Title
Conservation of gene expression patterns in nervous system evolution

Permalink
https://escholarship.org/uc/item/8bv5d2d1

Author
Esteves, Francisco Freire

Publication Date
2012

Peer reviewed|Thesis/dissertation
Conservation of gene expression patterns in nervous system evolution

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

in

Biology

by

Francisco Freire Esteves

Committee in Charge:

Ethan Bier, Chair
Fred H. Gage
Timothy Q. Gentner
Ralph J. Greenspan
Harvey J. Karten
William James McGinnis

2012
The Dissertation of Francisco Freire Esteves is approved, and it is acceptable in quality and form for publication of microfilm and electronically:

_______________________________________________

_______________________________________________

_______________________________________________

_______________________________________________

_______________________________________________

_______________________________________________

Chair

University of California, San Diego

2012
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures and Tables</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>vi</td>
</tr>
<tr>
<td>Abstract of the Dissertation</td>
<td>ix</td>
</tr>
<tr>
<td>Introduction to the Dissertation</td>
<td>1</td>
</tr>
<tr>
<td>I. Cross-taxa identification of brain sub-structures by <em>in-situ</em> hybridization:</td>
<td></td>
</tr>
<tr>
<td>The Dentate Gyrus</td>
<td>6</td>
</tr>
<tr>
<td>II. Evolution of BMP regulation of <em>msx</em> genes</td>
<td>27</td>
</tr>
<tr>
<td>Final Discussion of the Dissertation</td>
<td>57</td>
</tr>
<tr>
<td>References</td>
<td>60</td>
</tr>
</tbody>
</table>
LIST OF FIGURES AND TABLES

Figure 1.1 - Schematic representation of the proposed subdivisions of the avian hippocampus in comparison with the mammalian hippocampus……9

Figure 1.2 – Examples of conserved genes with high and/or restricted expression in the mouse hippocampus………………………………………………….19

Figure 1.3 - RT-PCR analysis of genes predicted to be enriched in the avian hippocampus………………………………………………………………………..20

Figure 1.4 - Comparative expression of Prox1 in the mouse and chick hippocampus delineates the borders of the Dentate Gyrus in birds……………………………………………………………………………………………………22

Figure 1.5 - Markers that are expressed in mouse Dentate Gyrus, consistently show up in the same region in chicken…………………………………24

Figure 1.6 - Comparative expression of stem cell markers in the mouse and chick hippocampus........................................................................................................25

Table 2.1 – Primers used in this study............................................................................................39

Figure 2.1 - The msh CRM faithfully recapitulates msh response to BMPs……41

Figure 2.2 - Identification of BMP responsive sites on the msh CRM…………44

Figure 2.3 - Identification of ME activation sites…………………………………………47

Figure 2.4 - BMPs repress msh in Drosophila…………………………………………………48

Figure 2.5 – Identification of Vertebrate msx CRMs……………………………………………………51

Figure 2.6 – Differential response of Drosophila and vertebrate msh/msx genes to BMPs ..........................................................53
ACKNOWLEDGEMENTS

I would like to acknowledge, my mentor, Professor Ethan Bier for his generosity, guidance, inspiration and genuine kindness without which none of this work would have been possible. He will forever be my measure of what a truly passionate, rigorous, curious and creative scientific mind is. Also, as a role model for how a group leader, in a often times frustrating endeavor, can rekindle the drive to push forward, time after time, in the people under him. I thank him for this amazing opportunity and for teaching me how to be a scientist.

My deepest gratitude to my other mentor, Professor Harvey J. Karten, for embodying what it means to be an outstanding scholar in experimental sciences. Several of my most enjoyable science moments in graduate school happened sitting at a microscope with Harvey.

I would also like to acknowledge past and present members of the Bier Lab for their mentoring, support and camaraderie: Professor Mieko Mizutani for the patience and humor when teaching me the technical skills I came to rely on; Orna Cook, Annabel Guichard, Margie Smelkinson and Beatriz Cruz-Moreno for invaluable help in the lab time after time; fellow past and current graduate students in the lab Long Do and Valentino Gantz for companionship, advice, support and reality checks in science and in life, in sum, true friendship.

Chapter I, in part, is a reprint of the material as it appears in Long H Do, Francisco F Esteves, Harvey J Karten, Ethan Bier. “Booly: a new data integration platform”, BMC Bioinformatics, vol 11 pp 513, 2010. The dissertation author was an investigator and author on this paper.

Chapter II is part of collaborative efforts and it would not have been possible without the help of Erika Kague in Shanon Fisher’s lab from University of Pennsylvania
who helped me with the creation of zebrafish transgenic animals and Alexander “Helge” Springhorn in Georgios Pyrowolakis lab from University of Freiburg who helped me with the EMSAs and with helpful bioinformatic analysis and discussions.

I would have not been able to withstand the rigors of graduate school without my many friends in the program and outside of school here in San Diego as well as my friends and the WestSideKru in Portugal. Thank you guys.

I also am indebted to my teachers and fellow students at Aoinagi for being the support that made UCSD home.

And finally, a thank you to my parents for the unconditional Love and support always.
VITA

2005  Licenciatura, Biologia Microbiana e Genetica, Faculdade de Ciencias, Universidade de Lisboa

2012  Doctor of Philosophy, University of California, San Diego

FIELDS OF STUDY

Major Field: Biology

Studies in Genetics and Molecular Biology
Professor Ethan Bier

Studies in Neuroscience and Neuroanatomy
Professor Harvey J. Karten
Several features of the nervous system have been highly conserved during evolution. For example, gene regulatory networks controlling the formation of a centralized nervous system consisting of segregated motor and sensory neurons in the spinal cord, as well as genes defining recognizable cell types that play similar physiological roles in particular brain regions, are conserved. There is also compelling evidence that the telencephalon of both birds and reptiles has
large pallial areas similar to the mammalian brain but it is unclear whether cell
types within particular regions are also conserved. Finally, there are also
similarities in the neuronal circuits mediating sensory processing suggesting a
common evolutionary origin at the level of connectivity. In the first chapter of my
dissertation I explore the issue of conserved cell types by attempting to elucidate
whether the hippocampus and its principle subdivisions are conserved between
birds and mammals through the analysis of avian expression patterns of
conserved genes that are expressed in restricted domains of the mammalian
hippocampus. I provide evidence in this study suggesting that several of
hippocampal gene homologues are expressed in coinciding regions of the
chicken brain thought to correspond to the avian hippocampus based on
physiological criteria.

Strong evidence for evolutionary conservation in neuronal patterning
comes from the three main rows of neuroblasts being determined by the
expression of three conserved transcription factors (\textit{vnd/nkx2.2}, \textit{ind/gsh} and
\textit{msh/msx} in \textit{Drosophila}/\textit{vertebrates}). These neural identity genes are expressed
in the same relative order and orientation with respect to the dorsal-ventral axis
during early nervous system development. While these conserved suites of gene
expression strongly suggest a common ancestral origin for these patterning
systems, it is unclear whether the original regulatory mechanisms establishing
these patterns have been similarly conserved during evolution. The second
chapter addresses this question by comparing BMP-mediated regulation of cis-
regulatory modules controlling the expression of the \textit{Drosophila msh} CRM and
zebrafish and mouse msx CRMs in the early dorsal nerve chord. This comparison suggests that while overall gene expression patterns have been conserved across these species, the mechanisms employed to generate these patterns may have evolved differentially in two distinct lineages.
Introduction to the Dissertation

Studies of neuronal projection, activity pattern and gene expression all suggest that several features of the nervous system are indeed ancient. For instance, the gene circuitry that leads to the formation of a centralized nervous system consisting of segregated motor and sensory centers appears to be conserved broadly across bilaterians (animals with right-left symmetry at some point in development)\(^1\) and there is evidence that conserved recognizable neuronal cell types play a similar role in particular brain circuits such as the hypothalamus\(^2\).

An example illustrating this ancient conservation is the sequence of events that leads to the differentiation of neural tissue itself. In early bilaterian embryos the ectoderm, one of the three primary germinal tissues, is subdivided into neural and non-neural (epidermal) domains through the interplay between highly conserved secreted morphogens, known as Bone Morphogenetic Proteins (BMPs), and their equally conserved antagonists. Although the exact details of this interplay can vary\(^3\), evidence supports\(^4-6\) the view that these initial molecular events represent the ancestral ectodermal gene regulatory network primary chain of events for all bilaterian animals and perhaps all Eumetazoa\(^7\).

