The 3-Dimensional Landscape of the Mouse HoxA Cluster Conforms to Collinear Function

By

Matthew Pinchuk

Department of Biochemistry

McGill University

Montreal, Quebec, Canada

January 2009

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

Abstract

Now that the genome has successfully been sequenced, the next aim will be to characterize and annotate features. With the development of high resolution tools, such as Chromosome Conformation Capture (3C) technology, the 3-dimensional architecture of large linear spans of DNA can be determined. We have established that the mouse HoxA gene cluster is structured in such a way as to contain four distinct looping regions with a central interacting rosette core in its resting pluripotent state. This central core feature might directly affect the spatiotemporal regulation of the Hox genes during normal development as well as differentiation via retinoic acid administration in the P19 embryonal carcinoma cell line. The DNA looping might be mediated by cis-regulatory elements, which could function in a cooperative manner to regulate HoxA gene expression. Regulatory elements might include the previously identified retinoic acid response elements (RAREs) known to mediate specific developmental cues. Interestingly, Hox genes expressed earlier and more anteriorly are within loops containing higher amounts of identified RAREs as well as other conserved non-coding sequences that also appear to play a potential role in orchestrating the collinear mechanism of action. Loops containing the more 5' located genes have fewer regulatory elements and therefore may respond later to developmental cues.

Résumé

Maintenant que le génome été séquencé avec succès, le but suivant sera de caractériser et annoter les caractéristiques regulataire. Avec le développement d'instruments de haute résolution, comme la technologie du Capture de Conformation Chromatin (3C), l'architecture 3- dimensionnelle de grandes durées linéaires d'ADN peut être déterminée. Nous avons établi que le groupe de gène de HoxA dans la souris est structurée d'une telle façon pour contenir quatre régions boucles distinctes avec un centre de rosette réagissant réciproquement central dans son état de reposant pluripotent. On croit que cette caractéristique centrale de base directement affecte le règlement spatiotemporel des gènes Hox pendant le développement normal aussi bien que la différentiation via l'administration acide rétinoïque dans le P19 ligne de cellule de carcinome. Les éléments cis-de-contrôle responsables de définir ces boucles ont l'air de fonctionner dans une manière coopérative dans laquelle le pluripotent HoxA l'état est plein d'assurance de réagir à l'administration acide rétinoïque dans une manière colinéaire contrôlée d'expression de gène. Auparavant identifié les éléments de réponse acide rétinoïque (ERARs) sont organisés pour tenir compte de la réorganisation dynamique du groupe sur la réception de signaux spécifiques du développement. Les gènes de Hox ont exprimé tôt et sont plus antérieurement dans les boucles qui contiennent de plus hautes quantités de ERARs identifié aussi bien que d'autres ordres de non-codification conservés qui ont aussi l'air de jouer un rôle potentiel dans le fait d'orchestrer le mécanisme colinéaire d'action.

Acknowledgments

First and foremost I would like to thank my family for always believing that I was capable of achieving anything and always showing their support in every aspect of my life, including my academics.

I would like to thank Carol for being there to take the edge off when things got stressful and for just being a friend.

Maria, you were there for me whenever I needed you. You helped me get through the most difficult situation I have ever experienced and don't know how I will ever repay you. I would not have made it this far without you.

Thank you to Dr. Dostie for taking a chance on me. I want you to know that I still have the upmost respect and appreciation towards you.

Thanks to McGill for the funding through the chemical biology scholarship award that allowed our research to continue.

Thank you to everyone else that was involved in this work and my life.

M.S.J.G.A.K.M.D

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

List of Abbreviations

3C	Chromosome conformation capture
5C	Chromosome conformation capture carbon copy
Abd	Abdominal
Alu	Arthrobacter luteus element
Antp	Antennapedia
A-P	Anterior-posterior
BAC	Bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
BLAT	Blast like alignment tool
Bmp8b	Bone morphogenetic protein 8b
BX-C	Bithorax complex
CBP	CREB binding protein
CCS	Chromatin conformation signatures
CDX	Caudal type Homeobox
CENP-A	Centromere protein A
CNS	Conserved non-coding sequence
CREB	cAMP response element binding
CT	Chromosome territories
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
EC	Embryonal carcinoma
EED	Embryonic ectoderm development
EGCG	Epigallocatechin gallate
EKLF	Erythroid Krüppel-like Factor
ENCODE	Encyclopedia of DNA elements
ES	Embryonic stem
EST	Expressed sequence tag
Evx1	Even skipped Homeobox homolog 1
Exd	Extradenticle
EZH2	Enhancer of zeste homolog 2
FGF	Fibroblast growth factor
FISH	Fluorescent in situ hybridization

GATA Globin transcription factor
GCR Global control region
GD Gene desert
GR Glucocorticoid receptor
GRN Gene regulatory networks
hGH Human growth hormone
Hibadh 3-hydroxyisobutyrate dehydrogenase
HKMT Histone lysine methyl transferase
HMGA1 High mobility group AT-hook 1
HMTase Histone methyl transferase
HOTAIR HOX antisense intergenic RNA
Hox Homeobox family
HP1 Heterochromatin protein 1
ICR Imprinting control region
IF Interaction frequency
Igf2 Insulin-like growth factor 2
Jazf1 Juxtaposed with another zinc finger gene 1
Jmjc Jumonji domain -containing
LCR Locus control region
LINES Long interspersed repeated DNA
LMA late maturity α-amylase
Lnp Lunapark
Mash-1 Mammalian achate schute Homolog *1*
Meis1 Myeloid ecotropic viral integration site
MF Mouse fragment
MGMT O6-methylguanine-DNA methyltransferase
miRNA Micro RNA
MLS Microphthalamia with linear skin defects
MOF Males absent on the first
MOZ Myeloid ecotropic viral integration site
ncRNA Non-coding RNA
ors Olfactory receptors
PAD4 Peptidylarginine deiminase 4
Pbx Pre-B-cell leukemia homeobox
PCAF P300/CBP-associated factor
PcG Polycomb group proteins
PCR Polymerase chain reaction
PMA Para-methoxyamphetamine
PRC1/2 Polycomb repressive complex
PRE Polycomb response element
PRMT Protein arginine methyltransferase
RA Retinoic acid

Rad	Ras associated with diabetes
RAR	Retinoic acid receptor
RARE	Retinoic acid receptor element
RNA	Ribonucleic acid
Rsk	Ribosomal S6 Kinase
RXR	Retinoid x receptor
SET	Su(var)3-9) Enhancer of zeste (E(z)) trithorax (Trx)
shRNA	Small hairpin RNA
Sir2	Silent information regulator 2
siRNA	Short interfering ribonucleic acid
Smad1	Similar to mothers against decapentaplegic homolog 1
STAT	Signal Transducers and Activator of Transcription
SUMO	Small Ubiquitin-like Modifier
SUZ12	Suppressor of zeste 12 homolog
TALE	Three amino acid loop extension
Tax1bp1	T-cell leukemia virus type I binding protein 1
TCR	T-cell receptor
TF	Transcription Factor
TH2	T-Cell Helper 2
Ubx	Ultra bithorax
UCSC	University of California Santa Cruz
UTR	Untranscribed region
UTX	Ultra trithorax
UV	Ultraviolet
WNT	Wingless and INT
Xi	Inactive X Chromosome

Section 1: Introduction

1.1. General Introduction

The mechanism of transcription is a tightly regulated process, which ensures that the correct expression of appropriate genes is coordinated to follow developmental, environmental and cell cycle cues. This being considered, it is not entirely surprising that more and more complex modes of gene regulation are being discovered. To begin, linear regulatory elements such as promoters and enhancers activate genes based on the availability of specific required transcription factors [1]. In addition, the condensation state of chromatin influences the ability of said transcription factors and RNA polymerase to access the genes needed to be transcribed. However, there appears to be even more players involved with this crucial biological mechanism.

The most recently emerging form of regulation, that appears to be fundamental to proper gene expression, is the spatial organization of target sequences within the nucleus [2-4]. The overall chromatin architecture is dynamic, and to identify 3-dimensional aspects of its organization over a given period of time, whether it is during the developmental or the cell cycle process, would allow for a more complete understanding of how these processes relate to gene expression. Discovering function within organization could also prove to be very informative and lead to even more discovery.

We know that regulatory elements can be quite distant and spread out with respect to the genes they regulate, so how then do these elements coordinate in a manner to facilitate gene regulation? In fact, recent evidence has shown that distal elements are capable of forming direct

physical contacts with their target genes [5-7]. Therefore, it is very likely that the 3-dimensional spatial arrangement of genes and their control elements by way of direct spatial clustering and the subsequent formation of chromatin loops allows for proper regulation to endure [8, 9]). Interactions involving trans-acting (inter-chromosomal) regulation between genes and their regulatory elements have been observed as well [10, 11], further adding to this already dazzling mode of regulation. Furthermore, these long-range interactions have been found in a variety of organisms and suggest that this mechanism of gene regulation is of a conserved nature.

The evidence for this mechanism is vast. Take the beta-globin locus for example. The beta-globin locus contains 5 developmentally regulated genes in human that encode variants of the beta-chain for the all important hemoglobin protein. The regulation of all of these genes is carried out by a single control region, known as the Locus Control Region (LCR), which is located 25kb upstream of the most proximal ε-globin gene. The LCR is capable of regulating genes up to 80kb away and it was originally believed that this was carried out by the interaction of protein factors bound at the gene and LCR that came together [12] and was later confirmed with the advancement of technological tools that were capable of detecting these types of interactions [5]. Importantly, the control of globin genes by the LCR was also found to be developmentally regulated; the LCR allows for gamma-globin expression during fetal development and beta-globin expression during the adult stages [13]. This confirmed for the first time that long-range control and communication between distant genes and control elements, but there are a number of other examples that have been explored. One gene cluster in particular is now emerging as an exceptional example of these properties, the *Hox* cluster.

The Hox genes are a family of evolutionary conserved genes that are expressed during embryonic development in a highly coordinated manner and remain transcribed in almost all adult tissues. The mammalian *Hox* genes are organized into 4 distinct clusters segregated onto separate chromosomes, with letter distinctions from A to D. Each cluster is structured in a highly organized manner with genes arranged and numbered along the chromosome in ascending order from 3' to 5'. Homologous genes from each cluster are given the same number distinction.

The most important feature of the *Hox* genes is the mechanism by which they are regulated. The collinear mode of gene regulation is best described as the spatio-temporal regulation of transcription determined by the linear sequence of genes in the cluster, where the more 3' located genes are expressed earlier and more anteriorly than the 5' genes. A number of regulatory elements have been identified within the clusters and some even located outside of the immediate clusters, however the exact way in which these elements regulate the collinear mode of action has not been determined. That being said, there does exist some findings that support the collinear mechanism. For instance, evidence of chromosomal looping within *Hox* clusters and its relation to their regulation has been observed in drosophila and mouse[14]. These looping conformations are believed to be directly responsible for coordinating the spatial and temporal expression of the Hox genes and these structures are themselves thought to be formed through the interaction of specific cis-regulatory elements.

1.2. Hox Genes

What are Hox genes?

Hox genes are members of the Homeobox containing gene family that were first discovered in 1983 while examining the genome of the fruit fly, *Drosophila melanogaster* [15]. All Homeobox containing genes encode a transcription factor that is involved in the regulation of patterns of development in animals, fungi and plants. Homeobox genes are defined by the presence of a characteristic 183 bp DNA sequence known as the homeobox, which codes for the relatively conserved 61 amino acid section of the protein, known as the homeodomain [16]. The homeodomain is capable of binding to DNA in a sequence specific manner based on its secondary structure, which consists of three very well defined alpha helices and a flexible fourth helix. Helices 2 and 3 form the commonly observed helix-turn-helix motif which enables helix 3 to interact directly with the major groove of DNA and also allows for the loop preceding the helix-turn-helix to interact with the DNA backbone. The flexible amino terminus of this domain is also capable of interacting with the minor groove of the DNA to further increase the stability of the protein-DNA interaction at gene specific promoters. Through this interaction, Hox proteins are capable of effectively recruiting other transcription factors and transcription initiators to enable the expression of their targets.

Evolution and conservation of Hox

The 39 mammalian *Hox* genes are believed to have arisen by multiple duplication events beginning with genes that were most closely related to the 3' end coding sequences of the now

sequenced clusters. A set of protoHox1/2 and protoHox3 genes present in early metazoan are believed to be the original forms of *Hox* genes which duplicated to generate *Hox* predecessors, from which allowed for further duplications to occur[17] and the eventual formation of clusters.

The original ancestral cluster most likely contained five total genes. These genes are the ancestors of the drosophila genes labial (lb), proboscipedia (pb), Deformed (Dfd), Antennapedia (Antp), and Abdominal B (Abd-B). The mammalian counterparts to these genes are paralogue groups 1, 2, 4, 6-8, and 9, respectively. Eventually in mammals, the *Hox* genes formed into 4 homologous clusters now classified as Hox A-D, with paralogues being numbered in descending order of transcription, with Hox1 paralogues located at the 3' end of the clusters (figure 1).

Why these duplication events may have occurred is still somewhat unresolved. However, it is believed that an adaptive model could explain these events, including the conservation between *Hox* genes. This model proposes that gene sequences responsible for vital organismal functions are kept constrained with respect to change. Developmental genes, such as the *Hox* genes, may then duplicate and the newly duplicated genes may undergo unconstrained change until they also become vital to the developmental process. Evolution of developmental genes is therefore fueled by gene duplication and the *Hox* clusters may very well have arisen for this reason.

The paralogous genes share a high degree of sequence similarity, expression pattern and are found to be partially redundant in function [18, 19]. HoxA1 substitution by HoxB1 via knockdown methods in a mouse model showed that these two genes do in fact have identical function [18]. The evidence showing overlapping function of Hox genes in mammals is particularly strong for the paralogous groups. In fact, the encoded homeodomains are nearly identical within each group, differing by zero to six amino acids at most. For instance, the human

HOXD4 (Dfd homolog) or the mouse HoxB6 (Antp homolog) expressed under the control of a heat-shock promoter in transgenic flies lacking either of these homologous genes resulted in phenotypes that suggested the vertebrate proteins function similarly to their fly counterparts [20, 21]. However some Hox genes have acquired specificity; *Hox* A and D clusters in particular have been implicated in uniquely regulating mammalian limb growth and formation [22].

Hox proteins have similar in vitro DNA target-binding specificities. This characteristic partly explains why loss of function of two or more adjacent, paralogous expressed *Hox* genes results in increasingly devastating phenotypes compared with the single mutants, even revealing phenotypes that were not evident after a single mutation of any of the genes involved [23, 24]. In fact, single loss of function mutation of a paralogous group mostly results in a hypomorphic phenotype. These observations suggest functional redundancy among the *Hox* genes, where the inactivation of one gene can be compensated by the activity of the other *Hox* genes that are slightly similar with respect to sequence and expression domains.

Cluster organization

Hox genes are found in 4 different types of clusters: type-O (organized), type-D (disorganized), type-S (Split) and type-A (Atomized)[25]. Vertebrate *Hox* genes are usually found in type-O (or Organized) clusters and are defined by their tight (~100kb) cluster size with genes arranged in sequential fashion without interspersed non-*Hox* genes. This type of cluster has been identified in chicken, zebrafish, xenopus, and newt to name a few organisms [26]. Other cluster designations, such as type-D, have been identified in the sea urchin [25]. The type-D cluster is larger in size compared to type-O clusters and contains genes in opposite

transcriptional orientations and unexpected locations with respect to their paralogous groups [27]. The deuterostome sea urchin *S. purpuratus* genome contains almost all of the *Hox* homologs, but in a single scrambled cluster [28]. Type-S clusters, such as the quintessential example in Drosophila, are characterized by their chromosomal breakpoint that splits the cluster into segments. The Antennapedia (ANT-C) and Bithorax (BX-C) complexes are found at distinct loci and are often mistakenly considered to be a single cluster [29]. Furthermore, the ANT-C complex has interspersed genes within it, resulting in even less clustering. Type-A clusters, such as those found in Oikopleura, are the most lacking in 'cluster 'characteristics. Genes in atomized clusters are usually found to be by themselves; however some do remain in pairs [17].

Organisms that contain the three latter types of clusters have generally derived a mode of embryogenesis and lineage-dependant mechanisms for the determination of cell fate which do not rely on highly organized gene clusters such as in the type-O cluster [30]. These developmental strategies make temporal collinearity completely unnecessary [22] because they eliminate the need for delaying the caudal expansion and the posterior restriction of posterior Hox genes [31]. For example, C. elegans do not need the Hox genes contained within the same cluster because they display a completely un-segmented body plan. In fact, the 'clustering' of the Hox genes is an absolute requirement for their temporal collinearity. In contrast, observations of the spatial regulation over Hox genes have shown that highly organized clusters are not entirely necessary. Type-A clusters, in which genes are the most disorganized, have still been observed to follow, to some extent, spatial collinear distribution [32]. In addition, when single Hox genes from type-O clusters were introduced randomly in transgenic mice in vivo, they could recapitulate part of their spatial expression patterns. Ultimately, this indicates that cluster

organization is dispensable for establishing some of the expected rostral-to-caudal expression boundaries, at least within a certain spatial window.

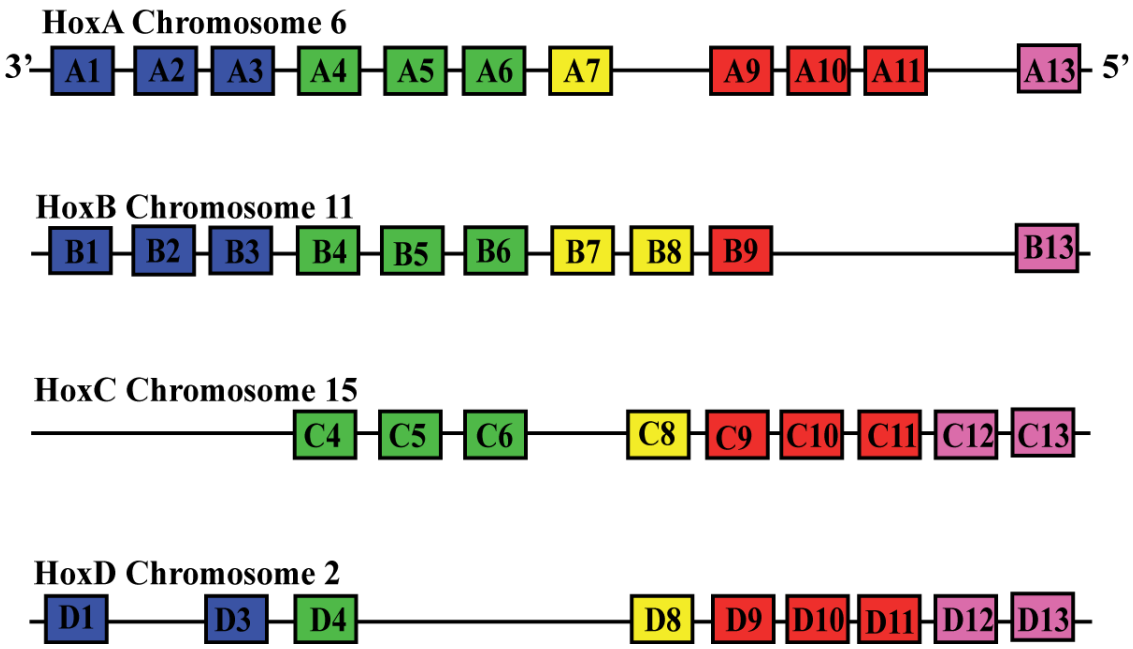


Figure 1. Mouse *Hox* Gene Clusters

The organization of the mouse *Hox* clusters. The 39 *Hox* genes are divided among 4 different clusters denoted as HoxA-D. *Hox* genes are shown as colored boxes in their respective order on the chromosome in which they are present. Paralogous mouse genes are shown color-coded.

Hox gene function

The *Hox* genes encode a family of transcription factors, which are essential for embryonic body patterning during development. The main function of the *Hox* genes is their role in regulating the anterior-posterior (A-P) body axis and the morphological features along it [33-35]. These genes have also been implicated in regulating the growth of more specific anatomical features such as limb and genitalia [22, 36]. Importantly, paralogous *Hox* genes also have similar or redundant functions as well as overlapping expression domains during development. Each paralogous group shares a color in Figure 1, which identifies which genes are associated to each group based on the segmentation pattern of the developed mammal, specifically, at which rhombomeres they are expressed.

A variety of *Hox* downstream targets have been identified, however, the majority of the genes and target sequences involved are still unknown. The identification of downstream effectors is hampered largely by the complex network of *Hox* gene targets together with the short and degenerate DNA sites where the Hox proteins bind [33]. Early experiments involving yeast one-hybrid system assays to identify regulatory elements that mediated a *Hox* response showed very limited success [37]. The unfortunate results of these early experiments are believed to be due to the absence of additionally required DNA-binding proteins that normally interact with Hox transcription factors [38-40]. In contrast, some targets of Hox function have been indentified using in vivo enhancer trap techniques[33]. For example, the Distal-less (Dll) gene, another homeobox transcription factor and developmental regulator was discovered in this manner[41]. More recent efforts have involved the implementation of chromatin immunoprecipitation (ChIP) techniques that have identified the specific regulatory elements

within Hox regulated genes such as scarbrous (sca), Transcript 48 (T48) and centrosomin (cnn) [42-44].

