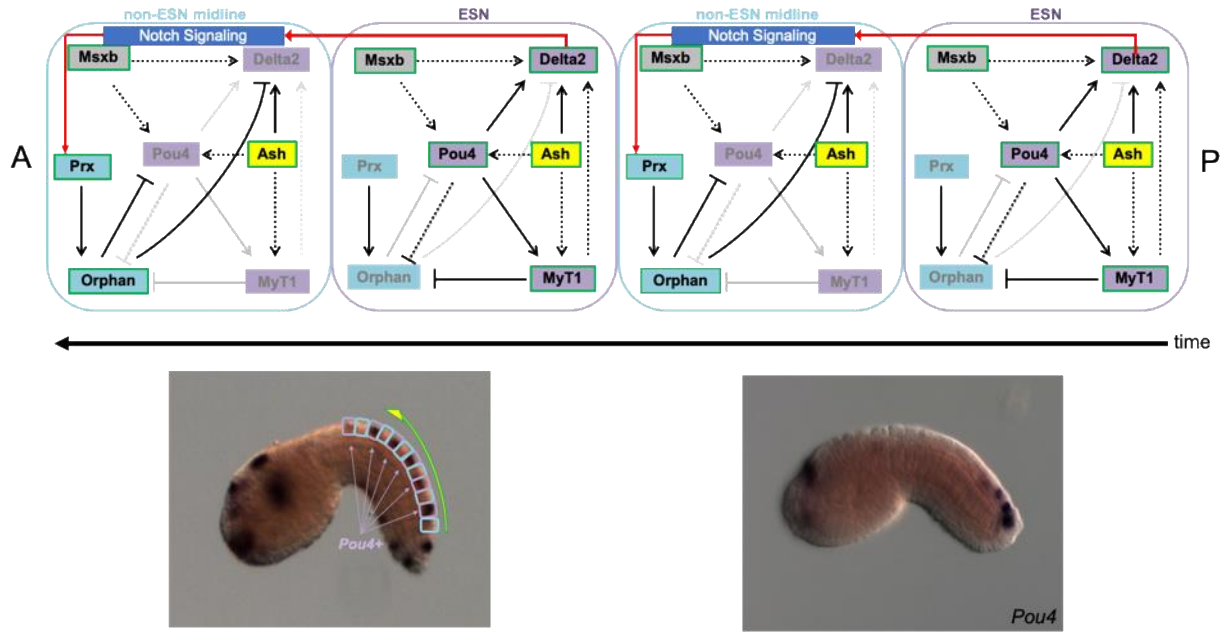


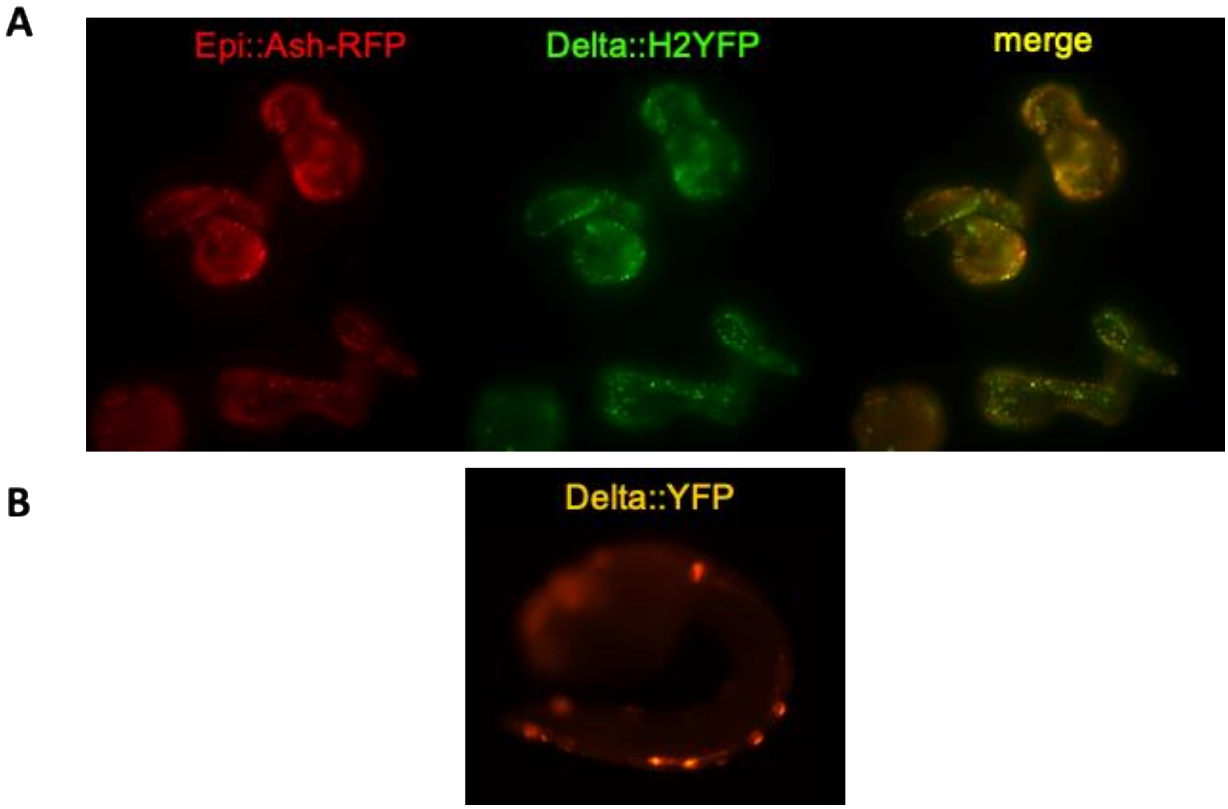
Figure 3.8: Hypothetical model explaining spatial dynamics of ESNs in the dorsal midline requiring MyT1 repressive activity and Ash activation of Delta2. *Ash a-like 2* (yellow) is expressed posteroanteriorly, activating *Delta2* and others (ESN fate), except in the subsequent cell where *Ash a-like 2* expression encounters Delta2-activated Notch signaling (generating non-ESN midline fate). In each "cell" shown, genes active are outlined in green with black text, whereas genes not active are not outlined and have grey text. See Figure 7 for description of arrows and T's. Bottom right: image of a wildtype St. 19-20 tailbud larva probed for expression of *Pou4*, which is expressed in a few cells of the dorsal midline at that stage, plus the developing BTNs. Bottom left: Stage 21-22 tailbud larva probed for expression of *Pou4*. *Pou4*+ cells are boxed purple with purple arrows, non-ESN midline cells boxed in blue. Green and yellow arrow signifies anteroposterior pattern of *Ash a-like 2* expression. Embryo images and model shown as anterior "A" to the left, posterior "P" to the right.