Since genes determining the positions of neuronal subsets have been highly conserved, an important question is whether these genes determine conserved cell identities. If such a link between genes and cell identity exists, it would enable the identification of functionally equivalent cell groups across
different organisms by looking for conserved cell fate specifying markers and their effectors. Such a link between positional and cell-type specification is not a forgone conclusion. For instance, to pattern the anterior-posterior (AP) axis of the animal body plan, a set of highly conserved genes, the homeotic (HOX) genes, with highly conserved relative gene expression patterns throughout bilaterian animals (reviewed in 8), give rise to cell lineages that have little in common (e.g., bone skeleton versus chitin exoskeleton in regard to limb formation in chordates and arthropods, respectively). Thus, HOX genes may be simply providing an abstract positional code without any conserved link to conserved cell type specification. In contrast, differentiation of the animal body plan along the dorso-ventral (DV) axis is a process that displays clear conservation in terms of cell types formed at different positions (e.g., within the spinal chord ventral cell types expressing Nkx2.2 homologues give rise to motor neurons while Msx1/2 expressing dorsal cells give rise to sensory neurons). The process of changing the “effector” genes a “master” gene regulates along evolutionary time can be referred to as “regulatory churning”. This model justifies the differential degree of downstream divergence between the two major genetic systems governing the A/P versus D/V axes of the animal body plan, by noting that the DV axis spans relatively few, well conserved, tissue types, while the AP axis covers several different subtypes of many different tissues (reviewed in 9).

Two questions arise from the above considerations. First, are conserved gene-regulatory systems more intrinsically tied to cell identities for defined tissue types, and in that case, can they be used to predict equivalent structures in
different animals? Second, are the molecular mechanisms that establish the patterns of gene expression conserved as well?

Regarding the first question, the highly concordant patterns of gene expression in the spinal chord suggest that there is indeed a conserved relationship between gene expression patterning and cell specialized neuronal cell types. Perhaps most notably, the suite of genes specifying several adjacent neurogenic domains of Platynereis dumerilii, an annelid, are orthologous and expressed with similar relative distribution as in vertebrate development. An important unresolved question, however, is whether a degree of conservation between gene expression and cell-type specification holds for the far more complex brain as well. There is some evidence that this is the case for cells within neurosecretory centers. For instance, in earthworms and fish vasotocinergic extraocular photoreceptors seem to be specified by the same genes, have the same relative position within the brain in regard to molecular landmarks and have at least one common transcription factor (rx3 for example) regulating neurotransmitter production. It may even be the case that there are corresponding brain structures between invertebrates and vertebrates. For example, there are similarities between the gene expression profiles of pallial neurons in vertebrates with mushroom body neurons in annelids. Also, the honey bee gene, mahya, is expressed predominantly in the centers of learning and memory of the insect brain (mushroom bodies) not unlike its mammalian counterparts which are expressed predominantly in the Dentate Gyrus and CA fields of the hippocampus.
In the first chapter of my thesis I explore the issue of conserved cell types by attempting to elucidate whether the hippocampus and its principle subdivisions are conserved between birds and mammals. These experiments are motivated in part by the apparent differences in forebrain organization between mammals (where high level information processing takes place in multilayered cortical structures) and birds (in which comparable neuronal circuits are organized in discrete nuclei). Harvey Karten, one of my mentors, has proposed that despite the superficial differences in the apparent organization of the forebrain, highly similar cell types forming nearly identical circuits similarly process visual and auditory information in birds and mammals suggesting that the so-called pallial structures of the avian forebrain correspond to particular regions of the neocortex in mammals. Although a wealth of existing evidence and the logic of evolutionary mechanisms already support the Karten hypothesis, the identification of conserved patterns of gene expression in corresponding cell types would add yet further support to this important idea. The choice of the hippocampus as a region to investigate was based on the importance of this region in forming new memories and on bioinformatic analysis carried out collaboratively between a fellow graduate student, Long Do, and myself.

To address the second question, of whether conserved regulatory relationships require conserved mechanisms, I took advantage of the well-characterized and conserved relationship between BMPs and genes involved in patterning the spinal chord. In this system there is already solid evidence linking
transcription factor expression domains and neuronal cell identity along the DV axis\textsuperscript{13}. I have focused on one element of the gene regulatory network acting to establish DV positions and cell identities, namely the regulation of Msx genes, which define the dorsal-most cell fates of the spinal chord including sensory neurons and neural crest derivatives. One reason for focusing on BMP-mediated regulation of Msx genes is that there is evidence that BMPs are clearly involved in patterning the dorsal regions of both the vertebrate and invertebrate CNS, yet it is not clear whether they act by similar or opposite mechanisms to do so. Thus, in vertebrates it has been proposed that BMPs may act positively in a dose dependent fashion to activate expression of target genes including Msx, while in Drosophila, genetic evidence suggests that graded BMP signaling acts differentially and represses expression of the orthologous neural patterning genes. Also, the cis-regulatory module (CRM) of the Drosophila msh homolog has been identified and in my collaborative studies with Shannon Fisher’s group we have identified both zebrafish and mouse CRMs that direct dorsal expression of Msx genes in transgenic zebrafish embryos, permitting a mechanistic comparison of the vertebrate and invertebrate regulatory elements.
I.

Cross-taxa identification of brain sub-structures by

\textit{in-situ} hybridization: The Dentate Gyrus
1.1 Introduction

Until the later half of the 20th century, comparative anatomists thought of the sauropsid (reptiles and birds) cerebrum as being fundamentally striatal in nature with limited pallial areas, i.e., these brains consisted of hypertrophied basal ganglia with small cortex-like areas. This view was established in the works of Edinger and Ariens-Kappers, among others, by gross morphological comparison and cytological observation. A more recent reevaluation of anatomical relationships and conserved neuronal circuitry has lead to a revised view in which the telencephalon of both birds and reptiles consists of large pallial areas similar in function to the mammalian neocortex\textsuperscript{14-16}. Furthermore, such pallial areas are present in these three classes of vertebrates as well as in more ancestral-like vertebrates such as fish\textsuperscript{17}. This new perspective is consistent with a growing appreciation of avian intelligence in which characteristics previously thought to be restricted to primates (relative size of the forebrain, complex social groups, long gestation periods, and high abstract cognitive ability) are now recognized to exist in certain groups of birds, particularly among the Corvid (crows and kin) and Psittacine (parrots) orders. Displays of behaviors such as spatial memory and food-caching, social learning and tool use are examples of well studied avian intelligence (review in\textsuperscript{18}), which in many cases has proven comparable to levels of performance observed in the great apes. Given the great evolutionary divergence between the avian and mammalian lineages (e.g. \( \approx 310 \) MYA), an important question is whether the high level cognitive abilities observed
in species within these vertebrate branches arose independently by convergent evolution or rather by retention of ancestral properties that generate common modes of information processing by homologous cell types forming conserved neuronal circuits.

One brain area in which to investigate the question of brain homology at the cellular level is the hippocampus. The hippocampus is considered to be the brain region responsible for memory formation and learning (review in 19). The existence or location of the avian equivalent of the mammalian hippocampus remains controversial. Based on morphological and developmental considerations, Ariens-Kappers20 and Kuhlenbeck21 proposed originally that the dorsomedial region of the avian telencephalon may be the avian hippocampus. Subsequent studies have solidified the homology of the region to that of the hippocampal formation in mammals using cytoarchitectural, behavioral, immunohistochemical, hodological, and electrophysiological evidence16, 22-25, however, there are few noticeable morphological landmarks. Moreover, the apparent uniformity of this area makes it difficult even to distinguish where the hippocampus ends and the adjacent parahippocampal area begins (these subdivisions were proposed by Karten and Hodos26). In contrast, the mammalian hippocampus proper is cytologically distinguishable from the surrounding tissue and its subdivisions are evident upon microscopic analysis.

Several studies have proposed differing subdivisions of the avian hippocampus and their possible relationships to those in the mammalian brain, but a consensus remains elusive. For the purpose of this dissertation, I will use
the regional terms adopted by Atoji and Wild for the subdivision of the avian hippocampus in order to analyze the most relevant and recent literature addressing this problem (Fig. 1.1).