Interestingly, a number of Hox targets are the *Hox* genes themselves. Deformed (Dfd) maintains its own expression in the maxillary and mandibular segments through the interaction of an autoregulatory element [45]. Antp is regulated by three different *Hox* genes, including itself [46]. Antp autoregulates itself in abdominal neuronal cells and is prevented from this regulation by the competitive interaction at the same enhancer site by Ubx and Abd-A.

Large scale analysis of downstream targets has answered a few questions concerning *Hox* function that were not previously answered using the techniques mentioned above. It was previously believed that *Hox* genes only affected regulatory genes, especially transcription factors, however, microarray analysis of *Hox* targets has identified a number of realisator terminal differentiation genes [47]. The identification of a large number of unique downstream targets from this technique led to another key finding; the regulation over most *Hox* targets only occurred at one of two developmental stages, embryological and adult. This crucial fact of *Hox* function created the need to determine how cellular context affected *Hox* mechanism and led to the understanding that Hox proteins gain the ability to regulate their targets through the interaction with known cell- and/or tissue-specific transcription factors [48]. Of these cofactors, the first identified from drosophila genetic screens was *extradenticle* (exd) [49] and preceded the identification of its mammalian homolog, Pbx1. It has been shown that both of these proteins are fully capable of physically interacting with Hox proteins and in doing so, increase their binding affinity and specificity to a large extent (reviewed in [50]). Whereas Pbx1 is largely responsible for partnering paralogue groups 1-10 [51], the Meis1 protein can dimerize with groups 11-13 [52] and functions similarly to Pbx1.

In general, the Hox genes function as both transcriptional activators and repressors. For instance, Pbx/Hox dimers can induce the transcription of sonic hedgehog during limb development [53] while Exd/Ubx is found to repress transcription of distalless (dll) [54]. Other than direct binding to DNA promoter elements, Hox proteins have also been observed to bind to CREB binding protein (CBP) and therefore modulate histone acetyltransferase activity, which makes the effect on target genes even more difficult to elucidate.

Hox and development

The development of the mammalian vertebral column, a key structure in differentiating the A-P axis, is derived from mesodermal structures termed somites. The somites themselves form in a sequential manner according to their exit from the gastrulating embryo. This timing event, otherwise known as the segmentation clock [55], is directly related to the position of specific somites along the A-P axis and must be directed in a manner that allows for proper spatial and temporal development. The eventual differentiation of somites into specific vertebrae is thus dependant on their axial position in the developing embryo and indicates the importance of timing during somite development. This entire process is largely guided by the unique combination of specifically expressed *Hox* genes in each developing somite[56].

Mutations in *Hox* clusters, either partial duplication or deletions, can result in homeotic transformations in which body segments or structures are observed to develop in irregular locations. Important alterations in morphologies due to improperly expressed *Hox* genes are a consequence of deregulated control over their expression and confirm that the orchestration of these genes must be very tightly controlled. The level of organization of the clusters, specifically

their linear arrangement along the chromosome, is what allows these genes to be properly expressed with respect to time and space. This will be elaborated upon in the next section.

In contrast to differentiation, some studies have examined the role of *Hox* in cellular proliferation [57]. When cell division begins, proteins must be assembled onto replication origins to establish the initiation of DNA synthesis. The selection and activation of these replication origins is the key process in controlling chromosome replication during cell proliferation. It has been shown that HOXC10 binds to a 74-bp sequence within the human DNA replication origin associated with the Lamin B2 gene in Cos7 cells [57] and is also highly expressed in HeLa S3cells [58] and monocytic U937 cells, which all grow indefinitely. Further evidence establishing HOXC10's role in proliferation is the fact that it is degraded early in mitosis [59], indicating that it is no longer necessary to mark the origin. Together, these results indicate that HOXC10 is expressed not only during differentiation, as expected for a homeoprotein, but also in response to proliferative stimuli.

The exact role of HOXC10 during proliferation remains unclear. However, it is known that transcription factors may participate in regulating replication origin firing. In fact, actively transcribed regions of the genome are replicated early in S-phase, suggesting a connection between transcription and DNA replication pathways, although the most likely effect that TFs provide for replication involves the altering of chromatin into a configuration that allows for the assembly of replication complexes at the origin [60]. Chromatin remodeling at origins of replication through the binding of specific TFs has been seen in the activation of *Saccharomyces cerevisiae* origins as well [61]. Therefore, it can be inferred that HOXC10 binding to the origin area does contribute to establish a chromatin structure suited for the different functions of this particular region of the genome, including origin activation and/or transcriptional activation [57].

Collinearity

As stated above, the vertebrate *Hox* genes are arranged in a cluster such that their ordered sequence reflects not only their timing of expression during embryogenesis but also their expression domains with regard to the A-P axis. The term "collinear expression" , when used to describe *Hox* regulation, refers to simultaneous control over spatial and temporal expression patterns and was first explained by Lewis in 1978 [62] while examining the genetics of the fruit fly Bithorax homeotic gene complex. The concept was eventually found to play a significant role in vertebrate *Hox* gene regulation[63]. In general, the most 3' genes (A1, A2, A3, etc.) from each cluster are expressed earlier and more anterior than the 5' located genes (A13, A11 etc.)[64, 65].

In terms of its developmental implication, temporal collinearity of *Hox* gene expression is first observed in the posterior streak region at the boundary between embryonic and extra embryonic tissues in mouse [66]. For each gene, expression spreads anteriorly to reach the anterior primitive streak, or node region, where paraxial and axial mesoderm are formed. This progressive mode of expression has been described as anterior propagation, forward spreading or rostral expansion [63, 65, 66] and is the direct result of the collinear mechanism.

Collinear Regulation

Many studies have examined the types of regulation governing these genes and suggest a variety of possibilities. Chromatin remodeling, short and long-range promoters and tissue specific regulators [65, 67] are all viable options to consider and may in fact act in conjunction with one another to completely fulfill the requirement of the collinear mechanism.

With regards to regulation of *Hox* expression via signaling events, a few key signaling processes such as FGF, WNT, and all-trans retinoic acid (RA) have been examined.
FGF is required to initiate mesoderm formation through the primitive streak and cause the regression of this developmental feature; therefore it is believed that FGF could also initiate *Hox* gene expression [68]. WNT signaling may also affect the progression of anterior *Hox* genes based on the observations that it is also a regulator of primitive streak formation [69]. Hox regulation through FGF, WNT or RA is thought to be propagated through cdx, a transcription factor that is often found upregulated in the presence of these signals. RA is capable of activating the RARE located in the promoter region of cdx1 and causes subsequent transcriptional activation of the TF, which can then further activate other homeobox family members [70]. The WNT and FGF pathways also show a strong correlation to cdx activation. For instance, Wnt- 3a expression overlaps with that of Cdx1 in the caudal embryo, and is involved in specification of the posterior embryo [71]. Also, in Xenopus, it was found that overexpression of bFGF can induce anterior expression of both the Xenopus homologue of Cdx4 and certain Xenopus Hox genes normally restricted to the posterior embryo [72]. However, of these factors, the most compelling evidence indicates that RA induction of *Hox* genes is the most likely to promote the collinear mechanism of action.

RA has been shown to elicit an anterior shift of 32 *Hox* genes expression domains in somites [56]. More specifically, RA is capable of inducing early 3' transcription followed by later 5' expression during development [73, 74]. Retinoic acid response elements (RAREs) located in various genomic regions are the primary candidates, which are responsible for enabling the RA induction of the *Hox* genes. These elements have been found to be present upstream (*HoxA1* and *HoxB1*), downstream (*HoxA4, HoxB4*, and *HoxD4*), or involved with the

regulation of multiple gene groups (*HoxB4, HoxB5, HoxB6*) of *Hox* clusters. The most highly characterized RAREs are involved with the regulation of *HoxA1, HoxA5, HoxB1, HoxB4*, and *HoxD4* [74-80].

Several in vitro studies have shown that RA is capable of binding to two families of transcription factor heterodimers and directly regulate gene expression [81]. These studies identified RA as a high affinity ligand for RA receptors (RAR-α, RAR-β, and RAR-γ) and a low-affinity ligand for their heterodimers partners, the retinoid X receptors (RXR-α, RXR-β, and RXR-γ). When RA binds to the RAR partner of the RAR/RXR heterodimers that are bound to a regulatory DNA element, it stimulates a cascade of events resulting in the displacement of transcriptional co-repressors and subsequent recruitment of co-activators that induce transcription [82].

Because of the various developmental roles that RA signaling plays, numerous birth defects have been associated with its deregulation. Acute promyelocytic leukaemia (APL) is caused by fusion of retinoic acid receptor-α (RARα) with promyelocytic leukaemia (PML) and promyelocytic leukaemia zinc finger (PLZF). These aberrant proteins bind to RAREs and recruit HDACs, histone modifying enzymes and transcriptional activators. This binding has high affinity and prevents the response to physiological concentrations of retinoids, which induce the normal differentiation and development of myeloid cells [83, 84]. Therefore, as an outcome of these mutations, Hox genes can themselves be direct targets of deregulation.

Although RA may initiate the collinear events associated with Hox activation, the true nature of this mechanism relies on the specific layout of the clusters. Human genome sequencing has revealed that the clusters contain a very low density of Alu sequences, an interspersed repeat, suggesting that specific cis-regulatory elements incorporated in the cluster are evolutionarily

intolerable to interrupting insertions [85]. Kmita and Duboule have determined that these features are associated with the progressive activation of the *Hox* genes within their clusters. For example, the relocation of *HoxD9/LacZ* and *HoxD11/LacZ* transgenes to the 5' extremity of the cluster, just after *HoxD13*, delayed the expression of their expression making them follow a pattern of expression that closely resembled the displaced *HoxD13* gene[67]. Not only does this show how crucial the organizational characteristic of the *Hox* clusters are, but it may also indicate that a repressive mechanism is directly affected by the positional characteristic of each *Hox* gene within the cluster.

Limb Development and Hox Regulation

The development of the vertebrate limb has also been found to be under *Hox* control and examination of its developmental process has aided in discovering more features of the collinear mode of action and Hox regulation [86]. The HoxD genes are initially activated from the 3' end towards the 5' end in a time specific manner. However, this homogeneous expression is only observed up to HoxD9. The late 5' genes are restricted to progressively more posterior limb cells and are believed to be under the regulation of a second player in this particular collinear mechanism.

First, the location of a region rich in enhancer sequences located 5' to the *HoxD* genes [87] has been shown to effect the expression of the centromeric end of the cluster, due to distance and sequence specific properties that create a preference for *HoxD13* activation, with sequentially less activation for the more telomeric genes[86]. Identified by Kondo et al. by way

of multiple insertions at the 5' end of the HoxD cluster as well as sequential deletions within the cluster, the enhancer rich region was found to be essential for early collinear regulation [88].

The digit enhancer, as it is commonly called because of its role in late limb development, is also capable of controlling the expression of two additional genes, Evx2 and Lunapark (Lnp), two genes, which are located just outside of the HoxD cluster. It is embedded within a genomic area that contains a number of conserved neural enhancers that do not regulate HoxD genes. The set of digit and neural enhancers are known as the Global Control Region (GCR).

nc*RNA control over Hox*

Various theories regarding the intricately complex control over global *Hox* regulation initiation and maintenance have been proposed in an attempt to explain this biological mechanism. One such theory describes the transcription of neighboring *Hox* genes in the developing vertebrate body being progressively activated as a result of concomitant chromatin modifications [67]. In general, repressors and/or silencing factors are released in a 3' to 5' manner thus alleviating the inactivation of the *Hox* genes. The RA signaling machinery explained above may possibly be involved at this level of control and appears to rely heavily on another major area of Hox research, epigenetics. A more detailed review of these mechanisms will be discussed in later sections.

Further examination of *Hox* regulation has revealed that a family of non-coding RNAs (ncRNA) can participate in post-translational regulation. Approximately 22nt ncRNAs, known as miRNAs, have been shown to direct the cleavage of specific *Hox* coding transcripts, usually by

direct binding to the non-coding sequences 3'UTR [89]. *HoxA, HoxB* and *HoxC* clusters all encode specific conserved miRNAs between genes 9 and 10 that are capable of binding to and functionally eliminating the mRNAs of HoxB8, HoxC8, HoxD7 and HoxA7, which themselves are responsible for regulating the expression of other Hox genes in certain situations [90].

This form of regulation has also been proposed to function in the collinear mechanism of Hox expression. For instance, miR-181 is expressed later than its target, HoxA11, and therefore could function by limiting the expression domain of A11. This effect is known as posterior prevalence and can be described as the dominance that the more posterior genes exert over the anterior genes with regards to phenotype [91]. This concept directly explains why loss of function mutations of *Hox* genes results in a posterior structure transformation.

Prior to 2007, only cis-acting ncRNAs had been implicated in regulating *Hox* genes, but now the identification and characterization of trans-acting ncRNAs has come to fruition. Rinn et al identified and characterized hundreds of ncRNAs from the human *Hox* clusters in 11 fibroblast cell lines that were isolated from distinct positions along the A-P and proximal-distal axes in order to determine the differential expression of ncRNAs in different cell types. Many of these transcripts have previously been identified and found to regulate the expression of their neighboring genes [92, 93]; however, an identified transcript from Rinn et al seems to regulate an entirely different *Hox* gene cluster in Trans. The 2158kb HOTAIR transcript expressed in the HoxC cluster can silence transcription in the HoxD locus. This was verified by HOTAIR knock down and the observation that repression of a 40kb region of HoxD is relieved in this state.

HOTAIR's association with the polycomb repressive complex 2 (PRC2; described below, seems to validate these findings in terms of establishing a repressive mechanism. The interaction

between ncRNAs and the PRC2 further proves that a major mechanism of Hox regulation is in fact epigenetic.

Hox and disease

A major factor contributing to disease states involving Hox genes is the creation of fusion genes. These genetic products can cause the deregulation and overexpression of *Hox* genes and result in such developmental defects like acute myeloid leukaemia (AML). For example, when the mixed lineage leukaemia gene (MLL), MLL1, is fused with a *Hox* regulator, aggressive leukemias in children and adults are commonly observed. Transcription of *HoxA7*, 9 and 10 from the UBX (A7) and ABD (A9 and A10) group of *Hox* genes, is coordinately activated in hematopoietic stem cells and overexpression of these genes has been documented in AML with the understanding that MLL maintains their transcription after differentiation [94]. *HoxB3* and *4* mRNA levels have also been observed to increase in AML bone marrow tissue [95], but the most significant increase in expression levels has been confirmed by many groups to be from *HoxA9* [95, 96].

Normally, MLL is able to interact with the HoxA promoters and cause the addition of an activating acetylation mark on H3 and H4 [97]. In the cancerous cell, the fusion of MLL with other genes results in the loss of the acetylating and activating SET domain, and although this would theoretically cause the down-regulation of *Hox* genes, it is not the case [94]. Some other, not yet fully described, mechanisms are at work.

MLL translocations have revealed ~40 fusion partners of which most contain the N-terminus of MLL (non-SET containing end) with the C-terminus of the partner. When

investigators examined the role of the MLL-ENL fusion product on *Hox* expression in a mouse model, they found an increase in transcription for *HoxA4-11,* with *A9* showing the highest degree of increase compared to controls [98]. They also concluded that in the presence of MLL-ENL but when *HoxA9* and *A7* are deleted, there is no observed leukaemia result[99]. The induction of MLL–ENL also caused an increase in H3K79me levels on HoxA9 and Meis1 promoters [100], indicating that increased transcriptional activity of these genes is most likely directed through epigenetic modifications.

Reciprocal chromosomal translocations involving individual *Hox* genes that result in the creation of fusion products can also cause enhanced cellular differentiation of hematopoietic stem cells. *HoxA9* and *HoxD13* genes have been found fused to the *NUP98* nucleoporin gene in some cases of acute myeloid leukaemia (AML) [101, 102]. Nup98-HoxA9 fusion protein includes the *HoxA9* DNA binding homeodomain fused with the N-terminal *Nup98* sequence and has recently been shown to upregulate a number of *Hox* genes in both the A and B clusters which in turn results in an AML disease progression [103].

Taken together, it is of no surprise that *HoxA9* is the single most highly correlated gene (out of 6817) for poor prognosis in AML [96]. In fact, *HoxA9* and the important Hox regulator, *MEIS1,* are commonly co-expressed in myeloid cell lines and in samples from AML patients [104]. In murine models of leukaemia caused by co-overexpression of a *Hox* gene with a TALE family member cofactor like*Meis1* or *Pbx1*, it has shown that specific complexes are required to initiate and maintain the leukaemia state. In this model, the collaborating cofactor acts as an accelerator of the leukemic process and the *HOX* gene defines the identity of the leukemia [105].

HOX gene deregulation has also been documented in lymphoid leukemias. Gene expression analysis showed that the whole *HoxA* gene cluster was dramatically deregulated in T-cell acute lymphocytic leukemia samples harboring the *TCRβ-HOXA* rearrangement [106]. *HOXA* genes were also found to be upregulated in *MLL* and *CALM-AF10*-related T-cell acute lymphoblastic leukemia's cases, strongly suggesting that *HOXA* genes are oncogenic in these leukemias as well [107].

In addition to the well-documented involvement of Hox genes in leukaemia, they have also been observed to play a significant role in breast cancer. Neoplastic growth in mammary epithelial cells occurs with the increase in expression of human growth hormone (hGH). hGH has itself been found to cause an increase in the expression and transcriptional activity of *HoxA1*[108] and overexpression resulting from this association in mammary carcinomas causes an up-regulation of Bcl-2, an anti-apoptotic factor, and therefore an increase in total cell numbers. The overexpression of *HoxA1* can also enhance the anchorage-independent cell proliferation capabilities of these mammary cells and therefore initiates their oncogenic transformation [109].

Other interesting *Hox* related diseases have to do with resultant anatomical malformations. Mutations in *HoxD13* have been found in humans affected by synpolydactyly, an inherited human abnormality of the hands and feet [110], which results in the duplication of fingers (polydactyly) and webbing between them (syndactyly). Affected individuals show an expansion of an alanine stretch within the coding sequence of *HoxD13* [110] that results in 7, 8 or 10 additional residues being expressed. This effects exon 1 of the gene product and sufficiently disables the protein from its natural function. Brachydactyly, another type of

malformation, involves the shortening of the digits as a result of a conversion mutation of Ile314 to a Leu within the HoxD13 homeodomain.

A third disorder that affects the developing limb is known as hypodactyly, in which a loss of digits is observed. The loss of digital arch formation in animals that have been examined is the result of a deletion within exon 1 of *HoxA13*[111]. The deletion causes a transformational shift that completely abolishes *HoxA13* function. This disorder, originally found in animal models, has been linked to humans and is known as hand-foot-genital syndrome[112]. There have been many reported causes for hand-foot-genital syndrome, from nonsense mutations in exon 2 that cause a truncation [112] to expansions of polyalanine regions similar to those found in *HoxD13*[113].

Many 3' Hox genes from the A and B clusters are normally expressed in fetal lungs [114] and their expression is naturally reduced as development progresses. Some of these genes, like *HoxA5*, continue to be expressed throughout development and are believed to be required for lung maturation [115]. When these genes are abnormally expressed, resulting malformations can occur. For instance, when *HoxB5* is found expressed at high levels beyond the early stages of lung development, the lungs remain as they were during fetal development resulting in disorders such as bronchopulmonary sequestration [116] and congenital cystic adenomatoid malformation[117]. Even abnormalities that do not pertain to malformation can result from abnormal expression of many *Hox* genes. Such disorders include emphysema, primary pulmonary hypertension and lung carcinomas [118, 119].

There are many types of disease that are due to improperly expressed *Hox* genes, and many of the mechanisms normally controlling these genes are found to be either non-functional or

deregulated themselves. A major branch of control over *Hox* genes is the subject of our next section and has been pointed out in this section.

1.3. Chromatin and Epigenetics

Chromatin structure

The packaging of DNA begins with stretches of 146bp segments wrapped around octomeric complexes of histone proteins two times to form the fundamental component of chromatin, the nucleosome core particle. The identified major histone proteins H1, H2A, H2B, H3 and H4 are all involved in chromatin formation. The nucleosome octomer is composed of a central (H3-H4) tetramer that is flanked by 2 H2A-H2B dimers [120]. The histone protein H1 and its homologous protein H5 are involved in higher-order chromatin structure. The linkage between each nucleosome is approximately 50bp, and when observed through electron microscopy, resembles “beads on a string”. This fiber is referred to as the 10nm chromatin fiber and further compaction of the nucleosomes forms the 30nm fiber. This 30nm fiber, otherwise known as euchromatin, is highly dynamic. It is capable of unwinding to lower order structures to allow for RNA pol II binding to occur and transcription to initiate or it may condense further to form transcriptionally silent heterochromatin [121].

