

The Genetics of Cranial Sensory Ganglia Development and Evolution

Thesis presented for the degree of Doctor of Philosophy in Zoology

Vasileios Papadogiannis
Wolfson College
Hillary 2020

The Genetics of Cranial Sensory Ganglia Development and Evolution

Thesis presented for the degree of Doctor of Philosophy in Zoology

Vasileios Papadogiannis
Wolfson College
Hillary 2020

Abstract

Vertebrates perceive the world through sensory neurons of their cranial nerves that connect primary sensory receptors to the brain. The bodies of these neurons are organized in cranial sensory ganglia (CSG) in the head. CSG are considered a vertebrate novelty and they develop from neurogenic placodes and neural crest. The genetic network underlying their early development has been relatively well studied and placodal cell homologs have been suggested in other chordates. However, knowledge has been lacking on the evolution of different ganglia and their neurons. Undertaking a comparative approach, cyclostomes and tunicates are used here to look into the molecular identity and evolution of CSG sensory cells. First, conserved pan-vertebrate markers of the ganglia are identified by exploring the expression and specificity of candidate genes in lamprey embryos. Two of these marker gene families are the starting ground for a comparative investigation of sensory cell types between vertebrates and other chordates. Conserved expression and regulation of the estrogen related receptor (*ERR*) between lamprey and gnathostomes is presented. It is revealed that a single *ERR* homolog acquired a role in the development of the vestibuloacoustic ganglion in early vertebrates. Invertebrate sensory expression highlights *ERR* as an ancient sensory cell marker in bilaterian animals, suggesting a mode of evolution at cell type level. The *Hmx* homeobox gene family is then identified to ubiquitously mark all CSG and possess conserved genomic architecture across vertebrates. A unique enhancer pair is reported, derived by tandem duplication, connected to ancestral regulation of expression in the central and peripheral nervous systems. *Hmx* expression and function in *Ciona* embryos, combined with vertebrate data, provide evidence of homology between CSG and tunicate sensory neurons. Finally, a review is carried out, discussing how placodal and CSG components gradually evolved and assembled under new genomic organization and control in vertebrates.

Dedication

To them, who kept wishing on falling stars for me, on our street of dreams:

For my mother, Alexandra, who taught me to only do what I love and love what I do and passed on to me her beautiful smile and curiosity.

You could only take me half the way, but that was the most important part of the journey.

For my grandmother, Maria, who taught me to never give up and who was the bravest, strongest and most caring person I know.

You made sure I kept smiling until the end of the road… and beyond.

Ευχαριστώ για όλα

Acknowledgements

This thesis is the destination of a big adventure, the final product of a long line of work spanning over four years. I truly consider myself very lucky to have walked this path with some incredible people, collaborators and friends. There is not the slightest exaggeration in saying that I could not have reached here without their help, support and company.

Most importantly, I want to express my gratitude to my supervisor, **Professor Sebastian Shimeld**, for showing endless patience and believing in me, time and time again. He continuously encouraged me by saying the “yes”s that allowed me to explore my unorthodox ideas, while being wisely vigilant to give me the “no”s I did not want to hear, but that were necessary nonetheless. As the project he offered me in October 2015 now concludes, I sincerely hope my work is worthy of his support, opening up more doors than the one I close behind.

I am grateful to my college (Wolfson), my department (Zoology) and NaturalMotion for my scholarship, which allowed me to pursue this project and live these unforgettable past years in Oxford. In addition, I want to thank the Hannah Elizabeth Jenkinson Fund for funding one of the most important parts of my project and I am also thankful to the Genetics Society for funding my attendance to the 6th and 7th meeting of the European Society for Evolutionary Developmental Biology.

I am also sincerely grateful to our collaborators, whose willingness to help and contribution brought down the barriers of troubling times and empowered my progress. I want to thank **Professor Andrea Streit** and the members of her group **Mark Hintze** and **Mohi Ahmed** for welcoming me in Guy’s Hospital London, patiently training me and allowing me to use their facilities for in-ovo electroporations in chick. I want to thank **Professor Marianne Bronner** in Caltech for offering the facilities needed and **Dorit Hockman** and **Hugo Parker** for eagerly carrying out sea lamprey embryo microinjections. Finally, I want to thank **Professor**

Tatjana Sauka-Spengler for offering her lab at WIMM in Oxford and her student **Martyna Lukoševičiūtė** for eagerly doing zebrafish microinjections for me.

I also want to thank all the dear friends and colleagues with whom I had the joy and honour to share my time in Zoology. **Sonia Trigueros**, whose knowledge on everything and joyful spirit lit up our days, the amazing people of the Shimeld group, **Gui, Harrie, Matthew, Brigid** and **Claire,** who I was very lucky to have supervised, **Professor Peter Holland**, whose door was always open when I asked for help, and the incredible people in his group **Steve, Amy, Serena** and **Roddy**.

I also want to give special thanks to some friends and colleagues whose help in and out of the lab was invaluable. I want to thank my lab buddy **Gui**, with whom I shared this journey from the very first day. He was there when we drove and collected animals singing songs on the way, when I failed and needed someone to say “next time man”, when I needed small or bigger help in the lab; I am thankful to him, for not once can I remember him saying no. I also want to thank **Silvia** for teaching me a fraction of her impressive *Ciona* skills and secrets. She may have only joined our lab for an “awfully” short time, but she quickly became a long-time friend. Finally, a big thank you to **Adam** for sharing endless science conservations, amazing food and music and his incredible bioinformatics skills and guidance. He made sure that, no matter what, I always remember to “blend the goblet” and do “cookie earning things”.

I must also say that Oxford has been much more than my research and I cannot thank enough the beautiful people out of the lab who shared this time with me, transforming it into something magical. This includes so many friends from Wolfson and Oxford that cannot all be mentioned. However, I must say a special thank you to **Marilena**, **Maria, Sharat**, **Galbha**, **Aayush** and **Oana** for becoming our multicultural Greek-Balkan-Indian family in Oxford, crafting wonderful memories together, sharing laughter and tears and welcoming the three beautiful babies in our lives, Anaaya, Ahrian and Saske. Finally, I want to thank my brothers back home, **Nikos, Kostas, Nikos and Panayotis** who have stood by me and supported me greatly all these years and especially in the tough times of my writing year.

The biggest thank you must be left for my family, whose unwavering support and love have followed me everywhere I go. I want to deeply thank my dad **Thanasis**, a fighter in life, who taught me that combining love, hard work and patience for what you do will always lead to something great. I cannot thank enough my dad's bother, uncle **Giorgos**, an incredibly bright man, who supported us immeasurably in every way. A big thank you to the rest of our family, who have all been next to us in our difficult times of need and are still here to celebrate the great days ahead. Finally, I need to offer my endless love and gratitude to my mother and grandmother, who are the pillars of everything I ever built and to whom this work is dedicated.

I could not finish without thanking my partner in life and science alike, my closest friend and the caring eyes that always keep watch over me, **Despoina**. Thank you for walking this decade long road together. These years in Oxford have been a life changing experience, filled with countless memories and beloved friends who became family and you stand in the centre of it all. Thank you for actively supporting me in big and small decisions and pushing me ever upwards. Your care nourished me through difficult long days, your company guided me through my darkest hours, your love became the light that always shows me where home is.

Abbreviations

AP	Anteroposterior
AS	Atrial Siphon
ASP	Atrial Siphon Primordium
BTN	Bipolar Tail Neurons
CNE	Conserved Non-coding Element
CNS	Central Nervous System
CSG	Cranial Sensory Ganglia
DRG	Dorsal Root Ganglia
DV	Dorsoventral
ESN	Epidermal Sensory Neurons
Fb	forebrain
G	Geniculate
Hb	Hindbrain
Mb	Midbrain
N	Nodose
NC	Neural Crest
NP	Neural Plate
Ol	Olfactory
OS	Oral Siphon
OSP	Oral Siphon Primordium
OV	Otic Vesicle
P	Petrosal
PNS	Peripheral Nervous System
PPR	Pre-Placodal Region
SC	Spinal Cord
Tg	Trigeminal
Tmm	Trigeminal Maxillomandibular
Top	Trigeminal Ophthalmic
VA	Vestibuloacoustic

Contents

CONTENTS 7

CHAPTER 1 GENERAL INTRODUCTION 13

1.1 PROLOGUE – AIMS AND SCOPE OF THIS STUDY 13

1.2 THE VERTEBRATE NERVOUS SYSTEM 14

1.2.1 SIGNALLING NETWORKS AND CNS PATTERNING 15

1.2.2 VERTEBRATE CNS STRUCTURE AND COMPARTMENTALISATION 17

1.3 VERTEBRATE SENSORY NERVES AND GANGLIA 19

1.3.1 PLACODE INDUCTION AND DEVELOPMENT 22

1.3.2 PLACODE IDENTITY AND SPECIFICATION 24

1.3.3 CRANIAL GANGLIA AND THEIR MARKERS 26

1.4 BASALLY DIVERGING VERTEBRATES: THE CYCLOSTOMES 29

1.4.1 THE LAMPREY CNS 32

1.4.2 LAMPREY CRANIAL PLACODES AND SENSORY NERVES 32

1.5 CHORDATE RELATIVES OF VERTEBRATES – CSG AND PLACODE ORIGINS 36

1.5.1 THE ORIGIN OF ECTODERMAL TERRITORIES AND PLACODAL PROGENITORS 37

1.5.2 PLACODE HOMOLOGS IN TUNICATES AND AMPHIOXUS 40

1.6 GENOMIC SEQUENCE EVOLUTION AND DEVELOPMENTAL INNOVATIONS 43

1.6.1 GENE COPY CHANGES AND WHOLE GENOME DUPLICATIONS 44

1.6.2 NON-CODING SEQUENCE CONSERVATION AND CONSERVED REGULATION 47

1.7 THESIS LAYOUT 48

CHAPTER 2 MATERIALS AND METHODS 49

2.1 GENE AND CNE IDENTIFICATION AND SEQUENCE ANALYSES 49

2.1.1 GENE IDENTIFICATION AND SEQUENCE SOURCES 49

2.1.2 ALIGNMENTS 50

2.1.3 SYNTENY ANALYSIS 51

2.1.4 CNE IDENTIFICATION 51

2.2 GENE FRAGMENT AND CNE CLONING 52

2.2.1 LAMPREY SEQUENCE CLONING 52

2.2.2 INVERTEBRATE SEQUENCE CLONING 53

2.3 EMBRYOS AND IN SITU HYBRIDISATION 54

2.3.1 EMBRYOS 54

2.3.2 IN SITU HYBRIDISATION 55

2.4 TRANSGENICS, IMAGING AND CONTROLS 56

2.4.1 VERTEBRATE TRANSGENICS 56

2.4.2 IMMUNOHISTOCHEMISTRY 57

2.4.3 IMAGE ANALYSIS 58

2.4.4 KRT4 BACKGROUND 58

2.5 *CIONA* TRANSGENICS AND BULK EMBRYO RNASEQ 61

2.5.1 *CIONA HMX* ORF CLONING 61

2.5.2 *CIONA* TRANSGENICS AND RNASEQ 61

2.5.3 RNASEQ DATA PROCESSING AND DIFFERENTIAL EXPRESSION ANALYSIS 62

2.5.4 *CIONA* CNE TRANSGENICS (CARRIED OUT BY ALESSANDRO PENNATI) 63

CHAPTER 3 THE MOLECULAR IDENTITY OF CRANIAL SENSORY GANGLIA 64

3.1 CHAPTER INTRODUCTION 64

3.1.1 EXPERIMENTAL APPROACH 66

3.1.2 AIMS AND MARKER SELECTION....67
3.1.3 WORK SUMMARY....73
3.2 RESULTS....74
3.2.1 GENES WITHOUT IDENTIFIED CSG EXPRESSION....76
3.2.2 *PRDM12* EXPRESSION IN TRIGEMINAL AND GENICULATE....80
3.2.3 LAMPREY *ERR* MARKS THE VA....82
3.2.4 *SHOX* EXPRESSION IN EPIBRANCHIAL AND TRIGEMINAL GANGLIA....86
3.2.5 LAMPREY *HMX* PARALOGS MARK ALL CSG....88
3.3 DISCUSSION....92
3.3.1 THE MOLECULAR SIGNATURE OF LAMPREY CSG....92
3.3.2 CSG MARKER CONSERVATION AND LACK OF SPECIFICITY....95

CHAPTER 4 EXPRESSION AND REGULATION OF ANCIENT SENSORY MARKER *ERR* IN VERTEBRATES....98

4.1 CHAPTER INTRODUCTION....98
4.1.1 THE ERR FAMILY: EVOLUTION AND ROLES IN VERTEBRATE DEVELOPMENT....99
4.1.2 INVESTIGATING ANCESTRAL EXPRESSION AND REGULATION....100
4.1.3 CHAPTER SUMMARY....101
4.2 RESULTS....102
4.2.1 LAMPREY *ERR* EXPRESSION IN THE VA AND CNS....102
4.2.2 RELATIONSHIPS AND EVOLUTION OF *ERR* LOCI....104
4.2.3 *ERRA* AND *ERRΓ* CONSERVED SYNTENY AND NON-CODING ELEMENTS....107
4.2.4 CROSS-SPECIES ACTIVITY OF *ERRA/ERRΓ* CNE IN CNS AND PNS....111
4.2.5 THE PRE-VERTEBRATE ANCESTRY OF *ERR* EXPRESSION....117
4.3 DISCUSSION....120
4.3.1 RELATIONSHIPS OF VERTEBRATE *ERR* LOCI....120
4.3.2 COMMON ORIGIN OF VA EXPRESSION AND REGULATION OF *ERRA* AND *ERRΓ*....121

4.3.3 *ERR* EXPRESSION IN SENSORY CELLS PREDATES VERTEBRATES.....123

4.3.4 CONCLUSIONS.....127

CHAPTER 5 *HMX* CONSERVATION IDENTIFIES A SHARED ORIGIN FOR VERTEBRATE AND TUNICATE SENSORY NEURONS.....130

5.1 CHAPTER INTRODUCTION.....130

5.1.1 *HMX* CSG EXPRESSION AND GENOMIC ORGANISATION IN VERTEBRATES.....132

5.1.2 CHAPTER SUMMARY.....133

5.2 RESULTS.....134

5.2.1 HMX GENE EXPRESSION MARKS ALL VERTEBRATE CRANIAL SENSORY GANGLIA.....134

5.2.2 CLUSTERED ORGANISATION AND NON-CODING CONSERVATION OF VERTEBRATE *HMX* LOCI.....137

5.2.3 *HMX* CNE ARE ENHANCERS THAT REGULATE EXPRESSION IN CNS AND CSG.....143

5.2.4 *HMX* EXPRESSION IN INVERTEBRATE CHORDATES.....146

5.2.5 *HMX* DRIVES SENSORY NEURON IDENTITY IN TUNICATES.....147

5.3 DISCUSSION.....151

5.3.1 UNIQUE CONSERVATION AND LINKAGE IN VERTEBRATE *HMX* LOCI.....151

5.3.2 *HMX* EXPRESSION, REGULATION AND FUNCTION TRACE THE SHARED ORIGIN OF CSG AND BTN.....153

5.3.3 CONCLUSIONS AND FUTURE PROSPECTS.....154

CHAPTER 6 GENERAL DISCUSSION.....157

6.1 CSG MARKERS HIGHLIGHT DIFFERENT LEVELS OF HOMOLOGY AND CO-OPTION.....157

6.1.1 GENES MARKING SPECIFIC SENSORY FATES PROVIDE SUPPORT FOR CELL LEVEL HOMOLOGY.....158

6.1.1(A) ANCESTRAL ROLE OF *ERR* IN HEARING AND VISION.....159

6.1.1(B) *PRDM12* AS A SOMATOSENSORY MARKER.....160

6.1.2 PLACODAL AND CSG MARKERS WITH BROAD EXPRESSION 161

6.2 NOVEL SENSORY ORGANISATION BUILT BY AN EXPANDED GENOMIC ARSENAL.....163

6.2.1 A GENETIC BLUEPRINT UNDERLYING CSG DEVELOPMENT FIXED IN STEM-VERTEBRATES..163

6.2.2 A STEP-WISE TRANSITION FROM SENSORY CELL TYPES TO CSG 165

6.3 THE ANCESTRY OF BTN AND CSG NEURONS.....167

6.3.1 THE BTN/CSG SPECIFICATION NETWORK IS SHARED BETWEEN VERTEBRATES AND TUNICATES.....168

6.3.2 BTN AND PLACODE RELATED COMPONENTS ARE UNCOUPLED IN TUNICATES.....172

6.4 CONCLUSIONS AND FUTURE PROSPECTS 174

REFERENCES.....177

Chapter 1 General Introduction

Work Declaration

The literature review presented in this chapter, confocal data collection and immunostaining of embryos presented in fig1.1 were carried out by myself (Vasileios Papadogiannis).

1.1 Prologue – Aims and scope of this study

This research explores the molecular identity of developing cranial sensory ganglia (CSG), in an evolutionary context. Before reviewing the necessary background and framework behind this research, it would be useful to clarify and narrow down the scope and boundaries of this exploration.

CSG develop through a series of induction, specification and morphogenetic steps from ectodermal placodes and neural crest. This work looks into the genetics of ganglia, the molecular identity of which has been less clear, rather than earlier/placodal phases, which are better studied and described. Molecular identity hereby refers to a combination of genes with specific expression and role in differentiating CSG neurons. Drawing from literature on jawed vertebrates, transcription factors with demonstrated function in sensory nerve/structure development were primarily selected.

The overarching theme of this research is the study of the evolutionary origins of CSG development, based on comparative genetics. Using lampreys and ascidians as outgroups, a stepwise search unfolds, looking for conserved expression, regulation and function of candidate CSG markers. The main focus here is tracing cell type evolutionary relationships and revealing ancestral expression patterns connected to sensory neuron specification. Therefore, analysing the mechanistic details of CSG specification at cell or gene level and providing a comprehensive regulatory network of CSG development are both out of the scope of this introduction.

In this chapter, a review is carried out of the literature describing CSG and nervous system development, the organisms used for this comparative study and their relationships and our current understanding of vertebrate gene and genome evolution. Focus is put on the genetic programmes behind the patterning and specification of the nervous system that underlie CSG development, as well as on CSG ontogeny itself.

1.2 The vertebrate nervous system

The development of the vertebrate central nervous system (CNS) has been extensively studied and its development well described in gnathostomes. The various morphogenetic movements that take place are driven by the interplay of several regulatory networks with dynamic changes of gene expression. As in other bilaterians, different patterning mechanisms along the anteroposterior (AP) and dorsoventral (DV) axes have a pivotal role in this process. Peripheral nervous

system (PNS) and CNS development is closely tied, sharing many similarities, while the former is also dependent on signals from the latter. Thus, at this point, it is useful to summarise the main events and networks underlying vertebrate CNS development.

1.2.1 Signalling networks and CNS patterning

The establishment of signalling centres and expression patterns is pivotal to the stepwise process that is neural development. Many molecules and signalling pathways have important roles in both early and later development, as well as in both CNS and PNS development. The BMP, Wnt and FGF pathways are some of the earliest components in this process. The interplay between the above, as well as the RA, Shh, Notch and Nodal signalling pathways orchestrate cell fate decision events at various stages (Sanes et al., 2006; Squire et al., 2012).

Neural induction starts early in development. Early ectoderm is primed to differentiate to a neural fate. However, BMP signals supress this path, promoting an epidermal fate instead. In the gastrula, this anti-neuralising BMP action is blocked by BMP antagonists (e.g. originating from Spemman's organizer in amphibians), such as noggin, chordin and follistatin. In contrast, FGFs and Wnts have been shown to help create the early neural potential of the ectoderm, partially by inhibiting BMP signal transducers (Squire et al., 2012). Other genes with very early expression that also promote a neural fate include *Zic* and *Sox* (Sanes et al., 2006). This process drives the induction of the neural plate dorsally and above the developing notochord (Sanes et al., 2006; Squire et al., 2012).

As bilaterians, vertebrate embryos are polarised. The anterior of the embryo gives rise to a number of distinctive vertebrate head characters, including the cranium, brain, as well as the cranial sensory organs and nerves. Hence, the establishment of the AP and DV axes is another crucial event in nervous system development. At late gastrulation, Wnt signals have a central posteriorizing role. The anterior endoderm expresses Wnt signalling inhibitors, such as Cerberus and DKK1. These anti-Wnt signals help promote head and neural tube formation by creating a Wnt gradient, with higher Wnt concentrations at the posterior and lower at the anterior (Ciani and Salinas, 2005; Squire et al., 2012). Accordingly, a BMP gradient with high lateral and low medial concentrations drives early mediolateral axis formation (Liu and Niswander, 2005; Squire et al., 2012).

As discussed above, BMPs regulate early neural induction and promote non-neural over neural ectodermal fate. Noggin and other BMP inhibitors are also expressed at the lateral edges of the neural plate, the neural folds, restricting BMPs laterally. This pattern appears sufficient and necessary for inducing neurulation and neural tube closure (Liu and Niswander, 2005; Ybot-Gonzalez et al., 2007). Subsequently, BMP expression dorsally and Shh expression ventrally control dorsal and ventral neural tube fates respectively, by restricting the domain of each other's influence (Liu and Niswander, 2005). In addition, RA signalling from the somites and FGFs from the posterior are also important for the development of the extending neural tube (Maden, 2007). Finally, Wnt signalling has important function in regulating dorsal spinal cord neural progenitor proliferation (Ciani and Salinas, 2005).

Guided by the gradients and patterns set up by early signalling pathways, the newly formed neuroepithelium starts to express a suite of transcription factors that will define the identity of the various CNS segments along the AP axis.

1.2.2 Vertebrate CNS structure and compartmentalisation

Chordates, including vertebrates, have a dorsal nerve cord, as opposed to the ventral cord of protostomes. Hemichordates, on the other hand, have a ventral and a dorsal cord, complicating the inference of the ancestral deuterostome state. While the exact positioning of the nerve cord in ancestral bilaterians is still debated, a single origin of bilaterian nervous systems has been suggested (Holland et al., 2013). This is also supported by the conserved expression of many patterning genes between protostomes and deuterostomes (Arendt et al., 2016b; Holland et al., 2013). The tripartite brain of vertebrates is characterised by its conserved compartmentalisation. The vertebrate brain is separated in three major compartments, positioned serially, anterior to posterior: the forebrain, the midbrain and the hindbrain. The spinal cord then follows posteriorly from the hindbrain (Kiecker and Lumsden, 2005; Squire et al., 2012).

The hindbrain is characteristically segmented, in 7 or 8 segments named rhombomeres (r1-r8), separated by sharp boundaries of reduced cell proliferation (Eickholt et al., 2001; Kiecker and Lumsden, 2005). In general, segments of the brain that give rise to particular structures later in development have been called neuromeres. Based on molecular markers, the forebrain (procencephalon) and midbrain (mesencephalon) have been suggested to be divided in prosomeres and

mesomeres respectively. In forebrain, these neuromeres would give rise to the telencephalon and diencephalon (including the thalamus, prothalamus, prethalamus and epithalamus) and in midbrain, the areas of the tectum, tegmentum and isthmus (or midbrain hindbrain boundary - MHB) arise (Kiecker and Lumsden, 2005; Sanes et al., 2011; Squire et al., 2012). Hindbrain rhombomeres have well defined limits and neuronal populations occupy distinct territories. In contrast, between forebrain and midbrain subdivision, cells can generate descendants that occupy more than one territory. For this reason, the neuromeric view of these compartments has been debated (Kiecker and Lumsden, 2005).

CNS zones are defined by more or less rigid expression boundaries of many transcription factors, particularly homeobox genes, organised by various signalling centres. In the hindbrain, expression of *Hox* and *Krox* patterns different rhombomeres, conferring specific molecular identity to the neurons of each segment (Keynes and Lumsden, 1990; Sanes et al., 2011). These transcription factors act as selectors that define the nature of these neuronal populations (Kiecker and Lumsden, 2005). While *Hox* expression does not extend anterior to r2, other genes pattern the midbrain and forebrain as well. *Otx* marks both forebrain and midbrain, while *Emx* is specific to forebrain. The zona limitans intrathalamica (ZLI) develops via lineage restriction as a *WNT8B* positive and *LFNG* negative compartment to become a restricted *Shh* signalling zone that acts as a boundary between the thalamus and prethalamus in the forebrain. Posterior to the ZLI, *IRX3* is restricted to the thalamus. In contrast to the ZLI, the MHB is a very sharp boundary separating the midbrain from the hindbrain, marked by a sharp line of *FGF8* expression. *IRX3* expression that starts at the ZLI ends at the MHB, while

Otx expression also ends here. Posterior to the MHB, *Gbx* and *IRX2* mark R1 in the hindbrain, followed by the expression of various *Hox* genes that serially pattern the rest of the hindbrain (Kiecker and Lumsden, 2005; Squire et al., 2012).

The expression of homeobox transcription factors along the vertebrate CNS AP axis has been shown to have deep evolutionary conservation. More examples include *Rx, Pax6* (also in particular rhombomeres) and *Six3* expressed in the forebrain and *Pax2/5/8,* expressed in midbrain and hindbrain (Holland et al., 2013). Similar to AP patterning, many of these homeobox genes and others, such as *NKX, Pax* and *Irx* genes, are involved in patterning the CNS along the DV axis, with comparable roles in both brain and spinal cord. The expression of these DV markers is largely organised by Shh signals from the flour plate and BMP signalling from the dorsal neural tube, as discussed earlier (Lupo et al., 2006; Squire et al., 2012).

1.3 Vertebrate sensory nerves and ganglia

The vertebrate head houses various sensory organs. Their sensory cells and neurons arise from two different, but interacting embryonic populations, the neurogenic placodes and neural crest. The evolution of this sensory arrangement in the vertebrate head has been suggested to have been crucial for the passage to an active predatory lifestyle in ancestral vertebrates (Gans and Northcutt, 1983; Northcutt and Gans, 1983). The sensory nerves of the head and trunk relay information from the various sensory systems to the CNS. The bodies of sensory neurons of the PNS are clustered in ganglia in the head (CSG) and along the spinal cord (dorsal root ganglia – DRG) (Squire et al., 2012). During development, the

source of DRG is solely neural crest. In contrast, neurogenic placodes and neural crest contribute cells to the developing CSG (Le Douarin et al., 1986). Neural crest cells arise from the edges of the closing neural tube, from the neural plate border. They are pluripotent cells that migrate and then differentiate to very different cell types in both trunk and head (Steventon et al., 2014). In contrast, ectodermal placodes appear as thickenings of the ectoderm after neural tube closure and are neurogenic, with the exception of the adenohypophyseal and lens placodes (Schlosser, 2006). Neural crest and neuronal progenitors of placode origin interact throughout CSG development. In the absence of neural crest, CSG still develop from placode derived cells, but they are displaced and projections form abnormally or fail to connect at all (Begbie et al., 1999). Neural crest cells act to organise placodal neurons via a chase and run mechanism early on. This process is reciprocally important for proper neural crest migration as well (Steventon et al., 2014). Neural crest cells have also been shown to form corridors that restrict placode neuronal progenitor migration and subsequent axonal process formation (Freter et al., 2013). This project focuses on CSG neuronal populations arising from placodes.

The paired sensory ganglia and placodes of the head are highly conserved in all vertebrates, conveying different senses from the periphery to the CNS. The olfactory placode produces chemosensory neurons that mediate olfaction. The trigeminal placodes (comprising the profundal/ophthalmic and maxillomandibular) generate the somatosensory neurons of the trigeminal ganglia that sense temperature, touch and pain in the head. The otic placode produces both the mechanosensory hair cells of the ear and the sensory neurons of the

vestibuloacoustic ganglion, which transfer hearing and balance information. Comparably, lateral line placodes produce the hair cells and neurons of the lateral line ganglia that detect water vibrations or electric fields in aquatic vertebrates. Epibranchial placodes give rise to the visceronsensory neurons of the geniculate (also contains somatosensory neurons that innervate the pinna), petrosal and nodose ganglia, which innervate taste buds and chemosensory neurons along the digestive tract and lungs (Northcutt and Gans, 1983; Patthey et al., 2014). Finally, a small paratympanic placode that generates paratympanic organ hair cells and neurons has been described in amniotes and hypobranchial placodes ventral to the branchial basket have also been identified in amphibians (O'Neill et al., 2012; Schlosser and Northcutt, 2000).

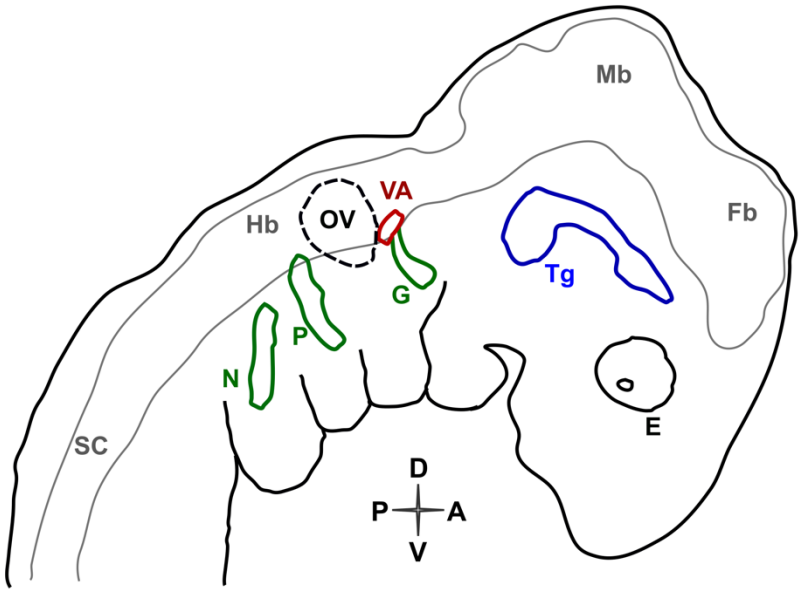


Figure 1.1 The cranial sensory ganglia. CSG in HH17 (Hamburger and Hamilton, 1951) chicken embryo: trigeminal(Tg) in blue, vestibuloacoustic (VA) in red, epibranchial ganglia -geniculate(G), petrosal(P), nodose(N)- in green. Additional abbreviations: E, Eye. OV, Otic Vesicle. Fb, Forebrain. Mb, Midbrain. Hb, Hindbrain. SC, Spinal Cord.

1.3.1 Placode induction and development

After neurulation, placodes and neural crest occupy distinct territories and differ in their migration and differentiation. However, early in their development, both populations originate from adjacent regions in the gastrula. Neural crest arises from the region that is immediately proximal to the neural plate. In turn, all placodes are understood to derive from the pan-placodal primordium or pre-placodal region (PPR), a horse-shoe shaped region that is adjacent to the neural crest territory (Grocott et al., 2012; McCabe and Bronner-Fraser, 2009; Pieper et al., 2011). How these territories develop is less understood, with two main models proposed. One model supports both territories arise from a common region, the neural plate border (neural plate border state model). Alternatively, it is suggested that neural crest is induced from a neural competence territory and the PPR from a non-neural ectodermal territory (binary competence model) (Schlosser, 2006; Schlosser et al., 2014). Regardless of the exact model underlying early induction, the two populations then occupy distinct adjacent territories and begin to express placode or crest specific markers.

The combination and timing of different signals from the anterior neural plate, dorsolateral mesoderm and non-neural ectoderm are crucial for pre-placodal ectoderm induction. Foremost, a constant FGF background from the mesoderm is necessary. In this high FGF environment, high BMP signalling during gastrula stages and suppression of BMPs by inhibitors in the neurula drive placodal vs neural crest induction. Similarly, low Wnt signalling also appears to favour pre-placodal ectoderm fate (Litsiou et al., 2005; McCabe and Bronner-Fraser, 2009;

Saint-Jeannet and Moody, 2014). Placodes were traditionally identified as thickenings of the ectoderm. Tracing their complicated development had been a difficult task until the last decades, when tissue grafting, in-situ hybridisation and dye labelling techniques facilitated tracking individual ectodermal cell populations (Baker and Bronner-Fraser, 2000; Le Douarin et al., 1986; Pieper et al., 2011; Schlosser and Northcutt, 2000). Even still, it was not until the first pan-placodal markers were found that hypotheses on very early placode induction could be tested.

The PPR is defined by the expression of members of the *Six* and *Eya* homeobox families (particularly *Six1*, *Six4*, *Eya1*). From this crescent shaped territory all placodes arise, universally marked by *Six* and *Eya* genes. Interestingly, some markers of specific placodes (e.g. trigeminal, otic, epibranchial) also already start being expressed in the PPR (Schlosser, 2006). These cell populations appear to be partially intermingled. However, their relative positions already correspond to the relative positions the placodes will assume after they separate. Thus, only limited cell movement and migration of placodal populations are documented in *Xenopus*, which is likely to be the case in other vertebrates (Pieper et al., 2011). Some debate exists on this topic and this may be due to the difficulty of tracing cell migration and movement, timing and inter species differences (Saint-Jeannet and Moody, 2014; Streit, 2002).

1.3.2 Placode identity and specification

While the PPR already expresses various placodal markers in the early neurula, commitment to a specific fate is not yet fixed. At this stage, cells will assume fates based on their new positions after experimentally rotating the PPR along the AP axis, leading to normal placode order. This capacity is progressively lost and rotating the PPR later will lead to a reversal in placode order, with cells differentiating according to their original positions. Thus, early signals along the AP axis induce placodal fates on the anterior and posterior PPR, but responsiveness to these signals is then lost (McCabe and Bronner-Fraser, 2009; Saint-Jeannet and Moody, 2014). This AP distinction is supported by molecular markers, complementing anatomical and functional studies. Two placodal domains (equivalence domains) are defined by the expression of two distinct sets of transcription factors, giving rise to placodes that also express these markers. The anterior domain expresses *Pax6, Six3* and *Six6* and gives rise to the adenohypophyseal, the olfactory and lens placodes. The posterior domain is marked by *Pax2, Pax8*, *Irx1/2/3* and the otic and epibranchial placodes emerge from this domain. *FoxI1* and *Dlx3* are also important in the posterior domain. Later *FoxI1* is crucial for epibranchial fate determination and epibranchial marker expression. In contrast, *FoxI1* suppression by *Pax2* and *Pax8* is necessary for normal otic development. The trigeminal placode develops in between these two domains, with the ophthalmic/profundal trigeminal characteristically marked by *Pax3* expression. Ectopic *Pax3* expression posteriorly represses otic/epibranchial fate (Dude et al., 2009; Grocott et al., 2012; Lee et al., 2003; Saint-Jeannet and Moody, 2014). Interestingly, two homeobox genes that mark non-overlapping AP

domains in the CNS are also important regulators of placodal domains accordingly. *Otx2* is necessary in the anterior for adenohypophyseal, olfactory, lens and trigeminal development, while *Gbx2* is required in the posterior for otic and epibranchial determination (Grocott et al., 2012; Saint-Jeannet and Moody, 2014). A pattern is already becoming apparent at this stage, as anterior marker *Otx2* and the posterior *Irx* family both mark the trigeminal. Placode groups are marked by combinations of transcription factors and a degree of overlap exists between the expression domains of these markers. Transcription factor expression progressively becomes more restricted to placodal sub-sets. However, there are various degrees of marker expression overlap between different placodes and ganglia later.

Similar to the PPR, the two early AP domains are induced by various signals. Subsequently, emerging placodes and ganglia are similarly specified by signalling networks and transcription factors originating from ectodermal, mesodermal and endodermal tissue. Once again, the basic developmental networks behind PPR, as well as neural tube induction, are at play during placode specification and segregation. FGF signalling remains important throughout placodal development, for both anterior and posterior placodes. However, timing is crucial, with both *FGF* upregulation and downregulation taking place according to stage. Hence, early on, FGF signals from the CNS (anterior neural ridge) repress lens fate (proposed to consist the placodal ground state in the anterior domain), promoting olfactory fate. In contrast, later, FGFs from the optic vesicle promote lens differentiation (Grocott et al., 2012; McCabe and Bronner-Fraser, 2009; Saint-Jeannet and Moody, 2014). FGF signalling is well described in its role to promote posterior placode specification

as well. FGF8 in combination with at least one other FGF member (the identity of which differs according to species) is necessary for otic/epibranchial specification. In this case, length of FGF exposure, in conjunction with other signals, is important. Initially, FGFs from the neural tube and head mesoderm drive a common posterior placode fate. Long, sustained FGF signalling then promotes epibranchial differentiation. Instead, shorter FGF exposure originating from the neural tube, likely combined with Wnt and Notch signalling, guide the complex otic development (Grocott et al., 2012; McCabe and Bronner-Fraser, 2009; Saint-Jeannet and Moody, 2014). BMP signalling is important for olfactory and lens specification. BMPs from the anterior neural ridge promote olfactory and repress lens fates, while prolonged BMP exposure of ectodermal origin is later necessary for lens specification. Shh signalling is mainly required for adenohypophyseal specification. Finally, trigeminal specification requires PDGF (necessary for ophthalmic lobe development) and Wnt signals (Grocott et al., 2012; McCabe and Bronner-Fraser, 2009; Saint-Jeannet and Moody, 2014).

1.3.3 Cranial ganglia and their markers

After ectodermal placodes develop, they become the main source of the neuronal progenitors that will form CSG (D'Amico-Martel and Noden, 1983). Placodes consist transient structures; they release neuronal progenitors or post-mitotic cells (in the case of the Ophthalmic placode) that assemble to respective ganglia (Begbie et al., 2002). In general, peripheral gangliogenesis involves both migration and differentiation of neural populations. In CSG these two processes are coordinated. By the time migrating cells reach their final positions, they lose their naïve state and

ability to divide. In contrast, in neural crest derived DRG, progenitors can be found in the ganglia after migration and continue to produce new sensory neurons in the ganglia (Blentic et al., 2011; Krispin et al., 2010). As new sensory neurons are added to CSG, they progressively acquire distinct identities, characteristic of each ganglion. Eventually CSG neurons innervate a sensory and a CNS target, giving rise to the various cranial nerves. In a manner comparable to earlier placode stages, CSG differentiation and maturation is also driven and marked by specific transcription factors (Begbie et al., 2002; Blentic et al., 2011; Schlosser, 2006)

As placodes become established and begin producing neuronal progenitors, genes involved in early induction are gradually switched off (e.g. *SoxB1*, *Irx*) (Schlosser, 2006). The processes of CSG neuronal determination and differentiation are marked and driven by a later wave of transcription factors. Many such sensory differentiation genes belong to the basic-helix-loop-helix (bHLH) group, members of which have roles in cell fate determination in various tissues (e.g. *MyoD* in muscle). Vertebrate bHLH genes with roles in neurogenesis are related to Dropsophila proneural bHLH genes, such as *Achaete-scute* and *Atonal* (Schlosser, 2006; Sommer et al., 1996). These include neurogenins (*NGNs*) and their downstream targets *NeuroD*, *Atoh* (Atonal-like) and *ASCL/ASH* (Achaete-scute like) (Ma et al., 1998; Sommer et al., 1996). *NeuroD*, *Atoh* and *ASH* genes are expressed in CNS and PNS, including different placode derivatives (Chitnis and Kintner, 1996; Krüger et al., 2006; Lee et al., 1995; Nakada et al., 2004). The role of *Atoh* and *ASH* genes is mostly associated with neuronal differentiation, rather than determination. *NeuroD* is able to induce ectopic neuron formation in ectoderm, but its endogenous expression suggests its main role lies during terminal

differentiation (Chitnis and Kintner, 1996; Lee et al., 1995; Nakada et al., 2004). *COE* (Collier/Olf-1/EBF) genes also code for bHLH proteins that mark olfactory and trigeminal ganglia in different gnathostomes (Bally-Cuif et al., 1998; Dubois et al., 1998; Dubois and Vincent, 2001).

NGNs are some of the earliest markers of sensory neuron determination and were used to track and describe CSG differentiation and morphogenesis (Begbie et al., 2002). *NGN1* and *NGN2* expression marks differentiating trigeminal, vestibuloacoustic and epibranchial ganglia in mouse, chicken, frog and zebrafish (Andermann et al., 2002; Begbie et al., 2002; Fode et al., 1998; Ma et al., 1998; Perron et al., 1999). Furthermore, *NGNs* are both necessary and sufficient for sensory lineage determination (Andermann et al., 2002; Begbie et al., 2002; Ma et al., 1998). Interestingly, the site of expression and function of each of the two neurogenins differs between species. In mouse, *NGN1* marks the trigeminal and otic, while *NGN2* marks the epibranchial ganglia (Fode et al., 1998; Ma et al., 1998). In chicken, *NGN2* is only expressed in the ophthalmic trigeminal placode and *NGN1* is expressed in all other placodes (Begbie et al., 2002). Similarly, *NGN1* expression alone in zebrafish marks all developing CSG (Andermann et al., 2002). This shows that the family shares this role in CSG neuron determination, but *NGN1* and *NGN2* appear to be interchangeable.

Downstream of *NGNs*, another set of homeobox genes marks terminal sensory differentiation. *Isl1 (Islet1)* is a general marker of CNS and PNS neuronal differentiation, expressed in all general sensory neurons and shown to be necessary for proper eye, ear and trigeminal development along with *POU4f1/Brn-*

3a, *POU4f2/Brn-3b* (Deng et al., 2014; Sun et al., 2008; Wu et al., 2015). *Brn-3a* and *Phox2* mark relative anterior and posterior CSG populations respectively. *Brn-3a* marks the trigeminal and VA, expressed in maxillomandibular and VA cells after they have delaminated. In the ophthalmic branch of the trigeminal, *Brn-3a* is expressed within the placode as well. More posteriorly, *Phox2a* and *Phox2b* are specific markers of epibranchial ganglia (Begbie et al., 2002). The two genes are co-expressed, while the onset of expression of each one differs, with *Phox2a* preceding *Phox2b*. As the epibranchial placodes form serially on the clefts between the branchial slits, *Phox2* expression comes up first in geniculate, then in petrosal and eventually in nodose placodes (*Phox2A*), then marking ganglion neurons which delaminate and differentiate (*Phox2A* and *Phox2B*) (Begbie et al., 2002). Strikingly, these distinct ganglia populations - somatosensory neurons positive for *Brn-3a* and viscerosensory *Phox2B* positive neurons - appear to be related. A *Brn-3a* expressing somatosensory type has been suggested to consist the ground state for posterior ganglion specification. *Phox2B* acts as a developmental switch, essentially conferring the epibranchial/ viscerosensory identity. In the absence of *Phox2B*, epibranchial neurons acquire a somatosensory identity, express trigeminal markers and innervate respective sensory and CNS targets of this fate (D'Autréaux et al., 2011).

1.4 Basally diverging vertebrates: The cyclostomes

In the preceding sections, I have described the main events of vertebrate CNS and PNS development and explored the molecular code framing these. The literature mentioned up to here derives from jawed vertebrates, in which nervous system

development and gene expression are better studied. The earliest diverging vertebrate lineage, the cyclostomes or jawless vertebrates, have important differences to gnathostomes, but share the same core developmental plan, particularly when considering the nervous system. They lack interlocking jaws, paired fins and myelination and they have evolved their own system of immunity (Osório and Rétaux, 2008; Richardson et al., 2010). Nevertheless, jawed and jawless vertebrates share the same CNS and PNS organisation and have comparable nervous system development driven by highly similar patterning mechanisms and gene expression. All cranial placodes are present in cyclostomes, as well as neural crest that exhibits comparable migratory patterns. Owing to these similarities and their key phylogenetic position, cyclostomes consist an invaluable model for understanding vertebrate nervous system evolution (Osório and Rétaux, 2008; Shimeld and Donoghue, 2012).

Cyclostomes include two surviving lineages, the lampreys and the hagfish. Lampreys were originally thought as the sister group to gnathostomes, with hagfish positioned more basally. However, cyclostome monophyly is now strongly supported by various molecular studies based on ribosomal RNA, gene coding and microRNA sequences (Delarbre et al., 2002; Furlong and Holland, 2002a; Heimberg et al., 2010; Kuraku et al., 1999; Mallatt and Sullivan, 1998; Mallatt and Winchell, 2007; Takezaki et al., 2003; S. Yu et al., 2008). What is more, the two groups share the same VLR system of immunity, which evolved independently from gnathostome TCR/Ig based adaptive immunity (Boehm et al., 2012; Cooper and Alder, 2006). Cyclostomes possess a cranium and both lampreys and hagfish are now recognised to possess cartilaginous elements, homologous to vertebrae. They

also retain an adult notochord, which consists an important component of the cyclostome skeleton. Hagfish prey on polychaetes and are opportunistic deep-sea scavengers, feeding on carcasses of whales and other large sea organisms. They are capable of producing large quantities of slime from specialised glands that help them escape predators. They are less studied than lampreys, as they are harder to capture and grow in a laboratory environment (Shimeld and Donoghue, 2012).

Lampreys have filter feeding larvae and several species are ectoparasites of teleost fish as adults. However, more than half of the species known do not have a parasitic adult phase and remain filter feeders throughout their lives. Instead of jaws, they possess an oral disk with concentric rows of teeth, which is used to attach to prey (Richardson et al., 2010). Different species that have been exploited experimentally include *Petromyzon marinus* (sea lamprey), *Lethenteron camtchaticum* (arctic or Japanese lamprey), *Lampetra fluviatilis* (European river lamprey) and *Lampetra planeri* (brook lamprey). Sea lampreys are anadromous, while freshwater lampreys also climb upstream to nest and breed at the end of their life cycle (Osório and Rétaux, 2008). Embryo development is slow, with embryos taking more than a month before they become fully developed larvae. This consists an advantage of the lamprey embryo as a system for studying developmental events that happen in a much more limited timeframe in jawed vertebrate model systems, such as zebrafish and chicken (Shimeld and Donoghue, 2012).

1.4.1 The lamprey CNS

The lamprey central nervous system has typical vertebrate segmentation. The hindbrain is divided into rhombomeres marked by similar patterning as in gnathostomes. The expression of several marker families, such as *Otx*, *Pax6, Pax2/5/8*, *Dlx1/6, EMX, TTF1*, reveal the tripartite organisation of the lamprey brain (Watanabe et al., 2008). Similarly, markers such as *Lhx1/5, Lhx2/9, Pax3/7* and *Hh* occupy conserved domains in the forebrain and midbrain, marking homologous areas or suggested neuromeres (Osorio et al., 2005). Homologous *Hh* expression in the floor plate also hints to conserved patterning mechanisms driving brain segmentation. *Nkx2.1* expression marks the hypothalamus, same as in jawed vertebrates. Differences are also noted, such as the lack of a *Nkx2.1* / *Shh* domain in the ventral telencephalon, suggesting this as a novel gnathostome feature (Osorio et al., 2005; Watanabe et al., 2008). The lamprey MHB is marked by conserved *FGF8/17* and *Pax2/5/8* expression. Similar to gnathostomes, the MHB consists the limit of the anteriorly expressed *Otx,* while posterior *Gbx* expression begins here (Watanabe et al., 2008). Finally, the lamprey hindbrain is marked by conserved *Hox, Krox*, *Irx* and *Hnf* domains (Jimenez-Guri and Pujades, 2011; Murakami et al., 2005).

1.4.2 Lamprey cranial placodes and sensory nerves

Early neural plate development in the lamprey has been relatively understudied, hindered by the difficulty in working with early embryos due to their high yolk content. Nevertheless, in *Petromyzon*, early *BMP2/4*, *Notch/Delta* and *Wnt*

expression has been shown to be similar to other vertebrates, with the former more broadly expressed in non-neural ectoderm, while *Wnt6* marks the developing neural plate and later the neural plate border (Sauka-Spengler et al., 2007). Other neural plate border and neural crest markers have also been shown to have conserved expression and roles early in development such as *Msx1/2, Dlx3/5, Pax3/7, Zic*, *AP-2* and *PRDM1* (Meulemans and Bronner-Fraser, 2002; Nikitina et al., 2011; Sauka-Spengler et al., 2007).

In contrast, there is a better understanding of cranial nerve organisation and neural crest development in the arctic lamprey and CSG development in the sea lamprey. Lampreys possess all twelve vertebrate cranial nerves and spinal nerves. As already mentioned, lamprey cranial sensory nerves arise from the respective ectodermal placodes, while neural crest contributes to cranial ganglia and dorsal root ganglia in the same fashion as described in gnathostomes. Lampreys have paired lateral eyes, paired ears, a lateral line system, as well as a pineal and parapineal organ (developing from the epiphysis) that can also sense light (Richardson et al., 2010). Unlike the three semi-circular canals found in the gnathostome ear, lampreys only possess two vertical canals, while only a single one is found in the hagfish torus-like ear structure. Nevertheless, early otic development and hair cell formation in lamprey are comparable to jawed vertebrates, with conserved expression of markers like *Ptc, Tbx1* and *follistatin* (Hammond and Whitfield, 2006). What is more, careful imaging of the sea lamprey labyrinth revealed the existence of horizontal vertical ducts that have evolved independently in lampreys (Maklad et al., 2014). The lamprey cranial nerves develop with stereotypical vertebrate positioning. Similar to melanocytes, nerves

and ganglia are primarily spaced in between myotomes, suggesting interaction and guidance by migrating neural crest is at play. The anterior profundal and ophthalmic trigeminal and facial nerves are more heavily branched than postotic nerves. The first two lamprey myotomes undergo an expansion in the preotic area, above and below the eye. In this process, preotic nerves are entrenched below the myotomes, resulting in a deeper position of these nerves compared to gnathostomes (Kuratani et al., 1997).

CSG development in lamprey is highly similar to jawed vertebrates, with some notable exceptions. Placode and neural crest derived neurons migrate to form CSG (Modrell et al., 2014). In the anterior, the ophthalmic and maxillomandibular trigeminal are separate, as in several other vertebrates. However, in lamprey, the two branches of the maxillomandibular (innervating the upper and lower lip respectively) arise from two distinguishable precursor domains that are separated by ophthalmic precursors. More posteriorly, the geniculate, anterior lateral line and vestibuloacoustic ganglia are fused to a contiguous ganglion complex. Postotically, the petrosal follows first with the posterior lateral line ganglia positioned dorsally to the petrosal. Finally, a chain of nodose ganglia forms on top of the clefts of the developing branchial arches. The expression of only a few lamprey CSG markers is described, which corresponds to conserved CSG domains. Lamprey *Pax3/7*, the single homolog to gnathostome *Pax3* specifically marks the placodally derived ophthalmic, but not maxillomandibular trigeminal. This suggests an ancient origin of the two, the identity of which must have been established in the common vertebrate ancestor (Modrell et al., 2014). *NGN* is a conserved, specific marker of placodes and CSG derivatives as well, while other CNS sites are also conserved

between the two lineages. A single neurogenin, *NgnA,* is expressed in all CSG in *L. planeri*, while *NgnB* is not found outside the CNS. As in gnathostomes, this supports the need for at least one member of the family in the development of each CSG (Lara-Ramírez et al., 2015). Epibranchial ganglia are specifically marked by *Phox2* expression, consistent with the role of the family in epibranchial specification in gnathostomes, while lamprey *Ascl1* is also expressed in trigeminal CSG (Häming et al., 2011).

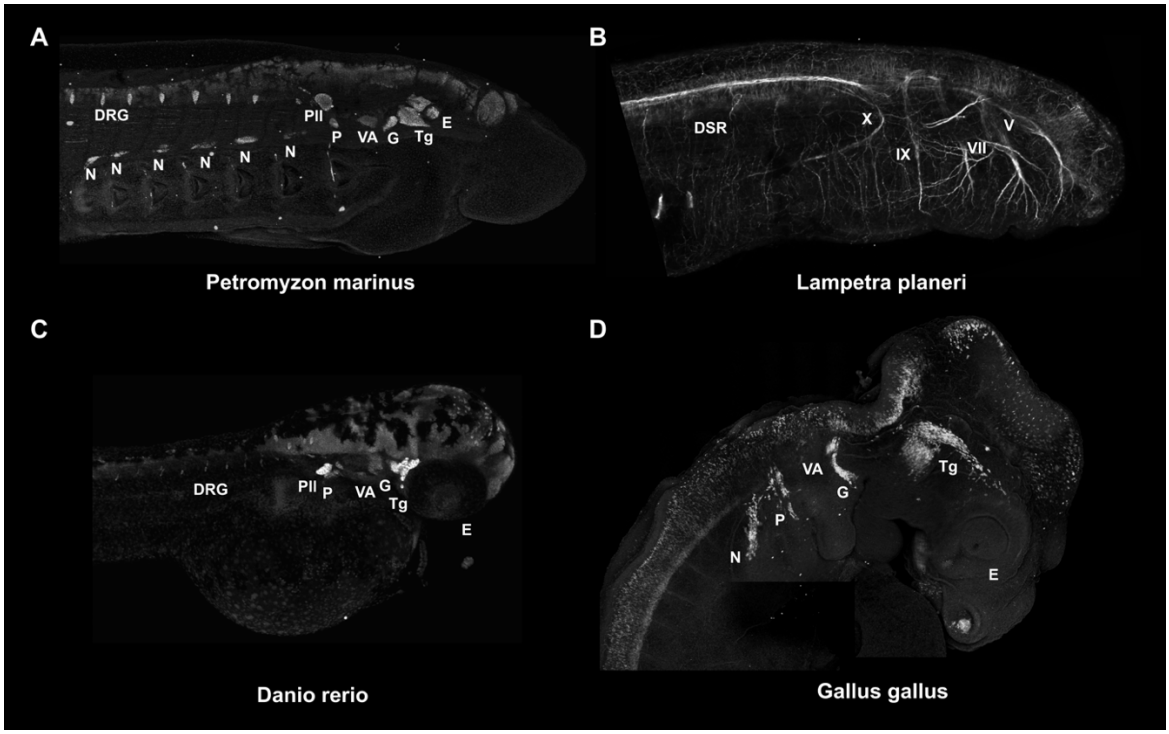


Figure 1.2 CSG and respective cranial nerves. A. CSG in stage 28 (Tahara, 1988) sea lamprey embryo (*Petromyzon Marinus*) **B.** Cranial nerves in stage 26 (Tahara) brook lamprey embryo (*L. planeri*) **C.** CSG in 48hpf zebrafish embryo (*Danio Rerio)* **D.** CSG in HH17 (Hamburger and Hamilton, 1951) chicken embryo (*Gallus gallus*). Embryos are oriented in the same direction (anterior-right, dorsal-top). Embryos were stained with antibodies against Hu/ELAV (A, C, D) or acetylated Tubulin (B) and all embryos were also stained with DAPI. Abbreviations: Tg, Trigeminal. E, Eye. G, Geniculate. VA, Vestibuloacoustic. P, Petrosal. PII, Posterior Lateral-Line. N, Nodose. DRG, Dorsal Root Ganglia. V, Cranial Nerve V. VII, Cranial Nerve VII. IX, Cranial Nerve IX. X, Cranial Nerve X. DSR, Dorsal Root of Spinal Nerve.