SUPPLEMENTAL INFORMATION

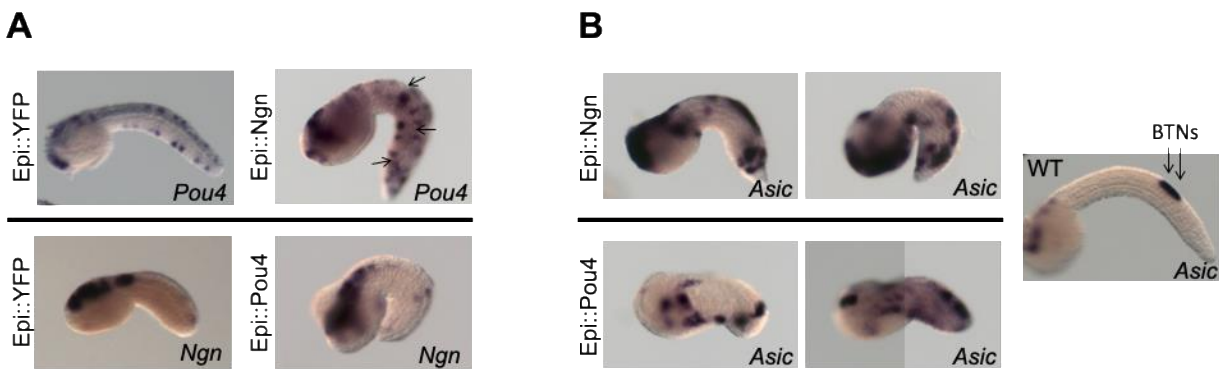
Supplementary Table 3.1: Gene models and primer sequences used to generate ISH probes and overexpression transgenes, and gRNA target sequences used in CRISPR-Cas knockdown experiments.

