The Establishment and Stabilization of Anterior-Posterior Identity
In the Hindbrain:
On the Regulation of the Segmentation Gene $MafB$

by

Angela Sing

A thesis submitted in conformity with the requirements
for the degree of Doctorate of Philosophy
Graduate Department of Molecular Genetics
University of Toronto

© Copyright by Angela Sing 2009
Abstract

The Establishment and Stabilization of Anterior-Posterior Identity in the Hindbrain: On the Regulation of the Segmentation Gene *MafB*

Angela Sing

Doctorate of Philosophy 2009

Department of Molecular Genetics, University of Toronto

In vertebrates, the embryonic hindbrain is transiently subdivided along its anterior-posterior (A-P) axis into 8 well defined segments termed rhombomeres (r1-8). Each rhombomere represents a true cellular compartment in transcriptional profile, lineage restriction and neuronal organization. Thus, the vertebrate hindbrain provides a beautiful model for studying mechanisms of anterior-posterior patterning, signal transduction and interpretation, initiation and maintenance of transcriptional profiles, cell sorting and border formation. The *Kreisler/MafB* gene, which encodes a basic leucine zipper (bZIP) transcription factor that regulates some *Hox* genes, is one of the first genes to be expressed segmentally in the hindbrain, and is subject to a dynamic and complex regulatory process. However, unlike the *Hox* genes, *Kreisler/MafB* is not located within a large cluster of genes and therefore provides a simple system for dissecting the molecular mechanisms involved in hindbrain compartmentalization. In dissecting the mechanisms that govern *Kreisler/MafB* regulation, we have identified the S5 regulatory element that directs early *MafB* expression in the future r5-r6 domain. We have found a binding site within S5 that is specific for the Variant Hepatocyte Nuclear Factor 1 (vHNF1) to be essential, but not sufficient for early induction of r5-r6-specific expression. Thus, early inductive events that initiate *MafB* expression are clearly distinct from later acting ones that modulate its expression levels. Using mouse mutants, we have shown that *MafB* is dependent on the M33 polycomb protein and other mechanisms of chromatin remodeling. We then utilized transgenic flies and mice as well as binding assays to identify and validate a PcG/trxG response element (PRE), PRE1 which acts to reorganize the surrounding chromatin, regulating S5-dependent expression. To our knowledge, PRE1 is the first validated vertebrate PcG/trxG response element. Thus, PRE1 provides a springboard for further exploration of the mechanisms governing chromatin remodeling.
Acknowledgements

For specific contributions to each project, please refer to the end of each corresponding chapter. There are so many people to whom I owe thanks throughout the eight years of this project that I cannot express in these few and feeble words my full appreciation. Know that you are in my thoughts and I am eternally grateful.

I am deeply indebted to my thesis supervisor and mentor Dr. Sabine Cordes for offering me the “soft sell” on this project, highlighting the riskiness, while conspicuously hinting at its potential. I thank her for her years of enthusiasm, inspiration and support. I am especially grateful for her encouragement, advice and good ideas in these last months of writing, though to me, the finish always seemed out of reach.

To all past and present members of the Cordes Lab, you have all left an impression, an influence, and at times, an inspiration. I thank you for your camaraderie, helpful discourse, debate and the infinite little favours that make science possible. In particular, I would like to thank Virginal Sadl in the early years and Anson Sing and Joanna Yu in the later years for being invaluable resources of information and help. I thank you for keeping science fun at times when it became tedious, the distractions, and your friendship.

To my thesis committee members Drs. C.C. Hui and H.D. Lipshitz: I am grateful for your years of support, encouragements, advice and much needed and persistent urging to finish in the final years.

Benjamin Spergel, my 24/7 on-call tech support and Graphics-GOD, I thank you for sharing your many talents, time and support. For being my stalwart and seeing that I finished in spite of myself, you have my admiration, my gratitude and love.

Lastly, I wish to thank my parents Hanc and Ellen Sing, for being the solid foundation on which I stand, whether it be financial support, transport, foodstuffs or their ever quiet encouragement, I cannot overstate my gratitude to them. They have always lived and taught by an example we should all strive for. To them, I dedicate this thesis.
# Table of Contents

**ABSTRACT**  
ii

**ACKNOWLEDGEMENTS**  
iii

**TABLE OF CONTENTS**  
iv

**LIST OF FIGURES**  
vii

**LIST OF TABLES**  
xix

**CHAPTER 1: INTRODUCTION**  
1

## The Establishment and Stabilization of Anterior-Posterior Identity in the Hindbrain

### A) Segmentation and Patterning in the Vertebrate Hindbrain

- Hindbrain segmentation and cellular compartmentalization 2
- Ephrins and Eph receptor signaling in rhombomeric cell sorting 6
- Role of retinoic acid in patterning the posterior hindbrain 7
- Hox genes and posterior patterning 10
- RA regulation of Hox gene expression 12
- Integrated signaling pathways from the isthmic organizer 13
- The hindbrain segmentation gene MafB/Kreisler/Krml1 14
- MafB mouse mutants 17

### B) Mechanisms of Chromatin Remodeling in the Developing Vertebrate

- Integrated mechanisms of epigenetics in gene silencing 18
- DNA methylation and gene transcription 19
- Chromatin structure and histone modifications 21
- Chromatin modifications in normal physiological processes and disease 23
- Cellular memory and transcriptional silencing by the Polycomb group proteins 24
- Genetic identification of the Polycomb and trithorax group of proteins 25
- Chromatin modifications by Polycomb repressive complexes 1 and 2 26
- PcG silencing during mammalian development has consequences of aberrant activation 30
- Transcriptional regulation by trxG proteins 31
- Genetic identification of trithorax group proteins 33
- TrxG proteins in mammals 33
- Diverse functions of trxG proteins 35
- Polycomb and Trithorax Response Elements 37
- Defining a PRE/TRE 38
- Reprogramming of PRE/TREs throughout development 40
- Identifying PRE/TREs in mammals 41

**Goal of this Thesis**  
45
CHAPTER 2: THE vHNF1 HOMEODOMAIN PROTEIN ESTABLISHES EARLY RHOMBOMERE IDENTITY BY DIRECT REGULATION OF MafB EXPRESSION 46

Abstract 46

Introduction 47

Materials and Methods 48

Results 53
  Identification of an early-acting rhombomere-specific enhancer for MafB 53
  An evolutionarily conserved consensus site is present in the MafB enhancer 55
  vHNF1 is expressed transiently in r5 and r6 58
  vHNF1 interacts functionally with a single site in the MafB enhancer 60
  The vHNF1 binding site is essential, but not sufficient for early MafB expression 62

Discussion 65

Acknowledgments 70
  Footnote on previously published materials 70

CHAPTER 3: A VERTEBRATE POLYCOMB RESPONSE ELEMENT GOVERNS SEGMENTATION OF THE POSTERIOR HINDBRAIN 71

Abstract 71

Introduction 72

Materials and Methods 75

Results 78
  MafB protein is ectopically expressed in r3 of kr/kr and kr/+ embryos 78
  Analysis of the kr X-ray induced inversion suggest the presence of a PRE 80
  Chromatin remodeling agents reactivate MafB expression in the posterior hindbrain 82
  The M33 polycomb gene regulates MafB expression 83
  Sequence analyses identifies a potential polycomb response element 86
  The PRE1 region can silence gene expression in Drosophila 90
  Specific PcG and trxG proteins modulate PRE1 function 93
  PRE1 can recruit Pc and PH binding in Drosophila 97
  Analysis of PRC2 binding in the mouse F9 embryonic carcinoma cell line 97
  PRE function in transgenic mice 101

Discussion 104
  PRE1 may restrict gene expression to the posterior hindbrain 104
  PRE1 may regulate MafB initiation 105
  Interactions of specific PcG and trxG proteins with PRE1 107
## List of Figures

### Chapter 1

1.1 Organization of cranial motor nerves and ganglia in the mouse 3  
1.2 Maintenance of rhombomere identity depends on transcriptional profile 5  
1.3 Integrated signals from the anterior and posterior convey positional information on the developing hindbrain 9  
1.4 The hindbrain segmentation gene *MafB/Kreisler/Krml1* and its mutant 16  
1.5 DNA is tightly condensed into chromatin 22  
1.6 Histone tail modifications 22  
1.7 The concept of cellular memory 32

### Chapter 2

2.1 Identification of a regulatory element required for early *Kreisler* expression 54  
2.2 Evolutionary sequence conservation within the rhombomere-specific S5 enhancer 56  
2.3 RAR/RAR and RAR/RXR binding to the S5 enhancer 57  
2.4 Whole mount RNA in situ hybridization detected vHNF1 expression in the hindbrains of -014 somite stage mouse embryos 59  
2.5 vHNF1 binds to a single site within the S5 element 61  
2.6 vHNF1 can activate transcription via the S5-vHNF1 site 63  
2.7 vHNF1 binding site is essential, but not sufficient, for r5- and r6-specific expression of Kreisler in 0-10 somite stage embryos 64  
2.8 Models for vHNF1 dependent activation of rhombomere-specific *Kreisler* expression 69
### Chapter 3

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Effects of the $kr$ and $kr^{enu}$ mutations on MafB expression</td>
<td>79</td>
</tr>
<tr>
<td>3.2</td>
<td>Rhombomere-specific expression of <em>Neuronatin</em> is altered by the $kr$ inversion</td>
<td>81</td>
</tr>
<tr>
<td>3.3</td>
<td>Chromatin remodeling agents and $M33$ levels affect $MafB$ expression in normal and $kr/kr$ embryos</td>
<td>84</td>
</tr>
<tr>
<td>3.4</td>
<td>PRE1 and hcPRE1 exhibit hallmarks of a Polycomb Response Element</td>
<td>87</td>
</tr>
<tr>
<td>3.5</td>
<td>PRE1 contains a conserved sequence motif in Polycomb Response Elements identified in <em>Drosophila</em></td>
<td>89</td>
</tr>
<tr>
<td>3.6</td>
<td>Diagram of reporter construct used to generate transgenic flies</td>
<td>91</td>
</tr>
<tr>
<td>3.7</td>
<td>PRE1 acts as a developmental silencer in <em>Drosophila melanogaster</em></td>
<td>92</td>
</tr>
<tr>
<td>3.8</td>
<td>PRE1-dependent silencing is relieved by mutations in PcG genes</td>
<td>94</td>
</tr>
<tr>
<td>3.9</td>
<td>PRE1-dependent silencing is enhanced by mutations in trxG genes</td>
<td>95</td>
</tr>
<tr>
<td>3.10</td>
<td>Effects of PcG/trxG mutations on PRE1-dependent silencing is specific</td>
<td>96</td>
</tr>
<tr>
<td>3.11</td>
<td>PRE1 transgene recruits ectopic Polycomb and polyhomeotic binding in <em>Drosophila</em></td>
<td>98</td>
</tr>
<tr>
<td>3.12</td>
<td>Binding profiles of the PRC2 component SUZ12, RNA polymerase II, and methylated forms of Histone 3 at PRE1 in F9 cells</td>
<td>100</td>
</tr>
<tr>
<td>3.13</td>
<td>hcPRE1 silences induction by S5 in transgenic mice</td>
<td>102</td>
</tr>
<tr>
<td>3.14</td>
<td>Schematic model of the proposed role of PRE1 in MafB expression</td>
<td>103</td>
</tr>
<tr>
<td>3.15</td>
<td>Example of mating scheme to test effect of PcG mutants form chromosome III on PRE1-dependent silencing</td>
<td>112</td>
</tr>
<tr>
<td>3.16</td>
<td>Original mating scheme to test effect of PcG mutants from chromosome II on PRE1-dependent silencing</td>
<td>113</td>
</tr>
<tr>
<td>3.17</td>
<td>Alternative mating scheme to test effect of PcG mutants from chromosome II on PRE1-dependent silencing</td>
<td>114</td>
</tr>
<tr>
<td>3.18</td>
<td>Mating scheme to test effect of mutation in $ph$ from X chromosome on PRE1-dependent silencing</td>
<td>115</td>
</tr>
<tr>
<td>3.19</td>
<td>Mating scheme to test effect of mutation in <em>Pho</em> from chromosome IV on PRE1-dependent silencing</td>
<td>116</td>
</tr>
</tbody>
</table>
List of Tables

T1.1 Core PcG genes in model systems 27
T1.2 Biochemical functions of trxG proteins 34

T2.1 Transgenic embryos with r5 and r6-specific LacZ expression 52

T3.1 Sequence comparison of hcPRE1 across vertebrate species 88
T3.2 Statistical analysis of the effect of PRE1-dependent silencing in transgenic flies 116
T3.3 Statistical analysis of the effect of PcG/trxG mutants on PRE1 silencing 117
T3.4 Statistical analysis for ChIP assays 118
T3.5 Lines of *Drosophila melanogaster* used in analysis of PRE1 dependent silencing 119
Chapter 1: Introduction

The Establishment and stabilization of anterior-posterior identity in the hindbrain

A) Segmentation and Patterning in the Vertebrate Hindbrain

The developing vertebrate hindbrain is transiently subdivided along its anterior-posterior (A-P) axis into 8 segments termed rhombomeres. Each rhombomere represents a true cellular compartment in transcriptional profile, lineage restriction and neuronal organization. Key to providing positional information to each rhombomere are a small number of segmentation genes, including the Hox family of homeotic genes whose expression precedes the formation of rhombomeric borders and function to specify rhombomeric identity. Furthermore, neurogenesis in the hindbrain originates from small groups of cells that obey rhombomeric borders while mature neurons can be traced back to their rhombomeric origins in cell tracing experiments. These include motor, sensory as well as interneurons. Early on, much of what we know about the vertebrate hindbrain comes from experiments in segmentation in the chick hindbrain. Later, important advances have also been made in both zebrafish and mouse; the mouse was particularly crucial in deciphering the genetics and molecular mechanisms involved in this highly complex process. The developing hindbrain therefore is a highly productive area for studying mechanisms of anterior-posterior patterning, signal transduction and interpretation, initiation and maintenance of transcriptional profiles, cell sorting and border formation. The Kreisler/MafB gene, which encodes a basic leucine zipper (bZIP) transcription factor that regulates some Hox genes, is one of the first genes to be expressed segmentally in the hindbrain, and is subject to a dynamic and complex regulatory process, however, unlike the Hox genes, Kreisler/MafB is not located within a large cluster of genes and therefore may provide a simple system for dissecting and understanding the molecular mechanisms involved in hindbrain compartmentalization and anterior-posterior patterning.
Hindbrain segmentation and cellular compartmentalization

In many organisms, ranging from plants to invertebrates to vertebrates, the developing biological system depends on segmentation to lay down precise positional information in order to establish complex multimeric structures. Segmentation allows for the use of a common blueprint by allocating specific precursor cells a defined identity, so that they can differentiate into their successive parts or structures. This process is perhaps best understood in the *Drosophila melanogaster* embryo, where morphological studies have revealed a visibly segmented structure which is controlled by the homeotic genes. Additionally, the ease of genetic screens in the fly has allowed researchers to visualize the role of various genes responsible for this process. For example, pair-rule genes such as *even-skipped* are expressed in striped patterns of seven bands perpendicular to the A-P axis, each stripe is then refined by segment polarity genes such as *en不了解*

The vertebrate hindbrain is perhaps the most fly-like region in the developing chordate embryo (Lumsden, 2004). Early in its development, the hindbrain is subdivided along its rostral-caudal axis into 8 bulge-like segments called rhombomeres. Like segmentation in the fly embryo, rhombomeres form by subdivision, rather than by budding from a growth zone then organized into pairs like the somites. Also, there appears to be a two segment repeat pattern of alternate rhombomeres as evidenced by the appearance of specific neurons. For example the cranial nerve axons can be seen leaving even-numbered rhombomeres (r2, r4 and r6) (Figure 1.1B) (Cordes, 2001). Early translocation experiments in the chick performed by Sarah Guthrie clearly demonstrate this alternating pattern of like rhombomeres that are defined by their transcriptional profiles (Figure 1.2) (Guthrie and Lumsden, 1991). Each rhombomere represents a true cellular compartment in that cells from each rhombomere share a distinct lineage, cell affinity and genetic identity. Once the compartments are formed, there is a physical and clearly visible border between the rhombomeres. Moreover, cells from neighboring rhombomeres do not mix (with the exception of migrating cells at the floorplate), thereby restricting movement of cells within each rhombomere. This is best illustrated by the cell lineage tracing studies in the chick. There, Scott Fraser, Roger Keynes and Andrew Lumsden labeled individual cells *in ovo* with an intracellular vital dye, then allowed them to multiply and form cell clones over one to two developing days.
Figure 1.1. Organization of cranial motor nerves and ganglia in the mouse.

(A) A dorsal view of the mouse hindbrain at 10.0 days post coitum (dpc), stained with anti-neurofilament (2H3) antibody. Rhombomeres are clearly visible at this time and are labelled r1–r8. Axonal staining can be seen in r1, and in the cranial nerve axons leaving even-numbered rhombomeres (r2, r4 and r6) and at the inter-rhombomeric boundaries. ov, otic vesicle. (B) A schematic diagram shows the cranial motor nerves (m) and cranial ganglia (g) of the mouse. Cell bodies of the motor nerves are at different mediolateral positions within the hindbrain, whereas those of the sensory ganglia are located outside the hindbrain and derived from neural crest cells. IV, trochlear nerve; V, trigeminal nerve; VI, abducens nerve; VII, facial nerve; VIII, acoustic nerve; IX, glosso-pharyngeal nerve; X, vagus nerve; XI, spinal accessory nerve; XII, hypoglossal nerve. The III, IV, VI and XII nerves are somatic motor nerves and have ventrally exiting axons. Axons from the branchiomotor (V, VII, IX, X and XI) and visceromotor (III, VII, IX and X) nerve axons exit the neural tube dorsally at distinct exit points located in the even numbered rhombomeres. Sensory ganglia and motor nerves use the same exit points.

CVA, contralateral vestibular neurons; fp, floorplate; mb, midbrain; sc, spinal cord. Adapted from Lumsden and Keynes, 1989 for the mouse according to Fritsch, 1998. (C) A sagittal view of a 10.5 dpc mouse embryo stained with anti-neurofilament antibody shows the distinct appearance of each cranial ganglion.

Cells were seen to freely mix with neighboring unlabelled cells within the same rhombomere, but never moved from one rhombomere to another. In contrast, when clones are marked before boundary formation, cells were seen to straddle rhombomeric boundaries (Fraser et al., 1990).

Restrictive boundaries allow cells from each rhombomere to freely move within its compartment, receive and interpret the same signals, and for some cells eventually pursue the same developmental fates. For example, cell tracing experiments have shown that the branchio-motor neurons which exit from r2 (trigeminal), r4, (facial) and r6 (glossopharyngeal) can be traced back to their respective rhombomeric origins (Figure 1.1) (Lumsden and Keynes, 1989).

However, it would appear that although boundary formation contributes to cell lineage restriction, it is itself a consequence of rather than the sole cause of segmentation. When rhombomeric border cells are ablated, the gap formed is quickly filled in with no mixing between adjacent rhombomeres (Guthrie and Lumsden, 1991). Furthermore, boundaries appear to form wherever cells from odd and even rhombomeres are juxtaposed, while cells from like rhombomeres, for example r2 and r4 intermingle (Figure 1.2) (Guthrie et al., 1993). Finally, classic chick studies have shown that when hindbrains are treated with retinoic acid prior to boundary formation, posterior rhombomeric boundaries disappear as evidenced by cell morphology, gene expression and accumulation of axons normally found at these boundaries. However, even in the absence of borders, there was no change in the dispersal of precursor cells, no mixing of cells between adjacent rhombomeres and no change in the segmentally distributed motor nuclei even though Hox gene expression is now altered (Nittenberg et al., 1997).

While the adhesive and self sorting properties of cells within rhombomeres are incompletely understood, a variety of experiments, the first of which were performed by Wilkinson et al. in the chick, point towards a role for ephrins and Eph receptors in cell-contact-dependent signaling in this sorting mechanism (Wilkinson, 2001).
Figure 1.2. Maintenance of rhombomere identity depends on transcriptional profile. Classical chick transplantation experiments by Sarah Guthrie demonstrated that like rhombomeres (in this case two even rhombomeres r2 and r4) when transplanted together would generate a large compartment that is a mixture of r2 and r4 cells. However when an even and an odd rhombomere (in this case r3 and r6) are juxtaposed, the cells with a different complement of segmentation genes segregate out into two distinct compartments.
Ephrins and Eph receptor signaling in rhombomeric cell sorting

The process by which odd and even rhombomeres acquire and maintain differential affinities involve the members of the receptor tyrosine kinase, Ephs and their membrane-bound ligands, ephrins. These play an important role in preventing cells from neighboring rhombomeres from intermingling once the boundaries have formed by mediating cell contact-dependent repulsion. Evidence of this has been suggested by their complementary expression patterns: the receptors (EphA4, EphB2 and EphB3) are expressed in odd rhombomeres (r3 and r5), and their ligands (ephrin-B1, B2, B3) are found in even rhombomeres (r2, r4, and r6) (Xu et al., 2000). This ensures that Eph receptors and their corresponding ephrins can interact at the boundary interfaces and promote cell repulsion, thus preventing cells from one rhombomere from entering a neighbouring one. Xu et al. further demonstrated the role Eph and ephrins play in rhombomere cell sorting by generating mosaic rhombomeres. They found that activating Eph receptors in a population of cells within ephrin expressing rhombomeres lead to sorting of Eph+ cells to the already formed boundaries. In contrast, mosaic activation of ephrins in even rhombomeres where they are normally expressed results in a scattering of the ectopic cells (Xu et al., 1999). These data implicate Eph/ephrin signaling in the segmental restriction of cell intermingling at the borders, sharpening and maintaining them.

In addition to the cell repulsion events at the rhombomeric borders, there is evidence in zebrafish that suggests adhesive forces may also be involved. In loss of function experiments using antisense morpholinos against EphA4, embryos not only had disturbed rhombomeric borders as predicted but, in rhombomeres that normally express EphA4, mosaics of EphA4MO sort out from normal cells, suggesting EphA4 promotes cell adhesion within the rhombomeres in which it is expressed (Cooke et al., 2005). Finally, there is evidence that fibroblast growth factor receptor 1 (FGFR1) can modify these cell adhesive properties (Trokovic et al., 2003).

Upstream control of Eph/ephrin expression may depend on earlier acting transcription factors that also share an even or odd rhombomeric identity such as the zinc finger protein Krox20 which is expressed in two stripes in the future r3 and r5. There is evidence from Krox20 null mice as well as ectopic overexpression of Krox20 in chicks
that it may be a key player in establishing odd-rhombomere identity (Wilkinson, 1995). *MafB/Kreisler* is another early expressed gene found in r5 and r6. It is directly responsible for the activation of *Hoxa3* and *Hoxb3* (Manzanares et al., 1999a; Manzanares et al., 1997), as well as a repressor of *Hoxb1* (Giudicelli et al., 2003). Loss of *MafB* in the hindbrain leads to a complete loss of r5, a mis-specified r6 and a general loss of segmentation posterior to the r3/r4 boundary (Giudicelli et al., 2003; McKay et al., 1994; Sadl et al., 2003). Early analysis of these *MafB* hindbrain-specific null mutants (*kr*) revealed a downregulation of Eph receptors (Manzanares et al., 1999b; McKay et al., 1994). This was later confirmed in the zebrafish *MafB* mutant *valentino* (*val*) where EphB4a is normally expressed in the MafB domain of r5 and r6, while ephrin-B2a, which encodes a ligand for EphB4a, is expressed in surrounding rhombomeres r4 and r7. In *val* mutant embryos, EphB4a expression is downregulated and ephrin-B2a expression is upregulated between r4 and r7 (Cooke et al., 2001). Taken together, these data implicate *MafB* as a master regulator of rhombomere identity, establishing the transcriptional profiles of *Hox* genes as well as repulsive/adhesive cues in the posterior hindbrain, thus playing a major role in regulating segmentation in the hindbrain.

Clearly, there is repulsive and perhaps adhesive signaling between Ephrins and their receptors the Ephs and they may play a crucial role in boundary formation and or maintenance. However the actual process of cell sorting, both the aggregation of like and repulsion of unlike cells remain dependent on the transcriptional programs initiated by early segmentation genes such as *MafB* and *Krox20*, rather than any known feedback signaling from the Eph/ephrins.

**Role of retinoic acid in patterning the posterior hindbrain**

One of the most basic and best understood examples of simple signals establishing a gradient which in turn gives rise to anterior-posterior identity comes once again from the *Drosophila* embryo. Well before the egg is even fertilized, the oocyte is polarized by the maternal effect genes, expressed first as mRNAs and then translated upon fertilization. Gradients of *Bicoid* and *hunchback* from the anterior, and *Nanos* and *Caudal* from the posterior set up the A-P axis of the embryo. The transcription factors Bicoid, Caudal and Hunchback then regulate the expression of later acting gap genes and pair-rule genes which further specify each segment.
In the vertebrate, the most important posteriorizing signal is the active derivative of Vitamin A, retinoic acid (RA). Deficiencies in RA as seen in vitamin A deficient (VAD) animals have devastating effects in the posterior hindbrain and branchial arch region, presumably by directly or indirectly altering specific subsets of Hox gene expression (Gale et al., 1999; White et al., 2000). Excess RA causes a dose-dependent posteriorization, where the hindbrain is expanded at the expense of the midbrain and forebrain (Avantaggiato et al., 1996). RA signaling is transduced via a number of retinoic acid receptors RARs and RXRs which act as ligand-dependent transcription factors that, when activated, are translocated into the nucleus where they bind DNA as a heterodimer at specific elements termed RA response elements (RARE). The expression of several segmentation genes, particularly the posterior Hox genes is induced in this manner (Marshall et al., 1994). There are three types of RAR receptors (α, β and γ) and three types of retinoid X receptors (RXR-α, -β and -γ). Knockout studies have shown that there are functional redundancies between the subtypes which are cell-type and promoter dependent. And although all three RARs can functionally substitute each other in the induction of RA target genes and cell differentiation, depending on the context, one RAR can cell-specifically override the activity of the other RARs (Taneja et al., 1996).

Early in embryogenesis, RA is produced in the posterior mesoderm by the enzyme Retinaldehyde dehydrogenase 2 (Raldh2) and is thought to diffuse up through the spinal cord and hindbrain. A second cytochrome p450 class enzyme, Cyp26C1 degrades RA in the rostral hindbrain and is expressed at early developmental stages E8.0-E8.5 from the prospective r2 and r4, in the first branchial arch and along the lateral surface mesenchyme adjacent to the anterior hindbrain at the isthmus (Figure 1.3) (Tahayato et al., 2003). Thus, a model emerges where Raldh2 and Cyp26C1 act in concert to establish an RA gradient from the posterior to anterior. This gradient would then be responsible for the expression of 3' Hox genes that respond more quickly to lower RA concentrations than more 5' Hox genes (Figure 1.3) (Papalopulu et al., 1991). Support for this model comes from mouse knockouts of Raldh2 (Niederreither et al., 2000) and disruption of RARγ receptors (Dupe and Lumsden, 2001), both leading to anteriorization of the hindbrain. That is, loss of posterior Hox gene expression and retaining expression of rostral Hox genes only in the most posterior hindbrain.
Figure 1.3. Integrated signals from the anterior and posterior convey positional information on the developing hindbrain.
At the 6 somite stage, early signals emanating from the isthmic organizer including FGF8 establish the anterior hindbrain. Once r4 has formed, FGFs are expressed from r4, further patterning the caudal hindbrain. From the posterior mesenchyme, RALDH2 synthesizes RA which diffuses up the spinal cord and hindbrain. A second cytochrome p450 class enzyme, Cyp26C is expressed from the prospective r2 and degrades RA in the rostral hindbrain, creating a RA gradient from the posterior to anterior. Early segmentation genes including Kreisler and Krox20 as well as Hox genes respond at varying degrees to the different local concentrations of RA, resulting in their segmental expression patterns.
RA signaling also affects other early acting segmentation genes such as *Krox20* (Dupe and Lumsden, 2001) and *MafB* (Grapin-Botton et al., 1998). The expression of both genes is lost in the absence of RA signaling, and can be shifted anteriorly when retinoid concentrations are artificially increased. Thus, just as in the *Drosophila* embryo, the alteration of the gradient of a single morphogen can have an enormous impact on local transcriptional states as well as future developmental fates.

**Hox genes and posterior patterning**

The most widely studied segmentation genes are the *Hox* family of homeotic genes. Early in metazoan development, this family of homeobox transcription factors specifies the anterior-posterior axis and segment identity. Homeotic genes contain a homeobox which encodes the helix-turn-helix homeodomain that binds at specific enhancers of downstream genes ultimately responsible for terminal differentiation. *Hox* genes have been highly conserved throughout metazoan evolution. In mammals, there are 39 *Hox* genes organized into four clusters located on four different chromosomes, labeled *HoxA*, *HoxB*, *HoxC* and *HoxD*. Within each cluster, there are 13 paralogous groups classified by both sequence conservation and position on the chromosome (Scott, 1993). Intriguingly, there is a correlation between the position of a *Hox* gene within the cluster and its temporal and spatial expression pattern. The genes located most 3’ in the cluster are expressed earlier in more anterior regions (e.g. the hindbrain), while genes more 5’ are expressed later and more posterior (Duboule, 1998; Duboule and Dolle, 1989; Graham et al., 1989). This colinearity aspect of *Hox* gene expression has been documented in invertebrates such as *Drosophila* and in all vertebrates. The mechanisms controlling this process are poorly understood, although the clustered organization of *Hox* genes would suggest that this is crucial for coordinated expression. Support for this comes from analysis of *HoxB* chromatin structure and nuclear organization in mouse gastrula, as well as E9.5 embryos where there is a progressive opening of chromatin structure from 3’ to 5’, which corresponds to the sequential activation of *HoxB* gene expression (Chambeyron et al., 2005).

Another significant characteristic of the *Hox* clusters is that they are capable of highly efficient auto- and cross-regulation via enhancer sharing. Nearly two decades ago, autoregulatory loops were first recognized in the fly by the McGinnis and Kaufman labs
for the *Drosophila* homeotic genes *labial (lab)* and *Deformed (Dfd)*, where high affinity binding sites for each protein were found within their respective enhancers (Chouinard and Kaufman, 1991; Regulski et al., 1991). Later, it was found that the mouse homologs of *lab* and *Dfd* (*Hoxb1* and *Hoxb4* respectively) also depend on autoregulatory mechanisms in the activation of their expression (Gould et al., 1997; Popperl et al., 1995). Gould et al. also identified an enhancer located between *Hoxb3* and *Hoxb4* that could work efficiently with both promoters. Additionally, they found *Hox* binding sites within this enhancer, that when bound by *Hoxb4* or *Hoxb5* could activate these promoters demonstrating cross-regulation (Gould et al., 1997). The Krumlauf group went on to show this relationship between other *Hox* genes lying in the same cluster (Sharpe et al., 1998).

In addition to enhancer sharing, *Hox* genes lying within a cluster also benefit from the use of global enhancers. For example, limb development requires the coordinated expression of several *HoxD* genes. By using a targeted enhancer-trap approach, Spitz et al. identified a DNA segment which they termed a global control region (GCR) capable of directing gene expression over distances of several hundred kilobases at the mouse *HoxD* cluster (Spitz et al., 2003). Adding to the immense complexity of gene regulation at *Hox* clusters is the contribution of downstream targets feeding back into the equilibrium of *Hox* genes expressed, as in the case of the murine *retinoic acid receptor β*, *Hoxb4* and *Hoxd4* feedback loop (Serpente et al., 2005). Finally, it is clear that the positioning of *Hox* genes is crucial not only to their spatial patterns of expression, but also to the timing of their activation, since translocation of *Hox* genes to ectopic locations sometimes maintains their spatial expression pattern, but their temporal expression is misregulated (Kmita and Duboule, 2003; Spitz et al., 2005).

