

Open Research Online

The Open University's repository of research publications and other research outputs

Induction of 'Hox' genes and genome wide identification of Hox binding sites in mice

Thesis

How to cite:

De Kumar, Bony (2013). Induction of 'Hox' genes and genome wide identification of Hox binding sites in mice. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2013 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk



Induction of '*Hox*' genes and genome wide identification of Hox binding sites in mice

Bony De Kumar, B.Sc. (Ag) M.Sc.

A thesis submitted in fulfillment of the requirements of the Open University
for the Degree of Doctor of Philosophy

The Stowers Institute for Medical Research
An Affiliated Research Centre of the
Open University
Kansas City, USA

September 2013

DATE OF SUBMISSION: 27 SEPTEMBER 2013

DATE OF AWARD: 27 NOVEMBER 2013

The Open University
14 JAN 2014
The Library

DONATION

X 574.87322 2013

Consultation copy

Dedication

To

My Papa and Mummy

Acknowledgements

Though this thesis work is largely an individual work, I could never reach this level of excellence without help, guidance, discussion and mentorship of many of my colleagues, friends, well-wishers and teachers. Their valuable discussion helped me to get better insight into my data and helped me to see things in a different perspective. Colleagues from core facilities were instrumental in helping me to learn and apply evolving technologies. This helped me to implement the latest available technology in a proficient manner.

First, I would like to thank my mentor and director of study, Dr. Robb Krumlauf for providing me with lab space, funding and an interesting thesis problem. His faith in my abilities was constant force driving me towards my research goal in spite of many personal obstacles that I faced in my life during course of my research. His infectious enthusiasm and hunger for knowledge helped motivate me and learn to set high expectations. His amazing ability to tell complex stories in a simple way and to help me do so during my thesis training have been a valuable life learning experience for me.

I would like to thank my supervisor Dr. Ron Conaway for his help and guidance especially in biochemical experiments. I am thankful to Ron and Joan Conaway for providing guidance and helping me in troubleshooting problems pertaining to biochemical experiments. I am thankful to them for allowing me to work in their lab to learn basic protein related techniques.

It is with immense pleasure that I thank my committee members Dr. Zulia Zeitlinger and Dr. Marco Blanchette for their constructive suggestions during the entire period of my doctoral research. Discussions in committee meeting were extremely helpful in keeping my focus on the given problem. Their valuable suggestions were helpful in improving data representation and presentation.

I would like to thank Dr. Leanne Wiedemann and Shelly Hornbuckle for their all help during my doctoral research. Their systematic and organized efforts were vital in submitting reports on time. I would like to thank Leanne for her constructive opinions during my lab meetings which greatly helped me to improve my presentations, especially using PowerPoint slides. I appreciate Leanne's help during writing of my thesis.

It gives me immense pleasure to thank Ariel Paulson, our analyst, for his valuable help. During my doctoral research, I learned a lot from Ariel about implementation of statistics and approaches used in genomics data analysis. He really worked hard to bring raw data sets to an insightful state using computational analysis such that we are able to understand our problem in a more comprehensive manner.

I would like to thank Dr. Mark Parrish, who as a good friend and colleague, was always willing to help me and give best suggestions. It would have been a lonely in lab without discussions with him. I would like to acknowledge contribution of our lab manager Carrie Scott for helping and teaching me chicken electroporation experiments. I would like to thank her for managing lab inventories and ordering things as quickly as possible. It gives me immense pleasure to thank Aaron Gottschalk, who as good friend and teacher was added immense value to my life. He taught me basic principles of biochemical experiments and techniques related to proteins. Our visits to claret lounge and intense discussion (scientific and non-scientific) will be always my cherished memories.

I would like to thank members of tissue culture core and molecular biology for their valuable help. Tissue culture core was helpful in maintain our cell culture line without differentiation. Sequencing of large number of samples and q-PCRs were not possible without help of molecular biology.

I would like to thank Christof Nolte for stimulating discussion and helping me with his immense knowledge of the literature which was an important component in my progress during my doctoral research. I would like to thank Tara Alexander and Kieran Pemberton for their help during my early days in lab. I would like to thank my other lab members Youngwook Ahn, Jennifer McEllin, Hugo Parker and Marina Yurieva for making this lab a wonderful place to work.

I would like to thank my friends' Santharam Katta and Amol Ranjan for supporting me at all times, thick and thin. I am thankful to my friends Arnob, Malini, Rushi, Shachi, Vijay, Alex, Justin and Ram for their help. I am thankful to my friends and their families of Deepu Jacob and Jyaesh Vijay for their support and care. It gives me immense pleasure to thank Jayan Aachayan and Dolly Aunty for being with our family all the time with valuable help and guidance.

I thank Jim and Virginia Stowers and Stowers Institute for such wonderful working place.

Finally, I thank my family- My parents, in laws, sister, brother, wife and son for their support during my doctoral research. Encouragement from my father and mother was always a reason to keep moving ahead during my doctoral research. Their blessings were always my strength. I thank my wife for standing with me in worst phase of my life. She was encouraging and always provided strength to continue. I thank my son, Advait, whose smile was reason for me to live and progress.

Abstract

Hox genes encode a family of transcription factors that play highly conserved regulatory roles in specifying the properties of tissues in developing embryos. Very little is known about how HOX proteins control the cellular and developmental processes governing morphogenesis through regulation of down-stream target genes. The goal of this research was to investigate on a genome-wide basis, the rules and principles which underlie the binding of different HOX proteins to target sites and understand the basis for their distinct specificities. I utilized the programmed differentiation of mouse embryonic stem cells into a neural fate with retinoids and genomic technologies to systematically investigate binding properties of two HOX proteins, HOXA1 and HOXB1 and their co-factors PBX and MEIS. I analyzed the induction properties of the cells and the transcriptional dynamics and epigenetic states in Hox clusters to explore the differentiation process. An extensive and dynamic pattern of transcriptional activity indicates that Hox clusters generate a large number of non-coding RNAs which may impact their activation and chromatin states. Global identification of HOXB1, HOXA1, PBX and MEIS binding regions by chromatin immune precipitation and high throughput sequencing (ChIP-seq) has generated insight into many potential Hox target genes. HOXA1 binding peaks generally overlapped with those of PBX and MEIS, supporting their roles as HOX co-factors. The sites bound by HOXB1 uncovered new classes of binding motifs. Regulatory assays demonstrated that many of these novel motifs functioned as neuronal enhancers. Many HOXB1 binding peaks have closely associated REST motifs and bind the REST repressor complex, which is important in neuronal differentiation. The close association of REST and HOXB1 binding sites provides a mechanism for coordinating cell differentiation programs in neurogenesis. This research has uncovered novel properties of HOX proteins and their co-factors that underlie their role as master regulators of patterning and morphogenesis.

Table of Contents

Acknowledgements.....	v
Abstract	vii
Table of Contents.....	ix
Table of Figures	xiii
Table of Tables.....	xvii
Table of Abbreviations.....	xix
Chapter 1 Introduction	1
1.1 History.....	7
1.2 Evolution of Hox genes	9
1.3 Function of Hox genes in mammals-lessons from gain and loss-of-function mutants in mice.....	19
1.3.1 Group 1 paralogs.....	22
1.3.2 Group 2 paralogs.....	26
1.3.3 Group 3 paralogs.....	27
1.3.4 Group 4 paralogs.....	27
1.3.5 Axial skeleton and other Hox paralog groups	29
1.3.6 Limb development and Hox paralogs	30
1.4 DNA binding properties of HOX proteins and roles of co-factors in specificity	31
1.4.1 Hox co-factors	35
1.4.2 Role of Co-factors in determining specificity of HOX binding.....	38
1.4.3 HOX proteins-Determinants of its own specificity	42
1.5 Down-stream targets of Hox proteins.....	46
1.6 Thesis problem statement.....	52
1.6.1 Aim 1: Study activation of Hox clusters in neuro-differentiation using an ES cell model system	53
1.6.2 Aim 2: Genome-wide identification of HOX-response elements.....	54
Chapter 2 Methods.....	55
2.1 Induction of KH2 cells with retinoic acid.....	55
2.2 RNA preparation for Affymetrix microarray analysis.....	55
2.3 Analysis of Affymetrix data	56
2.4 Agilent tiling Array	57
2.4.1 Design of custom made tiling array.....	57
2.4.2 Library preparation and hybridization.....	57

2.4.3	<i>Analysis of Expression</i>	58
2.5	Directional mRNAseq library preparation and sequencing	58
2.6	Analysis of RNA sequencing data	58
2.7	Chromatin Immuno-precipitation	59
2.7.1	<i>Library preparation and hybridization for ChIP on Chip</i>	60
2.7.2	<i>Analysis of ChIP on ChIP</i>	60
2.7.3	<i>ChIP-Sequencing</i>	60
2.7.3.1	Library preparation for ChIP-Sequencing	60
2.7.3.2	Analysis of ChIP Sequencing	61
2.7.3.3	Analysis for over represented motifs	61
2.7.3.4	Feature mapping	61
2.7.3.5	Nearest-Neighbors	62
2.7.3.6	Functional analysis	62
2.7.3.7	Heat maps	62
2.8	Quantitative PCR and analysis.....	62
2.8.1	<i>Quantitation of Hox genes using TLDA cards</i>	62
2.8.2	<i>Quantitation of Non-coding transcripts</i>	63
2.9	RA Gavage of 10 dpc female CD-1 mice	63
2.10	Template binding assay	64
2.11	MudPIT.....	64
Chapter 3 Study of activation of Hox clusters in neuro-differentiation using an ES cell model system		
..... 65		
3.1	Result	68
3.1.1	<i>Analysis of genome-wide gene expression profile during RA induced murine ES cell differentiation</i>	68
3.1.2	<i>Induction of Hox genes during neuro-ectodermal differentiation</i>	72
3.1.3	<i>Analysis of transcriptional activity in Hox cluster</i>	73
3.1.4	<i>Temporal changes in Epigenetic properties and retinoid receptor occupancy along Hox cluster upon RA induced neuro-ectodermal differentiation</i>	79
3.1.5	<i>RA Response of Non-coding transcripts</i>	90
3.2	Discussion.....	96
Chapter 4 Characterization of genome-wide binding of Hoxb1 in differentiating ES cells.....		
..... 105		
4.1	Result	107
4.1.1	<i>Generation of KH2 ES Cells with Epitope tagged HOXB1</i>	107

4.1.2	<i>Genome-wide identification of HOXB1 binding sites</i>	107
4.1.3	<i>Dynamics of HOXB1 occupancy in differentiating ES Cells and 9.5dpc embryo</i>	113
4.1.4	<i>Identification of over-represented motifs in genome-wide binding site</i>	118
4.1.1	<i>HOXB1 as activator for REST regulated genes</i>	123
4.1.2	<i>HRE2 (Hox Response Element 2) is a repressive Hoxb1 element</i>	131
4.1.3	<i>Modular Motif–Long cis finder motif</i>	138
4.2	Discussion	141
Chapter 5 Interactors, co-factors and combinatorial binding		143
5.1	Result.....	145
5.1.1	<i>Generation of Epitope tagged Hoxa1 cells line</i>	145
5.1.2	<i>Dynamics of HOXA1 occupancy in 24hrs RA and Doxycycline induced ES Cells</i> ..	146
5.1.3	<i>Dynamic changes in Histone modification on future HOXA1 sites</i>	152
5.1.4	<i>Identification of over-represented motifs in genome-wide binding site</i>	155
5.1.5	<i>Combinatorial binding of Paralogus group 1, Co-factors and REST</i>	161
Chapter 6 Summary discussion		163
Chapter 7 References		175
Appendix I. Genes Up-regulated in Hoxa1 mutant		212
Appendix II. Genes Down-regulated in Hoxa1 mutant		213
Appendix III. Genes Down- regulated in Hoxb1 mutant		214
Appendix IV. Genes Up- regulated in Hoxb1 mutant		215
Appendix V. Comparative list of genes changed in Hoxa1 and Hoxb1 mutant		216
Appendix VI. Direct Hoxb1 identified by Gavalas lab		218
Appendix VII. Protein bound on Hoxb1-ARE		219
Appendix VIII. Publications		221

Table of Figures

Figure 1-1. Origin and divergence of Hox cluster	10
Figure 1-2 Cladogram depicting Hox gene chromosomal organization for representative animals.	12
Figure 1-3 <i>Cis</i> -regulatory modules within the Abd-B locus of the bithorax complex involved in co-linearity of Hox genes in <i>Drosophila</i>	15
Figure 1-4 Nested Expression of Hox genes in hindbrain.	21
Figure 1-5 Transcriptional regulation of <i>Hoxb1</i> through auto and cross-regulatory mechanism	25
Figure 1-6 Consensus bipartite HOX/PBC sites in the enhancer and their properties	39
Figure 1-7 Organization of PREP MEIS –(PM) and PBX-HOX (PH).....	40
Figure 1-8 Model summarizing the different regulatory interactions between HoxB genes	47
Figure 3-1 Flow chart explaining design of experiment used in current study.....	68
Figure 3-2 Heat map showing global change in gene as analyzed on affymetrix Mouse Genome 2.0 arrays	70
Figure 3-3 Gene ontology (GO) term analysis of differentially expressed genes	71
Figure 3-4 Changes in Hox genes upon RA induction as analyzed on affymetrix Mouse Genome 2.0	74
Figure 3-5 Quantitative changes in Hox gene expression upon RA induced differentiation	75
Figure 3-6 Transcriptional activity in Hox clusters upon various length of RA treatment ..	76
Figure 3-7 Quantitative analysis of known genes and non-coding transcripts using RNA-Seq.....	78
Figure 3-8 Changes in epigenetics state and RA receptor occupancy of Hox A and B cluster during RA induced neuro-ectodermal differentiation.....	80
Figure 3-9 Changes in epigenetics state and RA receptor occupancy of Hox C and D cluster	81
Figure 3-10 Dynamic changes in H3K27Me3 marks on Hox gene promoters and over Hox cluster.....	83
Figure 3-11 Occupancy of RAR receptors and changes in transcriptional and epigenetic properties in region encompassing HoxA1 and HoxA2	86
Figure 3-12 Novel non-coding transcripts “heater” identified upstream of HoxA cluster ...	87
Figure 3-13 Occupancy of RAR receptors and changes in transcriptional and epigenetic properties in region encompassing HoxB4 and HoxB5.....	89

Figure 3-14 RARE upstream of Heater transcripts region	91
Figure 3-15 Heater – An extensively transcribed region with non-coding transcripts.....	93
Figure 3-16 mHOT arm- non-coding transcripts between HoxA1 and HoxA2	94
Figure 3-17 B2iT transcript- rapidly induced non-coding transcripts between HoxB4 and HoxB5.....	95
Figure 3-18 Clustering of differentiation time course based on gene expression profile.....	98
Figure 3-19 Early and late up regulated genes during RA induced differentiation of ES Cells.....	99
Figure 3-20 Early and late down regulated genes during RA induced differentiation of ES Cells.....	100
Figure 3-21 Dynamic chromatin state based on H3K4Me3 and H3K27Me3 in differentiating ES Cells.....	102
Figure 3-22 Relative predictive values	104
Figure 4-1 Induction of <i>Hoxa1</i> and <i>Hoxb1</i> in differentiating ES cells upon RA induction	106
Figure 4-2 Constructs for expression of epitope tagged HOXB1 in KH2 cells	108
Figure 4-3 Induction of HOXB1 expression doxycycline and RA in three different ES Cell clones with epitope tagged <i>Hoxb1</i> in Col II locus.....	109
Figure 4-4 Distribution of HOXB1 occupied genomes near pre-defined genomic features	111
Figure 4-5 Comparison of direct targets of HOXB1 in <i>Hoxa1</i> and <i>Hoxb1</i> mutant.....	112
Figure 4-6 Go Term analysis of nearby genes with a HOXB1 bound region.	114
Figure 4-7 Gene-interaction network of genes with a nearest neighbor HOXB1 binding peaks	115
Figure 4-8 Genome-wide occupancy of HOXB1 in differentiating ES Cells.	116
Figure 4-9 Enhancer behavior of HOXB1 occupied region in chicken and mice	118
Figure 4-10 Examples of HOXB1, PBX and MEIS Co-occupancy	120
Figure 4-11 Distribution of K-mers in HOXB1 bound genomic regions.....	124
Figure 4-12 Distribution of REST motifs in HOXB1_REST co-occupied peaks.....	125
Figure 4-13 Co-occupancy of REST and HOXB1 in differentiating ES cells	126
Figure 4-14 Binding properties of Full and Half sites.....	128
Figure 4-15 Novel REST motifs with different spacers between two half sites	129
Figure 4-16 Minimum distance of REST, HOX and PBX motifs in HOXB1-REST Co- occupied peaks.....	130
Figure 4-17 Enrichment of H3K27Ac, H3K4Me1 and REST binding on genomic region co- occupied by HOXB1 and REST.....	131

Figure 4-18 Logo of HRE2.....	132
Figure 4-19 Enhancer activity of HRE2 in chicken embryos.....	133
Figure 4-20 HOXB1 binds specifically to HRE2 and Known ARE region.	135
Figure 4-21 Occupancy of HOXB1, PBX, MEIS and modified histones on HRE2 containing HOXB1 binding sites.....	136
Figure 4-22 Distribution of HRE2 in HOXB1 bound genomic regions.....	138
Figure 4-23 Logo of Long Motif.....	139
Figure 4-24 Functional assay of long motif for enhancer activity.....	140
Figure 5-1 Construct used for expression of epitope tagged HOXA1.....	145
Figure 5-2 Distribution of HOXA1 occupied peak with respect to nearby pre-defined genomic features.....	147
Figure 5-3 Enriched GO terms from nearest neighbor genes from HOXA1 binding genomic regions.....	149
Figure 5-4 Gene interaction network for genes with nearby HOXA1 bound region.....	151
Figure 5-5 Genome-wide occupancy of HOXA1 after 24hrs of RA Treatment of ES cells	153
Figure 5-6 Relative spatial distribution of ATTA (HOX) and HOX-PBX bipartite motifs in HOXA1 bound region.....	156
Figure 5-7 Co-occupancy of HOXA1 with cofactors and other Hox genes.....	158
Figure 5-8 Genome-wide occupancy of HOXA1, HOXB1, PBX and MEIS in differentiating ES Cells.....	160
Figure 5-9 Genome-wide occupancy of HOXA1, HOXB1, PBX and MEIS in differentiating ES Cells.....	162
Figure 6-1 Combinatorial binding of various anterior Hox genes, Cofactors, REST and associated epigenetic modification.	172

Table of Tables

Table 1-1 Evolutionary status of Hox cluster in various species	13
Table 1-2 Pbx mutants in Fish and mouse recapitulate Hox loss-of-function mutant.....	37
Table 4-1 Identified over-represented motifs using Meme.....	121
Table 4-2 Occurrence and Enrichment of Pre-defined K-mers in HOXB1 occupied region	122
Table 4-3 Proteins recruited by HRE2.....	134
Table 4-4 Proteins recruited by long motif(Jabet et al., 1999; Knoepfler et al., 1999; Phelan and Featherstone, 1997)	141
Table 5-1 Over-represented motifs in HOXA1 bound region	154
Table 5-2 Enrichment of pre-defined K-Mers in HOXA1 bound region	155
Table 5-3 Co-enrichment of K-mers.....	157

Table of Abbreviations

ANT-C	Antennapedia complex
AP axis	Anterior posterior axis
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
BX-C	Bithorax complex
CMV	Cytomegalovirus
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxy ribonucleic acid
DR	Direct repeat
DTT	Dithiothreitol
EDTA	Ethylene Diamine Tetra Acetic acid
FACS	Fluorescence-activated cell sorting
FGF	Fibroblast Growth factor
FPKM	Fragments Per Kilo base of transcript per Million mapped reads
GCR	Global control region
GO term	Gene ontology term
GTF	Gene Transfer Format
IGV	Interactive genome viewer
lincRNA	Long non coding ribonucleic acid
miR	micro ribonucleic acid
mRNA	Messenger ribonucleic acid
MudPIT	Multi-dimensional protein identification tool
NAA	Non-Essential Amino Acid
ncRNA	Non-coding ribonucleic acid
NP-40	Nonyl phenoxypolyethoxylethanol-40
PBS	Phosphate Buffer Saline
PcG	Polycomb group
PCR	Polymerase Chain reaction
PG	Paralogous group
RA	Retinoic acid
RAR	Retinoic acid receptor
RARE	Retinoic acid response element

RIN	RNA integrity number
RNA	Ribonucleic acid
RXR	Retinoic acid X receptor
TALE proteins	Three amino acid loop extension proteins
TLDA card	Taqman low density Array card
P300	E1A-Binding Protein, 300kD
REST	RE1-Silencing Transcription Factor
NRSF	Neuron Restrictive Silencer Factor
TSS	Transcription Start Site

Chapter 1 Introduction

Homeotic genes (HOM/HOX genes) encode homeodomain-containing transcription factors that confer segmental identity along the primary body axis (Lewis, 1978; McGinnis and Krumlauf, 1992). These conserved genes are implicated in mechanisms controlling the regionalization of the body plan of all bilaterally symmetrical animals (de Rosa et al., 1999; Duboule, 2007). The functions of Hox homeotic genes are conserved in invertebrates and vertebrates, as extensive studies on Hox gene function have revealed that they play a common role in specification of regional diversity along the anteroposterior (AP) axis (Carroll, 1995; Krumlauf, 1994). Hox genes are required for elaboration of the main body plan and also appendage development (Zakany and Duboule, 2007; Zakany et al., 2004). In vertebrates, for example, functional studies have demonstrated that the Hox genes play key roles in patterning regional properties along the axial skeleton, central nervous system, limbs, gut and genitalia (Alexander et al., 2009; Burke and Nowicki, 2003; Dollé et al., 1991; Kondo et al., 1996; Mallo et al., 2010; Podlasek et al., 1997; Wellik and Capecchi, 2003; Zakany and Duboule, 2007).

All HOX proteins share a characteristic 60 amino acid motif, called the homeodomain (Gehring et al., 1994a; McGinnis et al., 1984; Scott and Weiner, 1984). These function as sequence-specific DNA binding domains associated with functional roles for the HOX proteins in transcriptional regulation through direct activation and/or repression of downstream target genes. The homeodomain folds into 3 helices. Helix 2 and 3 generate a helix-turn-helix conformation, characteristic of transcription factors that bind to the major groove of DNA (Gehring et al., 1994b). The N-terminal region preceding helix 1 contacts nucleotides of the minor groove of the target DNA (Passner et al., 1999). The third helix (the recognition helix) binds to a consensus TAAT motif, which is conserved in nearly all sites recognized by homeodomain proteins (Berger et al., 2008; Otting et al., 1990).

Recent advances in large-scale sequencing and comparative genomics have provided valuable insight on the gene complements and organization of Hox clusters in a wide range of animal model systems spanning different phyla. This knowledge has advanced our understanding on the evolution of Hox genes and lays the ground work for a better understanding of the molecular basis of many aspects of patterning processes during development in the animal kingdom (Carroll, 1995; Hoegg and Meyer, 2005; Meyer, 1998). The data from genomic analyses, along with genetic studies in different model

organisms has also provided insight into human congenital defects, such as synpolydactyly and hand-foot-genital syndrome (Goodman et al., 2000; Muragaki et al., 1996).

In general, mammalian genomes contain 39 Hox genes organized into four complexes (HoxA, HoxB, HoxC and HoxD) per haploid set, located on four different chromosomes (Boncinelli et al., 1988; Hoegg and Meyer, 2005; Krumlauf, 1994; Scott, 1992). A sequence alignment between *Drosophila* HOM-C and vertebrate Hox complexes suggests that the four mammalian Hox complexes arose from a single ancestral cluster by gene and chromosome duplications during evolution (Duboule and Dollé, 1989; Graham et al., 1989). A recent study has shown that Hox clusters can be fragmented, reduced or expanded in many animals and suggesting this might be associated with a role in bringing about morphological changes during evolution (Duboule, 2007; Holland, 2013; Lemons and McGinnis, 2006). By virtue of their roles in specifying segmental identities along the A/P axis, Hox genes play a major role in morphological diversification during evolution (Alexander et al., 2009; Carroll, 1995; Carroll, 2005; Mallo et al., 2010; Wellik, 2009). There is evidence that variation in the body plan among the arthropods is due to variation in the expression and regulation of Hox genes (Hughes and Kaufman, 2002; Ronshaugen et al., 2002). The sequence variations in homeodomain proteins also contribute to changes in target specificity and are subject to evolutionary changes (Egger et al., 1994; Li and McGinnis, 1999; Ronshaugen et al., 2002).

A unique feature of clustered Hox genes in many animals is the relationship between their chromosomal organization, expression and their function in time and space during development. This is termed co-linearity. The clustered organization of the Hox genes exhibits a direct relationship to their temporal and spatial modes of expression (Duboule and Dollé, 1989; Graham et al., 1989; Lewis, 1978; Simeone et al., 1990). For example, the AP boundaries of Hox gene expression along the primary AP axis of developing embryos generally correlates with the relative location of individual genes within a cluster. Hence, the linear arrangement of the genes on the chromosome is somehow translated into a similar order of expression along embryonic axes during development. Since all Hox genes in vertebrate Hox clusters have the same 5' to 3' orientation with respect to transcription, the property of co-linearity results in genes located in more "3' regions" of a cluster having more anterior expression domains than those in "5' regions". This ordered relationship of arrangement of genes and its relative expression has been termed spatial co-linearity (Kmita and Duboule, 2003). In vertebrates there is evidence for temporal co-linearity, wherein the most 3' genes exhibit the earliest onsets of expression, followed by the sequential activation of more 5' genes (Duboule,

1998). The molecular basis of co-linearity is poorly understood. There is evidence for long-range regulatory interactions (global control regions), modulation of chromatin structure and accessibility, sharing of regulatory regions and complex deployment of local acting *cis*-elements (Kmita and Duboule, 2003). It appears that different vertebrate Hox clusters may use distinct combinations of these types of mechanisms to establish their nested and ordered co-linear domains of gene expression and mechanisms for a given cluster may vary in different tissues.

One potential insight into mechanisms of co-linearity arises from *in vivo* analyses of the response of Hox genes to growth factors (FGF) and inducing signals (RA) (Bel-Vialar et al., 2002; Conlon and Rossant, 1992; Isaacs et al., 1998; Marshall et al., 1992; Pownall et al., 1998). Several studies have demonstrated that teratocarcinoma and embryonic stem cells can be induced to differentiate upon treatment with retinoic acid (RA). During this differentiation process there appears to be a co-linear activation of Hox genes, such the 3' Hox genes are sequentially activated before 5' members of the clusters (Papalopulu et al., 1991b; Simeone et al., 1990; Simeone et al., 1991). This response is believed to reflect the underlying signaling mechanisms related to how axial domains of Hox expression are established through the action of dynamic signaling centers during elongation of the body axis in vertebrate embryogenesis (Deschamps and van Nes, 2005; Diez del Corral and Storey, 2004; Young et al., 2009). Hence, understanding the RA mediated Hox response in ES cells is highly relevant for understanding how the ordered domains of Hox expression in neural development are generated in part by retinoid signals (Gavalas, 2002; Gavalas and Krumlauf, 2000; Gould et al., 1998; Itasaki et al., 1996; Marshall et al., 1994; Serpente et al., 2005b; Studer et al., 1998a). Furthermore, knowledge on the mechanisms which regulate the ordered Hox response to RA in ES cells offers a possibility for understanding aspects of co-linearity in the early embryo. The presence of RA response elements and the ability of Hox genes to respond to RA in non-vertebrate systems suggest that the ability to respond to RA may be an ancient regulatory feature of Hox clusters (Holland and Holland, 1996; Manzanares et al., 2000; Pani et al., 2012).

In vertebrates, the activation and maintenance of axial Hox expression in the three germ layers in development is achieved through a balance of opposing signaling cascades (e.g. RA and Fgf/Wnts) (Deschamps and van Nes, 2005; Diez del Corral and Storey, 2004; Young et al., 2009). Retinoic signaling recruits co-activators (p300/CBP), co-repressors (SMART/N-CoR) and other protein complexes which have inputs into regulation of epigenetic states and modifiers that control chromatin confirmation and accessibility. Direct input of retinoids on transcriptional activity has been described through binding of

heterodimeric complexes of retinoid receptors to specific target sites, called RAREs (Retinoic acid response elements), in genomic sequence (Chambon, 1994). These RAREs tend to have a short direct repeat sequence motif with a spacer of two (DR2) or five (DR5) nucleotides. One of the three Retinoic Acid receptors (RARs), alpha, beta or gamma, can partner with Retinoid X receptors (RXRs) to form heterodimers that bind with high affinity to DR2 and DR5 RAREs. In absence of RA, these hetero-dimeric receptors can bind DNA and recruit co-repressors, such as N-CoR and SMRT, to maintain a repressed or closed chromatin conformation through deacetylation of lysine residues in histone tails (Glass and Rosenfeld, 2000). In the presence of RA, ligand binding induces conformational changes that disrupt N-CoR/SMART complexes and activation of target genes can occur through recruitment of co-activators: SRC family proteins, p300/CBP, P/CAF complex and CARM1. This recruitment can also facilitate chromatin conformational changes through extensive chromatin remodeling.

The proper domains of nested or Co-linear Hox expression in the developing nervous system depend upon retinoid signaling. This has been demonstrated in a number of vertebrate model systems (mouse, avian and zebra fish) through perturbations in the synthesis or degradation of retinoids. Avian embryos (quail) raised on a retinoid deficient diet display shifts in the boundaries of Hox expression in the hindbrain (Gale et al., 1999; Maden, 2002). Mouse and zebra fish mutants in the enzyme RALHD2, which converts retinol to retinoic acid, display a wide variety of defects associated with changes in patterns of Hox expression in the CNS and other tissues (Begemann et al., 2001; Grandel et al., 2002; Linville et al., 2004; Molotkova et al., 2005; Niederreither et al., 1999; Niederreither et al., 2000; Oosterveen et al., 2004). Hox genes have also been shown to regulate *Raldh2* and *RAR-Beta* setting up feedback loops that reinforce positive cross-talk between Hox expression and retinoid signaling (Serpente et al., 2005a; Vitobello et al., 2011). In a complementary manner, disruptions in the activity of the cytochrome p450 enzymes Cyp26a-c, which degrade retinoids, are known to stimulate or expand domains of Hox expression in the CNS during development (Hernandez et al., 2007; Molotkova et al., 2005; Sirbu et al., 2005; White and Schilling, 2008). Cellular retinoid binding proteins have been shown to be important for maintaining the proper balance or levels of retinoids in hindbrain patterning (Cai et al., 2012). In addition to the direct effects of retinoids on Hox expression, RA modulates the expression domains of the *Cdx* transcription factors and these in turn bind to *Cdx* sites in the Hox clusters to regulate axial expression (Houle et al., 2000; Houle et al., 2003; Lohnes, 2003; van de Ven et al., 2011; van Rooijen et al., 2012; Young et al., 2009).

In mice Hox genes are among first to respond *in vivo* upon exposure to exogenous levels of RA through the administration of retinoic acid to pregnant females. Ectopic levels of RA have been shown to dramatically alter Hox expression patterns in developing nervous system inducing an anterior expansion of many genes (Conlon and Rossant, 1992; Folberg et al., 1997; Marshall et al., 1996). In chick embryos, ectopic RA provided by transplanted beads alters segmental patterns and neuronal differentiation dependent upon Hox genes (Guidato et al., 2003a; Guidato et al., 2003b). Depending upon the stage of treatment, exposure to ectopic levels of RA has been shown to expand or contract domains of Hox expression in the somites and derivatives of the vertebral column, which shows that the competence of the Hox gene clusters to respond to RA varies over time (Kessel, 1992). *In vitro* studies in various mouse and human embryonic carcinoma cells indicate that genes are progressively induced upon RA treatment in Co-linear fashion (Papalopulu et al., 1991a; Simeone et al., 1990; Simeone et al., 1991). *Cis*-regulatory analyses have found that part of the basis for the response of Hox genes to RA is mediated through the presence of RAREs identified and characterized within Hox clusters (Alexander et al., 2009; Tümpel et al., 2009). RAREs which contribute to regulation have been identified adjacent to mammalian: *Hoxd4* (Moroni et al., 1993), *Hoxb4* (Gould et al., 1998), *Hoxa4* (Packer et al., 1998), *Hoxb5* (Oosterveen et al., 2003a; Sharpe et al., 1998), *Hoxa1* (Dupé et al., 1997; Langston and Gudas, 1992) and *Hoxb1* (Marshall et al., 1994; Ogura and Evans, 1995a; Ogura and Evans, 1995b; Studer et al., 1998b; Studer et al., 1994).

With respect to functional specificity, studies on targeted mutagenesis of Hox genes in mice reveal that some biological functions can be unique to each HOX protein while others may be shared with other HOX proteins (Maconochie et al., 1996; Mallo et al., 2010; Wellik, 2007). For example, analyses of *Hoxa1*, *Hoxb1* and *Hoxb2* mouse mutants indicates that these genes have a common role in patterning the identity of rhombomere 4 in the hindbrain and in formation of facial motor neurons through cross-regulatory mechanisms (Alexander et al., 2009; Gavalas et al., 2003; Studer et al., 1998b; Tvrdik and Capecchi, 2006). While many paralogous Hox genes display an ability to functionally compensate for each other (redundancy) there is also strong evidence for distinct functions (Wellik, 2007). It is unclear to what extent these unique functional roles relate to differences in domains of expression between the genes or to subtle variations in the HOX proteins themselves and their down-stream target loci.

In light of their common origin, HOX proteins have very similar homeodomains and overall structures which bind to simple sequences with relatively similar sequence preferences (Berger et al., 2008). Hence, their individual specificity for target sites *in vivo*

is likely to be modulated through the involvement of cofactors or interacting proteins. Regulatory studies in *Drosophila*, mouse and zebra fish have demonstrated that DNA binding specificity and target recognition of HOX proteins is influenced by interactions with co-factors. Evidence to support this comes from studies of two major classes of HOX co-factors: PBC and MEIS, which indicate they have an ancient role in potentiating the binding specificity of HOX proteins (Hudry et al., 2012; Slattery et al., 2011). For example, in the mouse hindbrain activation of Hox genes and other key regulator transcription factors are transiently induced by signaling pathways such as RA and FGF. Their continuing expression is maintained through both epigenetic mechanisms (Polycomb and Trithorax) but auto- and cross-regulation among Hox genes themselves also has a major input into maintenance of Hox expression (Tümpel et al., 2009). *Cis*-regulatory analyses of these interactions have uncovered a number of *in vivo* Hox response elements. Characterization of these has shown that they frequently contain bipartite recognition sites for heterodimers of HOX and PBX proteins.

PBC and MEIS are classes of TALE (Three amino acid extension)-homeodomain containing proteins that can have HOX-dependent and HOX-independent functions (Longobardi et al., 2013; Penkov et al., 2013). The most commonly characterized PBC proteins in mice are the Pbx family and the most studied member of the MEIS class are the MEIS and PREP proteins (Penkov et al., 2013). Many of the HOX binding sites or HOX response elements identified *in vivo* at gene enhancers and promoters were found in association with adjacent PBX and/or MEIS/PREP binding sites (Manzanares et al., 2001). Mutating the PBX and MEIS binding sites in Hox gene targets (which include Hox genes themselves) can prevent expression (Ferretti et al., 2000; Gould et al., 1997; Maconochie et al., 1997; Manzanares et al., 2001).

The hexapeptide region adjacent to the homeodomain of HOX proteins has been found to be an important site of interaction for the PBC DNA-binding partners and targeted mutations in the hexapeptide domains of HOX proteins prevents them from binding to PBX (Chan et al., 1996; Hudry et al., 2012; Medina-Martinez and Ramirez-Solis, 2003). Alterations of the hexapeptide domain of Hoxb8 result in dominant homeotic transformations similar to that observed in *Hox7* and *Hox9* null mice, however, the expression of these genes was unaffected. This suggests that in the absence of Hoxb8/Pbx interactions, target genes of other Hox proteins were being abnormally activated. Together, these studies highlight the importance of co-factors, such as PBX and MEIS, in guiding HOX proteins to their appropriate and tissue-specific targets essential for their *in vivo*

functions. However, they also illustrate critical gaps in understanding genuine Hox target genes/sites and the underlying mechanisms controlling differential Hox specificity *in vivo*.

Understanding how the Hox genes are coupled to these signaling gradients and how different outcomes are mediated by the different members of the Hox family of proteins is central to building knowledge on control of morphogenetic processes. Genetic studies have provided valuable and general insight into aspects of the specific and overlapping functions regulated by Hox proteins. Each Hox protein may have specific DNA binding properties through which it exerts its unique function, but how then is the binding of other Hox proteins on common targets achieved to explain their shared or overlapping functions? It is relatively easy to understand different functional roles for related Hox genes if these are generated by differences in their spatial or temporal expression domains. However, the rules or principles which underlie the binding of similar or different Hox proteins to the same or distinctly different target sites, when they are expressed, is poorly understood. Hence, understanding what types of sites Hox proteins bind *in vivo*, and what governs the binding specificities of Hox proteins critical for their regulatory function is a fundamentally unsolved problem. Invoking the action of cofactors such as PBX, MEIS (TALEs) provides valuable insight to the problem. However, are these the only cofactors for Hox proteins and how do they participate in specificity of different Hox proteins?

In this chapter, I will try to deal with current status of Hox gene binding specificity through understanding historical perspective of discovery of Hox gene using single probe, illustrating unique and redundant functions of Hox through its loss of function phenotype and finally biochemical and structural properties of Hox genes and Hox-Cofactor interaction and its role in binding specificity and functional outcome. At the end of this chapter, I will try to summarize known downstream target genes and DNA motifs to update with current status of our understanding.

1.1 History

This section of introduction will deal with history of discoveries of Hox genes and their function in vertebrate and invertebrates. This section is not merely fascinating documentation of discoveries but will give a glimpse of how one single probe led to discoveries of genes containing homeobox in various organisms. This is relevant for this thesis since it gives an idea about sequence similarities of Hox genes within and between organisms. Bridges described the first example of a homeotic transformation in *Drosophila* in 1913 and this mutation, known as *bithorax*, was later mapped by Ed Lewis

in the bithorax complex (Lewis, 1978; Lewis, 1994). The mutant was fascinating since it generated a four wing fly due to transformation of the third thoracic segment into a second thoracic segment identity. This may have been the first regulatory gene described in eukaryotes. In 1966, Walter Gehring identified a spontaneous mutation in *Drosophila* where antennae on head were transformed into legs and named it *Nasobemia* (Gehring, 1966). Genetically this mutation was later mapped to a chromosomal location and the locus was called *Antennapedia*. The first indication of a repeat sequence which would become the homeobox arose from cross-hybridization of an *antennapedia* cDNA probe to other loci outside of the *antennapedia* locus. This cross-hybridizing region was later identified as the *fushi tarazu* gene (Garber et al., 1983). Bill McGinnis and coworkers showed that *ultrabithorax* also cross-hybridized with *antennapedia* and *fushi tarazu*. This demonstrated a shared repeat region in multiple genes is associated with homeotic patterning defects and was termed the homeobox. Upon cloning this region, they found that homeobox repeat of these three genes are highly similar in sequence (75-77% identity) and encode a 60 amino acid module, termed the homeodomain (McGinnis et al., 1984). Independently Scott and Weiner (Scott and Weiner, 1984) reached the same conclusion and co-discovered the homeobox using analogous techniques to study homeotic loci being investigated in Tom Kaufman's group.

In Walter Gehring's lab, Mike Levine and Ernst Hafen (Hafen et al., 1983; Levine et al., 1983) developed *in situ* hybridization protocols for these transcripts and were the first to reveal the spatially-restricted expression pattern of a gene, linking it to segmentation. This was exciting because the domain of expression they observed was consistent with the affected region of the embryo in the mutant phenotype. Andres Carrasco in collaboration with Eddy de Robertis cloned a vertebrate homolog of a homeobox containing gene in *Xenopus* (Carrasco et al., 1984) and at the same time in Walter Gehring's lab, Bill McGinnis and Frank Ruddle cloned a mouse homeobox gene (McGinnis et al., 1984). This evidence for evolutionary conservation of a potentially important protein regulatory domain laid the foundation which stimulated efforts by many groups to study these homeobox genes in diverse organisms. The use of sequence conservation of genes identified in *Drosophila* became an effective means to clone a wide variety of developmental regulatory genes in vertebrates.

In 1984, the actual function of Hox genes and encoded proteins was still a matter of much speculation. It was assumed they could encode transcription factors but experimental evidence was lacking. Informatics analysis revealed that a significant similarity existed between the homeodomain and yeast mating type proteins, which function as a

transcriptional repressor (Strathern et al., 1981). Further structural similarity with prokaryotic proteins lead to speculation that homeodomain might be a helix turn helix protein (Shepherd et al., 1984). These observations lead to a widely accepted working hypothesis that homeodomain-containing proteins functioned as gene regulatory factors. In vitro assays in cultured cells provided biochemical support for this model (Biggin and Tjian, 1989; Jaynes and O'Farrell, 1988; Thali et al., 1988). In vivo support arose from analyses of *fushi tarazu*, showing that the gene contained an auto-regulatory region which bound FTZ to integrate feedback interactions, conclusively illustrating this homeodomain protein acted as a transcription factor (Schier and Gehring, 1992).

The amazing conservation in the Co-linear organization, expression and function of Hox/HOM-C clusters between *Drosophila* and mammals (Akam, 1989; Duboule and Dollé, 1989; Graham et al., 1989), reviewed in (Duboule, 1994; Duboule and Morata, 1994; Kmita and Duboule, 2003; McGinnis and Krumlauf, 1992) suggested there might be a common gene toolkit for regulating the basic body plans of animals. This opened up many opportunities for investigating the basis for similarities and differences in animal evolution and development, probing regulation of gene clusters in time and space, and understanding the control of morphogenic processes. Studying the complex mechanisms involved in regulation of Hox genes has become a problem of wide general interest. The levels, timing and spatially-restricted domains of Hox expression are critical for their functional roles in patterning, differentiation and development. To achieve these highly modulated patterns of expression it appears that the Hox gene clusters have utilized or exploited nearly every known gene regulatory mechanism available to establish appropriate expression dynamics. Hence, they have become an important system for studying mechanisms of gene regulation through *cis*-elements, auto-, cross- and long-range regulatory elements, epigenetic and chromatin remodeling, nuclear location, chromosome looping, inputs from non-coding RNA, differential translation of mRNAs and stability of RNAs. Furthermore, the HOX protein family has also become an important model to explore mechanisms and specificity of DNA binding properties inherent in a large family of transcription factors. It is no wonder that the rich and fascinating history along with diverse roles of Hox genes in animal evolution has made this important family of proteins a treasured gem for developmental biologists.

1.2 Evolution of Hox genes

This section of thesis will give idea about evolution of Hox genes and other transcripts (including non-coding) from Hox cluster. This discussion is important to

understand evolutionary constraints operating in the Hox cluster. Such constraints lead to large scale conservation in Hox and non Hox genes including non-coding transcripts and their synteny across vertebrate and non-vertebrate species. Because Hox genes and clusters arose by duplication and divergence from a common ancestor, the resulting similarity between Hox genes/proteins within a species and between species in combination with conserved roles in axial patterning links them as a key node for studies on the evolution of animals (Fig. 1-1). Many properties of Hox clusters, co-linearity, and posterior prevalence, and direct auto-, para- and cross-regulation and long-range or global regulation appear to be common features of the regulatory landscape of Hox genes among many species. In this context, the multiple rounds of whole genome-wide duplications in animal evolution have generated multiple Hox complexes and a series of paralogous genes. In this process some clusters and paralogs have diverged, differentially maintaining properties of the ancestral cluster and they have evolved new

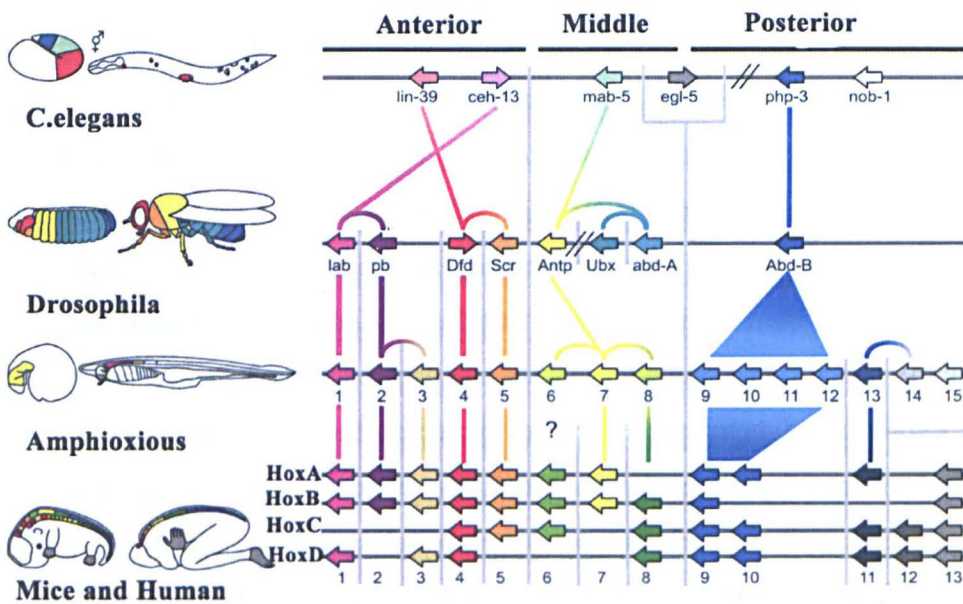


Figure 1-1. Origin and divergence of Hox cluster

The organisms are ordered to show origin, divergence and duplication of Hox cluster in various organisms. The figure is not supposed to indicate that *Drosophila* is descended from *Caenorhabditis*. Vertical gray lines delineate sequence similarity groups. Colored lines linking Hox-genes indicate sequence similarity. The colors are used to represent groups of similar sequences, except for the ‘non-colors’ white and gray. These ‘non-colors’ indicate proteins with considerable sequence divergence to any other sequence in the model organisms we compare (Modified from Hueber et al., 2010).

roles (Hoegg and Meyer, 2005; Holland, 2013; Kuraku and Meyer, 2009; Nolte et al., 2012). This makes understanding the evolution of Hox clusters and their conserved

versus unique roles important from an evolutionary perspective, important for understanding the underlying regulatory processes and mechanisms associated with generation of diversity in the basic body plan during evolution (Carroll, 1995). The majority of reviews and analyses on the evolutionary origins and expression of Hox genes and clusters have focused on the coding genes or domains transcribed in Hox clusters. However, recent advances in genomics, such as hybridization based tiling arrays and Next-Generation sequencing methods have unearthed a treasure box of sense and anti-sense non-coding sequences (miRs and lincRNAs) transcribed in Hox clusters and flanking regions that are believed to participate in regulation of their temporally and spatially restricted patterns of expression. In this section of the introduction I will review knowledge on the evolution of Hox clusters and try to integrate emerging analysis of transcriptome of Hox clusters to include micro-RNAs and other non-coding transcripts.

The earliest example for the presence of Hox-like genes can be seen in Cnidarian (Fig.1-2) (Chiori et al., 2009; Chourrout et al., 2006; Finnerty et al., 2004; Kamm et al., 2006; Matus et al., 2006). The analysis of cnidarians suggests that a protohox locus contains at least one anterior and posterior paralogs and lacks central paralogs (Quiquand et al., 2009).

The first time at least one gene from all three anterior, central and posterior Hox paralogs can be seen is in Acoels (Hejnal and Martindale, 2009; Moreno et al., 2009). The majority of arthropods have seven to eight Hox genes, indicating there were at least 10 genes in an ancestral arthropod cluster (Powers et al., 2000). Unlike arthropods, molluscs were identified with 5-11 Hox genes, carrying two paralog 5 genes. Ancestral echinoderms and hemichordates had at least 12 genes and they are seen in a single cluster with expression in a spatially Co-linear fashion (Fig. 1-2) (Freeman et al., 2012; Popodi and Raff, 2001). The *Drosophila* Hox complex is split into two regions, 700Kb apart on the

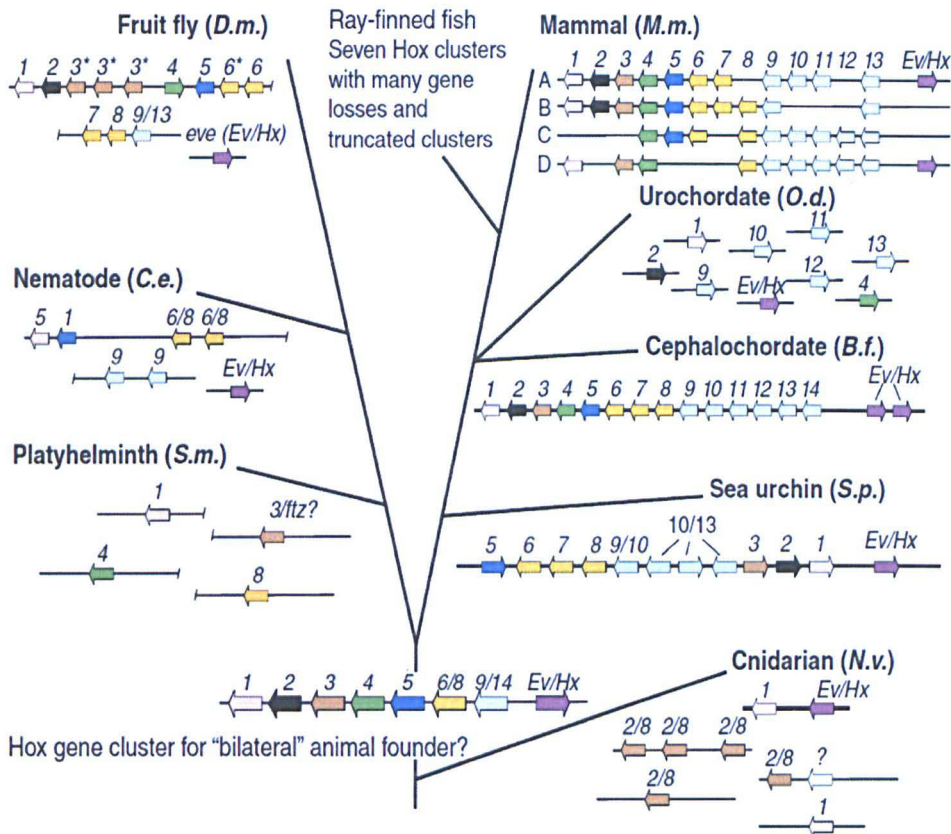


Figure 1-2 Cladogram depicting Hox gene chromosomal organization for representative animals.

At the base is shown a cnidarian (*Nematostella vectensis*), which has a dispersed genomic organization of Hox genes and lacks posterior Hox paralogs. The left branch displays fragmented Hox clusters for the lophotrochozoan flatworm *Schistosoma mansoni* and the ecdysozoan fruit fly (*Drosophila melanogaster*) and nematode (*Caenorhabditis elegans*). The right (deuterostome) branch portrays the rearranged but coherent Hox cluster of the sea urchin *Strongylocentrotus purpuratus*, the “prototypical” Hox cluster of *Branchiostoma floridae* (a cephalochordate), the dispersed genomic organization of the Hox genes of a urochordate (*Oikopleura dioica*), and the quadruplicated Hox clusters of a mammal (*Mus musculus*), which remain coherent but have experienced losses of multiple paralogs. Similar to the mammals but not shown diagrammatically, the ray-finned fish have multiple duplicate. Adapted from Derek Lemons and William McGinnis, Science 313, 1918 (2006)

same chromosome (Carroll, 1995; Lewis, 1978; Powers et al., 2000). These two complexes are known as the Bithorax (BX-C) and Antennapedia (ANT-C) complexes. The Antennapedia complex consists of *Labial (lab)*, *Proboscipedia (Pb)*, *Sex-comb related (Scr)*, *Deformed (Dfd)* and *Zen (Zn)*. The Bithorax complex consists of three genes, *Ultrabithorax (Ubx)*, *Abdominal A (abdB)* and *Abdominal B (abdB)*. Interestingly, the Bithorax complex is split in many *Drosophila* species without compromising spatial expression patterns. Furthermore, the different *Drosophila* species which display a split in BX-C do so at different places (Clark et al., 2007; Stark et al., 2007). BX-C controls the

identity of the posterior two-thirds of a *Drosophila* embryo. *Ubx*, *AbdA* and *AbdB* are expressed in a co-linear spatial and temporally-restricted manner in developing fly embryos.

Table 1-1 Evolutionary status of Hox cluster in various species

	Organism	No. of Hox genes	Dispersed/ clustered	Expressed	Remarks
Cnidarian	Nematostella	7	4 linked		(Chourrout et al., 2006)
	Clytia hemisphaerica	4			(Chiori et al., 2009)
	Hydra	6	Dispersed		(Chourrout et al., 2006)
Acoel	Symsagittifera roscoffensis	3	Dispersed	Nested	(Moreno et al., 2009)
	Convolutriloba longifissura	3	Dispersed	?	(Hejnol and Martindale, 2009)
	<i>C. elegans</i>	6	Dispersed		(Aboobaker and Blaxter, 2003)
	<i>Nereis virens</i>	11	Clustered		(Andreeva et al., 2001)
	<i>Platynereus</i> sp	9	Linked; one gene dispersed		(Hui et al., 2012)
	<i>Drosophila</i> sp	10	Clustered		(Lewis, 1978)
Urochordata	<i>Ciona intestinalis</i>		Dispersed		(Ikuta et al., 2004)
Urochordata	<i>Oikopleura dioica</i>		Dispersed	Spatially co-linear expression in nervous system Co-linear	(Seo et al., 2004)
Urochordata	<i>Balanoglossus misakiensis</i>	12	Clustered		(Urata et al., 2009)
	Sea urchin	10	Clustered		(Arenas-Mena et al., 2000)
	<i>Amphioxus</i>	15	Clustered	Co-linear	(Amemiya et al., 2008)
	<i>Petromyzon marinus</i>		Two		(Smith et al., 2013a)

Ubx is expressed from parasegment 5 to 12-13, while *AbdA* and *AbdB* are expressed in parasegments 7-12 and parasegments 10-14 respectively (Karch et al., 1985; Lewis, 1978; Maeda and Karch, 2006; Mihaly et al., 2006). The overlapping expression

patterns of BX-C genes are generated by a 300kb genomic region divided into 10 *cis*-regulatory domains, namely *abx/bx*, *bxd/Pbx* and *iab-2-9* and these are separated by insulators (Beachy et al., 1985; Karch et al., 1990; Mihaly et al., 2006; Mihaly et al., 1998; Struhl and White, 1985; White and Wilcox, 1984, 1985; Zavortink and Sakonju, 1989). The *Drosophila* HOM-C cluster contains genes which encode non-HOX proteins, which is different than other insects or the vertebrate Hox clusters. Similar to the vertebrate Hox clusters, large numbers of non-coding transcripts are seen to arise from the *Drosophila* HOM-C. In fact, Ed Lewis in his original analysis postulated that many *cis*-regulatory regions were regulatory RNAs because he found evidence for transcripts, but this concept was not favored by the community for a long time (Lewis, 1978).

Interestingly, these non-coding transcripts exhibit spatial co-linearity in expression and transcription and they never breach insulator boundaries. Intriguingly, abolition of insulator function is observed when transcription proceeds through insulator elements and leads to segmental transformation (Busturia et al., 1989; Cavalli and Paro, 1998; Lipshitz et al., 1987; Rank et al., 2002) (Fig.1-3).

The *Tribolium* Hox cluster spans ~ 756 kb of genomic DNA. It contains all eight *Drosophila* Hox genes and contains the Hox-derived genes *fushi tarazu* and *zen*. *Tribolium* contains two paralogs of *zen*. The order of related genes is same as other insects but all genes are oriented in same direction with respect to transcription, unlike *Drosophila*. The *Tribolium* Hox cluster has microRNA genes, namely *miR-10* and *miR-iab-4*. Through the *Tribolium* Hox cluster is constrained and preserves an overall organization similar to vertebrates, surrounding regions show no syntenic relationship with an ancestral cluster or the vertebrate Hox clusters.

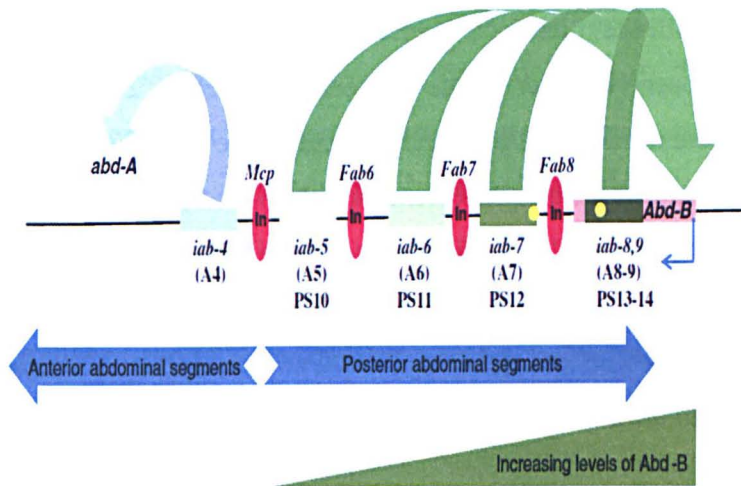


Figure 1-3 Cis-regulatory modules within the Abd-B locus of the bithorax complex involved in co-linearity of Hox genes in *Drosophila*

The Abd-B gene is activated from PS10-14 in the developing embryo. The levels of Abd-B are the lowest in PS10 and highest in PS14, and this is believed to be either due to the strength of the iab's (shown in shades of green rectangles) or the distance of iabs from the Abd-B promoter. The iab-5 enhancer drives the expression of Abd-B in PS10, iab-6 in PS11, iab-7 in PS12 and iab-8, 9 in PS13-14. Each of these enhancer domains is demarcated by insulator elements (red oval) like Mcp, Fab6, Fab7 and Fab8. The iabs are also known to contain PRE elements (yellow circle) as experimentally identified for the iab-7 and iab-8 PREs. The earlier iab elements such as the iab-4 act upon the abd-A gene in anterior PS.

(Adapted from Chopra, 2011)

The *Tribolium* Hox cluster is devoid of any non-HOX protein coding genes. Three non-coding transcripts were identified in *Tribolium*. One has been found between *ptl/Tc-Antp* and *Utx/Tc-Ubx* and other in first intron *Utx/Tc-Ubx* (Shippy et al., 2008).

Mosquito Hox genes are found in 700kb genomic region and form a single cluster, as in *Tribolium*. The anopheles Hox genes are arranged from left to right and consist of *lab, pb, Zen, Dfd, Scr, ftz, Antp, Ubx, abdA*. This Hox cluster does not contain a second *Zen* gene (Devenport et al., 2000; Powers et al., 2000).

In grasshopper (*Schistocerca gregaria*) *sgzen, sgdax, Scr, abdA* and *abdB* have been identified in a single cluster spanning more 700kb, but less than 2Mb (Ferrier and Akam, 1996). *Manduca sexta* have orthologs of *Antp, Ubx* and *abd-A* (Nagy et al., 1991; Zheng et al., 1999). In *Bombyx Mori* the Hox genes form a single cluster with the exception that *lab* split from the other genes. This silk worm Hox cluster is ~4X the size relative to *Drosophila*. *Bmzen, Dfd, Scr, Antp, Ubx, abd-A* and *abd-B* are present in silk worm and the cluster seems to be highly similar to *Tribolium* (Yasukochi et al., 2004).

In hemichordates, *Saccoglossus kowalevskii* and *Ptychodera flava* Hox clusters have been cloned and analyzed in detail (Freeman et al., 2012). Both hemichordates have 12 Hox genes in single cluster with the microRNA *miR10* positioned between members of paralogous groups 4 and 5, as observed in vertebrate clusters (Nolte et al., 2012; Smith et al., 2013b). *Hox1-10* paralogs are oriented in the same manner as the hypothetical ancestral Hox cluster, in terms of order and transcriptional orientation. However, the last two genes (paralogous group 11/13 like) are inverted. In *Saccoglossus kowalevskii* all Hox genes except *Hox9/10* have two exons and the homeobox resides in single exon. The average intergenic distance is 36kb and 34 kb in *Saccoglossus kowalevskii* and *Ptychodera flava* respectively. In *Ptychodera flava* all Hox genes have two exons except *Hox5* and *Hox7*. The *Saccoglossus kowalevskii* Hox cluster has two sense and antisense non-coding transcripts. *Saccoglossus kowalevskii* Hox genes are expressed in co-linear fashion in anterior posterior axis as in vertebrate nervous system and protostomes ectoderm (Freeman et al., 2012; Lowe et al., 2003; Pani et al., 2012).

Among Urochordates, the *Ciona sp* Hox cluster has been studied extensively. *Ciona* has a dispersed and disintegrated Hox cluster spanning around 5 MB genomic regions (Ikuta et al., 2004; Natale et al., 2011). The *Ciona* Hox cluster contains many non-HOX protein coding genes and has lost 4 members of Hox gene family. *Ci-Hox1*, -3, -5, -10 and -11 exhibit spatial co-linear expression patterns while temporal co-linearity is lost in these organisms. Many *Ciona* Hox genes are not expressed in nervous system and the *Ciona* Hox genes, *Hox 2*, *Hox4* and *Hox13* do not show expression in ectoderm (Ikuta et al., 2004; Locascio et al., 1999).

The *Amphioxus sp*, a cephalochordate, Hox cluster is considered as “archetypal” for chordates. The *Branchiostoma floridae* (Amphioxus) Hox cluster consists of 14 Hox genes in single cluster. The AmphiHox cluster contains very little evidence for invasion of repeat elements. The *Hox1-4* and *Hox6* members are expressed in a co-linear fashion in CNS (Garcia-Fernandez and Holland, 1994; Holland et al., 1992; Wada et al., 1999). *Hox1*, *Hox3* and *Hox4* are involved in controlling the fate of neurons in AP axis. In the AmphiHox cluster there is conservation in positioning of paralog groups 1-10 however there appears to be a selective expansion of the most posterior paralogous groups, which has been proposed to reflect “deuterostome posterior flexibility” (Gibbs et al., 2004). In *B. lanceolatum* *Hox6* is not expressed in co-linear fashion as in case of *B. floridae*. In lamprey and shark, a *Hox14* paralog is expressed in posterior ectoderm while in case of Amphioxus *Hox14* is expressed in anterior cerebral vesicles (Amemiya et al., 2008).

Intriguingly, regulatory studies testing AmphiHox genomic regions in vertebrate embryos using transgenic reporter constructs reveals that retinoic acid response elements may be positioned in similar places adjacent to the group 1 and group 4 paralogs (Holland and Holland, 1996; Manzanares et al., 2000; Wada et al., 2006). These RARE-based control modules might represent conserved regulatory features of an ancient cluster present before the emergence of vertebrates. In accord with this idea, studies by Chris Lowe and colleagues using a hemichordate, the acorn worm *Saccoglossus kowalevskii*, have shown that the A/P domain map of transcription factors (including Hox) and signaling ligands can surprisingly be used to align the body plans of hemichordates and chordates (Lowe et al., 2003; Pani et al., 2012). This implies that the axial signaling centers must have evolved long ago in a common chordate ancestor. Hence, the conserved positions of RAREs in AmphiHox clusters might reflect an ancient *cis*-signature of retinoid signaling associated with generating ordered domains of Hox expression in animal evolution.

Unlike other vertebrate Hox clusters which contain 39 genes in four separate clusters (see sections below), teleosts and ray-finned fishes have undergone an extra round of genome-wide duplication and some other fish display additional rounds (Hoegg et al., 2007; Hoegg et al., 2004; Kuraku and Meyer, 2009; Meyer and Van de Peer, 2005; Taylor et al., 2003). As a result, along with the genome duplication and loss of paralogs for duplicated clusters, the fishes exhibit a wide range of variation in both the number of Hox genes and number of clusters. In all fishes, at least seven Hox clusters are identified. *Danio rerio* (Zebra fish), *Takifugu rubripes*, *spheroids nephelus* (Puffer fish) and *oryzias latipes* (Medaka) have 48, 51, 50 and 33 Hox genes respectively (Amores et al., 1998; Amores et al., 2004; Aparicio et al., 2002; Hurley et al., 2005; Naruse et al., 2000).

Mammalian genomes contain 39 Hox genes organized in four separate clusters and there is a remarkable degree of similarity with other vertebrates (Fig 1-1 and 1-2). Ancestral microRNAs such as *miR10* are conserved in these species. No non-HOX protein coding genes are present in the mammalian Hox clusters. With the exception of region between *Hoxb9* and *Hoxb13*, the four mammalian Hox clusters are devoid of repeats and there is substantial synteny extending to other genes adjacent to the Hox clusters. This might be a reflection of regulatory constraints, since clustered Hox genes are regulated in part through long-range interaction from elements outside clusters (Kmita and Duboule, 2003). A classic example of this is the GCR region in HoxD cluster important in modulating expression in limbs (Noordermeer and Duboule, 2013; Spitz and Duboule, 2008; Spitz et al., 2003; Tarchini and Duboule, 2006). Mammalian Hox clusters are subject to complex auto-, para- and cross-regulation by the Hox genes/proteins themselves and

many of the *cis*-regulatory regions integrating these inputs are highly conserved in mammals and vertebrates (Alexander et al., 2009; Tümpel et al., 2009). In addition, the ability to respond to morphogenic signals such as RA through RAREs appears to be a property that evolved during chordate ancestry. The positions of these RAREs appears to be highly conserved in the mammalian Hox clusters and important for their regulation in diverse tissues (Alexander et al., 2009; Manzanares et al., 2000; Nolte et al., 2013; Tümpel et al., 2009).

Mammalian Hox clusters show extensive transcriptional activity during development. Both coding and non-coding regions are transcribed from both strands during development (Gupta et al., 2010; Rinn et al., 2007; Sasaki et al., 2007; Sessa et al., 2007; Zhang et al., 2009). The general functional significance of such non-coding transcription is not yet clear. However, evidence from several groups suggests an important role for non-coding transcription in regulation of Hox genes (Gupta et al., 2010; Ma et al., 2010; Ma et al., 2007; Maamar et al., 2013; Rinn et al., 2007). Rinn and coworkers showed that intergenic transcripts are associated with active Hox genes. About 74% of all intergenic transcripts are transcribed from the opposite strand of Hox genes (Rinn et al., 2007). Nearly 90% of non-coding transcripts correlate in *cis* with its 3' Hox counterpart (Rinn et al., 2007). Like Hox genes, non-coding RNAs (ncRNAs) also vary their expression along the developmental axis of the body and follow co-linearity (Rinn et al., 2007). Out of 231 Hox ncRNA, about 147 (64%) are differentially expressed. There are 48 Hox ncRNAs differentially expressed in coordination with their neighboring Hox genes along proximal-distal axis (Rinn et al., 2007). All 41 transcribed regions from both Hox and ncRNAs induced in the distal sites belong to paralogous group 9-13 and those repressed in distal region belongs to paralogous group 1-6. This recapitulates the evolutionary origin of two domains from *Drosophila*, namely *Ubx* and *antennapedia*. The expression patterns of 90% of ncRNAs are coordinately induced with their 3' Hox genes while only 10% expressed differentially (Rinn et al., 2007).

Sessa and co-workers identified many intergenic transcripts in the human Hox cluster using *in silico* method and analyzed six transcripts for expression in human adult and embryonic tissues (Sessa et al., 2007). These non-coding transcripts were shown to be differentially spliced and are part of a long transcript. In many cases they were shown to be associated with the active state of nearby Hox genes. Over expression of antisense or sense ncRNAs do not show any effect while *in vitro* differentiation of human carcinoma cells by retinoic acid treatment facilitated expression of non-coding transcripts in a co-linear fashion with nearby Hox genes (Sessa et al., 2007). They further showed that the intergenic

transcripts of HoxA genes are targets of PcG complex PRC2 and characterized by PcG-specific histone modification patterns. Upon induction with retinoic acid, specific changes in histone modification take place due to loss of interaction with *Eed/Ezh2-Pcg* complex.

Mainguy and co-workers identified 15 blocks transcribed antisense to Hox genes. Antisense transcripts represent a maximum of 38% of spliced transcripts (38.46% for HoxA, 33.11% for HoxB, 13.16% for HoxC and 34.84% for HoxD). Furthermore, at least three sense-antisense transcripts are conserved between human and mouse which suggests a putative functional role (Mainguy et al., 2007).

Sasaki and coworkers identified a non-coding RNA, *HIT18844*, upstream of *Hoxa13*; they found to be transcribed from the opposite strand to the HoxA genes (Sasaki et al., 2007). They also found another ncRNA in the vicinity. They identified ncRNAs in intergenic regions between *Hoxa1* and *Hoxa2*, *Hoxa3* and *Hoxa4* and between *Hoxa6* and *Hoxa5*. In humans and mice these non-coding RNAs are conserved and transcribed from the opposite strand. The expression pattern of these non-coding RNA from HoxA locus reflects those of neighboring Hox genes and follows co-linearity patterns. *HIT18844* contains an extremely conserved block of 265bp across species. The conserved block maps to a position 1.8 kb upstream of *Hoxa13* gene and 1568 bases from the 5' terminus of *HIT18844* in humans (Sasaki et al., 2007).

The summary above indicates that the complex body plan in bilaterians was achieved in part through coupling Hox genes and their activity to axial patterning and exploiting the diversity in number of Hox genes and clusters which arose through duplication and divergence in animal evolution. These changes in number of Hox genes were achieved through genome-wide duplications in association with localized deletion and /or duplication of Hox genes within Hox clusters.

1.3 Function of Hox genes in mammals-lessons from gain and loss-of-function mutants in mice

This section will review loss of function phenotypes of Hox genes in CNS and Axial skeleton. This is important to understand functional genetic interactions between and among paralogous Hox genes. This further helps to understand hierarchy of interactions and unique and redundant role of various Hox genes in determining anterior posterior polarity in developing CNS and Axial skeleton. The restricted and ordered Hox expression patterns in bilaterians are defined as a 'zootype' which contributes to generation of morphologically distinct body segments. Alteration of Hox gene expression through gain- or loss-of-function perturbations have been associated with changes or loss in segmental

identity, as most clearly observed by phenotypes in the axial skeleton of vertebrates (Mallo et al., 2010; Wellik, 2009; Wellik and Capecchi, 2003). In this section I will describe evidence on the functional roles of Hox genes based largely on analyses arising from genetic studies in mice. The ability to make targeted mutations in genes and their regulatory regions has facilitated a comprehensive body of work on the Hox gene family. While published studies have detailed functional roles in many tissues, organs, developmental stages and in the adult I will primarily focus the description on phenotypes in head and anterior neural tissues, as this is most relevant to the experimental studies I will describe in the results chapters of the thesis.

The murine Hox genes are expressed in paraxial mesoderm, lateral mesoderm, neuroectoderm, neural crest, and endoderm and contribute to the control of craniofacial development, hindbrain patterning, the axial and abaxial skeleton, limb, gut and genitourinary tract during embryonic development.

Hox gene expression can be first detected in posterior part of the primitive streak which spreads anteriorly rostral to node in developing chicken and mouse embryos (Deschamps et al., 1999; Deschamps and van Nes, 2005; Forlani et al., 2003; Iimura and Pourquie, 2006). In mesoderm and its derivatives, after ingression from the primitive streak, a progressive series of Hox genes are activated and expression extends anteriorly until there is in general a set of fixed anterior boundaries by ~12.5dpc of development. In neuroectoderm and mesoderm, early and late expression boundaries are different and this highlights the dynamic changes in activity between 9.5 dpc to 12.5 dpc of development. The establishment of Hox expression in neuroectoderm and mesoderm are independent utilizing separate regulatory elements (Alexander et al., 2009; Whiting et al., 1991). In both tissue types, expression can be influenced through cell-cell contacts. Posterior cells expressing Hox genes in early stages of gastrulation migrate and make contact with anterior cells to extend Hox expression more anteriorly (Deschamps and van Nes, 2005). The RA, Fgf and Wnt major signaling pathways critical or controlling axial elongation also couple Hox genes to this process in order to provide the nested domains of expression which patterns these tissues as they are generated (Diez del Corral et al., 2002; Diez del Corral et al., 2003; Mallo et al., 2010; Young et al., 2009).

In hindbrain the expression of Hox genes is tightly coupled to the process of segmentation and Hox genes play an important role in modulating segmental identity (Alexander et al., 2009; Lumsden and Krumlauf, 1996; Tümpel et al., 2009). Hox genes are also expressed in ordered patterns in cranial neural crest and this provides a code for patterning the branchial regions of the vertebrate head (Hunt et al., 1991; Murakami et al.,

2005; Santagati and Rijli, 2003; Trainor and Krumlauf, 2000a; Trainor and Krumlauf, 2000b, 2001) (Fig. 1-4). Hox gene expression patterns established in rhombomeres are not passively translated into arches indirectly by the migration of neural crest cells. Rather they are mediated through independent *cis*-regulatory elements (Maconochie et al., 1999). However, Hox genes impact the ability of neural crest cells to migrate from hindbrain segments and enter the branchial arches (Gavalas et al., 2001).

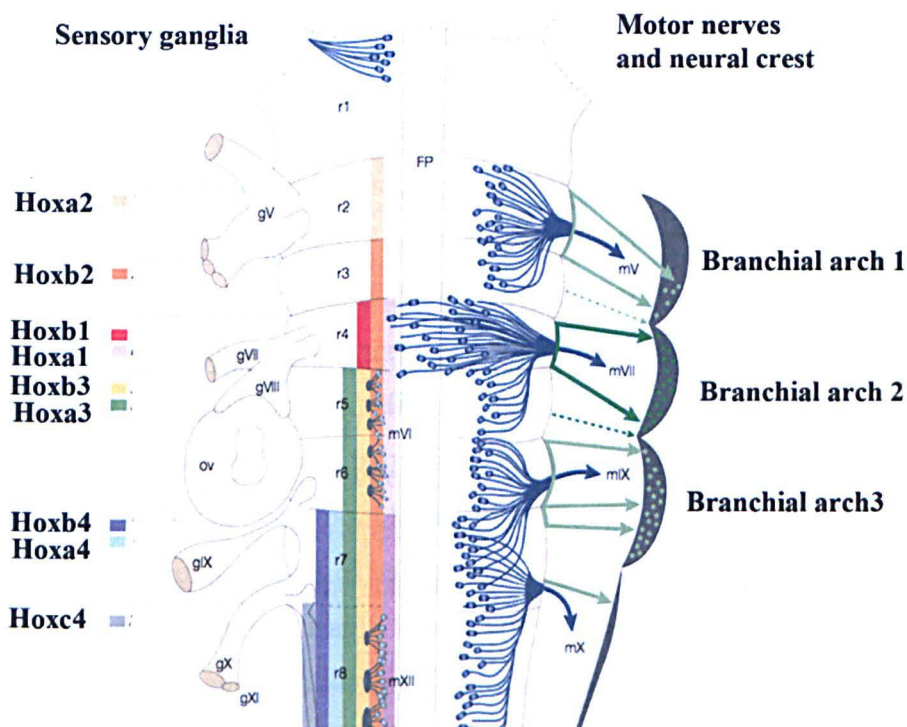


Figure 1-4 Nested Expression of Hox genes in hindbrain.