Gene	KH gene ID	Fwd	Rev
Prx	KH.C1.414.	ATGACAGAACATTCAACAGCAAGAAG	TTGTAAGAATCCTCTTTGAAACGTCATACC
AP-2-like	KH.C7.43.	ATGAGTGATATTCGAATTCTGTCC	GATACTGGTATGAGGCCTGATGAC
Neurogenin	KH.C6.129.	ATGTTGGATTTTTC TTCAAATAAG	GCCAAACTACTCCATGCATTTACTCCGA
S1313	KH.S1313.1.	ATGGCATATCATGATAACTTACGTCACAG	GTTTACGGATTGTTCTAGCCATCC
nas-15	KH.L96.44.	CGTGACTGGCAAAGGCTG	TTGATTTTTGTCAACTTTCCAGTTGGG
EvxB	KH.C3.836.	ATGATCGAGAACGAGTTACTACCATTG	CCTTCCAAACATCATCATCAATTCC
SoxC	KH.C7.523.	ATGGCAAGTACATCAAGACTTACAAAG	CGGACAGCAGATGTCGTCCC
Emx	KH.L142.14.	TTCGAGATATTTTAGTTTTTAAAAATAGATTTAAG	GGGAAATTCGTTCAAGCGTG
Dlx.b	KH.C7.243.	TGTAACATAAAGCATTCTAGTTTAAAATGC	GGAGATCGTCATCAGTATTTCCAACTTC
Smad6/7	KH.C3.230.	ATGTTGACGCTGTATAGAAAAAGAC	AGAATGCTGAGATGGGTTAAATAGAATTTC
EMC2	KH.C7.157.	ATGGTTAAAGTTGTTCGAAAATCATC	AACTGCTTGCAACCTTGTC
ATFc	KH.C1.26.	ATGATGAGTATGTTGCCTCATATAGAGG	ATCGTTTTGTTTGTGAACGTACAG
C/EBP	KH.C3.176.	ATGCTGTGAAGAAAAGTCGTGAAAAATCACG	GCTTTTTCAAATGGAGGAATTAATGTTTC
Pax1/9	KH.C11.488	ATGTTGTCGTGTCTACCACAAAG	CGTTGATGACGTCTGACCATTAG
Pax3/7	KH.C10.150.	ATGATGCATCCAGGGTCTAATTTC	ATATGCATGCTGAACGCTC
ZF163	KH.L153.63.	ATGGAGTCAATAAAATGTGCAAAGAAAAC	GTTATGTGTCTCTTCATGCTGTTTTCTG
Hmx3	KH.S563.4.	ATGAAGAGATATCTAAGCAGTTCTGAACG	AACCACGGGATTCCCC
NFIL3	KH.C5.520.	ATGAATAATAACAACTTAAATAATGATGACTTTCC	GGTTAGTCTGTGTGAGAGTTAAAATTATG
FoxC	KH.L57.25.	CTATAATGACAATGCAAATCCC	CCTGATGAGTCAGGTCTTGTAAGC
Barx-C	KH.L41.33.	ATGTCGAACGGTTGTGACG	GAAAGCAGGGGATCATCAACG
AP-4	KH.C14.448.	ATGGCTTTCTACGCGAAGAC	GTACCGCACTCTAACCCAC
Bicaudal-C	KH.C1.971.	ATGGCGGTGAACTCGAATG	GTTAGTCAGAGTGGTCGTTGG
Jun	KH.C5.610.	ATGGAAGTAACGTTTTACCACGACG	CCCAACAGCAGGTCACCTTT
NF-YB	KH.S1818.1.	GTATATATTCGTTGCCTGGCTTTAGTG	ATCTCTGTACTTTTGAAGGAAAACTTTCAG
Sprouty	KH.C2.825	ACGCTACTAAGCATTAAGCCATTTTAAATACC	GGTCTTCGTTGATTTATCGCTTTTTAAATCC
Lhx3	KH.S215.4.	ATGCAGACCGGAAGTGAGTTTCATC	ACTTGACCACGTGACACATTTCCAA
Fox7	KH.C14.230.	ATGTTCGGCAATAAACGTGGAG	TTTTGGTTCTGAGTCAAAATCGAGC
Tlx	KH.L84.30.	ATGCATCAAACCAGGCCCATGCCAGCG	TTGTGCATTCTGAAGGCGGGCTTGTAG
Lhx1 (LIM)	KH.L107.7.	ATGGAAGGACCTTTATGTAGCAGATG	ATCTATTGTTTACTGGAATGCAGCG
Pou4	KH.C2.42.	ATGTTTACTAACATGCTTGCTCCA	ATTTAATGGGGACGTGATTATG
MyT1	KH.C1.274.	ATGGATACAGCGGTCCACCACTCTGT	CAGTCGTGCCCGATGCCTGGGTG
Ash a-like 2	KH.C9.13	ATGGCGACCGGAAGTGACGAACC	GATTGTGGTTTAACCAATCAGCGTCATTGGC
Orphan-bHLH	KH.C7.269.	ATGGTTAAAGCGAGCCCGATCAAAG	ATAATTCCAATTCCAGAACGCGAGAG
Delta2	KH.L50.6.	ATGAGCATCAAGCTTATATTACTTCTC	CTCCCCGCTTACGATGAACGTTAC
Asic	KH.C1.215.	ATGATGATACAACCCAACGACAACGAAAGG	GGCATGGTGAACCCGTTGTTCATATC
HairyB (WRPW)	KH.C3.312.	ATGTTGTTTCCTCAGCATTGCTCACC	TTACCATGGTCTCCATACTGGATCAGAAT

gRNA	Target Sequences
Ash a-like 2-173	TCACTAATCTGGCCTCCAAG
Ash a-like 2-652	CGACCGGAAGTGACGAACCG
Delta-606	GTTCCTCCATTGGAGCACCG
Delta-1153	GGAGGTAGAGACATCAAACG
Pou4-0	ACGTTCATATTATTATCTAT
Pou4-423	GCGTATTATAATATCTACAG
MyT1-173	GTTACCTCCGTCATGCACTC
MyT1-454	CACGTCGCTATATGGCACTG