1.5 Chordate relatives of vertebrates – CSG and placode origins

The comparative study of CSG development across vertebrates revealed high conservation down to the most basally splitting lineage, the cyclostomes. This, along with lines of evidence from other organ systems and genetics, created a view of the ancestral vertebrate already possessing all CSG, as well as the basic organisation of the vertebrate head. This led to the suggestion of the “new head” hypothesis; placodes, neural crest and a centralisation of the senses in the vertebrate head consist a crucial component of a predatory lifestyle. In turn, this new adaptation to a predatory life would have allowed vertebrates to thrive by outcompeting other groups in early marine ecosystems (Gans and Northcutt, 1983; Northcutt and Gans, 1983). If the last common ancestor of living vertebrates already possessed a complete set of placodes, the question then remains of where these structures and their constituent cell types originated from.

Chordate relatives of vertebrates, the cephalochordates and the urochordates, have been crucial in the search to find anatomical, developmental and genetic links to sensory structures and cell types homologous to CSG and their cells. A lot of work on this endeavour dates back not only to classical genetics, but even to classical embryology work of the late 19th and early 20th century. However, it was not until the genomics era, when compiled data from many chordate and deuterostome genomes elucidated chordate relationships, that earlier findings could be properly understood and connected. The traditional view was that cephalochordates (also known as amphioxus) are the sister group of vertebrates, while the highly divergent urochordates (or tunicates) are basal in the chordate tree.

However, the position of tunicates as the sister group of vertebrates is now well established, while the urochordate body plan is understood to be highly derived secondarily (Delsuc et al., 2006; Lemaire, 2009) (Lemaire et al., 2008). In addition, while cephalochordates diverged earlier during chordate evolution, they are thought to have retained a lot of ancestral/basal characters (Holland et al., 2004). Nonetheless, gene expression and morphological studies have suggested that tunicates have structures of homologous origin to vertebrate placodes, painting the picture of a gradual acquisition of placodal cell types during chordate evolution (Graham and Shimeld, 2013).

In this section, I will be covering current knowledge behind tunicate and amphioxus sensory development and discuss our understanding of evolutionary connections between sensory cells in these groups to vertebrate placodes. The main focus here will be put on conserved patterning mechanisms and transcription factor expression that provide evidence for homology. There are also notable differences in patterning, expression and cell fate determination among chordates, but a detailed description of these will not be given here. For a more extensive analysis, Schlosser and colleagues provide two exhaustive reviews covering similarities and differences of ectodermal patterning, sensory domains and sensory cell types among chordates (Patthey et al., 2014; Schlosser et al., 2014).

1.5.1 The origin of ectodermal territories and placodal progenitors

The distinction between neural and non-neural ectoderm has been suggested to consist an ancient bilaterian trait. A BMP gradient set up by a combination of BMPs

and BMP antagonists is crucial for dorsoventral patterning across animals (Arendt et al., 2008). Similar to vertebrates, it drives the establishment of transcription factor expression and the formation of the neural plate in amphioxus (Kozmikova et al., 2011). While the dorsal neural plate, characteristic of all chordates, is also shared by tunicates, patterning and gene expression have diverged, reflecting their derived state (Lemaire, 2009). Different signals involved in neural AP patterning even predate the evolution of vertebrates. Wnt signalling is ancient in bilaterians, with roles in ectodermal patterning in protostomes and deuterostomes alike (Schlosser et al., 2014). Patterning by RA signalling is likely a chordate character, as it is involved in different AP domains in tunicates and it controls the AP positioning of amphioxus epidermal sensory neurons (ESN) (Schlosser et al., 2014; Schubert et al., 2004). FGF signalling is involved in the determination of anterior and posterior neural and epidermal domains in tunicates, but no such role is seen in amphioxus, suggesting it acquired an AP patterning function in Olfactores (comprising vertebrates and tunicates) instead (Schlosser et al., 2014).

In amphioxus, conserved expression of several genes distinguishes dorsal (*SoxB1*, *Zic*) to ventral (*AP2*, *Dlx*, *Msx*, *Vent*) ectoderm. Furthermore, the neural plate border is marked by the conserved expression of genes like *Pax3/7*, *Zic*, *Snail* and to some extent *Msx* and *Dlx*. However, the expression of these transcription factors appears to pattern ventral versus dorsal domains in the neural plate itself. In contrast, most genes associated with the vertebrate neural crest or PPR territories are not specifically enriched in this region (Schlosser et al., 2014; Yu et al., 2008). Ventrally, the non-neural ectoderm does not appear to be segregated further to domains marked by particular markers. Instead, expression of genes associated with

vertebrate placodes, including *Six1/2*, *Hu/Elav*, *POU-IV*, *SoxB1*, *Coe* and *Islet* mark primary and secondary epidermal sensory cells (ESN) found all over the epidermis. These sensory cells develop from progenitors that arise from a ventral neurogenic field, in response to BMP and Notch signalling, representing a cephalochordate specific mechanism (Candiani et al., 2006; Lu et al., 2012; Mazet et al., 2004; Meulemans and Bronner-Fraser, 2007; Satoh et al., 2001; Schlosser et al., 2014; Schubert et al., 2004).

Patterning mechanisms have diverged in ascidians and morphogen gradients along the body axes are largely absent. Instead, cell to cell interactions and segregation of signals in particular cell fates drive the formation of the conserved chordate plan in a different way (Lemaire, 2009). The posterior and anterior neural plate is shaped in this fashion. FGF signalling has important roles in controlling specific cell fate gene expression, especially in the anterior border of the neural plate, where it inhibits *FoxC* and promotes *Zic* expression (Wagner and Levine, 2012). *Zic* and *SoxB* mark the neural ectoderm early on, while *Dlx* expression marks non-neural ectoderm, similar to amphioxus and vertebrates (Irvine et al., 2007; Miya and Nishida, 2003; Wada and Saiga, 2002). Again, similar to vertebrates and amphioxus, *Pax3/7, Zic, Msx* and *Snail* mark the neural plate border (Imai et al., 2006; Wada and Saiga, 2002). There are some cells that have been suggested to consist neural crest homologs in *Ciona*. Two populations that contribute to the pigmented organs of the sensory vesicle develop via mechanisms comparable to neural crest. The first is a precursor lineage of the ocellus (the light sensing organ), which develops from the neural plate border. These cells express various neural crest markers, while they become migratory upon *Twist* overexpression (Abitua et

al., 2012). The second lineage also contributes pigmented cells to the sensory vesicle, expresses *HNK-1*, like neural crest derived pigment cells in vertebrates, but is of mesodermal origin instead (Jeffery et al., 2008). Finally, bipolar tail neurons (BTN) have also been suggested to consist a neural crest homolog. BTN express *Islet*, *POUIV* and *NGN*, while they resemble DRG in their caudal positioning in the body and their developmental origin from migratory progenitors that arise from the neural plate border (Stolfi et al., 2015). Nevertheless, as in amphioxus, no territory showing neural crest marker enrichment in the tunicate neural plate border is present (Schlosser et al., 2014). In contrast, the early non-neural ectoderm has regions marked by the expression of pan-placodal markers of the *Six* and *Eya* families. *Six1/2* and *Eya* are expressed in non-neural ectoderm bordering the anterior neural plate border, where they persist when the oral siphon primordium and palps (*eya*) develop. In addition, they also mark the forming atrial siphon primordia, along with *Six4/5* (Graham and Shimeld, 2013; Mazet et al., 2005). This is discussed further below.

1.5.2 Placode homologs in tunicates and amphioxus

As discussed earlier, amphioxus does not possess early domains with placodal marker enrichment. Individually, as they emerge, subsets of sensory cells express transcription factors shared by some placodal populations. No organised structures exist that carry out similar functions or that are marked by comparable marker expression to CSG. There are only two populations that arise in the anterior suggested homologous to anterior placodal domains, the olfactory and the adenohypophysis (Graham and Shimeld, 2013; Schlosser et al., 2014). The anterior

non-neural ectoderm of amphioxus produces various chemosensory cells marked by *Pax6, Six3/6 and Pitx*. Again, no organised olfactory structures develop and these primary sensory cells remain scattered in the epidermis. They also do not express *Six1/2, Six4/5* or *Eya* (Kozmik et al., 2007). In contrast, *Six1/2, Eya, Pax6, Six3/6, Pitx* and *Islet* all mark Haetschek's pit, suggested to be homologous to the adenohypophysis (Boorman and Shimeld, 2002a; Candiani et al., 2008; Jackman et al., 2000; Kozmik et al., 2007). Finally, as described, groups of ESN across the whole epidermis develop from a ventral neurogenic domain. Amphioxus ESN express various markers of sensory cells originating from different placodes in vertebrates. These cells possibly represent mechanoreceptors and chemoreceptors of at least three different types (Kaltenbach et al., 2009).

Unlike amphioxus, tunicate embryos have ectodermal regions which express markers of specific placodes. The populations of these domains do not show high proliferation rates and migratory capabilities like placodal derived sensory progenitors, but they do produce similar sensory cell types (Graham and Shimeld, 2013; Schlosser et al., 2014). The most characteristic of these populations are the oral and atrial siphon primordia, which develop as thickenings of the ectoderm, express pan-placodal markers *Six1/2* and *Eya* and produce cell populations of the two siphons of the adult tunicate. The oral siphon primordium is positioned at the anterior and it remains in contact with the adjacent neural tube throughout development. It is marked by *Six1/2, Six3/6, Eya* and *Pitx*, while various sensory types, such as photoreceptors, primary and secondary sensory cells of the coronal organ arise from here. It has been suggested to be homologous to the adenohypophysis (Boorman and Shimeld, 2002a; Mazet et al., 2005). Accordingly,

the atrial siphon primordia have been suggested to be homologous to the posterior vertebrate placodal domain and particularly the otic placode. This is based on their posterior positioning, expression of *Six1/2, Six4/5, Eya*, *Pax2/5/8* and *Foxl*, induction by FGF and the fact that the cupular and capsular sensory organs develop from here (Kourakis and Smith, 2007; Mazet and Shimeld, 2005). Furthermore, palp sensory cells have been suggested to be homologous to olfactory placode populations of vertebrates. Palp sensory cells are positioned anterior to the oral siphon primordium, develop from the non-neural ectoderm bordering the anterior neural plate and express *Eya, POUIV* and *COE* (Schlosser et al., 2014).

This comparison of ectodermal domains of the tunicate embryo and vertebrate placodes provides one basis for inferring homology by taking account of positional correlation along the AP axis. However, similar sensory cell types can arise from non-correlated domains as well. For example, the oral siphon primordium gives rise to hair cells of the coronal organ. These hair cells have been shown to be potentially homologous to the hair cells of the vertebrate ear, based on their genetic identity and functional characteristics (Manni et al., 2006; Rigon et al., 2018). Similar is the case for various photoreceptors that arise in different embryonic (ocellus and otolith) and adult (oral siphon) tunicate tissues (Graham and Shimeld, 2013; Shimeld et al., 2005). Thus, similar sensory cells can also stem from different developmental origins, with homology at cell level, instead of tissue level (Arendt et al., 2016a).

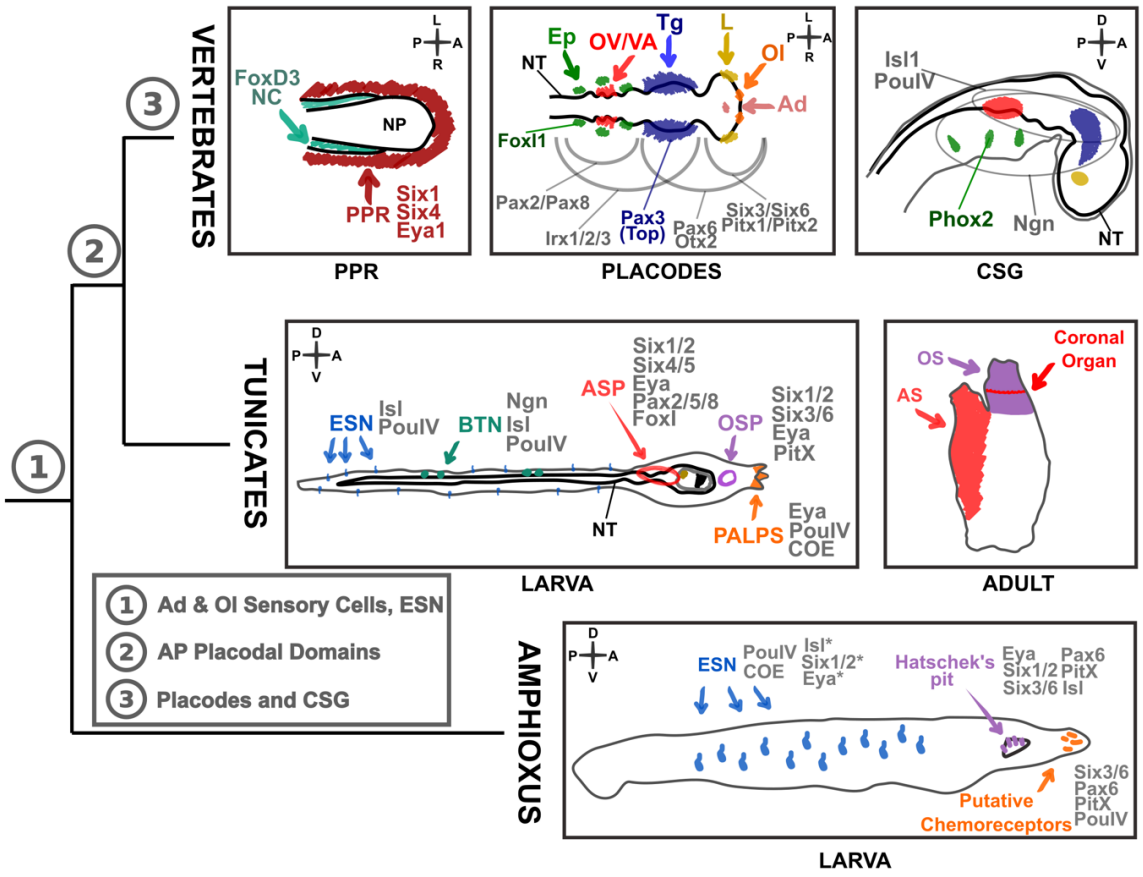


Figure 1.3 Graphical summary of placode and CSG marker genes in chordate sensory populations. For details on genes involved in vertebrate placode and CSG development see sections 1.3 and 1.4. For details on tunicate and amphioxus sensory development, suggested homologies to vertebrate sensory cells and relevant gene expression see section 1.5. Amphioxus genes annotated with an asterisk are expressed in a subset of ESN. Additional abbreviations: PPR, Pre-Placodal Region. NC, Neural Crest. NP, Neural Plate. NT, Neural Tube. OV, Otic Vesicle. Ep, Epibranchial. L, Lens. Ol, Olfactory. Ad, Adenohypophysis. Top, Trigeminal Ophthalmic. ESN, Epidermal Sensory Neurons. BTN, Bipolar Tail Neurons. ASP, Atrial Siphon Primordium. AS, Atrial Siphon. OSP, Oral Siphon Primordium. OS, Oral Siphon. AP, Anteroposterior.

1.6 Genomic sequence evolution and developmental innovations

In the previous sections I explored the sensory cells and structures of vertebrates and their chordate relatives. I discussed the studies comparing these cells and systems, focusing on conservation of gene expression and function. A good understanding of gene homology is crucial in this process, but assigning the type

of gene homology between groups is not always straightforward. For example, differential gene gain and loss can complicate comparisons. In addition to gene conservation, conserved function and regulation also provide evidence of homology, as we saw when comparing chordate sensory cells. The genomes of vertebrates, tunicates and amphioxus have changed in different ways, reflecting their evolutionary history and adaptations. This final section of the introduction will focus on genomic changes that shaped vertebrate evolution, providing novel sequences and fuelling developmental innovations.

1.6.1 Gene copy changes and whole genome duplications

The evolution of vertebrate innovations may be coupled with genomic changes that took place early in vertebrate evolution. Early studies on genomic DNA content lead Susumu Ohno to suggest that whole genome duplications played a key role in the evolution of vertebrate innovations (Ohno, 2013). Molecular phylogenetics, early gene mapping studies and eventually the sequencing of the amphioxus genome eventually revealed that two round of whole genome duplication (2R hypothesis) followed by subsequent re-diploidisations and gene loss lead to current paralog counts (Cañestro et al., 2013; Dehal and Boore, 2005; Furlong and Holland, 2002b; Hokamp et al., 2003; Holland et al., 1994; Putnam et al., 2008). Following 2R, differential gene gains and losses further shaped the genomes of different classes of vertebrates. Ray finned fish (*Actinopterygii*) underwent an additional round of whole genome duplication (3R) (Meyer and Van De Peer, 2005). The gain of new gene copies has been associated with key developmental changes and the emergence of new structures. Indeed, after 2R, paralogs of transcription factor

families involved in development (TransDev genes) were relatively highly retained (Cañestro et al., 2013; Putnam et al., 2008). Subsequent neo-functionalisation and sub-functionalisation of different copies could then allow for the evolution of new cell types via the re-wiring of existing regulatory networks. In this fashion, whole genome duplications allowed for the diversification of existing cell types and the evolution of new organisation and function (Cañestro et al., 2013; L. Z. Holland, 2013). The distinction between segregation of existing ancestral functions among new paralogs versus specialisation is also important for understanding cell type evolution. In depth functional genomic analysis of the European amphioxus (*Branchiostoma lanceolatum*) and comparison to vertebrates provided evidence of widespread specialisation of 2R TransDev paralogs. Thus, TransDev gene families with multiple 2R paralogs and broadly expressed in vertebrate embryos were found to be restricted in their ancestral expression. Interestingly, many such specialised 2R paralogs also have regulatory environments with increased complexity, suggesting that fine tuning of expression by binding of multiple transcription factors, silencing and other mechanisms are at play (Marlétaz et al., 2018).

Cyclostomes share a comparable number of paralogs to gnathostomes (Kuraku, 2013). The sequencing of the lamprey germ line genome clearly supported that two rounds of whole genome duplications also occurred in lampreys (Smith and Keinath, 2015). However, studies of different gene families have shown that it is difficult to assign clear one to one orthology between lamprey and gnathostome paralogs. Different lamprey paralogs are comparable in numbers and clearly group within the respective gene families in reconstructed phylogenies, but often fail to group further with specific gnathostome paralogs (Kuraku, 2013). This lack of clear

orthology, in conjunction with differences in paralog numbers in many families lead to the hypothesis that only one round of whole genome duplication is shared between cyclostomes and gnathostomes (Smith and Keinath, 2015). However, recently, careful reconstruction of pre-2R vertebrate linkage groups and mapping to jawless and jawed vertebrate genomes, clearly supports that both whole genome duplications are shared (Sacerdot et al., 2018). Instead, loss of paralogs, in combination with differential re-diploidisation in the two groups after 2R may explain orthology and synteny problems (Furlong and Holland, 2002b; Kuraku, 2013).

The sequencing of the Florida amphioxus (*Branchiostoma floridae*) genome allowed for the reconstruction of the ancestral chordate linkage groups, a proxy to synteny in the ancestral chordate chromosomes. Comparing amphioxus and vertebrate genomes then revealed quadruple conservation of synteny, offering strong support for the 2R hypothesis as mentioned above (Putnam et al., 2008). In addition, analysis of transcription factor family diversity in amphioxus and other chordates revealed little gene loss in amphioxus (Paps et al., 2012). This is in line with the view of amphioxus as slowly evolving and retaining basal chordate characters (Holland et al., 2004). In stark contrast, studies of the *Ciona intestinalis* genome detected loss of many genes with roles in development, loss of synteny and lineage specific duplications. This reflects the secondary loss or reduction of chordate characters and the different body plan of the adult tunicate (Dehal et al., 2002; Holland and Gibson-Brown, 2003; Hughes and Friedman, 2005).

1.6.2 Non-coding sequence conservation and conserved regulation

Outside the coding genome, conserved non-coding elements (CNE) were originally identified that are shared among gnathostomes. Strikingly, a number of these CNE are shared between gnathostomes and lampreys, meaning these elements have remained conserved for at least 450-510 million years (McEwen et al., 2009). The majority of vertebrate CNE neighbour TransDev genes, while CNE of protostome groups have also been shown to have independently evolved around similar loci (Woolfe et al., 2005). Reporter construct studies in lamprey and zebrafish have shown CNE regulate the expression of various developmental transcription factors. In addition, they are able to function across species, confirming their sequence conservation reflects regulatory conservation (McEwen et al., 2009). Furthermore, at least half of vertebrate CNE are duplicated, found in paralogous loci. These duplicated CNE can drive expression in overlapping but not identical domains, corresponding to those of respective paralogous genes (McEwen et al., 2006). Therefore, CNE provide evidence of conserved ancestral regulation shared by all vertebrates, which first arose and became fixed in the last common ancestor, prior to whole genome duplications. A very small number of vertebrate CNE are shared by amphioxus (Putnam et al., 2008). This led to hypothesising an ancient chordate origin of some of these regulatory blocks that then evolved to a much larger fixed set in vertebrates. Tunicates do not show any signs of shared CNE with vertebrates or amphioxus (McEwen et al., 2009).

1.7 Thesis layout

The main aim of this work is to study the molecular identity of developing CSG and provide insight on their evolutionary origins. For this reason, a comparative approach is undertaken that traces cell type relationships across vertebrates and their chordate relatives, by looking at the expression and regulation of conserved molecular markers. Chapter 3 details the first step in this process, the identification of conserved pan-vertebrate CSG markers. To achieve this, the lamprey is used to characterize the conservation and specificity of marker expression. Two of the identified marker families consist the starting ground for a comparative study of CSG cell types between vertebrates and other chordates. Chapter 4 explores the role of the estrogen related receptor (*ERR*) as an ancient sensory cell marker in bilaterian animals. It is shown that a single *ERR* homolog acquired a role in VA development in vertebrates, supported by conserved regulation and expression between lamprey and gnathostomes and a mode of evolution at cell type level is discussed. Chapter 5 focuses on the *Hmx* homebox gene family and how its unique genomic architecture in vertebrates is connected to ancestral regulation of expression in CSG. It then provides evidence of homology between CSG and tunicate sensory neurons, showing that *Hmx* is a specific marker and a fate determinant of the latter. Finally, chapter 6 reviews how this study advances our understanding of CSG cell type evolution and considers the next steps in this comparative exploration.

Chapter 2 Materials and Methods

2.1 Gene and CNE identification and sequence analyses

2.1.1 Gene identification and sequence sources

The identification of cyclostome and invertebrate sequences was carried out using a reciprocal BLAST (Camacho et al., 2009) approach, by analyzing multiple sources of data. For lamprey this constituted *Lampetra planeri* transcriptome data (Lara-Ramírez et al., 2017), genome assemblies for *Petromyzon marinus* (Smith et al., 2018, 2013), a *P. marinus* transcriptome assembly built in-house from Illumina GAII data available on SRA, the *Lethenteron camtschaticum* genome assembly (Mehta et al., 2013) , and an *L. camtschaticum* transcriptome assembly kindly provided by Juan Pascual-Anaya. For hagfish I searched the *Eptatretus burgeri* genome Eburgeri_3.2 genome assembly. Transcriptomes were searched using annotated *Homo sapiens* protein sequences (longest isoforms available), retrieved from the NCBI RefSeq database for each gene family. Transcript sequences corresponding to the protein queries were then matched back to the NCBI RefSeq database to provide additional support for their identity. Transcripts filtered through this process were subsequently mapped to the respective genomes and consolidated into individual genetic loci. Sequences for the inshore hagfish (*Eptatretus burgeri*), common limpet (*Patella vulgata*) and keelworm (*Spirobranchus lamarcki*) (Kenny et al., 2015) were identified via the same approach. For *P. vulgata* and *S. lamarcki*

previously published transcriptome and genome data were used (Kenny et al., 2015; Kenny and Shimeld, 2012; Werner et al., 2013).

For the *Hmx* family, three genes were identified in each dataset as described in chapters 3 and 5. In *P. marinus* (genome version Pmar_germline 1.0/petMar3) these are located on scaffold_00015 represented by gene models PMZ_0020818-RA, PMZ_0048148 and PMZ_0028877-RA. An additional *P. marinus* scaffold, scaffold_00813, also contained a gene model (PMZ_0038761-RA) with an *Hmx* type homeobox. However, when the sequence of this locus was examined, it was found to have >99.5% identity to part of the *Hmx* locus from scaffold_00015 (supplementary appendix section s4). Hence, it is either a very recent duplication of sequence from scaffold_00015, or an artefact of the genome assembly process, and I have not considered it further.

2.1.2 Alignments

Sequence alignments were built using MUSCLE (Edgar, 2004) and trimmed using TRIMAL with a gap threshold of 0.5. Maximum parsimony (raxml (Stamatakis, 2014)) and Bayesian (Mr Bayes (Altekar et al., 2004; Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003)) analysis were employed (JTT model, 1000 bootstrap) to construct phylogenetic trees. All alignments not presented in other chapters are included in the supplementary appendix.

Protein sequences included in this analysis retrieved from the NCBI database are listed in the supplementary appendix. Homeodomain sequences used to construct

the phylogenetic trees of fig3.11 and fig3.14 were retrieved from the HomeoDB[2] database (Zhong et al., 2008; Zhong and Holland, 2011).

2.1.3 Synteny analysis

ERRγ synteny was investigated in the genomes of different annotated vertebrate species listed in figure 3.4. For jawed vertebrate genomes, the following genome assemblies were searched using the UCSC genome browser: *Homo sapiens* (GRCh38.p12, NCBI ID: 5800238), *Gallus gallus* (GRCg6a, NCBI ID: 1668981), *Xenopus tropicalis* (Xenopus_tropicalis_v9.1, NCBI ID: 768701), *Danio rerio* (Zebrafish Build 11/ GRCz11, NCBI ID: 1104621), *Lepisosteus oculatus* (LepOcu1, NCBI ID: 327908). Additionally, the following genomes were used to identify *ERR-ER* locations in different species shown in fig 3.3: *Oryzias latipes (*UCSC oryLat2), *Anolis carolinensis* (AnoCar2.0, accession #GCA_000090745.1), *Branchiostoma floridae* (JGI v2.0), *Lottia gigantea* (JGI Lotgi1), *Saccoglossus kowalevskii* (Skow_1.1 NCBI ID: 95611). Lamprey genes were identified and mapped as described in section 2.1.1.

2.1.4 CNE identification

To identify CNEs I first compared jawed vertebrate loci using the Conserved Non-coDing Orthologous Regions (CONDOR) database (Woolfe et al., 2007) . This identified a small number of elements in the *ERRγ* and *Hmx3-Hmx2* loci that were conserved across jawed vertebrates. I extended this to lamprey and hagfish (for *Hmx*), using sequence similarity searches to search specifically for these conserved

elements in these lineages. For *ERR*, the last two 3' most introns of *L. camtchaticum ERRa* and *ERR*γ from Human, Mouse, Chicken, *Anolis*, *Xenopus,* Zebrafish and elephant shark (Callorhinchus milii- 6.1.3 calMil1, NCBI ID: 85971) were aligned using BLAST, and VISTA plots were built using mVISTA (Frazer et al., 2004) to reveal conservation peaks. Alignments and molecular phylogenetic analyses were undertaken using MUSCLE and RAxML using the Maximum Likelihood method. 1000 bootstrap replicates were used to assess node confidence.

2.2 Gene fragment and CNE Cloning

2.2.1 Lamprey Sequence Cloning

Lamprey gene fragments were cloned via PCR amplification from *L. planeri* cDNA and CNE sequences from *L. planeri* genomic DNA, using the primers shown in Table 2.1. All clones were verified by sanger sequencing. Gene fragments were cloned in the pCR™II (Invitrogen™) vector using TA Cloning™. Lamprey CNE fragments were cloned using restriction ligation cloning or the Cold Fusion system (System Biosciences) into respective GFP reporter vectors for testing in chicken (barcode 1 tagged pTK vector (Chen and Streit, 2015)), lamprey (Parker et al., 2014a) and zebrafish (Ac/Ds-E1b-EGFP vector - Adgene 102417). More specifically, for lamprey transgenics, CNE sequences were cloned into the HLC vector with a zebrafish *krt4* minimal promoter (Parker et al., 2014b). ERR8t12 and dHMXcne were first cloned in the pCR™II vector and then transferred to the krt4-HLC vector after restriction digest of both with BamHI-XhoI and ligation. The Cold Fusion system (System Biosciences) was used to directly clone uHMXcne in krt-

HLC, using the primers listed in table 2.1, that have homologous ends to the vector digested with KpnI and XhoI. For chicken transgenics, ERR8t12 was cloned from pCR™II to the tag1-pTK vector (see section 2.1) using restriction-ligation cloning with Spe1-ECORV. For zebrafish transgenics, ERR8t12 was cloned from pCR™II to the Ac/Ds-E1b vector using the Cold Fusion system (System Biosciences) with the primers tcgagtttacgtaccgctagGCCCACACACCGCCTAATA and tatcgccgcaagcttgctagGCATCCGAGAGGAATGCTCA, that have homologous ends to the vector.

2.2.2 Invertebrate Sequence Cloning

The *Ciona* intestinalis *Hmx/Nkx5* locus was already annotated (Wada et al., 2003), though the gene model was incomplete. Since no ESTs mapped to this gene, no clones were available in arrayed plasmid libraries. I hence first cloned a fragment of the gene in the pGEM®-T vector (Promega), using the primers shown in table 2.1, and used this for in situ hybridisation. I then used homology to *Ciona savignyi*, coupled with RNAseq data mapped on ANISEED (Brozovic et al., 2018), to identify the full open reading frame. This was amplified and cloned using the primers shown below, in section 2.5.1. *Branchiostoma lanceolatum Hmx* was identified by searching the genome (Marlétaz et al., 2018). The in-situ probe was cloned by PCR from 24-36 hours post fertilisation larvae cDNA. The Ciona *ERR* probe was derived from the *Ciona* release 1 gene collection library, clone coordinate R1CiGC03o05.

	Gene/Region	Forward Primer	Reverse Primer	Size
L. planeri Probes	ERRA	CGAGTACATGCTCAACGCCA	GTCCTGAAGCTTTTGCACCG	847
	ERRB	CCTCGTCCATGACGAAGTCC	GTGCAGACCAAGTGCGAGTA	702
	ERRC	GCAAGTGCGAGTACATGCTG	AGGAAAAGTTTGTGCATGGGC	1012
	HMXA	GACGCAGGTGAAGATCTGGT	GGATGGACAAAAAGCGGCAG	909
	HMXB	CCGCGCTGGCATTTCTTATT	GTAGCGCTTCATGTCGAACG	780
	HMXC	ACGTCGTCCGGGTTTGATAA	CGGAAGACGACAACGAAGGA	517
	DRGX	GCCCTGTCGCGATGTTTTAC	TGAACACGTCTGGGTAGTGC	216
	PRDM12	CGAGAAGTTGGAGTTCGTCGT	TTAGCTTCTCGGGAGGAGTCG	513
	GAD	CTCGCAGGCTCAGAGGTATG	CAGATCCTGGTGGACTGTCG	1072
	SHOXA	AACTCGGTGTGCGGTTGAT	CTGATCGGCACCACGAGTC	437
	SHOXB	GGAGATCCCTTCCATTTGGAC	GGTTTCAGAACAGGAGAGCCA	164
	PROXA	CACACAGTCCCCGACTAACC	AGCGTTGATCGCCTTGAAGA	958
	PROXB	TGTATCACGTCGCAGCTCATC	TCCATCTTGCAGATCACCTTGT	311
	PROXC	CTGAGGAGCATCGAGCAAGA	GTCCGTGCGAACTTCTCCAT	395
	PROXD	CGGACATGTCGCCTTACTCC	TTTTCCGCGGACGACAAAAG	549
Invertebrate Probes	ERR Patella	TCGCGATGGGTTAGGAATGA	CAACGTCATGGGGGAGACAT	972
	ER Patella	TATCCGCCGGTCTAGTCGT	CGTCTGCGCAAGTGTTATGA	912
	ERR Spirobranchus	CAACCTGCGTGCGTACATTC	GCACAGGGTAACTGCGAGAT	1062
	HMX Ciona	AGATCTGGTTTCAGAACCGACG	GGGATTCCCCTCTTTGCGTA	560
Lamprey CNE	ERR8t12	GCCCACACACCGCCTAATA	GCATCCGAGAGGAATGCTCA	2978
	dHMXcne	CATCATCAGCGCGCGTATTT	TTCCTGCACACAACTCTCGC	2078
	uHMXcne	TTTCTCTATCGATAGGCAGCAAG CGATGTATGCAA	GATCGCAGATCTCGACGAGTAC GCGAAGAGGTGG	1582

Table 2.1. PCR primers used to amplify new sequences

2.3 Embryos and In situ hybridisation

2.3.1 Embryos

Naturally spawned *Lampetra planeri* embryos were collected from a shallow stream in the New Forest, UK, under a Permission granted by Forestry England. They were cultured in filtered river water at 16°C and processed for in situ hybridisation as previously described (Lara-Ramírez et al., 2015). Adult *Ciona intestinalis* were collected from Northney Marina, UK, and adult *P. vulgata* from Tinside, Plymouth,

UK. Both were maintained in a circulating sea water aquarium at 14°C under constant light. *C. intestinalis* gametes were liberated by dissection, fertilised in vitro and embryos allowed to grow to the desired stage before fixation and storage. Methods for fixation, storage and in situ hybridisation were as previously described (Boorman and Shimeld, 2002a). Briefly, embryos were fixed with 4% paraformaldehyde (PFA) and stored at -20°C after gradual transfer to Methanol (*L. planeri*) or Ethanol (invertebrate embryos). Adult *Branchiostoma lanceolatum* were collected near Banyuls-sur-Mer, France by C. Patthey and spawning was induced by heat stimulation (Fuentes et al., 2007, 2004). Embryos were grown for 36hr at 19°C in natural sea water. Fixation was performed for 2hr on ice in 4% PFA in MOPS buffer containing 0.1M MOPS, 1mM EGTA, 2mM $MgSO_4$ and 500mM NaCl.

2.3.2 In Situ Hybridisation

Lamprey *in situ* hybridisation was performed as described (Lara-Ramírez et al., 2015), *C. intestinalis* and *B. lanceolatum in situ* hybridisation was performed as described (Boorman and Shimeld, 2002a). Lamprey embryo clearing was either by transferring to 25% then 50% glycerol in PBS, or by dehydrating with 25%, 50%, 75% and 100% ethanol and transferring to benzyl benzoate/benzyl alcohol 2:1 (BABB). Differential interference contract (DIC) microscopy was used to observe and photograph embryos (Zeiss Axioskop microscope).

2.4 Transgenics/transfection, imaging and controls

2.4.1 Vertebrate transgenics/DNA transfection

Lamprey transient transgenesis was performed in *L. planeri* (chapter 4) or *P. marinus* embryos (chapter 4 and 5) as previously described (Parker et al., 2014b, 2014a). Briefly, injection mixes consisting of 20ng μl^{-1} reporter plasmid, 1x CutSmart buffer (NEB), and 0.5U μl^{-1} I-SceI enzyme (NEB) in water were incubated at 37°C for 30 minutes and then micro-injected at a volume of approximately 2nl per embryo into lamprey embryos at the one-cell stage. Embryos were then raised and screened for GFP reporter expression. Transient transgenic reporter assays may generate mosaicism in reporter expression patterns, with variation in levels and domains between embryos. 100 embryos were screened for each construct at two stages (Tahara 25 and 27).

Chicken in-ovo electroporations were performed as described (Abelló et al., 2007), targeting the otic region at stage HH10 and allowing embryos to grow to HH15 and HH17. Ds red control plasmids with a ubiquitous promoter were co-electroporated with the GFP reporter constructs to visualise all transgenic cells. For Zebrafish transgenics, the Ac/Ds transposition system was used as described (Chong-Morrison et al., 2018), injecting zygotes and allowing embryos to grow to 24 hpf. Chicken embryo sections were made by embedding in gelatin and sectioning with a cryostat, followed by immunohistochemistry as described below to detect GFP. DsRed fluorescence persisted post-fixation and no further staining was necessary.

Representative GFP-expressing lamprey embryos were first imaged live to record GFP fluorescence, using a Zeiss SteREO Discovery V12 microscope and a Zeiss Axiocam MRm camera with AxioVision Rel 4.6 software. All embryos were fixed in 4% paraformaldehyde and immunostaining was carried out as described below. Before imaging, embryos were counterstained with DAPI.

2.4.2 Immunohistochemistry

Fixed embryos were initially washed 3 times in PBS-TritonX100 (PBS with 0.5% Triton X100) for 30 minutes, washed 2 times in block solution (PBS-TritonX100 with 10% heat treated sheep serum) for 30 minutes and blocked in block solution for longer than 2 hours, and up to overnight. Embryos were then incubated overnight with the primary antibody solution followed by 5 PBS-TritonX100 and 2 block solution washes for 30 minutes. Secondary antibody solution incubation was then carried out for longer than 4 hours up to overnight, followed by 5 PBS-TritonX100 washes for 30 minutes each. Embryos were finally transferred to 25% then 50% glycerol in PBS to clear and visualised in an optical fluorescent (Zeiss Axioskop) or confocal (Olympus FV1000) microscope.

Sections were initially incubated in PBS at 45°C for 30 minutes to remove gelatin. Then 5 washes in PBS-TritonX100 for 15 minutes were followed by blocking for 1 hour at 4°C and ovenight incubation in primary antibody, 5 washes in PBS-TritonX100 for 15 minutes, incubation with secondary antibody for more than 2 hours and 5 washes in PBS-TritonX100 for 15 minutes before transferring to glycerol and sealing.

Antibodies used were: Chicken Polyclonal Anti-GFP antibody (Abcam AB13970) (used in lamprey and zebrafish) visualised with Goat Anti-Chicken IgY H&L Alexa Fluor® 488 (Abcam AB150169), mouse anti-GFP (Invitrogen A-11120) (used in chicken) visualised with Goat anti-Mouse IgG H+L Alexa Fluor® 488 (Invitrogen A-11001), Mouse HuC/HuD Monoclonal Antibody (Invitrogen 16A11) visualised with anti-Mouse alexa594 (Abcam AB150116) and anti-acetylated Tubulin (Sigma T7451) visualised with anti-Mouse alexa594 (Abcam AB150116).

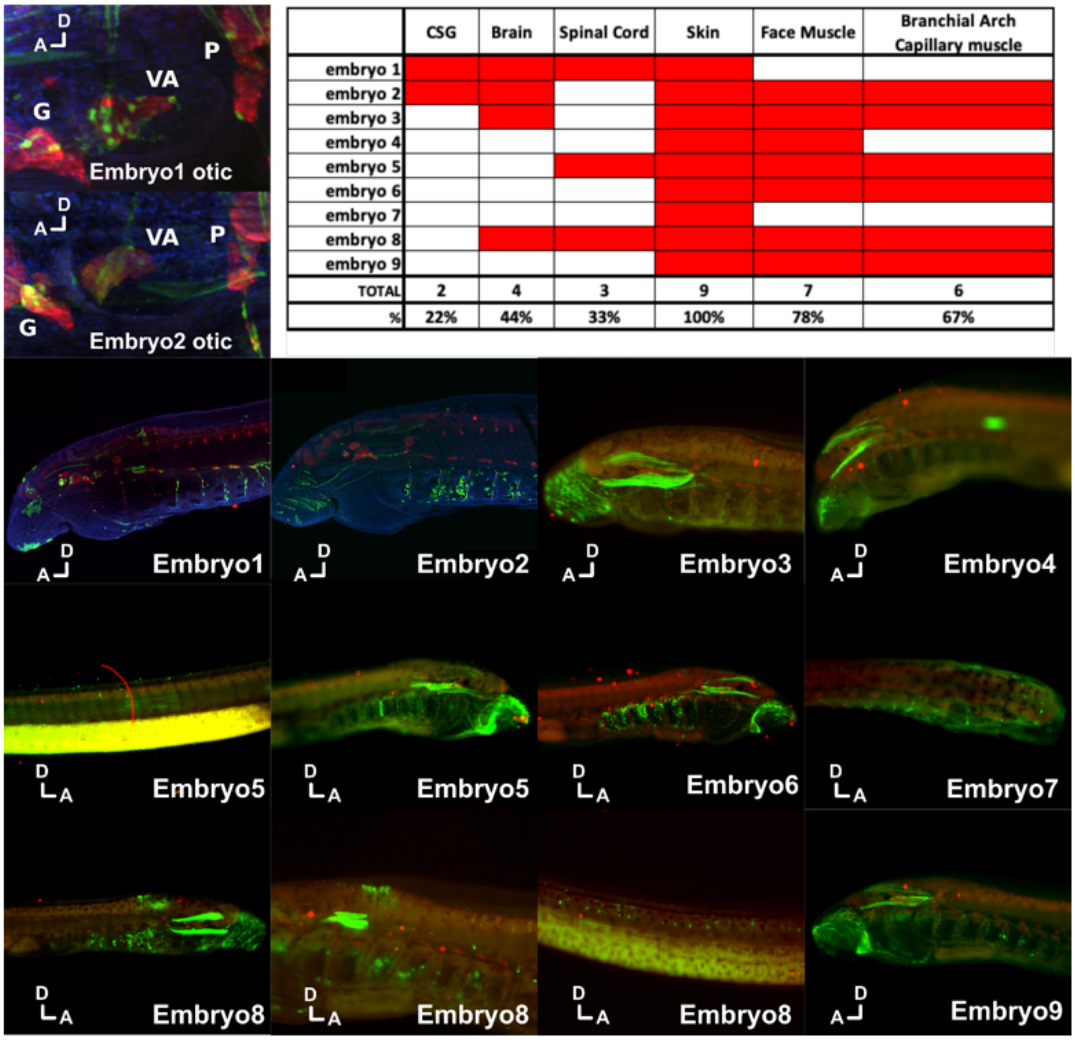
2.4.3 Image analysis

Analysis and processing of fluorescence microscopy data were carried out using imageJ v. 1.52g. Confocal or optical z stacks and 3D projects were created using maximum intensity projection. Some images are composites of multiple individual images of the same embryo focused at slightly different levels to follow expressing cell populations as the embryo moved in and out of focus. These were assembled in PowerPoint by overlapping individual images (Figure 4.1 C, D).

2.4.4 Krt4 background

Confocal microscopy was able to reveal GFP expressing cells not possible to image in live embryos. Since lamprey transgenesis is a relatively new technique and previous studies have not assessed levels of background at this resolution, I analysed reporter activity in embryos injected with the plasmid vector HLC-krt4 (Figure 2.1), focusing on ganglia and CNS expression that might overlap with

endogenous *ERR* or *Hmx* gene expression staining and confound interpretation. This revealed GFP expression in skin cells, head muscle and branchiomeric muscle. I also saw occasional expression in CNS and ganglia. Spinal cord expression was confined to isolated cells and distinct from the consistent column of expression generated by the *Hmx* enhancers (see Figure 5.5). Brain expression appeared in the dorsal hindbrain and midbrain and was also distinct from *Hmx* or *ERR* expression and enhancer activity. In two embryos (1 and 2 below: also in high magnification in top left focused on the otic area) close examination revealed scattered cells around some cranial ganglia, including a few co-expressing Hu/ELAV. These differed from those labelled with Hmx enhancers in that GFP staining did not penetrate into axons.



	CSG	Brain	Spinal Cord	Skin	Face Muscle	Branchial Arch Capillary muscle
embryo 1						
embryo 2						
embryo 3						
embryo 4						
embryo 5						
embryo 6						
embryo 7						
embryo 8						
embryo 9						
TOTAL	2	4	3	9	7	6
%	22%	44%	33%	100%	78%	67%

Figure 2.1 Assessment of background deriving from the vector used to generate lamprey transgenics. Embryos were injected with vector only (which includes the zebrafish krt4 minimal promoter and reporter gene but no cloned enhancer), allowed to develop then fixed and labelled for DNA (DAPI, blue), GFP (green) and Hu/ELAV (red) before analysis by confocal microscopy. Each embryo was scored for expression in multiple tissues, as shown in the table at the top right of the picture, with a focus on tissues overlapping with *Hmx* or *ERR* expression. D and A indicate dorsal and anterior orientations for each image. CSG, cranial sensory ganglia. G, geniculate ganglion. VA, vestibuloacoustic ganglion. P, petrosal ganglion.

2.5 *Ciona* transgenics and bulk embryo RNAseq

2.5.1 *Ciona Hmx* ORF cloning

The plasmid vector containing the *epiB* promoter driving *YFP* (*epiB::YFP*) was kindly provided by Bob Zeller (Joyce Tang et al., 2013) . The full *Hmx* open reading frame was amplified by PCR and cloned downstream of the *epiB* promoter, replacing *YFP* and creating *epiB::Hmx. Ciona intestinalis Hmx* was amplified in two sections, a 5' region using the primers TAAAATAGTAAAATGGTACCTATGACGTCACTGTGCCAATTG and TTCCCCTTCTGACGTAGGGA, and a 3' section using the primers TCCCTACGTCAGAAGGGGAAG and ACCGGCGCTCAGCTGGAATTATGATTGTCTCACACCACGGAA. This resulted in two fragments: each had a homologous arm overlapping the other fragment and a homologous arm overlapping one end of the vector digested with KpnI and EcoRI. The Cold Fusion system (System Biosciences) was used to insert these into the vector via recombination, fusing the 5' end of the resulting full *ciHMX* ORF with the 3' end of the *epiB* promoter. Integrity of the resulting construct was confirmed with sequencing.

2.5.2 *Ciona* Transgenics and RNAseq

Constructs were electroporated into C*. intestinalis* zygotes as previously described (Corbo et al., 1997). I first confirmed that these constructs drove their respective

transgenes into the epidermis as expected, using YFP live imaging and *Hmx* in situ hybridisation respectively. I then electroporated parallel batches with either *epiB::YFP* only (control) or *epiB::YFP* and *epiB::Hmx* (*Hmx* overexpression) constructs. Embryos were grown to the tailbud stage when YFP was visible, identifying transgenic embryos which were then manually selected and processed for RNA extraction. Three full biological replicates were performed on embryo batches derived from different fertilisations, with each biological replicate combining RNA from at least 50 individual embryos. All six RNA samples were sequenced by Illumina HiSeq4000 following polyA selection, yielding approximately 28 million paired end 75bp reads per sample.

2.5.3 RNAseq data processing and differential expression analysis

For differential gene expression analysis, fastQC was used to assess sequencing quality, Trimmomatic (Bolger et al., 2014) to trim off adapters and Sickle (Joshi and Fass, 2011) to trim low quality reads. Remaining reads were then mapped to the *Ciona robusta* genome (KH2012) using STAR (Dobin et al., 2013). Differential expression analysis was carried out using the DESeq2 R package (Love et al., 2014), using an adjusted p value threshold of 0.01. Finally, a minimum FPKM threshold of 2 was also applied to exclude very lowly expressed transcripts. This yielded a list of genes significantly (adjusted $p<0.01$) up or downregulated in the *Hmx* overexpression treatment compared to the control. Gene lists deriving from *C. intestinalis* single cell sequencing were extracted from the supplementary files of published literature (Horie et al., 2018; Sharma et al., 2019) , and cross-correlated with the up and down regulated gene lists to provide the annotation of data shown

in Figure 5.7B. The list of annotated differentially regulated target genes produced by this analysis is included in section S7 of the supplementary appendix.

2.5.4 *Ciona* CNE transgenics (Carried out by Alessandro Pennati)

To test CNE activity in *Ciona*, the 2kb 5' to *C. intestinalis Hmx*, lamprey uCNE or lamprey dCNE were cloned into the reporter vector pCES (Harafuji et al., 2002) using the primers shown in Table 2.2. Constructs were electroporated into *C. intestinalis* zygotes as above, and embryos stained for β-galactosidase activity as described (Chen et al., 2014).

	Region	Forward Primer	Reverse Primer
Ciona CNE	Ciona CNE	CACTATAGAACTCGACGTCCAACTAAATCGTCGAAG	CTGACCATGGGGATCCTTTTCGGCGGTTCGTTCTC
	Lamprey uCNE into pCES	CACTATAGAACTCGAGCAGCAAGCGATGTATGC	GAGTCGACCTGCAGGCGAGTACGCGAAGAGGTGG
	Lamprey dCNE into pCES	CACTATAGAACTCGACATCATCAGCGCGCGTATTTAAG	GAGTCGACCTGCAGGTTCCTGCACACAACTCTCGC

Table 2.2. PCR primers used to clone CNE sequences in the pCES vector.

Chapter 3 The Molecular Identity of Cranial Sensory Ganglia

Work Declaration

Experimental work and data analysis presented in this chapter have been performed by myself (Vasileios Papadogiannis).

3.1 Chapter Introduction

Vertebrates share morphological characters that are unique to the group, discussed in better detail in chapter 1. The vertebrate head comprises many innovations, suggested to have evolved as part of an adaptation to a predatory lifestyle (Gans and Northcutt, 1983; Northcutt and Gans, 1983). All vertebrates share the conserved tripartite brain and CSG that develop from placodes and neural crest (Watanabe et al., 2008). While invertebrate chordates lack these novelties, increasing evidence supports that many of these new structures evolved by elaboration of older chordate characters (Holland, 2009; Shimeld and Donoghue, 2012). Anatomical studies were the first to highlight the characters shared by all vertebrates, as well as homologies between those characters and structures in other chordates. However, the dissection of developmental gene expression really allowed for in depth comparisons, even at cell type level. In the case of sensory system evolution, acquiring a suite of placodal marker genes was crucial in this search, as covered more extensively in sections 1.4 and 1.5.

A highly conserved molecular network drives the formation of the derivatives of the neural plate (CNS) and the neural plate border (PNS) in vertebrates including lampreys (Murakami et al., 2004; Sugahara et al., 2016). Early PNS development, involving placodes and neural crest, is well described, as seen in section 1.3. Determination and differentiation of CSG has been analysed to a good degree, while late CSG identity is less understood (Blentic et al., 2011). CSG neurons derive from neurogenic placodes, which develop under a conserved network in all vertebrates (McCabe and Bronner-Fraser, 2009; Schlosser, 2006). Neural crest cells also migrate and colonise emerging CSG (Schlosser et al., 2014). CSG neuronal progenitors, as they emerge in placodes, express neural and sensory marker genes that are conserved across taxa (Grocott et al., 2012). Also, as seen in section 1.3, placodal markers are gradually restricted to particular placodal subsets, but a degree of overlap is always retained. In this way, the molecular identity of different placodes is characterised by combinations of marker genes.

Different CSG may comprise neuronal populations of multiple types and innervate sensory systems with diverse functions and cell types. The identity of these neurons is determined in a multi-step process. Early on, overlapping transcription factors are expressed sequentially in multiple placodes, suggesting a combinatorial and gradual mode of fate determination (Schlosser, 2006). Genes with a later onset of expression and roles in differentiation eventually complement earlier markers in developing CSG, as seen in section 1.3.3. Thus, the above could suggest that the molecular identity of CSG may be characterised by a pattern of overlapping transcription factor domains comparable to what is seen at placodal stages.

3.1.1 Experimental Approach

The work covered in this chapter aims to identify conserved CSG markers and interrogate their specificity. The starting point for this comparison was a dataset of marker genes of developing CSG neurons in chicken embryos (Patthey et al., 2016). In that study, transcriptomes were obtained from dissected, FACS-sorted neurons from trigeminal, vestibulo-acoustic and epibranchial chick embryo ganglia. Differential expression analysis then allowed for a selection of markers with high specificity in each ganglion at the time of dissection. A subset of differentially expressed genes was finally short-listed and validated via in situ hybridisation, mainly selecting for transcription factors involved in PNS development.

Here, the expression of candidate markers is explored in lamprey embryos, to assess pan-vertebrate conservation. The lamprey was selected as an evolutionary model because of its phylogenetic position as part of the earliest diverging vertebrate group. As jawed and jawless vertebrates split more than 450 million years ago, conservation of expression between these lineages is informative on the ancestral state and on inferring what patterns consist vertebrate characters (Shimeld and Donoghue, 2012). Thus, in principle, it can be inferred that CSG marker expression shared between lampreys and a model system like the chick should be primitive for vertebrates.