Clearly, spatial and temporal regulation of *Hox* gene expression is highly complex, involving interpretation of signaling molecules, global and shared enhancers, auto- and cross-regulation between *Hox* genes, feedback loops, and mechanisms of chromatin remodeling. However, their highly coordinated organization within clusters makes it more difficult to tease out regulatory mechanisms responsible for their precise spatial and temporal expression patterns.
RA regulation of Hox gene expression

Several studies have shown that RA-induced patterning acts in part via Hox gene mediation. The addition of exogenous retinoids resulted in dramatic anterior shifts of Hox gene expression boundaries, followed by mis specification of rhombomeric identities (Conlon and Rossant, 1992; Marshall et al., 1992; Morriss-Kay et al., 1991). Conversely, decreasing RA signaling in the hindbrain resulted in a posterior shift of Hox gene expression domains (Gale et al., 1999; Kolm et al., 1997). Direct dependence for activation of some Hox genes on RA was revealed by the presence of RAREs located 3’ and/or 5’ to the coding regions. RAR/RXR bind as specific heterodimers at RAREs which consist of two direct repeats separated either by 2 or 5 nucleotide spacers (DR2 or DR5) and give rise to distinct DNA binding affinities. Hoxa1 (Frasch et al., 1995; Langston and Gudas, 1992), Hoxb1 (Marshall et al., 1994; Studer et al., 1994), Hoxb4 (Gould et al., 1998), and Hoxd4 (Morrison et al., 1997; Nolte et al., 2003; Zhang et al., 2000) have all been found to contain RAREs located 3’ and/or 5’ to their promoters. These elements are highly conserved across vertebrates indicating their crucial role in Hox gene regulation. The significance of these RAREs was shown in vivo when a single knockout of RAREs from Hoxa1 (Hoxa1ΔRARE) and Hoxb1 (Hoxb1ΔRARE) resulted in milder molecular and morphological defects than their full-gene knockout counterparts. Specifically, the phenotypes of Hoxa1ΔRARE or Hoxb1ΔRARE were consistent with early disruptions in expression of Hoxa1 or Hoxb1 nulls, but none of the later disruptions in expression were observed (Dupe et al., 1997; Gavalas et al., 1998; Studer et al., 1994).

Several transgenic lines have been generated carrying these RARE-containing elements driving a LacZ reporter. These studies have shown not only that RAREs are important in establishing normal Hox gene expression, but that they are also capable of inducing ectopic expression in response to RA treatment. Furthermore, multiple RAREs are responsible for different regions of expression. For example, Hoxb1 is normally expressed between E9.0 and E9.5 as a precise stripe in r4. Although a 3’ DR2 RARE was found to be required for early posterior expression up to the future r3/r4 border (Marshall et al., 1994), an additional 5’ RARE represses expression in r3 and r5 in a RA-dependent manner (Studer et al., 1994). This last example illustrates that although enhancers containing RAREs are essential for induction of expression at the anterior borders, they
themselves are not sufficient for maintaining expression throughout development. Later acting elements must be present to mediate auto- and cross-regulation between Hox genes and other segmentation genes to further refine early expression patterns.

Retinoid signaling is therefore directly responsible for establishing rostral 3’ Hox genes. However, if and how RA regulates expression of the less sensitive caudal 5’ Hox genes is not as well understood. There is evidence that activation of caudal Hox genes requires FGF signaling and in some cases, Hox induction requires interactive feedback loops between FGFs, RARs and Hox genes (Bel-Vialar et al., 2002).

**Integrated signaling pathways from the isthmic organizer**

In addition to RA from the paraxial mesoderm in the tail, there is a second important signaling centre that is found in all vertebrate species that is responsible for anterior-posterior patterning in the hindbrain: the isthmic organizer that lies at the junction between the midbrain and hindbrain. In the mouse, shortly after gastrulation (E7.5) several genes are expressed within this region which contribute to the positioning and growth promoting activities of the isthmic organizer. The homeobox-domain-containing transcription factors orthodenticle homologue 2 (Otx2) in the anterior and gastrulation brain homeobox 2 (Gbx2) in the posterior are first expressed in a complementary fashion around the future mid-hindbrain border (Simeone et al., 1992; Wassarman et al., 1997). It is the meeting point of these two gene products that determines the future position of the isthmic organizer. Other early signals expressed that further define the organizer include paired box 2 (Pax2) and the secreted signal Wnt1, which is followed shortly by engrailed 1 (En1), En2 and Pax5 (Asano and Gruss, 1992; Davis and Joyner, 1988; Davis et al., 1988; Rowitch and McMahon, 1995). By the 6 somite stage (E8-8.5) the secreted factor fibroblast growth factor 8 (FGF8) is switched on and restricted to the Gbx2 positive side (Figure 1.3). It is Fgf8 and Wnt1 that will ultimately orchestrate the maintenance and activities of the isthmic organizer, since complete or partial knockouts of either gene leads to a loss of both midbrain and anterior hindbrain structures (r1-r3) (McMahon et al., 1992; Meyers et al., 1998; Reifers et al., 1998). However only FGF8 is sufficient to act as the organizer and induce ectopic mesencephalic and metencephalic markers as evidenced by beads soaked with FGF8 and implanted in the caudal hindbrain (Irving and Mason, 2000; Martinez et al., 1999).
FGF8 also serves to position *Hoxa2* in r2 by repressing expression anterior to r1/r2 border (Irving and Mason, 2000). The posterior hindbrain (r4-r8) is regulated by later acting segmentation genes such as *Krox20, Hoxa1* and *MafB/Kr* which act upstream of the *Hox* genes. These are the among the earliest to be segmentally expressed and are thought to be induced in part by FGFs. Faustino Marin and Patrick Charnay investigated the role of FGFs in the regulation of *Krox20* and *MafB* by applying exogenous FGFs onto early chick embryos in r7 and r8, a region devoid of FGFs. They found that not only could FGFs locally induce *Krox20* and *MafB*, but that an FGF receptor inhibitor could endogenously repress the two genes as well (Marin and Charnay, 2000). R4 is the first rhombomere to form and appears to function as a second signaling centre patterning the caudal hindbrain and anterior spinal cord. It does so in part by expressing *Fgf3* and *Fgf8* (Figure 1.3), which in zebrafish have been shown to be required for r5 and r6 formation (Maves et al., 2002; Walshe et al., 2002; Wiellette and Sive, 2003). In the mouse, FGF8 appears to be the principle FGF involved with r5 and r6 development, while FGF3 is necessary for later stages of development of the inner ear (Frohman et al., 1993; McKay et al., 1996).

More recently, the FGF receptors (*Fgfr1, Fgfr2* and *Fgfr3*) have been the focus of much investigation. Using conditional mutants of various FGF receptors, the Partanen group have convincingly shown that loss of *Fgfr1*, normally expressed abundantly in the midbrain and r1 is essential for normal midbrain and r1 development, though the defects were much less severe than *Fgf8* mutants (Jukkola et al., 2006; Trokovic et al., 2003). This suggested that perhaps there is some redundancy and *Fgfr2* and *Fgfr3* may also be mediating FGF signaling. Having generated the mice carrying different combinations of mutations in these three receptors, they found that all three FGF receptors served to transduce signals form the isthmic organizer, regulating cell survival and pattern in the developing midbrain and r1 (Saarimaki-Vire et al., 2007).

**The hindbrain segmentation gene MafB/Kreisler/Krml1**

The large Maf proteins are a distinct group of regulators of cellular differentiation belonging to a larger group of evolutionary conserved basic leucine zipper domain (bZIP) containing transcription factors (Figure 1.4). Members including *MafA, MafB, c-Maf,* and NRL act as key regulators of differentiation in a wide range of tissues including brain,
kidney, bone and the gut. The bZIP structure comprises of a basic region comprising of the minimal 18 amino acid peptide that is required for specific DNA recognition and binding (Dlakic et al., 2001) and a leucine zipper which allows for homo or heterodimerization with other compatible bZIP proteins (Kataoka et al., 1994).

*MafB/Kreisler/Krml1* was identified by positional cloning in the mouse as the causal gene in the classical segmentation mouse mutant *kreisler* (*kr*) (Cordes and Barsh, 1994). *MafB* expression is conserved in mice, rats, chicken and zebrafish. In mice, it first appears as a thin stripe in the future r5 region beginning at E8.0. By E8.5 the expression domain expands into r6. As the embryo develops, hindbrain expression retracts to r6 as the neural crest (originated from r6) expression is initiated at E9.0, and is completely gone from the hindbrain by E9.5. Later domains of expression include the eye lens, kidneys, pancreas, bone, and blood. Thus far, analyses of *MafB* null mice has uncovered key roles in hematopoiesis as well as kidney development (Cordes and Barsh, 1994; Kim et al., 1999; Sadl et al., 2002; Sakai et al., 1997; Sieweke et al., 1996). In zebrafish, *MafB* was found to be the causal gene in the *valentino* (*val*) mutation, where *val* mutant hindbrains are shortened by one rhombomere (Moens et al., 1996). Outside of vertebrates there are no known orthologs of *MafB/Kreisler*, though there are a number of incidental homologs which are not expressed in a segmental fashion including *traffic jam* (*tj*) which controls gonad morphogenesis in the fly (Li et al., 2003). Additionally, *zerknult* (*zen*), a homolog of *Hoxa3* and *Hoxb3*, the direct targets of *MafB* (Manzanares et al., 1999a) does not play a role in early *Drosophila* segmentation. *Kreisler* is the only Maf family member that has been shown to have a role in segmentation.

In the mouse, *MafB* is found on the distal arm of chromosome 2 and is the subject of complex and dynamic mechanisms of regulation. At the onset of my thesis research, no transcription factors that directly regulate *Kreisler* in r5 and r6 had been identified. We and others have since shown that the regulation of *MafB* involves both anterior (FGFs) and posterior (RA) signaling (Hatch et al., 2007; Kim et al., 2005), mechanisms of autoregulation (Huang et al., 2000; Moens et al., 1998; Sadl et al., 2003) and chromatin remodeling (Chapter 3).
Figure 1.4. The hindbrain segmentation gene MafB/Kreisler/Krml1 and its mutants. (A) Diagram of a normal mouse chromosome 2. (B) Chromosome 2 containing the kr inversion. In the kr allele, the inversion breakpoints (bpt) are located 51kb 5’ of the MafB (bpt2) and 45kb upstream of the Neuronatin (Nnat) (bpt1) transcriptional start sites (Cordes and Barsh, 1994; John et). Previous transgenic analyses have located an r5 and r6 specific enhancer 21 kb 5’ upstream of MafB transcriptional start (Sadl et al., 2003). PRE1 denotes the region containing the Polycomb response element: region A is translocated with the MafB gene, while region B is left behind in the kr inversion. The disrupted PRE1 region is denoted as ΔPRE1 in the kr inversion allele. (C) MafB protein is a basic leucine zipper domain (bZIP) transcription factor. kr^{enu} refers to the chemically induced point mutation at Asn^{248} to a Ser while va^{p37} refers to the nonsense mutation in zebrafish resulting in a premature stop prior to the transactivation domain. (D-F) Dorsal views of whole mount RNA in situ hybridization on 14 somite stage embryos. MafB was expressed in r5 and r6 of normal embryos (D). (E) MafB expression could not be detected in kr/kr embryos by RNA in situ hybridization. (F) In kr^{enu} homozygotes, MafB RNA was present in a smaller domain that was roughly the width of one rhombomere rather than two.
We have found that vHNF1 (variant Hepatocyte Nuclear Factor 1) directly regulates initiation of MafB via a single binding site found in the r5-r6 specific enhancer S5 which is 19 kb upstream of the MafB start site. This binding site appears to be conserved in chick and zebrafish. With the use of transgenic experiments, we have shown that although vHNF1 is essential for MafB induction, it alone is not sufficient to drive r5-r6 expression, nor is the S5 enhancer able to maintain later expression patterns. FGFs have been implicated in posterior hindbrain patterning and may contribute to the regulation of early expression of MafB. It is of note that no vHNF1 homolog has been found to play a similar role in fly segmentation.

**MafB mouse mutants**

There are several mutants of MafB in the mouse including a lethal targeted knock-out (mafb-), and a lethal GFP knock-in (Mafb-) (Blanchi et al., 2003; Moriguchi et al., 2006). Of particular interest to this work are the original kr inversion and the chemically induced kr<sup>emu</sup> mutants (Figure 1.4). The classical X-ray induced kr mutant is a near-perfect chromosomal inversion of 2.9 Mbp (3.87 cM) which does not disrupt the gene (Cordes and Barsh, 1994). Homozygous mutants do not express MafB in the hindbrain (by in situ hybridization), although later regions of expression such as the kidney and eye lens are normal. Adult mice are viable, fertile, deaf and exhibit the hallmark circling behaviour from which the mutant derives its name (Deol, 1964). In the kr<sup>emu</sup> mutant N-ethyl-N-nitrosourea (ENU) was used to induce a point mutation that changes a conserved asparagine (amino acid position 248) found in the DNA binding domain of all known bZIP proteins into a serine (Figure 1.4C). MafB is still expressed in an r5+r6-like domain in a slightly condensed hindbrain, but the protein can no longer bind DNA (Sadl et al., 2003). Death occurs shortly after birth most likely due to apnea from a failure in the rhythmic centre (Blanchi et al., 2003).

MafB is one of the earliest genes to be expressed segmentally in the developing embryo and is subject to complex and dynamic mechanisms of gene regulation. Unlike the Hox genes that lie in large clusters and are subject to cross- and inter-regulatory mechanisms, MafB is more easily accessed and manipulated making it ideal for studying transcriptional events as they happen in the ever changing environment of the developing hindbrain. These include early induction models such as RA signaling, as well as
maintenance events such as the mechanisms of chromatin remodeling. As with the discovery of segmentation genes, the initial molecules identified in vertebrates as important in the regulation of segmentation were orthologs of genes previously identified in *Drosophila*. Two major gene families emerged as key players in both fly and mouse segmentation. These were the Polycomb and trithorax groups of proteins which act to regulate gene expression via chromatin remodeling. However, the precise mechanism by which they govern segmentation genes remained largely a mystery.

B) Mechanisms of Chromatin Remodeling in the Developing Vertebrate

Transgenic experiments have shown that signaling molecules and transcription factors play a key role in establishing early identities in development via specific and complex regulatory regions. More recently it has become apparent that chromatin remodeling plays a pivotal role in stabilizing and maintaining these transcriptional programs. Because of multilevels of cross talk, feed forward and feed back events that take place in regulating segmentation genes in development, any perturbations, however small might result in significant consequences. The PcG/trxG proteins buffer gene expression from many such disturbances in this highly dynamic and sensitive system.

**Integrated mechanisms of epigenetics in gene silencing**

Large and complex eukaryotic genomes are highly compacted into chromatin, presenting the cell with several obstacles to DNA replication and gene transcription. Epigenetics refers to heritable alterations in gene expression that are not caused by changes in the DNA sequence. Epigenetic modifications that affect gene expression include DNA methylation, post-translational histone modifications, binding by PcG/trxG complexes, and RNA-mediated gene silencing. These epigenetic mechanisms often work in concert to open and close chromatin, facilitating access to transcriptional and DNA replicating machinery. Interplay between different epigenetic marks is essential for development, cell differentiation, protection against foreign genomes, and interpretation of intrinsic and environmental signals (Jaenisch and Bird, 2003). Deregulation of these mechanisms can lead to a variety of human diseases, most notably cancer (recently
reviewed by (Kokavec et al., 2008; Li and Zhao, 2008; Liu et al., 2008; Vaissiere et al., 2008).

**DNA methylation and gene transcription**

DNA methylation of the vertebrate genome is heritable; these epigenetic marks are preserved over multiple cycles of cell division, and are therefore considered a form of cellular memory. The major target of DNA methylation in animals occurs on the C5 position of cytosine bases in CpG dinucleotides. In mammals, there are two different types of DNA methyltransferases. Early in development, DNA methyltransferase 3a and 3b (Dnmt3a and Dnmt3b) add de novo methyl groups and establish DNA methylation patterns. DNA methyltransferase 1 (Dnmt1) is responsible for maintaining methylation through many cycles of cell division (Bird, 2002). Vertebrate methylation is dispersed over much of the genome, a pattern referred to as global methylation which in humans account for 70-80% of all CpG dinucleotides (Ehrlich et al., 1982). However, these data mask the fact that in mammals DNA methylation patterns vary in time and space. For example, early in mouse development, methylation levels decline to ~30% of normal levels (Kafri et al., 1992). By the time of implantation, de novo methylation restores levels back to normal.

Reducing the levels of DNA methylation by either genetic manipulation or by treatment with DNA methyltransferase inhibitors such as 5-aza-deoxycytidine (5-AzadC) lead to suppression of some forms of tumors in mouse models of cancer. In contrast, formation of other tumor types is enhanced by low levels of DNA methylation. It is clear that normal patterns of DNA methylation are crucial to the health of mammals. In humans, mutations in Dnmt3b are found in ICF syndrome, a developmental defect characterized by hypomethylation of pericentromeric repeats (Ehrlich et al., 2008).

Another striking feature of DNA methylation patterns in vertebrates is the presence of CpG islands. These largely unmethylated GC-rich regions are located at the 5’ ends of many genes and have been shown to play a key role in development. For example, the interleukin-2 gene must lose CpG methylation at its promoter region in order to be turned on during T-cell differentiation (Bruniquel and Schwartz, 2003). Additionally, during X-chromosome inactivation of female mammalian embryos, de novo
methylation of many CpG islands on the X chromosome is essential for the absolute silencing of genes on the inactivated chromosome whereas the active X chromosome in the same cell remains resistant (Wolf et al., 1984). Another example where DNA methylation plays an important role in development is genomic imprinting. Here, DNA methylation interferes with gene expression by hindering the binding of transcription factors to GC-rich sequence motifs at the promoters of target gene. Evidence for this comes from the study of the H19/Igf2 locus in mice. The CCCTC-binding factor (CTCF) is associated with transcriptional domain boundaries and acts to insulate a promoter from the influence of neighbouring enhancers (Bell et al., 1999). A maternally derived copy of the Igf2 gene is silenced due to CTCF binding between its promoter and a downstream enhancer. Contrastingly, the CpG-rich CTCF binding domain of the paternal locus is heavily methylated, preventing binding, thereby allowing activation of Igf2 by the downstream enhancer (Lewis and Murrell, 2004).

The above examples would imply that CpG methylation is a default state of the genome and CpG islands are excluded from global methylation in order for active gene transcription to be possible. However, it has become clear that DNA methylation patterns may be determined by modifications on the chromatin. For example, in both the fungus Neurospora crassa and the plant Arabidopsis thaliana, there is strong evidence that DNA methylation depends on methylation of histone H3 on lysine 9 (H3K9me) (Jackson et al., 2002; Tamaru and Selker, 2001). Furthermore, loss of the mammalian PcG protein Enhancer of Zeste Homolog 2 (EZH2), a histone lysine methyltransferase (HKMT) specific for the H3K27 residue, was shown to cause CpG demethylation at EZH2-target gene promoters. Vire and colleagues also used chromatin immunoprecipitation (ChIP) to show that DNMT binding to EZH2-repressed genes depended on the presence of EZH2, suggesting that it acts as a recruitment platform for DNMTs (Vire et al., 2006). Other chromatin associated factors involved with DNA methylation include the SWI/SNF-like proteins. For example, the SNF2-like proteins ATRX in humans and Lsh2 in mouse were shown to have effects on global DNA methylation patterns (Dennis et al., 2001; Gibbons et al., 2000).
Chromatin structure and histone modifications

The enormous DNA double helix is never found without its many associated proteins within the nuclei of plant or animal cells. There are several specially evolved packaging and architectural proteins that fold DNA into highly condensed chromatin. Chromatin consists of DNA, histones and architectural proteins such as heterochromatin protein 1 (HP1). Each core nucleosome is made up of 147 bp of genomic DNA wrapped around two copies each of the four core histones: H2A, H2B, H3 and H4. The linker histone, H1 binds the intervening DNA (Figure 1.5). Generally, DNA is wound tightly around the core histones, a process maintained by dynamic association with the architectural proteins as well as a variety of posttranslational modifications that regulate chromatin accessibility and gene expression. Histone modifications that affect chromatin accessibility and function include histone acetyltransferases (HATs) (Roth et al., 2001), histone deacetylases (HDACs) (Marks et al., 2004), histone methyltransferases (Trievel, 2004), and histone demethylases (Klose et al., 2006). In general, histone acetylation opens up chromatin allowing for gene transcription, while histone methylation is associated with closed chromatin and gene silencing (Figure 1.6). Exceptions to this include H3 methylation on lysine 4 (H3K4me) and lysine 36 (H3K26me) which are hallmarks of actively transcribed chromatin (Millar and Grunstein, 2006). Other histone modifications that add to the complexity of chromatin regulation include poly-ADP ribosylation (Althaus et al., 1994), phosphorylation (Loury and Sassone-Corsi, 2004), ubiquitination (Zhang, 2003), and sumoylation (Shiio and Eisenman, 2003).

Recently evidence has been mounting that chromatin, like the nucleus that contains it is organized into well-defined domains termed chromosome territories (Cremer et al., 1982). The positioning of chromosome territories and gene loci are tightly regulated and specific for tissue and cell-type as well as developmental stage (Misteli, 2005). There are two classes of chromatin: The transcriptionally active and decondensed euchromatin and the highly condensed inactive heterochromatin. Euchromatin is characteristically associated with active histone markers such as trimethylated lysine 4 on H3 (H3K4me3) and acetylated lysine 9 on H3 (H3K9ac). Correspondingly, heterochromatin like the inactive centromeric regions are commonly marked by H3K9me3 and H4K20me3 (Millar and Grunstein, 2006).
Figure 1.5. DNA is tightly condensed into chromatin. Chromatin consists of DNA, histones and architectural proteins. Nucleosome structure consists of 147 bp of genomic DNA (brown cords) wrapped around two copies each of the four core histones: H2A, H2B, H3 and H4. The linker histone, H1 (orange) binds the intervening DNA and heterochromatin protein HP1 (red) is bound to modified histone H3 tails (H3K9me3).

Figure 1.6. Histone tail modifications. Histone H3 and H4 tail methylation (top) and acetylation (bottom) at lysine residues. Generally, histone acetylation opens up chromatin allowing for active gene transcription (green), while histone methylation is associated with closed chromatin and gene silencing (red). Exceptions include H3K4me and H3K26me which are hallmarks of actively transcribed chromatin.
Facultative heterochromatin refers to silenced regions that can be activated and is typically marked by H3K27me3. These characteristic histone signatures have been used to investigate the global chromatin states in different cell types, most notably in ES cells, to profile potentially active and inactive genes en masse. Surprisingly, while many genes did indeed show clear active or inactive histone modification signatures consistent with their transcription state, some, which were known to be important for ES cell differentiation showed bivalent marks (Gan et al., 2007). Thus, it has been argued that a bivalent histone signature may identify key cell fate determining genes that are poised to be activated or inactivated.

Therefore, histone methylation marks are generally, but not exclusively associated with active and inactive chromatin, while other genes, which will be discussed in more detail below, can exhibit bivalent marks; that is histone signatures associated with both active and inactive chromatin.

**Chromatin modifications in normal physiological processes and disease**

The subject of chromatin remodeling mechanisms involved in normal physiological processes and disease will be addressed in detail in later sections. Of particular interest is the impact of such processes in the development and degeneration of the nervous system such as when heterochromatin is rearranged to be more condensed during neuronal differentiation of embryonic stem (ES) cells (Meshorer, 2007), or the global changes that occur in chromatin structure of brain nuclei in Alzheimer’s patients that alter the ability of neurons to generate sufficient levels of gene products to maintain normal cortical function (Lukiw and Crapper McLachlan, 1990).

In psychiatric disorders, the methylated DNA-binding proteins MeCP2 and MBD2 appears to play a role in the pathology of GABAergic dysfunctions in mice and psychiatric disorders such as schizophrenia, bipolar disorder and autism in humans. The glutamic acid decarboxylase67 (GAD67) and Reelin genes are both reduced in the prefrontal cortex of schizophrenia patients and in methionine-induced mouse models of schizophrenia (Dong et al., 2005). The alterations occur at the promoters of GAD67 and Reelin where they are hypermethylated, resulting in the elevated recruitment of MeCP2 and MBD2. This is particularly relevant as DNA methylation is often linked with histone
deacetylation and vice versa, and Dong and colleagues were able to reverse promoter hypermethylation by treating mouse models with the HDAC inhibitor valproate, an effective atypical antipsychotic, resulting in an increase of GAD67 and Reelin transcription (Dong et al., 2007). Once again, the intricate relationship between histone modifications on DNA methylation at discrete gene loci, which in turn influence specific transcriptional programs, is highlighted.

In humans, the relationship between DNA methyltransferase inhibitors and human Reelin and GAD67 expression has been recently explored in NT-2 neuronal precursor cells. The data showed that Reelin and GAD67 mRNAs are induced in the same dose- and time-dependent manners when treated with the DNA methyltransferase inhibitors doxorubicin (a potent inhibitor of DNMT1), 5-AzadC and zebularine (nucleoside analogs that when incorporated into DNA, forming covalent bonds with DNA methyltransferase, thus inhibiting their function), further lending support to the concept that inhibition of DNA methylation and/or downregulation of DNMT1 protein leads to reactivation of human Reelin and GAD67 gene expression (Kundakovic et al., 2007). Similar experiments were later carried out using HDAC inhibitors that showed a reduction in DNMT1, DNMT3A, and DNMT3B protein levels as well as DNMT enzyme activity, again leading to induction of the Reelin and GAD67 mRNAs (Kundakovic et al., 2009).

The above examples clearly illustrate how chromatin structure has the potential to become a rewarding target for altering transcriptional states that lead to changes in cell state in normal development and diseased models. Focus has centered on HDAC inhibitors and DNA demethylating agents in particular as potential drugs for both human neurological disorders and cancer, with much effort devoted to deciphering the exact mechanisms of action.

**Cellular memory and transcriptional silencing by the Polycomb group proteins**

Early in development, the identity and function of a given cell type is determined by the transcriptional programs of key regulators such as the *Hox* genes that are established in response to signaling molecules such as retinoic acid. In order for development to proceed normally, these gene expression profiles, once established, are
maintained faithfully over many generations of cell division. The Polycomb Group (PcG) proteins, a widely conserved group of transcriptional repressors best known for their maintenance of \textit{Hox} gene expression patterns from fruit flies to humans, play key roles in setting transcriptional states. Together with the counteracting trithorax group (trxG) proteins that maintain active states of transcription, the PcG/trxG proteins act as large protein-protein complexes that are recruited to their target genes by specific DNA sequences termed Polycomb response elements (PRE). These two families of proteins form the molecular basis for cellular memory by faithfully maintaining transcriptional states determined early in embryogenesis (Figure 1.7). This cellular memory program involves changes to the chromatin, via a complex interplay of some of the mechanisms mentioned in previous sections including: histone modifications, DNA methylation and alterations in nuclear architecture.

**Genetic identification of the Polycomb and trithorax group of proteins**

In the course of studying the function of the \textit{Hox} family of segmentation genes, many trans-acting regulators were isolated. One of these, Polycomb (Pc) was identified and genetically analyzed by Pam and Ed Lewis to be an important repressor of the \textit{bithorax} complex (Lewis, 1978). \textit{Pc} heterozygotes displayed typical homeotic transformations, specifically; males have additional sex combs on the second and third legs, whereas wild-type males only carry sex combs on the first leg. Homozygotes were embryonic lethal, but showed classic homeotic transformations of the most posterior abdominal segment. Later, it was discovered that the PcG proteins were required for the maintenance of \textit{Hox} repression rather than being involved with the spatial/temporal establishment of \textit{Hox} gene expression. The antagonistic positive acting regulators, the trxG proteins were also identified in genetic screens performed in the fly. Thus there were now two opposing groups of trans-acting regulators in the maintenance of \textit{Hox} gene expression, the PcG and trxG proteins (Kennison, 1995).

PcG orthologs identified in mammals such as \textit{Bmi1} and \textit{M33} also proved to be key regulators of \textit{Hox} gene expression and showed homeotic transformations of the vertebrae (Core et al., 1997; van der Lugt et al., 1996). These orthologs proved to be highly conserved, both genetically and functionally. Furthermore, PcG genes have been found in two other model organisms. In the nematode \textit{Caenorhabditis elegans}, PcG
members were identified in screens for maternal effect sterile (mes). Mutants were shown to be involved in X-chromosome silencing in the hermaphrodite germ line (Fong et al., 2002). In the flowering plant Arabidopsis thaliana, PcG genes were identified in several genetic screens involving developmental processes. For example, CURLY LEAF (CLF) was identified to be a homolog of the fly PcG gene Enhancer of zeste (E(z)) and was shown to be necessary for the stable repression of floral homeotic genes (Goodrich et al., 1997). Thus, the PcG complexes and quite possibly the basic molecular mechanisms by which they act appear to be preserved in all eukaryotes.

**Chromatin modifications by Polycomb repressive complexes 1 and 2**

PcG proteins are biochemically classified into two major Polycomb repressive complexes, PRC1 and PRC2. Members of PRC2 are conserved between plants and animals, whereas PRC1 proteins have only been found in flies and vertebrates. This would imply conservation but also specificity in the establishment of a system of cellular memory. In the fly and vertebrates, the maintenance of silencing requires the sequential binding and action of both complexes. PRC2 is first recruited to silencing targets, and modifies histones by methylating H3K27 and/or H3K9. These typically silent chromatin markers then recruit PRC1 components which then induce structural changes to the chromatin in order to maintain the silenced state.

In Drosophila, four core proteins have been consistently purified from PRC2: the SET domain histone-methyltransferase Enhancer of Zeste (E(Z)), the WD40 protein Extra Sex Combs (ESC), the histone-binding protein Nurf-55/Caf1, and Suppressor of Zeste12 (SU(Z)12) (Levine et al., 2004) (Table 1.1). The E(Z) subunit of PRC2 contains a SET domain that confers histone lysine methyltransferase (HKMT) activity to the complex (Rea et al., 2000), specifically methylating H3K27, a marker for silencing strongly associated with PREs (Cao et al., 2002). Because the PRC1 member Pc specifically interacts with methylated H3K27 via its chromodomain, it is thought that E(Z) methylation of H3K27 recruits PRC1 to PREs. ESC is a small protein that contains five WD40 repeats and serves as a platform for protein-protein interactions. ESC physically interacts with both Nurf-55/Caf1 and E(Z) in all model systems analyzed, making it structurally important in maintaining the complex. Also, although HKMT activity is attributed to E(Z), it alone cannot methylate H3K27, but requires the presence
### Table 1.1. Core PcG genes in model systems

<table>
<thead>
<tr>
<th>Protein Domains</th>
<th>D. melanogaster</th>
<th>H. sapiens</th>
<th>M. musculus</th>
<th>A. thaliana</th>
<th>C. elegans</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PcG DNA-binding proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pho</em></td>
<td>Zinc finger Pleiohomeotic</td>
<td>YY1</td>
<td>YY1, YY2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Phol</em></td>
<td>Zinc Finger Pleiohomeotic-like</td>
<td>YY2</td>
<td>YY1, YY2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Psq</em></td>
<td>BTB-POZ domain Pipsqueak</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Dsp1</em></td>
<td>HMG domain Dorsal Switch Protein1</td>
<td>HMGB2</td>
<td>HMGB2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PRC2 core complex (E(z)-Esc complex)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Esc</em></td>
<td>WD40 domain Extra sex combs Enhancer of zeste</td>
<td>EED</td>
<td>EZH2</td>
<td>Ezh1/Enx2</td>
<td>FIE</td>
</tr>
<tr>
<td><em>E(z)</em></td>
<td>SET domain</td>
<td>EED</td>
<td>EZH1</td>
<td>Ezh2/Enx1</td>
<td>CLF</td>
</tr>
<tr>
<td><em>Su(z)12</em></td>
<td>Zinc finger VEFS box RING Suppressor of zeste12</td>
<td>SUZ12</td>
<td>mSUZ12</td>
<td></td>
<td>FIS2</td>
</tr>
<tr>
<td><strong>PRC1 core complex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pc</em></td>
<td>Chromodomain Polycomb</td>
<td>DCBX2/HPC1/M33 CBX4/HPC2 CBX8/HPC3/PC3</td>
<td>Cbx2/M33 Cbx4/MPC2 Cbx6 Cbx8/MPC3 Cbx7 NPCD</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ph</em></td>
<td>Zinc finger SAM Polyhomeotic</td>
<td>PHC1/EDR1/HPH1 PHC2/EDR2/HPH2 PHC3/EDR3/HPH3 BMI1</td>
<td>Edr1/Mph1/Rae2 8 Edr2/Mph2 Bmi1 Rnf110/Zfp144/Mel18</td>
<td></td>
<td>SOP-2</td>
</tr>
<tr>
<td><em>Psc</em></td>
<td>Zinc finger HTH domain Posterior Sex Combs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>dRing/Sce</em></td>
<td>RING finger dRing/Sex combs extra</td>
<td>RNF2/RING1B RING2</td>
<td>Ring1/Ring1a</td>
<td>Ring1/Ring1b</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from (Allis et al., 2007; Schwartz and Pirrotta, 2007); Cavalli lab: The Polycomb and Trithorax Page, 2009 [online])
of both ESC and SU(Z)12, which may have modulating functions (Müller et al., 2002). This modulating activity was suggested by the differential targeting by various isoforms of the ESC homolog Embryonic endoderm development (EED), a component of mammalian PRC2 (Kuzmichev et al., 2004). Alternate translational start sites are responsible for the different isoforms which in turn determines the specificity of mammalian PRC2 for either H1K26 or H3K27 methylation (Kuzmichev et al., 2004). However the biological relevance of H1K26 methylation is unclear. Like ESC, SU(Z)12 is an essential component for HKMT activity of the PRC2 complex. Additionally, SU(Z)12 together with Nurf55 (which contains six WD40 repeats that physically interact with ESC) are sufficient for nucleosome binding, anchoring the HKMT active E(Z) component to target chromatin (Nekrasov et al., 2005).