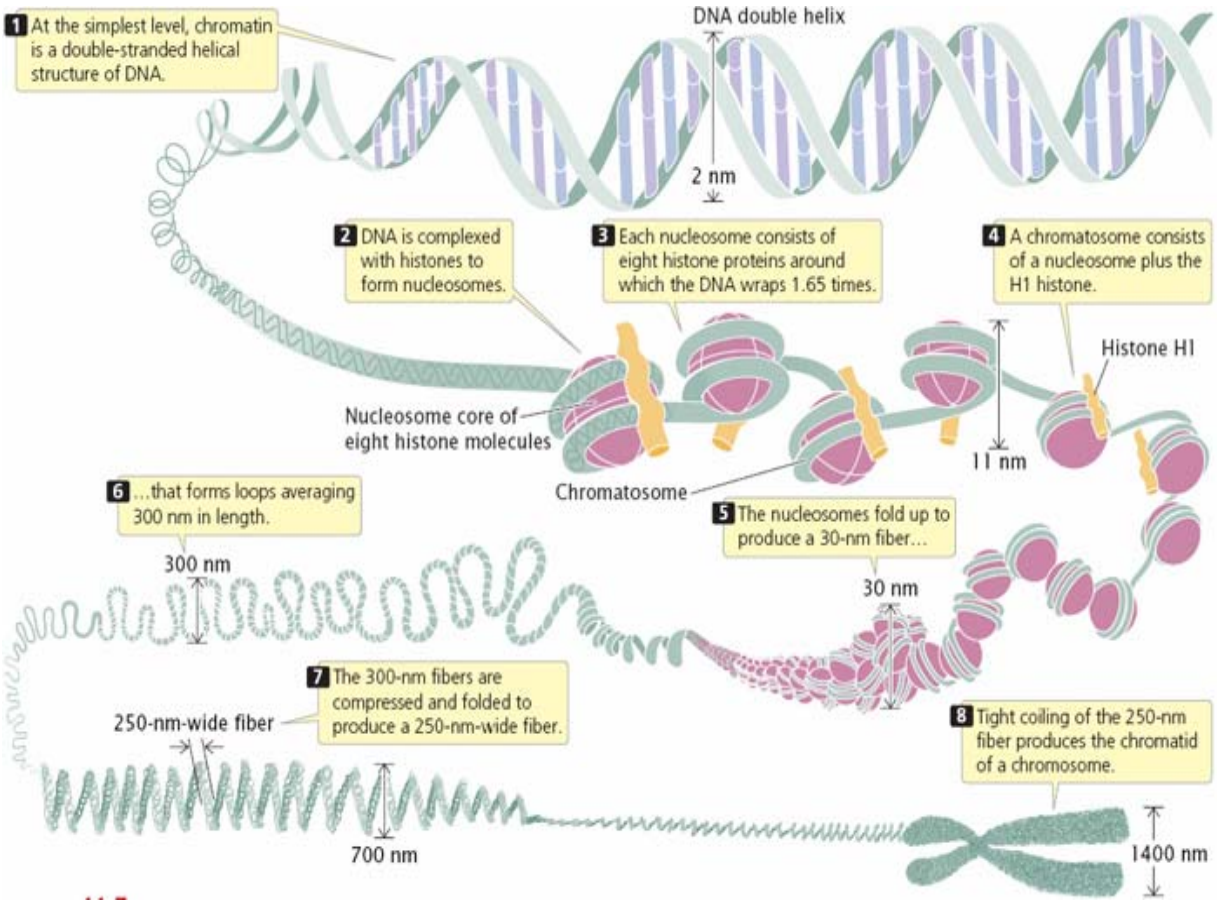


Figure 2 Chromatin levels of Condensation.

Beginning with (1) the double-stranded helical structure of DNA, the condensation begins through (2) the complexing to histone proteins to form (3) the nucleosome. The inclusion of Histone H1 forms (4) the chromosome. The nucleosomes structures then fold up to from (5) the 30-nm fiber which is then subsequently folded into (6) 300nm loops that themselves fold into (7) the 250nm fiber. The coiling of this dense 250nm fiber results in (8) the final chromatid of the chromosome.

Histone H3

There exist a number of histone variants that have each been implicated in unique cellular functions. For instance, CENP-A is a highly conserved histone H3-like variant that is found at centromeric locations in mammals [122]. It has been established that the loss of this histone variant is embryonic lethal[123]. CENP-A is basically required to mark the centromere for kinetochore assembly, a step involved in the coordination of the separation of sister chromatids during mitosis [124].

Another H3 variant, H3.3, is constitutively expressed during the cell cycle and is found associated with genes that are poised for transcription as well as genes that are actively being transcribed. It is believed that this variant may serve to replace H3, for a brief time, at active genes as nucleosomes reform behind the RNA polymerase [125].

Histone H2A

Histone macroH2A is a divergent variant of H2A and has been implicated in the processing and maintenance of inactive X chromosomes in female mammals [126]. Interestingly, this is mediated by the interaction of a non-coding RNA (ncRNA) known as Xist, which we will discuss shortly. MacroH2A is also found at other chromosomal locations and has been implicated in playing a more general role in gene silencing[127]. Another H2A variant, H2A-Bbd, is 42% identical to H2A and has been associated with transcriptionally active regions of the genome.

Finally, the core histone variant H2A.Z is found in a range of organisms and has been shown to be essential for viability in a number of them [128, 129]. This particular variant

localizes to the transcriptionally active macronucleus of Tetrahymena [130] and although found throughout drosophila and yeast genomes, it is depleted from silenced regions and elevated in the intergenic regions of inducible genes [131]. H2A.Z is usually situated at the promoter region and enables the genes to be rapidly induced, however, once transcription has occurred, these histones are lost from the promoters [131]. In contrast, H2A.Z has also been found to associate to the pericentric heterochromatin in early mouse embryos, suggesting that this variant could play an important role in heterochromatin formation and function during the development process [132].

Histone H1

As for the linker histone H1, a number of variants have been found in comparison to the core histones [133]. As mentioned above, these proteins are involved in stabilizing the higher order chromatin structure and may function as repressors of transcription by limiting the access of transcriptional activators to the chromatin[134]. For instance, the well known H1 variant, H5, is involved with global transcriptional repression once deposited during the terminal stages of erythrocyte differentiation in chickens [135]. The mouse has at least 8 H1 variants that all play specific roles in regulating the transcription of tissue specific genes. For example, H1t is specifically required for the maintenance of an open chromatin structure in spermatocytes and spermatids [136].

Imprinting

The imprinting of genes involves the attainment of an extremely tight chromatin conformation through DNA hypermethylation in one allele of a gene early in the germline that results in monoallelic gene expression. Approximately 90 genes are believed to be imprinted in

humans, and many of these genes are found in a clustered genomic region[137]. Imprinting control regions (ICRs) are known regulatory elements that are absolutely required for imprinting to occur. Deletion of these regions results in the loss of imprinting abilities [138].

The H19/Igf2 cluster is a very well characterized example of what is called the insulator model of genetic imprinting. The imprinting center 1 (IC1), this clusters ICR, is positioned between the two genes. This location enables it to regulate the interaction between the two gene promoters as well as their shared enhancer located downstream of H19[138]. The imprinting of these genes depends on the methylation of the IC1 on the paternal allele, thereby keeping the maternal one in an unmethylated state[139].

There are also a number of identified conserved sequences at the IC1 and other ICR loci that bind to the insulator protein CCCTC-binding factor (CTCF) when the ICR is unmethylated [140]. Interestingly, this binding factor was originally found to play a similar role in the beta-globin locus, which we will discuss in detail shortly [141]. In terms of the mechanism employed by CTCF in the H19/Igf2 example, the protein actually prevents *de novo* methylation at the maternal allele thereby also preventing downstream enhancers from activating Igf2. This of course allows for the transcription factors to then affect H19 instead. In contrast, when looking at the methylated IC1 in the paternal allele, the opposite expression pattern is observed.

The H19/Igf2 imprinting mechanism has also led to the suggestion that chromosomal looping is involved. Some groups strongly believe that there is a direct interaction between the shared enhancers and the Igf2 promoter on the paternal chromosome; however, there seems to be less agreement when examining the interactions on the maternal chromosome. For instance, it is believed that a tight loop forms around the Igf2 gene causing a silencing effect that is mediated

by the ICR contacting a matrix attachment region as well as a differentially methylated region of the gene locus [142]. Another theory resides in experiments in which the ICR was observed to form an interaction with the enhancers that resulted in Igf2 silencing [143], but this mechanism does not explain why H19 will be maternally expressed. Still, other results indicate that interactions between the enhancer is seen with coding and promoter regions of H19 up until the ICR [144]. All of these potential mechanisms for the involvement of chromatin loops were discovered through chromosome conformation capture techniques, which we will detail in the upcoming sections.

Another well-characterized example of imprinting is the X-chromosome inactivation involved in gene dosing compensation. The complete silencing of one of the X chromosomes in females is required to attain the accepted level of gene expression for genes located on this chromosome. The mechanism involved relies on the assistance of ncRNAs such as Tsix and Xist [145]. The Xist ncRNA is observed to coat the chosen X chromosome very early in development, which leads to chromosome wide gene silencing. Xist can associate with another important epigenetic factor; a protein from the Polycomb family, EED. This protein recruits HDACs [146] and is therefore believed to aid in the initiation of X inactivation [147].

Other than H3K27 modifications, other H3 modifications are also very common in X chromosome inactivation. H3K4 dimethylation is decreased on the Xi of female somatic cells, but is present at high concentrations in male meiotic cells[148]. In contrast, increased H3K9 methylation normally associated with heterochromatin formation is observed throughout the Xi [149]. These observations reveal that distinct epigenetic patterns for X chromosomes of female somatic cells and male germ line cells exist. Patterns are indeed what truly makes epigenetics

unique and the units that make up these patterns and there combinations along specific regions of chromatin will now be discussed.

Epigenetics

Epigenetics, the transmission of genetic data apart from the actual DNA sequence, from one cell to its daughters as well as from one organism to its offspring is currently one of the most interesting topics in the scientific community. Embryonic development and epigenetics, at first glance, appear to be counterproductive and unlikely to cooperate in regard to organismal growth. Development relies upon fast, dynamic processes that require very frequent changes in gene expression, while epigenetics, either up or down regulates transcription over an extended period of time. However, these two biological mechanisms do operate synergistically in various complex ways to attain incredible biological feats.

The earliest and most abundant epigenetic modification found in mammals is the methylation of cytosine residues within the dinucleotide CpG. The pattern and content of methylation is not only species specific, but also tissue specific. In humans it has been determined through high performance liquid chromatography (HPLC), that 0.75-1% of the genome contains 5-methylcytosine bases [150]. Large regions of CpG dinucleotides, denoted as islands [151], exist throughout the genome and can frequently be found at the 5' end of many genes. These promoter located CpG islands remain unmethylated for genes that are regularly transcribed and allow for an open chromatin configuration to persist with association of hyper-acetylated histones.

More recent findings have led to the understanding that higher order chromatin structure is defined by a set of epigenetic histone modifications such as methylation, phosphorylation and ubiquination[120] in conjunction and cooperation with DNA methylation[152]. In fact, the histone code hypothesis states that the specific combination of histone modifications form a language that specifies the structural state of chromatin [153]. These types of posttranslational histone modifications are "read" by effector proteins that are capable of carrying out the code's instructions, which results in either tightly packed chromatin, known as heterochromatin, or loosely packed euchromatin. The overall effect of the code is the correct regulatory control over the genome and therefore individual cell identity.

Methylation

Methylation of histone protein tails is generally associated with transcriptional repression, but there are of course situations where methylation results in active transcription. The methyl group donation from AdoMet is catalyzed by lysine methyltransferases (HKMTs) and protein arginine methyltransferases (PRMTs) on specific amino acid residues [154].

Residues from histones H3 and H4 are the most commonly found methylated substrates. The major sites of Lys-methylation being K4, K9, K27, K36, K79 from H3 and K20 from H4[155]. Mono-, di- and trimethylation marks can be found at these residues and add to the functional diversity at each site. For example, dimethylation at H3K4 occurs at both inactive and active genes, whereas trimethylation is exclusive to active genes [156]. Furthermore, condensed heterochromatin is enriched in trimethylation of H3K9, K3K27, and H4K20 [157] and silencing

of euchromatin loci caused by histone deacetylation involves the recruitment of specific K9 histone methyltransferases (HMTs).

Histone Methylating Enzymes

With the exception of Dot1/Dot1L , all HKMTs have the evolutionally conserved SET (SU(VAR)3-9, enhancer-of-Zeste, Trithorax) domain that is responsible for lysine methylation activity[158].

Many of these enzymes have specific cellular function. For example, the histone methyl transferase (HMT) enzyme Su(var)3-9 from drosophila, as well as its homolog clr4 in yeast, are involved with transcriptional silencing and heterochromatin formation [159]. The mammalian homolog, Suv39H1, directly affects the methylation status of K9 in histone H3 [160] and when this is disrupted, the HP1 protein, a chromatin-binding nuclear protein that localizes to heterochromatic regions in *Drosophila* and higher eukaryotes, is incapable of binding and there heterochromatin formation is hindered [161]. This function is intimately tied to histone methylation and chromatin remodelling and involves the cooperation of HMTases and histone deacetylases (HDACs).

PcG (ex. E(Z)) and trxG (ex.TRX) proteins, originally identified in drosophila, are two of the major epigenetic modifiers of *Hox* gene regulation. The PcG, or polycomb group proteins, are capable of negatively affecting *Hox* gene expression by efficiently and effectively methylating its H3K27 target [162], whereas the trxG, or trithorax group proteins, are

responsible for coordinating the methylation of H3K4 and therefore activating gene expression[163].

The PcG proteins come together to form multiprotein complexes such as the polycomb repressive complex 1 and 2 (PRC1/2). The repressive mechanism of *Hox* regulation begins with the trimethylation of target lysines by the PRC2 [160]. This marking then allows for the recruitment of PRC1 to promote the condensation of chromatin through such epigenetic modifications as ubiquitination [164, 165].

Markings present at *Hox* loci are not exclusively found at their promoters but are observed to be distributed along ~100kb domains within all four clusters [166]. The exploration into the relationship between chromatin modification and *Hox* gene activation was examined by chromatin immunoprecipitation of the *HoxD4* locus in mouse ES and P19 pluripotent cells. The identification of activating epigenetic marks began at the 3' end of the gene and progressed towards the 5' end when cells were induced with RA [167]. It should be noted that specific histone modifications are not always predictive of gene activity and other requirements may need to be met as well. For instance, the recruitment of transcription factors such as CREB binding factor and RNA PolII to the Hox genes was also observed to follow a 3' to 5' recruitment pattern. In addition, it is believed that the removal of the repressive methyl modifications added by the PRC1/2 complexes is also likely involved with the activation process. We will discuss the demethylation of Hox genes shortly.

The coordination of the two SET containing histone methylases of the PcG and TRX groups is crucial for partitioning the genome into transcriptionally active and repressive states. They work in coordination to maintain the spatial pattern of *Hox* gene expression after they are established in early embryonic development [168, 169].

When examined in a broader spectrum, methylation status can reveal other informative details. In embryonic stem (ES) cells, lineage specific genes that are either repressed or transcribed at low levels are denoted as "bivalent," bearing marks of active (H3K4me, H3K9ac) and repressive (H3K27me3) chromatin modifications[170]. These observations have led to the hypothesis that dual markings enable developmental genes to be repressed in ES cells while remaining poised for quick activation. Bickmore et al. examined the H3K4 methylation markings at three time points in ES cells treated with retinoic acid[171]. They observed that the H3K4 methyl modification increased at the *HoxB1* locus after 4 days of RA treatment and subsided by day 10. Interestingly, the same epigenetic marking was observed at *HoxB9* at the same time; however this gene was not expressed until day 6 of RA treatment indicating that additional players are indeed required for controlling *Hox* expression.

Bickmore's group also investigated the potential "opening" of the chromatin structure by applying FISH to ES cell nuclei, prior to and after RA induction. After a few days of RA treatment, the FISH signals from *HoxB1* and *HoxB9* were distant enough to be resolved[171]. This suggests that chromatin decondensation, through epigenetic modifications, is fundamental to the process of *Hox* gene regulation and may be the initial step in the collinear mechanism. Further analysis of this model revealed that sequential looping out of the two examined *HoxB* genes from their chromosomal territories (CT) followed the decondensation of the chromatin. This two-step model fits very nicely into previously proposed mechanisms, but requires more corroborating evidence in order to substantiate the hypothesis. More importantly, it remains to be established as to which steps are involved in controlling which loops exit their CTs and which do not with respect to time.

The Dot1/Dot1L non-SET domain containing proteins have a slightly different preference for lysine methylation. They are primarily found to methylate residue K79 that is located in the core or globular domain of H3 as opposed to the more frequently modified histone tails [172, 173]. This particular mark is observed in about 90% of all H3 proteins in yeast and corresponds to the amount of euchromatin in its genome [173]. Therefore, Dot1 appears to play the role of anti-silencing agent. It is believed that the mechanism in yeast involves the ability of the K79me mark to repel the normal silencing machinery, the Sir Protein complex, from the genetic loci [173]. This marking is also observed in eukaryotes, however it has not been established whether it plays a similar role but it has been reported that the DNA repair enzyme 53BP1 can recognize this particular modification [174] indicating that DNA repair mechanisms also rely on epigenetic modification.

Arginine methylation of histones is largely associated with gene activation. Enzymes such as CARM1 and PRMT1 are part of the transcriptional activation machinery of nuclear hormone regulated genes [175] mainly because the modification is easily reversed to allow for the resetting of these highly inducible genes. For this reversal to occur, enzymes such as PAD4 carry out the complete deamination of the Arg residues instead of simply removing the methyl marking [176]. However, the mechanism of the conversion of the citrulline product of this reaction back to an arginine residue is still unknown.

Until recently the discovery of enzymes capable of removing methyl modifications has remained elusive. In 2004, histone demethylases were first described [177], and were shown to have the opposite effect on transcription when compared to their counterparts, the HMTs. The histone demethylase LSD1, originally identified as part of the repressor complex Co-REST, is

responsible for H3K4 demethylation, which leads to transcriptional inactivation. LSD1 can convert the mono- and di- methylated H3 forms to an unmethylated state, but it cannot enzymatically remove trimethylation markings. In contrast to these findings, when LSD1 forms a complex with androgen receptors, it demethylates H3K9 and activates transcription [178]. Other histone demethylases, such as jumonji (JHDM2A), are responsible for H3K9 demethylation [157, 179], whereas JHDM1 has the ability to convert active chromatin marks such as H3K36me2 to a repressive unmodified state [180].

With regards to *Hox* regulation, UTX and JMJD3 appear to play critical roles during the differentiation of stem cells as well as in animal development [181, 182]. These proteins contain a JmjC peptide sequence that has been shown to have catalytic activity in other observed histone demethylases, like those previously mentioned enzymes. ChIP experiments have revealed that JMJD3 and UTX are in fact directly associated with a number of *Hox* gene transcriptional start sites, and their presence correlates with the transcriptional activity of the promoters [181]. When these genes are knocked down using shRNA, an elevation of the H3K27me3 at *Hox* genes was observed along with transcriptional repression. It is believed that these proteins require additional support to initiate transcription. In fact, interactions with MLL2/3 (TrxG) have been observed [182, 183], thus allowing for a possible two-fold mechanism that allows for a rapid response of the target genes. H3K27me3 modifications can be found spanning large areas in the *Hox* clusters, and it remains unknown whether UTX, which is located specifically at promoters, is capable of affecting these widely distributed marks.

Acetylation

Histone acetylation is an epigenetic modification that is highly associated with an open chromatin conformation and the higher order folding properties of chromatin. These modifications are mediated by histone acetyl transferases (HATs). The acetyl group not only neutralizes the charge on histones and increases chromatin accessibility, but also acts as a signal for the binding of trans-acting factors by providing a suitable binding surface [184] similar to that observed for some of the methylation markings mentioned in the previous section. Which of these two features is more important in the regulation of gene expression has yet to be determined.

The main superfamilies of HATs are GNAT (Gcn5-related N-Acetyl Transferase) and the MYST families. Of these, the most highly studied HAT is the yeast Gcn5, which plays a central role in the acetylation of histones H3 and H2B [185]. It has been implicated in global acetylation and the creation of a more accessible chromatin environment [186], which allows for the formation of the pre-initiation-complex and a subsequent increase in global transcription. Homologous proteins to Gcn5 have been found in many organisms [187] including human and mouse.

The recombinant p300/CBP HAT protein can acetylate all 4 core histone tails. Interestingly, it is involved with the regulation of transcription related proteins such as TFIIE, p53 and HMGA1 [188]. p300/CREB-binding protein-associated factor (PCAF) can also acetylate histones and transcription factors and has been observed to play an important role in such processes as myogeneis [189], nuclear-receptor-mediated activation [190] and growth-factor-signaled activation [191].

Many HATs are also involved in the DNA repair machinery. Hat1 in particular was shown to aid in the repair of DNA double stranded breaks by being directly recruited to sites of DNA damage [192]. Another HAT that partakes in the DNA repair mechanism is Tip60, which contains several putative functional domains such as a chromodomain, a zinc-finger-like domain and a HAT domain that acetylates nucleosomal histones during DNA repair.

Members of the MYST family of HATs have also been found to play specific cellular roles. For example, the MOZ HAT is directly involved in the oncogenic transformation that leads to leukemia. It is usually found fused to certain transcription factors such as CBP [102], which leads to high acetyltransferase function. We will discuss this situation further in the upcoming section “epigenetics and disease”.

In direct functional contrast to HATs, a group of proteins exists that are capable of removing the acetylation markings. The histone deacetylases (HDACs) in humans have a very high degree of conservation and are subdivided into two main classes, I and II, based on their homology to the yeast proteins. In higher eukaryotes, HDACs 1, 2, 3 and 8 belong to class I and are similar to the yeast Rpd3 HDAC. Class II HDACs 4, 5, 6, 7, 9 and 10 are related to the Hda1 yeast homolog. HDACs 6 and 10 are rather unique since they contain two catalytic domains [193]. Proteins from both classes can catalyze the removal of the acetyl group with the aid of a zinc cofactor , and are normally found in complexes such as NuRD and Sin3 in order to allow for DNA binding [194]. Class III HDACs, also known as Sirtuins, are quite different from the other classes. They are related to the yeast Sir2/Hst family and do not primarily use histones as substrates [195].