Supplementary Figure 3.1: Ectopic expression of Ash a-like 2 activates a Delta2::YFP transgene. Top row: EpiB::Ash-RFP and Delta2::YFP were coelectroporated and embryos developed for 16hpf resulting in embryos exhibiting cells containing RFP and YFP expression. Bottom row: Stage 23 embryo demonstrating the Delta2::YFP transgene is, without other influences, fidelitous to the Delta2-expressing lineages (i.e., YFP expression in motor ganglion, CESNs and some cells of the nerve cord). Top row, hatched swimming larval stage; bottom row, stage 24 embryo. Coelectroporated transgenes remain together in the cells in which they are active [16], i.e. if one cell carries a plasmid it will also carry the other coelectroporated plasmid. Epi::Ash a-like 2 resulted in the activation of the Delta2::YFP transgene in all epidermal cells carrying the plasmids, further indicating that *Ash a-like 2* is an activator of *Delta2*.



Supplementary Figure 3.2: *Neurogenin* can activate *Pou4* but *Pou4* cannot activate *Neurogenin*; both *Neurogenin* and *Pou4* activate the BTN marker gene *Asic*. A) top row, *Pou4* expression patterns after electroporation with either Epi::YFP (L) or Epi::Ngn (R). Bottom row, *Ngn* expression patterns after electroporation with either Epi::YFP (L) or Epi::Pou4 (R). **B)** *Asic* expression after electroporation of Epi::Ngn (top row, both images) or Epi::Pou4 (bottom row, both images). Rightmost image: wild type stage 23 embryo probed for *Asic* expression, arrows indicate *Asic* expression in the posterior BTNs. Experimental embryos are stage 22. All embryos are positioned anterior to the left and dorsal to the top.

REFERENCES

1. Arendt, D., M.A. Tosches, and H. Marlow, *From nerve net to nerve ring, nerve cord and brain—evolution of the nervous system.* Nature Reviews Neuroscience, 2016. **17**(1): p. 61.

2. Schlosser, G., *A Short History of Nearly Every Sense-The Evolutionary History of Vertebrate Sensory Cell Types.* Integr Comp Biol, 2018. **58**(2): p. 301-316.

3. McGlone, F. and D. Reilly, *The cutaneous sensory system.* Neuroscience & Biobehavioral Reviews, 2010. **34**(2): p. 148-159.

4. Fritzsch, B. and H. Straka, *Evolution of vertebrate mechanosensory hair cells and inner ears: toward identifying stimuli that select mutation driven altered morphologies.* Journal of Comparative Physiology A, 2014. **200**(1): p. 5-18.

5. Océane Tournière, D.D., Gemma Sian Richards, Kartik Sunagar, Yaara Y Columbus-Shenkar, Yehu Moran, Fabian Rentzsch, *NvPOU4/Brain3 functions as a terminal selector gene in the nervous system of the cnidarian Nematostella vectensis.* biorxiv, 2020.

6. Sebe-Pedros, A., et al., *Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq.* Cell, 2018. **173**(6): p. 1520-1534 e20.

7. Huang, E.J., et al., *Brn3a is a transcriptional regulator of soma size, target field innervation and axon pathfinding of inner ear sensory neurons.* Development, 2001. **128**(13): p. 2421-32.

8. Xiang, M., et al., *Requirement for Brn-3c in maturation and survival, but not in fate determination of inner ear hair cells.* Development, 1998. **125**(20): p. 3935-46.

9. Leys, S.P., *Elements of a 'nervous system' in sponges.* J Exp Biol, 2015. **218**(Pt 4): p. 581-91.

10. Gazave, E., et al., *NK homeobox genes with choanocyte-specific expression in homoscleromorph sponges.* Development genes and evolution, 2008. **218**(9): p. 479.

11. Costa, A., et al., *Atoh1 in sensory hair cell development: constraints and cofactors.* Semin Cell Dev Biol, 2017. **65**: p. 60-68.

12. Costa, A., et al., *Generation of sensory hair cells by genetic programming with a combination of transcription factors.* Development, 2015. **142**(11): p. 1948-1959.