Anatomy of hindbrain is also shown. Sensory ganglia and motors nerves are also marked in the figure (Kiecker and Lumsden, 2005)

Hox genes also play a role in pharyngeal endoderm (Manley and Capecchi, 1995; Manley and Capecchi, 1998). Hence, the hindbrain, cranial neural crest, endoderm and mesoderm have coordinated and nested patterns of Hox expression which are critical for proper patterning of head and craniofacial development. With respect to these roles for Hox genes in head development, as a consequence of spatial and temporal co-linearity primarily only Hox genes in paralogy groups 1-4 are expressed in the relevant anterior tissues. In mouse *Hoxd1* and *Hoxc4* are not expressed in neural tissues. Hence, in mouse only 10 genes from these paralogy groups appear to be critical for patterning head development at these early developmental stages and I will focus my discussion on the evidence for their roles.

1.3.1 Group 1 paralogs

Studer and Colleagues (1996) showed that loss of *Hoxb1* in developing mice embryos leads to failure to maintain r4 identity in hindbrain. In the hindbrain markers of r4 such as *Wnt8*, *Hoxb3* and *CRABP1* fail to be properly up-regulated in mutant embryos. At the same time r2 markers, *sek1* and *r2-HPAP* are expressed ectopically in the r4 region. This suggests that there is a transformation from of r4 to r2, whereby r4 adopts the identity of r2 (Studer et al., 1996). This results in abnormal posterior migratory behavior of the r4 derived facial neurons which instead migrate laterally in a manner similar to trigeminal neurons in r2, adding support for a homeotic transformation.

Goddard and coworkers (1996) have shown that loss-of-function of *Hoxb1* is characterized by facial paralysis similar to Bell's palsy or Moebius syndrome. This phenotype can be attributed to complete absence or 2-3 fold reduction of somatic motor branches of VIIth nerve. Neurons present along medial margins of r4 and r5 are also absent or reduced in *Hoxb1* mutant (Goddard et al., 1996). Similar migratory defects in facial motor neurons were seen with this allele compared with that described by Studer and colleagues (Studer et al., 1996). The changes to the facial motor nerves in these studies alter facial muscle integrity, which is greatly compromised in *Hoxb1* mutant mice. Zygomatic, buccinators, depressor anguli oris and caudal digastric muscles were greatly reduced while temporalis and masseter muscles were abnormal. Nerves innervating the stapes and stapedius muscles are reduced and dissociated from muscles. This alters their ability to feed and contributes to neonatal lethality.

Lufkin and coworkers (Lufkin et al., 1991) found that *Hoxa1* null mice die at birth. Closure of neural tube is delayed in some embryos. In these mutant embryos r4-r6 is not patterned properly presumably due to loss of r5. Motor nuclei of cranial nerves of trigeminal (V), ambiguous (IX, X), dorsal vagus and hypoglossal are normal while abducens (VI) is absent in *Hoxa1* null mice. Motor nuclei of facial nerve (VII) were greatly reduced. Sensory ganglia of trigeminal (V) is normal and geniculate (VII) is always present however the spiral ganglia (VIII) is absent in *Hoxa1* null mice. Vestibular (VIII), superior and jugular (IX, X proximal) ganglia are greatly reduced in mutants while the petrosal and nodose (IX, X distal) ganglia were fused and missing their connection with brain stem. The Vth and XIIth cranial nerves are normal while the VIth is absent. The inter-ganglionic portion of IXth and Xth cranial nerves are missing while VIth cranial nerve shows reduced branching. In *Hoxa1* null mice the otic pit is formed adjacent to r4 and by 9.5dpc, the otocyst is small and displaced laterally and rostrally. The Organ of Corti

is absent and the otic capsule is fenestrated but the middle and external ear are normal. In a different *Hoxa1* null mouse (Chisaka et al., 1992) which vary with respect to the extent of deletion of the coding region and different transcripts generated the putative null mice survive until 3.5 days after birth.

Mice carrying mutations in the retinoic acid response elements adjacent to *Hoxa1* and *Hoxb1* have been generated and used in combination with full null mutants of these paralogs to explore genetic cooperation between the genes (Dupé et al., 1997; Gavalas et al., 1998; Marshall et al., 1994; Studer et al., 1998b). A double 3' RARE mutant *Hoxa1*^{3'RARE}/*Hoxb1*^{3'RARE} shows loss of hyomandibular branch of facial motor nerve with partial penetrance. Double homozygous null embryos for *Hoxa1* and *Hoxb1* show either fusion of V and VII/VIII ganglia or loss of connectivity between hindbrain and VII/VIII ganglia in 10.5dpc embryos (Gavalas et al., 1998; Rossel and Capecchi, 1999). Facial and trigeminal neurons are present in overlapping domains. The VII/VIII ganglia and inferior X ganglia show reduced neurofilament staining. The r5 like facial motor axon projections are lacking. In *Hoxa1*^{null}/*Hoxb1*^{3'RARE} there is a fully penetrant fusion of V and VII/VIII ganglia and as in the double mutant for *Hoxa1* and *Hoxb1*, the facial and trigeminal neurons are present in an overlapping domain. Facial branchiomotor neurons fail to migrate and consistent with the absence of r5 facial motor axon projections are lacking (Gavalas et al., 1998).

In *Hoxa1*^{null}/*Hoxb1*^{3'RARE} r4 is further reduced compared with *Hoxa1*^{null}. Expression of neural crest markers in the second arch at 9.5dpc is greatly reduced or absent and by 10.5dpc the second arch is completely gone. The inner, middle and external ear is either underdeveloped or completely absent. At 17.5 dpc, the styloid process and stylohyoid muscle is completely absent while lesser horns of hyoid bones are absent. This phenotype is same as double homozygous mutant of *Hoxa1* and *Hoxb1*. These genetic studies clearly illustrate that the paralogous group 1 genes, *Hoxa1* and *Hoxb1*, synergize in correct patterning of hindbrain, cranial nerves and second pharyngeal arch (Gavalas et al., 1998).

In zebra fish, the knockdown of the *Hoxb1b* gene (functional equivalent to mouse *Hoxa1*), results in disruption of r4 while disruption of *Hoxb1a* gene (functional equivalent to Mouse *Hoxb1*) is involved in specification of facial nerves originating from r4. A double knock out of these two zebra fish genes does not show the same degree of severity in phenotypes as those in mice. The r4 and r5 segments are present but reduced in size (McClintock et al., 2002). This may reflect roles for other zebra fish Hox genes in patterning these regions.

In *Xenopus*, there is evidence that all three paralogous 1 gene namely *Hoxa1*, *Hoxb1* and *Hoxd1* are important for correct patterning and formation of hindbrain, unlike mice and zebra fish (Choe and Sagerstrom, 2004; Kolm and Sive, 1995; Moens et al., 1998; Moens and Prince, 2002; Prince et al., 1998). A triple knockout of paralogous 1 genes leads to complete loss r4/r5 markers and *Krox20*. But presence of at least one paralogous 1 gene leads to mere alteration in expression domain of *Krox20*. The triple knockdown shows transformation of r2-r7 segments to r1-like identity, which is similar to *Pbx* and *Meis* phenotype in zebra fish (McNulty et al., 2005; Waskiewicz et al., 2002).

Ectopic application of retinoids leads to an anterior expansion of *Hoxa1* and *Hoxb1* (Conlon and Rossant, 1992; Marshall et al., 1992). The hindbrain phenotypes arising from RA treatments resemble those observed with ectopic expression of Hox group 1 paralogs in a number of vertebrate model systems (Alexandre et al., 1996; Choe and Sagerstrom, 2004; Choe et al., 2011; Moens and Prince, 2002; Zhang et al., 1994). The RAREs adjacent to *Hoxa1* and *Hoxb1* provide direct input from retinoid signaling and these genes display the most rapid response to RA induction (Fig.1-5). Hence in the early mouse embryo (7.75-9.0 dpc), the group 1 paralogous Hox genes appear to be a major input for RA signaling in the hindbrain. Retinoids increase r4 identity at the expense of r3 and r5 (Tvrdik and Capecchi, 2006) which mirrors phenotypes generated by loss-of-function mutants in RA degrading enzymes (Abu-Abed et al., 2001; Sakai et al., 2001).

Previous genetic and molecular studies have clearly demonstrated that *Hoxa1* and *Hoxb1* both are required to trigger an auto-regulatory loop whereby *Hoxb1* feedbacks to maintain its expression in r4 through a HOX response element located in its 5' flanking region (Fig.1-5) (Pöpperl et al., 1995; Studer et al., 1998a; Studer et al., 1996). The initial domains of expression of *Hoxa1* and *Hoxb1* required for this are established in the neural tube via the 3' RAREs adjacent to each gene (Dupé et al., 1997; Studer et al., 1998b). *Hoxa1* is activated by RA slightly earlier than *Hoxb1*. While there is evidence for both unique and overlapping functions from the mutational analyses in mice there is a question as to whether these differences are the result of differences in the expression patterns or in the proteins. In an elegant genetic analysis to test for the degree of overlaps in the functions of *Hoxa1* and *Hoxb1* Tvrdik and Capecchi have shown that *Hoxa1* is able

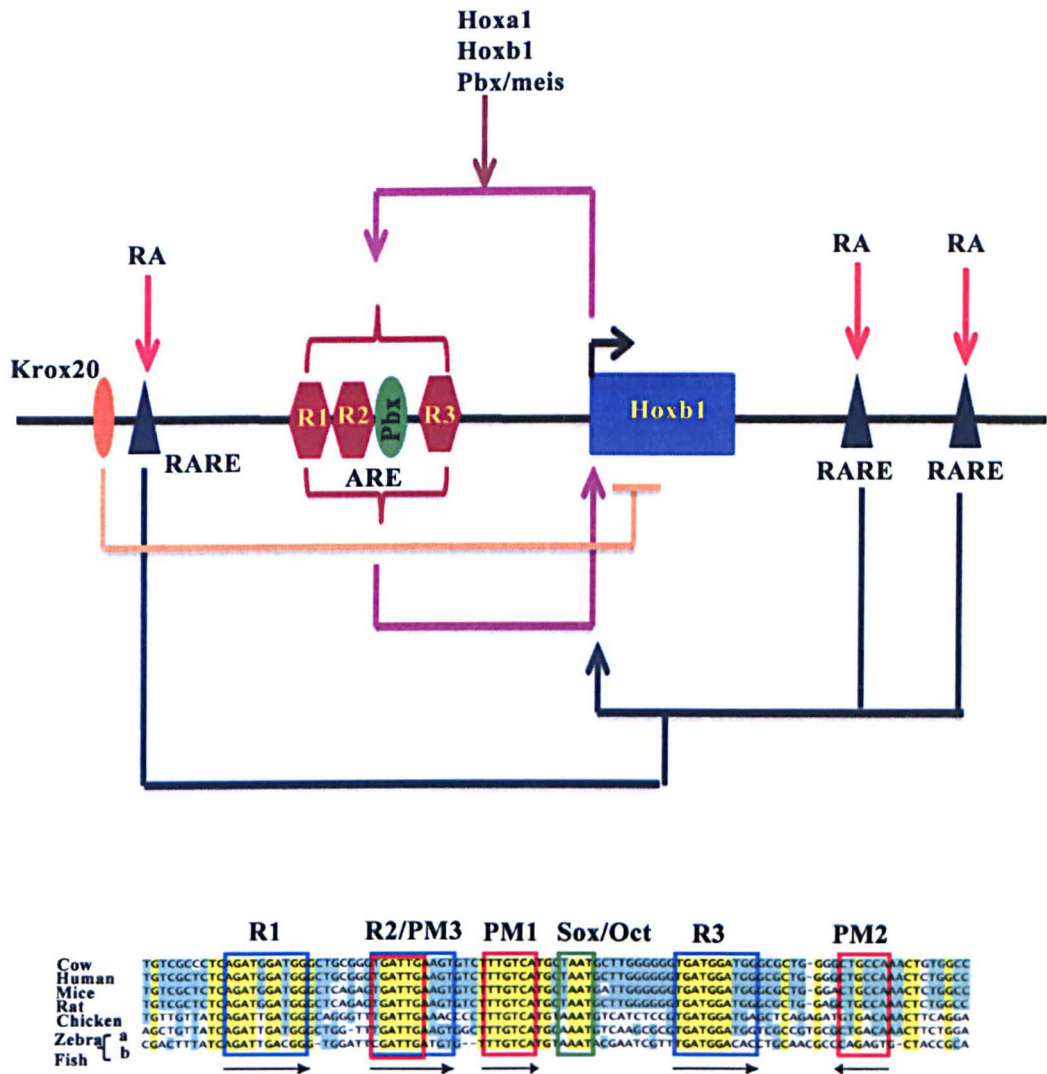


Figure 1-5 Transcriptional regulation of *Hoxb1* through auto and cross-regulatory mechanism

Hoxb1 auto-regulatory element is shown as combination of three conserved block namely R1, R2 and R3. Sequence alignment and conservation across species of this region is shown below. Modified from Tümpel et al, Curr Top Dev Biol. 2009;88:103-37 and Ferretti et al Mol. and Cellular Biology 2005 8541-8522

to replace the role of *Hoxb1* if the *Hoxb1* auto-regulatory control region is placed in the *Hoxa1* locus (Tvrdik and Capecchi, 2006). In essence, if *Hoxa1* is provided with an auto-regulatory elements so it can stimulate its own expression in r4 of the mouse hindbrain in combination with the co-factors like Pbx and Meis, this is sufficient to replace the r4-dependent function of *Hoxb1* (Fig.1-6). This leads to the speculation that Hoxa1 and Hoxb1 have equivalent protein functions in the hindbrain but the genes have sub-divided the regulatory elements accounting for the observed synergy and differences in phenotypes between the two genes (Tvrdik and Capecchi, 2006). It is not clear whether their functions in other tissues or regions of the CNS are equivalent.

1.3.2 Group 2 paralogs

The *Hoxa2*^{-/-} null mutation displays a relatively minor set of defects in the hindbrain (Gavalas et al., 1997). Ganglionic neural crest cells are normal but mesoectodermal neural crest cells of second pharyngeal arch are dramatically altered (Gendron-Maguire et al., 1993; Rijli et al., 1993). This is interpreted as an anteriorization phenotype due to transformation (though incomplete) of second pharyngeal arch into a first pharyngeal arch fate. Consistent with this there are first pharyngeal arch derivatives in the second branchial arch. Gain-of-function of *Hoxa2* in a number of vertebrate models displays the opposite transformation (Grammatopoulos et al., 2000; Hunter and Prince, 2002; Pasqualetti et al., 2000). It is interesting to note that in morpholino experiments in zebra fish, to obtain the same phenotypes observed in mice both *Hoxa2* and *Hoxb2* genes must be interfered with indicating both have a common role in neural crest patterning (Hunter and Prince, 2002). More recent studies in the mouse with compound deletions of the HoxA and HoxB clusters have also uncovered synergy between *Hoxa2* and *Hoxb2* in patterning cranial structures (Minoux et al., 2009; Vieux-Rochas et al., 2013). Together, these studies have led to a hypothesis that Hox paralogy group 2 genes (*Hoxa2* and *Hoxb2*) function as selector genes for second branchial arch identity and patterning of the vertebrate jaw (Hunter and Prince, 2002; Minoux et al., 2009; Santagati et al., 2005; Santagati and Rijli, 2003; Vieux-Rochas et al., 2013).

Hoxb2 null mutants exhibit facial paralysis and closure defects of the ventral thoracic wall, leading to sternum defects. *Hoxa2* and *Hoxb2* cooperate in regulating D-V patterns of neural differentiation in the hindbrain (Davenne et al., 1999). The *Hoxb2*^{-/-} phenotype is similar to a mixture of *Hoxb1* and *Hoxb4* null phenotypes. Interestingly, the facial paralysis shown by *Hoxb2* mutant mice is due to failure to form the somatic motor component of the VIIth (facial) nerve, similar to *Hoxb1* mutant embryos. This may arise because *Hoxb2* and *Hoxa2* are direct targets of *Hoxb1* (Maconochie et al., 1997; Tümpel et al., 2007) and they also feedback onto later aspects of *Hoxb1* expression during neurogenesis (Gavalas et al., 2003; Pattyn et al., 2003). Hence it is consistent with the known regulatory relationships between these genes. The sternum defect in *Hoxb2* mutants is similar to *Hoxb4* null mutant phenotypes (Barrow and Capecchi, 1996; Ramirez-Solis et al., 1993).

1.3.3 Group 3 paralogs

The hindbrain is largely unaffected in single *Hoxa3* or *Hoxd3* mutant embryos. *Hoxa3* null mice show complete lack of r6 and r7 derived NCC and defects in pharyngeal endoderm (Manley and Capecchi, 1995; Manley and Capecchi, 1998; Manley and Capecchi, 1997). *Hoxd3* mutants show no effect in NCC derivatives, but display skeletal abnormalities with defects in morphogenesis of the atlas (Manley and Capecchi, 1997). In *Hoxa3/Hoxd3* double mutants there is increased severity of *Hoxd3* phenotype, which indicates partial redundancy of paralogous group 3 genes in axial segmentation (Chisaka and Capecchi, 1991; Condie and Capecchi, 1994; Manley and Capecchi, 1995). In addition, in the double mutants, somatic motor neurons in r5 are completely missing and a new population of V2 neurons appears in that position (Gaufo et al., 2003; Gaufo et al., 2004). Ectopic expression of *Hoxa3* in the chick hindbrain also causes altered specification of somatic motor neurons (Guidato et al., 2003a). These data are consistent with a shared role for Hox paralogy group 3 genes in patterning somatic motor neurons. Intriguingly, it has been shown that in the *Hoxa3/Hoxb3* double mutant embryos *Hoxb1* is ectopically activated in the r6 region (Gaufo et al., 2003). This suggests that the group 3 Hox genes may act to repress *Hoxb1* activity. This is supported by a regulatory study, indicating that *Hoxb3* can bind and repress *Hoxb1* via the auto-regulatory response element *Hoxb1* and *Hoxa1* use to stimulate activity (Wong et al., 2011). This illustrates that Hox proteins can serve as both activators and repressors via the same *in vivo* Hox response elements.

1.3.4 Group 4 paralogs

Paralogous group 4 genes are important for development of cervical vertebra but there are minimal perturbations of the nervous system. Loss of *Hoxb4* and *Hoxd4* alter the expression of the *RAR-Beta* gene and there is evidence for a direct cross-regulation of these genes to set up feedback loops in modulation of retinoid signaling in the nervous system (Serpente et al., 2005b). There is also newly emerging unpublished data that *Hoxb4* and *Hoxd4*, are required for the proper formation of the r6/r7 boundary through non-cell autonomous roles in regulation of cell segregation and subsequent boundary cell specializations (Prin et al unpublished; Alex Gould personal communication).

Loss of three paralogous 4 genes namely *Hoxa4*, *Hoxb4* and *Hoxd4* leads to anteriorization and all cervical vertebra are transformed into more anterior structures resembling the first (atlas) and second (axis) cervical vertebra (Horan et al., 1995a; Horan et al., 1995b). Individual mutant genes show milder transformations. Loss of *Hoxa4* shows

partial anterior transformation where C3 is transformed to a C2 like characters and C7 resembles a thoracic vertebra with ribs (Horan et al., 1994). In *Hoxb4* mutants there is a transformation of C2 to C1 like character and split sternum (Ramirez-Solis et al., 1993). The *Hoxd4* mutant phenotype partially resembles that seen in *Hoxa4* and *Hoxb4* mutants as C2 is transformed to a C1 fate, as in case of *Hoxb4*, and C7 is transformed to a thoracic vertebra with ribs, as in case of *Hoxa4* (Horan et al., 1995b). C1, C2 and C3 also show broadening of neural arch. These results indicate that there are some distinct functional roles for each of the Hox4 paralogs but that co-operatively or functional overlaps among analogous group4 genes plays a key role in axial patterning (Horan et al., 1995b). This may indicate that relative levels of expression of these genes are important, such that phenotypes appear when levels drop below a threshold.

Hoxc4 is different than other group4 genes. In *Hoxc4* transformations are seen in thoracic vertebra, instead of cervical vertebra. Loss of *Hoxc4* leads to transformation to anterior segment identity in thoracic region. T11, T10, T8, T3 and T2 are transformed into more anterior segments and attain T10, T9, T7, T2 and T1 identities. The *Hoxc4* mutant also shows phenotypes in gut, as the lumen of esophagus has a partial or complete blockage. This indicates that some paralog group 4 genes can have unique roles in defining segmental identities in unrelated segments (Boulet and Capecchi, 1996; Saegusa et al., 1996).

Summarizing the various loss-of-function mutants of anterior Hox genes illustrates that overlapping and unique functions in combination with cross-regulation between the Hox genes themselves makes a definitive interpretation of Hox gene function in hindbrain development a challenging task. In *Drosophila*, Hox genes function as segment identity or selector genes and other genes (Gap and Pair-Rule) regulate the formation of segments. Loss-of-function mutants show clear correlations between the expression domain and altered identity of the respective segments in developing embryos. In contrast, in the mouse hindbrain, *Hoxa1* can function as a segmentation gene essential for the formation of r5 and also displays roles in the regulation of r4 identity. Hence, Hox genes regulate multiple aspects of the segmentation process in vertebrates. One of the challenges in interpreting phenotypes in mice or other vertebrates where a Hox gene has been disrupted is that it may exert its functional activity in multiple tissues or stages which are necessary to coordinate the proper formation and patterning of structures. When processes are altered the result can be abnormalities which are often not easily interpreted as transformations. Many authors describe what appear to be anteriorization phenotypes or suppression of certain traits and consider this to be evidence for homeotic transformation. However,

caution must be exercised in speculating about changes based on analogy to roles in *Drosophila*. It seems clear that the vertebrate Hox genes regulate morphogenesis of tissues and cells which can impact the ultimate character or identity of a structure. However, the process of duplication and divergence of Hox genes coupled with differences in the nature of embryonic development provide a means whereby Hox genes may have evolved additional functions in regulating developmental processes in the evolution of vertebrates. It will be interesting and exciting to search for these putative roles.

1.3.5 Axial skeleton and other Hox paralog groups

The other major system in which Hox genes show properties of segmental identity genes is in patterning the axial skeleton as noted above for the group 4 paralogs. The axial skeleton is derived from paraxial mesoderm and lateral plate mesoderm. Skeletal structures are formed by fusion of serially homologous structures called somites arranged either side of neural tube. Somitic sclerotome differentiate and undergoes segmentation through fusion of posterior half of one somite with anterior half of next somite. Sternum arises from fusion of lateral plate mesoderm at midline. Both these structures are independently targeted by Hox genes during segmentation.

In the anterior regions the Hox genes in paralogous groups 1-4 pattern the most anterior skeletal components as noted above. In more posterior regions, paralogous group (PG) 6 genes are important for specification of cervical /thoracic region boundary (Wellik, 2009). The anterior expression boundary of PG6 genes are correlated to the region of cervical /thoracic transition. Loss-of-function of paralogous group 6 does not show a complete loss of ribs. Instead ribs are small and show abnormal rib morphology. Only first rib is missing in these animals and ribs associated with vertebra 2 to 4 shows distal fusion (Vinagre et al., 2010). However, ectopic expression of *Hox6* genes results in the formation on ribs on every vertebral segment, indicating that they are able to specify a thoracic fate which in turn develops ribs. In contrast, ectopic expression of group 10 genes represses thoracic fates and represses rib formation (Guerreiro et al., 2013; Vinagre et al., 2010). Compound mutants of PG10 Hox genes lead to transformation of lumbosacral region into thoracic like character and gain of associated rib like structure. Ectopic expression of *Hox10* genes in paraxial mesoderm lead to complete loss of rib (Carapuco et al., 2005; McIntyre et al., 2007; Wellik and Capecchi, 2003). *Hox13* genes are involved in axial extension and specification of axial skeleton. Ectopic expression of *Hox13* genes under control of *cdx2* promoter leads to premature termination of axial skeleton (Young et al., 2009).

These gain- and loss-of-function type experiments have led to the proposal that posterior Hox genes have roles in defining the general fields of vertebral components, in regulating segmental identity of vertebrae and in regulating axial elongation required to generate the tissues that will contribute to vertebrae (Guerreiro et al., 2013; Mallo et al., 2010; Vinagre et al., 2010; Wellik, 2007; Wellik, 2009; Wellik and Capecchi, 2003; Young et al., 2009).

PG5 and PG6 compound mutants lead to severe phenotypes, indicating cooperatively among in rib morphogenesis. Interestingly, the anterior expression boundary of PG5 genes matches that of PG6. Apart from this, loss of PG5 genes has more profound effects, including transformation of C3 to C7 into C2 axis morphology. Non-allelic non-complementation suggests that two or more non-paralogous groups may be involved in common pathway and their relative levels of expression is important is correct patterning (Burke et al., 1995; McIntyre et al., 2007; Young et al., 2009).

The PG9 genes are primarily involved in regional patterning and control the production of sternal versus floating ribs. Complete loss of PG9 leads to 13-14 attached ribs and complete loss of floating ribs. Expression pattern of these genes are mainly seen in lateral mesoderm instead of somites which is consistent with the loss of function mutant (McIntyre et al., 2007). *Hox11* genes are involved in specification of sacral vertebra. Loss-of-function mutants show complete absence of these structures. Ectopic expression of *Hox11* genes in presomitic mesoderm leads to fusion of adjacent vertebra and resembles sacral vertebra and called as “sacralization”. (Carapuco et al., 2005; Wellik and Capecchi, 2003)

1.3.6 Limb development and Hox paralogs

Hox genes play important roles in development of appendages (Andrey et al., 2013; Gonzalez et al., 2007; Woltering and Duboule, 2010; Zakany and Duboule, 2007). Limb development is highly regulated by spatial and temporally co-linear expression of HoxA and HoxD genes in both developing fore and hind limbs (Tarchini and Duboule, 2006; Tarchini et al., 2005; Zakany et al., 2007). Though HoxC genes are also expressed in developing hind limb they are not expressed in co-linear fashion.

Hoxd13 mutant mice exhibit truncation of most of the metacarpal and metatarsal bones in fore limb and hind limb (Dollé et al., 1993). In contrast, *Hoxa9* or *Hoxd9* mutants show growth and morphogenesis defects in humerus but no detectable effects on hind limb elements (Fromental-Ramain et al., 1996). Similar to *Hox9* genes, *Hoxa11*, *Hoxd11* and *Hoxd12* mutants show phenotypic defects in forelimb (Davis and Capecchi, 1994; Davis

and Capecchi, 1996; Favier et al., 1995; Small and Potter, 1993). The lack of hind limb phenotypes is interpreted as a redundant function of HoxC genes in hind limb which is absent in fore limb during development.

Trans-heterozygous mutant phenotypes indicate that posterior HoxD genes, namely *Hoxd11*, *Hoxd12* and *Hoxd13* act on the same developmental pathway in a coordinated manner during limb development and 5' genes have more influence than 3' genes. Shortening of radius and ulna in *Hoxa11/Hoxd11* mutants and in *Hoxa10/Hoxd11* embryos indicates redundancy and coordination among paralogous genes and between adjacent genes in region where they are co-expressed (Davis and Capecchi, 1996; Wellik, 2009). This is similar to that observed in hindbrain development and axial patterning.

Based on this extensive series of published studies, in general the most identifiable homeotic transformations associated with loss-of-function of Hox genes are anterior in nature. Transformed segments attain the morphology or characteristics of more anterior segments. Furthermore these changes in morphology are in good agreement with the established domains of anterior expression boundaries of the respective individual Hox genes. This indicates that it is at the anterior expression boundary where Hox genes exert the greatest effects on modulating regional changes. This suggests that posterior Hox genes exert a putative repressive mechanism known as “posterior prevalence” to restrict anterior gene function. This is a key yet poorly understood property involved in generating the “Hox Code” underlying regional specification. However, the loss-of-function studies described above indicate that functional dominance of posterior genes over anterior genes is not absolute. The relative levels or concentration of individual HOX proteins in a given segment is important. Hence, models of posterior prevalence also need to integrate “quantitative” features. Hox genes clearly function as selector genes in vertebrates but they have a wide variety of additional roles in morphogenesis and patterning. Compound mutations of paralogous group genes generally indicate functional compensation and synergism between genes in same paralogous group.

1.4 DNA binding properties of HOX proteins and roles of co-factors in specificity

This section of the thesis reviews current status of understanding about factors determining DNA binding specificity of Hox proteins, This section aims to understand DNA binding specificity from the perspective of DNA sequence, Hox proteins and Cofactors. From above discussion it is clear that many paralogous Hox genes display an ability to functionally compensate for each other (redundancy) there is also strong evidence for distinct functions (Wellik, 2007). This implies that many HOX proteins might

have common targets and function in similar pathways. However, it is unclear to what extent these unique functional roles relate to differences in domains of expression between the genes or to subtle variations in the HOX proteins themselves and their down-stream target loci.

An important question arises from this experimental data. Do different HOX proteins regulate common versus different targets? What is the nature of their specificity? Mice contain 39 Hox genes, many have overlapping expression patterns and their loss-of-function mutations partially phenocopy each other. In light of their common origin, HOX proteins have very similar homeodomains and overall structures which bind to simple sequences with relatively similar sequence preferences (Berger et al., 2008). Hence, their individual specificity for target sites *in vivo* is likely to be modulated through the involvement of co-factors or interacting proteins. Evidence to support this comes from studies of two major classes of HOX: co-factors, PBC and MEIS which indicate they have an ancient role in potentiating the binding specificity of HOX proteins (Hudry et al., 2012; Slattery et al., 2011).

PBC and MEIS are classes of TALE-homeodomain containing proteins that can have Hox-dependent and Hox-independent functions (Longobardi et al., 2013; Penkov et al., 2013). The most commonly characterized PBC proteins in mice are the Pbx family and the most studied member of the MEIS class are the MEIS and PREP proteins (Penkov et al., 2013). Many of the HOX binding sites or Hox response elements identified *in vivo* at gene enhancers and promoters are found in association with adjacent PBX and/or MEIS/PREP binding sites (Fig.1-6) (Manzanares et al., 2001). Mutating the PBX and MEIS binding sites in down-stream Hox targets (which include Hox genes themselves) can prevent expression (Ferretti et al., 2000; Gould et al., 1997; Maconochie et al., 1997; Manzanares et al., 2001).

The hexapeptide region adjacent to the homeodomain of HOX proteins has been found to be an important site of interaction for the PBC DNA-binding partners and targeted mutations in the hexapeptide domains of HOX proteins prevents them from binding to PBX (Chan et al., 1996; Hudry et al., 2012; Medina-Martinez and Ramirez-Solis, 2003). Alterations of the hexapeptide domain of *Hoxb8* result in dominant homeotic transformations similar to that observed in *Hox7* and *Hox9* null mice, however, the expression of these genes was unaffected. This suggests that in the absence of Hoxb8/Pbx binding, target genes of other HOX proteins were being abnormally activated.

Together, these studies highlight the importance of co-factors, such as PBX and MEIS, in guiding HOX proteins to their appropriate and tissue-specific targets essential for

their *in vivo* functions. However, they also illustrate critical gaps in understanding the underlying mechanisms controlling differential HOX specificity *in vivo*. This raises several important issues which I will summarize:

1. What role do co-factors play in determining specificity of binding?
2. Are distinct *cis*-sequences a key to differential HOX specificity? In other words, how do tethered binding modules and their underlying binding sites dictate rules for specific binding of different HOX proteins?
3. How do differences in amino acid sequence of HOX proteins contribute to their binding and functional specificity?

These issues are in general highly relevant not only to HOX proteins but many other families of transcription factors. Hence, the general rules and knowledge gained from analysis of HOX proteins should be relevant for other types of factors or provide a useful basis for comparison. Most transcription factors are members of larger protein families. In mice there are 19 T-box transcription factors and all of them recognize 5'-TCACACC-3' motif for binding. The 39 mammalian HOX proteins recognize AT-rich binding sites, while the ~60 basic helix loop helix (bHLH) proteins recognize 5'-CACGTG-3' or E-box (Berger et al., 2008; Conlon et al., 2001; Jones, 2004). In *Drosophila* more than 50 homeodomain containing proteins bind to a six base pair core binding sequence as 5-TAATTG-3' and 5'-AATTA-3' (Noyes et al., 2008) (Fig.1-6). All invertebrate and vertebrate HOX proteins show binding affinity to such AT-rich sequences in *in vitro* monomeric binding assays. This issue becomes more complex because there are differences in *in vitro* and *in vivo* properties of these proteins. In *Drosophila*, Antennapedia (Antp) recognizes 5'-[C/T][C/A] ATTA-3' and binds DNA with high affinity while labial and Pb bind to 5'-nTGATTGATnnn-3'. Dfd and Scr prefer 5'-TGATTAATnn-3'. Such sequences are abundant in genome and one such sequence can be found in every 500bp pairs on a genome-wide basis. In *in vitro* binding studies, Ubx and Antp binding properties are indistinguishable. *In vivo* Antp is involved in control of leg versus antenna choice while Ubx is involved in haltere vs wing choice. Does the underlying binding sequence have a role in binding specificity that results in these distinct functional roles? Do co-factors modify binding properties of HOX proteins and lead to specific interactions masked in the *in vitro* binding assays? One of the reasons for flexibility in binding may be attributed to Glutamine residue at 50th position in homeodomain. Its flexible nature allows it to make contact with different bases at the 5' end of 5'-ATTA-3' (Billeter et al., 1996; Gehring et al., 1994a) whereas homeodomains with a different amino acid at 50th position show altered preference for these bases (Treisman et al., 1989). Occasionally, lower

affinity non-consensus sequences can interact with other regions of homeodomain to modulate binding interactions (Austin and Biggin, 1995; TenHarmsel and Biggin, 1995).

Biggin and McGinnis proposed two models to explain HOX-cofactor interactions on DNA. They called these the “co-selective model” and the “widespread binding model” respectively. In the widespread binding model, HOX proteins bind to HOX-response elements without the aid of co-factors. Many of these binding sites might not relay a functional output. However, co-factor binding could alter the ability of HOX proteins to regulate target genes. In this case, the HOX proteins are switched to an active mode from a neutral or repressed state. Once activated, they could serve to stimulate transcription depending upon their context and other recruited proteins. Evidence to support this model comes from the fact that EXD/PBX proteins are required along with HOX proteins (Dfd) for activation of target genes. However, repression by Dfd does not have a requirement for EXD (Biggin and McGinnis, 1997; Peifer and Wieschaus, 1990; Pinsonneault et al., 1997).

Additional evidence comes from structures that do not appear to depend upon cofactors for Hox control of development of the Haltere. Hox regulated structures, especially distal appendages in arthropods, including *Drosophila* and in vertebrates, do not need *Pbx/Exd* or *hth/Meis* for their proper development (Casares and Mann, 2000; DiMartino et al., 2001; Gonzalez-Crespo et al., 1998; Gonzalez-Crespo and Morata, 1995, 1996). Galant and coworkers have shown that individual Ubx binding sites in *sal cis*-regulatory region define overall strength of repression in an EXD independent manner (Galant and Carroll, 2002). This raises the possibility that HOX proteins can regulate their targets through monomeric binding without using co-factors like EXD/PBX or HTH/MEIS. However, these studies do not rule out the possibility that another co-factor(s), apart from EXD/PBX or HTH/MEIS, may be involved in this process (Galant et al., 2002). In another example, *Antp* contains 41 Ubx binding sites in the P2 *cis*-regulatory element and all of them are required for repressive function (Appel and Sakonju, 1993). It appears that HOX proteins can use a series of weak binding sites in an additive manner to achieve binding required for regulatory activity. In such cases, multiple sites may increase the overall strength of binding through cooperatively or increase the chances of occupancy. Galant and colleagues reported that as few as three sites are sufficient to complete repression in target genes (Galant et al., 2002). These examples form the basic experimental support idea the widespread binding model.

In the “co-selective model”, HOX proteins do not significantly bind with high affinity to any response element without the aid of co-factors such as EXD/PBX and *hth/Meis*. This model employs a bipartite site for HOX-binding and *Exd/Pbx* binding, such

that specificity is imparted by the composite or HOX-PBX Bipartite site (Fig.1-6). Binding of PBX to the motif may be required for HOX binding to the other half. These sites are also known to contain PBX-HOX-MEIS ternary complexes (Ferretti et al., 2005; Ferretti et al., 2000; Ferretti et al., 1999). Strong support for this model has arisen from regulatory studies in *Drosophila* and vertebrates. The identification and characterization of HOX-response elements associated with auto-, para- and cross-regulatory interactions between Hox genes have revealed that bipartite HOX-PBX sites are commonly used for HOX binding and functional activities (Alexander et al., 2009; Gould et al., 1998; Mann and Chan, 1996; Mann et al., 2009; Manzanares et al., 2001; Slattery et al., 2011; Tümpel et al., 2009). The prevalence of these bipartite elements in so many HOX-response elements in different species and the deep utilization of the PBX and MEIS TALE proteins as co-factors with HOX proteins in bilaterians (Hudry et al., 2012; Hudry et al., 2011; Slattery et al., 2011) provide strong support for the co-selective model (Biggin and McGinnis, 1997). A few important examples will be discussed in detail in a later section.

1.4.1 Hox co-factors

EXD was first TALE protein to be identified as HOX co-factor. Mutation in *Exd* causes homeotic transformation in *Drosophila* without altering Hox expression (Peifer and Wieschaus, 1990; Rauskolb et al., 1993; Rauskolb et al., 1995). This leads to a suggestion that *Exd* worked in parallel with Hox genes but later was clearly identified in some cases to also have a role as co-factor essential for specificity. Biochemical studies using a limited number of HOX response elements suggest that HOX protein specificity can be defined through TALE-HOX interactions. The major evidence of PBX as a co-factors come from two separate lines of experiments. One, most of the known HOX response elements identified to be important *in vivo* have HOX-PBX bipartite sites and these have been demonstrated to be important for function (Chan et al., 1994a; Maconochie et al., 1997; Manzanares et al., 2001; Pöpperl et al., 1995; Tümpel et al., 2007). This aspect is discussed in detail in later part of this discussion. A second line of evidence comes from mutant analysis in mice and zebra fish. *Pbx* mutants in mice and zebra fish can recapitulates Hox loss-of-function mutant phenotype in the hindbrain and other tissues (Moens and Selleri, 2006; Popperl et al., 2000; Selleri et al., 2004; Vitobello et al., 2011; Waskiewicz et al., 2001; Waskiewicz et al., 2002). Details can be seen in table 2. Furthermore, *Pbx* and *Hox* genes have been genetically shown to interaction as partial knockdown of *Hoxb1a* in *Pbx4* heterozygotes shows synergistic effects in control of motor neuron migration (Cooper et

al., 2003). Similarly paralogous group1 Hox genes show synergistic interactions with zygotic *Pbx4* in specification of rhombomeres (Waskiewicz et al., 2002).

In the vertebrate genome, multiple *Pbx* genes (namely *Pbx1*, *Pbx2*, *Pbx3* and *Pbx4*) have been identified through sequence similarity and genetic screens. Analyses in zebra fish have revealed that ectopic expression of any *Pbx* can rescue *Pbx4* mutant phenotype. Interestingly, *Drosophila Exd* is equally competent to rescue the *Pbx4* mutant phenotype in zebra fish. This indicates that various PBX proteins have a high degree of functional conservation within and across species. PBX1, PBX2 and PBX3 have identical binding properties in biochemical assays (Chang et al., 1995; Popperl et al., 2000).

The three amino acid loop extension in PBX/EXD interacts with the Tryptophan containing hexapeptide containing motif present N-terminal to homeodomain of anterior and middle Hox paralogs. In the absence of a HOX homeodomain, EXD can interact with other motifs, such as the 15 amino acids down-stream of the Pbx homeodomain. Interestingly, PBX/EXD-HOX interactions have been shown to be involved in repression as well as activation of target genes (Rauskolb and Wieschaus, 1994). Hence, PBX/EXD-HOX interactions may define binding specificity but not its functional outcome. In some contexts, mutating the hexapeptide motif activates otherwise repressed genes. This suggests that HOX co-factors may dictate functional outcomes in a context-dependent manner (Chang et al., 1995; Galant et al., 2002; Merabet et al., 2003; Neuteboom and Murre, 1997; Piper et al., 1999).

Meis is another TALE protein that serves as a HOX co-factor. Vertebrates have two Meis proteins namely MEIS1 and MEIS 2 and two other related proteins called PREP1 and PREP2. *Drosophila* has a homologous protein called HTH (Homothorax) (Rieckhof et al., 1997; Ryoo et al., 1999). There have been binding sites found for PBX-MEIS hetero-dimers adjacent to many HOX-PBX bipartite sites in HOX response elements (Ferretti et al., 2005; Ferretti et al., 2000).

It has been shown that HOX, PBX and MEIS can bind and form ternary complexes using these adjacent binding sites. It has been shown that MEIS can be immunoprecipitated with HOX and PBX. Meis can be found as component of HOX-PBX complex or can act as in absence of PBX with posterior Hox genes (Chang et al., 1997; Ferretti et al., 2000; Jacobs et al., 1999).

Meis proteins are known to be important for nuclear localization of PBX (Berthelsen et al., 1999; Waskiewicz et al., 2001). It has been postulated that controlling the nuclear-cytoplasmic distribution of EXD/PBX may be the most important function of MEIS. However, it has been shown that MEIS is also involved in stability of PBX

mediated complexes (Waskiewicz et al., 2001). It is important to mention at this point that PBX partners with large number of homeodomain proteins (Engrailed, Distaless, Pax) and other transcription factors, including MyoD, while MEIS has limited interaction partners. This has led to the idea that PBX might have many more genome-wide binding site reflecting its role as a common co-factor compared with MEIS or HOX proteins themselves.

Table 1-2 Pbx mutants in Fish and mouse recapitulate Hox loss-of-function mutant

<i>Pbx</i> mutant		Phenotype	Reference	Resembles Hox mutant	References
<i>Pbx4</i> ^{-/-}	Zebra Fish	Facial motor neuron migration defect	(Cooper et al., 2003)	<i>Hoxb1</i> ^{-/-} (Mice) <i>Hoxb1a</i> MO (Zebra Fish)	(McClintock et al., 2002; Studer et al., 1996)
		Mis-targeting of trigeminal motor axons		<i>Hoxa2</i> ^{-/-} (Mice)	(Gavalas et al., 1997)
<i>MzPbx4</i> ^{-/-} <i>Pbx2</i> MO		Absence of r4 and r5	(Waskiewicz et al., 2002)	<i>Hoxa1</i> ^{-/-} (Mice)	(Wright, 1993)
<i>Pbx1</i> ^{-/-}	Mice	2 nd branchial arch transformation	(Selleri et al., 2001)	<i>Hoxa2</i> ^{-/-} (Mice)	(Gendron-Maguire et al., 1993; Rijli et al., 1993)
		Cervical vertebral transformation		<i>Hoxa3</i> ^{-/-} , <i>Hoxd3</i> ^{-/-} , <i>Hoxa4</i> ^{-/-} , <i>Hoxa6</i> ^{-/-} (Mice)	(Condie and Capecchi, 1994; Kostic and Capecchi, 1994)
		Rib malformation		<i>Hoxa9</i> ^{-/-} (Mice), <i>Hoxb9</i> ^{-/-} (Mice)	(Chen and Capecchi, 1997)
		Proximal limb malformation		<i>Hoxa10</i> ^{-/-} (Mice)	(Favier et al., 1996)
		Impaired hematopoiesis	(DiMartino et al., 2001)	<i>Hoxa9</i> ^{-/-} (Mice)	(Lawrence et al., 1997)
		Incomplete descent and fusion of thymus	(Manley et al., 2004)	<i>Hoxa3</i> ^{-/-} , <i>Hoxb3</i> ^{-/-} , <i>Hoxd3</i> ^{-/-} (Mice)	(Manley and Capecchi, 1997)
		Spleen agenesis	(Brendolan et al., 2005)	<i>Hoxa11</i> (<i>tlx</i>) ^{-/-} (Mice)	(Dear et al., 1995; Dear et al., 1993; Roberts and Tabin, 1994)
		Pancreas hypoplasia	(Kim et al., 2002b)	<i>Pdx1</i> (<i>Ipfl</i>) ^{-/-}	(Jonsson et al., 1994)
<i>Pbx3</i> ^{-/-}		Congenital apnea	(Rhee et al., 2004)	<i>Rnx</i> ^{-/-}	(Shirasawa et al., 2000)

Adapted from (Moens and Selleri, 2006)

1.4.2 Role of Co-factors in determining specificity of HOX binding

Very few genome-wide studies have been done to investigate binding of HOX proteins in a tissue-specific or developmental context (Donaldson et al., 2012; Huang et al., 2012; Jung et al., 2010; Sorge et al., 2012). Insight into binding specificity mostly comes from analysis of individual Hox regulatory elements characterized from the Hox clusters of *Drosophila* and mice. Hence, current models of HOX specificity are derived from binding information provided by these elements and structural studies. A few well characterized elements are discussed in detail below.

The *Deformed* auto-regulatory region is a well characterized enhancer. A 5 Kb upstream region of *Dfd* in *Drosophila* Hom-C complex serves as an auto-regulatory region was identified. Four moderate- to high-affinity binding sites for DFD protein, with the two highest affinity sites sharing a 5'-ATCATTA-3' consensus sequence, were identified within a 920 bp segment. Mutation of this region led to loss of binding activity in *in vitro* assays and absence of embryonic regulatory activity in transgenic assays. This region contains a binding site for Deformed and co-factors. It seems that clustering of HOX and co-factors binding sites might be one mechanism to enforce spatial requirement and achieve specificity (Kuziora and McGinnis, 1988; Regulski et al., 1991; Zeng et al., 1994).

In the vertebrate HoxB cluster, a well characterized r4 auto-regulatory region involved in maintenance of *Hoxb1* in r4 has described (Fig.1-5) (Ferretti et al., 2005; Pöpperl et al., 1995). In a 331 bp highly conserved region there are four key sets of sites. The first block contains a 5'-TAAT-3' motif but mutation of this region alone has no effect on r4 expression or RA response. Three remaining conserved block contain Hox-Pbx bipartite sites 5'[T/A]GAT[T/A]GA[T/A]G-3'. Deletion of these conserved blocks abolishes r4 expression while mutation in first two blocks does not have any significant effect on r4 expression. Mutation in third conserved block leads to reduced reporter expression in r4. Interestingly, mutating one more conserved block with region 3 abolishes

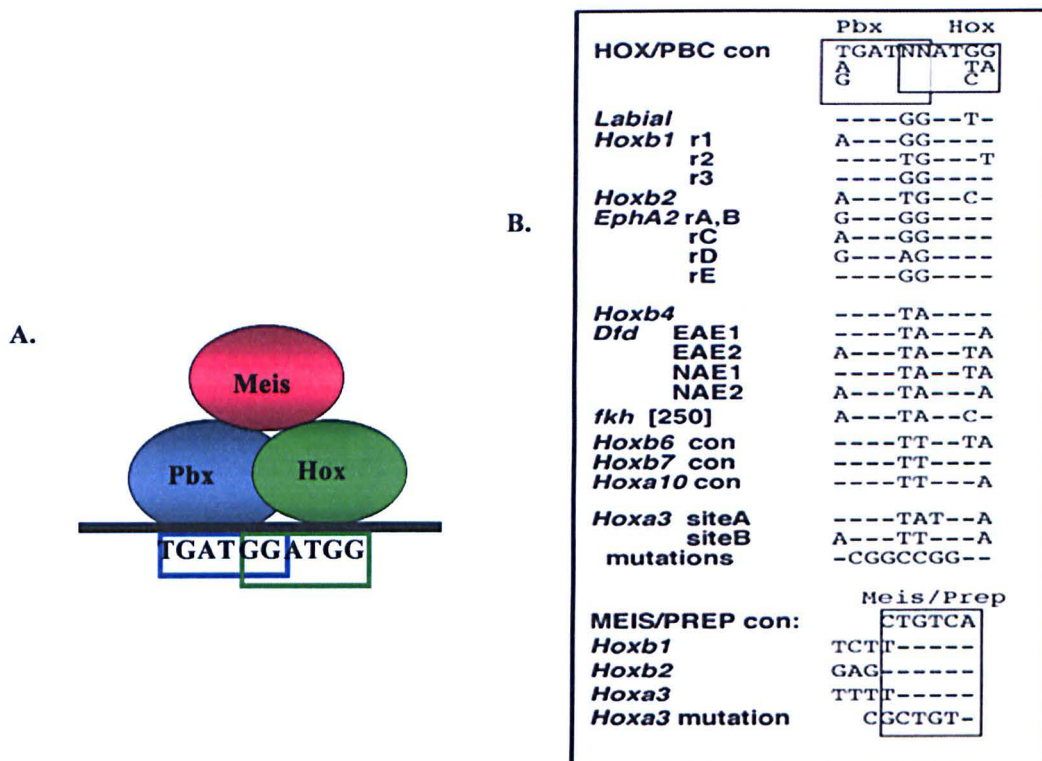


Figure 1-6 Consensus bipartite HOX/PBC sites in the enhancer and their properties (A). Formation of MEIS, PBX, HOX ternary complex on HOX-PBX Bipartite site. (B) List of characterized HOX/PBC sites and MEIS/PREP sites found in target genes aligned with those detected in *Hoxa3*. Mutated sequences used in binding and transgenic assays are indicated. Adapted from Manzanares et al Dev 128, 3595-3607(2001)

all r4 expression. Three copies of a 20bp region containing the third conserved block (R3) faithfully generate r4 expression in transgenic reporter assays. This suggests that this R3 region is sufficient to generate r4 expression pattern and plays central role in r4 auto-regulation of *Hoxb1*. It has been shown that this auto-regulation is dependent upon *labial* group of genes in when tested in mice and *Drosophila* (Pöpperl et al., 1995; Studer et al., 1998b). This interaction is through co-operative binding of HOXB1 with PBX (Marshall et al., 1994; Pöpperl et al., 1995; Pöpperl and Featherstone, 1992). Ferretti and colleagues mapped at least three prep-Meis TALE sites near the Hox-Pbx bipartite site and demonstrated that these are essential for functional activity of the enhancer (Ferretti et al., 2005). Berthelsen and coworkers have shown assembly of PBX1, PREP1 and HOXB1 trimeric complex on the R3 element of *Hoxb1* (Berthelsen et al., 1998).

The Hox-Pbx bipartite site is also an integral part of the *Hoxb1* responsive enhancers of *Hoxb1*, *Hoxb2*, *Hoxa2*, *Hoxa3* and *Hoxb4* in mice and *labial* in *Drosophila melanogaster*. Interestingly most of them also contain a combination of PBX/PREP-MEIS (PM) and PBX-HOX bipartite (PH) sites suggesting that ternary complexes of HOX-PBX and MEIS may be a common feature on HOX response elements (Ferretti et al., 2000;

Gould et al., 1997; Manzanares et al., 2002; Ryoo et al., 1999). Though the *Hoxb1*, *Hoxa2* and *Hoxb2* enhancers generate r4 restricted expression pattern in transgenic assays the specific organization and numbers of the PM and PH sites varies between them and also varies in the same gene between species. As described above *Hoxb1* contains three PH site with two nearby PM site while the *Hoxb2* enhancer has only single adjacent PM and PH site. PM and PH sites are separated by 17 and 8 bp respectively in *Hoxb1* and *Hoxb2* enhancers and form trimeric complexes *in vitro*. Complete abolition of r4 expression is observed upon mutating the PM site from *Hoxb2* enhancer (Ferretti et al., 2000) while mutation of the PM1 alone in the *Hoxb1* enhancer failed to abolish the r4 expression pattern (Fig.1-7) (Ferretti et al., 2005).

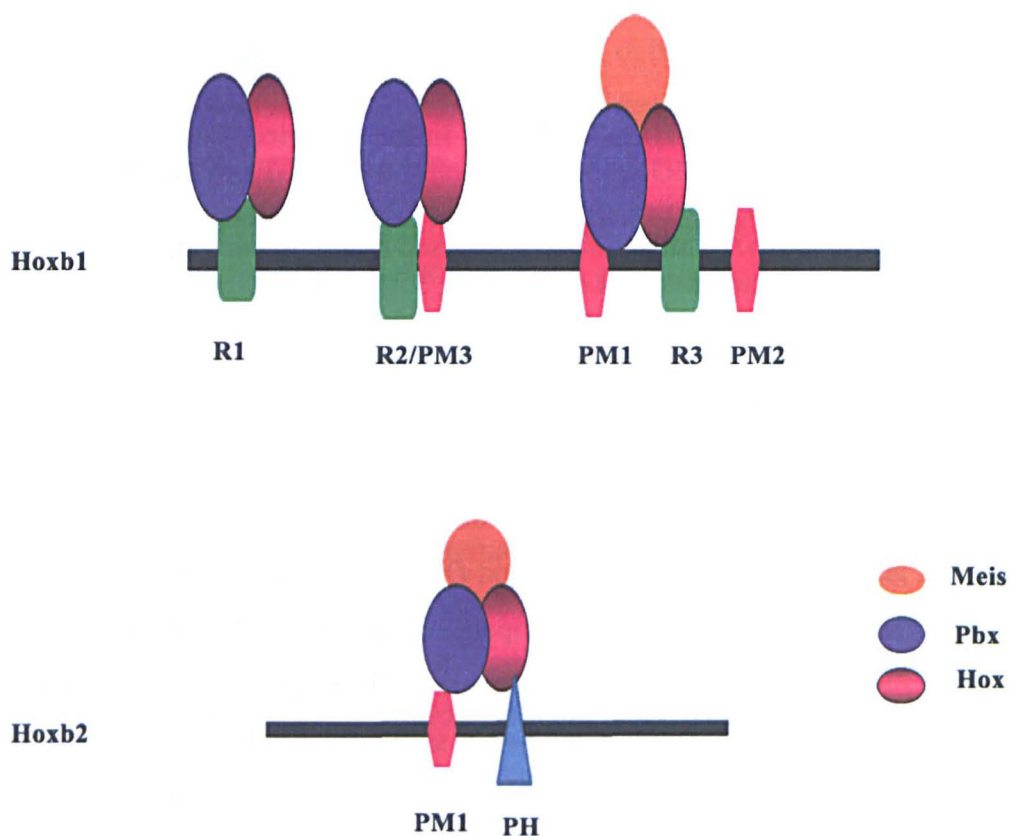


Figure 1-7 Organization of PREP MEIS –(PM) and PBX-HOX (PH) *Hoxb1* and *Hoxb2* enhancers. (A) Schematic representation of r4-*Hoxb1* and r4-*Hoxb2* enhancers including the PM and PH sites binding to the various PREP-PBX complexes.

This confusion on the role of PM sites was resolved by the discovery of a second PM site in the *Hoxb1* r4 enhancer region. Ferretti and colleagues have shown that a PM2 site from *Hoxb1* can form ternary complexes in *in vitro* assays and drive robust transgene reporter expression in chicken hindbrain (Ferretti et al., 2000). Transgene reporter assays of 622 bp regulatory region from *Hoxb1* containing all of the PM and PH sites indicates

that both PM and PH sites are important for enhancer activity in chicken and mice. Alteration of PM2 has stronger effects on reporter activities compared to PM1 but when both the PM1 and PM2 sites are altered reporter expression is completely abolished or highly reduced. Interestingly, the R2/PM3 motif in the enhancer appears to interfere with formation of ternary complex. This negative influence of R2/PM3 site may help to restrict expression of *Hoxb1* enhancer in r4. Binding of the PREP-MEIS complex may inhibit assembly of a ternary complex on R3. This arises due to steric hindrance created by occupancy of PREP-MEIS on R2PM3.

Synergy between the PM and PH sites in *Hoxb2* enhancer defines its activity. Ferretti and colleagues argued that differential affinity of PM1; PM2 and R2PM3 sites determine formation of ternary complex on R3. Increased levels of PBX1 upon RA induction may change binding affinity of PBX/PREP-MEIS complexes on PM sites through increased availability of PBX for interaction on R3. These changes may alter configurations from repression to activation. Fine balance in this mechanism is achieved through tethered PREP-MEIS sites (Ferretti et al., 2005; Ferretti et al., 2000). It has also been argued that Pbx and Meis may interact with this *Hoxb1* r4 enhancer first to open chromatin and that the availability of Hoxb1 changes the interactions to switch this to an active state (Choe et al., 2009).

Similar to R3 site, a similar site is involved in *labial* mediated auto-regulation in *Drosophila*. Just two base pair alteration in this site changes its specificity from Labial to Deformed (Chan et al., 1994a; Chan et al., 1997). Changing the central base of the bipartite Hox-Pbx site from GG to TA resulted in *deformed* or *Hoxb4* like expression pattern in mice and *Drosophila*. Interestingly, similar binding sites with GG as central base pairs can be found in the *Drosophila deformed* gene and in the mouse *Deformed* ortholog, *Hoxb-4* (Fig.1-6). These results indicate that DNA sequence play central role to HOX specificity and probably provides a platform for context-specific interaction of HOX proteins with its EXD/PBX and HTH/MEIS co-factors (Joshi et al., 2007; Mann et al., 2009; Ryoo et al., 1999; Slattery et al., 2011).

Tümpel and colleagues identified a conserved region in *Hoxa2* intron containing three bipartite (PH-1–PH3) sites and single PBX-PREP (PM1) site. This region can act as an r4 enhancer in chick electroporation assays. Ectopic expression of *Hoxb1* can *trans-activate* the *Hoxa2* enhancer in chicken embryos (Tümpel et al., 2007).

Jacob and colleagues presented analyses that the Abd co-factors PBX and MEIS may compete with each other to generate a hierarchy of heterodimers or they may cooperatively interact to make higher order DNA binding complex (Jacobs et al., 1999).

They also verified that HOXB1 binds to the *Hoxb2* enhancer along with PBX AND MEIS. MEIS appears to be important for specificity of the ternary complex binding to DNA. They argued that PBX-MEIS are able to bind DNA without a stringent requirement for the half site. The amino terminal may be enough for heterodimer formation between TALE proteins, which leaves their homeodomains free to interact with DNA in various configurations. This model would allow assembly of ternary complexes on DNA consisting of a site for HOX and PBX with a distinct flanking MEIS site. Despite the differences in models, their work further supports the idea that inclusion of MEIS in Hox-PBX interaction complexes appears to help in increasing specificity.

Manzanares and colleagues identified a *Maf b-Zip* DNA binding motif (KrA site) in the *Hoxa3* enhancer region. They demonstrate that this site is essential for Kreisler binding and early activation of *Hoxa3* in r5 and r6 (Manzanares et al., 1999). Later it was identified that this KrA site is within a longer conserved block in *Hoxa3* enhancer which was able to drive strong reporter expression in both early and late stages. Mutation of the KrA site within this enhancer abolishes early expression, suggesting that Kreisler is essential for early expression. However, the mutated *Hoxa3* enhancer is capable to direct segmental expression in late stages. Manzanares and colleagues identified two HOX-PBX bipartite sites and a PREP- MEIS binding site within this conserved block. Using *in vitro* binding assays they showed that the *Hoxa3* PBC-B site can bind HOXB3, HOXD3 and HOXA3 in increasing efficiency. Competition with the *Hoxa3*/PBC-A /B site was effective at blocking other known Hox/PBC sites from *Hoxb1* and *Hoxb2* to inhibit binding activity. Furthermore, multimerized Hox/PBC-B sites (five copies) were able to direct reporter expression in r6/7 and r5. The Hox/PBC site at *Hoxa3* differs slightly with other known Hox-Pbx bipartite sites and contains TA or TT in the center instead of GG (Fig.1-6). Comparing results from binding and transgenic assay with sequence information it appears that subtle changes in binding site can lead to changes in nature of binding through the same set of HOX proteins and co-factors to alter domains of expression in a gene specific manner (Manzanares et al., 1999).

1.4.3 HOX proteins-Determinants of its own specificity

Greer and colleagues swapped complete coding regions between *Hoxa3* and *Hoxd3*. These experiments have shown that *Hoxa3* and *d3* are functionally equivalent. This analysis supported the idea that Hox genes are functionally equivalent and different functions arise from differences in temporal and spatial domains of expression (Greer et al., 2000). This has been interpreted as an “equivalency model” suggesting that the

quantity of HOX protein is important in determining functions and not variations in the proteins themselves (Duboule, 2000).

In analogous types of experiments, Zhao and Potter reported that a homeobox swap between *Hoxa11* and *Hoxa13*, generating chimeric *Hoxa11* (A1113hd) alleles, results in mice that develop normally and give rise to normal skeletons, kidneys and male reproductive tracts. While in limb and the female reproductive tract development, *Hoxa11*^{13hd} was acting as a dominant-negative allele. The uterus was transformed into cervix/vagina. Interestingly, the *Hoxa13* expression domain overlaps these structures in female reproductive tract (Zhao and Potter, 2001).

Swapping the homeodomain of *Hoxa10* into *Hoxa11* (*Hoxa11*^{10hd}), leads to hypomorphic phenotype in appendicular skeleton, kidney and reproductive tracts but show no defects in development of axial skeleton. Swapping homeodomain of *Hoxa4* into *Hoxa11* (*Hoxa11*^{4hd}) generates animal similar to *Hoxa11* null phenotype. Surprisingly *Hoxa11*^{4hd} mice have normal axial skeletons (Zhao and Potter, 2002). These domain swap experiments imply tissue-independent and tissue-specific roles of different homeodomains. It further indicates that defining segmental identity may have been a common or primitive function of the homeodomain acquired before functional divergence between different paralogous groups.

All HOX proteins contain a conserved hexapeptide motif –YPWM. The importance of this motif varies in different HOX proteins and contexts (Mann et al., 2009). *Hoxa1* and *deformed* require the YPWM motif for interaction with *Exd/Pbx* (Green et al., 1998; Joshi et al., 2010). While *UBX* and *AbdA* do not require the YPWM motif, instead they interact with a distinct six amino acid motif (Known as *UbdA*) C-terminal to homeodomain (Galant et al., 2002; Merabet et al., 2003; Merabet et al., 2007; Saadaoui et al., 2011; Tour et al., 2005). Furthermore, *Ubx* and *AbdA* have conserved C-terminal residues important for their *in vivo* function (Chan et al., 1994b).

Phalen and Featherstone have shown that the N-terminal arm residues are important for monomeric and heterodimeric binding specificity of HOX proteins (Phelan and Featherstone, 1997). They confirmed that this position is contacted by the HOX N-terminal arm and differs between HOX proteins. Though, it is important to mention that N-terminal difference has moderate to no effect on binding affinity *in vitro*. In the case of HOXA1, Arg5 in N-terminal arm makes contact with minor groove and the hexapeptide plays an important role in interaction with *Pbx*. In HOXD9 and HOXD10 the hexapeptide has diverged from the consensus. Unlike HOXA1 and HOXD4 proteins, HOXD9 and HOXD10 bind to 5'-TTAT-3' and 5'-TAAT-3' motifs in monomeric binding assays and

can bind to 5'-TTAT-3' in heterodimeric binding. The residues responsible for heterodimeric binding can be mapped to Lys-3, Lys-6 and Lys-7. This adds support for the idea that N-terminal residues can alter specificity of monomeric and heterodimeric binding of HOX proteins from different paralogous groups. One more intriguing aspect of this study comes from the surprising observation that the R3-*labial* binding site is not the site with highest heterodimeric binding affinity. This raises a question regarding the relative balance between specificity and affinity. Site selection in an *in vivo* context may be more weighted towards specificity rather than to affinity (Phelan and Featherstone, 1997). Neuteboom and Murre selected similar affinity binding sites for Hoxc6, Hoxb7 and Hoxb8 and Hox-Pbx-1 using the PCR selection method. This analysis provided further support for the idea that specificity is more weighted versus affinity (Neuteboom and Murre, 1997).

Lelli and colleagues reported that a Tryptophan motif is important for interaction with Exd. In case of AbdA an additional Tryptophan containing motif (TDWM) and C-terminal sequences are important for specific interaction with *Exd*-dependent targets. Furthermore, Ubx is not depended on C or N terminal region for context-specific regulation of target genes. Altering the UbdA domain severely affects the binding property of the homeodomain in monomeric or heteromeric (HOX-PBX) binding assays. Lelli and colleagues argued that presence of the extra Tryptophan containing motif in posterior HOX proteins may be basis of posterior dominance of posterior genes over anterior genes (Lelli et al., 2011). Interestingly many proteins, other than HOX proteins, such as Engrailed and MyoD, also use Tryptophan to interact with EXD and PBX. In the case of Scr a single Tryptophan residue in YPWM motif is sufficient for interaction with EXD-dependent function (Knoepfler et al., 1999; Peltenburg and Murre, 1996).

Slattery and coworkers demonstrated that HOX proteins acquire novel binding specificity when they bind together with co-factors. HOX proteins and co-factor association yields a new specificity and binding site recognition, which is not shown by either of them in monomeric binding. In other words, though HOX proteins may show analogous monomeric binding properties on similar AT-rich sequences, their "latent specificity" is unlocked by co-factor association (Slattery et al., 2011). They defined "latent specificity" as "*differences in the amino acid sequences of transcription factors within the same structural family may only impact DNA recognition when these factors bind with co-factors.*" They argued that this mechanism is distinct from co-operativity, where binding kinetics is important and co-factor association interferes with binding energetics (Slattery et al., 2011)

The source of latent specificity may be the N terminal and linker sequences of HOX proteins. Binding of Exd limits freedom of the YPWM or hexapeptide region of HOX proteins. In the case of the Scr homeodomain, two Arginines at 3rd and 5th positions are located to the minor groove by YPWM and Exd interactions. An adjacent Glycine to 3rd position Arg, which is unique to paralogous group 2 Hox genes, is required for this interaction. An RQR motif with Arginine in 3rd position is a unique feature of group 2 proteins and may favor a conformation allowing insertion of both Args into the minor groove (Joshi et al., 2007). In the case of class 3b genes (namely *Ubx*, *AbdA* and *AbdB*), the homeodomain contains an Arginine at the 2nd position. Crystal structures of HOXA9-PBX complexes reveal this Arginine makes contacts with the minor groove through water mediated hydrogen bonds (LaRonde-LeBlanc and Wolberger, 2003; Mann et al., 2009). This suggests that small changes in one or more amino acids in homeodomain are utilized by cofactors to modulate DNA binding specificity.

All preferred binding sequences form a narrow groove and Arg5 makes contact or is located near to this region based on available crystal structures (Rohs et al., 2009). Interestingly minor groove topologies show distinct features based on interactions with HOX proteins. Anterior HOX proteins (group 1 and group 2) prefer narrow minor grooves while posterior proteins achieve specificity through wider minor groove. Different HOX proteins bind to distinct DNA sequences but seem to have a similar overall DNA topology/structure (Slattery et al., 2011). Slattery and coworkers argued that the presence of TpR motif tends to widen the minor groove in middle of the binding site to accommodate Arg3 and Arg5 while TpA in group 3 proteins prevents insertion of Arg3.

The crystal structure of human HOXB1-PBX1-DNA ternary complex shows Hoxb1 and Pbx1 bind to opposite faces of DNA. HOXB1 also makes contact with PBX using the hexapeptide. The hexapeptide makes contact in a small pocket generated by the three residue insertion in helix 3 and a part of C-terminal region of PBX (Piper et al., 1999). Jabet and colleagues have shown that a disorganized C-terminal region of Pbx-1 undergoes formation of a fourth helix upon binding to DNA. Binding of the Pbx1 homeodomain to DNA constrains flexibility of C-terminus and forces it to acquire a helical R
-PBX interaction and completes the hexapeptide pocket favoring cooperative interaction with HOX proteins (Jabet et al., 1999). PBX and MEIS co-factors as well as HOX proteins themselves contribute in concert to modulating binding specificity. Small variations in choice of amino acid in the hexapeptide or homeodomain in turn lead to recognition of slightly different DNA sequences. In this context, DNA sequence per se seems to be less

important than topology generated by a specific sequence. HOX proteins might bind to different DNA sequences if they are capable of generating a similar overall topology. Many binding and structural studies seem consistent with this possibility. It is also true that the TALE co-factors play a very important role in determining functional output of HOX binding. These co-factors determine stability of binding complex, recruitment of co-activators and co-repressors to affect the final outcome of binding.

1.5 Down-stream targets of Hox proteins

This Section of the introduction deals with current understanding of downstream targets regulated by various Hox genes. This section highlights genes identified as regulated by Hox genes through various studies. This section also deals with some interesting recent findings related to identification of genome-wide binding sites for various Hox proteins through combination of ChIP and next generation sequencing.

Many characterized down-stream targets of Hox proteins are Hox genes themselves. HOX proteins regulate their own loci and other Hox genes through auto- and cross-regulatory mechanisms (Fig.1-5 and Fig.1-8).

Apart from Hox genes, other transcription factors are known to be targets of HOX proteins. Guazzi and coworkers have shown that *Otx2* gene expression is controlled by RA through cross-regulatory mechanism. The 5'flanking region from the *Otx2* promoter responds to anterior HOX proteins (Hoxb1, Hoxb2 and Hoxb3) but not to posterior HOX proteins (Hoxc6 and Hoxd8) (Guazzi et al., 1998). Theokli and coworkers identified *Irx5* as a Hoxb4 downstream target. Interestingly, *Hoxb4* and *Irx5* show over-lapping expression patterns in developing mice and *Xenopus* CNS (Theokli et al., 2003). *GATA3* and *GATA2* are Hoxb1 targets. *Hoxb1* null mice show loss of *GATA2* and *GATA3* expression in the ventral region of r4. Interestingly, *GATA3* null mice phenocopy aspects

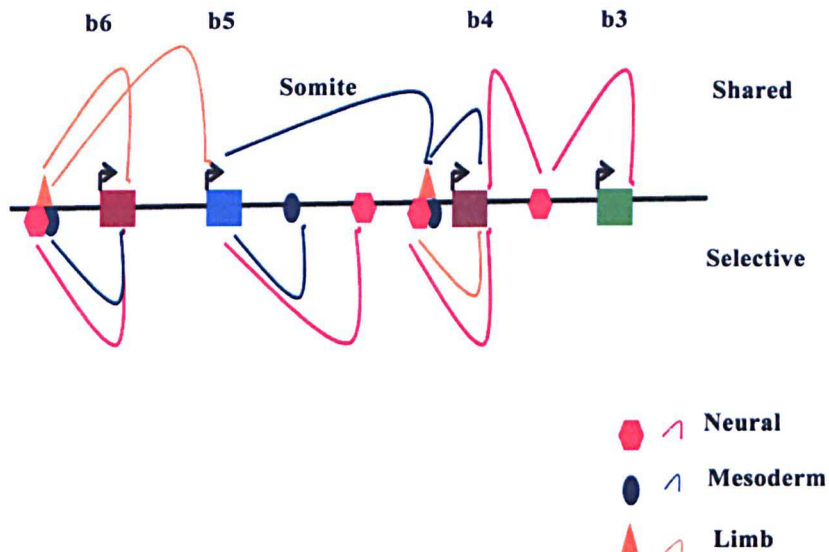


Figure 1-8 Model summarizing the different regulatory interactions between HoxB genes The different shapes and arrow marks the position of the enhancers and colored looped arrows above the complex note shared interactions of these enhancers between promoters, while those below the complex mark enhancer/promoter interactions that are selective. Modified from Sharpe et al The EMBO J. 17(6), 1788-1798 (1998)

of the *Hoxb1* mutant phenotype (Pata et al., 1999). Human *Norepinephrine transporter gene (hNET)* has been described as *Hoxa5* target. In *in vitro* studies, *Hoxa5* seems to recognize a HBS sequence within *hNET* promoter. *In vitro* binding and transfection studies indicate that these HOX binding sequences (HBS) play critical role in expression of *hNET* (Kim et al., 2002a).

HOX proteins regulate many genes related to adhesion, signaling and repulsion. Among them some of the interesting candidates include *N-CAM*, *R-cad*, *APP*, *Osteopontin*, *Ephrin*, and *Pcp-2*. The *N-CAM* promoter contains HOX response sequences and integrates positive and negative input from various HOX proteins. Interestingly, when these elements are assayed out of context from their promoter, they fail to induce reporter activity. *Hoxb9* and *Hoxc6* act as activators on these sequences while *Hoxb8* represses activity of a linked reporter (Holst et al., 1994; Jones et al., 1993; Jones et al., 1992; Jones, 2004; Wang et al., 1996). *Cad-6* shows overlapping expression pattern with *Hoxa1*. In transgenic assays, the *Hoxa1* mutant shows rhombomeric and stage specific defects in *Cad6* expression (Inoue et al., 1997). Furthermore, through differential hybridization Shen and coworkers identified *Cad-6* as downstream target of *Hoxa1* (Shen et al., 2000). Another cell adhesion molecule, β -amyloid precursor protein (*APP*) gene, is repressed by *Hoxc8*. In transfection studies, a transgene reporter with *APP* promoter comes under repressive effects of *Hoxc8*. It is important to note that *APP* promoter region contains at least 20 uncharacterized HOX

binding sequences (Violette et al., 1992). *Osteopontin* is considered to be part of *Hoxc8* and *Hoxc9* gene regulatory network. *Hoxc8* and *Hoxc9* repress *osteopontin* in progenitor osteoblast cells. A HOX binding sequence in the promoter is believed to mediate this response and Smad signaling overcomes this repressive effect during differentiation (Shi et al., 2001; Shi et al., 1999). An uncharacterized tumor suppressor gene *mgl-1* is under transcriptional control of *Hoxc8*. *mgl-1* and *hoxc8* have complimentary expression pattern suggesting a repressive relationship (Tomotsune et al., 1993).