There are also a number of PRC2-associated factors that are not found in all preparations of PRC2. The Rpd3 gene encodes a HDAC that has been implicated in PcG-mediated silencing (Tie et al., 2001). RPD3 was found to be required for silencing by PRC2 at the fly Ubx PRE in vivo and mutations in rpd3 enhance PcG phenotypes, though on their own, they do not show the typical homeotic transformations (Tie et al., 2003). Similarly, Sir2 and its human homolog SirT1 were found in a subgroup of PRC2 complexes to have HDAC activity and play an important role in modifying chromatin states in development and cancer (Furuyama et al., 2003; Kuzmichev et al., 2005). Although not a part of all PRC2 complexes, the last two examples illustrate how powerful a partnership of HKMT and HDAC activities can be in reinforcing transcriptionally silent chromatin.

PRC1 also contains four core PcG proteins: Polycomb (PC), Polyhomeotic (PH), Posterior Sex Combs (PSC) and Ring1 (dRing2/SCE) (Table 1.1) (Francis et al., 2001). Levine and colleagues purified a similar complex containing the homologs of these four core proteins in mammalian cells (Levine et al., 2002). The PC protein contains a chromodomain at its amino-terminus that has homology to HP1, the fly protein required for heterochromatin formation (Paro and Hogness, 1991), and has been shown to bind methyl groups at H3K27 and H3K9 (Bannister et al., 2001; Fischle et al., 2003). As mentioned above, it is thought that PRC1 is recruited to silencing targets via PC recognition of methylation marks left by HKMT activity of PRC2.
The gene product of *Psc* has an amino terminus that is conserved with the vertebrate oncogene *bmi1* and the tumor suppressor gene *mel-18*. This domain consists of a ring finger motif, generally thought to mediate protein-protein interactions. Similarly, *ph* which is duplicated in the fly into two highly homologous genes *ph-p* and *ph-d* share a conserved single zinc finger and a SAM domain. This domain is also involved in both homo- and heteromeric protein-protein interactions. These findings would suggest a way by which these large protein-protein complexes might form and be localized to their silencing targets.

Unlike many of the other members of PcG proteins, dRING1 was not initially recognized as a PcG protein by mutational analysis in *Drosophila*, but was identified because of the remarkable conservation of PcG genes between flies and mammals. Targeted mutation of RING1a in the mouse led to classic homeotic transformations, and through sequence comparison was found to be the ortholog of the PRC1 member *Sex combs extra (Sce/dRING)* (Satijn and Otte, 1999). In mammals, Ring1A and Ring1B have been associated with H2A ubiquitylation on the inactive X chromosome, a silencing mark dependent on Ring1 proteins (de Napoles et al., 2004; Fang et al., 2004). More recently, Cao and colleagues used *bmi1*-(mammalian PSC homolog) and *Ring1A*-null mouse embryonic fibroblasts (MEF) to show that they were both required for H2AK119 ubiquitylation *in vivo* which is an important mark downstream of H3K27 methylation for *Hox* gene silencing (Cao et al., 2005).

Aside from covalent histone modifications, the exact sequence of events in which PRE-bound PRC1s interact with the promoter to prevent transcription is still largely unknown. One important aspect of efficient transcriptional activation involves the efficient escape of RNA polymerase from the promoter. In reporter constructs with the hsp26 heat shock promoter, Dellino et al. showed that PcG binding pauses the transcriptional machinery at the promoter rather than preventing initial binding of RNA polymerase and associated transcription factors (Dellino et al., 2004). Additionally, *in vitro* purified PRC1 core complex has been shown to counteract ATP-dependent nucleosome remodeling and compact chromatin structure (Francis et al., 2004; Francis et al., 2001).
**PcG silencing during mammalian development has consequences of aberrant activation**

Much of the previous discussion has focused on molecular mechanisms of PcG proteins in *Hox* gene regulation in *Drosophila*. It has already been mentioned that mutations in murine PRC1 members also exhibit homeotic transformations of the axial skeleton. For example, loss of the mouse PC protein M33 results in a deregulation of *Hox* genes and the consequential appearance of additional vertebrae (Core et al., 1997). Interestingly, these skeletal malformations worsen when treated with RA, indicating an increased RA sensitivity which suggests a role for M33 in regulating access to *Hox* gene RAREs during development (Core et al., 1997). *M33* null mutant mice also display severe combined immunodeficiencies (Raaphorst, 2005), demonstrating the important role PcG proteins play in maintaining cell identity. This is particularly apparent in self-renewing cell populations such as blood cells, as most blood-cell lineages are characterized by very specific cell-type transcription programs. In PcG knockout mice, B- and T-cell precursor populations are produced normally, indicating that once again PcG proteins are not involved with establishing lineage-specific gene expression programs, but are required for locking in the lineage choice throughout development. Over time, any anomalies in cellular differentiation are amplified and thus can develop into leukemias.

Besides control of *Hox* gene expression programs and cell-lineage restriction, PcG proteins play a major role in controlling proliferation in vertebrates. The murine *bmi1* gene was originally identified as an oncogene, that induces mouse lymphomagenesis (van Lohuizen et al., 1991). *Bmi1* protein controls the cell regulators p16INK4a and p10ARF and together with Mel-18 (another homolog of PSC) are negative regulators of the INK4a-ARF locus (Jacobs et al., 1999). While the importance of PcG function becomes apparent over time, an assay that would allow more direct and immediate scoring of aberrant transcriptional programs would provide a powerful tool for analyzing PcG/trxG function.

Several parallels can be drawn between PcG maintenance of *Hox* gene expression patterns and other epigenetic phenomena, such as X inactivation and imprinting in mammals. In female XX cells, PcGs are involved with a series of chromatin
modifications required for silencing the inactive X chromosome (Heard et al., 2004). Specifically, the maintenance of a silent paternal X chromosome in extra-embryonic tissues is dependent on components of PRC2, like the ESC homolog Eed, or the E(Z) homolog Enx1, which mark the inactive X with H3K27me3, H3K9me2 and H4K20me. Similarly, Bmi1 is required for H2AK119ub1 ubiquitinylation (Heard et al., 2004; Hernandez-Munoz et al., 2005). In mouse, PRC2 has also been implicated in silencing of the paternal allele of imprinted genes (Umlauf et al., 2004). As in previous cases, this phenomenon involves the histone marks H3K27me3, and H3K9me2. Interestingly, in eed mutant mice, several imprinted loci from six unlinked imprinting clusters showed biallelic expression (Mager et al., 2003). The DNA methylation status at these loci was altered in these mutant mice suggesting the importance of DNA methylation by PcGs in the maintenance of imprinting marks.

**Transcriptional regulation by trxG proteins**

During cellular differentiation, the identity and function of a given cell type set up by transcriptional programs must be maintained. The previous section addressed how key genes are kept in the “OFF” state by the PcG proteins. The trxG family of proteins maintains master regulatory gene expression programs in the “ON” state. In general, the epigenetic mechanisms involved in maintaining active gene expression patterns appear to be more complex than those of the PcG proteins.

The trxG family of proteins includes a large number of members with many diverse functions. In addition to activating genes by promoting its transcription, trxG proteins sometimes activate a given gene by repressing a second gene. Some trxGs function via regulating chromatin structure in opposition to the mechanisms used by the PcGs such as placing opposing covalent modifications on histones. Other trxG proteins function as part of the transcription machinery. Because trxG proteins appear in a variety of complexes with diverse functions and roles, they require more levels of control and regulation.
Figure 1.7. The concept of cellular memory.
A simplified schematic illustration highlighting the role of trxG and PcG complexes in maintaining heritable transcriptional states of gene expression.
Genetic identification of trithorax group proteins

Like the founding members of the PcG proteins, the trxG proteins were first identified as modulators of Hox gene expression in Drosophila, where mutants displayed dramatic homeotic transformations. Early members of the trxG family trithorax (trx), absent, small or homeotic 1 and 2 (ash1/2), and female-sterile homeotic (fsh) were identified in screens for mutations that mimic loss-of-function Hox mutations in Drosophila (Kennison, 1995). Mutations of the founding member trx caused partial transformation of halteres to wings (due to decreased Ubx); first to second legs (due to decreased Scr); and posterior abdominal segments to more anterior identities (due to decreased abdA and AbdB). Other screens focused on suppressors of Pc (Su(Pc)) mutations, that is, seeking genes that when reduced, would counteract the reduced silencer (Kennison and Tamkun, 1988). The ATP-dependent chromatin remodeler brahma (brm) came out of such screens. There are currently at least 16 trxG members with a variety of functions including chromatin remodeling, histone methyltransferase, and transcription factor (Table 1.2).

TrxG proteins in mammals

Virtually all Drosophila trxG proteins have functional counterparts in mammals. For example, the TRX homolog mixed-linkage leukemia (MLL) was first discovered because translocations at chromosome 11q23 caused lymphoblastic (ALL) or myeloid (AML) leukemias (Ziemin-van der Poel et al., 1991). Furthermore, heterozygous MLL mutant mice displayed transformations of the axial skeleton due to failure to maintain active transcription of Hox genes and a posterior shift of their anterior borders of expression (Yu et al., 1995). As with MLL mutations, aberrations in other mammalian trxG proteins are often associated with cancers. For example, brahma-related gene 1 (BRG1), the human ortholog of brm physically interacts with the retinoblastoma tumor suppressor protein (Dunaief et al., 1994). Accordingly, mice heterozygous for brg1 are prone to a number of cancers (Bultman et al., 2000). Thus, trxG proteins play highly conserved functions in gene expression and development from flies to mammals.
### Table 1.2. Biochemical functions of trxG proteins

<table>
<thead>
<tr>
<th>Known function</th>
<th>Protein Domain</th>
<th>D. melanogaster</th>
<th>H. sapiens</th>
<th>M. musculus</th>
<th>Yeast</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP-dependent chromatin remodeling</td>
<td>SNF2 HELICc Bromodomain</td>
<td>Brahma (BRM)</td>
<td>BRG1/HBRM</td>
<td>BRG1/SMARCA4</td>
<td>Swi2/Sf2 Sth1</td>
</tr>
<tr>
<td>(SWI/SNF complex)</td>
<td>BRIGHT</td>
<td>OSA</td>
<td>BAF250</td>
<td>ARID1B</td>
<td>Swi1/Adr6</td>
</tr>
<tr>
<td></td>
<td>SWIRM SAINT</td>
<td>Moira (MOR)</td>
<td>BAF155 BAF170</td>
<td>SMARCC1 SMARCC2</td>
<td>Swi3 Rsc8</td>
</tr>
<tr>
<td></td>
<td>CHROMO</td>
<td>Kismet (KIS)</td>
<td>CHD7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histone methyltransferases</td>
<td>PHD SET</td>
<td>Trithorax (TRX)</td>
<td>MLL1 MLL2 hSET1</td>
<td>WBP7 MLL1</td>
<td>Set1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Absent, small or homeotic 1 (ASH1)</td>
<td>hASH1</td>
<td>ASH1L</td>
</tr>
<tr>
<td>Mediator subunits</td>
<td>Kohtalo (KTO)</td>
<td>TRAP230/MED12L</td>
<td>Srb8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Skuld (SKD)</td>
<td>TRAP240/THRAP2</td>
<td>Srb9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcription factor</td>
<td>Zinc finger BTP/POZ</td>
<td>Trithorax-like (TRL)/ GAGA factor</td>
<td>BTBD14B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth factor receptor</td>
<td>Breathless (BTL)</td>
<td>FGFR3</td>
<td>Sallimus (SLS)</td>
<td>Titin</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>PHD SPRY</td>
<td>ASH2</td>
<td>hASH2L</td>
<td>ASH2L</td>
<td>Bre2</td>
</tr>
</tbody>
</table>

Adapted from (Allis et al., 2007; Schwartz and Pirrotta, 2007); Cavalli lab: The Polycomb and Trithorax Page, 2009 [online])
Diverse functions of trxG proteins

The process of eukaryotic transcription involves complex and numerous interactions and multiple stages. It is thought that transcriptional activation requires appropriate opening of the DNA helix, binding of sequence specific transcriptional activating factors, recruitment and assembly of the many components of the general transcription machinery at the promoter, the efficient escape of RNA polymerase from the promoter, and efficient elongation of RNA polymerase through the gene. Different members of the trxG proteins can regulate transcription at any of these steps including acting as transcription factors as in the case of Trithorax-like (Trl). Mostly, trxGs affect transcriptional activation by affecting nucleosome placement and chromatin structure which are known to inhibit binding of transcription factors and general transcription machinery.

One major group of trxG proteins belongs to the ATP-dependent chromatin remodeling proteins that are related to the SWI2/SNF2 complex found in the yeast Saccharomyces cerevisiae. The fly trxG gene brm and its mammalian counterparts Brg1 and Hbrm were found to be highly related to SWI2/SNF2 (Kwon et al., 1994; Tamkun et al., 1992; Wang et al., 1996). These genes were first identified in yeast screens for genes involved in mating-type switching (switch (swi)) and sucrose-fermentation (sucrose-nonfermenting (snf)) and then were found to be required for the activation of many inducible yeast genes by counteracting chromatin repression (Holstege et al., 1998; Kruger et al., 1995; Sudarsanam and Winston, 2000). SWI/SNF proteins act in large complexes that depend on ATP hydrolysis from the SWI2/SNF2 subunit to bind to nucleosomal DNA (Cote et al., 1994; Imbalzano et al., 1994; Kwon et al., 1994), while other members of the complex mediate interactions with regulatory proteins and the chromatin substrate in order to alter the structure of chromatin or positioning of nucleosomes (Phelan et al., 1999). Another trxG gene found to be a key member of the ATP-dependent remodeling complex is moira (mor), and human homologs of both mor and brm were found to interact directly to form the functional core of the SWI/SNF chromatin remodeling complex (Phelan et al., 1999).

The SWI/SNF remodeling complex utilizes the energy from ATP hydrolysis in order to alter chromatin structure such that transcription factors and transcriptional
machinery can access DNA that is normally bound by nucleosomes. Depending on the gene, there are a number of mechanisms by which it does this: nucleosome sliding (moving nucleosome along DNA leaving an exposed region), histone exchange (exchanging a variant histone for normal one), nucleosome eviction (leaving a large open region of DNA) and altering nucleosome structure (creating a loop on the surface of the nucleosome allowing access) (Wang, 2003). OSA, another evolutionarily conserved trxG protein occurs only in a subset of SWI/SNF complexes, while others contain the Polybromodomain protein, suggesting that this may be a targeting system for a particular SWI/SNF complex to specific set of genes (Mohrmann and Verrijzer, 2005).

Another class of trxG proteins is distinguished by the presence of SET domains that confer HKMT activity. TRX and ASH1 for example share homology with the PcG proteins Suppressor of variegation (SU(VAR)3-9) which methylates H3K9 and E(Z) which methylates H3K27 (Levine et al., 2004; Rea et al., 2000; Ringrose et al., 2004). As one might expect, TRX and ASH1 and their mammalian orthologs can methylate H3K4, a covalent histone mark tightly associated with active genes from yeast to flies to mice (Beisel et al., 2002; Byrd and Shearn, 2003). Thus, it would appear that histone tail methylation is also important for maintenance of active transcription.

In addition to HKMT activity, there is also evidence that TRX and its human counterparts MLL1, MLL2, and hSET1 are involved in histone acetylation as well. In humans, MLL is associated with the MOF acetyltransferase, which acetylates lysine 16 of histone H4 (H4K16ac), another histone mark linked with gene activation (Dou et al., 2006; Dou et al., 2005). Histone acetylation works synergistically with H3K4 methylation to regulate active states. Furthermore, acetylation can also prevent methylation at H3K9 and H3K27 the general histone marks of repression.

Other mechanisms by which trxG proteins direct active transcription include interacting with the “Mediator” complex as in the case with the trxG members *skuld (skd)* and *kohtalo (kto)* (Janody et al., 2003). This is a large complex that functions at the interface between gene-specific activator proteins and recruitment of the pre-initiation complex that contains RNA polymerase II (Lewis and Reinberg, 2003). Other trxG proteins such as human Ring3 are nuclear protein kinases implicated in cell cycle
progression and leukemogenesis, though its exact mechanism of action is still unclear (Denis and Green, 1996). Another trxG gene *sallimus (sls)*, also known as *titin* may be involved in regulating higher-order chromatin structure by affecting chromosome condensation and segregation, thus helping to maintain the integrity and elasticity of the sarcomere (Machado and Andrew, 2000).

It is clear that the trxG proteins consist of a large family of proteins with diverse roles and multiple modes of action. It is possible that these dynamic regulatory processes are integrated as with the PcG proteins. For instance, the histone marks left by TRX or ASH1 might regulate or target the activities of other trxG members that are involved in ATP-dependent chromatin remodeling (such as *brm*). Consistent with this model, BRM and other subunits of the SWI/SNF complexes contain bromodomains and chromodomains that are able to interact with covalently modified histone tails. This model provides a reasonable mechanism by which histone modifications can be maintained over many cycles of cell division, perpetuating an open chromatin state.

**Polycomb and Trithorax Response Elements**

As with transcription factors, such as the segmentation gene products that are recruited to specific cis-regulatory sequences in the early stages of development, the PcG/trxG proteins are recruited to Polycomb and trithorax response elements (PRE/TRE). These specialized elements range in size from ~300 to over 1000 bp, and their sequence requirements are poorly understood and appear to be heterogeneous. Several studies have shown that PRE/TREs recruit PcG and trxG binding prior to the action of the initiating factors and maintain active (trxG) or silenced (PcG) transcriptional states over many cell divisions long after the transient initiator has gone (Cavalli and Paro, 1998a; Maurange and Paro, 2002; Orlando et al., 1998; Poux et al., 1996). As described previously, the PcG and trxG proteins were first identified by their effect on the expression of Hox segment selector genes in *Drosophila*. It was not surprising then that the first PREs were identified to be cis-acting sequences important in the regulation of the fly Hox gene clusters *Bithorax (BX-C)* and *Antennapedia (ANTP-C)*. These sequences were found to prevent the activation of Hox reporter genes outside of their normal expression domains (Chan et al., 1994; Simon et al., 1993). Because this repression relied on the function of PcG proteins, these elements were termed Polycomb response
elements. This was confirmed by chromatin immunoprecipitation (ChIP) analyses at these loci that showed highly enriched binding of Pc and other PRC1 subunits at their PREs (Orlando et al., 1998; Strutt et al., 1997; Strutt and Paro, 1997). More recently, ChIP in both tissue culture cells and in fly embryos have shown that subunits of both PRC1 and PRC2 as well as other PRC associated proteins such as Pleiohomeotic (PHO) have been found to bind specifically at PREs of Hox and other target genes (Cao et al., 2002; Klymenko et al., 2006; Papp and Muller, 2006). PREs are known to act at a distance to repress expression from different promoters, for example from within a Hox gene cluster. This has been shown in reporter assays where a single bithoraxoid (bxd) PRE from the Ultrabithorax (Ubx) Hox gene can block transcriptional activation from a number of promoters (Dellino et al., 2004; Sengupta et al., 2004).

**Defining a PRE/TRE**

Although several studies have predicted the presence of several hundred PcG/trxG regulated loci in the *Drosophila* genome, only a handful of functionally tested PRE/TREs have been found in the fly (Chinwalla et al., 1995; Zink and Paro, 1989). To date, no PRE/TREs have been identified in either vertebrates or plants. Defining a conserved sequence in order to design search algorithms for PRE/TREs has proved challenging. One complication comes from the wide range in size of functionally characterized PREs; these can range from a few hundred to several thousand base pairs. PRE/TREs also contain several binding sites for several different DNA binding proteins and there are usually more than one PRE/TRE at any given target locus (Muller and Kassis, 2006; Ringrose and Paro, 2001; Ringrose and Paro, 2007; Schwartz and Pirrotta, 2007).

Despite the complexity and the existence of other binding motifs in some PREs, two consensus sequence motifs, occurring in clusters have been identified to play a role in all *Drosophila* PRE function. The first motif (GCCAT) is bound by the PHO and Pleiohomeotic-like (PHOL) proteins. PHO and PHOL function in PcG targeting and are found associated with both PRC1 and PRC2 (Poux et al., 2001). *Drosophila* PREs also contain clusters of GAGAG motifs, which bind GAGA factor (GAF) (encoded by the *Trithorax-like* (Trl) gene) and Pipsqueak (PSQ). These are also known to associate with PcG complexes and appear to function in both gene silencing and activation (Bejarano and Busturia, 2004; Hodgson et al., 2001; Horard et al., 2000; Strutt et al., 1997). Binding
motifs for Zeste and dorsal switch protein 1 (DSP1) have also been found in several PREs (Dejardin and Cavalli, 2004; Dejardin and Cavalli, 2005b; Hagstrom et al., 1997).

Functionally, all PRE/TREs display pairing-sensitive repression of adjacent reporter genes that is dependent on the presence of PcG/trxG proteins, and can recruit PcG/trxG proteins to its site of transgene insertion creating a novel PcG/trxG binding site. Each functionally identified PRE/TRE fulfills at least 1 of the 3 criteria tested by these respective assays: First, the maintenance assay tests a PRE’s ability to maintain a pattern of expression of a LacZ reporter gene throughout development in a PcG/trxG dependent manner. Second, the pairing-sensitive assay tests whether a PRE can repress expression of a mini-white reporter. Finally, the salivary gland assay tests a PRE’s ability to create additional PcG binding sites on polytene chromosomes at the site of transgene insertion. This last test can be visualized using immuno-fluorescent in situ hybridization (iFISH) techniques (Mihaly et al., 1998). Note that although insulators and silencers may be able to silence reporter genes in these assays, as in the case of the imprinting element from the mouse H19 locus (Lyko et al., 1997; Pannell et al., 2000a), they do not function in a PcG/trxG-dependent manner, nor do they recruit PcG/trxG binding.

There have been several attempts to define specific consensus sequences that make up a PRE in Drosophila, leading to bioinformatics algorithms designed to predict novel PREs (Brown et al., 1998; Fiedler and Rehmsmeier, 2006; Fritsch et al., 1999; Hur et al., 2002; Kassis et al., 1989; Mihaly et al., 1998; Ringrose et al., 2003; Saurin et al., 2001; Strutt et al., 1997). Generally, these algorithms search for the presence of short motifs such as those specific for the DNA binding proteins PHO, GAF/Psq and Zeste which have been found in all PREs identified. However, they also occur with some frequency in many random DNA sequences. More specialized search engines scan DNA sequences surrounding developmentally regulated genes for the presence of GAF and PHO binding motifs either alone or in combination with one another at variable distances using established PREs as a training set, though even these failed to identify all of the functionally defined PREs (Fiedler and Rehmsmeier, 2006; Ringrose et al., 2003). The most recent data would suggest that PRE/TREs are far more complex than clusters of binding motifs alone, but may rely on other factors such as local nucleosome structure, chromatin state as well as the presence of non-coding RNAs (Ringrose and Paro, 2007).
Reprogramming of PRE/TREs throughout development

In flies, it is clear that PRE/TRE elements play critical roles in the differentiation of embryonic stem (ES) cells as well as maintaining the pluripotency of ES cells due to their potential to switch between activated and silenced states at a given promoter (Boyer et al., 2006; Bracken et al., 2006; Ringrose and Paro, 2007). Previous sections have described in detail how aberrations in PcG/trxG function can lead to misregulation of key developmental genes and ultimately to various cancers. In follows then, that mutations within PRE/TREs of these target genes will also lead to a loss of identity and the ability to switch between proliferation and differentiation (Ringrose and Paro, 2007; Valk-Lingbeek et al., 2004).

PRE elements lock early developmental programs into cellular memory, but as transitions occur throughout development, there must be ways to overcome PcG silencing and switch target genes to an active state. A variety of intriguing models have been proposed. However, none have been completely tested, nor has a single model been shown to apply to all PREs. One example can be found in Drosophila where transcriptional activity was observed to circumvent PcG-mediated silencing. That is, silencing is prevented when transcription occurs through PREs and an active state is maintained for a given transgenic reporter construct (Schmitt et al., 2005). By transcribing through the PRE, the chromatin is most likely remodeled such that an active state is generated, for example by the loss of repressive histone methylation or acquisition of active histone acetylation. Without the appropriate histone silencing marks, a pre-loaded PRE-PRC1 complex cannot establish a stable silenced state. Another possible mechanism may involve trxG proteins preventing PcG binding, thus opening target genes to an active state. More recently, non-coding RNAs (ncRNA) have been proposed to also play a role in overriding PcG-silencing.

In mammals, epigenetic states also switch in the course of development. For example, in mouse ES cells, many PcG target genes become derepressed during retinoic-acid-induced differentiation (Boyer et al., 2006). Additionally, other signaling pathways including activated integrins, T-cell receptor and platelet derived growth factor (PDGF) cause an export of the PRC2 complex from the nucleus to the cytoplasm, where it
interacts with the signal response pathway (Hobert et al., 1996; Su et al., 2005; Witte et al., 2004).

**Identifying PRE/TREs in mammals**

To date, no functional PRE/TREs have been defined in mammals though it is clear that PcG/trxG proteins are involved with all the processes previously described in *Drosophila*. Besides regulating Hox gene expression profiles, PcG/trxG complexes also maintain stem cell pluripotency in both embryonic and adult stem cells. Examples include involvement with maintaining hematopoietic stem cells (HSCs) and blood cell lineage development in the bone marrow. Mice mutant in PRC1 genes *bmi1/mel-18*, *mph1/rae28*, and *M33* all suffer from various defects in the hematopoietic system, including increased cell proliferation and reduction in B and T cells (Zardo et al., 2008). PcG complexes are also required for neural stem cell (NSC) renewal and differentiation as indicated by the neuronal defects observed in *bmi* mouse mutants (Bruggeman et al., 2005). Clearly PcG/trxG complexes play key roles in global developmental transitions such as switching between pluripotency and self-renewal programs to differentiating into specific cell lineages. Therefore the question remains, are mammalian PcG/trxG complexes recruited to PRE/TREs? If so, are their mechanisms of action comparable to those in the fly?

Searching for mammalian PRE/TREs has proved challenging for a number of reasons. First, aside from functional tests, it has been difficult to define a consensus sequence of what makes a PRE/TRE, even in flies. Predictive programs such as jPREdictor used the known PREs from *Drosophila* as a training set, then scanned the fly genome for potential PRE/TREs that had multi-motif clusters of known PRE motifs such as GAF, Dsp1, Zeste and PHO binding motifs (Fiedler and Rehmsmeier, 2006; Ringrose et al., 2003). However, this program failed to detect 3 of the 11 test PREs in the original training set (*iab-8, Scr, Mcp*), with non PRE/TREs like *white* and *hsp23* scoring higher. Also, being bound by PcG/trxG complexes does not necessarily define a PRE/TRE. For example, three genome-wide studies using ChIP and DNA adenine methyltransferase identification (DamID) which analyze protein binding profiles have found several hundred PcG/trxG-enriched loci in the *Drosophila* genome (Negre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). However, the binding profiles from these three studies
only showed partial overlap. Although these inconsistencies may be a reflection of genuine differences in PcG/trxG binding at different developmental stages or cell types, it was curious that few functionally confirmed PRE/TREs showed enrichment of PcG/trxG binding in these studies (Ringrose et al., 2003). In fact, only 6-27% of bound sites were previously predicted to contain PRE/TREs based on the presence of clustered consensus motifs (Negre et al., 2006; Ringrose et al., 2003; Schwartz et al., 2006; Tolhuis et al., 2006).

Similar complications exist in defining PRE/TREs in mammalian systems. This task is made more difficult by the lack of mammalian homologs to several of the sequence-specific DNA binding proteins (e.g. PHOL, Zeste, and PSQ) that act on PRE/TREs in *Drosophila*. However, three recent genome-wide PRC1/2 binding profile studies have been performed in mouse (Boyer et al., 2006), and human (Lee et al., 2006) ES cells, and in human embryonic fibroblasts (Bracken et al., 2006). All three studies used ChIP on microarrays (ChIP-on-chip) and have identified over 500 sites that are enriched with several PcG proteins as well as other histone signatures that are linked to PcG binding (e.g. H3K27 methylation). However none of these loci have yet been functionally analyzed for PRE function. Support for the validity of some of these target sites has come from analyses that showed binding of multiple PRC2 components to the same region and deregulated expression of some EED or SUZ12 bound targets in cells lacking the respective PcG gene (Boyer et al., 2006; Lee et al., 2006). Lee et al. noted that several of the loci bound by the PcG protein SUZ12 overlap with many of the highly conserved noncoding elements (HCNE) previously identified and predicted to hold PRE/TREs (Lee et al., 2006; Woolfe et al., 2005). Unfortunately, on a global scale, the overlap between HCNE regions and the PcG binding is low. Furthermore, Bernstein et al. showed that the H3K27 methylation patterns typically produced by the PcG protein EZH2 were enriched at these loci, though the highest peaks of methylation were not in HCNEs, thus precluding HCNEs themselves from containing authentic PREs (Bernstein et al., 2006). These studies have also been limited by the scarcity of ChIP quality antibodies.

Further frustrating the search for mammalian PREs was that the majority of PcG binding sites identified in vertebrates reside within 1 kb of promoters, while in flies only
30% are within 2 kb of the promoter and the rest are several 10s of kbs away. These apparent differences reflect a sampling bias caused by the predominant use of undifferentiated vertebrate ES cells, in contrast to the pseudo-differentiated cell types and embryos examined in the fly. Also, by using mouse and human microarrays that, with one exception, span maximally -8 to +2 kb of known or predicted promoters, these studies may have missed critical regions where PRE/TREs lie (Boyer et al., 2006). Nevertheless, these studies point to a prevalence of PcG target sites near promoters rather than the more distant location of most validated PREs in Drosophila. These observations could be reconciled by the existence of two distinct types of PcG target sites: PcG loaded promoters and classically defined PREs. PcG loaded promoters appear to be involved in regulating cell differentiation. This is reflected in the genome-wide Drosophila ChIP and DamID studies, where many of the PcG targets at promoters that are involved in cellular differentiation were excluded (Negre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). The rationale is that these studies were carried out not in ES cells, but in embryos where cellular differentiation has occurred and PcG proteins were removed from their target sites at the promoters upon activation (Chen et al., 2005).

In Drosophila, the best characterized PRE/TREs are located within the Drosophila Hox clusters. Hox gene translocation experiments in the mouse also suggest the presence of PRE/TREs (Kmita et al., 2002; Roelen et al., 2002; Spitz et al., 2005). Therefore, efforts have focused on inspecting PcG targets in human and murine Hox gene clusters. Functional analyses of PREs in the fly and on activity dependent nucleosomal positioning in the mouse Hoxd4 gene suggest that PcG complexes and PRC1, in particular, may bind to discrete sequence platforms, while H3K27me3 marks cover large silenced genomic regions broadly (Beisel et al., 2007; Kobrossy et al., 2006). Studies of PcG targets within entire Hox clusters suggest that the cell type analyzed and its differentiation state play a significant role in whether PcG complexes blanket entire Hox clusters, bind specific peaks embedded within these blanketed regions or do not bind at all. For example, analyses of SUZ12, EZH2 and H3K27me3 binding showed broad blanketing of the 200 kb region containing the HoxA gene cluster in mouse F9 cells and human Ntera2 cells, which resemble F9 cells, but not in the SW480 and MCF7 adult tumor lines (Squazzo et al., 2006). Another study observed binding of the PC homolog...
CBX8, EZH2 and H3K27me3 to posteriorly expressed genes of the \textit{HoxA} cluster (\textit{Hoxa8-a13}) in undifferentiated, but not in RA treated human NT2/D1 cells, in which \textit{HoxA} genes show faithful regulation upon differentiation induced by RA; while \textit{Hoxa1-5} genes showed the reverse occupancy (Bracken et al., 2006). In this latter study, binding was seen in sharper peaks while other larger regions are blanketed. Lee et al. also saw blanketing of \textit{Hox} genes in human embryonic stem cells (Lee et al., 2006). Clearly vertebrate \textit{Hox} gene clusters provide a valuable but challengingly complex substrate for the identification and analysis of individual PRE/TREs. Indeed, none of the potential PcG target sites have been validated as true PREs. Therefore, segmentally expressed genes like \textit{MafB}, which are not part of clusters or subject to complex cross-regulatory programs, may prove a more suitable target for the identification and analysis of novel PRE/TREs.
Goal of this Thesis

The general purpose of this thesis is to further the understanding of how anterior-posterior identity is established and stabilized in a developing system. I have used the compartmentalized structure of the mouse embryonic hindbrain, where transcriptional events can be observed as they occur on a cell by cell basis, as a model to study the regulation of one gene, MafB/Kreisler. By focusing on the regulation of a single early acting segmentation gene I have found that a rhombomere-specific enhancer, S5, can interpret the powerful posteriorizing signal retinoic acid as well as transcription factors such as vHNF1. By using a variety of approaches I have shown that although S5 can drive vHNF1-dependent early expression in r5 and r6, there appears to be other activating and repressing transcriptional networks involved, as well as mechanisms of chromatin remodeling that govern accessibility to inductive enhancers such as S5.