Phosphorylation, ubiquitination and SUMOylation

Histone phosphorylation is a less examined form of histone modification when compared to methylation and acetylation. Nonetheless, it has been shown to be responsible in the mediation of such cellular events like transcription, DNA repair and chromosome condensation [196]. The phosphoryl group is generally added to the histone tails of H3 at serine and threonine residues. The Rsk-2 catalyzed phosphorylation of serine 10 of H3 is associated with active transcription and has been observed to increase drastically in quiescent fibroblast cells once treated with EGF [197]. Similar to acetylation, the phosphate groups neutralize the basic charges residing on the histone tails, thereby reducing their affinity for DNA and allowing for an opening of the structure.

In terms of chromosome condensation, phosphorylation of H2A and H3 at serine 10 seems to be pivotal for mitotic condensation [196]. Mutation of this residue in *Tetrahymena* causes abnormal chromosomal condensation and improper separation during anaphase. When the enzymes coordinating the addition of the phosphate, such as yeast Ipl1 and *Aspergillus* NIMA, are disrupted through mutation or deletion, the normal process of chromosome condensation and segregation is subsequently also disrupted [198].

Phosphorylation has also been linked to the activation of DNA repair mechanisms. For instance, the phosphorylation of histones H2A as well as the mammalian variant H2A.X at Ser139 by PI3K-like family kinases is observed to increase upon exposure to DNA-damaging agents [199].

The addition of ubiquitin to histone tails is another known element of the epigenetic code. The 76 amino acid ubiquitin protein is usually found associated with proteins that are destined for destruction through the proteosome pathway; however, the addition of this modification can also be a signal for a specific function. The addition of this substrate is carried out by three separate enzymes, beginning with the activation of ubiquitin by the ubiquitin-activating enzyme E1, which transfers a phosphate to it from ATP. This is followed by the conjugation to the ubiquitin conjugating enzyme E2 and the subsequent transfer to the target lysine residue by a ubiquitin-protein isopeptide ligase E3[200].

In mammals, histone H2A and H2B are found mono-ubiquinated at residues K119 and K120 respectively [201]. A number of mammalian enzymes have been found that may carry out these functions. For example, Mdm2, a well-characterized negative regulator of p53, has been shown to interact with H2A and H2B histones and cause increased ubiquitination when it is overexpressed, albeit not at residues K120 of H2B [202]. Rad6 or its homolog Ubc2, can monoubiquitinate K123 of H2B, and in many species is associated within a complex which enables it to carry out this specific function [203]. It also appears that Rad6 is recruited to promoters through the interaction between activators like the PAF complex and then associates with RNA PolII as transcription begins [201]. Rad6 associates with other proteins when it is required to work in the DNA repair mechanism, its usual process.

The main function of this type of epigenetic marking seems to be its role in coordinating the crosstalk between itself and other histone modifications, including H3 methylation, that are themselves involved with transcriptional activation. H2B ubiquitination specifically affects di- and tri-methylation of H3K4 and H3K79 in which the absence of the ubiquitin does not allow for

the methylation of the histone [204]. The monoubiquitination of human H2A is mediated by Ring1B, an enzyme that associates with the polycomb repressive complex 1, itself a verified *Hox* regulator. In fact PRC1 has been shown to contain E3 ubiquitin ligase activity specific to H2A [165], indicating the strong connection between different histone modifiers.

Ubiquitination is a reversible reaction, and the enzymes capable of carrying out the deubiquitination, such as Ubp8 and Ubp10, do so by ubiquitin-specific protease action. Ubp8 is a component of the SAGA complex, which itself contains acetyltransferase activity[205], further establishing the crosstalk between modifications. Interestingly, Ubp10 can function without the aid of the SAGA complex and appears to act on a distinct pool of ubH2B, unlike Ubp8 [206]. Instead of assisting in transcriptional activation like Ubp8, Ubp10 is found interacting with silencing protein Sir4 and participates in Sir-mediated telomeric and rDNA silencing [201].

In conjunction with ubiquitin, SUMOylation, the addition of a small ubiquitin-related modifier (SUMO), is another less commonly observed histone marking. Budding yeast have all four core histones SUMOylated in a manner that depends genetically on intact sumoylation machinery. Although the modification is found at very low levels when compared to other modifications, it is present in multiple modified forms, which suggests that multi-site modifications or poly-SUMO chains are formed and likely have distinct functions. This modification is not distributed equally on all the histone proteins. For instance, the H2A.Z histone variant, which is normally associated with active gene transcription, is found less modified by SUMO. Mass spectrometry analysis of protease digests identified lysines 6 and 7 in H2B and lysine 126 in H2A as sites of modification. Mutagenesis studies also implicated K16/17 in H2B as well as all five lysines in the N-terminal region of H4 [207].

All of these characteristics of SUMOylation suggest that it can occur at multiple lysine residues without strict sequence requirements. The relative overlap of identified sites for SUMO and acetyl addition indicates that the two are counter-regulatory, with SUMOylation causing repression and acetylation increasing transcription. This brings us to an important element of epigenetics: crosstalk.

Epigenetic crosstalk

There appears to be three cross regulatory mechanisms involved in histone modifications relating to gene transcription. Firstly, initial histone modifications sometimes increase the activity of other histone modifying enzymes. For instance, phosphorylation of H3S10 in yeast promotes the acetylation of H3K14 by the Gcn5 acetyltransferase [208] and mammalian acetylation of H3K18 and H3K23 promotes the methylation of R17 also on this histone [209]. The second mechanism involves the coordination of multiple histone-modifying enzymes into the same protein complex. The mammalian MLL3/4 Set1-H3K4 methyltransferase can coordinate the removal of repressive methyl markings and the subsequent formation of activating methyl-histone complexes on H3 [210].

Of particular interest in explaining the second mechanism are the Polycomb group (PcG) complexes, as mentioned above, which contain a number of histone modifying enzymes. PRC1 contains E3 ubiquitin ligase activity that is specific for H2A ubiquitination and this modification is localized to Pc targets like the *Hox* genes [165]. The ubiquitination by PRC1 occurs downstream of H3K27 methylation and has been shown to inhibit MLL mediated di- and tri-

methylation of H3K4, thereby silencing gene expression [211]. PRC2, otherwise known as the E(Z)/ESC complex contains the E(Z) H3K27 methyltransferase and allows for the recruitment of PRC1 to the methylated substrate for further modification [212].

The third mechanism involves a connection between histone modification and the cleavage of the N-terminal tail of histones that provides a way to achieve irreversible histone modification. Although still quite new, some evidence points towards a possible function for this type of crosstalk. The differentiation of mouse ES cells led to the observation that forms of H3 were present in the chromatin architecture of the genome that lacked N-terminal tails [213]. These cleaved H3 peptides were found to have both activating and repressive marks in its N-terminus: methylation of K27 and acetylation of K18 increased cleavage, whereas K23 acetylation reduced it. This mechanism could allow for the release of polycomb proteins from H3K27 methylation thereby allowing for the derepression of genes involved in the differentiation pathway of ES cells.

Combinations of different epigenetic modifications are crucial for specifying distinct biological functions. For instance, H4K8 acetylation, H3K14 acetylation, and H3S10 phosphorylation are all often associated with transcription. Conversely, tri-methylation of H3K9 and the lack of H3 and H4 acetylation correlate with transcriptional repression in higher eukaryotes. Particular patterns of histone modifications also correlate with global chromatin dynamics, as di-acetylation of histone H4 at K4 and K12 is associated with histone deposition at S phase, and phosphorylation of histone H2A (at S1 and T119) and H3 (at T3, S10 and S28) appear to be hallmarks of condensed mitotic chromatin. There are already a great number of identified combinations of modifications and mechanisms of recognition that have been

discovered and examined; however, there may still be new combinations of histone modifications that have different unknown functions.

1.4 Epigenetics and disease

DNA methylation

When examining a cancer cell, there are a number of variables that can be looked at more closely to determine what caused this cellular state to be. A very dramatic transformation occurs in the cancerous cell in terms of its epigenetic profile.

Tumor suppressor genes are regularly found hypermethylated at their CpG island promoter regions, and are therefore rendered inactive. For example, the $p16^{INK4\alpha}$ gene is found hypermethylated in a variety of human cancers and cell lines [214]. Some tumors display hypermethylation at genetic loci pertaining to the DNA repair machinery. When MGMT (O6-methylguanine-DNA methyltransferase) is found in this state, the removal of groups at the O6 position of guanine is not possible and eventually may lead to K-ras and p53 mutations [215, 216]. When aberrant methylation exists at crucial hormonal receptor genes such as the progesterone receptor, breast cancer cells are left unresponsive to the steroid hormones that normally regulate cellular development [217], which further establishes the cancerous condition.

Interestingly, not only are hypermethylation of CpG islands observed at specific genetic loci in cancerous tissues but an increase in global hypomethylation is also found. This can lead to increased chromosomal instability thereby enabling the reactivation of transposable elements and also causing the loss of imprinting. In fact, when DNA methyltransferase 1 is knocked out in mouse ES cells, there is a significant increase in genetic deletions [218]. Similar outcomes are

also observed in patients with germline mutations in DNA methyltransferase 3b [219]. With regards to transposon reactivation, long interspersed nuclear elements (LINES) and Alu elements become transcribed and may then translocate after the methylation status of these regions drops [220], which can lead to further disruption of the genome. When imprinted genes such as those present at the H19/IGF-2 locus become hypomethylated, anti-apoptotic growth factors become severely overexpressed and result in certain childhood cancers [221].

A number of DNA methylation inhibitors have been developed, and they can be classified into nucleoside and non-nucleoside type compounds. Nucleoside analogs are metabolized by kinases that convert the nucleosides into nucleotides for incorporation into DNA and/or RNA[222]. 5-azacytidine (5-aza-CR) and 5-aza-20-deoxycytidine (5-aza-CdR decitabine) have long been known to contain activity that allows for the inhibition of DNA methylation [223]. It has been reported that inhibition of methylation induced by 5-aza-CR reactivates the expression of genes that have been repressed by DNA methylation [224, 225]. 5-Aza-CR has anti-proliferative activity against several types of cancer cells and is used for the clinical treatment of acute myeloid leukemia and myelodysplastic syndrome (MDS) [226]. FCDR (5-fluoro-20-deoxycytidine) and zebularine (1-beta-D-ribofluranosyl-2(1H)-pyrimidinone), two cytidine analogues, have been developed as new DNMT inhibitors that display an increased half-life and potency in comparison to aza-CR [227]. Zebularine is capable of causing the demethylation and subsequent reactivation of the hypermethylated and silenced p16 gene in human bladder tumour cells grown in nude mice [228]. Cytidine analogues, such as these, are mechanism based inhibitors even though they get incorporated into nucleic acids; they form a covalent link with the cysteine residue in the active site of DNMTs and fully abolish its activity.

Non-nucleoside DNA methylation inhibitors function by binding directly to the catalytic site of the DNMT enzyme, without being incorporated into DNA [229]. The small-molecule inhibitor, RG108, has been developed by Lyko and co-workers [229] to target the catalytic pocket of human DNMT1 and has been found to effectively inhibit DNMTs in vitro. Of more relevance is the identification of another small-molecule non-nucleoside DNA methylation inhibitor, Procainamide [230, 231]. This analog was approved by the FDA for the treatment of cardiac arrhythmias and was reported to display anti-proliferative activity against human cancer cells [230, 232]. The DNA-demethylating action of procainamide is thought to be mediated by its binding to GC-rich DNA sequences [232] and interfering with enzyme-substrate recognition. EGCG, a major polyphenol from green tea, can inhibit DNMT activity competitively and is capable of reactivating silenced genes such as p16INK4a and hMLH1 in several human cancer cell lines [233]. MG98 has also been reported to function as an antisense oligonucleotide of human DNMT1, which prevents the translation of DNMT1 mRNA [234].

Not only have tools been designed to treat these particular cellular situations, but new techniques and improvements in detection procedures implemented during diagnosis are also being cultivated. The fact that normal cells contain an unmethylated CpG island at genes that are susceptible to hypermethylation led to the understanding that these locations could be regarded as biomarkers for malignant transformation. This observation allowed for the development of methylation-specific PCR [235], which can be used to detect these features. This is extremely useful because the PCR detection of such a state cannot be masked by normal cell contamination and hypermethylation occurs early in tumor progression, which allows for all important early detection and therefore an early and more positive prognosis.

Histone acetylation

HATs in cancer can function either as tumour suppressors or as tumour activators depending on different tumour types and tumour development stages due to the fact that cancer itself is a very complex, dynamic process. A global loss of H4K16 acetylation has been linked to tumourigenesis, and in humans the MYST family HAT member, MOF, is responsible for the majority of genome-wide H4K16 acetylation [236]. Other HATs that are believed to be tumour suppressors are CBP and p300 [237], which show decreased expression during chemical hepatocarcinogenesis. Mutations in p300/CBP are associated with different human cancers and other human diseases such as colorectal, gastric, and epithelial carcinomas [238]. In addition, the transcriptional activity of the p53 tumour suppressor can be upregulated by HATs p300[239], PCAF[240] and TIP60 [241]. As already mentioned, the misregulation of HAT targeting and activity has been linked to leukemogenesis. For example, in acute myeloid leukemia (AML), the CBP gene is translocated and fused to either the Monocytic Leukemia Zinc finger (MOZ) gene, MORF [242], or to MLL [243].

The MLL-ENL translocation in particular is an important type of translocation regarding the deregulation of the *Hox* genes. When this specific translocation and fusion of the MLL gene occur, HoxA9 gene expression is left unchecked and overexpressed, often resulting in leukemia [244].

In terms of potential therapeutic agents in the defense against misregulated HATs, there are a number of viable candidates. HAT chemical inhibitors such as bisubstrate analogues made of a histone substrate peptide covalently linked to a CoA motif at the lysine site, are some of the first ever identified compounds capable of effective inhibition. Most of these compounds however are unable to permeate the cell and further work has been carried out to develop new

structures that have a higher cellular permeability [245]. Some small molecules have also been discovered as inhibitors of HAT activity, such as anacardic acid [246]. Anacardic acid works as weak non-specific inhibitors of p300/CBP and PCAF and is capable of easily permeating the cells in culture. Interestingly, CTPB, the amide derivative of anacardic acid, enhances HAT activity of p300 fourfold, but not that of PCAF [247]. Garcinol is another potent cell permeable inhibitor of histone acetyltransferases p300 and PCAF both in vitro and in vivo [248]. Curcumin (diferuloylmethane), a purified plant ingredient, was also reported to inhibit the HAT activity of p300/CBP but not that of PCAF [249]. Curcumin has been shown to have anticancer activity, but the molecular mechanisms by which it inhibits cell growth and induces apoptosis are not clearly understood [250].

Giannis' group developed MB-3, a small, cell-permeable inhibitor of the human GCN5 HAT, which showed only weak inhibition of CBP and GCN5 but does provide a lead structure from which more potent GCN5 inhibitors can be developed [251]. Cinnamoyl compounds are small molecule inhibitors of p300, and one of its derivatives (2c) was proven the most potent anti-p300 agent. Derivative 2c was shown to be active in mammalian cells and demonstrated the down-regulation of histone H3 acetylation [252].

There is much evidence that has been presented to implicate the role of Class I and Class II HDAC members in malignant transformations [253]. The aberrant recruitment of HDACs has been identified in leukemias as well as lymphomas. Mechanistically, when not functioning as they should, they contribute to the altered gene expression found in these diseased states [254] and the loss of acetylation at H4K16 is a common hallmark of human cancer[255] and further implicates these enzymes. For instance, silencing of the AML1 gene, a gene required for the differentiation of hematopoietic cells, which functions as both a transcriptional activator and

repressor, can lead to leukemic transformation when fused to ETO, itself a strongly interacting HDAC gene [256]. Thus, HDACs represent emerging targets in the treatment of cancer.

Hydroxamic acid HDAC inhibitors such as Trichostatin A (TSA), bind to the zinc ion in the catalytic domain of the enzyme, thereby preventing the deacetylation of histones [257]. TSA was initially isolated from *Streptomyces hygroscopicus* and was first shown to be a potent inducer of differentiation and cell-cycle arrest but was later reported to possess anti-HDAC activity [258]. TSA can act synergistically with the demethylating agent 5-aza-CdR described earlier in a mouse cancer model [259] to enhance the anti-tumour properties of both drugs. Suberoylanilide hydroxamic acid (SAHA) was originally designed for use as a differentiation agent, but it has been found to be active and safe in patients with solid and haematological malignancies due to aberrant HDAC activity, like those seen in patients with Hodgkin's disease [260].

Thus far, the exact mechanisms through which HDAC inhibitors mediate antitumor activity are still not quite clear. Different models have been proposed to explain the antitumor effects of HDAC inhibitors [222]. One possible mechanism is that the hyperacetylation of histones induced by HDAC inhibitors leads to genomic instability, which ultimately triggers the cell-cycle checkpoint [261] and therefore stops tumour progression.

As is the case for deregulated methylation status, innovative techniques for early detection of improper acetylation can also be developed. The identification of biomarkers is becoming essential for good prognosis in many cancers. One such technique that may eventually be used in such a way is the subject of our next section.

1.5 Chromosome Conformation Capture (3C)

In this section, we will begin by presenting the intricacies of 3C technology. This will of course be followed by various studies highlighting its use.

Through the work of various groups, we are now aware that higher order chromatin organization plays a vital role in genome function [262]. For example, the identification, via electron micrographs, of chromatin looping features with sizes of approximately 90kbp have been shown to interact with the nuclear matrix and aggregate into rosette type structures with around 18 loops during cellular mitosis [263]. Originally it was believed that the genomic distance between two genetic markers followed a random walk behavior [264], however, further evidence showed that this relationship is based on a power law [265], which led to the postulation of a random walk/giant loop model [266].

The random walk/giant loop model describes the configuration of chromosomes as forming 2 to 5 Mbp loops at the 30nm fiber level that still display a random walk relationship between genomic separation and genomic markers (refer to Figure 2). This proposal, although reasonable for large genomic distances, did not explain separations that are less then 4Mbp. To explain the relationship observed in these smaller distances, another model was proposed, the multi-loop-subcompartment model (MLS).

The MLS model depicts the 30nm fiber as a folded rosette consisting of many small loops that are connected by a variety of linker strands (refer to Figure 2). With regard to gene regulation, the compaction and looping of chromatin as well as the usual associated transcription elements like enhancers and insulators, influence expression over large genomic distances [267].

An important example of this is the mode of regulation for the human β-globin locus. The globin locus control region does in fact cause the looping out of the DNA between itself and its target in order to make contact and initiate transcription with the β-globin gene [5]. Similar events have been observed in the Th2 cytokine locus and the interferon gamma gene[6].

As mentioned above, chromatin condensation and its general organization in the nucleus are both crucial factors that regulate gene expression. Chromosomal territories and transcription factories are recently discovered features of DNA condensation and nuclear organization that reveal how chromatin may be regulated; however, to observe these features it is not possible to achieve a resolution that is indicative of small scale chromatin arrangements. The following technique is now emerging as an extremely powerful tool that is used to observe significantly small scale chromosomal configurations and changes.

Chromosome Conformation Capture (3C) is a PCR based technology that allows for high resolution detection of chromatin interactions based on the relative frequency of contacts between genetic loci *in vivo* (see Figure 2 A). In general, 3C involves the detection of head to head, 5' to 5' or 3' to 3', interactions between fragments of linear DNA sequences. The 3C technique can be used to determine overall chromosome architecture from intact cells and is very powerful at identifying long-range interactions between DNA fragments, such as looping interactions between genes and regulatory elements (Figure 2 C). It was originally developed and used to study the conformation of the yeast *Saccharomyces cerevisiae* chromosome III [268]. Since then its use has been implemented in the discovery and/or verification of long-range cis-interactions, such as that found in the human and mouse β-globin loci [5, 7, 13, 269].

The 3C protocol relies on formaldehyde fixation to covalently crosslink DNA and protein in order to create a snapshot of chromatin interactions in 3-dimensional space. Formaldehyde cross-linking functions by utilizing the amino and imino groups of proteins and DNA to form covalent bonds within and between said macromolecules.

Following fixation, the reaction proceeds with restriction enzyme digestion, which itself has specific requirements; the selection of the enzyme for this step will depend on the genomic area to be analyzed. For regions spanning 20Kb-500kb, it is necessary to use a six base cutter which create fragments of approximately 4Kb. EcoRI, HindIII and BamHI are all very good candidates for this procedure due to their high degree of enzymatic efficiency and have been shown to have no preference for cutting at open chromatin regions. It is crucial to have efficient digestion; at least 80% of the cross-linked DNA should be cut. The selection of enzyme to use is also dependent on the linear distribution of cut sites in the area of interest. The restriction cut sites should be evenly spaced and should not contain fragments larger then 15Kb or smaller then 500bp. Fragments not within this range result in skewed unacceptable interaction frequencies that are either too high or too low as a result of biased re-ligation. The fragments should also dissect relevant regulatory elements such as gene bodies, promoters, enhancers, insulators, etc. in a manner to designate each fragment a single element if possible.