Signaling molecules are also known as targets of Hox genes. One of well-studied example is the *Ephrin* genes. Group 1 genes namely *Hoxa1* and *Hoxb1* show overlapping expression with *EphA2* in primitive streak during gastrulation. In later stages of development, *EphA2* and *Hoxb1* show restricted expression in r4. HOX-PBX bipartite sites have been characterized in *EphA2* r4- enhancer and have shown to be trans-activated by HOXA1/HOXB1-PBX heterodimers (Chen and Ruley, 1998). Compound mutation of *Hoxa1* and *Hoxb1* show reduced *EphA2* expression which further strengthens the idea *EphA2* is a downstream target of HOX proteins. *Hoxa2* null mice show lack of expression of *Eph7* in r3 and alteration in other rhombomeres (Chen and Ruley, 1998; Gavalas et al., 1998; Studer et al., 1998a; Taneja et al., 1996).

Purkinje cell specific (*Pcp*) is a signaling molecule regulated by HOX proteins. The proximal promoter of the *Pcp* gene contains a short regulatory region with HOX binding consensus sequences. Homeodomain domain proteins are capable of binding to this sequence and in cell culture assays. HOXA5 and HOXB7 are capable of transactivating reporter gene under regulation of this sequence. Furthermore, Engrailed seems to exert repressive action on *Pcp* gene through this element (Baader et al., 1998; Sanlioglu et al., 1998). Morsi El-Kadi and coworkers identified *XRap1* as downstream target of *Hoxb4* using differential display. In *in vivo* studies, it has been shown that *Hoxb4* binds to regulatory region at 3' UTR of *XRap1* to mediate transcriptional repression (Morsi El-Kadi et al., 2002). A protein inhibitor named Serpin SPI3 was characterized as downstream target of *Hoxb5*. These genes show overlapping expression in developing mice embryo. *In vitro* studies revealed that *SPI3* promoter contains *Hoxb5* binding regulatory region (Kato et al., 2001).

Using microarray analyses, Zhao and Potter identified many keratin genes, transcription factors (*Six1*, *HFH-1* and *Elf-5*), desmosomal adhering proteins (*plakophilin 1* and *Desmocolin 2*) and other genes (*Ask2*, *DHCR7*, *p73H*, *connexin*, *Galectin*, *Pgp-1*, *Scca2* and *Fv1*) as downstream targets of *Hoxa13* (Zhao and Potter, 2001). Williams and colleagues identified 68 genes as downstream target of *Hoxa13* in mouse embryonic

fibroblast ectopically expressing *Hoxa13*. Out of 68 genes, 50 genes were up regulated while 18 genes were down regulated. Go Term related analysis revealed extracellular matrix and basement membrane genes were highly enriched. Up-regulated genes included membrane/cytoskeleton proteins, metabolic enzymes, secreted/extra cellular matrix and signal transduction/growth control. Down regulated genes listed membrane/cytoskeleton proteins, secreted/extra cellular matrix and signal transduction / growth control (Williams et al., 2005).

Donaldson and colleagues in their genome-wide binding studies identified canonical *Wnt* signaling as target for *Hoxa2* by identifying binding sites of *Hoxa2* near genes involved in *Wnt*-signaling pathway. They have shown that expression *fzd4* and *Wnt-β-catenin* is lost in *Hoxa2* mutant (Donaldson et al., 2012). Gene ontology (GO) analysis showed enrichment of terms related to skeletal system morphogenesis, mesenchymal and epithelial cell proliferation and middle ear and artery development along with *Wnt* signaling receptors. This is consistent with the role for *Hoxa2* in patterning neural crest derived structures in the head. In this genome-wide study, Hox and Hox-Pbx motifs displayed an enrichment in *Hoxa2* binding peaks (Donaldson et al., 2012).

Identification of *Hoxb4* downstream targets in a cell line model of primitive hematopoietic progenitor cells by Lee and coworker identified 465 genes as putative direct or indirect targets (Lee et al., 2010). Chip-seq analysis on these cell lines revealed occupancy of *Hoxb4* at 1910 promoter regions including *CD34*, *Sox4*, *Gp49a*, *Laptm4b* and *B220*. Go term analysis of putative target genes shows enrichment of terms related to hematopoiesis, cell adhesion and immune phagocytosis. Out of 465 peaks, at least 71 genes have *Hoxb4* occupancy in their promoter. The highest ranked genes with *Hoxb4* occupancy in their promoter includes *Rab38*, *Clec4e*, *rpl3*, *zfp521*, *tm2d2*, *fcgr2b*, *als2*, *Mfge8*, *scap2*, *rab19* (Lee et al., 2010).

Makki and Capecchi identified *Hoxa1* downstream targets through comparing gene expression in micro-dissected samples from prospective r3-r5 in wild type and mutant embryos (Makki and Capecchi, 2011). They identified 137 down regulated and 162 up regulated genes. Go term analysis of differentially expressed genes have shown enrichment of terms related to development of embryonic organ, hindbrain, inner ear, vasculature, hematopoietic or lymphoid organ and cardiac muscle tissues. Go term analysis also shows enrichment of terms related to differentiation of neurons and muscles, cell migration, regulators of apoptosis, retinol metabolism, *Wnt* and *Tgf beta* signaling. Some interesting downstream candidates are *Dfna5*, *Foxd3*, *Lhx5*, *Sema3c*, *Hnf1b*, *spry4*, *Fzd8*, *wnt10b*,

tbx15, pax8, Zic1, Hoxd3, apob, Clic5, lefty1, nodal, Hnf4a etc (Makki and Capecchi, 2011).

Rohrschneider and coworkers identified multiple downstream target for zebra fish homolog of *Hoxb1*; *Hoxb1a*. They identified 216 ESTs differentially expressed in mutant zebra fish r4. Out of 216 ESTs, 113 were down regulated while 103 were up regulated in mutant. Genes identified in this screen were broadly classified as transcription-translation related, cytoskeletal and cell adhesion matrix, growth factor or receptors and signaling molecules. Some targets identified (which are also reported by others) are *evi1, fabp7a, eaat2, ca2, calret, col7a11, pou4f2, Zic4, rar α , Zic1* and *Meis2*. They further identified *prickle1b* as downstream target of *Hoxb1a* and shown its requirement in migration of facial neuron (Rohrschneider et al., 2007).

Lei and coworkers identified 34 genes regulated by *Hoxc8* in *C57BL/6J* mouse embryonic fibroblast over-expressing *Hoxc8*. Among 34 genes 18 genes were up regulated and 16 were down regulated. Genes related to cell adhesion, migration, metabolism, apoptosis and tumor genesis were considerably enriched. Genes like *Gas1, Zac1, CARP, Serpinf1, NAD(P)H* and *Nqo1* involved in apoptosis were down-regulated upon *Hoxc8* over-expression. The cell adhesion molecule NCAM was down-regulated while *Cdh11* and *Emb* were up-regulated. *Osteopontin* was down regulated in these cells (Lei et al., 2005).

McCabe and colleagues identified transcriptional network for *Hoxc6* using prostate cancer cell line. They identified 468 genome-wide binding sites for *Hoxc6*. Further comparing with expression profile of prostate from *Hoxc6* deficient mice lead to identification of 31 directly regulated target genes. They reported that many developmental ligands and receptors such as *Igfbp3, Bmp7, Runx1, Fgfr2* and *Pdgfra* are under direct control of *Hoxc6*. Several tumor suppressers, *Fgfr2, Cd44* Wnt antagonists- *Wif, Dkk3, Sfrp1* and *Sfrp2*, are activated by *Hoxc6* (McCabe et al., 2008).

Hedlund and coworkers identified 69 differentially expressing genes required for development and patterning of lumbar spinal cord which are dependent upon *Hoxd10*. Genes identified in this screen are involved in cell adhesion, cell cycle regulation, and some act as transcription factor. A large proportion of these genes are involved in cell-cell communication and oncogenesis. Many genes identified in this screen contain HOX response elements, HOX/PBX1 motifs or HOX/PBC motifs, near their promoter region (Hedlund et al., 2004).

Pavlopoulos and Akam (2011) identified 872 targets for *Ubx* in *Drosophila*. These genes were related to signaling, cuticularogenesis, veins and inter-veins, margin and bristles, growth and patterning, adhesion and morphogenesis and cytoskeleton. De Navas

and colleagues identified multiple classes of Ubx targets. The first class, such as *dpp* in visceral mesoderm and genes responsible for embryonic cuticle, respond equally to both isoforms of Ubx namely Ia and Iva. The second class of targets, *sal*, *wg* and *ara*, in the haltare disc is repressed more by Ia than Iva. Dpp in posterior visceral mesoderm is differentially controlled by Ia and Iva (de Navas et al., 2011). Agrawal and coworkers found 519 genome-wide binding sites for Ubx in their genome-wide binding studies. Nearest neighbor genes to this binding site were mainly enriched for various signaling pathways, transcription factors, cell adhesion, cytoskeleton, and cuticle formation. Interestingly, no Ubx consensus motif was seem to be enriched in these sequences. Surprisingly, GAGA factor, MAD, Adf-1, Grainyhead, DREF, Ovo, snail binding motifs were over represented motifs in these binding regions. Similarly, Choo and coworkers did not find enrichment of HOX/PBX or UBX/EXD motifs in their genome-wide UBX binding data. They did find Pho, Brk and Dref enriched binding motifs. In Pavlopoulos and Akam's data set and in genome-wide binding assays by Choo and coworkers, Ubx appears to regulate distinct genes at separate time point (Choo et al., 2011; Pavlopoulos and Akam, 2011). This indicates that an equivalency model of HOX function is an overly simplistic interpretation of HOX function. HOX proteins bind to variety of targets and show distinct targets evident from genome-wide binding data. Some common pathways like cell adhesion, cell migration and signaling (*Wnt* and *Tgf* beta) are often shared by various Hox genes. Meanwhile genome-wide expression analysis does indicate whether each HOX protein has a distinct or overlapping set of target genes. Many of these studies do not distinguish between direct and indirect effects.

This illustrates the need to systematically understand and compare HOX protein binding properties on a genome-wide basis. There is the question of whether common targets are regulated through binding on a common DNA sequence and whether differential affinity and/or co-factors determine HOX protein selection preferences. Furthermore, while HOX-PBC bipartite sites were thought of as primary binding sites from analyses of known auto and cross-regulatory elements this seems to be a rather over simplification to explain emerging binding properties of various HOX proteins. Genome-wide binding studies indicate that recruitment of HOX complexes on DNA fragment must be a much more complex process. Other transcription factors, co-activators and co-repressors may recruit HOX proteins directly or indirectly to DNA sites and chromatin structures. Hence, going forward it will be important to understand binding properties of a given HOX protein in comparison to paralogous and non-paralogous HOX proteins and cofactors. This requires a model system were expression can be qualitatively and/or

quantitatively controlled and the expression and binding properties of a given HOX proteins is amenable on a genome-wide basis under related experimental conditions.

1.6 Thesis problem statement

Regulatory proteins including transcription factors directly or indirectly bind DNA and facilitate regulation of gene expression in a temporal and spatial manner. These factors either recruit other proteins or are recruited by other protein factors for such regulatory activity. The collective knowledge of sequential binding of these factors and its effectors to DNA, together constitute a basis for formulating a Hox gene regulatory network. Though many aspects of Hox gene expression and their developmental functions are known, very little is known about genomic sequences where they are recruited as HOX protein complexes and what cofactors are associated with binding and regulatory activity.

One challenge is that the 39 HOX proteins have very similar structures, so their individual specificity is likely to be modulated by subtle differences in cofactors, interacting proteins, target sites or other processes. Therefore, identifying *in vivo* relevant cofactors, binding partners, co-activators or co-repressors and binding sites is essential for understanding their functional roles. Apart from such protein factors; recent advances in gene expression profiling suggests a post-transcriptional role of non-coding transcripts along Hox-clusters in regulation of Hox expression along AP axis. Addressing these poorly understood processes is important if we are to generate a better mechanistic understanding of the Hox gene regulatory network and protein interaction network associated with their conserved role in regulating morphogenesis.

Interestingly DNA binding specificity of Hox genes are a function of three independent factors namely DNA sequence (motifs), Co-factors and Hox genes. Loss of function studies on cofactors have convincingly proven that loss of cofactors alone can generate homeotic phenotype without altering Hox gene levels. This illustrates that Hox genes and its co-factors work in a parallel pathway in determining segmental identity of developing embryos. Limited studies on individual genomic elements have shown that cofactors like PBX and MEIS are important factors involved in site recognition and recruitment of Hox genes. Further tethered site of these cofactors helps to strengthen these interactions and provide another layer of regulatory control. But it will be important to know how cofactors like Pbx and Meis can influence genome-wide binding properties of various Hox genes? Are Pbx and Meis the only cofactors involved in determining DNA binding specificity of Hox genes? Do Hox genes have absolute requirement of Co-factor for DNA binding? Does Hox cofactors diversity explain DNA binding specificity and

diversity? There is undeniable evidence of importance of underlying DNA sequences and its role in specific recruitment of Hox proteins through Cofactors like PBX and Meis. (Details in 1.4). Small changes in nucleotide composition leads to loss of function in transgenic context and decreased binding affinity as studied through biochemical approaches. In this context it is important to understand that what kind of sequences are more often associated with binding of Hox-proteins? How these sequences differ among various Hox proteins? Do we see tethered Co-factor binding site near these binding regions? What role these tethered sites play in recruitment and functionality of Hox proteins?

Losses of function studies have shown that interaction between and among Paralogous Hox genes. Some function has been shown as redundant among Hox genes while others as unique function of individual Hox genes. This raises some interesting question regarding binding properties of Hox genes. Do Hox genes uses two different motif and cofactors for unique and redundant function in targets conferring redundant function ? Does redundant function of Hox genes requires co-occupancy of two Hox genes? Does binding of one Hox gene have functional consequence to binding of other Hox gene on same or different site?

These wide ranging questions were the focus of my thesis research .As a beginning to understand these wide ranging questions, we decided to generate an understanding of the binding properties of anterior HOX proteins in the group 1 paralogs (HOXA1 and HOXB1) along with their down-stream targets, and co-regulators. This will further help to understand and elucidate role of HOX proteins and their regulators along with its down-stream targets in the development of hindbrain, segmentation and patterning along AP axis. We decided to use ChIP and next generation sequencing technology to understand genome-wide binding properties of these proteins. We have used ES cell based system to get better control over the tissue type/ cell identity for this purpose. ES Cells are handy to compare and contrast binding properties of these proteins in cells with various cellular identity related to ectoderm, endoderm and mesoderm due to availability of robust protocol for differentiation into distinct cell types. To achieve above mentioned objectives, I set out this project with two main aims as given below.

1.6.1 Aim 1: Study activation of Hox clusters in neuro-differentiation using an ES cell model system

- a) Characterize the properties of neuro-differentiation process

- b) Analyze accessibility and epigenetic properties of Hox clusters during differentiation.
- c) Analyze binding of retinoid acid receptors (RARs & RXRs) to study their roles in *cis*-regulation of Hox genes during differentiation.

1.6.2 Aim 2: Genome-wide identification of HOX-response elements

- a) Identify Hoxb1 target sites in genome.
- b) Identify target sites in genome for some other HOX proteins.
- c) Functionally validate a sample of the potential HOX response elements in transient transgenic mouse and chick embryos and analysis in mutant embryos.

Chapter 2 Methods

2.1 Induction of KH2 cells with retinoic acid

KH2 cells at Passage 12 were cultured on gamma irradiated feeder cells with DMEM containing 15% fetal bovine serum, NAA and β -mercaptoethanol. Media was changed 3 hours before passaging them. Media was aspirated and washed twice with PBS. 2 ml of pre warmed Trypsin/EDTA solution were added and placed in incubator at 37⁰ C for 1 minute. During this period colonies float off when flicking the plate. Trypsin activity was stopped by adding 5ml of FCS-ES medium to flask. Colonies were dissociated into single cells by pipetting up and down for several times and pelleting the cells by centrifugation at 1000 rpm for 5 minutes. Media was aspirated and the cells were resuspended in appropriate volume of fresh ES cell medium. Cells treated for 12, 16 and 24 hours with RA were supplemented with differentiation media (DMEM + 10% Serum + NAA+ 0.03 μ M RA) while other time points 2 hours, 4 hours, 6 hours and 8 hours were changed to differentiation media next day morning. For all experiments, going beyond 16 hours media was changed after every 16 hours.

After required period of induction, Media was aspirated and washed twice with PBS. 2 ml of pre warmed Trypsin/EDTA solution were added and placed in incubator at 37⁰ C for 1 minute. During this, period colonies float off when flicking the plate. Trypsin activity was stopped by adding 5ml of FCS-ES medium to flask. Colonies were dissociated into single cells by pipetting up and down for several times. Additional 15 ml of Differentiation media (without RA) was added and plated into freshly gelatinized plates. For gelatinized plates, Culture plates were treated half an hour before seeding with 0.1% Gelatin. Gelatin was later aspirated just before seeding of cells. After half an hour media was aspirated and centrifuged for 5 min at 1000 rpm. Pellets were dislodged by gentle tapping and 2 ml trizol were added. These tubes were stored at -80 until RNA isolation.

2.2 RNA preparation for Affymetrix microarray analysis

RNA isolation was performed using Trizol and later purified by RNA Easy Kit (Quiagen). RNA was tested for integrity and concentration using the RNA 6000 Nano Assay and RNA LabChips on the Agilent Bioanalyzer 2100 (Agilent Technologies, Inc., Palo Alto, CA). Based on RNA Integrity Numbers (RIN) and ribosomal ratios, samples were chosen to proceed with RNA amplification starting with 200 ng total RNA. Labeled

cRNA targets were generated from total RNA samples using the MessageAmp III RNA Amplification Kit (Applied Biosystems / Ambion, Austin, TX.) according to the corresponding instruction manual. Aliquots of labeled cRNA were assessed using the RNA 6000 Nano Assay and RNA LabChips on the Agilent bioanalyzer 2100 (Agilent Technologies, Inc., Palo Alto, CA). Biotinylated and fragmented cRNA targets (15 µg) were hybridized to Mouse Genome 2.0 arrays using the GeneChip Fluidics Station 450 according to the manufacturer's standard protocol. Arrays were scanned with a GeneChip Scanner 3000 7G and the image data on each individual microarray chip was scaled to 150 target intensity, using the GeneChip Command Console Software (AGCC software v.1.1) (Affymetrix, Santa Clara, CA).

2.3 Analysis of Affymetrix data

Samples were hybridized to Affymetrix Mouse 430 v2 arrays. Three biological replicates were done for all samples. Microarray data was analyzed in R (2.11.1) using the *affy* (1.26.1) and *limma* (3.4.3) packages. Normalization was done using RMA. Annotation information was taken from Bioconductor annotation package *mouse4302.db* (2.4.1).

K-means clustering was done in R (2.13.2) with $k = 9$. K was selected using partition clustering in Partek 6.4 (which uses the Davies-Boudin index to identify optimal values for k). Only clusters with absolute value of cluster mean $> .5$ are shown. Enriched Biological Process GO terms for each cluster were found using R package *GeneAnswers* (1.8.0) with a hypergeometric test. GO term annotations were from the bioconductor package *org.Mm.eg.db* (2.5.0).

Hierarchical clustered heat map was generated from the $\log_2(\text{fold change})$ of the top 5000 probe sets from the Affymetrix expression data based on standard deviation of fold change. Data was clustered using hierarchical clustering using the *heat map* function in R (2.13.2). Columns were not clustered. \log_2 fold changes exceeding -6 to 6 were set to the scale limits for image rendering. Image was generated using the *image* function in R. Row names of the heat map were selected as follows - for genes with one unique probe set in the heat map, only the gene symbol is shown. For genes with more than one probe set, the probe set id is also given in parentheses. For probe sets without an annotated symbol, only the probe set id is shown.

2.4 Agilent tilling Array

2.4.1 Design of custom made tilling array

The probes were designed as 60mers overlapping by 20 bases, on both strands. Probes matching more than one place in the genome with 100% identity (aligned with Blat) were removed; So on average there is one probe every 22 bases. The sequences from which these probe sets were derived were selected from GenBank®, dbEST, and RefSeq. The sequence clusters were created from the UniGene database (Build 107, June 2002) and then refined by analysis and comparison with the publicly available draft assembly of the mouse genome from the Whitehead Institute for Genome Research (MGSC, April 2002). Design number 027887 "rek1" (2x105k agilent array).

Hox covered regions are (mouse genome version mm9)

HoxA chr6: 51960548-52269383

HoxB chr11: 95994533-96346113

HoxC chr15: 102626926-102985721

HoxD chr2: 74351893-74665608

This is also equivalent to:

A: 1000 bases into *Skap2* to 1000 bases past *Evx1*

B: 1000 bases past *Till6* to 1000 bases into *Skap1*

C: 1000 bases past *AK043982* to 1000 bases into *Smug1*

D: 1000 bases past *Lnp* to 1000 bases into *Mtx2*

2.4.2 Library preparation and hybridization

RNA isolation was performed using Trizol and later purified by RNA Easy Kit (Quiagen). Total RNA in the amount of 1ug was amplified according to Ambion's Message Amp II aRNA Amplification Kit, part number AM1751. Positive control RNA from Agilent's One Color RNA Spike-In Kit, part number 5188-5282, was used to monitor sample amplification and labeling as well as array hybridization. Amplified mRNA, referred to as aRNA, was quantified on a NanoDrop ND-1000 and a mass of 2ug was labeled with cy3 dye using Kreatech's ULS Fluorescent Labeling Kit, part number EA-023. Labeled cRNA was quantified on the NanoDrop ND-1000 and a 1.5ug mass of cy3 labeled cRNA was hybridized to custom Agilent 2x105K HOX tiling arrays.

Hybridizations were performed at 65C for 17 hours under standard conditions (1X Agilent blocking agent, and 1X Agilent hybridization buffer) and slides were washed successively

with Agilent GE wash buffer 1, at room temperature and with Agilent GE wash buffer 2, at 31C, prior to scanning. Microarray images were acquired with an Agilent High-Resolution DNA Microarray Scanner (G2505C). Hybridization, array washing, scanning and probe information extraction with Agilent's Feature Extraction Software (Version 10.5.1.1) were all performed according to Agilent's One-Color Microarray-Based Gene Expression Analysis Protocol, Version 6.0, December 2009 (Low Input Quick Amp Labeling).

2.4.3 Analysis of Expression

Agilent tiling arrays were hybridized in a single-color configuration and data was read into R. Agilent "gMeanSignal" was used as the measurement (mean green signal for each spot). Data was analyzed using the limma package. Data was normalized between arrays using scale normalization. Replicates were averaged and bedgraph files were created and visualized using IGV.

2.5 Directional mRNAseq library preparation and sequencing

RNA isolation was performed using Trizol and later purified by RNA Easy Kit (Quiagen). Quality was RNA was analyzed by Agilent 2100 Bioanalyzer. RNA with more than 9 RIN (RNA integrity number) was used in library preparation. Library generation for Directional mRNA-Seq was performed using the Small RNA Sample Prep Kit (Illumina, cat# FC-102-1010) with 10x v1.5 sRNA 3' Adaptor (Illumina, cat# 15000263) and mRNA-Seq Library Prep Kit (Illumina, cat# RS-100-0801) according to manufacturer's Directional mRNA-Seq Sample Prep protocol (part #15018460 Rev A Oct 10).

Briefly, 1 µg Total RNA was enriched for poly(A)+ RNA by oligo-(dT) selection. The Poly(A)+ RNA were then fragmented, and the ends repaired using phosphatase and PNK treatments. Illumina's v1.5 sRNA 3' adaptor was ligated to the blunted RNA segment, followed by the ligation of the kit's standard SRA 5' adaptor. The libraries were reverse transcribed and enriched by 15 rounds of PCR. The purified libraries were quantified with the High Sensitivity DNA assay on an Agilent 2100 Bioanalyzer. Libraries were sequenced single read with 36 nt sequencing on a GAIIx and fastq files were returned.

2.6 Analysis of RNA sequencing data

For each sample, reads were aligned to mm9 using Tophat 1.4.1 and the Ensembl 65 GTF, allowing uniquely-aligning reads only (parameters: **-g 1 -x 1 --segment-length 20**

--segment-mismatches 2). Initial BAM files were split into separate BAM files for + and – strand alignments. Strand-wise BAM files were quantitated with Cufflinks 1.3.0 and no GTF (parameters: **--max-mle-iterations 10000 --max-bundle-frags 100000000 -F 0.05 -a 0.05 --trim-3-dropoff-frac 0.01**). For each time point and strand, ab-initio cufflinks transcripts were condensed into loci using Cuff-compare 1.3.0, default parameters, using the above GTF file as the reference. Any such *ab-initio* loci which were detectable in only one technical replicate were removed. Genomic coordinates for the four hox regions were extended out to their proximal flanking genes, resulting in the following regions of interest:

chr6	51809165	52268372	HoxA
chr11	95995100	96620791	HoxB
chr15	102537210	102988383	HoxC
chr2	74491049	74716488	HoxD

All *ab-initio* loci from these regions were extracted and displayed as a heat map. Loci were arranged in genomic order. For each time point, heat map values are log₂ of the average FPKM of the Cuff compare transcripts assigned to that locus.

2.7 Chromatin Immuno-precipitation

ChIP was done according to Upstate protocol with certain modification. Cells were fixed by adding formaldehyde to media at a final concentration of 1% and by incubating at 37°C for 11 min. Crosslinking was stopped by 1mL 1.25 M glycine to each 10 mL and by incubating at room temperature for 5 min. Cell were washed twice with Ice cold PBS and collected by scraping. Cells were washed twice with ice cold PBS and snap frozen in liquid nitrogen. Washing onwards in each step Protease inhibitors were added. Cells were suspended in lysis buffer containing 1% SDS and incubated for 10 min in ice. Cells were sonicated for 25 min in bioruptor at high setting and 30 sec on-off cycle. Chromatin was pre cleared and incubated with antibody-bead complex for overnight. Beads were washed five times with different buffers and crossed linking was reversed by incubating at 65 °C for overnight. DNA was precipitated and quantified by Pico green assay.

Antibodies used

H3K4Me3	(Ab1012)	Abcam
H3K27Me3	(Ab6002)	Abcam
Pol II	(Sc889)	Santa cruz biotechnology inc.

RAR alpha	(Ab28767)	Abcam
RAR beta	(Ab53161)	Abcam
RAR gamma	(Ab97569)	Abcam
RXR alpha	(Sc-553X)	Santa cruz biotechnology inc
NcoR	(Ab24552)	Abcam
Suz12	(Ab12073)	Abcam

2.7.1 Library preparation and hybridization for ChIP on Chip

Input DNA in the amount of 100ng and IP DNA in the amount of 10ng were amplified and labeled according to the Agilent Genomic DNA Labeling Kit PLUS (Agilent part number 5188-5309, Agilent publication number G4481-90010). Custom Agilent 2x105K HOX tiling arrays were hybridized with a mixture of 4ug Cy3 labeled DNA and 4ug Cy5 labeled DNA probes. Hybridizations were performed at 65C for 24 hours under standard conditions (45 mg/mL Human Cot-1 DNA, 1X Agilent blocking agent, and 1X Agilent hybridization buffer) and slides were washed successively with Agilent ChIP wash buffer 1, at room temperature and then Agilent ChIP wash buffer 2, at 31C, prior to scanning. Microarray images were acquired with an Agilent High-Resolution DNA Microarray Scanner (G2505C). For image analysis Agilent Feature Extraction software (Version 10.5.1.1) was used.

2.7.2 Analysis of ChIP on ChIP

Agilent tiling arrays were hybridized in a two-color configuration and data was read into R (2.11.1). Data was analyzed using the limma (3.4.3) package. Data was normalized within arrays using loess normalization.

2.7.3 ChIP-Sequencing

2.7.3.1 Library preparation for ChIP-Sequencing

Following manufacturer's directions and starting at end repair step, short fragment libraries were made with 10ng of DNA per sample using the Illumina TruSeq library construction kit (Illumina, Cat. No. RS-122-2002). Adapters were diluted 1:3 in order to accommodate the lower starting amounts. The resulting libraries were purified using Agencourt AMPure XP system (Beckman Coulter, Cat. No. A63880), then quantified using a Bioanalyzer (Agilent Technologies) and a Qubit Fluorometer (Life Technologies). All libraries were pooled, re-quantified and run as high output mode 50 bp single-end lanes

on an Illumina HiSeq 2500 instrument, using HiSeq Control Software 2.0.5 and Real-Time Analysis (RTA) version 1.17.20.0. Secondary Analysis were done with CASAVA-1.8.2 was run to demultiplex reads and generate FASTQ files.

2.7.3.2 Analysis of ChIP Sequencing

Sequences were aligned using bowtie v0.12.8 with “-m 1 -v 2”, mm9 genome as parameter. Since each samples set shown very high variation in read counts, hence samples were downsized. On sets with multiple IPs per input, all IP and input BAMs were randomly down sampled to match the smallest BAM. MACS2 v2 v2.0.10.20120913 were used for calling peaks in samples using “callpeak -f BAM -g mm --nomodel -p 0.05” as parameter. Peaks were filtered using following parameters: width ≥ 200 and ≤ 1000 ; height ≥ 10 ; fold-change ≥ 5 ; p-value ≤ 0.05 . This was done to generate peaks of 200-1000bp width, with 5 fold enrichment. Unified peak sets across multiple IPs were generated with Bedtools’s mergeBed utility.

2.7.3.3 Analysis for over represented motifs

Over –represented motifs were analyzed using MEME (MEME Suite, v4.8.1), Cis-finder and Transfac (2012.2). Over represented motifs were assigned to various peak using FIMO(). Motifs were identified in MEME suite using “-nmotifs 20 -dna -revcomp -mod anr -maxsize 1000000 -minw 6 -maxw 14” as parameters. Peaks sets were parsed for Identified over-enriched motifs and pre-defined k-mers with FIMO using “-p 0.001 -q 1 --max-stored-scores 1E12” as parameters. Over represented Motifs were matched with known motif definition using TomTom (-thresh 0.1 -min-overlap 4). Enrichment of Known motif definition was analyzed using Transfac 2012.2. For each peak list, a background set of 10x random sequences with the same size and chromosomal distribution was generated, and not allowed to overlap any of the original peaks. Peak list and background were converted to fasta. FIMO searched for Transfac motifs (and grep for kmers) in peaks and background. Motif hit counts were tested with Fisher’s Exact Test in R. This was done a number of times, using a range of p-value cutoffs for detected motifs, from $p=1E-3$ to $p=1E-6$.

2.7.3.4 Feature mapping

Peaks were mapped to genomic features using Bedtools’s intersectBed and Ensembl 67 gene models. Hits were classified both by gene and as exonic, intronic,

intergenic, promoter, or tail region. Promoter and tail region are defined as the windows from TSS to 10kb upstream, and from TES to 1kb downstream, respectively.

2.7.3.5 Nearest-Neighbors

Intergenic peaks were annotated with nearest neighbors and distances using Bedtools's intersectBed on a bed file of intergenes. The intergenes were generated by applying Bedtools's complementBed to the set of Ensembl 67 protein-coding genes. The resulting intergenes were annotated with the flanking genes, so that peak matches are instantly annotated with neighbor genes.

2.7.3.6 Functional analysis

For each peak list, direct-hit genes and nearest-neighbor genes were tested for functional significance using GO biological process and KEGG pathway terms. The utility used was FatiClone (Ariel Paulson), a local software package which emulates the strategy of Babelomics's FatiGO utility. Briefly, instead of testing significance for only directly-annotated terms, each term from level of the GO tree (default, levels 3-9) is tested for significance. Since some of these terms have no direct gene annotations, hits to all child terms of the pending term are also counted, to boost sample size for Fisher's Test. Peak list neighbor-genes are always compared against the corresponding background-list neighbor genes.

2.7.3.7 Heat maps

Peak lists were converted to 1kb windows by flanking the peak midpoint by 500bp in each direction. Coverage depth for these 1kb windows was extracted from the bedGraphs for each IP and input, so that signal from any sample could be analyzed for any peak list. These coverage matrices were variously arranged, hierarchically clustered, and imaged as heat maps in R.

2.8 Quantitative PCR and analysis

2.8.1 Quantitation of Hox genes using TLDA cards

400ng aliquots of each RNA sample were used as the template in 20ul total volume reactions of ABI's High Capacity cDNA Reverse Transcription Kit. These 20ul reactions were combined with 80ul ultrapure water and 100ul of ABI 2x Gene Expression Master Mix. For each of the resulting reactions, 95ul was added to 2 lanes of a 48 assay Custom

TLDA Card from ABI. The TLDA cards were spun down, sealed, and cycled on an ABI 7900HT according to ABI's standard protocol.

Analysis of the fluorescence curves was done using ABI's SDS2.3 software. All curves that showed errors as determined by the SDS2.3 software or that were above 35 Ct were thrown out. The remaining Ct values were exported and analyzed using DataAssistv2.0. DataAssistv2.0 was used to determine the most stably expressed endogenous control genes. GAPDH and TBP were used as endogenous controls.

2.8.2 Quantitation of Non-coding transcripts

5µg total RNA was used as template in 20µl total volume reaction of Stratascript II (Invitrogen) reverse transcription kit using oligo dT primers. 20ul reaction was diluted 50 times and 2µl were used for qPCR and cycled an ABI 7900HT according to ABI's standard protocol. Analysis of the fluorescence curves was done using ABI's SDS2.4 software. All curves that showed errors as determined by the SDS2.4 software or that were above 35 Ct were thrown out. The remaining Ct values were exported and analyzed using the Biogazelle qBase plus version 2.4 software have been used to analyze normalized relative quantity using assays for Atp5b and GAPDH as endogenous controls. GAPDH and ATP5b were used as endogenous controls. Each primers pairs were standardized for linear range of amplification through standard curve analysis.

2.9 RA Gavage of 10 dpc female CD-1 mice

10 dpc female CD-1 mice were injected with all-trans retinoic acid. 20 µg RA/g body weight mixed with 160ul embryo tested mineral oil were used. After 8 hours of injection, embryos were harvested. RNA was isolated my trizol method.

5µg total RNA was used as template in 50µl total volume reaction of Stratascript II (Invitrogen) reverse transcription kit using oligo dT primers. 20ul reaction was diluted 10 times and 2µl were used for qPCR and cycled an ABI 7900HT according to ABI's standard protocol. Analysis of the fluorescence curves was done using ABI's SDS2.4 software. All curves that showed errors as determined by the SDS2.4 software or that were above 35 Ct were thrown out. The remaining Ct values were exported and analyzed using the Biogazelle qBase plus version 2.4 software have been used to generate normalized relative quantities using assays for *Atp5b*, *UBC*, *CanX* and *GAPDH* as endogenous controls.

2.10 Template binding assay

5' –biotynalated DNA fragments were bound to 200 μ L streptavidin dynal beads, washed, and resuspended in a final volume of 200 μ L buffer (20mM Hepes, PH 7.9, 0.05% NP-40, 10% Glycerol, 10mM MgCl₂, 2mM DTT, 0.5mM PMSF, 100ug/ml BSA, ddh₂O (75 fmol DNA/ μ L beads). 10 ul Nuclear extract was incubated with 90 μ L of reaction buffer containing 10 ul Flag IP buffer(10mM HEPES (pH7.9), 0.1M NaCl, 1.5mM MgCl₂, 0.05% TritonX-100 w/ protease inhibitor cocktail (Sigma P8340)), 20 ul Re-constitution, 1mM ATP,50uM ZnCl₂, sheared lamda DNA 45ng/ul reaction and Nacl (varies as per reaction requirement) for 10 min at 30oC at vibrating incubator at 1200rpm.5ul of DNA-bead complex is added to reaction and incubated for 20 min at 30oC at vibrating incubator at 1200rpm.

Template bound intermediates were separated from supernatant fractions through the use of magnetic racks. Following fractionation, beads were washed with 200 μ L of reaction buffer, transferred to a fresh microcentrifuge tube, and bound proteins were eluted with 1x SDS sample buffer. Template bound intermediates were analyzed by western blot through the use of infrared antibodies and imaging (Li-COR) or using MudPIT.

2.11 MudPIT

Multidimensional protein identification technology (MudPIT) is a non-gel-based shotgun proteomic technique, which combines on-line high-resolution liquid chromatography and tandem mass spectrometry. Peptides from a complex mixture are eluted in an iterative process from a multi-phase (RP/SCX/RP) microcapillary column directly into an electrospray ionization ion trap mass spectrometer (LTQ, ThermoFinnigan). Tandem mass spectra (MS/MS), which contain fragmentation patterns specific to amino acid sequences, are generated from peptides after they elute into the mass spectrometer. MS/MS spectra are assigned to peptides found in proteins sequence database by the SEQUEST algorithm. SEQUEST results are reassembled into protein information, filtered, sorted, and compared using DTASelect/Contrast.

Chapter 3

Study of activation of Hox clusters in neuro-differentiation using an ES cell model system

Though the function of HOX proteins is critical for regulation of developmental processes very little is known about genomic sequences and down-stream target loci where they are recruited as protein complexes. One challenge is that the 39 HOX proteins have very similar structures, so their individual specificity is likely to be modulated by subtle differences in co-factors or interacting proteins. My overall goal is to identify *in vivo* relevant HOX response elements to gain insight into the sites, binding partners and mechanisms of transcriptional regulation essential for their function in hindbrain.

Understanding the DNA binding specificity of HOX proteins is an important problem. Major obstacles in undertaking such studies are: lack of sufficient tissues for biochemical and genomic studies; unavailability of antibodies against HOX proteins, and need for the development of genomics and computational approaches to generate and analyze genome-wide binding data. Three major breakthroughs in this regard make it possible to enable such studies for my thesis. The advancement of Next Generation sequencing made it possible to sequence from the whole genome at considerable depth at reasonable costs. Further, standardization of chromatin immune precipitation (ChIP) assays lead to an explosion of such studies. Availability of defined embryonic tissues for biochemical studies and lack of antibodies were still major obstacles. The recent advancements in stem cell biology have provided a new set of reagents to study this problem through directed differentiation of ES cells. Combining the directed differentiation of ES cells with ChIP and Next gen sequencing (ChIP-seq) is now an effective tool for understanding binding specificity of HOX proteins in mice. The availability of programs for an advanced informatics pipeline facilitates the ability to handle large data sets to systematically study combinatorial binding properties and its output in terms of gene expression.

Several studies have demonstrated that teratocarcinoma and embryonic stem cells can be induced to differentiate upon treatment with retinoic acid (RA) into neuroectodermal fate. F9 EC cells differentiate upon treatment with RA (Strickland and Mahdavi, 1978; Strickland and Sawey, 1980). RA alone induces differentiation of F9 cells to primitive ectoderm, while in combination with dibutyryl cyclic AMP they differentiate

into parietal endoderm. Interestingly, RA induced embryoid bodies of F9 cells differentiate into visceral endoderm (Strickland and Mahdavi, 1978). P19 EC cells are closer to ES cells and treatment of P19 embryoid bodies with DMSO results in differentiation into cells with characters of cardiac and skeletal muscle cells (McBurney et al., 1982). In contrast RA treatment of P19 embryoid bodies leads to formation of cells resembling neurons, glia and fibroblast (Jones-Villeneuve et al., 1982). RA treatment of P19 monolayer culture yields endodermal and mesodermal derivatives (Mummery et al., 1986).

Mouse ES cells can be differentiated into variety of cell types upon treatment with RA or other factors. In most protocols, differentiation is induced in LIF and mercaptoethanol free media with reduced serum levels. ES cells can be differentiated with or without embryoid body formation. RA induction of embryoid bodies from day 0 to 2-5 days results in the formation of neurons or glial cells (Fraichard et al., 1995; Glaser and Brustle, 2005; Gottlieb and Huettner, 1999; Strubing et al., 1995) while treatment between 2-5 days yields presomitic mesoderm. Removal of RA between 5 to 7 days after embryoid body formation followed by treatment with adipogenic factors like insulin, triiodothyronine or thiazolidine and PPAR γ leads to adipocytes (Dani et al., 1997). Treatment with BMP-4 or TGF- β 3 results in formation of osteoblasts and chondrocytes (Kawaguchi et al., 2005). RA treatment of ES cell derived embryoid bodies after 5-9 days of formation results into ventricular cardiomyocytes (Wobus et al., 1997) while treatment with RA and dibutyryl cAMP yields contracting smooth muscles (Drab et al., 1997).

Positional identity of ES cell derived neural cells is specified as anterior forebrain but low concentrations of RA treatment posteriorize gene expression patterns to those of mid/hindbrain. Treatment with high concentrations of RA results in hindbrain to rostral spinal cord identity (Liu et al., 2001). Plasticity in the ability of ES cells to differentiate into various cell types, like neural, cardiac and adipocytes, cells of different origins (endoderm, ectoderm and mesoderm) under influence of different signaling molecules in a concentration and time-dependent manner provides various avenues to exploit this system for developmental studies. I wanted to explore this approach to address binding properties of HOX proteins in a model of neural tissue. The ES system had the potential to provide a uniform platform to compare and contrast binding properties of different HOX proteins under programmed differentiation conditions.

During neuro-ectodermal differentiation, Hox genes are among first few genes to be activated in response to RA. *Hox1*, *RAR-Beta* and *Cyp26a1* are known to be among the fastest responding genes. During this differentiation process there appears to be a co-linear

activation of Hox genes, such as 3' Hox genes are sequentially activated before 5' members of the clusters. Developing a detailed understanding of the dynamics of programmed ES cell differentiation in response to RA is important to establish an experimental system. My main rationale for characterizing ES cell differentiation was to understand when specific Hox genes are activated and to measure their relative levels of expression so I could conduct genomic experiments to identify binding sites of HOX proteins under appropriate conditions.

Many different ES cell lines are available and I selected one because it offered unique technical advantages. Genome-wide identification of HOX binding sites is hampered by a lack of appropriate antibodies for HOX proteins and limited availability of relevant embryonic tissues. To deal with this issue, I selected KH2-ES cells because they provide a convenient means for site-specific integration of cDNAs encoding epitope-tagged proteins at a single defined target site in the genome at a promoter under tight Tet control to modulate levels of expression (Beard et al., 2006). These lines are able to go germ line and can be grown in large cultures.

I performed a comprehensive characterization of the temporal dynamics of the neuro-differentiation process in KH2 cells. Transcriptional profiling of a detailed time course of differentiation in response to RA was done with a variety of platforms: Affymetrix arrays, RNA-Seq, Agilent high density Hox tiling arrays (designed by our informatics group) and ABI qt-PCR arrays (TLDA cards). These experiments enabled me to determine the precise order, timing and levels of gene expression of all 39 Hox genes; identify novel non-coding transcriptional activities in and around Hox clusters; and globally characterize rapid changes in gene expression during differentiation. This provides a basis to understand and compare ES differentiation with normal hindbrain and spinal cord development. I also used chromatin immune precipitation (ChIP) and high density Hox tiling arrays or next gen sequencing in combination with a variety of antibodies against active and repressive histone marks, RNA Pol II (N-term and CTD regions) and RARs & RXRs retinoid acid receptors. These results will be presented in this chapter and generate a detailed picture on the accessibility and dynamics of the epigenetic states of Hox clusters related to their transcriptional activity and identify new sites of potential direct input by RA signaling through occupancy of receptor binding.

To understand and calibrate the programmed differentiation process using RA I set the following objectives (Fig.3-1).

Aim 1: Study activation of Hox clusters in neuro-differentiation using an ES cell model system

- a) Characterize the properties of neuro-differentiation process through studying global changes in gene expression and local qualitative and quantitative changes in transcriptional activity (coding and non-coding) of Hox cluster
- b) Analyze accessibility and epigenetic properties of Hox clusters during differentiation using dynamics of activation and repressive marks and PolII occupancy over Hox clusters
- c) Analyze binding of retinoid receptors (RARs & RXRs) to study their roles in *cis*-regulation of Hox genes during differentiation.

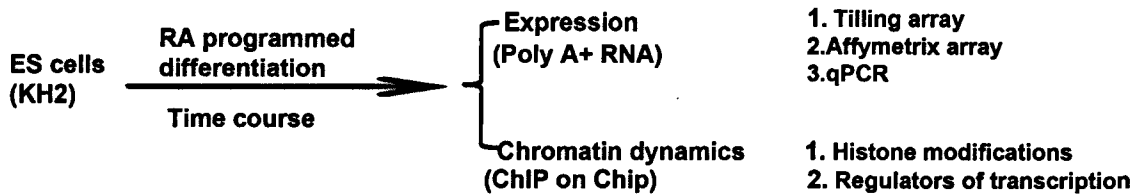


Figure 3-1 Flow chart explaining design of experiment used in current study
 KH2 cells were harvested after fixed length of RA treatment. After separation of feeder cells, RNA was isolated and divided into four portions. Each portion was used in for Tiling array, affymetrix, qPCR and RNA seq. For ChIP on Chip Cells were harvested as per protocol and isolated RNA was hybridized on custom Agilent 2x105K HOX tiling arrays after appropriate labeling

3.1 Result

3.1.1 Analysis of genome-wide gene expression profile during RA induced murine ES cell differentiation

I performed a detailed time-course of RA treatment of KH2-ES cells, harvested cells, isolated RNA and transcriptionally profiled the patterns of gene expression using microarrays on an Affymetrix platform.

We analyzed the Affymetrix data by RMA (Robust Multi-Chip Average) method using background correction, quartile normalization, and median polish. K-mean clustering was employed in whole data set in an unbiased manner and clusters were validated using Davies-Bouldin index. We further clustered genes showing 2 fold or more changes in gene expression level using hierarchical clustering. Clustering into 8 groups seems to be best fit. Among them, the first six clusters show maximum mean changes in gene expression (Fig.3-2). Four of these k-mean clusters have shown moderate to high up-regulation while two clusters represent genes with-down regulation. 4467 and 615 genes were down-regulated while 262, 1481, 4528, 2404 genes were respectively up-regulated with distinct induction profiles.

The data was further analyzed for over-representation of Gene Ontology (GO) terms to investigate the gene expression profiles and aspects of the differentiation process (Fig.3-3). Cluster 1 represented by 262 genes, shows enrichment of developmental processes: pattern specification, regionalization and anterior /posterior pattern formation. Cluster 3 is represented by 4528 genes and shows enrichment of terms related to neurogenesis, including: neurogenesis, neuronal differentiation and neuronal projections. This cluster contains the highest number of genes displaying dynamic differences over the time-course. Cluster 6 is represented by 615 down-regulated genes, enriched with terms of: stem cell differentiation, stem cell development, stem cell maintenance and negative regulation of cell differentiation.

Hence, analysis of GO term enrichment confirms that the differentiation process is able to drive cells toward neuro-ectodermal fates, as indicated by up-regulation of a large number of genes related developmental processes and neurogenesis and associated down-regulation of stem cell related genes. Further, I looked closely at GO term enrichments for classes of genes showing specific induction profiles (Fig.3-3). Cluster 1 consists of genes, which showed early induction after 2 hours of RA treatment and showed a constant increase up to 72 hours of RA treatment. These genes showed over-representation of GO terms related to anterior posterior pattern formation, regionalization, and pattern specification.

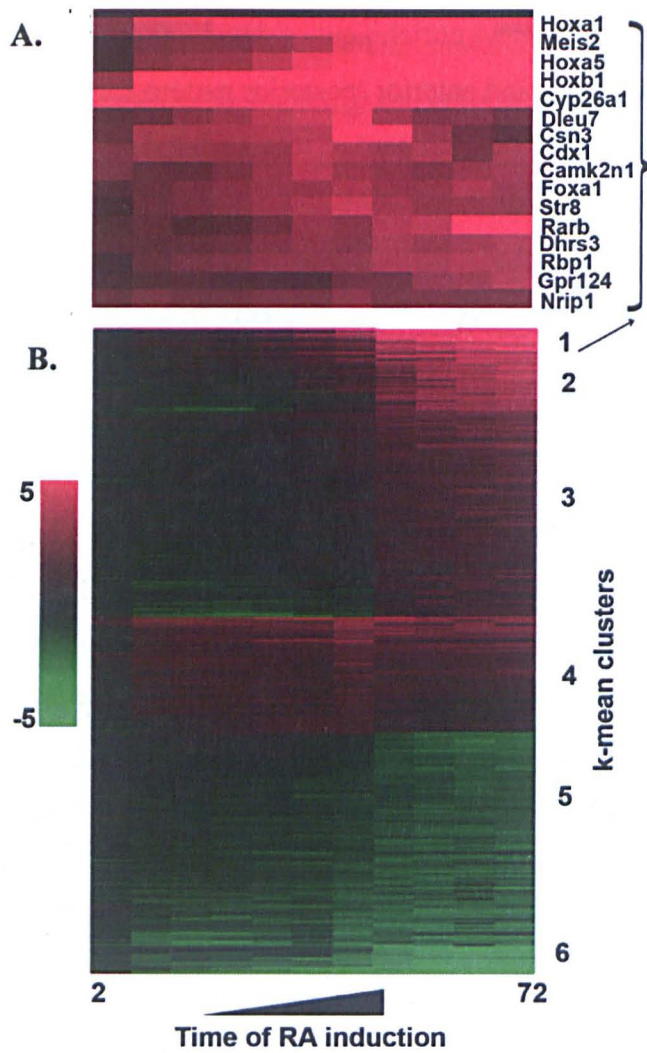


Figure 3-2 Heat map showing global change in gene as analyzed on affymetrix Mouse Genome 2.0 arrays

A. Upper panel shows change in expression profile of 15 rapidly induced genes. B. Presence of Hox genes and cofactors are noticeable. Middle panel shows Heat map of global change in gene expression upon RA induced differentiation. Expression data was clustered by k-mean clustering and only clusters with absolute value of cluster mean > 0.5 are shown Expression values are average value from three independent biological replicates

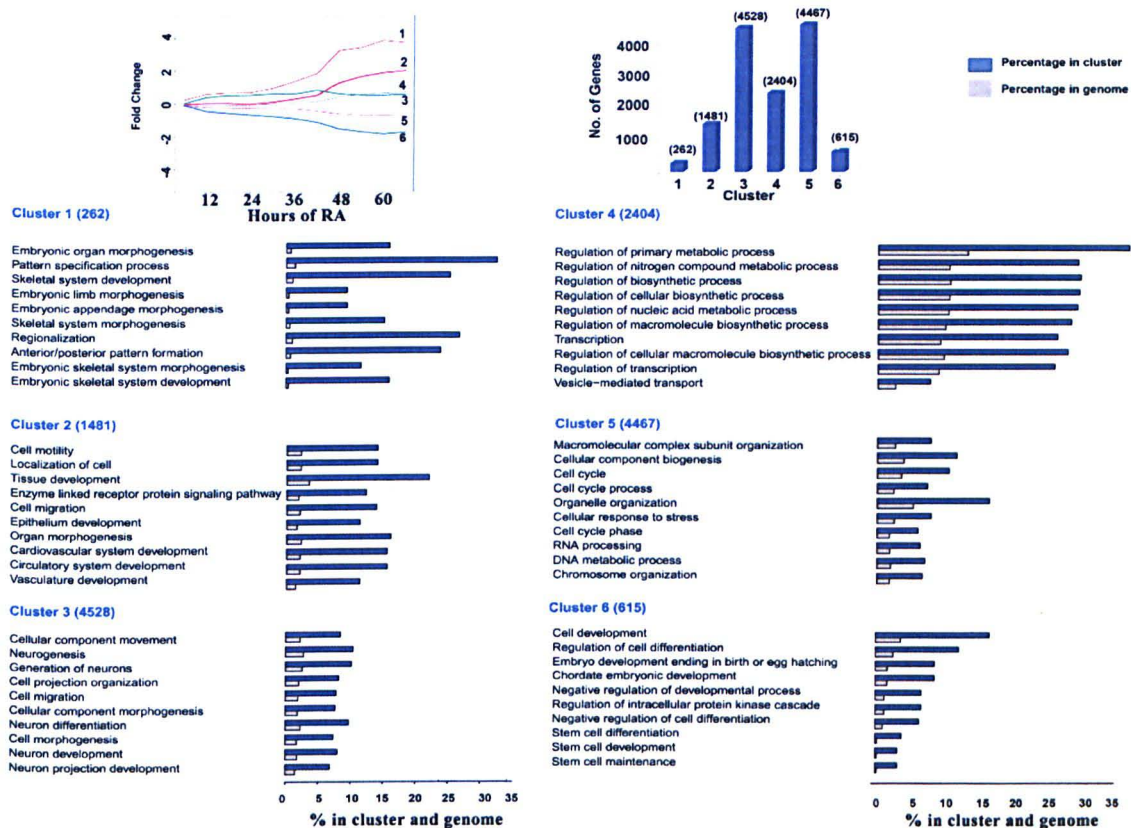


Figure 3-3 Gene ontology (GO) term analysis of differentially expressed genes
 Percentage of terms in cluster (Blue) and whole genome (gray) is shown side by side. Number of genes in each cluster has been shown in parenthesis. Upper right panel shows general induction profile of genes in each cluster. Clusters are marked from 1-6 in the graph. Y-Axis shows fold change compared to uninduced cells while X-Axis indicates time of RA treatment. Upper Left panel shows number of genes in each cluster. X-Axis indicates cluster while Y-Axis shows number of genes.

Cluster 2 includes genes which are up-regulated after 12 hours of RA induction, and their level keep increasing until 72 hours of RA treatment. These genes are mainly involved in tissue and anatomical structure development, organ morphogenesis and system development. Cluster 3 shows early induction upon RA treatment for two hours and show up-regulation until 4 hours of RA treatment and later remain at steady state levels of gene expression. They represent a group of genes involved in transcriptional regulation and anatomical structure morphogenesis. Cluster 4 contains genes involved in stem cell maintenance and stem cell development. They show constant down-regulation throughout the time-course of RA treatment.

The Affymetrix data indicates that 0.03 μ M of 9-*cis*-RA is capable of differentiating murine ES cell line KH2 into neuro-ectodermal fate. It is evident that most over-represented GO terms correspond to up-regulated or down-regulated genes indirectly or directly related with assigning a neuronal fate to these cells. Interestingly it seems that we were able to achieve a gradual differentiation process, as evident by the down-regulation

profile of stem cell maintenance and development genes (Cluster 4). Alternatively this may indicate that it takes some time to turn over the early stem cell transcriptional program to that of the differentiated state.

3.1.2 Induction of Hox genes during neuro-ectodermal differentiation

In embryos neuronal differentiation is controlled by anterior posterior patterning systems and requires posteriorizing signals. The group of earliest genes expressed upon RA treatment correlates with transcriptional regulation and pattern specification processes. These genes set the early pattern for commitment of these cells to a neuronal fate and many positive and negative regulators of transcription take part in this process. Consistent with this idea, major genes in Cluster 1 are: *cyp26A1*, *Hoxb1*, *Hoxa1*, *Hoxa5* and *Crabp2* (Fig.3-2). Among them transcription factors like *Hoxb1*, *Hoxa1* and *Hoxa5* are known to be involved in early pattern formation and neuronal development. Genes that show early down-regulation include negative regulators of neuronal fate commitment, segmentation specification, and ganglioside catabolic process. Genes with delayed induction profiles are also involved in pattern formation, while very late induced genes are involved in nervous system development and elaboration. Together this data reveals that RA induced differentiation of KH2 cell lines shows a general pattern of gene regulation similar to *in vivo* phases of the neural development process. Genes involved in transcription, transcription regulation are induced first along with genes for anterior posterior patterning. The next genes are involved in cell fate commitments while last genes induced are involved in nervous system development and specification.

Among the cluster of rapidly induced genes, Hox genes and co-factors are highly represented (Fig.3-2). Anterior (3') PG genes show rapid activation while posterior genes exhibit either weak or no activation. Surprisingly, a strict pattern of temporal co-linearity was not followed by many Hox genes during RA induced differentiation. The order of gene expression is staggered, especially in the HoxA cluster. To further explore this result and quantitate gene expression, we used custom made Applied Biosystems TLDA cards containing probes for the 39 murine Hox genes and 5 endogenous controls. Quantitative PCR confirms the Affymetrix results (Fig.3-4 and 3-5). The only exception is *Hoxb1* and a few other HoxB cluster genes. The qPCR shows *Hoxb2* as highest responding in the HoxB cluster and it is expressed around 2 fold higher than other HoxB genes. *Hoxb2* is the strongest expressing gene in all four Hox clusters in response to RA induction. The qualitative activation profile of the HoxA cluster matches the quantitative expression profile obtained through TLDA card analysis. HoxC responds in a delayed manner

compared to HoxA and HoxB in a co-linear fashion with exception of *Hoxc9*. *Hoxc9* is activated before *Hoxc6* and *Hoxc8* and is strongest induction profile. The HoxD cluster shows very little quantitative expression compared with the other three clusters (Fig.3-5). The Affymetrix expression data from HoxD indicates there is a very late delayed response of HoxD cluster and qPCR result indicates the levels of induction are very low.

3.1.3 Analysis of transcriptional activity in Hox cluster

A limitation of the Affymetrix and qPCR approaches are that they are confined to the standard selected probe sets used on the array platforms. This will miss transcripts from different regions of genes, unannotated transcripts and give a limited view of the overall changes in the transcription profiles.

We wanted to investigate the nature of the transcriptional profiles spanning the four Hox clusters in more detail. Therefore, we designed custom tiling array with probes covering both strands of DNA covering the full Hox clusters regions and large areas of their flanking DNA up to the adjacent non-Hox coding genes on the 5' and 3' sides of the clusters. These arrays could be used to more systematically explore transcription patterns within and around Hox clusters and correlate them with epigenetic profiles using ChIP-on-chip approaches. Therefore, I utilized RNA from the same time-course of programmed differentiation analyzed by Affymetrix and qPCR for transcriptional activity in Hox genomic regions using these custom high density tiling arrays.

Hybridization was done comparing harvested time-points with undifferentiated ES cells to calculate the fold levels of induction. A heat map indicating relative levels of transcription over time and along the chromosomal position reveals a dynamic pattern of transcription. It is evident that both strands from the Hox complexes are transcribed (Fig.3-6). The order and quantitative expression profile over coding regions matches with data obtained for these genes from the Affymetrix and qPCR analyses. However, we observed a large amount of transcriptional activity from non-coding regions in the Hox clusters from both sense and anti-sense strands. Noticeable activities were observed in region between *Hoxa1* and *Hoxa2*, *Hoxa3* and *Hoxa4*, *Hoxa4* and *Hoxa5*, *Hoxb6* and *Hoxb7* and *Hoxb7* and *Hoxb8*. Several of these correspond to novel or previously unknown transcriptions and we did not observe expression of previously characterized long non-coding RNAs (lincRNAs) *Hot-tip* and *Hotair*. This is likely to be due to the timing as posterior genes and the HoxC cluster are weakly expressed during the time-course in ES cells and they correlate with later programs in the embryo.



Figure 3-4 Changes in Hox genes upon RA induction as analyzed on affymetrix Mouse Genome 2.0

Rapid and robust induction of A and B locus and staggered induction of Hox genes in A locus are noticeable. Each Row represent one Hox gene and each block represent one Hox cluster. Each column represents one time point of RA treatment. It is worth mentioning that some genes are shown twice which signifies use of two different probe set in affymetrix array. These probes may or may not show same trend of expression change

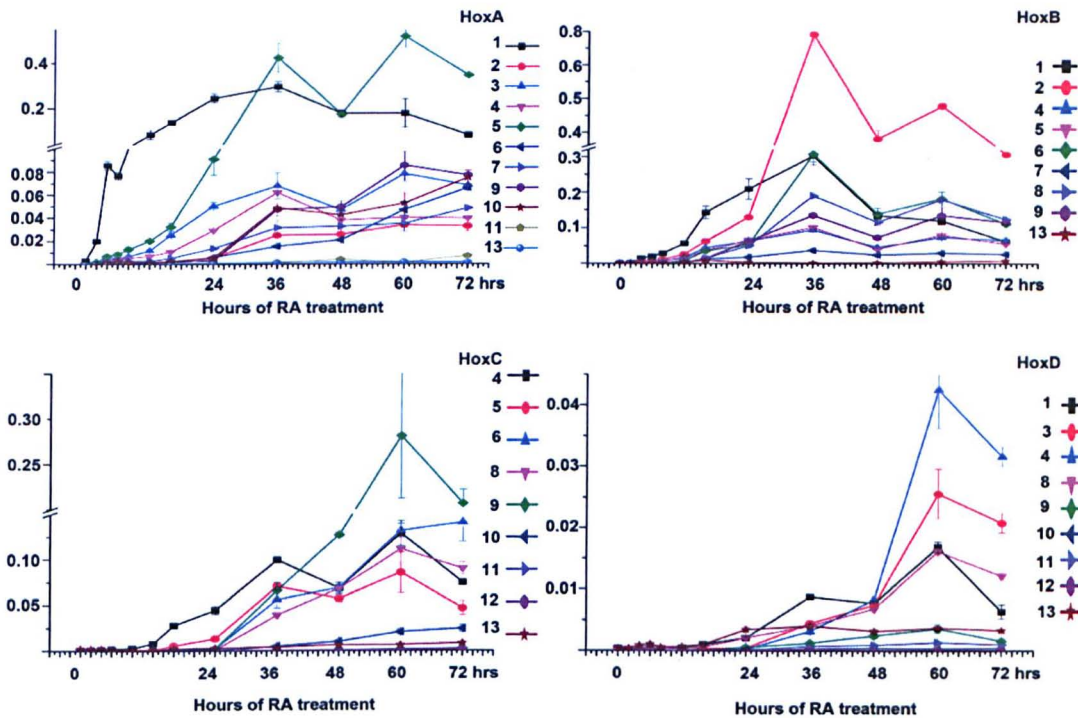


Figure 3-5 Quantitative changes in Hox gene expression upon RA induced differentiation

All data points are average of three biological and two technical replicates. X-axis indicates time of RA treatment while Y-axis shows non-scaled delta Ct value representing change in transcripts levels compared to housekeeping genes. Non-scaled Ct values were normalized against Ct values of *Gapdh* and *Tbp*. Y-axis in all clusters is in different scale. Difference in Y-Axis indicates differential response of individual cluster to RA treatment.

We were able to see rapid induction of *Hotarm* between *Hoxa1* and *Hoxa2* which was previously found in human cells. We named this transcript, *mHotarm* for clarity. We identified a novel transcript rapidly induced by retinoic acid treatment positioned between *Hoxb4* and *Hoxb5* on the coding strand and named it, B4-B5 intergenic transcript (*B²iT*). This *B²iT* transcript is

induced at 4hrs of RA treatment along with *Hoxb4* and *Hoxb5* and is expressed at similar levels.

We observed a striking and rapidly induced regions of extensive transcriptional activity positioned at ~50kb 3' upstream of *Hoxa1* in the HoxA cluster. A large region of approximately 15 kb is transcribed and there are transcripts from both strands of DNA. The timing indicates that this region is activated earlier or at the same time as *Hoxa1* making on

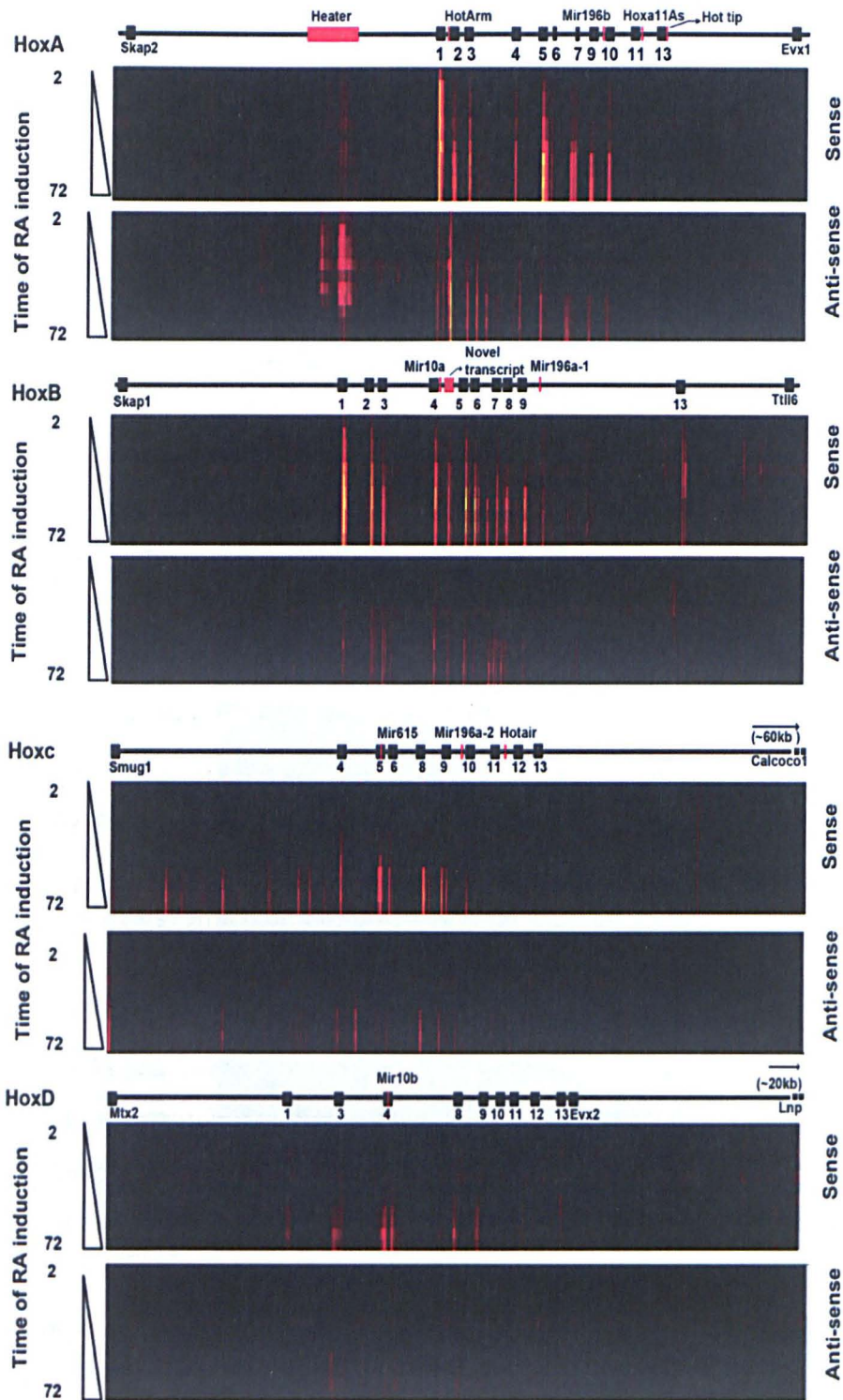


Figure 3-6 Transcriptional activity in Hox clusters upon various length of RA treatment

Cy3 labeled mRNA was hybridized to custom Agilent 2x105K HOX tiling arrays. Heat maps were made on IGV genome viewer 1.5. Upper and lower panels in each cluster represent sense and antisense strand respectively. Y-Axis indicates time of RA treatment while X-Axis represents genomic coordinates spanning between nearest 3' and 5' non – Hox gene neighbor in Hox cluster. Intensity of color (from red to yellow) indicates increasing levels of transcript.

of the earliest Hox cluster associated transcripts to be induced by retinoids. Our analysis shows that this is a complex transcriptional landscape and produces multiple Polyadenylated, spliced transcripts of various lengths, with different promoters and activity from both strands. We collectively named this collection of transcripts *Heater (Hox EARly Transcribed REgion)* to designate its early transcriptional activity from the HoxA cluster. Individual transcripts are labeled with prefix H (for heater) and followed by a number. We have identified 8 distinct spliced polyadenylated transcripts and a few discreet transcriptional units in this region using a combination of informatics and 5'RACE. Most of the transcripts identified in this region contain 2-3 spliced transcriptional units. While transcriptional activity in this region can be scored from both strands, the anti-sense strand is transcribed at a higher level compared to the sense strand. Many transcripts from the Heater region are also present in uninduced KH2 cells. We were not able to identify spliced transcripts from the most 5' end of this region, although high transcriptional activities are scored on tilling array.

The highest levels of intergenic transcriptional activities were observed in the HoxA cluster followed by the HoxB cluster. The HoxC cluster showed very little transcriptional activity while the HoxD cluster had no transcriptional activity in intergenic region in the time points examined. The region between *Hoxc4* and its upstream non-Hox gene, *smug1*, display a distinct region of intergenic transcripts. These transcripts take at least 24hrs of RA treatment for induction. We quantitated intergenic transcripts at a few time-points using RNAseq (FPKMs) by comparing intergenic transcripts in uninduced KH2 cells and RA induced cells at 24hrs. At least five transcripts were transcribed in uninduced KH2 at considerably high level. These intergenic regions are 52052832-52056774 (*Heater*), 52056859-52056523(*Heater*), 52062665-52065158(*linc1547*),

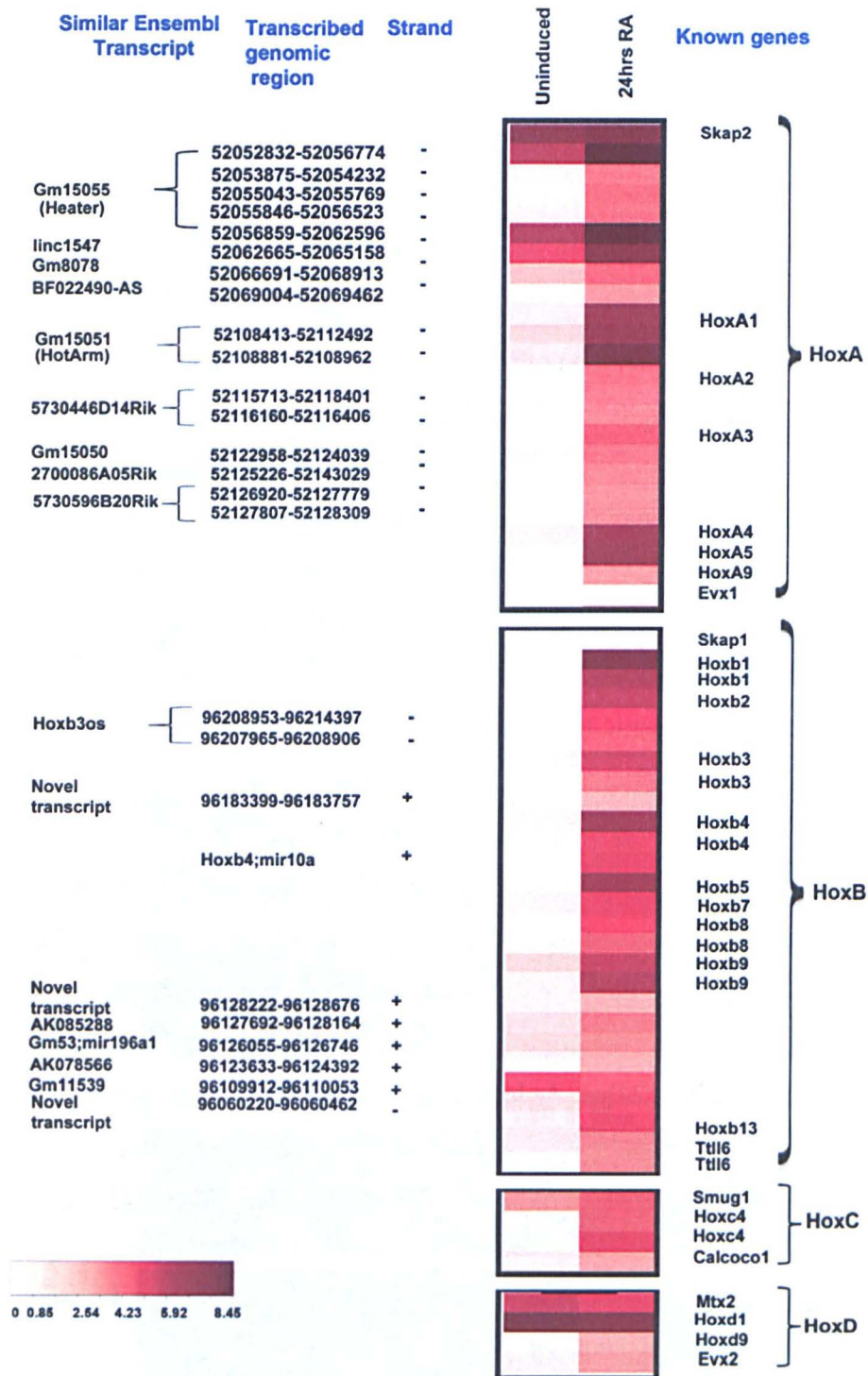


Figure 3-7 Quantitative analysis of known genes and non-coding transcripts using RNA-Seq

Directional mRNA seq libraries were sequenced by illumine GAIIX. Transcript structures were generated by cufflink and cuffcompare. Any such *ab-initio* loci which were detectable in only one technical replicate were removed. Average FPKM values from two biological replicates are used for generation of heat map. Right hand side of figure shows genomic coordinate and left hand side shows expression in uninduced and 24 hours of RA treated cells. Known Ref seq transcripts are named in rightmost panel as similar ensemble transcript while known UCSC genes are shown in left most panels as known genes. Transcripts from sense strand are indicated as + while antisense strand as -. Sense strand is equivalent to general direction of transcription of Hox genes in cluster.

52069004-52069462 (*Gm8078*) in chromosome 6 (HoxA cluster) and 961019912-96110053 in Chromosome 11(HoxB cluster) (Fig.3-7). The other *Heater* transcripts, *mHotarm*, *Hoxb3os* and intergenic region 96060220-96060462 were highly induced upon 24hours of RA treatments. Gm11539 (96109912-96110053) on chromosome 11 (HoxB cluster) was slightly down regulated upon RA treatment. Many other transcripts on both stands were observed at a low to moderate level of induction at 24hrs of RA treatment. In comparing expression levels of adjacent genes (3' and 5') to respective intergenic transcripts the expression level are different. This suggests that these are not run-through transcripts and are regulated in a manner independent of adjacent genes. This highlights an unexpected degree of complexity in the transcriptional activity in and around Hox clusters in addition to the coding regions that need to be considered in thinking about transcriptional regulation of the clustered Hox genes.