I have identified and characterized a Polycomb responsive element, PRE1 that affects accessibility to S5, the first of its kind outside of Drosophila. I have utilized assays and mutants in both flies and mice to show that MafB is governed by effectors of chromatin remodeling from histone deacetylases to Polycomb mutants, and that these are evolutionarily conserved. My goal was first to show that PRE1 is indeed a functional PRE as defined in the fly, and then to fully interrogate the region for characteristics that define a PRE in vertebrates. Ultimately, we would like to understand the underlying mechanisms of chromatin remodeling that govern accessibility to genes in a developing embryo. This includes the DNA consensus motifs and/or chromatin context that define a functional PRE, the protein complexes recruited, how PREs interact with the genes they control, and what reprogramming regimes might occur throughout development. Overall, this thesis strives to illustrate that in a highly dynamic system such as the developing embryonic hindbrain, there is a complex and well orchestrated system of events that include interpretation of signaling molecules, classical mechanisms of transcriptional control as well as multiple layers of regulation at the level of the chromatin, including PREs which act as entry sites for PcG/trxG chromatin remodeling.
Chapter 2: The vHNF1 homeodomain protein establishes early rhombomere identity by direct regulation of MafB expression

Abstract

The early transcriptional hierarchy that subdivides the vertebrate hindbrain into eight segments, the rhombomeres (r1-r8) is largely unknown. The MafB (Kreisler, Krml1, Val) gene is one of the earliest genes to be segmentally expressed in r5 and r6 and is essential for the development of the posterior hindbrain. We have identified the S5 regulatory element that directs early MafB expression in the future r5-r6 domain in 0-10 somite stage embryos. Variant Hepatocyte Nuclear Factor 1 (vHNF1/HNF1β/LF-3B) is transiently expressed in the r5-r6 domain of 0-10 somite stage embryos and a vHNF1 binding site within this element is essential, but not sufficient for r5- and r6-specific expression. Thus, early inductive events that initiate MafB expression are clearly distinct from later acting ones that modulate its expression levels. This site and some of the surrounding sequences are evolutionarily conserved in the genomic DNA upstream of the MafB gene among species as divergent as mouse, humans, and chickens. This provides the first evidence of a direct requirement for vHNF1 in the initiation of MafB expression, suggests the role of vHNF1 is evolutionarily conserved, and indicates that vHNF1 collaborates with other transcription factors, which independently bind to the S5 regulatory region, to establish the r5-r6 domain.
Introduction

Its visible physical compartmentalization and segment-specific gene expression patterns make the vertebrate hindbrain an ideal model for studying how positional information is laid down in development. The embryonic hindbrain is transiently subdivided into seven to eight physical compartments, known as rhombomeres (r1-r8), along its anterior-posterior axis. Each rhombomere represents a true embryonic compartment according to their morphology, genetic identity, and future cell lineage (Fraser et al., 1990; Lumsden and Krumlauf, 1996). Many gene expression patterns, including those of Hox genes, the most widely studied segmentation genes, are limited to specific subsets of rhombomeres (Murphy et al., 1989; Wilkinson et al., 1989). The MafB segmentation gene encodes a basic domain leucine zipper (bZIP) transcription factor, which is expressed in, and is essential for the formation of the fifth and sixth rhombomeres (r5 and r6) (Cordes and Barsh, 1994). Initial expression of MafB precedes the appearance of rhombomeric boundaries and other r5- and r6-restricted gene expression, and is required directly for the regulation of Hoxa3 in r5 and r6 and Hoxb3 in r5 (Cordes and Barsh, 1994; Manzanares et al., 1999a; Manzanares et al., 1997).

What determines the expression pattern of MafB itself is still not clear, but its early segment-specific expression is likely to occur via complex integration of anterior and posterior signals. Studies in the chick first showed that MafB expression, like that of Hox genes, was sensitive to posteriorizing signals such as retinoic acid (RA) and fibroblast growth factors (FGFs) (Grapin-Botton et al., 1998; Marin and Charnay, 2000; Papalopulu et al., 1991). Later analyses of the role of FGFs in zebrafish indicated that signals emanating from the future r4 domain also affected MafB expression either directly or indirectly. For example, in zebrafish transplantation studies, r4 appeared to induce MafB expression in a Fibroblast growth factor dependent manner (Maves et al., 2002; Walshe et al., 2002).

In order to identify upstream regulatory mechanisms that establish MafB expression domains, we have searched for rhombomere-specific regulatory elements in the MafB genomic region. Here we have isolated an early acting rhombomere-specific
enhancer of MafB. We have begun to dissect the regulatory elements required for MafB expression and have identified a single variant Hepatocyte Nuclear Factor 1 (vHNF1) binding site that alone, is essential, but not sufficient for r5-r6-specific expression. Recent data in zebrafish show that loss of function of vHNF1 leads to hindbrain defects similar to those observed in animals lacking MafB/Valentino and that MafB/Valentino expression is missing in vHNF1 mutants (Sun and Hopkins, 2001). In the zebrafish neural tube, vHNF1 is expressed coincidentally with that of MafB in the future r5 and r6 domains (Barbacci et al., 1999; Coffinier et al., 1999a; Ott et al., 1991; Sun and Hopkins, 2001). Here we show that vHNF1 is transiently expressed in the r5-r6 domain of the mouse in 0-10 somite stage embryos. Furthermore, vHNF1 protein is highly conserved between zebrafish and mouse (Sun and Hopkins, 2001), and the vHNF1 regulatory site in the MafB enhancer is also highly conserved across species as divergent as chickens, mice, and humans. Together these observations suggest a conserved role for vHNF1 in establishing aspects of hindbrain patterning in collaboration with other transcription factors.

**Materials and Methods**

**Mice**

Transgenic mice were generated by pronuclear injection into fertilized oocytes from super-ovulated FVB mice. For all constructs expression was analyzed in multiple founder embryos and/or in independent transgenic lines, as detailed in Table 2.1. Embryos were stained for LacZ as previously described (Whiting et al, 1991). Transgenic lines were maintained by crossing with CD1 mice. RNA *in situ* analyses were performed as described (Conlon and Rossant, 1991).

**Transgenic DNA constructs**

Overlapping BamHI and SacI digested fragments from the upstream region of MafB were isolated from genomic P1 and BAC clones obtained from Research Genetics, Inc. All fragments were cloned into the filled in NotI site of the hsp68LacZ expression construct, in which a minimal hsp68 promoter can direct LacZ expression in the presence of additional enhancers. For microinjection, SalI digestion followed by DNA
electrophoresis separated inserts from vector sequences, and inserts were purified using GeneClean beads (Bio101, Vista, CA). The specific mutation in the vHNF1 site of the S5 fragment was generated using the QuickChange mutagenesis system (Stratagene, LaJolla, CA). Multimerized binding sites were generated from doublestranded oligonucleotides of the vHNF1 binding site (5’- GCTATCCCGCTAATCATTAACCCAATGAGCATGTC-3’). The number of vHNF1 binding sites inserted into the 2xvHNF1 and 5xvHNF1 constructs was determined by DNA sequencing.

**Immunofluorescence**

An affinity purified anti-full-length MafB antibody was used in immunofluorescence experiments (Sadl et al., 2002; Sadl et al., 2003). Immunohistochemistry using this anti-MafB antibody was performed on 12 μm cryosections, which had been postfixed in 4% paraformaldehyde, rinsed in PBS, and blocked in 10% goat serum. Sections were incubated overnight with affinity purified primary antibody at a 1:5000 dilution at 4°C, rinsed three times in PBS, and incubated at room temperature for 1 hour at room temperature with Cy3-donkey anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA, USA). Slides were rinsed three times in PBS, mounted with VectaShield mounting medium (Vector Laboratories, Inc., Burlingame, CA), and examined microscopically.

**Whole mount in situ RNA hybridizations**

Whole mount in situ RNA hybridizations were performed as previously described in Conlon and Rossant, 1992 (Conlon and Rossant, 1992). Hoxb1 and MafB probes were generated as previously described (Frohman et al., 1993; Cordes and Barsh, 1994). The vHNF1 probe, generously provided by Dr. Silvia Cereghini, was prepared as described (Cereghini et al., 1992).

**Electrophoretic mobility shift assays**

vHNF1 from zebrafish, generously provided by Drs. Sun and Hopkins was in vitro transcribed and translated using the TNT reticulolysate system. DNA binding reactions using in vitro transcribed and translated vHNF1 and either the radiolabeled S5-2kb SacI to XbaI fragment or S5-vHNF1 doublestranded oligonucleotide (5’-GCTATCCCGCTAATCATTAACCCAATG) performed at room temperature for 20
minutes in a 20μl reaction containing 20 mM Hepes, pH7.9, 1mM EDTA, 20 mM KCl, 5 mM DTT, 4 mM MgCl2, 5% glycerol, 200 μg/ml poly (dIdC). Reactions were resolved by DNA electrophoresis on 4% non-denaturing polyacrylamide gels and analyzed by autoradiography. Mock TNT reactions served as negative controls and caused no mobility shifts. For competition experiments, double stranded oligonucleotides or gel purified DNA fragments were added in 3-100 fold molar excess of the radiolabeled probes at the start of the binding reactions. The oligonucleotide sequences used in these experiments were:

S5-vHNF1 (5’- GCTATCCCGCTAATCATTAACCCAATGAGCATGTC-3’),
S5-mutvHNF1 (5’-GCTATCCCGCGTATCAGGAACCCAATG-3’),
PE-vHNF1 (5’-TGTGATTAATGATCTACAGTTA-3’), and
DS34vHNF1 (5’-TGTGGTGTATGAGGTACAGTTA-3’) (Costa and Grayson, 1991; Tronche et al., 1999; Tronche et al., 1990). DNA fragments were obtained by digestion of the S5 region with XbaI and SacI, and purification of 1.3 and 2.0 kb fragments by gel electrophoresis and the Qiagen gel extraction kits.

Cell Culture and Transfection

The (PE)2-CAT, (S5)2-CAT, and (S5mut)2-CATreporter plasmids were generated similarly by inserting dimers of the PE, S5, and S5mut-vHNF1 sites described above into the BamHI site of pBLCAT2 (Seto et al., 1990). Sequences of all constructs were verified. The vHNF1 expression construct was generously provided by Gerald Crabtree. Cos-7 cells were maintained in Dulbecco’s modified Eagle medium (DMEM, high glucose) with 10% fetal bovine serum, penicillin (100U/ml) and streptomycin (100μg/ml) and transfected using the Lipofectamine Plus Reagent (Life Technologies, Inc., Rockville, Md), in keeping with manufacturer’s suggestions. Cells were seeded at 1.75 x 10^5 per 35 mm dish 18 to 24 hr prior to transfection. Total DNA concentration per dish was 1.5 μg, with the reporter plasmid accounting for 425 – 550 ng, the internal control (pCMVb-gal) 1/20 (75ng) of the total, and the expression plasmid or equivalent empty expression plasmids (pcDNA3). Each transfection mixture included 13μl Plus Reagent and 5 μl Lipofectamine Reagent diluted in Opti-MEM I (Life Technologies, Inc.,
Rockville, Md). CAT (Roche Diagnostics, Laval, Quebec, Canada) and β-galactosidase assays were performed 36 hours post-transfection. β-gal activity was used to normalize CAT activity.

**Sequence analyses**

Sequence alignments were performed using standard BLASTn searches and the ENSEMBL database initially, and later refined with the ECR browser (http://ECRbrowser.dcode.org) (Ovcharenko et al., 2004). Predicted vHNF1 transcription factor binding sites were identified via TRANSFAC database searches (http://www.gene-regulation.com) and rVISTA (http://rvista.dcode.org/) (Knuppel et al., 1994; Loots and Ovcharenko, 2004; Wingender et al., 1996). These analyses revealed homology between mouse chromosome 2: 160446648-160445202 (AL591665 nucleotides 119252-119836), human chromosome 20: 38768899-38769453 (AL035665 nucleotides 111036-111620), and chick chromosome 20: 4452924-4453176 (contig 18.362 nucleotides 8492-8745) of the current ENSEMBL database sequences.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Regulatory Elements</th>
<th># with r5-r6 expression</th>
<th>Transgenic mice generated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S5</td>
<td>4/4</td>
<td>1 transgenic line</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 founder embryos</td>
</tr>
<tr>
<td></td>
<td>S5-2kb</td>
<td>4/4</td>
<td>4 founder embryos</td>
</tr>
<tr>
<td></td>
<td>S5-0.9kb</td>
<td>0/2</td>
<td>2 transgenic lines</td>
</tr>
<tr>
<td></td>
<td>S5mut</td>
<td>0/5</td>
<td>3 transgenic lines</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 founder embryos</td>
</tr>
<tr>
<td></td>
<td>2x vHNF1</td>
<td>0/3</td>
<td>3 transgenic lines</td>
</tr>
<tr>
<td></td>
<td>5x vHNF1</td>
<td>2/5</td>
<td>3 transgenic lines</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 founder embryos</td>
</tr>
</tbody>
</table>

Table 2.1. Transgenic embryos with r5 and r6 specific LacZ expression
The regions of S5 used to drive expression of the hsp68LacZ reporter in transgenic lines are shown. The number of transgenic lines and founder embryos with LacZ expression in r5-r6 and the number of transgenic lines and transgenic founder embryos examined for each construct are indicated. vHNF1 site is indicated by white boxes. Point mutation is marked by an asterisk.
Results

Identification of an early-acting rhombomere-specific enhancer for *MafB*

The classical *kreisler (kr)* mutation was caused by a submicroscopic chromosomal inversion that does not disrupt the *MafB* coding sequence or transcription unit, but abolishes *MafB* expression in the prospective r5 and r6 region and leads to ectopic expression of *MafB* in r3 (Cordes and Barsh, 1994; Sadl et al., 2003). A simple model that might account for the loss of r5-r6-specific expression would be that a rhombomere-specific enhancer required for early r5-r6-specific *MafB* expression is located 5' of the transcriptional start and is thus removed by the *kr* inversion. To determine whether rhombomere-specific enhancers for *MafB* were located 5' of the *MafB* transcriptional start site, we began to examine genomic P1 clones that spanned the *kr* centromere distal inversion breakpoint and the *MafB* transcription unit (Cordes and Barsh, 1994) (Figure 2.1A). We subcloned overlapping BamHI and SacI fragments from the P1-222 and P1-730 clones into a reporter construct containing the minimal hsp68 promoter and a *LacZ* reporter gene, and began to test these for their ability to direct rhombomere-specific *LacZ* expression. In this manner, we identified a 3.2 kb SacI fragment from P1-222 referred to as S5, which directed *LacZ* expression in the prospective r5-r6 region in 0-10 somite stage embryos in one transgenic line and 3 transgenic founder animals (Figure 2.1, Table 2.1). *LacZ* expression disappeared completely from the hindbrain by the 10 somite stage. To determine whether rhombomere-specific *LacZ* expression coincided with *MafB* expression we stained embryos for *LacZ* and used immunofluorescence to localize *MafB* protein. In sections of 6-8 somite stage embryos, the domain containing light and slightly spotty *LacZ* expression coincided with the *MafB* expressing r5-r6 region (Figure 2.1C-F). Some faint *LacZ* speckles were scattered at the dorsal edges of rhombomeres 7 and 8 and sometimes in the trunk mesoderm. Because this *LacZ* speckling was only seen in two lines containing the S5 element or subdomains of S5 and has been seen in animals transgenic for the hsp68 minimal promoter-*LacZ* reporter alone, it is likely not caused by the S5 element. Thus, the S5 enhancer directs *LacZ* expression to the prospective r5 and r6 region between the 0-10 somite stages and is probably involved in initiating rhombomere-specific *MafB* expression. Surprisingly, the S5 element is located between the inversion breakpoint and the transcriptional start site of *MafB* and is not removed by
Figure 2.1. Identification of a regulatory element required for early MafB expression

(A) Schematic diagram of the genomic region 5' of the MafB transcription unit. The position of the S5 element is shown. P1 clones, P1-222 and P1-730, span the centromere-distal kresler inversion breakpoint (Cordes and Barsh, 1994). (B) A restriction map of the 3,244 bp Sacl fragment, S5, which directs LacZ expression in r5 and r6 of 0-10 somite stage embryos. The regions of highest identity between mouse and human is indicated by a light gray box, and the location of the vHNF1 site is indicated by a dark gray oval.

(C-F) The S5 element directs LacZ expression in r5 and r6 domain. MafB expression as detected by whole mount RNA in situ hybridization in an 8 somite stage embryo (C). This expression pattern coincides with that of LacZ directed by the S5-hspLacZ reporter in the future r5 and r6 (D). (E-F) Co-staining of longitudinal hindbrain sections with anti-MafB antibody (E) and LacZ (F). Immunofluorescence detects MafB protein (E) in r5 and r6 of an 8 somite stage transgenic S5-hspLacZ embryo, which also exhibits LacZ (F).
the kr inversion. Thus, a position effect caused by the kr inversion rather than the complete removal of rhombomere-specific enhancers may alter MafB expression in the hindbrains of kr/kr and kr/+ mice. Of course, it is possible that other rhombomere-specific regulatory elements that synergize with the S5 element to initiate MafB expression or act slightly later at the 10-20 somite stages to maintain and refine MafB expression may have been removed by the kr inversion.

**An evolutionarily conserved consensus site is present in the MafB enhancer**

Upon examination of subdomains within the 3.3 kb S5 region in transgenic mice as well as by electrophoretic mobility shift assay (EMSA) profiling binding of *in vitro* transcribed and translated (TNT) RAR and RXR, we localized the rhombomere-specific enhancer further to a 2 kb SacI-XbaI fragment, S5-2kb (Figure 2.3A and Table 2.1). We then used cross-species comparisons to search for conserved regions in this 2 kb segment (Ovcharenko et al., 2004). Standard Blastn searches initially identified a 121 bp stretch with 85% identity to regions upstream of the human MafB gene on chromosome 20 (Wang et al., 1999). Subsequent sequence alignment extended the highly conserved domain to a 584 bp region which shared 70.5% identity between mouse and human genomic regions located 16.5 and 17.6 kb, respectively, upstream of the MafB gene. Even more compellingly, a 176bp region 13kb upstream of the MafB gene in chicken genomic DNA shared 40% identity with both the S5 mouse and homologous human region (Figure 2.2A). Interestingly, the S5 region that shows high homology to chick is distinct from the 121 bp region with the highest homology to human genomic DNA. Overall, the conserved nature of this rhombomere-specific enhancer and the evolutionarily conserved role of MafB suggest that at least some regulatory mechanisms governing early MafB expression will be shared across species.

Previous transplantation experiments in the chick had implicated retinoic acid in the regulation of early MafB expression (Grapin-Botton et al., 1998; Marin and Charnay, 2000). However, analysis of this 2 kb sequence for the presence of candidate hindbrain patterning transcription factor binding sites could not detect any predicted consensus binding sites for RAR/RAR or RAR/RXR dimers. However using EMSA with *in vitro* transcribed and translated RAR and RXR proteins, we found that RAR/RAR homodimers bind to S5-2kb with high affinity, which was later localized to a 600bp region (probe 3).
A.  

mouse AATTATCAGA GAGTTCTACC GCAGTCCATT TTTGCTCTTG TCCCTTTTCC ATGTTAGGCGT  
Human AATTATCAGA GAGTCTACC AGCTGCTGCT TTTGCTCTTG TCCCTTTTCC ATGTTAGGCGT  

mouse TATAATGCC TTTGCTCTTG AGAGCTCTAC TAGACCCAG TGGAGCTCTAC TATAATGCC  
Human TATAATGCC TTTGCTCTTG AGAGCTCTAC TAGACCCAG TGGAGCTCTAC TATAATGCC  

mouse AGTCAAACTAC TTTAATTGTTA TGTCATGACT GTAGAGAGCTG CGCTCATAAA ATTGGCCAAA  
Human AGTCAAACTAC TTTAATTGTTA TGTCATGACT GTAGAGAGCTG CGCTCATAAA ATTGGCCAAA  

mouse GTTAAGAAAA AGAGATATCAAC AATGTGAGTT TTTGAGGCGA AGTCTCTAGGC  
Human GTTAAGAAAA AGAGATATCAAC AATGTGAGTT TTTGAGGCGA AGTCTCTAGGC  

mouse CAGCTCTCAA ATAGCCAGCT-- GAGGGCAAGT TTACTTCTCA AGCTGCTCAG CAGCTCTCAA  
Human CAGCTCTCAA ATAGCCAGCT-- GAGGGCAAGT TTACTTCTCA AGCTGCTCAG CAGCTCTCAA  

mouse G---TCCTC--A GGGAGGACCA CAGCCCTCCAG TCACCCTCAG GAGGGCAAGT CAGCTCTCAA  
Human G---TCCTC--A GGGAGGACCA CAGCCCTCCAG TCACCCTCAG GAGGGCAAGT CAGCTCTCAA  

mouse GCCCTTCAA TGGAGGAGA CTGGCCACAC TATTTATATAA GAGGGCAAGT CAGCTCTCAG  
Human GCCCTTCAA TGGAGGAGA CTGGCCACAC TATTTATATAA GAGGGCAAGT CAGCTCTCAG  

mouse GATTAGAGAAT CTGTTCTTAA ACCATTATAAC AGATGGGACCC GAGGAGAGG  
Human GATTAGAGAAT CTGTTCTTAA ACCATTATAAC AGATGGGACCC GAGGAGAGG  

mouse GATGCAAAAG GAAGCTGTCT CTCACTGCTTG CTCAAGAGCC TCAAGACAAA CCACTGTTG  
Human GATGCAAAAG GAAGCTGTCT CTCACTGCTTG CTCAAGAGCC TCAAGACAAA CCACTGTTG  

mouse GATTAGAGAAT CATCTCTCTTC TTCTTTAACC TACATGGAG GAGAGCTGGG GGGAGGGCTT  
Human GATTAGAGAAT CATCTCTCTTC TTCTTTAACC TACATGGAG GAGAGCTGGG GGGAGGGCTT  

mouse TATATTCCTTAC AAAA  
Human TATATTCCTTAC AAAA  

B.  

mouse GCTAATCTATTACC  
Human GCTAATCTATTACC  

Dog GCTAATCTATTACC  

chick GCTAATCTATTACC  

consensus MTTAATCTATTACC  

A T T

Figure 2.2. Evolutionary Sequence conservation within the rhombomere-specific S5 enhancer.  
(A) DNA homology searches identified evolutionary conserved region surrounding the vHNF1 site. The sequence alignments are shown here for mouse, human, and chick. The pink box marks the conserved vHNF1 binding site. Yellow boxes indicate identity shared between two species, while orange indicates shared identity between mouse, humans, and chick. (B) A highly conserved vHNF1 consensus binding site is found upstream of the MafB gene in mouse, human, dog, and chick genomic DNA. N- denotes any nucleotide, while M- denotes any nucleotide except for guanosine.  

56
Figure 2.3. Localization of RAR/RAR and RAR/RXR binding to the S5 enhancer
(A) The 3.3 kb full length S5 element was broken down into five smaller probes (1-5) using restriction endonucleases indicated. (B) EMSA testing binding of in vitro transcribed and translated (TNT) RAR and RXR homodimers (RAR or RXR) and RAR/RXR heterodimers (RAR/RXR) to S5 probes from (A). TNT mock made with empty protein expression plasmid. Lanes where binding has occurred is indicated by red arrows. (C) Full length probe Sacl-BglII probe broken down into three overlapping probes 3-1 to 3-3 using PCR. (D) EMSA shows RAR homodimer binding is localized to regions 3-2 and 3-3.
However as stated previously, this region does not contain any predicted canonical RAR/RAR homodimer binding sites (Figure 2.3B). Using overlapping probes, I refined the binding of RAR/RAR homodimers to 80 bp at the junction of probes 3-2 and 3-3 (Figure 2.3C-D). Furthermore, RAR/RXR heterodimer binding was also found within S5-2kb within 900 bp of the XbaI site (Figure 2.3B). No RAR/RAR or RAR/RXR consensus sites were found within this region, indicating that \textit{MafB} may be induced by RA indirectly. However the binding profile would suggest that this may represent a secondary entry site for RAR/RAR and or RAR/RXR binding which may contribute to finer control of S5-dependent activation of \textit{MafB} transcription.

Although no consensus RAREs were found within S5, closer examination of the S5-2kb region, revealed a predicted binding site for vHNF1 which is evolutionarily conserved between mouse S5 and the regions upstream of \textit{MafB} in chicken, dog, and human genomic DNA (Figure 2.2B) (Wingender et al., 1996).

\textit{vHNF1} is expressed transiently in r5 and r6

When we examined \textit{vHNF1} expression in the embryonic hindbrain, we found that \textit{vHNF1} expression coincides with \textit{MafB} induction and mirrors S5-directed LacZ expression precisely both spatially and temporally. Previous studies in zebrafish showed that \textit{vHNF1} is co-expressed with \textit{MafB} in r5 and r6. In the mouse hindbrain, \textit{vHNF1} expression has not been extensively described. In 0-10 somite stage mouse embryos (8.0-8.5 dpc), we found that \textit{vHNF1} mRNA expression in the neural tube extended from the future r5 to roughly the level of the heart, and abutted the Hoxb1-expressing future r4 domain (Figure 2.4). However, between the 10-14 somite stages, a gap between \textit{vHNF1} expression and the \textit{Hoxb1} expressing r4 domain becomes apparent as the anterior boundary of \textit{vHNF1} expression shifts posteriorly to the r6/r7 boundary. Such a posterior shift of expression has not been reported previously, and suggests that \textit{vHNF1} plays a role only in \textit{MafB} induction and not in its later expression.
Figure 2.4. Whole mount RNA in situ hybridization detected vHNF1 expression in the hindbrains of 0-14 somite stage mouse embryos (8.0-8.5dpc).

In 8 somite stage embryos, vHNF1 expression extends from the posterior hindbrain into the future spinal cord (A), while Hoxb1 is expressed specifically in r4 (B). Here, double staining with vHNF1 and Hoxb1 probes detects continuous hybridization from the r3/r4 boundary posteriorly into the future spinal cord in 4-6 somite (C, E, and F) and 8-10 somite (G, H) stage embryos. This suggests that the anterior boundary of vHNF1 expression is located roughly at the future r4/r5 boundary in 4-10 somite stage embryos. By contrast in 12-14 somite stage embryos (D, I, and J) a decided gap in expression between the Hoxb1 expressing r4 domain and the vHNF1 expression domain can be detected. Thus, presence of vHNF1 mRNA coincides with MatB expression only in 0-10 somite stage mouse embryos. Anterior arrows indicate the r4/r5 boundary, while posterior arrows coincide with the morphologically identifiable r6/r7 boundary in 12-16 somite stage embryos.
vHNF1 interacts functionally with a single site in the MafB enhancer

We examined whether this potential vHNF1 site, S5-vHNF1, could bind vHNF1 protein in electrophoretic mobility shift assays, and found that in vitro transcribed translated vHNF1 protein bound specifically to the S5-2kb region (Figure 2.5B). When we subdivided the 3.3 kb S5 region into 2.0 kb and 1.3 kb fragments by digestion with XbaI, only the 2.0 kb SacI-XbaI fragment, S5-2kb, which contained the predicted vHNF1 binding site, could bind and compete effectively for vHNF1 protein binding. To examine whether this was the only vHNF1 binding site present within S5 we mutated the S5-vHNF1 site into the mutS5-vHNF1 site (GC\textsuperscript{ta}ATCA\textsuperscript{tt}AAc to GC\textsuperscript{gt}ATCA\textsuperscript{gg}AAc). Introduction of similar mutations into the HNF1 binding site, that regulates hepatocyte-specific transthyretin expression, eliminates HNF1 dependent activation in hepatoma cells (Costa and Grayson, 1991; Tronche et al., 1999; Tronche et al., 1990). vHNF1(HNF1\textbeta) is highly homologous to the closely related HNF1\textalpha protein in regions involved in DNA binding, i.e. the N-terminal dimerization domain, POU-like domain, and the divergent homeodomain, and in all studies so far their DNA binding affinities have been identical (Baumhueter et al., 1988; Mendel et al., 1991; Rey-Campos et al., 1991; Tronche and Yaniv, 1992). Thus, these mutations in the S5-vHNF1 site were predicted to abolish vHNF1 binding. This mutated S5 fragment (mutS5) could not compete for vHNF1 binding to S5-2kb. Addition of 10- or 100-fold molar excess of double-stranded oligonucleotides containing the predicted S5-vHNF1 binding site or a known vHNF1 binding site (PE-vHNF1) competed for vHNF1 binding to S5 (Costa and Grayson, 1991; Tronche et al., 1999; Tronche et al., 1990). However, oligonucleotides, which contained either the mutated S5-vHNF1 site or a mutated PE-vHNF1 site (DS34vHNF1), which previously had been shown not to bind vHNF1, could not compete for vHNF1 binding. Finally, we demonstrated that vHNF1 could bind specifically to a S5-vHNF1 site containing oligonucleotide (Figure 2.5C).
Figure 2.5. vHNF1 binds to a single site within the S5 element.

(A) The sequences of the vHNF1 sites used in these studies. (B) In vitro transcribed/translated vHNF1 protein binds specifically to a single site within the radiolabeled 2kb SacI-XbaI region of S5, shown in Figure 2.1, in electrophoretic mobility shift assays. Lanes 2-20 contain 50ng of vHNF1 protein. Lane 1 contains only the radiolabeled probe. Competitor DNA present is shown above the lanes. 10 fold (Lane 3) and 100 fold (Lane 4) molar excess of the SacI-XbaI 1.3kb S5 fragment does not compete for vHNF1 binding, while effectively competes for vHNF1 binding, at 10 fold (Lane 5) and 100 fold (Lane 6) molar excess. 10x (Lane 7), 100x (Lane 8), and 1000x (Lane 9) molar excess of the entire S5 fragment can compete for vHNF1 binding, but the S5mut fragment (Lanes 10-12; 10x,100x, and 100x molar excess, respectively), which contains the mutated vHNF1 site shown in A, cannot. 10x and 100x molar excess of double-stranded oligonucleotides of the S5-vHNF1 site (Lanes 13,14) and the PE-vHNF1 site from the rat albumin enhancer (Lanes 15,16) bind vHNF1. The mutated S5 (Lanes 15 and 16) and DS34 (Lanes 19 and 20) vHNF1 sites cannot compete for vHNF1 binding. (C) A radiolabeled S5-vHNF1 site containing oligonucleotide binds vHNF1 protein (Lane 2) specifically. Lane 1 does not contain vHNF1 protein. All other lanes contain 50ng of vHNF1 protein. Addition of either 10, 100, or 1000 fold molar excess of S5-vHNF1 (Lanes 3-5, respectively) or PE-vHNF1 (Lanes 9-11) oligonucleotides compete for vHNF1 binding, but addition of 10, 100, or 1000 fold molar excess of S5mutvHNF1 (Lanes 6-8, respectively) or DS34-vHNF1 (Lanes 1-14) does not affect vHNF1 binding to the S5vHNF1 site.
In transfection studies performed in COS-7 cells, a dimer of the S5-vHNF1 binding site can direct vHNF1 dependent transcription as efficiently as a dimer of the previously described PE-vHNF1 binding site (Figure 2.6). By contrast, the mutated S5-vHNF1 binding site failed to induce vHNF1 dependent transcription. Thus, the S5-vHNF1 binding site can bind vHNF1 in vitro and direct vHNF1 dependent transcription in vivo.