13. Serrano-Saiz, E., et al., *BRN3-type POU Homeobox Genes Maintain the Identity of Mature Postmitotic Neurons in Nematodes and Mice.* Curr Biol, 2018. **28**(17): p. 2813-2823 e2.

14. Pickett, C.J. and R.W. Zeller, *Efficient genome editing using CRISPR-Cas-mediated homology directed repair in the ascidian Ciona robusta.* Genesis, 2018. **56**(11-12): p. e23260.

15. Delsuc, F., et al., *Tunicates and not cephalochordates are the closest living relatives of vertebrates.* Nature, 2006. **439**(7079): p. 965-8.

16. Zeller, R.W., M.J. Virata, and A.C. Cone, *Predictable mosaic transgene expression in ascidian embryos produced with a simple electroporation device.* Developmental dynamics: an official publication of the American Association of Anatomists, 2006. **235**(7): p. 1921-1932.

17. Imai, J.H. and I.A. Meinertzhagen, *Neurons of the ascidian larval nervous system in Ciona intestinalis: I. Central nervous system.* J Comp Neurol, 2007. **501**(3): p. 316-34.

18. Imai, J.H. and I.A. Meinertzhagen, *Neurons of the ascidian larval nervous system in Ciona intestinalis: II. Peripheral nervous system.* J Comp Neurol, 2007. **501**(3): p. 335-52.

19. Crowther, R.J. and J.R. Whittaker, *Serial repetition of cilia pairs along the tail surface of an ascidian larva.* J Exp Zool, 1994. **268**(1): p. 9-16.

20. Pasini, A., et al., *Formation of the ascidian epidermal sensory neurons: insights into the origin of the chordate peripheral nervous system.* PLoS Biol, 2006. **4**(7): p. e225.

21. Rigon, F., et al., *Developmental signature, synaptic connectivity and neurotransmission are conserved between vertebrate hair cells and tunicate coronal cells.* J Comp Neurol, 2018. **526**(6): p. 957-971.

22. Hurless, V.L., *Investigating Ciliated Sensory Neurons and Molecular Tools in the Invertebrate Chordate, Ciona intestinalis*. p. 1 online resource (xviii, 101 pages).

23. Terakubo, H.Q., et al., *Network structure of projections extending from peripheral neurons in the tunic of ascidian larva.* Dev Dyn, 2010. **239**(8): p. 2278-87.

24. Torrence, S.A. and R.A. Cloney, *Nervous-System of Ascidian Larvae - Caudal Primary Sensory Neurons.* Zoomorphology, 1982. **99**(2): p. 103-115.

25. Imai, K.S., et al., *Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks.* Development, 2004. **131**(16): p. 4047-58.

26. Roure, A., P. Lemaire, and S. Darras, *An otx/nodal regulatory signature for posterior neural development in ascidians.* PLoS Genet, 2014. **10**(8): p. e1004548.

27. Waki, K., K.S. Imai, and Y. Satou, *Genetic pathways for differentiation of the peripheral nervous system in ascidians.* Nat Commun, 2015. **6**: p. 8719.

28. Roure, A. and S. Darras, *Msxb is a core component of the genetic circuitry specifying the dorsal and ventral neurogenic midlines in the ascidian embryo.* Dev Biol, 2016. **409**(1): p. 277-287.

29. Roure, A. and S. Darras, *Msxb is a core component of the genetic circuitry specifying the dorsal and ventral neurogenic midlines in the ascidian embryo.* Developmental biology, 2016. **409**(1): p. 277-287.

30. Joyce Tang, W., J.S. Chen, and R.W. Zeller, *Transcriptional regulation of the peripheral nervous system in Ciona intestinalis.* Dev Biol, 2013. **378**(2): p. 183-93.

31. Chen, J.S., et al., *An expanded Notch-Delta model exhibiting long-range patterning and incorporating MicroRNA regulation.* PLoS Comput Biol, 2014. **10**(6): p. e1003655.

32. Chen, J.S., M.S. Pedro, and R.W. Zeller, *miR-124 function during Ciona intestinalis neuronal development includes extensive interaction with the Notch signaling pathway.* Development, 2011. **138**(22): p. 4943-53.

33. Akanuma, T., et al., *Notch signaling is involved in nervous system formation in ascidian embryos.* Development genes and evolution, 2002. **212**(10): p. 459-472.

34. Di Gregorio, A., et al., *Tail morphogenesis in the ascidian, Ciona intestinalis, requires cooperation between notochord and muscle.* Dev Biol, 2002. **244**(2): p. 385-95.

35. Candiani, S., et al., *Ci-POU-IV expression identifies PNS neurons in embryos and larvae of the ascidian Ciona intestinalis.* Dev Genes Evol, 2005. **215**(1): p. 41-5.

36. Mall, M., et al., *Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates.* Nature, 2017. **544**(7649): p. 245-249.