3.1.4 Temporal changes in Epigenetic properties and retinoid receptor occupancy along Hox cluster upon RA induced neuro-ectodermal differentiation

I next wanted to compare how the transcriptional activity in genomic regions spanning the Hox clusters correlated with epigenetic changes in the state of chromatin taking place in these same regions during neuro-ectodermal differentiation. Towards this goal I performed CHIP-on-chip assays over a time-course of RA programmed differentiation. Antibodies were used against: 1) H3K27Me3, as a repressive mark; 2) Suz12, a member of the PRC2 complex, a mark of chromatin silencing; 3) N-CoR, a co-repressor associated with retinoic acid receptors; 4) H3K4Me3, as an indicator of an active state; 5) Pol II, as a mark of active transcription and 6) RAR alpha, beta, gamma and RXR, as an indicator of binding of retinoic acid receptor hetero dimeric complexes. We analyzed cells before induction with RA, and after 4, 8 and 24hrs of RA treatment (Fig.3-8, Fig.3-9).

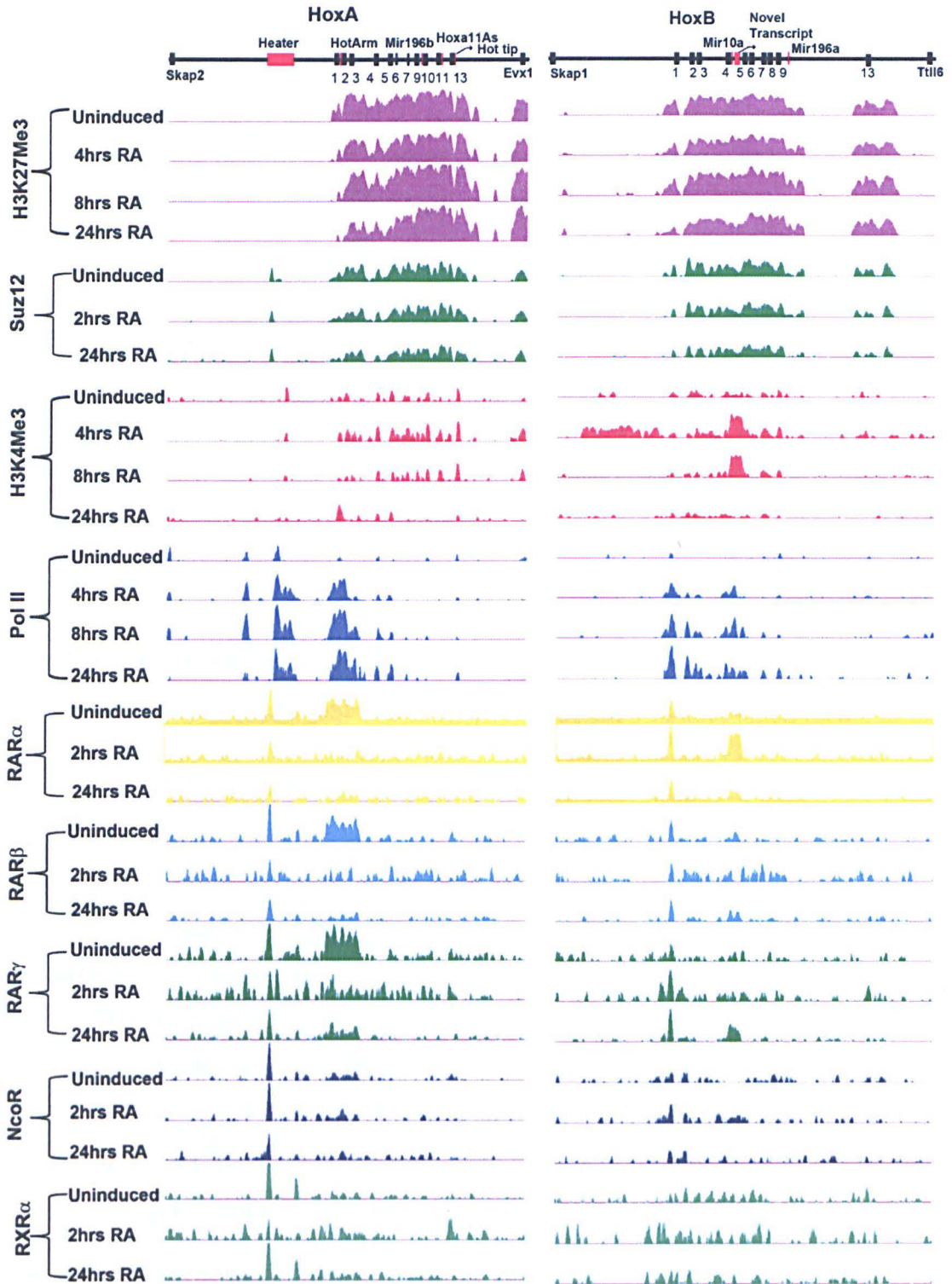


Figure 3-8 Changes in epigenetics state and RA receptor occupancy of Hox A and B cluster during RA induced neuro-ectodermal differentiation

UCSC genome tracks showing dynamics of various histone modification (H3K4Me3), Histone modifiers (Suz12) occupancy of receptors (RAR alpha, beta and gamma, RXR alpha) Pol II and co-repressors (NcoR) in HoxA(right) and HoxB (Left) cluster. X-Axis represents genomic coordinates spanning between nearest 3' and 5' non-Hox gene neighbor in Hox cluster. Y-Axis is same of individual antibody. Tracks were configured by using windowing function as mean and smoothing windows as 10 pixels in UCSC genome browser. Many major changes in Pol II occupancy and gain of H3K4Me3 are related to non-coding transcripts.

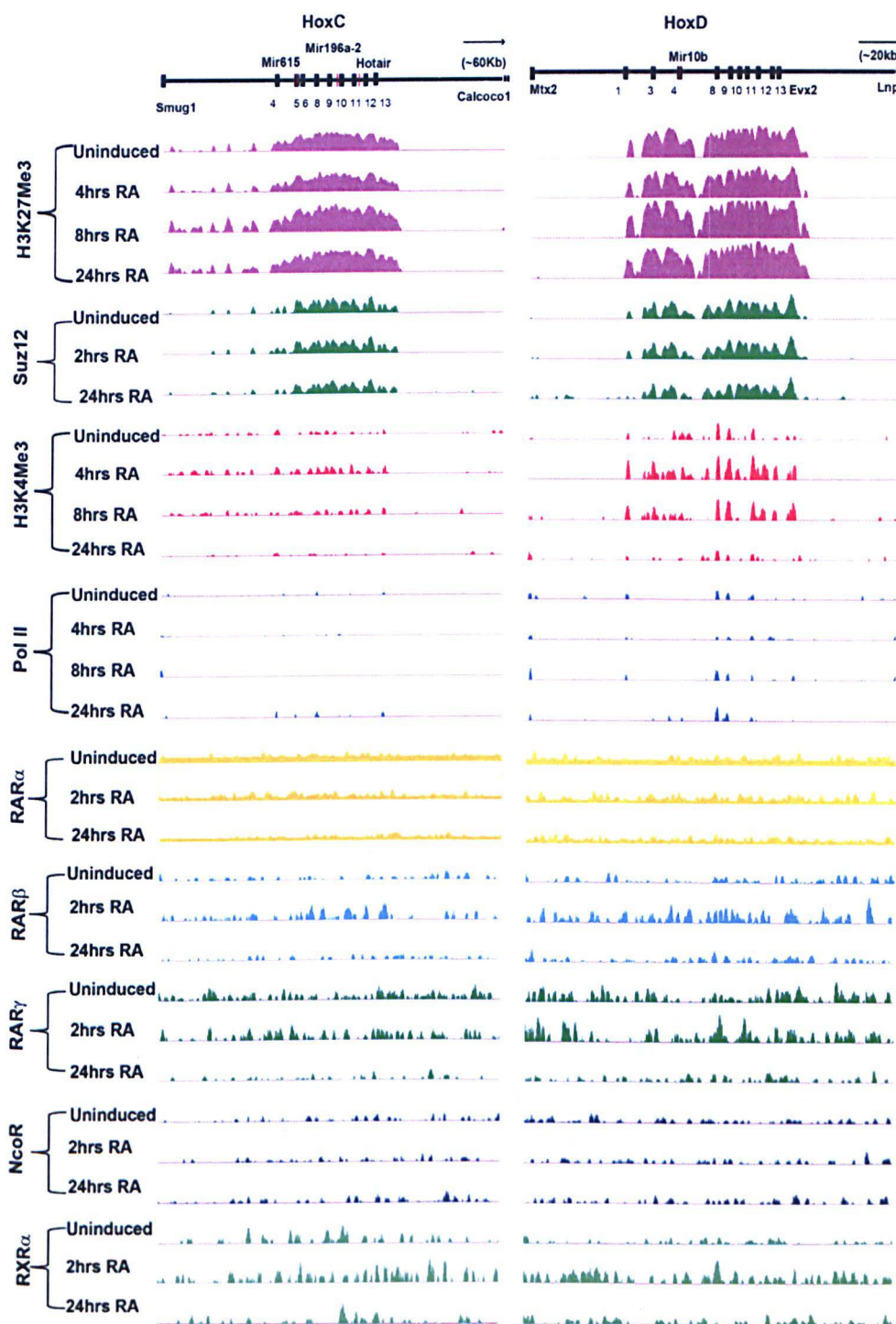


Figure 3-9 Changes in epigenetics state and RA receptor occupancy of Hox C and D cluster

UCSC genome tracks showing dynamics of various histone modification (H3K4Me3), Histone modifiers (Suz12) occupancy of receptors (RAR alpha, beta and gamma, RXR alpha) Pol II and co-repressors (NcoR) in HoxC(right) and HoxD (Left) cluster. X-Axis represents genomic coordinates spanning between nearest 3' and 5' non-Hox gene neighbor in Hox cluster. Y-Axis is same of individual antibody. Tracks were configured by using windowing function as mean and smoothing windows as 10 pixels in UCSC genome browser Very little changes are seen in these clusters during RA induced differentiation .Bivalent state of HoxD locus is most noticeable feature.

In undifferentiated ES cells the distribution of the H3K27Me3 repressive mark is widely spread over the entire HoxA and HoxB complexes. This is consistent with the need to maintain Hox genes in a silent or inactive state to prevent them from inducing differentiation of these pluripotent cells. Over the 24 hr period of RA treatment there is a slight decrease in the levels of H3K27Me3 over the most 3' genes, *Hoxa1* and *Hoxb1*. However, this epigenetic mark remains over the majority of these two clusters despite the fact that the majority of the HoxA and HoxB genes are expressed based on the Affymetrix, qPCR and tiling array analyses described above. This clearly indicates that in ES cells the H3K27Me3 mark does not need to be completely eliminated as a precursor to facilitating expression of genes from these two Hox clusters. These repressive marks are slowly erased upon increased length of RA treatment. Genes are activated upon RA treatment as rapid as 2 hrs after treatment but it takes a longer amount of time for removal of repressive mark. Performing similar analysis for H3K27Me3 on all four clusters and up to 36hrs shows the gradual removal of this epigenetic mark (Fig.3-10). On careful examination of the TSS of activated genes, we can observe rapid removal of repressive mark from TSS. Hence, removal from the whole cluster is slower compared to rate of gene activation. In further support of this idea, Suz12, which marks the PRC2 repressive complex, is also spread over the HoxA and HoxB clusters in both undifferentiated ES cells and 24hrs. These marks of a repressed state appear to be confined to the regions spanning the coding genes because there is no evidence for occupancy in 3' or 5' flanking regions immediately adjacent to the clusters. There is one exception, as there is peak of Suz12 occupancy over part of the *Heater* transcribed region, 50 kb upstream of *Hoxa1*.

With respect to active marks, there is evidence for the occupancy of Pol II over some genes in the HoxA cluster in undifferentiated cells. This polymerase is concentrated near the promoter region and not located over the whole transcription units of *Hoxa1*,

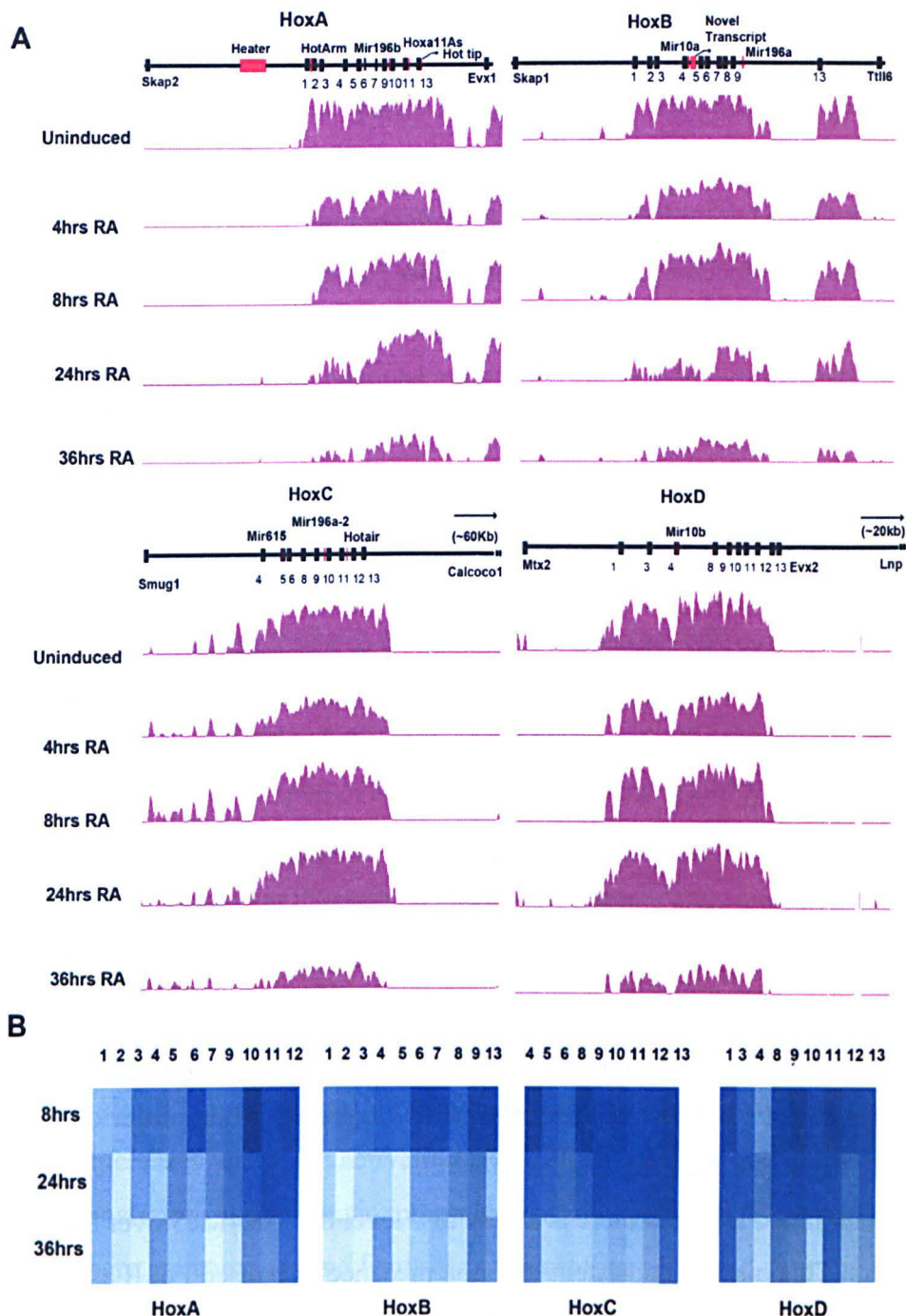


Figure 3-10 Dynamic changes in H3K27Me3 marks on Hox gene promoters and over Hox cluster

A. UCSC genome tracks showing dynamics of H3K27Me3 in all four hox clusters in differentiating ES Cells. X-Axis represents genomic coordinates spanning between nearest 3' and 5' non-Hox gene neighbor in Hox cluster. Y-axis is same for all tracks shown. Tracks were configured by using windowing function as mean and smoothing windows as 10 pixels in UCSC genome browser. Loss of repressive marks over Hox cluster is evident during RA induced differentiation. B. Heatmap showing relative levels of H3K27Me3 levels compared to uninduced ES cells in the region of 500+ bp around TSS (Transcription Start Site). Each column represents a Hox gene and each block represents a cluster. Each row represents individual length of RA treatment. Rapid loss of repressive marks over TSS and promoter region can be seen around active genes.

Hoxa3, *Hoxa5* and *Hoxa13*. This pattern is characteristic of paused Pol II, suggesting that polymerase has initiated but is paused and waiting for signals to potentiate elongation.

From my analysis of the genes most rapidly induced by RA in Affymetrix arrays, I noticed that *Hoxa1* was one of the most rapid. In collaboration with the Shilatifard group, we found that *Hoxa1* is induced by recruiting the Super Elongation Complex to the promoter to stimulate elongation of the paused polymerase (Lin et al., 2011). In addition, we observed that many of the most rapidly RA induced genes in the ES cells are similarly regulated by transcriptional elongation and not by initiation (Lin et al., 2011). This provides a mechanism to explain why some genes are able to be induced so rapidly and suggests that this might be important as a means of triggering and coordinating the differentiation process. The presence of Pol II over *Hoxa3*, *Hoxa5* and *Hoxa13* suggests that they too may be induced by regulation of elongation of paused polymerases. There is also evidence for paused Pol II over a part of the *Heater* transcribed region indicating that some of these transcripts may also be regulated by elongation in line with rapid induction of this region in response to RA.

In the presence of RA there is a rapid increase in occupancy of Pol II over the *Heater* and 3' genes of the HoxA and HoxB clusters at 4hrs and it increases through to the 24 hr time-point. Over the time course Pol II occupancy has changed along 3'-5' direction and begins to be detected over more 5' genes. This pattern clearly correlates with the observed patterns of Hox expression from transcriptional profiling. At the end of 24hrs of RA induction, we can see Pol II occupancy over *Hoxa1* to *Hoxa5* while posterior 5' HoxA are just beginning to recruit Pol II on their promoters.

In undifferentiated ES cells there is a very low level of H3K4Me3 over the HoxA and HoxB clusters. This may be consistent with the presence of paused Pol II as this mark in mammalian cells frequently correlates with Pol II initiation at the promoter. Upon 4hrs of RA treatment there is a rapid increase in H3K4Me3. In the HoxA cluster this appears as an increase in the general pattern detected in untreated cells. However, over the HoxB cluster in addition to an increase in the general pattern there is a large peak between *Hoxb4* and *Hoxb5* and a large region extending upstream, 3' of *Hoxb1*. These patterns were dynamic, as the upstream mark was lost at 8hrs and the peak between *Hoxb4* and *Hoxb5* began to decrease, disappearing by 24hrs. Gains of H3K4Me3 were progressively seen from anterior to posterior Hox genes upon increased length of RA treatment. The appearance of H3K4Me3 corresponds to current or future gene activation states. The

H3K4Me3 mark overlaps with much of the Pol II occupancy indicating the relative state of activation.

In examining Pol II occupancy in more detail, after 4 hours of RA induction Pol II is spread on a wider chromatin domain than expected if it was only present on the *Hoxa1* and *Hoxa2* promoters. The region spanning *Hoxa1* and *Hoxa2* also includes mHotarm, transcribed from the opposite strand (Fig.3-11). This indicates that Pol II is active on all of these transcriptional units at this early stage of induction, explaining the broad region of occupancy.

The region containing the *Heater* transcripts displays dynamic changes in the epigenetic state. Pol II occupancy can be seen on this region in uninduced ES cells. Uninduced ES cells also show occupancy of the elongation factors, AFF4 and Ell2. These factors along with another elongation factor Cdk9 increases their level of occupancy upon RA treatment, consistent with the idea that there is an increase in the rate of elongation from a paused polymerase upon treatment with RA. This region displays a bivalent state as it contains both H3K4Me3 and H3K27Me3 marks over the transcriptional units.

Unlike the pattern seen over the Hox clusters, the H3K27Me3 is more focused and does not cover a large domain in the *Heater* region. Pol II also shows focused occupancy along this region. Together, 5'-RACE results, Pol II occupancy and presence of H3K27Me3 and H3K4Me3 marks suggest that the series transcripts arising from the *Heater* region are generated from different promoters and vary with respect to the timing and mechanisms (paused versus non-paused) of activation (Fig.3-12).

The HoxB cluster generally behaves in a manner similar to the HoxA cluster. Pol II occupancy increases with gene activation and H3K4Me3 mark is increasingly gained from anterior to posterior (3'-5') direction upon increased length of RA treatment (Fig.3-8). However, there are some unique features in the profiles observed over the HoxB complex. It is interesting that in the initial response (4hrs) Pol II is rapidly recruited to the Hoxb1 and Hoxb4/Hoxb5 regions (Fig.3-8). It then appears over the Hoxb2 and Hoxb3 regions at 8hrs and 24hrs of RA treatment. These patterns do not directly correlate with the observed timing on the order of induction of the coding transcripts. There is no evidence for paused polymerase over Hoxb1, like that seen for Hoxa1, but there appears to be exceptionally rapid recruitment of Pol II over Hoxb1. This correlates with expression seen in embryos where Hoxa1 is the first Hox gene induced and Hoxb1 the second. It is tempting to speculate that this timing difference in embryos reflects the different modes of activation of these two PG1 genes.

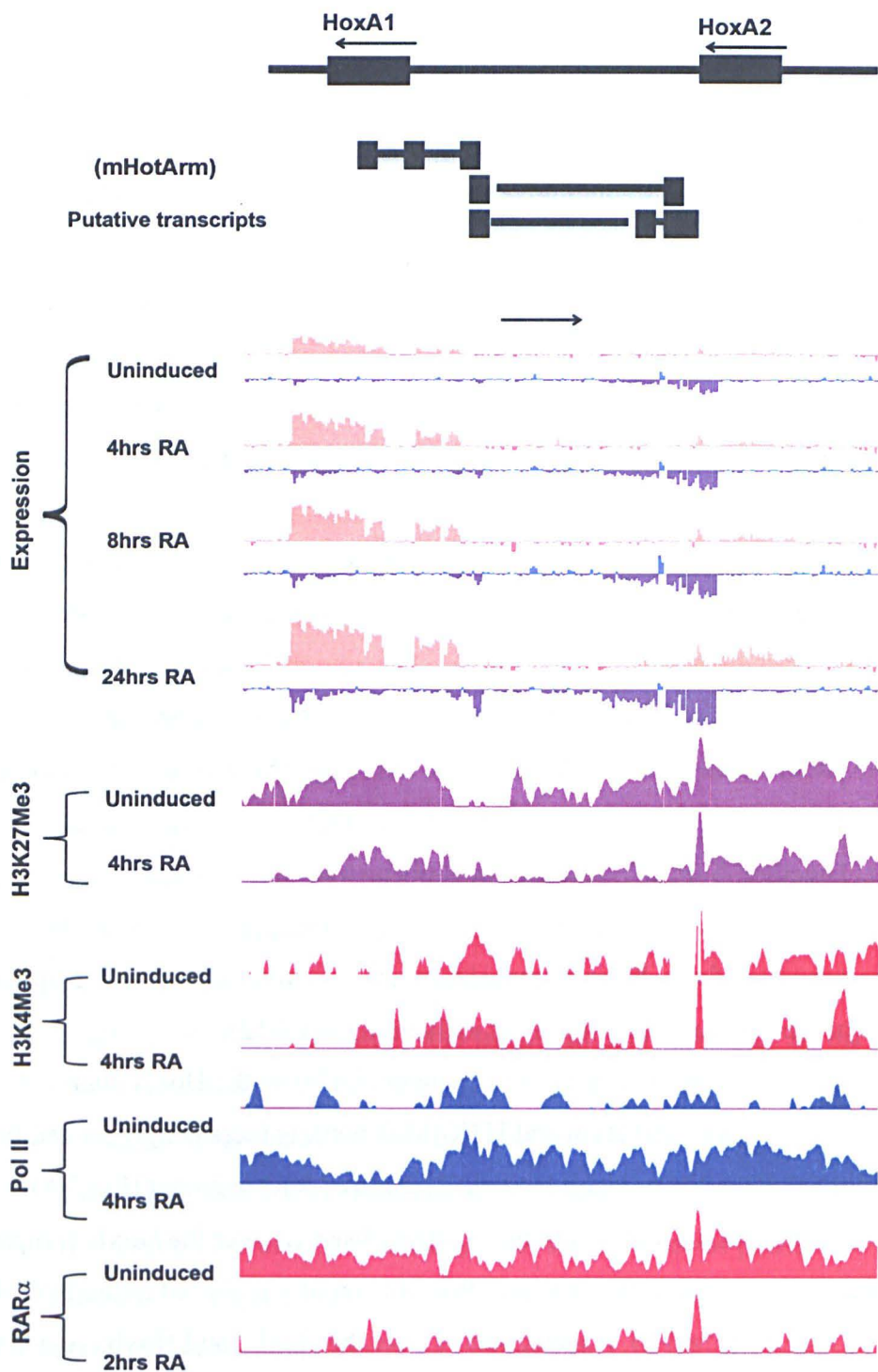


Figure 3-11 Occupancy of RAR receptors and changes in transcriptional and epigenetic properties in region encompassing HoxA1 and HoxA2

UCSC genome tracks showing region between Hoxa1 and Hoxa2. Expression profile indicates rapid expression of Hoxa1 and a non-coding transcript (mHotarm) upon RA treatment. Early changes in Histone marks (H3K4Me3 and H3K27Me3), Pol II, and retinoic acid receptor (RAR alpha) are also shown. Y axes for individual antibodies are same. Rapid induction of Hoxa1 and mHotarm can be seen. Multiple isoforms of Hotarm is shown schematically. Tracks were configured by using windowing function as mean and smoothing windows as 10 pixels in UCSC Genome browser

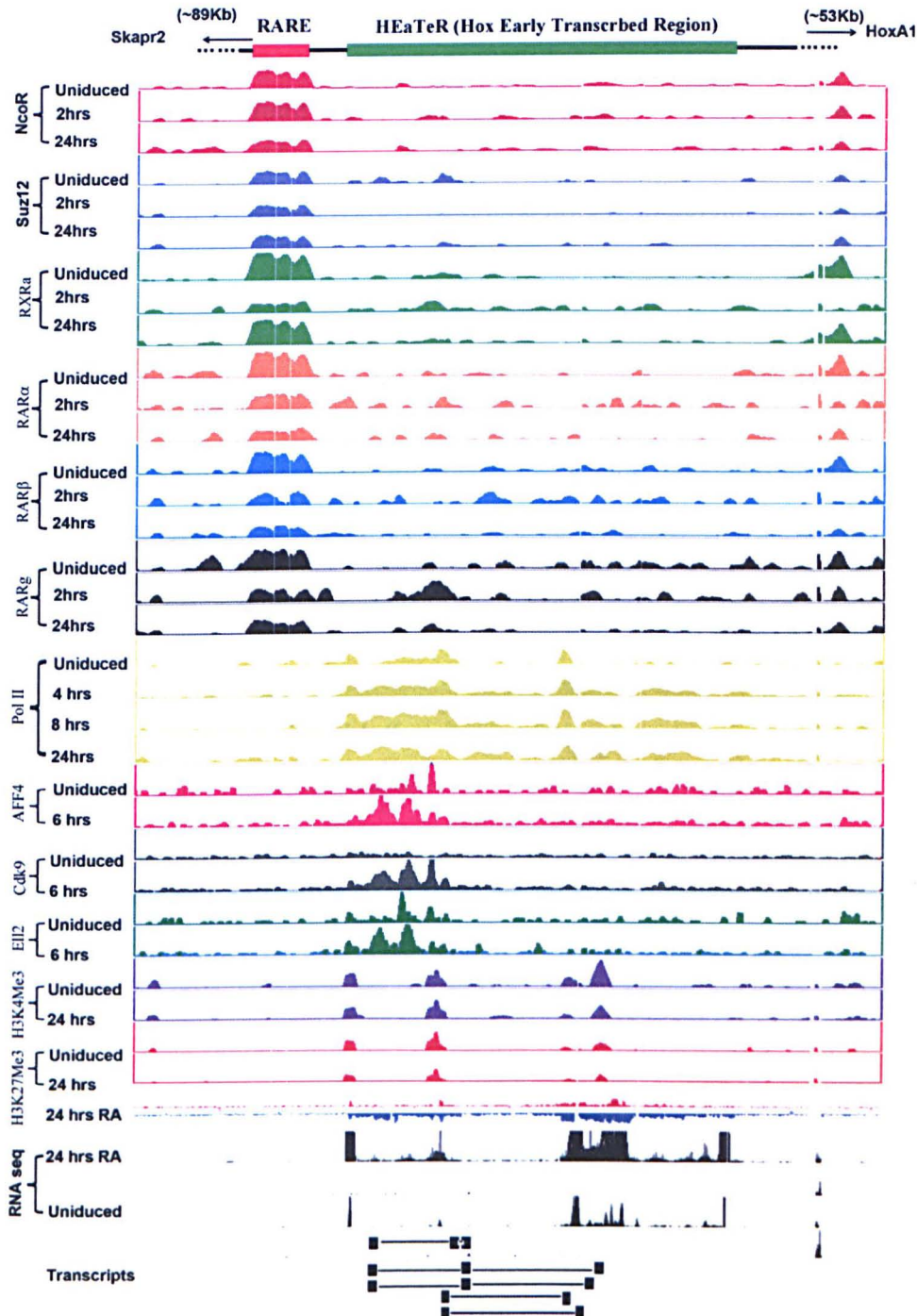


Figure 3-12 Novel non-coding transcripts “heater” identified upstream of HoxA cluster

UCSC genome tracks showing ~ 50 kb upstream of Hoxa1. RARE containing regulatory region is shown in red while non-coding transcribed region is shown in green. Distance for Hoxa1 and Skap2 is also shown in this figure. At the bottom, schematic representation of various non-coding transcripts inferred from RACE and informatics are also shown. Bivalent mark formed by H3K4Me3 and H3K27Me3 is noticeable at TSS of various non-coding transcripts. Gain transcription elongation factors upon RA induction and occupancy of RAR and RXR receptors 1.8Kb upstream is a significant finding. Occupancy of RAR receptors and suz12 emphasizes regulatory potential of RARE containing region. Tracks were configured by using windowing function as mean and smoothing windows as 10 pixels.

A major unexpected feature of this analysis in HoxB was the rapid gain of Pol II occupancy and appearance of the H3K4Me3 mark in region between Hoxb4 and Hoxb5. This is a striking change where the rapid gain of the H3K4Me3 mark is observed after 4hrs of RA treatment. Hoxb4 and Hoxb5 are not as rapidly induced as Hoxb1. In more detailed analyses we noted that this intergenic region which we named as B4-B5 intergenic transcript (B2iT) is transcribed (Fig.3-13). The epigenetic mark and the Pol II occupancy appear to correlate with this non-coding transcript and not those from the Hoxb4 or Hoxb5 coding regions. This illustrates that epigenetic states over the Hox clusters are likely to reflect the entire transcriptional profile and not just the coding regions.

The HoxC cluster does not contain any remarkable occupancy of Pol II or gain of H3K4Me3 marks over this time course. This most likely reflects the modest and slow response of this locus upon RA treatment. However, HoxD which also responds very slowly to RA displays bivalent marks over genes. We can see co-occurrence of the H3K4Me3 and H3K27Me3 near TSS. At 4 hrs and 8 hrs of RA treatment there is a modest gain in the level of H3K4Me3 but no detectable shift in Pol II (Fig.3-9). This seems to suggest that the HoxD cluster is ready for activation in later stages and illustrates the

To explore mechanisms of gene activation in Hox locus mediated during RA induced differentiation, we analyzed dynamic occupancy of the retinoid receptors RAR α , β , γ , RXR α and their associated co-repressor NcoR on Hox complex in uninduced KH2 cells and cells treated with RA for 2hrs and 24hrs. These experiments uncovered several surprising observations.

In the HoxA cluster a large binding region spanning from *Hoxa1* to *Hoxa3* displayed occupancy of all three RARs but not RXR in undifferentiated cells. There also appeared to be a low level of NcoR binding suggesting that RARs are already bound to this region and in the absence of ligand may be recruiting NcoR to repress activity. After 2 hrs of RA treatment there is a rapid change with a loss of occupancy and retention in some more focused regions. Associated with this RXR α begins to show occupancy in this region after 2hrs RA treatment and overlaps with the region of RARs occupancy. This suggests that these are the regions where heterodimeric complexes are being formed to stimulate transcription in the presence of ligand. At 24hrs there begins to be more evidence for occupancy of RAR α and RAR γ over the *Hoxa1-Hoxa3* region. RAR and RXR binding peaks can be seen upstream of the *Heater* region. This region also shows occupancy of Suz12 and NcoR. These patterns of occupancy remain over the time course of RA treatments. This suggests that the induction of transcripts from the *Heater* region might be mediated by the direct action of retinoid receptors.

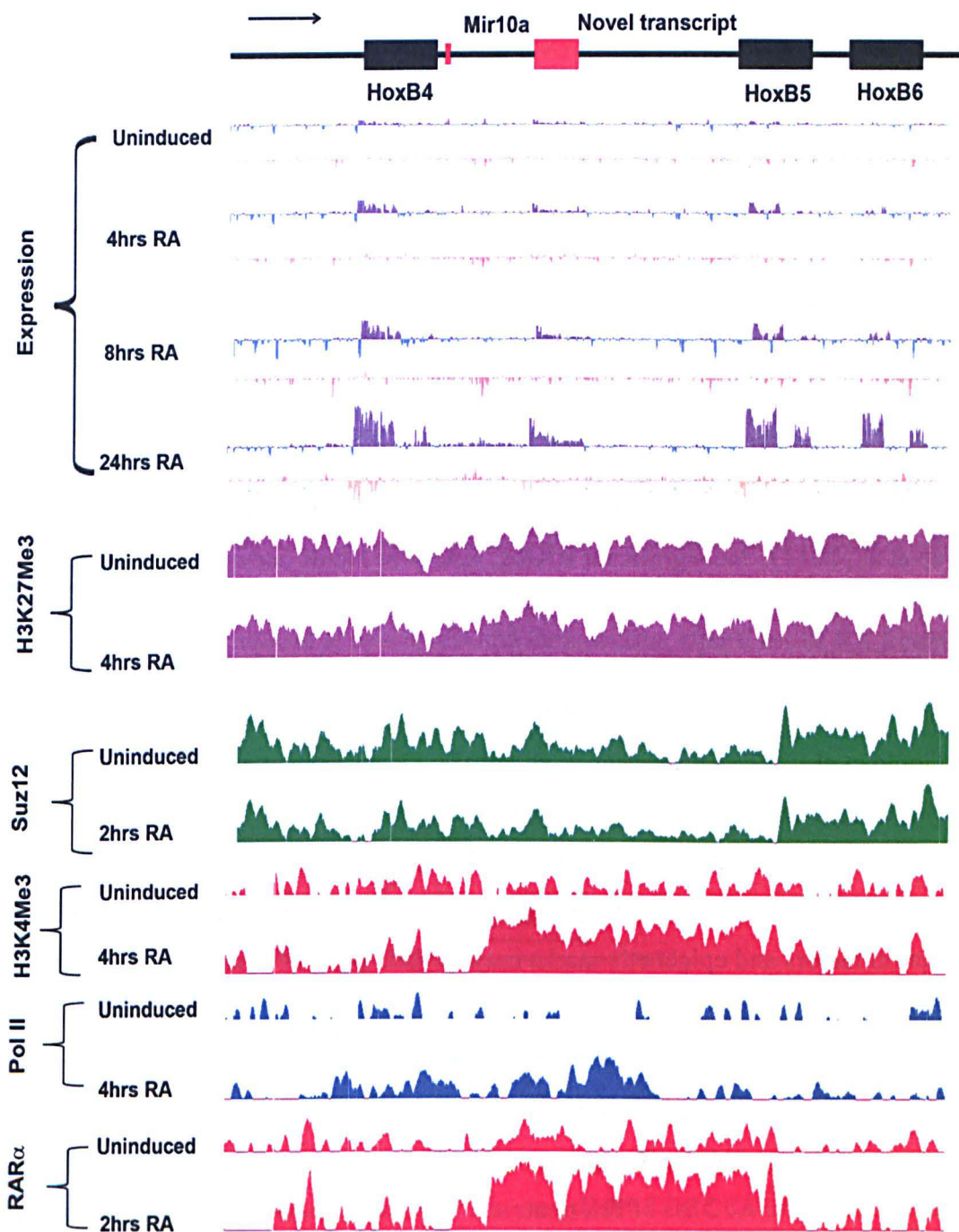


Figure 3-13 Occupancy of RAR receptors and changes in transcriptional and epigenetic properties in region encompassing HoxB4 and HoxB5

UCSC genome tracks showing region between HoxB4 and HoxB5. Expression profile indicates rapid expression of Non coding transcripts between Hoxb4 and Hoxb5 upon RA treatment. Early changes in Histone marks (H3K4Me3 and H3K27Me3), Pol II, suz12 and retinoic acid receptor (RAR alpha) are also shown. Y axes for individual antibodies are same. Tracks were configured by using windowing function as mean and smoothing windows as 10 pixels in UCSC Genome browser. Rapid gain of H3K4Me3 and pol II around this transcript can be seen. Position of Mir10a is also shown in this map.

variations in epigenetic stages when comparing the four different Hox clusters.

The patterns over the HoxB cluster are very different. In this cluster there is a well characterized set of RAREs near *Hoxb1* and we observe occupancy of all three RARs, RXR α and NcoR in a focused manner over this region in undifferentiated ES cells. Upon RA treatment, in contrast to what observations over *Hoxa1-Hoxa3*, there is very little change in the occupancy profiles. There does appear to be an increase in RAR binding.

As seen with the H3K4Me3 mark, a striking change was observed in the region between *Hoxb4* and *Hoxb5*. This region is devoid of RAR occupancy in uninduced KH2 ES cells, but rapidly gains RAR α after 2hrs of RA treatment and there are no changes seen the other RAR receptors over this region. The RAR α occupancy is lost after 24hrs of RA treatment and is replaced with RAR γ . This type of RAR α -RAR γ switch has been observed in the differential utilization of retinoid receptors during early versus late stages of embryonic development and ES cells differentiation (Gillespie and Gudas, 2007; Kashyap et al., 2011). No significant RXR α , RAR β or NcoR occupancy is detected in this region. The specific binding of RAR α and γ while RAR α -RAR γ switch indicates functional significance of this occupancy. It is noteworthy that this region harbors previously characterized RAREs and long range *cis*-regulatory elements (Gould et al., 1998; Gould et al., 1997; Oosterveen et al., 2003a; Oosterveen et al., 2003b; Sharpe et al., 1998) (Fig.3-8). These regulatory regions may be acting upon the novel non-coding region we have discovered in addition to previously characterized inputs into control of the adjacent *Hoxb4* and *Hoxb5* coding regions. This again highlights the fact that binding of transcription factors and epigenetic marks may have primary roles in regulating non-coding regions within the Hox clusters in addition to or in place of the coding transcripts. This makes it a challenge to definitively link changes in chromatin to specific transcriptional events.

Consistent with the weak response of the HoxC and HoxD clusters to RA there is very little occupancy. At 2 hrs of RA treatment there appears to be a modest occupancy of RAR β , RAR γ and RXR α but this is not sustained or increased at 24hrs (Fig.3-13).

3.1.5 RA Response of Non-coding transcripts

In light of the rapid response of the *Heater* region transcripts generated by RA treatments, the epigenetic changes and the binding profiles of the retinoid receptors it seemed important to look at this region in more detail. I therefore, closely looked at region upstream of many of the *Heater* RNAs with respect to RAR occupancy. A 1.8kb upstream

region displayed the most significant occupancy of RARs and RXR, so I searched this region for the presence of RAREs. Interestingly, I was able to identify two DR1 and one DR2 repeats along with a DR3 and ER6 motif in this region (Fig.3-14). These putative retinoid response elements were consistent with the hypothesis that RA may be directly regulating the activity of selected Heater transcripts and are associated with the rapid activation and appearance of transcripts from this region upstream of HoxA. A recent study has speculated that the *Heater* region may be important for potentiating the response of *Hoxa1* to retinoids (Maamar et al., 2013). To further explore roles of retinoid signaling in induction of *Heater*, *mHotarm* and *B²iT* transcripts, we performed RA gavage on 9.25dpc pregnant mice and harvested embryos after 8hrs. All *Heater*, *mHotarm* and *B²iT* transcripts displayed a response to RA treatment. They show at least a 2 fold change compared to control embryos. The *B²iT* transcript was up-regulated 6 fold. We also quantitated changes in transcript levels of all known Heater transcripts, *mHotarm* and *B²iT* transcripts over the ES cell differentiation time course. The *Heater* transcripts *H1*, *H5* and *H4* are rapidly induced and reach their peak expression level by 12 hrs of RA induction. *H2* and *H8* are rapidly induced at low levels and then decline after 6 hrs of RA treatment. The *H3* transcript is induced at a moderate level and maintains this throughout the time course.

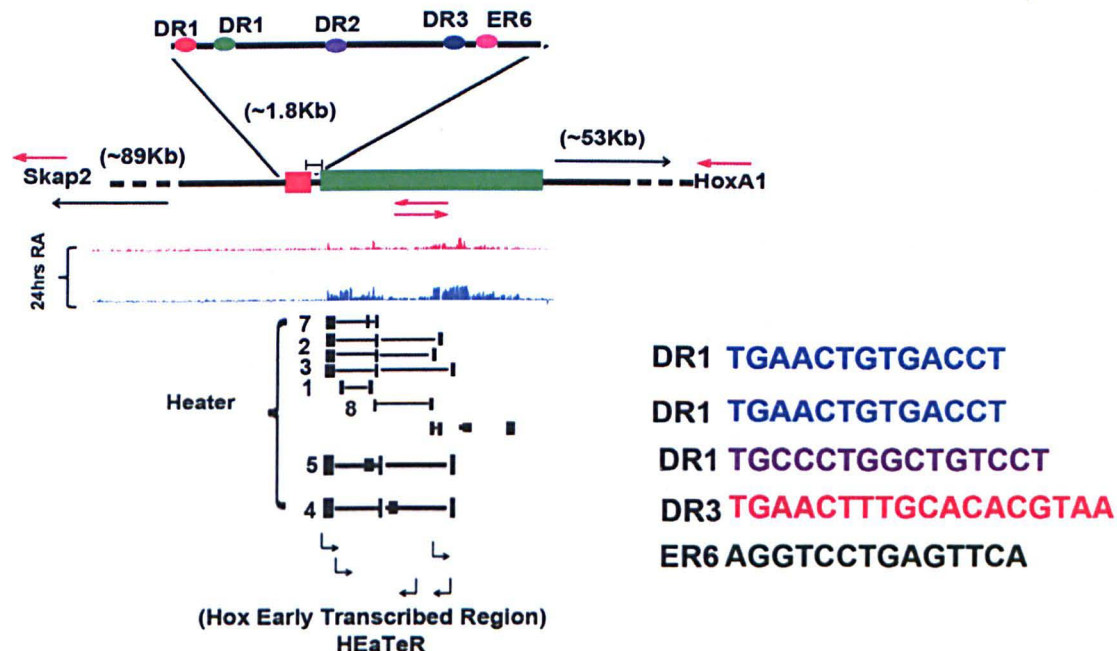


Figure 3-14 RARE upstream of Heater transcripts region

Region marked in Red indicates RARE harboring region 1.8 kb upstream of heaters. DRs are marked in different colors and their sequences are shown at the left hand side panel. Various transcripts and its putative directions are also shown in the figure.

Several of the transcripts show cyclic expression patterns whereby they rise and fall upon longer RA treatment. We quantitated the level of *mHotarm* and see that initially in ES cell there is already a low level of expression. The transcript is rapidly down regulated in first 4 hrs of RA treatment and then after 6 hrs there is a sharp up-regulation. It reaches maximum expression at 36 hrs of RA treatment. The *B²iT* transcript is slowly Up-regulated upon RA treatment and obtains a high level of expression in a window between 24 to 48 hrs of RA treatment. This data indicates that all these intergenic and non-coding transcripts have distinct expression profiles and show independent regulation at quantitative and qualitative levels (Fig.3-15, Fig.3-16 and Fig.3-17). This illustrates the dynamic nature of the non-coding transcriptional profiles of the Hox clusters and the potential it may have in shaping or responding to epigenetic modifications of chromatin in this region.

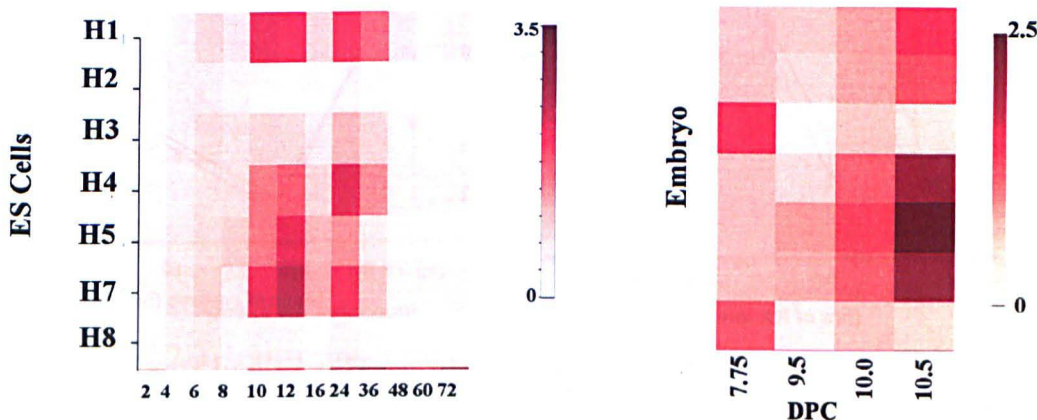
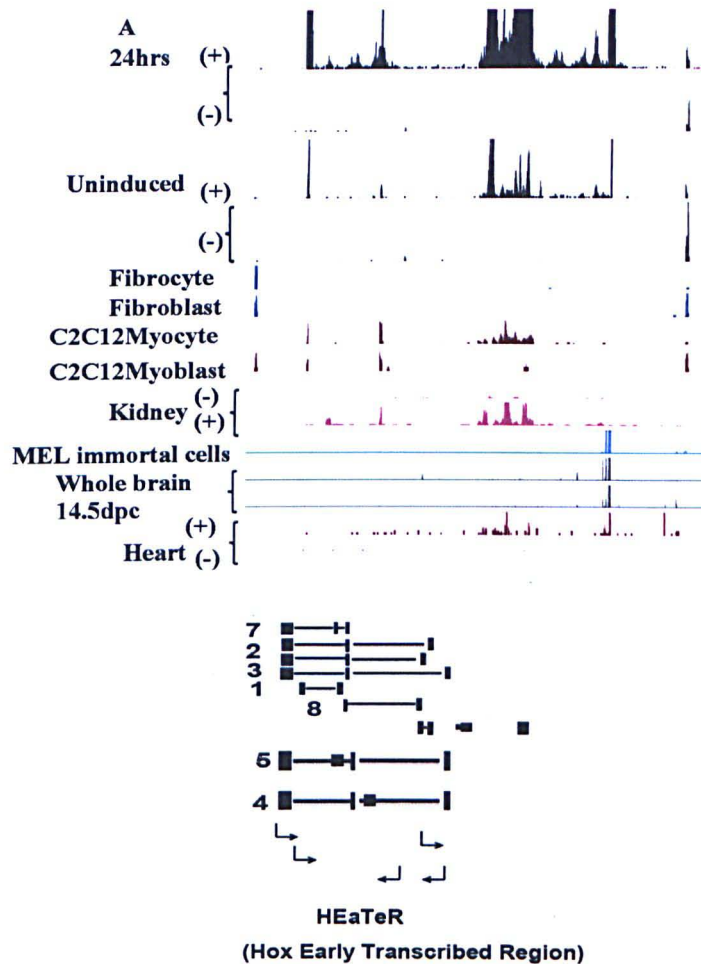


Figure 3-15 Heater – An extensively transcribed region with non-coding transcripts
A. UCSC genome track showing transcription upstream of *Hoxa1* and schematic representation of Heater transcripts. Evidence of Heater transcripts in various tissues such as fibroblast, C2C12 myocyte, C2C12 Myoblast, Kidney, MEL immortal cells, whole brain, 14.5dpc embryos from ENCODE data is shown in the tracks. **B & C** Quantitative changes in Heater transcripts during RA induced differentiation of ES Cells in developing embryos are shown as heatmap. Each row represents one transcript while column represents either length of RA treatment or dpc stage of embryos. Scaled Ct values were normalized against Ct values of *Gapdh* and *Atp5b*. All data points are average of at least two biological and two technical replicates.

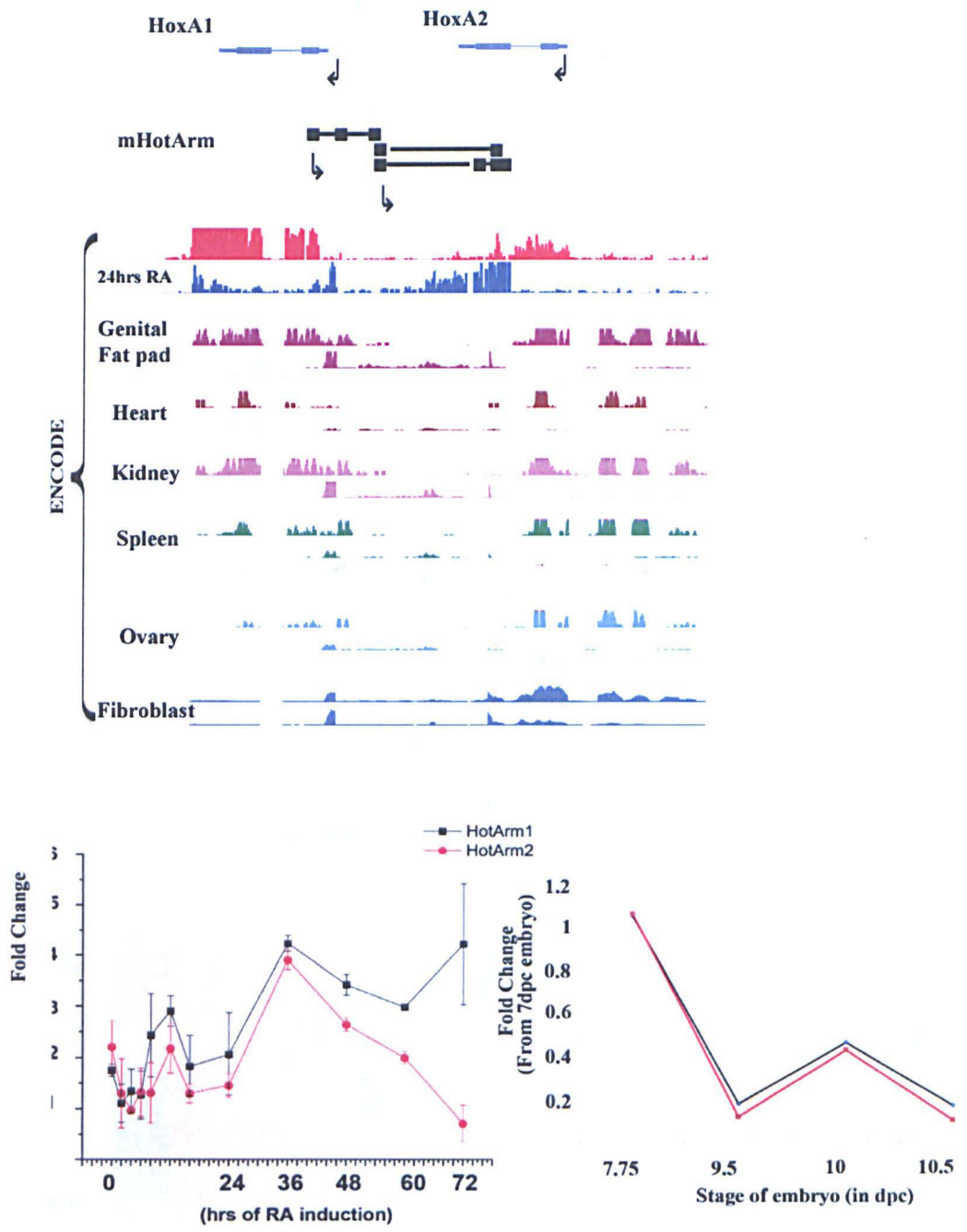


Figure 3-16 mHOT arm- non-coding transcripts between HoxA1 and HoxA2
A. UCSC genome track showing transcription in the region between Hoxa1 and Hoxa2. Schematic representation of various isoforms of non-coding transcript mHotarm is also shown. Evidence of mHotArm transcripts in various tissues such as Genital pad, Heart, Kidney, Spleen, ovary and fibroblast from ENCODE data is shown in the tracks. **B & C** Quantitative changes in two isoforms of Hotarm transcripts during RA induced differentiation of ES Cells (**B**) and in developing embryos (**C**) are shown. Y axis represents fold change compared to uninduced ES Cells (in case of differentiating ES Cells) and 7dpc embryonic expression (in case of developing embryos). Scaled Ct values were normalized against Ct values of *Gapdh* and *Atp5b*. All data points are average of at least two biological and two technical replicates.

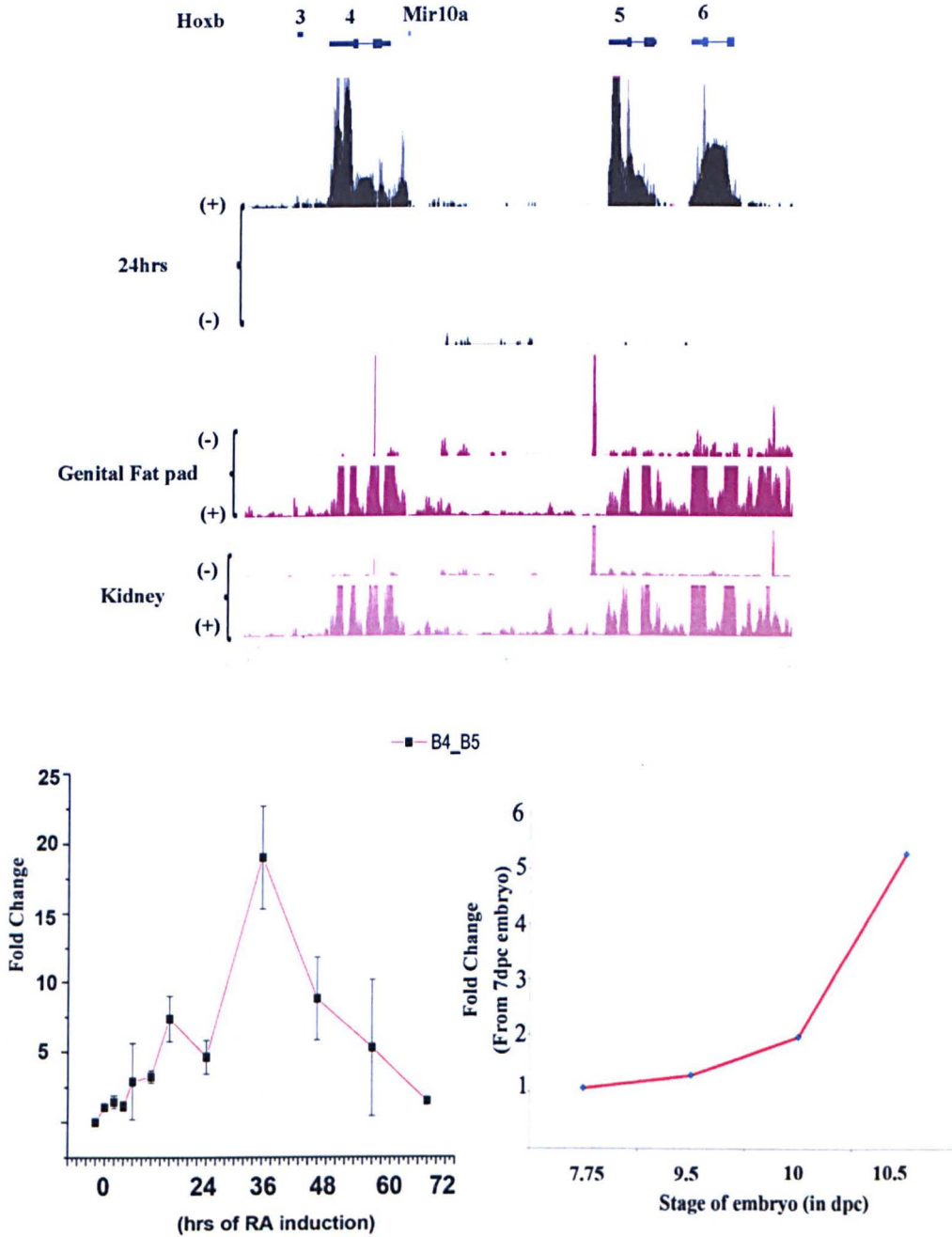


Figure 3-17 B2iT transcript- rapidly induced non-coding transcripts between HoxB4 and HoxB5

A. UCSC genome track showing transcription in the region between Hoxb4 and Hoxa5. Evidence of B2iT transcripts in various tissues such as Genital pad and Kidney from ENCODE data is shown in the tracks. B & C Quantitative changes in of B2iT transcripts during RA induced differentiation of ES Cells (B) and in developing embryos (C) are shown. Y axis represents fold change compared to uninduced ES Cells (in case of differentiating ES Cells) and 7dpc embryonic expression (in case of developing embryos). Scaled Ct values were normalized against Ct values of *Gapdh* and *Atp5b*. All data points are average of at least two biological and two technical replicates.

3.2 Discussion

Comparing the progressive changes in gene expression profiles it seems that there is an important transition point after 24hrs of RA treatment. This might be related to acquisition of a neural fate by 24 hours of RA treatment (Fig.3-18). This is illustrated by the shift in enriched GO terms after 24hrs of RA treatment. In early stages of treatments up to 24hrs, most GO terms are related to patterning. However, after 24hrs there is a shift to enrichment for differentiation and specific components associated with nervous system development. This is consistent with the idea that the differentiation process moves through steps that general progenitor populations for the expansion of lineage components and then elaboration of the structural components of the lineages forming a biphasic process (Fig.3-19, Fig.3-20). Go term analysis illustrate this process in a magnificent way during RA induced differentiation. Gene expression changes after 2-8 hours of RA treatment is mostly related to Pattern specification, metabolism and regulation of transcription while after 12-24 hours of RA treatment gene expression shifts to genes involved in regionalization, embryonic morphogenesis and morphogenesis of anatomical structure. This indicates that differentiation starts with rapid changes in metabolism and regulation of transcription followed by changes in genes involved in early specification of pattern., Go terms indicate that after 36-72 hours of RA treatment cells start taking definitive neuronal identity.

This analysis of the global gene expression changes during RA induced differentiation reveals that diverse mechanisms appear to be functioning in mediating the most rapid responses to RA treatment. Some of the fastest responding genes: *Hoxa1*, *Hoxb1* and *Hoxa5*, RA metabolizing enzyme *Cyp26A1* and RAR B appear to be induced by modulating mobilization of paused polymerase or the rapid induction of initiation of Poll II. Many of the rapid response genes have RAREs suggesting that these modes of induction are achieved through direct inputs on the RARE elements.

Our analysis has shown that there is a dynamic change in the profiles of occupancy of RARs over the HoxA cluster and that there are different patterns of changes over the HoxB cluster. Kashyap and coworkers have shown that RAR γ is a key mediator in RA signaling and responsible for activation of HoxA and HoxB cluster genes in cell culture (Kashyap et al., 2011). RA treatment leads to epigenetic reorganization of Hox clusters. The H3K27Me3 repressive mark is erased but this does not precede acquisition of active H3K4Me3 marks. Complete removal of the repressive marks takes a longer time suggests

that this process is uncoupled or not directly dependent upon activation by the mammalian MLL histone methyl transferases. It seems that generation of specific chromatin signatures and the presence of bivalent marks are not related to the switch in retinoid receptor occupancy patterns where RAR γ is preferred in later time points. In our studies and other published work the occupancy of Pol II shows some initial dramatic changes upon RA induction and then shows moderate increases and a spreading of occupancy as time progress. Active marks are gained more rapidly than repressive marks are lost.

We along with Lin and coworkers together showed that co-occupancy of AFF4 and ELL2 component are correlated with high level of gene expression in ES cells (Lin et al., 2011). This suggests that SEC (Super elongation complex) is frequently associated with high level of gene expression. The SEC complex shown to contains ELL1-3, EAF1-2, P-TEFb, AFF1, AFF4, AF9 and ENL. In the analysis of Hox clusters we were able to see bivalent marks with Pol II co-occupancy only on *Hoxa1*, *Hoxa3*, *Hoxa4* and *Hoxa7* while no *Hoxb* gene promoters showed this configuration. While *Hoxa1* and *Hoxb1* are the most rapidly induced genes upon RA induction of ES cells, Pol II seems to be paused only on the *Hoxa1* promoter in uninduced ES cells. AFF4 and ELL2 were seen to be bound to *Hoxa1* promoter but not on *Hoxb1*. Comparison of the initial induction rates of *Hoxa1* and *Hoxb1* indicates that *Hoxa1* is more rapid and correlates with the paused Pol II. Cdk9 is bound to both the *Hoxa1* and *Hoxb1* promoters, while p-TEFb is only occupied on *Hoxa1*. The presence of p-TEFb may then account for potentiation the rapid induction. RNAi of *Ell2* leads to a reduction in *Hoxa1* induction. Loss of ELL2 leads to loss of Pol II occupancy on the *Hoxa1* gene body without a reduction in promoter.

In contrast the gene body and promoter of *Hoxb1* show loss of PolII occupancy upon loss of ELL2. This suggests that *Hoxa1* might be a rapid direct target of SEC, while *Hoxb1* is initially indirectly regulated through cross-regulatory mechanisms using inputs from *Hoxa1* and later directly regulated through enhanced initiation and elongation. The genome-wide analysis identified 37 rapidly induced genes which contain paused Pol II.

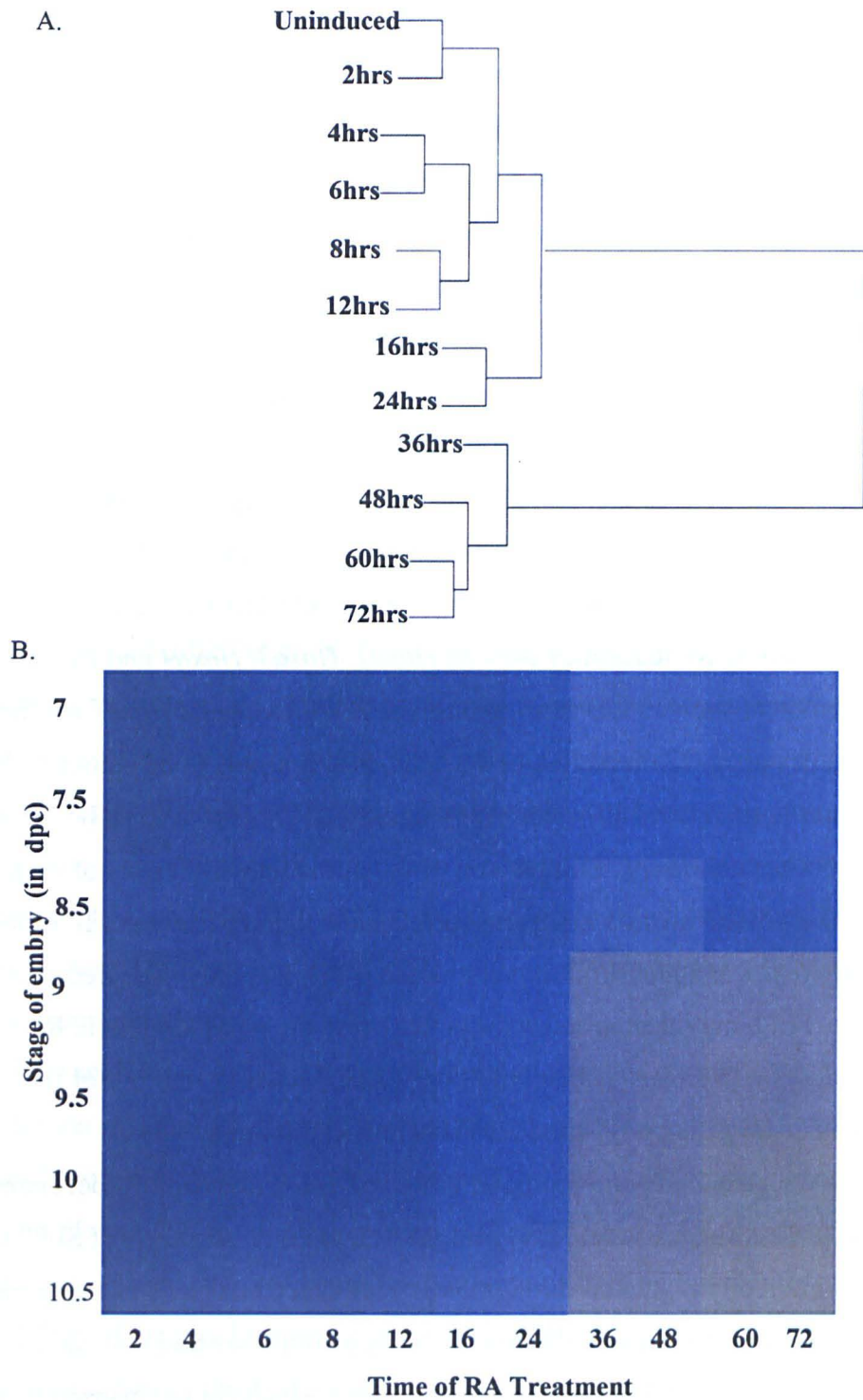
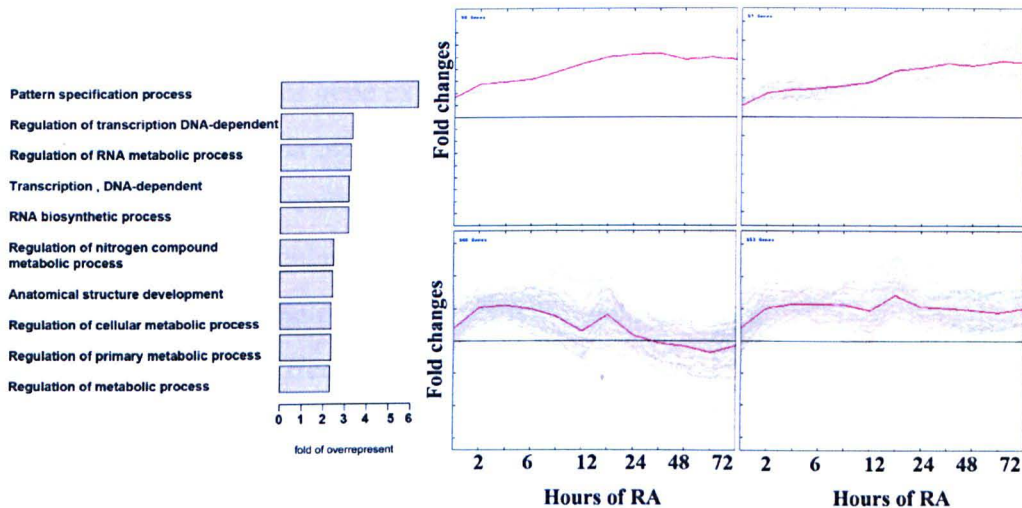
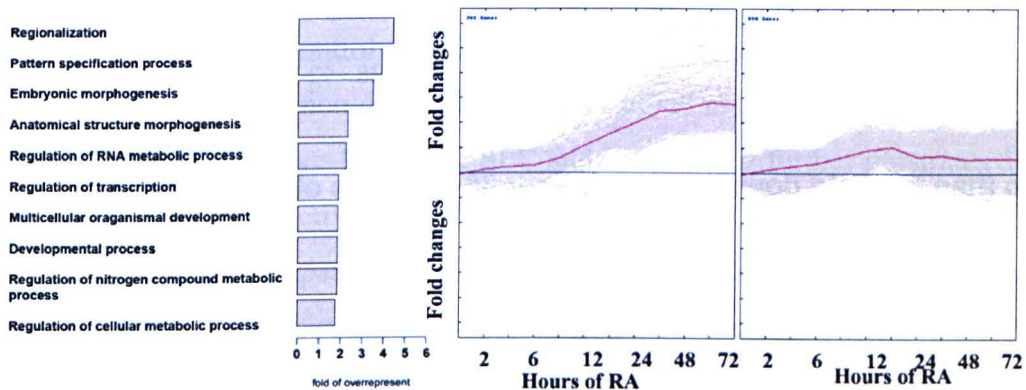


Figure 3-18 Clustering of differentiation time course based on gene expression profile
A. Dendrogram showing clustering of various RA treatment time points based on gene expression profile obtained through affymetrix. Two major clads can be seen in the figure representing change in cellular character after 24 hours of RA treatment. **B.** Correlation matrix shown as heatmap comparing embryonic expression with ES differentiation time course. It is evident that uninduced to 24hrs differentiation profiles are similar to each other while gene expression changes drastically after 24hours.

More than 2 fold upregulation at 2-8 hours of RA Induction



More than 2 fold upregulation between 12-24 hours of RA treatment



More Than 2 fold upregulation at 36-72 hours of RA treatment

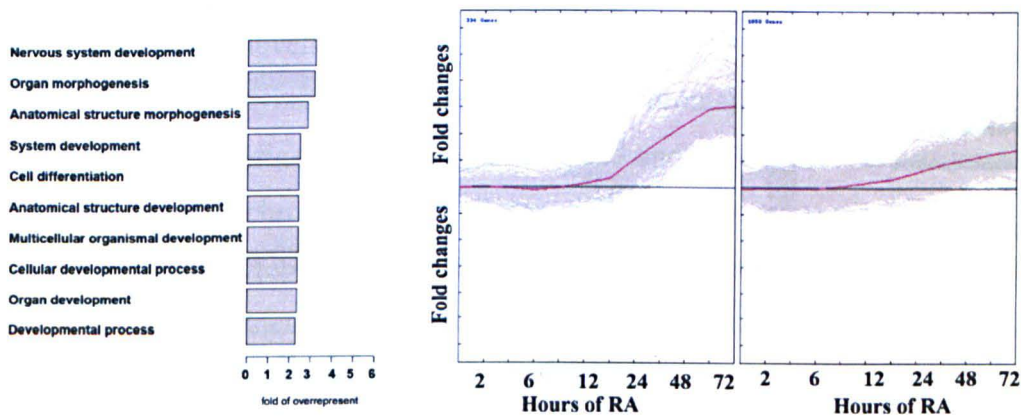
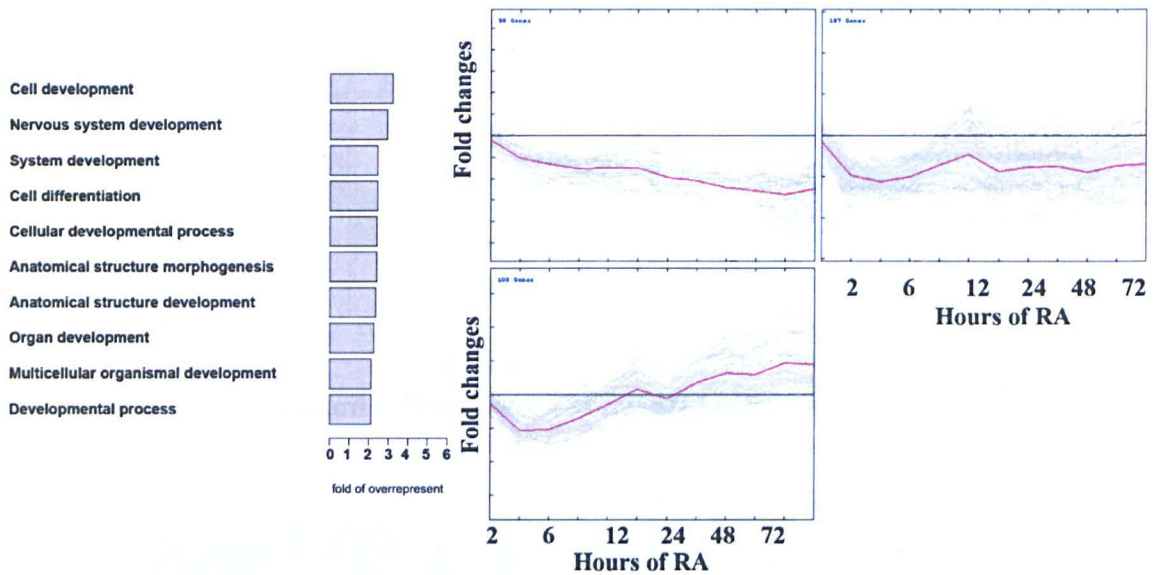


Figure 3-19 Early and late up regulated genes during RA induced differentiation of ES Cells

Enriched GO terms and general gene expression profile is shown for each group. Differentiation time course is binned into three categories namely up regulated by 2-8 hours of RA treatment, 12-24 hours of RA treatment and 36-72 hours of RA treatment. General expression profile of each category of genes is shown as line graph. Pink line represents general expression pattern while each gray line represents expression profile of individual genes in the bin. GO terms of all genes in a particular category is shown in right hand side panel.