3.1.2 Aims and Marker Selection

Two complementary subjects are addressed here. First, an effort is made to characterise the conserved molecular identity of developing CSG. Essentially, this allows us to infer ancestral gene expression. Second, marker specificity in each ganglion is interrogated. For this purpose, the CSG markers assessed here were selected from the aforementioned shortlisted and validated dataset by *Patthey* and colleagues (2016). As with the shortlisting process initially used in the preceding study, different criteria were considered. Most importantly, markers were chosen that showed high strength and specificity of expression in marked chick ganglia, as assessed by transcriptome data. A literature search was carried out to select genes with published information on CSG expression and roles in sensory or neural development. Finally, transcription factors were favoured, for their potential in regulating other downstream targets. Through this process, a combination of trigeminal, VA and epibranchial markers were chosen that are summarised in Fig.3.1. A short description and literature review for each of these genes follows.

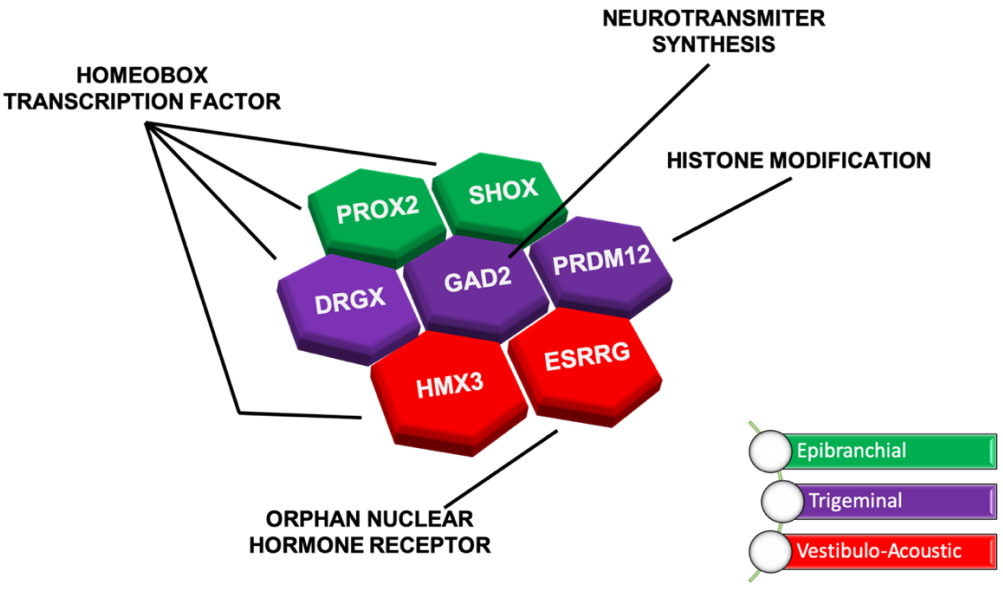


Figure 3.1. Selected chicken CSG markers screened for embryonic expression in lamprey in current study. Trigeminal Markers: *PRDM12, GAD2, DRGX*. Vestibulo-Acoustic Markers: *Hmx3, ESRRG(ERRγ)*. Epibranchial Markers: *Prox2, Shox*.

PRDM12 (Chicken Trigeminal Marker)

PR domain zinc finger protein 12 (*PRDM12)* is a member of the large PR domain containing transcription factor family, which has its origins in the metazoan common ancestor. All proteins of the family share a PRDF1 RIZ (PR) domain close to the N terminal, as well as a variable number of zinc fingers, while other domains can also be found in specific members. While PR domains are related to the catalytic SET domains with histone methyltransferase activity, most PRDM proteins lack intrinsic enzymatic activity. However, they often have the potential to recruit other histone-modifying enzymes to carry out their functions. In addition, they are also able to bind to DNA via their zinc finger motifs (Hohenauer and Moore, 2012). In vertebrates, many PRDM genes expanded and had subsequent structural rearrangements (Fumasoni et al., 2007). Most family members have been implicated in nervous system development, while *PRDM8, PRDM12, PRDM13* and *PRDM14* have roles in spinal cord DV pattering (Zannino and Sagerström, 2015). *PRDM12*, which encodes for a PR domain and three zinc finger motifs, is expressed in mouse and *Xenopus* brain and spinal cord where it has important roles in interneuron specification (Chen et al., 2015; Nagy et al., 2015; Thelie et al., 2015). *PRDM12* has a crucial role in promoting the specification of the pre-placodal ectoderm and suppressing neural crest identity (Matsukawa et al., 2015). The gene is important for sensory cell development and nociception in human, mouse and *Xenopus*, while late expression has been described in developing trigeminal and

DRG in mouse, *Xenopus* and zebrafish (Chen et al., 2015; Nagy et al., 2015; Zannino et al., 2014). In zebrafish, possible expression in other CSG has been noted, but it looks more restricted than in trigeminal and DRG (Zannino et al., 2014).

DRGX (Chicken Trigeminal Marker)

Dorsal Root Ganglia Homeobox, also known as DRG11 or PRRXl1, is a homeobox transcription factor of the PRD class. The protein has a paired related homeodomain, as well as an OAR domain close to the C-terminus, thought to act as a cofactor binding site. The gene has been shown to be expressed in the mouse and zebrafish CNS and PNS during embryogenesis (McCormick et al., 2007; Rebelo et al., 2007; Wang et al., 2007). Trigeminal expression commences early in zebrafish and persists, followed by dorsal root ganglion (DRG) and dorsal spinal cord expression. However, fainter DRGX expression is also seen in other CSG in zebrafish later in development (McCormick et al., 2007). Expression in trigeminal and DRG is also seen in mouse embryos, while similar to zebrafish, DRGX protein is also detected in epibranchial ganglia in mouse (Rebelo et al., 2007). Nevertheless, the gene has been shown to be crucial for mesencephalic trigeminal nucleus development in mice and is associated with trigeminal cell death, mechanosensation and nociception defects in knockout animals (Jacquin et al., 2008; Wang et al., 2007). Hence, it was considered a good candidate to test, in conjunction with trigeminal specificity in the chicken transcriptome dataset. Expression has also been documented in midbrain and hindbrain sensory neurons and transiently in spinal cord Rohon-Beard neurons (McCormick et al., 2007; Rebelo et al., 2006; Wang et al., 2007) .

GAD2 (Chicken Trigeminal Marker)

Unlike other transcription factor candidates, glutamate decarboxylase is an enzyme that takes part in gamma-aminobutyric acid (GABA) production, via the decarboxylation of glutamate. It is considered a strong marker of GABAergic neurons and vertebrates have two isoenzymes, GAD67 and GAD65, encoded by two separate genes *GAD1* and *GAD2* respectively. GAD65 has been previously correlated with nociception in the face. Thus the gene was selected as a candidate marker of trigeminal GABAergic neurons in particular (Bosma et al., 1999; Sun et al., 2005; Vit et al., 2009).

ERRγ (ESRRG) (Chicken Vestibuloacoustic Marker)

Estrogen related receptors (ERR) are orphan steroid hormone receptors similar in amino acid sequence to the estrogen receptors (ER). Four *ERR* paralogs have been identified to be ancestral in jawed vertebrates, Estrogen Related Receptor alpha (*ERRα)*, beta (*ERRβ)*, gamma (*ERRγ)* and delta (*ERRδ)*, while lineage specific losses or gains are seen in different vertebrate classes. Nevertheless, at least one copy of each of *ERRβ and ERRγ* is always present in all gnathostomes (Bertrand et al., 2004; Thornton et al., 2003). No known natural ligand has been identified binding to ERR; instead the protein is constitutively active, acting as a transcription factor by binding to DNA and regulating gene expression (Tremblay and Giguère, 2007). Family members are expressed in different embryo and adult tissues, including CNS, muscle and kidney (Bardet et al., 2004; Heard et al., 2000; Süsens et al., 2000). *ERRγ* is expressed in the VA and trigeminal in embryos, as well as in lateral line ganglia, the epiphysis and hindbrain in zebrafish (Bardet et al., 2005, 2004). The gene has also been associated to hearing loss in mice and

humans (Nolan et al., 2013). In zebrafish, *ERRα* is expressed in retina, hindbrain, epiphysis, muscle and pronephros and *ERRβ* in diencephalon, hindbrain, epiphysis, pronephros and spinal cord (Bardet et al., 2005, 2004). It should be noted that different names have been used to refer to ERR genes in different organisms, with *ESRRA*, *ESRRB*, *ESRRG and ESRRG-like* and *ERRα*, *ERRβ*, *ERRγ and ERRδ* respectively referring to the same genes. It is similar with the related ERs, with *ESR1*, *ER1* or *ER* referring to the gene coding for the Estrogen Receptor and *ESR2* or *ER2* referring to the gene coding for the Estrogen Receptor 2.

Hmx3 (Chicken vestibuloacoustic Marker)

The *Hmx* family, also known as *NKX5* or *NK5*, is part of the NK homeobox transcription factor subclass. *NK* and *Hmx* genes are present in all Metazoa, with a representative of the family described in calcisponges, *Trichoplax*, *Drosophila*, sea urchin and amphioxus (Fortunato et al., 2014; Garcia-Fernàndez, 2005; P. W. H. Holland, 2013; Martinez and Davidson, 1997; Monteiro et al., 2006; Wang et al., 2000). In gnathostomes four members have been described: *Hmx*1 or *NKX5.3*, *Hmx*2 or *NKX5.2*, *Hmx*3 or *NKX5.1* and *SOHo* (Sensory Organ HOmeobox), while *SOHo* has been lost in mammals (Adamska et al., 2001). *NK* genes are important for the specification of neural and mesodermal lineages across the Bilateria, with different members expressed during CNS and heart development (Garcia-Fernàndez, 2005). The *Hmx* family has a role in CNS development in both *Drosophila* and vertebrates, with specific early expression in neural tissue (Wang et al., 2000). In addition to the CNS, family members are expressed in CSG in *Medaka*, zebrafish, *Xenopus* and mouse (Adamska et al., 2001; Feng and Xu, 2010; Kelly and El-Hodiri, 2016; Wang et al., 2004). In jawed vertebrates all four

Hmx paralogs are expressed in CSG, with their roles in the developing VA and lateral line ganglia better studied (Feng and Xu, 2010; W Wang et al., 1998; Wang et al., 2004, 2001; Weidong Wang et al., 1998). Hmx genes are expressed early in differentiating sensory cells in gnathostomes and the family appears to be crucial for sensory cell specification. In both zebrafish and mouse *Hmx3* knockouts, the delaminating vestibular neurons of the VA fail to differentiate, while the zebrafish lateral line system is also affected (Feng and Xu, 2010; Wang et al., 2004). *Hmx3* was selected as a candidate, because of its expression and role during ear development, as well as the fact that different *Hmx* family members are expressed in various CSG. Hence, it remained possible the family had broader expression in the vertebrate ancestor, with one paralog acquiring VA specificity.

Shox (Chicken epibranchial Marker)

Short stature homeobox (Shox) genes are a relatively understudied homeodomain containing family. *Shox2* is described as expressed in Trigeminal and DRG in mice and has been previously shown to exhibit roles in cerebellar, facial motor nucleus, facial nerve and mechanosensory nerve development, as well as being associated with motor coordination (Rosin et al., 2015b, 2015a; Scott et al., 2011). Therefore, it consists an interesting candidate, as *Shox* is a strong epibranchial marker in chicken, while *Shox2* is seen in trigeminal and DRG in mice.

Prox2 (Chicken epibranchial Marker)

Prospero-related homeobox genes are known to play important roles in embryogenesis in vertebrates, while the *Drosophila* homolog *Prospero* is expressed during neuronal cell differentiation. Both vertebrate *Prox* genes are expressed in

nervous system tissues during development and in the retina. *Prox2* is known to be expressed in the developing epibranchial ganglia in mouse and zebrafish (Nishijima and Ohtoshi, 2006; Pistocchi et al., 2008).

3.1.3 Work summary

The work presented and discussed in this chapter provides evidence of conserved molecular identity of CSG across all vertebrates. Family paralogs of selected candidates were mapped in lamprey and their embryonic expression was assessed. No embryonic expression was identified for one family (*DRGX*) and no CSG expression was detected for two families (*GAD*, *Prox*). At least one paralog from four of these families (*ERR*, *PRDM12*, *Hmx* and *Shox*) shows CSG expression. What is more, expression is conserved in the respective ganglia they mark in chicken. However, specificity for these sites is not confirmed, with all families (except for *ERR*) expressed in at least one other ganglion. This characterisation of conserved CSG identity is the first step in a search for the origins of CSG sensory cell types that unfolds in chapter 4 and chapter 5.

3.2 Results

Marker genes were identified, cloned and their expression was assessed in brook lamprey embryos. Transcripts matching each gene family were retrieved from transcriptomes of the brook lamprey (*Lampetra planeri*), Japanese lamprey (*Lethenteron camtschaticum*) and sea lamprey (*Petromyzon marinus*). A reciprocal BLAST approach was employed, using the human ortholog (longest RefSeq isoform) for each chick marker as a query. Sequences were mapped to the Japanese lamprey genome to retrieve gene loci. Alignments with jawed vertebrate paralogs and molecular phylogenies were built to test the orthology of identified family members. Expression was assessed in Tahara stage 23-26 brook lamprey embryos via in situ hybridisation. A full description of methods is provided in chapter 2. A list of identified loci for gene families studied is provided in Table 3.1, which includes loci positions in the Japanese lamprey genome and a summary of mapped transcripts. Full transcript sequences are included in the supplementary appendix.

GENE	Japanese Lamprey Scaffold	Exon Start-End (exons separated by comma)	Lethenteron Transcripts	Petromyzon Transcripts	Lampetra Transcripts
ERRA	KE993693	1791057-1791340,1793741-1793862,1810156-1810283,1811841-1812007,1824702-1825001,1828303-1828651,2499899-2500096,2500945-2501086,2501304-2501453,2501887-2502130,2505287-2505358,2507384-2507503,2507689-2507873	comp164469_c5_seq8	Locus_85892_Transcript_1_2_Confidence_0	TRINITY_DN42311_c3_g1_i1, TRINITY_DN42311_c1_g1_i2
ERRB	KE993744	843865-844112,857262-857536,916856-917019,966749-967003,974208-974429	comp164469_c5_seq15, comp164469_c5_seq22	ENSPMAT00000002705	TRINITY_DN42311_c1_g1_i2
ERRC	KE993792	904283-904545,910377-910660,912602-912768,912932-913091,933312-933441,937196-937641	comp164469_c5_seq18, comp164469_c5_seq9, comp164469_c5_seq16	Locus_22390_Transcript_2_2_Confidence_0	TRINITY_DN68209_c0_g1_i1
ERRD	KE993920	783074-783324,790370-790460,795345-795461,797807-797912			
HMXA	KE993680	3621732-3623066	comp133801_c0_seq1	-	TRINITY_DN47845_c12_g2_i1
HMXB	KE993680	3601305-3602299,3602831-3603375	comp139112_c0_seq2	-	TRINITY_DN23440_c0_g1_i1
HMXC	KE993680	3583350-3583943,3585649-3586230	-	ENSPMAT00000002780	TRINITY_DN46446_c0_g1_i1
DRGX	KE993696	889253-889359,890184-890282,903033-903119,908264-908301	comp189895_c0_seq1	-	TRINITY_DN98271_c0_g1_i1
PRDM12	KE993733	2516019-2516637,2527260-2527472,2529183-2529340,2531377-2531493,2533841-2533938	comp102901_c0_seq1, comp102901_c1_seq2	Locus_133939_Transcript_1_1_Confidence_1	TRINITY_DN6857_c0_g1_i1, TRINITY_DN11919_c0_g1_i1
GAD	KE993764	982961-983086,1001737-1002012,1010940-1011004,1019863-1020073,1020297-1020544,1023597-1023692,1026608-1026721,1027352-1027575,1028437-1028517,1029273-1029330,1029743-1029868,1030926-1030991,1032502-1032583,1033387-1033539,1034364-1034473,1034701-1035565	comp164555_c0_seq1, comp164555_c0_seq2	Locus_8851_Transcript_1_1_Confidence_1	TRINITY_DN40566_c2_g1_i1, TRINITY_DN49542_c0_g1_i1
SHOXA	KE994509	16045-16188,19695-19767,55976-56030	comp313800_c0_seq1	-	TRINITY_DN69692_c0_g1_i1
SHOXB	KE996193	13816-13846,14942-15034,19647-19709	-	-	TRINITY_DN87225_c0_g1_i1
SHOXC	KE993821	1481516-1481632,1484041-1484119,1484453-1484514	-	ENSPMAT00000009315	-
PROXA	KE993693	1423105-1423385,1425457-1428550,1430491-1431228,1432083-1432281,1433969-1435062	comp164363_c3_seq7, comp164363_c3_seq6, comp164363_c3_seq5	Locus_57986_Transcript_1_1_Confidence_1, Locus_1_Transcript_4008_59 68_Confidence_1, Locus_1_Transcript_4007_59 68_Confidence_1	TRINITY_DN43202_c0_g1_i1
PROXB	KE993810	1247247-1247882,1269872-1270073,1277916-1278028,1290605-1290839,1292136-1292455	comp240049_c0_seq1, comp110109_c0_seq1	Locus_169599_Transcript_1_1_Confidence_1, Locus_1_Transcript_4007_59 68_Confidence_1, Locus_153806_Transcript_1_1_Confidence_1	TRINITY_DN86424_c0_g1_i1, TRINITY_DN43202_c0_g1_i2
PROXC	KE993797	1347242-1347487,1348122-1348435,1372254-1372362,1382488-1382709,1382723-1382789,1394933-1395363	comp17112_c0_seq1, comp208620_c0_seq1, comp152513_c0_seq2	ENSPMAT00000007312, ENSPMAT00000007316	TRINITY_DN65978_c0_g1_i1, TRINITY_DN88612_c0_g1_i1
PROXD	KE993801	135472-135723,137779-137993,189514-189624,215131-215331,249927-250199	comp320497_c0_seq1, comp152513_c0_seq1	Locus_124952_Transcript_1_1_Confidence_1	TRINITY_DN16830_c1_g1_i1, TRINITY_DN93079_c0_g1_i1

Table 3.1 New lamprey loci. Candidate CSG marker gene loci in the Japanese lamprey (*L. camtschaticum)* genome. Selected transcripts matching each locus are provided, retrieved from transcriptomes of the brook lamprey (*Lampetra planeri*), Japanese lamprey (*Lethenteron camtschaticum*) and sea lamprey (*Petromyzon marinus*). Details on the identification method are provided in chapter 2. Full transcript sequences are included in the supplementary appendix.

3.2.1 Genes without identified CSG expression

DRGX

One locus corresponding to *DRGX* was found (protein sequence included in homeodomain phylogeny presented in Fig. 3.11). No expression was seen for *DRGX* in embryos of any stage. The restricted probe size of 216bp that could be made may have impacted the ability to detect expression. Only two transcripts could be mapped to this locus (table 3.1). Using other vertebrate proteins was only able to identify the conserved homeobox sequence, already part of the sequence mapped by these transcripts. Attempts to design primers to expand the cloned region outside this 216bp window also failed.

GAD2

A single lamprey *GAD* gene was identified, with the protein grouping robustly with the gnathostome GAD1 sequences (93 bootstrap in Fig 3.2). Expression of lamprey *GAD* was seen along the central nervous system, in the brain and spinal cord (Fig 3.3C). No signal was detected in early embryos, with expression becoming evident after stage 24 (Fig 3.3B). No CSG expression was seen in the stages screened here.

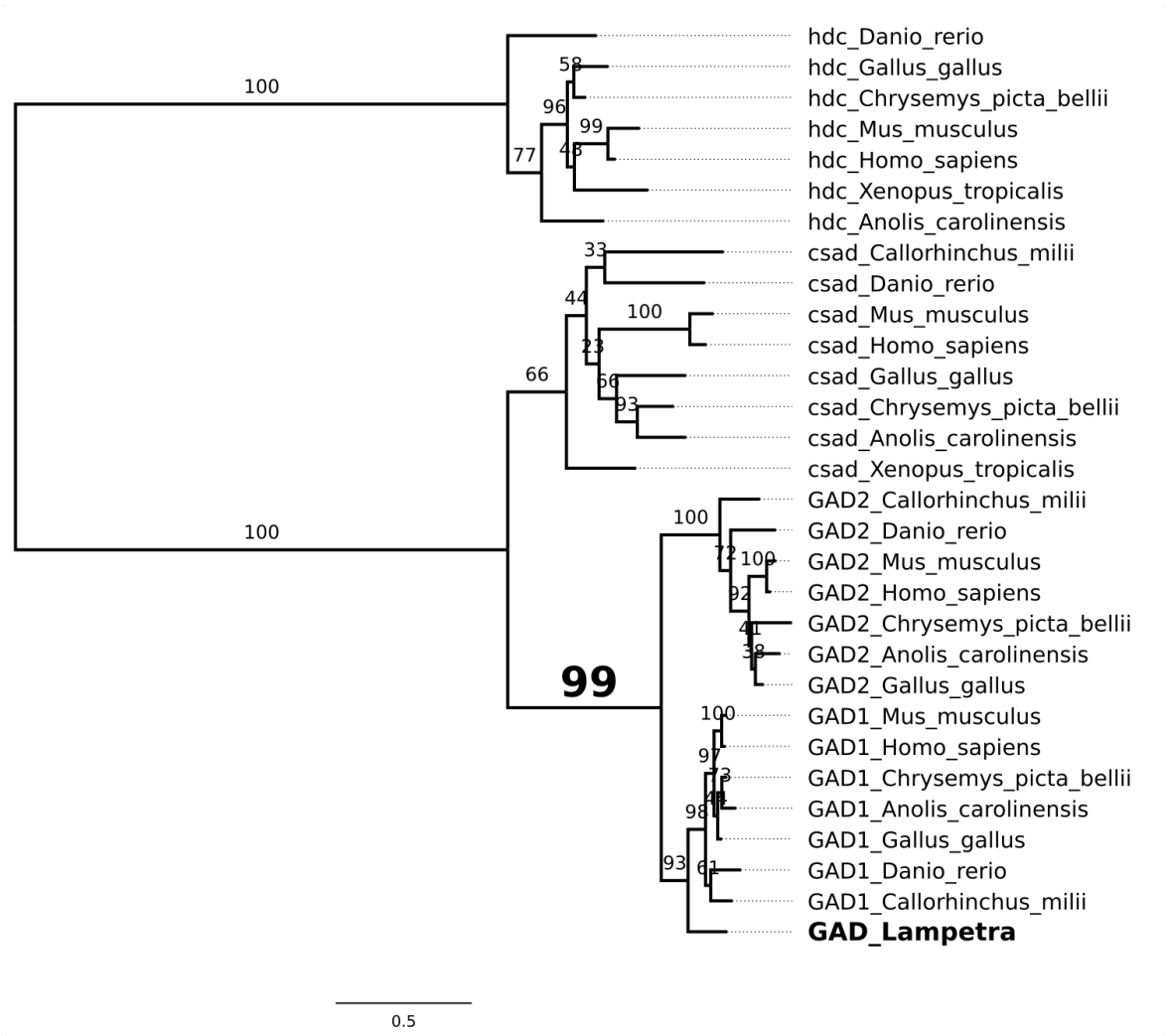


Figure 3.2. Maximum Likelihood Tree (JTT, 1000 bootstrap) for GAD. Histidine Decarboxylase (HDC) and Cysteine Sulfinic Acid (CSAD) proteins were used as outgroups. CSAD, HDC, GAD1 and GAD2 amino acid sequences were obtained from various vertebrates. As seen in the tree, GAD sequences formed a strongly supported branch (99% support), with further strong grouping for vertebrate GAD2 (100%) and GAD1 (93%) sequences. The single identified lamprey GAD sequence groups strongly within the GAD1 clade.



Figure 3.3. *GAD* expression in lamprey embryos. Signal was not seen in embryos younger than stage 23 **(A)** and became more widespread after stage 24 **(B)**. By stage 25, expression was seen in different areas in CNS **(C)**.

Prox

Four *Prox* loci were found in lamprey. Lamprey Prox protein sequences were identified by the extremely conserved Prox homeodomain (Fig 3.4). Embryonic expression was found for *ProxA*, with dispersed staining around the head (Fig 3.5). No expression was seen for any other *Prox* paralog. The small size of the region for *ProxB* that could be cloned may have limited the potential of the *ProxB* RNA probe.

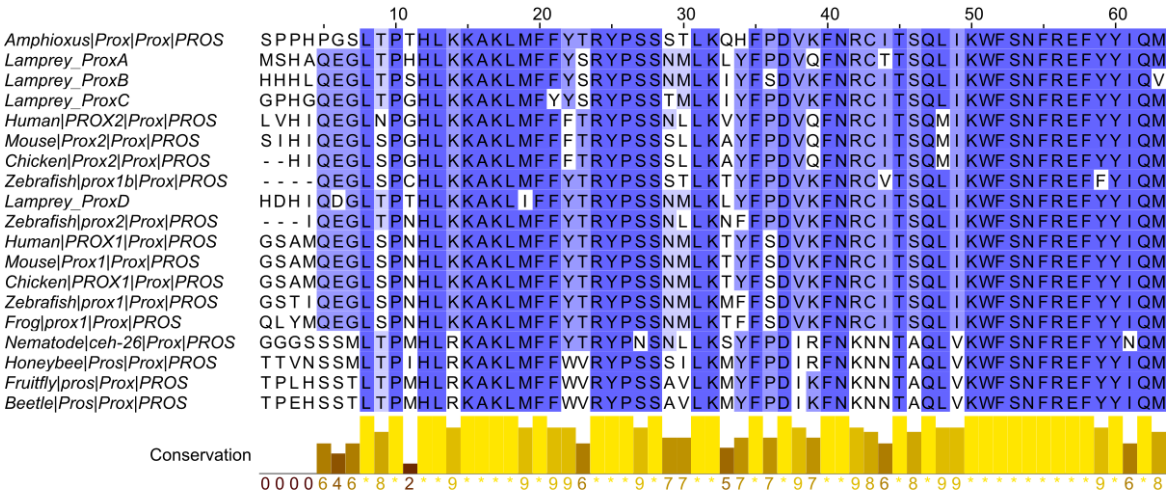


Figure 3.4. Alignment of Prox homeodomain sequences from different animals. Lamprey Prox candidates are clearly identified by the extremely conserved Prox homeodomain sequence.



Figure 3.5. *ProxA* expression in lamprey embryos. Expression is already seen at stage 24 **(A)** in areas of ectoderm in the head, with more widespread expression becoming more prominent in ventral areas by stage 25 **(B)** and extending further back in the body by stage 26 **(C)**. No expression in cranial sensory ganglia was seen.

3.2.2 *PRDM12* expression in trigeminal and geniculate

A single *PRDM12* lamprey paralog was identified, encoding for a PR domain with high sequence similarity to the gnathostome *PRDM12* PR domain. Sequences of other *PRDM* family members were sampled from different vertebrates, representing different clades of the family tree (Fumasoni et al., 2007), from *PRDM14 (*most closely related to *PRDM12) to PRDM13 (*most distantly related to *PRDM12*). The phylogeny built confirms the identity of identified lamprey paralog as *PRDM12* (Fig3.6).

L. planeri PRDM12 is expressed in CNS and PNS at stage 23. Expression in CNS is seen in midbrain, hindbrain and spinal cord (Fig. 3.7). Expression in the trigeminal is seen from an early stage, with strong signal in the developing ganglion at stage 23 (Fig. 3.7A). Furthermore, expression is also evident in DRG (Fig.3.7D), as well as in the geniculate and in the otic vesicle (Fig. 3.7E).

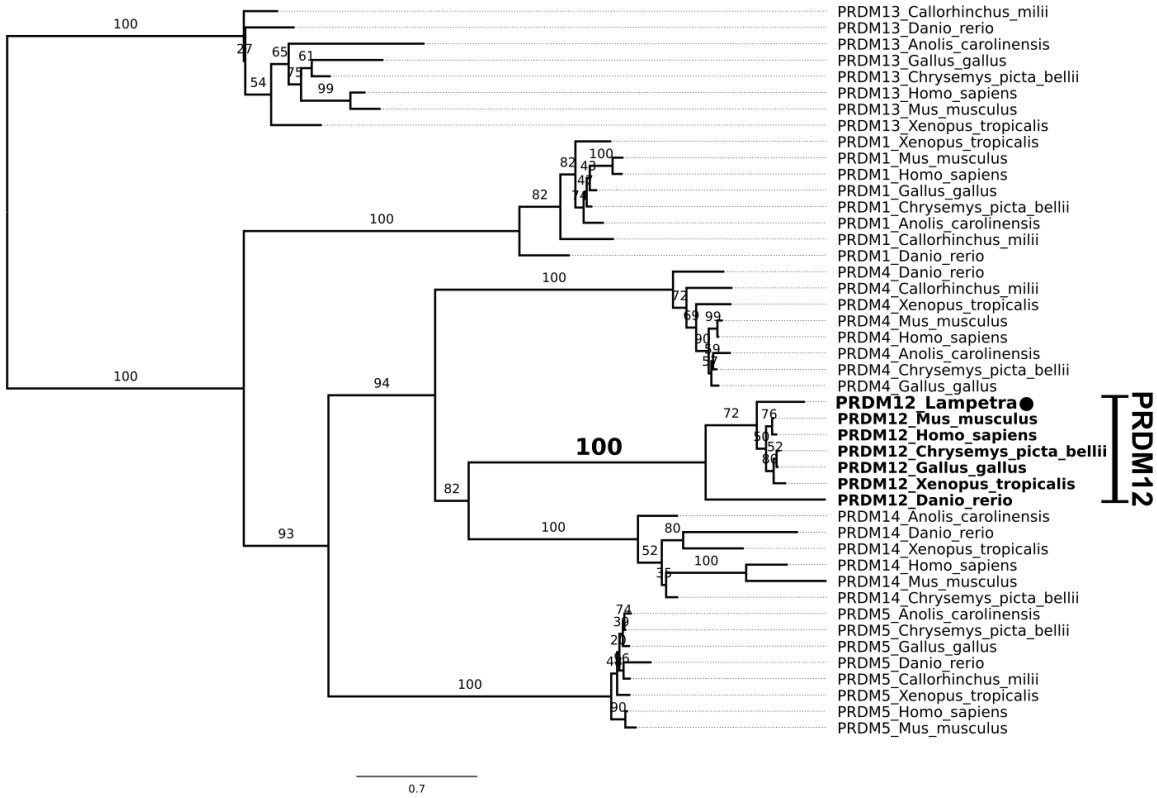


Figure 3.6. PRDM12 phylogeny (maximum likelihood analysis, JTT, 1000 bootstrap). Vertebrate PRDM1, PRDM4, PRDM5, PRDM13, PRDM14 and PRDM12 proteins were used to reconstruct the phylogeny, sampling across the PRDM family tree (Fumasoni et al., 2007). Lamprey PRDM12 groups within the strongly supported PRDM12 clade (100 bootstrap).

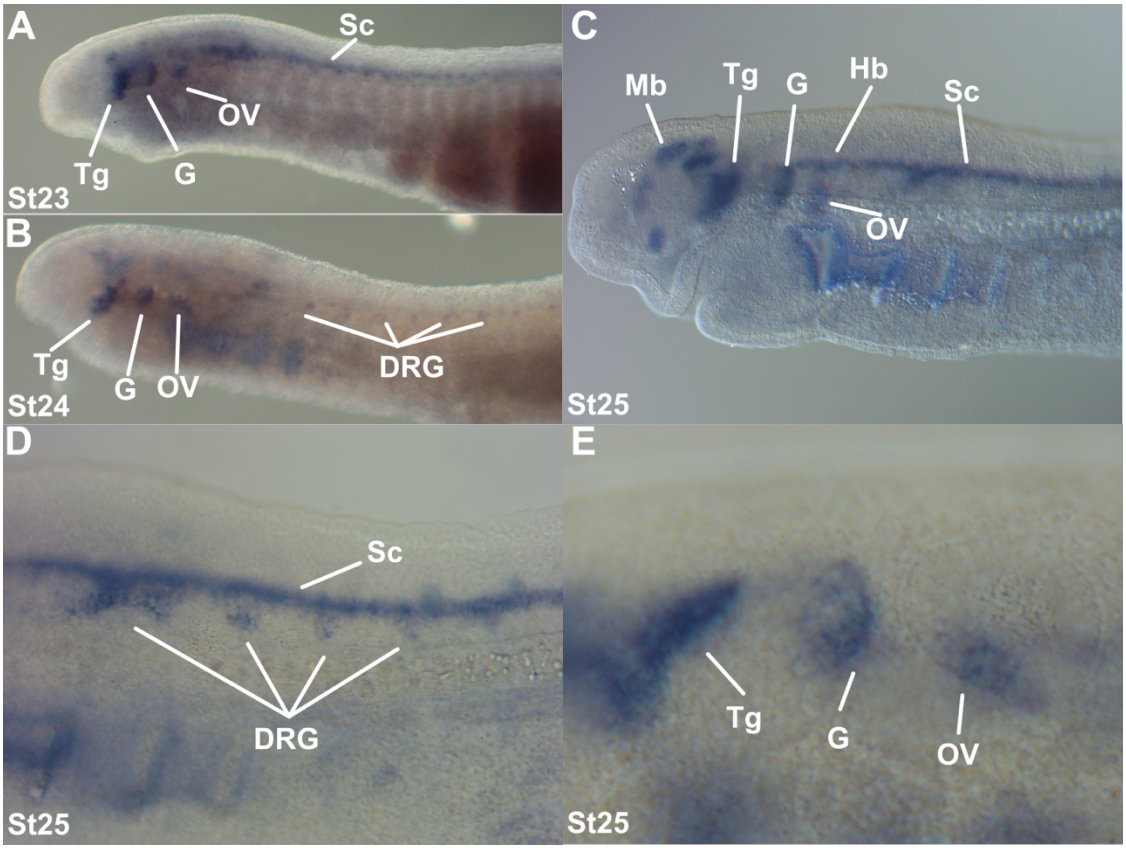


Figure 3.7. PRDM12 expression in lamprey embryos. A-C. The gene is expressed at stage 23 and persists in later stages in midbrain, hindbrain, spinal cord, trigeminal (strong staining in Tmm, see 3.7E. Staining also likely extends to Top, but closer analysis and sectioning is needed to clarify this), geniculate, otic vesicle and dorsal root ganglia. **D.** Close up of spinal cord and dorsal root ganglia in stage 25 embryo. **E.** Cranial ganglia expression anterior to the otic vesicle in stage 25 embryo. Abbreviations: DRG,Dorsal root ganglia. G,Geniculate. Mb,Midbrain. Hb,Hindbrain. OV,Otic Vesicle. SC,Spinal Cord. Tg,Trigeminal. Tmm,Trigeminal Maxillomandibular. Top,Trigeminal Ophthalmic.

3.2.3 Lamprey *ERR* marks the VA

As ERRs are similar in sequence to ERs, paralogs for both families were investigated in lamprey. Four *ERR* and three *ER* paralogs were identified. In the reconstructed phylogeny for ERR and ER, using the related RXR as an outgroup, lamprey ERR and ER proteins group within the respective ERR and ER family clades with high confidence (Fig. 3.8). No consistent one-to-one orthology to specific jawed vertebrate paralogs is seen for paralogs of either family. For this

reason and in order to avoid confusion between cyclostome and gnathostome paralogs, the lamprey genes were named *ERRa, ERRb, ERRc, ERRd and ERa, ERb, ERc*.

ERRa and *ERRc* showed embryonic expression, with *ERRc* expressed only in an area of the hindbrain and the mandibular arch (Fig. 3.9). While *ERRb* sequences were recovered from transcriptome data, no ISH signal was detected. No *ERRd* sequences were matched in any of the transcriptome datasets screened and I was not able to amplify this gene by PCR. *ERRa* is expressed in various CNS sites and in the VA (Fig. 3.10). Expression is seen in the epiphysis, midbrain and hindbrain at stage 24 (Fig. 3.10A) and becomes apparent in the olfactory placode and spinal cord by stage 25 (Fig. 3.10B, 3.10C). Starting at stage 24, strong and specific expression is seen in the VA (Fig. 3.10D).

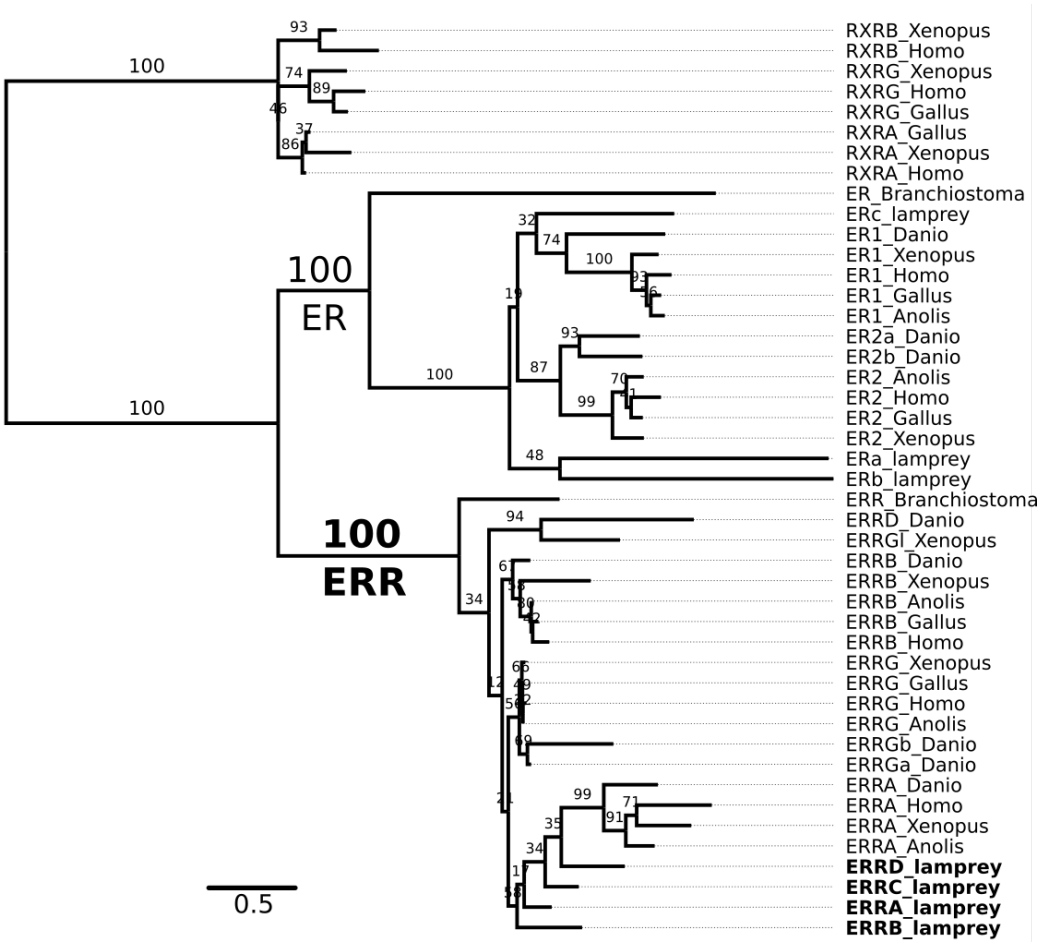


Figure 3.8. Phylogeny (maximum likelihood analysis, JTT, 1000 bootstrap) of estrogen related receptor (ERR) proteins. The estrogen receptor (ER) and retinoid receptor X (RXR) were used as outgroups. The phylogeny supports a clear placement of the four lamprey *ERR* (*ERRa, ERRb, ERRc, ERRd*) within the ERR family, but no further relationship to specific jawed vertebrate paralogs is seen.

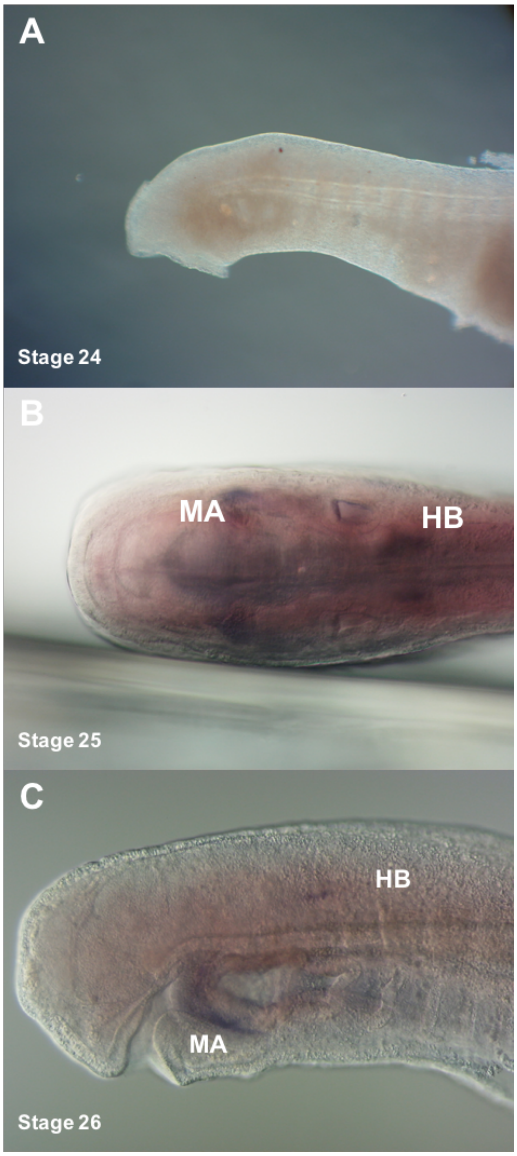


Figure 3.9. *ERRc* expression in lamprey embryos. There was no visible expression until after stage 24 **(A)**, with expression first seen in a zone in the hindbrain and in the mandibular arch in stage 25 **(B)**. Expression remains in later stages **(C)**. No expression in the cranial sensory ganglia was seen. Embryos are oriented with anterior to the left, A and C show lateral views (left side visible) and B shows a dorsal view of the head. Additional Abbreviations: MA, Mandibular arch.

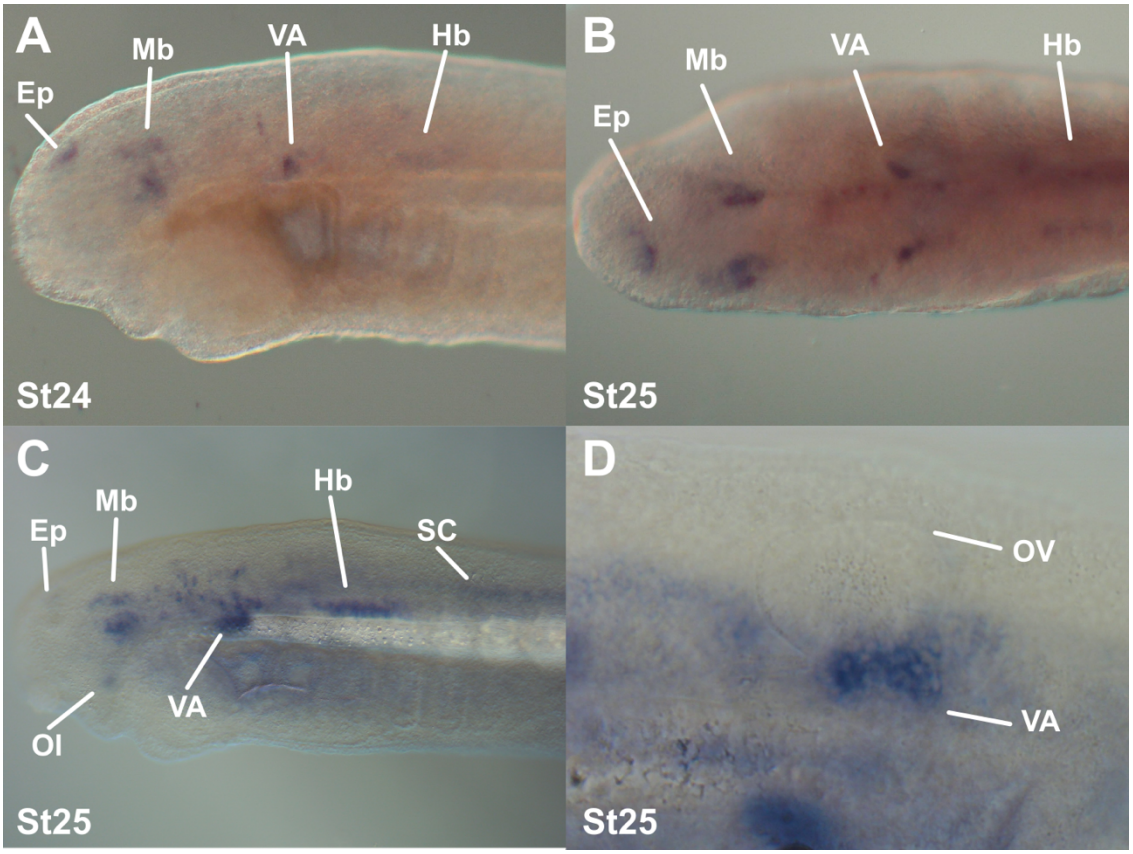


Figure 3.10. *ERRa* expression in lamprey embryos. A. Expression at stage 24 in midbrain, hindbrain, epiphysis, and vestibuloacoustic. B. Stage 25 embryo dorsal view revealing bilateral expression in brain and the vestibuloacoustic. C. Expression in the spinal cord and olfactory is evident by stage 25, added to previous domains. D. Close up of otic vesicle and the vestibuloacoustic of stage 25 embryo. Additional abbreviations: Ep,epiphysis. Ol,olfactory. VA,vestibuloacoustic.

3.2.4 *Shox* expression in epibranchial and trigeminal ganglia

Three lamprey *Shox* homologs (*ShoxA*, *ShoxB*, *ShoxC*) were found, though lacked clear orthology to jawed vertebrate *Shox* or *Shox2* (Fig. 3.11). *ShoxA* and *ShoxB* sequences were cloned from cDNA derived from embryo RNA, but it was not possible to amplify *ShoxC*. No embryonic expression was detected for *ShoxB* by in situ hybridisation. *ShoxA* is expressed in forebrain, midbrain and hindbrain from stage 23 (Fig. 3.12A), with signal persisting and becoming stronger in the same domains later (Fig. 3.12B, 3.12C). Expression becomes evident in CSG at stage

26, in geniculate, petrosal, nodose and trigeminal, but not in the VA (Fig. 3.12D). The signal is fainter than CNS expression and is better revealed under high contrast (Fig. 3.12E).

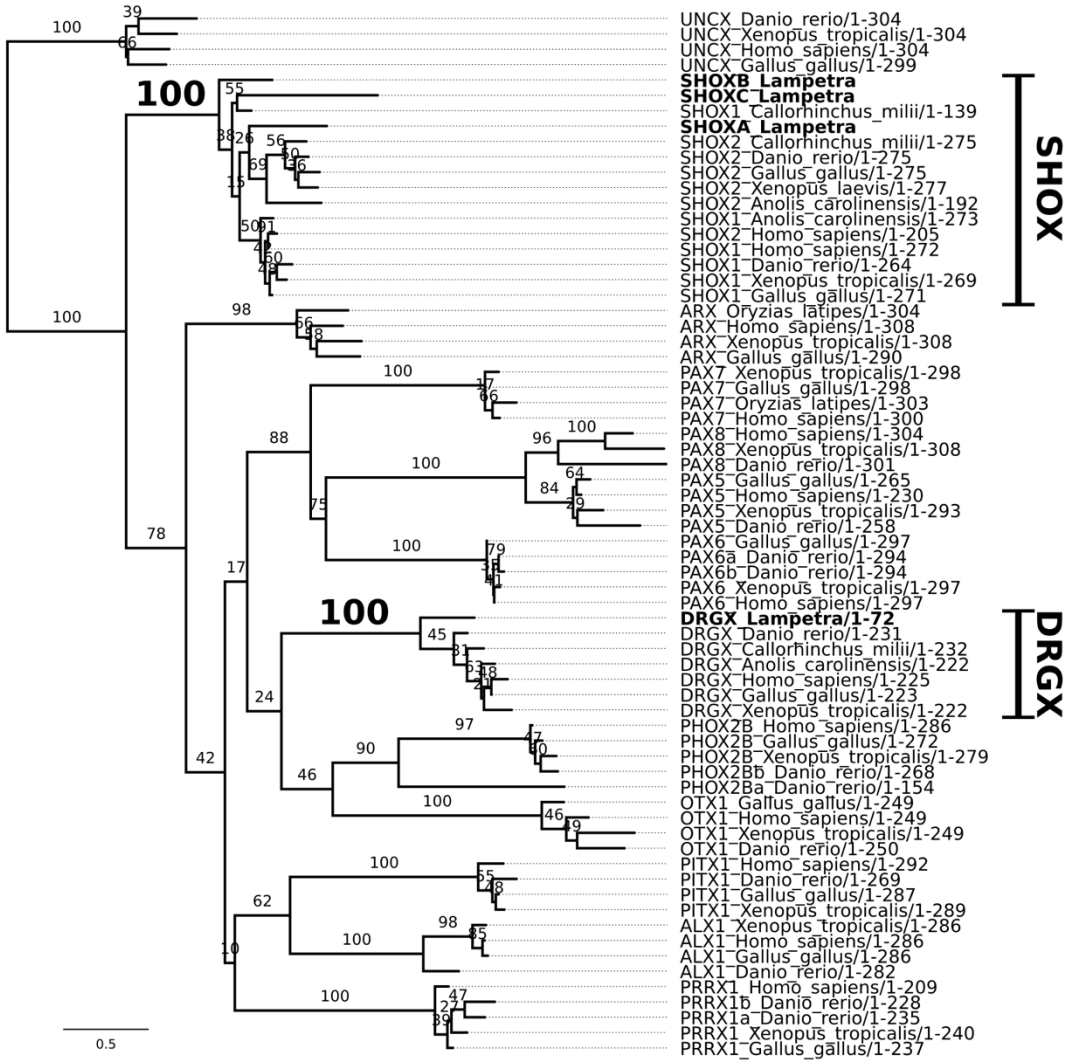


Figure 3.11. PRD homeobox class phylogeny (maximum likelihood analysis, JTT, 1000 bootstrap). Homeodomain sequences acquired from the HomeoDB[2] database. Different PRD class homeodomain sequences were aligned to assess orthology of lamprey *DRGX* and *Shox* candidates. Lamprey proteins consistently group within expected families, with no clear orthology to jawed vertebrate paralogs seen for *Shox*.

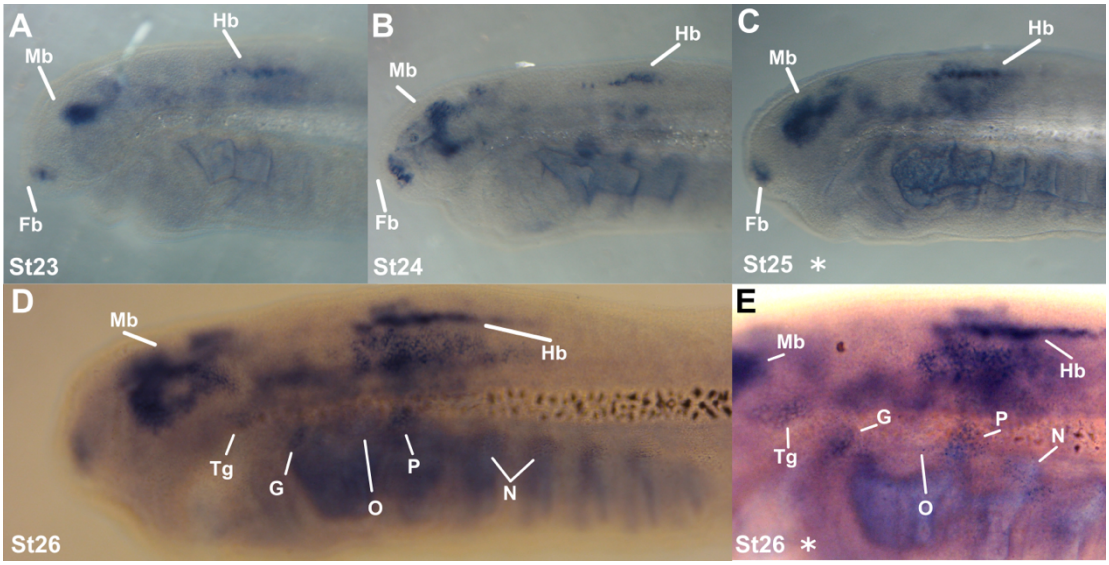


Figure 3.12. *ShoxA* expression in lamprey embryos. A. Expression in CNS is seen from as early as stage 23, becoming stronger and broader in the same forebrain, midbrain, hindbrain domains at later stages **(B,C)**. **D,E.** Faint expression in cranial ganglia becomes apparent in stage 26 embryos in geniculate, petrosal, nodose and trigeminal. Trigeminal staining is clearly seen in the maxillomandibular. It likely extends to the ophthalmic, but it is not possible to assess this properly without sectioning. Additional abbreviations: P,petrosal. N,nodose.

3.2.5 Lamprey *Hmx* paralogs mark all CSG

The search for *Hmx*/*Nkx5/NK5* family members in lamprey revealed three genes in a single scaffold. They share an extremely conserved homeobox region, identical at nucleotide level, with very divergent sequence outside this area (figure 3.13, full alignment included in the supplementary appendix). The homeodomain sequence was used to confirm orthology by aligning to other NK subclass homeodomain sequences from jawed vertebrates followed by molecular phylogenetic analysis. Lamprey *Hmx* clearly resolve within the *NK5* family, but one to one orthology to gnathostome paralogs is not observed (Fig. 3.14). For this reason, lamprey genes were named *HmxA*, *HmxB*, *HmxC*, avoiding conflict with gnathostome *Hmx1, Hmx2, Hmx3.* Hmx gene relationships are further explored in chapter 5.