The vHNF1 binding site is essential, but not sufficient for early MafB expression

To determine whether this S5-vHNF1 binding site was required for rhombomere-specific LacZ expression, we made transgenic mice using the entire S5 region with the vHNF1 binding site mutated. No LacZ activity could be observed in the prospective r5 or r6 domains in any of these lines (Figure 2.7B). Thus, this vHNF1 site is essential for rhombomere-specific LacZ expression. However, this vHNF1 site alone cannot direct high level rhombomere-specific LacZ expression. We could not detect any r5-r6-specific LacZ expression in embryos transgenic for constructs containing two copies of the S5-vHNF1 binding sites linked directly to the hsp68-LacZ reporter (Figure 2.7C and D). However, in two transgenic founder lines containing five copies of the S5-vHNF1 site, we detected very faint LacZ staining in the neural tube extending from r5 posterior to roughly the level of the heart in 0-10 somite stage embryos. This expression coincides with that of vHNF1 itself. These results indicate that vHNF1 is essential but not sufficient for rhombomere-specific MafB expression. Therefore, other factors must collaborate with vHNF1 to allow for rhombomere-specific transcriptional activation of MafB. Because a 934 bp SacI-EarI fragment that contains the highly homologous region and the vHNF1 site is not sufficient for r5-r6-specific LacZ expression (Table 2.1, data not shown), some of these additional factors do not bind S5 near the vHNF1 site, are unlikely to associate directly with vHNF1, and have binding sites located outside the region with the highest homology between mouse, human, and chick genomic sequences.
Figure 2.6. vHNF1 can activate transcription via the S5-vHNF1 site. vHNF1 can activate transcription of reporter plasmids containing dimers of the PE-vHNF1 site or the S5-vHNF1 site to an equivalent extent upon transfection into COS-7 cells. By contrast, vHNF1 fails to activate transcription of a reporter construct containing a dimer of the mutated S5mut-vHNF1 site above the level seen for the empty reporter vector.
Figure 2.7. A vHNF1 binding site is essential, but not sufficient, for r5- and r6-specific expression of Kreisler in 0-10 somite stage embryos.

(A) LacZ expression can be seen in r5 and r6 in a 6 somite stage embryo transgenic for the S5-2kb construct. (B) In a 6 somite embryos transgenic for the S5mut construct, in which the vHNF1 site has been mutated to that shown in Figure 2.5, no rhombomere-specific LacZ expression can be seen. (C) A dimer (2xvHNF1) is not sufficient to direct rhombomere-specific expression of the hsp68LacZ reporter. However, in embryos transgenic for a quintuplet (5xvHNF1) of the S5vHNF1 site faint LacZ staining can be detected in the neural tube extending from r5 throughout the posterior hindbrain (D).
Discussion

These studies strongly suggest that vHNF1 acts as a key regulator of early hindbrain patterning by directly regulating initiation of MafB expression during early development. vHNF1 likely activates initiation of MafB expression directly via a single binding site that is present in the rhombomere-specific S5 enhancer located 16.5 kb upstream of the MafB transcriptional start site in the mouse. Portions of the S5 enhancer are evolutionarily conserved 17.5 kb upstream of the Human MafB gene and 13 kb 5' of the chicken MafB gene. S5 homologous regions are also present in equivalent positions in all other mammalian genomes, such as dog and rat, for which the sequence surrounding the MafB gene is currently available. Strikingly, the single vHNF1 binding site, for which the consensus sequence is NM(T/A)AATN(A/T)TTAAC(C/T) (Tronche and Yaniv, 1992), is present in the S5 homologous region in all of these instances, and in transcription factor binding site searches of the mouse and human genomic sequences surrounding the MafB gene this was the only vHNF1 binding site detected. As yet Xenopus tropicalis, Fugu, and zebrafish genomes are incomplete and the sequence surrounding their MafB genes is not yet available in the public domain. Thus, we can not draw any conclusions regarding the conservation of the S5 region and vHNF1 site for amphibians or teleosts. However, zebrafish and mouse vHNF1 proteins and, in particular, their DNA binding regions are highly conserved (Sun and Hopkins, 2001). Thus, it seems likely that a similar vHNF1 binding site will eventually be identified and shown to govern initiation of MafB(Val) expression directly in zebrafish, just as in the mouse.

An interaction between vHNF1 and MafB was first strongly suggested by studies in zebrafish showing that no MafB (Val) expression was detected in vHNF1-/- zebrafish and that ectopic expression of vHNF1 extends MafB (Val) expression ectopically (Sun and Hopkins, 2001). In mice, knock-out of vHNF1 is lethal prior to hindbrain formation, because vHNF1 is required for visceral endoderm development. In the mouse, an unusually small and rounded otic vesicle similar to that seen in the absence of kr/kr mutants is observed in vHNF1-/- embryos generated by tetraploid rescue, but MafB expression has so far not been examined in vHNF1-/- mouse embryos (Barbacci et al., 1999; Coffinier et al., 1999b). These observations taken together with the evolutionary
conservation of vHNF1 structure and MafB function in mouse and zebrafish, strongly suggest that vHNF1 acts upstream of \textit{MafB} in the mouse as well. Furthermore, the expression pattern of \textit{vHNF1} in the mouse is consistent with its putative role in hindbrain patterning, and, specifically, in \textit{MafB} induction. In the mouse neural tube, \textit{vHNF1} expression extends from the future r4/r5 boundary posteriorly to roughly the level of the heart starting at the 0 somite stage, and recedes posteriorly out of the r5-r6 domain by the 10-12 somite stages. \textit{S5} directed \textit{LacZ} expression is only transient and mirrors the appearance and disappearance of \textit{vHNF1} expression in the future r5-r6 domain precisely. Thus, the present studies provide the first evidence of a direct requirement for vHNF1 in regulating initiation of \textit{MafB} expression.

However, vHNF1 alone is not sufficient to drive r5-6 expression nor is the enhancer containing vHNF1 sufficient for later maintenance of \textit{MafB} expression patterns. vHNF1 contains a dimerization domain, a POU-specific box, and a divergent homeodomain, but lacks the activation domains that are present in the closely related HNF1\textalpha protein (Mendel et al., 1991). vHNF1 can form extremely stable homo- and heterodimers that bind DNA with HNF1\textalpha (Mendel et al., 1991; Rey-Campos et al., 1991). \textit{HNF1\textalpha} expression overlaps to some degree with that of \textit{vHNF1} in liver, kidney, pancreas and intestine, but not in the hindbrain (Barbacci et al., 1999; Cereghini et al., 1992; Coffinier et al., 1999b; Ott et al., 1991). Furthermore, mice lacking \textit{HNF1\textalpha} are viable and have normal hindbrains (Lee et al., 1998; Pontoglio et al., 1996). Thus, vHNF1 and HNF1\textalpha may interact directly during these later developmental processes, but HNF1\textalpha is not likely the partner for vHNF1 in hindbrain development. In \textit{MafB} regulation, as in many other cases, vHNF1 relies on additional factors, which have their own distinct binding sites, to activate high level region- or tissue-specific gene expression.

Fibroblast growth factors (FGFs) have been implicated in posterior hindbrain patterning and play an as yet not well understood role in \textit{MafB} regulation. Studies in the chick first demonstrated that exogenous FGF2, FGF4, and FGF8 could increase the r3 and r5 domains at the expense of r4 (Marin and Charnay, 2000). In addition, local insertion of FGF4 or FGF8 loaded beads could induce \textit{MafB} expression in r7 and r8, but
only in regions in close proximity to the FGF source. Furthermore, inhibition of FGF signaling abolished MafB expression in r5 and r6. Later, Wiellette and Sive showed that FGF signaling acted similarly in patterning the posterior zebrafish hindbrain and that FGF signaling acted with vHNF1 to induce MafB expression (Wiellette and Sive, 2003). The above results have since been confirmed again in the fish by others (Hernandez et al., 2004).

As yet it is unclear whether FGF induces factors specific to the posterior hindbrain or activates more generally expressed FGF responsive transcription factors. However, a clue as to where vHNF1 collaborators and FGF inducible factors important for MafB regulation may normally reside may be gleaned from ectopic expression studies performed in zebrafish. Sun and Hopkins found that ectopic vHNF1 expression in zebrafish could only extend the MafB (Val) expression domain into r4, but not more anteriorly or posteriorly (Sun and Hopkins, 2001). By contrast, others reported that ectopic vHNF1 expression in zebrafish lead to additional low level MafB (Val) expression in r2 and r3 (Hernandez et al., 2004). Even in this latter case, MafB (Val) expression appeared significantly higher in r4 than in r2 and r3. A simple way to reconcile these differences is that intermediate to low levels of ectopic vHNF1 would readily activate expression of MafB (Val) in the presence of the proper collaborators in r4. Much higher vHNF1 levels are required for MafB (Val) activation in r2 and r3, because normal collaborators are absent or present at lower levels in these domains and, thus, expression depends on vHNF1 to a greater degree. Some support for this view comes from our findings that S5-vHNF1 binding site dimers cannot activate LacZ reporter expression, but quintuple binding sites lead to low level reporter gene expression. So too, we would expect unusually high levels of vHNF1 to eventually drive low levels of MafB (Val) expression in ectopic domains even in the absence of the ideal or normal collaborator(s). Once MafB (Val) expression is activated, autoregulatory activity of MafB (Val) can act to maintain and increase MafB (Val) expression in these ectopic domains. So, we postulated that the lower level MafB (Val) expression in r2 and r3 is due to very high ectopic vHNF1 levels, and that under normal conditions the molecule(s) that collaborate with vHNF1 to activate MafB expression are found at high levels or in an activated state in r4-r6.
In the simplest case a single collaborator or set of collaborators expressed in r4-r6 would be involved (Figure 2.8A). Of course, r4-derived FGF signals could activate molecules in r4-r6 that are present more ubiquitously in the neural tube. However, most early gene expression patterns in the developing hindbrain show sharp anterior expression boundaries and more diffuse posterior expression boundaries, like vHNF1 itself, and no r4-r6 specific early acting transcription factor has yet been identified. Thus, aside from a vHNF1 co-activator, another molecule or set of molecules would be required to repress \( MafB \) expression posterior of the r6/r7 boundary. The observation that the posterior boundary of \( MafB \) expression extends caudally into r7 and r8 in \( RAR\alpha/RAR\beta \) compound null mice lends support to the existence of such a repressor (Dupe et al., 1999). Thus, we favor this second model (Figure 2.8B). Further dissection of the regulatory sites within S5 will reveal the identity of these vHNF1 co-activators and repressors.

The analysis of mouse and zebrafish mutants has lead to the identification of individual transcription factors that determine specific anterior-posterior domains, but understanding links between these transcription factors and thus the precise transcriptional networks involved has required the dissection of regulatory elements. We have begun to elucidate the transcriptional network required to build r5 and r6. We have shown that in the mouse vHNF1 is likely to collaborate with other transcription factors to initiate MafB transcription directly in the future r5-r6 domain, and MafB in turn directly activates the expression of \( Hoxa3 \) and \( Hoxb3 \) (Figure 2.8C) (Manzanares et al., 1999a; Manzanares et al., 1997). So, it is quite likely that early signals, such as retinoic acid and FGFs which establish anterior-posterior positioning, govern vHNF1 expression directly. Indeed early studies in F9 cells demonstrate that vHNF1 can be induced by retinoic acid treatment and more recent analyses in zebrafish suggest that Retinoic acid induction of \( MafB \) occurs, at least in part, through vHNF1 (Hernandez et al., 2004; Kuo et al., 1991). Thus, further analyses of vHNF1 collaborators and regulators will provide us with a more complete picture of the molecular events that transform early positional signals into two precisely defined compartments.
Figure 2.8. Models for vHNF1 dependent activation of rhombomere-specific Kreisler expression

(A) The simplest molecular model for vHNF1 dependent activation of Kreisler predicts that a vHNF1 collaborator is expressed in the future r4-r6 domain. (B) However, other observations support a model in which at least one vHNF1 collaborator is expressed throughout the posterior hindbrain up to the r3/r4 boundary, and a repressor is expressed throughout the posterior hindbrain up to the r6/r7 boundary. (C) A schematic diagram shows the regulatory network that governs early r5 and r6 development. Retinoic acid (RA) and Fibroblast growth factors (FGFs) may regulate vHNF1 and/or Kreisler directly. vHNF1 directly regulates Kreisler expression. Kreisler in turn activates expression of r5 and r6-specific Hoxa3 expression, and in collaboration with Krox20, activates r5-specific Hoxb3 expression.
Acknowledgments

I would like to thank Dr. Florence Kim, Dr. Tomomi Kaneko, Melissa Bieman, Nicole Stallwood and Virginia Sadl for direct and indirect contributions to this work. To Drs. Xiaoxia Sun and Nancy Hopkins, we are grateful for the generous gift of the pCS2-vHNF1 construct and for assistance, and Dr. Janet Rossant for the hsp68LacZ construct. Sandra Tondat, Lois Schwartz, and the Samuel Lunenfeld transgenic core facility were essential for generating transgenic mice for these studies and their efforts are much appreciated.

Footnote on previously published materials

The bulk of this chapter was previously published in *Mechanisms of Development* where I am listed as co-first author (Kim et al., 2005). I was directly and solely responsible for Figures 2.1, 2.3 (unpublished data) and 2.4 and assisted or collaborated on data pertaining to Figures 2.5 and 2.7 which were mainly conducted by V. Sadl and F. Kim respectively. CAT assays in Figure 2.6 were done by T. Kaneko while sequence analyses in Figure 2.2 were made by S. Cordes.
Chapter 3: A vertebrate Polycomb response element governs segmentation of the posterior hindbrain

Abstract

Chromatin remodeling by Polycomb (PcG) and trithorax Group (trxG) proteins regulates gene expression in all metazoans. In Drosophila, PcG/trxG protein complexes reorganize chromatin by binding to DNA elements known as PcG/trxG response elements (PREs). PREs have been validated in Drosophila but not in any other organism. Here we have identified a vertebrate PRE, PRE1, that regulates expression of the mouse MafB/Kreisler segmentation gene. We show that reductions of the M33 Polycomb gene or partial removal of the PRE1-containing region shift MafB expression anteriorly. While the role of MafB in segmentation and the PRE1 sequence are not conserved between flies and mice, the interactions of PRE1 with PcG and trxG proteins are: In transgenic flies, PRE1 modulates reporter gene expression in a PcG/trxG dependent manner and can recruit PcG-protein binding. Overall our results support the view that PREs act as entry sites for PcG/trxG chromatin remodeling in both vertebrates and invertebrates.
Introduction

Chromatin remodeling mediated by the Polycomb group (PcG) and trithorax group (trxG) complexes acts to stabilize transcriptional states of cells situated in dynamic signaling environments and, thus, plays key roles in development, maintenance of stem cell pluripotency and the progression of cancer. PcG proteins lock the transcriptional state of key developmental genes into cellular memory (Dejardin and Cavalli, 2005a; Grimaud et al., 2006; Ringrose and Paro, 2001). By contrast, trxG proteins do not maintain continuous activity or repression of genes, but stimulate enhancer dependent expression by keeping specific enhancers ‘open’. Most known PcG and trxG genes encode chromatin-associated proteins, which bind as large protein complexes to specific DNA elements, known collectively as PcG/trxG response elements (PREs), to reorganize the surrounding chromatin. In Drosophila, PcG and trxG protein containing complexes have been shown to bind distinct, but often closely situated or overlapping, DNA elements. While gene translocation experiments in the mouse and genome-wide profiling of PcG target sites in mouse and human cell lines suggest that PREs exist in organisms other than the fly, so far no vertebrate PREs have been functionally validated (Boyer et al., 2006; Bracken et al., 2006; Kmita et al., 2000; Kondo and Duboule, 1999; Lee et al., 2006; Spitz et al., 2005; Squazzo et al., 2006).

Epigenetic mechanisms and their applications can be tremendously diverse between different organisms. Nonetheless, the core components of the two main polycomb complexes, the early acting Polycomb repressive complex 2 (PRC2) and the later acting Polycomb repressive complex 1 (PRC1) appear to be highly conserved (Dejardin and Cavalli, 2005a; Grimaud et al., 2006; Ringrose and Paro, 2001). PRC2 core subunits are present in animals and plants, while those of PRC1 are unique to flies and vertebrates. In the fly, the Polycomb repressive complex 2 (PRC2) contains as its core subunits the PcG proteins Enhancer of zeste (E(z)), Extra sex combs (Esc), SUZ12 and the histone binding protein NURF-55. E(z) and Esc proteins form a histone methyltransferase complex that methylates lysine 27 and lysine 9 of histone H3 and, thereby, creates an epigenetic mark, which is bound by the chromodomain of Polycomb (Pc) proteins in the fly and mouse, thereby recruiting the PRC1 complex (Cao et al.,
PRC1 consists of Pc, polyhomeotic (Ph), Posterior Sex Combs (Psc), and dRing. The PRC1 and PRC2 core complexes do not bind DNA, but pleiohomeotic (Pho), its vertebrate homologue Yin Yang1 (YY1) and related Pho-like (Phol) proteins can bind a specific DNA motif (Atchison et al., 2003; Brown et al., 2003; Brown et al., 1998) and at times, interact specifically with Pc and Ph subunits of PRC2 (Wang et al., 2004). While all known PREs contain Pho-binding sites, these alone cannot tether PcG proteins to DNA \textit{in vivo}, and, in Pho/Phol double-mutants, polytene chromosomes still stain for PcG binding sites (Brown et al., 2003).

The definitive feature of PREs identified in the fly is their ability to regulate expression of transgenes and endogenous genes in a manner dependent on PcG and trxG proteins. When placed in front of a reporter transgene and/or the mini\textit{white} gene, PREs act as silencers. This repression is lifted by mutations that inactivate PcG proteins or enhanced by mutations in trxG proteins. In addition, all validated PREs contain clusters of putative YY1 and GAGA factor binding sites (Ringrose et al., 2003). Within PREs, Pho-binding sites and juxtaposed Pc-binding elements (PBEs) appear to function as an integrated DNA platform for the synergistic binding of Pho and the multi-subunit polycomb core complex (PCC) \textit{in vitro} (Mohd-Sarip et al., 2006). Aside from the presence of YY1/Pho and GAGA factor binding sites, no unifying sequence requirements for PREs have emerged. Analyses of PREs in the fly and profiling of PcG target sites in mouse and human cell lines suggest that PcG/trxG proteins most often regulate developmentally relevant transcription factors (Boyer et al., 2006; Lee et al., 2006).

PcG and trxG proteins were first identified by their effect on embryonic segmentation and expression of the \textit{Hox} segment selector genes in \textit{Drosophila}, and the first PREs identified reside in the \textit{Hox} gene complexes of the fly (Akasaka et al., 1996; Chiang et al., 1995; Paro, 2000; Pirrotta, 1997; Pirrotta, 1998). In the mouse, mutations in PcG and trxG genes often affect segmentation of the embryonic vertebrate hindbrain, the best system for studying vertebrate developmental compartments. The embryonic hindbrain is transiently subdivided along its anterior-posterior axis into eight lineage-restricted compartments, known as rhombomeres (r1-r8) (Fraser et al., 1990; Lumsden and Krumlauf, 1996). Rhombomeres represent true compartments, as defined by genetic and cell lineage experiments. Many gene expression patterns, including those of the \textit{Hox}
segmentation genes, respect rhombomere boundaries and are limited to specific subsets of rhombomers (Capecchi, 1997; Lumsden and Keynes, 1989; Lumsden and Krumlauf, 1996). Transplantation experiments have shown that cells from adjacent rhombomeric domains cannot intermingle, but sort themselves according to their rhombomeric identity and transcriptional profiles (Guthrie and Lumsden, 1991; Lumsden and Guthrie, 1991). This property provides a direct and easily visualized record of transcriptional states on a cell-by-cell basis.

In the mouse and fly, mutations in PcG genes can shift segmentation gene expression anteriorly. Such an anterior shift in the expression of the \textit{Kreisler/MafB} segmentation gene is also seen in mouse embryos carrying the X-ray induced \textit{kreisler (kr)} mutation (Giudicelli et al., 2003; Sadl et al., 2003). The basic domain leucine zipper transcription factor \textit{MafB/Kreisler} is the earliest gene expressed in a segmental fashion in r5 and r6 (Cordes and Barsh, 1994) and in its absence, r5 does not form and r6 is mis-specified (Frohman et al., 1993; McKay et al., 1994). The \textit{kr} allele is caused by a near perfect inversion that removes a total of only 19 bp, does not disrupt the \textit{MafB} coding region or remove the S5 rhombomere-specific regulatory element (Cordes and Barsh, 1994), but nevertheless eliminates detectable \textit{MafB} expression in r5 and r6, and causes ectopic low level \textit{MafB} expression in r3 of \textit{kr/+} and \textit{kr/kr} embryos (Figure 3.1) (Giudicelli et al., 2003; Kim et al., 2005; Sadl et al., 2003). \textit{MafB} misexpression in animals carrying the \textit{kr} inversion must be caused by a cis-acting event, because it is not seen in embryos homo- or heterozygous for the severely hypomorphic \textit{krenu} allele (Figure 3.1) (Sadl et al., 2003). In \textit{kr/kr} mice, all other \textit{MafB} expression domains are normal (Eichmann et al., 1997) S.P. Cordes, unpublished data).

We postulated that the \textit{kr} inversion might anteriorize \textit{MafB} expression by affecting a PRE. Using further analyses of the nature of the \textit{kr} allele, of the effects of chromatin remodeling agents and reduction of the \textit{M33} polycomb gene levels on \textit{MafB} expression in normal and \textit{kr/kr} embryos, comparative sequence analyses, assays for PcG– and trxG- dependent PRE activity in \textit{Drosophila}, chromatin immunoprecipitation assays in F9 embryonic carcinoma cells and, finally, transgenic analyses in the mouse, we have identified and validated what to our knowledge is the first true PRE, referred to as PRE1, that regulates the rhombomere-specific \textit{MafB} expression.
Materials and Methods

Transgenic flies and β-galactosidase staining

PRE1 was excised from P1 clone P1-222 (Genome systems) as a 9kb SalI fragment extending from Chromosome 2: 160234094-160243533, and the hcPRE1 fragment, which extends from Chromosome 2: 160240462-160243530 was excised from the PRE1 by digestion with BamHI and BspHI. PRE1 and hcPRE1 were inserted into the NotI restriction site of the pUZ vector (Lyko et al., 1997). Constructs were injected into white1118 embryos according to standard procedures (Rubin and Spradling, 1982). Embryos were co-injected with 500 µg/ml of the pUZ constructs and 100 µg/ml of pHSPi helper plasmid (Steller and Pirrotta, 1985). All transgenic insertions were homozygosed or balanced over CyO or TM3. Flies positive for transgenes were identified by their ability to rescue white eye color. Insert orientation and copy number were verified by sequencing. All flies were raised at 25°C on standard medium. The hsGAL42-1 (Brand et al., 1994) driver line, which carries the GAL4 driver on chromosome 3, was used to measure initial β-galactosidase in transgenic lines before and after Cre excision. Quantification of β-galactosidase activity was performed a minimum of 6 times for at least 3 lines for each transgene using a CPRG assay (Lyko et al., 1997). Data is represented as a percentage of vector alone β-galactosidase activity per mg of whole protein from the same experiment. In experiments testing the effects of PcG/trxG mutants on PRE1-dependent silencing, the assay was adapted to 96 well plates by reducing test volume from 1 mL to 200 µL. In order to reveal effect of Trl mutant on further silencing LacZ, volume of lysate was increased by 2 fold. All fly lines used are listed in Table 3.5. For mating schemes in testing PcG/trxG mutant effects, see Figures 3.15-3.19. Statistical analysis was carried out with GraphPad Prism 4.0 (GraphPad Software, Inc., La Jolla, CA). For detailed statistical analyses see Table 3.2.

Immuno-Fluorescent in situ hybridization of Drosophila polytene chromosomes

Immuno-FISH experiments were performed as described in (Lavrov et al., 2004). Biotinylated FISH probes were made with the Bio-Nick nick-translation kit (Invitrogen, Carlsbad, CA) using the pUZ vector as a template. Immunostaining of polytene chromosomes was performed using affinity-purified rabbit polyclonal antibodies specific
for PH and Pc (generous gifts from R. Paro). Chromosomes were mounted for fluorescence microscopy using Vectashield with DAPI (Vector Labs, Burlington, Canada). Fluorescent images of stained chromosome squashes were captured with the Improvision Openlab system (Quorum Technologies Inc., Guelph, Canada). For each transgene, 10-20 chromosomes were analysed from two independent lines in at least three different experiments.

**Immunofluorescence of mouse embryonic sections**

E8.5 mouse embryos were harvested and 10 μm hindbrain longitudinal sections were prepared for immunofluorescence. An affinity purified anti–full-length MafB antibody was used in immunofluorescence experiments as previously described (Sadl et al., 2002).

**Whole-Mount in Situ RNA Hybridizations**

Whole-mount in situ RNA hybridizations were performed using the MafB/Kreisler probe as previously described (Cordes and Barsh, 1994; Frohman et al., 1993). kr/kr embryos were unambiguously genotyped as previously described (Sadl et al., 2003). Neuronatin antisense probes were made by digesting the Image clone ID4317155 plasmid with EcoRI, and then transcribing with T3 RNA Polymerase.

**Chromatin remodeling agents**

Between 7.25-8.75 dpc of their pregnancies, female mice received intraperitoneal injections of one of the following: saline, 5-aza-2'-deoxycytidine (0.5mg/kg) (Sigma-Aldrich, Oakville, Canada), Trichostatin A (Sigma-Aldrich) at 0.75mg/kg in DMSO, or DMSO. Embryos were dissected 6 to 12 hours later and prepared for in situ hybridization.

**Chromatin Immunoprecipitation (ChIP) and real-time quantitative PCR**

F9 embryonic carcinoma cells were cultured in Dulbecco’s modified Eagle’s medium supplemented with 10% (V/V) fetal bovine serum (Invitrogen), 2mM glutamine, 100units/ml of penicillin, and 100µg/ml of streptomycin. Cells were plated on 10 cm 0.1% (w/v) gelatin-coated tissue culture plates (1.5x10⁶ cells/plate) 48 hours before formaldehyde fixation. 24 hours after plating, cells were washed with fresh media and, 36 hours after plating, treated with 1µM all-trans retinoic acid (Sigma-Aldrich) or ethanol
carrier. At 48 hours, each plate (2.1 x 10^7 cells/plate) was washed once with ice cold PBS, fixed with 10 ml of 1% (v/v) formaldehyde in normal media for 10 min at room temperature, and lysed in SDS lysis buffer. A cell lysate of 3.2 x 10^7 cells/ml was sonicated at 30% for four 10 second intervals. Immunoprecipitations were performed with an EZ ChIP kit (Millipore, Etobicoke, Canada, #17-371) using one of the following antibodies: 1μg Anti-RNA Polymerase II (clone CTD4H8, Millipore #05-623B), 4μl Anti-Trimethyl-histone-H3 (Lys27) (Millipore #17-622), 1μg Anti-Histone H3, dimethyl (Lys4) (Abcam, Cambridge, MA, ab7766) or 2μg Anti-SUZ12 (Abcam, ab12073). Real time PCR assays were carried out on an Applied Biosystems 7900HT Real-Time PCR System using the SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA) using 40 cycles with an annealing/extension temperature of 60°C. The following primer pairs were used for qPCR:

*MafB* promoter (prF: 5'-CATCGCTAAAAGCGAGGCTCAG-3’ and prR: 5'-CTAGAGCCAAAGAGTCGGAGAG-3’),
S5 enhancer (S5 F: 5'-GTGGAGCCAAGTCTGAGC-3’ and S5 R: 5'-GCATAGGTCTCAGCCACCA-3’),
PRE1a (PRE1aF: 5’-CTTAATCCCCAATATTACTGCTGACT-3’ and PRE1aR: 5’-ATGTGATTATGGCTGTTTTCATC-3’),
PRE1b (PRE1bF: 5’-GCCTGGATTTATAAGAGTGCAAGT-3’ and PRE1bR: 5’-AGACAGATACCATGTTTTCAT-3’),
PRE1c (PRE1c F: 5’-GCTGGATTTAAAGAGTGAACAGT-3’ and PRE1c R: 5’-GGACAGCTCAGAAAAAGCT-3’),
PRE1d (PRE1d F: 5’-GTCCCGCAGAAAAAGTTTGAG-3’ and PRE1dR: 5’-GAGGAAACCCAACACTAATTTCAT-3’),
Mouse GAPDH promoter (mGAPDHpr F: 5’-GATGATGGAGGACGTGAGTGG-3’ and mGAPDHpr R: 5’-GGCTGCAGGAGAAGAAAATG-3’).

ChIP samples were quantified via the standard curve method, by measuring the CT values and interpolating the percentage input on the standard curve. Each experiment was performed three times from the cell stage and PCRs were performed in triplicate. ChIP assays were analyzed by a 2-way ANOVA using GraphPad Prism 4.0 (GraphPad Software, Inc.) (Table 3.3).
All mouse husbandry and procedures involving mice were performed in accordance with CACC guidelines. Transgenic mice were generated by pronuclear injection into fertilized oocytes from superovulated FVB mice. Founder animals were identified by Southern and used to generate lines. Embryos were stained for LacZ as previously described (Whiting et al., 1991). Transgenic lines were maintained by crossing with CD1 mice.

The 3.02 kb BamH1-BspH1 hcPRE1 fragment from a genomic P1–222 clone (Research Genetics, Invitrogen) was cloned into the filled in NotI site of the S5-hsp68LacZ expression construct described in Kim et al. (2005) to generate the hcPRE1-S5-LacZ reporter construct. The RARE site from Hoxa1 was inserted by annealing the primers FRT RARE F (5'-GAACGCGGCCGCGAAGTTCCTATACTTTCTAGAG AATAGGAACCTTCAGGTCTACCCGAAGTTCAAGA-3’) and FRT RARE R (5’- CTTGCGGCCGCGCTCGATCGAAGTTCCTATTCTCTAGAAAGTATAGGAACCTTC TTGAACCTTCGCTG-3’) and inserting these into the NotI site of hcPRE1-S5-LacZ to create the RARE-hcPRE1-S5-LacZ construct. For microinjection, SalI digestion followed by DNA electrophoresis separated inserts from vector sequences, and inserts were purified using GeneClean beads (Bio101, Vista, CA).

**Results**

**MafB protein is ectopically expressed in r3 of kr/kr and kr/+ embryos**

In normal 14 somite stage embryos, *MafB* is expressed in rhombomeres 5 and 6 (Figure 3.1A). Our lab had previously generated a chemically induced point mutation *kremu*, using ethynitrosourea (ENU) which changes a conserved asparagine in the DNA binding domain to a serine, rendering the resulting protein transcriptionally inactive (Sadl et al., 2003). In *kremu*/*kremu* homozygotes, *MafB* is expressed, though in a smaller “r5-r6-like” domain (Figure 3.1B). Previously, *MafB* expression was not detected in the hindbrains or cranial neural crest of E8.0-9.5 *kr/kr* embryos by RNA whole-mount *in situ* hybridization (Figure 3.1C). Later, *MafB* expression could be detected in the roofplate of
Figure 3.1. Effects of the \textit{kr} and \textit{kr}^{enu} Mutations on MafB Expression.
(A-C) Dorsal views of whole mount RNA \textit{in situ} hybridization on 14 somite stage embryos. \textit{MafB} was expressed in r5 and r6 of normal embryos (A). In \textit{kr}^{enu} homozygotes (B), \textit{MafB} RNA was present in a smaller domain that was roughly the width of one rhombomere rather than two. \textit{MafB} expression could not be detected in \textit{kr}/\textit{kr} embryos (C) by RNA \textit{in situ} hybridization. (D-G) Immunofluorescence with anti-MafB antibody on embryonic day 8.5 (E8.5) hindbrain longitudinal sections. MafB was nucleary localized to r5 and r6 in normal embryos (D), and in a reduced domain in \textit{kr}^{enu} homozygotes (E). MafB protein could not be detected in the posterior hindbrains of \textit{kr}/\textit{kr} embryos (F), but could be seen in r3. In \textit{kr}/+ embryos (G), MafB protein was detected in r5 and r6 and at lower levels in r3. (H-K) DAPI counterstained sections for D-G, respectively. All embryos shown in D-K are at the six- to eight-somite stages. Scale bar 200 $\mu$m.
E9.0 kr/kr, kr/+, and +/+ embryos at equivalent levels (S.P. Cordes, unpublished observations). Even though MafB mRNA could not be detected, expression of Hoxa3, a MafB target gene, was reported in r3 of kr/kr and kr/+ embryos (McKay et al., 1994; Manzanares et al., 1999b). In addition, Krox20 expression was detected in kr/kr embryos well past the 22-somite stage, a stage at which r3-specific Krox20 expression has disappeared in normal embryos (Frohman et al., 1993). It seemed possible that the chromosomal inversion responsible for the kr mutation might cause low-level ectopic Kreisler expression in r3. To test this hypothesis, we examined the hindbrains of kr/+ and kr/kr embryos for the presence of MafB protein by immunofluorescence. In kr/kr embryos, MafB protein was absent in the "r5-r6" domain and was present only in r3 (Figure 3.1F). In kr/+ embryos, MafB protein was present in r5 and r6 and in an additional domain of lower level expression in r3 (Figure 3.1G). No MafB protein could be detected in r3 of +/+, kr<sup>enu</sup>/+, and kr<sup>enu</sup>/kr<sup>enu</sup> embryos (Figure 3.1D, E). Thus, the X-ray-induced inversion responsible for the kr mutation causes ectopic expression of Kreisler in r3 by an as yet unknown cis-acting molecular mechanism.