More than 2 fold downregulation at 2-8 hours of RA treatment



More than two fold downregulation after 36-72 hours of RA treatment

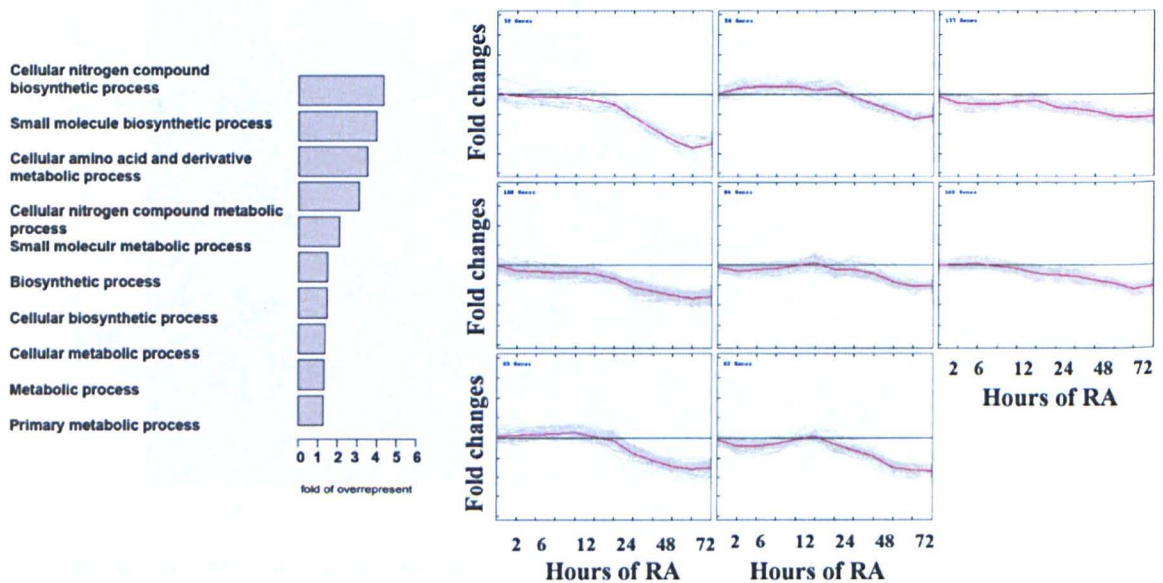


Figure 3-20 Early and late down regulated genes during RA induced differentiation of ES Cells

Enriched GO terms and general gene expression profile is shown for each group. Differentiation time course is binned into two categories namely down regulated by 2-8 hours of RA treatment, and 36-72 hours of RA treatment. General expression profile of each category of genes is shown as line graph. Pink line represents general expression pattern while each gray line represents expression profile of individual genes in the bin. GO terms of all genes in a particular category is shown in right hand side panel.

Many of these contain pre-loaded SEC in their promoters, so the presence of paused Pol II and SEC is a good correlative indicator for the rapid response of genes to differentiating signals. This was found to be true for serum induced HCT-116 human cells. The rapid and high level of gene expression of *Cyp26a1* surprisingly is not associated with paused Pol II but it recruits SEC and new initiation in an unusually rapid manner to facilitate coordinated and controlled induction (Lin et al., 2011). This implies that there may be a variety of molecular mechanisms through which promoters can mobilize bound polymerase or recruit and elongate new polymerase complexes.

We identified the *Heater* region upstream of *Hoxa1* and showed that it produced a number of different transcripts from both strands. This region is very rapidly transcribed in response to RA. Like *Hoxa1*, we can see occupancy of the SEC complex over this region. In uninduced cells, there is pre-loading of Pol II, AFF4 and Ell2 which rapidly increase occupancy upon RA treatment. The increase in Cdk9 occupancy upon RA treatment is consistent with a role for control of elongation in modulating transcription in this region. The two other region with non-coding transcripts studied in detail in this work, *B²IT* and *mHotarm*, do not display pre-loading of SEC and Poll II. This further illustrates differences control of transcriptional activation and induction of non-coding RNAs.

In mouse ES cells, paused Pol II has found to be associated with higher levels of H3K4Me3. It is narrowly distributed on poised genes but shows the same level of occupancy as highly transcribed genes. High levels of H3K4Me3 on these promoters cannot be attributed to CpG islands since even promoters without CpG islands show significant H3K4Me3 levels.

We analyzed the epigenetic status or chromatin state in differentiating ES cells using two epigenetic marks namely H3K4Me3 and H3K27Me3. It was possible to identify nine different chromatin states during differentiation using these two epigenetic modifications. Group I represents genes with no H3K4Me3 marks either on their promoter or upstream region. Posterior HoxC and HoxD genes come under this group. Group II genes have high H3K27Me3 and a low H3K4Me3. Other posterior Hox genes are in this group. Group III and Group IV have moderate H3K4Me3 and H3K27Me3. *Hoxa2*, *Hoxa3*, *Hox6*, *Hoxb4*, *Hoxb5* and *Hoxb6* are in this group. Group V genes are basically bivalent genes in uninduced ES cells and after 36 hours of RA treatment losses most of its H3K27Me3 marks. Group VI and group VII contains low

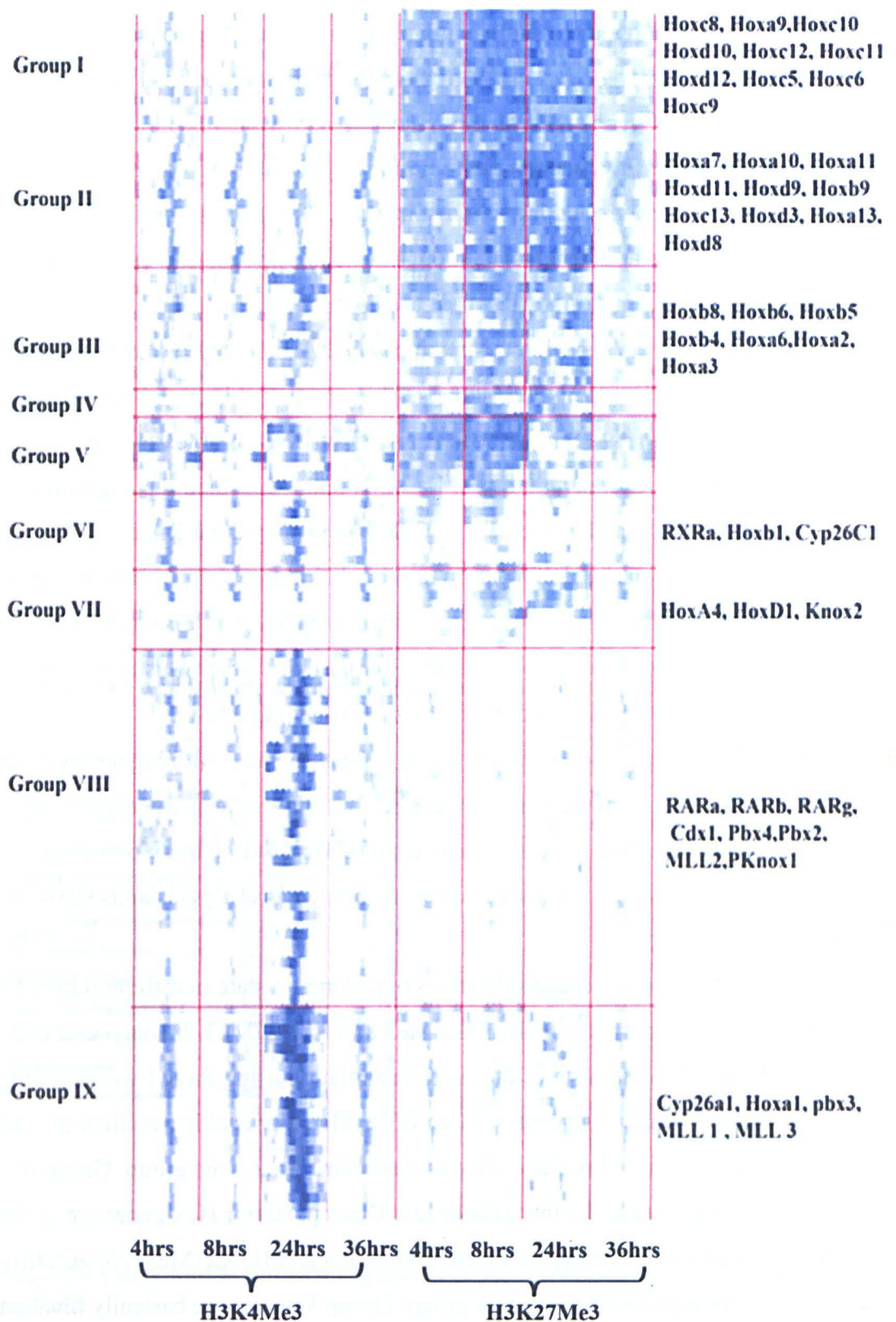


Figure 3-21 Dynamic chromatin state based on H3K4Me3 and H3K27Me3 in differentiating ES Cells Various Hox and Non hox genes are clustered based on H3K4Me3 and H3K27Me3 marks around + 500bp around TSS during ES Cell differentiation. Each row represents one gene while each column represents a time point and antibody. 9 distinct states are seen in Hox genes and its cofactors. Genes in each state are shown on left hand side.

H3K27Me3 with moderate to high H3K4Me3. *RARa*, *Hoxb1*, *Cyp26c1*, *Hoxa4* and *Krox20* show this class of chromatin states. Group VIII and Group IX genes have very strong H3K4Me3 with no H3K27Me3. This group represents some of the genes which are rapidly induced upon RA induction and consists of *RAR alpha*, *beta* and *gamma*, *MLL1-3*, *Pbx1- 3*, *Cyp26a1* and *Hoxa1* (Fig.3-21). The co-occurrence of these marks represents a bivalent state or balanced state.

In *Drosophila*, the balanced state is defined as co-occupancy of the repressive polycomb mark (H3K27me3) with paused Pol II while in mammals the bivalent domain is defined as co-occurrence of H3K27Me3 and H3K4Me3 near the TSS of a gene. In collaboration with the Zeitlinger lab combining our ES cells data with their analysis in *Drosophila*, we found that there is predictive value for future gene expression where there is co-occupancy of poised Pol II and H3K4Me3 (Gaertner et al., 2012). Poised Pol II shows high predictive values in a stage specific manner while H3K4Me3 shows high predictive value throughout differentiation of ES cells. In combination with the H3K27Me3 mark these properties predict a potential for late gene expression. This leads to a hypothesis that the bivalent and balanced states are related (Gaertner et al., 2012) (Fig.3-22).

The removal of the H3K27Me3 marks presents a paradox. The rate of removal of the repressive mark is slower over Hox complexes compared with the activation of coding and non-coding transcripts or the appearance of the activation marks. A closer look at promoter regions of these transcripts indicates that the rate of removal of the H3K27Me3 mark over the promoter is faster in general than over the gene body. This raises possibility that removal of H3K27Me3 at promoters might be done through active participation of trithorax and demethylases while over cluster these marks are removed passively during transcription. Some non-coding RNAs studied in this work also seem to have a bivalent state. Co-occupancy of H4K4Me3 and H3K27Me3 marks can be seen in *Heater* region. These bivalent states appear in uninduced ES cells and persist upon RA induction. Interestingly this region does not show any occupancy of Suz12. In contrast an enhancer region which contains RAREs 1.8 kb upstream of *Heater* shows co-occupancy of Suz12 with RARs and RXR. This indicates that bivalent state can prepare long non coding transcripts also for rapid induction in differentiating ES cells. In contrast, the other two non-coding RNAs studied in detail in this work are devoid of bivalent state in uninduced ES cells.

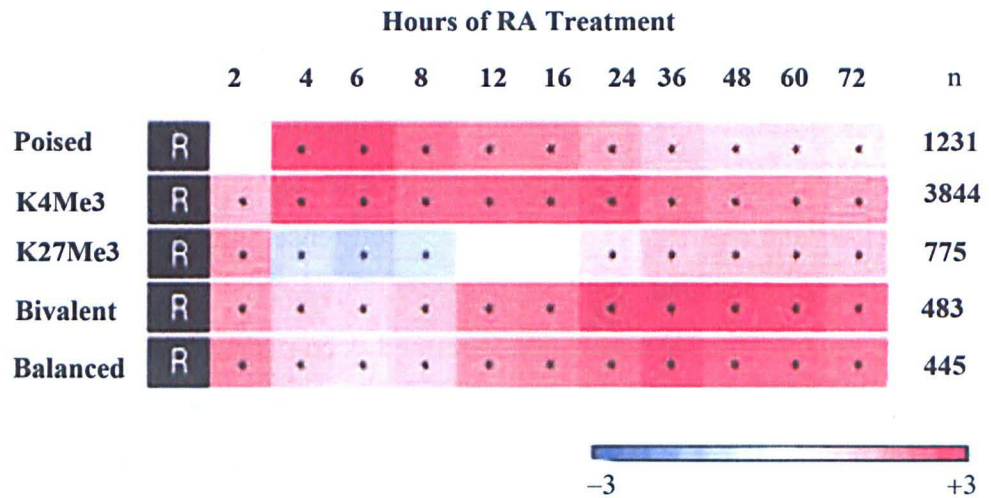


Figure 3-22 Relative predictive values

Relative predictive values of bivalent and balanced state are shown as heatmap. Relative predictive value is obtained as ratio expression of top 30% highest expressed genes and 30% least expressed genes with a particular histone mark or state. Positive value indicates great predictive values while negative values shows independence. Poised genes are expressed rapidly while balanced state predicts well with later time point of RA Induction. From Gaertner et al, 2012 Cell reports 1670-1683.

Another Interesting aspect to come out of this study is the RA response of non-coding RNAs. We have studied three different regions with non-coding transcription potential in detail. We found that these transcripts respond rapidly to RA treatment in embryos as well as in differentiating ES cells. Characterizing the basis for the RA response of *mHotarm* and *B²IT* will require further analysis. However, we identified a 2.5 kb region with multiple RAREs in a 1.8kb upstream of *Heater* region. This region is occupied by RARs, RXRs and NCOR. Recently Maamar and colleagues have shown that knockdown of at least three transcripts from this region can lead to increased *Hoxa1* levels in uninduced ES cells (Maamar et al., 2013). RA treatment of ES cells tends to alter this relationship. My analysis reveals that the transcription of the Heater region is much more complicated than appreciated by Maamar and colleagues and provides a basis for understanding how RA is able to trigger the non-coding RNAs. Direct stimulation by retinoids in combination with induction by stimulation of elongation on promoters with paused polymerase can generate a rapid response which may be essential for their putative role in regulation of *Hoxa1*. This further emphasizes the role of RA signaling in function of non-coding transcripts from this region.

Chapter 4 Characterization of genome-wide binding of *Hoxb1* in differentiating ES cells

Among 39 Hox genes in mice, Hox1 paralogous genes are evolutionarily related to labial gene in Drosophila. In Mice, paralogous group 1 genes are expressed first among Hox genes. *Hoxa1* and *Hoxb1* expressed in hindbrain but *Hoxd1* is absent from developing brain. *Hoxb1* first appears around 7dpc of mouse embryonic development at caudal primitive streak. At 8-8.5dpc, during neurulation ectodermal cells of R4 shows expression of *Hoxb1* which later expands to whole r4 by end of neurulation. High level of *Hoxb1* expression persist in r4 through 11.5dpc and can be seen in r4 neuroepithelium and second branchial arch; populated by r4 derived neural crest cells. In hindbrain, *Hoxb1* expression can be seen in cell bodies and axonal tracts of facial nucleus and cell associated with ventrolateral exit points of VIIth cranial nerve. At 10.5dpc, *Hoxb1* expression can be seen in whole embryonic trunk and limbs which extends to throughout the developing embryo extending till axial level demarcated by r4 expression boundary (Arenkiel et al., 2003; Arenkiel et al., 2004).

Hoxb1 maintains a very restricted r4 expression domain in developing mice embryo from 9.5dpc. R4 restriction of *Hoxb1* is regulated in developing embryos through a complex regulatory maze of regulatory mechanism involving positive induction, auto regulation (maintenance) and negative regulation (prevention of spread to r3 and r5). A conserved 3' RARE elements (DR2 type element) is responsible for establishment of Early *Hoxb1* expression domain in neuroectoderm, mesoderm and primitive streak. This element perceives RA signaling during embryonic development. Under *in vitro* condition, this element can bind RAR/RXR heterodimer (But not homodimers). Auto-regulatory element at 5' of *Hoxb1* containing three similar repeat regions with sites for PBX and Meis binding maintains *Hoxb1* expression in r4. This is achieved through auto-regulation of *Hoxb1* with Cross-regulatory inputs from *Hoxa1* and PBX. This auto-regulatory region is conserved among mouse, chicken and puffer fish. Another 5' DR2 RARE element helps in restricting Expression of *Hoxb1* in r4 through abolishing *Hoxb1* expression in r3 and r5. Like other RARE this also exclusively binds to RAR/RXR heterodimer.

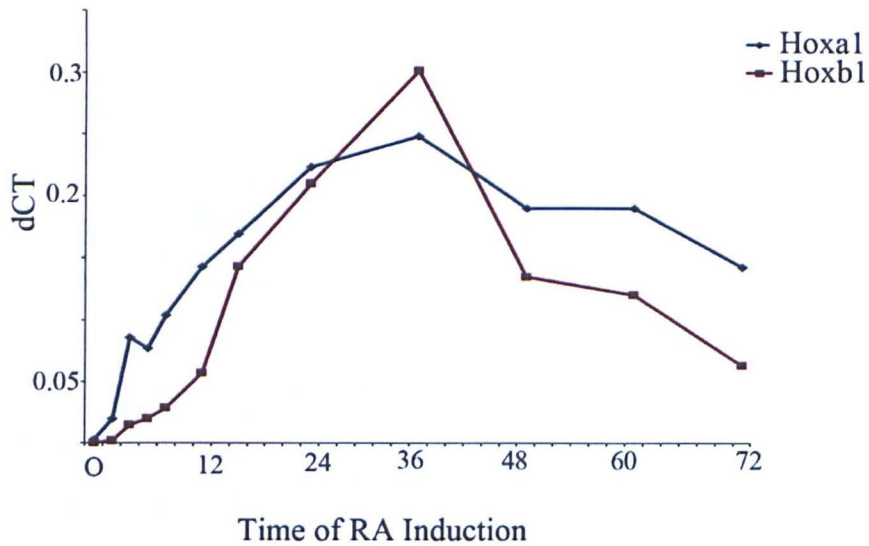


Figure 4-1 Induction of *Hoxa1* and *Hoxb1* in differentiating ES cells upon RA induction All data points are average of three biological and two technical replicates. X-axis indicates time of RA treatment while Y-axis shows non-scaled delta Ct value representing change in transcripts levels compared to housekeeping genes. Non-scaled Ct values were normalized against Ct values of *Gapdh* and *TBP*.

So it is clear that RA plays an important role in induction and restriction of *Hoxb1*. Discussions in previous chapter make it clear that *Hoxb1* along with *Hoxa1* are rapidly induced upon RA treatment of ES cells (Fig.4-1). Bami and Coworkers (Bami et al., 2011) have shown that ectopic expression of *Hoxb1* with RA induction is a sensitive system to identify model hox gene effectors using mouse ES Cell as model system. Their study indicated that *Hoxb1* activity mainly includes regulation of cellular response to retinoic acid signaling. Many other studies (Boylan and Gudas, 1991; Dani et al., 1997; Jones-Villeneuve et al., 1982; Kashyap et al., 2011) along with our studies indicate that RA induced differentiation of ES cells can be used as excellent model to study downstream targets of *Hoxb1*. This system is convenient and more controlled. Further, changing differentiation protocol will give comparable results in different tissue types. ES cell based system have further advantage of availability of large amount of cells and extracts which can be used for ChIP and *in vitro* biochemical studies to understand underlying mechanism governing Hox gene specificity.

In current study, we decided to analyze genome-wide binding property of *Hoxb1* in differentiating mouse ES Cells. We analyzed genome-wide binding to understand nature of over enriched motifs, their putative role and combinatorial binding with Hox cofactors and other Hox genes. We were further interested to analyze changes in expression of genes

having a binding of *Hoxb1* in *Hoxb1* mutant hindbrain. To achieve this objective, we set aside following specific aims

- Generate clones of KH2 cells containing epitope tagged *Hoxb1* validate normal karyotype and Dox induction profile.
- Use Chip-seq with epitope-tagged antibody to identify genome-wide sites for binding of *Hoxb1* at several stages of differentiation.
- Identification and comparative analysis of genome-wide targets of the Hox-cofactors *Pbx* and *Meis*.
- Examine whether *Hoxb1* binding sites are located near genes whose expression change in *Hoxb1* mouse mutants.

4.1 Result

4.1.1 Generation of KH2 ES Cells with Epitope tagged HOXB1

We have generated three different ES cells line with different epitope tagged version of *Hoxb1*. In Collaboration with Mark Parrish, I created three different tagged versions of KH2 cells. *Hoxb1* with His-flag epitope and *Hoxb1* with triple flag-Myc were cloned in pBS31 (Fig.4-2). These constructs were tested for its intact function after adding epitope using chicken embryos. These constructs were Co-electroporated with lacZ driven by minimal beta globin promoter under influence *Hoxb1*ARE . Epitope tagged version of *Hoxb1* is capable of activating reporter system in chicken neural tube which indicates that *Hoxb1* is functional and not compromised by addition of epitopes. KH2 were engineered through lipofection and epitope tagged *Hoxb1* were inserted in Col II locus and put under control of doxycycline inducible promoter. A separate construct with *Hoxb1*-triple flag-Myc was created in pCMS vector (Fig.4-2). Using this construct, KH2 lines were made through random integration. Epitope tagged *Hoxb1* was under the control of CMV promoter in this line. All cell lines were tested for karyotype stability. FACS Calibur was used to analysis of DNA content and to get indirect inference of karyotype stability. Mark Parrish made a transgenic mice with epitope tagged *Hoxb1* (*Hoxb1*-His-Flag) using BAC transgenic technology (Parrish et al., 2011)

We analyzed induction kinetics of epitope tagged *Hoxb1* through combining FACS and immune-staining. Cells were induced with 9-*Cis*-RA (0.033 μ M) and 0.5mg/ml Doxycycline for various length of time. Cells were harvested, fixed and immuno-stained

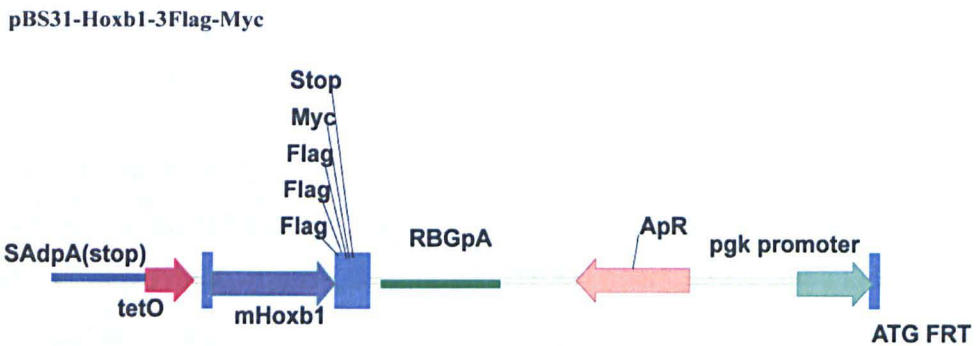
4.1.2 Genome-wide identification of HOXB1 binding sites

I tried Anti-Flag monoclonal and polyclonal antibodies, Ant-Myc antibody and Commercial HOXB1 antibody for ChIP experiments. These results were compared. Myc-

A.



B



C

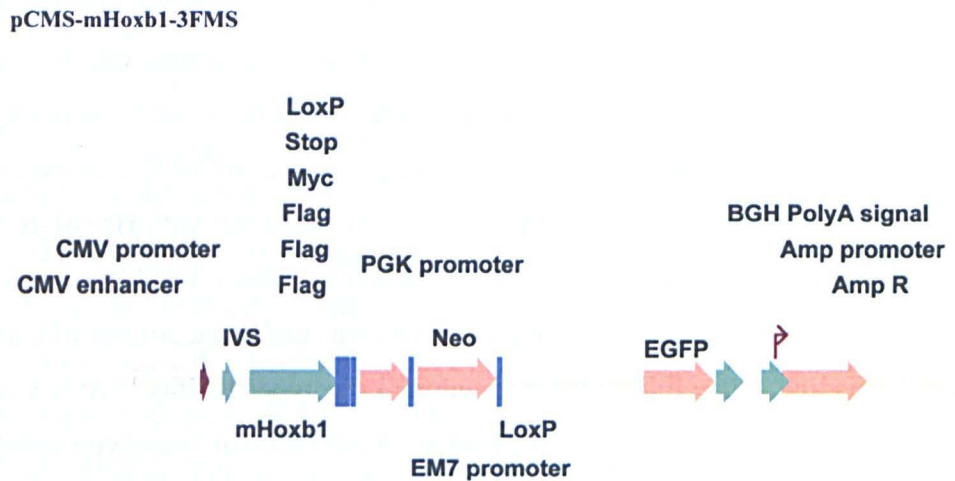


Figure 4-2 Constructs for expression of epitope tagged HOXB1 in KH2 cells

A. Mouse *Hoxb1* cDNA with 6X His-Flag epitope tag is cloned in pBS31 vector backbone using homologous recombination at FRT site. This cDNA comes under control, of tetO and inducible with doxycycline. **B.** Mouse *Hoxb1* cDNA with 3XFlag-Myc epitope tag is cloned in pBS31 vector backbone using homologous recombination at FRT site. This cDNA comes under control of tetO and inducible with doxycycline. **C.** Mouse *Hoxb1* cDNA with 3XFlag-Myc epitope tag is cloned in pCMS vector backbone. *Hoxb1* expression is under control of CMV promoter.

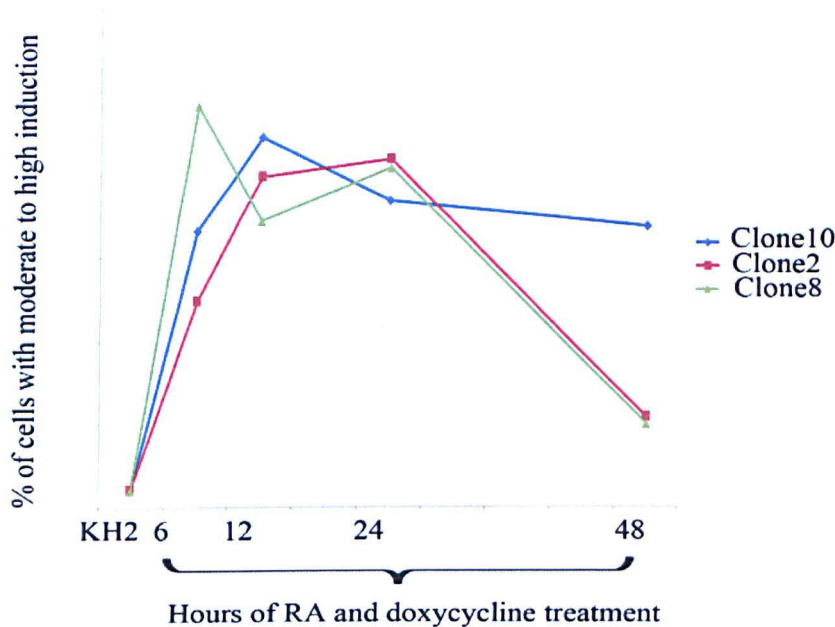


Figure 4-3 Induction of HOXB1 expression doxycycline and RA in three different ES Cell clones with epitope tagged *Hoxb1* in Col II locus Cells were induced with Dox for various length of time and stained with anti-flag antibody. Stained cells were counted using flow cytometer. Cells were binned into three category based on fluorescence intensity and moderate to high intensity cells were quantified and showed in this graph as % of induced cells. Three independent clones *Hoxb1* with epitopes were studied for their induction kinetics.

antibodies were inefficient and difficult to elute bound DNA fragments from beads. I observed that monoclonal anti-flag M2 antibody is less noisy and can be easy used for comparative studies with other Hox genes and cofactors. So, current work discusses genome-wide binding site obtained through two separate ChIP experiments done in two separate cell lines with monoclonal anti-flag M2 antibody. I did each ChIP experiment in triplicate. Finally, we generated peaks set with 200-1000bp width and 5 fold enrichment.

These Peaks sets were analyzed for identification of over represented motifs, enriched K-mers or K-mer pairs and analyzing co-occupancy with other Hox and Cofactor binding.

721 peaks were identified from three time point of RA and doxycycline induced ES Cell ChIP experiments. We analyzed distribution of HOXB1 bound peaks with respect to pre-defined genomic landscape. 19% of peaks were found in exonic region while 28% peaks were in intronic regions of genome. 53% of peaks were intergenic in which 17% of peaks were within 10 Kb regions from TSS (Transcription Start Site). Less than 1% peaks

were found within 1Kb of 3'UTR. This distribution indicates that HOXB1 binding is not biased towards any particular region of genome (Fig.4-4).

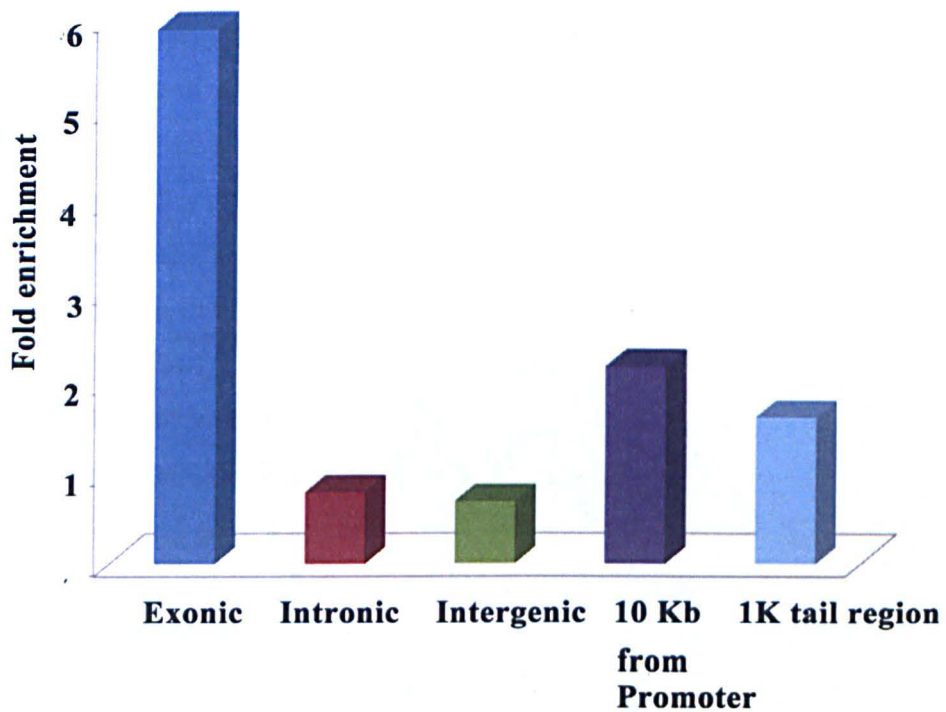
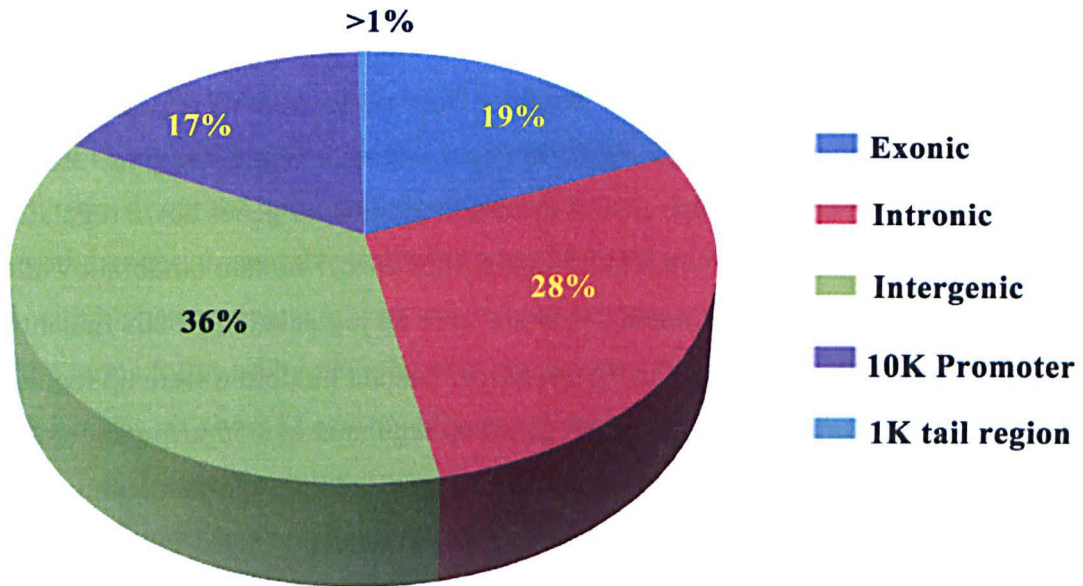


Figure 4-4 Distribution of HOXB1 occupied genomes near pre-defined genomic features Pie chart shows distribution of HOXB1 bound peaks with respect to pre-defined genomic features like Exon, intron, genomic region within 10kb from TSS, within 5' UTR and intergenic region. Left panel shows enrichment of Occurrence of these peaks in a given pre-defined genomic feature. Enrichment is calculated as ratio of % of peaks in a given pre-defined feature to % of that pre-defined feature in the mouse genome.

573 genes were identified as nearest neighbors to these peaks. I analyzed change of expression of these genes in *Hoxb1* and *Hoxa1* mutant from data generated by Marina Yurieva in our lab. 66 genes with nearest neighbor ChIP peaks were up regulated while 86 genes were down regulated in 9.5dpc *Hoxb1* mutant hindbrain. While 430 genes show no change in gene expression in 9.5dpc *Hoxb1* mutant hindbrain. 49 genes down regulated in *Hoxb1* mutant hindbrain were down regulated in 9.5dpc *Hoxa1* mutant hindbrain while 27 genes down regulated in *Hoxb1* mutant hindbrain were up regulated in *Hoxa1* mutant hindbrain. 35 genes down regulated in 9.5dpc *Hoxb1* mutant hindbrain were up regulated in 9.5dpc *Hoxa1* mutant hindbrain while 30 genes up regulated in 9.5dpc mutant hindbrain were down regulated in 9.5dpc *Hoxa1* mutant hindbrain (Fig.4-5). I further compared *Hoxb1* mutant data from (Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006)'s lab and compared with Marina's mutant data. I identified 31 down regulated and 11 up regulated genes consistent in both data sets. These 42 targets seem to be direct target of *Hoxb1* functions (Appendix I-V). This analysis doesn't suggest that *Hoxb1* has only 42 direct targets. Further, comparing our ChIP results with ES cell based differentiation data from Gavalas lab (Gouti and Gavalas, 2008), we identified 84 new possible direct targets of *Hoxb1* gene (Appendix VI).

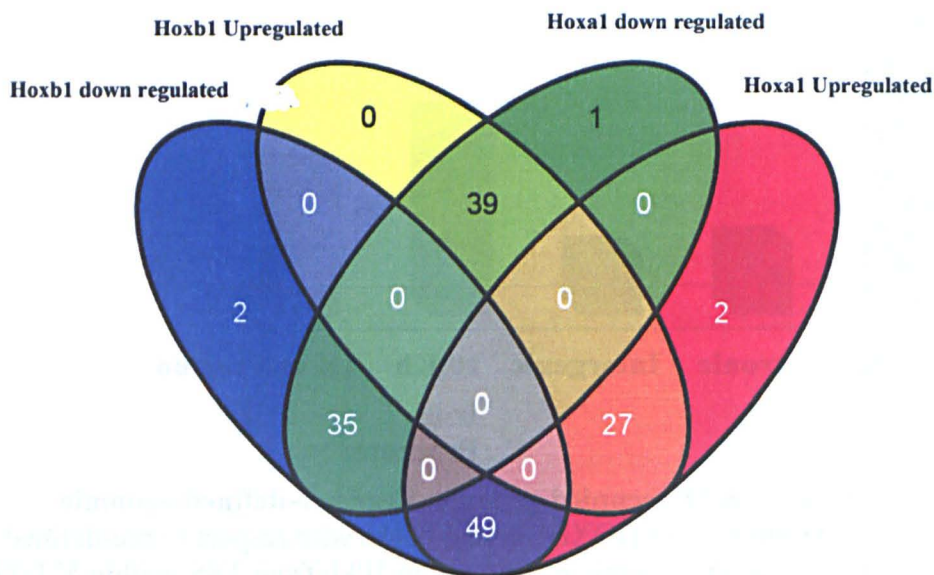


Figure 4-5 Comparison of direct targets of HOXB1 in *Hoxa1* and *Hoxb1* mutant
 Direct targets were identified as genes changed it expression in *Hoxb1* and /or *Hoxa1* mutant embryos with a nearest binding peaks of HOXB1. HOXB1 occupancy was identified from ChIP experiments on 6 hrs, 12hrs and 24hrs RA induced KH2 cell with HOXB1-3XFlag-Myc. Data from Tavrik et al, 2003 and Marina's unpublished work were compared with nearest neighbor gene from HOXB1 occupied region. (Appendix I-V)

We further analyzed enrichment of Gene ontology terms (GO Terms) for nearest neighbor gene. We took stringent cut-off criteria. We restricted FDR at less than 1% and 7% for biological processes and cellular components. P value was restricted at less than e^{-4} . We further discarded terms with less than 10 genes. This was to avoid false chance of getting higher enrichment due to low number of genes in the bin. Biological process GO terms with highest enrichment were related to neurogenesis. Hindbrain was highly over enriched terms followed by neuromuscular processes, Neuron projection, Cell morphogenesis involved in neuron differentiation and axiogenesis. Cation channel complex, Ion channel complex, axon, Post synaptic membrane, neuron projection were Cellular component related GO term enriched in our analysis (Fig.4-6).

I was interested to know, if these genes are part of one or more related pathways? To understand this, I analyzed nearest neighbor genes for their involvement in related pathways using a program called as "String". I identified a network involving 178 genes co-regulated by direct binding of HOXB1 (Fig.4-7). This network has three main hub. Group 1 hub consists of genes like *Kcnq1*, *Dpp6*, *Lrp11*, *Kcnh1* etc. Interestingly *Kcnh1* gene mutation results in reduced first branchial arch and cranio-facial defect as phenotype. *Dpp6* and *Lrp11* are part of *Bmp4* and *wnt* signaling pathway which were known to be target of many Hox genes. Group 2 hub consists of genes such as *Bai3*, *Ppp2r2b*, *atb2b2*, *Slc32a1*, *Chga* etc. *Ppp2r2b* have been already shown as direct target of HOXB1 from genetic studies in Zebra fish. Third major hub was consisting of genes like *Neurod4*, *Adra2a*, *Pdyn*, *Pau4f1*, *Kirb1c*, *Pnoc*, *Nms*, *Giar1*, *Runx1t1* etc. *Neurod4* knock out results in large scale neuronal phenotype including abnormal neuron differentiation. Partial and complete postnatal lethality is observed. I analyzed enrichment of GO terms among these 178 genes forming gene interaction network. Many interesting genes related neuronal development and differentiation were over represented. Most over represented terms were regulation of system process, locomotory behavior, cell migration and Cell-cell signaling.

4.1.3 Dynamics of HOXB1 occupancy in differentiating ES Cells and 9.5dpc embryo

We first identified statistically significant peaks from individual time point. Peak list from these three time point were merged to make a comprehensive Hoxb1 binding sites in differentiating ES Cells. Using this comprehensive coordinates we generate a heat map showing binding dynamics in all three time point (Fig.4-8). 500bp either side from middle of peak were plotted as heat map. Representation as heat map was more informative than numerical comparison of MACS called peaks.

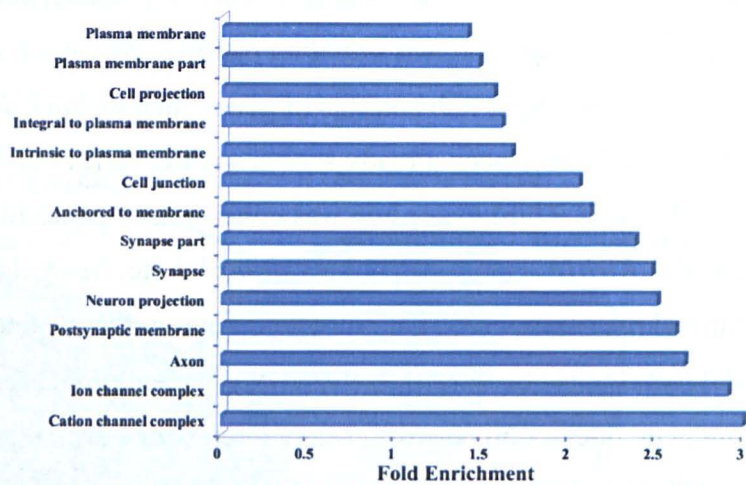
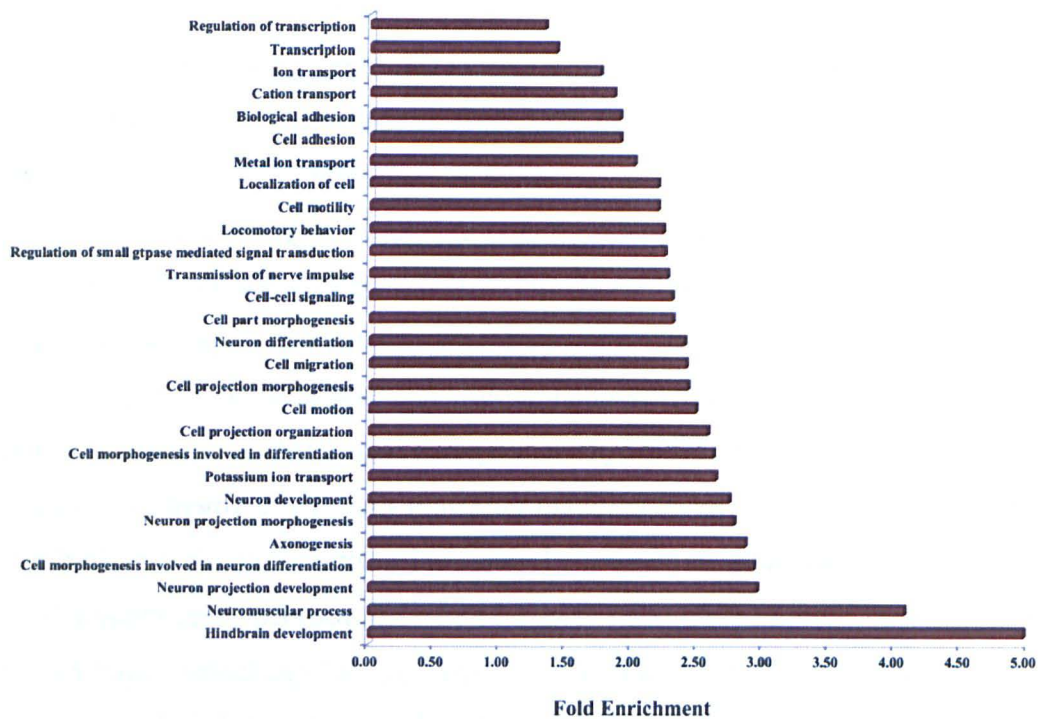


Figure 4-6 Go Term analysis of nearby genes with a HOXB1 bound region.

HOXB1 occupancy was identified from bound peaks obtained in ChIP experiments on 6 hrs, 12hrs and 24hrs RA induced KH2 cell with HOXB1-3XFlag-Myc. Nearest neighbor genes were identified and enrichment of GO (Gene ontology term) were analyzed. **A.** Biological Process. **B.** Cellular Component. Enriched terms were selected with FDR <1 for Biological processes and FDR <7 for Cellular components

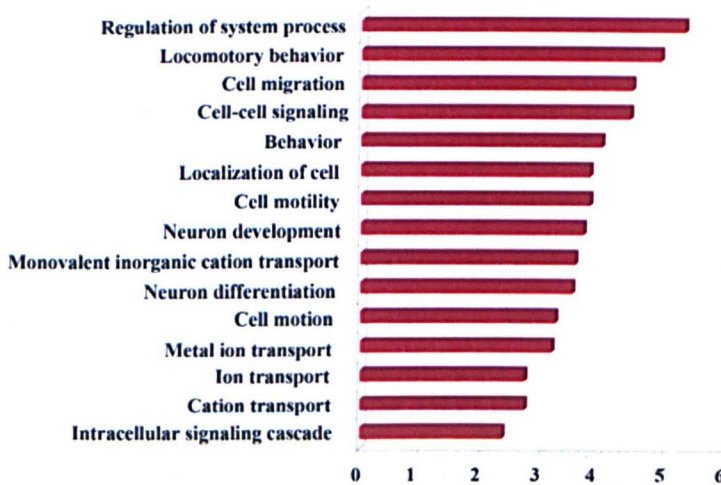
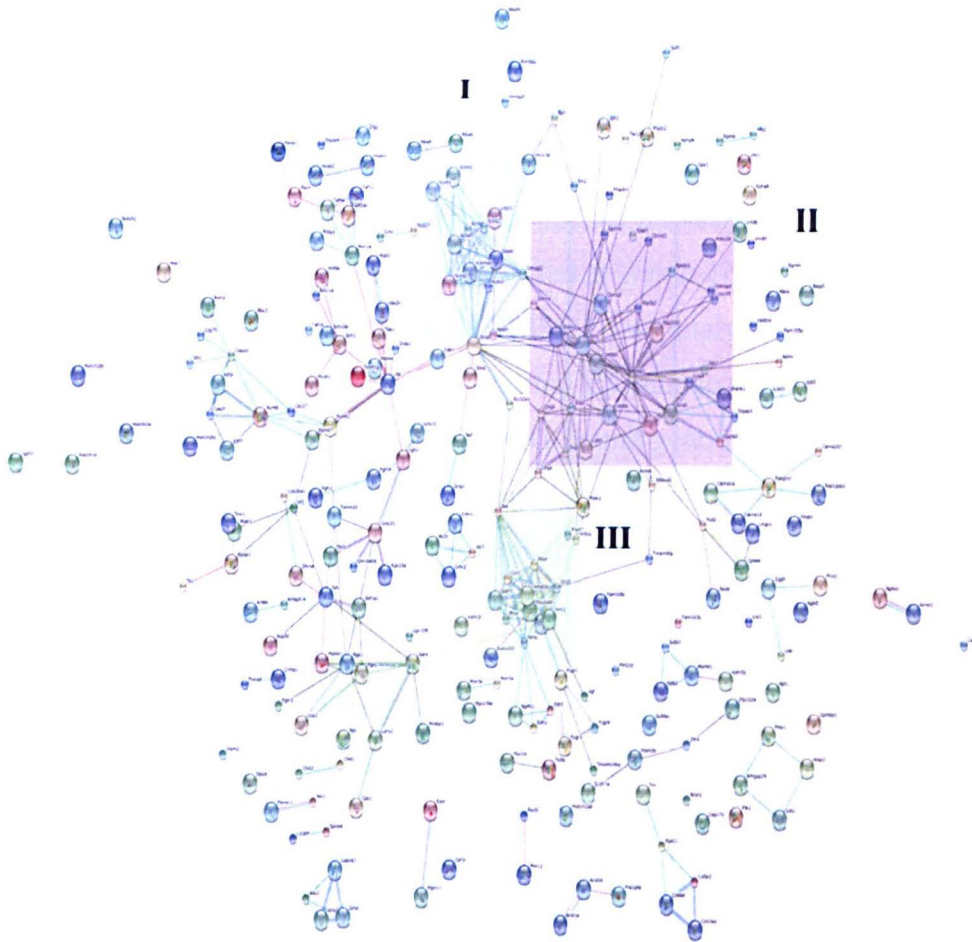


Figure 4-7 Gene-interaction network of genes with a nearest neighbor HOXB1 binding peaks

A Identification of gene interaction network using String ver 9.05. Nearest neighbor genes with Hoxb1 bound region were analyzed for evidence of known and predicted interactions. Direct (physical) and indirect (Functional) interactions were analyzed using information like genomic context, conserved expression pattern, high throughput experiments and publications. Highly enriched network with three dense hubs formed by 178 genes were identified and highlighted within green, blue and purple boxes. B. Enriched GO term from genes forming interaction network. 178 genes showing interaction were analyzed for enrichment of specific GO term.

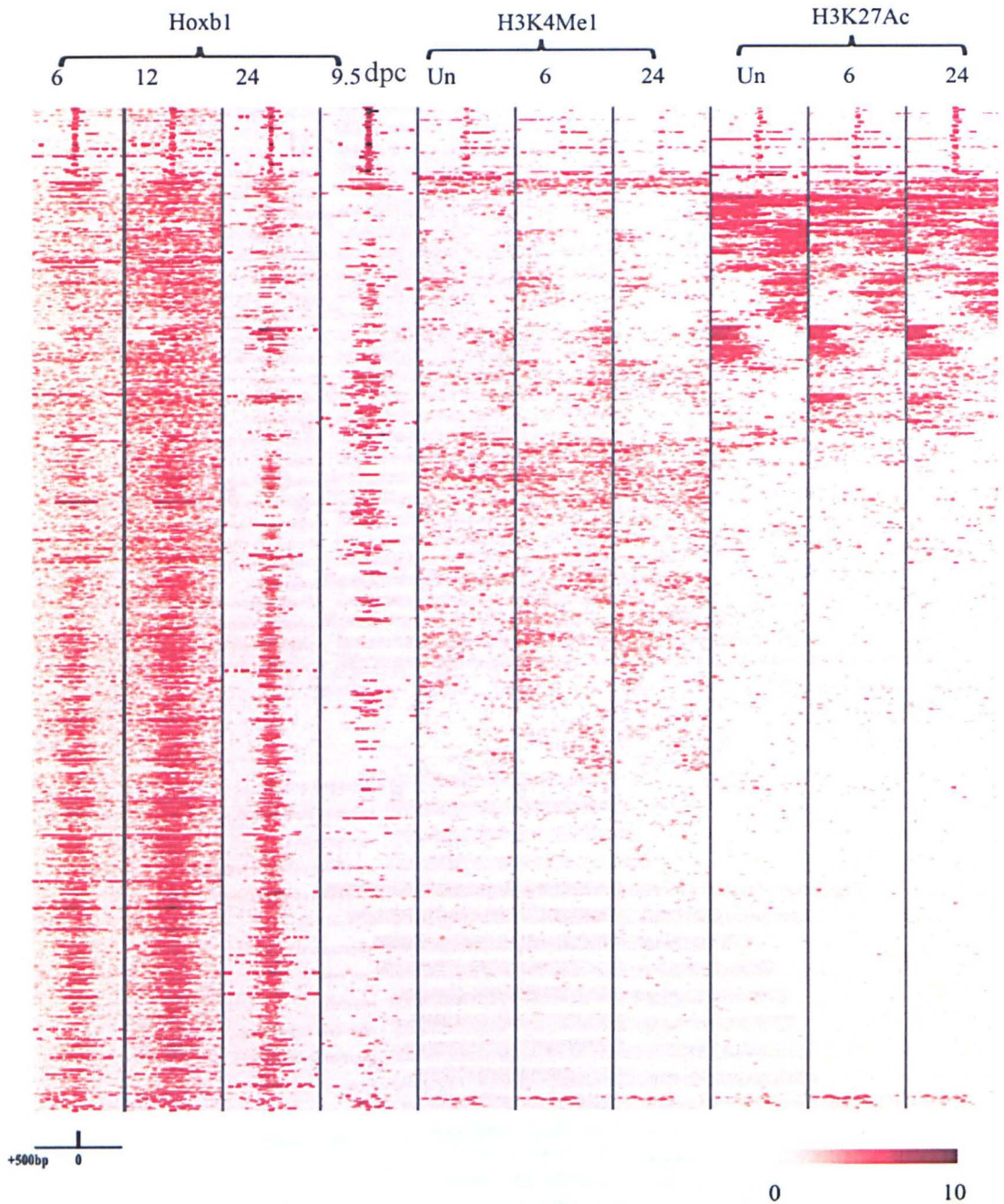


Figure 4-8 Genome-wide occupancy of HOXB1 in differentiating ES Cells.

All HOXB1 occupied regions from 6hrs, 12hrs and 24hrs of RA induced ES cells are shown on Y-axis. Every row is a genomic coordinate with HOXB1 occupancy in at least one time point. Each column represents onetime point antibody combination. Heatmap is hierarchically clustered. Each genomic coordinate is centered on midpoint of HOXB1 peaks and region including 500bp either side of midpoint is shown in heatmap. Occupancy of HOXB1 in these genomic coordinates in 9.5dpc embryos are also shown in this figure. Marks of active (H3K27Ac) Poised enhancer (H3K4Me1) are also seen along with Hoxb1 occupancy.

Heat map shows more elaborate binding dynamics in a more comprehensive way. We can see at least 1/3 peaks present only in early time points like 6 hours and 12 hours of RA and doxycycline induction. But majority of binding sites are occupied by HOXB1 in all three time points. Among these at least 50% binding sites are seems to be occupied by HOXB1 in 9.5dpc embryos. We further looked for histone modifications at these genomic loci. Since histone 3 at enhancers are modified at lysine 27 with acetylation (H3K27Ac) and lysine 4 with mono methylation (H3K4Me1), we analyzed these two histone modifications in uninduced and differentiating ES cells to understand nature of HOXB1 sites. It is evident that at least 2/3 of HOXB1 bound genomic loci are modified for either H3K27ac or H3K4Me1 or both. This indicates towards inference that these genomic loci are genuine enhancer. Genomic loci occupied more in early time point compared with 24hours of RA induction have stronger H3K27Ac marks. It is worth noting that H3K27Ac marks active enhancers. Peaks showing stronger occupancy at 24hours of RA induction seems have predominately H3K4Me1 marks. At least 20% peaks are seen only in ES cells and show no or little occupancy in 9.5dpc embryos. These genomic loci are also devoid of any H3K27Ac or H3K4Me1 marks. Very few genomic loci have H3K27Ac or H3K4Me1 histone marks below threshold. This enhancer may be special regulatory elements or specific for ES Cells or have different nature in terms of histone modifications. This needs further investigation. We found 57 much punctuated binding peaks. These peaks are seen as compact area with high level of Hoxb1 binding with no binding around flanking region. They are greater than 100 bp in length. We analyzed its genome-wide distribution pattern. These peaks are distributed in exons, introns, intergenic region in an unbiased fashion. Further, they do contain some repeat within these peak regions but it is not same repeat all the time. Due to these observations it seems that it is less likely that these are artifact. But at this point, we are not sure about relevance of such punctate binding region. Interestingly they show H3K27ac and H3K4Me1 modifications. H3K4Me1 is seems to be decreasing in this region as length of RA induction increases (Fig.4-8).

I cloned three separate HOXB1 bound region (Chr12:55600160-591; Chr14:63681200-2000; Chr5:121473017-497) in upstream of mcherry driven by minimal beta-globin promoter. This construct was electroporated in neural tube of 4-9 somite stage chicken embryos. After electroporation eggs were incubated in an incubator with 37°C, 5% CO₂ and at 85% humidity. After 16 hours of incubation, extra embryonic tissues were cleaned and embryos were imaged under fluorescence microscope for expression of mcherry. I found that all three genomic regions bound by HOXB1 can drive mCherry expression in chick neural tube (Fig.4-9).This indicates that these genomic regions capable

of driving mCherry expression in neural tube and hence can be classified as neural enhancer element. I further looked for any HOXB1 bound region reported as enhancer in vista enhancer browser. I found one region on Chromosome 11 (87235852-7215) reported as enhancer through mice transgenic experiment. This genomic region was able to drive reporter LacZ expression in neural tube and heart. Out of 15 embryos tested, 10 were showing neural tube expression while 4 shown expression in heart.

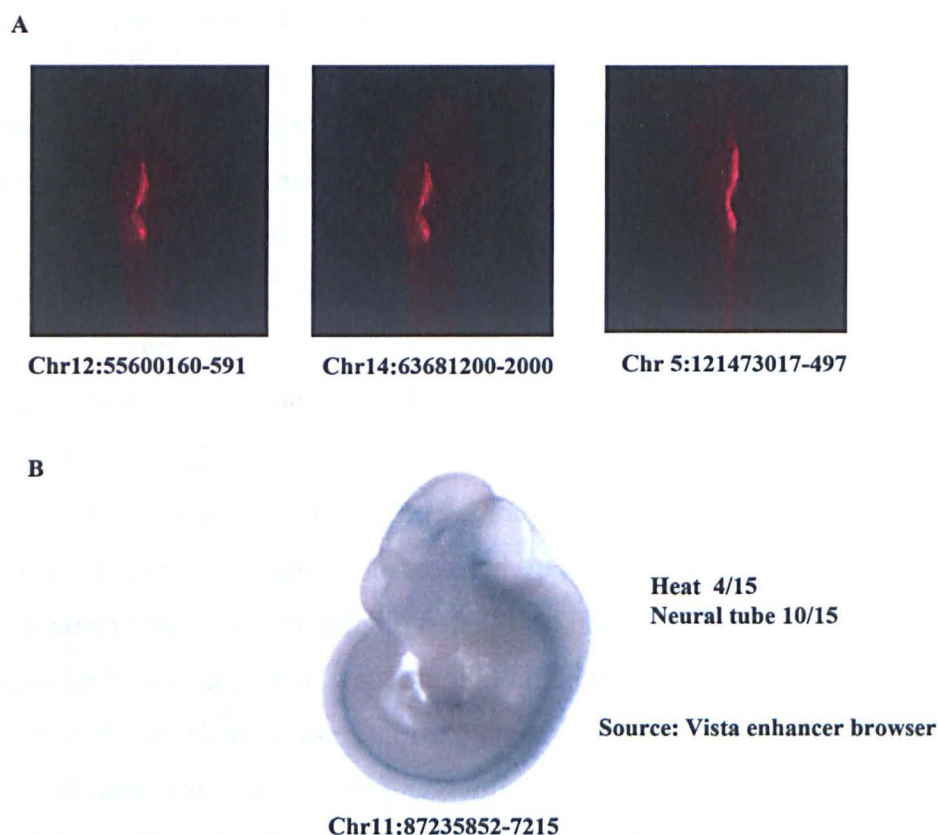


Figure 4-9 Enhancer behavior of HOXB1 occupied region in chicken and mice
A. Three selected regions bound with HOXB1 were tested for enhancer function in Chicken. Selected regions were cloned upstream of a reporter mcherry driven by beta globin minimal promoter. Enhancer mediated expression can be seen in chicken neural tube upon electroporation. **B.** Example from VISTA enhancer browser of enhancer functions of HOXB1 bound region in neural tube and heart in mice.

4.1.4 Identification of over-represented motifs in genome-wide binding site

Auto and cross regulation of HOX genes through HOX-PBX bipartite site is well documented in literature. Many examples are discussed in detail including mechanism in chapter 1. To understand co-occupancy of HOXB1, PBX and MEIS, I did ChIP-Seq with anti-Pbx (Sc-888) and Anti-Meis (Sc-25412) antibodies in 24 hours RA treated ES Cells. As described in introduction of this chapter, auto-regulatory pathway to maintain r4 expression of Hoxb1 is a well characterized enhancer acting through Hox-Pbx bipartite

sites. So, we first tried to see HOX, PBX and MEIS recruitment on this site in our ChIP experiment at 24 hours RA induced Cells. Strangely, we could not see any HOXB1 occupancy on this region. But we could see strong recruitment of PBX and MEIS in this region. We further looked at *Hoxb2* upstream region and found co-occupancy of HOXB1 along with PBX and MEIS (Fig.4-10). There are many examples where we can see co-occupancy of HOXB1 with PBX with or without MEIS in genome-wide search.

We analyzed sequence under the peak region to identify over-enriched motifs. Such analysis is always indicative about important motifs through which TF in question is recruited on DNA. Such analysis also throws light on tethered motifs assisting in recruitment of TFs or important in providing functional specificity. We first identified top 20 over represented motifs from top 100 rank peaks from each time point. These 20 motifs from each time point were compared with each other using a homemade script and metamotifs were created based on similarity. These metamotifs were tested for enrichment using FIMO on all peaks at each time point separately. Cut-off P-value for background was calculated using Kill-curves and set at 0.001 in current study. Enrichment P-values were evaluated for their significance and reported. We are not reporting fold enrichment values since metamotifs fold enrichment might be under or over reported due to degenerate nature of motifs. These metamotifs were compared with Transfec TF database using TomTom and similarity with any known transcription factors were inferred (Table 4-1). We identified four motifs with known transcription factor binding definition while identified another 8 novel motifs.

Among four motifs with known definition, two belongs to REST binding site, one for MEIS/PREP and last one for KROX binding site. We were surprised to find that REST is top ranked over represented motifs and HOX-PBX bipartite site is missing from over-represented motifs.

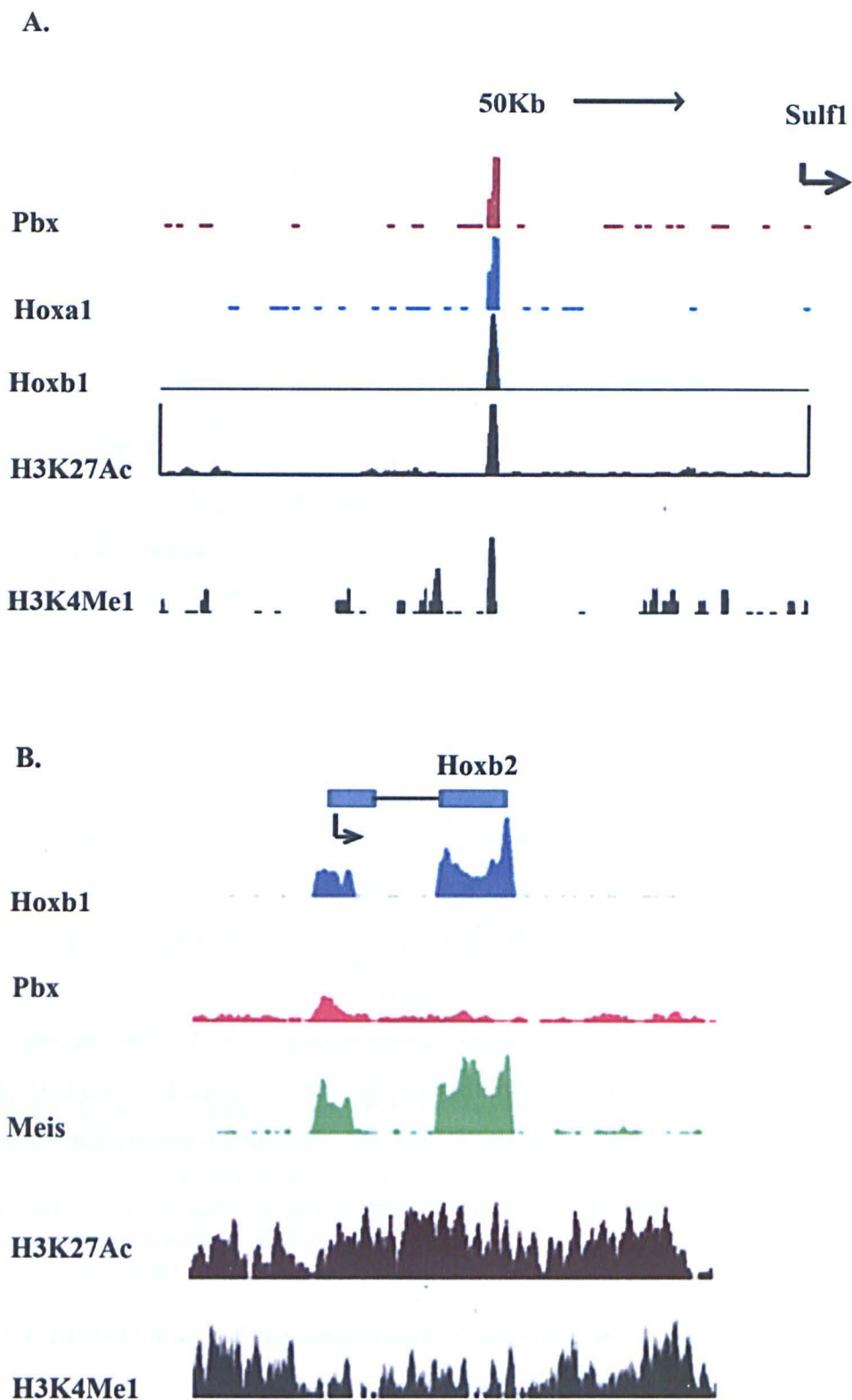


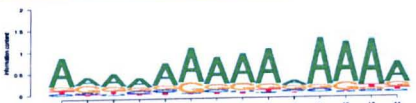











Figure 4-10 Examples of HOXB1, PBX and MEIS Co-occupancy

UCSC genome browser snapshot of region showing co-occupancy of HOXB1 and Tale proteins -PBX and MEIS. Height of peak corresponds to read coverage in this region. Y-Axis is different for each antibody. Two different genomic loci show HOXB1 and PBX co-occupancy. Distinct enhancer marks of H3K4Me1 and H3K27Ac can also be seen over both regions

Table 4-1 Identified over-represented motifs using Meme

Over-represented motifs were identified from 100 top ranking HOXB1 bound regions and their enrichment P-value were calculated from all peaks at individual RA treatment time point. Known transcription factor binding to these motifs are shown in TF column.

SN	Motifs	6hrs	12hrs	24hrs	TF
		Enrichment P -Value			
1		2.49E-13	7.1E-19	4.9E-38	REST
2		1.02E-07	1.2E-19	1.25E-25	REST
3		1.86E-05	0.00011	9.9E-10	
4		4.43E-13	5.40E-26	3.5E-45	Meis/Prep
5		0.00010	0.000115	5.3E-20	
6		1.25E-06	8.0E-09	2.6E-13	
7		0.00061	3.6E-05	1.2E-13	
8		3.617E-10	3.1E-11	7.1E-25	
9		0.03112	0.000440	0.00020	
0	1 	0.3418	0.136649	0.00094	
1	1 	8.393E-07	2.17E-8	2.4E-8	Krox, Sp1
2	1 	0.225710	0.30803	0.00420	

To address question of enrichment of previously known HOX-PBX bipartite site and ATTA site, we analyzed enrichment of these K-mer using homemade script. All HOXB1 peak sequences were analyzed for various Kmers of HOX-PBX bipartite site, ATTA and ATTG (Both Hox binding site and TGACAG (Meis/Prep site). Enrichment and P-values were calculated against an identical background sequence. (Table 4.2)

Table 4-2 Occurrence and Enrichment of Pre-defined K-mers in HOXB1 occupied region

Enrichment of kmers in all HOXB1 bound regions after 6, 12 and 24hr of RA treatment are shown in this table. A random background with similar nucleotide distribution was used to calculate enrichment and significance. No previously defined Hox binding kmer shown any significant enrichment in HOXB1 bound region.

NAME	Occurrence	LOG2FC	ADJ.P
ATTA	473	-0.01396	0.541389
WRATNNATKR	89	-0.70815	3.00E-06
ATTAG	220	-0.12613	0.193434
TGACAG	117	-0.0908	0.429016
TGATNNAT	74	-0.55077	0.001446
TGATTGACAG	1	-0.58496	0.565784
TGATNNATKR	17	-0.94376	0.007031
TGATTGAT	9	0.099536	0.541389

We found at least 2/3 peaks containing ATTA or ATTAG sequences but compared to background they are under enriched. Similarly with HOX-PBX site can be seen in around 8% of peaks while MEIS K-mers is seen in 15% Peaks. But all these K-mers are under enriched. We analyzed enrichment of Co-occurrence of pairs of K-mers but no pairs found to be significantly enriched. Though these sites are under enriched in this data set but we were interested to know positional distribution of these K-mers in HOXB1 occupied peaks. All K-mers show broad distribution and distributed from 0 ± 200 bp range from middle of the peak. No Kmers showed any tendency to over represent near center of the peak. Interestingly, in case of peaks with MEIS binding K-mer, ATTA or ATTAG shows more frequent occupancy near middle of the peak. Similarly ATTA or ATTAG K-mers shows close constraint co-occupancy near HOX-PBX bipartite site (Fig.4-11).

4.1.1 *HOXB1 as activator for REST regulated genes*

REST binding motifs seems to be highly enriched motif in *Hoxb1* binding region in all time point. This raises an important question – Do *REST* also physically bind to these genomic region? *REST* motif enrichment in *Hoxb1* binding region indicates that *Hoxb1* and *REST* play a crucial role in regulating gene expression at least subset of genes during neuronal developmental in r4. Mapp and colleagues demonstrated that *REST* is important for migration of facial branchiomotor neuron migration. Interestingly, *Hoxb1* mutant in mice shows defective facial brachiomotor neuron migrate(Mapp et al., 2011)ion. We first look at spatial distribution of REST-motifs in the HOXB1 binding region. To study this, we marked center of the peaks as 0 position and generated a plot showing distribution of REST-motif with reference to center of the HOXB1 binding peak. Plot was finally sorted for motifs with farthest position in right hand side to farthest position in left hand side (Fig4-12).

At least 20% of peaks are seen at the center of the peaks while remaining at least 70% of peaks have REST motifs within 100bp from center of the peak. 10% peaks have a *REST* -motifs between 100-200bp from center of the peak. This indicates that large number of *REST* -motifs form core of the HOXB1 binding peaks, hence seems to be functionally relevant. It will be important to mention here that summit of the peak is assumed at center of the peak but in small instances they may be different. But we feel that this small deviation doesn't change interpretation of our result. Due to these correlations, we decided to look into co-occupancy of *REST* and HOXB1 on *REST*- motif containing genome region. I did CHIP-Seq for *REST* occupancy in uninduced and 24Hours RA induced ES Cells. Peaks were identified with same parameters as for HOXB1 and data were compared. We analyzed occupancy of *REST* in *REST*- motif containing HOXB1 binding regions. At least 30% of HOXB1 peaks contains a full or half *REST*-motifs. Interestingly most of *REST*-motifs containing HOXB1 peaks are co-occupied with REST in uninduced ES Cells and after 24hours of RA induction (Fig.4-13).

We analyzed spatial distribution of *REST* and HOXB1 peaks and found that they are on top of each other and bound on same genomic region (Fig.4-13). During metamotif generation, we found that *REST* half sites are also enriched in Meme analysis. This raises an important question that is half sites capable of recruiting *REST*? We divided HOXB1-

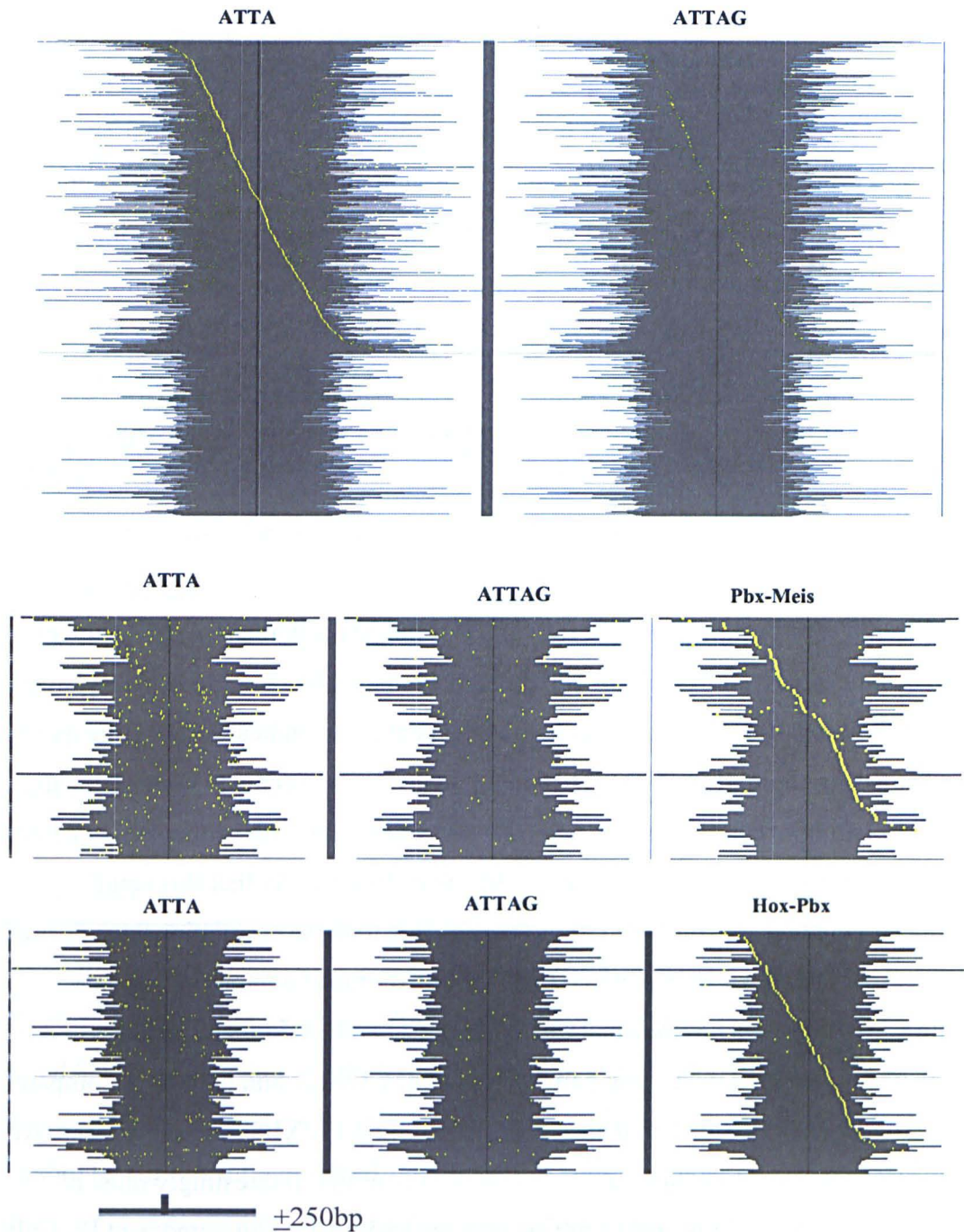


Figure 4-11 Distribution of K-mers in HOXB1 bound genomic regions.

Position of motif is shown with respect to center of HOXB1 peak. Each yellow dot represents presence of one kmer. Each row represent genomic region bound by Hoxb1 and shadowed area represents span of each peak. Peak with given kmer at farthest right from center is shown at the top while Peak with kmer at farthest left from center is shown at the bottom. Other kmers are shown relative to kmer based on which data is sorted. **A.** Co-occurrence of ATTA and ATTAG k-mers in HOXB1 bound region. Kmers are sorted based on position of ATTA K-mer. **B.** Co-occurrence ATTA, ATTAG and *Pbx-Meis* motif in HOXB1 bound region. HOXB1 bound region with *Pbx-Meis* kmer are shown. Data is sorted based on position of *Pbx-Meis* K-mer. **C.** Co-occurrence ATTA, ATTAG and *Hox-Pbx* motif in HOXB1 bound region. HOXB1 bound region with *Hox-Pbx* kmer are shown. Data is sorted based on position of *Hox-Pbx* K-mer.