Figure 1.1 - Schematic representation of the proposed subdivisions of the avian hippocampus in comparison with the mammalian hippocampus. In all figures, top is dorsal, bottom is ventral. A: Giemsa stain of a sagital section of a mouse brain. Hippocampus highlighted inside the red rectangle. B: Schematic of the subdivisions of the mouse hippocampus. C: Giemsa stain of a sagital section of a chicken brain. Hippocampal areas are highlighted inside the red rectangle. D: Subdivisions and comparison of the avian hippocampus to mammalian hippocampus as proposed by Szekely and Krebs. E: Subdivisions and comparison as proposed by Atoji and Wild. DL- Dorso lateral; DM– Dorso Medial; PoMaPa- Cell poor area, Magnus Cellular area and Parvo cellular area.
Erichson and Krebs published two companion papers in 1991 in which they attempted to characterize the subdivisions of the pigeon hippocampus by means of immunohistochemistry. They elected as markers several neuropeptides and neurotransmitter-related enzymes. However, there are several concerns with this choice of markers. First, the distribution of neuropeptides varies among species of mammals and possibly birds and these tend to evolve rapidly in primary sequence. Also, it is unclear if the antibodies used were raised against the avian proteins or mammalian counterparts. Nevertheless, their data suggest that: the V-layer (V) corresponds to the Ammons’ Horn; the dorso-medial (DM) region to the hillar region and granule layers of the Dentate Gyrus; and the cell poor, magnocellular and parvocellular areas (PoMaPa) correspond to the alvear layer and fimbria-fornix of the mammalian hippocampus (area 1 in Erichson et al., 1991) as well as to granule cells (area 5 in Erichson et al., 1991) and the dorso-lateral (DL) area to the subiculum and entorhinal cortex.

In 1996 Szekely and Krebs published a study on the efferent connectivity of the zebra finch hippocampus using an anterograde tracing method in which injections were made in different regions of the hippocampus and telencephalon. Three injection sites were tested: DL, near the parahippocampal sulcus, DM around PoMaPa, and the ventricular arm of the V. The authors concluded that: the lateral portion of the DM and DL correspond to the subicullum (the main output area in mammals) due to predominance of projections outside the hippocampus (to the Arcopallium, basal ganglia and diencephalon); the PoMaPa areas of the DM correspond to the Dentate Gyrus because of the intrinsic nature
of the connections, which is typical of the mammalian Dentate Gyrus (although extensive projections to the septum can be seen in their figure 7a); and finally the V corresponds to the CA fields as it projects contralaterally (but also striking are the projections to the ipsilateral DM/DL and the exclusion of the PoMaPa areas as can be seen in their figure 9a).

An electrophysiological profile of the pigeon hippocampus was published in 2002 by Siegel et al. Using single unit recordings, they sampled electrical activity throughout the hippocampus of moving pigeons, both at rostral and caudal locations. They grouped their measurements in two types: faster-firing units with shorter spike widths and slower-firing units with longer spike widths. At the level of the V, as well as in the DM/DL region, there seems to be a predominance of slower-firing units. A significant difference was found at the DM/DL level, rostrally, over half of the units displayed a slow-firing pattern, while caudally 75% of units were of the slower-firing type. The location of the rapid-firing units seems to be widespread throughout the hippocampus. These results led the authors to propose that the V region corresponds to the mammalian CA fields because the activity profile showed some rhythmic bursting activity similar to the theta rhythm of pyramidal cells of CA3 and CA1. Regarding the dorsal regions of the hippocampus, the authors propose that there is a difference between the rostral and the caudal portions. The caudal portion having little or no activity between occasional bursts would correspond to the mammalian Dentate Gyrus, while the rostral portion would correspond to the subiculum as similar activity patterns unique to this area were recorded in birds and mammals.
Kahn et al. generated additional connectivity data of the pigeon hippocampus in 2003 where both retrograde and anterograde tracing methods were employed to observe the internal connectivity. They concluded that the DL corresponds to the entorhinal cortex, the dorsal part of the DM to the Dentate Gyrus and the V to the Ammons’ Horn. However, it is striking that in this study the injection sites often included one or more of the subregions that the authors were trying to characterize. Five out of seven injection sites overlapped different areas. From the remaining two injection sites it could be concluded that: there are afferents connections to the medial arm of the V from the lateral arm and from both the DL and DM (their figure 6); and the dorsal part of the DM close to the parahippocampal sulcus projects to both the V and outside the hippocampus (their figure 7). In summary, these connectivity data suggest that the DM is the major source of efferent pathways from the avian hippocampus.

Finally, Atoji and Wild in 2002 and 2004, carried out tracing experiments corroborated by kainic acid lesions. They complemented previous intrahippocampal connectivity information by injecting the tracers into the parahippocampal area and several nuclei of the septum in an attempt to characterize the subdivisions of the avian hippocampus using an external reference (the septum). In mammals, the septum receives projections from the CA1-CA3 and subiculum and projects back to the hippocampus. In previous studies, hippocampal-septal connections had already been shown to exist in the avian brain. From their results, they propose that: the DM region corresponds to a mosaic of subiculum and CA since this area projects outside the
hippocampus and reacts to kainic acid lesions the same way the mammalian CA does (i.e., atrophy of pyramidal neurons and gliosis in the CA and subiculum at low doses, but not in the Dentate Gyrus); the DL corresponds to the entorhinal cortex since it is the site of entry of sensory input; and the equivalent of the Dentate Gyrus would be the V because of the predominance of intrinsic connections (see figures 7-12 of Atoji et al, 2002) in spite of the existence of reciprocal connections that have not been reported in mammals. The proposal of the V as being the Dentate Gyrus is in direct contradiction with the literature previously mentioned.

One approach to resolving the above debate is to identify patterns of conserved expression since such gene expression signatures clearly delineate conserved groups of cell types within the spinal chord\textsuperscript{30} and other areas of the brain\textsuperscript{2}. Prior gene expression studies of the avian brain have largely focused on development\textsuperscript{31} and behavior dependent gene expression\textsuperscript{32}. In contrast, few efforts have been made to characterize homologous anatomical regions in the adult brain through these means.

The fish hippocampus is currently thought to comprise the posterior zone of the dorsal telencephalic area and the ventral part of the lateral zone of the telencephalic area\textsuperscript{17}. This notion is derived from neuroanatomical, developmental and behavioral performance in fish with lesioned brains\textsuperscript{33}. Another argument is that this putative hippocampus is a region of cell proliferation\textsuperscript{34}, however, in zebrafish cell proliferation in the adult brain is so widespread that this argument
carries little significance. To our knowledge, no attempts have been made to
draw subdivisions in the hippocampus of fish.

With these considerations in mind, I set out to clarify the level of
conservation in the vertebrate brain by identifying potentially homologous brain
structures in the brains of mammals (mouse), birds (chicken) and fish (zebrafish).
My main objective was to identify corresponding subregions (i.e., Dentate Gyrus
and CA fields) of the mammalian hippocampus in birds by comparing the
expression patterns of several highly conserved genes in the brain of Gallus
gallus using in situ hybridization. My preliminary findings suggest that there is
indeed a region of the avian brain expressing orthologues of several mammalian
hippocampus-specific genes. Interestingly, this region is more restricted than
proposed by any of the previous studies.
1.2 Materials and Methods

Selection of gene candidates for identifying brain structures

I took advantage of a tool developed by Long H. Do, the Booleome Database (https://booleome.ucsd.edu) and generated a list of 140 genes that are highly conserved and expressed at high levels in the mammalian hippocampus or have restricted expression in the hippocampus. From these lists I selected 13 initial candidates to optimize in situ hybridization protocols. I prioritized for transcription factors and genes that have very distinct hippocampal patterns.

Probe synthesis

Probes were synthesized from EST clones (acquired from ArkGenomics) for the desired genes. RNA probes were synthesized as in [35] with a 20 minute fragmentation time to produce an average probe fragment size of around 150-200 bases. However, since the signal seemed unspecific and/or the background high, I removed the fragmentation step to ameliorate this issue.

In-situ hybridization

I carried out in situ hybridization on 30 μm thick 2-day old chicken brain slices. The in-situ hybridization protocol used was adapted from the methods described in Kosman et al.[35]. Stringency of hybridization for a given probe (with a specific melting temperature, Tm) depends mainly on the temperature at which the hybridization reaction is carried. Stringency of nucleic acid hybridization can also be achieved by changing the formamide content, a denaturation agent, of
the hybridization solution instead of changing the temperature or concomitant with the temperature change. I raised the temperature from the original 55°C to 65°C for probes of varying lengths of the same gene and the results indicate a reduction of background when longer probes are used. The results presented are all from hybridizations carried out at 65°C.