37. Bellefroid, E.J., et al., *X-MyT1, a Xenopus C2HC-type zinc finger protein with a regulatory function in neuronal differentiation.* Cell, 1996. **87**(7): p. 1191-202.

38. Romm, E., et al., *Myt1 family recruits histone deacetylase to regulate neural transcription.* J Neurochem, 2005. **93**(6): p. 1444-53.

39. Satou, Y., et al., *Gene expression profiles in Ciona intestinalis tailbud embryos.* Development, 2001. **128**(15): p. 2893-904.

40. Satou, Y., et al., *An integrated database of the ascidian, Ciona intestinalis: towards functional genomics.* Zoolog Sci, 2005. **22**(8): p. 837-43.

41. Weirauch, M.T., et al., *Determination and inference of eukaryotic transcription factor sequence specificity.* Cell, 2014. **158**(6): p. 1431-1443.

42. Lee, C. and C.H. Huang, *LASAGNA-Search: an integrated web tool for transcription factor binding site search and visualization.* Biotechniques, 2013. **54**(3): p. 141-53.

43. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.

44. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.

45. Fritzsch, B., K.W. Beisel, and N.A. Bermingham, *Developmental evolutionary biology of the vertebrate ear: conserving mechanoelectric transduction and developmental pathways in diverging morphologies.* Neuroreport, 2000. **11**(17): p. R35-44.

46. Buniello, A., et al., *Headbobber: a combined morphogenetic and cochleosaccular mouse model to study 10qter deletions in human deafness.* PLoS One, 2013. **8**(2): p. e56274.

47. Sud, R., et al., *Transcriptional regulation by Barhl1 and Brn-3c in organ-of-Corti-derived cell lines.* Brain Res Mol Brain Res, 2005. **141**(2): p. 174-80.

48. Wang, W., T. Van De Water, and T. Lufkin, *Inner ear and maternal reproductive defects in mice lacking the Hmx3 homeobox gene.* Development, 1998. **125**(4): p. 621-634.

49. Ikuta, T., et al., *Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development.* Proceedings of the national Academy of Sciences, 2004. **101**(42): p. 15118-15123.

50. Liu, W., J.H. Wang, and M. Xiang, *Specific expression of the LIM/Homeodomain protein Lim-1 in horizontal cells during retinogenesis.* Developmental dynamics: an official publication of the American Association of Anatomists, 2000. **217**(3): p. 320-325.

51. Zhao, Y., et al., *LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum.* Proceedings of the National Academy of Sciences, 2007. **104**(32): p. 13182-13186.

52. Tam, P.P., et al., *Regionalization of cell fates and cell movement in the endoderm of the mouse gastrula and the impact of loss of Lhx1 (Lim1) function.* Developmental biology, 2004. **274**(1): p. 171-187.

53. Perea-Gómez, A., et al., *HNF3beta and Lim1 interact in the visceral endoderm to regulate primitive streak formation and anterior-posterior polarity in the mouse embryo.* Development, 1999. **126**(20): p. 4499-4511.

54. Zhang, J., et al., *Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2.* Nature, 1996. **381**(6579): p. 238-241.

55. Bassett, E.A., et al., *AP-2α knockout mice exhibit optic cup patterning defects and failure of optic stalk morphogenesis.* Human molecular genetics, 2010. **19**(9): p. 1791-1804.

56. Imai, K.S., et al., *Tfap2 and Sox1/2/3 cooperatively specify ectodermal fates in ascidian embryos.* Development, 2017. **144**(1): p. 33-37.

57. Lu, C.C., et al., *Developmental profiling of spiral ganglion neurons reveals insights into auditory circuit assembly.* Journal of Neuroscience, 2011. **31**(30): p. 10903-10918.

58. Simeone, A., et al., *Nested expression domains of four homeobox genes in developing rostral brain.* Nature, 1992. **358**(6388): p. 687-690.

59. Hartmann, B., et al., *Expression, regulation and function of the homeobox gene empty spiracles in brain and ventral nerve cord development of Drosophila.* Mechanisms of development, 2000. **90**(2): p. 143-153.

60. Jarvis, L.A., et al., *Sprouty proteins are in vivo targets of Corkscrew/SHP-2 tyrosine phosphatases.* Development, 2006. **133**(6): p. 1133-1142.

61. Zhao, G., et al., *Negative regulation of lens fiber cell differentiation by RTK antagonists Spry and Spred.* Experimental eye research, 2018. **170**: p. 148-159.

62. Shim, K., et al., *Sprouty2, a mouse deafness gene, regulates cell fate decisions in the auditory sensory epithelium by antagonizing FGF signaling.* Developmental cell, 2005. **8**(4): p. 553-564.