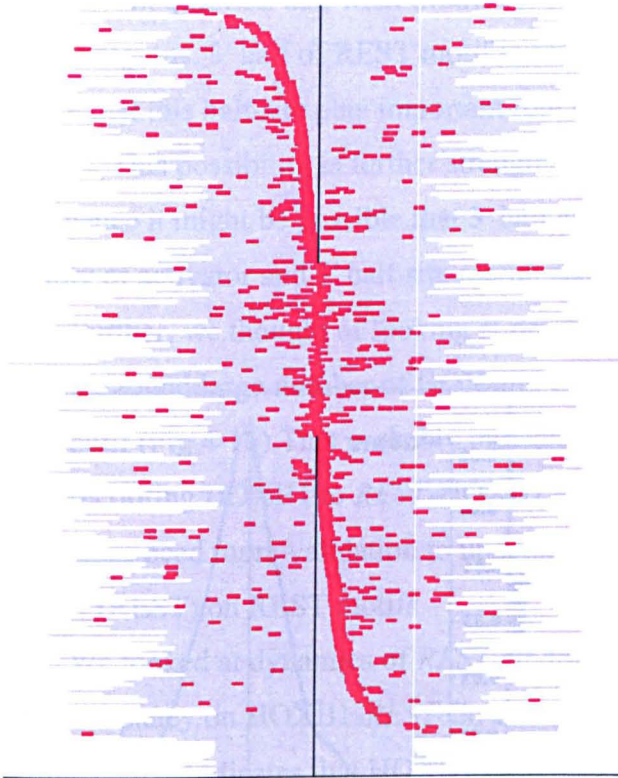


Figure 4-12 Distribution of REST motifs in HOXB1_REST co-occupied peaks

Position of motif is shown with respect to center of HOXB1 peak. Each red dot represents presence of one REST motif. Each row represent genomic region bound by HOXB1 and has a at least one REST motif. Shadowed area represents span of each peak. Peak with at least one REST motif at farthest from center is shown at the top while Peak with farthest left kmer from center is shown at the bottom.

REST co-occupied peaks based on presence of various configurations of half and full sites and generated a heat map of showing binding and its intensity (Fig.4-14). To our surprise, half sites were able to recruit *REST* in these sites. There is no difference in dynamics of *REST* recruitment on these half site or two half sites spatially separated, or a full site with half site or a single full site (Fig.4-14). We further identified that two half site can have a 3-5bp spacers. Canonical *REST* motifs have a 2 degenerate spacer nucleotide. Hence this discovery seems to be novel in nature (Fig.4-15).

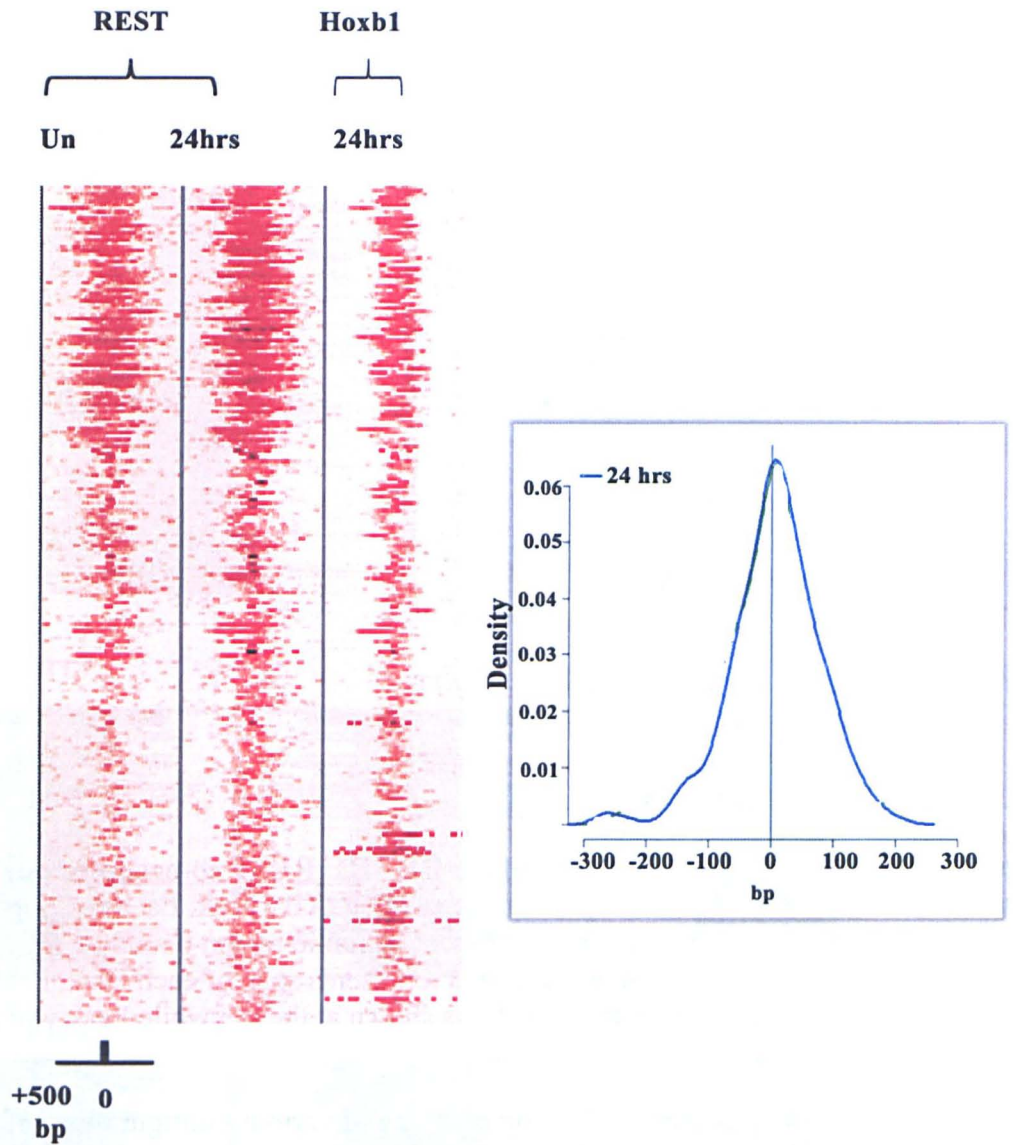


Figure 4-13 Co-occupancy of REST and HOXB1 in differentiating ES cells

All HOXB1 bound genomic regions after 6 hrs, 12 hrs and 24hrs of RA treatment with a REST motif is shown in this figure. Each row is a HOXB1 bound region. Occupancy of REST in Uninduced and 24hrs RA induced KH2 cells on these genomic regions are shown along with HOXB1 occupancy after 24hrs of RA and Doxycycline treatment. Peaks are centered on center of HOXB1 bound region and ± 500 bp region is shown as heatmap. Heatmap is sorted on intensity of REST binding in uninduced ES cells. Graph on left panel shows spatial distribution of REST peaks and HOXB1 peaks in 24 hours RA induced ES cells.

Discovery of REST motifs at the center of the HOXB1 occupied region raises a more relevant question that what motifs are responsible for recruitment of HOXB1 to these genomic regions. 3' half of REST motif is similar to MEIS/PREP/TGFII motif. It may be possible that this half site play important role in recruitment of HOXB1 through its cofactors. This possibility is further strengthened by the fact that 5' half site can recruit *REST* itself, hence it might be possible that 3' half site of full *REST* site is important for assembly of activator and 5' half site recruit repressor complex.

Further, we thought of looking at presence of HOX or PBX binding site near REST Motif. We found large number of HOX and PBX binding sites around 50-100bp from REST motif (Fig.4-16). This probably indicates that these tethered HOX or PBX site may help in recruiting HOXB1 to *REST*-HOXB1 Co-occupied peaks. But both these observations need more validation to understand recruitment of activators (HOXB1) and repressor (REST) on REST motifs.

We looked at dynamics of *REST* binding in differentiating ES Cells. We compared REST occupancy on HOXB1-*REST* Co-occupied peaks before and after HOXB1 binding. Our time course indicates that HOXB1 proteins are available as early as 6 hours of RA Induction. So, we compared *REST* binding on genomic region co-occupied by *REST* and HOXB1 in uninduced ES Cells and 24hours RA Induced. To our surprise, there no change in *REST* Occupancy among these two time points (Fig.4-13).

This does not indicate that HOXB1 occupancy is depends neither upon loss of *REST* binding nor HOXB1 occupancy result in loss of *REST* from its occupied sites. We do believe that this statement needs more validation. It may be possible that ChIP is capturing general heterogeneity in cell population. Site with *REST* may be free from HOXB1 occupancy and Vice-versa. To resolve this further, we are planning for a sequential ChIP with *REST* and HOXB1.

We analyzed histone modification status of HOXB1-*REST* Co-occupied peaks. It seems that around 80% of peaks have either H3K4Me1 or H3K27Ac or both. Very few peaks have active enhancer marks of H3K27Ac while most of these have H3K4Me1 modification. H3K27Ac do not change much over time course of RA induced differentiation. But majority of binding region with H3K4Me1 start with no or low H3K4Me1 marks but over time course of differentiation more and more regions gain enhancer related marks of H3K4Me1. One more striking observation was related to highly *REST* bound peaks. These peaks are mostly devoid of any enhancer related histone marks. This probably indicates sub functionalization within REST bound regions (Fig.4-17).

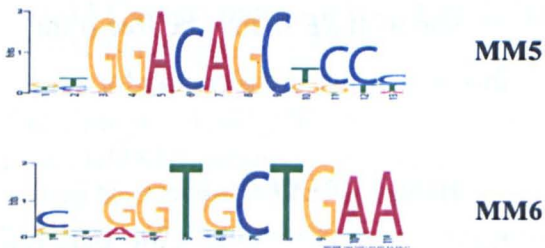
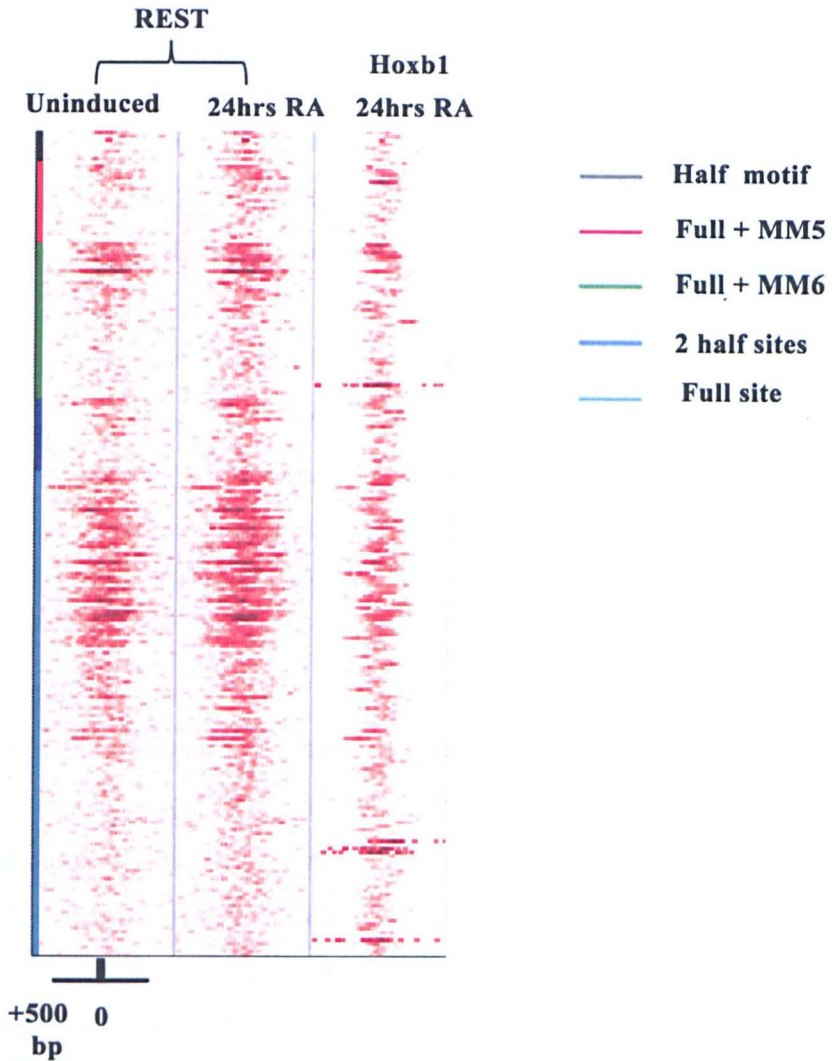
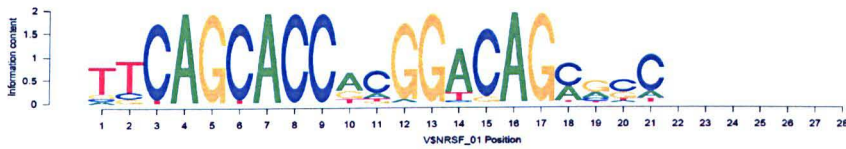


Figure 4-14 Binding properties of Full and Half sites

All HOXB1 bound genomic regions after 6 hrs, 12 hrs and 24hrs of RA treatment with a REST motif is shown in this figure. Occupancy of REST in Uninduced and 24hrs RA induced KH2 cells on these genomic regions are shown along with HOXB1 occupancy after 24hrs of RA and Doxycycline treatment. Peaks are centered on center of HOXB1 bound region and ± 500 bp region is shown as heatmap. Two conserved half sites of REST motif is shown as MM5 and MM6. Bound regions are stratified into various combinations of full and half sites. Each group is color coded and shown at right side of the heatmap. No sorting was done. Half and Full sites have equal affinity for REST binding and Hoxb1 binding.

REST full Motif



Novel REST Motifs with varying spacer length between 3' and 5' conserved halves

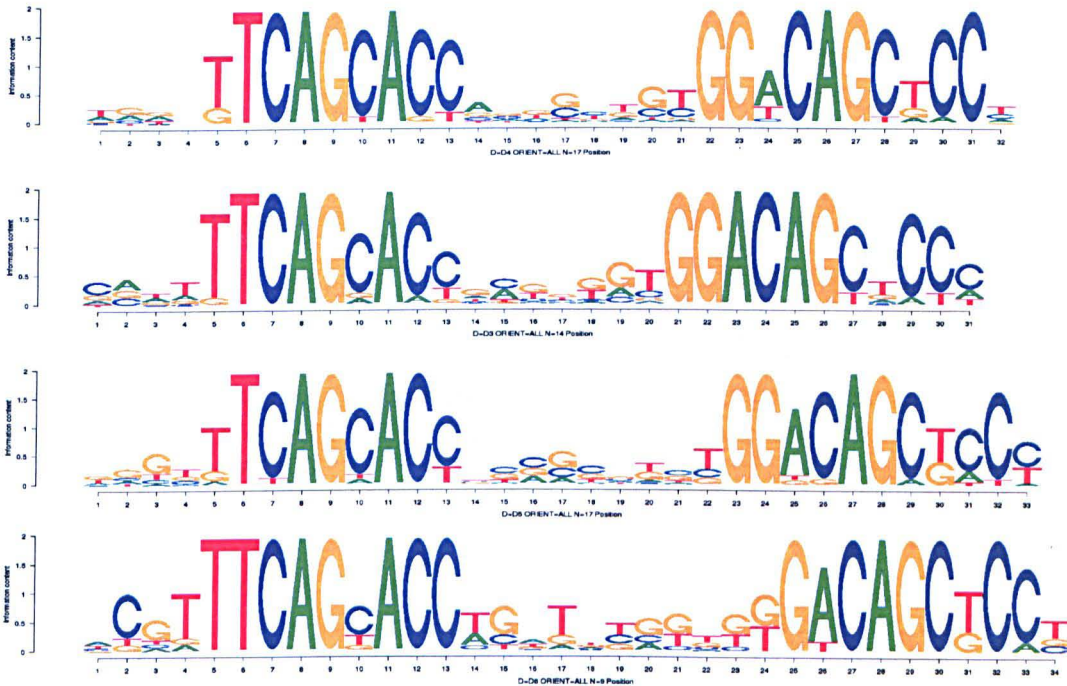


Figure 4-15 Novel REST motifs with different spacers between two half sites
 Canonical REST full site is shown at the top. Two base pair spacer between two conserved halves can be seen. Various novel REST motifs with varying spacer length (more than 2 bp) with two flanking conserved halves can be seen.

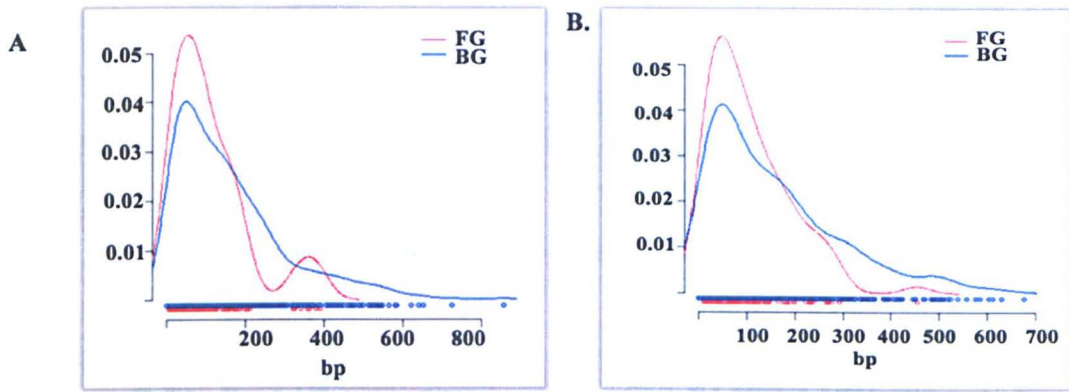


Figure 4-16 Minimum distance of REST, HOX and PBX motifs in HOXB1-REST Co-occupied peaks

Minimum distance between REST motif and K-mers were calculated in HOXB1 and REST co-occupied genomic regions and plotted as cumulative frequency distribution function. X-Axis shows distance between nearest REST-Kmer pair. A random background peak set was also generated to test random distribution pattern and shown as blue line. **A.** Distribution of distance between nearest REST-ATTA motif Hoxb1-REST Co-occupied peaks. **B.** Distribution of distance between nearest REST- HOX-PBX motif Hoxb1-REST Co-occupied peaks

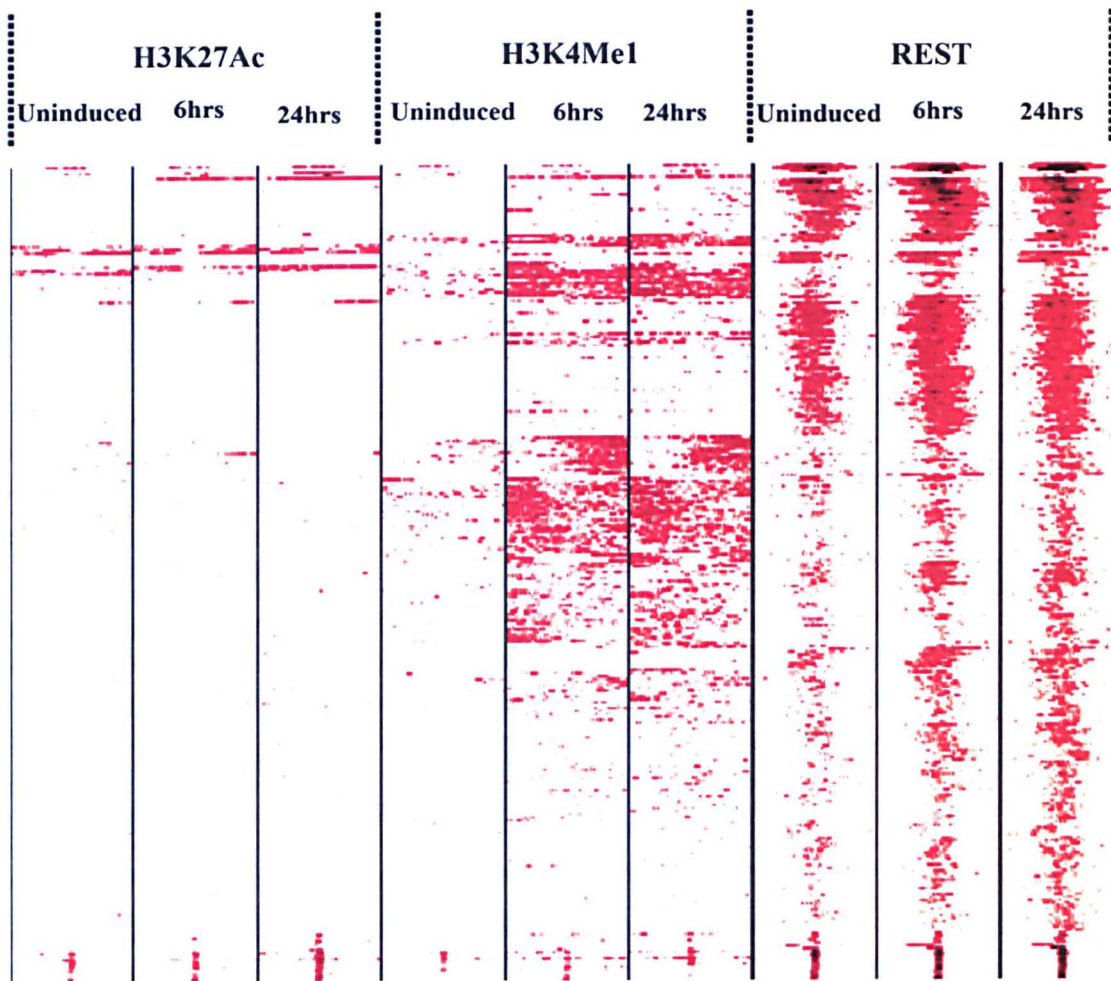


Figure 4-17 Enrichment of H3K27Ac, H3K4Me1 and REST binding on genomic region co-occupied by HOXB1 and REST

All HOXB1 bound genomic regions after 6 hrs, 12 hrs and 24hrs of RA treatment with a REST motif is shown in this figure. Each row is a HOXB1 bound region. Occupancy of REST in uninduced, 6hrs and 24hrs RA induced KH2 cells on these genomic regions are shown. Peaks are centered on center of HOXB1 bound region and ± 500 bp region is shown as heatmap. HOXB1 occupancy is not shown in this figure. Enhancer specific histone modification namely H3K27Ac and H3K4Me1 in uninduced ES cells and after 6 and 24hours of RA treatment in these regions are also shown.

4.1.2 HRE2 (Hox Response Element 2) is a repressive Hoxb1 element

We identified 126 top ranking common peaks between 24 hours induced ES cells and 9.5dpc embryos. These peaks were subjected to identification of over represented motifs. Among many over represented motifs, we found a novel motif and named it as HRE2 (Hox Response Element 2). This motif was initially named as Meme motif2 (MM2) or C2 and later (Old meme motif) OM2 in records (Fig.4-18).

HRE2

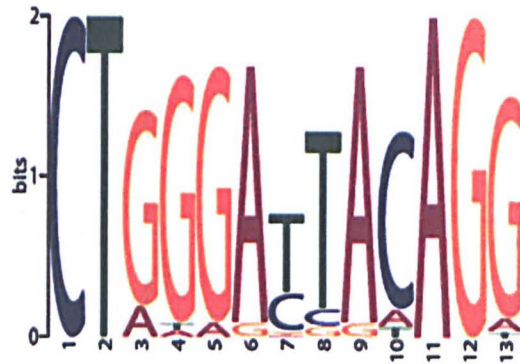


Figure 4-18 Logo of HRE2.HRE2 contains an ATTA motif at its center.

I cloned HRE2 (6X) repeated 6 times in upstream of lacZ driven by minimal beta-globin promoter. This construct was electroporated in neural tube of 4-9 somite stage chicken embryos. After electroporation eggs were incubated in an incubator with 37°C, 5% CO₂ and at 85% humidity. After 16 hours of incubation, I fixed it and stained with X-gal. After staining, embryos were scored for lacZ expression in neural tube. I found that HRE2 can drive lacZ expression in chick neural tube. Out of 15 embryos tested 13 showed LacZ expression in the neural tube (Fig.4-19). This indicates that the 13bp long HRE2 is capable of driving lacZ expression in the neural tube and hence can be classified as neural enhancer element. Further, we co-electroporated *Hoxb1* expression construct (CMV driven *Hoxb1*) and minimal beta-gal promoter with HRE2 driven LacZ and assayed lacZ expression as described above. To our surprise, no embryos expressed lacZ (Fig.4-19). This indicates that HRE2 is a regulatory element and integrates repressive action of *Hoxb1* into regulatory part of the genome. We used *Hoxb1* auto-regulatory region for same assay as positive control.



HRE2-Lacz



HRE2-LacZ + Cmv-Hoxb1

Figure 4-19 Enhancer activity of HRE2 in chicken embryos

Multimerized (6X) HRE2 were tested for enhancer function in Chicken. Multimerized (6X) HRE2 were cloned upstream of a reporter LacZ driven by beta globin minimal promoter. Enhancer mediated expression can be seen in chicken neural tube upon electroporation. Loss of LacZ expression can be seen upon co-electroporation of CMV-driven *Hoxb1* with Multimerized (6X) HRE2 reporter system indicating HRE2 as a putative repressor sequence.

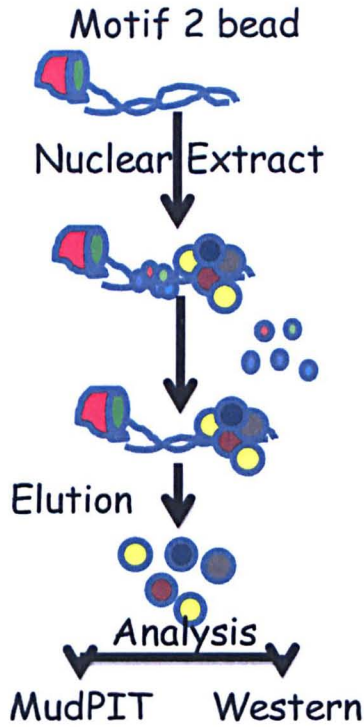
This raise question how *Hoxb1* is recruited on this motif? This is not a typical *Hox-Pbx* site, so what else co-occupies this motifs and how they are functionally important? To answer these questions, we developed a novel assay in collaboration with Conaway lab members. This assay was named the “template binding assay” by the Conaway lab. In this assay, 6X HRE2 was biotinylated at the 5' end. This biotinylated DNA fragment was immobilized on ferro-magnetic streptavidin beads. Streptavidin bead- biotinylated DNA fragment was incubated with nuclear extract. After 30 min incubation at 30°C, unbound proteins were washed with wash buffers. Template bound proteins were eluted using a SDS-buffer. These eluted proteins were analyzed by western or MudPIT. A random fragment is used as control to standardize reaction for specific binding. Elutes from random fragments were also used as negative control in MudPIT studies. I characterized protein bound exclusively to HRE2 but not to random fragment. Known contaminants were removed before analysis. Proteins with dNSAF value of at least 1/100 th of highest dNSAF value were selected. I found that HOXB1 can bind specifically on HRE2 and *Hoxb1* auto-regulatory element (ARE) at 150mM salt concentration using Western hybridization (Fig.4-20). I used poly-dicd or sheared lambda DNA as competitors. Irrespective of the nature of the competitor, HRE2 can bind *Hoxb1* specifically between 150mM and 200mM salt concentration. This assay indicates that HRE2 can recruit *Hoxb1*.

Table 4-3 Proteins recruited by HRE2

Proteins bound to HRE2 were analyzed using MuDPIT(Multi-dimensional Protein identification tool). Protein bound on a random DNA fragment were used as negative control.Hoxb1 ARE region were used as positive control. Proteins with dNSAF value of at least 1/100 th of highest dNSAF value were selected. List of transcription factors and their dNSAF value is shown below.Hoxb1 was not identified in MudPit analysis but was seen in western.

Protein	dNSAF
Pbx1	0.00029
Pbx2	0.00022
Meis1	0.00016
Med15	0.00014
Med18	0.00027
Med19	0.00015
Med24	0.00022
Med27	0.00018
Med28	0.0001
Med8	0.0004
Iws1	0.00014
Sox2	0.00022
Dppa4	0.0001
Yy1	0.00013
Brd7	0.0001
Cdk1	0.00162
Chaf1a	0.00019
Parp2	0.00016
Pcgf2	0.00022
Polr2e	0.00018
Polr2g	0.00022

A.



B.

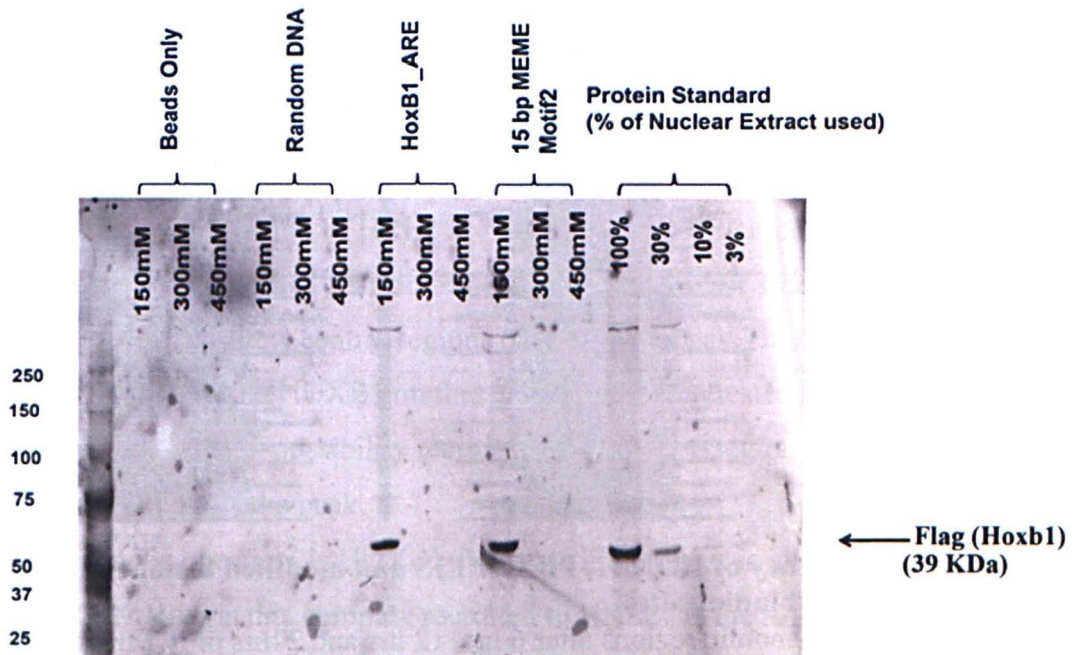


Figure 4-20 HOXB1 binds specifically to HRE2 and Known ARE region.

Template binding assay was done understand physical binding of HOXB1 on HRE2 and ARE (Hoxb1 auto-regulatory region). Nuclear extract was prepared from KH2-cell line with HOXB1-His-Flag tag after treatment with Doxycycline and RA for 24 hour. Western with Anti-flag antibody was used to identify HOXB1 on HRE2 and ARE. Random DNA fragment and beads only controls were used as negative control. A. Schematic diagram of template binding assay B. Specific binding of HOXB1 over HRE2 and ARE region can be seen at 150 mM salt concentration.

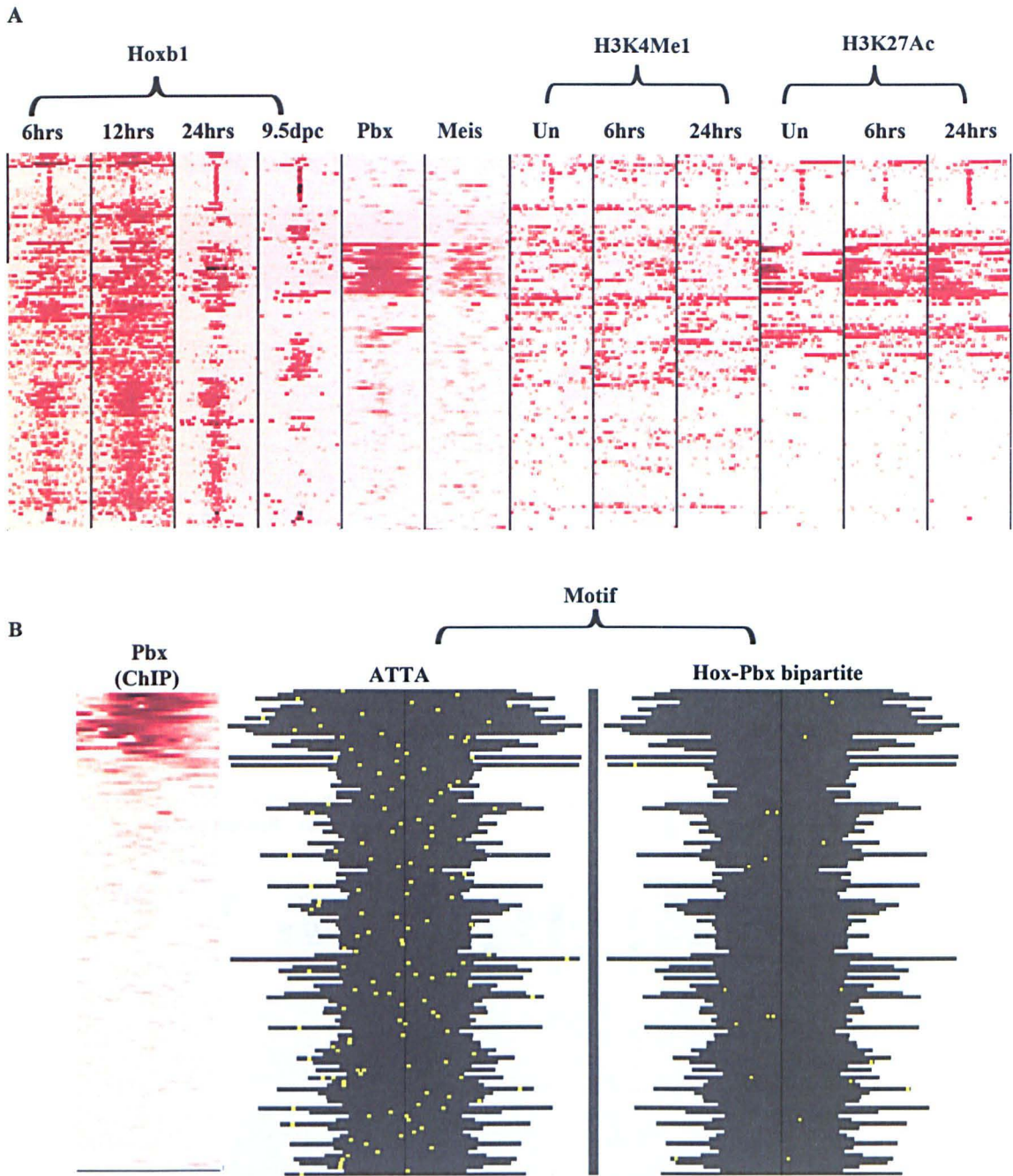


Figure 4-21 Occupancy of HOXB1, PBX, MEIS and modified histones on HRE2 containing HOXB1 binding sites

All HOXB1 bound genomic regions after 6 hrs, 12 hrs and 24hrs of RA treatment with a HRE2 motif is shown in this figure. Each row is a HOXB1 bound region with HRE2 motif. Peaks are centered on center of HOXB1 bound region and ± 500 bp region is shown as heatmap. HOXB1 occupancy is not shown in this figure. Occupancy of PBX and MEIS are also seen. . Enhancer specific histone modification namely H3K27Ac and H3K4Me1 in uninduced ES cells and after 6 and 24hours of RA treatment in these regions are also shown. A. Co-relation between Pbx binding and H3K27Ac histone marks is striking. This indicates Putative role of *Pbx* in activation of HRE2 containing enhancers B. Distribution of Hox and Hox-Pbx motifs in HRE containing regions Shaded area shows extent of HOXB1 bound peaks. Data is sorted based on intensity of PBX binding.

We further analyzed protein bound on HRE2 by using MudPIT. To our surprise, HRE2 can recruit PBX1 and 2, MEIS1, SOX2, many mediator subunits (MED18, MED19, MED24, MED28, and MED8), ISWI and PARP2 (Table 4.3). We have used Hoxb1 ARE region as positive control. We identified PBX1 and PBX2 binding on this region (Appendix VII). It is known that ARE region function through recruitment of PBX and MEIS (Ferretti et al., 2005). Incidentally, we could not find any Hoxb1 binding on HRE2 using MudPIT which was confirmed through western hybridization. This inconsistency might be due to small size of the Hoxb1 protein or its relative abundance of tryptic digestion sites.

We have observed that many HRE2 containing HOXB1 bound regions are co-occupied with *Pbx*. One specific example worth mentioning is *NPAS4* (Neuronal PAS domain protein 4). This gene is essentially down regulated in neuronal differentiation. This gene shows HOXB1 occupancy on intron1. Interestingly, this region is co-occupied with PBX. Presence of HRE2 in this region makes more sense, since during differentiation into neuronal lineage, this gene needs to be down regulated and HRE2 mediated HOXB1 and PBX recruitment is essential for this functional outcome.

We analyzed distribution of HRE2 in Hoxb1 occupied HRE containing region (Fig.4-22). It seems that HRE2 is distributed between 250bp either side of center of the peak. HRE2 occupies very few central regions of the Hoxb1 peak. This raises the question about mode of recruitment of HOXB1 on these peaks. I am interested to look for the nature of any other motif which can possibly recruit Hoxb1 to these region and show a distribution along center of the peak.

This raises the question, is HOXB1-PBX-MEIS complex is involved in regulatory property of HRE2? How is this complex recruited to HRE2? Is the ATTA sequence at the middle of this motif responsible for recruitment of the HOXB1-PBX-MEIS complex? Is Sox2 responsible to bring HOXB1-PBX-MEIS complex into this template? The answers to these questions need further experimentation.

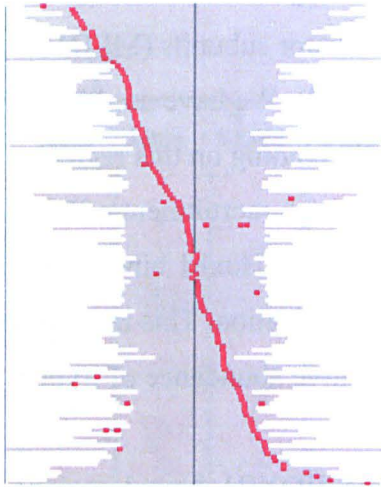


Figure 4-22 Distribution of HRE2 in HOXB1 bound genomic regions

Position of HRE2 motif is shown with respect to center of HOXB1 peak. Each red dot represents presence of one HRE2 motif. Each row represent genomic region bound by HOXB1 and has at least one HRE2 motif. Shadowed area represents span of each peak. Peak with at least one HRE2 motif at farthest from center is shown at the top while Peak with farthest left kmer from center is shown at the bottom. HRE2 show wider range of distribution in HOXB1 peaks. Presence of mostly single motifs is another striking aspect of this motif.

We analyzed enhancer related histone modifications in the HRE2 containing regions. It seems that most of these regions do contains H3K4Me1 marks. But to our surprise, HOXB1 bound region co-occupied with *Pbx* have exclusive histone H3K27Ac marks. This is a histone mark corresponding to active enhancers, suggesting that occupancy of PBX is essential for a HRE2 containing enhancer to achieve active status (Fig.4-22). Further, I wish to speculate that HOXB1 binding proceeds to PBX occupancy and binding of PBX results in repressive action of HRE2 mediated by *Hoxb1*. This speculation needs more validation and experiments

4.1.3 Modular Motif–Long cis finder motif

I identified a 41 bp long modular overrepresented in 126 previously described common Hoxb1 bound regions in 9.5 embryos and 24hrs of RA induced ES Cell samples (Fig.4-23).

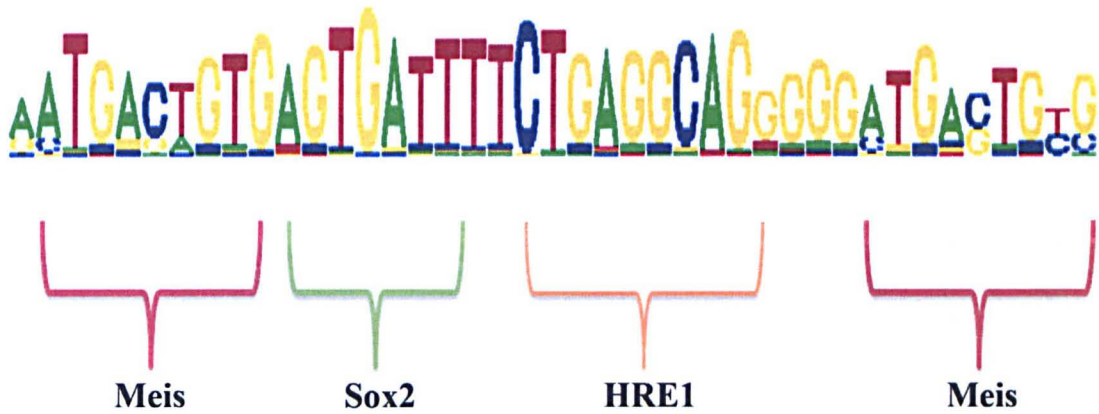


Figure 4-23 Logo of Long Motif Modular nature is highlighted clearly. It is evident that this long motif is made up of three different motifs

This is modular motif which contains a flanking MEIS binding site with a core SOX2 binding motif with a novel motif which was earlier identified as HRE1 (or Meme motif 1). To test enhancer activity of this motif, I cloned this long motif 6 times in tandem repeat upstream of lacZ reporter driven by minimal beta-globin promoter. This reporter construct was electroporated in 4-9 somite stage chicken embryos and after 16 hours of incubation lacZ expression were examined in neural tube. It was interesting to see that this 41bp DNA motif is capable to drive lacZ expression in neural tube. When murine *hoxb1* was expressed along with this Motif enhanced lacZ reporter through co-electroporation, to our surprise we found complete repression of lacZ reporter linked with Long motif (Fig.4-24). This is an indication that like previous described HRE2 long motif is also a repressive element. One major difference between HRE2 and long motif was nature of motif. HRE2 is short core motif while long motif consist of many core motifs like HRE2. This probably indicate that modular motifs as well as tethered motifs can play role in recruitment of Hoxb1. Using previously described template binding assay, we analyzed protein showing binding affinity to long motif. To our surprise, fewer proteins binds to long motif compared with HRE2. Like HRE2, the long motif is bound by PBX, MEIS and SOX. It is important to know at this point that the long motif is built of MEIS and SOX binding

motifs. Further large number of mediators and chromatin remodelers are also seen recruited on long motif (Table 4.4). This indicates a probable repressive action and recruitment of Hoxb1.



Long-motif-LacZ



**Long-motif-LacZ
+
Cmv-Hoxb1**

Figure 4-24 Functional assay of long motif for enhancer activity Multimerized (6X) long motif were tested for enhancer function in Chicken. Multimerized (6X) long motif was cloned upstream of a reporter LacZ driven by beta globin minimal promoter. Enhancer mediated expression can be seen in chicken neural tube upon electroporation. Loss of LacZ expression can be seen upon co-electroporation of CMV-driven *Hoxb1* with Multimerized long motif- reporter system indicating long motif as a putative repressor sequence.

shown role of HOX-PBX complexes in conferring binding specificity to HOXB1. Various HOX-PBX sites were found to be part of auto and Cross-regulatory elements of Hox genes within Hox locus. In our study, we were able to find many examples of HOX-PBX and HOX-PBX-MEIS co-occupancy and presence of Hox-Pbx bipartite site. But, to our surprise Hox-Pbx bipartite site mediated recruitment of HOXB1 is not seems to be major mechanism. Only 10% of HOXB1 binding site seems to have co-occupancy of HOX and PBX. Interestingly, we identified a novel repressive motif functioning through HOX- PBX interaction. We observed a positive co-relation between PBX –MEIS binding and gain of H3K27Ac. This may suggest that binding of PBX

able 4-4 Proteins recruited by long motif(Jabet et al., 1999; Knoepfler et al., 1999; Phelan and Featherstone, 1997)

Proteins bound to long motifs were analyzed using MuDPIT(Multi-dimensional Protein identification tool) .Protein bound on a random DNA fragment were used as negative control.*Hoxb1* ARE region were used as positive control. Proteins with dNSAF value of at least 1/100 th of highest dNSAF value were selected. List of transcription factors and their dNSAF value is shown below.

NCBI_Gene	dNSAF
Pbx1	0.00023
Meis1	0.00015
Sox2	0.00017
Med18	0.00013
Med24	0.00022
Med27	0.00011
Med8	0.0004
Chaf1a	0.00019
Iws1	0.00024
Llgl2	0.00023
Parg	0.00011

probably changes enhancer from poised to active status. This may further suggest that PBX may be acting as switch to convert HRE containing HOXB1 binding site from repressor to activator. Both questions need more experimentation and validation for correct answers.

4.2 Discussion

Hoxb1 is expressed in an r4 restricted manner in developing mice embryos. The limited amount of tissue and difficulty in purification of these specific cells were major limitation in experimental approaches to understand genome-wide binding properties and source of binding specificity of HOXB1 in r4.. We have therefore generated a doxycycline inducible epitope tagged *Hoxb1* ES cell line. This is a very handy system since many currently available protocols are capable of directed differentiation of ES cells into specific cell types; hence it provides a valuable tool to generate comparable data of Hoxb1 binding in different cell types. This further facilitates understanding the rationale behind site selection in different tissue types and the factors governing binding specificity. Bami and

coworkers demonstrated the sensitivity of such an inducible system and identified many novel targets which were not identified in previous studies (Bami et al., 2011).

Various genetic studies and in vitro binding assays identified the Hox-Pbx bipartite site as an important Hox binding sites. Further, in vitro studies and structural studies have Biggest surprise from this study was identification of REST as most enriched motif. REST is not just enriched as motif in Hoxb1 binding peaks but also physical present on these sites. REST is a repressor protein which transiently represses neuronal genes in stem cells and progenitors while permanently represses in non-neuronal tissues (Ballas et al., 2005; Ballas and Mandel, 2005). REST with Co-REST recruits chromatin modifiers to achieve transient and long term repression of neuronal genes. These chromatin modifiers include *mSin3A*, *SUV39H1*, *HP1*, *Scps* (Small CTD phosphatases), *HMTaseK4* and *MeCP2*. Establishment and maintenance of REST mediated repression is well studied but activation of REST repressed genes are less understood. It is generally believed that REST complex leaves enhancer to allow activation of neuronal genes. Role of an activator is also envisaged in this process but no activators were identified till date. It seems that our study suggest that HOXB1 might be the illusive activator involved in activation of transiently REST bound Neuronal genes. Further it seems that these genes do not need complete removal of repressors from enhancer. We can see continued occupancy of REST on these enhancers upon HOXB1 binding. This may suggest that binding of an activator is independent of release of repressors. Binding of activator is sufficient to achieve activation. Further removal of repressor is achieved through passive process. I would like to speculate that removal might be achieved through diluting out repressor occupancy in few cycles of cell division.

To summarize our results it seems that Pbx and Meis mediated recruitment of HOXB1 on genomic location is not primary mechanism. Rather at least REST and HRE2 containing peak data indicate that Pbx is not required for recruitment of Hoxb1 on many genomic loci. This is contrary to many existing model where chromatin remodeling activities of TALE proteins are considered as essential component to facilitate HOXB1 binding. HRE2 data further indicate that Hoxb1 can bind without helping hand of Pbx but probably Pbx recruitment alters behavior of Hoxb1 bound enhancers. Further, transcription factors other than TALE proteins might help to bring Hoxb1 to DNA. Template binding assay and MudPIT indicates that probably SOX2 might be the candidate in case of HRE2. It was a speculation that Hox genes can act activators and repressors. Many Known Hox response elements were activators in nature. In current study, we identified a novel 13bp long HRE which perceive repressive signal from HOXB1.

Chapter 5 Interactors, co-factors and combinatorial binding

Specificity of Hox gene binding is attributed to Co-factors and interactors (Ferretti et al., 2005; Ferretti et al., 2000; Ferretti et al., 1999; Jabet et al., 1999; Mann and Chan, 1996). Interestingly, functional outcome of Hox genes are believed to be through combinatorial binding of Hox genes along anterior posterior body axis. This combinatorial code generated by single hox gene with cofactors or by group of hox genes with their cofactors is generally referred as “Hox code”. So it is always unwise to attempt understanding of a single hox gene in isolation. It is true that one single gene interacts with number of other hox genes directly or indirectly which makes such a study a very big challenge at the level of logistics and analysis. In this context, it is at least possible to study a given Hox genes with its other paralogous gene partner(s) and known cofactor (s). Paralogous hox gene is important since numbers of functional redundancies are reported by many authors among various hox genes and their paralogous hox genes. In current study, we attempted to understand *Hoxb1* genome-wide binding properties through comparing and contrasting genome-wide binding properties of group1 paralogous gene *Hoxa1*, Co-factors (*Pbx* and *Meis*) and a novel interactor *REST*. Our study doesn't include *Hoxd1* since it is not expressed in mice hindbrain.

Hoxa1 is earliest expressed Hox gene in mice. *Hoxa1* transiently expressed between 7.5 dpc to 8.dpc in presumptive hindbrain. *Hoxa1* expression domain extends from posterior end to presumptive hindbrain and restricted at future r3 by 7.5dpc of mice embryo development. *Hoxa1* expression is earliest indicator of hindbrain development. Loss of function phenotypes are dealt in detail in chapter1. But in short, Loss of *Hoxa1*^{-/-} leads to reduced number of rhombomere which results in alteration of inner ear and some cranial nerves. In *Hoxa1*^{-/-}, r4 and r5 and their derivatives are severely reduced and pups die shortly after birth. Gavalas and coworkers established that regulatory hierarchy between *Hoxa1*, b1 and b2 is important for control of early hindbrain patterning, patterns of neurogenesis and nature of differentiating neurons (Gavalas et al., 2003). Further, Gavalas and coworkers have shown that *Hoxa1* and *Hoxb1* play a synergistic role in second arch patterning and generation of cranial neural crest. Ectoderm specific regulatory mutant of *Hoxb1* in *Hoxa1*^{-/-} background resulted in reduced presumptive rhombomere 4. R4 lost the ability to generate neural crest cells in these mutants (Gavalas et al., 1998). These results indicate the importance of studying *Hoxb1* binding properties with respect to *Hoxa1* genome-wide binding properties.

Among many characterized co-factors of Hox genes, TALE (Three amino acid loop extension) protein are well studied group of proteins. TALE group of proteins consist of six protein families namely IRX, MKX, *MEIS*, PBC, PKNOX and TGIF. TALE family of transcription factors have implicated in many developmental processes and pathogenesis. *Pbx* is extensively studied among TALE protein due it its association as co-factor of Hox genes. Drosophila homolog *exd* (extradenticle) is functionally similar and act as cofactor of *Dfd*, *labial* and many more fly hox homolog(Mann and Chan, 1996). Vertebrate genome has 4 *Pbx* proteins namely PBX1, 2, 3 and 4. Many authors have shown that Physical interaction of HOX with *Pbx* raises their binding specificity hence improves *in vivo* site selection. This is achieved through heterodimer formation of PBX-Hox proteins on DNA in a sequence specific manner. This aspect is reviewed in detail in chapter1. Though Sequence specificity is the outcome of HOX-PBX heterodimer formation on specific DNA sequence, but functional outcome of this binding is solely determined by associated HOX genes. PBX lacks intrinsic ability of transcriptional activation or repression but is able to recruit co-activators and co-repressors. DNA binding specificity of HOX-PBX heterodimer is further fine- tuned through interaction with MEIS/PREP proteins. PBX interacts with HOX and MEIS/PREP using two different domains and a tethered binding site near to *HOX PBX* binding site (details in chapter 1). This interaction results in formation of ternary complex and helps in fine tuning binding specificity. MEIS regulate PBX further by helping its nuclear transport. These arguments clearly support that HOXB1 occupancy is essential to be compared and contrasted with genome-wide occupancy of PBX and MEIS proteins to fully understand mechanism of genome-wide site selection by HOXB1.

Third partner we were interested to study was REST proteins. In ES cells, REST is occupied on RE-1 site with cofactors like CO-REST, Sin3A, HDAC and MeCP2. Chromatin at these locus are enriched with di and tri methylation of K4 on histone 3. This indicates that in stem cells, neuronal genes are inactive but is in the permissive poised state. During terminal differentiation, REST is dismissed from RE-1 site and further transcriptionally down-regulated. In this way, REST plays a key role in regulation of neuronal genes in neuronal and non-neuronal tissues. It is clear for data shown and discussion in chapter 3 that HOXB1 might be the illusive activator of REST bound neuronal genes. This further raises our interest to understand unique nature of sequence and bound proteins on these HOXB1-REST Co-occupied genomic regions and their role in gene regulation

To address these questions, we set forward the following specific aims

1. Identify target sites in the genome for *Hoxa1*, *Pbx*, *Meis* and REST in differentiating ES cells.
2. Generate clones of KH2 cells containing epitope tagged *Hoxa1* and validate normal karyotype and Dox induction profile.
3. Characterize genome-wide occupancy of *Hoxa1* with epitope-tagged antibodies and protein specific antibodies of *Pbx*, *Meis* and REST in differentiating ES cells
4. Compare and contrast the genome-wide binding profiles of all *Hoxa1*, *Pbx*, *Meis* and REST to investigate shared and unique sites of interaction.

5.1 Result

5.1.1 Generation of Epitope tagged *Hoxa1* cells line

In collaboration with Mark Parrish, I have generated an ES cells line with triple flag_Myc with *Hoxa1* in KH2 cells. *Hoxa1* with triple flag-Myc was cloned in pBS31 (Fig.5-1). KH2 were engineered through lipofection and epitope tagged *Hoxa1* were inserted in Col II locus and put under control of doxycycline inducible promoter. All cell lines were tested for karyotype stability. FACS Calibur was used to analysis of DNA content and to get indirect inference of karyotype stability. Cells were induced with RA and doxycycline and induction was confirmed through western hybridization (data not shown).

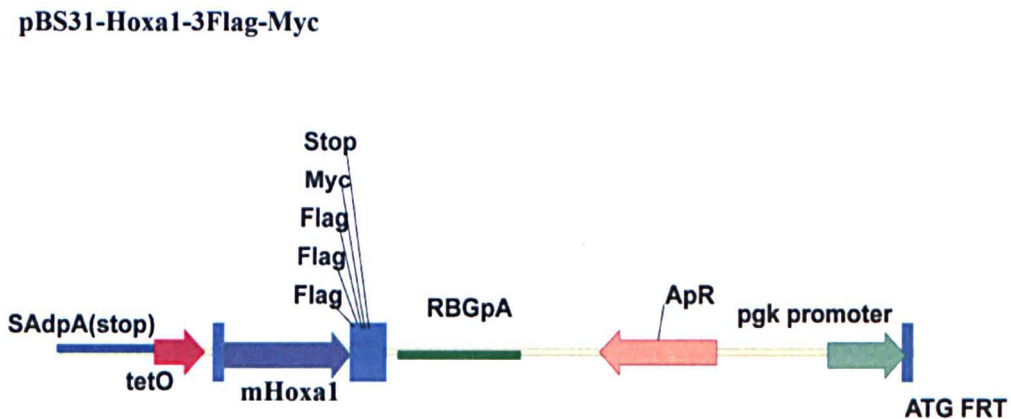


Figure 5-1 Construct used for expression of epitope tagged HOXA1

Mouse *Hoxa1* cDNA with 3 X Flag-Myc epitope tags is cloned in pBS31 vector backbone using homologous recombination at FRT site. This cDNA comes under control, of tetO and inducible with doxycycline.

5.1.2 Dynamics of HOXA1 occupancy in 24hrs RA and Doxycycline induced ES Cells

Two separate biological samples were induced with RA and doxycycline for 24 hours. HOXA1 bound chromatin was immuno-precipitated using M2 Anti-flag antibody. I did each ChIP experiment in triplicate and samples were pooled. Finally, we generated peaks set with 200-1000bp width and 5 fold enrichment. These Peaks sets were analyzed for identification of over represented motifs, enriched K-mers or K-mer pairs and analyzing co-occupancy with other Hox and Cofactor binding.

Our analysis revealed 529 genomic regions occupied by HOXA1 after 24 hours of RA and Doxycycline induction of ES cells. 3% of total occupied region fall within exonic regions of known genes. While total 57% peaks are present in intergenic region. Among these peaks, at least 8% peaks are within 10kb from the TSS (Transcription Start Sites). 39% of HOXA1 occupied are present in introns of known genes (Fig.5-2). This suggests that Hoxa1 occupancy is highly biased for intergenic and intronic region unlike in case of HOXB1. Only 2 peaks were present within 1kb at 5'UTR.

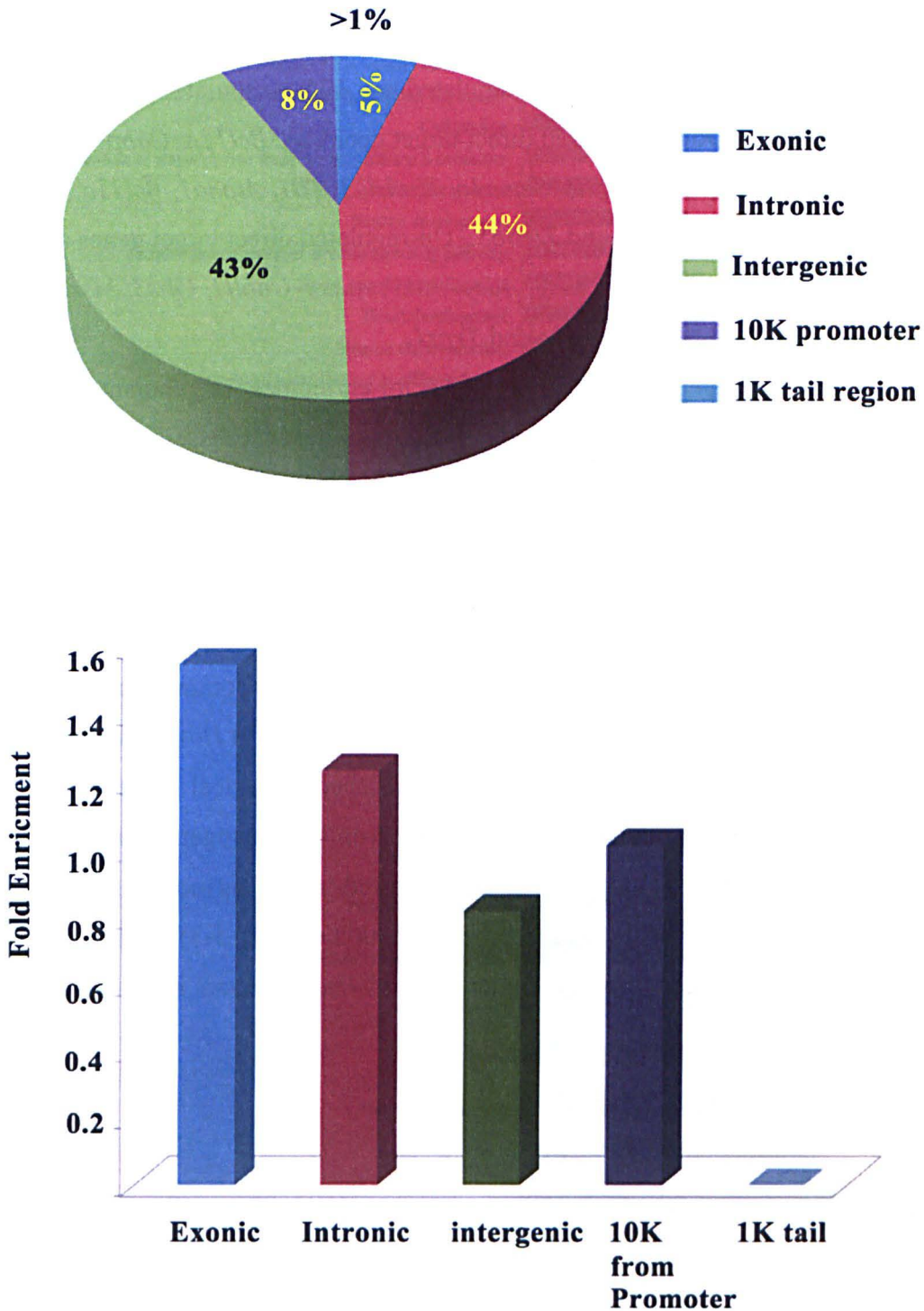


Figure 5-2 Distribution of HOXA1 occupied peak with respect to nearby pre-defined genomic features Pie chart shows distribution of HOXA1 bound peaks with respect to pre-defined genomic features like Exon, intron, genomic region within 10kb from TSS, within 5' UTR and intergenic region. Left panel shows enrichment of co-occurrence of these peaks in a given pre-defined genomic feature. Enrichment is calculated as ratio of % of peaks in a given pre-defined feature to % of that pre-defined feature in the mouse genome.

529 *Hoxa1* bound region have 484 nearest neighbor known gene. This suggests that most of the nearest neighbor genes have a single HOXA1 binding region. We compared up-regulated and down-regulated gene list from Marina's unpublished data of comparison

with *Hoxa1* *+/+* and *-/-* mice hind brain gene expression and Capecchi's lab (Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006) with our HOXA1 ChIP-Seq data to identify direct targets of *Hoxa1*. I identified 21 down regulated and 19 Up-regulated direct target genes for this comparison. Down regulated direct target genes are *Bcl11a*, *Cabp7*, *Dner*, *Exoc4*, *Hoxb1*, *Krt15*, *Lhx5*, *Mafb*, *Rgmb*, *Sema3c*, *Tbc1d23*, *Tll1*, *Atp8a1*, *Bcl11a*, *Col3a1*, *Irf2bp2*, *Kirrel3*, *Nr6a1*, *Sor11*, *Sox2* and *Zfp365*. Up-regulated direct target genes are *Alpk2*, *Ankrd1*, *Apob*, *Hoxa2*, *Lefty2*, *Tinag*, *Adcy7*, *Ak7*, *Auts2*, *Cobll1*, *Glis3*, *Msx1*, *Myo7a*, *Nme5*, *Nr2f2*, *Olig3*, *Otx2*, *Sema3c* and *Trps1*.

I analyzed 484 nearest neighbor genes for enrichment of gene ontology terms (GO terms). I restricted FDR at less than 1% for biological processes and cellular components. P value was restricted at less than e-4 and e-3 for biological processes and cellular component. I further discarded terms with less than 10 genes. This was to avoid false chance of getting higher enrichment due to low number of genes in the bin.

Most enriched five terms were mechanoreceptor differentiation, hindbrain development, ear development, inner ear development and cell fate commitment. It is important to mention at this point that one of the major differences in *Hoxb1* and *Hoxa1* phenotype is ear defect in *Hoxa1* mutant mice. Other enriched biological processes include pattern specification process, neuron development, embryonic morphogenesis and neuron differentiation. Enrichment for cellular component revealed 'Axon' as most enriched term. Other enriched terms were cell projection; neuron projection etc (Fig.5-3). These observations are in concurrence to observed phenotype of *Hoxa1* mutant mice (details in chapter1).

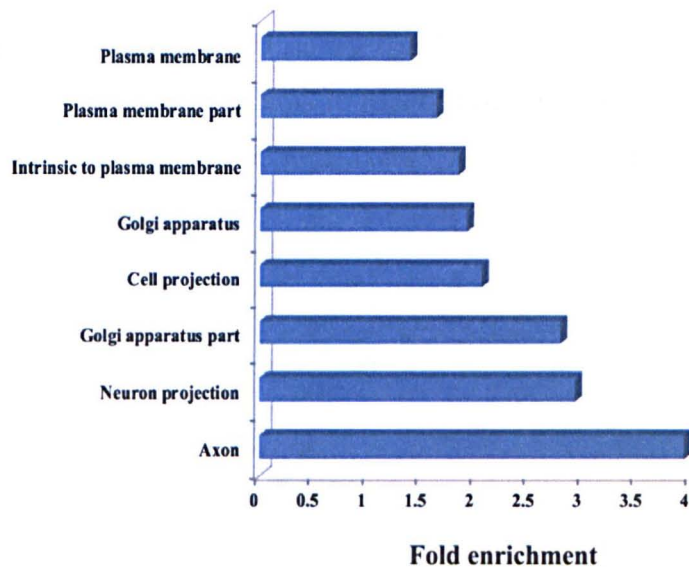
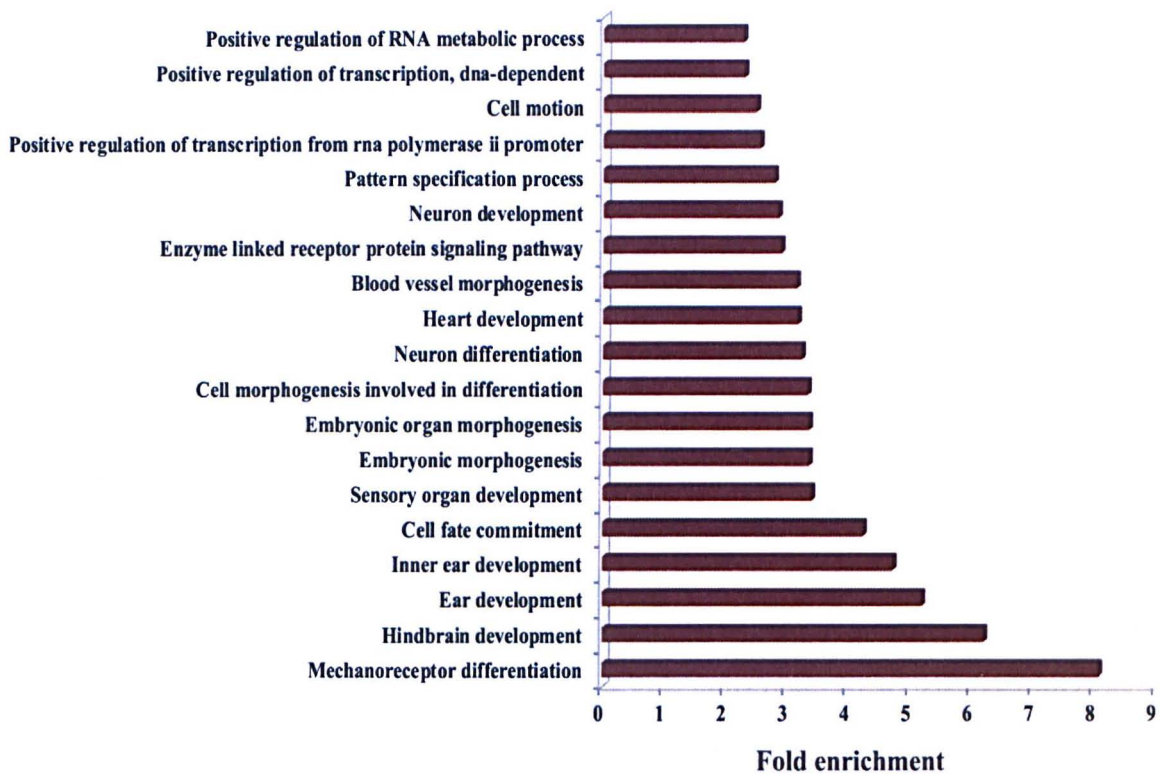


Figure 5-3 Enriched GO terms from nearest neighbor genes from HOXA1 binding genomic regions

HOXA1 occupancy was identified from bound peaks obtained in ChIP experiments on 24hrs RA induced KH2 cell with HOXA1-3XFlag-Myc. Nearest neighbor genes were identified and enrichment of GO (Gene ontology term) were analyzed. A. Biological Process. B. Cellular Component. Enriched terms were selected with FDR <1

I analyzed 484 genes for their relatedness with a common pathway through studying their relatedness in terms of gene-gene interaction network. I generated gene interaction network using these 484 genes. 2 different networks were identified in this network. First network was a relatively big network with 194 related genes (Fig.5-4). These 194 genes were mostly related with inner ear receptor cell differentiation, neuron fate commitment, mesenchymal cell differentiation, mesenchyme development, dorsal/ventral pattern formation, hindbrain development, ear development, cell fate commitment, axon guidance, sensory perception of sound and inner ear development. This large network can be further subdivided arbitrarily into five busy hubs. In Hub 1 have 18 genes and *Rgs9* and 7, *Ak7*, *Cenpc1*, *cdca2*, *Vme5* et.ac are some important genes. At least 5 genes in this hub is related to G-protein coupled receptor protein signaling pathway. one gene *AK7* shown to be up regulated in microarray studies described above.

Hub2 consist of 14 interesting genes. This hub consists of many important transcription factors. This hub consists of genes like *Pax3*, *Rgmb*, *Sox9*, *Nog*, *Gli3*, *Irx3*, *Msx2*, *Msx1*, *Olig2*, *Ptx2* and *lefty*. Among these *Rgmb* gene is down regulated while *Gli3*, *Msx1*, *Otx2* and *Lefty* were up regulated in previously described microarray experiments.

Hub3 consist mainly of Hox genes, Cofactors and other developmentally important transcription factor. This hub consist of *Hoxa2*, *hoxb2*, *Hoxb4*, *Meis2*, *Pbx3*, *Tcf21*, *Robo1*, *Sema3c*, *Nln1*, *Unc5a*, *Epha4*, *Epha2*, *Epha3*, *Epha8* and *Efna5*.

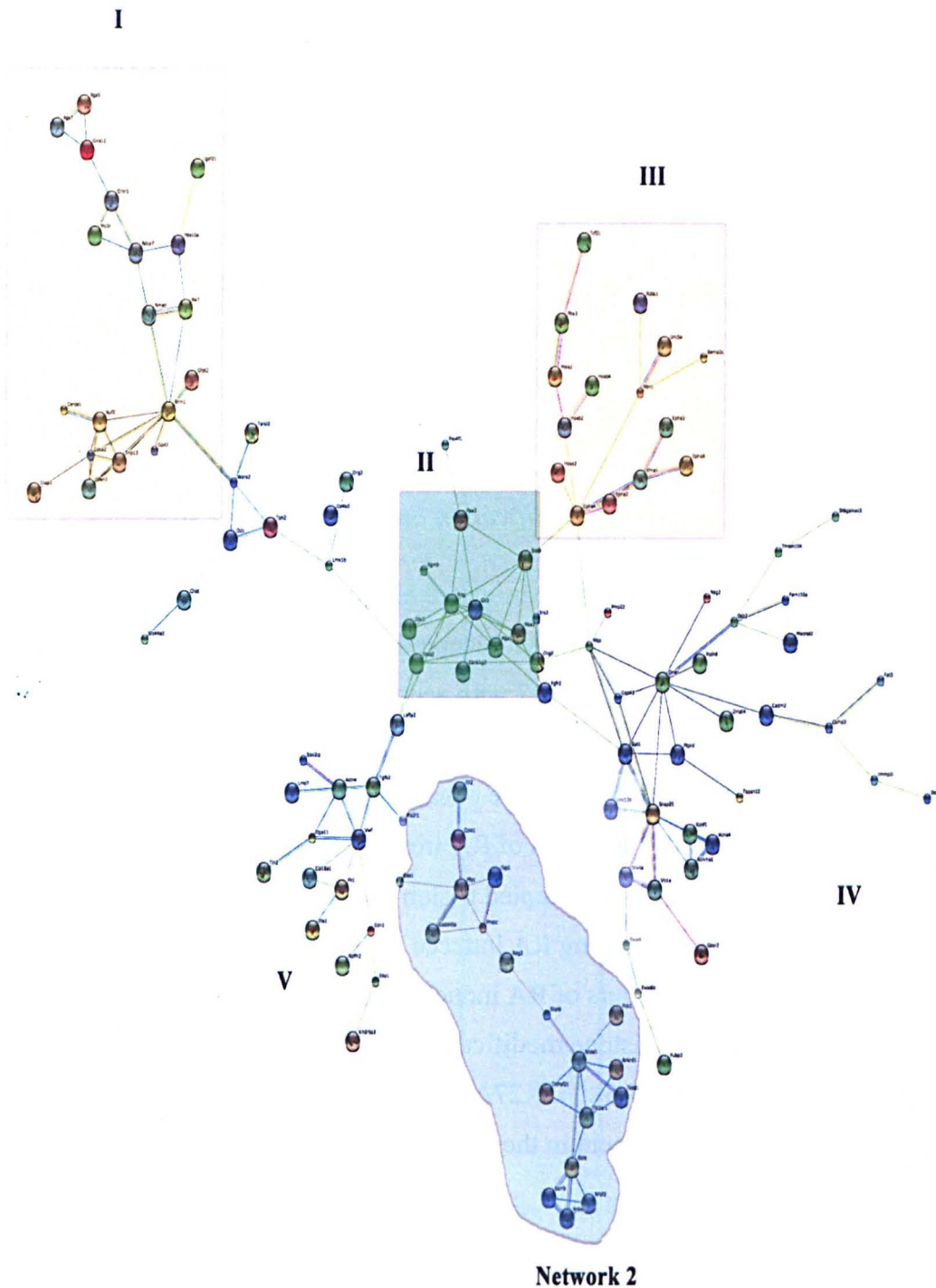


Figure 5-4 Gene interaction network for genes with nearby HOXA1 bound region
 A Identification of gene interaction network using String ver 9.05. Nearest neighbor genes with HOXA1 bound region were analyzed for evidence of known and predicted interactions. Direct (physical) and indirect (Functional) interactions were analyzed using information like genomic context, conserved expression pattern, high throughput experiments and publications. Highly enriched network with three dense hubs formed by 119 genes were identified Two networks can be seen in this figure. Network one can be further sub divided into five distinct hub. B. Enriched GO term from genes forming interaction network.119 genes showing interaction were analyzed for enrichment of specific GO term.

Hoxa2 is up regulated while Sema3c is down regulated in *Hoxa1* mutant hindbrain. Interestingly, many epharins are part of this network and it is known that r4 and r5 is reduced and merged to r6. This cell sorting defect may be due to changes in this gene interaction hub.

Hub 4 consists of 24 genes. These genes are mainly related to plasma membrane related terms and GO term analysis shows enrichment of term related to transmission of nerve impulse. These genes are also related to SNARE interactions in vesicular transport.

Hub 5 consists of 15 genes. These genes are related to adhesion. At least one gene; Col3a1 from this hub is found to be down regulated in mutant hindbrain. KEGG pathway analysis shows that genes in this hub are involved in adherens junction, focal adhesion and cardiomyopathy.