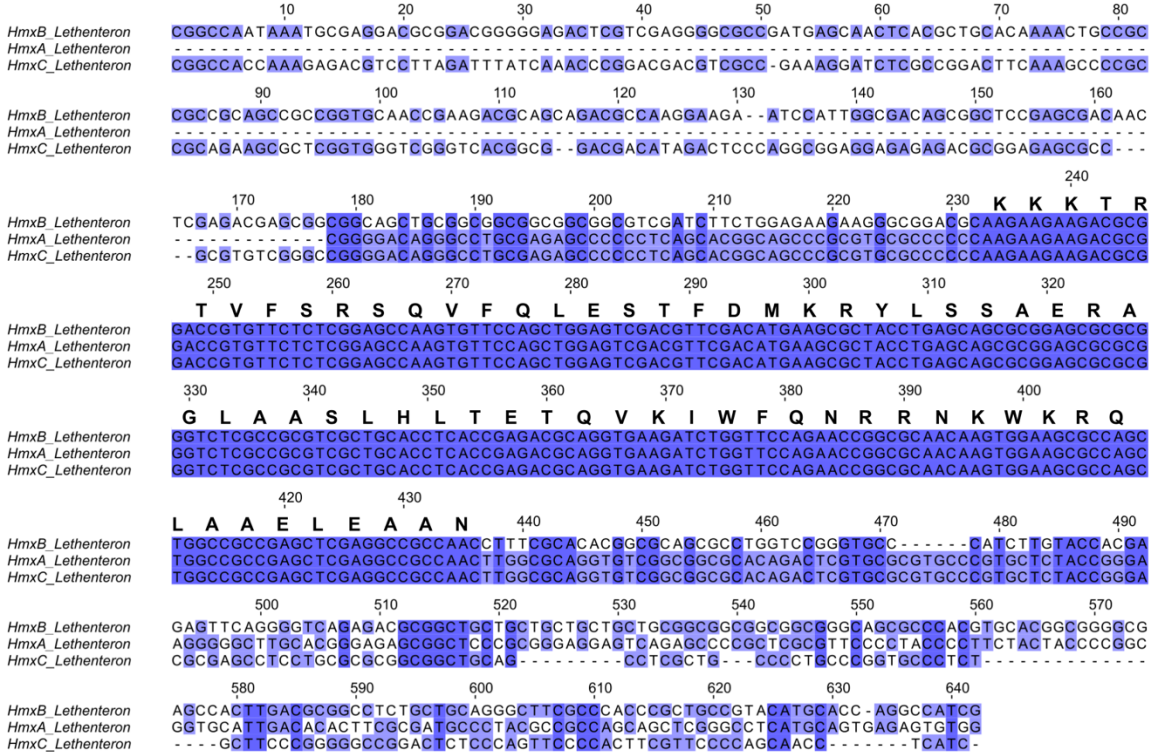


Figure 3.13. *L. camtschaticum Hmx* gene sequence alignment. All three genes share a homeobox sequence identical at nucleotide level, annotated with the encoded homeodomain sequence. *HmxA* and *HmxC* share additional sequence similarity upstream and downstream the homeobox region, not found in *HmxB*. No further sequence is shared outside this area by any of the three genes. Full alignment is included in the supplementary appendix. A more detailed analysis on the relationship of the three lamprey *Hmx* genes is carried out in chapter 5.

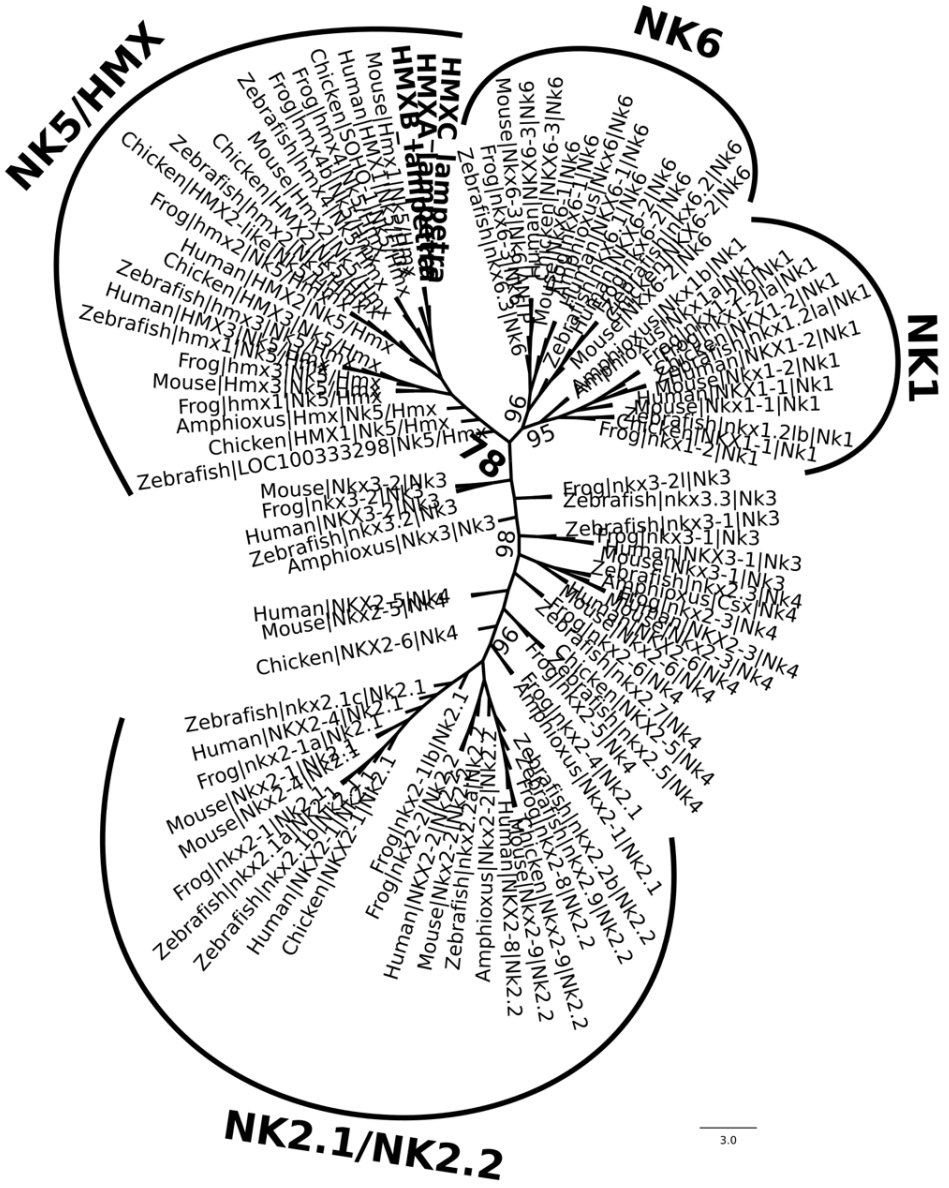


Figure 3.14. NK homeobox subclass phylogeny (maximum likelihood analysis, JTT, 1000 bootstrap). Homeodomain sequences acquired from the HomeoDB[2] database and lamprey Hmx were aligned. *L. planeri* Hmx clearly group within the *NK5* family (78 bootstrap), with no further orthology to jawed vertebrate *Hmx* genes seen. A more focused analysis on the Hmx/NK5 family is provided in chapter 5, using full Hmx protein sequences to dissect relationships within this group.

The three lamprey *Hmx* are co-expressed in the same domains in CNS and PNS, but *HmxA* and *HmxC* show fainter signal when stained for the same duration. Signal intensity does not correlate to probe size, while all probes are also sufficiently large

(HmxA 909bp, HmxB 780 bp, HmxC 517 bp). The overlap in expression is unlikely to be an effect of cross-hybridisation, as probes were made for non-conserved sequences. *HmxB* is already expressed at stage 23 (not shown here, see fig 5.1) and by stage 24 all CSG show strong *HmxB* signal (Fig. 3.15). This pattern persists in older embryos. Staining is also found in the CNS, in the hypothalamus, hindbrain and spinal cord. I hypothesise that a domain of expression ventral to the branchial arches possibly corresponds to hypobranchial ganglia, only observed before in *Xenopus*.

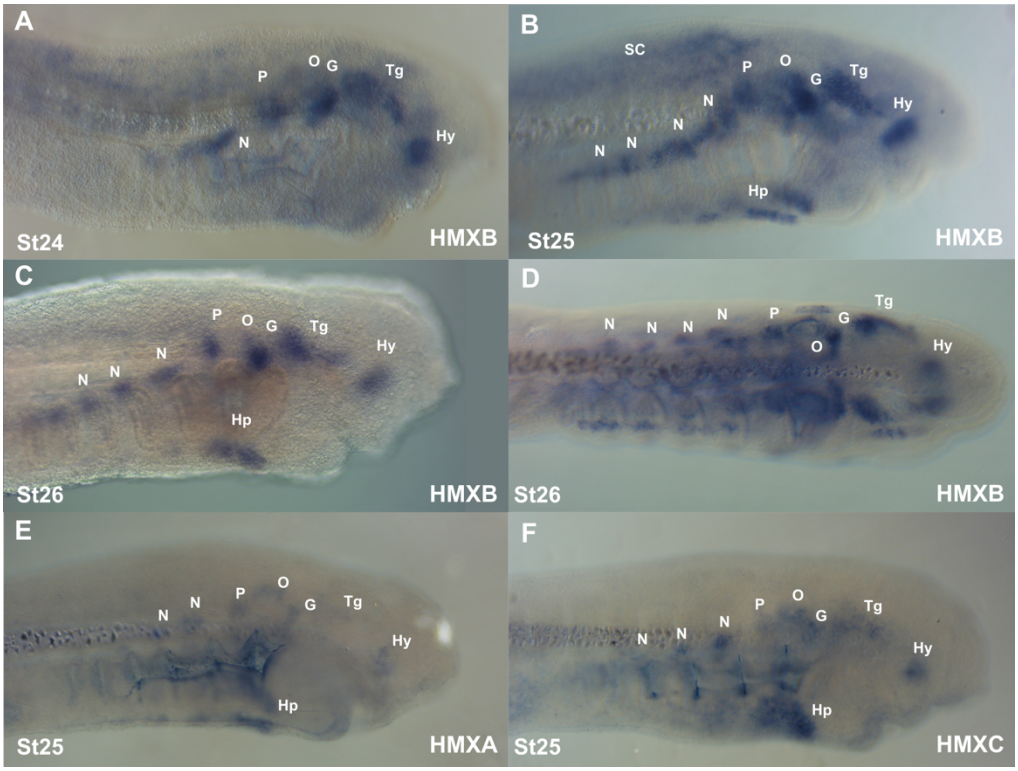


Figure 3.15. *Hmx* expression in lamprey embryos. **A.** At stage 24 *HmxB* is already expressed in trigeminal, vestibuloacoustic and epibranchial ganglia, as well as in hypothalamus. **B.** Expression is also seen in hindbrain, spinal cord and in hypobranchial ganglia by stage 25. **C,D.** All cranial ganglia are strongly stained for *HMB* at stage 26. *HmxA* **(E)** and *HmxC* **(F)** are co-expressed in overlapping domains as *HmxB*. Additional abbreviations: Hp, putative Hypobranchial Ganglia. Hy,Hypothalamus.

3.3 Discussion

The lamprey gene families studied here have paralog numbers comparable, but not identical, to jawed vertebrates. This is a characteristic pattern seen when comparing cyclostome and gnathostome gene families, as discussed in section 1.6 of chapter 1. As in gnathostomes, *PRDM12* and *DRGX* are represented by a single member in lamprey. A single lamprey *GAD* paralog is also found, which groups strongly with the gnathostome *GAD1* sequences. This implies either lampreys have lost *GAD2*, or I have failed to find it in available genome and transcriptome data. The *ERR* family has four unlinked genes in each vertebrate lineage, which is compatible with the possibility that they arose as a result of the vertebrate whole genome duplications. In the case of *Prox*, four members have been found in lamprey, but only two (*Prox1* and *Prox2*) are found in jawed vertebrates. Finally, the three lamprey *Hmx* genes are closely linked in a single cluster. Jawed vertebrate *Hmx* are also found in two closely-linked pairs, *Hmx2-Hmx3* and *Hmx1-SOHo*. This locus topology was suggestive of a duplication event in the family preceding whole genome duplications. *ERR* paralog relationships and homology are further explored in chapter 4, while models on the evolution of *Hmx* loci are discussed in chapter 5.

3.3.1 The molecular signature of lamprey CSG

Collating the expression patterns of studied genes, lamprey CSG can be identified by distinct expression phenotypes. The trigeminal and geniculate are both identified as *PRDM12*$^{+}$*ERRa*$^{-}$*Hmx*$^{+}$*ShoxA*$^{+}$. The VA, identified as *PRDM12*$^{-}$

ERRa^{+}*Hmx*$^{+}$*ShoxA*$^{-}$, is characteristically marked by *ERRa*. Nodose and petrosal share the same expression pattern for studied marker genes, as *PRDM12*$^{-}$*ERRa*$^{-}$ *Hmx*$^{+}$*ShoxA*$^{+}$ (Fig. 3.16). In addition, lamprey CSG are *Hmx* positive, as opposed to neural crest derived DRG, which do not express *Hmx* and are uniquely marked by *PRDM12* instead.

As discussed in section 1.4.2, few conserved CSG markers have been well described in lamprey before. *Pax3/7,* as the single homolog of gnathostome *Pax3* and *Pax7*, is the most specific, only marking the ophthalmic branch of the trigeminal (Modrell et al., 2014). In turn, *Phox2* is a conserved pan-vertebrate marker of epibranchial ganglia (Häming et al., 2011). The specificity of these markers correlates with their crucial function during fate determination of respective placodes and CSG. Ectopic expression of *Pax3* suppresses otic/epibranchial fate, while in absence of *Phox2* epibranchial ganglia acquire a somatosensory identity (D'Autréaux et al., 2011; Dude et al., 2009). The role of these markers in determination in placodes is well described, as opposed to later CSG differentiation and identity. However, both *Pax3* and *Phox2* continue to be expressed later in development and it is possible that they may also have such later roles. Finally, lamprey *NgnA* marks all CSG, with the neurogenin family being important for both determination and differentiation of CSG neurons in jawed vertebrates (Begbie et al., 2002). Interestingly, lamprey *Hmx* expression is comparable to that of *Ngn* in CSG. However, lamprey and gnathostome *Ngn* also marks DRG, while *Hmx* is specific to CSG. Thus, a conserved, specific role of *Hmx* during CSG development

is indicated. This role of *Hmx* in CSG and sensory neuron development is explored in both lamprey and non-vertebrate chordates in chapter 5. Compiling the expression of *NgnA, Pax3/7* and *Phox2* and new marker genes studied here provides distinct molecular signatures that identify each type of sensory ganglion, summarised in fig3.16.

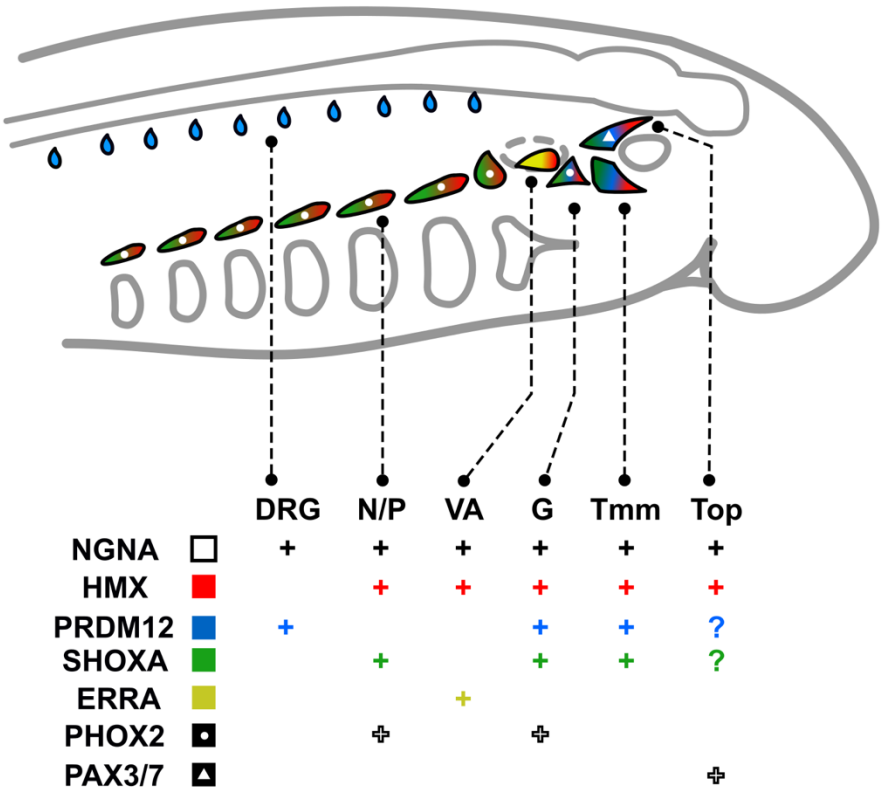


Figure 3.16. Lamprey CSG marker expression summary. Each sensory ganglion is identified by the expression of a combination of markers: Top by [*NgnA, Hmx, Pax3/7*], Tmm by [*NgnA, Hmx, PRDM12*, *ShoxA*], G by [*NgnA, Hmx, PRDM12*, *ShoxA*, *Phox2*], VA by [*NgnA, Hmx, ERRa*] and N/P by [*NgnA, Hmx, ShoxA*, *Phox2*]. All CSG are marked by *NgnA* and *Hmx* expression, while DRG are marked by *NgnA* and *PRDM12* expression.

3.3.2 CSG marker conservation and lack of specificity

Comparative analysis of marker expression reveals that CSG share a conserved molecular identity across vertebrates. The trigeminal is marked by *PRDM12,* the *VA* by *ERR and Hmx* and the epibranchial ganglia by *Shox,* in both lamprey and gnathostomes (based on chick marker expression and other literature described in section 3.1.2). However, specificity of marker expression is not supported. Unlike gnathostome *Hmx3*, which is specifically expressed in the VA and lateral line, no lamprey *Hmx* paralog shares this specificity in the VA, as all lamprey *Hmx* are expressed in all CSG. The expression of *ShoxA* in both epibranchial and trigeminal ganglia resembles *Shox2*, which is in turn expressed in geniculate and trigeminal in mouse (Rosin et al., 2015b). It remains possible that *ShoxB* or *ShoxC* are specifically expressed in epibranchial ganglia, since I was unable to detect their expression. Nevertheless, expression of the family in trigeminal in both lamprey and gnathostomes suggests ancestral *Shox* expression in all CSG, except for the VA. In the case of *PRDM12*, it is likely that expression in the geniculate could be lamprey specific as it has not been identified in other taxa. However, overall, the emerging pattern for these markers is reminiscent of the *Ngn* family, in that these families are conserved CSG markers, but individual paralogs have been employed differentially in different ganglia in different species. Hence most of these genes are conserved markers of CSG, but not of specific CSG cell fates. This is consistent with what is known about placodal markers, as very few have been described showing specificity. Instead, the majority of markers are expressed in combinations of placodes (Schlosser, 2006).

The development of CSG is a long and complicated process. Induction, determination and specification are marked by the expression of various genes (Pieper et al., 2011; Saint-Jeannet and Moody, 2014; Schlosser, 2006). Placodal progenitors appear only transiently, then delaminate, sometimes migrate, and differentiate producing the different neuronal populations of each CSG (Blentic et al., 2011). Because of the complexity of this multi-stage development, the number of cells expressing particular markers could vary depending on timing and differ across species. For this reason, it is possible to hypothesise that the specificity of expression of selected markers seen in chick could be stage dependent. Likewise, it could stem from cell number differences across CSG populations. Finally, the stages screened here in lamprey were chosen because CSG determination and differentiation takes place in this window. Hence, it can be hypothesised that early markers such as *Hmx* and *PRDM12* are more likely to be connected to fate determination. Accordingly, slightly later and specific *ERR* expression may mark differentiating VA cells.

3.3.2 Conclusions

The marker genes studied here were selected based on CSG expression in chicken and knowledge of their role in developing or adult CSG in gnathostomes. By extension, conservation of marker expression across vertebrates is likely indicative of conserved function in CSG. Lamprey data shown here identify a set of pan-vertebrate markers and open up the possibility of carrying out comparisons with other animal groups. Importantly, this can be used to compare sensory cell types across chordates. Indeed, expression of CSG markers presented here has been

largely unexplored in other chordates. To my knowledge, the only available data have been amphioxus embryo *PRDM12, Shox* and *ERR* expression. Strikingly, while *PRDM12* and *Shox* are only found in CNS, *ERR* is found in CNS and in photoreceptors and scattered epidermal sensory neurons during development (Bardet et al., 2005; Jackman and Kimmel, 2002; Thelie et al., 2015). At the start of this study no ascidian expression data for any of the identified marker genes were available. Comparative analysis is explored further in the following chapters. Chapter 4 focuses on *ERR*. It provides evidence of ancient regulation of homologous *ERR* loci across vertebrates and supports a role of the family in sensory cell development across bilateria. Chapter 5 then explores how conserved expression, regulation and function of *Hmx* between vertebrates and tunicates traces the origins of CSG evolution.

Chapter 4 Expression and regulation of ancient sensory marker *ERR* in vertebrates

Work Declaration

Experimental work and data analysis presented in this chapter have been performed by myself (Vasileios Papadogiannis), apart from experiments and data described otherwise in this declaration. Sea lamprey (*P. marinus)* embryo microinjections were performed by Dr Dorit Hockman. Zebrafish embryo microinjections were performed by Martyna Lukoševičiūtė. In situ hybridisation data was kindly provided by Dr Silvia Mercurio for *C. intestinalis ERR* and Dr Cedric Patthey for *B. lanceolatum ERR.* SEM micrographs of limpet trochophores were kindly provided by Professor Sebastian Shimeld.

4.1 Chapter Introduction

CSG are considered a vertebrate novelty. Covered in more detail in section 1.3 of the general introduction, they develop from neurogenic placodes, with cranial neural crest also contributing (Patthey et al., 2016, 2014; Steventon et al., 2014). The molecular network behind the early specification of placodes has been studied in vertebrate model systems, revealing conservation of genes and mechanisms (Grocott et al., 2012; McCabe and Bronner-Fraser, 2009; Pieper et al., 2011; Schlosser, 2006). However, potential conservation of the molecular nature of the ganglia later in development, and of the sensory cell populations placodes produce, are less well studied. Chapter 3 explored this, starting from a transcriptomics

dataset of CSG marker genes in chicken embryos (Patthey et al., 2016). By assessing expression of marker families in lamprey, conserved markers across vertebrates were identified. This chapter focuses on one of the genes that emerged from this study as marking the vestibuloacoustic ganglion (VA), the Nuclear Hormone Receptor (NHR) transcription factor *Estrogen Related Receptor Gamma* (known as *ERRγ* or *ESRRG,* depending on the species).

4.1.1 The ERR family: evolution and roles in vertebrate development

ERRγ is one of four vertebrate paralogs in the Estrogen Related Receptor (ERR) gene family, with a single *ERR* gene present in most invertebrates. This pattern of gene family membership is common for transcription factor gene families, and has usually resulted from the two rounds of genome duplication (2R) early in vertebrate evolution (section 1.5 in chapter 1) (Dehal and Boore, 2005; Putnam et al., 2008), though this has not been explicitly tested for the *ERR* family. The ERR family is named after its sequence similarity to the Estrogen Receptor (ER) family to which it is closely related (Baker, 2008; Bertrand et al., 2004). *ER* has been independently lost by several different invertebrate groups, while to date there are no identified lineages where *ERR* has been lost (Thornton et al., 2003). While ER requires ligand-binding to activate transcription of target genes, ERR is an ‘orphan’ NHR. It does not need a ligand to bind DNA, no naturally occurring ligands have been identified, and it can act as a typical transcription factor by binding to Estrogen Related Receptor Elements (ERREs) and regulating the expression of target genes (Heard et al., 2000; Tremblay and Giguère, 2007).

The vertebrate ERR paralogs have been quite well studied, with expression identified in multiple tissues including the nervous system, eye, muscle and kidney (Bardet et al., 2004; Heard et al., 2000; Süsens et al., 2000). In developing embryos this includes expression of *ERRγ /ESRRG* in cranial ganglia including the trigeminal, VA and (in zebrafish) lateral line ganglia, and the locus is associated with hearing loss in humans and mice (Bardet et al., 2005, 2004; Lorke et al., 2000; Nolan et al., 2013). Outside of the vertebrates, the developmental expression and role of ERR genes have been little investigated, since most studies focus on its potential regulation in adults by environmental contaminants such as bisphenol-A (Coward et al., 2001; Takayanagi et al., 2006). Current knowledge of potential roles in embryogenesis are limited to a provisional description of expression in the developing *Drosophila* gut (Thurmond et al., 2019), and to description of expression in the central and peripheral nervous systems (CNS and PNS) of amphioxus (Bardet et al., 2005).

4.1.2 Investigating ancestral expression and regulation

Amphioxus predates the evolution of CSG (Patthey et al., 2014; Schlosser et al., 2014). *ERR* PNS expression in this lineage hence suggests two hypotheses: that the ancestral *ERR* may have had a role in PNS development, and that in jawed vertebrates one paralog, *ESRRG/ ERRγ*, inherited this as expression in the VA. This could imply cell-type homology between the vertebrate VA and amphioxus *ERR*-expressing PNS cells. Alternatively, PNS expression may have evolved independently in different lineages. To investigate this, the expression and regulation of ERR genes was studied in vertebrates and additional outgroups.

Chapter 3 already explored expression in cyclostomes, as they are the earliest diverging extant vertebrate group. While they lack jawed vertebrate characters, they share many developmental processes and structures with other vertebrates (Green and Bronner, 2014; Shimeld and Donoghue, 2012). CSG and cranial nerves are present in lampreys, developing from homologous neurogenic placodes with neural crest contribution (Kuratani et al., 1997; Modrell et al., 2014). As discussed in section 1.4 of chapter 1, evidence from broad genome comparison suggests two rounds of genome duplication (2R) preceded the separation of lamprey and jawed vertebrate lineages (Sacerdot et al., 2018). Although one to one orthology between genes in the two lineages can be difficult to assign (Kuraku, 2013), some conserved non-coding elements (CNEs) are shared between homologous genes in the two groups (McEwen et al., 2009), and these have been shown to include enhancers for important developmental genes with conserved embryonic expression (Parker et al., 2011).

4.1.3 Chapter Summary

The work presented in this chapter first resolves the evolutionary history of the *ERR* gene family. Close genomic linkage of *ERR* and the *ER* genes is shown in several invertebrate genomes, illustrating their origin by tandem duplication of an ancestral NHR gene. Linkage is also preserved for some vertebrate *ERR* and *ER* genes and this, and other syntenic genes, allow for robust comparison of lamprey and jawed vertebrate *ERR* loci. Embryonic expression of lamprey *ERR* genes revealed one that is expressed in the VA. Here, genomic analysis identifies CNEs within this gene that are found within *ERRγ/ESRRG* in jawed vertebrates. The function of one group

of CNEs is tested in transgenic lamprey, zebrafish and chicken, showing it drives expression in CNS and PNS and hence illustrating a conserved network regulates *ESRRG* across vertebrates. Finally, *ERR* expression in additional invertebrate species is addressed to polarise evolutionary change.

4.2 Results

4.2.1 Lamprey *ERR* expression in the VA and CNS

In chapter 3, embryonic expression of lamprey *ERR* was characterised. Four *ERR* genes were identified in lampreys, which were named *ERRa* to *ERRd* to avoid direct attribution to jawed vertebrate *ERRα* to *ERRδ*. In-situ hybridisation showed *ERRa* is expressed in the CNS, including the midbrain, hindbrain, epiphysis, olfactory system and spinal cord, and in the PNS in the VA. To better characterise VA expression, the ISH screen was expanded to include later stages and is presented here in Fig.4.1. *ERRa* embryonic expression commences at Tahara stage 24 (Tahara, 1988) in the epiphysis, midbrain and the VA. By stage 25 expression is also seen in the olfactory system, forebrain, hindbrain and in the spinal cord. From stage 27 onwards, expression persists in all aforementioned domains, but with more widespread signal in the hindbrain and spinal cord and stronger signal in the VA (Fig. 4.1C, D, F). Conserved expression in the VA and most CNS domains is shared by lamprey *ERRa* and gnathostome *ERRγ*. Hence, it was hypothesised the two paralogs may have inherited this role from an ancestral gene originally expressed in the VA in stem-vertebrates. However, in chapter 3, it was shown that one to one orthology of lamprey and gnathostome *ERR* paralogs cannot be

assigned. Therefore, other sources of data were necessary to infer the ancestral condition and address whether *ERR*a and *ERRγ* VA expression has shared origin.

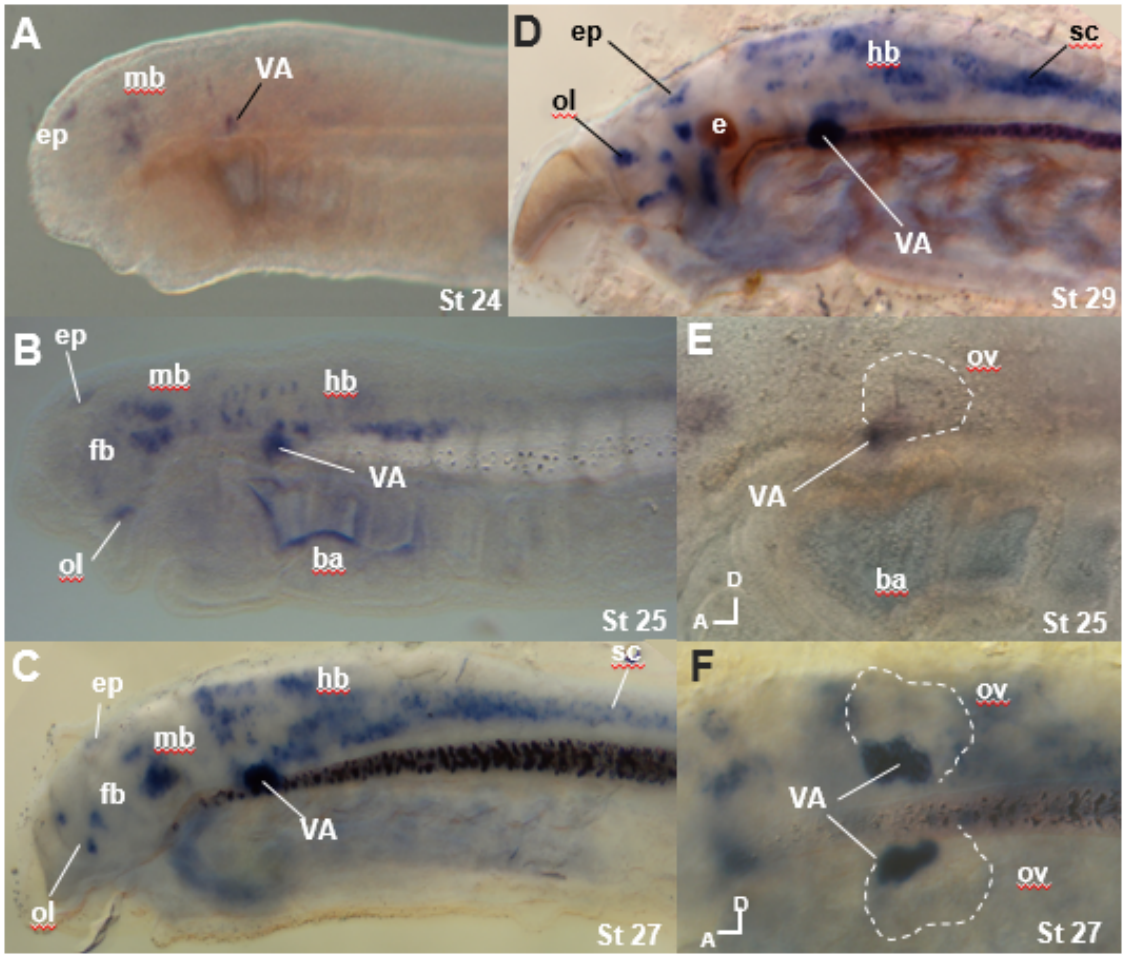


Figure 4.1. Lamprey *ERRa* Expression. Embryos are in left lateral view with anterior to the left, except F which is a dorsal view. **A.** Lamprey *ERRa* expression at stage 24 in the epiphysis (ep), midbrain (mb) and vestibuloacoustic (VA). **B.** At stage 25, expression is also seen in in the ventral and dorsal hindbrain (hb), olfactory (ol) and forebrain (fb). Apparent labelling in the branchial arches (ba) in this specimen is background due to trapping. **C.** At stage 27 expression in the spinal cord (sc) is also visible. VA expression is very strong by this stage. **D.** Expression at stage 29 persists in all previous domains. **E.** Close-up of the otic region in a stage 25 embryo showing staining in the VA, positioned at the ventral and anterior side of the otic vesicle (ov; white dotted outline). **F.** Stage 27 embryo in dorsal view and focused on the VA region, white dotted outline marks the position of the otic vesicles. Additional abbreviations. E, eye.

4.2.2 Relationships and evolution of *ERR* loci

A broader molecular phylogenetic analysis (Fig. 4.2) was carried out to better examine orthology and paralogy relationships, including lamprey, hagfish (*Eptatretus burgeri*), gnathostome and invertebrate data. ERR and ER sequences were included for a range of vertebrate and invertebrate taxa, plus the more distantly related Retinoid X Receptor family ((RXR/USP: (Mangelsdorf et al., 1992)) as outgroups. This analysis includes the taxa of the analysis carried out in chapter 3, but expands this with the addition of ER and ERR proteins from more vertebrates and invertebrates, aiming to better resolve paralog relationships. The results show previously defined jawed vertebrate paralogs (ER1 and ER2; ERRα, ERRβ, ERRγ and ERRδ) form well-supported clades within each family. Lamprey and hagfish ERR and ER sequences group within the ERR and ER families respectively and, while lamprey and hagfish sequences generally group together, the tree does not resolve sufficiently to determine orthology between these sequences and jawed vertebrate sequences.

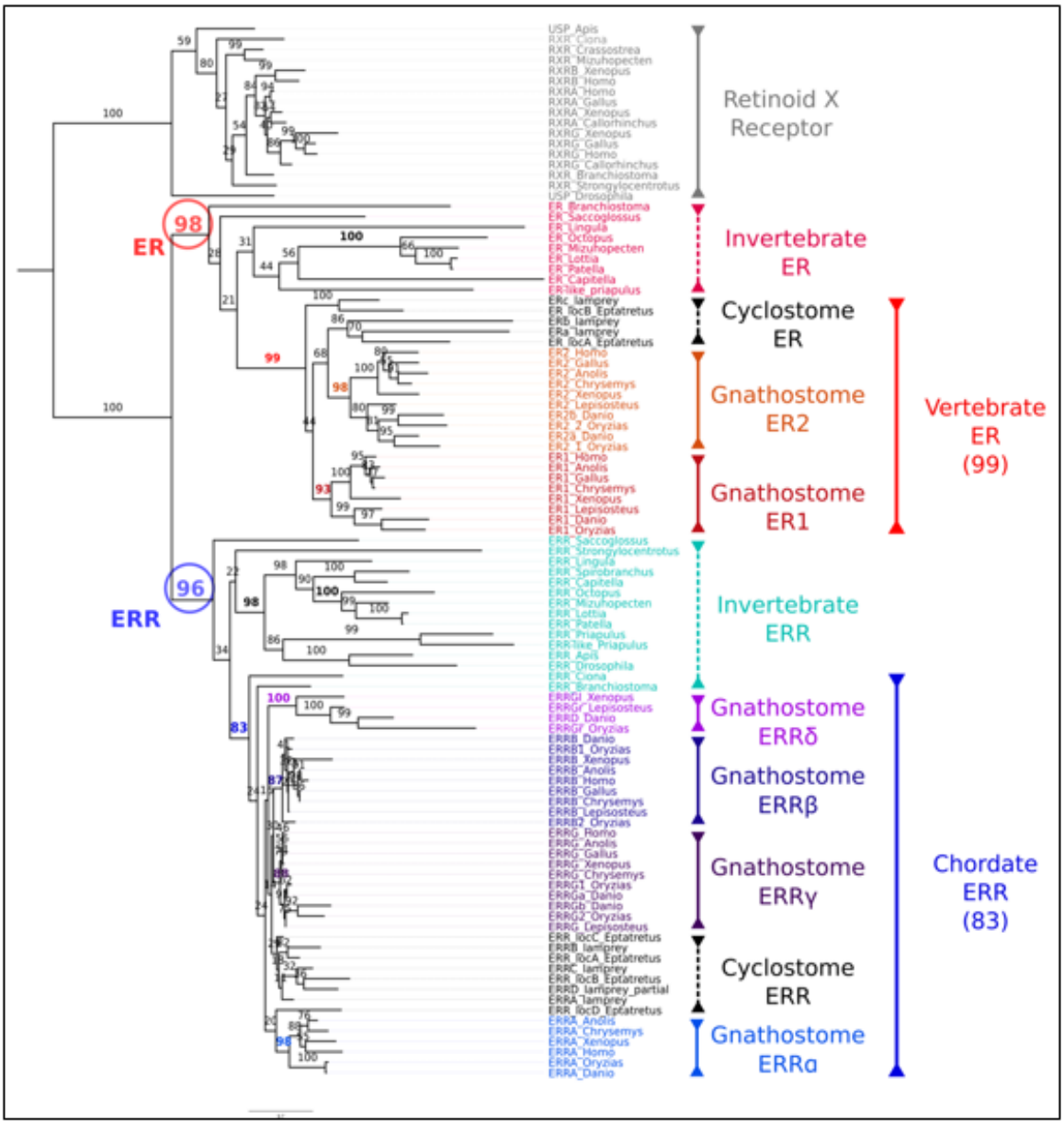


Figure 4.2. Molecular phylogenetic analysis of *ERR* and *ER* sequences. Phylogenetic tree using maximum likelihood analysis (JTT) with 1000 bootstrap replicates (showing as percentage support values on nodes) of ERR and ER family members. The RXR family is used as an outgroup. Different taxonomic groups are colour coded and highlighted with bars on the right side, with continuous lines indicating monophyly and dashed lines indicating non-monophyletic groups. All sequences described in this study fall robustly within their expected families: ER (98%, red circle), ERR (96%, blue circle), RXR (100%). Bootstrap support values for chordate, vertebrate and gnathostome clades are coloured according to the respective group name colour.

During the analysis of lamprey *ERR* and *ER* loci, it was noted that one *ERR* paralog and one *ER* paralog were in the same genomic scaffold, about 600kb apart. On

examining synteny in other vertebrate and invertebrate taxa where both genes are present and good genome assembly available, *ERR* and *ER* genes were found in linkage elsewhere (Fig. 4.3A, B). The distance between genes is small in some invertebrates, for example around 40Kb in the hemichordate *Saccoglossus kowalevskii*. In jawed vertebrates, distances are larger, usually in the order of several Mb. Paralog pairs *ERRγ-ER1* and *ERRβ-ER2* often (but not always) appear in linkage, while the *ERRα* and *ERRδ* loci appear unlinked to ER loci in all jawed vertebrates examined, with the exception of *Xenopus* where *ERRδ* is found close to *ER2*.

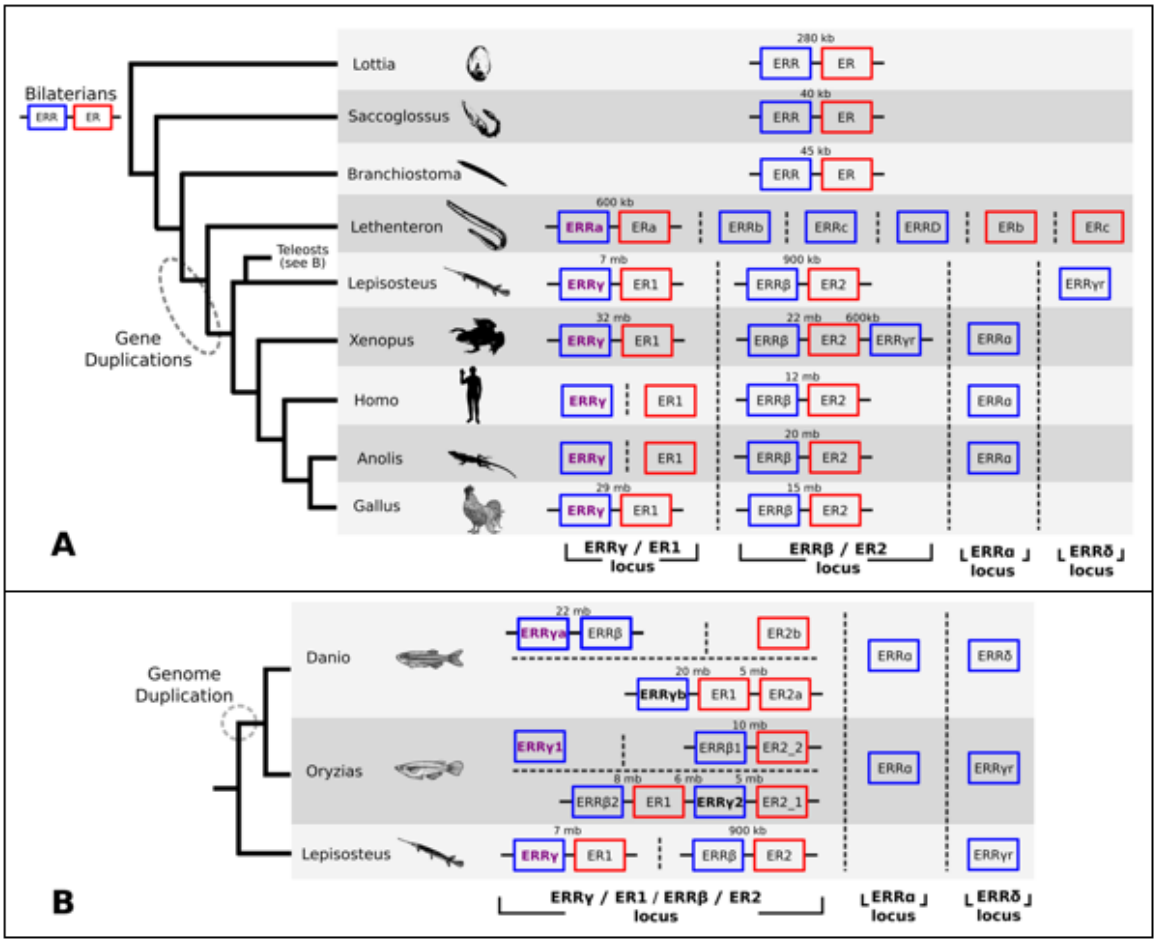


Figure 4.3. The evolutionary history of *ERR-ER* gene linkage. *ERR* genes are in blue and *ER* genes in red. A black line connecting the gene boxes indicates the loci lie in proximity in the same scaffold/chromosome, with intergenic distances shown. Dotted

vertical lines represent a change in scaffold/chromosome. Phylogenetic relationships of species presented are on the left. **(A)** shows the broader context across animals and spanning the 2R genome duplications. **(B)** is focused on the ray-finned fish, amongst which teleosts have undergone another genome duplication**.** Pictures from https://commons.wikimedia.org.

4.2.3 *ERRa* and *ERRγ* conserved synteny and non-coding elements

As discussed above, while sequence analysis failed to demonstrate one-to-one relationships between lamprey and jawed vertebrate *ERR* genes, the expression of lamprey *ERRa* was most comparable to that of jawed vertebrate *ERRγ.* What is more, the linkage seen for pairs *ERRγ-ER1* and *ERRβ-ER2* in many jawed vertebrates hinted to the possibility of conserved ancestral synteny. Synteny around each gene provided evidence that all four *ERR* loci in both lamprey and jawed vertebrate lineages had evolved by block duplication of a single ancestral *ERR* locus (Fig. 4.4A). This is consistent with the *ERR-ER* linkage described above, and consistent with an origin via the 2R genome duplications. Local synteny also provided support for orthology between the lamprey and jawed vertebrate paralogs expressed in the VA, with *ERRa* and *ERRγ* flanked by the single copy genes *EPRS, SPATA17* and *USH2A* in both lineages (Fig. 4.4A,B). This suggests that although *ERRa* and *ERRγ* do not resolve as orthologs in molecular phylogenetic analysis, they are sited in orthologous loci.

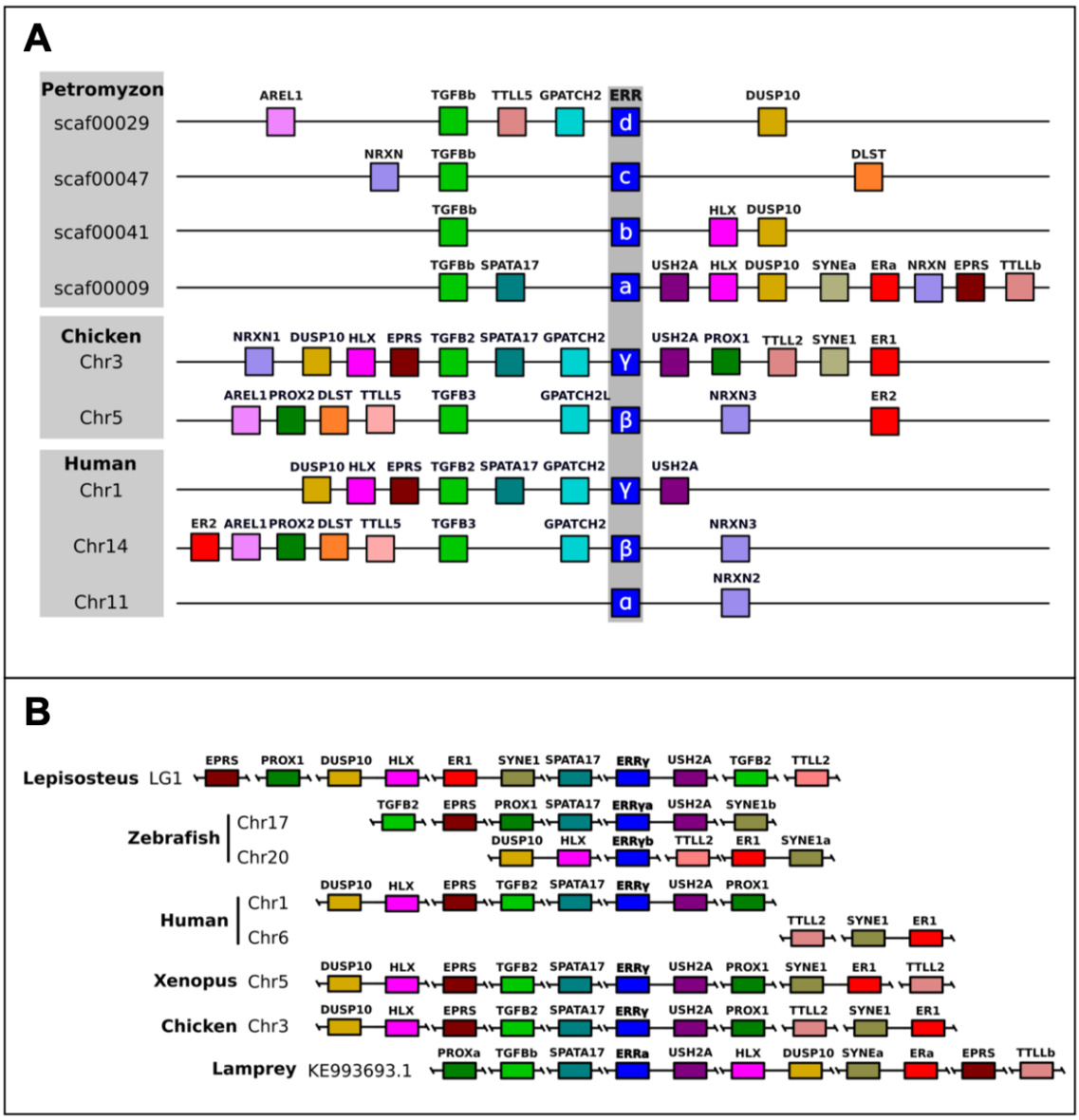


Figure 4.4. *ERR* synteny. A. Local synteny around *ERR* loci in lamprey, human and chicken. Paralogs are colour coded and only genes shared between loci are shown. Distances are not to scale, but gene order is as observed. **B.** Synteny comparison of jawed vertebrate *ERRγ* loci with lamprey *ERRa*.

Similar expression and shared ancestry raised the question if they might also be associated with conserved enhancer elements involved in regulating conserved gene expression. As discussed in section 1.6.2 of chapter 1, a number of conserved non-coding elements have been shown to be shared between gnathostomes and lamprey. Many such CNE regulate the expression of developmental genes and they

also show cross-species recognition, as lamprey CNE can drive reporter expression in zebrafish (McEwen et al., 2009, 2006). A jawed vertebrate CNE regulating *ERRγ*, EL161, has been previously identified as conserved between zebrafish, medaka, human and chicken (Roberts et al., 2014) and is located in the most 3' intron of *ESRRGa* in Zebrafish. The authors demonstrated that the zebrafish element drives reporter activity in expression domains of *ESRRGa* that include the epiphysis and parts of the forebrain and hindbrain. Here, multiple additional CNEs were identified, all located in the two most 3' introns (Fig. 4.5). These introns lie between the last three exons, which encode the highly conserved ligand binding domain of the nuclear hormone receptor. Intron-exon boundaries map to the same positions in the ligand binding domains in both lineages, demonstrating the ancestry of this gene structure (Section S6 of supplementary appendix). The newly identified CNEs were named CNE1-10. Four of these, plus EL161, are conserved in lamprey (Fig. 4.5). Conserved synteny and non-coding elements between *ERRa* and *ERRγ* are summarised in Fig. 4.5B. CNE sequences and alignments can be found in the supplementary appendix.

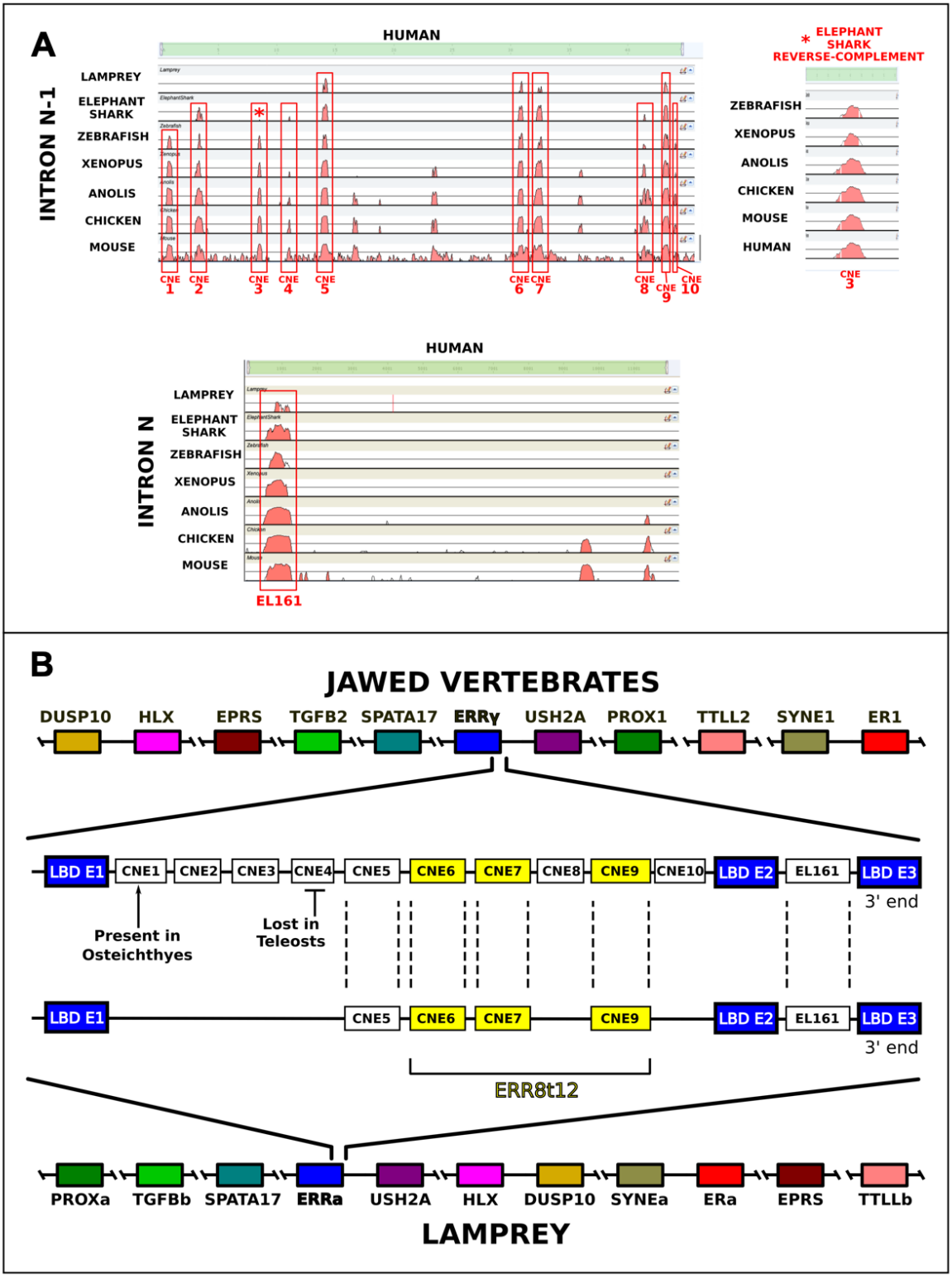


Figure 4.5. *ERRa – ERRγ* shared CNEs. A. Vista comparisons of the 3' end of human *ERRγ*, showing conservation across vertebrate genomes and identification of CNEs. The inset shows a similar analysis using the putative shark CNE3 sequence in reverse orientation **B.** Schematic representation of CNEs identified in jawed vertebrate *ERRγ* and lamprey *ERRa* and local synteny. Colour coding corresponds to colours used in Fig.4.4. The ERR8t12 cluster is highlighted in yellow

4.2.4 Cross-species activity of *ERRa/ERRγ* CNE in CNS and PNS

As these data suggest that jawed vertebrate *ERRγ* and lamprey *ERRa* should be considered orthologs, it was hypothesised the shared CNEs reflect ancestral regulatory elements responsible for expression in territories shared between the genes in jawless and jawed vertebrates. To investigate this, a region with multiple CNEs in close proximity was cloned from lamprey, (named *ERR8t12* and containing CNE6, CNE7 and CNE9; Fig. 4.5B), and its ability to drive reporter expression in transgenic lamprey embryos was tested. Transgenesis was achieved using microinjection of GFP reporter constructs in lamprey zygotes, employing a mega-nuclease system for genomic integration. Description of the method is provided in chapter 2. Reporter expression viewed in live embryos overlapped with domains of the endogenous *ERRa* expression, with 36.3% of injected embryos showing expression in the brain, and 40.3% in more posterior neural tube (Fig. 4.6A,B; Fig. 4.S1). Additionally, expression of the reporter was seen in the branchial arches and head muscles in many embryos (Fig. 4.S1). Since precise sites of expression in the CNS were difficult to see in live embryos, and CSG are difficult to distinguish from underlying CNS, fixed embryos were stained using antibodies against GFP and the neural differentiation marker Hu/ELAV and detailed expression was recorded via confocal microscopy (Fig. 4.6 C, D, E, Fig 4.7). This gave close to single cell level resolution, and revealed GFP in sites that could not be seen in live embryos. It confirmed CNS expression mirrored the expression of *ERRa*, with reporter expression in the olfactory system and epiphysis (Fig. 4.6C, D), midbrain (Fig. 4.6C) and spinal cord (Fig. 4.6E). Outside the CNS, expression was also seen in scattered

epidermal cells, in the eye, and infrequently in few cells in cranial ganglia (Fig. 4.7). The frequency of expression in cranial ganglia was lower than that detected in the CNS, and it was not found consistently in any one ganglion. As low, infrequent activity in CSG is also driven by the krt4 promoter alone in control embryos injected with empty constructs (see chapter 2), I was not able to conclude that this reflected *ERR8t12* activity in lamprey cranial ganglia.