I optimized both colorimetric and fluorescent detection methods. Three signal development schemes have been used. Alkaline Phosphatase conjugated antibodies, Peroxidase conjugated antibodies and fluorescent dye conjugated antibodies. In the first scheme, the phosphatase-conjugated antibody binds directly to the hapten labeled probe and then the signal is revealed by tissue exposure to a chromogenic solution (5-bromo-4-chloro-3-indolyl-phosphate/nitro blue tetrazolium). The second detection scheme, peroxidase, offers increased amplification since the enzyme that catalyzes the chromogen, 3,3'-diaminobenzidine, is attached to streptavidin molecules. Streptavidin, in turn, binds to a biotinilated secondary antibody, which recognizes a primary antibody bound to the haptens in the probe. The third detection method involves fluorescently labeled antibodies.

I used as positive controls brain markers such as calbindin (Calb-1) and tyrosine hydroxilase (Th), and the sense probes as negative controls. lacZ probes served as a general purpose negative control.
Data collection

For histochemical stains, high-resolution pictures were taken using an Aperio ScanScope system and fluorescent stains were collected using a LEICA SP2 confocal microscope.
1.3 Results

Selection of candidate genes to define homologous sub-regions of the avian hippocampus

To select the candidate genes that might serve as a conserved signature for the different parts of the hippocampus across taxa, I collaborated with a fellow graduate student, Long Do. Using Booly, a powerful tool for integration of different types of biological data, we narrowed down genes that might be enriched in the avian hippocampus by mining the Allen Institute Brain Atlas for genes that are expressed selectively in the mouse hippocampus or at much higher levels compared to other brain regions (Fig. 1.2). Next, we integrated this list with mouse BLAST hits against chicken, fish and fruit-fly, as well as mouse Gene Ontology annotations. A first order prediction of the Karten hypothesis is that conserved genes that are restricted and highly expressed in the hippocampus of mice, should also be enriched in the regions of the avian brain that have been functionally characterized as being the hippocampus. To test this prediction we performed RT-qPCR on different areas of the avian brain of a 1-day old chick (Fig. 1.3). Our results show that approximately 50% of genes analyzed were enriched in the chick hippocampal regions. These results are described and discussed in detail in Do et al, 2010.
Figure 1.2 – Examples of conserved genes with high and/or restricted expression in the mouse hippocampus.
Figure 1.3 - RT-qPCR analysis of genes predicted to be enriched in the avian hippocampus. Relative fold change of selected genes in the Hippocampus and other areas of the Chick Brain, based on two or more individual independent experiments, which were highly concordant. Housekeeping genes GAPDH and actin were used as controls for normalization. From Do et al, 2010
Prox1 expression defines a restricted region of the avian brain that may correspond to the Dentate Gyrus.

The homeobox transcription factor Prox1 has restricted expression in the mouse brain and is enriched in a particular subdivision of the hippocampus. In the adult mouse brain, Prox1 seems to be almost exclusively expressed in the Dentate Gyrus and in the thalamus. I found prox1 to be equally restricted in the avian brain. Two regions are highly enriched in the 2-day old chick: the thalamus and the v-region of the hippocampus. Moreover, the prox1 signal in the avian telencephalon did not extend all the way rostrally, instead being located more caudally while spanning most of the medial-temporal axis (Fig. 1.4). This expression pattern favors a model of the avian hippocampus where a portion of the v-region corresponds to the Dentate Gyrus.
Figure 1.4 - Comparative expression of Prox1 in the mouse and chick hippocampus delineates the borders of the Dentate Gyrus in birds. A-D: Mouse brain prox1 in situ hybridization. Data from the Allen Brain Atlas. E-H: Chicken brain prox1 in situ hybridization.
The v-region of the Avian Hippocampus has Dentate Gyrus-like gene expression patterns.

Additional genes expressed in the avian hippocampus resemble the mouse expression profiles. For instance, netrin-G2 and synaptotagmin 7 (Fig. 1.2) are all expressed in the Dentate Gyrus and throughout the mammalian hippocampus. The chick homologues are also detected in several regions of the avian hippocampus and have strong expression in the v-region (Fig. 1.5B and E). On the other hand, slit2 in mammals is expressed selectively in the CA3 but not in the Dentate Gyrus. Similarly in the chicken, slit2 is not expressed in the v-region but is in the DL regions of the hippocampus (Fig. 1.5F), further supporting the hypothesis that a portion of the v-region corresponds to Dentate Gyrus and that CA-like regions may exist as separate adjacent groups of cells.
Figure 1.5 - Markers that are expressed in mouse Dentate Gyrus, consistently show up in the same region in chicken. A,D: netrin G2 in situ hybridization. B,E: synaptotagmin 7 in situ hybridization. C,F: slit homologue 2 in situ hybridization.
Stem-cell markers are widely expressed in the early adult avian brain

In mammals, the subgranular zone (SGZ) of the Dentate Gyrus is one of the few regions (the other being the subventricular zone), where neurogenesis occurs in adult brains. Accordingly, stem-cell markers such as *sox2*, *musashi1* and *prominin1* are detected in the zone in the SGZ in mouse. However, in the early avian brain, the chick homologues have widespread expression and cannot be reliably used to identify discrete homologous structures (Fig. 1.6).

Figure 1.6 - Comparative expression of stem cell markers in the mouse and chick hippocampus.
1.4 Conclusions

My results suggest that a portion of v-region in the avian hippocampus is homologous to the mammalian Dentate Gyrus given the similar gene expression patterns of the majority of genes tested (9 out of 15 genes). To confirm this hypothesis, it would be informative to perform a developmental series of prox1 expression in the chick embryo, thus elucidating the embryonic origins of this area. It would then be possible to compare such results with equivalent available data in the mouse, thus establishing whether these adult homologous regions indeed arise from embryonic homologous structures. With this more refined definition of the location of the avian Dentate Gyrus, it would be interesting to repeat connectivity studies with more precise injection sites using a bidirectional tracer such as biotinylated dextran amines. Regarding neurogenesis in the mammalian Dentate Gyrus versus the avian Dentate Gyrus, a Bromodeoxiuridine (BrdU, a marker for DNA synthesis) study could help clarify if neurogenesis is restricted, like in the adult mammalian brain, or instead more widespread in the early adult avian brain. It would also be interesting to examine the expression of a series of chick homologs of mammalian genes expressed selectively in the three primary subdivisions of the CA region of the hippocampus and follow up with tracing experiments to see whether the same hodological relations exist in the avian brain.
Chapter I, in part, is a reprint of the material as it appears in Long H Do, Francisco F Esteves, Harvey J Karten, Ethan Bier. “Booly: a new data integration platform”, BMC Bioinformatics, vol 11 pp 513, 2010. The dissertation author was an investigator and author on this paper.
II.

Evolution of BMP regulation of *msx* genes
2.1 Introduction

The trunk CNS of both invertebrates and vertebrates consists of three primary rows of neuroblasts that are determined by the expression of three conserved transcription factors. In Drosophila and vertebrates these genes are: vnd/nkx2.2, ind/gsh and msh/msx, and they are expressed in the same relative order and orientation with respect to the dorsal-ventral axis during the early nervous system development of all bilaterally symmetrical animals investigated thus far\textsuperscript{37, 38}. This conserved pattern of neural identity genes is reminiscent of that of HOX genes along the anterior-posterior (AP) axis. While these conserved suites of gene expression strongly suggest a common ancestral origin for the axial patterning systems, it is unclear whether the original regulatory mechanisms establishing these patterns have been similarly conserved during evolution. Indeed, with regard to HOX genes, there is little evidence for conservation of cis-regulation of these genes other than apparently similar cross-regulatory interactions (e.g., posterior dominance). The situation along the DV axis is also unclear. In Drosophila the maternal signal Dorsal is necessary for establishing DV patterning and specification of ventral versus medial regions of the neuroectoderm, but in vertebrates WNT signaling seems to perform the equivalent function. One common patterning system common to flies and vertebrates is that mediated by Bone Morphogenetic Proteins (BMPs). In both classes of organisms, BMP4-related proteins are expressed in the epidermal ectoderm abutting the dorsal border of the neuroectoderm. BMPs are thought to
exert a common function in the early epidermal ectoderm (i.e., suppressing expression of neural genes) and they act subsequently in a dose dependent fashion to pattern dorsal versus medial regions of the neuroectoderm. It is not known, however, whether BMPs act similarly in flies and vertebrates to accomplish this patterning. Indeed, according to current paradigms, BMPs may act by opposite means in Drosophila (graded repression of target genes) and vertebrates (graded activation of target genes).