**Analysis of the kr X-ray induced inversion suggest the presence of a PRE**

The kr inversion breakpoints are located ~51 kb upstream of the MafB transcriptional start and ~45 kb upstream of the transcriptional start site of the Neuronatin (Nnat) gene, which encodes a novel secreted protein of unknown function (Figure 3.2A). Previously, we have identified a retinoic acid responsive enhancer termed S5, 21 kb upstream of MafB start site, which drives expression of MafB in rhombomeres (r) 5 and 6 in the early embryonic hindbrain. Nnat is expressed in r3 and r5 early during hindbrain patterning (Figure 3.2C, F, I) (Wijnholds et al., 1995) and is paternally imprinted (Evans et al., 2001; Kagitani et al., 1997; Kikyo et al., 1997). Thus, analyzing Nnat expression might provide further clues as to what regulatory elements might have been affected by the kr inversion. The imprinting control region (ICR) of Nnat lies 15 kb upstream of its start site (John et al., 2001). Neither S5 or the ICR of Nnat are disrupted by the kr inversion and are translocated intact with their respective transcriptional units (Figure 3.2B). Imprinting of Nnat was unaffected in kr/+ animals derived from kr/+ dams. However, in kr/+ progeny derived from kr/kr males, Nnat expression was reduced in r3 and now extended throughout r5 and r6 and was present at high levels into the dorsal
Figure 3.2. Rhombomere-specific expression of Neuronatin is altered by the kr inversion.

(A) Diagram of a normal mouse chromosome 2. (B) Chromosome 2 containing the kr inversion. In the kr allele, the inversion breakpoints (bpt) are located 51kb 5' of the MafB (bpt2) and 45kb upstream of the Neuronatin (Nnat) (bpt1) transcriptional start sites (Cordes and Barsh, 1994; John et al., 2001). Previous transgenic analyses have located an r5 and r6 specific enhancer 21kb 5' of MafB and the Nnat imprinted control region (ICR) within 15kb of the Nnat transcriptional start (Sadl et al., 2003; John et al., 2001). PRE1 denotes the region containing the Polycomb response element: region A is translocated with the MafB gene, while region B is left behind in the kr inversion. The disrupted PRE1 region is denoted as ΔPRE1 in the kr inversion allele.

(C-K) Dorsal views of whole mount RNA in situ hybridization with Nnat. In +/- or maternally derived kr/+ animals, Nnat is first expressed in r3 at 10-12 somites (s) (C) and in r3 and r5 at 16-18s (F). By 20-22s (I), expression is restricted to r5, the somites (s) and presumptive post mitotic neuronal cell types (*). In kr/+ (D, G, J) and kr/kr (E, H, K) animals carrying a paternal kr allele, Nnat expression in r3 is reduced and posterior hindbrain expression extends from r5 into r6 and r7. These mutant embryos also exhibit neural crest (nc) expression, whereas somite and sensory placode expression is lost. Ectopic expression in the posterior hindbrain is marked with arrows and reduced expression in r3 with an arrowhead. Scale bars are all 100μm.
aspect of r7 in 10-18 somite embryos (Figure 3.2D, G). By the 20-22 somite stage, Nnat was downregulated in r5 but still present in r6, and at low levels in r7 (Figure 3.2J). By contrast, at this stage MafB expression is extinguished in r5 and r6 (Cordes and Barsh, 1994) and Nnat is expressed only in r5 in normal embryos (Figure 3.2I). Notably, the posterior boundary of Nnat expression at all of these stages was very diffuse. Timing of Nnat expression differed from normal Nnat and MafB expression and persisted past the 22 somite stage. Finally, in both kr/+ and kr/kr embryos, ectopic Nnat expression, such as in r6-specific neural crest (nc) cells resembled that of MafB in normal embryos (Figure 3.2G-H, J-K). In kr/kr embryos, Nnat expression in the posterior hindbrain was similarly restricted to a narrower region, as would be expected in embryos in which r5 was lost and r6 mis-specified (Figure 3.2E, H, K). Thus, the kr inversion disrupts the rhombomere-specific expression patterns of two genes, Nnat and MafB, in an incompletely reciprocal manner leading to anteriorization of MafB expression and a posterior shift of Nnat expression. Some of these observations could be explained by the translocation of neural crest specific and rhombomere-specific enhancers from MafB into the vicinity of the Nnat gene and vice versa. However, the timing and overall posterior and anterior shift of these gene expression patterns is also consistent with translocation of a PRE from the MafB locus into the vicinity of the Nnat gene.

**Chromatin remodeling agents reactivate MafB expression in the posterior hindbrain**

To assess whether MafB expression in kr/kr mutants is silenced as a consequence of chromatin remodeling, we treated kr/kr embryos with chromatin remodeling agents and examined reactivation of r5 and r6-specific MafB expression (Figure 3.3D-E). We found that MafB expression could be detected in r3 and in the posterior hindbrains of kr/kr embryos that had been exposed to the histone deacetylase (HDAC) inhibitor trichostatin A (TSA) in utero at E7.75-8.25 (Figure 3.3E), but not in vehicle treated kr/kr control embryos or in kr/kr embryos exposed to TSA after E8.5. By contrast, treatment of kr/kr embryos with the DNA methyltransferase inhibitor 5-aza-2'deoxycytidine (5-AzadC) led to higher levels of MafB expression in r3 (Figure 3.3D), but did not reactivate expression in the posterior hindbrain. Thus, under permissive chromatin conditions, the S5 enhancer, which is located about -21 kb from the MafB transcriptional start site (and any other rhombomere-specific regulatory regions not separated from the MafB
transcriptional unit by the kr inversion), is sufficient to activate MafB expression in the posterior hindbrains of kr/kr embryos.

**The M33 polycomb gene regulates MafB expression**

If a PRE affects access to rhombomere-specific MafB enhancers, then PcG proteins should regulate MafB expression. To test whether PcG genes might affect rhombomere-specific MafB expression, we chose to examine the consequences of reducing the levels of the M33 gene, which is a murine homologue of the *Drosophila Polycomb* gene. M33 was a particularly relevant candidate MafB regulator, because in M33-/- embryos RA responsiveness is enhanced and expression of a MafB target gene, Hoxb3, is shifted anteriorly at 8.0 dpc (Akasaka et al., 1996; Core et al., 1997; Takihara et al., 1997). Using RNA in situ hybridization on whole embryos as well as immunofluorescence with an anti-MafB antibody (Figure 3.3F-P, data not shown), we found that MafB expression was affected in a variable manner in both M33-/- and M33-/+ embryos. In some embryos, MafB-expressing (MafB+) cell clusters could be detected anterior of the r4/r5 boundary (Figure 3.3F, H-I, L-K, arrow in L’), in others cell clusters lacking MafB expression (MafB-) could be found in r5 and r6 (Figure 3.3G-K, arrowhead in K’). Often MafB expression was suppressed in the dorsal portion of rhombomeres 5 and 6 (brackets in Figure 3.3G, J, M, and G’). Finally, some M33-/- embryos showed ectopic MafB activation in r3 (* in Figure 3.3K) similar to that seen in kr/kr and kr/+ embryos (Figures 3.1, Sadl et al., 2003). Thus, MafB expression is directly or indirectly governed by M33.

We reasoned that, if restricting MafB expression to specific rhombomeres requires PcG/trxG dependent chromatin remodeling, then reducing levels of the PcG proteins involved might reactivate MafB expression in the posterior hindbrains of kr/kr embryos. Therefore, we next tested whether reducing M33 gene dosage might alter MafB expression in mice homo- or heterozygous for the kr inversion. Overall MafB expression in kr/+; M33-/+ and kr/+; M33-/- embryos (Figure 3.3L-N) was as variable as had been seen in M33+/+ and M33-/-, animals that did not carry the kr inversion. In kr/kr; M33+/embryos, four out of five embryos showed no reactivation of MafB expression (Figure 3.3O), but strikingly one out of five showed reactivation of MafB expression in the r5-6 domain (Figure 3.3P). This reactivated domain of MafB expression did not appear
Figure 3.3. Chromatin remodeling agents and M33 levels affect MafB expression in normal and kr/kr embryos.
(A-P) Dorsal views of whole mount in situ hybridization with MafB RNA. Normal (+/+) embryos at E8.0 (A) and E8.5 (B) show expression in the future r5 and r6 domain, but kr/kr embryos (C) express MafB only at very low levels in r3.
(D) In kr/kr embryos treated with 5-Azadeoxycytidine (AzadC), MafB expression was enhanced in r3. (E) Treating with trichostatin A (TSA) reactivated MafB expression in r5-r6 and enhanced ectopic r3 expression. Continued on following page.
Figure 3.3. continued. (L-P) M33 shows a genetic interaction with MafB in compound mutants of M33 and kr. In kr/+; M33+/- embryos (L-N), MafB expression was as variable in r5 and r6 as in M33+/- embryos. As in kr/kr embryos (C), 4 out of 5 kr/kr;M33+/- compound mutant embryos (O) showed no expression in r5-r6. (P, P’) Reactivation of MafB expression was seen in 1 out of 5 kr/kr;M33+/- embryo. In G’, K’, L’ and P’, loss of MafB in cell clusters within r5-r6 is indicated with arrowheads; cell clusters with ectopic MafB expression are marked with arrows; and loss of MafB expression at the dorsal edges of the hindbrain is marked with brackets. Ectopic expression of MafB in r3 is designated by an asterisk. G’, K’, L’ and P’ show higher magnification views of G, K, L and P. Scale bar 100μm.
completely normal; it did not have the crisp stripe appearance observed in normal embryos, but rather groups of \textit{MafB}+ cells were present anterior of the presumptive r4/r5 boundary and \textit{MafB}- cells were present within the r5-r6 domain (Figure 3.3P,P’). Thus, decreasing \textit{M33} levels can lead to reactivation of \textit{MafB} from the \textit{kr} inversion allele. This observation is consistent with a model where the \textit{kr} inversion removes portions of a PRE, which normally facilitates access of PcG/trxG protein complexes to a rhombomere-specific enhancer, such as S5, and thereby promotes precise segmental gene expression, while leaving other regions with repressive functions (Figure 3.2B). It was not surprising that we observed reactivation in only one of five embryos since loss of one copy of \textit{M33} would facilitate access, but not open the locus entirely. Also, \textit{M33} is one of several PC homologs in the mouse, and is itself a member of a complex of multiple polycomb proteins. Thus, one would expect that in most cases, there would be redundant function from other PRC1 members.

**Sequence analyses identifies a potential polycomb response element**

In \textit{Drosophila}, all validated PREs contain binding sites for the pleiohomeotic (PHO) PcG protein and GAGA factor (GAF) (Figure 3.4A). A few bioinformatics programs have been developed to identify sequences that might contain possible PREs, but none of these identified any potential PREs 5’ of \textit{MafB}. However, close inspection of the region spanning the breakpoint revealed several interesting features. First, a 474 bp region near the distal breakpoint (bpt2) shares 92% sequence identity between mouse and human genomic \textit{MafB} DNA and 82% between mouse and chick DNA (Figure 3.4C). This conserved region contains no known or predicted transcripts, but contains consensus binding sites for Pho/YY1 and GAGA factor. Further sequence analysis that compared this highly conserved region dubbed Human homology region 1 (Huhom1) with other vertebrates revealed a minimum of 92% identity between all mammals and 80-82% identity with birds and marsupials (Table 3.1). Strikingly, a palindromic double PHO binding site is found within this highly conserved region, which contains conserved core elements found in several PREs in the fly (Figure 3.5) (Brown et al., 1998; Horard et al., 2000; Mihaly et al., 1998). In particular, this PHO core within Huhom1 is most similar to the \textit{en(a)}, one of two PREs found 5’ of the \textit{engrailed} gene and the \textit{Mcp} PRE which regulates the \textit{bithorax} complex (Figure 3.5). A comparison of this PHO binding site
Figure 3.4. PRE1 and hcPRE1 Exhibit Hallmarks of a Polycomb Response Element.
(A) DNA motifs found in PRE/TREs. Pleiohomeotic/Pleiohomeotic-like (Pho/Phol) core consensus; GAGA factor binding site (GAF); Extended Pho consensus (MPho) according to Mihaly et al., 1998; Extended GAGA site (G10).
(B) Occurrence of motifs in PRE1 and hcPRE1. Highlighted in yellow are Human homology region 1 (HuHom1) which is 82% sequence identical between mouse/chick and 92% between mouse/human DNA. Human homology region 2 (HuHom2) is 80% identical between mouse/human. kr distal breakpoint (bpt2) marked by arrow. A double palindromic Pho site within HuHom1 is indicated by *.
(C) Comparison of highly conserved sequence present in the hcPRE1 region (marked in yellow in (B)) from mouse, human and chick.
<table>
<thead>
<tr>
<th>Species</th>
<th>Chromosome</th>
<th>Huhom1 identity (%)</th>
<th>Sequence alignment of PHO core and surrounding nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>2</td>
<td>92.3</td>
<td>CAGAGAG---ACCAGCCATGGCTTCTCCTTTAG</td>
</tr>
<tr>
<td>Human</td>
<td>20</td>
<td>92.3</td>
<td>CAGAGGCTGA-AGCTGCC---GGCTTCTCCTCCCAA</td>
</tr>
<tr>
<td>Chimp</td>
<td>20</td>
<td>92.3</td>
<td>CAGAGGCTGA-AGCTGCC---GGCTTCTCCTCCCAA</td>
</tr>
<tr>
<td>Orangutan</td>
<td>20</td>
<td>92.5</td>
<td>CAGAGGCTTG-AGCTGCC---GGCTTCTCCTCCCAA</td>
</tr>
<tr>
<td>Rhesus</td>
<td>10</td>
<td>93.4</td>
<td>CAGAGGCTTG-AGCTGCC---GGCTTCTCCTCCCAA</td>
</tr>
<tr>
<td>Marmoset</td>
<td>Contig2966</td>
<td>93</td>
<td>CAGAGGCTTG-AGCTGCC---GGCTTCTCCTCCCAA</td>
</tr>
<tr>
<td>Rat</td>
<td>3</td>
<td>97.3</td>
<td>CAGAGGAG---ACCAGCCAAAGGCTTCTCCTCCAG</td>
</tr>
<tr>
<td>Guinea pig</td>
<td>45</td>
<td>93.4</td>
<td>TTGCAAG---GCAGCCAGGGCTTCTCCTCAA</td>
</tr>
<tr>
<td>Cat</td>
<td>164456</td>
<td>93</td>
<td>CAGAGGACAG---CCAGCCAGGGCTTCTCCTCAA</td>
</tr>
<tr>
<td>Dog</td>
<td>24</td>
<td>94</td>
<td>CAGAGGACGAGCCAGCCAGGGCTTCTCCTCCAG</td>
</tr>
<tr>
<td>Horse</td>
<td>22</td>
<td>93.8</td>
<td>CTGAGACTCGCCAGAGCCAGGGCTTCTCCTCCAG</td>
</tr>
<tr>
<td>Cow</td>
<td>13</td>
<td>92</td>
<td>CGGAGAGA---</td>
</tr>
<tr>
<td>Opossum</td>
<td>1</td>
<td>80.3</td>
<td>None</td>
</tr>
<tr>
<td>Platypus</td>
<td>Ultra337</td>
<td>82.9</td>
<td>None</td>
</tr>
<tr>
<td>Chick</td>
<td>20</td>
<td>82.3 in 288bp</td>
<td>None</td>
</tr>
<tr>
<td>Xenopus</td>
<td>38</td>
<td>85.6 in 108bp</td>
<td>None</td>
</tr>
<tr>
<td>Xenopus</td>
<td>9</td>
<td>62.5 in 112bp</td>
<td>None</td>
</tr>
<tr>
<td>Xenopus</td>
<td>853</td>
<td>72 of 25 bp</td>
<td>CAGCTTTGAGCCAGCCATGGCTTCTCCTCC</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>Small</td>
<td>stretches &lt;31 bp of identity throughout genome</td>
<td>None</td>
</tr>
</tbody>
</table>

**Table 3.1. Sequence comparison of hcPRE1 across vertebrate species.**

Unless otherwise stated, sequence identity is indicated as a percentage of the complete 474 bp between mouse and other vertebrate species of Human homology region 1 (Huhom1) upstream of presumed *MafB* homolog. PRE consensus core as defined by Mihaly et al., 1998 (Figure 3.5) with surrounding sequence is compared across the species. Conservation is highlighted in yellow.
**Figure 3.5.** PRE1 contains a conserved sequence motif in Polycomb Response Elements identified in *Drosophila.*

All of the PREs listed were identified by at least one of three functional assays: Maintenance assay, pairing-sensitive assay and salivary gland assay. The PRE consensus is highlighted in yellow. Other nucleotides shared with PRE1 are in green. Position 1 refers to the first nucleotide in the YY1 consensus shown at the bottom of the figure. Adapted from Mihaly et al., 1998.
across vertebrates indicated that although the PHO binding site is not conserved amongst all mammals, the surrounding regions are highly conserved (Table 3.1). Such conservation suggests that this region may play an important functional role that must be preserved during evolution, though the exact site of binding and thereby the complement of PcG/trxG proteins recruited may impart specificity to the PRE (Loots and Ovcharenko, 2004). Taken together, these observations suggest that this region might play an important functional role in the regulation of MafB expression, possibly acting as a binding site for PcG/trxG complexes. For these reasons and those that will become clear below, we dubbed the ~9 kb spanning the distal inversion breakpoint PRE1.

The PRE1 region can silence gene expression in Drosophila

To test whether this region might act as a PRE we used Drosophila-based in vivo assays, which have allowed for confirmation and identification of a variety of fly PREs, the H19 imprinting region from mouse, a retroviral silencer and the Fab7 silencer of the bithorax complex in Drosophila (Brenton et al., 1998; Cavalli and Paro, 1998b; Lyko et al., 1997; Pannell et al., 2000b). We generated flies transgenic for PRE1-LacZ reporter constructs consisting of LoxP sites flanking the 9kb candidate PRE1 followed by a LacZ reporter, which was under the control of six Gal4 responsive UAS elements (Figure 3.6). A mini-white cassette within the construct acted both as a transformation marker and as a qualitative reporter. Silencers often suppress the expression of the adjacent mini-white gene to some degree, and the lines carrying the PRE1 element had pale orange eyes rather than the brick red of control lines (Figure 3.7A). To quantify this effect, we crossed five independent lines mapped to 3 separate chromosomes of PRE1-LacZ transgenic flies with 'driver' lines that expressed the yeast Gal4 transcriptional activator in salivary glands of wandering third instar larvae, and performed biochemical assays on whole larvae to assess silencing of the LacZ reporter (Figure 3.7B). All five independent PRE-LacZ lines showed repression of both LacZ and mini-white. When the PRE1 sequences were removed from these transgenic lines by Cre-recombinase-mediated excision, a substantial recovery of β-galactosidase activity was seen and the eyes of these flies were darker than those of their PRE1 carrying parents and grandparents (Figure 3.7A, B). This was not seen in control lines carrying the r5+r6 specific S5 enhancer. Thus, PRE1 acts as a silencer in these assays. We mapped the region containing silencing activity using three
Figure 3.6. Diagram of reporter construct used to generate transgenic flies. (A) The loxP-flanked candidate PRE is 5' of a LacZ reporter driven by five copies of the GAL4 responsive UAS element (UAS_s). A mini-white cassette acts as a marker of integration and a secondary reporter of silencing activity. (B) Cre recombinase mediated excision can remove the PRE.
Figure 3.7. PRE1 acts as a developmental silencer in *Drosophila melanogaster*. (A) In eyes of transgenic pUZPRE1 flies, *mini-white* expression is repressed, but recovered upon Cre-recombinase mediated PRE1 excision (pUZΔPRE1). Eye colors of pUZΔPRE1 and pUZ vector control flies are comparable. (B) Both full-length PRE1 (pUZPRE1) and hcPRE1 (pUZhcPRE1) suppress LacZ expression. Five independent pUZPRE1 and pUZhcPRE1 lines are shown in orange. Companion Cre-excised lines are shown in red. In S5 control lines (pUZS5), LacZ was not suppressed before or after excision. Each bar represents percentage of control LacZ activity over mg of whole protein ±SE.
overlapping 3-4kb fragments to a 3kb region containing the conserved PRE1 region and the distal inversion breakpoint, henceforth referred to as hcPRE1 (Figure 3.4).

**Specific PcG and trxG proteins modulate PRE1 function.**

Studies in flies have suggested that distinct subsets of PcG proteins might function to regulate anteriorly expressed *Hox* genes (Gellon et al., 1997; Rusch and Kaufman, 2000). Thus, only a subset of PcG/trxG genes might govern PRE1 dependent chromatin remodeling. Using these studies as a guide, we tested mutations in *Posterior Sex Combs (Psc)*, *Polycomb (Pc)*, *Pleiohomeotic (Pho)*, *polyhomeotic (ph)*, *trithorax-like (Trl)*, *moira (mor)*, *brahma (brm)*, and, as a negative control, the *Histone2a variant (His2Av)*, which affects general chromatin compaction (Swaminathan et al., 2005), on PRE1 dependent silencing in PRE1-LacZ or hcPRE1-LacZ transgenic fly larvae. In larvae, loss of a single copy of *Pc*, *ph*, or *Pho* resulted in significant reactivation of *LacZ* expression in either PRE1-LacZ or hcPRE1-LacZ lines (Figure 3.8B-E). In larvae lacking one copy of *Psc*, LacZ levels could not be assayed due to lethality of the GFP-GAL4 driver on chromosome III (Figure 3.16 and Figure 3.17). Instead, expression of the adjacent mini-*white* reporter gene was found to be reactivated in adult fly eyes (Figure 3.8A). In contrast, loss of a single copy of the activating trxG gene *Trl* resulted in increased suppression of *LacZ* (Figure 3.9). Mutations in *moira*, *brahma*, or *His2Av* had no detectable effect on LacZ activity (Figure 3.10 A-C). Thus, just as in the fly, only a subset of PcG/trxG proteins regulates PRE1 dependent gene expression, which lends further support to the specificity of PcG/trxG proteins with individual PREs.
Figure 3.8. PRE1-dependent silencing is relieved by mutations in PcG genes. (A) A mutation in *Posterior sex combs* (*Psc*) affects PRE1-dependent silencing of adult eye color. LacZ activity of pUZPRE1 (B, D-E) and pUZhcPRE1 (C) transgenic larvae that were either wildtype or heterozygous for PcG mutant alleles were assayed. Each bar represents relative LacZ activity per mg of whole protein of one larva. Red bars represent larvae heterozygous for the mutant, whereas yellow bars are wildtype sibling controls. Heterozygosity of mutations in *Pc, pho, ph* or *Psc* significantly alleviates PRE1-dependent silencing (*p*<0.001).
Figure 3.9. PRE1-dependent silencing is enhanced by mutations in trxG genes.
Loss of Trl further suppresses LacZ activity (p<0.001). Each bar represents relative LacZ activity per mg of whole protein of one larva. Red bars represent larvae heterozygous for the Trl, whereas yellow bars are wildtype sibling controls.
Figure 3.10. Effects of PcG/trxG mutations on PRE1-dependent silencing is specific.
Reducing Brahma (Brm) (A), moira (mor) (B) or His2Av (C) does not significantly alter PRE1-dependent silencing. Each bar represents relative LacZ activity per mg of whole protein of one larva. Red bars represent larvae heterozygous for the mutant, whereas yellow bars are wildtype sibling controls.
PRE1 can recruit Pc and PH binding in Drosophila

To test whether PRE1 and hcPRE1 could recruit PcG proteins, we combined fluorescent in situ hybridization with immunostaining (iFISH) of polytene chromosomes of PRE1-LacZ or hcPRE1-LacZ transgenic flies (Figure 3.11). Using this assay, Negre et al. previously showed that a Fab-7 transgene could recruit PcG protein binding and also validated additional PREs identified by Chromatin immunoprecipitation (ChIP) with anti-PC and PH antibodies (Negre et al., 2006). In flies homozygous for the PRE1-LacZ (Figure 3.11A, C) or hcPRE1-LacZ (Figure 3.11E, G) transgenes, we observed a novel binding site of both PC and PH to the transgene insertion sites. However, upon Cre-mediated excision of the PRE1 sequences, these new PC and PH binding sites were lost (Figure 3.11B, D). The smaller 3kb hcPRE1 region could also effectively recruit both PC and PH binding to its insertion site (Figure 3.11E, G) and again Cre-dependent excision of hcPRE1 sequences abrogated PH and PC binding (Figure 3.11F, H). Thus, in Drosophila PRE1 acts a definitive PRC1 protein binding target.

Analysis of PRC2 binding in the mouse F9 embryonic carcinoma cell line

Next we tested whether PRE1 binds PcG proteins in mouse cells. Regrettably, at the time of this analysis, none of the existing M33 antibodies were of sufficient quality for chromatin immunoprecipitation experiments. So, we chose to analyze the binding of the PRC2 core component SUZ12 in F9 embryonic carcinoma cells, because a high quality antibody made these analyses possible and because only SUZ12 binding has been profiled in the whole human genome in human ES cells. All other PRC1/2 profiling studies have been performed using more limited promoter microarrays and so could not have examined PRE1 occupancy. Because it is difficult to analyze SUZ12 binding in the very small population of MafB expressing cells in r5 and r6, we examined SUZ12 occupancy of PRE1 in F9 embryonic carcinoma cells, in which retinoic acid can induce the expression of anterior Hox genes, including Hoxa1 and the MafB target genes, Hoxa3 and b3. After establishing that retinoic acid could be induce high levels of MafB protein expression by Western analysis (data not shown), we examined binding of SUZ12, H3K27me3, H3K4me2 and RNA polIII in vehicle and retinoic acid treated F9 cells by Chromatin immunoprecipitation (ChIP), followed by subsequent real-time PCR analyses.
Figure 3.11. PRE1 transgene recruits ectopic Polycomb and polyhomeotic binding in *Drosophila*.
Immuno-fluorescent *in situ* hybridization (FISH) was performed on polytene chromosomes from transgenic flies using antibodies against Polyhomeotic (ph) (A-B, E-F) or Polycomb (Pc) (C-D, G-H). Transgene insertion sites for pUZPRE1 (PRE1) (A-D) and pUZhcPRE1 (hcPRE1) (E-H) were visualized with biotinylated pUZ DNA probe. FISH signal was detected even when Cre-recombinase had excised the PRE1 (ΔPRE1) (B, D) or the hcPRE1 (ΔhcPRE1) (F, H) element, but only in the presence of the PRE1 (A, C) or hcPRE1 (E, G) could binding of ph (A, E) or Pc (C, G) be detected. In each experiment, immunostaining (green), FISH (red) and an overlay (orange) are shown. In immunostaining and merge panels, the position of the transgene is marked by arrows.
We detected a peak of SUZ12 binding at PRE1, which was significantly decreased upon RA treatment. We did not detect SUZ12 binding in the vicinity of the \textit{MafB} promoter, near the S5 enhancer or –70 kb from the transcriptional start site (Figure 3.12B). These observations are consistent with PRE1 specific binding of SUZ12. PRC2 complexes catalyze H3K27-specific trimethylation. While the binding of SUZ12 was present only at PRE1, the H3K27me3 signature blanketed the \textit{MafB} locus in untreated F9 cells more generally, showing a less dramatic peak of binding at the PRE1b subdomain of PRE1 and the highest binding at the inactive \textit{MafB} promoter. Upon RA treatment the H3K27me3 signature was decreased dramatically at the \textit{MafB} promoter, as would be expected. However, the peak centered around PRE1b remained and enriched H3K27me3 binding was detected in the PRE1d subdomain of PRE1 and in the vicinity of the S5 enhancer (Figure 3.12C).

Members of trxG proteins, including Mll and ASH2 catalyze formation of the H3K4me2 and H3K4me3 marks (Steward et al., 2006; Wysocka et al., 2005). RA treatment did not alter the levels of the H3K4me2 signature at the established \textit{MafB} promoter, the S5 enhancer or the PRE1a-c regions. Strikingly, upon RA treatment, a dramatic decrease in the H3K4me2 mark was detected in the PRE1d subdomain, which appeared reciprocal to the increase in H3K27me3 binding seen here upon RA treatment (Figure 3.12D). Overall these findings fit with previous analyses in flies, mice and human cell lines of PcG target sites. However, the counterintuitive reciprocal H3K4me2 and H3K27me3 signatures in the PRE1d subdomain was unexpected and may in the future provide further clues as to how PRE1 specifically and perhaps vertebrate PREs in general may function.

Finally, RNA polII did not bind to PRE1, and showed equivalent levels of binding in RA treated and vehicle treated cells at the established \textit{MafB} promoter. Surprisingly, RNA polII binding near the S5 enhancer was dramatically increased upon RA treatment (Figure 3.12E). These findings may suggest that an RA inducible alternative promoter for \textit{MafB} might be located in the vicinity of S5.
Figure 3.12. Binding profiles of the PRC2 component SUZ12, RNA polymerase II, and methylated forms of Histone 3 at PRE1 in F9 cells.

(A) Schematic of MafB locus with Chromatin Immunoprecipitation (ChIP) primers. 1a-1d represent overlapping regions of PRE1, highly conserved region (hc) is highlighted in grey and arrow denotes breakpoint 2 (bpt2). Control primers include -70 kb upstream (-70k) of MafB start, S5 induction element (S5) and MafB promoter region (pr). (B-E) ChIP experiments with anti-SUZ12 (B), anti-H3K27me3 (C), anti-H3K4me2 (D) and anti RNAPolII (E) antibodies. Chromatin samples were taken from F9 cells treated with either retinoic acid (RA) (black bars) or control media (grey bars) for 12 hours. Bound DNA was quantified by real time PCR. Mouse GAPDH promoter primers were used as an external control. The data are represented as a percentage of input DNA prior to immunoprecipitation (mean ± SE). A 2-way ANOVA was performed giving significant values of P<0.0001 (***), P<0.001 (**) and P<0.05 (*).
PRE function in transgenic mice

PREs are thought to act as silencers in their ground states. So, we examined the effect of hcPRE1 on S5-directed \( \text{LacZ} \) expression in transgenic mice. The rhombomere-specific S5 enhancer directed \( \text{LacZ} \) expression in r5 and r6 in 12 of 12 mouse lines reporter transgene (Figure 3.13A) (Kim et al., 2005). We generated 5 transgenic mouse lines with PRE1 positioned 5’ of the S5 enhancer-\( \text{LacZ} \) reporter construct, and, in each line, rhombomere-specific \( \text{LacZ} \) expression was suppressed (Figure 3.13B). Because loss of M33 increases the embryo’s sensitivity to retinoic acid, we reasoned that the presence of a single retinoic acid receptor binding element (RARE) might facilitate PRE1-dependent chromatin remodeling, and allow access to the S5 enhancer accessible and rhombomere-specific expression. For these experiments, we chose a well-defined RARE from the mouse \( \text{Hoxa1} \) gene (Frasch et al., 1995; Langston and Gudas, 1992) which shows remarkable evolutionary conservation: it is found in the \( \text{Hoxa1} \) genes of animals ranging from mice, humans and chickens to the cephalochordate \( \text{Amphioxus} \), the closest invertebrate relative of vertebrates. \( \text{Hoxa1} \) is normally expressed up to the r3/r4 boundary in E7.5-8.5 embryos. When this RARE is removed from the endogenous mouse \( \text{Hoxa1} \) gene, overall \( \text{Hoxa1} \) levels and retinoic acid sensitivity in E8.25-8.5 embryos decrease, but no other effects on expression are seen (Gavalas et al., 1998; Studer et al., 1998). In RARE-hcPRE1-S5-LacZ mice, \( \text{LacZ} \) was expressed in r5 and r6 (Figure 3.13C). However, when the RARE, which was flanked by \( \text{Frt} \) sites, was removed by crossing RARE-hcPRE1-S5-LacZ mice with CMV-FLPe mice, which express Flipase ubiquitously, r5-r6 specific \( \text{LacZ} \) expression was lost (Figure 3.13D). Thus, PRE1 in its ground state acts as a silencer, which can be reset in a retinoic acid dependent manner.
Figure 3.13. hcPRE1 silences induction by S5 in transgenic mice.
LacZ staining of E8.25 (8 somite stage) transgenic embryos. (A) S5LacZ transgene directs LacZ expression in r5-r6 starting at 0 somites (S5) in 12 out of 12 transgenic lines. (B) When hcPRE1 is placed upstream of S5 (hcPRE1-S5), LacZ expression is suppressed in embryos from 5 out of 5 transgenic lines. (C) When an early acting RARE (retinoic acid responsive element) from Hoxa1 (Frasch et al., 1995) was placed upstream of hcPRE1 to generate the RARE-hcPRE1-S5, LacZ expression was recovered in all embryos from two of six transgenic lines, but removing the RARE element with Flipase ΔRARE-hcPRE1-S5) extinguished LacZ expression in r5 and r6 (D).
Figure 3.14. Schematic model of the proposed role of PRE1 in MafB expression. PRC1/2 components act in r3 and r4 at very low levels of retinoic acid (RA) to stabilize MafB repression by chromatin remodeling. In r5 and r6, activators, including vHNF1 drive MafB expression, while in the neural tube posterior of r6, unknown inhibitory mechanisms govern S5 dependent repression of MafB.
Discussion

Previously Polycomb response elements have only been authenticated in *Drosophila*, although their presence has been suspected in other organisms. Here, using experiments in mouse cell lines, in mouse mutants and in transgenic analyses in mice and flies, we have identified and validated a PRE, PRE1 in the mouse that meets the criteria for a functional PRE as defined in *Drosophila*. Our studies have revealed some interesting features of PRE1, as a prototypical vertebrate PRE, and may guide identification of further vertebrate PREs and ultimately understanding of their mechanisms of action.