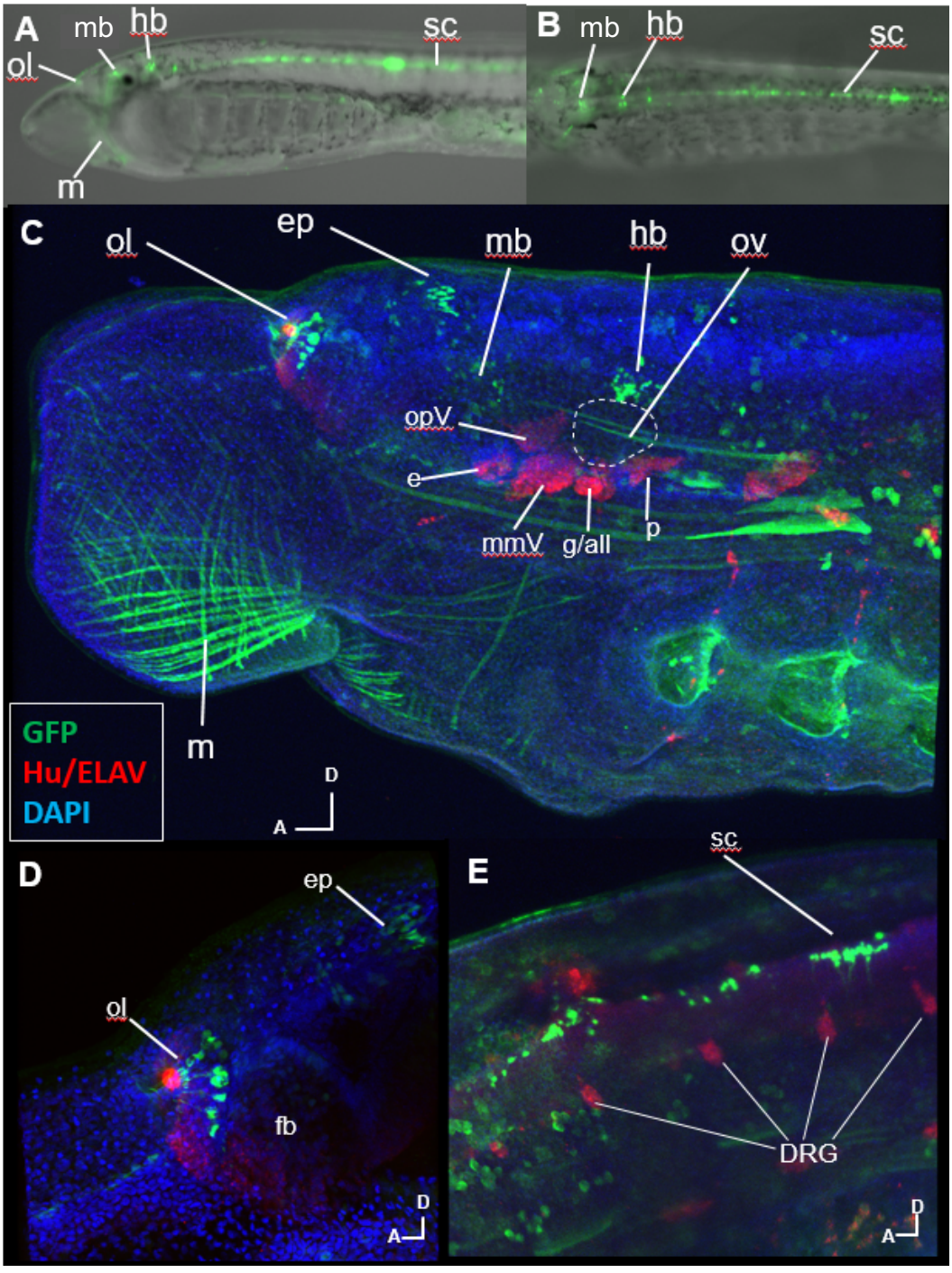


Figure 4.6. *ERR8t12* CNE activity in lamprey embryos. A, B. Live stage 29 *P. marinus* embryo in lateral and dorsal view respectively, showing reporter expression in olfactory system (ol), forebrain (fb), hindbrain (hb) and spinal cord (sc), as well as head muscle (m). For additional images see Fig. 4.S1. **C-E**. Confocal reconstructions of stage 29 *L. planeri* embryo labelled for GFP (reporter, green), Hu/ELAV (differentiated neurons, red) and DAPI (nuclei, blue). Embryo orientations are shown in each image (A, anterior, D, dorsal). **C.** Stage 29 head showing reporter localisation in olfactory, epiphysis, midbrain (mb) and hindbrain, as well as cranial muscle. The otic vesicle is outlined and cranial ganglia identified. **D.** Close up of the olfactory and epiphysis (ep) in the forebrain region. **E.** Spinal cord reporter localisation, alongside the dorsal root ganglia (DRG) which are labelled by Hu/ELAV. Additional abbreviations: e, eye. g/all, geniculate/anterior lateral line ganglia. mmV, maxilomandibular trigeminal ganglion. opV, ophthalmic trigeminal ganglion. p, petrosal ganglion.

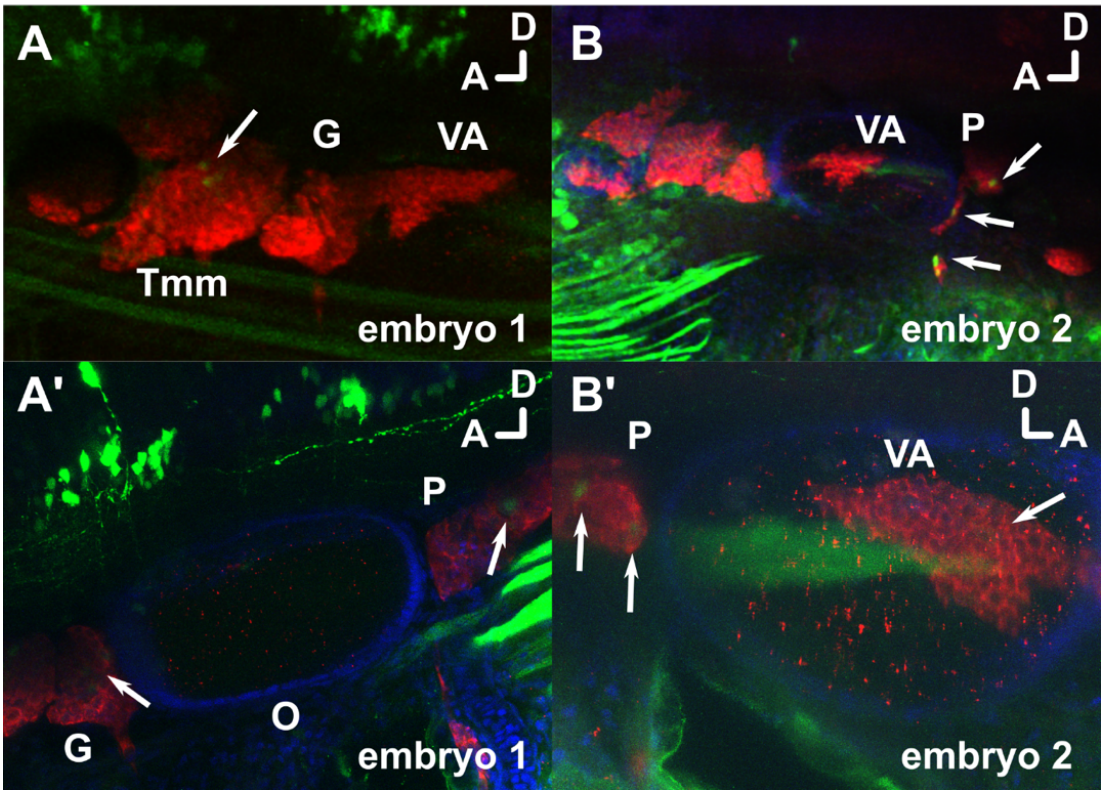


Figure 4.7. Cranial sensory ganglia reporter expression in *ERR8t12::GFP* transgenic lamprey embryos (*L. planeri)*. A and **A'** show different depth scans of the same embryo. A. few labelled cells are visible in the maxillomandibular ganglion (Tmm) in A, and one cell in the geniculate (G) and one in the petrosal/anterior lateral line ganglia (P) in A' (white arrows). **B** and **B'** show a different embryo scanned from left and right sides respectively (hence AP orientation in reversed in B' compared to B). The white arrows indicate labelled cells in the petrosal and vestibuloacoustic ganglia. Additional abbreviation: O, otic vesicle.

Embryos are labelled for GFP (*ERR8t12* activity, green), Hu/ELAV (neurons, red) and DAPI (DNA: blue).

If sequence conservation reflects functional constraint due to conserved upstream regulation, the ability of *ERR8t12* to drive reporter expression should extend to other vertebrates. To explore this, the ability of *ERR8t12* to drive specific expression of reporter genes in transgenic zebrafish and transfected chicken embryos was tested. A full description of zebrafish and chicken transgenesis/transfection approaches is given in chapter 2. Transgenic zebrafish were generated using the Ac/Ds transposition system. Embryos were stained with antibodies against GFP and Hu/ELAV and confocal imaging revealed *ERR8t12* activity in forebrain, hindbrain, eye, the otic vesicle and VA (Fig. 4.8). Chicken embryo in-ovo electroporation allows transgenes to be targeted to particular tissues in defined developmental windows. The *ERR8t12* element was cloned into a modified form of pTK-EGFP (Chen and Streit, 2015). Embryos were electroporated at stage HH10 targeting transgene DNA to the head ectoderm, including cells from where the otic vesicle and VA develop. A vector containing a constitutively active promoter linked to Ds-Red was co-electroporated to confirm which cells received transgene DNA, embryos grown to stage HH17 or HH21, and reporter gene activity was evaluated. Imaging and sectioning of embryos showed *ERR8t12* was able to drive reporter expression specifically in both the otic vesicle and VA at both stages, but not more broadly in head ectoderm (Fig. 4.9).

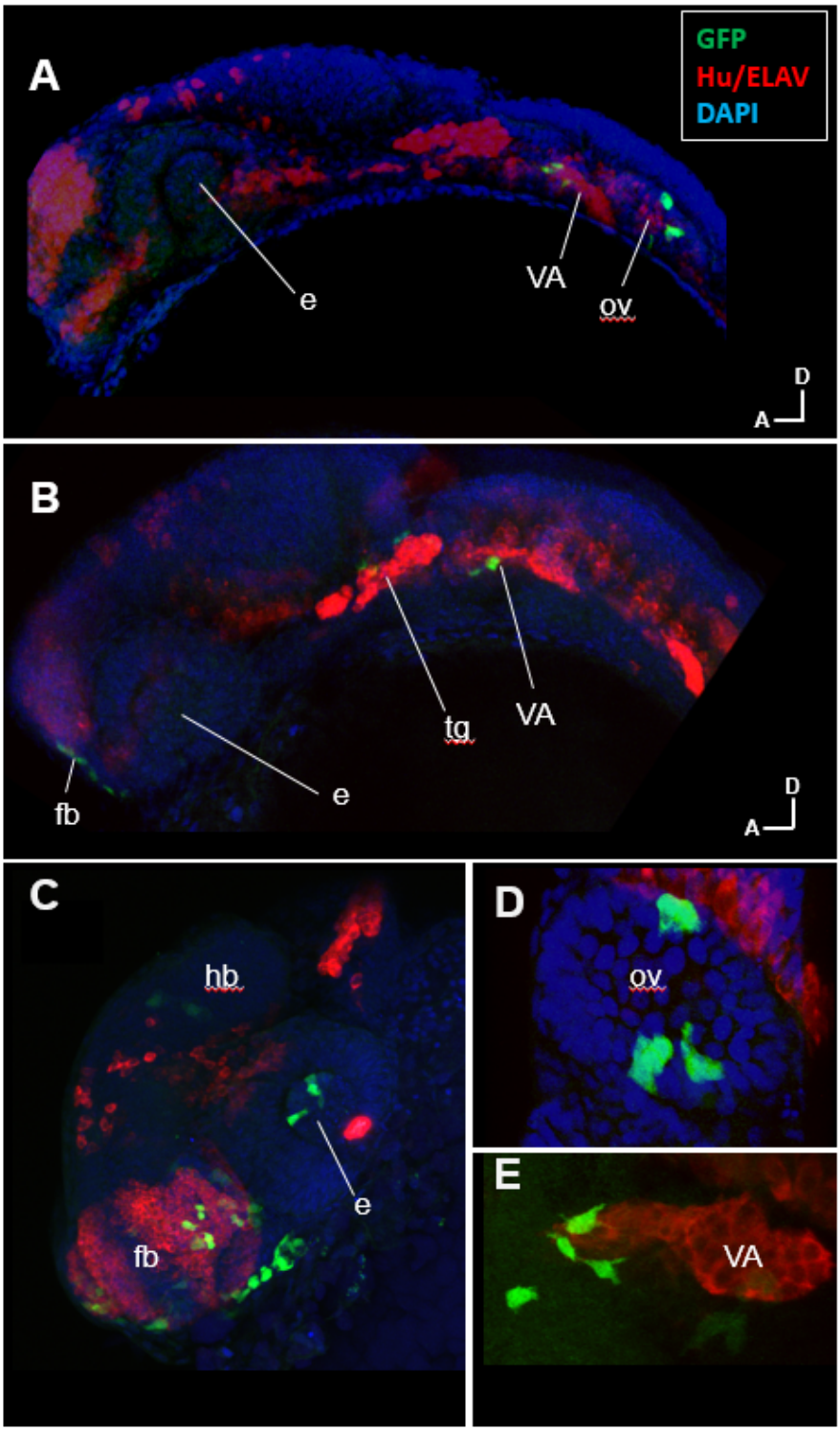


Figure 4.8. *ERR8t12* CNE activity in zebrafish embryos. Confocal reconstructions labelled for GFP (reporter, green), Hu/ELAV (differentiated neurons, red) and DAPI (nuclei, blue). Embryo orientations are shown in each image (A, anterior, D, dorsal). **A.** Embryo with GFP expression in otic vesicle (ov) and vestibuloacoustic (VA) **B.** Embryo showing reporter localisation in forebrain (fb) VA, and trigeminal (tg). **C.** Close-up of the head region with forebrain, eye (e) and hindbrain (hb) GFP visible. **D.** Close-up of otic vesicle region of the embryo in (A). **E.** Close-up of VA region of the embryo in panel (A).

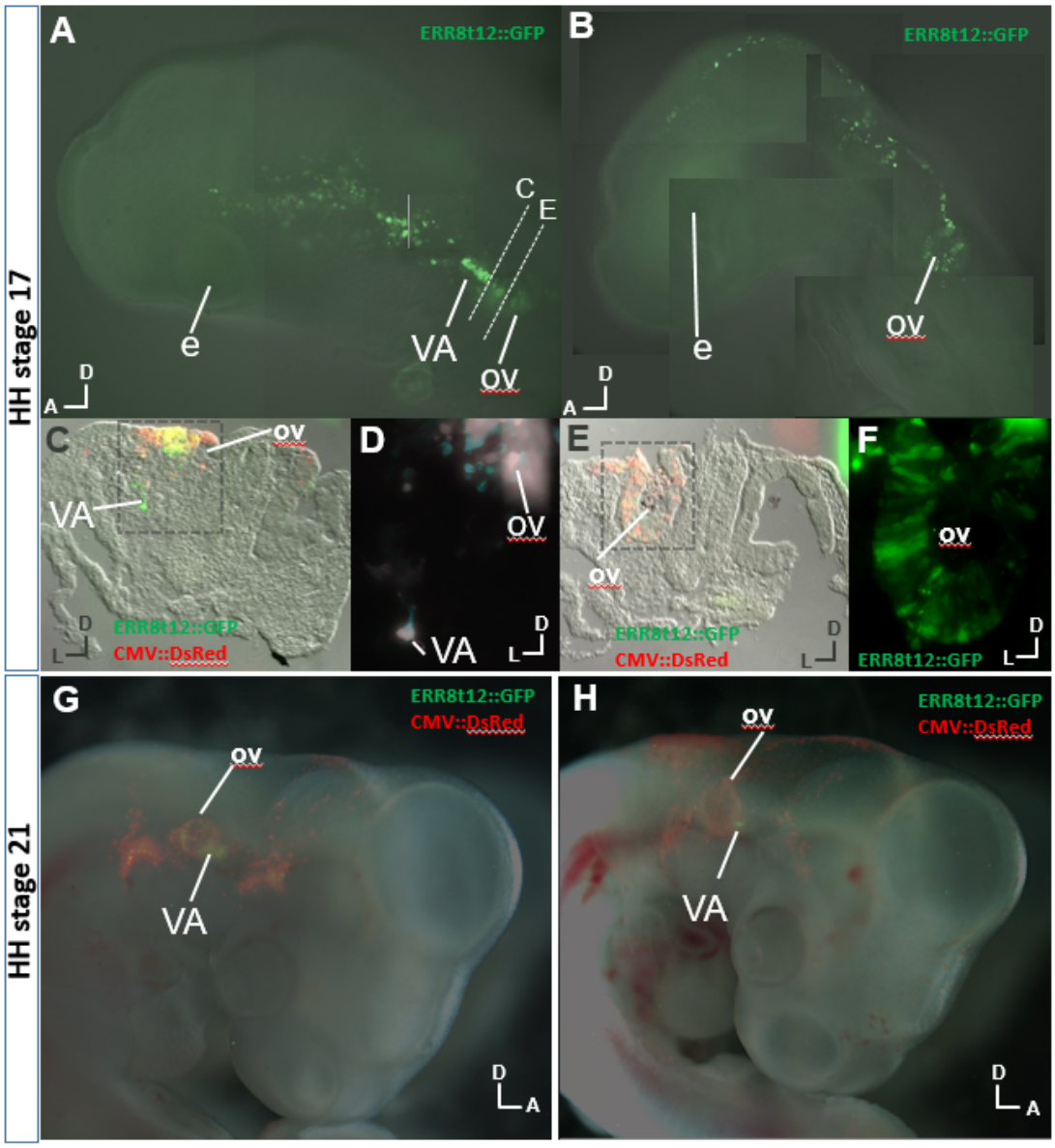


Figure 4.9. *ERR8t12* CNE activity in chicken embryos. A. HH15 embryo showing *ERR8t12* activity (green) in the otic vesicle (OV) and vestibuloacoustic (VA). **B.** HH17 embryo showing *ERR8t12* activity in the otic vesicle **C.** Transverse section at the level of the VA of the embryo in (A), as shown by the dotted line in (A), and overlaying DsRed and GFP. The framed region is shown in **(D)**, with an overlay of multiple sequential focal planes encoding depth as change in colour. Isolated neurons in the bottom left of the area, labelled VA. **E.** Transverse section at the level of the otic vesicle of in the embryo in (A), as shown by the dotted line. **F.** Close-up of the region framed in dotted lines in (E), with overlay of multiple sequential focal planes revealing green fluorescence in the otic vesicle. **G, H**. HH stage 21 embryos showing *ERR8t12* activity (green) in the otic vesicle (OV) and vestibuloacoustic (VA). DsRed driven by CMV in the control construct is also shown. Orientations dorsal (D) and anterior (A) or lateral (L) are shown.

Taken together, these experiments show that lamprey *ERRa* and jawed vertebrate *ERRγ* are expressed in the CNS and VA and share a suite of CNEs, including the *ERR8t12* block which is able to replicate aspects of *ERRa/ERRγ* gene expression in transgenic lamprey, zebrafish and chicken embryos.

4.2.5 The pre-vertebrate ancestry of *ERR* expression

The expression of *ERR* in both CNS and PNS in the amphioxus *Branchiostoma floridae* suggests expression of *ERR* by these tissues may predate vertebrates (Bardet et al., 2005). However, *ERR* expression in *Drosophila* has only been reported from the endoderm (Thurmond et al., 2019). Since further interpretation was hampered by a lack of data on *ERR* from other invertebrates, *ERR* genes were examined in other chordates and in a protostome. *ERR* loci were identified in a second amphioxus species, *Branchiostoma lanceolatum*, in the tunicate *Ciona intestinalis*, and in the gastropod mollusc *Patella vulgata (*the common limpet), confirming the identity of the genes by molecular phylogenetics (included in phylogeny in Fig. 4.2). These loci were searched for the CNEs identified in vertebrates, no conserved sequence in either the orthologous introns or elsewhere were found. *ERR* in situ hybridisation in *B. lanceolatum* confirmed both CNS and PNS expression, as well as expression in axial muscles (Fig. 4.10A-D). As seen in *B. floridae* (Bardet et al., 2005), PNS expression is confined to a small subset of cells with restricted AP and DV location.

In embryos of *C. intestinalis*, *ERR* is also expressed in axial muscle and in the sensory vesicle (part of the CNS), but not by PNS cells (Fig. 4.10E, F). Sensory

vesicle cells expressing *ERR* form part of the ocellus (the visual organ of the larva), plus two cells that form part of the lining of the vesicle in which the otolith, the larval intertial/gravity sensor, lies. These cells are known as ‘antenna cells’ and connect the otolith to the brain, probably relaying information about its displacement (Esposito et al., 2015; Ryan et al., 2017). Finally, in the trochophore larva of the limpet, *ERR* is expressed in the mantle edge surrounding the shell field, and by cells of the apical organ (Fig. 4.10G, I, J). The apical organ is a ciliated sensory organ with a central tuft of long cilia, and comparison to scanning electron microscopic images of embryos at the same developmental stage (Fig. 4.10H, K) shows expression is not in the central cells bearing elongated cilia, but in cells lying next to them.

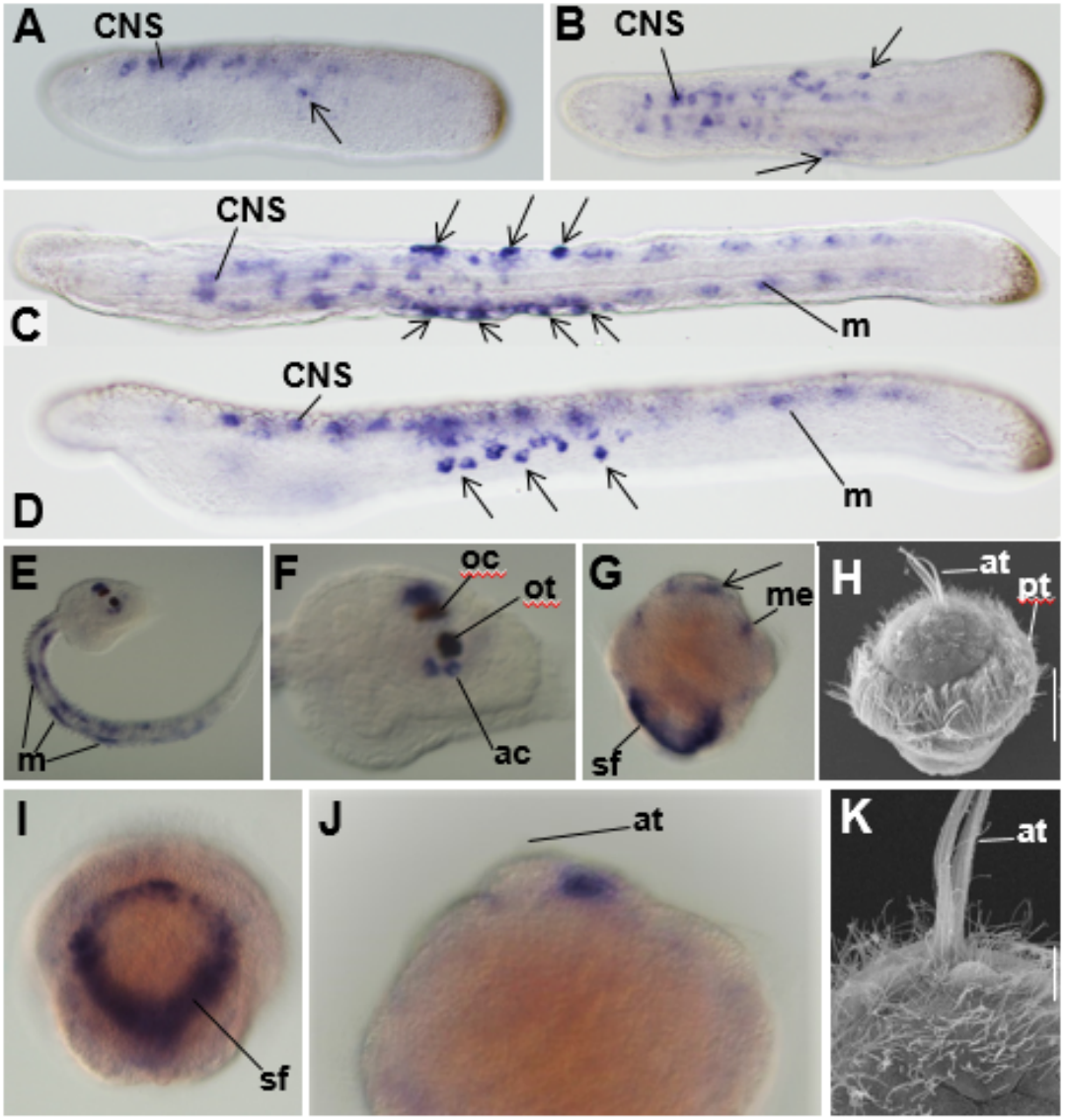


Figure 4.10. ***ERR*** **expression in amphioxus ascidian, and mollusc embryos. A-D.** Amphioxus (Branchiostoma lanceolatum) 32hr embryo (A,B) and early (42hr) larva (C,D) in lateral (A,D) and dorsal (B,C) views, anterior to the left. Expression is seen in the neural tube (CNS) and PNS (arrows), as well as in muscle (m) in the larva. **E, F**. *Ciona intestinalis* late tailbud stage. (F) is a close up of the head of the specimen in (E). Expression is seen in muscle and in ocellus (oc) and antenna cells (ac) by the otolith (ot) in the head. **G.** Limpet trochophore showing expression around the shell field (sf) and some anterior mesoderm (me) as well as in cells by the apical tuft (arrow). Close ups of the shell field and apical tuft (at) in different embryos are shown in **(I)** and **(J)**. **H, K**. SEMs of limpet trochophores at the same stage, focused on the apical tuft and the surrounding ciliated cells. Additional abbreviation: pt, prototroch. Scale bar on (H) is 50μm, on (K) is 10μm.

4.3 Discussion

4.3.1 Relationships of vertebrate *ERR* loci

The ER and ERR genes are thought to have originated from a single ancestral gene early in animal evolution (Bertrand et al., 2004). The analysis presented here suggests this occurred by tandem duplication, explaining why the two genes are linked in many lineages. However, the genes are not always linked and, when linked, distances between them vary and are generally very large in vertebrates. Persistent linkage is therefore probably the result of the chance maintenance of synteny in some lineages, rather than of a regulatory or functional constraint requiring their co-localisation. Molecular phylogenetic analysis generally resolved ER and ERR gene families well, as seen in previous analyses (Bertrand et al., 2004; Filowitz et al., 2018). While cyclostome sequences were placed within the vertebrate branch for each family as expected, they did not resolve with specific jawed vertebrate paralogs. A lack of clear orthology between jawed vertebrate and cyclostome loci has been documented and discussed previously (Kuraku, 2013), and examples include the *Pitx*, *Sox, Fox* and *KCNA* gene families among others (Boorman and Shimeld, 2002b; Kuratani et al., 2008; Qiu et al., 2011; Wotton and Shimeld, 2011). This prompted considering additional lines of evidence for gene relationships.

A starting point for this study was the hypothesis that *ERRγ* has a conserved role in the development of the VA. The expression of lamprey *ERRa* in the VA supports this, and the two genes also share similar expression in the CNS. To further

compare these loci, synteny and non-coding element conservation was considered, since these can provide a different type of data on gene relationships. Local synteny showed sufficient similarity between *ERR* loci in lamprey and jawed vertebrates to conclude they all evolved by block duplication of a single ancestral locus. While this does not prove this happened via 2R genome duplications, it is consistent with 2R and this is likely the simplest explanation. Amongst this broad similarity in locus organisation across *ERR* genes, the *ERRa* and *ERRγ* loci are most similar to each other, sharing flanking single-copy syntenic genes. They also share a suite of CNEs located in the same introns, with the same order and orientation relative to each other and the surrounding exons. Similar CNEs were not identified around other *ERR* paralogs. All of these features are indicative of orthology, though do not formally prove it since in principle they could have been inherited from a single pre-duplication *ERR* locus and lost by other paralogs. It is safe though to conclude that the similarity between these loci in lamprey and jawed vertebrates demonstrates this is an ancestral state for some, and possibly all, vertebrate *ERR* genes.

4.3.2 Common origin of VA expression and regulation of *ERRa* and *ERRγ*

A follow-on prediction from this conclusion is that the similar expression patterns of lamprey *ERRa* and jawed vertebrate *ERRγ* are controlled by their shared CNEs. The analysis of CNE function focused on the *ERR8t12* cluster in the second-to-last intron of the genes. *ERR8t12* activity in transgenic lamprey embryos showed overlap with multiple *ERRa* expression domains in the CNS, including the olfactory system, epiphysis, and along the spinal cord. This confirms the function of *ERR8t12*, demonstrating it contains regulatory information and is not some other

form of conserved genomic element. In addition, frequent reporter activity was seen in skin, head muscle and branchial arch, which do not express *ERRa*. These ectopic sites are probably due to the vector, which included a minimal promoter from a lamprey gene encoding keratin protein. In addition to these sites, confocal imaging showed reporter expression in cranial ganglia. Cranial ganglia expression suggests *ERR8t12* could also regulate *ERRa* expression in the PNS, though on its own this should be treated with caution since expression was of low frequency, was similar to that seen in transgenic embryos generated using only the reporter vector (data presented in chapter 2). However, *ERR8t12* cross-species activity provides evidence that this element can drive expression into the otic vesicle and VA.

Conserved sequence implies the CNEs in *ERR8t12* have been under evolutionary constraint, and raised the possibility they may function across species. When this was tested, it was revealed that *ERR8t12* shows VA and otic activity in chicken and zebrafish embryos. CNS activity was not tested in chicken embryos as reporter constructs were targeted to the ectoderm from which the placodes develop, but was seen in zebrafish embryos. Some variability between species was seen, though this this is not unexpected, as the *ERR8t12* region is large (2977 bp) and includes non-conserved sequences as well as those defined as CNEs. Nevertheless, the lack of activity in lamprey VA and otic is noteworthy. The large *ERR8t12* region could include sequence that acts as a repressor in this domain. Infrequent activity was recorded in these tissues (fig4.7), hinting at this possibility, though it could be vector background (see chapter 2 and section 4.2.4 above). Isolating and testing the three conserved CNE in this area could assess this. Alternatively, it is possible that the elements drive otic and VA activity in different developmental stages than the ones

assessed here. Finally, some of the other CNE may drive otic and VA activity in lamprey instead. In this scenario, a change in the upstream network in jawed vertebrates or lamprey would have led to gain or loss of *ERR8t12* activity in this domain respectively. This could be supported by differences in lamprey and jawed vertebrate ear development, such as the lack of *Otx* expression in the lamprey ear (Hammond and Whitfield, 2006; Maklad et al., 2014). Overall, the activity of *ERR8t12* in lamprey and two model vertebrates, with some shared expression between the latter, demonstrates *ERR8t12* has a conserved regulatory role, driving gene expression into morphologically conserved regions of developing vertebrate embryos. Sequence analysis also identified a CNE in the most 3' intron of *ERRa* and *ERRγ*. This sequence is included in the *EL161* CNE previously shown to be conserved within teleost fish and capable of driving reporter expression into the CNS in zebrafish (Roberts et al., 2014). Combined, these data show that the expression of this subset of *ERR* genes in extant vertebrates was inherited from the expression of one ancestral *ERR* gene of stem-vertebrates. This ancestral expression was regulated by a suite of conserved regulatory elements present before the divergence of living vertebrates. This shows the gene regulatory network driving *ERR* expression into these conserved sites evolved before the separation of lamprey and jawed vertebrate lineages and is hence a vertebrate character.

4.3.3 *ERR* expression in sensory cells predates vertebrates

One previous study suggests that *ERR* expression in CNS and PNS sensory cells may be older still, since both were described in amphioxus (Bardet et al., 2005). Relevant data on other invertebrate *ERR* genes are limited, however, as most

invertebrate studies have focused on the potential relationship between ERR and the environmental pollutant bisphenol-A, to the neglect of embryo and functional study (Bannister et al., 2013; Morales et al., 2018; Park and Kwak, 2010). To fill this gap, ERR expression was evaluated in the two chordates, and in a protostome outgroup. Tunicates, the sister group to vertebrates, show many aspects of conservation with vertebrates, including in development of the nervous system (Davidson, 2007; Hotta et al., 2000; Imai et al., 2009; Mazet et al., 2005). *ERR* CNS expression was confirmed in the tunicate *Ciona,* but PNS expression could not be identified. Expression is seen in the sensory vesicle in cells of both sensory organs, the ocellus and the otolith. Expression in the ocellus is interesting, as the organ has been suggested to be homologous to the vertebrate retina (Kusakabe et al., 2001). Different *ERR* paralogs have been reported in the retina in different vertebrates. Retinal *ERRγ* expression is seen in human, mouse and in zebrafish, where it specifically marks the ganglion cell layer (Bertrand et al., 2007; Heard et al., 2000; Hermans-Borgmeyer et al., 2000). *ERRδ* is also found in the zebrafish retina (Bardet et al., 2004). Finally, *ERRβ* is also expressed in both retinal ganglion cells and is necessary for rod photoreceptor survival in mouse (Onishi et al., 2010; Real et al., 2008). Furthermore, amphioxus *ERR* is expressed in the frontal eye, suggested to be a homolog of the vertebrate eye (Vopalensky et al., 2012) and the photoreceptor cells of Hesse (Bardet et al., 2005). These data support *ERR* expression in the eye and CNS likely consists an ancestral chordate character. Expression in the eye could have been missed in the lamprey ISH screen, as it appears to be very specific and in few cells. Future sectioning and staining of eyes from embryos of different stages would be necessary to follow this up.

Expression of *Ciona ERR* in two CNS cells associated with the otolith is also intriguing. The *Ciona* otolith functions as a gravity sensor (Tsuda et al., 2003), helping the swimming larva swim downwards when the time comes for it to settle and metamorphose. The system is thought to work as an inertial sensor, detecting displacement of the otolith itself. *Ciona ERR* expression near the otolith is in cells known as 'antenna cells', the neurons that relay sensation of otolith displacement to the processing part of the brain (Ryan et al., 2017). This has a clear functional parallel to otic and VA function in vertebrates and suggests an ancient association of *ERR* with inertial sensing. However, it is important to note the otolith does not appear to be formally homologous to the otic or ear structures when classical definitions of homology are used, and antenna cells are not formally homologous to VA cells under these criteria. This is because in *Ciona* these cells develop from within the CNS, not from ectodermal placodes that lie outside the CNS, as in vertebrates. Placode homologs may well be present in *Ciona* (Horie et al., 2018; Mazet et al., 2005), but *ERR* expression is not identified in their derivatives. Homology in these cells would hence be at the level of the cell type, in the manner proposed by Arendt and colleagues (Arendt, 2008; Arendt et al., 2016a), and not connected to position in the body or cell lineage.

In two species of amphioxus *ERR* expression is found in epidermal sensory neurons (part of the PNS) as well as in the CNS, which includes the frontal eye and photoreceptor cells of Hesse previously discussed. Amphioxus epidermal sensory neurons are born in the ventral embryonic ectoderm as scattered precursors (Lu et al., 2012), and delaminate to lie between the epidermis and underlying mesendoderm, leaving their sensory cilium projecting through the epidermis to the

outside world (Mazet et al., 2004; Schubert et al., 2004). They are comparable to vertebrates as peripheral sensory cells, but not directly to more complex structures like placodes and/or ganglia due to their scattered, ventral embryonic origin. Previous study of neural marker gene expression in amphioxus has shown there are many more epidermal sensory neurons than those marked by *ERR* expression, with *ERR* specific to a subset of cells on the middle of the animal with restricted AP and DV location (Candiani et al., 2006; Mazet et al., 2004; Satoh et al., 2001). These cells are alongside the part of the amphioxus neural tube homologous (on the basis of Hox expression) to the hindbrain. The function of these and other amphioxus sensory cells has not been elucidated, though it is possible they are mechanosensory, as the animal is known to respond to touch on its flanks (Parker, 1908). In addition, some studies have also looked at the structure and morphology of amphioxus ESN, attempting to assess their likely function. Two main type of ESN have been described, type I (with two variants) and type II. Type II ESN are secondary sensory cells with extensively branched microvilli surrounding their cilium. Originally suggested to be mechanoreceptors, this has been debated, as their morphology could be indicative of a chemoreceptor adaptation (Bone and Best, 1978; Lacalli and Hou, 1999). In contrast, there is consensus that type I ESN, primary sensory cells with unbranched microvilli, most likely consist mechanoreceptors (Baatrup, 1981; Lacalli and Hou, 1999).

ERR expression in limpet embryos suggests sensory neural expression may be an even more ancient character, despite *Drosophila ERR* appearing to be confined to the endoderm. The common limpet produces a typical planktonic trochophore as found in several phyla including annelids, nemerteans and molluscs. Trochophores

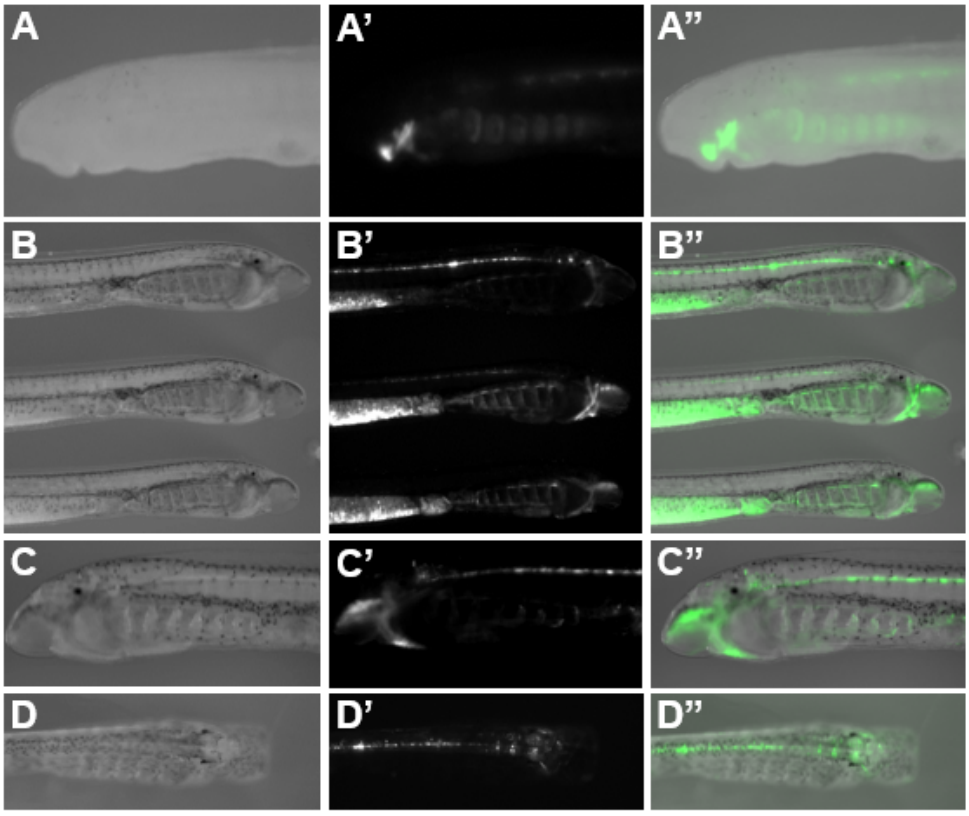
have an apical organ built from ciliated sensory cells, and have closer ecological comparison to the planktonic dispersive larvae of *Ciona* and amphioxus than the derived embryo and larva of *Drosophila*. What is more, the invertebrate ciliated larva and the apical organ have been suggested to be ancestral in neuralian animals, based on shared gene expression and patterning by Wnt signalling. Specifically, the developing apical organ of many invertebrates has been shown to be characteristically marked by the conserved expression of several genes, including *Nkx3*, *Irx* and *Hox* (Marlow et al., 2014). *ERR* expression in the limpet apical organ supports the interpretation that *ERR* expression in a subset of sensory cells may be primitive for bilaterian phyla. In some animals with a trochophore larva the apical organ mediates multiple stimuli including mechanosensation and chemosensation (Marlow et al., 2014), while it has been shown to be involved in the sensing of settlements cues in the gastropod *Phestilla sibogae* (Hadfield et al., 2000). Given the different cell types observable in the limpet apical organ (Figure 4.10H, K) it is likely it also mediates multiple stimuli, though the functions of individual cells are unstudied. So, while we can infer an ancient sensory role for ERR in the limpet, we cannot assign this to a specific sensory modality.

4.3.4 Conclusions

This study suggests that *ERR* has an ancient association with the formation of sensory cell types. In invertebrate chordates this manifests in expression in a subset of PNS cells in amphioxus, and in neurons of the gravity sensing organ in *Ciona*. Homology here is at the level of cell type (Arendt, 2008; Arendt et al., 2016a), and similarity between amphioxus and vertebrates suggests *Ciona* is the derived

lineage. Early in vertebrate evolution *ERR* evolved a set of CNEs in its 3' introns which regulate conserved expression in CNS and VA. Whether this predated or postdated the 2R genome duplications cannot be determined, but subsequently one ERR gene in all living vertebrate lineages has retained this CNE architecture, and CNS and VA expression, over the many hundreds of millions of years of evolution that followed.

Chapter 4 Supplementary Data



E. Scoring of GFP in live transgenic lamprey embryos. Expression in head mesoderm and musculature (for example as in A and C above) were not scored as these are ectopic sites that do not overlap with endogenous gene expression. Ganglia and finer detail in the brain were difficult to observe in live embryos and are explored by confocal imaging in Fig. 4.6 . 295 embryos were generated in total, data are show by embryo and by expression site

Expression by embryo

Location of expression	Number of embryos
Brain, posterior neural tube, branchial arches	80
Posterior neural tube, branchial arches	67
Posterior neural tube, brain	27
Posterior neural tube only	25
Branchial arches only	29
No expression	67
Total scored	295

Expression by site

Expression site	Number of embryos	Percentage of total (295)
Branchial arches	176	59.7%
Posterior neural tube	119	40.3%
Brain	107	36.3%

Figure 4.S1. *ERR8t12* GFP scoring in *P. marinus*. (Data kindly provided by Dr Dorit Hockman).

Chapter 5 *Hmx* conservation identifies a shared origin for vertebrate and tunicate sensory neurons

Work Declaration

Experimental work and data analysis presented in this chapter have been performed by myself (Vasileios Papadogiannis), apart from experiments and data described otherwise in this declaration. Sea lamprey (*P. marinus)* embryo microinjections were performed by Dr Hugo J. Parker and *B. lanceolatum Hmx* in situ hybridisation data were kindly provided by Dr Cedric Patthey. In addition, some of the figures and text are included in a publication currently under revision for resubmission on which I am the first author.

5.1 Chapter Introduction

Most animals have a clear AP axis, with clustering of sensory systems at the anterior. In vertebrates these systems are served by CSG, collections of sensory neurons which connect receptor cells in the sense organs to the brain. Combined, the ganglia mediate most cranial sensation, including touch and pain via trigeminal ganglia, sound and balance via the VA ganglia, gustation and internal gas sensing via epibranchial ganglia and, in many aquatic vertebrates, sensation of water-borne vibration via lateral line ganglia. These variable populations of sensory neurons all share similar development, covered in detail in section 1.4 of the general introduction. CSG cells derive from two sources. Placodes provide neural

progenitors, which delaminate and differentiate as they gather to form the developing ganglia. Neural crest cells migrate from the dorsal neural tube and contribute the glia and some neurons (mainly in the trigeminal) of the ganglia (Schlosser et al., 2014).

Cranial sensory ganglia are one of the evolutionary characters that appear specific to vertebrates, and form part of the influential 'New head Hypothesis'. This essentially proposed that a suite of characters in the vertebrate head relating to feeding, sensation and neural processing underlay the evolution of the vertebrates and their ecological transition from filter feeders to predators (Gans and Northcutt, 1983; Northcutt and Gans, 1983). However, the closest extant non-vertebrate chordate relatives of vertebrates lack CSG, as discussed in chapters 1 and 3 (Patthey et al., 2014; Schlosser et al., 2014). The consequent anatomical gulf between surviving lineages is substantial, raising significant obstacles to understanding how characters like CSG evolved. Comparison of gene expression and regulatory networks involved in sensory development of different chordates has provided valuable information to this end. It has become obvious that CSG cell types predate vertebrates, but no organised CSG-like structures exist in invertebrate chordates. Amphioxus lacks organised sensory structures, with scattered epidermal sensory neurons showing CSG marker gene expression instead. In addition, these sensory cells have a different developmental origin to CSG, deriving from a ventral neurogenic field. In tunicates, regions and individual cell types have been suggested to be homologous to placodal domains and their derivatives. Nevertheless, most of these regions give rise to primary sensory cells,

large cell proliferation is not observed and clear homology to CSG neurons has not been found.

5.1.1 *Hmx* CSG expression and genomic organisation in vertebrates

The work presented and discussed in chapter 3 provided a list of pan-vertebrate CSG markers, describing expression in lamprey and comparing this to chicken and other gnathostomes. One group of these marker genes, *Hmx* (also known as *Nk5* or *Nkx5*), are a relatively little studied family of homeobox genes (background information in section 3.1.4). As seen in section 3.2.5, three *Hmx* paralogs were identified in lamprey, closely linked in a single cluster. Lamprey *Hmx* paralogs are expressed in all CSG and the CNS from an early stage. They ubiquitously mark the trigeminal, VA and epibranchial ganglia. As seen in chapter 3, in jawed vertebrates, *Hmx3* is a VA and lateral line marker. However, other *Hmx* paralogs have been shown to be expressed in trigeminal and epibranchial ganglia, while all gnathostome *Hmx* are also expressed in the CNS (Adamska et al., 2001; Wang et al., 2000). What is more, experimental manipulation of different *Hmx* results in defects in ganglia development, showing *Hmx* function is crucial for normal CSG development (Wang et al., 2004, 2001; Weidong Wang et al., 1998). Interestingly, as in lamprey, *Hmx* genes also have an unusual clustered organisation in the genomes of model jawed vertebrates. Two linked pairs can be found, *Hmx3-Hmx2* and *Hmx1-SOHo* (Adamska et al., 2001). Therefore, comparable *Hmx* genomic organisation and expression in CSG and CNS is found in both jawed vertebrates and lamprey. This data suggested that an ancient *Hmx* gene cluster may have been involved in CSG development in the vertebrate common ancestor, before the split

of the cyclostome and gnathostome lineages. In turn, the possibility remained that expression in sensory cells could predate vertebrates, as *Hmx* in other chordates was completely unexplored.

5.1.2 Chapter Summary

In this chapter, a thorough comparison unfolds, looking into homology, genomic organisation and expression of cyclostome and gnathostome *Hmx* paralogs. *Hmx* sequences from different gnathostomes, lamprey and hagfish were used to build a comprehensive phylogeny of vertebrate *Hmx* genes. This data revealed cyclostome and gnathostome paralogs derived from two ancestral *Hmx* genes. An extensive literature review was also performed to compile evidence of *Hmx* expression in CSG in different gnathostomes. Taking this and lamprey expression into account, it becomes evident that *Hmx* expression in CSG is ancestral. Furthermore, close genomic analysis, focusing on non-coding conservation, identified a pair of CNE in *Hmx* loci that are shared by all vertebrates. Strikingly, phylogenetic analysis supports these are homologous to each other, sharing a large conserved sequence core. Together, gene and CNE data support a tandem duplication in stem-vertebrates producing an ancestral cluster containing two *Hmx* genes and two CNE. Transgenesis in lamprey showed both CNE drive expression in the CNS, while one drives expression in CSG.

Tracing the origin of *Hmx* expression and regulation in CSG in the vertebrate ancestor prompted an examination of *Hmx* genes in other chordates. While amphioxus *Hmx* is only expressed in CNS, *Ciona Hmx* expression was found in

both CNS and Bipolar Tail Neurons (BTN), a sensory neuron population developing from the edge of the neural plate. As vertebrate *Hmx* genes are necessary for CSG development, the role of *Hmx* in *Ciona* BTN was investigated. Using overexpression in the epidermis followed by RNAseq, it was shown that *Hmx* is able to upregulate the BTN differentiation program, while supressing epidermis and anterior sensory identity. Overall, these data demonstrate an ancient, conserved regulatory network controlling *Hmx* gene expression in vertebrates and an ancestral role of *Hmx* in the development of a specific subset of PNS cells in the ancestor of tunicates and vertebrates. It is inferred that CSG and BTN are homologous sensory populations that evolved from these ancestral placode-like PNS cells.

5.2 Results

5.2.1 Hmx gene expression marks all vertebrate cranial sensory ganglia

First, *Hmx* family expression in different CSG was examined in jawed vertebrates, in order to compare to lamprey *Hmx* expression, previously characterised in chapter 3. Expression data for *Hmx* genes compiled from model jawed vertebrate systems (mouse, chick, *Xenopus*, zebrafish and medaka) are summarised in Fig. 5.1A, while literature sources are listed in Fig. 5.2. Jawed vertebrates have four Hmx genes; *Hmx1-3* and a fourth gene known as *SOHo*, which was lost in mammals. Published in situ hybridisation data (Adamska et al., 2001; Feng and Xu, 2010; Kelly and El-Hodiri, 2016; Kiernan et al., 1997; Quina et al., 2012; Takahashi et al., 2003; Wang et al., 2000), plus transcriptome data from developing sensory neurons of the chicken (Patthey et al., 2016), show at least one *Hmx* gene is expressed by cells in

all cranial ganglia in all species, with the exception of the epibranchial ganglia in fish, where data are missing. These data are compared to lamprey *Hmx* expression in CSG in Fig. 5.1. As presented in chapter 3, the three lamprey *Hmx* showed identical patterns of expression, ubiquitously marking CSG (Fig. 5.1B-G). *HmxB* shows the strongest expression, which starts early in development, already marking CSG at stage 23 (Fig. 5.1B). By stage 24 CSG are strongly and specifically marked, while brain expression also appears, followed by expression in spinal cord by stage 25 (Fig. 5.1C, D). Based on these observations it was hypothesised that *Hmx* expression in all cranial ganglia was likely to reflect the ancestral vertebrate state. While gnathostome paralogs are expressed in CSG subsets, the *Hmx* family marks all CSG in both cyclostomes and gnathostomes.