Establishing the Drosophila embryonic DV axis

In Drosophila the maternal gradient of Dorsal, a protein related to the Rel/Nf-κB family, is responsible for the differentiation of three primary embryonic tissues (reviewed in 39): mesoderm, neurogenic ectoderm, and dorsal ectoderm.

This subdivision into primary tissue types is established by the maternally deposited transcription factor Dorsal. Dorsal is distributed in a nuclear concentration gradient as a consequence of the action of a maternal cascade of serine proteases. DV target genes respond to the Dorsal gradient via Dorsal binding sites present in their Cis-Regulatory Modules (CRMs). Dorsal binding sites are not intrinsically activators or repressors; instead they act either way depending on the molecular context. Nonetheless, one can distinguish two functional types of Dorsal binding sites, high-affinity and low-affinity binding sites. The presence and number, as well as the combinations between the two kinds of sites, allows the establishment of at least four thresholds of gene activity along the Drosophila embryonic DV axis 40-42. The first threshold (the highest Dorsal
concentration in the ventral most region), establishes the differentiation of the mesoderm by inducing the expression of *twist* (*twi*) and *snail* (*sna*); the second threshold activates genes such as *vnd* and *rho* in the ventral third of the neurogenic ectoderm; the third threshold activates *ind* in medial regions of the neuroectoderm; and the fourth threshold delimits the dorsal border of the neuroectoderm throughout which *short-gastrulation* (*sog*) (a *Chordin* homologue) is activated by low amounts of Dorsal. Responding to this same threshold is *decapentaplegic* (*dpp*, an orthologue of BMP2/4), which is repressed within the neuroectoderm and therefore expressed only in the dorsal epidermal ectoderm where Dorsal is virtually absent.

Dorsal epidermal cell fates are determined by expression of genes including *dpp*, *tolloid* (*tld*) and *twisted gastrulation* (*tsg*). These genes subdivide the dorsal region of the embryo into two principle domains: most dorsally, an extraembryonic tissue known as amnioserosa and more ventrally, the epidermal ectoderm proper. This subdivision is generated by the graded activation of the Dpp pathway, with higher levels giving rise to the amnioserosa dorsally. The formation of the Dpp activity gradient requires four other extracellular proteins, Sog, Tsg, Tld and Scw (encoded by *screw*, another BMP-related protein) as well as the serine-threonine kinase BMP-receptors Saxophone (Sax), Punt (Put) and Thick Veins (Tkv).

Dpp is produced in dorsal cells but, being a secreted signal, it can diffuse and generate cellular responses at long range. High levels of Dpp signaling are restricted to dorsal regions of the embryo due to the action of Sog. Sog binds to
BMP heterodimers preventing BMPs from docking to its receptors. Since Sog is distributed in a graded fashion as a consequence of its diffusion from the lateral neuroectoderm to the dorsal ectoderm, it generates a ventral to dorsal inhibitory gradient. Moreover, another role for Sog as a BMP carrier has been proposed in which Dpp bound Sog may “shuttle” BMPs by carrying them dorsally and thus creating a maximum activity peak of Dpp along the dorsal midline\textsuperscript{48}. Tsg is an essential cofactor for Sog-Dpp binding in \textit{Drosophila} \textsuperscript{49, 50} and Tolloid is a metalloprotease that cleaves Sog and renders Dpp free to engage its receptors. Scw is uniformly expressed in the early embryo\textsuperscript{51} and dimerises with Dpp. The Dpp-Scw dimer binds to heteromeric serine-threonine kinase receptor complexes in the following fashion: Dpp binds to Tkv and Put, while Scw binds Sax and Put. Ligand binding also triggers receptor-mediated endocytosis that leads to degradation of Dpp. Like in other morphogen systems, ligand production and degradation rates play a role in gradient formation and this case is no exception \textsuperscript{52, 53}.

**BMP responsive elements**

When the BMP receptors are activated by a ligand, the cytoplasmic receptor-Smad protein Mothers Against Dpp (Mad) is phosphorylated (pMad) and translocates into the nucleus bound with Medea (Med) (the Smad4 homologue) where these proteins act as transcription factors to mediate BMP regulation of target genes. Mad and Medea bind DNA as a heteromeric complex and regulate genes through interactions with binding sites composed by a Mad (GC-rich) site
separated, by a variable length spacer, from a Med (Smad Binding Element or SBE) site. One of the best characterized such sites is the brinker (brk) Silencer Element (SE) which has a spacer length of 5 nucleotides. Brk encodes a transcriptional repressor protein and reciprocally, the brk gene is repressed by Dpp signaling. Brk in turn acts as a repressor of other BMP regulated target genes, thereby playing a central role in establishing zones of high BMP signaling (in cells with no Brk) and low BMP signaling (in Brk expressing cells).

Brk exerts its repression of BMP target genes by competitively binding DNA containing Mad motifs. When bound, Brk recruits co-repressors, Groucho and/or CtBP, to inhibit different Dpp targets. Repression of brk through its SE requires the presence of the zinc-finger protein Schnurri (shn), which is provided both maternally and is expressed zygotically in dorsal epidermal regions of the early embryo. Hence, genes that are repressed by BMPs are expected to have a Mad/Shn/Med complex bound to their CRM. In contrast, genes that are directly activated by BMPs, such as the inhibitory Smad daughters-against-dpp (dad), contain activating elements (AE) in their CRMs. These elements also share the bipartite configuration (GC-rich/spacer/SBE) but have configurations (spacing and sequence constraints) that do not allow for Shn binding and lead instead to the recruitment of other transcriptional co-factors.

**Neuroectodermal patterns of gene expression along the Drosophila D/V axis**

Development of the Drosophila central nervous system (CNS) from a ventral bilaterally symmetric neuroectoderm begins with the delamination of
neural and glial precursors, called neuroblasts, and develops from three columns of neuroectodermal cells alongside the dorsalventral axis \(^{60, 61}\). The expression domains of three homeobox genes, *muscle segment homeobox (msh)*, *intermediate neuroblasts defective (ind)*, and *ventral nervous system defective (vnd)*, define these columns. These genes are referred to as neural identity genes since mutants lacking the function of these genes display defects in cell fate specification and/or survival of cells deriving from their respective expression domains\(^{13, 60, 62}\).

Neural identity genes are expressed sequentially prior to gastrulation, where *vnd* is the most ventral one and the first to appear. A particular threshold of Dorsal concentration leads to the expression of *vnd* \(^{63}\) in the ventral neurogenic ectoderm. The ventral border of *vnd* expression is achieved by repression from *twi* along with the Dorsal activation. This regulation is enacted by clustered *twi* and Dorsal sites in an intronic CRM\(^{39}\). In the intermediate regions of the neuroectoderm, lower Dorsal levels are no longer sufficient to activate *vnd*, but can still activate *ind* expression. In addition, maintenance of *ind* expression requires EGFR signaling. EGF ligands (Vein and Spitz) are produced in ventral *vnd* expressing cells and diffuse dorsally to activate *ind*. The ventral border of *ind* expression is defined via repression by *vnd* (Sna also represses all neural genes in the mesoderm). This regulatory network, suggested by genetic experiments\(^{64}\), was confirmed by analysis of the *ind* CRM \(^{42, 47}\). The *ind* CRM also contains Smad/Smad sites which when mutated, lead to dorsal expansion of reporter gene constructs. Another important determinant of *ind* expression is the
Capicua \((cic)\) repressor, which is down-regulated by EGFR signaling and together with BMPs contributes to determining the position of the dorsal border of \textit{ind} expression\(^65\).

The regulation of the dorsal most neural identity gene \textit{msh} is not completely resolved yet. Genetic data indicates that it is independent of a Dorsal for activation and that it is repressed directly by \textit{snail}, \textit{vnd}, \textit{ind} in ventral regions and by BMPs\(^64, 66, 67\) in dorsal regions. However, the mechanism by which BMPs act on the \textit{msh} CRM remains unknown.