**PRE1 may restrict gene expression to the posterior hindbrain**

PRE1 appears to facilitate expression in the posterior hindbrain, while repressing it in anterior rhombomeres (Figure 3.14). We have shown that serendipitously, in the *kr* inversion, much but not all of hcPRE1 has been removed from the *MafB* gene and placed upstream of the imprinted r3-and r5-specific *Nnat* gene (Figure 3.2). Hence, analysis of rhombomere-specific *Nnat* and *MafB* expression in *kr/+* and *kr/kr* embryos provided an assay for some possible effects of PRE1 during hindbrain patterning. An obvious caveat to our interpretation of the effects of the *kr* allele is that the reciprocal exchange of the upstream regions of *MafB* with those of *Nnat* may have transferred some uncharacterized regulatory elements, such as a weak r3-specific regulatory element from *Nnat* or an early acting inductive RARE from *MafB*, in addition to PRE1. Nonetheless, we observed that translocation of PRE1 eliminates *MafB* expression in r5 and r6, and causes premature and more wide-spread expression of *Nnat* from r5 through r7. Furthermore, we saw reciprocal and inverse effects of this translocation on anterior *MafB* and *Nnat* expression. These observations would suggest that PRE1 may facilitate gene expression in the posterior hindbrain and repress it in anterior rhombomeres.

In addition to the effects of the *kr* inversion, *M33* reduction also leads to the appearance of *MafB*+ cell clusters anterior of its normal r5+r6 domain. Anteriorization due to the loss of a the PcG protein M33 would fit well with earlier observations from both *Drosophila* and mice, where in general, the loss of some PcG genes causes an anterior shift of Hox gene expression patterns (Akasaka et al., 1996; Core et al., 1997; Takihara et al., 1997).
Furthermore, M33 and perhaps other PcG proteins that mediate RA responsiveness at RAREs may play an integral, possibly direct role in this process. This model would help explain why RAREs, such as that which regulates early *Hoxa1* expression, but do not appear to play particularly striking roles in regulating rhombomere-specific *Hox* genes expression in transgenic analyses are still highly conserved, and play critical roles within their endogenous context. This would suggest that one of the roles of such cryptic RAREs may be that they are involved in directing chromatin remodeling.

**PRE1 may regulate MafB initiation**

While enhancers that regulate rhombomere-specific regulation of *Hox* genes and *MafB* have been explored, this is the first study that has focused on the validation of a polycomb response element governing hindbrain segmentation. In order to understand how PRE1 might act during hindbrain patterning, it is important to review regulation of rhombomere-specific expression of *MafB*. In the mouse, rhombomere-specific gene expression of *MafB*, like that of *Hox* genes, involves sequential action of distinct regulatory elements and molecules. Retinoic acid (RA) emanating from the tailbud and FGFs originating from anterior neural organizers help induce rhombomere-specific *MafB* transcription (Figure 1.3) (Grapin-Botton et al., 1998; Marin and Charnay, 2000). In 0-10 somite stage mouse embryos, the S5 enhancer which lacks consensus RA receptor binding sites, regulates *MafB* expression (Kim et al., 2005). Thus, early RA dependence may rely on sites outside of S5 or may be indirect. From the 10 to the 20 somite stage, an autoregulatory mechanism, possibly acting via *MafB* binding sites located within 2 kb of the transcriptional start site maintains *MafB* expression (Huang et al., 2000).

Several observations point toward a role of chromatin remodeling and PRE1 in early *MafB* initiation. First, in the absence of PRE1, the reversible histone deacetylase inhibitor TSA (Yoshida et al., 1990) can only reactivate *MafB* expression from the *kr* allele when administered between E7.75-8.25 (presomitic to 8 somite stages) the time when the homeodomain transcription factor vHNF1 acts to induce *MafB* expression and is transiently expressed in the future r5+r6 region (Kim et al., 2005). Thus, PRE1 may normally facilitate access to S5. Second, an RARE from the *Hoxa1* gene can overcome PRE1 mediated repression of S5 dependent r5- and r6-specific *LacZ* expression in
transgenic mice (Figure 3.13). So, both M33 and PRE1 might control access to S5 during early MafB induction.

The timing of PRE1 function may derive from the composition of the PRC1 complexes that interact with PRE1 and the trxG recruiting properties of transcription factors involved in MafB regulation. Analyses of mouse mutations in M33 and the polyhomeotic homologue Rae28, both of which are expressed ubiquitously during early mouse neural development, suggest that perhaps specific PcG subunits interact with PRE1. As we have shown here, M33 is required for early MafB expression. By contrast, loss of the polyhomeotic homologue Rae28 does not affect MafB or early Hoxb3 expression, but acts later on to maintain Hoxb3 expression (Takahara et al., 1997). Similarly, studies of vertebrate neuronal differentiation indicate that different trxG containing complexes and, with regards to SWI/SNF containing complexes, a specific combination of subunits rather than the simple presence of individual subunits provides functional specificity to trxG complexes, thereby governing distinct developmental stages of neural development and differentiation (Lessard et al., 2007; Wu et al., 2007). Thus, as also suggested by our analyses in Drosophila, PRE1 may interact with specific subsets of PcG and trxG proteins. Additionally, transcription factors that regulate rhombomere-specific MafB expression and their ability to recruit trxG complexes, may determine when PRE1-binding PcG complexes can be displaced. In transgenic animals, only in the presence of an RARE, but not on their own, can vHNF1 and its auxiliary cofactors relieve PRE1-dependent repression. Several non-mutually exclusive explanations for these results are: (1) RA receptors may act before vHNF1 to set the ‘open’ or ‘closed’ state of PRE1; (2) because M33 loss increases RA responsiveness of Hox gene expression (Bel-Vialar et al., 2000), a more direct link between M33 dependent gene activation and RA receptors and RAREs may exist; and (3) RA receptors can interact directly with SWI/SNF proteins, and thus, may recruit these trxG proteins to PRE1 and support gene expression (Debril et al., 2004). By contrast, no interactions of any trxG protein with vHNF1 or its homologues have been reported. It remains to be seen whether PRE1-dependent repression can be relieved by transcription factors other than RA receptors or ones that do not interact with SWI/SNF proteins. In the future, defining minimal PRE1 sequences in vivo and examining the composition of PRE1 binding PcG
complexes may help establish what features contribute to this spatial and temporal specificity of PRE1 action.

**Interactions of specific PcG and trxG proteins with PRE1**

While our analyses of mice carrying the *kr* allele or lacking the *M33* Polycomb gene hinted at the presence of a PRE, *Drosophila*-based assays have provided direct evidence that PRE1 and the hePRE1 subdomain in particular, mediate PcG- and trxG-dependent gene expression, and directly recruit Ph and Pc protein binding in *Drosophila*. To our knowledge, this is the first vertebrate regulatory region that has been shown to regulate gene expression in the *Drosophila* in a PcG- and trxG-dependent manner and the only one shown to actually recruit PcG protein. In addition, in mouse F9 embryonic carcinoma cells the binding of the PRC2 core component SUZ12 to the endogenous PRE1 was relieved upon MafB induction with RA (Figure 3.12B). These results underscore the high level of conservation among the PcG/trxG machinery and regulatory elements.

Specifically, we have shown that PRE1 binds core components of PRC1 and 2, and potentially also trxG complexes in *Drosophila* and in mouse F9 cells. PcG proteins that act as core components of the PRC1 complex, including Posterior Sex Combs, Polyhomeotic and Polycomb, regulated PRE1 dependent gene silencing and Ph and Pc could be recruited to PRE1 transgene insertion sites in *Drosophila* (Figures 3.8 and 3.11). Additionally, in the F9 mouse cell line, the PRC2 component SUZ12 was recruited specifically to PRE1 and displaced upon RA-dependent *MafB* induction (Figure 3.12B). Finally, homologs of trithorax proteins and probably the trithorax protein itself act as histone methyltransferases for lysine 4 of histone H3 (Milne et al., 2002; Roguev et al., 2001), (Nagy et al, 2002, Nakamura et al, 2002). While no ChIP-quality antibodies against mouse trithorax homologues exist, indirect evidence for an interaction of trithorax proteins with PRE1 comes from our finding that the H3K4me2 signature was altered specifically at PRE1 (Figure 3.12D). The functions of Trx and these PRC1 components appear to be interlinked with the methylation state of histone tails and DNA. Thus, PRE1 and possibly other vertebrate PREs may act as entry points for PRC1/2 and basic trxG components, which act on histone methylation states and, less directly, on DNA methylation states.
While all core components of PRC1 tested in flies interacted with PRE1, trxG protein interactions with PRE1, as with PREs in general, showed greater selectivity. Brahma and Moira, which encode components of the SWI/SNF-related ATP-dependent Brahma chromatin remodeling complex (Crosby et al., 1999; Tamkun et al., 1992) which alters the position of nucleosomes or their interactions with DNA, did not affect PRE1-dependent gene expression in Drosophila (Figure 3.10). We chose to examine the effects of mutations in moira and brahma, because they were among the few trxG genes that affected the expression of Deformed, a parologue of mouse Hox4 genes, which are expressed in and up to the r6/r7 boundary in the hindbrain. Our results suggest that the roles of trxG proteins may not have been preserved in any simple fashion along the anterior-posterior axes of animals. In addition, transcription factors not present in Drosophila may have binding sites within or near PRE1, and these might be able to recruit SWI/SNF nucleosome remodeling complexes specifically in vertebrates. For example, RA receptors, which are not present in Drosophila, might interact directly with SWI/SNF proteins, and thereby recruit trxG proteins to PRE1.

**PRE1 as a unique entry site for PcG/trxG proteins**

All of our observations of PRE1 interactions with PcG and TrxG complexes in Drosophila and of SUZ12 in F9 cells are consistent with previous functional analyses of PREs in Drosophila which suggested that PcG complexes bind to discrete sequence platforms (Beisel et al., 2007; Kobrossy et al., 2006). At present, the best characterized PREs are located within the Drosophila Hox clusters, and Hox gene translocation experiments in the mouse have suggested the presence of PREs in vertebrates (Kmita et al., 2002; Roelen et al., 2002; Spitz et al., 2005). Thus, efforts have focused on inspecting human and murine Hox gene clusters for possible PcG target sites. However, in contrast to our observation of a discrete peak of SUZ12 binding within the PRE1 region of MafB, PcG complexes were found to blanket entire Hox clusters, bind specific peaks embedded within these blanketed regions or not bind at all, depending on the type and differentiational state of the cell line analyzed (Bracken et al., 2006; Lee et al., 2006; Squazzo et al., 2006). For example, SUZ12 and EZH2 binding broadly blankets the HoxA gene cluster in F9 cells, embryonic human Ntera2 cells, and human embryonic stem cells, but is absent in adult tumor lines (Lee et al., 2006; Squazzo et al., 2006). Thus,
vertebrate Hox clusters provide a valuable but challengingly complex substrate for the identification PREs and none of the potential PcG target sites within vertebrate Hox clusters have been validated as true PREs. By contrast, segmentally expressed genes that, like MafB, are not part of clusters and show distinct peaks of PcG binding may prove more suitable for the identification and analysis of further PREs.

**Comparison of PRE1, predicted vertebrate PcG target sites and Drosophila PREs**

Recent genome-wide profiling strategies have relied on chromatin immunoprecipitation (ChIP) using antibodies against PRC1 and 2 components and various histone signatures to identify potential PcG target sites in mouse and human cell lines. The PRC1/2 components examined in this manner have been limited by the scarcity of ChIP-quality antibodies. Nonetheless, support for the validity of some target sites has come from analyses that showed binding of multiple PRC2 components to the same region and deregulated expression of some Eed or Suz12 bound targets in cells lacking the respective PcG gene (Boyer et al., 2006; Lee et al., 2006). The majority of PcG-binding sites identified in vertebrates reside within 1 kb of promoters, while in Drosophila only 30% are within 2 kb of the promoter and the rest are several tens of kilobases away. These apparent differences may reflect a sampling bias caused by the predominant use of undifferentiated vertebrate ES cells in contrast to the pseudo-differentiated cell types and embryos examined in the Drosophila, and by the use of mouse and human microarrays that, with one exception, span maximally -8 to +2 kb of known or predicted promoters (Boyer et al., 2006).

**Histone profiles hint at mechanisms governing PRE1 function**

Our analysis of histone signatures may elucidate additional transcriptional mechanisms regulating PRE1 function and MafB expression. Intriguingly, the decreased H3K4me2 signature and increased H3K27me3 mark at the PRE1d subdomain in response to RA treatment (Figure 3.12) may provide clues to how PRE1 may function, and may fit well with possible transcriptional interference mechanisms as seen in the bxd PRE governing Ubx transcription (Petruk et al., 2006). In Drosophila, transcription of non-coding untranslated RNAs at PREs plays an important, incompletely understood, and perhaps not universal role in PRE-dependent silencing. Petruk et al. found that transcription of untranslated RNAs from the bxd PRE repress Ubx expression in vivo.
(Petruk et al., 2007; Petruk et al., 2006). Thus, the highly conserved non-coding region might flag a region containing a non-coding RNA important in PRE1 dependent silencing.

Is PRE1 an archetypal PRE or a vertebrate innovation?

Hindbrain segmentation is one of the key innovations underlying vertebrate head evolution. Thus, comparative genomic analyses of the hcPRE1 sequence might provide insights into the sequence requirements for vertebrate PREs and also address whether PRE1, like hindbrain segmentation, is a vertebrate innovation or a modification of an archetypal PRE. The hcPRE1 region is present upstream of all mammalian MafB orthologues and also of the chick MafB gene (Figure 3.4), but was not recognized by homology searches in the genome of zebrafish, in which the neuronal hindbrain architecture, while segmented, is significantly less complex (Table 3.1). The origin of PRE1 cannot be addressed in Drosophila since the Hox3 gene, zerknult (zen), and homologues of MafB are not expressed segmentally in the fly (Falciani et al., 1996)(S. Cordes, unpublished observations). However, future analyses of the cephalochordate Amphioxus, the closest living invertebrate relative of vertebrates, may prove useful. While not visibly segmented, the Amphioxus nerve cord posterior of the cerebral vesicle approximately down to the level of the eighth somite corresponds to the vertebrate hindbrain (Holland and Holland, 1999; Williams and Holland, 1998). Amphioxus is the only invertebrate in which RA signaling is interlinked with Hox gene expression and anterior-posterior patterning (Castro and Holland, 2003; Garcia-Fernandez and Holland, 1994; Holland et al., 2008; Marletaz et al., 2006; Wada et al., 1999). Importantly, anterior expression of the Amphioxus Hox3 gene extends up to somite 4, a region roughly equivalent to the r4/r5 boundary respected by the Hox3 class of genes in vertebrates (Schubert et al., 2006). Expression of the MafB Amphioxus homologue, AmphiMafB remains to be examined (Holland et al., 2008), but AmphiKrox20, the Amphioxus homologue of Krox20, a vertebrate segmentation gene expressed coincidentally with MafB in r5, shows no expression in the Amphioxus nerve cord (Knight et al., 2000). Thus, Amphioxus may be hovering at a point in evolution just before hindbrain segmentation has emerged. BLAST searches of the Amphioxus genomic sequence did not identify any sequences related to hcPRE1, and thus it is possible that PRE1 and the
role of MafB in anterior-posterior patterning may have emerged together in vertebrates. Ultimately, analyses of the functional PRE1 sequence requirements and the role of AmphimafB in anterior-posterior patterning may resolve the origins of PRE1 more definitively.

**Implications for neurodevelopment disorders**

Finally, our observations have important implications for human disorders. Here the self-sorting behavior of rhombomeric cells has revealed that reducing M33 levels causes variable MafB expression on a cell-by-cell basis (Figure 3.3). This, to our knowledge, is the first report that a ~50% decrease of a PcG gene can cause patchy expression of a key neural patterning gene. Several important neural circuits, including the rhythmic circuitry, which governs breathing, are disrupted in kr/+ and kr/kr mice and often affected in autistic patients, and neuronal subtypes such as the majority of mood modulating serotonergic neurons, are born within the hindbrain (Chatonnet et al., 2002), reviewed in (Cordes, 2005; Fortin et al., 2003). Thus, uneven gene expression during early hindbrain development might sometimes have dramatic neurobiologic consequences. Precedent for epigenetic involvement in neurobiologic disorders comes from the finding that mutations in the X-linked gene encoding methyl-CpG binding protein 2 (MeCP2), which encodes a multifunctional protein that binds methylated DNA sequences, are responsible for the postnatal neurodevelopmental disorder, Rett syndrome (RS; MIM 312750). While possible links between variations in PcG/trxG proteins and neurodevelopmental disorders have not been explored, our findings suggest that variations in PcG gene levels or disruptions of specific PREs might conceivably contribute to such disorders or the variability in their presentation among individuals.

Previously, our in vivo understanding of PcG and PRE mediated cellular memory has come from studies performed exclusively in Drosophila, largely on the PREs contained within the fly Hox gene clusters. Now with the first vertebrate PRE in hand and because PRE1 acts to regulate chromatin remodeling of a developmentally critical gene that is not part of a cluster, we are poised to address some issues that are at the heart of understanding how PcG based chromatin remodeling operates in vertebrates and in animals, in general.
Figure 3.15. Example of mating scheme to test effect of PcG mutants from chromosome III on PRE1-dependent silencing.

Individual larva that carried the PRE1 transgene (UASII) on chromosome II, GAL4 driver (CyOGFP) and a copy of either Polycomb mutant (Pc) or control allele (Ki) were tested for LacZ activity in the CPRG assay. Larvae positive for the GAL4 driver were identified by the presence of GFP. Equal numbers of males and females were collected and tested for each genotype. Curved arrows denote self crossing. See Table 3.3 for exact genotypes of lines used.
Figure 3.16. Original mating scheme to test effect of PcG mutants from chromosome II on PRE1-dependent silencing.

Individual larva that carried the PRE1 transgene (UAS<sup>III</sup>) on chromosome III, GAL4 driver (TM3GFP) and a copy of either Polysexcomb mutant (Psc) or control allele (Elp) were tested for LacZ activity in the CPRG assay. Larvae positive for the GAL4 driver were identified by the presence of GFP. Unfortunately, no GFP larvae could be recovered, so an alternative scheme was used (Figure 3.18) to assay for eye colour. Curved arrows denote self crossing. See Table 3.3 for exact genotypes of lines used.
Collect adult males and assay for eye colour.

**Figure 3.17. Alternative mating scheme to test effect of PcG mutants from chromosome II on PRE1-dependent silencing.**

Three day old adult males were collected to assay for eye colour. Each fly carried the PRE1 transgene on chromosome III (UAS III), and either mutant Psc or sibling control (Sp) on II. See Table 3.5 for exact genotypes of lines used.
Figure 3.18. Mating scheme to test effect of mutation in *ph* from X chromosome on PRE1-dependent silencing.

Individual female larva that carried the PRE1 transgene (UAS), GAL4 driver (CyOGFP) and a copy of *ph* were tested against males that did not carry the mutation for LacZ activity in the CPRG assay. Larvae positive for GAL4 driver were identified by presence of GFP. Equal numbers of males and females were collected and tested for each genotype. Curved arrows denote self crossing. See Table 3.3 for exact genotypes of lines used.
Figure 3.19. Mating scheme to test effect of mutation in pho from chromosome IV on PRE1-dependent silencing.

Individual female larva that carried the PRE1 transgene (UAS\(^X\)) on X, GAL4 driver (CyOGFP) on II for LacZ activity in the CPRG assay. Larvae positive for GAL4 driver were identified by presence of GFP. Presence of pho mutant allele was determined by PCR on larval tails. See Table 3.3 for exact genotypes of lines used.
### Table 3.2. Statistical analysis of the effect of PRE1-dependent silencing in transgenic flies (Figure 3.7).

Graphpad prism analysis using 2 way ANOVA shows $P<0.0001$ for PRE1 and hcPRE1 lines for whether PRE is present or not, S5 has a $P>0.05$ and is not significant. Post test follows for each individual line.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>P value</th>
<th>Mann Whitney U</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pc on PRE1</td>
<td>0.0002</td>
<td>0</td>
<td>***</td>
</tr>
<tr>
<td>Pc on hcPRE1</td>
<td>0.0030</td>
<td>4.0</td>
<td>**</td>
</tr>
<tr>
<td>Ph</td>
<td>0.0159</td>
<td>0</td>
<td>*</td>
</tr>
<tr>
<td>Pho</td>
<td>0.0050</td>
<td>16.00</td>
<td>**</td>
</tr>
<tr>
<td>Trl</td>
<td>0.0001</td>
<td>0.00</td>
<td>***</td>
</tr>
<tr>
<td>Brm</td>
<td>0.2238</td>
<td>16.00</td>
<td>ns</td>
</tr>
<tr>
<td>Mor</td>
<td>1.0000</td>
<td>18.00</td>
<td>ns</td>
</tr>
<tr>
<td>His2Av</td>
<td>0.4559</td>
<td>20.00</td>
<td>ns</td>
</tr>
</tbody>
</table>

### Table 3.3. Statistical analysis of the effect of PcG/trxG mutants on PRE1 silencing

Graphpad prism analysis using nonparametric ranked sum with Mann Whitney test with two tailed p value with a CI of 95%.
<table>
<thead>
<tr>
<th>Antibody</th>
<th>Primer</th>
<th>noRA mean</th>
<th>SE</th>
<th>n</th>
<th>RA mean</th>
<th>SE</th>
<th>n</th>
<th>P-value for RA</th>
<th>post test</th>
<th>Signif</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUZ12</td>
<td>GAPDH</td>
<td>1.05%</td>
<td>0.44%</td>
<td>3</td>
<td>0.80%</td>
<td>0.33%</td>
<td>2</td>
<td>0.7622</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>-70kb</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3</td>
<td>0.3784</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1a</td>
<td>0.21%</td>
<td>1.17%</td>
<td>6</td>
<td>0.07%</td>
<td>0.06%</td>
<td>8</td>
<td>P &lt; 0.0001</td>
<td>P &lt; 0.001</td>
<td>y</td>
</tr>
<tr>
<td></td>
<td>PRE1b</td>
<td>1.97%</td>
<td>0.03%</td>
<td>3</td>
<td>0.41%</td>
<td>0.04%</td>
<td>3</td>
<td>0.0102</td>
<td>P &lt; 0.05</td>
<td>y</td>
</tr>
<tr>
<td></td>
<td>PRE1c</td>
<td>1.74%</td>
<td>0.21%</td>
<td>3</td>
<td>0.74%</td>
<td>0.53%</td>
<td>3</td>
<td>0.0005</td>
<td>P &gt; 0.05</td>
<td>y</td>
</tr>
<tr>
<td></td>
<td>PRE1d</td>
<td>1.04%</td>
<td>0.55%</td>
<td>2</td>
<td>0.20%</td>
<td>0.00%</td>
<td>3</td>
<td>0.0816</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3</td>
<td>0.8914</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>kprp</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3</td>
<td>0.337</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>GAPDH</td>
<td>8.56%</td>
<td>0.76%</td>
<td>3</td>
<td>10.75%</td>
<td>0.23%</td>
<td>3</td>
<td>0.0001</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>-70kb</td>
<td>24.62%</td>
<td>0.67%</td>
<td>3</td>
<td>32.59%</td>
<td>5.89%</td>
<td>3</td>
<td>0.0988</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1a</td>
<td>10.21%</td>
<td>2.86%</td>
<td>2</td>
<td>24.92%</td>
<td>9.84%</td>
<td>2</td>
<td>0.8399</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1b</td>
<td>36.35%</td>
<td>7.96E-02</td>
<td>2</td>
<td>46.64%</td>
<td>6.35%</td>
<td>3</td>
<td>0.0802</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1c</td>
<td>16.88%</td>
<td>1.94%</td>
<td>3</td>
<td>21.24%</td>
<td>4.03%</td>
<td>3</td>
<td>0.0007</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1d</td>
<td>2.71%</td>
<td>2.21%</td>
<td>3</td>
<td>39.59%</td>
<td>11.00%</td>
<td>2</td>
<td>P &lt; 0.0001</td>
<td>P &lt; 0.01</td>
<td>y</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>31.77%</td>
<td>1.49%</td>
<td>3</td>
<td>61.68%</td>
<td>4.48%</td>
<td>2</td>
<td>0.0023</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>kprp</td>
<td>68.25%</td>
<td>0.13951</td>
<td>3</td>
<td>9.80%</td>
<td>4.46%</td>
<td>3</td>
<td>0.2097</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>GAPDH</td>
<td>9.72%</td>
<td>1.82%</td>
<td>3</td>
<td>8.57%</td>
<td>0.98%</td>
<td>3</td>
<td>0.0397</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>-70kb</td>
<td>1.23%</td>
<td>0.07%</td>
<td>3</td>
<td>2.15%</td>
<td>0.15%</td>
<td>2</td>
<td>0.0246</td>
<td>P &lt; 0.001</td>
<td>y</td>
</tr>
<tr>
<td></td>
<td>PRE1a</td>
<td>2.07%</td>
<td>1.58%</td>
<td>2</td>
<td>0.71%</td>
<td>0.51%</td>
<td>2</td>
<td>0.5319</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1b</td>
<td>0.83%</td>
<td>0.16%</td>
<td>3</td>
<td>1.09%</td>
<td>0.15%</td>
<td>2</td>
<td>0.877</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1c</td>
<td>4.38%</td>
<td>1.64%</td>
<td>3</td>
<td>6.34%</td>
<td>2.81%</td>
<td>3</td>
<td>0.8223</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1d</td>
<td>8.18%</td>
<td>1.63%</td>
<td>2</td>
<td>1.21%</td>
<td>0.65%</td>
<td>2</td>
<td>0.0267</td>
<td>P &lt; 0.01</td>
<td>y</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>4.38%</td>
<td>1.64%</td>
<td>3</td>
<td>6.34%</td>
<td>2.81%</td>
<td>3</td>
<td>0.3622</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>kprp</td>
<td>8.00%</td>
<td>3.41%</td>
<td>3</td>
<td>3.89%</td>
<td>0.00%</td>
<td>1</td>
<td>0.0506</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td>RNAPolII</td>
<td>GAPDH</td>
<td>25.94%</td>
<td>5.67%</td>
<td>3</td>
<td>22.01%</td>
<td>4.43%</td>
<td>3</td>
<td>0.6891</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>-70kb</td>
<td>0.60%</td>
<td>0.43%</td>
<td>3</td>
<td>2.90%</td>
<td>1.33%</td>
<td>3</td>
<td>0.6642</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1a</td>
<td>0.08%</td>
<td>1.11E-04</td>
<td>2</td>
<td>0.11%</td>
<td>0.00%</td>
<td>1</td>
<td>0.5522</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1b</td>
<td>2.27%</td>
<td>0.81%</td>
<td>2</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3</td>
<td>0.7382</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1c</td>
<td>0.95%</td>
<td>0.43%</td>
<td>3</td>
<td>0.81%</td>
<td>0.59%</td>
<td>2</td>
<td>0.0387</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>PRE1d</td>
<td>3.09E-04</td>
<td>1.75E-04</td>
<td>3</td>
<td>0.14%</td>
<td>0.14%</td>
<td>3</td>
<td>0.7295</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>0.67%</td>
<td>0.57%</td>
<td>2</td>
<td>38.96%</td>
<td>13.49%</td>
<td>3</td>
<td>P &lt; 0.0001</td>
<td>P &lt; 0.05</td>
<td>y</td>
</tr>
<tr>
<td></td>
<td>kprp</td>
<td>9.80%</td>
<td>3.08%</td>
<td>2</td>
<td>8.80%</td>
<td>0.00%</td>
<td>1</td>
<td>0.2435</td>
<td>P &gt; 0.05</td>
<td>n</td>
</tr>
</tbody>
</table>

Table 3.4. Statistical analysis of ChIP data (Figure 3.12). Graphpad prism analysis using 2 way ANOVA showing P-values when comparing RA to no RA. Bonferroni post test values shown with 95% CI of difference.
<table>
<thead>
<tr>
<th>Name of Line</th>
<th>Also Known As/stock #</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUZ</td>
<td>transgenic construct backbone; UZ/15; P[U/15]-2; pU/15-2</td>
<td></td>
<td>Paro</td>
</tr>
<tr>
<td>Crey'/CyO</td>
<td>766</td>
<td>y[1] w[67c23] P{y[+mDint2]=Crey}1b; nol[ScO]/CyO</td>
<td>Bloomington</td>
</tr>
<tr>
<td>Crey'/TM3</td>
<td>851</td>
<td>y[1] w[67c23] P{y[+mDint2]=Crey}1b; D[^]/TM3, Sb[1]</td>
<td>Bloomington</td>
</tr>
</tbody>
</table>

**Cre Recombinase expressing lines**

<table>
<thead>
<tr>
<th>GAL4 drivers and GFP lines</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1032.hx</strong></td>
</tr>
<tr>
<td>Egfr/CyO; Ki/TM3G</td>
</tr>
</tbody>
</table>

**PcG/trxG mutants**

| Psc[^1]/CyO | Psc; posterior sex combs; #Z708 | cn Psc[IIN48] bw sp/ CyO | Nusslein-Volhard laboratory |
| pho[^b] | pho; pleiohomeotic; 1140 | pho[b]/In(4)ci[D], ci[D] pan[ciD] | Bloomington |
| Ph-d[^503] | ph[^d]; polyhomeotic distal; 1411 | y[1] ph-d[503] w[1] f[36a]/FM7c, sn[+]/Dp(1;2;Y)w[+] | Bloomington |
| Trl[^2325] | Tri; Trithorax-like; 12088 | w[1118]; P{w[+mC]=lacW}Trl[s2325]/TM3, Sb[1] | Bloomington |
| His2Av | His2Av[^05146]; Histone H2A variant; 11650 | ry[506]; P{ry[+t7.2]=PZ}His2Av[05146]/TM3, ry[RK] Sb[1] Ser[1] | Bloomington |

Table 3.5. Lines of *Drosophila melanogaster* used in analysis of PRE1-dependent silencing.
Acknowledgements

For assistance, extra hands and a cool head with mice and flies, I offer much thanks to Joanna Yu and Anson Sing. I am very grateful to A. Karaiskakis, Drs. H.D. Lipshitz, D. Pannell and W. Tadros for teaching me the essentials in all things fly related, from basic husbandry to fancy genetic schemes. My thanks to L. Ming and Dr. G. Cavalli for advice regarding immuno-FISH and Dr. J. Ellis and U. Djuric for guidance with regards to ChIP and qPCR. We are also grateful for the generous gifts of the M33 mice from Dr. M. Djabali and PC and Ph antibodies from Dr. R. Paro. Sandra Tondat, Lois Schwartz, and the Samuel Lunenfeld transgenic core facility were essential for generating transgenic mice for these studies and their efforts are much appreciated.

Footnote on previously published material

Figure 3.1 was previously published in Developmental Dynamics where I am listed as co-first author (Sadl et al., 2003). The bulk of chapter 3 has been submitted to Cell and is under review for publication (September 2008). I am directly and solely responsible for all data collected and presented there including supplementary materials not submitted for publication (Figures 3.5, 3.15-19, Tables 3.1-3.5).
Chapter 4: Discussion

Much of this work has highlighted the fact that the embryonic hindbrain and its clearly defined rhombomeres is a beautiful model for studying mechanisms of anterior-posterior patterning in the vertebrate. Interest in the vertebrate hindbrain came initially from the well-studied role of the *Hox* genes in segmentation and how well conserved these mechanisms were from *Drosophila* to vertebrates. However, the regulation of *Hox* genes in the vertebrate diverges from that in the fly embryonic segments. Our work in the development of the posterior hindbrain, specifically, the regulation of the *MafB*/*Kreisler* segmentation gene is a good example of this.