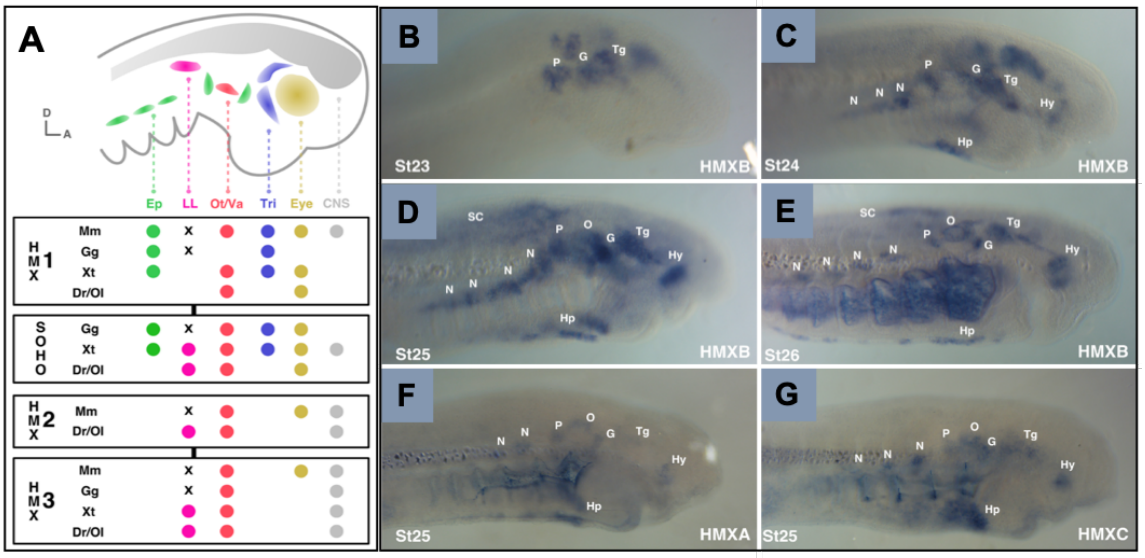
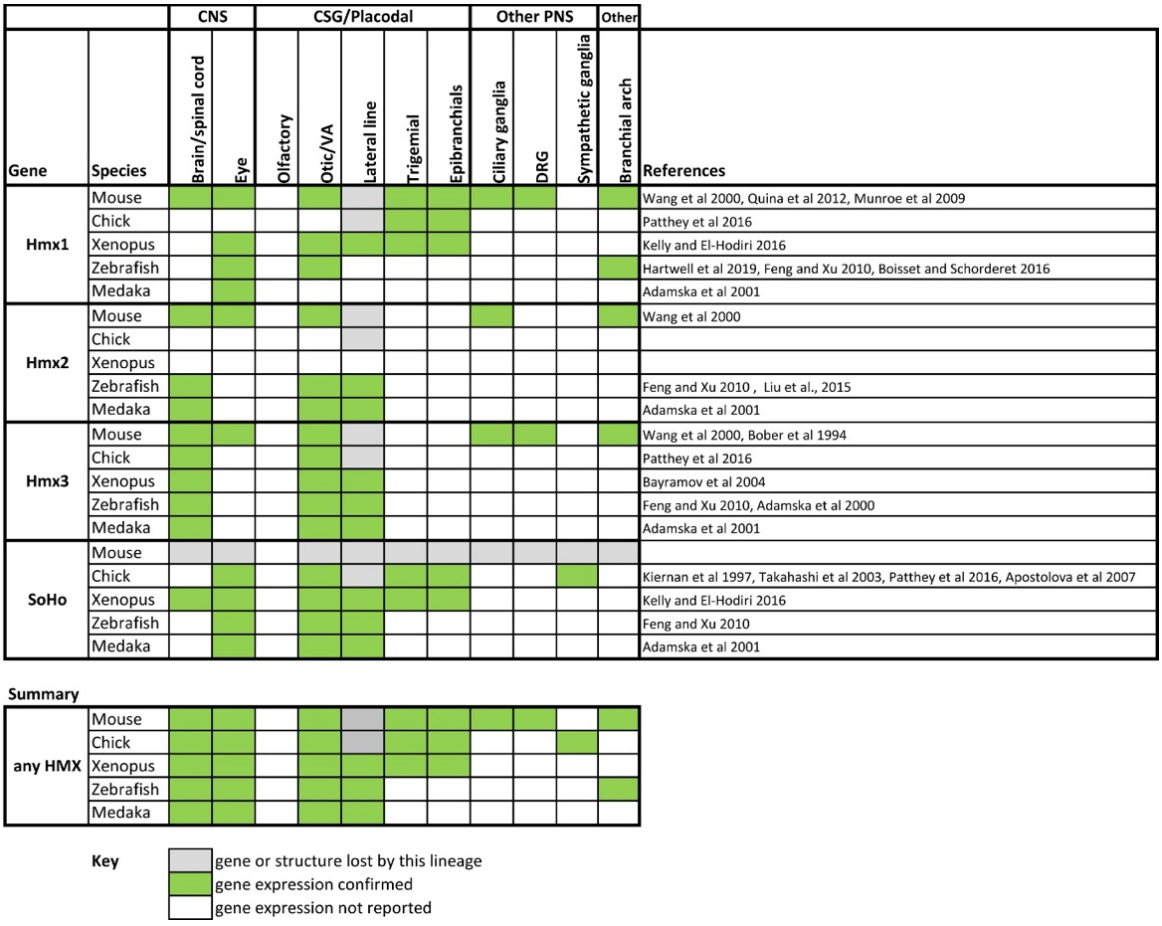


Figure 5.1. *Hmx* expression marks vertebrate cranial sensory ganglia. A. Schematic depiction of *Hmx* expression in jawed vertebrate cranial placodes/ganglia. Species shown are: Mm, *Mus musculus*. Gg, *Gallus gallus*. Xt, *Xenopus tropicalis*. Dr, *Danio rerio*. Ol, *Oryzias latipes.* Where species are missing it means either the gene has not been analysed (*Hmx2* in Gg and Xt) or the gene has been lost (*SOHo* in Mm). Other abbreviations are: Ep, Epibranchial. LL, Lateral Line. Ot/Va, Otic/Vestibuloacoustic. Tg, Trigeminal. CNS, central nervous system. X indicates lateral line ganglia have been lost by these species. The olfactory is not shown as Hmx has not been reported from this placode. **B-G** Expression of lamprey *HmxB* in different developmental stages. **C.** Expression starts in cranial sensory ganglia at stage 23. **D.** Hypothalamus expression in also seen by stage 24.

An expression domain below the branchial basket corresponds to the position of the hypobranchial ganglia. **F.** Hindbrain and spinal cord are strongly marked after stage 25. Expression of *HmxA* (**G**) and *HmxC* (**H**) becomes clear by stage 25 in cranial sensory ganglia and midbrain domains of *HmxB.* Panels D,F,G show the same embryos as Fig3.15 panels B,E and F respectively. Panels B,C and E show new embryos, adding st23 expression and complementary lateral views. Abbreviations. N, Nodose. P, Petrosal. G, Geniculate. O, Otic. Tg,Trigeminal. SC, Spinal Cord. Hy, Hypothalamus. Hp, Hypobranchial.



		CNS		CSG/Placodal					Other PNS			Other	
Gene	Species	Brain/spinal cord	Eye	Olfactory	Otic/VA	Lateral line	Trigemial	Epibranchials	Ciliary ganglia	DRG	Sympathetic ganglia	Branchial arch	References
Hmx1	Mouse	■	■		■		■	■	■	■		■	Wang et al 2000, Quina et al 2012, Munroe et al 2009
	Chick						■	■					Patthey et al 2016
	Xenopus		■		■	■	■	■					Kelly and El-Hodiri 2016
	Zebrafish		■		■							■	Hartwell et al 2019, Feng and Xu 2010, Boisset and Schorderet 2016
	Medaka		■										Adamska et al 2001
Hmx2	Mouse	■	■		■				■			■	Wang et al 2000
	Chick												
	Xenopus												
	Zebrafish	■			■	■							Feng and Xu 2010 , Liu et al., 2015
	Medaka	■			■	■							Adamska et al 2001
Hmx3	Mouse	■	■		■				■	■		■	Wang et al 2000, Bober et al 1994
	Chick	■			■								Patthey et al 2016
	Xenopus	■			■	■							Bayramov et al 2004
	Zebrafish	■			■	■							Feng and Xu 2010, Adamska et al 2000
	Medaka	■			■	■							Adamska et al 2001
SoHo	Mouse												
	Chick		■		■		■	■			■		Kiernan et al 1997, Takahashi et al 2003, Patthey et al 2016, Apostolova et al 2007
	Xenopus	■	■		■	■	■	■					Kelly and El-Hodiri 2016
	Zebrafish		■		■	■							Feng and Xu 2010
	Medaka		■		■	■							Adamska et al 2001

Summary

any HMX	Mouse	■	■		■		■	■	■	■		■
	Chick	■	■		■		■	■			■	
	Xenopus	■	■		■	■	■	■				
	Zebrafish	■	■		■	■						■
	Medaka	■	■		■	■						

Key
- gene or structure lost by this lineage
- gene expression confirmed
- gene expression not reported

Figure 5.2. Literature survey of jawed vertebrate *Hmx* expression domains.

5.2.2 Clustered organisation and non-coding conservation of vertebrate *Hmx* loci

Paralogs in many vertebrate gene families fall into broader paralogous genome regions, reflecting their origin via two rounds of whole genome duplication early in vertebrate evolution (Dehal and Boore, 2005), and with teleost fish sometimes having additional paralogs due to a third genome duplication in this lineage (Meyer and Van De Peer, 2005). This is more broadly covered in section 1.6 of the general introduction. Vertebrate *ERR* paralogs evolved in this fashion as discussed in chapter 4. In contrast, in the case of the vertebrate *Hmx* genes, previous work suggested their evolutionary history may be more complex than this, since *Hmx2* and *Hmx3* are closely linked (Wang et al., 2004), as are *Hmx1* and *SOHo.* This resulted in a model in which tandem duplication created an ancestral two-gene cluster, followed by duplication of this to give the two paralogous two-gene clusters (Adamska et al., 2001).

Lamprey *Hmx* paralogs are also closely linked in a single cluster, as discussed in chapter 3. Hagfish *Hmx* genes were additionally identified and mapped, to ask if this organisation is conserved in cyclostomes. In both cyclostome groups, three genes are found in a single cluster (Fig. 5.3C), supporting this is the ancestral condition. These data are compatible with the model of a tandem duplication of *Hmx* in ancestral vertebrates. Still, other evolutionary models could also account for this gene organisation in jawless and jawed vertebrates having independently evolved. In order to clarify this further, a molecular phylogenetic analysis of vertebrate *Hmx* proteins from cyclostomes and gnathostome was carried out. The analysis carried

out in chapter 3 used the homedomain sequence from vertebrate and invertebrate taxa to confirm new Hmx sequence identity. However, the Hmx homeodomain is highly similar across different Hmx members (see alignments in supplementary appendix) and the phylogeny of chapter 3 did not provide good support for member relationships. For this purpose, full Hmx proteins from different vertebrates are used here instead. The reconstructed phylogeny shows that vertebrate *Hmx* paralogs are separated in two groups. Cyclostome *HmxA* and *HmxC* are most closely related to each other and group together with gnathostome *Hmx2* and *SOHo,* while cyclostome *HmxB* groups with *Hmx3* and *Hmx1* (Fig. 5.3A). This clearly supported that a tandem duplication yielded an ancestral two-gene cluster with an *Hmx1/Hmx3/HmxB* ancestor linked to an *Hmx2/SOHo/HmxA/HmxC* ancestor. This duplicated to yield the two two-gene clusters of jawed vertebrates, while in the jawless vertebrate lineage the *HmxA/C* ancestor likely duplicated in tandem to give the current three-gene state.

Conserved genomic organisation and conserved expression in CSG prompted me to ask if gene regulation might be conserved. Genome comparisons across vertebrates have identified many conserved CNEs, and in some instances this conservation extends to lampreys (McEwen et al., 2009). Comparison of the *Hmx1-SOHo* locus between species did not identify any lamprey CNEs. Comparison of the *Hmx3-Hmx2* locus identified two CNEs also conserved in cyclostomes, which I named uCNE and dCNE since they lie upstream of Hmx3 and between *Hmx3* and Hmx2 respectively (Fig. 5.3C). In lamprey and hagfish, uCNE lies upstream of *HmxA*, and dCNE between *HmxB* and *HmxC* (Fig. 5.3C).

Conservation among uCNE sequences across species is quite extensive, as is conservation among dCNE sequences (Fig. 5.3D, 5.4A, B). While such conservation is relatively unusual between jawed and jawless vertebrates it is not unique (for example, (Parker et al., 2014a)), but their size at around 1Kb each (Fig. 5.4A, B) is rare (McEwen et al., 2009). However, the uCNE and dCNE sequences were also found to be homologous to each other, sharing a high level of similarity over several hundred base pairs of the central part of each element (Fig. 5.3D, 5.4A, B). This indicates that the CNEs evolved by tandem duplication from a single ancestral CNE. Since CNE duplication could in principal have happened once in the common ancestor, or independently in different vertebrates, I analysed their evolution using molecular phylogenetics (Fig. 5.3B). uCNE and dCNE sequences across vertebrates clustered into respective monophyletic groups, showing they are orthologous and derive from a single ancestral 'udCNE' that predates the radiation of vertebrates (Fig. 5.3E). This is in line with the model of a tandem duplication in the ancestral vertebrate *Hmx* locus. The two ancestral CNE and *Hmx* genes originated from the duplication of an initial udCNE-*Hmx* block in the ancestor of all vertebrates.

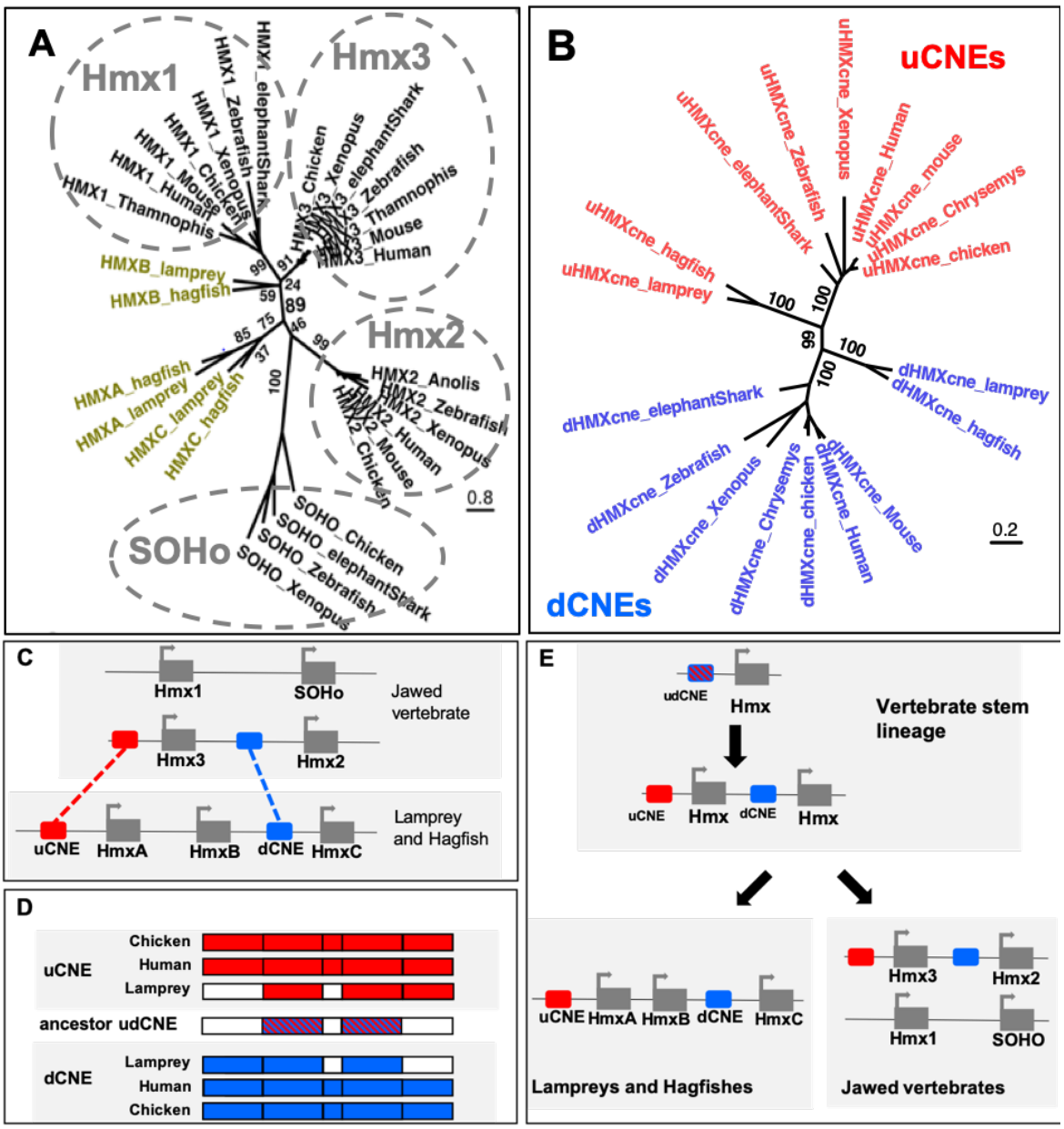
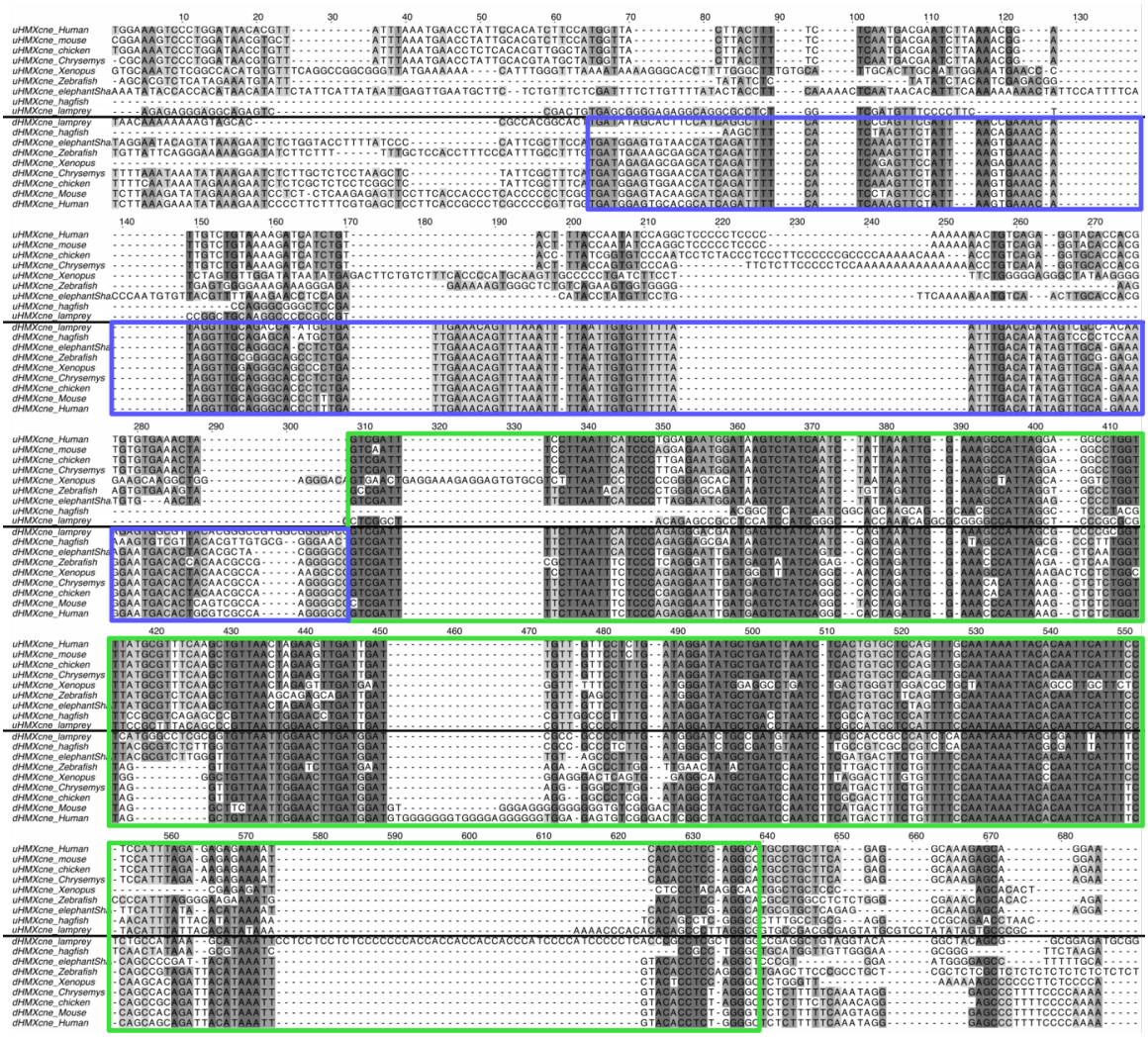
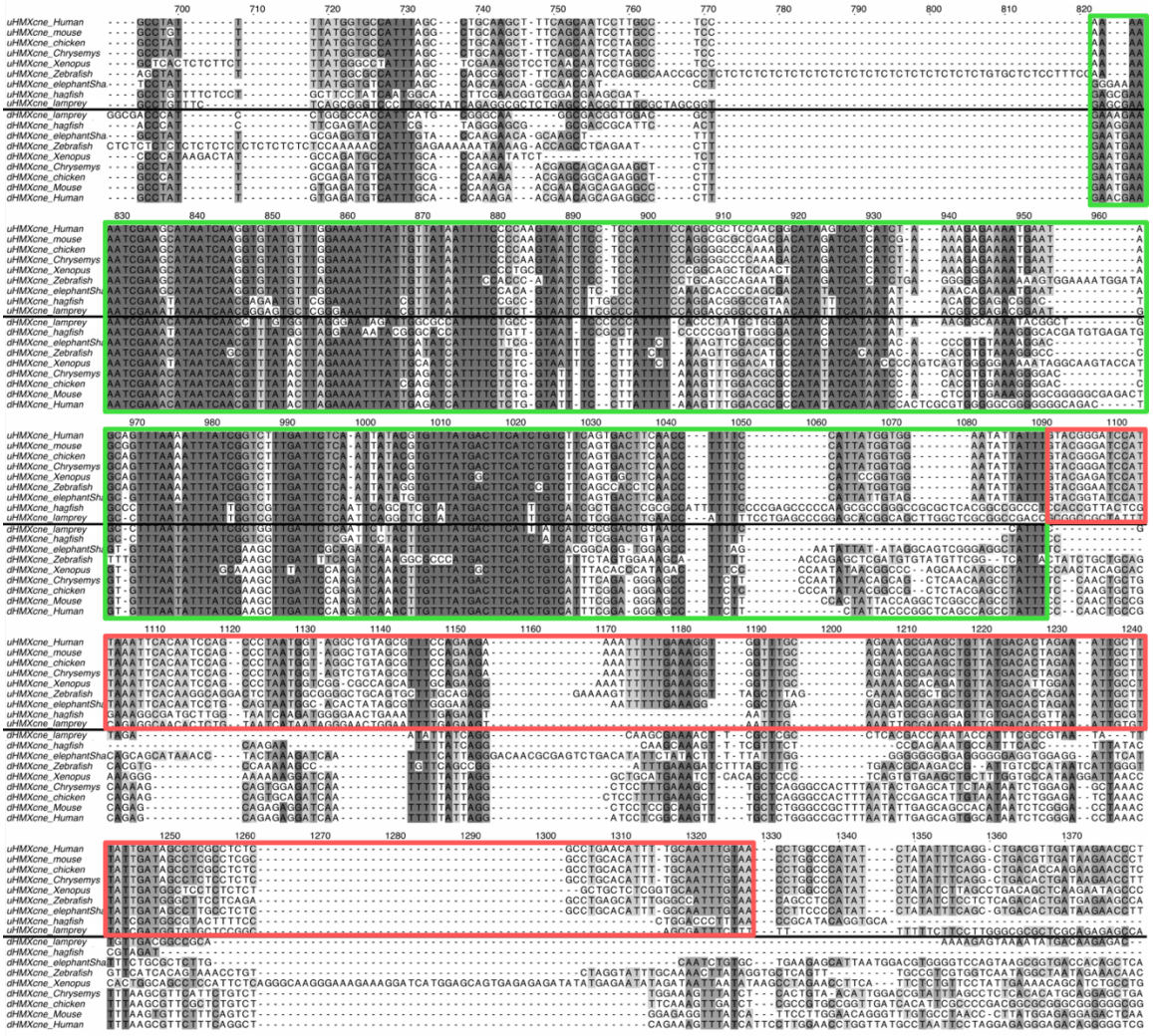


Figure 5.3. ***Hmx*** **gene and CNE phylogeny and evolution. A.** Phylogeny of vertebrate HMX proteins constructed using maximum likelihood (JTT, values indicate percentage support from 1000 bootstraps). **B.** Phylogeny of Hmx CNE sequences, showing uCNE and dCNE sequences form monophyletic groups (GTR+Γ, 1000 bootstraps). **C.** Schematic maps of Hmx cluster and CNE organisation in jawed and jawless vertebrate genomes. **D.** Comparison of sequence similarity between uCNE and dCNE sequences in lamprey, human and chicken. **E.** A model for the evolution of the Hmx locus. The current arrangements in vertebrates are inferred to have evolved from a single ancestral cluster with both uCNE and dCNE elements present, which itself evolved from a one-gene state with a single ancestral udCNE.



Figures 5.4. *Hmx* CNE sequence alignment. A: nt positions 1-690. First part of global alignment of uCNE and dCNE sequences from a range of vertebrates including lamprey and hagfish. The conserved core (first part here) is marked in green and dCNE specific alignment in blue.



Figures 5.4. *Hmx* CNE sequence alignment. B: nt positions 691-1379. Second part of global alignment of uCNE and dCNE sequences from a range of vertebrates including lamprey and hagfish. The conserved core (second part) is marked in green, and uCNE specific alignment in red.

5.2.3 *Hmx* CNE are enhancers that regulate expression in CNS and CSG

CNEs are usually inferred to be regulatory elements, with constraint on sequence evolution deriving from the need to maintain interaction with sequence-specific transcription factors of an upstream gene regulatory network. I speculated that the conservation of expression in CSG and/or in CNS across vertebrates might be linked to the evolution and extreme conservation of these CNEs. However, other genomic features might also lead to sequence constraint. Hence, to establish that these CNEs were regulatory elements, and to determine whether they have tissue-specific activities, their function was tested in lamprey embryos through transgenesis (Fig. 5.5A-G'). This was carried out using microinjection of reporter constructs in zygotes, implementing a mega-nuclease genomic integration system (methodology is covered in chapter 2). Both CNEs were able to drive tissue specific expression of the GFP reporter into regions of the embryo in which *Hmx* genes are expressed. Specifically, both lamprey uCNE and dCNE drive expression in the CNS, while confocal imaging revealed that uCNE additionally drives expression in CSG derivatives. Domains of reporter expression in the CNS were similar for both elements. As assessed by comparison to anatomical maps (Kuratani et al., 1997), uCNE showed activity in multiple cranial nerves, including CSG derivatives (nerves IX, X – see Fig5.5 and close ups in Fig5.S1). It also showed activity in anterior dorsal root ganglia nerves and motor neurons in which *Hmx* expression was not identified in the in-situ hybridisation experiments (Fig. 5.5F-G'), suggesting additional regulatory elements may serve to restrict its activity in vivo. Taken together, this analysis of lamprey *Hmx* gene expression and regulation shows that expression in CSG is a pan-vertebrate character, as are the uCNE and dCNE

regulatory elements that together drive expression in these structures and the CNS. As presented above these CNEs evolved from a single ancestral udCNE by tandem duplication, and parsimoniously this occurred when the original *Hmx* gene duplicated to give the two-gene cluster. Therefore, it can be inferred that this ancestral udCNE drove expression into the CNS as this property is shared by both of its descendants. Since ganglia expression was identified for only one element, this property might be ancestral or might have evolved after the duplication of the CNE.

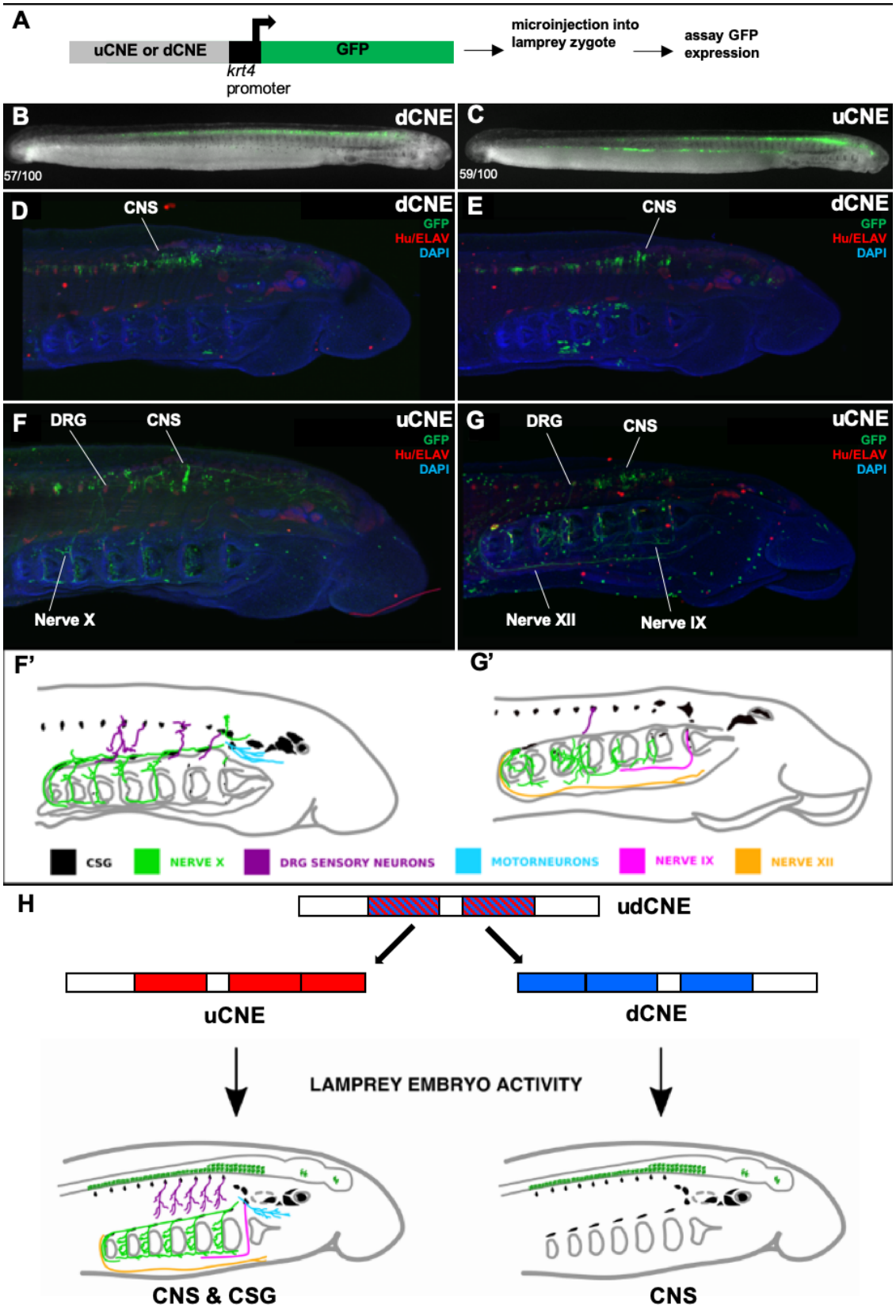


Figure 5.5. Hmx CNE enhancer activity in transgenic lamprey embryos. A. Experimental strategy for detecting reporter activity in lamprey embryos. **B, C.** Representative embryos showing *dCNE::GFP* and *uCNE::GFP* activity in the CNS. dCNE CNS activity was seen in 53% of stage 25 embryos and 57% of stage 27 embryos, uCNE

CNS activity was seen in 57% of stage 25 embryos and 59% of stage 27 embryos. n≥100 embryos in each case. **D, E.** Confocal reconstructions of lamprey embryos transgenic for *dCNE::GFP*, showing activity (green) in forebrain, midbrain, hindbrain and spinal cord. **F, G.** Confocal reconstructions of lamprey embryos transgenic for *uCNE::GFP*, showing activity in the CNS, and in peripheral nerves. In D-G ganglia and other neurons have been stained in red using an antibody against Hu/ELAV. **F', G'** Schematic tracing of embryos shown in C and D respectively, with nerve CNE reporter activity traced and colour coded. **E**. Schematic overview of Hmx CNE comparative structure and activity. See Fig.5.4 for CNE sequence alignments.

5.2.4 *Hmx* expression in invertebrate chordates

Having established that Hmx expression in CSG dates back at least to the base of vertebrates, it was next sought to investigate the ancestral state for HMX genes in chordates. The two invertebrate chordate lineages were explored, represented by the amphioxus *Branchiostoma lanceolatum* and the tunicate *Ciona intestinalis*. A single *Hmx* gene was identified in both species, but no evidence was found in either for sequence conservation of any of the vertebrate *Hmx* CNEs. To this end, sequence similarity search was carried out in the genome of each species via BLAST and more locally in the upstream and downstream region of *Hmx* using BLAST and MUSCLE, but no significant matches to dCNE or uCNE were retrieved. However, a large region 2kb upstream of *Hmx* in *Ciona* showed significant conservation between *C. robusta* and *C. savignyi* in the genome annotations of *C. robusta* (KH2012). This hinted to a possible tunicate specific conserved regulatory element. This has been recently investigated and the results of this experiment are included and discussed in section 6.3.1 of chapter 6. In amphioxus *Hmx* expression was found to be restricted to the CNS (Fig. 5.6A), including at stages where the numerous PNS cells scattered across the flank of the embryo are known to be present, as discussed in section 4.3.3 of chapter 4 (Fig. 5.6B; (Mazet et al., 2004; Schubert et al., 2004)).

Unusually for *Ciona*, the *Hmx* gene identified in the genome (Wada et al., 2003) had no EST data associated with it, and its spatial expression had not been determined (Brozovic et al., 2018). Cloning the gene and characterising its expression via in situ hybridisation showed that *Ciona Hmx* is restricted to a few cells in the brain, and to specific cells flanking the posterior neural tube (Fig. 5.6C-E). These posterior cells were identified as bipolar tail neurons (BTNs). These peripheral neural cells arise from the border between neural plate and epidermal ectoderm, from where they delaminate and migrate to become embedded under the ectoderm next to the neural tube (Stolfi et al., 2015). They are specified by an AP transcription factor code (Horie et al., 2018), and go on to make axonal connections between the tail fin primary sensory neurons and the brain (Ryan et al., 2016).

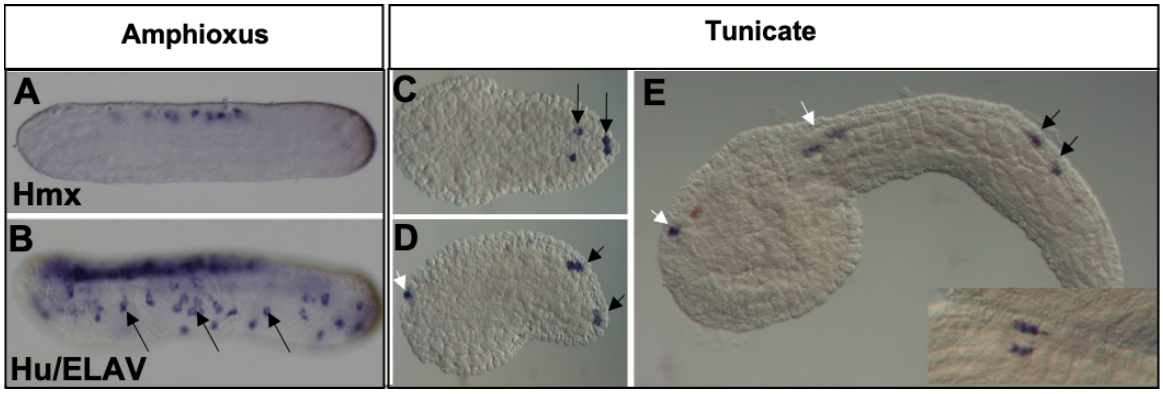


Figure 5.6. *Hmx* expression in invertebrate chordates. A, B. In amphioxus *Hmx* is confined to the CNS at stages where peripheral neurons are already specified, as revealed by *Hu/ELAV* expression (arrows on B). **C-E.** In *Ciona Hmx* is expressed by BTN cells lateral to the posterior neural plate (black arrows), and by cells of the anterior of the neural plate (white arrows). The inset on E shows a dorsal view of *Hmx* expressing BTNs lying alongside the neural tube.

5.2.5 *Hmx* drives sensory neuron identity in tunicates

The properties of BTNs have seen them likened to neural crest derived DRG cells, and to placode derived cranial ganglion cells (Horie et al., 2018; Stolfi et al., 2015). The characterisation of *Hmx* expression in these cells supports the latter, and the specific expression of *Hmx* by the BTNs and not by any other *Ciona* PNS cells, coupled with Hmx expression across the vertebrate cranial ganglia, suggests it may have an instructive role in generating these cells. Data from vertebrates support this, with *Hmx* manipulation yielding a range of sensory phenotypes (Quina et al., 2012; Wang et al., 2004, 2001; Weidong Wang et al., 1998). To explore this further, it was reasoned that if *Ciona Hmx* has an upstream control function in generating BTNs from a more general PNS progenitor, then its ectopic expression should generate supernumerary BTNs. Thus, *Ciona Hmx* was cloned under the control of the *epiB* promoter, which drives expression in non-neural epidermal cells (Joyce Tang et al., 2013). This construct was electroporated into *Ciona* zygotes and RNAseq was used to compare the transcriptome of *Hmx* overexpressing embryos to that derived from control embryos electroporated in parallel with a YFP construct (Fig. 5.7A, B). A detailed description of experimental methodology and follow up analysis is provided in chapter 2.

With an adjusted p value of 0.01 and an FPKM threshold of 2, differential expression analysis gave 272 genes that were significantly upregulated in *Hmx* overexpression embryos, and 414 genes that were significantly downregulated. Upregulated genes included known BTN marker genes such as *Islet* and *Neurogenin* (Stolfi et al., 2015). As individually these genes are also expressed by other neural cells, published single cell sequencing data were also leveraged (Horie et al., 2018; Sharma et al., 2019), which included identification of gene sets for BTNs, epidermis,

CNS and other sensory cells. These data were used to assess whether gene sets identified as enriched for these cell populations were found in the significantly *Hmx*-upregulated or downregulated gene sets (Fig. 5.7B, C). This comparison revealed 6 epidermal markers present in the *Hmx*-upregulated set, versus 100 in the *Hmx*-downregulated set. This shows *Hmx* overexpression in this tissue leads to a suppression of epidermal identity. Of the BTN markers, 38 were found in the *Hmx*-upregulated set, with 21 showing a greater than 2fold increase in expression. In contrast, no BTN markers were seen in the Hmx-downregulated set. Horie and colleagues (Horie et al., 2018) also demonstrated that manipulating the functions of early neural plate border patterning genes could transform BTNs into palp cells, a type of anterior sensory cell. If *Hmx* is involved in driving BTN identity downstream of these patterning genes as hypothesised, then it should also repress palp identity. This was indeed the case, with 31 palp specific markers in the *Hmx*-downregulated set, and only 4 in the *Hmx*-upregulated set. Overall, this experiment provided clear evidence that *Ciona Hmx* is a specific marker of the BTN lineage of peripheral sensory neurons, and that it is sufficient to convert other ectodermal cells to a BTN identify, including repression of palp and epidermal identity.

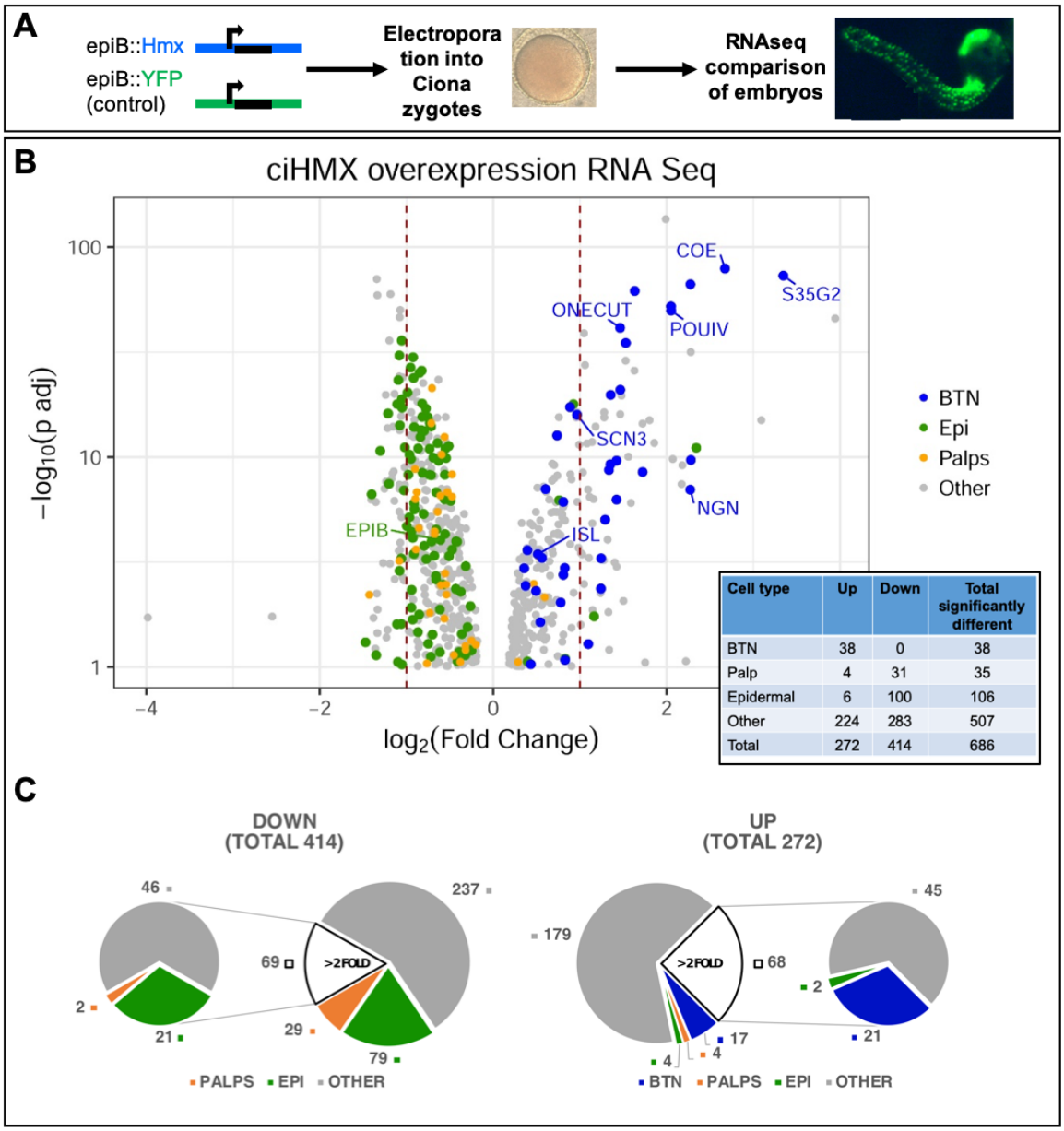


Figure 5.7. *Ciona Hmx* Overexpression RNAseq. A. Schematic of the *Ciona Hmx* overexpression experiment. **B.** Volcano plot of genes up or down regulated after *Ciona Hmx* overexpression. Selected genes are named, and colour coding reflects genes identified as enriched in a particular cell type in the single cell sequencing data of Horie et al (2018) and Sharma et al (2019). The inset table shows these quantified for BTN, palp and epidermal cell types. **C.** Pie charts showing proportion of downregulated and upregulated target markers. In each larger chart, the markers showing greater than 2fold change in expression are summed together in the “>2FOLD” category, which is then analysed separately in each smaller chart.

5.3 Discussion

5.3.1 Unique conservation and linkage in vertebrate *Hmx* loci

This study connects the evolution of the *Hmx* gene family with that of the peripheral sensory nervous system. All vertebrate CSG express one or more *Hmx* genes, including in members of the earliest diverging lineages of extant vertebrates, the lampreys. Focusing on the *Hmx3/Hmx2* locus, it is shown this is regulated by a unique set of CNE derived by tandem duplication. Comparative genomics and molecular phylogenetics show this duplication happened before the divergence of the jawed and jawless vertebrate lineages, and probably coincident with the first tandem duplication of the locus, that created the ancestral linked gene pair. Since this gene pair also subsequently duplicated to create paralogous pairs on separate chromosomes, this probably all happened before the 2R genome duplications that predate the radiation of jawed and jawless vertebrates (Sacerdot et al., 2018). This dates the duplication of the CNE to at least 450-510 million years ago (Shimeld and Donoghue, 2012). This is the only known example of a CNE which has duplicated in tandem then retained such levels of sequence conservation in both duplicates over such a long time, and it speaks to extreme constraint on the primary nucleotide sequence. Other examples of duplicated CNE in literature involve elements that lie in paralogous loci, a result of the 2R genome duplications (McEwen et al., 2006). Similar to the *Hmx* CNE, it has been suggested that such CNE became fixed prior to the genome duplications in the ancestral vertebrate locus. After the duplications, each paralogous CNE became associated with one respective paralogous gene. These duplicated paralogous CNE drive expression in overlapping, but not identical

domains, resembling the endogenous expression of the paralogous genes they are associated with. In this case, it is possible to assume that constraints of responding to a largely shared upstream regulatory network keep these CNE similar. Unlike other duplicated CNE, the *Hmx* CNE are a result of a tandem duplication and have been retained in a single locus.

The simplest assumption to explain the retention of both *Hmx* CNE in jawed vertebrates would be that one CNE is associated with *Hmx3* and the other with *Hmx2*. This is comparable to what is suggested for other duplicated CNE, but both lie in a single locus instead. The lamprey cluster houses the two CNE and three *Hmx* genes, creating a slightly more complicated case. The phylogeny presented in fig5.3 and the extended homeobox sequence similarity seen in fig3.13 support *HmxA* and *HmxC* are a result of a cyclostome specific tandem duplication. Hence, it could be possible that one or both are regulated by the CNE. Instead, it is also possible to hypothesise that the CNE synergise to regulate the genes found in the locus. Jawed vertebrate *Hmx3* and *Hmx2* show significant overlap in expression. The same is also seen for the three lamprey genes, while the domains of the CNE in the CNS largely overlap. Furthermore, the overlap in expression of *Hmx3 and Hmx2* with *Hmx1* and *SOHo* is also notable, since the latter lack related CNE. This implies that conserved *Hmx* expression across jawed vertebrate embryos is also regulated by other mechanisms. While this study shows the two CNE exhibit enhancer activity, this does not exclude the elements also have other functions as well. For example, they could also be involved in silencing the locus in other cell types and stages by guiding epigenetic modifications or 3D chromatin conformation changes.

5.3.2 *Hmx* expression, regulation and function trace the shared origin of CSG and BTN

The paralogous elements share a conserved sequence core with each other, but also share additional conserved sequence with orthologous elements in other vertebrates. In transgenic lampreys they differ in their ability to drive reporter expression, with both exhibiting activity in similar regions of the CNS, but only the uCNE in cranial ganglia. Initially it was speculated that duplication of the CNE may have been associated with the evolution of expression in the cranial ganglia, but other data reject this hypothesis. *Hmx* in *Ciona* is expressed by the BTN cells and is sufficient to drive BTN identity when ectopically expressed in other ectodermal cells. BTNs are lateral neural plate border derived cells that become part of the PNS and share evolutionary ancestry with vertebrate peripheral neural cells. BTN progenitors arise from the region lateral to the neural plate, migrate, differentiate and settle in the tail ectoderm, connecting primary epidermal sensory cells to the brain. They were originally suggested to be homologous to neural crest cells, owing to their similarity to DRG neurons, since they migrate and settle posteriorly to the brain in the tail. However, whether they are homologous to placodes, to neural crest, or to both, has been debated (Horie et al., 2018; Stolfi et al., 2015). Placodal progenitors also arise from the neural plate border and produce CSG neurons that migrate and differentiate before they settle. Since vertebrate *Hmx* genes mark placode derived cells, the data presented here support placode and BTN homology, but irrespective show a role for *Hmx* in development of a subset of peripheral neural sensory cells predates vertebrates.

Whether this role dates further back into bilaterian evolution is more equivocal. *Hmx* expression was not found in the PNS of amphioxus, throughout the period when these cells are born, specify and differentiate. However, amphioxus PNS cells are born in the most ventral ectoderm in a region of high Bmp signalling (Lu et al., 2012; Mazet et al., 2004; Schubert et al., 2004), also distinct from the development of *Ciona* BTN and vertebrate ganglia cells from the neural plate border. These data could suggest a role for *Hmx* in the PNS does not date further back in evolutionary time. However, *Hmx* is an old gene family, dating to the earliest animals (Fortunato et al., 2014). In *Drosophila melanogaster,* CNS expression has been reported, but not PNS expression (Thurmond et al., 2019; Wang et al., 2000). In *Caenorhabditis elegans* neural expression of the *Hmx* ortholog *mls-2* has been reported, including in the AWC neurons, which are sensory (Kim et al., 2010). A provisional report in the annelid *Platynereis dumerilii* suggests both CNS and PNS expression, though where in the PNS was not described (Saudemont et al., 2008), while in the sea urchin *Stronglyocentrotus purpuratus*, no neural expression was noted (Martinez and Davidson, 1997). Hence, while neural expression has been described from animals from four phyla and is likely ancestral at least for bilaterians, evidence for conserved PNS expression outside the chordates is lacking but suffers from scarcity of appropriate study.

5.3.3 Conclusions and future prospects

This study provides data that build into a model linking *Hmx* gene family evolution and the chordate PNS. Assuming amphioxus reflects the ancestral chordate

condition, early chordates had a single *Hmx* gene with CNS expression. In the ancestor of the tunicates and vertebrates this acquired expression in a subset of cells deriving from the lateral neural plate border: the cellular ancestors of the BTNs in *Ciona* and CSG neurons in vertebrates, in both of which *Hmx* has retained a regulatory role (this study; (Wang et al., 2004, 2001; Weidong Wang et al., 1998)). Significant change then occurred in the vertebrate stem lineage. *Hmx* duplicated in tandem along with its large regulatory element, which is inferred to have driven expression in both CNS and PNS. The duplicated regulatory elements diverged to some extent, but came under high constraint in both jawed and jawless vertebrates resulting in the extensive sequence similarity seen between orthologous and paralagous CNEs in living vertebrates. Additional gene duplications have occurred in both jawless and jawed vertebrate lineages, but have not changed this regulatory architecture. This ‘locking in’ of genomic regulatory architecture suggests a need to respond to a complex upstream regulatory network, and correlates with the evolutionary origin of vertebrates themselves. Deciphering the detail of this network, and how it relates to *Hmx* networks in *Ciona* and amphioxus, will be will be a challenging task, but one with the potential to illustrate how this evolutionary transition occurred.

Chapter 5 Supplementary Data

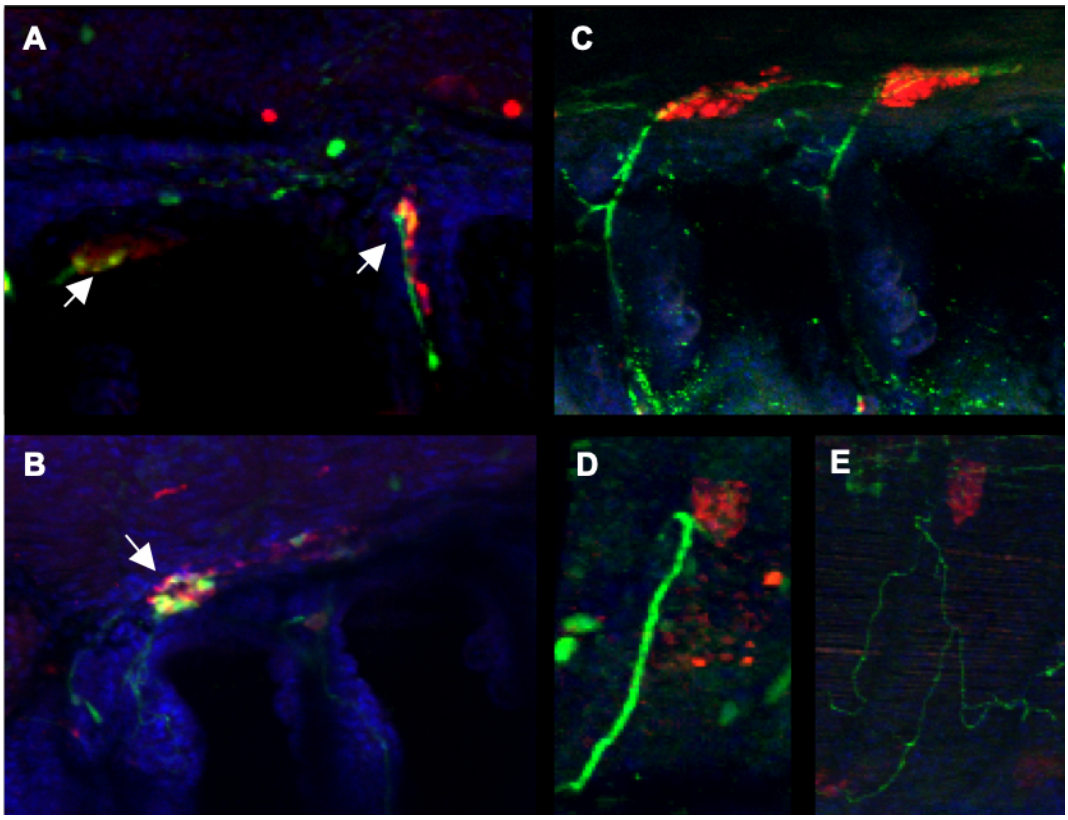


Figure 5.S1. Hmx CNE enhancer activity close up views. A.B. GFP and Hu/ELAV co-staining seen in CSG in the anterior (A) and posterior (B) branchial basket of uCNE injected embryo shown in Fig5.5G. **C.** GFP staining in cranial nerve X at the level of the epibranchial ganglia in uCNE embryo injected embryo shown in Fig5.5F. **D.E.** GFP staining in spinal nerves at the level of DRG in Fig5.5G embryo (D) and Fig5.5F embryo (E).