**Regulation of \textit{msh/msx} genes by BMPs**

In flies, BMPs act to repress expression of neural genes during both neural induction and subsequent neuroectodermal patterning. In contrast, in vertebrates, the prevailing view is that while BMPs act as they do in flies to repress expression of neural genes early during neural induction within epidermal regions\(^37\), they then switch function and activate expression of orthologous neural identity genes in dorsal and lateral regions of the neural tube (e.g., the \textit{msh} orthologs \textit{Msx1/2}). Thus, in mice, \textit{msx} genes seem to be activated by BMP signaling\(^68\), since ectopic BMP signaling leads to \textit{msx} ventral expansion on the neural tube. In contrast, in \textit{Drosophila}, the absence of BMPs leads to \textit{msh} expanding dorsally into non-neural domains. However, in zebrafish, there is evidence that BMPs act in a bimodal fashion with intermediate levels of BMP being necessary for activating \textit{Msx} genes and both low and high levels of BMPs repressing or failing to activate these target genes\(^69\). Similarly, in amphioxus, a
basal chordate, msx expression again is spatially tied with BMPs and seems simultaneously induced and repressed when subjected to ectopic BMP signaling. After the evolutionary split between protostomes and deuterostomes, the most basal deuterostome phylum is the Echinoderms. In these animals, msx is not detected close to the source of BMPs, differing from chordates and arthropods, but seems to be activated by high levels of BMP signaling in non-neural regions. To shed light on possible evolutionary scenarios for BMP regulation of Msx genes in different lineages, (e.g., repression versus activation of Msx) I performed a comparative analysis of vertebrate (zebrafish) and invertebrate (Drosophila) CRMs driving expression in the dorsal CNS.
2.2 Materials and Methods

Bioinformatics

To look for the presence of SE and AE sites I curated neuroectoderm relevant binding site consensus sequences from the literature referenced and used Gene Palette\textsuperscript{71} to visualize putative TFBS distribution on the \textit{msh} CRM. Genome wide information for all the species and multiple species conservation information was accessed through the UCSC genome browser (http://genome.ucsc.edu).

Drosophila Reporter Gene analysis

The \textit{msh} CRM was first described in Von Ohlen et al., 2009\textsuperscript{72}. All primers used in this study and the corresponding constructs generated can be found in Table 1. All constructs were subcloned in pCR-TOPO vectors and subsequently cloned into the [P]acman vector\textsuperscript{73} as NotI and KpnI restriction fragments. Site-directed mutagenesis PCR methods were adapted from\textsuperscript{74}.

The fly strain used to inject the all constructs has genotype PBac{yellow[+]-attP-3B}VK00002 and injections were outsourced to BestGeneInc (http://www.thebestgene.com/).

Zebrafish Reporter gene analysis

The primers used to isolate the zebrafish msxB CRM and the mouse gene msx1 CRM can be found on table 1. Constructs were cloned into pENTR-TOPO
and transferred to pTol2 by Gateway Recombination and injected in Zebrafish embryos as described in\(^\text{75}\). The Zebrafish genomic DNA was a gift from Albert Kim in the Traver Lab and the mouse genomic DNA was a gift from Robert Teachenor in the Murre Lab.

**Fly strains**

The *dpp* mutant stock is Flybase stock number 2061. The 8x HS-dpp stock and its use are described in Biehs *et al* 1996\(^\text{76}\). Shnurri mutant allele is *shn*\(^{04738}\).

To generate the *dl dpp st2-dpp* embryos, females that are *Dpdpp/+; dl\(^1\) cn\(^1\) sca\(^1\)/ dpp\(^{h46}\) wg\(^{sp}\) dl\(^1\) were crossed to *yw/Y; dpp\(^{h46}\) wg\(^{sp}\) st2-dpp\(^+\),w+/CyO* males.

**Zebrafish strains**

The zebrafish lines containing the HS-BMP and HS-CHD were crossed to stable transgenic lines containing the msxB-CRM construct. Erika Kague performed all fish husbandry, injections and heat shock treatments at the Fisher lab.

**In situ hybridization**

The *in-situ* hybridization protocol used was adapted from the methods described in Kosman *et al*, 2006\(^\text{35}\). *In situ* hybridization in fly embryos was carried
out at 55C while in zebrafish embryos was carried out at 65C. DNA template to generate the msxB probe was a generous gift from the Riley lab.

Data collection

For histochemical stains, images were acquired using a Zeiss optical microscope, fluorescent stain images were collected using a LEICA SP2 confocal microscope.
<table>
<thead>
<tr>
<th>Construct Name</th>
<th>DNA Templates</th>
<th>Primer Pair (5’-&gt;3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>msh-CRM</td>
<td>BDGP w/ genomic DNA</td>
<td>msh1- GCTTTTGGGCTAAGCTCACAGCAGG msh2- GCAGCTGAAGCTGGAAACTGAA</td>
</tr>
<tr>
<td>D1</td>
<td>BDGP w/ genomic DNA</td>
<td>D1F- GCTTTTGGGCTAAGCTCACAGCAGG msh2</td>
</tr>
<tr>
<td>SE1</td>
<td>pTopo-mshCRM</td>
<td>SE1f- TGGTCTGGAAATTCGCCCTTAACCATTGAGCCACTCAGAGGTAGTTTTC TGGCATCT SE1R- AGAATGCCAGAAAACCTACTCTGAGTGCTGAATGGTTAAGGGCGAAT TCCGACACA</td>
</tr>
<tr>
<td>SE2</td>
<td>pTopo-mshCRM</td>
<td>SE2f- CACCCACAGGGTGTAACATTCCGAGAGTTGGTGTTTTTGGG SE2r- CCCAAAAGCAAACCTCTCGGTAGTTCCGAATATGTTAACACCCTGGGTG</td>
</tr>
<tr>
<td>SE1SE2</td>
<td>pTopo-SE1</td>
<td>SE1f SE1r</td>
</tr>
<tr>
<td>SE2*</td>
<td>pTopo-mshCRM</td>
<td>SE2<em>f- CCAGGGTGTGGCGCCTCAGAGGTCACTCAGAGGGTGTTTTTGGG SE2</em>r- CCCAAAAGCAAACCTCTCGGTAGTTCCGAATATGTTAACACCCTGGGTG</td>
</tr>
<tr>
<td>D2</td>
<td>pTopo-mshCRM</td>
<td>D2f- CTAAGCTCACAGCAGGATTGCACATGTGTGCTCTCAGGG D2R- GGACACATAGTGCAATCCTGAGCTGTAGTCC</td>
</tr>
<tr>
<td>D3</td>
<td>pTopo-mshCRM</td>
<td>D3f- GCACATATGTGTCCCTCAGGGCAGGATTCTGAGTGTTG</td>
</tr>
<tr>
<td>D4</td>
<td>BDGP w/ genomic DNA</td>
<td>msh1 D4r- ACAAGACACTCGAATCTGCC</td>
</tr>
<tr>
<td>AE2</td>
<td>pTopo-mshCRM</td>
<td>AE2f- GGACACCCCTACCTGCGATGAGTAATACTCGAATACTGAGGTGGGTGATG CC AE2r- GCCACACCCACTCGAATCTCGAATACTCGAATACTGAGGTGGGTGATG</td>
</tr>
<tr>
<td>SE1SE2AE</td>
<td>pTopo-SE1SE2</td>
<td>AE2f AE2r</td>
</tr>
<tr>
<td>mxB-CRM</td>
<td>Zebrafish genomic DNA</td>
<td>mBr- CACACAATTCTTAACACACTCCGTCC</td>
</tr>
<tr>
<td>mxe1-CRM</td>
<td>Black 6 mice genomic DNA</td>
<td>mxe1- AACTTGTTAATAAGGCAAGGCCAGCTGGGAGCCAGGGCAGG</td>
</tr>
<tr>
<td>mxe2-CRM</td>
<td>Black 6 mice genomic DNA</td>
<td>mxe2- GGTGGTCTCGGGAATGAGAAAAAGGCAGGGCAGG</td>
</tr>
</tbody>
</table>
2.3 Results