63. Wijchers, P.J., et al., *Cloning and analysis of the murine Foxi2 transcription factor.* Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 2005. **1731**(2): p. 133-138.

64. Ogasawara, M., et al., *Gene expression profiles in young adult Ciona intestinalis.* Development genes and evolution, 2002. **212**(4): p. 173-185.

65. Peeters, K., et al., *Molecular defects in the motor adaptor BICD2 cause proximal spinal muscular atrophy with autosomal-dominant inheritance.* The American Journal of Human Genetics, 2013. **92**(6): p. 955-964.

66. Saffman, E.E., et al., *Premature translation of oskar in oocytes lacking the RNA-binding protein bicaudal-C.* Molecular and cellular biology, 1998. **18**(8): p. 4855-4862.

67. Topczewska, J.M., et al., *Sequence and expression of zebrafish foxc1a and foxc1b, encoding conserved forkhead/winged helix transcription factors.* Mechanisms of development, 2001. **100**(2): p. 343-347.

68. Berry, F.B., et al., *Functional interactions between FOXC1 and PITX2 underlie the sensitivity to FOXC1 gene dose in Axenfeld–Rieger syndrome and anterior segment dysgenesis.* Human molecular genetics, 2006. **15**(6): p. 905-919.

69. Hadrys, T., et al., *Nkx5-1 controls semicircular canal formation in the mouse inner ear.* Development, 1998. **125**(1): p. 33-39.

70. Thaler, J.P., et al., *LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions.* Cell, 2002. **110**(2): p. 237-249.

71. Hertzano, R., et al., *Lhx3, a LIM domain transcription factor, is regulated by Pou4f3 in the auditory but not in the vestibular system.* European Journal of Neuroscience, 2007. **25**(4): p. 999-1005.

72. Merlo, G.R., et al., *Multiple functions of Dlx genes.* International Journal of Developmental Biology, 2004. **44**(6): p. 619-626.

73. Panganiban, G., *Distal-less function during Drosophila appendage and sense organ development.* Developmental dynamics: an official publication of the American Association of Anatomists, 2000. **218**(4): p. 554-562.

74. Mansouri, A., *The role of Pax3 and Pax7 in development and cancer.* Critical Reviews™ in Oncogenesis, 1998. **9**(2).

75. Mazet, F., et al., *Pax gene expression in the developing central nervous system of Ciona intestinalis.* Gene Expression Patterns, 2003. **3**(6): p. 743-745.

76. Hudson, C., S. Lotito, and H. Yasuo, *Sequential and combinatorial inputs from Nodal, Delta2/Notch and FGF/MEK/ERK signalling pathways establish a grid-like organisation of distinct cell identities in the ascidian neural plate.* Development, 2007. **134**(19): p. 3527-3537.

77. Imai, K.S., et al., *Regulatory blueprint for a chordate embryo.* Science, 2006. **312**(5777): p. 1183-7.

78. Stolfi, A., et al., *Migratory neuronal progenitors arise from the neural plate borders in tunicates.* Nature, 2015. **527**(7578): p. 371-4.

79. Xiang, M., et al., *Role of the Brn-3 family of POU-domain genes in the development of the auditory/vestibular, somatosensory, and visual systems.* Cold Spring Harb Symp Quant Biol, 1997. **62**: p. 325-36.

80. Xiang, M., et al., *The Brn-3 family of POU-domain factors: primary structure, binding specificity, and expression in subsets of retinal ganglion cells and somatosensory neurons.* J Neurosci, 1995. **15**(7 Pt 1): p. 4762-85.

81. Certel, S.J., et al., *Regulation of central neuron synaptic targeting by the Drosophila POU protein, Acj6.* Development, 2000. **127**(11): p. 2395-405.

82. Finney, M., G. Ruvkun, and H.R. Horvitz, *The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors.* Cell, 1988. **55**(5): p. 757-69.

83. Murata, J., K. Ikeda, and H. Okano, *Notch signaling and the developing inner ear*, in *Notch Signaling in Embryology and Cancer*. 2012, Springer. p. 161-173.

84. Lewis, J. *Notch signalling and the control of cell fate choices in vertebrates*. in *Seminars in cell & developmental biology*. 1998. Elsevier.

85. Imai, K.S., et al., *Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system.* Development, 2009. **136**(2): p. 285-93.

86. Ellis, H.M., D.R. Spann, and J.W. Posakony, *extramacrochaetae, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins.* Cell, 1990. **61**(1): p. 27-38.

87. Ellmeier, W. and A. Weith, *Expression of the helix-loop-helix gene Id3 during murine embryonic development.* Developmental Dynamics, 1995. **203**(2): p. 163-173.

88. Fisher, A.L., S. Ohsako, and M. Caudy, *The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain.* Molecular and cellular biology, 1996. **16**(6): p. 2670-2677.