Chapter 6 General Discussion

The work presented in the preceding chapters approached the research questions set in the general introduction in a step wise fashion. The lamprey was the first stop in the search for the ancestral molecular signature of vertebrate sensory development. The exploration of the genomic evolution, expression and regulation of *ERR* and *Hmx* offered two complementary channels for comparing CSG cells to invertebrate sensory populations. To better consider how these data come together to advance our understanding of CSG evolution and development, I will revisit the main themes and questions, adding new information and discussing how it complements or changes existing ideas.

Work Declaration

The review and discussion in this chapter were carried out by myself (Vasileios Papadogiannis). *C. intestinalis* CNE transgenics data shown in fig6.1A, B and C were kindly provided by Alessandro Pennati and Professor Sebastian Shimeld.

6.1 CSG markers highlight different levels of homology and co-option

The different CSG marker genes characterised in lamprey in chapter 3 are heterogeneous in their expression onset and differ in the degree of expression site specificity. These genes were originally identified as expressed in CSG neurons in chicken (Patthey et al., 2016) and can be placed in two categories based on their

developmental timing. On one hand, expression of earlier marker genes that are activated at placodal stages can persist later. Such examples include *Pax3* and *Otx2* in the trigeminal, *Phox2B* in the epibranchial ganglia and also members of the *Hmx* family. On the other hand, markers of late sensory identity complement the earlier markers in the mature neuronal transcriptome. The shortlisting process employed was used to select candidates with regulatory potential. Thus, markers of mature general neuronal identity were avoided, focusing on transcription factors with specific described developmental roles instead. Hence, the screening window used in chicken, coupled with this selection process should have enriched the candidate list for markers of CSG determination and specification. Genes also differed in how specifically they marked different CSG. As seen in the case of *ERR*, more restricted expression can correlate with a specific sensory cell fate. In contrast, *Hmx* broader expression was related to a general role in sensory determination. Thus, these complementary subsets were useful for identifying homology at different levels, each tracing the origin of different CSG components.

6.1.1 Genes marking specific sensory fates provide support for cell level homology

A few genes associated with a role in the determination and specification of specific sensory fates have been shown to have conserved expression in lamprey, as discussed in section 1.4 of chapter 1. *Pax3/7* marks the profundal somatosensory neurons (Modrell et al., 2014) and *Phox2* genes mark epibranchial viscerosensory neurons (Häming et al., 2011). These genes continue to mark respective ganglia later in development and were identified as specifically expressed in chick

trigeminal and epibranchial ganglia respectively. Two of the marker genes studied in chapter 3 also show conserved expression in specific sensory fates, *ERR* in a subset of VA neurons and *PRDM12* in somatosensory neurons.

6.1.1(a) Ancestral role of *ERR* in hearing and vision

ERR was the most specific of all markers, with unique expression in the VA, but further study is needed to characterise the exact cell populations marked in lamprey. However, expression in chicken is found in the dorsolateral section of the VA, meaning only a subset of VA neurons are positive for *ERR*. In contrast, the chicken ventromedial VA is marked by *Hmx3* expression (Patthey et al., 2016). Additionally, specific expression is also seen in cell populations of the retina of many jawed vertebrates. Expression in photoreceptors, as well as other sensory cells in amphioxus and tunicates alike, suggests an old role of *ERR* in building such sensory cells. The conserved expression and regulatory landscape of the *ERRγ/ERRa* locus reveal *ERR* expression became strictly controlled in vertebrates. An additional genomic change provides another interesting connection between *ERRγ/ ERRa* and both ear and eye development. *USH2A* is consistently closely linked to *ERRγ* in all jawed vertebrates and the only lamprey *USH2A* paralog is also linked to *ERRa* (fig 4.5, fig 4.6). Mutations of *USH2A* result in Usher syndrome type IIA in humans, a congenital disorder characterised by hearing and vision impairment. *USH2A* encodes a structural protein found in the ankle links connecting stereocilia of hair cells (Adato et al., 2005). Its embryonic expression is described in *Medaka*, where it strongly marks the ear and eye, while knockouts in zebrafish cause hearing and vision problems (Aller et al., 2013; Han et al., 2018). The

combination of conserved genomic linkage of *ERRa/ERRγ* and *USH2A* with expression of both in parts of the ear and eye is intriguing. In *Ciona*, *USH2A* has been described in epidermal sensory neurons, suggesting an ancestral connection to sensory cells (Hurless, 2015). However, linkage of *USH2A* and *ERR* is not seen in any invertebrates. This suggests this happened early in vertebrate evolution and has been retained since, similar to the suite of *CNE* explored in chapter 4. Conserved expression, regulation and genomic linkage to a gene involved in ear/eye development all frame the connection of *ERRγ/ERRa* to vertebrate auricular and ocular development.

6.1.1(b) *PRDM12* as a somatosensory marker

PRDM12 has a more complicated expression pattern in lamprey PNS than *ERRa*. Expression in otic vesicle has been reported before and does not extend to the VA. It is probably unrelated to neuronal function and could be related to the patterning of the vesicle. Geniculate expression as seen in lamprey is not seen in any other vertebrate. This domain could be a lamprey specific acquisition and it is not possible to formulate a robust hypothesis without further careful dissection of the cells marked. However, the geniculate contains visceral neurons that innervate taste buds, but also somatosensory neurons that innervate the pinna of the ear. It has been previously shown in mouse that these populations are distinct and are characterised by different transcriptomic profiles. Specifically, only visceral neurons express *Phox2b,* while somatosensory neurons are marked by somatosensory markers *Brn3a*, *Brn3b* and *Prrxl1* instead (Dvoryanchikov et al., 2017). Interestingly, *Hmx1* is also expressed in these somatosensory populations and loss of *Hmx1*

leads to defects in the development of the geniculate (Quina et al., 2012). Strong expression in the trigeminal and DRG confirms that the role of the gene in somatosensory development is highly conserved across vertebrates. It has been previously shown that *PRDM12* is crucial for proper development of somatosensory neurons and nociception. Recent work has characterised this function in detail in mouse DRG, showing *PRDM12* is necessary for nociceptor fate specifically in both the trigeminal and DRG (Bartesaghi et al., 2019; Desiderio et al., 2019). This shows an ancestral role in the development of this sensory fate developing from both placode and neural crest progenitors. Limited expression data from amphioxus shows *PRDM12* is expressed in ESNs, providing evidence that somatosensory neuron expression may predate vertebrates (Thelie et al., 2015). No domains or structures homologous to the trigeminal or DRG are found outside vertebrates. However, it has been suggested that subsets of ESNs in amphioxus may be homologous to subsets of vertebrate somatosensory neurons (Patthey et al., 2014). Comparable to *ERR*, these data would support a role of *PRDM12* in homologous sensory types, rather than in structures of homologous origin. Additionally, *PRDM12* has been described to supress the expression of neural crest specifier *Foxd3* specifically in the PPR, promoting placodal over neural crest fate (Matsukawa et al., 2015). Hence, the gene appears to have been co-opted in different stages of PNS development in vertebrates as well.

6.1.2 Placodal and CSG markers with broad expression

A few genes with broader and earlier expression across many placodes or CSG have been useful in uncovering deeper homology at a developmental level. This is

covered in more detail in section 1.5 of the general introduction, so I will only briefly recapitulate. These genes have been useful for drawing connections between invertebrate chordate domains or cells and placodal populations with comparable developmental trajectories. The expression of pan placodal markers *Six1/2, Six4/5* and *Eya* linked the tunicate atrial siphon primordia to the posterior placodes. Similarly, *Six1/2* and *Six3/6* (along with *Pitx)* expression supports Haetschek's pit in amphioxus and the oral siphon primordium may be homologous to the adenohypophysis. Expression of these broad placodal markers in amphioxus ESNs has also led to the hypothesis that subsets of ESNs may be related to placode derivatives. Similar to how these early markers highlighted placode related domains, *Hmx* expression and function created a link between CSG and their progenitors to tunicate BTNs. This is also supported by the expression and function of *Ngn* in BTNs*,* which is another broad CSG marker (Gibboney et al., 2019). Vertebrate *Hmx* or *Ngn* paralogs collectively mark all CSG, having an earlier expression onset in sensory neuron progenitors, as these arise in placodes. Earlier *Hmx* expression has not been described, so it is unknown if vertebrate *Hmx* paralogs are expressed in the PPR as well. It is interesting to note that lamprey *ShoxA* expression is also similar to *Hmx*, covering all CSG, apart for the VA. While *Shox* was identified as an epibranchial marker gene in chicken, *Shox2* is also expressed in mouse trigeminal and DRG (section 3.1 of chapter 3), suggesting ancestral expression of the family in both epibranchial and trigeminal. *Shox* is completely unexplored in tunicates, while only CNS expression has been described in amphioxus, similar to *Hmx* (Jackman and Kimmel, 2002). Therefore, *Shox* consists another interesting candidate to investigate further outside vertebrates.

6.2 Novel sensory organisation built by an expanded genomic arsenal

Central to this research is the concept that vertebrates share a number of innovations not found in other animals, including the “new head”, which encompasses many different innovative characters. This is covered more extensively in sections 1.5 and 3.1. Briefly, different novel structures are part of the new vertebrate head, including the tripartite brain and organised sensory systems, wired by the cranial nerves. Placodes and neural crest, also unique vertebrate characters, build the sensory ganglia and related nerves. Cell types of variable embryonic origin are coordinated developmentally and contribute to these new head structures. At this level, looking at the cellular, genetic and patterning components at play, similarities between vertebrates and invertebrates become evident. In contrast, the origin of the complex ensembles of these blocks are more difficult to trace in evolution.

6.2.1 A genetic blueprint underlying CSG development fixed in stem-vertebrates

Conserved expression of CSG markers explored in chapter 3, along with expression of other ganglion and placodal markers (e.g. *Six1, Six4, Eya1, Pax2, Pax8, Pax3, Phox2a and Phox2b* – see sections 1.3 and 1.4 of chapter 1*)*, highlight the recruitment of many transcription factors in building and organising the new cephalic sensory arrangements in early vertebrates. This organisation is fixed early and the inherited expression and role of these TransDev genes is retained in different vertebrate classes. As discussed in section 1.6.1, early duplications in

many of these gene families offered paralogs that were subsequently used differentially in different ganglia and placodal domains. With an expanded repertoire of TransDev genes, early vertebrates could experiment with making new specialised receptors and sensory neurons. The evolution of CSG would have included an increase and expansion of sensory components, coupled with rearrangement and sophistication of patterning and specification mechanisms to build and assemble these components under a new framework. As discussed in section 6.1, genes involved in PPR and placode phases have been useful for understanding this latter process of how cell types are specified and organise into CSG. Developmental marker genes of later sensory identity can instead be useful for inferring where these constituent cell types originated from.

This conserved developmental blueprint is used in all extant vertebrates to build CSG from placodes and neural crest. These populations are patterned and controlled by a highly conserved, fine-tuned network, as reviewed in sections 1.3 and 1.4 of chapter 1. Only a few studies in jawed vertebrates have looked into how early placodal markers regulate later CSG markers. Even though this is largely unexplored in lamprey, enhancer conservation coupled with conserved expression suggest that the structure of this network is largely shared by all vertebrates. The study of *ERR* and *Hmx* CNE from chapters 4 and 5 and published work on CNE regulating *Meis* expression (Parker et al., 2011) reveal these elements can drive cross-species activity in CSG specifically. This shows that upstream regulation of these genes is indeed shared between species to a degree that the elements are recognised. Overall, expression and regulation are sufficiently conserved within the vertebrate lineage to infer that the genetic network underlying CSG development

became fixed in stem-vertebrates and has subsequently remained highly conserved across extant classes.

6.2.2 A step-wise transition from sensory cell types to CSG

As reviewed in the first chapter (sections 1.2.1, 1.4.2 and 1.5.1), patterning mechanisms driving nervous system development are shared between vertebrates and different outgroups. BMP signalling is involved in DV patterning and Wnt signalling in AP patterning across bilaterians. RA signalling has only been described to be implicated in nervous system AP patterning in chordates, controlling the positioning of both CNS and ESN fates in amphioxus. Finally, FGF signalling appears to have been recruited after cephalochordates split. In a gradual transition, different patterning networks were recruited in different ancestors. This gave rise to developmental mechanisms that were re-purposed to build a new body and head by regulating the expanded arsenal of vertebrate TransDev paralogs, as discussed above. Many of these sensory and placodal markers were already involved in making sensory cells before vertebrates. True placodes are not found in invertebrates, but placode like domains are seen in tunicates. In a step-wise recruitment process, more placode like features are found in urochordates than in cephalochordates. The unfocused, dispersed expression of some marker genes in amphioxus (*Six1/2, Six4/5, Six3/6, Eya, Pax2/5/8)* is found in specific fields in tunicates, as discussed in section 1.5 of chapter 1. Outside these fields, additional sensory lineages are found in tunicates marked by similar expression and with comparable development to placodal derivatives.

Nevertheless, cell types marked by similar gene expression as placodal cells are present in amphioxus and in protostomes, even though these do not develop in a comparable fashion to placodes. Ancient sensory markers reveal that many sensory cells across bilaterian animals have shared origin and likely belong to homologous types (Arendt et al., 2016a; Schlosser, 2018). This includes photoreceptors, mechanoreceptors and olfactory cells (Arendt, 2008, 2003). These primary sensory cells have been repeatedly re-purposed in different animal lineages and evolved into lineage specific cells in many cases. As an example, the ciliary and rhabdomeric photoreceptors of the vertebrate eye and pineal evolved from ancestral ciliary and rhabdomeric cells respectively. Corresponding photoreceptors of the amphioxus frontal eye and the tunicate ocellus also evolved from these ancestral types (Kusakabe et al., 2001; Vopalensky et al., 2012). Similarly, the hair cells of the vertebrate ear and their homologs in the tunicate coronal organ evolved from a hair cell precursor after the split of cephalochordates (Manni et al., 2006; Rigon et al., 2018). In turn, that early hair cell type would have evolved from a mechanoreceptor precursor, which possibly also gave rise to a subset of amphioxus ESN marked by comparable gene expression (e.g *ERR*). The cells marked by *ERR* expression in the tunicate sensory vesicle would have originated from such precursors of ancestral chordates. While these may have different development than *ERR* positive cells of the vertebrate ear or eye, they are likely homologous cell types that evolved from similar ancient sensory populations. In addition to primary sensory cells, tunicates also possess the BTN, which are secondary sensory neurons and are marked by pan-ganglionic marker *Hmx*. Overall, the data from the study of both *ERR* and *Hmx* families add to the idea that CSG cell types predate vertebrates.

6.3 The ancestry of BTN and CSG neurons

As discussed more extensively in section 5.3 of chapter 5, BTNs have been previously likened to neural crest derived DRG neurons. This has been based on several characters: the expression of sensory cell markers in BTNs that are also expressed in DRG, the developmental origin and migratory behaviour of their progenitors, as well as their position posterior to the brain (Stolfi et al., 2015). BTNs have been considered a proto-neural crest like population and contrasted to proto-placode like palp cells. These two sensory populations arise from the edge of the neural plate under regulation by a shared early upstream network and BTN can be transformed to palp cells (Horie et al., 2018). The data presented in chapter 5 provide crucial insight on the evolutionary connection between BTN and CSG neurons instead. The comparison of the expression and function of *Hmx* in *Ciona* and vertebrates rules out that BTN consist a neural crest homolog, highlighting the characters shared by BTN and CSG neurons. BTN develop in a comparable fashion to CSG progenitors from the edges of the neural plate, express *Islet* and *POUIV* and their specification is driven by *Hmx* (this study) and *Ngn* (Gibboney et al., 2019), similar to CSG neurons. Moreover, recent follow-up work complements this data and provides evidence of a conserved upstream network regulating *Hmx* expression in BTN and CSG alike. This connection of BTN and vertebrate sensory neurons allows us to consider how placodes and CSG evolved.

6.3.1 The BTN/CSG specification network is shared between vertebrates and tunicates

The function of *Hmx* in the specification of both BTN and CSG suggested the expression and role of *Hmx* in developing secondary sensory neurons is an ancestral character to vertebrates. As shown in chapter 5, there is extended conservation of *Hmx* gene structure and CNE sequence across extant vertebrates and it was inferred that an ancestral udCNE would have regulated *Hmx* expression in CSG in stem-vertebrates. Hence, this extreme genomic conservation places the origin of the regulatory network *Hmx* is part of at the root of vertebrate evolution. Adding this to the conserved function of *Hmx* in BTN and CSG, it was considered possible that this network could predate vertebrates. This hypothesis was also supported by the connection of *Hmx* and *Ngn* during BTN specification, briefly mentioned above. Chapter 5 data shows upregulation of *Ngn* by *Hmx,* but it was also recently shown that overexpression of *Ngn* also leads to upregulation of *Hmx (*supplementary data of (Gibboney et al., 2019)*)*. Interestingly, ectopic expression of *POUIV* in the epidermis, has also been previously documented to lead to upregulation of both *Ngn* and *Hmx*, as assessed by bulk embryo RNAseq (Hurless, 2015). Hence, the network that drives both BTN and CSG specification appears to be at least partially conserved between vertebrates and tunicates.

Alessandro Pennati, another member of our lab, recently tested this experimentally. Comparison of different genomes showed 2kb of sequence 5' to the *Hmx* transcription start site to be conserved between Ciona species (data not shown). This large conserved element was hypothesized to consist a tunicate CNE, with

sequence conservation reflecting evolutionary constraint deriving from a role in gene regulation. Testing this in transgenic Ciona (using a β-galactosidase reporter construct -Fig. 6.1A (Chen et al., 2014)) revealed that the 2kb fragment was able to drive robust and specific reporter expression in BTN and in part of the CNS, recapitulating aspects of endogenous *Hmx* expression (Fig. 6.1B). Neither this regulatory element nor any other sequence in the tunicate *Hmx* locus bears any sequence similarities to the vertebrate uCNE or dCNE. However, this regulatory element is comparable to the two vertebrate CNE, as it consists a large block of conserved sequence proximal to the gene that regulates expression in both CNS and PNS. For this reason, lamprey CNEs were electroporated in *Ciona* (same vector and approach as previously described) to test for cross species recognition (Fig. 6.1C). Strikingly dCNE was able to recapitulate Ciona *Hmx* expression in BTNs and CNS (Fig. 6.1C1-3). BTN expression was specific and robust (Fig. 6.1C1, C3). CNS expression extended along the majority of the neural tube (Fig. 6.1C1), encompassing cells in the anterior Ciona CNS that express *Hmx*, but resembling the more extensive CNS expression of *Hmx* genes and reporters in amphioxus and vertebrates (compare Fig. 6.1C with Fig. 5.1 and Fig. 5.6).

Cross-species specific activity of vertebrate dCNE in *Ciona* confirms a deeply conserved network regulates *Hmx* during PNS and CNS development, as previously hypothesized. Similarities and differences to the activity of the tunicate regulatory element highlight aspects of both conserved and derived characters between the two groups. Importantly, differences in regulatory activity parallel the secondary reduction of the tunicate embryonic plan and nervous system, as the tunicate element only drives expression in the anterior CNS. The ability of lamprey

dCNE to drive expression more broadly suggests that the upstream network is sufficiently conserved, with dCNE responding to transcription factors in the posterior CNS that the tunicate element does not recognize anymore. Interaction with this network likely became much more tightly controlled in stem vertebrates, leading to the conservation of genomic structure and ancestral udCNE sequence. Subsequently, possibly driven by a need for robustness in expression site and timing, dCNE and uCNE both retained the ancestral interaction interface. The two CNE also acquired individual gains or losses of binding sites and other features (e.g. sites for epigenetic marking, chromatin interactions) leading to sequence divergence outside the conserved core and the differences in activity recorded between the two in lamprey. The lack of dCNE activity in lamprey CSG may be explained by such divergence from the ancestral element. Acquisition of repressor sites, or epigenetic suppression of the element could have led to this. These vertebrate specific constraints would be absent in *Ciona* enabling activity of dCNE in the PNS. This analysis is summarized in Fig.6.1D, providing an example of a model explaining how these elements could have evolved in vertebrates and tunicates from an ancestral CNE regulating *Hmx* in CNS and sensory neurons. This is only given to assist in conceptualizing the preceding analysis and more work is necessary to understand how *Hmx* regulation evolved.

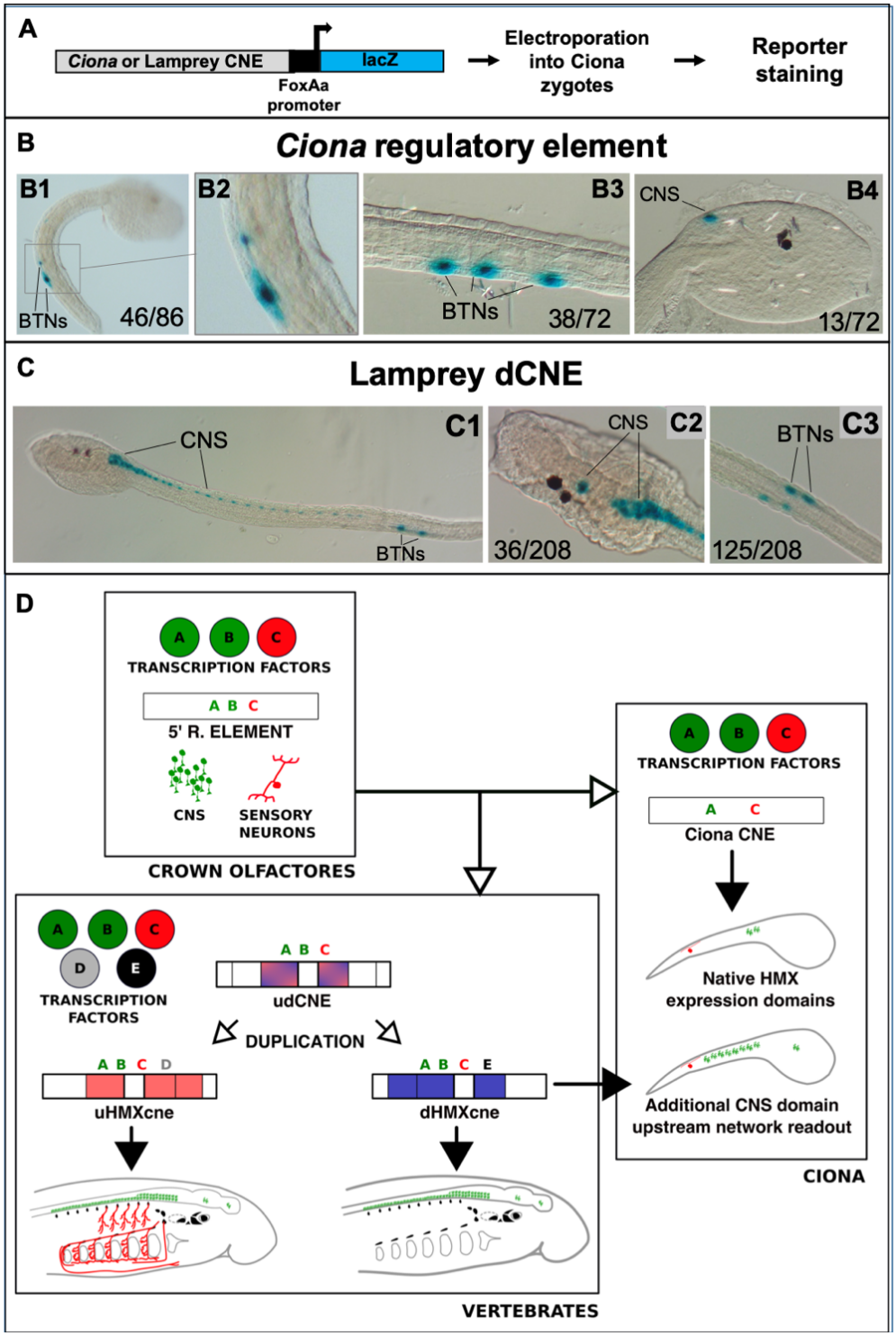


Fig. 6.1. CNE activity in *Ciona*. A. The *Ciona* regulatory element or the lamprey CNE were cloned in the pCES vector using the β-galactosidase reporter (Chen et al., 2014). **B.** Ciona Hmx CNE activity in Ciona embryos, visualised by lacZ staining. B1 shows a tailbud stage embryo with stain in BTNs, seen in close up in B2. No CNS stain was observed at this stage. B3 and B4 show BTN and CNS staining respectively in early larvae. Numbers indicate the number of embryos showing staining in the indicated structures, out of the total number of surviving embryos. (Data kindly provided by Alessandro Pennati and Professor Sebastian Shimeld) **C.** Transgenic *Ciona* larvae stained for lamprey dCNE reporter activity. Numbers indicate the number of times expression in the cells identified was seen, out of total surviving larvae. (Data kindly provided by Alessandro Pennati and Professor

Sebastian Shimeld) **D.** Hypothetical model of *Hmx* regulatory element evolution in tunicates and vertebrates. An ancestral regulatory element drove *Hmx* in different parts of the CNS, regulated by transcription factors A and B, and in sensory neurons by transcription factor C. Ability to respond to factor B was lost in *Ciona* leading to the restricted CNS expression pattern of native *Hmx*. In vertebrates, uCNE and dCNE inherit responsiveness to factors A, B, C. Additional factors could restrict each to particular domains, by enhancing uCNE activity (factor D) and/or repressing dCNE (factor E) in CSG. In contrast to the native regulatory element, dCNE is able to respond more broadly to the upstream network in *Ciona*, leading to the extended CNS expression pattern that is comparable to vertebrate *Hmx.*

6.3.2 BTN and placode related components are uncoupled in tunicates

In a final attempt to bring the data of chapters 3,4 and 5 together and formulate a hypothesis on placode and CSG neuron evolution, I would like to consider the analysis from the preceding sections together.

While CSG cell types are found in other animals, these populations were brought together under conserved organisation only in vertebrates. First, I will briefly mention some key points from the review of section 1.5 of chapter 1. Sensory cells related to placodal derivatives are present in amphioxus, but these are scattered and there are no sensory structures of higher organisation. Then, in addition to individual sensory cells marked by various sensory markers, tunicates have some domains marked by pan-placodal markers of the *Six* and *Eya* families. While the atrial and oral siphon primordia are marked by PPR markers, they do not show large scale cell proliferation and produce primary sensory cells. Additionally, coronal hair cells develop from the oral siphon primordium, which does not correlate to the developmental origin of vertebrate ear hair cells from the VA. Finally, cells of the CNS derived sensory vesicle are marked by *ERR,* while various epidermal sensory neurons of *Ciona* also express markers such as *POUIV* and *Islet*. Therefore, these

different sensory types similar to a range of placodal derivatives are not built in the same way as in vertebrates.

Unlike other placodal like populations in tunicates, BTN are the only cells that are directly comparable to vertebrate sensory ganglia neurons, as they are secondary sensory cells connecting sensory receptors to the brain. Both populations develop from the edge of the neural plate under the same conserved network, driven by *Hmx* and *Ngn* (see section 5.3 and 6.2 above). However, there is no ISH data to suggest BTN progenitors express *Six-Eya,* they do not arise in placode like thickenings, such as the atrial and oral siphon primordia and they do not connect to derivatives of the primordia in larvae. While more thorough analysis should investigate this in better detail, expression domains of neither *Hmx* nor *Ngn* seem to overlap with *Six-Eya* zones, based on available data (this study, (Mazet et al., 2005)). Thus, BTN, as CSG-like secondary neurons, appear uncoupled both developmentally and functionally (in terms of connectivity) to other placode related cells.

Early developmental networks in tunicates are known to be derived, especially when it comes to patterning mechanisms. It is possible that early BTN induction and specification differ from placodal CSG progenitors because the upstream network has diverged and become disconnected from *Six* and *Eya.* However, tunicate *Six* and *Eya* members mark the progenitors of the atrial and oral siphon primordia at the stage when BTN progenitors are already specified. Hence, *Six* and *Eya* act in the specification of other cells at these same stages. Also, amphioxus ESNs express *Six* and *Eya,* but not *Hmx.* Thus, it is possible that BTN/CSG like secondary

sensory neurons, ancestrally specified via the *Hmx/Ngn* network, came under *Six-Eya* control after vertebrates split from tunicates. In CSG, these neurons now develop from the large proliferative zones of the placodes, organised by the various patterning mechanisms to connect the large variety of sensory receptors to the new complex brain. In either case, a large increase in complexity was achieved in vertebrates by bringing pre-existing parts together, including different sensory receptors, secondary sensory neurons that develop from the neural plate border and different patterning and specification pathways.

6.4 Conclusions and future prospects

In the different chapters of this thesis I explored the molecular identity of emerging CSG and used their conserved developmental signature to track the evolution of vertebrate sensory cells. The third Chapter characterized four conserved gene markers of different ganglia. *PRDM12* and *ShoxA* were not investigated further and there is potential interest in looking into the expression and regulation of both of these genes, as explained in sections 6.1.1(b) and 6.1.2. *PRDM12* is described as a specific marker of somatosensory neurons in jawed vertebrate models. Investigating the role, expression and regulation of the gene in lamprey and amphioxus in better detail can help understand what populations are marked and if there is an ancestral role in nociceptor development specifically. The *Shox* family is much less studied in vertebrates and largely unexplored in other chordates. Based on this and its expression in various CSG in different vertebrates, it consists a good candidate for further research.

The work on *ERR* expression and regulation showed *ERRa/ERRγ*, as homologous loci, mark the VA across all vertebrates and uncovered the complex landscape of conserved enhancers that regulates these large loci. These enhancers can be a useful tool for future work on the populations marked by the gene in CNS and the VA. Further research can also look into the role of the gene in eye and ear development in vertebrates and extend this to the sensory populations marked in tunicates and amphioxus. The unique linkage to *USH2A* in vertebrates is also interesting for this purpose. A characterisation of *USH2A* expression during development in lamprey, jawed vertebrates, tunicates and amphioxus is first necessary. It would also be interesting to assess if the *ERRa/ERRγ* and *USH2A* loci interact, potentially via the CNE of *ERRa/ERRγ,* something that could possibly be assessed via chromatin interaction and epigenetic data. Finally, a dissection of the activity of individual CNE that are part of this suite and their transcription factor binding site composition would be helpful for understanding how *ERRa/ERRγ* is regulated.

The study of the expression, regulation and function of *Hmx* in sensory neurons was revelatory about the evolution of CSG neurons and provided multiple tools and directions for future work. First and foremost, it provided a robust connection between BTN and CSG neurons, bringing existing data into perspective. Testing the activity of the regulatory elements in *Ciona* was the first step in dissecting this connection further. Cross-species activity of dCNE is a potent tool for investigating the specification network of BTN and CSG, especially in conjunction with the endogenous regulatory element. Careful binding site analysis, in comparison to RNAseq data, can help understand what are the conserved transcription factors

that regulate these elements in BTN and CSG. Additionally, *Hmx* loss of function experiments could also complement the overexpression data and allow for comparison with similar vertebrate data. The uCNE and dCNE enhancer pair is the first reported case of a conserved duplicated CNE pair. Characterisation of the activity of the two enhancers in other vertebrates can help understand their role in CNS and CSG better. Transcription factor binding site analysis, epigenetic and chromatin interaction data are then crucial for elucidating their mechanism of action and the reason why such extended sequence has remained conserved for so long. Finally, I would also argue that *Hmx* should be tested for its ability to produce vertebrate sensory neurons in vivo or in vitro. I base this on the conserved early function of *Hmx* in sensory neuron specification across vertebrates and its sufficiency to induce BTN specification in tunicates. In addition, *Hmx* appears to be connected to a regulatory network with *Ngn,* which is used as a neural differentiation factor in mammalian cell cultures. Hence, I would suggest that *Hmx* may have the potential to drive sensory neuron specification from non-neural ectodermal progenitors, which could be tested either by ectopic expression in vivo, or in stem cell derived ectodermal cells in cell culture.

References

Abelló, G., Khatri, S., Giráldez, F., Alsina, B., 2007. Early regionalization of the otic placode and its regulation by the Notch signaling pathway. Mech. Dev. 124, 631–645. https://doi.org/10.1016/j.mod.2007.04.002

Abitua, P.B., Wagner, E., Navarrete, I.A., Levine, M., 2012. Identification of a rudimentary neural crest in a non-vertebrate chordate. Nature 492, 104–107. https://doi.org/10.1038/nature11589

Adamska, M., Wolff, A., Kreusler, M., Wittbrodt, J., Braun, T., Bober, E., 2001. Five Nkx5 genes show differential expression patterns in anlagen of sensory organs in medaka: Insight into the evolution of the gene family. Dev. Genes Evol. 211, 338–349. https://doi.org/10.1007/s004270100162

Adato, A., Lefèvre, G., Delprat, B., Michel, V., Michalski, N., Chardenoux, S., Weil, D., El-Amraoui, A., Petit, C., 2005. Usherin, the defective protein in Usher syndrome type IIA, is likely to be a component of interstereocilia ankle links in the inner ear sensory cells. Hum. Mol. Genet. 14, 3921–3932. https://doi.org/10.1093/hmg/ddi416

Aller, E., Sánchez-Sánchez, A. V., Chicote, J.U., García-García, G., Udaondo, P., Cavallé, L., Piquer-Gil, M., García-España, A., Díaz-Llopis, M., Millán, J.M., Mullor, J.L., 2013. Analysis of the Ush2a Gene in Medaka Fish (Oryzias latipes). PLoS One 8. https://doi.org/10.1371/journal.pone.0074995

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415. https://doi.org/10.1093/bioinformatics/btg427

Andermann, P., Ungos, J., Raible, D.W., 2002. Neurogenin1 defines zebrafish cranial sensory ganglia precursors. Dev. Biol. 251, 45–58. https://doi.org/10.1006/dbio.2002.0820

Arendt, D., 2008. The evolution of cell types in animals: Emerging principles from molecular studies. Nat. Rev. Genet. 9, 868–882. https://doi.org/10.1038/nrg2416

Arendt, D., 2003. Evolution of eyes and photoreceptor cell types. Int. J. Dev. Biol. 47, 563–571.

Arendt, D., Denes, A.S., Jékely, G., Tessmar-Raible, K., 2008. The evolution of nervous system centralization. Philos. Trans. R. Soc. B Biol. Sci. 363, 1523–1528. https://doi.org/10.1098/rstb.2007.2242

Arendt, D., Musser, J.M., Baker, C.V.H.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., Wagner, G.P., 2016a. The origin and evolution of cell types. Nat. Rev. Genet. 17, 744–757. https://doi.org/10.1038/nrg.2016.127

Arendt, D., Tosches, M.A., Marlow, H., 2016b. From nerve net to nerve ring, nerve cord and brain-evolution of the nervous system. Nat. Rev. Neurosci. 17, 61–72. https://doi.org/10.1038/nrn.2015.15

Baatrup, E., 1981. Primary Sensory Cells in the Skin of Amphioxus (Branchiostoma lanceolatum (P)). Acta Zool. 62, 147–157. https://doi.org/10.1111/j.1463-6395.1981.tb00624.x

Baker, C.V.H., Bronner-Fraser, M., 2000. Establishing neuronal identity in vertebrate neurogenic placodes. Development 127, 3045–3056.

Baker, M.E., 2008. Trichoplax, the simplest known animal, contains an estrogen-related receptor but no estrogen receptor: Implications for estrogen receptor evolution. Biochem. Biophys. Res. Commun. 375, 623–627.

https://doi.org/10.1016/j.bbrc.2008.08.047

Bally-Cuif, L., Dubois, L., Vincent, A., 1998. Molecular cloning of Zcoe2, the zebrafish homolog of Xenopus Xcoe2 and mouse EBF-2, and its expression during primary neurogenesis. Mech. Dev. 77, 85–90. https://doi.org/10.1016/S0925-4773(98)00144-0

Bannister, R., Beresford, N., Granger, D.W., Pounds, N.A., Rand-Weaver, M., White, R., Jobling, S., Routledge, E.J., 2013. No substantial changes in estrogen receptor and estrogen-related receptor orthologue gene transcription in Marisa cornuarietis exposed to estrogenic chemicals. Aquat. Toxicol. 140–141, 19–26. https://doi.org/10.1016/j.aquatox.2013.05.002

Bardet, P.L., Obrecht-Pflumio, S., Thisse, C., Laudet, V., Thisse, B., Vanacker, J.M., 2004. Cloning and developmental expression of five estrogen-receptor related genes in the zebrafish. Dev. Genes Evol. 214, 240–249. https://doi.org/10.1007/s00427-004-0404-6

Bardet, P.L., Schubert, M., Horard, B., Holland, L.Z., Laudet, V., Holland, N.D., Vanacker, J.M., 2005. Expression of estrogen-receptor related receptors in amphioxus and zebrafish: Implications for the evolution of posterior brain segmentation at the invertebrate-to-vertebrate transition. Evol. Dev. 7, 223–233. https://doi.org/10.1111/j.1525-142X.2005.05025.x

Bartesaghi, L., Wang, Y., Fontanet, P., Wanderoy, S., Berger, F., Wu, H., Akkuratova, N., Bouçanova, F., Médard, J.J., Petitpré, C., Landy, M.A., Zhang, M.D., Harrer, P., Stendel, C., Stucka, R., Dusl, M., Kastriti, M.E., Croci, L., Lai, H.C., Consalez, G.G., Pattyn, A., Ernfors, P., Senderek, J., Adameyko, I., Lallemend, F., Hadjab, S., Chrast, R., 2019. PRDM12 Is Required for Initiation of the Nociceptive Neuron Lineage during Neurogenesis. Cell Rep. 26, 3484-3492.e4.

https://doi.org/10.1016/j.celrep.2019.02.098

Begbie, J., Ballivet, M., Graham, A., 2002. Early steps in the production of sensory neurons by the neurogenic placodes. Mol. Cell. Neurosci. 21, 502–511. https://doi.org/10.1006/mcne.2002.1197

Begbie, J., Brunet, J.F., Rubenstein, J.L.R., Graham, A., 1999. Induction of the epibranchial placodes. Development 126, 895–902.

Bertrand, S., Brunet, F.G., Escriva, H., Parmentier, G., Laudet, V., Robinson-Rechavi, M., 2004. Evolutionary genomics of nuclear receptors: From twenty-five ancestral genes to derived endocrine systems. Mol. Biol. Evol. 21, 1923–1937. https://doi.org/10.1093/molbev/msh200

Bertrand, S., Thisse, B., Tavares, R., Sachs, L., Chaumot, A., Bardet, P.L., Escrivà, H., Duffraisse, M., Marchand, O., Safi, R., Thisse, C., Laudet, V., 2007. Unexpected novel relational links uncovered by extensive developmental profiling of nuclear receptor expression. PLoS Genet. 3, 2085–2100. https://doi.org/10.1371/journal.pgen.0030188

Blentic, A., Chambers, D., Skinner, A., Begbie, J., Graham, A., 2011. The formation of the cranial ganglia by placodally-derived sensory neuronal precursors. Mol. Cell. Neurosci. 46, 452–459. https://doi.org/10.1016/j.mcn.2010.11.010

Boehm, T., McCurley, N., Sutoh, Y., Schorpp, M., Kasahara, M., Cooper, M.D., 2012. VLR-Based Adaptive Immunity. Annu. Rev. Immunol. 30, 203–220. https://doi.org/10.1146/annurev-immunol-020711-075038

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bone, Q., Best, A.C.G., 1978. Ciliated Sensory Cells in Amphioxus (Branchiostoma).

J. Mar. Biol. Assoc. United Kingdom 58, 479–486. https://doi.org/10.1017/S0025315400028137

Boorman, C.J., Shimeld, S.M., 2002a. Pitx homeobox genes in Ciona and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. Evol Dev 4, 354–365. https://doi.org/10.1046/j.1525-142X.2002.02021.x

Boorman, C.J., Shimeld, S.M., 2002b. Cloning and expression of a Pitx homeobox gene from the lamprey, a jawless vertebrate. Dev. Genes Evol. 212, 349–353. https://doi.org/10.1007/s00427-002-0249-9

Bosma, P.T., Blázquez, M., Collins, M. a, Bishop, J.D., Drouin, G., Priede, I.G., Docherty, K., Trudeau, V.L., 1999. Multiplicity of glutamic acid decarboxylases (GAD) in vertebrates: molecular phylogeny and evidence for a new GAD paralog. Mol. Biol. Evol. 16, 397–404.

Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., Louis, A., Naville, M., Nitta, K.R., Piette, J., Reeves, W., Scornavacca, C., Simion, P., Vincentelli, R., Bellec, M., Aicha, S.B., Fagotto, M., Gueroult-Bellone, M., Haeussler, M., Jacox, E., Lowe, E.K., Mendez, M., Roberge, A., Stolfi, A., Yokomori, R., Brown, C.T., Cambillau, C., Christiaen, L., Delsuc, F., Douzery, E., Dumollard, R., Kusakabe, T., Nakai, K., Nishida, H., Satou, Y., Swalla, B., Veeman, M., Volff, J.N., Lemaire, P., 2018. ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. Nucleic Acids Res 46, D718–D725. https://doi.org/10.1093/nar/gkx1108

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics

10, 1–9. https://doi.org/10.1186/1471-2105-10-421

Candiani, S., Holland, N.D., Oliveri, D., Parodi, M., Pestarino, M., 2008. Expression of the amphioxus Pit-1 gene (AmphiPOU1F1/Pit-1) exclusively in the developing preoral organ, a putative homolog of the vertebrate adenohypophysis. Brain Res. Bull. 75, 324–330. https://doi.org/10.1016/j.brainresbull.2007.10.023

Candiani, S., Oliveri, D., Parodi, M., Bertini, E., Pestarino, M., 2006. 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 216, 623–633. https://doi.org/10.1007/s00427-006-0083-6

Cañestro, C., Albalat, R., Irimia, M., Garcia-Fernàndez, J., 2013. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. Semin. Cell Dev. Biol. 24, 83–94. https://doi.org/10.1016/j.semcdb.2012.12.008

Chen, J., Streit, A., 2015. A medium-scale assay for enhancer validation in amniotes. Dev. Dyn. 244, 1291–1299. https://doi.org/10.1002/dvdy.24306

Chen, W.C., Pauls, S., Bacha, J., Elgar, G., Loose, M., Shimeld, S.M., 2014. Dissection of a Ciona regulatory element reveals complexity of cross-species enhancer activity. Dev. Biol. 390, 261–272. https://doi.org/10.1016/j.ydbio.2014.03.013

Chen, Y.-C., Auer-Grumbach, M., Matsukawa, S., Zitzelsberger, M., Themistocleous, A.C., Strom, T.M., Samara, C., Moore, A.W., Cho, L.T.-Y., Young, G.T., Weiss, C., Schabhüttl, M., Stucka, R., Schmid, A.B., Parman, Y., Graul-Neumann, L., Heinritz, W., Passarge, E., Watson, R.M., Hertz, J.M., Moog, U., Baumgartner, M., Valente, E.M., Pereira, D., Restrepo, C.M., Katona, I., Dusl, M., Stendel, C., Wieland, T., Stafford, F., Reimann, F., von Au, K., Finke, C., Willems, P.J.,

Nahorski, M.S., Shaikh, S.S., Carvalho, O.P., Nicholas, A.K., Karbani, G., McAleer, M.A., Cilio, M.R., McHugh, J.C., Murphy, S.M., Irvine, A.D., Jensen, U.B., Windhager, R., Weis, J., Bergmann, C., Rautenstrauss, B., Baets, J., De Jonghe, P., Reilly, M.M., Kropatsch, R., Kurth, I., Chrast, R., Michiue, T., Bennett, D.L.H., Woods, C.G., Senderek, J., 2015. Transcriptional regulator PRDM12 is essential for human pain perception. Nat. Genet. 47, 803–808. https://doi.org/10.1038/ng.3308

Chitnis, A., Kintner, C., 1996. Sensitivity of proneural genes to lateral inhibition affects the pattern of primary neurons in Xenopus embryos. Development 122, 2295–2301.

Chong-Morrison, V., Simoes, F.C., Senanayake, U., Carroll, D.S., Riley, P.R., Sauka-Spengler, T., 2018. Re-purposing Ac/Ds transgenic system for CRISPR/dCas9 modulation of enhancers and non-coding RNAs in zebrafish. bioRxiv 450684. https://doi.org/10.1101/450684

Ciani, L., Salinas, P.C., 2005. WNTs in the vertebrate nervous system: From patterning to neuronal connectivity. Nat. Rev. Neurosci. 6, 351–362. https://doi.org/10.1038/nrn1665

Cooper, M.D., Alder, M.N., 2006. The evolution of adaptive immune systems. Cell 124, 815–822. https://doi.org/10.1016/j.cell.2006.02.001

Corbo, J.C., Levine, M., Zeller, R.W., 1997. Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. Development 124, 589–602.

Coward, P., Lee, D., Hull, M. V, Lehmann, J.M., 2001. 4-Hydroxytamoxifen binds to and deactivates the estrogen-related receptor gamma. Proc. Natl. Acad. Sci. U. S. A. 98, 8880–4. https://doi.org/10.1073/pnas.151244398

D'Amico-Martel, a, Noden, D.M., 1983. Contributions of placodal and neural crest cells to avian cranial peripheral ganglia. Am. J. Anat. 166, 445–468. https://doi.org/10.1002/aja.1001660406

D'Autréaux, F., Coppola, E., Hirsch, M.R., Birchmeier, C., Brunet, J.F., 2011. Homeoprotein Phox2b commands a somatic-to-visceral switch in cranial sensory pathways. Proc. Natl. Acad. Sci. U. S. A. 108, 20018–20023. https://doi.org/10.1073/pnas.1110416108

Davidson, B., 2007. Ciona intestinalis as a model for cardiac development. Semin. Cell Dev. Biol. 18, 16–26. https://doi.org/10.1016/j.semcdb.2006.12.007

Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3. https://doi.org/10.1371/journal.pbio.0030314

Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D.M., Harafuji, N., Hastings, K.E.M., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I.A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H.G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D.N., Haga, S., Hayashi, H., Hino, K., Imai, K.S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B.I., Makabe, K.W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M.M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P.D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N., Rokhsar, D.S., 2002. The draft

genome of Ciona intestinalis: Insights into chordate and vertebrate origins. Science (80-.). 298, 2157–2167. https://doi.org/10.1126/science.1080049

Delarbre, C., Gallut, C., Barriel, V., Janvier, P., Gachelin, G., 2002. Complete mitochondrial DNA of the hagfish, Eptatretus burgeri: The comparative analysis of mitochondrial DNA sequences strongly supports the cyclostome monophyly. Mol. Phylogenet. Evol. 22, 184–192. https://doi.org/10.1006/mpev.2001.1045

Delsuc, F., Brinkmann, H., Chourrout, D., Philippe, H., 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965–968. https://doi.org/10.1038/nature04336

Deng, M., Yang, H., Xie, X., Liang, G., Gan, L., 2014. Comparative expression analysis of POU4F1, POU4F2 and ISL1 in developing mouse cochleovestibular ganglion neurons. Gene Expr. Patterns 15, 31–37. https://doi.org/10.1016/j.gep.2014.03.001

Desiderio, S., Vermeiren, S., Van Campenhout, C., Kricha, S., Malki, E., Richts, S., Fletcher, E. V., Vanwelden, T., Schmidt, B.Z., Henningfeld, K.A., Pieler, T., Woods, C.G., Nagy, V., Verfaillie, C., Bellefroid, E.J., 2019. Prdm12 Directs Nociceptive Sensory Neuron Development by Regulating the Expression of the NGF Receptor TrkA. Cell Rep. 26, 3522-3536.e5. https://doi.org/10.1016/j.celrep.2019.02.097

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dubois, L., Bally-Cuif, L., Crozatier, M., Moreau, J., Paquereau, L., Vincent, A., 1998. XCoe2, a transcription factor of the Col/Olf-1/EBF family involved in the specification of primary neurons in Xenopus. Curr. Biol. 8, 199–209.

https://doi.org/10.1016/S0960-9822(98)70084-3

Dubois, L., Vincent, A., 2001. The COE - Collier/Olf1/EBF - Transcription factors: Structural conservation and diversity of developmental functions. Mech. Dev. 108, 3–12. https://doi.org/10.1016/S0925-4773(01)00486-5

Dude, C.M., Kuan, C.Y.K., Bradshaw, J.R., Greene, N.D.E., Relaix, F., Stark, M.R., Baker, C.V.H., 2009. Activation of Pax3 target genes is necessary but not sufficient for neurogenesis in the ophthalmic trigeminal placode. Dev. Biol. 326, 314–326. https://doi.org/10.1016/j.ydbio.2008.11.032

Dvoryanchikov, G., Hernandez, D., Roebber, J.K., Hill, D.L., Roper, S.D., Chaudhari, N., 2017. Transcriptomes and neurotransmitter profiles of classes of gustatory and somatosensory neurons in the geniculate ganglion. Nat. Commun. 8, 760. https://doi.org/10.1038/s41467-017-01095-1

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

Eickholt, B.J., Graham, A., Lumsden, A., Wizenmann, A., 2001. Rhombomere interactions control the segmental differentiation of hindbrain neurons. Mol. Cell. Neurosci. 18, 141–148. https://doi.org/10.1006/mcne.2001.1014

Esposito, R., Racioppi, C., Pezzotti, M.R., Branno, M., Locascio, A., Ristoratore, F., Spagnuolo, A., 2015. The ascidian pigmented sensory organs: Structures and developmental programs. Genesis 53, 15–33. https://doi.org/10.1002/dvg.22836

Feng, Y., Xu, Q., 2010. Pivotal role of hmx2 and hmx3 in zebrafish inner ear and lateral line development. Dev. Biol. 339, 507–518. https://doi.org/10.1016/j.ydbio.2009.12.028

Filowitz, G.L., Rajakumar, R., O'Shaughnessy, K.L., Cohn, M.J., 2018. Cartilaginous

Fishes Provide Insights into the Origin, Diversification, and Sexually Dimorphic Expression of Vertebrate Estrogen Receptor Genes. Mol. Biol. Evol. 35, 2695–2701. https://doi.org/10.1093/molbev/msy165

Fode, C., Gradwohl, G., Morin, X., Dierich, A., LeMeur, M., Goridis, C., Guillemot, F., 1998. The bHLH protein NEUROGENIN 2 is a determination factor for epibranchial placode-derived sensory neurons. Neuron 20, 483–494. https://doi.org/10.1016/S0896-6273(00)80989-7

Fortunato, S.A.V., Adamski, M., Ramos, O.M., Leininger, S., Liu, J., Ferrier, D.E.K., Adamska, M., 2014. Calcisponges have a ParaHox gene and dynamic expression of dispersed NK homeobox genes. Nature 514, 620–623. https://doi.org/10.1038/nature13881

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. VISTA : computational tools for comparative genomics. Nucleic Acids Res. 32, 273–279. https://doi.org/10.1093/nar/gkh458

Freter, S., Fleenor, S.J., Freter, R., Liu, K.J., Begbie, J., 2013. Cranial neural crest cells form corridors prefiguring sensory neuroblast migration. Dev. 140, 3595–3600. https://doi.org/10.1242/dev.091033

Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., Jimenez-Delgado, S., Oliveri, D., Candiani, S., Hirsinger, E., D'Aniello, S., Pascual-Anaya, J., Maeso, I., Pestarino, M., Vernier, P., Nicolas, J.-F., Schubert, M., Laudet, V., Geneviere, A.M., Albalat, R., Garcia Fernandez, J., Holland, N.D., Escriva, H., 2007. Insights into spawning behavior and development of the european amphioxus (Branchiostoma lanceolatum). J. Exp. Zool. Part B Mol. Dev. Evol. 308B, 484–493. https://doi.org/10.1002/jez.b.21179

Fuentes, M., Schubert, M., Dalfo, D., Candiani, S., Benito, E., Gardenyes, J., Godoy,

L., Moret, F., Illas, M., Patten, I., Permanyer, J., Oliveri, D., Boeuf, G., Falcon, J., Pestarino, M., Fernandez, J.G., Albalat, R., Laudet, V., Vernier, P., Escriva, H., 2004. Preliminary observations on the spawning conditions of the European amphioxus (Branchiostoma lanceolatum) in captivity. J. Exp. Zool. 302B, 384–391. https://doi.org/10.1002/jez.b.20025

Fumasoni, I., Meani, N., Rambaldi, D., Scafetta, G., Alcalay, M., Ciccarelli, F.D., 2007. Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates. BMC Evol. Biol. 7, 187. https://doi.org/10.1186/1471-2148-7-187

Furlong, R.F., Holland, P.W.H., 2002a. Bayesian Phylogenetic Analysis Supports Monophyly of Ambulacraria and of Cyclostomes. Zoolog. Sci. 19, 593–599. https://doi.org/10.2108/zsj.19.593

Furlong, R.F., Holland, P.W.H., 2002b. Were vertebrates octoploid? Philos. Trans. R. Soc. B Biol. Sci. 357, 531–544. https://doi.org/10.1098/rstb.2001.1035

Gans, C., Northcutt, R.G., 1983. Neural Crest and the Origin of Vertebrates: A New Head. Science (80-.). 220, 268–273. https://doi.org/10.1126/science.220.4594.268

Garcia-Fernàndez, J., 2005. The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. 6, 881–92. https://doi.org/10.1038/nrg1723

Gibboney, S., Kim, K., Razy-krajka, F., Wang, W., Stolfi, A., 2019. Neurogenin regulates effectors of migratory neuron cell behaviors in Ciona.