The *msh* CRM responds to BMPs

The *msh* CRM (henceforth referred to as ME) was first identified as a 700 base pair element by Von Ohlen *et al.*, 2009\(^7\). In that study, the authors showed that Ind could inhibit reporter gene constructs by binding directly to Ind sites in ME, but they did not examine BMP-mediated regulation of this CRM. As is the case for the endogenous *msh* gene (Fig. 2.1A), the ME-*lacZ* expression expands throughout the dorsal region of the embryo in *dpp* mutants (Fig. 2.1B). In order to determine whether Dpp regulates *msh* directly or indirectly, I undertook an analysis of BMP regulation of the ME element. I began by examining genome wide DNA binding data for several transcription factors mediating the BMP response in blastoderm stage embryos (available on the UCSC genome browser - http://genome.ucsc.edu/). These chromatin immune precipitation (ChIP) data indicate clear binding of Mad, Medea and Shn within the ME region (Fig. 2.1D). Consistent with evidence for Shn binding to sequences within the ME, homozygous zygotic *shn* mutant embryos exhibited a partial dorsal expansion of *msh* (Fig. 2.1C).
The msh CRM contains BMP-responsive sites

Encouraged by the genome-wide ChIP data, I employed two approaches to identify BMP responsive sites with the ME. First, I scanned the ME for known consensus binding sites for Mad, Med, Shn, and Brk. The two best characterized BMP responsive elements are known as the Silencer Element (SE), which binds a trimeric complex comprised of Mad/Med/Shn (GNCGNC(N)₅GNCTG), and the activator element (AE), which binds Mad/Med hetromers (GGCGCCA(N)₄GNCV). Brk binding sites (TGGCGYY) overlap with a subset of SE elements⁵⁵. Although there were no perfect consensus SE, AE, or Brk sites within the ME, I identified
several candidate sites with either single base-pair mismatches to the SE or AE elements or variable spacer \((N)_{5-6}\) length. Based on this initial analysis, in collaboration with Helge (Alexander) Springhorn in the Pyrowolakis lab, we defined two candidate SE sites (SE1 and SE2) and eight candidate AE sites to test for direct DNA binding of Mad/Med or Mad,Shn,Med complexes in vitro using the Electrophoretic Mobility Shift Assay (EMSA).

The SE1 and SE2 candidate silencer sites (Fig 2.2A) both conform to the modified consensus of GNYGNC(N)\(_5\)GNCTG (where Y can be either C or T). EMSA experiments using DNA probes corresponding to SE1 and SE2 oligonucleotide sequences revealed that Mad-Shn-Med and Mad/Med complexes could be assembled at both sites (Fig. 2.2B). Modifying the SE2 site by one base-pair (T to C) to adhere to the strict SE consensus, thus resembling a perfect SE site present within the Ind CRM\(^{42}\), exhibited yet greater binding to Mad/Shn/Mad complexes, consistent with such sites having the optimal spacing and thus higher binding affinity for the trimeric repressive complexes (Fig. 2.2B, SE2\(^*\)). Also, mutation of the Med (SBE) portion of the SE2 abolished binding of all BMP responsive complexes \textit{in vitro}. None of the candidate AE or Brk sites bound Mad/Med or Brk respectively (see below, however, regarding effects of mutating or deleting two of the best candidate AE sites).

I next analyzed the role of the two putative SEs \textit{in vivo}. I mutated the Mad and Med binding motifs for both of these sites (which abolish all BMP responsive DNA binding) in the context of the ME-lacZ reporter gene and inserted these mutant constructs into a single chromosomal integration site using the PhiC31
transgenesis system\textsuperscript{77}. I also generated a series of small deletions covering virtually the entire ME (i.e., all but 36 bp). Deletion of the 5’ 100bp of ME containing both SEs lead to a dorsal expansion of the reporter gene (Fig 2.2C), although transgene expression was also weaker within its normal neuroectodermal domain suggesting that contributing activation sites were also present within this region. Targeted mutation of the SE1 and SE2 sites individually, led to modest dorsal expansion of reporter gene expression in both cases. Deleting both sites simultaneously resulted in increased dorsal expansion which, however, was still less complete than that observed for the ME element in a \textit{dpp} mutant background. Consistent with the EMSA data indicating that mutating the SE2 site to conform to the strict SE consensus, incorporation of this mutation into the ME resulted in greatly reduced reporter gene expression \textit{in vivo}.
Figure 2.2 - Identification of BMP responsive sites on the msh CRM. A: Schematic representation of ME with the location of putative SE sites and 12 Insect species conservation. B: Sequence of probes used to detect mad-shn-med binding to regions of the ME in vitro. C: Transgenic embryos with mutated ME constructs stained for lacZ mRNA. SE1 and SE2 mediate partial dorsal repression.
The *msh* CRM contains activation sites that resemble BMP responsive sequences.

While no AE sites were identified using the strict AE consensus GGCGCCA(N)₄GNCV, and our collaborators could not detect binding to any of the eight candidate sites differing in single positions from the consensus or with varied spacer length, I nonetheless tested the in vivo effect of mutating one of two sites matching the modified GNCGNC(N)₆GNCV consensus since it localizes to highly conserved regions. Mutation of this (AE2) site resulted in greatly reduced transgene expression (Fig. 2.3B). Similarly, a 150 bp internal deletion of ME including the related AE1 site also severely reduced reporter gene expression. Given the requirement for the AE2 site for activation of the ME element, I next asked whether this function was also important for ectopic dorsal expression caused by loss of SE-mediated repression. As in the case of the AE2 single mutant, I observed reduced expression of reporter gene expression in the AE2, SE, SE2 triple mutant without any of the dorsal expansion caused by the SE1, SE2 double mutant.

These results prompted me to investigate the possibility of an activation role for BMPs in *msh* regulation. To test if there is a threshold at which Dpp enhances rather than suppress *msh* expression, I generated embryos that lack Dorsal and whose only source of Dpp is one copy of *dpp* driven by the *eve* stripe 2 CRM. In these embryos, *msh* expression was highly restricted and severely reduced compared to embryos without Dorsal and *dpp* null (compare Fig. 2.4A and D). The Dorsal and *dpp* mutant embryos express *msh* ubiquitously and
adding progressive amounts of Dpp (by varying copy number) led to a progressive decrease in msh expression, arguing against an activation role for Dpp in early embryonic msh regulation.
Figure 2.3 - Identification of ME activation sites. A: Wild-type ME expression and Schematic representation of ME depicting SEs, putative AEs, one Ind binding site and a Snail site. Cons - 12 Insect species Conservation. B-F: ME deletions and their respective expression pattern.
Figure 2.4 – BMPs repress msh in Drosophila. In all panels left is anterior. A: dpp<sup>−</sup> embryo without Dorsal showing ubiquitous expression of msh. B: msh in a dpp heterozygous embryo without Dorsal. C: msh in a dpp heterozygous embryo without Dorsal and with Dpp being expressed under the control of the eve stripe 2 CRM. D: dl<sup>−</sup> dpp<sup>+</sup> embryo with Dpp being expressed under the control of the eve stripe 2 CRM showing restricted expression of msh.
Identification of early embryonic vertebrate msx CRMs

The above analysis of the ME element in *Drosophila* supports prior genetic evidence that BMPs act directly to repress expression of *msh*. Since there is evidence for both positive and negative regulation by BMPs in vertebrates, I searched for the CRM driving expression of the zebrafish homolog *msxB* and the murine homolog *Msx1*. These experiments were performed in collaboration with Erika Kague in Shannon Fisher’s lab, wherein I made all the DNA constructs, she injected them into zebrafish to generate transient or stable transformed lines, and I performed the *in situ* gene expression analysis in embryos she sent me carrying the various constructs. As a first step in identifying the zebrafish *msxB* CRM, I surveyed the UCSC genome browser for highly conserved regions. This analysis revealed that a 2.5 Kb region of DNA immediately upstream of the *msxB* coding region had two peaks of strong sequence conservation as well as histone three lysine four (H3K4) single and triple methylation patterns typical of transcription factor accessible chromatin states (Fig.2.5A). When this fragment of DNA was tested for GFP-reporter expression it recapitulated the early *msxB* expression in dorsal neuroectodermal/neural crest progenitor cells (Fig.2.5C). This early neural plate expression pattern then resolved into narrower dorsal stripes during later neural tube stages.