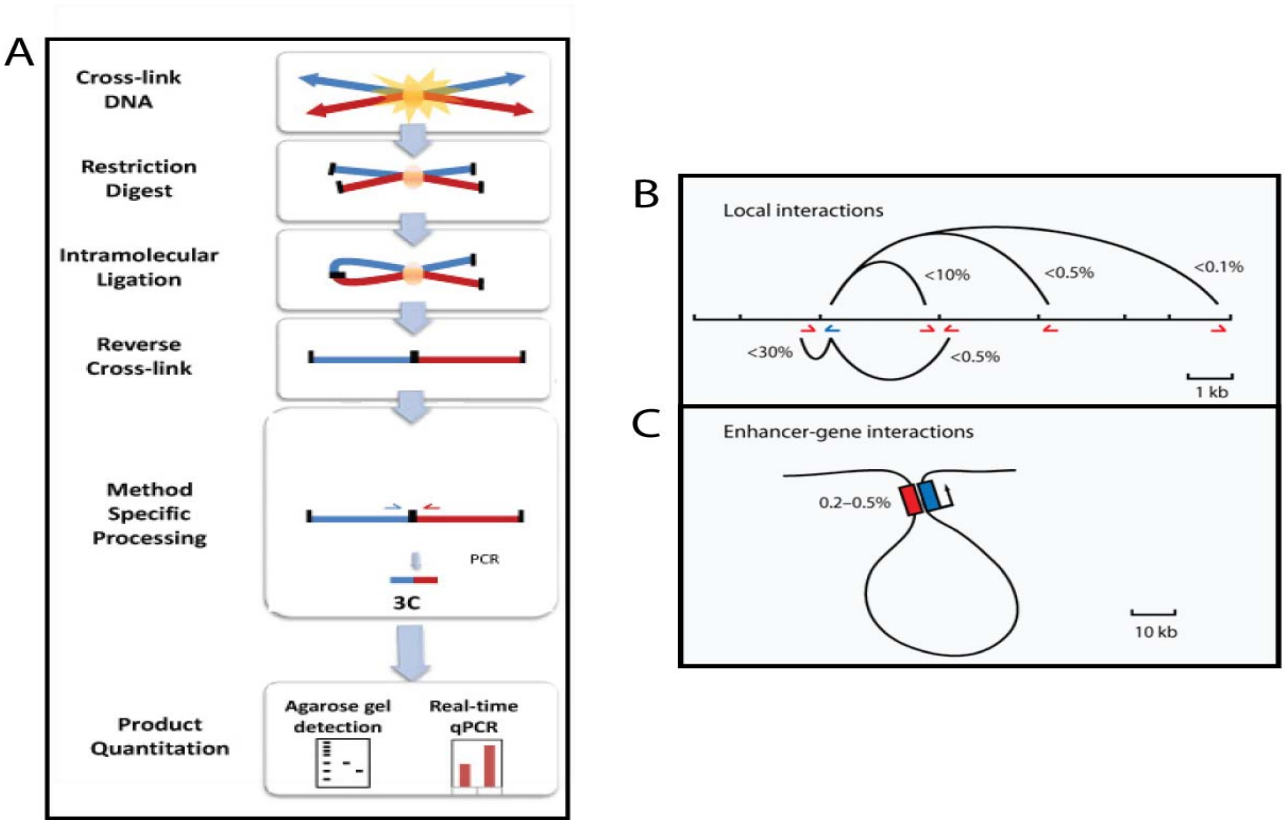


Figure 3. Chromosome Conformation Capture (3C).

A) A schematic representation of the Chromosome Conformation Capture technology. The process begins with formaldehyde fixation which results in DNA/Protein cross-linking, followed by restriction enzyme digestion. Digested material is then re-ligated and treated with proteinase to reverse the cross-linking. Samples are purified and then analyzed using specifically designed 3C primers and PCR followed by agarose gel detection. Optional real-time quantification is also possible. B) The relative percentages of observed local interactions for a fixed point 3C experiment based on genomic distance. Typical values for ligation frequencies (in % alleles) of a 'fixed' restriction fragment with a given variable restriction fragment are indicated. Fixed primer is in blue, variable primers are in red. C) The percentage of gene-enhancer interactions observed for distances of 30-100kb.

The digested/cross-linked chromatin is then ligated under dilute conditions that favor intramolecular ligation products between linked DNA regions. The dilute conditions are necessary to avoid ligation between fragments, which are not cross-linked. Of these types of artifacts, the first that is excessively ligated contains fragments that are linear neighbors. These may result from incomplete restriction digestion and can accumulate to comprise up to 30% of all the contacts [270]. This occurs simply because neighboring fragments are the most likely candidates to form interactions relative to their proximity within the nucleus. These products are still considered head to head interactions, also described as being re-ligated in opposite orientations; head to tail interactions, also described as being re-ligated in identical orientations or remaining as they were prior to digestion and re-ligation, are not detectable and become irrelevant when examining neighboring fragments. The second over represented product results from a fragment re-ligating to it. This circularization can occur for 5-10% of the products formed during this step and is an undetectable product in the 3C approach [270]. This artifact does however reduce the likelihood that more relevant contacts could have been formed and detected.

The proteins are then digested and the DNA library is purified. The resulting 3C library can then be used to carry out a set number of experiments to determine the chromosomal architecture of the region of interest.

The experimental cellular library must be used in conjunction with a control library generated form BAC DNA that encompasses the region of interest. The BAC 3C template is constructed in a similar fashion to the cellular library with some very minute alterations in the protocol. This library is mainly used to control for primer pair efficiency but can also be helpful in the testing of primer pairs prior to cellular library experiments. Interaction frequency, the unit

of measurement of 3C, is based on the signal intensity of the cellular library over the BAC control library.

A few specific requirements must be followed when selecting the BACs involved. First, the BACs should not overlap excessively. The ratio of BAC DNA to that of the DNA present in the cellular library in a case such as this would cause increased primer pair efficiency in the overlapping region thereby lowering the observed interaction frequency. Second, gaps between BAC DNA should be kept to a minimum in order to maximize the use of all potential fragments in the full region of interest. These two standards are sometimes difficult to meet, and if they are inevitable outcomes during design, then it must be understood when analyzing the acquired data.

Once complete, both libraries must be titrated using identical primer pairs, which we will discuss shortly, to evaluate the appropriate volume of each that will be used to carry out 3C analysis. This step enables the detection to remain semi-quantitative

Contacts between fragments are then detected by PCR. This step requires the design of specific primers that are dependent on the restriction enzyme selected for library construction. The primers are designed to be unidirectional at sites that correspond to approximately 100bp upstream of the restriction cut sites. The primers must be between 28-32bp, have approximately 50% GC content and contain a 3'GC clamp in order to allow for strong hybridization at the Taq DNA polymerase-binding region. These primers must be unique in the genome and therefore should be BLASTed and BLATed using NCBI and UCSC databases respectively. The primers are tested on BAC control templates to determine if they are specific (i.e. multiple PCR products) and if they produce the appropriate transcript size.

The PCR step requires that all potential contacts be measured in triplicate (i.e. 3 control and 3 experimental) in order to ensure the validity of observed interactions. All products are quantified using ethidium bromide UV detection on agarose gel and an image capture and analysis tool. The ratio of experimental cellular intensity to BAC control intensity is the calculated contact interaction frequency. Each primer pair is calculated for this interaction and plotted on a graph representing the distance between the two fragments/primers versus the interaction frequencies. Normally, fixed point experiments are done in which one fragment primer remains constant and the others vary. The observed pattern is examined for trends indicative of increasing and decreasing interactions. The data can also be plotted in a heat map format.

An alternative plot, termed a "compaction profile", can be used to determine the overall condensation state of the library by measuring interactions from a gene desert region. A normal compaction profile should show interactions decreasing significantly at approximately 20kb distances. If many interactions persist after 20kb, then the library is said to be condensed, which may be due to the cellular nature of the library or because of excessive cross-linking. In contrast, if the interactions are observed to be too low for short genomic distances, then the library was most likely not cross-linked effectively. This analysis also enables the experimenter to normalize experimental cellular libraries that display varying interaction frequencies between each other.

Importantly, when examining a potential gene-enhancer interaction using this technique, it can be concluded with a degree of certainty that a contact either exists or does not. For this purpose, the 3C experimental procedure is very effective. In contrast, the overall structural architecture of an examined region of interest by 3C is left in the hands of the experimenter to

interpret the data and construct a representative model. This task can be daunting and tedious; however with a step wise approach consisting of viewing the overall structure in segments or by examining the most highly interacting fragments followed by those that remain, the experimenter can indeed construct a suitable model. The visualization of such a complex structure consisting of a large array of fragments is also very difficult to represent and our group is currently developing computer software tools to try and image 3C data sets. With this type of tool and the capabilities of the 3C technique, large chromosomal regions of interest can be 3-dimensionally modeled to provide intuitive information regarding structure-function relationships.

3C and 3-D

As mentioned above, 3C was originally used to determine the 3-dimensional folding of yeast chromosome III. By mapping the interaction between the chromosomal telomeres, a somewhat circular chromosomal configuration was observed [268]. Since then the technique has been implemented in the detection of long-range looping interactions in several gene clusters as well as the identification of multiple interactions between enhancers and their target genes.

The alpha and beta-globin loci are the main model systems for long-range gene regulation. The beta-globin locus contains 5 genes in humans that encode variants of the beta-chain of hemoglobin. The expression of these genes is controlled by an element known as the locus control region (LCR). The LCR is located at an upstream distance of 25kb from the most proximal ε-globin gene. The LCR is approximately 20kb in size with a number of DNAse hypersensitivity sites that represent open chromatin regions where transcription factors such as GATA-1 can bind and initiate activation of the beta-globin genes [271]. As mentioned earlier, 3C analysis confirmed that the LCR interacts directly with its target genes [5] and further

analysis established that the interaction was developmentally controlled [13]. The factors mediating this interaction were subsequently identified though the use of cell systems containing specific transcription factor deletions. The 3C experiments in which transcription factors were removed revealed that GATA-1 and the cofactor FOG1 are required for the looping interaction to occur [7, 272]. Other proteins, such as Ikaros, have been identified this way as well. The Ikaros protein was shown to mediate the switch from LCR control over the gamma-globin gene to the adult beta-globin gene [272].

The alpha-globin locus, another cluster of genes involved with hemoglobin production, is regulated by a set of elements that have been identified at specific DNAse I hypersensitivity sites located 40-60kb upstream of the genes. Similar to the beta-globin locus, GATA1 and EKLF are required to mediate the direct interaction between the regulatory elements and their targets through long-range looping formations [273]. For both the alpha- and beta-globin locus it appears that large multiprotein complexes play crucial roles in the establishment of long-range looping interactions.

Another region that has also been observed to function via long-range looping interactions is the T helper type 2 (TH2) cytokine locus [6]. Although the layout of the locus may resemble that of beta-globin, with three genes being coordinated by a downstream LCR, the mechanism is slightly more complex with multiple elements associating with each other in a cell type specific manner. The TH2 LCR is located ~15kb upstream of the Il13 gene and ~60kb downstream of the Il5 gene [274], sandwiched between 2 of the 3 genes that it regulates (Il4 being the third) and embedded within the highly expressed RAD50 gene. 3C analysis revealed that the three gene promoters all interact with each other without regard for cell type. This

technique also showed that the LCR interacts with these promoters and that the interaction was strongest for cells of the TH2 lineage that regularly express these genes. 3C also identified STAT6 and GATA-3 as essential transcription factors in maintaining the LCR-cytokine gene interactions [6].

The TH2 3C data revealed many features of 3-D conformation and gene transcription. The fact that the LCR is always found associated with its target genes suggests that the interaction generates a poised state that remains capable of rapidly inducing gene transcription.

As mentioned in the description of the 3C protocol and capabilities, trans-interactions can also be observed with this technique. 3C analysis of mammalian X chromosome inactivation, which was also described earlier, has revealed that in order for the process to occur, the two inactivation centers present in female somatic cells interact to determine which chromosome will become randomly inactivated [275]. Although the precise mechanism is still unclear, it has been established that this interaction is necessary to ensure that there are in fact two X chromosomes in the cell and that one of them must be inactivated.

Another example of trans-interaction identified by 3C was the association of the TH2 RHS7 DNAseI hypersensitivity site from chromosome 11and the interferon gamma gene on chromosome 10 [276]. This interaction was specifically identified in cells that require the coordination of these two loci.

One final example of this form of interaction was found when examining the mechanism by which olfactory neurons select and express only one of a number of olfactory receptors (Ors)

[11]. 3C enabled the researchers to observe the *cis* and *trans* interactions of the H-enhancer with the chosen OR gene.

Of particular interest to our work is the 3C study of the *drosophila* homeotic genes and their regulation by PcG proteins. The Bithorax complex in *drosophila* contains the Ubx, AdbA and AbdB genes. It also contains elements that have been found to associate with polycomb proteins. These elements, Fab-7 in particular, have been observed to interact with one and other, even though they are dispersed throughout the genome. Clusters of polycomb response elements (PRE) are usually found in subnuclear bodies enriched in PcG proteins. In addition, the long-range interaction of Fab-7 elements enhances the repression of the target Abdominal B gene [277]. 3C confirmed the interaction between the AbdA promoter and all the PREs and boundary elements present within the 340kb region of BX-C. All of the other homeotic genes in this region also displayed this same interaction. This multiple loop system is believed to be required for the full repression of these genes.

Recently the 3C procedure was used to evaluate the interaction between the glucocorticoid receptor (GR) binding sites and distant promoters. The identification of a looping feature between the GR induced Lcn2 GR response element and the Ciz1 gene, located 30kb upstream, has led to the suggestion that this element is capable of regulating both genes in a cell specific manner [278].

3C has also been used in new and different ways to determine less commonly investigated topics. For instance, one group used replicating minichromosomes as probes to examine whether transcription occurs in factories [279]. They found that although a cell can contain thousands of minichromosomes, all of the minichromosomal RNA is concentrated in just

a few foci. Specifically, they determined that templates that contain similar promoters, such as the U2 promoter which they used, are more likely to be associated in local settings then those with different promoters, like the CMV with U2 elements as described in the experiment [279].

Additional tools have been developed based on this technology. Some that add new abilities and some that increase the throughput of the procedure.

1.6 Chromosome conformation capture carbon copy (5C)

Chromosome conformation capture carbon copy (5C) technology is another experimental tool that is capable of attaining high-resolution data pertaining to chromosome structure [269]. However this method of analysis is capable of collecting enough data to model a large region of interest after only a single experiment. The ability to multiplex primers as well as detect PCR products via microarray analysis are the main features that differentiate this novel technique with that of the original 3C.

The 5C protocol begins in the same way as that of 3C. Construction of a 3C library through fixation, digestion and re-ligation; however that is where the similarities end. Instead of carrying out the PCR step using fragment specific primers of just 2 fragments, ligation mediated amplification (LMA) is used [280].

Primers of approximately 50 nucleotides are designed to anneal at both ends of each fragment, instead of the 30-nucleotide primers that anneal 100bp upstream of fragment cut sites for 3C. 5C primers are constructed in a forward and reverse manner depending on which end of each fragment would like to be examined; normally consecutive fragments will have opposite primer orientation, but not always. These primers also contain overhanging tails with universal

T7 and T3 primer binding sites. As a rule of thumb, forward primers contain the T7 sequence at their 5' end and the reverse primers have the T3 sequence at their 3' end tails. These 5C primers will therefore anneal adjacently when two specific 3C fragments are found interacting in a 3C library. Due to the primer design in this experiment, only a forward and a reverse primer will be identifiable as a contact. The primers are then ligated using Taq ligase and every new 5C primer pair is amplified using the universal T7 and T3 primers to form the 5C library. A key feature of these steps is that thousands of 5C primers are used in a multiplexed setting, giving rise to a significant amount of possible contacts.

1.7 CNSs

A very small fraction of the human and mouse genomes are recognized as coding for functional products. Only about 2.2% of the human genome encodes mRNA [281]. This leaves an enormous amount of sequence that has not been allocated a specific function, and yet has been deemed necessary by the evolutionary process. Of this incredible amount of non-coding DNA reside distinct segments of conservation, conserved non-coding sequences (CNSs).

Chromatin attachment sites, miRNA genes or splice regulatory regions may be found in such highly conserved non-coding sequences [282, 283]. But these sequences most likely contain large amounts of non-coding segments which are primarily responsible for events such as the regulation of the complex mammalian genome during development [284]. The first large-scale study of CNSs compared the sequence similarity between human and mouse, and revealed that a significant degree of conservation was found in a 100kb region of the T-cell receptor gene family[285]. This finding forced the need to define these regions. Therefore, Loots et al.

described regions of at least 100bp in length and a sequence similarity of 70% or higher to be considered CNSs [286].

It has been well established that animal development is controlled by *cis*-regulatory DNA elements, such as enhancers and silencers [85, 287]. These elements are usually found clustered together to collectively function in giving genomic instructions for development and together form gene regulatory networks (GRNs). However, not much information has been elucidated about GRNs in vertebrates. Initial computer analysis identification of such sequences is not easy considering that current knowledge of their syntax or grammar is limited. In contrast, computational approaches for protein-coding exon identification have been established due to understanding of their characteristic sequence features, evolutionary conservation, and the availability of cDNAs and expressed sequence tags (ESTs). Some groups (i.e. ENCODE) are now working to annotate the entire genome in order to establish these networks.

The completion of a number of vertebrate genome sequences [85, 288-290] and the development of genomic alignment, visualization, and analytical bioinformatics tools [291], has made the comparison of large genomic regions possible. These achievements have therefore also allowed for the discovery of putative *cis*-regulatory elements. Comparing DNA sequences between different species provides a means of identifying common signatures that may contain possible functional significance.

When identifying CNSs, the choice of organisms for which the comparison will be made is extremely crucial. There have been a number of successful pair-wise and multiple-species sequence comparisons that have been carried out to identify novel enhancer elements in mammalian genomes [292-294]. An extremely useful approach that increases the resolving

power of comparative analysis is to use multi-species alignments combining both closely related and highly divergent organisms [295]. The large evolutionary distances compensate for the slowest-evolving neutral DNA, thereby significantly improving the signal to noise ratio in genomic alignments. Although CNS are less likely to be found among highly divergent species [295], there are a number of examples where comparison between human and pufferfish *(Fugu rubripes)* gene regions has identified highly conserved non-coding sequences that have been shown to have function *in vivo* [296-298]. The human- *Fugu* common ancestor existed around 450 million years ago [299], predating the emergence of the majority of all extant vertebrates. This implies that any CNSs between these two species are likely to be fundamental to vertebrate life. Not surprisingly, all of the reported findings of CNSs between these two species have been associated with genes that play critical roles in development. This suggests that some aspects of developmental regulation are common to all vertebrates and that whole-genome comparisons may be particularly powerful in identifying regulatory networks of this kind.

Recently, a multi-species comparison of the *HoxA* gene cluster was carried out [300]. The species chosen ranged from tilapia to human and included the distant *fugu.* All of the species were separated by more than 500 million years of evolution. Many putative regulatory elements that are known *Hox* regulators were recovered using this group's method and new elements were found to be almost completely conserved. Many of the new elements were found to be identical to other previously known regulatory sites for binding proteins.

As for the putative regions they found, the first part of the *HoxA11* intron seems to be the most highly conserved, at about 80% sequence similarity among the species examined. The fragment actually represents the consensus homeodomain binding site of HB1, which is also

located in the intron of the mouse *Hox*A4 and 7 genes [301]. These three *Hox* genes are expressed in different regions and at different times of the developing embryo. It is believed that the location of HB1 elements in the *HoxA* cluster might be directly related to this timing and distribution pattern.

Another region that displays some interest is the CNS directly upstream of *HoxA7*. C*is*-regulation of the *Hox*A7 gene has previously been described by Knittel et al. as an enhancer located 1.6 kb upstream of the coding sequence in human and mouse [302]. The authors carrying out the cross-species analysis hypothesized that another proximal regulatory element could cooperate in the expression of *Hox*A7 and the 185 bp stretch with more than 84% sequence identity found by Meyer et al. could be that region. The comparison revealed several completely conserved sequences within this fragment, characterized by the short motif GTAAA, further strengthening this belief.

A region immediately upstream of the *HoxA5* gene is also found to be between 70% and 85% conserved and contains the RARE elements "box c" and "box d" that were originally identified in human and mouse [303]. These particular elements are found in all of the genes of the paralogous group 5, and are known regulatory binding sites in the mouse *HoxA5* gene [303]. A 154bp stretch located upstream of *HoxA4* was found to be 85% conserved and also contained a RARE element that is part of the *HoxA4* promoter [304].

Of the many conserved sequences identified by Meyer et al., most matched previously described transcription factor binding sites like that of the nuclear factor *NF1* binding sites [305], abdominal B (*Abd*B) homeobox gene binding sites [306], *Cdx*A homeobox gene binding sites [307], and murine homeodomain binding sites [308].

One of the more significant findings was that of the *Krx20* binding site, which was found in humans, mouse, pufferfish, and tilapia clusters. *Krx20* binding sites are normally involved in the regulation of *HoxA2* during development [309]. This site is followed by a 12 bp long conserved sequence motif called "box a", which is highly similar to "box1", the corresponding element associated with *Krx20* binding sites in the *HoxB* cluster [310].