Graham, A., Shimeld, S.M., 2013. The origin and evolution of the ectodermal placodes. J. Anat. 222, 32–40. https://doi.org/10.1111/j.1469-7580.2012.01506.x

Green, S.A., Bronner, M.E., 2014. The lamprey: A jawless vertebrate model system for examining origin of the neural crest and other vertebrate traits. Differentiation

87, 44–51. https://doi.org/10.1016/j.diff.2014.02.001

Grocott, T., Tambalo, M., Streit, A., 2012. The peripheral sensory nervous system in the vertebrate head: A gene regulatory perspective. Dev. Biol. 370, 3–23. https://doi.org/10.1016/j.ydbio.2012.06.028

Hadfield, M., Meleshkevitch, E., Boudko, D., 2000. The apical sensory organ of a gastropod veliger is a receptor for settlement cues. Biol. Bull. 198, 67–76. https://doi.org/10.2307/1542804

Hamburger, V., Hamilton, H.L., 1951. A series of normal stages in the development of the chick embryo. J. Morphol. 88, 49–92.

Häming, D., Simoes-Costa, M., Uy, B., Valencia, J., Sauka-Spengler, T., Bronner-Fraser, M., 2011. Expression of sympathetic nervous system genes in Lamprey suggests their recruitment for specification of a new vertebrate feature. PLoS One 6, e26543. https://doi.org/10.1371/journal.pone.0026543

Hammond, K.L., Whitfield, T.T., 2006. The developing lamprey ear closely resembles the zebrafish otic vesicle: Otx1 expression can account for all major patterning differences. Development 133, 1347–1357. https://doi.org/10.1242/dev.02306

Han, S., Liu, X., Xie, S., Gao, M., Liu, F., Yu, S., Sun, P., Wang, C., Archacki, S., Lu, Z., Hu, X., Qin, Y., Qu, Z., Huang, Y., Lv, Y., Tu, J., Li, J., Yimer, T.A., Jiang, T., Tang, Z., Luo, D., Chen, F., Liu, M., 2018. Knockout of ush2a gene in zebrafish causes hearing impairment and late onset rod-cone dystrophy. Hum. Genet. 137, 779–794. https://doi.org/10.1007/s00439-018-1936-6

Harafuji, N., Keys, D.N., Levine, M., 2002. Genome-wide identification of tissue-specific enhancers in the Ciona tadpole. Proc. Natl. Acad. Sci. 99, 6802–6805. https://doi.org/10.1073/pnas.052024999

Heard, D.J., Norby, P.L., Holloway, J., Vissing, H., 2000. Human ERRgamma, a third

member of the estrogen receptor-related receptor (ERR) subfamily of orphan nuclear receptors: tissue-specific isoforms are expressed during development and in the adult. Mol. Endocrinol. 14, 382–92. https://doi.org/10.1210/mend.14.3.0431

Heimberg, A.M., Cowper-Sallari, R., Sémon, M., Donoghue, P.C.J., Peterson, K.J., 2010. MicroRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. Proc. Natl. Acad. Sci. U. S. A. 107, 19379–19383. https://doi.org/10.1073/pnas.1010350107

Hermans-Borgmeyer, I., Süsens, U., Borgmeyer, U., 2000. Developmental expression of the estrogen receptor-related receptor γ in the nervous system during mouse embryogenesis. Mech. Dev. 97, 197–199. https://doi.org/10.1016/S0925-4773(00)00422-6

Hohenauer, T., Moore, A.W., 2012. The Prdm family: expanding roles in stem cells and development. Development 139, 2267–2282. https://doi.org/10.1242/dev.070110

Hokamp, K., McLysaght, A., Wolfe, K.H., 2003. The 2R hypothesis and the human genome sequence. J. Struct. Funct. Genomics 3, 95–110.

Holland, L.Z., 2013. Evolution of new characters after whole genome duplications: Insights from amphioxus. Semin. Cell Dev. Biol. 24, 101–109. https://doi.org/10.1016/j.semcdb.2012.12.007

Holland, L.Z., 2009. Chordate roots of the vertebrate nervous system: expanding the molecular toolkit. Nat. Rev. Neurosci. 10, 736–746. https://doi.org/10.1038/nrn2703

Holland, L.Z., Carvalho, J.E., Escriva, H., Laudet, V., Schubert, M., Shimeld, S.M., Yu, J.-K., 2013. Evolution of bilaterian central nervous systems: a single origin?

Evodevo 4, 27. https://doi.org/10.1186/2041-9139-4-27

Holland, L.Z., Gibson-Brown, J.J., 2003. The Ciona intestinalis genome: When the constraints are off. BioEssays 25, 529–532. https://doi.org/10.1002/bies.10302

Holland, L.Z., Laudet, V., Schubert, M., 2004. The chordate amphioxus: An emerging model organism for developmental biology. Cell. Mol. Life Sci. 61, 2290–2308. https://doi.org/10.1007/s00018-004-4075-2

Holland, P.W.H., 2013. Evolution of homeobox genes. Wiley Interdiscip. Rev. Dev. Biol. 2, 31–45. https://doi.org/10.1002/wdev.78

Holland, P.W.H., Garcia-Fernàndez, J., Williams, N.A., Sidow, A., 1994. Gene duplications and the origins of vertebrate development. Dev. Suppl. 1994, 125–33.

Horie, R., Hazbun, A., Chen, K., Cao, C., Levine, M., Horie, T., 2018. Shared evolutionary origin of vertebrate neural crest and cranial placodes. Nature 560, 228–232. https://doi.org/10.1038/s41586-018-0385-7

Hotta, K., Takahashi, H., Asakura, T., Saitoh, B., Takatori, N., Satou, Y., Satoh, N., 2000. Characterization of Brachyury-downstream notochord genes in the Ciona intestinalis embryo. Dev. Biol. 224, 69–80. https://doi.org/10.1006/dbio.2000.9765

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Hughes, A.L., Friedman, R., 2005. Loss of ancestral genes in the genomic evolution of Ciona intestinalis. Evol. Dev. 7, 196–200. https://doi.org/10.1111/j.1525-142X.2005.05022.x

Hurless, V., 2015. Investigating Ciliated Sensory Neurons and Molecular Tools in the

Invertebrate Chordate, Ciona intestinalis. UC San Diego. https://doi.org/ProQuest ID: Hurless_ucsd_0033D_15060. Merritt ID: ark:/20775/bb8534284m.

Imai, K.S., Levine, M., Satoh, N., Satou, Y., 2006. Regulatory blueprint for a chordate embryo. Science (80-.). 312, 1183–1187. https://doi.org/10.1126/science.1123404

Imai, K.S., Stolfi, A., Levine, M., Satou, Y., 2009. Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. Development 136, 285–293. https://doi.org/10.1242/dev.026419

Irvine, S.Q., Cangiano, M.C., Millette, B.J., Gutter, E.S., 2007. Non-overlapping expression patterns of the clusteredDll-A/B genes in the ascidianCiona Intestinalis. J. Exp. Zool. Part B Mol. Dev. Evol. 308B, 428–441. https://doi.org/10.1002/jez.b.21169

Jackman, W.R., Kimmel, C.B., 2002. Coincident iterated gene expression in the amphioxus neural tube. Evol. Dev. 4, 366–374. https://doi.org/10.1046/j.1525-142X.2002.02022.x

Jackman, W.R., Langeland, J.A., Kimmel, C.B., 2000. Islet Reveals Segmentation in the Amphioxus Hindbrain Homolog. Dev. Biol. 220, 16–26. https://doi.org/10.1006/dbio.2000.9630

Jacquin, M.F., Arends, J.J.A., Xiang, C., Shapiro, L.A., Ribak, C.E., Chen, Z.-F., 2008. In DRG11 Knock-Out Mice, Trigeminal Cell Death Is Extensive and Does Not Account for Failed Brainstem Patterning. J. Neurosci. 28, 3577–3585. https://doi.org/10.1523/JNEUROSCI.4203-07.2008

Jeffery, W.R., Chiba, T., Krajka, F.R., Deyts, C., Satoh, N., Joly, J.S., 2008. Trunk lateral cells are neural crest-like cells in the ascidian Ciona intestinalis: Insights into the ancestry and evolution of the neural crest. Dev. Biol. 324, 152–160.

https://doi.org/10.1016/j.ydbio.2008.08.022

Jimenez-Guri, E., Pujades, C., 2011. An ancient mechanism of hindbrain patterning has been conserved in vertebrate evolution. Evol. Dev. 13, 38–46. https://doi.org/10.1111/j.1525-142X.2010.00454.x

Joshi, N.A., Fass, J.N., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].

Joyce Tang, W., Chen, J.S., Zeller, R.W., 2013. Transcriptional regulation of the peripheral nervous system in Ciona intestinalis. Dev. Biol. 378, 183–193. https://doi.org/10.1016/j.ydbio.2013.03.016

Kaltenbach, S.L., Yu, J.-K., Holland, N.D., 2009. The origin and migration of the earliest-developing sensory neurons in the peripheral nervous system of amphioxus. Evol. Dev. 11, 142–151. https://doi.org/10.1111/j.1525-142X.2009.00315.x

Kelly, L.E., El-Hodiri, H.M., 2016. Xenopus laevis Nkx5.3 and sensory organ homeobox (SOHo) are expressed in developing sensory organs and ganglia of the head and anterior trunk. Dev. Genes Evol. 226, 423–428. https://doi.org/10.1007/s00427-016-0555-2

Kenny, N.J., Namigai, E.K.O., Marlétaz, F., Hui, J.H.L., Shimeld, S.M., 2015. Draft genome assemblies and predicted microRNA complements of the intertidal lophotrochozoans Patella vulgata (Mollusca, Patellogastropoda) and Spirobranchus (Pomatoceros) lamarcki (Annelida, Serpulida). Mar. Genomics 24, 139–146. https://doi.org/10.1016/j.margen.2015.07.004

Kenny, N.J., Shimeld, S.M., 2012. Additive multiple k-mer transcriptome of the keelworm Pomatoceros lamarckii (Annelida; Serpulidae) reveals annelid trochophore transcription factor cassette. Dev. Genes Evol. 222, 325–339.

https://doi.org/10.1007/s00427-012-0416-6

Keynes, R., Lumsden, A., 1990. Segmentation and the origin of regional diversity in the vertebrate central nervous system. Neuron 4, 1–9. https://doi.org/10.1016/0896-6273(90)90438-L

Kiecker, C., Lumsden, A., 2005. Compartments and their boundaries in vertebrate brain development. Nat. Rev. Neurosci. 6, 553–564. https://doi.org/10.1038/nrn1702

Kiernan, A.E., Nunes, F., Wu, D.K., Fekete, D.M., 1997. The expression domain of two related homeobox genes defines a compartment in the chicken inner ear that may be involved in semicircular canal formation. Dev. Biol. 191, 215–229. https://doi.org/10.1006/dbio.1997.8716

Kim, K., Kim, R., Sengupta, P., 2010. The HMX/NKX homeodomain protein MLS-2 specifies the identity of the AWC sensory neuron type via regulation of the ceh-36 Otx gene in C. elegans. Development 137, 963–974. https://doi.org/10.1242/dev.044719

Kourakis, M.J., Smith, W.C., 2007. A conserved role for FGF signaling in chordate otic/atrial placode formation. Dev. Biol. 312, 245–257. https://doi.org/10.1016/j.ydbio.2007.09.020

Kozmik, Z., Holland, N.D., Kreslova, J., Oliveri, D., Schubert, M., Jonasova, K., Holland, L.Z., Pestarino, M., Benes, V., Candiani, S., 2007. Pax-Six-Eya-Dach network during amphioxus development: Conservation in vitro but context specificity in vivo. Dev. Biol. 306, 143–159. https://doi.org/10.1016/j.ydbio.2007.03.009

Kozmikova, I., Smolikova, J., Vlcek, C., Kozmik, Z., 2011. Conservation and diversification of an ancestral chordate gene regulatory network for dorsoventral

patterning. PLoS One 6. https://doi.org/10.1371/journal.pone.0014650

Krispin, S., Nitzan, E., Kassem, Y., Kalcheim, C., 2010. Evidence for a dynamic spatiotemporal fate map and early fate restrictions of premigratory avian neural crest. Development 137, 585–595. https://doi.org/10.1242/dev.041509

Krüger, M., Schmid, T., Krüger, S., Bober, E., Braun, T., 2006. Functional redundancy of NSCL-1 and NeuroD during development of the petrosal and vestibulocochlear ganglia. Eur. J. Neurosci. 24, 1581–1590. https://doi.org/10.1111/j.1460-9568.2006.05051.x

Kuraku, S., 2013. Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. Semin. Cell Dev. Biol. 24, 119–127. https://doi.org/10.1016/j.semcdb.2012.12.009

Kuraku, S., Hoshiyama, D., Katoh, K., Suga, H., Miyata, T., 1999. Monophyly of Lampreys and Hagfishes Supported by Nuclear DNA–Coded Genes. J. Mol. Evol. 49, 729–735. https://doi.org/10.1007/PL00006595

Kuratani, S., Kusakabe, R., Wada, H., Ohtani, K., Kobayashi, M., Yao, T., 2008. Expression of Sox and fibrillar collagen genes in lamprey larval chondrogenesis with implications for the evolution of vertebrate cartilage. J. Exp. Zool. Part B Mol. Dev. Evol. 310B, 596–607. https://doi.org/10.1002/jez.b.21231

Kuratani, S., Ueki, T., Aizawa, S., Hirano, S., 1997. Peripheral development of cranial nerves in a cyclostome, lampetra japonica: Morphological distribution of nerve branches and the vertebrate body plan. J. Comp. Neurol. 384, 483–500. https://doi.org/10.1002/(SICI)1096-9861(19970811)384:4<483::AID-CNE1>3.0.CO;2-Z

Kusakabe, T., Kusakabe, R., Kawakami, I., Satou, Y., Satoh, N., Tsuda, M., 2001. Ci-opsin1, a vertebrate-type opsin gene, expressed in the larval ocellus of the

ascidian Ciona intestinalis. FEBS Lett. 506, 69–72. https://doi.org/10.1016/S0014-5793(01)02877-0

Lacalli, T.C., Hou, S., 1999. A reexamination of the epithelial sensory cells of amphioxus (Branchiostoma). Acta Zool. 80, 125–134. https://doi.org/10.1046/j.1463-6395.1999.80220005.x

Lara-Ramírez, R., Patthey, C., Shimeld, S.M., 2015. Characterization of two neurogenin genes from the brook lamprey lampetra planeri and their expression in the lamprey nervous system. Dev. Dyn. 244, 1096–1108. https://doi.org/10.1002/dvdy.24273

Lara-Ramírez, R., Poncelet, G., Patthey, C., Shimeld, S.M., 2017. The structure, splicing, synteny and expression of lamprey COE genes and the evolution of the COE gene family in chordates. Dev. Genes Evol. 227, 319–338. https://doi.org/10.1007/s00427-017-0591-6

Le Douarin, N.M., Fontaine-Pérus, J., Couly, G., 1986. Cephalic ectodermal placodes and neurogenesis. Trends Neurosci. 9, 175–180. https://doi.org/10.1016/0166-2236(86)90055-X

Lee, J.E., Hollenberg, S.M., Snider, L., Turner, D.L., Lipnick, N., Weintraub, H., 1995. Conversion of Xenopus ectoderm into neurons by neuroD, a basic helix-loop-helix protein. Science (80-.). 268, 836–844. https://doi.org/10.1126/science.7754368

Lee, S.A., Shen, E.L., Fiser, A., Sali, A., Guo, S., 2003. The zebrafish forkhead transcription factor Foxi1 specifies epibranchial placode-derived sensory neurons. Development 130, 2669–2679. https://doi.org/10.1242/dev.00502

Lemaire, P., 2009. Unfolding a chordate developmental program, one cell at a time: Invariant cell lineages, short-range inductions and evolutionary plasticity in ascidians. Dev. Biol. 332, 48–60. https://doi.org/10.1016/j.ydbio.2009.05.540

Lemaire, P., Smith, W.C., Nishida, H., 2008. Ascidians and the Plasticity of the Chordate Developmental Program. Curr. Biol. 18, 620–631. https://doi.org/10.1016/j.cub.2008.05.039

Litsiou, A., Hanson, S., Streit, A., 2005. A balance of FGF, BMP and WNT signaling positions the future placode territory in the head. Development 132, 4051–4062. https://doi.org/10.1242/dev.01964

Liu, A., Niswander, L.A., 2005. Bone morphogenetic protein signalling and vertebrate nervous system development. Nat. Rev. Neurosci. 6, 945–954. https://doi.org/10.1038/nrn1805

Lorke, D.E., Susens, U., Borgmeyer, U., Hermans-Borgmeyer, I., 2000. Differential expression of the estrogen receptor-related receptor gamma in the mouse brain. Brain Res Mol Brain Res 77, 277–280.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, T.-M., Luo, Y.-J., Yu, J.-K., 2012. BMP and Delta/Notch signaling control the development of amphioxus epidermal sensory neurons: insights into the evolution of the peripheral sensory system. Development 139, 2020–2030. https://doi.org/10.1242/dev.073833

Lupo, G., Harris, W.A., Lewis, K.E., 2006. Mechanisms of ventral patterning in the vertebrate nervous system. Nat. Rev. Neurosci. 7, 103–114. https://doi.org/10.1038/nrn1843

Ma, Q., Chen, Z., Barrantes, I.D.B., De La Pompa, J.L., Anderson, D.J., 1998. Neurogenin1 Is Essential for the Determination of Neuronal Precursors for Proximal Cranial Sensory Ganglia. Neuron 20, 469–482.

https://doi.org/10.1016/S0896-6273(00)80988-5

Maden, M., 2007. Retinoic acid in the development, regeneration and maintenance of the nervous system. Nat. Rev. Neurosci. 8, 755–765. https://doi.org/10.1038/nrn2212

Maklad, A., Reed, C., Johnson, N.S., Fritzsch, B., 2014. Anatomy of the lamprey ear: Morphological evidence for occurrence of horizontal semicircular ducts in the labyrinth of Petromyzon marinus. J. Anat. 224, 432–446. https://doi.org/10.1111/joa.12159

Mallatt, J., Sullivan, J., 1998. 28S and 18S rDNA sequences support the monophyly of lampreys and hag fishes. Mol. Biol. Evol. 15, 1706–1718. https://doi.org/10.1093/oxfordjournals.molbev.a025897

Mallatt, J., Winchell, C.J., 2007. Ribosomal RNA genes and deuterostome phylogeny revisited: More cyclostomes, elasmobranchs, reptiles, and a brittle star. Mol. Phylogenet. Evol. 43, 1005–1022. https://doi.org/10.1016/j.ympev.2006.11.023

Mangelsdorf, D.J., Borgmeyer, U., Heyman, R.A., Yang Zhou, J., Ong, E.S., Oro, A.E., Kakizuka, A., Evans, R.M., 1992. Characterization of three RXR genes that mediate the action of 9-cis retinoic acid. Genes Dev. 6, 329–344. https://doi.org/10.1101/gad.6.3.329

Manni, L., Mackie, G.O., Caicci, F., Zaniolo, G., Burighel, P., 2006. Coronal organ of ascidians and the evolutionary significance of secondary sensory cells in chordates. J. Comp. Neurol. 495, 363–373. https://doi.org/10.1002/cne.20867

Marlétaz, F., Firbas, P.N., Maeso, I., Tena, J.J., Bogdanovic, O., Perry, M., Wyatt, C.D.R., de la Calle-Mustienes, E., Bertrand, S., Burguera, D., Acemel, R.D., van Heeringen, S.J., Naranjo, S., Herrera-Ubeda, C., Skvortsova, K., Jimenez-Gancedo, S., Aldea, D., Marquez, Y., Buono, L., Kozmikova, I., Permanyer, J.,

Louis, A., Albuixech-Crespo, B., Le Petillon, Y., Leon, A., Subirana, L., Balwierz, P.J., Duckett, P.E., Farahani, E., Aury, J.M., Mangenot, S., Wincker, P., Albalat, R., Benito-Gutiérrez, È., Cañestro, C., Castro, F., D'Aniello, S., Ferrier, D.E.K., Huang, S., Laudet, V., Marais, G.A.B., Pontarotti, P., Schubert, M., Seitz, H., Somorjai, I., Takahashi, T., Mirabeau, O., Xu, A., Yu, J.-K., Carninci, P., Martinez-Morales, J.R., Crollius, H.R., Kozmik, Z., Weirauch, M.T., Garcia-Fernàndez, J., Lister, R., Lenhard, B., Holland, P.W.H., Escriva, H., Gómez-Skarmeta, J.L., Irimia, M., 2018. Amphioxus functional genomics and the origins of vertebrate gene regulation. Nature 564, 64–70. https://doi.org/10.1038/s41586-018-0734-6

Marlow, H., Tosches, M.A., Tomer, R., Steinmetz, P.R., Lauri, A., Larsson, T., Arendt, D., 2014. Larval body patterning and apical organs are conserved in animal evolution. BMC Biol. 12, 7. https://doi.org/10.1186/1741-7007-12-7

Martinez, P., Davidson, E.H., 1997. SpHmx, a sea urchin homeobox gene expressed in embryonic pigment cells. Dev. Biol. 181, 213–222. https://doi.org/10.1006/dbio.1996.8461

Matsukawa, S., Miwata, K., Asashima, M., Michiue, T., 2015. The requirement of histone modification by PRDM12 and Kdm4a for the development of pre-placodal ectoderm and neural crest in Xenopus. Dev. Biol. 399, 164–76. https://doi.org/10.1016/j.ydbio.2014.12.028

Mazet, F., Hutt, J.A., Milloz, J., Millard, J., Graham, A., Shimeld, S.M., 2005. Molecular evidence from Ciona intestinalis for the evolutionary origin of vertebrate sensory placodes. Dev Biol 282, 494–508. https://doi.org/10.1016/j.ydbio.2005.02.021

Mazet, F., Masood, S., Luke, G.N., Holland, N.D., Shimeld, S.M., 2004. Expression of AmphiCoe, an Amphioxus COE/EBF Gene, in the Developing Central Nervous System and Epidermal Sensory Neurons. Genesis 38, 58–65.

https://doi.org/10.1002/gene.20006

Mazet, F., Shimeld, S.M., 2005. Molecular evidence from ascidians for the evolutionary origin of vertebrate cranial sensory placodes. J. Exp. Zool. Part B Mol. Dev. Evol. 304, 340–346. https://doi.org/10.1002/jez.b.21054

McCabe, K.L., Bronner-Fraser, M., 2009. Molecular and tissue interactions governing induction of cranial ectodermal placodes. Dev. Biol. 332, 189–195. https://doi.org/10.1016/j.ydbio.2009.05.572

McCormick, L.J., Hutt, J.A., Hazan, J., Houart, C., Cohen, J., 2007. The homeodomain transcription factor drg11 is expressed in primary sensory neurons and their putative CNS targets during embryonic development of the zebrafish. Gene Expr. Patterns 7, 289–296. https://doi.org/10.1016/j.modgep.2006.09.003

McEwen, G.K., Goode, D.K., Parker, H.J., Woolfe, A., Callaway, H., Elgar, G., 2009. Early evolution of conserved regulatory sequences associated with development in vertebrates. PLoS Genet. 5. https://doi.org/10.1371/journal.pgen.1000762

McEwen, G.K., Woolfe, A., Goode, D., Vavouri, T., Callaway, H., Elgar, G., 2006. Ancient duplicated conserved noncoding elements in vertebrates: A genomic and functional analysis. Genome Res. 16, 451–465. https://doi.org/10.1101/gr.4143406

Mehta, T.K., Ravi, V., Yamasaki, S., Lee, A.P., Lian, M.M., Tay, B.-H., Tohari, S., Yanai, S., Tay, A., Brenner, S., Venkatesh, B., 2013. Evidence for at least six Hox clusters in the Japanese lamprey (Lethenteron japonicum). Proc. Natl. Acad. Sci. 110, 16044–16049. https://doi.org/10.1073/pnas.1315760110

Meulemans, D., Bronner-Fraser, M., 2007. The Amphioxus SoxB Family: Implications for the Evolution of Vertebrate Placodes. Int. J. Biol. Sci. 33, 356–364. https://doi.org/10.7150/ijbs.3.356

Meulemans, D., Bronner-Fraser, M., 2002. Amphioxus and lamprey AP-2 genes: implications for neural crest evolution and migration patterns. Development 129, 4953–4962.

Meyer, A., Van De Peer, Y., 2005. From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). BioEssays 27, 937–945. https://doi.org/10.1002/bies.20293

Miya, T., Nishida, H., 2003. Expression Pattern and Transcriptional Control of SoxB1 in Embryos of the Ascidian Halocynthia roretzi. Zoolog. Sci. 20, 59–67. https://doi.org/10.2108/zsj.20.59

Modrell, M.S., Hockman, D., Uy, B., Buckley, D., Sauka-Spengler, T., Bronner, M.E., Baker, C.V.H., 2014. A fate-map for cranial sensory ganglia in the sea lamprey. Dev. Biol. 385, 405–416. https://doi.org/10.1016/j.ydbio.2013.10.021

Monteiro, A.S., Schierwater, B., Dellaporta, S.L., Holland, P.W.H., 2006. A low diversity of ANTP class homeobox genes in Placozoa. Evol. Dev. 8, 174–182. https://doi.org/10.1111/j.1525-142X.2006.00087.x

Morales, M., Martínez-Paz, P., Sánchez-Argüello, P., Morcillo, G., Martínez-Guitarte, J.L., 2018. Bisphenol A (BPA) modulates the expression of endocrine and stress response genes in the freshwater snail Physa acuta. Ecotoxicol. Environ. Saf. 152, 132–138. https://doi.org/10.1016/j.ecoenv.2018.01.034

Murakami, Y., Pasqualetti, M., Takio, Y., Hirano, S., Rijli, F.M., Kuratani, S., 2004. Segmental development of reticulospinal and branchiomotor neurons in lamprey: insights into the evolution of the vertebrate hindbrain. Development 131, 983–995. https://doi.org/10.1242/dev.00986

Murakami, Y., Uchida, K., Rijli, F.M., Kuratani, S., 2005. Evolution of the brain developmental plan: Insights from agnathans. Dev. Biol. 280, 249–259. https://doi.org/10.1016/j.ydbio.2005.02.008

Nagy, V., Cole, T., Van Campenhout, C., Khoung, T.M., Leung, C., Vermeiren, S., Novatchkova, M., Wenzel, D., Cikes, D., Polyansky, A.A., Kozieradzki, I., Meixner, A., Bellefroid, E.J., Neely, G.G., Penninger, J.M., 2015. The evolutionarily conserved transcription factor PRDM12 controls sensory neuron development and pain perception. Cell Cycle 4101, 1799–1808. https://doi.org/10.1080/15384101.2015.1036209

Nakada, Y., Hunsaker, T.L., Henke, R.M., Johnson, J.E., 2004. Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification. Development 131, 1319–1330. https://doi.org/10.1242/dev.01008

Nikitina, N., Tong, L., Bronner, M.E., 2011. Ancestral network module regulating prdm1 expression in the lamprey neural plate border. Dev. Dyn. 240, 2265–2271. https://doi.org/10.1002/dvdy.22720

Nishijima, I., Ohtoshi, A., 2006. Characterization of a novel prospero-related homeobox gene, Prox2. Mol. Genet. Genomics 275, 471–478. https://doi.org/10.1007/s00438-006-0105-0

Nolan, L.S., Maier, H., Hermans-Borgmeyer, I., Girotto, G., Ecob, R., Pirastu, N., Cadge, B. a., Hübner, C., Gasparini, P., Strachan, D.P., Davis, A., Dawson, S.J., 2013. Estrogen-related receptor gamma and hearing function: evidence of a role in humans and mice. Neurobiol. Aging 34, 2077.e1-2077.e9. https://doi.org/10.1016/j.neurobiolaging.2013.02.009

Northcutt, R.G., Gans, C., 1983. The Genesis of Neural Crest and Epidermal Placodes: A Reinterpretation of Vertebrate Origins. Q. Rev. Biol. 58, 1–28. https://doi.org/10.1086/413055

O'Neill, P., Mak, S.S., Fritzsch, B., Ladher, R.K., Baker, C.V.H., 2012. The amniote

paratympanic organ develops from a previously undiscovered sensory placode. Nat. Commun. 3. https://doi.org/10.1038/ncomms2036

Ohno, S., 2013. Evolution by gene duplication. Springer Science & Business Media.

Onishi, A., Peng, G.H., Poth, E.M., Lee, D.A., Chen, J., Alexis, U., De Melo, J., Chen, S., Blackshaw, S., 2010. The orphan nuclear hormone receptor ERRβ controls rod photoreceptor survival. Proc. Natl. Acad. Sci. U. S. A. 107, 11579–11584. https://doi.org/10.1073/pnas.1000102107

Osorio, J., Mazan, S., Rétaux, S., 2005. Organisation of the lamprey (Lampetra fluviatilis) embryonic brain: Insights from LIM-homeodomain, Pax and hedgehog genes. Dev. Biol. 288, 100–112. https://doi.org/10.1016/j.ydbio.2005.08.042

Osório, J., Rétaux, S., 2008. The lamprey in evolutionary studies. Dev. Genes Evol. 218, 221–235. https://doi.org/10.1007/s00427-008-0208-1

Paps, J., Holland, P.W.H., Shimeld, S.M., 2012. A genome-wide view of transcription factor gene diversity in chordate evolution: Less gene loss in amphioxus? Brief. Funct. Genomics 11, 177–186. https://doi.org/10.1093/bfgp/els012

Park, K., Kwak, I.-S., 2010. Molecular effects of endocrine-disrupting chemicals on the Chironomus riparius estrogen-related receptor gene. Chemosphere 79, 934–941. https://doi.org/10.1016/j.chemosphere.2010.03.002

Parker, G.H., 1908. The Sensory Reactions of Amphioxus. Proc. Am. Acad. Arts Sci. 43, 415. https://doi.org/10.2307/20022358

Parker, H.J., Bronner, M.E., Krumlauf, R., 2014a. A Hox regulatory network of hindbrain segmentation is conserved to the base of vertebrates. Nature 514, 490–3. https://doi.org/10.1038/nature13723

Parker, H.J., Piccinelli, P., Sauka-Spengler, T., Bronner, M., Elgar, G., 2011. Ancient Pbx-Hox signatures define hundreds of vertebrate developmental enhancers.

BMC Genomics 12, 637. https://doi.org/10.1186/1471-2164-12-637

Parker, H.J., Sauka-Spengler, T., Bronner, M., Elgar, G., 2014b. A Reporter Assay in Lamprey Embryos Reveals Both Functional Conservation and Elaboration of Vertebrate Enhancers. PLoS One 9, e85492. https://doi.org/10.1371/journal.pone.0085492

Patthey, C., Clifford, H., Haerty, W., Ponting, C.P., Shimeld, S.M., Begbie, J., 2016. Identification of molecular signatures specific for distinct cranial sensory ganglia in the developing chick. Neural Dev. 11, 3. https://doi.org/10.1186/s13064-016-0057-y

Patthey, C., Schlosser, G., Shimeld, S.M., 2014. The evolutionary history of vertebrate cranial placodes – I: Cell type evolution. Dev. Biol. 389, 82–97. https://doi.org/10.1016/j.ydbio.2014.01.017

Perron, M., Opdecamp, K., Butler, K., Harris, W.A., Bellefroid, E.J., 1999. X-ngnr-1 and Xath3 promote ectopic expression of sensory neuron markers in the neurula ectoderm and have distinct inducing properties in the retina. Proc. Natl. Acad. Sci. U. S. A. 96, 14996–15001. https://doi.org/10.1073/pnas.96.26.14996

Pieper, M., Eagleson, G.W., Wosniok, W., Schlosser, G., 2011. Origin and segregation of cranial placodes in Xenopus laevis. Dev. Biol. 360, 257–275. https://doi.org/10.1016/j.ydbio.2011.09.024

Pistocchi, A., Bartesaghi, S., Cotelli, F., Del Giacco, L., 2008. Identification and expression pattern of zebrafish prox2 during embryonic development. Dev. Dyn. 237, 3916–20. https://doi.org/10.1002/dvdy.21798

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutiérrez, È., Dubchak, I., Garcia-Fernàndez, J., Gibson-Brown, J.J., Grigoriev, I. V., Horton,

A.C., De Jong, P.J., Jurka, J., Kapitonov, V. V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W.H., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071. https://doi.org/10.1038/nature06967

Qiu, H., Hildebrand, F., Kuraku, S., Meyer, A., 2011. Unresolved orthology and peculiar coding sequence properties of lamprey genes: The KCNA gene family as test case. BMC Genomics 12. https://doi.org/10.1186/1471-2164-12-325

Quina, L.A., Tempest, L., Hsu, Y.W.A., Cox, T.C., Turner, E.E., 2012. Hmx1 is required for the normal development of somatosensory neurons in the geniculate ganglion. Dev. Biol. 365, 152–163. https://doi.org/10.1016/j.ydbio.2012.02.022

Real, M.A., Heredia, R., Dávila, J.C., Guirado, S., 2008. Efferent retinal projections visualized by immunohistochemical detection of the estrogen-related receptor beta in the postnatal and adult mouse brain. Neurosci. Lett. 438, 48–53. https://doi.org/10.1016/j.neulet.2008.04.044

Rebelo, S., Chen, Z.-F., Anderson, D.J., Lima, D., 2006. Involvement of DRG11 in the development of the primary afferent nociceptive system. Mol. Cell. Neurosci. 33, 236–46. https://doi.org/10.1016/j.mcn.2006.07.013

Rebelo, S., Reguenga, C., Osório, L., Pereira, C., Lopes, C., Lima, D., 2007. DRG11 immunohistochemical expression during embryonic development in the mouse. Dev. Dyn. 236, 2653–60. https://doi.org/10.1002/dvdy.21271

Richardson, M.K., Admiraal, J., Wright, G.M., 2010. Developmental anatomy of lampreys. Biol. Rev. 85, 1–33. https://doi.org/10.1111/j.1469-185X.2009.00092.x

Rigon, F., Gasparini, F., Shimeld, S.M., Candiani, S., Manni, L., 2018. Developmental

signature, synaptic connectivity and neurotransmission are conserved between vertebrate hair cells and tunicate coronal cells. J. Comp. Neurol. 526, 957–971. https://doi.org/10.1002/cne.24382

Roberts, J.A., Miguel-Escalada, I., Slovik, K.J., Walsh, K.T., Hadzhiev, Y., Sanges, R., Stupka, E., Marsh, E.K., Balciuniene, J., Balciunas, D., Muller, F., 2014. Targeted transgene integration overcomes variability of position effects in zebrafish. Development 141, 715–724. https://doi.org/10.1242/dev.100347

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Rosin, J.M., Kurrasch, D.M., Cobb, J., 2015a. Shox2 is required for the proper development of the facial motor nucleus and the establishment of the facial nerves. BMC Neurosci. 16, 39. https://doi.org/10.1186/s12868-015-0176-0

Rosin, J.M., McAllister, B.B., Dyck, R.H., Percival, C.J., Kurrasch, D.M., Cobb, J., 2015b. Mice lacking the transcription factor SHOX2 display impaired cerebellar development and deficits in motor coordination. Dev. Biol. 399, 54–67. https://doi.org/10.1016/j.ydbio.2014.12.013

Ryan, K., Lu, Z., Meinertzhagen, I.A., 2017. Circuit Homology between Decussating Pathways in the Ciona Larval CNS and the Vertebrate Startle-Response Pathway. Curr. Biol. 27, 721–728. https://doi.org/10.1016/j.cub.2017.01.026

Ryan, K., Lu, Z., Meinertzhagen, I.A., 2016. The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling. Elife 5, 1–34. https://doi.org/10.7554/elife.16962

Sacerdot, C., Louis, A., Bon, C., Berthelot, C., Roest Crollius, H., 2018. Chromosome evolution at the origin of the ancestral vertebrate genome. Genome Biol 19,

253104. https://doi.org/10.1101/253104

Saint-Jeannet, J.P., Moody, S.A., 2014. Establishing the pre-placodal region and breaking it into placodes with distinct identities. Dev. Biol. 389, 13–27. https://doi.org/10.1016/j.ydbio.2014.02.011

Sanes, D.H., Reh, T.A., Harris, W.A., 2011. Development of the Nervous System: Third Edition. Dev. Nerv. Syst. Third Ed. 1–341.

Sanes, D.H., Reh, T.A., Harris, W.A., 2006. Development of the Nervous System.

Satoh, G., Wang, Y., Zhang, P., Satoh, N., 2001. Early development of amphioxus nervous system with special reference to segmental cell organization and putative sensory cell precursors: a study based on the expression of pan-neuronal marker gene Hu/elav. J Exp Zool 291, 354–364. https://doi.org/10.1002/jez.1134

Saudemont, A., Dray, N., Hudry, B., Le Gouar, M., Vervoort, M., Balavoine, G., 2008. Complementary striped expression patterns of NK homeobox genes during segment formation in the annelid Platynereis. Dev. Biol. 317, 430–443. https://doi.org/10.1016/j.ydbio.2008.02.013

Sauka-Spengler, T., Meulemans, D., Jones, M., Bronner-Fraser, M., 2007. Ancient Evolutionary Origin of the Neural Crest Gene Regulatory Network. Dev. Cell 13, 405–420. https://doi.org/10.1016/j.devcel.2007.08.005

Schlosser, G., 2018. A short history of nearly every sense — The evolutionary history of vertebrate sensory cell types. Integr. Comp. Biol. 58, 301–316. https://doi.org/10.1093/icb/icy024

Schlosser, G., 2006. Induction and specification of cranial placodes. Dev. Biol. 294, 303–351. https://doi.org/10.1016/j.ydbio.2006.03.009

Schlosser, G., Northcutt, R.G., 2000. Development of neurogenic placodes in Xenopus laevis. J. Comp. Neurol. 418, 121–146.

https://doi.org/10.1002/(SICI)1096-9861(20000306)418:2<121::AID-CNE1>3.0.CO;2-M

Schlosser, G., Patthey, C., Shimeld, S.M., 2014. The evolutionary history of vertebrate cranial placodes II. Evolution of ectodermal patterning. Dev. Biol. 389, 98–119. https://doi.org/10.1016/j.ydbio.2014.01.019

Schubert, M., Holland, N.D., Escriva, H., Holland, L.Z., Laudet, V., 2004. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). Proc. Natl. Acad. Sci. 101, 10320–10325. https://doi.org/10.1073/pnas.0403216101

Scott, A., Hasegawa, H., Sakurai, K., Yaron, A., Cobb, J., Wang, F., 2011. Transcription Factor Short Stature Homeobox 2 Is Required for Proper Development of Tropomyosin-Related Kinase B-Expressing Mechanosensory Neurons. J. Neurosci. 31, 6741–6749. https://doi.org/10.1523/JNEUROSCI.5883-10.2011

Sharma, S., Wang, W., Stolfi, A., 2019. Single-cell transcriptome profiling of the Ciona larval brain. Dev. Biol. 448, 226–236. https://doi.org/10.1016/j.ydbio.2018.09.023

Shimeld, S.M., Donoghue, P.C.J., 2012. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). Development 139, 2091–2099. https://doi.org/10.1242/dev.074716

Shimeld, S.M., Purkiss, A.G., Dirks, R.P.H., Bateman, O.A., Slingsby, C., Lubsen, N.H., 2005. Urochordate βγ-crystallin and the evolutionary origin of the vertebrate eye lens. Curr. Biol. 15, 1684–1689. https://doi.org/10.1016/j.cub.2005.08.046

Smith, J.J., Keinath, M.C., 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res. 25, 1081–1090. https://doi.org/10.1101/gr.184135.114

Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., Morgan, J.R., Buxbaum, J.D., Sachidanandam, R., Sims, C., Garruss, A.S., Cook, M., Krumlauf, R., Wiedemann, L.M., Sower, S.A., Decatur, W.A., Hall, J.A., Amemiya, C.T., Saha, N.R., Buckley, K.M., Rast, J.P., Das, S., Hirano, M., McCurley, N., Guo, P., Rohner, N., Tabin, C.J., Piccinelli, P., Elgar, G., Ruffier, M., Aken, B.L., Searle, S.M.J., Muffato, M., Pignatelli, M., Herrero, J., Jones, M., Brown, C.T., Chung-Davidson, Y.W., Nanlohy, K.G., Libants, S. V., Yeh, C.Y., McCauley, D.W., Langeland, J.A., Pancer, Z., Fritzsch, B., De Jong, P.J., Zhu, B., Fulton, L.L., Theising, B., Flicek, P., Bronner, M.E., Warren, W.C., Clifton, S.W., Wilson, R.K., Li, W., 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat. Genet. 45, 415–421. https://doi.org/10.1038/ng.2568

Smith, J.J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M.C., Parker, H.J., Cook, M.E., Hess, J.E., Narum, S.R., Lamanna, F., Kaessmann, H., Timoshevskiy, V.A., Waterbury, C.K.M., Saraceno, C., Wiedemann, L.M., Robb, S.M.C., Baker, C., Eichler, E.E., Hockman, D., Sauka-Spengler, T., Yandell, M., Krumlauf, R., Elgar, G., Amemiya, C.T., 2018. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. Nat. Genet. 50, 270–277. https://doi.org/10.1038/s41588-017-0036-1

Sommer, L., Ma, Q., Anderson, D.J., 1996. neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. Mol. Cell. Neurosci. 8, 221–241. https://doi.org/10.1006/mcne.1996.0060

Squire, L.R., Berg, D., Bloom, F., Du Lac, S., Ghosh, A., Spitzer, N., 2012. Fundamental Neuroscience: Fourth Edition 1–1127.

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steventon, B., Mayor, R., Streit, A., 2014. Neural crest and placode interaction during the development of the cranial sensory system. Dev. Biol. 389, 28–38. https://doi.org/10.1016/j.ydbio.2014.01.021

Stolfi, A., Ryan, K., Meinertzhagen, I.A., Christiaen, L., 2015. Migratory neuronal progenitors arise from the neural plate borders in tunicates. Nature 527, 371–374. https://doi.org/10.1038/nature15758

Streit, A., 2002. Extensive cell movements accompany formation of the otic placode. Dev. Biol. 249, 237–254. https://doi.org/10.1006/dbio.2002.0739

Sugahara, F., Pascual-Anaya, J., Oisi, Y., Kuraku, S., Aota, S., Adachi, N., Takagi, W., Hirai, T., Sato, N., Murakami, Y., Kuratani, S., 2016. Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. Nature 531, 97–100. https://doi.org/10.1038/nature16518

Sun, Y., Dykes, I.M., Liang, X., Eng, S.R., Evans, S.M., Turner, E.E., 2008. A central role for Islet1 in sensory neuron development linking sensory and spinal gene regulatory programs. Nat. Neurosci. 11, 1283–1293. https://doi.org/10.1038/nn.2209

Sun, Z., Wang, H.B., Laverghetta, A., Yamamoto, K., Reiner, A., 2005. The distribution and cellular localization of glutamic acid decarboxylase-65 (GAD65) mRNA in the forebrain and midbrain of domestic chick. J. Chem. Neuroanat. 29, 265–281. https://doi.org/10.1016/j.jchemneu.2005.02.003

Süsens, U., Hermans-Borgmeyer, I., Borgmeyer, U., 2000. Alternative splicing and expression of the mouse estrogen receptor-related receptor γ. Biochem. Biophys. Res. Commun. 267, 532–535. https://doi.org/10.1006/bbrc.1999.1976

Tahara, Y., 1988. Normal stages of development in the lamprey, Lampetra reissneri (Dybowski). Zoolog. Sci. 5, 109–118.

Takahashi, H., Shintani, T., Sakuta, H., Noda, M., 2003. CBF1 controls the retinotectal topographical map along the anteroposterior axis through multiple mechanisms. Development 130, 5203–5215. https://doi.org/10.1242/dev.00724

Takayanagi, S., Tokunaga, T., Liu, X., Okada, H., Matsushima, A., Shimohigashi, Y., 2006. Endocrine disruptor bisphenol A strongly binds to human estrogen-related receptor γ (ERRγ) with high constitutive activity. Toxicol. Lett. 167, 95–105. https://doi.org/10.1016/j.toxlet.2006.08.012

Takezaki, N., Figueroa, F., Zaleska-Rutczynska, Z., Klein, J., 2003. Molecular phylogeny of early vertebrates: Monophyly of the agnathans as revealed by sequences of 35 genes. Mol. Biol. Evol. 20, 287–292. https://doi.org/10.1093/molbev/msg040

Thelie, A., Desiderio, S., Hanotel, J., Quigley, I., Van Driessche, B., Rodari, A., Borromeo, M.D., Kricha, S., Lahaye, F., Croce, J., Cerda-Moya, G., Ordono Fernandez, J., Bolle, B., Lewis, K.E., Sander, M., Pierani, A., Schubert, M., Johnson, J.E., Kintner, C.R., Pieler, T., Van Lint, C., Henningfeld, K.A., Bellefroid, E.J., Van Campenhout, C., 2015. Prdm12 specifies V1 interneurons through cross-repressive interactions with Dbx1 and Nkx6 genes in Xenopus. Development 142, 3416–3428. https://doi.org/10.1242/dev.121871

Thornton, J., Need, E., Crews, D., 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science 301, 1714–7.

https://doi.org/10.1126/science.1086185

Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., Kaufman, T.C., Calvi, B.R., Perrimon, N., Gelbart, S.R., Agapite, J., Broll, K., Crosby, L., Santos, G. dos, Emmert, D., Gramates, L.S., Falls, K., Jenkins, V., Matthews, B., Sutherland, C., Tabone, C., Zhou, P., Zytkovicz, M., Brown, N., Antonazzo, G., Attrill, H., Garapati, P., Holmes, A., Larkin, A., Marygold, S., Millburn, G., Pilgrim, C., Trovisco, V., Urbano, P., Kaufman, T., Calvi, B., Czoch, B., Goodman, J., Strelets, V., Thurmond, J., Cripps, R., Baker, P., 2019. FlyBase 2.0: the next generation. Nucleic Acids Res. 47, D759–D765. https://doi.org/10.1093/nar/gky1003

Tremblay, A.M., Giguère, V., 2007. The NR3B Subgroup: An Overrview. Nucl. Recept. Signal. 5, nrs.05009. https://doi.org/10.1621/nrs.05009

Tsuda, M., Sakurai, D., Goda, M., 2003. Direct evidence for the role of pigment cells in the brain of ascidian larvae by laser ablation. J. Exp. Biol. 206, 1409–1417. https://doi.org/10.1242/jeb.00235

Vit, J.P., Ohara, P.T., Sundberg, C., Rubi, B., Maechler, P., Liu, C., Puntel, M., Lowenstein, P., Castro, M., Jasmin, L., 2009. Adenovector GAD65 gene delivery into the rat trigeminal ganglion produces orofacial analgesia. Mol. Pain 5, 1–11. https://doi.org/10.1186/1744-8069-5-42

Vopalensky, P., Pergner, J., Liegertova, M., Benito-Gutierrez, E., Arendt, D., Kozmik, Z., 2012. Molecular analysis of the amphioxus frontal eye unravels the evolutionary origin of the retina and pigment cells of the vertebrate eye. Proc. Natl. Acad. Sci. U. S. A. 109, 15383–15388. https://doi.org/10.1073/pnas.1207580109

Wada, S., Saiga, H., 2002. HrzicN, a new Zic family gene of ascidians, plays essential

roles in the neural tube and notochord development. Development 129, 5597–5608. https://doi.org/10.1242/dev.00156

Wada, S., Tokuoka, M., Shoguchi, E., Kobayashi, K., Di Gregorio, A., Spagnuolo, A., Branno, M., Kohara, Y., Rokhsar, D., Levine, M., Saiga, H., Satoh, N., Satou, Y., 2003. A genomewide survey of developmentally relevant genes in Ciona intestinalis. II. Genes for homeobox transcription factors. Dev Genes Evol 213, 222–234. https://doi.org/10.1007/s00427-003-0321-0

Wagner, E., Levine, M., 2012. FGF signaling establishes the anterior border of the Ciona neural tube. Dev. 139, 2351–2359. https://doi.org/10.1242/dev.078485

Wang, C.-Z., Shi, M., Yang, L.-L., Yang, R.-Q., Luo, Z.-G., Jacquin, M.F., Chen, Z.-F., Ding, Y.-Q., 2007. Development of the mesencephalic trigeminal nucleus requires a paired homeodomain transcription factor, Drg11. Mol. Cell. Neurosci. 35, 368–376. https://doi.org/10.1016/j.mcn.2007.03.011

Wang, W., Chan, E.K., Baron, S., Van de Water, T., Lufkin, T., 2001. Hmx2 homeobox gene control of murine vestibular morphogenesis. Development 128, 5017–5029.

Wang, W., Grimmer, J.F., Van De Water, T.R., Lufkin, T., 2004. Hmx2 and Hmx3 homeobox genes direct development of the murine inner ear and hypothalamus and can be functionally replaced by Drosophila Hmx. Dev. Cell 7, 439–453. https://doi.org/10.1016/j.devcel.2004.06.016

Wang, W., Lo, P., Frasch, M., Lufkin, T., 2000. Hmx: An evolutionary conserved homeobox gene family expressed in the developing nervous system in mice and Drosophila. Mech. Dev. 99, 123–137. https://doi.org/10.1016/S0925-4773(00)00488-3

Wang, W, Van De Water, T., Lufkin, T., 1998. Inner ear and maternal reproductive defects in mice lacking the Hmx3 homeobox gene. Development 125, 621–34.

Wang, Weidong, Van De Water, T., Lufkin, T., Warchol, M.E., Corwin, J.T., 1998. Inner ear and maternal reproductive defects in mice lacking the Hmx3 homeobox gene. Development 125, 621–34.

Watanabe, A., Hirano, S., Murakami, Y., 2008. Development of the lamprey central nervous system, with reference to vertebrate evolution. Zoolog. Sci. 25, 1020–1027. https://doi.org/10.2108/zsj.25.1020

Werner, G.D.A., Gemmell, P., Grosser, S., Hamer, R., Shimeld, S.M., 2013. Analysis of a deep transcriptome from the mantle tissue of Patella vulgata Linnaeus (Mollusca: Gastropoda: Patellidae) reveals candidate biomineralising genes. Mar. Biotechnol. 15, 230–243. https://doi.org/10.1007/s10126-012-9481-0

Woolfe, A., Goode, D.K., Cooke, J., Callaway, H., Smith, S., Snell, P., McEwen, G.K., Elgar, G., 2007. CONDOR: a database resource of developmentally associated conserved non-coding elements. BMC Dev Biol 7, 100. https://doi.org/10.1186/1471-213X-7-100

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y.J.K., Cooke, J.E., Elgar, G., 2005. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 3. https://doi.org/10.1371/journal.pbio.0030007

Wotton, K.R., Shimeld, S.M., 2011. Analysis of lamprey clustered Fox genes: Insight into Fox gene evolution and expression in vertebrates. Gene 489, 30–40. https://doi.org/10.1016/j.gene.2011.08.007

Wu, F., J.Kaczynski, T., Sethuramanujam, S., Li, R., Jain, V., Slaughter, M., Mu, X., 2015. Two transcription factors, Pou4f2 and Isl1, are sufficient to specify the retinal ganglion cell fate. Proc. Natl. Acad. Sci. U. S. A. 112, E1559–E1568.

https://doi.org/10.1073/pnas.1421535112

Ybot-Gonzalez, P., Gaston-Massuet, C., Girdler, G., Klingensmith, J., Arkell, R., Greene, N.D.E., Copp, A.J., 2007. Neural plate morhogenesis during mouse neurulation is regulated by antagonism of Bmp signalling. Development 134, 3203–3211. https://doi.org/10.1242/dev.008177

Yu, J.-K., Meulemans, D., McKeown, S.J., Bronner-Fraser, M., 2008. Insights from the amphioxus genome on the origin of vertebrate neural crest. Genome Res. 18, 1127–1132. https://doi.org/10.1101/gr.076208.108

Yu, S., Zhang, W., Li, L., Huang, H., Ma, F., Li, Q., 2008. Phylogenetic analysis of 48 gene families revealing relationships between Hagfishes, Lampreys, and Gnathostomata. J. Genet. Genomics 35, 285–290. https://doi.org/10.1016/S1673-8527(08)60041-7

Zannino, D. a., Sagerström, C.G., 2015. An emerging role for prdm family genes in dorsoventral patterning of the vertebrate nervous system. Neural Dev. 10, 24. https://doi.org/10.1186/s13064-015-0052-8

Zannino, D.A., Downes, G.B., Sagerström, C.G., 2014. Prdm12b specifies the p1 progenitor domain and reveals a role for V1 interneurons in swim movements. Dev. Biol. 390, 247–260. https://doi.org/10.1016/j.ydbio.2014.02.025

Zhong, Y.-F., Butts, T., Holland, P.W.H., 2008. HomeoDB: a database of homeobox gene diversity. Evol. Dev. 10, 516–518. https://doi.org/10.1111/j.1525-142X.2008.00266.x

Zhong, Y., Holland, P.W.H., 2011. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. Evol. Dev. 13, 567–568. https://doi.org/10.1111/j.1525-142X.2011.00513.x