Using a similar strategy to that used to identify the zebrafish *msxB* CRM (Fig.2.5B), I identified a 5 Kb region upstream of the mouse *Msx1* gene which when introduced into zebrafish embryos drove an expression pattern similar to
that of the endogenous fish \textit{msxB} pattern (Fig. 2.5C). These results suggest that both these CRMs contain sufficient information to correctly position the expression domain and are responding to positioning cues in a conserved fashion. Moreover, it underscores the high degree of cis-regulatory conservation across vertebrates as the fish and mammalian lineages diverged over 400 MYA.
Figure 2.5 – Identification of Vertebrate msx CRM. A: Zebrafish msxB genomic locus. B: Mouse Msx1 genomic locus. C: msxB, msxB-CRM and Msx1-CRM expression in zebrafish embryos.
Fly and vertebrate *msx* respond to BMPs differently even though their relative expression pattern is conserved.

To confirm that the zebrafish CRM identified is indeed sensitive to BMPs as *msxB* is known to be, stable transgenic animals carrying the 2.5 Kb *msxB*-GFP reporter construct were crossed to lines carrying either Heat Shock-Chordin (HS-CHD) or Heat Shock-BMP (HS-BMP) constructs. When the BMP antagonist Chordin was induced by heat treatment, *msxB*-GFP reporter expression was strongly suppressed, as was the endogenous *msxB* expression. Reciprocally, in HS-BMP embryos, endogenous and reporter gene expression was intensified (Fig. 2.6) compared to control embryos that were subjected to the same conditions but did not carry the heat shock constructs.
Figure 2.6 – Differential response of *Drosophila* and vertebrate *msh/msx* genes to BMPs.
2.4 Conclusions

The binding site composition of Drosophila msh revealed in this study supports a direct role for BMP repression acting via the SE1 and SE2 sites. No deletions other than the one containing these two sites gave rise to dorsal expansion, and no deletions other than those containing the SE1,2 or AE1,2 elements resulted in reduced gene expression. It is possible that some deletions that had no net effect on ME-lacZ expression eliminated a combination of both activation and repression sites to account for the fact that expression of the intact reporter construct, like that of the endogenous msh gene, exhibited a more complete dorsal expansion in a dpp- mutant background. The nature of such hypothetical repressor sites is unknown but presumably they should bind to Mad, Medea and Shnurri or possibly yet unknown BMP mediators alone or in conjunction with other transacting factors. Since my deletion analysis left 36 bp of the ME untested, it is also possible that a yet uncharacterized TFBS necessary for dorsal repression is present at these poorly conserved regions.

I also identified sites that are necessary for msh activation. These sites, while resembling Mad-Medea bipartite binding sites did not bind these proteins in vitro. Given the suggested role of AE sites in positively regulating BMP response genes and the evidence that BMPs can at least in part act positively to promote msx gene expression in vertebrates, it is tempting to speculate that these sites may once have been BMP responsive and were then co-opted by different transcription factors in the course of evolution to maintain msh expression in a
BMP independent fashion. Identifying these transcriptional activators will be the goal of future experiments.

In conclusion, my analysis in *Drosophila* strongly suggests that BMPs pattern the neuroectoderm primarily via repression while in zebrafish, a positive activity may also contribute to regulating orthologous *msx* genes. Thus, *msh* expands dorsally in the absence of Dpp while ectopic Dpp expression reduced *msh* gene expression. In the case of the zebrafish *msxB* gene the reverse was true. In a situation of reduced BMP signaling (HS-CHD embryos), *msxB* was not detected while its expression was ubiquitously induced in the presence of ectopic BMP activity (HS-BMP embryos). Further analysis of the zebrafish *msxB* CRM by mutagenesis of candidate BMP responsive sites as performed for the *Drosophila msh* CRM, will reveal whether such sites mediate repression, activation or some combination of both activities. Similar analysis of the CRM driving expression of the echinoderm *Msx* homolog, which is expressed in regions of peak BMP activity in a BMP-dependent fashion will also be informative since in this case one might predict finding only positively acting AE-like BMP-responsive sites.

It would also be interesting to understand how flexible the ancestral metazoan state was by investigating the relationship between BMPs and *msx* genes in basal metazoans such as jellyfish. In these diploblastic animals, although the BMP-*msx* relationship has not been tested, BMP2/4\(^7\) and *msx*\(^78\) homologues are expressed in adjacent regions in the course of development, like in the majority of triploblastic animals.
Chapter II is part of collaborative efforts and it would not have been possible without the help of Erika Kague in Shanon Fisher’s lab from University of Pennsylvania who helped me with the creation of zebrafish transgenic animals and Alexander “Helge” Springhorn in Georgios Pyrowolakis lab from University of Freiburg who helped me with the EMSAs and with helpful bioinformatic analysis and discussions.
Final Discussion of the Dissertation

I started with the notion that several features of the nervous system common to bilaterians were ancient and set out to investigate the extent to which conserved gene expression holds at the anatomical level despite of seemingly gross morphological differences (in Chapter I). In addition I examined molecular mechanisms that might underlie this conservation across large evolutionarily distances (in Chapter II).

With regard to the first question of whether homologous regions of the brain can be identified between mammals and birds, Kuhlenbeck in 1939 had already suggested a region corresponding to the chicken Dentate Gyrus to be located in the ventral posterior regions of the avian pallium (v-region of the hippocampus) based on developmental comparisons. My study lends molecular support to this view and refines more precisely where the exact borders of the Chicken Dentate Gyrus reside. In addition, the fact that homologous gene sets are expressed in a developmentally homologous region raise the question of whether cell types in these two regions are the same, and if so, would circuits within this region be functionally equivalent to the mammalian Dentate Gyrus as well? Furthermore, my gene expression studies have given additional credence to the Karten hypothesis since I demonstrated that the same genes are expressed in telencephalic regions of both mammalian and avian brain with conserved expression patterns, thus supporting the view that cells in these regions may represent fully homologous structures.
In conclusion, in these studies I addressed the issue of identifying conserved brain structures at the cellular level with an Evo/Devo strategy and successfully identified a homologous structure, and possibly neurons, across different taxa using the expression of highly conserved genes as guide. This kind of approach can be extended to other model organisms to investigate how ancient the basic cellular circuitry needed to form complex behaviors truly is.

One of the potential pitfalls of the gene expression pattern approach is that even if the genes themselves have been conserved, their downstream targets may have changed and thus gene expression patterns might not represent meaningful evolutionary signatures of functional conservation (e.g. HOX genes and their targets). Thus additional studies of downstream targets of transcription factors expressed in the Dentate Gyrus (e.g., Prox1) will be informative as well as additional functional comparisons of neural projection patterns and physiology between cells displaying conserved gene expression signatures.

Another related question is whether the upstream regulatory mechanisms controlling similar patterns of gene expression are also conserved in different branches of phylogeny. A striking example of gene expression pattern conservation is observed in the early development of the nervous system in *Drosophila* versus vertebrates with regard to the neuroblast identity genes. Since BMPs are the only common patterning system retained by both *Drosophila* and vertebrates in the early nervous system, and one of the most ancient regulatory factors are transcription factors in the *msh/msx* family, comparing cis-regulation
of msh/msx by BMPs provided an excellent paradigm for studying whether cis-regulatory processes are maintained across distant taxa. My analysis of the Drosophila msh embryonic CRM suggests that BMPs are acting mainly through two binding sites that mediate repression, while activation sites, although resembling BMP responsive sites, are unlikely to be responsive to BMP signaling. In contrast, I identified an embryonic zebrafish msxB CRM that responds to BMPs by being activated in the presence of BMP signaling and being silenced when BMP signaling is suppressed. These results suggest that the regulatory mechanisms may be opposite in these lineages, yet resulting in similar expression patterns. Further analysis of the zebrafish msxB CRM by identifying and mutating BMP-responsive sites will be necessary to confirm this hypothesis.

In Drosophila, Evo/Devo studies of the even-skipped stripe 2 CRM have suggested that regulatory mechanisms that lead to a particular gene expression pattern are extremely flexible, i.e., the same pattern can be achieved in multiple ways. Accordingly, in the BMP/msx case, it seems that natural selection may have operated similarly to maintain relevant gene expression patterns that fulfill a particular function while allowing the upstream mechanisms generating that pattern to change over time. If this view is correct, then similar Evo/Devo approaches to identifying conserved brain cell types across different taxa should be fruitful.
References