Many of the novel CNSs found in the above mentioned study were located immediately upstream or downstream of the *Hox* genes and these locations are normally associated with specific functional roles [311]. For instance, promoters are located immediately 5'of the *Hox* genes and RAREs are located 3' of the regulated gene (e.g., [312]). Of all the regions, the largest conserved areas were found located between two genes and were actually quite distant (1–5 kb) from both. Because *cis*-regulatory regions in *Hox* clusters are located in positions that are intermediate between the genes they regulate, these large conserved areas are the most significant. For example, an element named H8/7–6 FCS [313] was shown to exist in all four clusters of mammals and shark that they compared, and Meyers group also showed that this element is present in the *Hox*A cluster of fishes. Meyers concluded that the conservation of the nucleotide sequence and relative position in all clusters examined so far makes this element an excellent candidate for an evolutionary conserved *cis*-regulatory element [300].

The functional analysis of CNSs, especially those identified in the *HoxA* cluster, needs to be carried out to verify the cis-regulatory nature of these elements. Using "enhancer" essays with transgenic mice have provided some data on the function of non-coding DNA around developmental genes (e.g., [314]), although this method is extremely slow. An alternative approach uses transient expression in zebrafish embryos [315], which allows for an increase in

experimental throughput. However, some of these elements may not exactly function as enhancers of expression and other types of assays are required to determine their functionality. The 3C method of analysis in conjunction with sequence conservation analysis could be a viable option to determine possible structural roles for these CNSs and is the method we have applied in our research.

1.8 P19 Embryonal Carcinoma Cell line

Cancer cells in modern day biological laboratories are an extremely useful tool when examining various cancer mechanisms as well as a means to discover new therapeutic targets. This is best exemplified by the cervical cancer cells of Henrietta Lacks (HeLa), which led to her death, and yet in the hands of researchers allowed for the development of methods to defend ourselves against these life threatening cells [316].

Developmental biology has had its own special kind of cell type: embryonal carcinomas (EC) cells. This cell type is derived from teratocarcinomas, a spontaneously developed malignant tumor found in humans and mouse testes as a result of defective germ cells. These cells are capable of altering their malignant phenotype to a non-malignant one simply by utilizing the process of differentiation. They can be induced to differentiate through the administration of certain chemical inducers or by transplantation in mouse embryos to extrauterine sites where they may develop into a number of tissue types. These malignant tumors also contain a population of stem cells. These undifferentiated stem cells can be isolated and grown indefinitely in culture [317]. Human Tera-2, mouse F9 and P19 cells are all very commonly used in the study of developmental biology [318, 319].

The mouse P19 cell line was established to be a cell line that is heterozygous for X-linked alleles [320]. In its design, 7.5 day old embryos from the crossing of C3H/He females with males carrying an X-chromosome that contained a number of variant alleles were transplanted into the testis of a C3H/He mouse. The primary tumor that developed from the transplantation contained the undifferentiated P19 EC cells, which were verified for pluripotency by injecting them into blastocysts of a different mouse strain. The P19 cells, of which had a unique karyotype of 40:XY, were found in all three germ layers of the developing mouse[321].

The differentiation process of these types of cells (EC) in culture depends greatly on the formation of embryoid bodies, a group of non-adhering aggregates which resemble the inner cell mass of a developing embryo [322]. The outer layer of the embryoid bodies begin to differentiate earliest and follow an endoderm-like cell lineage, whereas the inner cells remain undifferentiated until further cues are given. With the use of specific drugs in conjunction with the formation of embryoid bodies, cellular derivatives of all three germ layers are able to be induced.

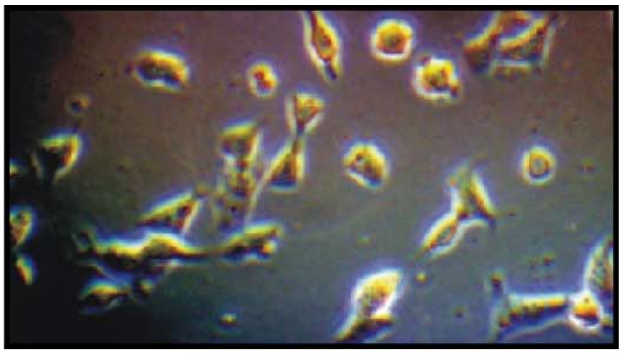
Of greatest significance, these pluripotent cells are capable of differentiating into neuronal and glial cells in the presence of all-*trans*-retinoic acid (see Figure 4) [323]. The neuronal cells differentiated in this way have been reported to be extremely similar to that of mammals in some aspects such as the morphology [321], functional synaptic formations [324] and the expression of neuron-specific genes for neurotransmitters [325] and proteins [323, 325].

As seen in the above sections, numerous groups have employed the use of the P19 cell line when examining cell fate. RA treatment in this cell line is capable of causing irreversible differentiation after only 4 hours of administration, however, markers for differentiation are only observed after approximately 3 days following treatment. RA treated P19 cell aggregates express

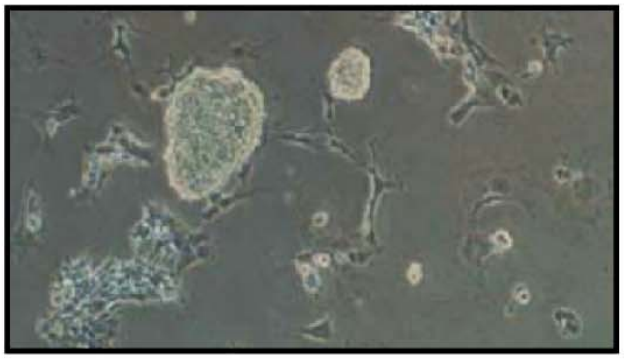
a few basic helix-loop-helix (bHLH) genes such as the *Mash-1,* a mammalian homolog of *achaete-scute, Math-1* and *NeuroD,* the mammalian homolog of *atonal*, and *Nscl-2* [326]. The products of these basic helix-loop-helix genes have been well known to function as positive regulatory molecules on neural fate determination and differentiation. P19 EC cells have therefore been thought to considerably reflect the early events of neuronal differentiation *in vivo*, and are widely used as an *in vitro* model system that is eminently suitable for analyzing the regulatory mechanisms of the mammalian neuronal differentiation pathway.

With respect to the actual mechanism governing the P19 neuronal differentiation process and the role that RA plays, not much is known. However, it can be said that RA initiates a cascade of cellular events responsible for affecting the cellular lineage pathway of this cell line. Although many possible steps have been elucidated, the specific sequence has yet to be determined. One extremely likely initial event to occur would be the binding of RA to its target receptor. This step relates to *Hox* gene function and connects the study of the *Hox* clusters with regard to *Hox* regulation to that of neuronal development.

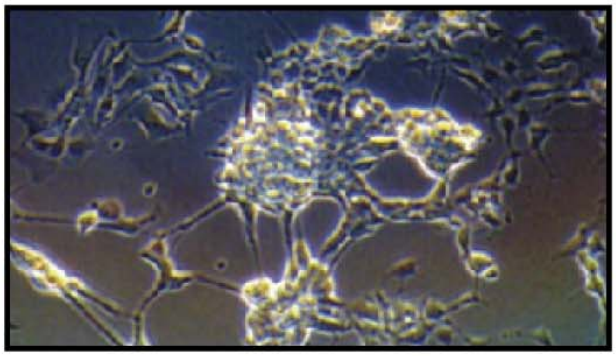
Pluripotent P19 Cells



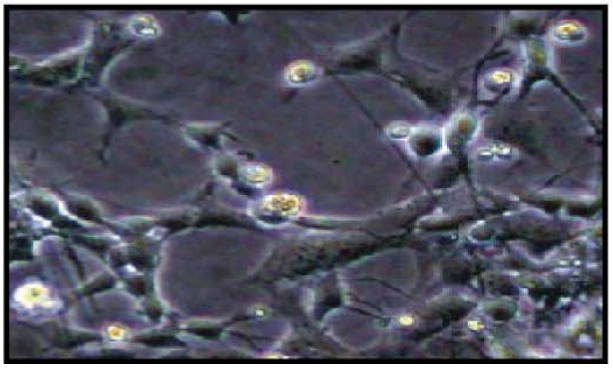
Embryoid Bodies



Differentiating P19 Cells



Neurons



Differentiation

Figure 4. P19 Differentiation.

The mouse P19 cell line undergoes neuronal differentiation upon treatment with retinoic acid. The cultured pluripotent cells begin this process first by clustering in suspension and forming embryoid bodies (amalgamations of cells). These bodies can then proceed through the cell lineage pathway and eventually develop into neuronal and glial cells.

Section 2: Results

2.1. Annotation and Primer Design

To ascertain the overall chromatin conformation in the mouse HoxA gene cluster, we began by fully annotating each of the 4 clusters. We selected and annotated BACs that span each Hox cluster as well as all of the Hox genes, including their intrinsic elements (5'UTR, 3'UTR, introns and exons), using the MacVector annotation software. This provided us with sequence specific information regarding each feature of the cluster. Even though we only examined the HoxA cluster, all additional Hox clusters were also annotated for future comparative and exploratory experiments. The annotation of all 3C fragments was based on the restriction digestion pattern of the HoxA region by the EcoRI endonuclease. This enzyme was chosen based on its efficiency and the relative uniform separation of cut sites within the examined area. Importantly, only the most highly observed gene isoforms were used in the annotation based on the data compiled by the ENCODE consortium. Only HoxA genes *HoxA1*, *HoxA3*, and *HoxA7* display splicing variants, however, for our purposes these alternative transcripts did not affect the data we obtained since they were still contained within the same fragments.

Gene expression primers were selected for all 39 Hox genes based on specific design requirements (20-22bp, 50% GC content, 5' GC clamp, etc.). The 3C primer design of all 4 gene clusters and gene desert region were designed using the web based Primer3 software. These primers had similar requirements that needed to be met (28-32bp). The verification of sequence specific primers was essential during the design process to ensure that artifacts and random products would be minimized in the PCR steps of both gene expression profiling and 3C

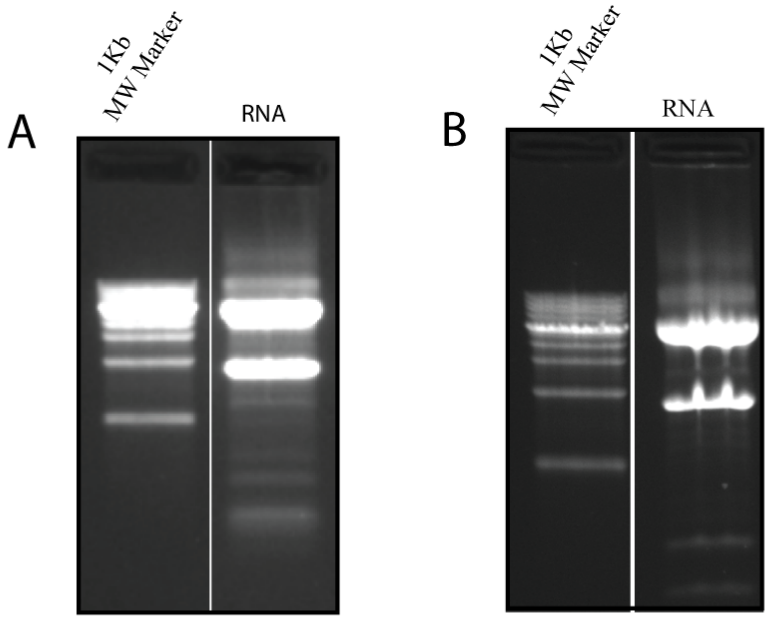
analysis. The uniqueness of each primer was verified by BLAST and BLAT analysis. A list of all designed primers can be seen in Table 1 of the appendix.

2.2. *Hox* gene Expression profile

To assess the potential effect on gene transcription due to chromosome conformation and to establish a baseline of expression in the HoxA cluster, we were required to determine the expression profiles of the Hox genes. The mouse *HoxA* gene expression profile was analyzed from undifferentiated mouse P19 cells by RT-PCR. RNA extraction (Figure 4) was followed by DNASe treatment to avoid DNA contamination (Figure 4). All major RNA species were observed on agarose gel stained with ethidium bromide and visualized by UV: Large 28s rRNA, small 18s rRNA, tRNA and mRNA trails. The isolated and purified P19 RNA was reverse transcribed and analyzed for gene expression at a non-retinoic acid-treated time point. The HoxA expression profile generated for these cells can be seen in Figure 4. The results obtained by end-point PCR for the expression of the HoxA genes from the mouse P19 embryonal carcinoma cell line indicate that the majority of the early 3' Hox genes are slightly, or not at all expressed, in the pluripotent cell state (Figure 4). These results are comparable to a high throughput method of gene expression profiling, previously carried out by Reese and Ramos-Valle, in which the expression for all Hox genes was determined in the absence and presence of retinoic acid at 48h and 96h [327]. We also examined all other Hox clusters and observed similar results in which low levels or no expression was observed (data not shown). Very little expression was observed in this cellular state and allowed us to establish a baseline (undifferentiated) state of expression to compare and contrast with our future 3C data. It is important to mention that there are a few instances of a representative signal in the minus RT lane of some of the HoxA genes. This may

have been due to incomplete removal and/or digestion of DNA contaminants, or simply from the contamination by DNA from an outside source.

The expression of HoxA genes was also measured after 3 and 4 day RA induction of P19 cells (Figure 5). This was done to verify which HoxA genes would respond to RA induction, in order to correlate gene expression whit chromatin architecture. The expression of many of the 3' *Hox* genes increased upon retinoic acid treatment owing to the activation potential of this compound on these genes. Some late *HoxA* genes were also observed to respond to RA treatment after the 3 and 4 day exposure to the signaling molecule (*HoxA 9, 10,* and *13*). Although the method of end-point PCR is not a quantitative tool, the overall effect of gene activation can be inferred based on the results obtained: the actin control in the untreated cells compared to the treated cells had a much lower intensity and the *Hox* signals were much greater in the treated cell. This data is corroborated by numerous expression profiles carried out by various other groups under similar conditions [327-329]. In addition, the induction of the *Hox* genes, specifically the early genes (1-9) have been previously shown in drosophila to be under the developmental control of retinoic acid (RA) [93]. Many of these genes have been well characterized and found to contain elements that respond to RA treatment to initiate gene transcription, mainly elements that are bound by the RAR homodimer [330-332]. From our results, we can therefore conclude that the activation of early *Hox* genes is initiated after at least 3 days of RA treatment, but this is likely an underestimation of the activating potential of RA. We may now also assume that the regulation of the *HoxA* cluster in response to RA effects early and late genes in a different manner and may imply certain early and late chromatin conformations.


A
1Kb
MW Marker
RNA
B
1Kb
MW Marker
RNA

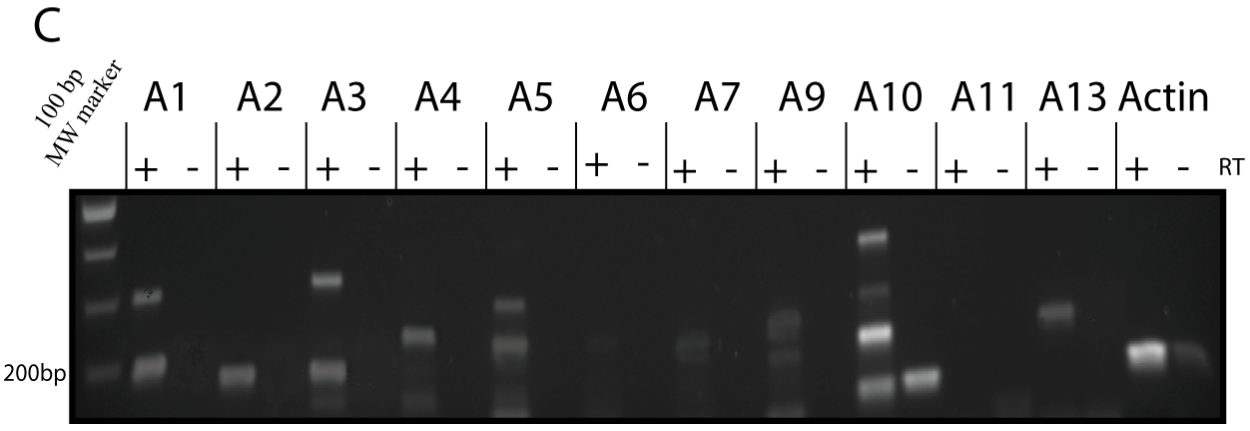

C
100 bp
MW marker
A1
A2
A3
A4
A5
A6
A7
A9
A10
A11
A13
Actin
+ - + - + - + - + - + - + - + - + - + - + - RT
200bp

Figure 5. Undifferentiated P19 HoxA expression profile.

A) RNA isolated from undifferentiated P19 cells collected 48 hours after passaging. **B**) DNAse treated P19 RNA. **C**) The expression profile of the HoxA gene cluster in undifferentiated P19 cells. Primers were designed to form products of approximately 200bp. Positive and negative reverse-transcription (RT) PCR products were detected on 1% agarose gel. Early genes (HoxA1-5) display low levels of expression when compared to the actin control. HoxA7, A9, and A13 also showed some slight levels of expression. Artifacts were observed in the + RT lanes of A1, A3 and A10. –RT products can also be seen in some of the lanes indicative of inefficient DNAse treatment.

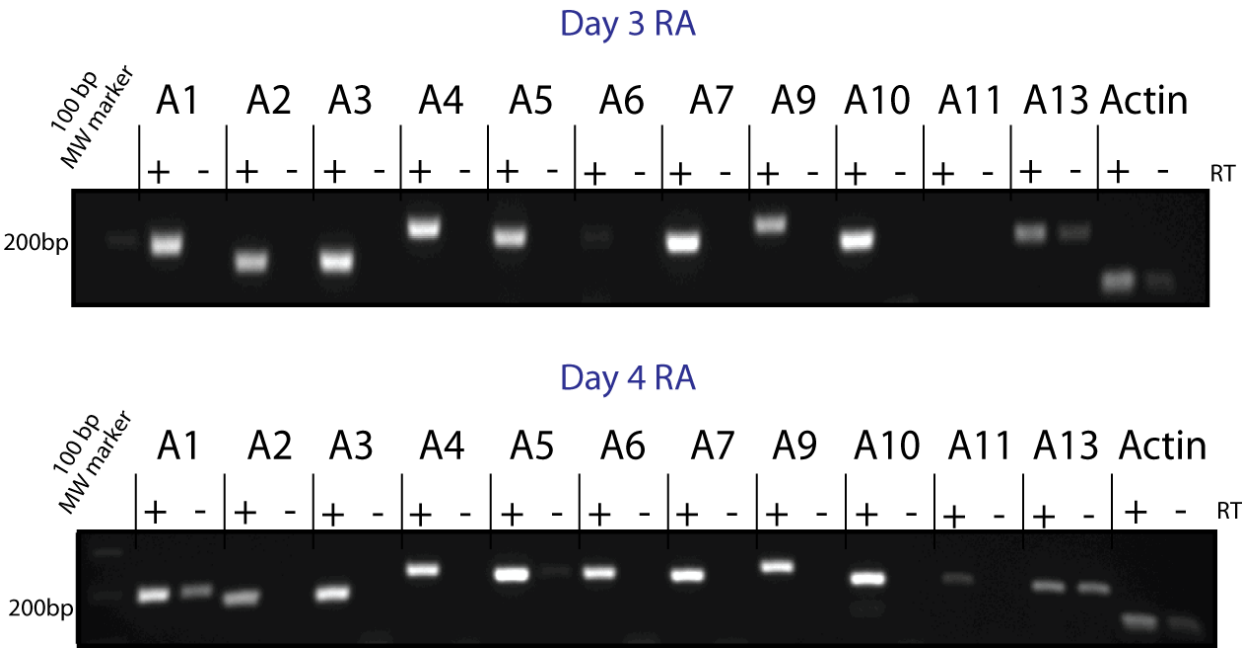


Figure 6. P19 HoxA gene expression after retinoic acid treatment.

The expression profile for the P19 HoxA gene cluster was analyzed after 3 and 4 days of retinoic acid (RA) administration. The expression levels of almost all of the genes within the cluster increased after 3 days of RA treatment. HoxA6 did not show expression before day 4 of treatment. HoxA11 did not show an increase with RA treatment. –RT products can be seen in lanes A1 and A13 as a result of inefficient DNAse treatment or possible contamination.

2.3. 3C Library Construction

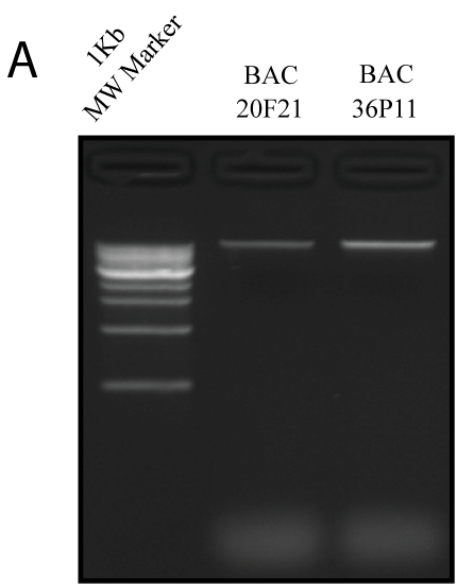
We next used Chromosome Conformation Capture (3C) technology to map the architecture of the *HoxA* gene cluster from undifferentiated mouse P19 cells in order to create the first high resolution 3-dimensional map of this cluster and to identify any legitimate structures for further investigation and characterization. The conformation could also be used to correlate its features with the HoxA gene expression profile in untreated and RA-treated cells.