5.1.3 Dynamic changes in Histone modification on future HOXA1 sites

We first checked presence of enhancer related histone modification over *Hoxa1* bound region (Fig.5-5). We analyzed HOXA1 occupied region for Histone 3 for modified lysine 4 with mono methylation and lysine 27 with acetylation. Interestingly at least 50% future HOXA1 occupied region have mono methylation at lysine 4 of histone. While around 15% of these sites are also having active enhancer marks, H3K27Ac. H3K4Me1 marks can be seen more frequently as length of RA treatment increases. After 24 hours of RA treatment, around 90% of HOXA1 occupied region shows H3K4Me1. Active enhancer marks also show considerable gain during RA induced differentiation. Around 40% of HOXA1 occupied region after 24 hours of RA induction have bivalent mark of H3K4Me1 and H3K27Ac. Presence of these histone modification indicates that HOXA1 bound regions are putative enhancer (Fig.5-5). H3K27Ac and H3K4Me1 in uninduced ES cells and after 6 and 24hours of RA treatment in these regions are also shown.

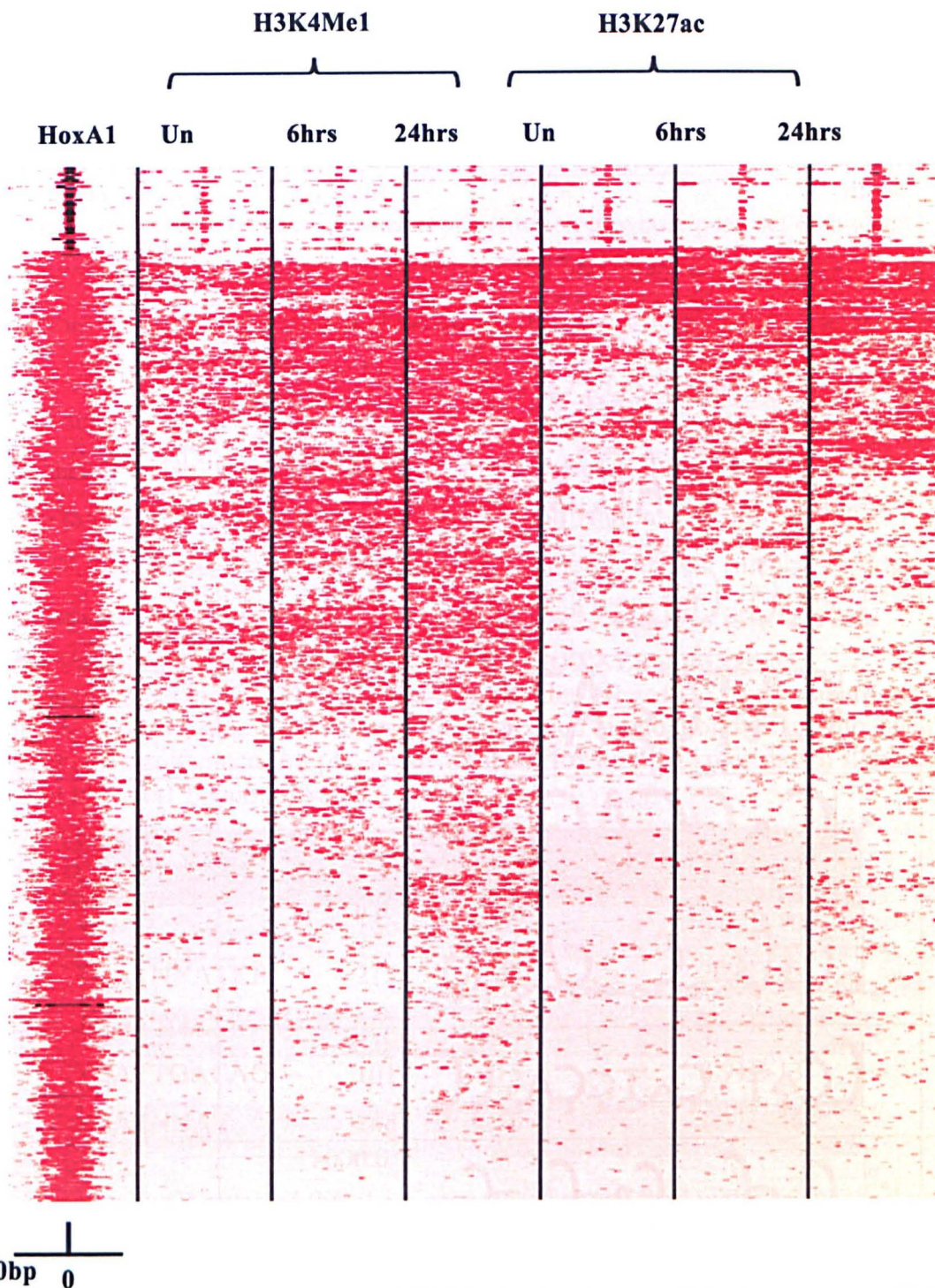











Figure 5-5 Genome-wide occupancy of HOXA1 after 24hrs of RA Treatment of ES cells

All HOXA1 occupied regions from 24hrs of RA induced ES cells are shown on Y-axis. Every row is a genomic coordinate with HOXA1 occupancy. Each column represents onetime point antibody combination. Heatmap is hierarchically clustered. Each genomic coordinate is centered on midpoint of HOXA1 peaks and region including 500bp either side of midpoint is shown in heatmap. Enhancer specific histone modification namely

Table 5-1 Over-represented motifs in HOXA1 bound region

Over-represented motifs were identified from 100 top ranking HOXA1 bound regions and their enrichment P-value were calculated from 24hrs RA and Dox treated ES cells. Known transcription factor binding to these motifs are shown in TF column.

SN	Motif	p-Value	Known TF factors
1		0.0004	
2		1.42E-58	Pbx
3		3.99E-10	Krox
4		3.95E-48	Pbx
5		2.67E-08	
6		0.015	
7		0.06	
8		0.0004	
9		1.06E-06	

5.1.4 Identification of over-represented motifs in genome-wide binding site

Next, we were interested to know what kinds of sequences are over enriched in these *Hoxa1* binding regions. We used MEME to identify over represented motifs. We selected top 100 peaks and identified highly over represented 20 motifs.

These motifs were tested for over-representation in whole peak set data using FIMO. Cut off p-value were set at less than 0.001. We identified nine over represented motifs (Table 4.1). We identified three known motifs and six novel motifs. Three known motifs were for two for *Pbx* and one for *Krox*.

To our surprise though these sequences are enriched with motifs for PBX, but we were not able to see any HOX-PBX bipartite site as enriched motif in these peak test. We decided to look for these sequences using K-mer identification. We analyzed pre-defined K-Mer in HOXA1 occupied peak. We set e-4 as P value cut-off and calculated enrichment against a random background set.

Table 5-2 Enrichment of pre-defined K-Mers in HOXA1 bound region

Enrichment of kmers in all HOXA1 bound regions after 24hr of RA treatment are shown in this table. A random background with similar nucleotide distribution was used to calculate enrichment and significance. All previously defined Hox binding kmer shown significant enrichment in *Hoxa1* bound region

NAME	FG.W	LOG2FC	ADJ.P
WRATNNATKR	186	2.41	5.02E-70
TGATNNATKR	212	2.46	2.82E-83
TGATTGAT	108	3.26	3.54E-56
TGATTGACAG	101	1.15	8.91E-13
TGATNNAT	148	2.15	1.31E-46

K-mer analysis revealed that these sequences are enriched with various versions of HOX-PBX bipartite site. Some K-mer variants of HOX-PBX bipartite sites are enriched more than 8 fold above background (Table 5-2). This observation leads us to analyze distribution of these K-mers in HOXA1 bound region. Interestingly, we observed that majority of HOXA1 bound region do contain a centrally located Hox-Pbx bipartite motif (Fig.5-6).

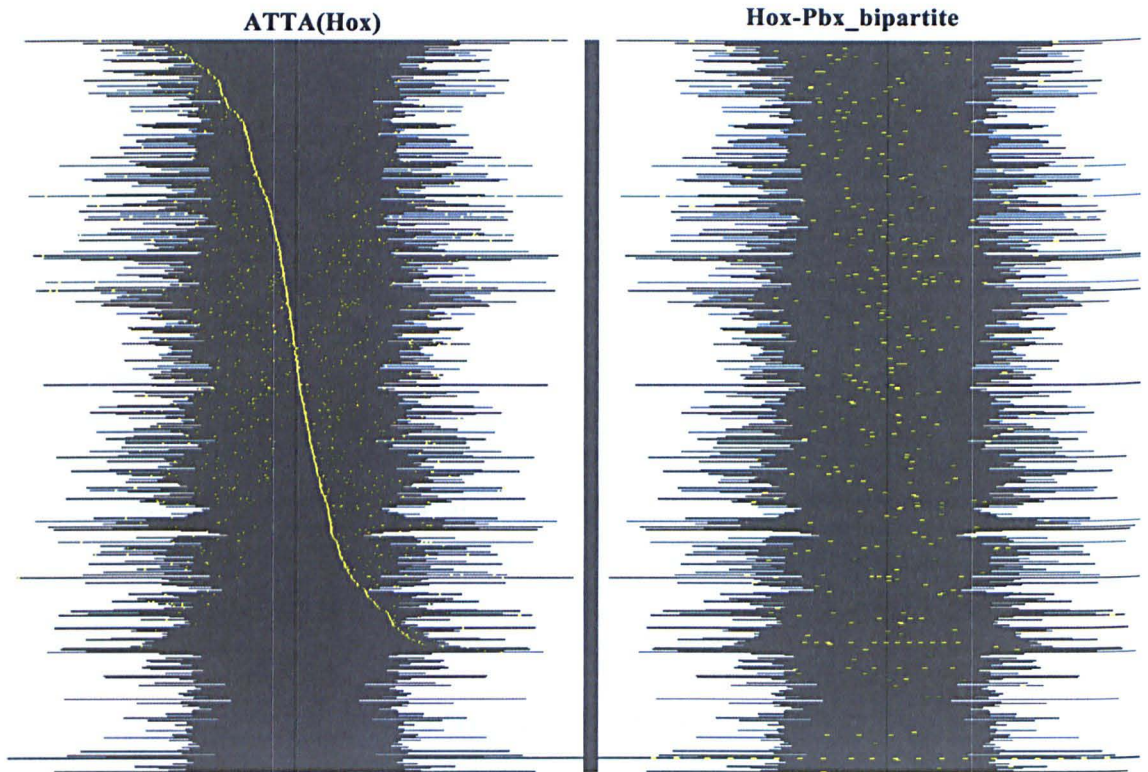


Figure 5-6 Relative spatial distribution of ATTA (HOX) and HOX-PBX bipartite motifs in HOXA1 bound region

Position of motif is shown with respect to center of HOXA1 peak. Each yellow dot represents presence of one kmer. Each row represent genomic region bound by HOXA1 and shadowed area represents span of each peak. Peak with given kmer at farthest right from center is shown at the top while Peak with kmer at farthest left from center is shown at the bottom. Other kmers are shown relative to kmer based on which data is sorted. Co-occurrence of ATTA and ATTAG k-mers in Hoxa1 bound region is shown in this figure. Kmers are sorted based on position of ATTA K-mer

Table 5-3 Co-enrichment of K-mers

Co-Enrichment of two kmers in all Hoxb1 bound regions after 4hr of RA treatment are shown in this table. A random background with similar nucleotide distribution was used to calculate enrichment. All previously defined Hox binding kmer shown significant enrichment in HOXA1 bound region

1st K-mer	2nd K-mer	occurrence	Log2FC
ATTAG	TGATNNATKR	106	2.99
TGACAG	TGATNNATKR	80	3.78
ATTA	TGATTGAT	76	3.86
TGACAG	TGATNNAT	124	2.46
TGACAG	WRATNNATKR	124	2.19
ATTAG	TGATTGAT	50	4.21
ATTAG	TGATNNAT	148	1.71
ATTA	TGACAG	197	1.27
TGACAG	TGATTGAT	39	4.7
ATTAG	WRATNNATKR	155	1.33
ATTAG	TGACAG	128	1.48

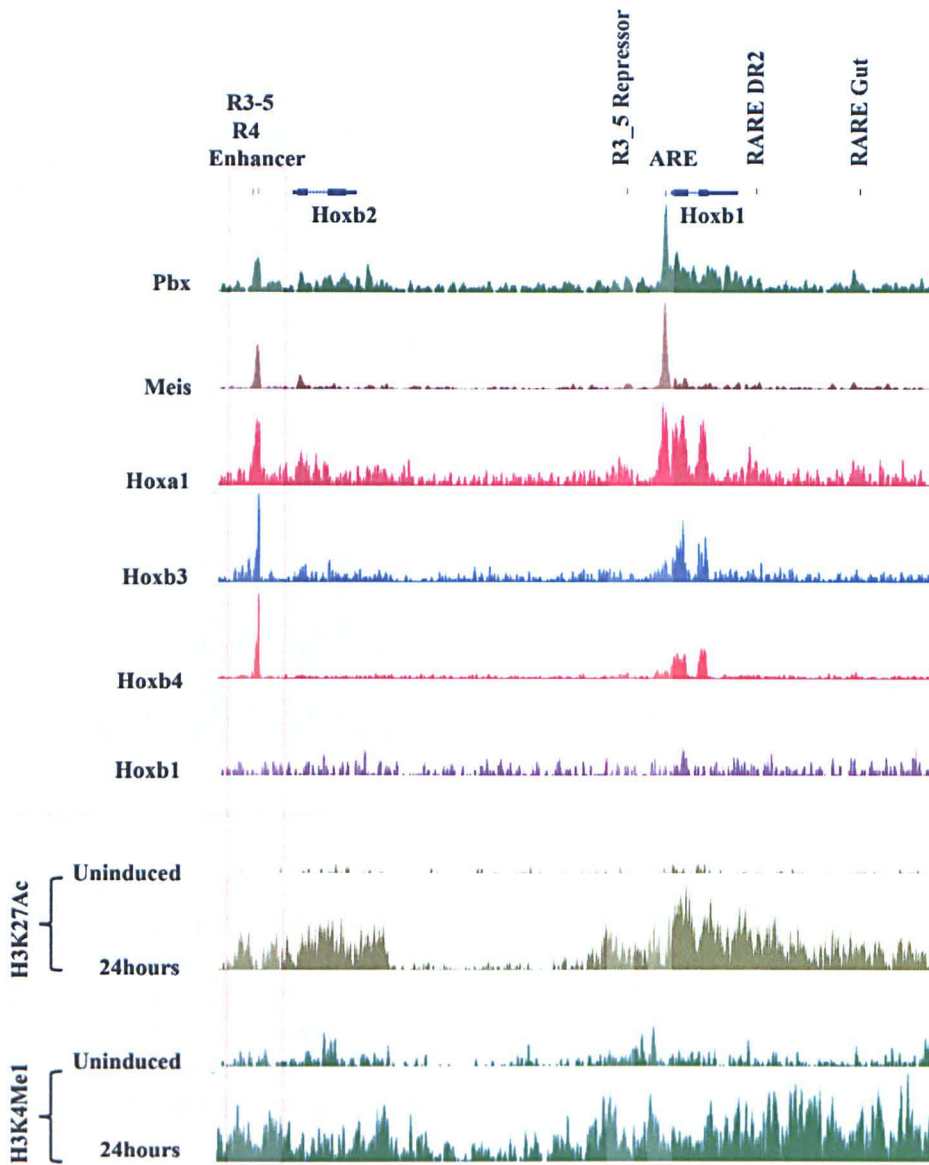


Figure 5-7 Co-occupancy of HOXA1 with cofactors and other Hox genes

UCSC genome browser snapshot of region showing co-occupancy of HOXA1 and Tale proteins -PBX and MEIS. Height of peak corresponds to read coverage in this region. Y-Axis is different for each antibody. Co-occupancy of HOXA1 with PBX and MEIS on *Hoxb1* ARE and *Hoxb2* R3_5 enhancer is noticeable. Distinct enhancer marks of H3K4Me1 and H3K27Ac can also be seen ver both regions.

We further analyzed significance of K-mers pairs and their combined enrichment. It seems that HOX-PBX-MEIS K-mer pairs are significantly enrichment in HOXA1 bound region (Table 5-3).

We looked into co-occupancy of HOXA1, PBX and MEIS in previously reported ARE (Auto-regulatory region) upstream of *Hoxb1*. This ARE contains HOX-PBX-bipartite site and shown to be important for recruitment of HOXA1, PBX and MEIS. We do see occupancy of HOXA1, PBX and MEIS on ARE region (Fig.5-7).

We can further see co-occupancy of HOXA1 with PBX, MEIS, HOXB3 and HOXB4 on R3_5 enhancer downstream of *Hoxb2*. This clearly indicates that HOXA1 shows strong association with PBX and MEIS at many previously functionally identified elements. We further looked at genome-wide physical Co-occupancy of *HOXA1* and TALE proteins. It was interesting to see that more than 90% of HOXA1 occupied regions are co-occupied with PBX and Meis (Fig.5-8). These results together indicate that TALE protein PBX and MEIS is an important cofactor for HOXA1 and *HOX-PBX* bipartite sites play important role in recruitment of HOXA1. All *Hoxa1* bound genomic region after 24hrs of RA Induction and all HOXB1 occupied regions from 6hrs, 12hrs and 24hrs of RA induced ES cells are shown on Y-axis. Every row is a genomic coordinate with either HOXA1 or HOXB1 occupancy in at least one time point. Each column represents onetime point antibody combination. Heatmap is hierarchically clustered. Occupancy of PBX and MEIS are also shown in these genomic region Each genomic coordinate is centered on midpoint of HOXA1 or HOXB1 peaks and region including 500bp either side of midpoint is shown in heatmap. Large numbers of HOXA1 bound regions are co-occupied with PBX and MEIS.

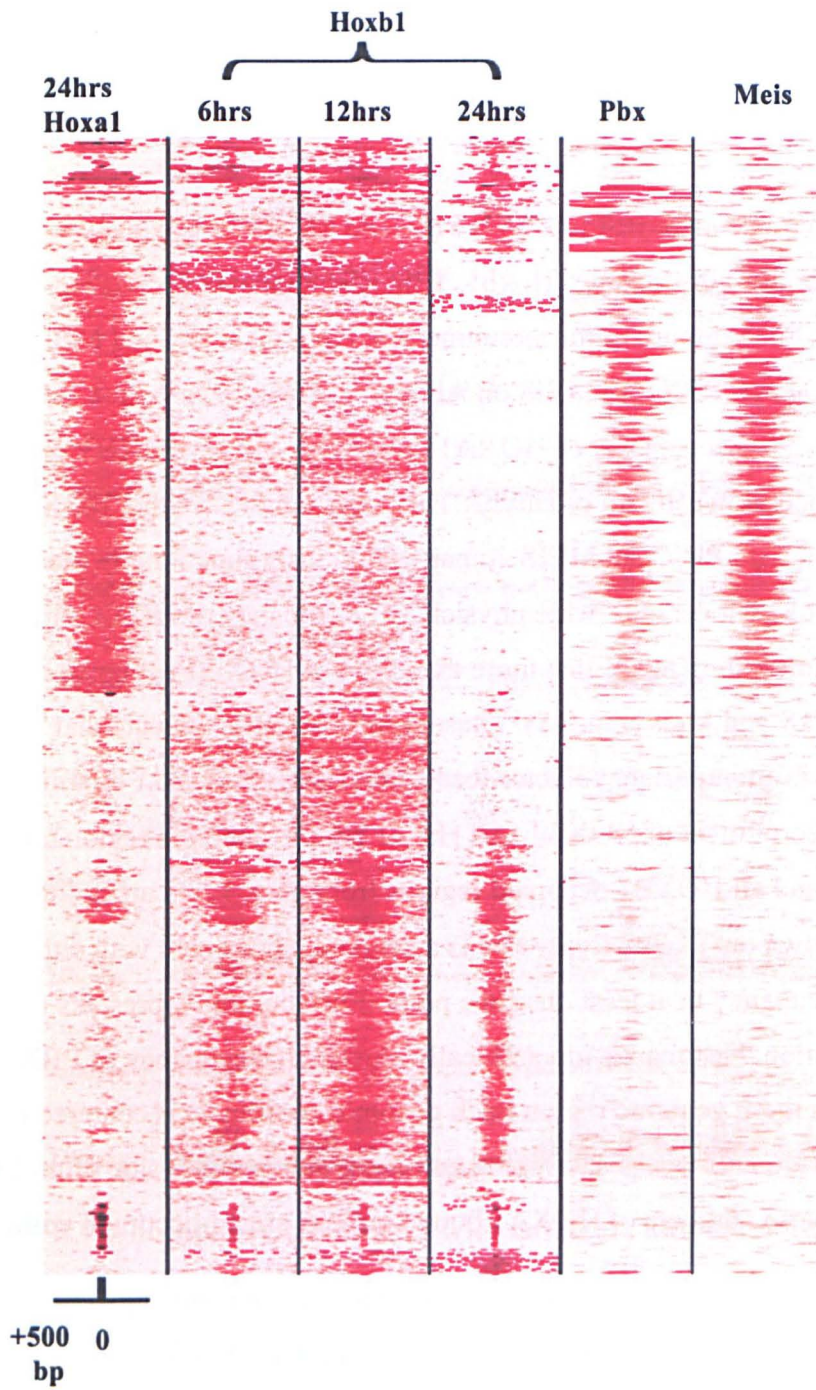


Figure 5-8 Genome-wide occupancy of HOXA1, HOXB1, PBX and MEIS in differentiating ES Cells

5.1.5 Combinatorial binding of Paralogus group 1, Co-factors and REST

As throughout this thesis general thread of emphasis was to understand genome-wide binding specificity of Hoxb1. But this needs to be understood with comparing and contrasting with Hoxa1. We first tried to understand Co-occupancy of HOXA1 and HOXB1 in a differentiating ES Cells. By looking at HOXA1 and HOXB1 genome-wide occupancy at 24hrs of RA induced ES cells, it seems that both paralogus genes have a very distinct genome-wide binding profile. Less than 5% HOXB1 bound regions are co-occupied with Hoxa1. Hoxa1 shows significant co-occupancy with PBX and MEIS (Fig.5-8). More than 90% HOXA1 bound regions are co-occupied with PBX and MEIS while around 10% of HOXB1 shows co-occupancy with PBX. Very poor occupancy of MEIS is seen over HOXB1-PBX co-occupied region. Small number of HOXB1 and HOXA1 co-occupied regions can be further classified into two distinct classes. One class shows co-occupancy of HOXB1, HOXA1, PBX and low levels of MEIS while other class shows HOXB1, HOXA1 and REST (Fig.5-9). Interestingly, this small subset of HOXA1 occupied region (Co-occupied with HOXB1) is only region devoid of TALE protein Co-occupancy. It was surprising to see that PBX and MEIS are not much co-occupied in HOXB1 bound region. But looking at this data along with genome-wide occupancy of REST, it is evident that at least one third of HOXB1 occupied regions are co-occupied with REST. These regions are mostly devoid of PBX and MEIS binding (Fig.5-9). Our results indicate that though HOXA1 and B1 are paralogus group 1 but they interacts with different co-factors and have a very different binding specificity. In this context, it may be possible that Co-occupancy of PBX and MEIS might be determinate of HOXB1 specificity while REST or REST associated proteins dictates genome-wide binding specificity for HOXB1.

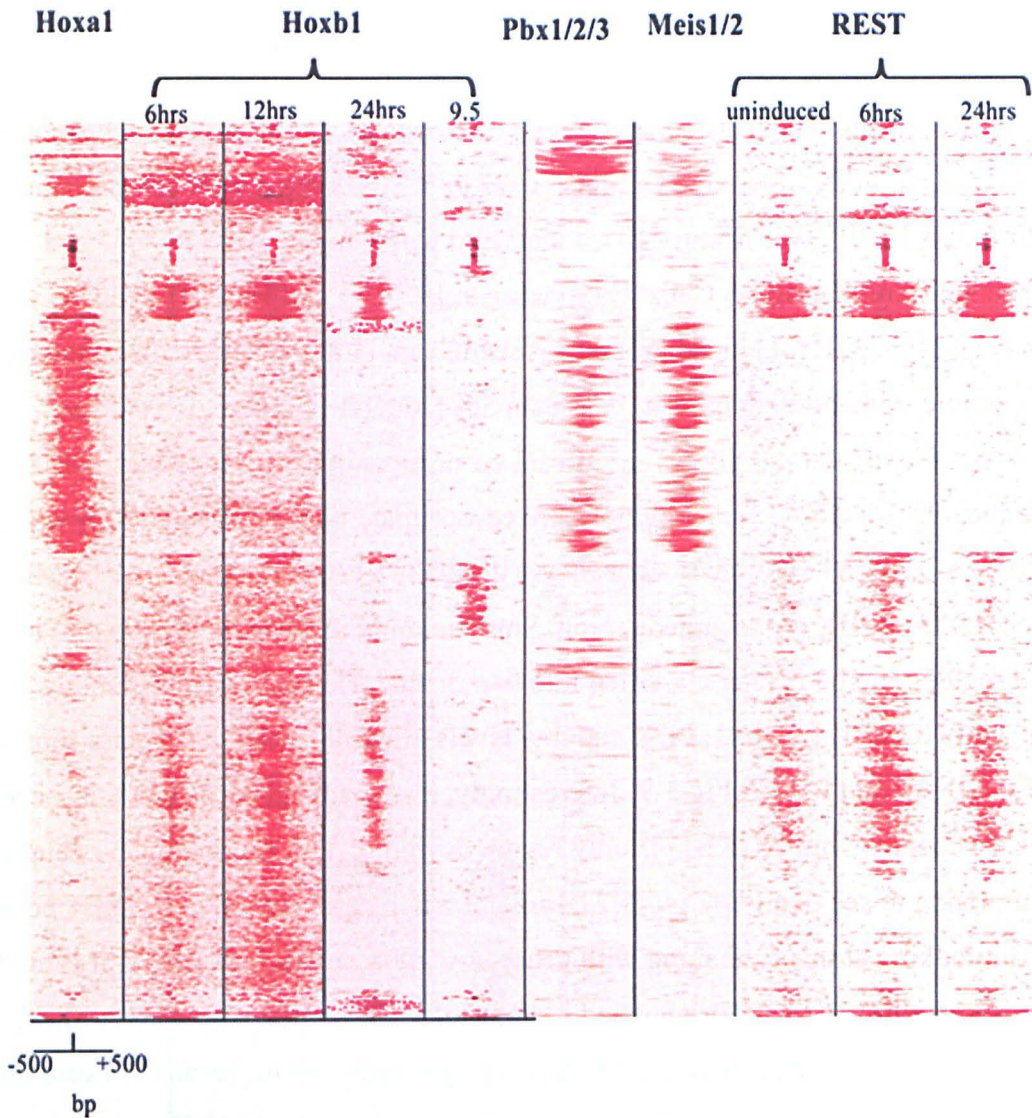


Figure 5-9 Genome-wide occupancy of HOXA1, HOXB1, PBX and MEIS in differentiating ES Cells

All Hoxa1 bound genomic region after 24hrs of RA Induction and all HOXB1 occupied regions from 6hrs, 12hrs and 24hrs of RA induced ES cells are shown on Y-axis. Every row is a genomic coordinate with either HOXA1 or HOXB1 occupancy in at least one time point. Each column represents one time point antibody combination. Heatmap is hierarchically clustered. Occupancy of PBX and MEIS are also shown in these genomic region. REST occupancy in uninduced ES cells and 6 and 24hrs RA Induced ES Cells are also shown. Each genomic coordinate is centered on midpoint of HOXA1 or HOXB1 peaks and region including 500bp either side of midpoint is shown in heatmap. Large numbers of HOXA1 bound regions are co-occupied with PBX and MEIS. Large numbers of HOXB1 bound regions are co-occupied with REST. PBX and REST binding are mutually exclusive.

Chapter 6 Summary discussion

Morphogenesis during embryonic development in mammals involves the control of basic processes, such as proliferation, growth, differentiation, migration, patterning, segmentation and survival. These processes are generally controlled by signaling cascades and transcription factors and much of the current efforts in developmental biology have been devoted towards understanding transcriptional and signaling regulatory networks that function in context dependent manners to elaborate the body plan.

Among these processes, segmentation plays an important role in determining anterior posterior patterning in the axial skeleton and hindbrain (Alexander et al., 2009). In these contexts, segmentation generates a reiterated series of distinct compartmentalized units along the body axis which then responds to local signals and intrinsic cues to adopt unique identities. Compartmentalization and the formation of boundaries in the neuroectoderm plays a key role in generation of morphologically distinct regions of the brain (Kiecker and Lumsden, 2005). Segmentation of hindbrain into 7 distinct lineage-restricted cellular compartments (rhombomeres) provides a ground plan that controls formation of both central and peripheral nervous system components, such as cranial motor nerves, sensory ganglion and neural crest cells and their derivatives. Segmentation of paraxial mesoderm leads to generation and formation of repetitive structures called “somites”. These somitogenesis generates what appear to be morphologically similar building blocks but they go on to generate the diverse vertebrae, bones and muscles of the axial skeleton. The somitogenesis and axial elongation are controlled through morphogen opposing gradients of RA and Fgfs and these signaling systems also play a role in patterning the A-P character of the adjacent CNS. Hence there is a close relationship between axial patterning and segmentation in both the CNS and paraxial mesoderm (Diez del Corral et al., 2003; Diez del Corral and Storey, 2004; Young et al., 2009). These morphogen gradients are capable of inducing the Hox homeotic selector genes in restricted manner and they play a key role in specifying positional identity in these two tissues (Alexander et al., 2009; Diez del Corral and Storey, 2004; Mallo et al., 2010).

Understanding how the Hox genes are coupled to these signaling gradients and how different outcomes are mediated by the different members of the Hox family of proteins is central to building knowledge on control of morphogenetic processes. Genetic studies have provided valuable and general insight into aspects of the specific and overlapping functions regulated by Hox proteins. However, these studies highlight a number of major

gaps in our knowledge. Each Hox protein may have specific DNA binding properties through which it exerts its unique function, but how then is the binding of other Hox proteins on common targets achieved to explain their shared or overlapping functions? It is relatively easy to understand different functional roles for related Hox genes if these are generated by differences in their spatial or temporal expression domains. However, the rules or principles which underlie the binding of similar or different Hox proteins to the same or distinctly different target sites, when they are expressed, is poorly understood. Hence, understanding what types of sites Hox proteins bind *in vivo*, and what governs the binding specificities of Hox proteins critical for their regulatory function is a fundamentally unsolved problem. Invoking the action of cofactors such as Pbx, Meis (TALs) provides valuable insight to the problem. However, are these the only cofactors for Hox proteins and how do they participate in specificity of different Hox proteins?

It is logical to assume that combinatorial binding, diverse cofactors, different classes of binding sites and variations in spatial and temporal expression all play important roles in final determination of Hox target genes. Investigation of these components have been hindered because of limiting amounts of *in vivo* tissue for biochemical or genomic studies, a lack of antibodies and the bioinformatic challenges of examining properties on a genome-wide basis rather than a few examples of target sites. Recent advances in sequencing, epitope-tagging, ChIP and ES cell technologies have provided a valuable opportunity to undertake such studies in cost effective manner. In this thesis research I decided to utilize these technologies to set up a system to systematically investigate the binding properties of two paralogous 1 group genes, namely Hoxa1 and Hoxb1 on a genome-wide basis. The goal was to identify binding sites by ChIP-seq and analyze these peaks to determine DNA binding specificity and characteristics at a genome-wide level. The hope was that this would also provide insight on the degree to which known cofactors such as PBX and MEIS or novel cofactors might influence binding of Hoxa1 and Hoxb1. I selected these two Hox proteins because they were known to have functional overlaps and are believed to share many common target genes. They would therefore serve as a model for the degrees of similarity in binding of two Hox proteins.

Programmed differentiation of ES cells into a neuro-ectoderm like character with retinoids was the system I chose to provide similar tissue contexts and sufficient material for genomic and biochemical assays. I selected KH2-ES cells because they provide a convenient means for site-specific integration of cDNAs encoding epitope-tagged proteins at a single defined target site in the genome at a promoter under tight Tet control to modulate levels of expression (Beard et al., 2006). I performed a comprehensive

characterization of the temporal dynamics of the neuro-differentiation process in KH2 cells. Transcriptional profiling of a detailed time course of differentiation in response to RA was done with a variety of platforms: Affymetrix arrays, RNA-Seq, Agilent high density Hox tiling arrays (designed by our informatics group) and ABI qt-PCR arrays. These experiments enabled me to determine the *precise* order, timing and levels of gene expression of all 39 Hox genes; identify novel non-coding transcriptional activities in and around Hox clusters; and globally characterize rapid changes in gene expression during differentiation (Chapter 3). The results from this work provided a basis to understand and compare ES differentiation with normal hindbrain and spinal cord development. I also used chromatin immune precipitation (ChIP) and high density Hox tiling arrays and/or next gen sequencing in combination with a variety of antibodies against active and repressive histone marks, RNA Pol II (N-term and CTD regions) and RARs & RXRs retinoid acid receptors. These results generated a detailed picture of the accessibility and dynamics of the epigenetic states of Hox clusters related to their transcriptional activity and identify new sites of potential direct input by RA signaling through occupancy of receptor binding.

The main rationale for these analyses was to understand when specific Hox genes are activated and to measure their relative levels of expression so I could conduct genomic experiments to identify binding sites of Hox proteins under appropriate conditions. However, these results have generated a wealth of interesting data and led to productive collaborations with other groups at the Stowers Institute. As expected from published work and the expression patterns during embryonic development, *Hoxa1* and *Hoxb1* are rapidly induced after RA treatment. I see transcripts as soon as 2 hours of RA treatment and this rapid activation can be attributed to RAREs present in the 3' regions of the *Hoxa1* and *Hoxb1* genes. In collaboration with the Shilatifard group, my analysis of the genes most rapidly induced by RA, including these Hox genes, led to the discovery that the majority of the rapidly induced loci are regulated by transcriptional elongation as opposed to initiation of transcription as previously believed (Lin et al., 2011). *Hoxa1* comes on more rapidly than *Hoxb1 in vivo* in both mouse embryos and ES cells. The ES experiments showed that *Hoxa1* has paused polymerase II (Poll II) over the *Hoxa1* transcription unit in uninduced ES cells which is rapidly elongated by the Super Elongation Complex (SEC) in response to RA. In contrast *Hoxb1* does not display paused Poll II, and its induction is mediated through new initiation. This highlights surprising differences in the nature of the RA response for these two genes. The ChIP-chip experiments in Chapter 3 using antibodies against RAR-alpha, beta and gamma, also uncovered surprising features of the RA response for *Hoxa1* and *Hoxb1*. The retinoid receptors are bound to the *Hoxa1* region and

removed by addition of RA while they remain present on *Hoxb1* in untreated and treated cells. This implies retinoid receptors might be involved in repressing the *Hoxa1* gene via SMART or NcoR and this repression is eliminated upon RA treatment. On *Hoxb1* the addition of RA may convert an inhibitory complex to an activating complex. In the future it would be interesting to understand the *cis*-elements and mechanism that underlie these differences. They may lie in the promoters or in the chromatin signatures. This also relates to the question of how paused Poll II appears on some promoters but not others. Analyses to compare and contrast these two related Hox group 1 genes might shed useful insight on these events.

Julia Zeitlinger's group discovered that a poised or balanced state, defined by aspects of Poll II and H3K27me3 occupancy, has predictive value in identifying genes that will be expressed in *Drosophila* mesoderm differentiation. Extending this to my ES cell data revealed that this balanced state exists in mammalian cells and also reflects genes prepared to be expressed (Gaertner et al., 2012). Hence, this ES system and programmed differentiation I established and characterized for my thesis research is becoming a valuable shared database of meta information that serves as an integrative hub for diverse interests on regulation of transcription and epigenetic states for the Krumlauf laboratory and for Institute groups.

One of the most surprising sets of findings to emerge from my characterization of transcriptional profiles of the Hox clusters in ES cells and their differentiated states was the extensive degree of transcription from both strands of the *HoxA* and *HoxB* clusters. There has been evidence for anti-sense transcription for many years and more recently microRNAs, such as *miR10*, and non-coding RNA, such as *HotAir* or *HotTip*. These RNAs are believed to play regulatory roles for both Hox and non-Hox transcripts in a variety of contexts. My data revealed that there was much more extensive transcription of non-coding transcripts from both the sense and anti-sense strands. The variety of these changes over time and many are present in the mouse embryo. Hence they represent a number of new candidates that need to be examined in relation to regulation of Hox genes. One of the most interesting of these is Heater, located 50 kb upstream of *Hoxa1*. It has multiple transcripts stimulated by multiple RAREs and there is evidence for regulation by paused Poll II on some of these transcripts. It appears that this region which responds to RA faster than any of the Hox genes might play a role in potentiating the expression of adjacent *HoxA* genes such as *Hoxa1*. There is some emerging evidence that this Heater region functions as a lincRNA (Maamar et al., 2013) and my characterization of this region

should form a useful starting point for examining its potential regulatory roles. This is something I feel worth pursuing in the future.

In characterizing these RA induced transcripts from both strands of the Hox complexes I have also collaborated with our Bioinformatics Core Team to analyze other aspects of transcription. Investigating data from RNA-seq provides some indication that splicing patterns are distinct at different stages. I am working with Marco Blanchette and Ariel Paulson to work on computational approaches to examine this data in more detail based on approaches Marco Blanchette developed to study differential splicing patterns in *Drosophila*. Hence, there is a wealth of information stored in the ES cell transcriptome data I generated that can be mined for useful insights on many aspects of transcription.

The Hox tiling arrays and detailed time course of differentiation in response to RA clearly showed that the four Hox clusters differ in their response to RA. *HoxA* and *HoxB* clusters show the strongest and most rapid RA responses. Based on published work it was expected that there would be a clear co-linear activation of successive genes in each cluster. In general there was some support for this expected trend, anterior (3') genes generally showed more rapid activation while posterior genes exhibited either weak or no activation. However, to our surprise, the order of gene activation does not follow a strict pattern of co-linearity. There was wide gene by gene variation and some of the most rapidly induced transcripts actually mapped to non-coding regions. With the extensive degree of non-coding transcription I realized that predicting a co-linear order of activation of transcription based only on the coding transcripts might not be informative.

Many models for the ordered activation of the Hox genes postulate changes in the epigenetic state which then lead to transcription. However, it is also possible that epigenetic marks follow transcription events and do not determine them. To understand role of epigenetic modification on Hox gene activation, I analyzed epigenetic changes at Lysine 27 tri-methylation (H3K27me3) a mark for Polycomb repression and at lysine 4 tri-methylation (H3K4me3) a mark for activation 3 at few selected time points during RA mediated differentiation of ES cells.. Gradual losses of H3K27me3 were observed upon increased duration of retinoic acid treatment in Hox A and HoxB locus. Genes were expressed before removal of this repressive mark indicating that it is not essential to eliminate it to permit expression. The non-co-linear activation of some of the Hox genes therefore, cannot be explained by the loss of the H3K27Me3 mark and gain of H3K4Me3 mark. An anterior to posterior gradient in changes of the repressive mark of H3K27Me3 are observed upon RA induced repression in that the anterior region of the cluster losses

the repressive mark earlier compared to posterior region. However, it takes more than 24 hours of RA treatment for complete loss of repressive mark H3K27Me3 during differentiation. The appearance of the active H3K4me3 mark is fairly rapid and dynamic indicating that it does reflect genes or regions which are active. However, some of these marks appear over the non-coding transcripts in the clusters. If one was unaware of the correlation of these epigenetic changes to non-coding transcripts they might be attributed to roles in regulation of adjacent coding genes. This illustrates the need for more careful analyses of the dynamics of the epigenetic state in Hox clusters and to which aspects of the transcriptome they reflect or regulate.

Many interesting questions arose from this ES study and I wished to utilize the system for genome-wide binding studies so I was unable to pursue many exciting areas. Interesting points worthy of future investigation are: Does the removal of H3K27Me3 on promoters and other regions of the cluster happen through two different mechanisms? What role does the extensive non-coding transcription play in modulation of epigenetic changes? What role do the epigenetic changes play in regulating non-coding transcription? Are there regulatory functions for many non-coding transcripts emanating from Hox clusters? If so are non-coding transcripts acting in *cis*-to modulate Hox expression or in *trans*? We identified non-coding RNAs responding to an RA signal. Are there specific non-coding RNAs generated in response to other signaling signals such as Wnts and Fgfs? I predict this will become a major area of study in the future as more and more roles for RNAs are uncovered in other contexts.

The characterization of the ES patterns of differentiation gave me an interesting insight into dynamics of Hox gene expression. I obtained a great deal of information about quantitative changes in Hox gene expression and also the expression of known co-factors such as PBX and MEIS during RA induced differentiation. This information was valuable in shaping the design for experiments to understand binding properties of HOXA1, HOXB1, PBX, MEIS and other Hox proteins. Based on knowledge gained from the dynamics of the ES cell differentiation time course, I selected three time points (6 hrs, 12hrs and 24hrs) of RA treatment to study Hoxb1 binding properties and 24 hour RA treatment for HOXA1, PBX and MEIS.

I intended to ask the following basic questions:

1. What defines a HOXB1 and HOXA1 binding site?
2. Are there different classes of binding sites for these two proteins?
3. Do the HOXA1 and HOXB1 proteins have distinct and/or overlapping target genes?

4. Are there different mechanisms for Hox binding?
5. What role do co-factors play in selection of targets?
6. Are there novel co-factors and interactors for HOXB1 and HOXA1 beyond PBX and MEIS?

Based on similar functional roles *in vivo* from genetic studies I expected that Hoxa1 and Hoxb1 would have very similar targets and depend upon PBX and MEIS as the major cofactors. I felt there were likely to be unique or distinct targets for these genes but in the main they would overlap. The results turned out very different and were very surprising. The majority of Hoxb1 binding sites are occupied as early as 6 hours of RA treatment. These sites show stronger occupancy of proteins as hours of RA treatment increases. A significant number of Hoxb1 genome-wide binding sites have overlapping PBX binding sites (only 10%) and there is evidence of previously for binding on previously defined HOX-PBX-MEIS interaction targets can be found in genome-wide analysis. However, the largest proportion of Hoxb1 sites do not display co-occupancy for PBX or MEIS suggesting that other types of sites and co-factors are involved in its interactions with DNA. Trying to understand this finding became a major aim of my analysis.

Computationally looking for over-represented sequences I uncovered new classes of binding sites which I termed Hox Binding Elements (HBEs *eg.* HRE2). HRE2 seems to be a Hoxb1 response element which is capable of exerting a repressive input from Hoxb1 in transgenic reporter assays. Since there is not a classic Hox binding site in the HRE2 motif I performed template affinity assays in combination with mass spec to sequence proteins capable of binding to it. The HRE2 is capable of recruiting of HOX, PBX and MEIS proteins in such assays suggesting that a HOX/PBX/MEIS complex might be recruited to these sites by other factors interacting with the motif. In support of this *in vitro* data a large number of HRE2 containing HOXB1 peaks are co-occupied with PBX based on PBX ChIP-seq data. Investigating this site and characterizing it in more detail is going to be a major focus of my research once I have completed my thesis. I plan to generate mutation in the sequence and evaluate binding requirements, interacting proteins and functional roles of these sites on *in vivo* target locations. This will serve as a good example in which to explore novel motifs or sites of interactions of Hox proteins and specificity between different Hox inputs. I plan to ask the following questions:

1. What recruits Hoxb1, PBX and MEIS to HRE2
2. Can Hoxb1 bind to HRE2 without PBX or MEIS?
3. If so what serves as a potential co-factor for HOXB1 in this case?

4. Does PBX binding generate the marks of enhancer activation with these elements seen with H3K27ac?

5. What is mechanism of enhancer activation upon PBX binding?

The largest numbers of Hoxb1 binding peaks are enriched with REST binding motifs. This is interesting because REST plays such a key role in coordinating the progression of neurogenesis and phenotypes in Hoxb1 mutants resemble in part those seen in REST zebra fish mutants. REST is a critical site to coordinate repression of neural genes and it appears that Hox proteins may function to oppose such repression to coordinate activation of selected genes in neurogenesis. I plan to continue studying this in much greater detail.

Not only are Hoxb1 peaks associated with over-represented REST binding motifs, ChIP-seq with REST in ES cells indicated that REST also occupies these sites in undifferentiated and differentiated ES cells. Surprisingly, REST occupancy remains unchanged even after Hoxb1 binding. This may suggest that complete removal of REST is not required for binding of activators and there may be a balance between activation and repression which potentiates progression through neurogenesis. However, there is also a possibility that the apparent REST and HOXB1 co-occupancy is due to experimental limitations. Heterogeneity in the cells may mean that with a subset of the cells has REST while another subset binds Hoxb1. To explore or eliminate this ambiguity, I am planning to do a sequential ChIP with REST and HOXB1.

The take home message from the *Hoxb1* analysis is that multiple motifs besides the simple ATTA or bipartite Hox/PBX sequences correlate with Hox binding. We see evidence for a variety of novel over-represented sequences or different classes of HOXB1 binding sites in genomic regions. In the initial studies aided by the Conaway lab it appears that other proteins may bind to these motifs and recruit HOXB1 or HOXB1/PBX/MEIS complex. I ran pilot MudPit experiments to examine proteins and going forward I plan to make a concerted effort to examine the sequences required for binding and the mechanistic basis for this recruitment. Is HOXB1 using new partners to bind directly to DNA or is it interacting with proteins already bound? The apparent wide variety in the sites bound by Hoxb1 and its mechanisms for interacting with DNA are important to understand if we are to be able to interrogate genomic sequences and make predictions on *in vivo* relevant sites of transcription factor interactions.

The unexpected nature of the data for genome-wide analysis of HOXB1 binding stimulated me to examine Hoxa1 binding to determine whether it was the same or

different. Loss of function phenotypes indicate that HOXA1 and HOXB1 may act on similar targets although their phenotypes are very different. This was previously thought to be related to differences in their expression but it could also be due to variations in target genes. To understand this better I analyzed genome-wide Hoxa1 occupancy after 24 hours of RA treatment in differentiated ES cells. The data clearly reveal that in general Hoxa1 and hoxb1 have a very different genome-wide occupancy on target peaks. Although they share a few binding sites, the majority of binding peaks are distinct for each class of protein. Unlike HOXB1, the majority of HOXA1 bound regions are enriched with HOX-PBX bipartite and classical PBX sites. My ChIP-seq data reveal that these sites are physically occupied by PBX and MEIS. There is minimal over-representation of REST sites associated with HOXA1 although there are some. REST shows co-occupancy with HOXA1 mostly on target where both HOXA1 and HOXB1 bind. Similarly, targets occupied with REST were not occupied with PBX and MEIS. This mutually exclusive set of targets probably implies some form of sub-functionalization in HOXB1 and HOXA1 binding. This could be achieved through differential interaction with two separate classes of TALE proteins or REST itself. For example, the HOXB1 responsive r4 enhancer (ARE, Auto-Regulatory Element) shows occupancy of HOXA1, PBX and MEIS. Surprisingly, HOXB1 does not occupy this site as much as HOXA1 at 24 hours of RA induction, raising the possibility that there may be temporal differences in the occupancy of different Hox proteins on HOX-PBX sites. It is worth noting that in developing embryos expression of *Hoxa1* is down-regulated before *Hoxb1* expression appears in r4. In future it would be interesting to study Hoxb1 binding properties in Hoxa1 loss of function mutant cell lines or in the presence of both proteins to see if there is a preference of one over another. I will undertake this study in future; since it will provide insight about mutual exclusiveness of these two paralogous group 1 genes. In addition I plan to make ES cells with tagged version of *Drosophila labial*. It is possible that labial might bind to both classes of sites and the two mouse proteins have segregated these binding properties. As an alternative approach I plan to generate chimeric versions of HOXA1 and HOXB1 swapping their N terminal domains to see if this alters the targets they bind on a genome-wide basis. These types of experiments have the potential to aid in understanding binding characteristics and specificities of proteins in a paralogous group.

A fundamental observation to emerge from this thesis study is the diversity in apparent Hox binding sites and the implication that cofactors or interacting proteins are intimately involved in potentiating their specificity on DNA and chromatin. As it is evident from results and discussions in Chapters 4 and 5, HOXA1 and HOXB1 genome-wide

binding data alone has limited value in terms of addressing the larger question of binding specificity. But by comparing and contrasting these patterns with epigenetic marks and binding of REST, PBX and MEIS interesting features and properties begin to emerge. It illustrates the potential of this system to investigate common and unique binding properties on a systematic basis for many of the Hox family of proteins. I want to understand the unique aspects each paralogous groups in terms of recruitment, binding kinetics, cofactors, interactors and co-activators or co-repressors. I have planned to expand the template binding assay coupled with MudPit to investigate proteins and complexes recruited on to ChIP-seq Hox sites.

With this experience in hand, I have already begun to compare this HOXA1 and HOXB1 data to the binding properties of other anterior Hox B genes, namely HOXB2, HOXB3 and HOXB4. I have cell lines carrying tagged versions of these proteins and I have performed ChIP-seq experiments. Preliminary analysis reveals that each Hox protein has a set of distinct targets and common target sites (Fig.6.1).

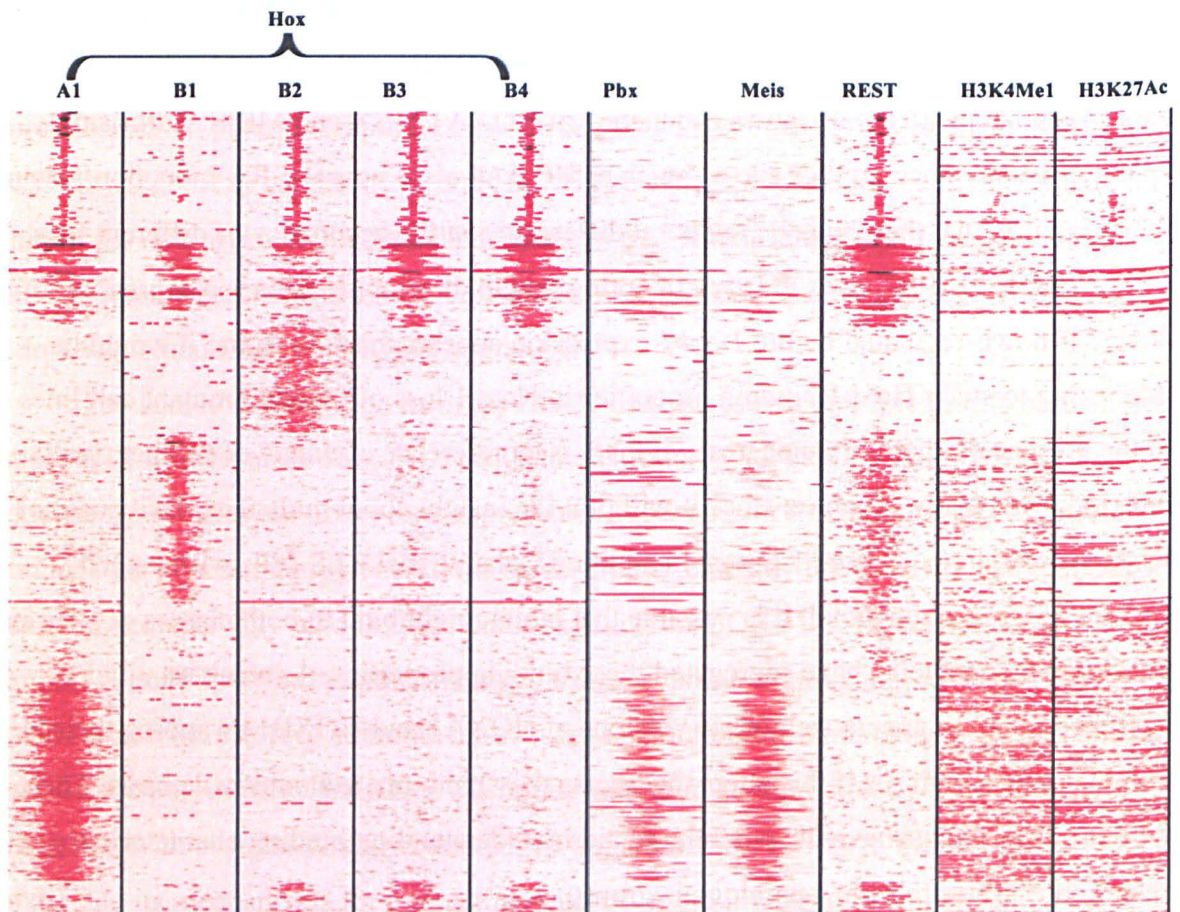


Figure 6-1 Combinatorial binding of various anterior Hox genes, Cofactors, REST and associated epigenetic modification.

All HOXA1,, B2, B3 and B4 bound genomic region after 24hrs of RA Induction and all HOXB1 occupied regions from 6hrs, 12hrs and 24hrs of RA induced ES cells are

shown on Y-axis. Every row is a genomic coordinate with at least occupancy by one or more Hox genes. Each column represents onetime point antibody combination. Heatmap is hierarchically clustered. Occupancy of Pbx and Meis are also shown in these genomic region. REST occupancy in in uninduced ES cells and 6 and 24hrs RA Induced ES Cells are also shown. Each genomic coordinate is centered on midpoint of HOXA1 or HOXB1 peaks and region including 500bp either side of midpoint is shown in heatmap. Large numbers of HOXA1 bound regions are co-occupied with PBX and MEIS. Large numbers of HOXB1 bound regions are co-occupied with REST. PBX and REST binding are mutually exclusive. Enhancer specific histone modification namely H3K27Ac and H3K4Me1 in uninduced ES cells and after 6 and 24hours of RA treatment in these regions are also shown.

I am optimistic this will serve to reveal new rules and properties which will aid in the approach to identify downstream target genes.

In future I am interested in knowing,

1. What is the difference between common Hox targets and distinct Hox targets?
2. Can all Hox proteins co-occupy a common site at same time or are their occupancy mutually exclusive, shedding light on posterior prevalence?
3. What determines the occupancy of a Hox protein?
4. Is it presence or absence of other Hox proteins, cofactors or target sites?
5. Their relative binding affinity, on versus off rates.
6. What kind of combinatorial binding code exists between Hox genes?
7. How this combinatorial code is influenced by Co-factors and interactors?
8. Nature of cofactors?
9. Their strength of interaction with common cofactors?

In summary, I established an effective system based on RA induced ES cell differentiation. This system not only allows the systematic study of changes in binding properties of Hox proteins and cofactors, but it also permits a means for biochemical dissection of regulatory mechanisms of action. My, current body of work characterized the genome-wide binding property of HOXA1 and HOXB1 proteins in differentiating ES cells and identified REST as a potential novel associated site to integrate interactions of HOXB1 and select other Hox proteins. This data has laid groundwork to investigate a role for HOXB1 in activation of REST occupied genomic targets during neurogenesis. Furthermore, while transcription factors can act as activators and repressors how this is achieved for Hox proteins is unclear. The identification of the HBE2 motif which appears to integrate a repressive action of Hoxb1 provides an opportunity to investigate this issue in more detail. It is interesting to note that Hoxb1 interacts with the HBE2 along with PBX and MEIS.

An implication of my data is that the two paralogous group 1 gene, HOXA1 and HOXB1, appear to have very limited overlapping targets and may regulate quite distinct gene sets. The ES differentiation based system is an efficient approach to dissect binding specificity of Hox genes and can be applied to other transcription factors. In future, integration of many more Hox proteins and other factors (*Sox*, *Krox*, *Prep*, *TgfII* and *Pax*) has the potential to build a better understanding on how these transcription factors exert their specificity and hopefully reveal the interesting properties that explain the combinatorial binding features that generate a functional "Hox Code".

Chapter 7References

Aboobaker, A., and Blaxter, M. (2003). Hox gene evolution in nematodes: novelty conserved. *Curr Opin Genet Dev* 13, 593-598.

Abu-Abed, S., Dolle, P., Metzger, D., Beckett, B., Chambon, P., and Petkovich, M. (2001). The retinoic acid-metabolizing enzyme, CYP26A1, is essential for normal hindbrain patterning, vertebral identity, and development of posterior structures. *Genes Dev* 15, 226-240.

Akam, M. (1989). *Hox* and *HOM*: homologous gene clusters in insects and vertebrates. *Cell* 57, 347-349.

Alexander, T., Nolte, C., and Krumlauf, R. (2009). Hox genes and segmentation of the hindbrain and axial skeleton. *Annu Rev Cell Dev Biol* 25, 431-456.

Alexandre, D., Clarke, J., Oxtoby, E., Yan, Y.-L., Jowett, T., and Holder, N. (1996). Ectopic expression of *Hoxa-1* in the zebrafish alters the fate of the mandibular arch neural crest and phenocopies a retinoic acid -induced phenotype. *Development* 122, 735-746.

Amemiya, C.T., Prohaska, S.J., Hill-Force, A., Cook, A., Wasserscheid, J., Ferrier, D.E., Pascual-Anaya, J., Garcia-Fernandez, J., Dewar, K., and Stadler, P.F. (2008). The amphioxus Hox cluster: characterization, comparative genomics, and evolution. *J Exp Zool B Mol Dev Evol* 310, 465-477.

Amores, A., Force, A., Yan, Y.-L., Joly, L., Amemiya, C., Fritz, A., Ho, R., Langeland, J., Prince, V., Wang, Y.-L., *et al.* (1998). Zebrafish *hox* clusters and vertebrate genome evolution. *Science* 282, 1711-1714.

Amores, A., Suzuki, T., Yan, Y.L., Pomeroy, J., Singer, A., Amemiya, C., and Postlethwait, J.H. (2004). Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res* 14, 1-10.

Andreeva, T.F., Kuk, C., Korchagina, N.M., C'Ike'm, M., and Dondya, A.K. (2001). [Cloning and analysis of structural organization of Hox genes in the Polychaete *Nereis virens*]. *Ontogenez* 32, 225-233.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 340, 1234167.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., *et al.* (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297, 1301-1310.

Appel, B., and Sakonju, S. (1993). Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in *Drosophila* embryos. *EMBO J* 12, 1099-1109.

Arenas-Mena, C., Cameron, A.R., and Davidson, E.H. (2000). Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. *Development* 127, 4631-4643.

Arenkiel, B.R., Gaufo, G.O., and Capecchi, M.R. (2003). Hoxb1 neural crest preferentially form glia of the PNS. *Dev Dyn* 227, 379-386.

Arenkiel, B.R., Tvrdik, P., Gaufo, G.O., and Capecchi, M.R. (2004). Hoxb1 functions in both motoneurons and in tissues of the periphery to establish and maintain the proper neuronal circuitry. *Genes Dev* 18, 1539-1552.

Austin, R.J., and Biggin, M.D. (1995). A domain of the even-skipped protein represses transcription by preventing TFIID binding to a promoter: repression by cooperative blocking. *Mol Cell Biol* 15, 4683-4693.

Baader, S.L., Sanlioglu, S., Berrebi, A.S., Parker-Thornburg, J., and Oberdick, J. (1998). Ectopic overexpression of engrailed-2 in cerebellar Purkinje cells causes restricted cell loss and retarded external germinal layer development at lobule junctions. *J Neurosci* 18, 1763-1773.

Ballas, N., Grunseich, C., Lu, D.D., Speh, J.C., and Mandel, G. (2005). REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* 121, 645-657.

Ballas, N., and Mandel, G. (2005). The many faces of REST oversee epigenetic programming of neuronal genes. *Curr Opin Neurobiol* 15, 500-506.

Bami, M., Episkopou, V., Gavalas, A., and Gouti, M. (2011). Directed neural differentiation of mouse embryonic stem cells is a sensitive system for the identification of novel Hox gene effectors. *PLoS One* 6, e20197.

Barrow, J., and Capecchi, M. (1996). Targeted disruption of the *Hoxb2* locus in mice interferes with expression of *Hoxb1* and *Hoxb4*. *Development* 122, 3817-3828.

Beachy, P., Helfand, S., and Hogness, D. (1985). Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature* 313, 545-551.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* 44, 23-28.

Begemann, G., Schilling, T.F., Rauch, G.J., Geisler, R., and Ingham, P.W. (2001). The zebrafish neckless mutation reveals a requirement for *raldh2* in mesodermal signals that pattern the hindbrain. *Development* 128, 3081-3094.

Bel-Vialar, S., Itasaki, N., and Krumlauf, R. (2002). Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the HoxB genes in two distinct groups. *Development* 129, 5103-5115.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., *et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 133, 1266-1276.

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., and Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. *Genes Dev* 13, 946-953.

Berthelsen, J., Zappavigna, V., Ferretti, E., Mavilio, F., and Blasi, F. (1998). The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity. *EMBO J* 17, 1434-1445.

Biggin, M.D., and McGinnis, W. (1997). Regulation of segmentation and segmental identity by *Drosophila* homeoproteins: the role of DNA binding in functional activity and specificity. *Development* 124, 4425-4433.

Biggin, M.D., and Tjian, R. (1989). A purified *Drosophila* homeodomain protein represses transcription *in vitro*. *Cell* 58, 433-440.

Billeter, M., Guntert, P., Luginbuhl, P., and Wuthrich, K. (1996). Hydration and DNA recognition by homeodomains. *Cell* 85, 1057-1065.

Boncinelli, E., Somma, R., Acampora, D., Pannese, M., D'Esposito, M., Faiella, A., and Simeone, A. (1988). Organization of human homeobox genes. *Hum Reprod* 3, 880-886.

Boulet, A.M., and Capecchi, M.R. (1996). Targeted disruption of *Hoxc4* causes esophageal defects and vertebral transformations. *Dev Biol* 177, 232-249.

Boylan, J.F., and Gudas, L. (1991). Overexpression of the cellular retinoic acid binding protein I (*CRABP-I*) results in a reduction in differentiation-specific gene expression in F9 teratocarcinoma cells. *J Biol Chem* 112, 965-979.

Brendolan, A., Ferretti, E., Salsi, V., Moses, K., Quaggin, S., Blasi, F., Cleary, M.L., and Selleri, L. (2005). A Pbx1-dependent genetic and transcriptional network regulates spleen ontogeny. *Development* 132, 3113-3126.

Burke, A.C., Nelson, C.E., Morgan, B.A., and Tabin, C. (1995). *Hox* genes and the evolution of vertebrate axial morphology. *Development* 121, 333-346.

Burke, A.C., and Nowicki, J.L. (2003). A new view of patterning domains in the vertebrate mesoderm. *Dev Cell* 4, 159-165.

Busturia, A., Casanova, J., Sanchez-Herrero, E., and Morata, G. (1989). Structure and function of the *bithorax* complex genes of *Drosophila*. In CIBA foundation symposium 144 (Berlin: Springer-Verlag), pp. 227-242.

Cai, A.Q., Radtke, K., Linville, A., Lander, A.D., Nie, Q., and Schilling, T.F. (2012). Cellular retinoic acid-binding proteins are essential for hindbrain patterning and signal robustness in zebrafish. *Development* 139, 2150-2155.

Carapuco, M., Novoa, A., Bobola, N., and Mallo, M. (2005). Hox genes specify vertebral types in the presomitic mesoderm. *Genes Dev* 19, 2116-2121.

Carrasco, A.E., McGinnis, W., Gehring, W.J., and De Robertis, E.M. (1984). Cloning of an *X. laevis* gene expressed during early embryogenesis coding for a peptide region homologous to *Drosophila* homeotic genes. *Cell* 37, 409-414.

Carroll, S.B. (1995). Homeotic genes and the evolution of arthropods and chordates. *Nature* 376, 479-485.

Carroll, S.B. (2005). Evolution at two levels: on genes and form. *PLoS Biol* 3, e245.

Casares, F., and Mann, R.S. (2000). A dual role for homothorax in inhibiting wing blade development and specifying proximal wing identities in *Drosophila*. *Development* 127, 1499-1508.

Cavalli, G., and Paro, R. (1998). The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* 93, 505-518.

Chambon, P. (1994). The retinoid signaling pathway; molecular and genetic analysis. *Semin Cell Biol* 5, 115-125.

Chan, S.-K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R.S. (1994a). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* 78, 603-615.

Chan, S.K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R.S. (1994b). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* 78, 603-615.

Chan, S.K., Popperl, H., Krumlauf, R., and Mann, R.S. (1996). An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *EMBO J* 15, 2476-2487.

Chan, S.K., Ryoo, H.D., Gould, A., Krumlauf, R., and Mann, R.S. (1997). Switching the *in vivo* specificity of a minimal *HOX*-responsive element. *Development* 124, 2007-2014.

Chang, C.-P., Shen, W.-F., Rozenfeld, S., Lawrence, H.J., Largman, C., and Cleary, M.L. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev* 9, 663-674.

Chang, C.P., Jacobs, Y., Nakamura, T., Jenkins, N.A., Copeland, N.G., and Cleary, M.L. (1997). Meis proteins are major *in vivo* DNA binding partners for wild-type but not chimeric Pbx proteins. *Mol Cell Biol* 17, 5679-5687.

Chen, F., and Capecchi, M.R. (1997). Targeted mutations in *hoxa-9* and *hoxb-9* reveal synergistic interactions. *Dev Biol* 181, 186-196.

Chen, J., and Ruley, H.E. (1998). An enhancer element in the *EphA2* (Eck) gene sufficient for rhombomere-specific expression is activated by *HOXA1* and *HOXB1* homeobox proteins. *J Biol Chem* 273, 24670-24675.

Chiori, R., Jager, M., Denker, E., Wincker, P., Da Silva, C., Le Guyader, H., Manuel, M., and Queinsec, E. (2009). Are Hox genes ancestrally involved in axial patterning? Evidence from the hydrozoan *Clytia hemisphaerica* (Cnidaria). *PLoS One* 4, e4231.

Chisaka, O., and Capecchi, M.R. (1991). Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox1.5*. *Nature* 350, 473-479.

Chisaka, O., Musci, T.S., and Capecchi, M.R. (1992). Developmental defects of the ear, cranial nerves and hindbrain resulting from targeted disruption of the mouse homeobox gene *Hox-1.6*. *Nature* 355, 516-520.

Choe, S.K., Lu, P., Nakamura, M., Lee, J., and Sagerstrom, C.G. (2009). Meis cofactors control HDAC and CBP accessibility at Hox-regulated promoters during zebrafish embryogenesis. *Dev Cell* 17, 561-567.

Choe, S.K., and Sagerstrom, C.G. (2004). Paralog group 1 hox genes regulate rhombomere 5/6 expression of *vhnf1*, a repressor of rostral hindbrain fates, in a meis-dependent manner. *Dev Biol* 271, 350-361.

- Choe, S.K., Zhang, X., Hirsch, N., Straubhaar, J., and Sagerstrom, C.G. (2011). A screen for *hoxb1*-regulated genes identifies *ppp1r14a* as a regulator of the rhombomere 4 Fgf-signaling center. *Dev Biol* 358, 356-367.
- Choo, S.W., White, R., and Russell, S. (2011). Genome-wide analysis of the binding of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in *Drosophila*. *PLoS One* 6, e14778.
- Chopra, V.S. (2011). Chromosomal organization at the level of gene complexes. *Cell Mol Life Sci* 68, 977-990.
- Chourrout, D., Delsuc, F., Chourrout, P., Edvardsen, R.B., Rentzsch, F., Renfer, E., Jensen, M.F., Zhu, B., de Jong, P., Steele, R.E., *et al.* (2006). Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* 442, 684-687.
- Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., *et al.* (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203-218.
- Condie, B.G., and Capecchi, M.R. (1994). Mice with targeted disruptions in the paralogous genes *Hoxa-3* and *Hoxd-3* reveal synergistic interactions. *Nature* 370, 304-307.
- Conlon, F.L., Fairclough, L., Price, B.M., Casey, E.S., and Smith, J.C. (2001). Determinants of T box protein specificity. *Development* 128, 3749-3758.
- Conlon, R.A., and Rossant, J. (1992). Exogenous retinoic acid rapidly induces anterior ectopic expression of murine *Hox-2* genes *in vivo*. *Development* 116, 357-368.
- Cooper, K.L., Leisenring, W.M., and Moens, C.B. (2003). Autonomous and nonautonomous functions for Hox/Pbx in branchiomotor neuron development. *Dev Biol* 253, 200-213.
- Dani, C., Smith, A.G., Dessolin, S., Leroy, P., Staccini, L., Villageois, P., Darimont, C., and Ailhaud, G. (1997). Differentiation of embryonic stem cells into adipocytes *in vitro*. *J Cell Sci* 110 (Pt 11), 1279-1285.
- Davenne, M., Maconochie, M.K., Neun, R., Pattyn, A., Chambon, P., Krumlauf, R., and Rijli, F.M. (1999). *Hoxa2* and *Hoxb2* control dorsoventral patterns of neuronal development in the rostral hindbrain. *Neuron* 22, 677-691.
- Davis, A.P., and Capecchi, M.R. (1994). Axial homeosis and appendicular skeleton defects in mice with a targeted disruption of *hoxd-11*. *Development* 120, 2187-2198.

Davis, A.P., and Capecchi, M.R. (1996). A mutational analysis of the 5' *HoxD* genes: dissection of genetic interactions during limb development in the mouse. *Development* 122, 1175-1185.

de Navas, L.F., Reed, H., Akam, M., Barrio, R., Alonso, C.R., and Sanchez-Herrero, E. (2011). Integration of RNA processing and expression level control modulates the function of the *Drosophila* Hox gene *Ultrabithorax* during adult development. *Development* 138, 107-116.

de Rosa, R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte, A., Akam, M., Carroll, S.B., and Balavoine, G. (1999). Hox genes in brachiopods and priapulids and protostome evolution. *Nature* 399, 772-776.

Dear, T.N., Colledge, W.H., Carlton, M.B., Lavenir, I., Larson, T., Smith, A.J., Warren, A.J., Evans, M.J., Sofroniew, M.V., and Rabbitts, T.H. (1995). The Hox11 gene is essential for cell survival during spleen development. *Development* 121, 2909-2915.

Dear, T.N., Sanchez-Garcia, I., and Rabbitts, T.H. (1993). The HOX11 gene encodes a DNA-binding nuclear transcription factor belonging to a distinct family of homeobox genes. *Proc Natl Acad Sci U S A* 90, 4431-4435.

Deschamps, J., van den Akker, E., Forlani, S., De Graaff, W., Oosterveen, T., Roelen, B., and Roelfsema, J. (1999). Initiation, establishment and maintenance of Hox gene expression patterns in the mouse. *Int J Dev Biol* 43, 635-650.

Deschamps, J., and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* 132, 2931-2942.

Devenport, M.P., Blass, C., and Eggleston, P. (2000). Characterization of the Hox gene cluster in the malaria vector mosquito, *Anopheles gambiae*. *Evol Dev* 2, 326-339.

Diez del Corral, R., Breitkreuz, D.N., and Storey, K.G. (2002). Onset of neuronal differentiation is regulated by paraxial mesoderm and requires attenuation of FGF signalling. *Development* 129, 1681-1691.

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., and Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* 40, 65-79.

Diez del Corral, R., and Storey, K.G. (2004). Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. *BioEssays* 26, 857-869.

- DiMartino, J.F., Selleri, L., Traver, D., Firpo, M.T., Rhee, J., Warnke, R., O'Gorman, S., Weissman, I.L., and Cleary, M.L. (2001). The Hox cofactor and proto-oncogene Pbx1 is required for maintenance of definitive hematopoiesis in the fetal liver. *Blood* 98, 618-626.
- Dollé, P., Dierich, A., LeMeur, M., Schimmang, T., Schuhbaur, B., Chambon, P., and Duboule, D. (1993). Disruption of the *Hoxd-13* gene induces localized heterochrony leading to mice with neotenic limbs. *Cell* 75, 431-441.
- Dollé, P., Izpisua-Belmonte, J.C., Brown, J.M., Tickle, C., and Duboule, D. (1991). *Hox-4* genes and the morphogenesis of mammalian genitalia. *Genes Dev* 5, 1767-1776.
- Donaldson, I.J., Amin, S., Hensman, J.J., Kutejova, E., Rattray, M., Lawrence, N., Hayes, A., Ward, C.M., and Bobola, N. (2012). Genome-wide occupancy links Hoxa2 to Wnt- β -catenin signaling in mouse embryonic development. *Nucleic Acids Res* 40, 3990-4001.
- Drab, M., Haller, H., Bychkov, R., Erdmann, B., Lindschau, C., Haase, H., Morano, I., Luft, F.C., and Wobus, A.M. (1997). From totipotent embryonic stem cells to spontaneously contracting smooth muscle cells: a retinoic acid and db-cAMP in vitro differentiation model. *FASEB J* 11, 905-915.
- Duboule, D. (1994). Temporal colinearity and phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development Supplement*, 135-142.
- Duboule, D. (1998). Vertebrate hox gene regulation: clustering and/or colinearity? *Curr Opin Genet Dev* 8, 514-518.
- Duboule, D. (2000). Developmental genetics. A Hox by any other name. *Nature* 403, 607, 609-610.
- Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development* 134, 2549-2560.
- Duboule, D., and Dollé, P. (1989). The structural and functional organization of the murine *HOX* gene family resembles that of *Drosophila* homeotic genes. *EMBO J* 8, 1497-1505.
- Duboule, D., and Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet* 10, 358-364.
- Dupé, V., Davenne, M., Brocard, J., Dollé, P., Mark, M., Dierich, A., Chambon, P., and Rijli, F. (1997). *In vivo* functional analysis of the *Hoxa1* 3' retinoid response element (3' RARE). *Development* 124, 399-410.

Ekker, S.C., Jackson, D.G., von Kessler, D.P., Sun, B.I., Young, K.E., and Beachy, P.A. (1994). The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *EMBO J* 13, 3551-3560.

Favier, B., Le Meur, M., Chambon, P., and Dollé, P. (1995). Axial skeleton homeosis and forelimb malformations in *Hoxd-11* mutant mice. *Proc Natl Acad Sci U S A* 92, 310-314.

Favier, B., Rijli, F.M., Fromental-Ramain, C., Fraulob, V., Chambon, P., and Dollé, P. (1996). Functional cooperation between the non-paralogous genes *Hoxa-10* and *Hoxd-11* in the developing forelimb and axial skeleton. *Development* 122, 449-460.

Ferretti, E., Cambroner, F., Tümpel, S., Longobardi, E., Wiedemann, L.M., Blasi, F., and Krumlauf, R. (2005). *Hoxb1* enhancer and control of rhombomere 4 expression: Complex interplay between PREP1-PBX1-HOXB1 binding sites. *Mol Cell Biol* 25, 8541-8552.