89. Kubo, A., et al., *Genomic cis-regulatory networks in the early Ciona intestinalis embryo.* Development, 2010. **137**(10): p. 1613-23.

90. Mahen, R., et al., *Comparative assessment of fluorescent transgene methods for quantitative imaging in human cells.* Mol Biol Cell, 2014. **25**(22): p. 3610-8.

91. Doudna, J. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Sci. 346, 1258096*. 2014.

92. Stolfi, A., et al., *Tissue-specific genome editing in Ciona embryos by CRISPR/Cas9.* Development, 2014. **141**(21): p. 4115-4120.

93. Daudet, N. and J. Lewis, *Two contrasting roles for Notch activity in chick inner ear development: specification of prosensory patches and lateral inhibition of hair-cell differentiation.* Development, 2005. **132**(3): p. 541-51.

94. Neves, J., et al., *Patterning and cell fate in the inner ear: a case for Notch in the chicken embryo.* Dev Growth Differ, 2013. **55**(1): p. 96-112.

95. Chen, J.S., *Role of the microRNA miR-124 in the regulatory network governingPNS development in Ciona intestinalis*. 2013, The Claremont Graduate University.

96. Heldin, C.H. and A. Moustakas, *Role of Smads in TGFbeta signaling.* Cell Tissue Res, 2012. **347**(1): p. 21-36.

97. Lebrun, J.J., et al., *Roles of pathway-specific and inhibitory Smads in activin receptor signaling.* Mol Endocrinol, 1999. **13**(1): p. 15-23.

98. Horie, R., et al., *Shared evolutionary origin of vertebrate neural crest and cranial placodes.* Nature, 2018. **560**(7717): p. 228-232.

99. Hurless, V.L., *Investigating Ciliated Sensory Neurons and Molecular Tools in the Invertebrate Chordate, Ciona intestinalis*. 2015, UC San Diego.

100. Lee, J., et al., *A Myt1 family transcription factor defines neuronal fate by repressing non-neuronal genes.* Elife, 2019. **8**: p. e46703.

101. Dawson, S.J., P.J. Morris, and D.S. Latchman, *A single amino acid change converts an inhibitory transcription factor into an activator.* J Biol Chem, 1996. **271**(20): p. 11631-3.

102. Hertzano, R., et al., *Transcription profiling of inner ears from Pou4f3 ddl/ddl identifies Gfi1 as a target of the Pou4f3 deafness gene.* Human molecular genetics, 2004. **13**(18): p. 2143-2153.

103. Scheffer, D.I., et al., *Gene expression by mouse inner ear hair cells during development.* Journal of Neuroscience, 2015. **35**(16): p. 6366-6380.

104. Boareto, M., D. Iber, and V. Taylor, *Differential interactions between Notch and ID factors control neurogenesis by modulating Hes factor autoregulation.* Development, 2017. **144**(19): p. 3465-3474.

105. NISHIDA, H., *Cell Division Pattern during Gastrulation of the Ascidian, Halocynthia roretzi: (cell division pattern/gastrulation/neurulation/ascidian embryo).* Development, growth & differentiation, 1986. **28**(2): p. 191-201.

106. Ogura, Y., et al., *Coordination of mitosis and morphogenesis: role of a prolonged G2 phase during chordate neurulation.* Development, 2011. **138**(3): p. 577-587.

107. Ogura, Y. and Y. Sasakura, *Developmental control of cell-cycle compensation provides a switch for patterned mitosis at the onset of chordate neurulation.* Developmental Cell, 2016. **37**(2): p. 148-161.

108. Yasugi, T., et al., *Drosophila optic lobe neuroblasts triggered by a wave of proneural gene expression that is negatively regulated by JAK/STAT.* Development, 2008. **135**(8): p. 1471-1480.

109. Bodmer, R., R. Carretto, and Y.N. Jan, *Neurogenesis of the peripheral nervous system in Drosophila embryos: DNA replication patterns and cell lineages.* Neuron, 1989. **3**(1): p. 21-32.

110. Maurer, K.A., A.N. Riesenberg, and N.L. Brown, *Notch signaling differentially regulates Atoh7 and Neurog2 in the distal mouse retina.* Development, 2014. **141**(16): p. 3243-3254.

111. Kay, J.N., B.A. Link, and H. Baier, *Staggered cell-intrinsic timing of ath5 expression underlies the wave of ganglion cell neurogenesis in the zebrafish retina.* Development, 2005. **132**(11): p. 2573-2585.

112. Quan, X.-J. and B. Hassan, *From skin to nerve: flies, vertebrates and the first helix.* Cellular and Molecular Life Sciences CMLS, 2005. **62**(18): p. 2036-2049.