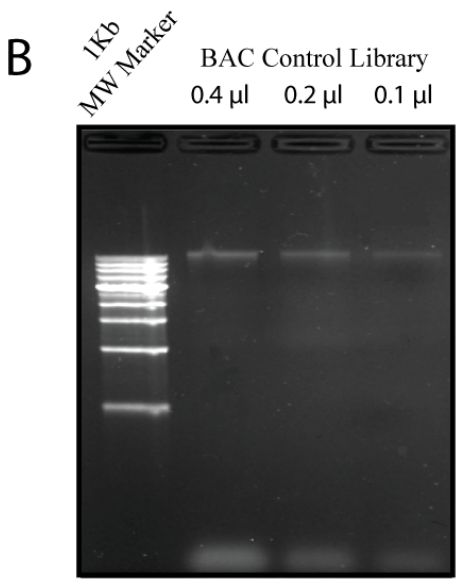
To create the 3C control library we were required to isolate BAC DNA for the region of interest. An array of BAC clones spanning 500kb of each *Hox* cluster was selected based on minimal overlap and no gaps between clones. A sample of the clones isolated for *HoxA* and *HoxC* can be seen in Figure 6. BAC DNA migrates as a tight band at approximately 10kb in a 0.8% agarose gel.

A total of 5 BACs were used in library construction: one each to encompass the ~200kb central region of all 4 Hox clusters (RP-23 library 20F21, 196F5, 36P11 and 101K1) and one to represent a transcriptionally silent gene desert region (450L4) from mouse chromosome 19 that is used to normalize our libraries to one another. BAC clones were quantified by quantitative real-time PCR (Q-PCR) of in order to determine the appropriate stoichiometric ratios for the cloned segments required in 3C control library construction (data not shown). As mentioned earlier, the ratio of BAC fragments must be 1:1 in order to avoid potential errors in analysis, such as abnormally high or low interactions. The agarose gel quality control check of the BAC 3C control library can be seen in Figure 6. This library migrated as a tight band slightly above the 10kb marker, which is expected because it is supposed to follow a similar migration pattern as an intact chromosome in the agarose gel. Quality control was assessed by measuring the intensity of

0.1µl, 0.2 µl and 0.4µl of the sample and observing that as volume doubled so did intensity. Many of our libraries failed to deliver any results due to unknown errors that occurred during their construction. We had suspected that one of the reagents we were using was possibly to blame, but we were not able to verify our claim. We focused purely on trying to construct a functioning library using the materials and protocol that we had available. We were successful about 40% of the time.

The 3C control library was titrated using primers GD5 and GD6 to ensure quality control and determine a starting PCR volume for 3C analysis. The agarose gel and graphical representation of this titration can also be seen in Figure 6. The selected volume is chosen based on the limit of detection due to saturation of the signal by UV analysis. The chosen signal/volume is just under saturation levels in order to allow for semi-quantitative analysis by 3C. A volume of 1/600µl was used in our experiments.


A
1Kb
MW Marker
BAC
20F21
BAC
36P11


B
1Kb
MW Marker
BAC Control Library
0.4 µl
0.2 µl
0.1 µl

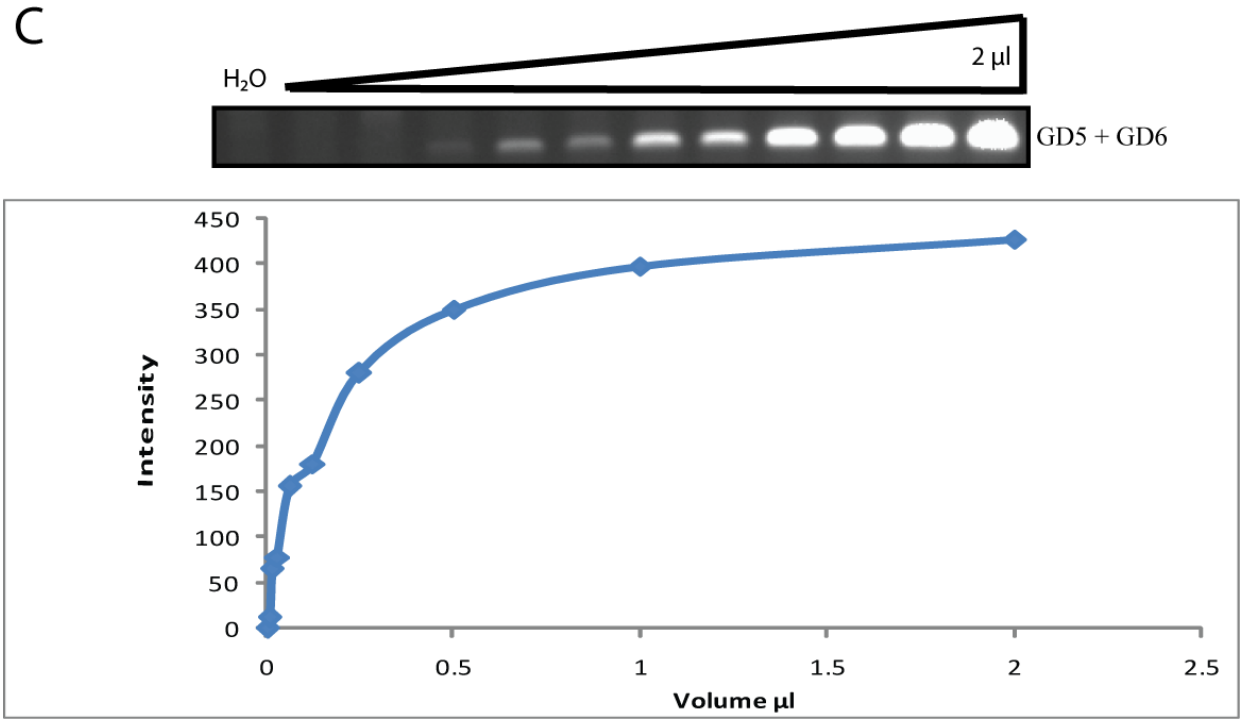

C
H2O
2 µl
GD5 + GD6
Intensity
Volume µl
0
50
100
150
200
250
300
350
400
450
0
0.5
1
1.5
2
2.5

Figure 7. 3C BAC control Library

A) Isolation of Bacterial Artificial Chromosomes (BACs) RP23- 36P11 and RP23- 20F21. These BACs were used in the construction of the 3C BAC control library. Three additional BACs were also used (not shown). **B**) Quality control examination of the 3C control library at 0.1µl, 0.2µl and 0.4µl volumes. The library migrates as a tight band just above 10kb in a 0.8% agarose gel. **C**) The control library was titrated using gene desert (GD) primers GD5 and GD6 up to a volume of 2µl. The agarose gel of the titration is displayed above the graphical representation. The selected volume for 3C experiments based on this data was 1/600 µl of control library.

Cellular DNA was obtained from exponentially growing undifferentiated P19 cells, verified by construction of a growth curve for this cell line (Figure 7). $5x10^7$ cells were used to produce the experimental 3C library according to the protocol for cellular 3C library formation discussed in the materials and methods. The agarose gel quality control of the cellular library can be seen in Figure 8. This 3C library is also observed to migrate above 10kb in a tight band with a small trail comprised of un-ligated fragments or smaller ligation products.

The 3C P19 cellular libraries, 3 in total for *HoxA* analysis, were each titrated using primers GD5 and GD6. The agarose gel and graphical representation of the titration of our first library can be seen in Figure 8. The resultant linear range starts to reach its limit at 1µl; therefore we selected a volume under this limit to remain within the linear range of detection.

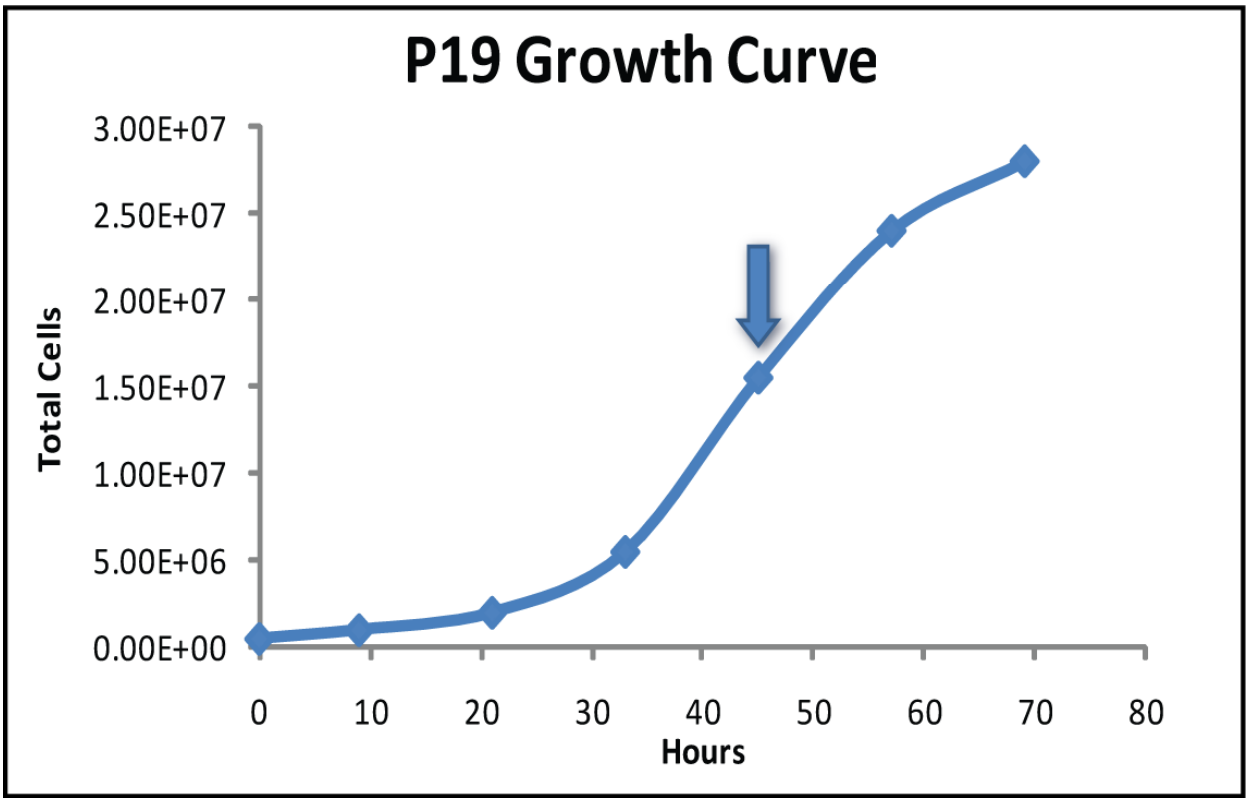
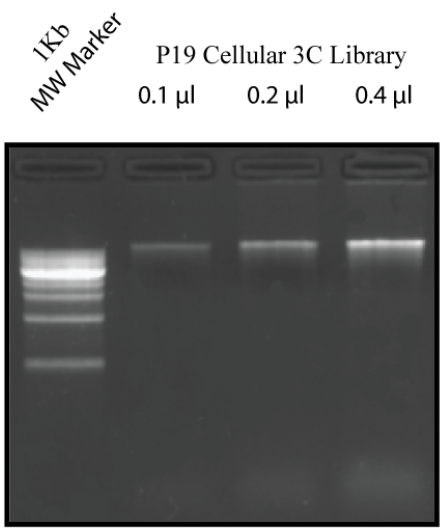


Figure 8. P19 Growth Curve

A growth curve representing cultured mouse P19 cells over time. The 48 hour time point during exponential growth was used in our 3C and gene expression experiments.

A


1Kb
MW Marker
P19 Cellular 3C Library
0.1 µl
0.2 µl
0.4 µl

B


H2O
2 µl
GD5 + GD6

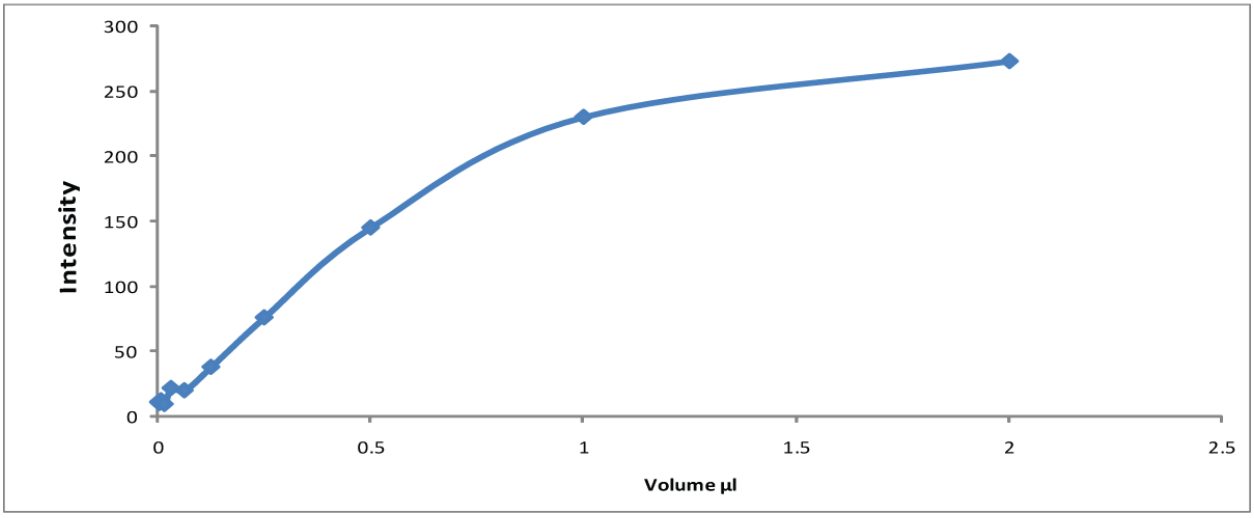

Intensity
300
250
200
150
100
50
0
0
0.5
1
1.5
2
2.5
Volume µl

Figure 9. Cellular P19 3C Library

A) The P19 cellular 3C library was assessed for quality control at 0.1µl, 0.2µl and 0.4µl volumes. The cellular library migrates as a tight band just above 10kb in a 0.8% agarose gel. B) The cellular library was titrated using gene desert primers GD5 and GD6 up to a volume of 2µl. The agarose gel of the titration is displayed above the graphical representation. Volumes used for 3C experiments varied depending on the library used due to the variability when constructing cellular libraries. The volume used from this library was 0.7µl.

2.4. 3C Analysis

HoxA gene cluster by 3C analysis was carried out using both control and experimental P19 libraries. A sample agarose gel of a fixed point experiment is presented in Figure 9. PCRs were done in triplicate for each primer pair to ensure the validity of our results. The graphical representation of this experiment is displayed in Figure 10 and is used to illustrate trends in interaction frequency between fragments; peaks representing high interaction relative to genomic separation. The culmination of all fixed point experiments are seen in the *HoxA* cluster heatmap of Figure 11. A number of missing interactions are seen in the heatmap. These blank regions are due either to the nature of the primer pairs not being compatible or simply because we did not have time to examine them. The calculated interaction frequencies between all analyzed fragments are a representation of their relative spatial interactions. A 2-D schematic representation of the data can be seen in Figure 11 and the major interactions are discussed below.

We further analyzed the interaction frequencies that were obtained by 3C analysis and compared this data to previously identified conserved non-coding sequences (CNS) [300] to see if any correlations could be found. All CNSs and their corresponding fragments are listed in Figure 12. All known RARE cis-acting elements as well as similar and newly discovered elements are listed with their respective fragments.

MF58 (*HoxA10*) is one of the most highly interacting fragments from this cluster. It displays very high interactions with fragments MF86, MF79, MF73, MF63 and MF53. We also

observed MF58 interacting slightly less with fragments MF70, MF67, and its neighbors MF57, MF59, MF60. This genomic region (MF58) has not been characterized as of yet, however, many of its corresponding interacting fragments have been extensively examined and found to contain various cis regulatory elements [302, 304]. We believe that further investigation of this fragment will reveal the identity of a cis-regulatory element simply due to its high degree of interaction, possibly a retinoic acid response element (RARE) or a similar type of element that can coordinate transcription factors and chromatin remodelers. In fact, a conserved non-coding sequence has been identified in this fragment, however its function remain unclear [300].

MF63 (*HoxA7* and intergenic *HoxA7-HoxA6*) was easily identified as the most highly interacting fragment. It displayed extremely high IFs (interaction frequency) with almost the entire cluster. The major interacting fragments for this region were found to be MF86, MF79, MF78, MF74, MF73, MF70, MF59, MF58, MF57 and its neighboring fragments MF62, MF65 and MF67. Considering that such high interactions occurred with this fragment (i.e. MF70 of 12.2) we can postulate that surrounding fragments (i.e. MF69 and MF71) displayed high interactions as a result of their neighbor "pulling" force and not necessarily due to their own interactions with MF63. Alternatively, this may have also been due to a cooperativity of multiple neighboring fragments augmenting the fixed point primer results. Of these major interacting loci, the currently identifiable regions in the literature that contain cis-acting regulatory elements are MF70, MF73 and MF86 [303, 304]. Elements of a *HoxA7* RARE have been identified in species including Human and mouse[330]. Interestingly, this fragment (MF63) has been found to contain a CNS that significantly resembles the Human *HoxA7* enhancer element characterized by Knittel et al. [302].

A region of MF70 has itself been characterized in the literature, identified as a conserved non-coding sequence responsible for *HoxA5* promoter activity and comprising a RARE [303] (Figure 12). The mouse RARE was shown to bind RARs and mediate RA activation. This may indicate why fragment MF70 is highly capable of interacting strongly with so many other fragments. Gillespie et al. have determined by chIP analysis that all RAR (RAR/RAR or RAR/RXR) are found associated to RAREs before and after RA treatment in F9 cells[333]. They also determined that transcriptional repressors such as SUZ12, a histone modifying enzyme and a member of the polycomb repressive complex 2 (PRC2), are associated with RARs prior to RA treatment and re-associate when RA is absent. RA binding causes a conformational shift in RAR and the subsequent release of bound factors from the repressive complex [334].This may be the major repressive region considering its central location in the linear and 3-D cluster as well as the homology of this specific RARE among the other *Hox* clusters[303].

MF85 displays relatively high IFs with some previously mentioned relevant fragments such as MF73, MF68, MF59 and MF58. Its neighbor, MF86, however also shows a high interaction with these respective fragments but interestingly also shows significant interaction with MF70, a fragment containing the *HoxA5* RARE. MF85 contains the *HoxA2* promoter region that has been well characterized in the literature[335]. The fact that fragment MF85 interacts with these regions may be, in part, due to the interaction of its more relevant neighbor MF86. This RARE containing fragment is capable of recruiting the RA receptor and repressive complexes similarly to the *HoxA5* RARE [312].

Our results indicate that the overall 3-D conformation of this cluster appears to contain 4 looping regions with a central highly interacting core feature.

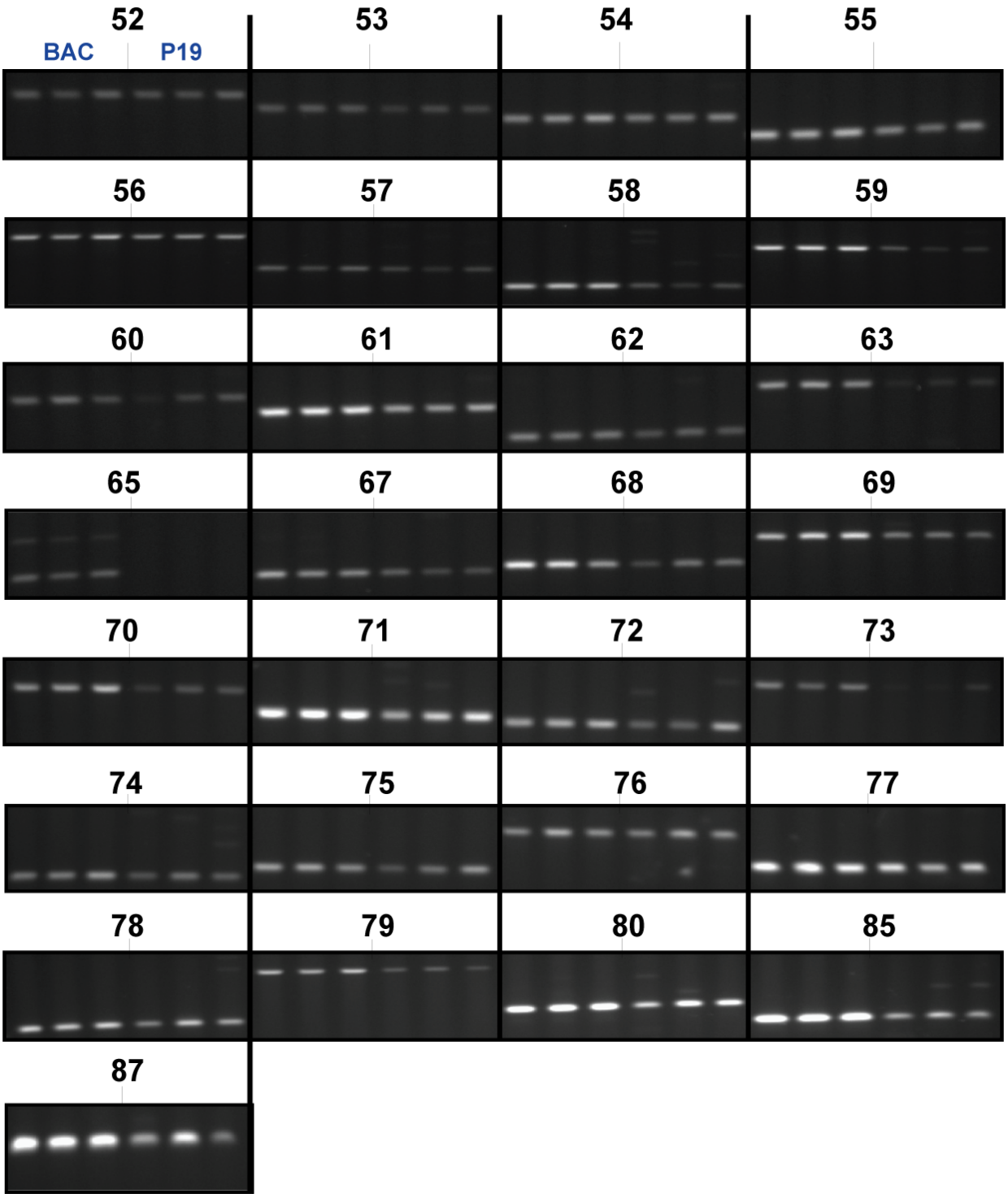

52
BAC
P19
53
54
55
56
57
58
59
60
61
62
63
65
67
68
69
70
71
72
73
74
75
76
77
78
79
80
85
87

Figure 10. Raw Data of MF86 3C Fixed point Experiment

An Agarose gel detection of all PCR products for the fixed point 3C experiments of mouse fragment (MF) MF86. All contacts are measured in triplicate to ensure that observed signals are acceptable. Fragment names are shown above each triplicate. BAC 3C control experiments are also done in triplicate and are observed adjacent to each cellular result. The majority of PCR products are 200bp, with some exceptions.