**Regulatory mechanisms governing *MafB* expression in the hindbrain**

In dissecting the transcriptional regulation of *MafB*/*Kreisler* expression, we have uncovered an r5-r6-specific enhancer, S5, that is dependent on the homeodomain protein vHNF1. We further showed that a second *cis*-regulatory element PRE1 acts as a functional polycomb response element that is likely to determine the accessibility to S5, and thus *MafB* activation. *MafB* is an early acting segmentation gene that directly regulates the activation of *Hoxa3* and *Hoxb3* (Manzanares et al., 1999a; Manzanares et al., 1997), and thus plays an integral role in precisely defining the identity of the posterior hindbrain.

*MafB* is a unique regulator of hindbrain patterning in that depending on the expression pattern of the target, there appear to be different modes of gene activation. For example, *Hoxa3* is expressed as a band in r5 and r6, while its paralog *Hoxb3* is only expressed in r5. Careful deletion analysis of the *Hoxa3* and *Hoxb3* enhancers revealed that there are MafB-specific binding sites in both genes, though there are differences in the number of binding sites and the way that MafB acts at these two control regions (Manzanares et al., 1999a). The *Hoxa3* regulatory region contains a specific r5-r6 enhancer that has a single high-affinity binding site for MafB that when multimerized, is sufficient to drive expression in r5 and r6 in transgenic mice (Manzanares et al., 1999a). In contrast, the *Hoxb3* r5 enhancer is bound by Krox20 (present in r3 and r5), which then synergizes with Mafb to activate *Hoxb3* expression (Manzanares et al., 1997; Manzanares et al., 2002). This would suggest that in the hindbrain, MafB acts at different enhancers.
which probably have a common origin, in distinct fashions, allowing for different expression domains.

We have shown that vHNF1 is necessary, but not sufficient to drive r5-r6 expression of MafB in the developing hindbrain, and thus, additional co-activator(s) may be involved. Several lines of evidence suggest that integrated signals like retinoic acid from the posterior and FGFs from the anterior play important roles in regulating vHNF1 and/or MafB directly. Inhibition experiments in zebrafish *neckless* mutants and targeted knockouts of mouse *RALDH2* have shown that RA signaling is required for the formation of the posterior hindbrain (r5-57) (Begemann et al., 2001; Niederreither et al., 2000). Early MafB induction is dependent on both RA and FGF signaling (Gratin-Botton et al., 1998; Marin and Charnay, 2000), and although no known retinoic acid responsive elements (RAREs) have been found within the r5-r6 specific S5 enhancer, we have found ETS-domain protein binding sites. Many ETS-related transcription factors are known to function downstream of FGF signaling (Sharrocks, 2001). Some have been implicated in regulating neuronal functions; specifically, the ETS genes *pea3* and *erm* are general targets of FGF8 signaling in zebrafish neuronal cell specification (Roehl and Nusslein-Volhard, 2001). Thus, it is possible that FGF induction of MafB may be mediated by ETS-domain proteins at the binding sites present in S5. Retinoid signaling may occur at RAREs outside S5 or RA induction maybe indirect.

**Possible collaborators of vHNF1 in early MafB induction**

Our studies have shown that vHNF1 is a co-activator that may interpret RA and/or FGF signals and be responsible for the clearly defined anterior border of MafB expression. However, the expression pattern of vHNF1 (i.e. from r5 posterior) suggests that a repressor may be involved in the down-regulation of MafB at the r6/r7 border. Further analyses of vHNF1 collaborators and regulators as well as other early expressed transcription factors in the posterior hindbrain should provide us with a more complete picture of how the precise band of MafB expression at r5-r6 is achieved. Possible candidates for posterior repressors include members of the *caudal*-related homeobox family of transcription factors *Cdx*. The family of *Cdx* genes resides in a cluster of homeobox genes closely related to the Hox cluster that function in the formation of the posterior body in both vertebrate and invertebrates (Lohnes, 2003). They have been found
to directly regulate the expression of posterior Hox genes through direct binding at discreet cis-regulatory elements (Gaunt et al., 2004; Isaacs et al., 1998). Additionally, inhibition of Cdx1 and Cdx4 induces ectopic expression of posterior hindbrain and anterior spinal cord markers, including MafB in zebrafish. This ectopic expression was shown to be dependent on both FGF and RA signals suggesting a role for Cdx in modifying tissue responsiveness to these powerful signals (Shimizu et al., 2006). Furthermore, Bel-Vialar et al have shown that genes in the HoxB cluster can be divided into two distinct groups based on their sensitivity to FGF and RA, and that this distinction is mediated by CDX activity (Bel-Vialar et al., 2002). They found that in the early chick neural tube, activation of 3’ HoxB genes (Hoxb2 and Hoxb3-Hoxb5) was sensitive to RA, but not to FGF treatment, while the 5’ genes (Hoxb6-Hoxb9) showed the reciprocal effect. Furthermore, they showed the effect of FGF treatment on 5’ genes was mediated in part by CDX activity which is expressed in the posterior hindbrain and spinal chord (Bel-Vialar et al., 2002). We have found putative Cdx-binding sites within S5. Exploration of these regions and a molecular dissection of the role of Cdx and other possible repressors could shed light on the dynamic and complex mechanisms that govern MafB transcription in the hindbrain.

**Identification of PRE1**

In general, the direct transcriptional regulation of Hox genes does not appear to be well conserved from flies to vertebrates. However the role of PcG/trxG proteins in establishing states of transcriptional competency and maintenance is highly conserved. PcG/trxG proteins were first identified in Drosophila for their role in the establishing and maintaining silenced or activated states of transcription of Homeotic genes. In the mouse, PcG/trxG proteins have also been shown to play an important role in the developing embryo, particularly in the regulation of Hox genes in the hindbrain. Here, we have shown that this is also true for MafB. PcG/trxG proteins operate as large protein-protein complexes via several mechanisms including DNA methylation, histone modification and direct or indirect ATP-dependent chromatin remodeling to name a few. However, these complexes are all thought to bind at specific DNA sequences termed Polycomb responsive elements (PREs).
Previously, PREs have only been found in the fly and although they have been predicted to exist in vertebrates, particularly within the *Hox* gene clusters, attempts at defining and locating functional PREs have been unsuccessful. Here, we have identified and validated PRE1 as the first functional polycomb responsive element found in vertebrates and characterized key features that may aid in the search for future PREs. Using the X-ray induced mutation *kr*, which is an inversion of the *MafB* locus along with its r5-r6 specific enhancer, we found not only a misregulation of *MafB* but also of the imprinted gene *Nnat* which is also translocated intact, most likely with its hindbrain specific enhancers in the *kr* allele. By analyzing the expression patterns of these two genes in mice, as well as the effect of chromatin remodeling agents and the Polycomb mutant *M33* on *MafB*, we hypothesized that the inversion disrupted a possible polycomb responsive element that regulates access to the rhombomeric enhancers of both *MafB* and *Nnat*.

We then tested this element, PRE1 in previously established *Drosophila* assays that were used to identify fly PREs. In these assays we showed that PRE1 acts as a strong silencer of reporter *LacZ* expression as well as of mini-*white*. Furthermore, we showed that this silencing was dependent on the presence of normal Pc, Psc, ph and Pho, while loss of the trithorax protein Trl resulted in further silencing. We then used immuno-FISH to show that the PRE1 transgene can recruit Pc and ph to its site of insertion creating an ectopic PcG binding site. These experiments have shown that mouse PRE1 exhibit the same properties of silencing as fly PREs when taken out of their endogenous context and inserted into another genome.

Finally we validated that PRE1 acts as a functional PRE in mice. With ChIP analysis using antibodies against the mammalian PcG member SUZ12, we found an enriched binding region within PRE1, which is relieved upon RA induction. Finally we showed that PRE1 can silence S5 induction of *LacZ* in transgenic mice, and that this affect can be overridden by the presence of an early-acting RARE, suggesting that PRE1-dependent silencing can be reset by RA induction. Our analysis of a vertebrate PRE in fly assays, followed by confirmation of function in the mouse would suggest that PRE function and perhaps mechanism of action is relatively conserved between *Drosophila* and chordates.
**PRE1 is the first functionally validated PRE identified outside of *Drosophila***

In *Drosophila*, there have been many attempts to define specific consensus sequences for PREs for the purpose of designing bioinformatics algorithms to predict novel PREs (Brown et al., 1998; Fiedler and Rehmsmeier, 2006; Fritsch et al., 1999; Hur et al., 2002; Kassis et al., 1989; Mihaly et al., 1998; Ringrose et al., 2003; Saurin et al., 2001; Strutt et al., 1997). Although these short motifs that are specific for DNA binding proteins such as PHO, GAF/Psq and Zeste, and their presence are necessary in all PREs identified, they also occur with some frequency in many random DNA sequences. jPREdictor is perhaps the most comprehensive bioinformatics programs designed to identify PREs and was developed by scanning DNA sequences surrounding developmentally regulated genes for the presence of GAF and PHO binding motifs either alone or in combination with one another at variable distances using the established PREs as a training set (Fiedler and Rehmsmeier, 2006; Ringrose et al., 2003). This program predicted 167 sites within the *Drosophila* genome, though many experimentally tested PREs were initially excluded by a more stringent protocol. Additionally, using the motif cluster method, 3 of the 11 training set PREs (iab-8, Mcp and Scr8.2Xba) could not be detected. Likewise, this program failed to detect PRE1 as a candidate PRE.

Further frustrating efforts to reliably predict PREs has been the failure to find striking similarities between experimentally tested PREs when their sequences are aligned (Ringrose and Paro, 2007). Genome wide screens that profile PRC1/PRC2 binding in flies have recently been published (Negre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006), and have given insights into the *in vivo* binding profiles of PcG/trxG proteins in different cells at different stages. Also, by comparing the data sets between the different screens, the “hits” from predictive programs, and experimentally defined PREs, we can gauge how well these predictive programs work, and thus, perhaps improve our understanding of the relationship between PcG/trxG binding and PRE function. For example, between 73% and 94% of bound sites found in the three screens lacked a PRE/TRE prediction from Ringrose et al (Ringrose et al., 2003). Similarly, mammalian genome-wide screens profiling PcG binding in mouse and human ES cells and human embryonic fibroblast (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006) have
informed two other predictive approaches which also use sequence analysis data (Eden et al., 2007). Once again, PRE1 was not predicted by these programs. The most likely explanation for this is that PcG proteins may bind to chromatin independently of PRE/TREs; they may loop from a true PRE/TRE site to a second site (Cléard et al., 2006), or bind via non-specific interactions with weak PRE/TRE-like sites. The latter would not be surprising as Polycomb is a highly mobile protein (Ficz et al., 2005). Also, the mammalian screens interrogated 1-5 kb regions spanning known transcriptional start sites (Eden et al., 2007), whereas the fly data resulted from genuine genome-wide searches for PcG binding and has showed that most PREs like PRE1 are distant from their target promoters.

**PcG binding and histone marker profiles give insight into mechanism of PRE1 function**

In the highly complex and ever changing environment of the developing nervous system, efficient regulation of gene expression is mediated in part at the chromatin level, often by post-translational modification of histone tails (Takizawa and Meshorer, 2008). These include acetylation, methylation and phosphorylation. Of particular interest are the histone H3 marks H3K4 and H3K27 methylation; H3K4 methylation activates transcription by recruiting nucleosome remodeling enzymes and histone acetylases (Pray-Grant et al., 2005; Santos-Rosa et al., 2003; Wysocka et al., 2005), while H3K27 methylation silences transcription by promoting a compact chromatin structure (Francis et al., 2004; Ringrose et al., 2004). This methylation is often catalyzed by trithorax and Polycomb-group proteins which impart mitotically heritable lineage-specific gene expression programs (Ringrose and Paro, 2004).

Interestingly, the H3K27me3 mark of silencing has sometimes been found in conjunction with the activating H3K4me2 mark in pluripotent cell lines at silenced genes that are important for various differentiation programs. Bernstein et al. used mouse embryonic stem cells and a number of differentiated lines to profile the histone H3 methylation marks for genes important in differentiation as well as those important for the maintenance of ES cell pluripotency. As in previous studies, they found that genes marked by H3K4me2 tend to be expressed at significantly higher levels than those associated with H3K27me3. An example of this can be found in the case of the *Ebf1*
gene, which encodes a transcription factor that is expressed at high levels in MEFs, MLFs and myoblasts, and plays an important role in their differentiation (Koli et al., 2004; Schraets et al., 2003). In these cell types Ebf1 is marked by large areas (>5 kb) of H3K4 methylation, while in neuroblastoma cells, where expression is undetectable, these same regions are associated with an expansive H3K27me3 domain (Bernstein et al., 2006). When Bernstein and colleagues looked in pluripotent ES cells however, an intriguing observation emerged around key developmental genes that showed low levels of expression, but were marked with a seemingly contradictory overall distribution of both H3K27 and H3K4 methylation (Bernstein et al., 2006). It has been suggested that these “bivalent” domains mark silenced genes that are poised for induction upon initiation of specific developmental pathways. Consistent with these initial observations in ES cells, most but not all pluripotent cell types have shown large regions of bivalent marks surrounding developmentally important genes. Atkinson et al. confirmed these observations within the HoxA cluster in NTera2SP12 and human ES cell line H9 (Atkinson et al., 2008), while others demonstrated the same phenomenon was found in mouse F9 cells (Gillespie and Gudas, 2007).

Embryonic carcinoma cells such as the F9 line of mouse teratocarcinoma cells used in this study are generally considered to be pluripotent cell types, and may be induced to differentiate into neuroectodermal lineages by the addition of all trans-retinoic acid. This is known to activate Hox gene expression (LaRosa and Gudas, 1988; Papalopulu et al., 1991), and we have shown that RA activates MafB expression. Our analysis of histone signatures of PRE1 and the MafB locus is consistent with many of these observations and may elucidate additional transcriptional mechanisms governing PRE1 regulation of MafB expression. First, bivalent H3 methylation signatures cover most, but not the entire PRE1 region, while the GAPDH promoter shows the typical enrichment of the activating H3K4me2 marker for an actively transcribed gene. Second, detection of retinoic acid dependent RNA PolII binding near the S5 enhancer suggests that an alternative RA inducible promoter may be located here. Also, SUZ12 binding is not detected at the S5 enhancer or at the known MafB promoter, but it is enriched within PRE1. This is consistent with the observation that, as in human embryonic stem cells, there is mutually exclusive binding of SUZ12 and RNA polymerase II to
developmentally specific genes (Lee et al., 2006). Third, polycomb repressive complexes may be involved with regulating accessibility to the RA inducible S5. There is precedent for PcG/trxG complexes controlling access to RA responsive enhancers. Atkinson et al (2008) used ChIP-on-CHIP histone profiling on the inactive HoxA cluster in undifferentiated EC cells. Although RAR/RXR heterodimer binding to DNA has been linked to hyperacetylation of histones around RAREs, examination of the HoxA cluster in undifferentiated EC cells showed that neither H3 nor H4 acetylation is enriched at these positions and upon differentiation, some but not all of these sites are enriched for acetylation (Atkinson et al., 2008). Thus, histone acetylation is not likely to control accessibility to RAREs during differentiation, leading to the more likely possibility that polycomb repressive complexes might be controlling access to the RAREs of the HoxA cluster in pluripotent cells. Support for this comes from large regions of H3K27 trimethylation within these clusters, which are removed in response to differentiation, perhaps leading to more permissive access to RAREs (Atkinson et al., 2008; Kirmizis et al., 2004). Whether this change in histone signature actually leads to upregulation of Hox gene expression is unknown since there is no evidence that a progressive loss of the H3K27 trimethyl signature is directly correlated to selective accessibility to RAREs, or increase in Hox gene expression.

We have not found a functional RARE within S5; however, it is curious that there is a strong peak of H3K27me3 binding that is enhanced upon RA induction while H3K4 methylation is not enhanced upon induction. The enrichment of H3K27me3 binding at the S5 enhancer in response to retinoic acid might be explained by the observation that MafB expression is repressed in the posterior neural tube. A repressor might bind directly to specific sequences in the S5 enhancer, and be induced by retinoic acid in F9 cells as well. Thus, RA-treated F9 cells, like all cultured cell types, may consist of a mixed population: cells that express the gene, in this case MafB, and others in which expression is inhibited. While bivalent marks at PREs are not well understood, their origins at enhancers can be more readily explained. The S5 enhancer induces expression in r5 and r6, while repressing expression posterior to r6. Upon RA induction, both activators and repressors that act on S5 can be induced depending on their relative levels in the cell. It is likely that the RA-treated F9 cells used in this study consisted of a mixture where some
were S5-dependently activated and others were S5-dependently repressed. In those that were repressing *MafB*, H3K27me3 may have been recruited to the S5 enhancer.

Another factor to consider in the analysis of histone signatures is the timing of histone mark expansion. It has been observed that enrichment of the long term active H3K4me3 mark takes a longer time to expand taking up to 10 days of RA induction (Atkinson et al., 2008). This could explain why activation of S5 by RA induction did not correlate with an increase in H3K4me3 binding.

Finally and intriguingly, the counterintuitive decrease in H3K4me3 signature and increase in the H3K27me3 mark at the PRE1d subdomain in response to RA treatment may provide clues as to how PRE1 functions, and may be explained by transcriptional interference mechanisms. First observed in prokaryotes then later in *Drosophila* and mammals, transcriptional interference (TI) occurs when transcription of one gene acts directly to prevent transcription of a neighboring gene *in cis* (Mazo et al., 2007; Shearwin et al., 2005). The most complex example of TI occurs in the fly bithorax complex (BX-C) which contains three *Hox* genes *Ultrabithorax* (*Ubx*), *Abdominal A* (*abd-A*), and *Abdominal B* (*Abd-B*). More than 20 years ago, noncoding RNAs (ncRNA) from the BX-C were detected in fly embryos (Akam et al., 1985; Cumberledge et al., 1990; Hogness et al., 1985; Lipshitz et al., 1987; Sanchez-Herrero and Akam, 1989). These untranslated RNAs were expressed in a spatially and temporally regulated fashion and in the same parasegments as the *Hox* genes they were hypothesized to regulate (Akam et al., 1985; Bae et al., 2002; Cumberledge et al., 1990; Drewell et al., 2002). It was postulated that these intergenic ncRNAs played a role in the opening of distinct chromatin domains. Indeed, it was within these intergenic regions of the BX-C that the first PREs were found to be binding platforms for PcG/trxG proteins, and validated as true cellular memory modules that regulated the repressed and activated states of *Hox* genes (Breiling et al., 1999; Grimaud et al., 2006; Ringrose and Paro, 2007). Furthermore, a strong correlation was found between the transcription of ncRNAs through the regulatory regions of the BX-C and relief of PcG-mediated silencing (Bender and Fitzgerald, 2002; Hogga and Karch, 2002; Rank et al., 2002; Schmitt et al., 2005). More recently, high-resolution FISH and transgenics lacking promoters for the *bxd* untranslated RNAs were used to show that different ncRNAs from these PREs repress the expression of the *Hox* genes.
they regulate, and were also expressed in a cell type and stage specific fashion (Petruk et al., 2007; Petruk et al., 2006).

In mammals, it has been suggested by several groups that the expression of ncRNAs within the *HoxA* cluster is a possible control mechanism of polycomb mediated gene silencing/activation (Sessa et al., 2007). The observation that several of the ncRNAs described (Rinn et al., 2007) coincide with H3K4me2-enriched regions of the *HoxA* cluster suggest that these ncRNAs may be involved in transcriptional interference (Atkinson et al., 2008). It remains to be tested whether these ncRNAs regulate *HoxA* expression via mechanisms of chromatin remodeling that involve transcriptional interference. Nevertheless, it is possible that the highly conserved non-coding region of PRE1 flags a region containing a non-coding RNA important in PRE1 dependent silencing. I have performed a limited RNA *in situ* hybridization using a small region of PRE1 on whole embryos and found no specific expression (not shown). However, we have transgenic data that suggests that PRE1 drives expression in mutually exclusive regions from *MafB* expression (S.P. Cordes, unpublished data). Further analysis using more targeted experiments such as RNA interference of more refined areas of PRE1 will clarify the possible role of TI in PRE1 function.

**Furthering our understanding of PRE1**

Ongoing experiments include an extended analysis of PcG protein binding to PRE1 and to the genomic region surrounding *MafB*. I am using additional antibodies against PRC1 (Bmi1, M33, Ring1b) and PRC2 (Ezh2) components to examine whether these, like SUZ12, bind PRE1 in F9 cells and whether this binding is reduced upon RA treatment. I will add further primer sets that are 2-3 kb apart to more thoroughly interrogate a region spanning from -10 kb upstream of PRE1 through the entire 9.4 kb of PRE1. I will also be investigating the region surrounding the breakpoint proximal to the *Neuronatin* gene. Concomitantly I will use ChIP-on-ChIP experiments to profile SUZ12 and additional PcG protein binding on a microarray spanning the entire *MafB* gene using *HoxD* locus as a positive control on chromosome 2. A possible caveat is that ChIP-on-ChIP analyses may not be as sensitive as our more directed qPCR approach suggesting that ChIP-on-ChIP experiments might miss potential PcG target regions.
Other experiments include a knockdown of SUZ12 using siRNA technology followed by ChIP to confirm that enrichment of SUZ12 binding at PRE1 is indeed specific. Previous studies using this technique has shown that siRNA-mediated removal of SUZ12 from colon cancer cells resulted in a loss of PRC2 member EZH2 binding at target promoters of genes whose expression was increased. The HKMT activity of PRC2 resides with EZH2, and correspondingly, H3K27 trimethylation was found to decrease upon SUZ12 knockdown (Kirmizis et al., 2004). We have shown that PRE1 is enriched for SUZ12 binding within hcPRE1, and that this region is also heavily marked by H3K27me3 (Figure 3.12). I am currently profiling EZH2 binding in this and surrounding regions in F9 cells and will test whether removal of SUZ12 will also disrupt recruitment of PRC2 component EZH2 and/or H3K27 methylation.

A role for PRE1 during normal development and disease

There are a number of global changes in chromatin structure throughout the cell cycle which can be easily visualized by fluorescent staining methods or assayed by restriction endonuclease digestion. For example, during neuronal differentiation of embryonic stem (ES) cells, there is a rearrangement of heterochromatin where the heterochromatin foci in terminally differentiated neurons appear to be more condensed than in their undifferentiated ES cell progenitors (Meshorer, 2007). Another example of global changes in chromatin structure occurs in Alzheimer’s patients where large-scale chromatin condensation has been observed. Brain nuclei of Alzheimer’s patients were less sensitive to digestion with micrococcal nuclease which digests the linker region between two neighbouring nucleosomes. Global chromatin condensation in Alzheimer’s disease patients suggests a reduction in transcription which alters the ability of neurons to generate sufficient levels of gene products to maintain normal cortical function (Lukiw and Crapper McLachlan, 1990). Therefore, it is clear from the above examples that global changes in chromatin structure occur when cells undergo large-scale morphological changes such as during neuronal differentiation and development, or in neurodegenerative diseases. However, local changes in chromatin structure at specific gene loci also play an important role in normal physiological processes and disease.

In the nervous system, the best understood example of local chromatin change playing a role in a psychiatric disorder is in the case of the glutamic acid decarboxylase67
(GAD67) and Reelin genes in schizophrenia. Both these genes are reduced in the prefrontal cortex of schizophrenia patients and in methionine-induced mouse models of schizophrenia (Dong et al., 2005). The promoters of GAD67 and Reelin are hypermethylated and have been found to have an elevated recruitment of the methylated DNA-binding proteins MeCP2 and MBD2. Since DNA methylation is often linked with histone deacetylation and vice versa, Dong and colleagues reasoned that treatment with the HDAC inhibitor valproate would reduce hypermethylation at the Reelin and GAD67 promoters, and in turn reduce MeCP2 binding. This was indeed what they found, not only did valproate treatment reverse promoter hypermethylation in a time-dependent manner, it increased transcription of GAD67 and Reelin (Dong et al., 2007). These experiments illustrate the intricate relationship between histone modifications on DNA methylation at discrete gene loci which in turn influences transcriptional programs.

Another intriguing example of chromatin architecture affecting CpG methylation and how it can persist over generations occurs in studies on the promoter at the gene encoding glucocorticoid receptor (GR). This study examined licking and grooming (LG) and arched-back nursing (ABN) behaviour in rat mothers and the hippocampal GR DNA methylation patterns of their pups (Weaver et al., 2004). Pups of mothers that displayed high levels of LG-ABN predictably performed better in hypothalamic-pituitary-adrenal stress tests as compared with offspring of mothers that showed low LG-ABN mothers. The surprising outcome was that these mice had different methylation patterns that persisted into adulthood. More intriguingly, the stress response and methylation levels were reversible upon cross-fostering between low LG-ABN and high LG-ABN mothers within the first 12 hours of birth. Furthermore, when treated with the HDAC inhibitor trichostatin A (TSA), the DNA methylation state at the GR promoter is reversed, restoring normal GR expression, and stress response (Weaver et al., 2004). This demonstrates that the environment can stimulate specific changes in chromatin state which affects DNA methylation at a given promoter. Not only do these chromatin marks persist over nearly a lifetime, but the system can be manipulated by administering chromatin remodeling drugs.

Our identification of PRE1 and the effect of mechanisms of chromatin remodeling on a key developmental gene MafB have important implications for human disorders as
well. By reducing the levels of the mouse polycomb protein M33, the self-sorting behavior of rhombomeric cells revealed variable *MafB* expression on a cell-by-cell basis (Figure 3.3). This, to our knowledge, is the first report that a ~50% decrease of a *PcG* gene can cause patchy expression of a key neural patterning gene. Several important neural circuits, including the rhythmic circuitry, which governs breathing, are disrupted in *kr/+* and *kr/kr* mice and often affected in autistic patients, and neuronal subtypes such as the majority of mood modulating serotonergic neurons, are born within the hindbrain (Chatonnet et al., 2002), reviewed in (Cordes, 2005; Fortin et al., 2003). Thus, uneven gene expression during early hindbrain development might sometimes have dramatic neurobiologic consequences.

While possible links between variations in *PcG/trxG* proteins and neurodevelopmental disorders have not been explored, our findings suggest that variations in *PcG* gene levels or disruptions of specific PREs might conceivably contribute to such disorders. There is however precedent for the ability of *PcG/trxG* proteins in counteracting the teratogenicity of chromatin remodeling drugs often used in neurobiologic disorders such as valproate. Polycomb-repressive complex 2 (PRC2)-mediated histone methylation plays an important role in aberrant cancer gene silencing and is currently a focus as a potential target for cancer therapy (Galmozzi et al., 2006). For example, a recent study found that the S-adenosylhomocysteine hydrolase inhibitor 3-Deazaaneplanocin A (DZNep) which induces efficient apoptotic cell death in cancer cells but not in normal cells, effectively depletes cellular levels of the PRC2 components EZH2, SUZ12, and EED. Loss of these PRC2 components was associated with and/or leads to a decrease in the gene silencing marker H3K27 methylation. Furthermore, using RNA interference (RNAi), genome-wide expression analysis, and chromatin immunoprecipitation (ChIP) studies, Tan et al. identified a prominent set of genes that when selectively repressed by PRC2 in breast cancer, can be reactivated by DZNep, lending more support for the use of such targeted drugs in cancer therapy (Tan et al., 2007).

Valproic acid is often used as an atypical complementary drug in the treatment of epilepsy, schizophrenia and bipolar disorder, but is also a potent teratogen whose mechanism of teratogenesis is poorly understood. It has been postulated that the HDAC
inhibitor valproate affects gene silencing pathways that are normally mediated by the PcG complex, contributing to its teratogenicity during anterior-posterior patterning in embryonic development. VPA may affect the gene silencing pathway mediated by the Polycomb group complex. The epigenetic mechanism of VPA teratogenicity on anterior-posterior patterning was recently explored by Okada and colleagues where pregnant mice were given teratogenic levels of valproate at E8.0. The expression patterns of a number of PcG/trxG genes were then measured by quantitative real-time RT-PCR. Two less teratogenic antiepileptic chemicals that are not thought to affect PcG/trxG action, valpromide and valnoctamide, were used in addition to vehicle control. It was found that valproate inhibits the expression of Eed, Ezh2, Zfp144, Bmi1, M33, Rnf2, and YY1 in the mouse embryos, while valpromide and valnoctamide do not. In contrast, the trxG gene MLL was not affected at comparable doses (Okada et al., 2004). Thus, while direct connections between variations in PcG/trxG proteins and neurodevelopmental disorders have not yet been explored, the above studies and our own observations would suggest a strong relationship between histone modifiers and variations in PcG gene levels. In addition, these alterations in PcG/trxG levels or disruptions of specific PREs might contribute to a number of such disorders or the variability in their presentation among individuals.

**PREs and the evolution of the segmented hindbrain**

Two key innovations underlie vertebrate head evolution: unsegmented cranial paraxial mesoderm and hindbrain segmentation. Is PRE1, like hindbrain segmentation, a vertebrate innovation or is it an adaptation or modification of an archetypal PRE? This question cannot be adequately addressed in *Drosophila*, where MafB homologues are present, but not expressed in a segmental fashion in the developing embryo (S. Cordes, unpublished observations). Furthermore, *Hoxa3* and *b3* are MafB targets during hindbrain segmentation, but the *Drosophila Hox3* gene, *zerknult* (*zen*), does not play a role in segmentation. In contrast to *Drosophila*, anterior-posterior region-specific expression of *Hox3* is seen in neurula stage embryos of the cephalochordate *Amphioxus*, which is considered the closest living invertebrate relative of vertebrates. While *Amphioxus* does not have a visibly segmented hindbrain, the nerve cord posterior to the cerebral vesicle
approximately down to the level of the eighth somite appears to correspond to the vertebrate hindbrain (Holland and Holland, 1999; Williams and Holland, 1998).

Unlike *Drosophila melanogaster*, in which the *Hox* gene cluster has been split into two clusters or vertebrates where it has been amplified, *Amphioxus* has a single *Hox* gene cluster, which is generally thought to most closely resemble the ancestral *Hox* gene cluster. Furthermore, the temporal and spatial colinear expression of *Hox* genes is preserved: with the most 3’ gene, *Hox1*, being expressed earlier and most anteriorly during *Amphioxus* patterning. Also, *Amphioxus* is the only invertebrate, in which retinoic acid signaling has been shown to be interlinked with *Hox* gene expression and anterior-posterior patterning (Castro and Holland, 2003; Garcia-Fernandez and Holland, 1994; Holland et al., 2008; Marletaz et al., 2006; Wada et al., 1999). Remarkably, even a DR5-type RA response element, similar to the one used in these studies, is present in the *Amphioxus Hox1* gene (Schubert et al., 2004; Schubert et al., 2006). Importantly, in *Amphioxus* anterior expression of *Hox3* extends up to the level of somite 4, a region roughly equivalent to the r4/r5 boundary that is respected by *Hox3* class genes in vertebrates (Schubert et al., 2006). A *Amphioxus MafB* homologue exists but, whether its expression overlaps with that of *Hox3* in the nerve cord, is unknown (Holland et al., 2008). However, the *Amphioxus* homologue of *Krox20*, *AmphiKrox20*, an early hindbrain patterning gene expressed coincidentally with *MafB* in r5, shows no equivalent anterior-posterior expression in the *Amphioxus* nerve cord (Knight et al., 2000). Thus, *Amphioxus* may be hovering at a point in evolution just before hindbrain segmentation emerged.

The *Amphioxus* genome sequence has recently been completed, and BLAST searches for regions with similarity to PRE1 upstream or near the *AmphiMafB* gene did not identify any matching sequence. However, the sequence requirements for PRE1 function remain to be determined and simple BLAST searches may not suffice to identify the equivalent functional region in more distant organisms. Also, while a PRE1-homologous region was identified upstream of all mammalian *MafB* orthologues, it was not recognized by homology searches in some vertebrate genomes, such as that of zebrafish, in which the neuronal hindbrain architecture is significantly less complex. Thus, it is unclear whether PRE1 and the role of *MafB* in anterior-posterior patterning...
have emerged together in vertebrates or whether an archetypal functional PRE1 region may have been present before the emergence of hindbrain segmentation. Ultimately, detailed analyses of the functional sequence requirements for PRE1 combined with comparative genomics and analyses of the role of MafB in anterior-posterior patterning of invertebrates such as Amphioxus as well as spiders and daphnia, which have also preserved the role of Hox3 in anterior-posterior patterning, may help to resolve the origins of PRE1.
References


methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell 121, 859-872.