Ferretti, E., Marshall, H., Pöpperl, H., Maconochie, M., Krumlauf, R., and Blasi, F. (2000). Segmental expression of *Hoxb2* in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins. *Development* 127, 155-166.

Ferretti, E., Schulz, H., Talarico, D., Blasi, F., and Berthelsen, J. (1999). The PBX-regulating protein PREP1 is present in different PBX-complexed forms in mouse. *Mech Dev* 83, 53-64.

Ferrier, D.E., and Akam, M. (1996). Organization of the Hox gene cluster in the grasshopper, *Schistocerca gregaria*. *Proc Natl Acad Sci U S A* 93, 13024-13029.

Finnerty, J.R., Pang, K., Burton, P., Paulson, D., and Martindale, M.Q. (2004). Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. *Science* 304, 1335-1337.

Folberg, A., Kovacs, E.N., and Featherstone, M.S. (1997). Characterization and retinoic acid responsiveness of the murine *Hoxd4* transcription unit. *J Biol Chem* 272, 29151-29157.

Forlani, S., Lawson, K.A., and Deschamps, J. (2003). Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. *Development* 130, 3807-3819.

Fraichard, A., Chassande, O., Bilbaut, G., Dehay, C., Savatier, P., and Samarut, J. (1995). In vitro differentiation of embryonic stem cells into glial cells and functional neurons. *J Cell Sci* 108 (Pt 10), 3181-3188.

Freeman, R., Ikuta, T., Wu, M., Koyanagi, R., Kawashima, T., Tagawa, K., Humphreys, T., Fang, G.C., Fujiyama, A., Saiga, H., *et al.* (2012). Identical genomic organization of two hemichordate hox clusters. *Curr Biol* 22, 2053-2058.

Fromental-Ramain, C., Warot, X., Lakkaraju, S., Favier, B., Haack, H., Birling, C., Dierich, A., Dollé, P., and Chambon, P. (1996). Specific and redundant functions of the paralogous *Hoxa-9* and *Hoxd-9* genes in forelimb and axial skeleton patterning. *Development* 122, 461-472.

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A.S., Gaudenz, K., De Kumar, B., Krumlauf, R., and Zeitlinger, J. (2012). Poised RNA Polymerase II Changes over Developmental Time and Prepares Genes for Future Expression. *Cell Rep* 2, 1670-1683.

Galant, R., and Carroll, S.B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 415, 910-913.

Galant, R., Walsh, C.M., and Carroll, S.B. (2002). Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* 129, 3115-3126.

Gale, E., Zile, M., and Maden, M. (1999). Hindbrain respecification in the retinoid-deficient quail. *Mech Dev* 89, 43-54.

Garber, R.L., Kuroiwa, A., and Gehring, W.J. (1983). Genomic and cDNA clones of the homeotic locus *Antennapedia* in *Drosophila*. *EMBO J* 2, 2027-2036.

Garcia-Fernandez, J., and Holland, P.W.H. (1994). Archetypal organisation of the *amphioxus Hox* gene cluster. *Nature* 370, 563-566.

Gaufo, G.O., Thomas, K.R., and Capecchi, M.R. (2003). Hox3 genes coordinate mechanisms of genetic suppression and activation in the generation of branchial and somatic motoneurons. *Development* 130, 5191-5201.

Gaufo, G.O., Wu, S., and Capecchi, M.R. (2004). Contribution of Hox genes to the diversity of the hindbrain sensory system. *Development* 131, 1259-1266.

Gavalas, A. (2002). ArRAnging the hindbrain. *Trends Neurosci* 25, 61-64.

Gavalas, A., Davenne, M., Lumsden, A., Chambon, P., and Rijli, F.M. (1997). Role of *Hoxa-2* in axon pathfinding and rostral hindbrain patterning. *Development* 124, 3693-3702.

Gavalas, A., and Krumlauf, R. (2000). Retinoid signalling and hindbrain patterning. *Curr Opin Genet Dev* 10, 380-386.

Gavalas, A., Ruhrberg, C., Livet, J., Henderson, C.E., and Krumlauf, R. (2003). Neuronal defects in the hindbrain of *Hoxa1*, *Hoxb1* and *Hoxb2* mutants reflect regulatory interactions among these Hox genes. *Development* 130, 5663-5679.

Gavalas, A., Studer, M., Lumsden, A., Rijli, F.M., Krumlauf, R., and Chambon, P. (1998). *Hoxa1* and *Hoxb1* synergize in patterning the hindbrain, cranial nerves and second pharyngeal arch. *Development* 125, 1123-1136.

Gavalas, A., Trainor, P., Ariza-McNaughton, L., and Krumlauf, R. (2001). Synergy between *Hoxa1* and *Hoxb1*: the relationship between arch patterning and the generation of cranial neural crest. *Development* 128, 3017-3027.

Gehring, W. (1966). [Cell heredity and changes of determination in cultures of imaginal discs in *Drosophila melanogaster*]. *J Embryol Exp Morphol* 15, 77-111.

Gehring, W.J., Affolter, M., and Burglin, T. (1994a). Homeodomain proteins. *Annu Rev Biochem* 63, 487-526.

Gehring, W.J., Qian, Y.-Q., Billeter, M., Furukubo-Tokunaga, K., Shier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., and Wurthrich, K. (1994b). Homeodomain-DNA recognition. *Cell* 78, 211-223.

Gendron-Maguire, M., Mallo, M., Zhang, M., and Gridley, T. (1993). *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. *Cell* 75, 1317-1331.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., *et al.* (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428, 493-521.

Gillespie, R.F., and Gudas, L.J. (2007). Retinoic acid receptor isotype specificity in F9 teratocarcinoma stem cells results from the differential recruitment of coregulators to retinoic response elements. *J Biol Chem* 282, 33421-33434.

Glaser, T., and Brustle, O. (2005). Retinoic acid induction of ES-cell-derived neurons: the radial glia connection. *Trends Neurosci* 28, 397-400.

Glass, C.K., and Rosenfeld, M.G. (2000). The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev* 14, 121-141.

Goddard, J., Rossel, M., Manley, N., and Capecchi, M. (1996). Mice with targeted disruption of *Hoxb1* fail to form the motor nucleus of the VIIth nerve. *Development* 122, 3217-3228.

- Gonzalez-Crespo, S., Abu-Shaar, M., Torres, M., Martinez, A.C., Mann, R.S., and Morata, G. (1998). Antagonism between extradenticle function and Hedgehog signalling in the developing limb. *Nature* *394*, 196-200.
- Gonzalez-Crespo, S., and Morata, G. (1995). Control of *Drosophila* adult pattern by extradenticle. *Development* *121*, 2117-2125.
- Gonzalez-Crespo, S., and Morata, G. (1996). Genetic evidence for the subdivision of the arthropod limb into coxopodite and telopodite. *Development* *122*, 3921-3928.
- Gonzalez, F., Duboule, D., and Spitz, F. (2007). Transgenic analysis of *Hoxd* gene regulation during digit development. *Dev Biol* *306*, 847-859.
- Goodman, F.R., Bacchelli, C., Brady, A.F., Brueton, L.A., Fryns, J.P., Mortlock, D.P., Innis, J.W., Holmes, L.B., Donnemfeld, A.E., Feingold, M., *et al.* (2000). Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. *Am J Hum Genet* *67*, 197-202.
- Gottlieb, D.I., and Huettner, J.E. (1999). An in vitro pathway from embryonic stem cells to neurons and glia. *Cells Tissues Organs* *165*, 165-172.
- Gould, A., Itasaki, N., and Krumlauf, R. (1998). Initiation of rhombomeric *Hoxb4* expression requires induction by somites and a retinoid pathway. *Neuron* *21*, 39-51.
- Gould, A., Morrison, A., Sproat, G., White, R.A., and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping *Hox* expression patterns. *Genes Dev* *11*, 900-913.
- Gouti, M., and Gavalas, A. (2008). *Hoxb1* controls cell fate specification and proliferative capacity of neural stem and progenitor cells. *Stem Cells* *26*, 1985-1997.
- Graham, A., Papalopulu, N., and Krumlauf, R. (1989). The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* *57*, 367-378.
- Grammatopoulos, G.A., Bell, E., Toole, L., Lumsden, A., and Tucker, A.S. (2000). Homeotic transformation of branchial arch identity after *Hoxa2* overexpression. *Development* *127*, 5355-5365.
- Grandel, H., Lun, K., Rauch, G.J., Rhinn, M., Piotrowski, T., Houart, C., Sordino, P., Kuchler, A.M., Schulte-Merker, S., Geisler, R., *et al.* (2002). Retinoic acid signalling in the zebrafish embryo is necessary during pre-segmentation stages to pattern the anterior-posterior axis of the CNS and to induce a pectoral fin bud. *Development* *129*, 2851-2865.

Green, N.C., Rambaldi, I., Teakles, J., and Featherstone, M.S. (1998). A conserved C-terminal domain in PBX increases DNA binding by the PBX homeodomain and is not a primary site of contact for the YPWM motif of HOXA1. *J Biol Chem* 273, 13273-13279.

Greer, J.M., Puetz, J., Thomas, K.R., and Capecchi, M.R. (2000). Maintenance of functional equivalence during paralogous Hox gene evolution. *Nature* 403, 661-665.

Guazzi, S., Pintonello, M.L., Vigano, A., and Boncinelli, E. (1998). Regulatory interactions between human HOXB1, HOXB2 and HOXB3 proteins and the upstream sequence of the *Otx2* gene in embryonal carcinoma cells. *J Biol Chem* 273, 11092-11099.

Guerreiro, I., Nunes, A., Woltering, J.M., Casaca, A., Novoa, A., Vinagre, T., Hunter, M.E., Duboule, D., and Mallo, M. (2013). Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc Natl Acad Sci U S A*.

Guidato, S., Barrett, C., and Guthrie, S. (2003a). Patterning of motor neurons by retinoic acid in the chick embryo hindbrain in vitro. *Mol Cell Neurosci* 23, 81-95.

Guidato, S., Prin, F., and Guthrie, S. (2003b). Somatic motoneurone specification in the hindbrain: the influence of somite-derived signals, retinoic acid and Hoxa3. *Development* 130, 2981-2996.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., *et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071-1076.

Hafen, E., Levine, M., Garber, R.L., and Gehring, W.J. (1983). An improved in situ hybridization method for the detection of cellular RNAs in *Drosophila* tissue sections and its application for localizing transcripts of the homeotic Antennapedia gene complex. *EMBO J* 2, 617-623.

Hedlund, E., Karsten, S.L., Kudo, L., Geschwind, D.H., and Carpenter, E.M. (2004). Identification of a Hoxd10-regulated transcriptional network and combinatorial interactions with Hoxa10 during spinal cord development. *J Neurosci Res* 75, 307-319.

Hejnal, A., and Martindale, M.Q. (2009). Coordinated spatial and temporal expression of Hox genes during embryogenesis in the acoel *Convolutriloba longifissura*. *BMC Biol* 7, 65.

Hernandez, R.E., Putzke, A.P., Myers, J.P., Margaretha, L., and Moens, C.B. (2007). Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development* 134, 177-187.

Hoegg, S., Boore, J.L., Kuehl, J.V., and Meyer, A. (2007). Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 8, 317.

Hoegg, S., Brinkmann, H., Taylor, J.S., and Meyer, A. (2004). Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J Mol Evol* 59, 190-203.

Hoegg, S., and Meyer, A. (2005). Hox clusters as models for vertebrate genome evolution. *Trends Genet* 21, 421-424.

Holland, L.Z., and Holland, N.D. (1996). Expression of AmphiHox-1 and AmphiPax-1 in amphioxus embryos treated with retinoic acid: insights into evolution and patterning of the chordate nerve cord and pharynx. *Development* 122, 1829-1838.

Holland, P., Holland, L., Williams, N., and Holland, N. (1992). An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution. *Development* 116, 653-661.

Holland, P.W. (2013). Evolution of homeobox genes. *Wiley Interdiscip Rev Dev Biol* 2, 31-45.

Holst, B.D., Goomer, R.S., Wood, I.C., Edelman, G.M., and Jones, F.S. (1994). Binding and activation of the promoter for the neural cell adhesion molecule by Pax-8. *J Biol Chem* 269, 22245-22252.

Horan, G.S., Kovacs, E.N., Behringer, R.R., and Featherstone, M.S. (1995a). Mutations in paralogous *Hox* genes result in overlapping homeotic transformations of the axial skeleton: Evidence for unique and redundant function. *Dev Biol* 169, 359-372.

Horan, G.S., Ramirez-Solis, R., Featherstone, M.S., Wolgemuth, D.J., Bradley, A., and Behringer, R.R. (1995b). Compound mutants for the paralogous *Hoxa-4*, *Hoxb-4*, and *Hoxd-4* genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed. *Genes Dev* 9, 1667-1677.

Horan, G.S., Wu, K., Wolgemuth, D.J., and Behringer, R.R. (1994). Homeotic transformation of cervical vertebrae in *Hoxa-4* mutant mice. *Proc Natl Acad Sci U S A* 91, 12644-12648.

Houle, M., Prinos, P., Iulianella, A., Bouchard, N., and Lohnes, D. (2000). Retinoic acid regulation of *Cdx1*: an indirect mechanism for retinoids and vertebral specification. *Mol Cell Biol* 20, 6579-6586.

- Houle, M., Sylvestre, J.R., and Lohnes, D. (2003). Retinoic acid regulates a subset of Cdx1 function in vivo. *Development* 130, 6555-6567.
- Huang, Y., Sitwala, K., Bronstein, J., Sanders, D., Dandekar, M., Collins, C., Robertson, G., MacDonald, J., Cezard, T., Bilenky, M., *et al.* (2012). Identification and characterization of Hoxa9 binding sites in hematopoietic cells. *Blood* 119, 388-398.
- Hudry, B., Remacle, S., Delfini, M.C., Rezsöházy, R., Graba, Y., and Merabet, S. (2012). Hox proteins display a common and ancestral ability to diversify their interaction mode with the PBC class cofactors. *PLoS Biol* 10, e1001351.
- Hudry, B., Viala, S., Graba, Y., and Merabet, S. (2011). Visualization of protein interactions in living *Drosophila* embryos by the bimolecular fluorescence complementation assay. *BMC Biol* 9, 5.
- Hueber, S.D., Weiller, G.F., Djordjevic, M.A., and Frickey, T. (2010). Improving Hox protein classification across the major model organisms. *PLoS One* 5, e10820.
- Hughes, C.L., and Kaufman, T.C. (2002). Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede. *Development* 129, 1225-1238.
- Hui, J.H., McDougall, C., Monteiro, A.S., Holland, P.W., Arendt, D., Balavoine, G., and Ferrier, D.E. (2012). Extensive chordate and annelid macrosyteny reveals ancestral homeobox gene organization. *Mol Biol Evol* 29, 157-165.
- Hunt, P., Gulisano, M., Cook, M., Sham, M.H., Faiella, A., Wilkinson, D., Boncinelli, E., and Krumlauf, R. (1991). A distinct Hox code for the branchial region of the vertebrate head. *Nature* 353, 861-864.
- Hunter, M.P., and Prince, V.E. (2002). Zebrafish hox paralogue group 2 genes function redundantly as selector genes to pattern the second pharyngeal arch. *Dev Biol* 247, 367-389.
- Hurley, I., Hale, M.E., and Prince, V.E. (2005). Duplication events and the evolution of segmental identity. *Evol Dev* 7, 556-567.
- Iimura, T., and Pourquie, O. (2006). Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression. *Nature* 442, 568-571.
- Ikuta, T., Yoshida, N., Satoh, N., and Saiga, H. (2004). *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc Natl Acad Sci U S A* 101, 15118-15123.

- Inoue, T., Chisaka, O., Matsunami, H., and Takeichi, M. (1997). Cadherin-6 expression transiently delineates specific rhombomeres, other neural tube subdivisions, and neural crest subpopulations in mouse embryos. *Dev Biol* 183, 183-194.
- Isaacs, H., Pownall, M., and Slack, J. (1998). Regulation of Hox gene expression and posterior development by the *Xenopus* caudal homolog Xcad3. *EMBO J* 17, 3413-3427.
- Itasaki, N., Sharpe, J., Morrison, A., and Krumlauf, R. (1996). Reprogramming *Hox* expression in the vertebrate hindbrain: Influence of paraxial mesoderm and rhombomere transposition. *Neuron* 16, 487-500.
- Jabet, C., Gitti, R., Summers, M.F., and Wolberger, C. (1999). NMR studies of the pbx1 TALE homeodomain protein free in solution and bound to DNA: proposal for a mechanism of HoxB1-Pbx1-DNA complex assembly. *J Mol Biol* 291, 521-530.
- Jacobs, Y., Schnabel, C.A., and Cleary, M.L. (1999). Trimeric association of Hox and TALE homeodomain proteins mediates *Hoxb2* hindbrain enhancer activity. *Mol Cell Biol* 19, 5134-5142.
- Jaynes, J.B., and O'Farrell, P.H. (1988). Activation and repression of transcription by homeodomain-containing proteins that bind a common site. *Nature* 336, 744-749.
- Jones-Villeneuve, E.M., McBurney, M.W., Rogers, K.A., and Kalnins, V.I. (1982). Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J Cell Biol* 94, 253-262.
- Jones, F.S., Holst, B.D., Minowa, O., De Robertis, E.M., and Edelman, G.M. (1993). Binding and transcriptional activation of the promoter for the neural cell adhesion molecule by HoxC6 (Hox-3.3). *Proc Natl Acad Sci U S A* 90, 6557-6561.
- Jones, F.S., Prediger, E.A., Bittner, D.A., De Robertis, E.M., and Edelman, G.M. (1992). Cell adhesion molecules as targets for Hox genes: neural cell adhesion molecule promoter activity is modulated by cotransfection with Hox-2.5 and -2.4. *Proc Natl Acad Sci U S A* 89, 2086-2090.
- Jones, S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biol* 5, 226.
- Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* 371, 606-609.
- Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., and Mann, R.S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131, 530-543.

Joshi, R., Sun, L., and Mann, R. (2010). Dissecting the functional specificities of two Hox proteins. *Genes Dev* 24, 1533-1545.

Jung, H., Lacombe, J., Mazzoni, E.O., Liem, K.F., Jr., Grinstein, J., Mahony, S., Mukhopadhyay, D., Gifford, D.K., Young, R.A., Anderson, K.V., *et al.* (2010). Global control of motor neuron topography mediated by the repressive actions of a single hox gene. *Neuron* 67, 781-796.

Kamm, K., Schierwater, B., Jakob, W., Dellaporta, S.L., and Miller, D.J. (2006). Axial patterning and diversification in the cnidaria predate the Hox system. *Curr Biol* 16, 920-926.

Karch, F., Bender, W., and Weiffenbach, B. (1990). *abdA* expression in *Drosophila* embryos. *Genes Dev* 4, 1573-1587.

Karch, F., Weiffenbach, B., Peifer, M., Bender, W., Duncan, I., Celniker, S., Crosby, M., and Lewis, E.B. (1985). The abdominal region of the bithorax complex. *Cell* 43, 81-96.

Kashyap, V., Gudas, L.J., Brenet, F., Funk, P., Viale, A., and Scandura, J.M. (2011). Epigenomic reorganization of the clustered Hox genes in embryonic stem cells induced by retinoic acid. *J Biol Chem* 286, 3250-3260.

Kato, K., Kishi, T., Kamachi, T., Akisada, M., Oka, T., Midorikawa, R., Takio, K., Dohmae, N., Bird, P.I., Sun, J., *et al.* (2001). Serine proteinase inhibitor 3 and murinoglobulin I are potent inhibitors of neuropsin in adult mouse brain. *J Biol Chem* 276, 14562-14571.

Kawaguchi, J., Mee, P.J., and Smith, A.G. (2005). Osteogenic and chondrogenic differentiation of embryonic stem cells in response to specific growth factors. *Bone* 36, 758-769.

Kessel, M. (1992). Respecification of vertebral identities by retinoic acid. *Development* 115, 487-501.

Kiecker, C., and Lumsden, A. (2005). Compartments and their boundaries in vertebrate brain development. *Nat Rev Neurosci* 6, 553-564.

Kim, C.H., Hwang, D.Y., Park, J.J., and Kim, K.S. (2002a). A proximal promoter domain containing a homeodomain-binding core motif interacts with multiple transcription factors, including HoxA5 and Phox2 proteins, and critically regulates cell type-specific transcription of the human norepinephrine transporter gene. *J Neurosci* 22, 2579-2589.

- Kim, S.K., Selleri, L., Lee, J.S., Zhang, A.Y., Gu, X., Jacobs, Y., and Cleary, M.L. (2002b). Pbx1 inactivation disrupts pancreas development and in *Ipfl1*-deficient mice promotes diabetes mellitus. *Nat Genet* 30, 430-435.
- Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301, 331-333.
- Knoepfler, P.S., Bergstrom, D.A., Uetsuki, T., Dac-Korytko, I., Sun, Y.H., Wright, W.E., Tapscott, S.J., and Kamps, M.P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx- Meis1/Prep1. *Nucleic Acids Res* 27, 3752-3761.
- Kolm, P., and Sive, H. (1995). Regulation of the *Xenopus* labial homeodomain genes, *HoxA1* and *HoxD1*: activation by retinoids and peptide growth factors. *Dev Biol* 167, 34-49.
- Kondo, T., Dollé, P., Zakany, J., and Duboule, D. (1996). Function of posterior HoxD genes in the morphogenesis of the anal sphincter. *Development* 122, 2651-2659.
- Kostic, D., and Capecchi, M.R. (1994). Targeted disruptions of the murine *Hoxa-4* and *Hoxa-6* genes result in homeotic transformations of components of the vertebral column. *Mech Dev* 46, 231-247.
- Krumlauf, R. (1994). *Hox* genes in vertebrate development. *Cell* 78, 191-201.
- Kuraku, S., and Meyer, A. (2009). The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. *Int J Dev Biol* 53, 765-773.
- Kuziora, M.A., and McGinnis, W. (1988). Autoregulation of a *Drosophila* homeotic selector gene. *Cell* 55, 477-485.
- Langston, A.W., and Gudas, L.J. (1992). Identification of a retinoic acid responsive enhancer 3' of the murine homeobox gene *Hox-1.6*. *Mech Dev* 38, 217-228.
- LaRonde-LeBlanc, N.A., and Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev* 17, 2060-2072.
- Lawrence, H.J., Helgason, C.D., Sauvageau, G., Fong, S., Izon, D.J., Humphries, R.K., and Largman, C. (1997). Mice bearing a targeted interruption of the homeobox gene *HOXA9* have defects in myeloid, erythroid and lymphoid hematopoiesis. *Blood* 89, 1922-1930.

- Lee, H.M., Zhang, H., Schulz, V., Tuck, D.P., and Forget, B.G. (2010). Downstream targets of HOXB4 in a cell line model of primitive hematopoietic progenitor cells. *Blood* 116, 720-730.
- Lei, H., Wang, H., Juan, A.H., and Ruddle, F.H. (2005). The identification of Hoxc8 target genes. *Proc Natl Acad Sci U S A* 102, 2420-2424.
- Lelli, K.M., Noro, B., and Mann, R.S. (2011). Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity. *Proc Natl Acad Sci U S A* 108, 21122-21127.
- Lemons, D., and McGinnis, W. (2006). Genomic evolution of Hox gene clusters. *Science* 313, 1918-1922.
- Levine, M., Hafen, E., Garber, R.L., and Gehring, W.J. (1983). Spatial distribution of Antennapedia transcripts during *Drosophila* development. *EMBO J* 2, 2037-2046.
- Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565-570.
- Lewis, E.B. (1994). Homeosis: the first 100 years. *TIG* 10, 341-343.
- Li, X., and McGinnis, W. (1999). Activity regulation of Hox proteins, a mechanism for altering functional specificity in development and evolution. *Proc Natl Acad Sci U S A* 96, 6802-6807.
- Lin, C., Garrett, A.S., De Kumar, B., Smith, E.R., Gogol, M., Seidel, C., Krumlauf, R., and Shilatifard, A. (2011). Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes Dev* 25, 1486-1498.
- Linville, A., Gumusaneli, E., Chandraratna, R.A., and Schilling, T.F. (2004). Independent roles for retinoic acid in segmentation and neuronal differentiation in the zebrafish hindbrain. *Dev Biol* 270, 186-199.
- Lipshitz, H.D., Peattie, D.A., and Hogness, D.S. (1987). Novel transcripts from the *Ultrabithorax* domain of the bithorax complex. *Genes Dev* 1, 307-322.
- Liu, J.P., Laufer, E., and Jessell, T.M. (2001). Assigning the Positional Identity of Spinal Motor Neurons. Rostrocaudal Patterning of Hox-c Expression by FGFs, Gdf11, and Retinoids. *Neuron* 32, 997-1012.

- Locascio, A., Aniello, F., Amoroso, A., Manzanares, M., Krumlauf, R., and Branno, M. (1999). Patterning the ascidian nervous system: structure, expression and transgenic analysis of the CiHox3 gene. *Development* 126, 4737-4748.
- Lohnes, D. (2003). The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. *BioEssays* 25, 971-980.
- Longobardi, E., Penkov, D., Mateos, D., De Florian, G., Torres, M., and Blasi, F. (2013). Biochemistry of the tale transcription factors PREP, MEIS and PBX IN vertebrates. *Dev Dyn*.
- Lowe, C.J., Wu, M., Salic, A., Evans, L., Lander, E., Stange-Thomann, N., Gruber, C.E., Gerhart, J., and Kirschner, M. (2003). Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* 113, 853-865.
- Lufkin, T., Dierich, A., LeMeur, M., Mark, M., and Chambon, P. (1991). Disruption of the *Hox-1.6* homeobox gene results in defects in a region corresponding to its rostral domain of expression. *Cell* 66, 1105-1119.
- Lumsden, A., and Krumlauf, R. (1996). Patterning the vertebrate neuraxis. *Science* 274, 1109-1115.
- Ma, L., Reinhardt, F., Pan, E., Soutschek, J., Bhat, B., Marcusson, E.G., Teruya-Feldstein, J., Bell, G.W., and Weinberg, R.A. (2010). Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. *Nat Biotechnol* 28, 341-347.
- Ma, L., Teruya-Feldstein, J., and Weinberg, R.A. (2007). Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 449, 682-688.
- Maamar, H., Cabili, M.N., Rinn, J., and Raj, A. (2013). linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes Dev* 27, 1260-1271.
- Maconochie, M., Krishnamurthy, R., Nonchev, S., Meier, P., Manzanares, M., Mitchell, P.J., and Krumlauf, R. (1999). Regulation of *Hoxa2* in cranial neural crest cells involves members of the AP-2 family. *Development* 126, 1483-1494.
- Maconochie, M., Nonchev, S., Morrison, A., and Krumlauf, R. (1996). Paralogous *Hox* genes: function and regulation. *Annu Rev Genet* 30, 529-556.
- Maconochie, M.K., Nonchev, S., Studer, M., Chan, S.K., Popperl, H., Sham, M.H., Mann, R.S., and Krumlauf, R. (1997). Cross-regulation in the mouse *HoxB* complex: the expression of *Hoxb2* in rhombomere 4 is regulated by *Hoxb1*. *Genes Dev* 11, 1885-1896.

- Maden, M. (2002). Retinoid signalling in the development of the central nervous system. *Nat Rev Neurosci* 3, 843-853.
- Maeda, R.K., and Karch, F. (2006). The ABC of the BX-C: the bithorax complex explained. *Development* 133, 1413-1422.
- Mainguy, G., Koster, J., Woltering, J., Jansen, H., and Durston, A. (2007). Extensive polycistronism and antisense transcription in the mammalian Hox clusters. *PLoS One* 2, e356.
- Makki, N., and Capecchi, M.R. (2011). Identification of novel Hoxa1 downstream targets regulating hindbrain, neural crest and inner ear development. *Dev Biol* 357, 295-304.
- Mallo, M., Wellik, D.M., and Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate body plan. *Dev Biol* 344, 7-15.
- Manley, N., and Capecchi, M. (1995). The role of Hoxa-3 in mouse thymus and thyroid development. *Development* 121, 1989-2003.
- Manley, N., and Capecchi, M. (1998). Hox group 3 paralogs regulate the development and migration of the thymus, thyroid and parathyroid glands. *Dev Biol* 195, 1-15.
- Manley, N.R., and Capecchi, M.R. (1997). Hox group 3 paralogous genes act synergistically in the formation of somitic and neural crest-derived structures. *Dev Biol* 192, 274-288.
- Manley, N.R., Selleri, L., Brendolan, A., Gordon, J., and Cleary, M.L. (2004). Abnormalities of caudal pharyngeal pouch development in Pbx1 knockout mice mimic loss of Hox3 paralogs. *Dev Biol* 276, 301-312.
- Mann, R., and Chan, S.-K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *TIG* 12, 258-262.
- Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. *Curr Top Dev Biol* 88, 63-101.
- Manzanares, M., Bel-Vialer, S., Ariza-McNaughton, L., Ferretti, E., Marshall, H., Maconochie, M.K., Blasi, F., and Krumlauf, R. (2001). Independent regulation of initiation and maintenance phases of *Hoxa3* expression in the vertebrate hindbrain involves auto and cross-regulatory mechanisms. *Development* 128, 3595-3607.
- Manzanares, M., Nardelli, J., Gilardi-Hebenstreit, P., Marshall, H., Giudicelli, F., Martinez-Pastor, M.T., Krumlauf, R., and Charnay, P. (2002). Krox20 and kreisler co-

operate in the transcriptional control of segmental expression of *Hoxb3* in the developing hindbrain. *EMBO J* 21, 365-376.

Manzanares, M., Trainor, P.A., Nonchev, S., Ariza-McNaughton, L., Brodie, J., Gould, A., Marshall, H., Morrison, A., Kwan, C.T., Sham, M.H., *et al.* (1999). The role of *kreisler* in segmentation during hindbrain development. *Dev Biol* 211, 220-237.

Manzanares, M., Wada, H., Itasaki, N., Trainor, P.A., Krumlauf, R., and Holland, P.W. (2000). Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. *Nature* 408, 854-857.

Mapp, O.M., Walsh, G.S., Moens, C.B., Tada, M., and Prince, V.E. (2011). Zebrafish *Prickle1b* mediates facial branchiomotor neuron migration via a farnesylation-dependent nuclear activity. *Development* 138, 2121-2132.

Marshall, H., Morrison, A., Studer, M., Popperl, H., and Krumlauf, R. (1996). Retinoids and Hox genes. *FASEB J* 10, 969-978.

Marshall, H., Nonchev, S., Sham, M.H., Muchamore, I., Lumsden, A., and Krumlauf, R. (1992). Retinoic acid alters hindbrain Hox code and induces transformation of rhombomeres 2/3 into a 4/5 identity. *Nature* 360, 737-741.

Marshall, H., Studer, M., Pöpperl, H., Aparicio, S., Kuroiwa, A., Brenner, S., and Krumlauf, R. (1994). A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*. *Nature* 370, 567-571.

Matus, D.Q., Pang, K., Marlow, H., Dunn, C.W., Thomsen, G.H., and Martindale, M.Q. (2006). Molecular evidence for deep evolutionary roots of bilaterality in animal development. *Proc Natl Acad Sci U S A* 103, 11195-11200.

McBurney, M.W., Jones-Villeneuve, E.M., Edwards, M.K., and Anderson, P.J. (1982). Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. *Nature* 299, 165-167.

McCabe, C.D., Spyropoulos, D.D., Martin, D., and Moreno, C.S. (2008). Genome-wide analysis of the homeobox C6 transcriptional network in prostate cancer. *Cancer Res* 68, 1988-1996.

McClintock, J.M., Kheirbek, M.A., and Prince, V.E. (2002). Knockdown of duplicated zebrafish *hoxb1* genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. *Development* 129, 2339-2354.

- McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J. (1984). A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* 37, 403-408.
- McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* 68, 283-302.
- McIntyre, D.C., Rakshit, S., Yallowitz, A.R., Loken, L., Jeannotte, L., Capecchi, M.R., and Wellik, D.M. (2007). Hox patterning of the vertebrate rib cage. *Development* 134, 2981-2989.
- McNulty, C.L., Peres, J.N., Bardine, N., van den Akker, W.M., and Durston, A.J. (2005). Knockdown of the complete Hox paralogous group 1 leads to dramatic hindbrain and neural crest defects. *Development* 132, 2861-2871.
- Medina-Martinez, O., and Ramirez-Solis, R. (2003). In vivo mutagenesis of the Hoxb8 hexapeptide domain leads to dominant homeotic transformations that mimic the loss-of-function mutations in genes of the Hoxb cluster. *Dev Biol* 264, 77-90.
- Merabet, S., Kambris, Z., Capovilla, M., Berenger, H., Pradel, J., and Graba, Y. (2003). The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. *Dev Cell* 4, 761-768.
- Merabet, S., Saadaoui, M., Sambrani, N., Hudry, B., Pradel, J., Affolter, M., and Graba, Y. (2007). A unique Extradenticle recruitment mode in the *Drosophila* Hox protein Ultrabithorax. *Proc Natl Acad Sci U S A* 104, 16946-16951.
- Meyer, A. (1998). Hox gene variation and evolution. *Nature* 391, 225, 227-228.
- Meyer, A., and Van de Peer, Y. (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* 27, 937-945.
- Mihaly, J., Barges, S., Sipos, L., Maeda, R., Cleard, F., Hogga, I., Bender, W., Gyurkovics, H., and Karch, F. (2006). Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. *Development* 133, 2983-2993.
- Mihaly, J., Hogga, I., Barges, S., Galloni, M., Mishra, R.K., Hagstrom, K., Muller, M., Schedl, P., Sipos, L., Gausz, J., *et al.* (1998). Chromatin domain boundaries in the Bithorax complex. *Cell Mol Life Sci* 54, 60-70.
- Minoux, M., Antonarakis, G.S., Kmita, M., Duboule, D., and Rijli, F.M. (2009). Rostral and caudal pharyngeal arches share a common neural crest ground pattern. *Development* 136, 637-645.

- Moens, C.B., Cordes, S.P., Giorgianni, M.W., Barsh, G.S., and Kimmel, C.B. (1998). Equivalence in the genetic control of hindbrain segmentation in fish and mouse. *Development* *125*, 381-391.
- Moens, C.B., and Prince, V.E. (2002). Constructing the hindbrain: Insights from the zebrafish. *Dev Dyn* *224*, 1-17.
- Moens, C.B., and Selleri, L. (2006). Hox cofactors in vertebrate development. *Dev Biol* *291*, 193-206.
- Molotkova, N., Molotkov, A., Sirbu, I.O., and Duester, G. (2005). Requirement of mesodermal retinoic acid generated by Raldh2 for posterior neural transformation. *Mech Dev* *122*, 145-155.
- Moreno, E., Nadal, M., Baguna, J., and Martinez, P. (2009). Tracking the origins of the bilaterian Hox patterning system: insights from the acoel flatworm *Symsagittifera roscoffensis*. *Evol Dev* *11*, 574-581.
- Moroni, M., Vigano, M., and Mavilio, F. (1993). Regulation of the human *HOXD4* gene by retinoids. *Mech Dev* *44*, 139-154.
- Morsi El-Kadi, A.S., in der Reiden, P., Durston, A., and Morgan, R. (2002). The small GTPase Rap1 is an immediate downstream target for Hoxb4 transcriptional regulation. *Mech Dev* *113*, 131-139.
- Mummery, C.L., Feijen, A., Moolenaar, W.H., van den Brink, C.E., and de Laat, S.W. (1986). Establishment of a differentiated mesodermal line from P19 EC cells expressing functional PDGF and EGF receptors. *Exp Cell Res* *165*, 229-242.
- Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B.R. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* *272*, 548-551.
- Murakami, Y., Uchida, K., Rijli, F.M., and Kuratani, S. (2005). Evolution of the brain developmental plan: Insights from agnathans. *Dev Biol* *280*, 249-259.
- Nagy, L.M., Booker, R., and Riddiford, L.M. (1991). Isolation and embryonic expression of an abdominal-A-like gene from the lepidopteran, *Manduca sexta*. *Development* *112*, 119-129.
- Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R., *et al.* (2000). A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics* *154*, 1773-1784.

- Natale, A., Sims, C., Chiusano, M.L., Amoroso, A., D'Aniello, E., Fucci, L., Krumlauf, R., Branno, M., and Locascio, A. (2011). Evolution of anterior Hox regulatory elements among chordates. *BMC Evol Biol* *11*, 330.
- Neuteboom, S.T., and Murre, C. (1997). Pbx raises the DNA binding specificity but not the selectivity of antennapedia Hox proteins. *Mol Cell Biol* *17*, 4696-4706.
- Niederreither, K., Subbarayan, V., Dollé, P., and Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat Genet* *21*, 444-448.
- Niederreither, K., Vermot, J., Schuhbaur, B., Chambon, P., and Dollé, P. (2000). Retinoic acid synthesis and hindbrain patterning in the mouse embryo. *Development* *127*, 75-85.
- Nolte, C., Ahn, Y., and Krumlauf, R. (2012). Evolutionary Developmental Biology: *Hox* Gene Evolution. In eLS Citable Reviews in Life Sciences (Chichester: John Wiley and Sons LTD).
- Nolte, C., Jinks, T., Wang, X., Martinez Pastor, M.T., and Krumlauf, R. (2013). Shadow enhancers flanking the HoxB cluster direct dynamic Hox expression in early heart and endoderm development. *Dev Biol* *383*, 158-173.
- Noordermeer, D., and Duboule, D. (2013). Chromatin architectures and Hox gene collinearity. *Curr Top Dev Biol* *104*, 113-148.
- Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* *133*, 1277-1289.
- Ogura, T., and Evans, R. (1995a). Evidence for two distinct retinoic acid response pathways for *Hoxb-1* gene regulation. *Proc Natl Acad Sci USA* *92*, 392-396.
- Ogura, T., and Evans, R.M. (1995b). A retinoic acid-triggered cascade of HOXB1 gene activation. *Proc Natl Acad Sci U S A* *92*, 387-391.
- Oosterveen, T., Meijlink, F., and Deschamps, J. (2004). Expression of retinaldehyde dehydrogenase II and sequential activation of 5' Hoxb genes in the mouse caudal hindbrain. *Gene Expr Patterns* *4*, 243-247.
- Oosterveen, T., Niederreither, K., Dolle, P., Chambon, P., Meijlink, F., and Deschamps, J. (2003a). Retinoids regulate the anterior expression boundaries of 5' Hoxb genes in posterior hindbrain. *EMBO J* *22*, 262-269.

- Oosterveen, T., van Vliet, P., Deschamps, J., and Meijlink, F. (2003b). The direct context of a hox retinoic acid response element is crucial for its activity. *J Biol Chem* 278, 24103-24107.
- Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J., and Wuthrich, K. (1990). Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J* 9, 3085-3092.
- Packer, A.I., Crotty, D.A., Elwell, V.A., and Wolgemuth, D.J. (1998). Expression of the murine *Hoxa4* gene requires both autoregulation and a conserved retinoic acid response element. *Development* 125, 1991-1998.
- Pani, A.M., Mullarkey, E.E., Aronowicz, J., Assimacopoulos, S., Grove, E.A., and Lowe, C.J. (2012). Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* 483, 289-294.
- Papalopulu, N., Clarke, J.D., Bradley, L., Wilkinson, D., Krumlauf, R., and Holder, N. (1991a). Retinoic acid causes abnormal development and segmental patterning of the anterior hindbrain in *Xenopus* embryos. *Development* 113, 1145-1158.
- Papalopulu, N., Lovell-Badge, R., and Krumlauf, R. (1991b). The expression of murine *Hox-2* genes is dependent on the differentiation pathway and displays a collinear sensitivity to retinoic acid in F9 cells and *Xenopus* embryos. *Nucleic Acids Res* 19, 5497-5506.
- Parrish, M., Unruh, J., and Krumlauf, R. (2011). BAC modification through serial or simultaneous use of CRE/Lox technology. *J Biomed Biotechnol* 2011, 924068.
- Pasqualetti, M., Ori, M., Nardi, I., and Rijli, F.M. (2000). Ectopic *Hoxa2* induction after neural crest migration results in homeosis of jaw elements in *Xenopus*. *Development* 127, 5367-5378.
- Passner, J., Ryoo, H., Shen, L., Mann, R., and Aggarwal, A. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* 397, 714-718.
- Pata, I., Studer, M., van Doorninck, J.H., Briscoe, J., Kuuse, S., Engel, J.D., Grosveld, F., and Karis, A. (1999). The transcription factor GATA3 is a downstream effector of *Hoxb1* specification in rhombomere 4. *Development* 126, 5523-5531.
- Pattyn, A., Vallstedt, A., Dias, J.M., Samad, O.A., Krumlauf, R., Rijli, F.M., Brunet, J.F., and Ericson, J. (2003). Coordinated temporal and spatial control of motor neuron and serotonergic neuron generation from a common pool of CNS progenitors. *Genes Dev* 17, 729-737.

- Pavlopoulos, A., and Akam, M. (2011). Hox gene Ultrabithorax regulates distinct sets of target genes at successive stages of *Drosophila* haltere morphogenesis. *Proc Natl Acad Sci U S A* 108, 2855-2860.
- Peifer, M., and Wieschaus, E. (1990). Mutations in the *Drosophila* gene extradenticle affect the way specific homeo domain proteins regulate segmental identity. *Genes Dev* 4, 1209-1223.
- Peltenburg, L.T., and Murre, C. (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J* 15, 3385-3393.
- Penkov, D., Mateos San Martin, D., Fernandez-Diaz, L.C., Rossello, C.A., Torroja, C., Sanchez-Cabo, F., Warnatz, H.J., Sultan, M., Yaspo, M.L., Gabrieli, A., *et al.* (2013). Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. *Cell Rep* 3, 1321-1333.
- Phelan, M.L., and Featherstone, M.S. (1997). Distinct HOX N-terminal arm residues are responsible for specificity of DNA recognition by HOX monomers and HOX.PBX heterodimers. *J Biol Chem* 272, 8635-8643.
- Pinsonneault, J., Florence, B., Vaessin, H., and McGinnis, W. (1997). A model for extradenticle function as a switch that changes HOX proteins from repressors to activators. *EMBO J* 16, 2032-2042.
- Piper, D.E., Batchelor, A.H., Chang, C.P., Cleary, M.L., and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* 96, 587-597.
- Podlasek, C.A., Duboule, D., and Bushman, W. (1997). Male accessory sex organ morphogenesis is altered by loss of function of Hoxd-13. *Dev Dyn* 208, 454-465.
- Popodi, E., and Raff, R.A. (2001). Hox genes in a pentamerous animal. *BioEssays* 23, 211-214.
- Pöpperl, H., Bienz, M., Studer, M., Chan, S., Aparicio, S., Brenner, S., Mann, R., and Krumlauf, R. (1995). Segmental expression of *Hoxb1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/Pbx*. *Cell* 81, 1031-1042.
- Pöpperl, H., and Featherstone, M. (1992). An autoregulatory element of the murine *Hox-4.2* gene. *EMBO J* 11, 3673-3680.

Popperl, H., Rikhof, H., Chang, H., Haffter, P., Kimmel, C.B., and Moens, C.B. (2000). *lazarus* is a novel pbx gene that globally mediates hox gene function in zebrafish. *Mol Cell* 6, 255-267.

Powers, T.P., Hogan, J., Ke, Z., Dymbrowski, K., Wang, X., Collins, F.H., and Kaufman, T.C. (2000). Characterization of the Hox cluster from the mosquito *Anopheles gambiae* (Diptera: Culicidae). *Evol Dev* 2, 311-325.

Pownall, M.E., Isaacs, H.V., and Slack, J.M. (1998). Two phases of Hox gene regulation during early *Xenopus* development. *Curr Biol* 8, 673-676.

Prince, V.E., Moens, C.B., Kimmel, C.B., and Ho, R.K. (1998). Zebrafish *hox* genes: expression in the hindbrain region of wild-type and mutants of the segmentation gene, *valentino*. *Development* 125, 393-406.

Quiquand, M., Yanze, N., Schmich, J., Schmid, V., Galliot, B., and Piraino, S. (2009). More constraint on ParaHox than Hox gene families in early metazoan evolution. *Dev Biol* 328, 173-187.

Ramirez-Solis, R., Zheng, H., Whiting, J., Krumlauf, R., and Bradley, A. (1993). *Hoxb-4* (*Hox-2.6*) mutant mice show homeotic transformation of a cervical vertebra and defects in the closure of the sternal rudiments. *Cell* 73, 279-294.

Rank, G., Prestel, M., and Paro, R. (2002). Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. *Mol Cell Biol* 22, 8026-8034.

Rauskolb, C., Peifer, M., and Wieschaus, E. (1993). *extradenticle*, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene *pbx1*. *Cell* 74, 1101-1112.

Rauskolb, C., Smith, K.M., Peifer, M., and Wieschaus, E. (1995). *extradenticle* determines segmental identities throughout *Drosophila* development. *Development* 121, 3663-3673.

Rauskolb, C., and Wieschaus, E. (1994). Coordinate regulation of downstream genes by *extradenticle* and the homeotic selector proteins. *EMBO J* 13, 3561-3569.

Regulski, M., Dessain, S., McGinnis, N., and McGinnis, W. (1991). High-affinity binding sites for the Deformed protein are required for the function of an autoregulatory enhancer of the Deformed gene. *Genes Dev* 5, 278-286.

Rhee, J.W., Arata, A., Selleri, L., Jacobs, Y., Arata, S., Onimaru, H., and Cleary, M.L. (2004). *Pbx3* deficiency results in central hypoventilation. *Am J Pathol* 165, 1343-1350.

Rieckhof, G.E., Casares, F., Ryoo, H.D., Abu-Shaar, M., and Mann, R.S. (1997). Nuclear translocation of extradenticle requires *homothorax*, which encodes an extradenticle-related homeodomain protein. *Cell* 91, 171-183.

Rijli, F.M., Mark, M., Lakkaraju, S., Dierich, A., Dollé, P., and Chambon, P. (1993). A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. *Cell* 75, 1333-1349.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., *et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311-1323.

Roberts, D.J., and Tabin, C. (1994). The genetics of human limb development [editorial] [see comments]. *Am J Hum Genet* 55, 1-6.

Rohrschneider, M.R., Elsen, G.E., and Prince, V.E. (2007). Zebrafish *Hoxb1a* regulates multiple downstream genes including *prickle1b*. *Dev Biol* 309, 358-372.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein-DNA recognition. *Nature* 461, 1248-1253.

Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature*.

Rossel, M., and Capecchi, M.R. (1999). Mice mutant for both *Hoxa1* and *Hoxb1* show extensive remodeling of the hindbrain and defects in craniofacial development. *Development* 126, 5027-5040.

Ryoo, H.D., Marty, T., Casares, F., Affolter, M., and Mann, R.S. (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. *Development* 126, 5137-5148.

Saadaoui, M., Merabet, S., Litim-Mecheri, I., Arbeille, E., Sambrani, N., Damen, W., Brena, C., Pradel, J., and Graba, Y. (2011). Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity. *Proc Natl Acad Sci U S A* 108, 2276-2281.

Saegusa, H., Takahashi, N., Noguchi, S., and Suemori, H. (1996). Targeted disruption in the mouse *Hoxc-4* locus results in axial skeleton homeosis and malformation of the xiphoid process. *Dev Biol* 174, 55-64.

Sakai, Y., Meno, C., Fujii, H., Nishino, J., Shiratori, H., Saijoh, Y., Rossant, J., and Hamada, H. (2001). The retinoic acid-inactivating enzyme CYP26 is essential for

establishing an uneven distribution of retinoic acid along the antero-posterior axis within the mouse embryo. *Genes Dev* 15, 213-225.

Sanlioglu, S., Zhang, X., Baader, S.L., and Oberdick, J. (1998). Regulation of a Purkinje cell-specific promoter by homeodomain proteins: repression by engrailed-2 vs. synergistic activation by Hoxa5 and Hoxb7. *J Neurobiol* 36, 559-571.

Santagati, F., Minoux, M., Ren, S.Y., and Rijli, F.M. (2005). Temporal requirement of Hoxa2 in cranial neural crest skeletal morphogenesis. *Development* 132, 4927-4936.

Santagati, F., and Rijli, F.M. (2003). Cranial neural crest and the building of the vertebrate head. *Nat Rev Neurosci* 4, 806-818.

Sasaki, Y.T., Sano, M., Kin, T., Asai, K., and Hirose, T. (2007). Coordinated expression of ncRNAs and HOX mRNAs in the human HOXA locus. *Biochem Biophys Res Commun* 357, 724-730.

Schier, A.F., and Gehring, W.J. (1992). Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. *Nature* 356, 804-807.

Scott, M.P. (1992). Vertebrate Homeobox Gene Nomenclature. *Cell* 71, 551-553.

Scott, M.P., and Weiner, A.J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax and fushi tarazu loci of *Drosophila*. *Proc Natl Acad Sci USA* 81, 4115-4119.

Selleri, L., Depew, M.J., Jacobs, Y., Chanda, S.K., Tsang, K.Y., Cheah, K.S., Rubenstein, J.L., O'Gorman, S., and Cleary, M.L. (2001). Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* 128, 3543-3557.

Selleri, L., DiMartino, J., van Deursen, J., Brendolan, A., Sanyal, M., Boon, E., Capellini, T., Smith, K.S., Rhee, J., Popperl, H., *et al.* (2004). The TALE homeodomain protein Pbx2 is not essential for development and long-term survival. *Mol Cell Biol* 24, 5324-5331.

Seo, H.C., Edvardsen, R.B., Maeland, A.D., Bjordal, M., Jensen, M.F., Hansen, A., Flaata, M., Weissenbach, J., Lehrach, H., Wincker, P., *et al.* (2004). Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* 431, 67-71.

Serpente, P., Tumpel, S., Ghyselinck, N.B., Niederreither, K., Wiedemann, L.M., Dolle, P., Chambon, P., Krumlauf, R., and Gould, A.P. (2005a). Direct crossregulation between retinoic acid receptor {beta} and Hox genes during hindbrain segmentation. *Development* 132, 503-513.

Serpente, P., Tümpel, S., Ghyselinck, N.B., Niederreither, K., Wiedemann, L.M., Dollé, P., Chambon, P., Krumlauf, R., and Gould, A.P. (2005b). Direct crossregulation between retinoic acid receptor β and Hox genes during hindbrain segmentation. *Development* 132, 503-513.

Sessa, L., Breiling, A., Lavorgna, G., Silvestri, L., Casari, G., and Orlando, V. (2007). Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. *RNA* 13, 223-239.

Sharpe, J., Nonchev, S., Gould, A., Whiting, J., and Krumlauf, R. (1998). Selectivity, sharing and competitive interactions in the regulation of *Hoxb* genes. *EMBO J* 17, 1788-1798.

Shen, J., Wu, H., and Gudas, L.J. (2000). Molecular cloning and analysis of a group of genes differentially expressed in cells which overexpress the Hoxa-1 homeobox gene. *Exp Cell Res* 259, 274-283.

Shepherd, J.C., McGinnis, W., Carrasco, A.E., De Robertis, E.M., and Gehring, W.J. (1984). Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins. *Nature* 310, 70-71.

Shi, X., Bai, S., Li, L., and Cao, X. (2001). Hoxa-9 represses transforming growth factor-beta-induced osteopontin gene transcription. *J Biol Chem* 276, 850-855.

Shi, X., Yang, X., Chen, D., Chang, Z., and Cao, X. (1999). Smad1 interacts with homeobox DNA-binding proteins in bone morphogenetic protein signaling. *J Biol Chem* 274, 13711-13717.

Shippy, T.D., Ronshaugen, M., Cande, J., He, J., Beeman, R.W., Levine, M., Brown, S.J., and Denell, R.E. (2008). Analysis of the *Tribolium* homeotic complex: insights into mechanisms constraining insect Hox clusters. *Dev Genes Evol* 218, 127-139.

Shirasawa, S., Arata, A., Onimaru, H., Roth, K.A., Brown, G.A., Horning, S., Arata, S., Okumura, K., Sasazuki, T., and Korsmeyer, S.J. (2000). Rnx deficiency results in congenital central hypoventilation. *Nat Genet* 24, 287-290.

Simeone, A., Acampora, D., Arcioni, L., Andrews, P.W., Boncinelli, E., and Mavilio, F. (1990). Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. *Nature* 346, 763-766.

Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F., and Boncinelli, E. (1991). Differential regulation by retinoic acid of the homeobox genes of the four *HOX* loci in human embryonal carcinoma cells. *Mech Dev* 33, 215-227.

- Sirbu, I.O., Gresh, L., Barra, J., and Duester, G. (2005). Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression. *Development* *132*, 2611-2622.
- Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., *et al.* (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* *147*, 1270-1282.
- Small, K.M., and Potter, S.S. (1993). Homeotic transformations and limb defects in Hox A11 mutant mice. *Genes Dev* *7*, 2318-2328.
- Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., *et al.* (2013a). Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet* *45*, 415-421, 421e411-412.
- Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., *et al.* (2013b). Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet* *45*, 415-421.
- Sorge, S., Ha, N., Polychronidou, M., Friedrich, J., Bezdán, D., Kaspar, P., Schaefer, M.H., Ossowski, S., Henz, S.R., Mundorf, J., *et al.* (2012). The cis-regulatory code of Hox function in *Drosophila*. *EMBO J* *31*, 3323-3333.
- Spitz, F., and Duboule, D. (2008). Global control regions and regulatory landscapes in vertebrate development and evolution. *Adv Genet* *61*, 175-205.
- Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* *113*, 405-417.
- Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N., *et al.* (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* *450*, 219-232.
- Strathern, J., Hicks, J., and Herskowitz, I. (1981). Control of cell type in yeast by the mating type locus. The alpha 1-alpha 2 hypothesis. *J Mol Biol* *147*, 357-372.
- Strickland, S., and Mahdavi, V. (1978). The induction of differentiation in teratocarcinoma stem cells by retinoic acid. *Cell* *15*, 392-403.
- Strickland, S., and Sawey, M.J. (1980). Studies on the effect of retinoids on the differentiation of teratocarcinoma stem cells in vitro and in vivo. *Dev Biol* *78*, 76-85.

Strubing, C., Ahnert-Hilger, G., Shan, J., Wiedenmann, B., Hescheler, J., and Wobus, A.M. (1995). Differentiation of pluripotent embryonic stem cells into the neuronal lineage in vitro gives rise to mature inhibitory and excitatory neurons. *Mech Dev* 53, 275-287.

Struhl, G., and White, R.A. (1985). Regulation of the Ultrabithorax gene of *Drosophila* by other bithorax complex genes. *Cell* 43, 507-519.

Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F., Chambon, P., and Krumlauf, R. (1998a). Genetic interaction between *Hoxa1* and *Hoxb1* reveal new roles in regulation of early hindbrain patterning. *Development* 125, 1025-1036.

Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F.M., Chambon, P., and Krumlauf, R. (1998b). Genetic interactions between *Hoxa1* and *Hoxb1* reveal new roles in regulation of early hindbrain patterning. *Development* 125, 1025-1036.

Studer, M., Lumsden, A., Ariza-McNaughton, L., Bradley, A., and Krumlauf, R. (1996). Altered segmental identity and abnormal migration of motor neurons in mice lacking *Hoxb-1*. *Nature* 384, 630-634.

Studer, M., Popperl, H., Marshall, H., Kuroiwa, A., and Krumlauf, R. (1994). Role of a conserved retinoic acid response element in rhombomere restriction of *Hoxb-1*. *Science* 265, 1728-1732.

Taneja, R., Thisse, B., Rijli, F.M., Thisse, C., Bouillet, P., Dollé, P., and Chambon, P. (1996). The expression pattern of the mouse receptor tyrosine kinase gene *MDK1* is conserved through evolution and requires *Hoxa-2* for rhombomere-specific expression in mouse embryos. *Dev Biol* 177, 397-412.

Tarchini, B., and Duboule, D. (2006). Control of *Hoxd* genes' collinearity during early limb development. *Dev Cell* 10, 93-103.

Tarchini, B., Huynh, T.H., Cox, G.A., and Duboule, D. (2005). *HoxD* cluster scanning deletions identify multiple defects leading to paralysis in the mouse mutant Ironside. *Genes Dev* 19, 2862-2876.

Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res* 13, 382-390.

TenHarmsel, A., and Biggin, M.D. (1995). Bending DNA can repress a eukaryotic basal promoter and inhibit TFIID binding. *Mol Cell Biol* 15, 5492-5498.

Thali, M., Muller, M.M., DeLorenzi, M., Matthias, P., and Bienz, M. (1988). *Drosophila* homoeotic genes encode transcriptional activators similar to mammalian OTF-2. *Nature* 336, 598-601.

Theokli, C., Morsi El-Kadi, A.S., and Morgan, R. (2003). TALE class homeodomain gene *Irx5* is an immediate downstream target for *Hoxb4* transcriptional regulation. *Dev Dyn* 227, 48-55.

Tomotsune, D., Shoji, H., Wakamatsu, Y., Kondoh, H., and Takahashi, N. (1993). A mouse homologue of the *Drosophila* tumour-suppressor gene *l(2)gl* controlled by *Hox-C8* in vivo. *Nature* 365, 69-72.

Tour, E., Hittinger, C.T., and McGinnis, W. (2005). Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development* 132, 5271-5281.

Trainor, P., and Krumlauf, R. (2000a). Plasticity in mouse neural crest cells reveals a new patterning role for cranial mesoderm. *Nat Cell Biol* 2, 96-102.

Trainor, P.A., and Krumlauf, R. (2000b). Patterning the cranial neural crest: Hindbrain segmentation and *Hox* gene plasticity. *Nat Rev Neurosci* 1, 116-124.

Trainor, P.A., and Krumlauf, R. (2001). Hox genes, neural crest cells and branchial arch patterning. *Curr Opin Cell Biol* 13, 698-705.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E., and Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell* 59, 553-562.

Tümpel, S., Cambroner, F., Ferretti, E., Blasi, F., Wiedemann, L.M., and Krumlauf, R. (2007). Expression of *Hoxa2* in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon *Hoxb1*. *Dev Biol* 302, 646-660.

Tümpel, S., Wiedemann, L.M., and Krumlauf, R. (2009). Hox genes and segmentation of the vertebrate hindbrain. In *HOX Genes*, O. Pourquié, ed. (Burlington: Academic Press), pp. 103-137.

Tvrđik, P., and Capecchi, M.R. (2006). Reversal of *hox1* gene subfunctionalization in the mouse. *Dev Cell* 11, 239-250.

Urata, M., Tsuchimoto, J., Yasui, K., and Yamaguchi, M. (2009). The *Hox8* of the hemichordate *Balanoglossus misakiensis*. *Dev Genes Evol* 219, 377-382.

van de Ven, C., Bialecka, M., Neijts, R., Young, T., Rowland, J.E., Stringer, E.J., Van Rooijen, C., Meijlink, F., Novoa, A., Freund, J.N., *et al.* (2011). Concerted involvement of *Cdx/Hox* genes and Wnt signaling in morphogenesis of the caudal neural tube and cloacal derivatives from the posterior growth zone. *Development* 138, 3451-3462.

van Rooijen, C., Simmini, S., Bialecka, M., Neijts, R., van de Ven, C., Beck, F., and Deschamps, J. (2012). Evolutionarily conserved requirement of Cdx for post-occipital tissue emergence. *Development* 139, 2576-2583.

Vieux-Rochas, M., Mascrez, B., Krumlauf, R., and Duboule, D. (2013). Combined function of HoxA and HoxB clusters in neural crest cells. *Dev Biol*.

Vinagre, T., Moncaut, N., Carapuco, M., Novoa, A., Bom, J., and Mallo, M. (2010). Evidence for a myotomal Hox/Myf cascade governing nonautonomous control of rib specification within global vertebral domains. *Dev Cell* 18, 655-661.

Violette, S.M., Shashikant, C.S., Salbaum, J.M., Belting, H.G., Wang, J.C., and Ruddle, F.H. (1992). Repression of the beta-amyloid gene in a Hox-3.1-producing cell line. *Proc Natl Acad Sci U S A* 89, 3805-3809.

Vitobello, A., Ferretti, E., Lampe, X., Vilain, N., Ducret, S., Ori, M., Spetz, J., Selleri, L., and Rijli, F.M. (2011). Hox and Pbx factors control retinoic acid synthesis during hindbrain segmentation. *Dev Cell* 20, 469-482.

Wada, H., Escriva, H., Zhang, S., and Laudet, V. (2006). Conserved RARE localization in amphioxus Hox clusters and implications for Hox code evolution in the vertebrate neural crest. *Dev Dyn*.

Wada, H., Garcia-Fernandez, J., and Holland, P.W. (1999). Colinear and segmental expression of amphioxus Hox genes. *Dev Biol* 213, 131-141.

Wang, Y., Jones, F.S., Krushel, L.A., and Edelman, G.M. (1996). Embryonic expression patterns of the neural cell adhesion molecule gene are regulated by homeodomain binding sites. *Proc Natl Acad Sci U S A* 93, 1892-1896.

Waskiewicz, A.J., Rikhof, H.A., Hernandez, R.E., and Moens, C.B. (2001). Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. *Development* 128, 4139-4151.

Waskiewicz, A.J., Rikhof, H.A., and Moens, C.B. (2002). Eliminating zebrafish pbx proteins reveals a hindbrain ground state. *Dev Cell* 3, 723-733.

Wellik, D.M. (2007). Hox patterning of the vertebrate axial skeleton. *Dev Dyn* 236, 2454-2463.

Wellik, D.M. (2009). Hox Genes and Vertebrate Axial Pattern. *Curr Top Dev Biol* 88, 257-278.

Wellik, D.M., and Capecchi, M.R. (2003). Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science* 301, 363-367.

White, R.A., and Wilcox, M. (1984). Protein products of the bithorax complex in *Drosophila*. *Cell* 39, 163-171.

White, R.A., and Wilcox, M. (1985). Distribution of Ultrabithorax proteins in *Drosophila*. *EMBO J* 4, 2035-2043.

White, R.J., and Schilling, T.F. (2008). How degrading: Cyp26s in hindbrain development. *Dev Dyn* 237, 2775-2790.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P.W., Stott, D., and Allemann, R.K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox-2.6* gene expression. *Genes Dev* 5, 2048-2059.

Williams, T.M., Williams, M.E., Kuick, R., Misek, D., McDonagh, K., Hanash, S., and Innis, J.W. (2005). Candidate downstream regulated genes of HOX group 13 transcription factors with and without monomeric DNA binding capability. *Dev Biol* 279, 462-480.

Wobus, A.M., Kaomei, G., Shan, J., Wellner, M.C., Rohwedel, J., Ji, G., Fleischmann, B., Katus, H.A., Hescheler, J., and Franz, W.M. (1997). Retinoic acid accelerates embryonic stem cell-derived cardiac differentiation and enhances development of ventricular cardiomyocytes. *J Mol Cell Cardiol* 29, 1525-1539.

Woltering, J.M., and Duboule, D. (2010). The origin of digits: expression patterns versus regulatory mechanisms. *Dev Cell* 18, 526-532.

Wong, E.Y., Wang, X.A., Mak, S.S., Sae-Pang, J.J., Ling, K.W., Fritsch, B., and Sham, M.H. (2011). Hoxb3 negatively regulates Hoxb1 expression in mouse hindbrain patterning. *Dev Biol* 352, 382-392.

Wright, C.V. (1993). Hox genes and the hindbrain. *Curr Biol* 3, 618-621.

Yasukochi, Y., Ashakumary, L.A., Wu, C., Yoshido, A., Nohata, J., Mita, K., and Sahara, K. (2004). Organization of the Hox gene cluster of the silkworm, *Bombyx mori*: a split of the Hox cluster in a non-*Drosophila* insect. *Dev Genes Evol* 214, 606-614.

Young, T., Rowland, J.E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.N., *et al.* (2009). Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev Cell* 17, 516-526.

Zakany, J., and Duboule, D. (2007). The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* 17, 359-366.

Zakany, J., Kmita, M., and Duboule, D. (2004). A dual role for Hox genes in limb anterior-posterior asymmetry. *Science* 304, 1669-1672.

Zakany, J., Zacchetti, G., and Duboule, D. (2007). Interactions between HOXD and Gli3 genes control the limb apical ectodermal ridge via Fgf10. *Dev Biol* 306, 883-893.

Zavortink, M., and Sakonju, S. (1989). The morphogenetic and regulatory functions of the Drosophila Abdominal-B gene are encoded in overlapping RNAs transcribed from separate promoters. *Genes Dev* 3, 1969-1981.

Zeng, C., Pinsonneault, J., Gellon, G., McGinnis, N., and McGinnis, W. (1994). Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in Drosophila embryos. *EMBO J* 13, 2362-2377.

Zhang, M., Kim, H.J., Marshall, H., Gendron-Maguire, M., Lucas, D.A., Baron, A., Gudas, L.J., Gridley, T., Krumlauf, R., and Grippo, J.F. (1994). Ectopic *Hoxa-1* induces rhombomere transformation in mouse hindbrain. *Development* 120, 2431-2442.

Zhang, X., Lian, Z., Padden, C., Gerstein, M.B., Rozowsky, J., Snyder, M., Gingeras, T.R., Kapranov, P., Weissman, S.M., and Newburger, P.E. (2009). A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* 113, 2526-2534.

Zhao, Y., and Potter, S.S. (2001). Functional specificity of the Hoxa13 homeobox. *Development* 128, 3197-3207.

Zhao, Y., and Potter, S.S. (2002). Functional comparison of the Hoxa 4, Hoxa 10, and Hoxa 11 homeoboxes. *Dev Biol* 244, 21-36.

Zheng, Z., Khoo, A., Fambrough, D., Jr., Garza, L., and Booker, R. (1999). Homeotic gene expression in the wild-type and a homeotic mutant of the moth *Manduca sexta*. *Dev Genes Evol* 209, 460-472.

Appendix I. Genes Up-regulated in Hoxa1 mutant

Genes	Log2 Fold change versus WT
Alpk2	0.430206882
Ankrd1	0.556436323
Apob	0.585545552
Hoxa2	0.239823291
Lefty2	0.228576264
Tinag	1.355139351
Adcy7	0.496604403
Ak7	0.962942235
Auts2	0.775514041
Cobll1	0.653284181
Glis3	0.754074854
Msx1	0.657403952
Myo7a	0.545005802
Nme5	0.532166848
Nr2f2	0.494757828
Olig3	0.799985541
Otx2	1.053918188
Sema3c	0.526695846
Trps1	0.55996234

Source (Makki and Capecchi, 2011; Tvrđik and Capecchi, 2006)

Appendix II. Genes Down-regulated in Hoxa1 mutant

Genes	Log2 Fold change versus WT
Bcl11a	-0.506260885
Cabp7	-0.388468035
Dner	-0.36839714
Exoc4	-0.708306122
Hoxa1	-0.89439078
Hoxb1	-0.201101209
Krt15	-0.244745645
Lhx5	-0.272753767
Mafb	-0.379421841
Rgmb	-0.865457647
Sema3c	-0.54280111
Tbc1d23	-0.547716869
Tll1	-0.997166602
Atp8a1	-0.575073316
Bcl11a	-0.506260885
Col3a1	-0.577315673
Hoxb1	-0.201101209
Irf2bp2	-0.587030844
Kirrel3	-1.065386201
Nr6a1	-0.48867364
Sorl1	-0.754517847
Sox2	-0.681208505

Source (Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006)

Appendix III. Genes Down- regulated in Hoxb1 mutant

Genes	Log2 Fold change versus WT
Arf2	-0.311018912
Axin2	-0.263918826
Bcat2	-0.344768424
Cdh22	-0.378238012
Celsr1	-0.353961635
Coro2b	-0.309950801
Cpz	-0.268378621
Ctnnd2	-0.3124323
Ednrb	-0.336418444
Gbas	-0.299753794
Hmga1	-0.263437727
Hoxb1	-2.708205235
Hoxb2	-0.32471397
Hoxb3	-0.316969943
Igsf8	-0.304005577
Lhx3	-0.315824574
Limd1	-0.262241219
Man2b2	-3.601136176
Mapt	-0.255002632
Neurod4	-0.315810251
Polr2a	-0.561229933
Ppp2r2b	-0.397155417
Ptprf	-0.445227771
Ptprs	-0.279734691
Rab6b	-0.290895746
Sema3f	-0.28370613
Sema6c	-0.279354704
Set	-0.494713267
Slc12a5	-0.61596762
Sox21	-0.377570737
Sulf2	-0.256581417
Taf4a	-0.280911457

Source (Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006)

Appendix IV. Genes Up- regulated in Hoxb1 mutant

Genes	Log2 Fold change versus WT
Bbx	0.339074451
Gpc6	0.261984464
Jag1	1.039695389
Lhx2	0.175526972
Limd1	0.126188307
Lrrfip1	0.269160081
Nrip1	0.284317397
Polr3k	0.387673608
Rpl23	0.224128679
Sst	0.620012925
Tcf3	0.247283603

Source (Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006)

Appendix V. Comparative list of genes changed in Hoxa1 and Hoxb1 mutant

Expression down-regulated in Hoxa1 ^{-/-}		Expression up-regulated in Hoxa1 ^{-/-}	
Down-regulated in Hoxb1 ^{-/-}	Up-regulated in Hoxb1 ^{-/-}	Down-regulated in Hoxb1 ^{-/-}	Up-regulated in Hoxb1 ^{-/-}
Serpib11	Fcgr3	Rhbg	Hhat
Rasgrp1	Plxdc2	Fcgr1	Unc80
Slc32a1	Mtx2	Edn2	Api5
Kcnb1	Dabl	Cps1	Barhl1
Pou3f1	Kif2c	Csmd2	Rin2
Fcgr4	Mynn	Csf3r	Taf4a
Pex5l	Scrt2	Icos	Acp6
Iqca	Taf12	Pax5	Epha8
Hrh3	Set	Arid5a	Ilf2
Hnf4a	Elf2	Gja10	Polr3k
Gpr149	Dlgap4	Pax1	Aff3
Lin28a	Kdm5b	Tm4sf1	Dbc1
Slc24a2	Mdm4	Gpr158	Tbc1d20
Calb1	Gpsm2	St6galnac3	Cdh20
Rims4	Chrb2	Matn1	Hmcn1
C1ql2	Rnf38	Vstm2l	Sulf1
Npl	Arhgef10l	Lrrc4c	Prex1
Ptchd2	Slc31a2	Mc3r	Fam123c
Hfe2	Dnajb14	Kcnh1	Fam5b
Raly	Ccnc	Slc23a2	Nfs1
Plekhb2	Tor1b	Satb2	Lphn2
Bai3	Elavl2	Zc3h12a	Arhgap29
Pecr	Hpca	Actr3	Scg2
Plekhf2	Nvl	Epha7	Pdyn
Pbx3	Stmn3	Lhx2	Shox2
Gjb5	Cbln4	Dpyd	Camta1
Notch2	Stmn2	Fam78b	Zfp238
Madd	Nvl	Pcsk2	Cyr61
Mareksl1	Stmn3	Cep170	Runx1t1
Mmp24	Cbln4	Ctxn2	Gpd2
Ccdc115	Stmn2	Otor	Neurod1
Chgb		Arid1a	Snap25
Med27		Fam163b	Ino80d
Sdhc		Pla2g4a	Clasp1
Dmrta2		Col19a1	Casc5
Ndst4		Epha7	Cerkl
Chst1		Lhx2	Cpne1
Shf		Dpyd	Shox2.1
Dtwd1		Fam78b	Rbm12
Plekhf2		Pcsk2	Pdyn

Pbx3		Cep170	Shox2
Gjb5		Ctnn2	Camta1
Notch2		Otor	Zfp238
Madd		Arid1a	Cyr61
Marcks11		Fam163b	Runx1t1
Mmp24		Pla2g4a	Gpd2
Ccdc115		Col19a1	Neurod1
Chgb			Snap25
Med27			Ino80d
Sdhc			Clasp1
Dmrta2			Casc5
Ndst4			Cerkl
Chst1			Cpne1
Shf			Shox2.1
Dtwd1			Rbm12
Mpzl1			
Ccdc108			
Tomm40l			
Trna1ap			
Igsf21			
Dlgap3			
Garnl3			
Vwc2l			
Psmb2			

(Source: Gouti and Gavalas, 2008; Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006)

Appendix VI. Direct Hoxb1 identified by Gavalas lab

After long induction		
Accn1	Lrrc4c	Tram111
Aplp1	Ly6h	Tusc3
Armc8	Mapt	Wnt7b
Atp2b2	Marcks	Zfp238
B3gnt1	Mpz11	Syt7
Bai3	Nap113	Thbs1
Bbx	Nek6	Gabrb3
Bcl11a	Neurod1	Gsta4
Cadps	Nptx2	H2-D1
Calb1	Nr4a2	H2-K1
Ccnc	Nrxn3	Hey2
Cdh20	Ntrk2	Hoxa2
Chga	Odz2	Hoxb1
Chgb	Oma1	Hoxb2
Chst2	Pax5	Hoxb3
Cntnap2	Pcdh20	Hs6st1
Crmp1	Pcsk2	Kirrel3
Ddah1	Plxdc2	Klf12
Dmrta2	Podxl	Lhx2
Dpyd	Pou3f1	Lmo1
Ednrb	Ppp2r2b	Lrp11
Fbxl16	Ptpu	Stmn3
Flrt2	Rab27b	Sulf1
Foxa2	Rasgrp1	Sert1
Spock1	Rcsd1	Shox2
St6galnac5	Rgmb	Snx6
St8sia1	Rtn1	Stmn2
Stk35	Scg2	Scg3

Source (Gouti and Gavalas, 2008)

Appendix VII. Protein bound on Hoxb1-ARE

NCBI_Gene	Relative abundance (dNSAF)
Eif5a	0.01171009
Yy1	0.00051015
Dppa4	0.00049398
Klf16	0.00038836
Pbx2	0.00035264
Pcgf2	0.00033253
Pbx1	0.00028967
Med28	0.00027382
Med27	0.0002612
Hoxb4	0.00025994
Med1	0.00025788
Med31	0.00024804
Med19	0.00019975
Med24	0.0001646
Numb	0.00016139
Foxk1	0.00015817
Med18	0.00015622
Iws1	0.00014847
Foxr1	0.00014571
Med15	0.00014414
Cdk9	0.00013102
Med17	0.00012517
Med8	0.00012124
Smad1	0.00011979
Hdac3	0.00011388
Parg	0.00010143
Zc3hc1	0.00010133
Bmi1	0.00010029

Appendix VIII. Publications