Fixed Point 3C of MF86

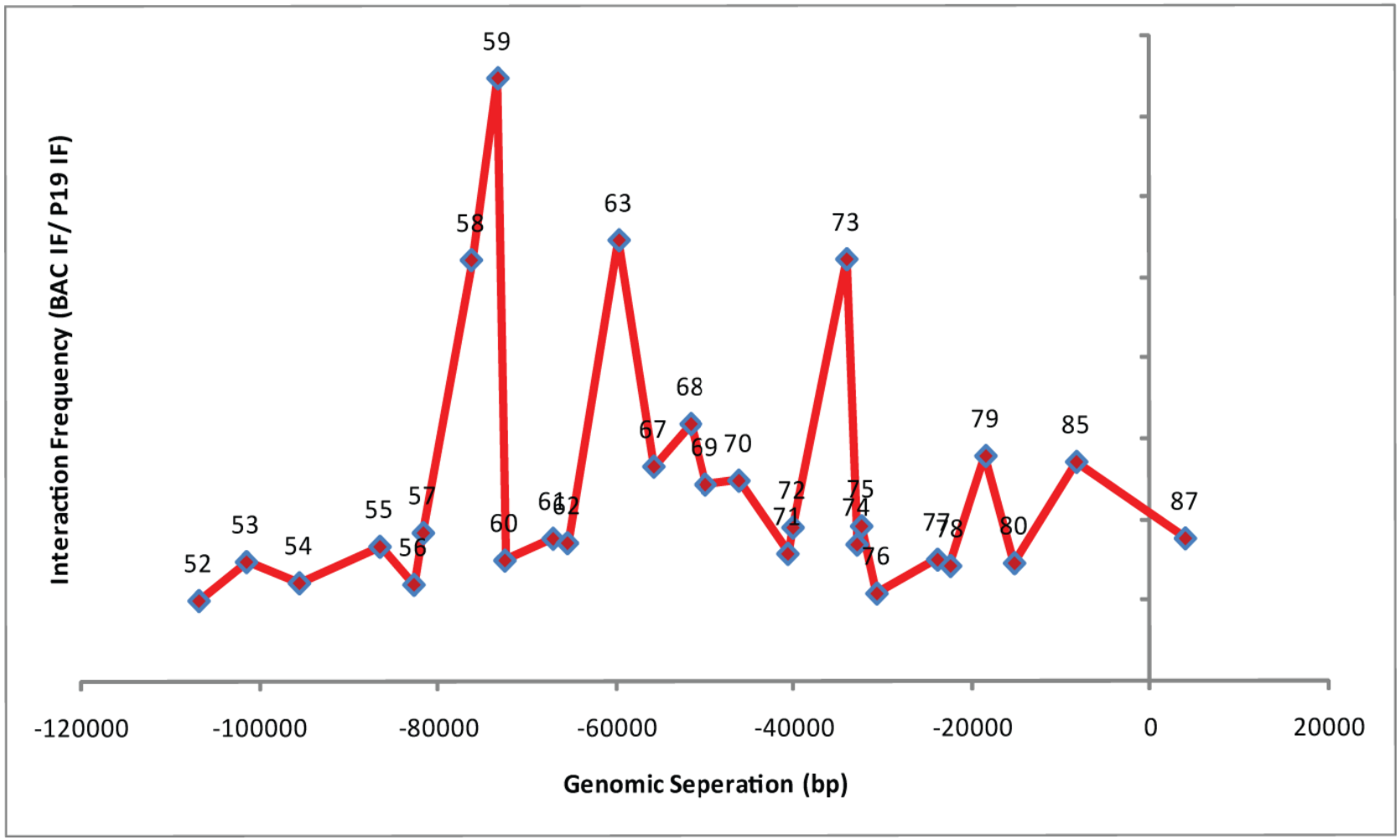


Figure 11. Analysis of MF86 3C Fixed Point Experiment

A graphical representation of the analysis for fixed point experiments of MF86 seen in Figure 11. The interaction frequency (IF) is plotted against genomic distance in which 0 represents the location of MF86. All fragments along the axis are listed above their respective interaction frequencies with MF86. Fragments MF59, MF63 and MF73 display higher than average IFs with fragment MF86, indicating the potential for genomic interaction.

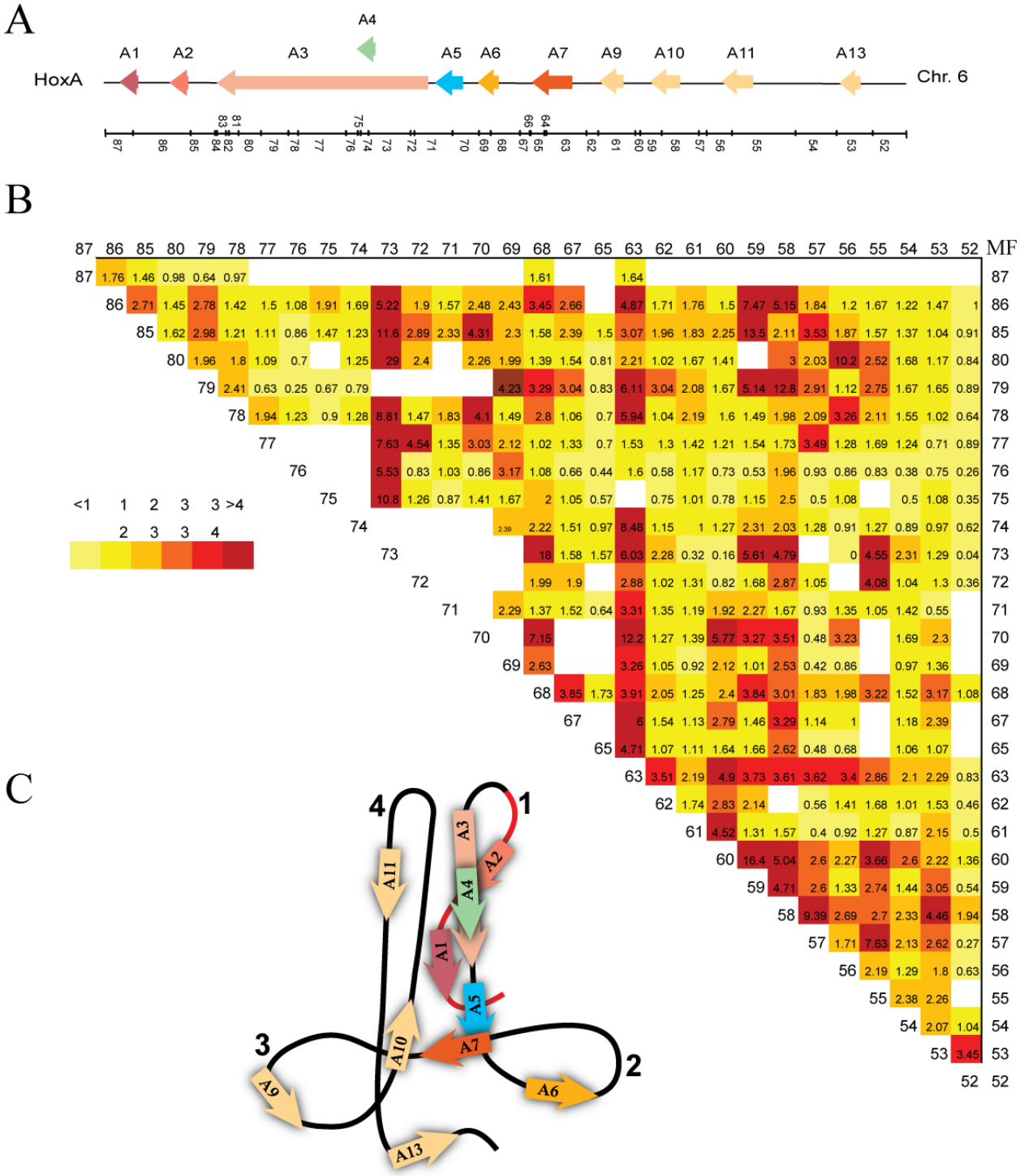

A
A4
A1
A2
A3
A5
A6
A7
A9
A10
A11
A13
HoxA
Chr. 6
B
MF
<1
1
2
3
3 >4
2
3
3
4
C
A11
A3
A2
A4
A1
A5
A10
A7
A9
A6
A13
1
2
3
4

Figure 12. Analysis of 3C experiments in the Mouse P19 HoxA cluster

A) A schematic representation of the annotated mouse HoxA gene cluster including the location of all fragments used for analysis. The schematic is drawn to scale. B) A heat map depicting the culmination of all fixed point experiments used in examining the conformation of the mouse HoxA gene cluster. Interaction frequencies (IFs) and their corresponding representative colors are indicated in the figure legend. Values for interaction frequency ranges were selected based on average IF over the entire cluster. Empty interaction boxes are observed due to insufficient data collection and weak primer pair efficiency for these sets. C) A schematic 2-D representation of the 3-D conformation of the mouse HoxA gene cluster based on the results obtained in the heat map (B). The gene region colored in red should be perceived as going into the plane of the paper, while the remainder of the cluster (black) is flush with the paper. Loops are denoted by number. Genes and genomic distances are not to scale.

Mouse Fragments and CNSs Identified Through the Comparative Approach

MF	Position	Length (bp)	Striped bass	Pufferfish	Zebrafish a	Zebrafish b	Horn shark	Human	Mouse	Over 95%	Literature	New/similar homeobox binding sites
53	1 kb upstream 13	63		86	73	82	78	79	80	1 X 10		NF-1 (Rossi et al. 1988)
53	Imm. upstream 13	188		83	65	63	66	75	71	0		New
55	13-11	192		89	26		60	68	67	2 X 10		New
55	Imm up stream 11	230		89	66	68	63	70	71	5 X 7-28		New
56	11-10	121		96	84	85		64	63	2 X 6-8		New
58	Imm. up stream 10	391	92	86		66	68	70	68	2 X 8-25		New
59	10-9 a	96	95	86	63		66	61	65	2 X 6-7		Abd B (Ekker et al. 1994); RNA pol. II cap signal (Bucher 1990)
59	10-9 b	95	98	98	89	81	79	73	72	1 X 24		Murine homeotic proteins b.s. (Catron et al. 1993)
61	Imm. up stream 9	191	94	87	63	56	69	61	63	2 X 5-6		Target sequences chicken CdxA (Margalit et al. 1993)
62	9-7 a	62	100	98	92	72	78	79	79	2 X 11-15		Abd B (Ekker et al. 1994)
63	9-7 b	276	96	98	71		71	70	69	3 X 6-11		c-ETS-1 protein b.s. (Woods et al. 1992)
63	Imm. upstream 7	185	95	88			79	78	78	3X 9–14		HoxA7 enhancer regulatory element, H. sapiens (Knittel et al. 1995)
67	7-5	163	81	78	77		78	81	81	2 X 8–11	H8/7-6 FCS (Kim et al. 2000)	
70	Imm. up stream 5	529	93	84	69		38	76	76	8 X 6–39	RARE (box c and box d), H. sapiens, M. musculus (Odenwald et al. 1989)	
71	5-4 a	280	99	94	77		82	83	83	7 X 9–33		Pax b.s., (Epstein et al. 1994); Ultrabithorax b.s. (Ekker et al. 1991); target sequences of chicken CdxA homeobox gene (Margalit et al. 1993)
71	5-4 b	63	97	98	83		83	81	83	2 X 9–19		Dof b.s. (Yanagisawa and Schmidt, 1999)
73	5-4 c	209	95	93	67		71	69	69	5 x 8–15		NF of C-EBP family (Grange et al. 1991)
73	5-4 d	239	92	89	81		84	79	78	7 X 7–24	RARE (HoxA4 promoter, H. sapiens, Doerksen et al. 1996)	
73	Imm. upstream 4	83	100	100	89		91	78	76	3 X 6–30	RARE (HoxA4 promoter, H. sapiens, Doerksen et al. 1996)	
73	4-3 a	78		91	69		76	72	66	2 X 6–7		Dof b.s. (Yanagisawa and Schmidt, 1999)
77	4-3 b	480		87	67		71	66	63	5 x 6–10		New
78	4-3 c	51		93	72		80	75	73	2 X 6–10		New
79	4-3 d	136		96	76		76	66	65	4 X 8–12		New
80	Imm. upstream 3	235		86	90		82	73	79	6 x 7–13		New
83	3-2 a	476		79			61	66	60	0		New
85	3-2 b	189		93	72	81	68	69	67	5 x 6–9		New
85	Imm. upstream 2	382		89	64	78	77	77	77	8 x 8–43	HoxA2 promoter, M. musculus (Tan et al. 1992)	
86	2-1	190		93			61	72	72	3 X10–11	RARE, M. musculus (Frasch et al. 1995)	New
86	Imm. upstream 1	352		78	60		59	64	61	2 X6–8	RARE, M. musculus (Frasch et al. 1995)	New
	Total	6233										

Column 1: position of CNS in the tilapia HoxA cluster. Column 2: length in bp of the CNS. Columns 3–9: percentage identity of the corresponding CNSs in the other genomes examined. Column 10: number (x) of occurrences and length in # of bp of highly conserved sequences with over 95% identity among all species. Column 11: reference for previously described CNSs in Hox clusters. Column 12: newly suggested CNSs, and reference for known binding sites that show a similar sequence.

Figure 13. Conserved Non-Coding Sequences and Mouse Fragments

List of conserved non-coding sequences (CNSs) identified by comparative analysis and the mouse fragment (MF) in which it is located. Previously identified functions of CNSs are listed in column 12. New and similarly identified sequences are listed in column 13.

Adapted with permission from "Evolutionary Conservation of Regulatory Elements in Vertebrate Hox Gene Clusters" Santini et. al. 2003

Section 3: Discussion

The mouse HoxA gene cluster contains within it important regulators of growth and development. Its own regulation, however, seems to elude us to a certain degree. The collinear mechanism of Hox gene expression remains mysterious in the sense that it works in this way but it is not known how. By examining the 3-dimensional landscape of the HoxA cluster in its resting state, we have determined a potential mechanism or level of control for regulating the collinear mechanism.

We began with determining the expression of the HoxA genes in the resting pluripotent state in mouse P19 cells to determine a connection between gene expression and chromosomal architecture. The resting state showed extremely low levels of gene expression and although the quantification of RA induction of Hox genes is not absolutely required for our current purposes, the baseline expression level of the HoxA cluster is very relevant. The low levels of HoxA gene expression observed in the pluripotent state of the P19 cells increases the likelihood of finding a highly condensed chromatic architecture of said cluster. Genes of this Homeobox family are generally kept repressed in such structures due to the increased ratio of repressive/condensing epigenetic modifications, such as previously mentioned H3K27/4, which themselves attract other chromatin modifiers to further condense the DNA[336].

The mapping of the spatial organization of the transcriptionally silent mouse HoxA cluster from P19 cells was carried out using the Chromatin Conformation Capture (3C) technology. The P19 cell line is a mouse neuronal differentiation system that has previously been used to examine the regulation of Hox gene clusters [337, 338] and is a well characterized

developmental model system. In general, the most highly interacting fragments appear to contain the major structural elements responsible for orchestrating the overall architecture of the cluster in this specific cellular state, that being the retinoic acid response elements (RAREs).

Overall, considering the high interaction between fragments containing identified RAREs, we propose a model for regulating the collinearity of *Hox* gene expression. The schematic 2-D representation of the observed 3-D conformation for the mouse transcriptionally silent HoxA gene cluster in Figure 10 places the majority of fragments containing a RARE, or elements similar to this feature, within the core of the rosette structure, the most densely interacting region. Due to the high concentration of RAREs and the cofactors that bind to it through association with the RARs, we propose that this microenvironment is responsible for regulating expression and therefore development through a co-operative binding theory. We believe that this region plays the role of central "control hub" in orchestrating collinearity by way of coordinating densely interacting RAREs. This mechanism is similar to that of the beta-globin LCR which seems to nucleate a "hub" or factory that ties the locus in loops and facilitates expression of globin-related genes [339]. Furthermore, other evidence points to factories such as these as specializing in the transcription of specific gene subsets [340].

The attenuation or repression of one of these RARE associated *HoxA* genes may cause a cooperative effect in which the other RAREs are brought within or retained in this central core structure by initial repressive complex binding and subsequent recruitment of more complexes by the increase in RARE concentration. This potential model is similar to that proposed by Lanzuolo et al. in which higher chromatin organization resulted from the interaction of Polycomb response elements (PRE) in the drosophila bithorax complex[341], however, the

resolution they obtained was not as high as in the 3C approach applied here. In general, this model may effectively lead to histone modifications of large linear domains in the cluster and full early cluster repression.

Four distinct loops are formed from the interaction of the gene cluster. Loop 1 contains *HoxA1-4*, loop 2 contains *HoxA 5-7*, loop 3 contains *HoxA 9*, and loop 4 contains *HoxA 10-13* (not including any paralogue genes absent from the *HoxA* cluster) (see Figure 10). The organization of these loops is representative of the timing of expression due to the characteristic collinearity of the cluster. Loop 1 genes are generally basally expressed in undifferentiated P19 cells (Figure 4) [327], and are expressed significantly at 24 hours post RA treatment, indicating that this repressive loop has likely opened up to allow for the appropriate transcriptional machinery to operate. Loop 2 genes are expressed at smaller relative levels to the earlier 3' genes and correspond to the second opening of a chromatin loop. The other loops do not contain identified RAREs and are therefore not incorporated into this mechanism of control, although they may contain regulatory elements that can operate in conjunction with factors involved in RA activation, such as PRC1/2 factors. The interacting fragment of *HoxA10* does not contain an identified RARE and so its regulation with respect to chromatin remodeling is not yet verified, but it would appear that this fragment will not likely release its interactions upon RA treatment with regards to this mechanism .

Along with the description of a possible level of control explained above, the role of RA in causing subsequent release of looping interaction with respect to A-P body patterning requires further characterization. RA is expressed along the anterior-posterior axis of the developing body in a gradient manner and is also observed in this way during limb development [86]. It has been

previously shown that *Hox* genes from embryonal carcinoma cells display differential activation based on RA concentration, with the 3' genes requiring lower concentrations of RA when compared to the 5' genes[342]. Our theory can be used to explain the concentration dependence of RA in *Hox* gene activation [342, 343]. An increase in RA concentration, whether it is observed in cell culture systems or in animal models in which gradients are formed based on body structure, could initiate a cascade of decondensation events based on the loops outlined in Figure 10 and the concentration of RAREs in these areas.

When considering this concentration dependence characteristic along with the distinct loops created by interacting cis-elements in the *HoxA* cluster and there sequential activation, we can hypothesize that increasing concentrations of RA will sequentially release the interactions thereby establishing a collinear mechanism of gene expression. The concentration of RA may possibly cause initial release of the most loosely associated loop, followed by the next and so on.

This may simply be the preliminary step in *Hox* gene activation. Another possibility, which we believe is more likely to function, is based on the number of RAREs involved in each looping contact that could potentially require a distinct RA concentration to be effective in abolishing the repressive interaction. This theory would imply that the contacts of loop 1 have the greatest amount of interacting RAREs followed by loop 2 and then loop 3. A loop containing a high degree of RAREs or other response elements would most likely react earlier to the presence of RA. The Hill coefficient is correlated with the number of active binding sites, and with the way in which dimers bind to them. The Hill coefficient is observed to increase by 20% for three active binding sites when compared to a single active site[344]. When examining the number of identified RAREs in the mouse *Hox A* cluster, we have determined that loop 1

contains 11 CNSs, loop 2 contains 7 CNSs and loop 3 contains 3 CNSs [79, 312, 345, 346] (see Figure 10 and 11). Although not all of the determined CNSs are confirmed RARE, they do share a significant amount of similarity to the consensus sequence of RAREs. Also, RAREs are comprised of two distinct "boxes" or sequences and the elements mentioned above may either have one or both currently identified. Nonetheless, our conclusion remains the same based on this data.

The molecular mechanisms governing collinearity are still poorly understood. At the ultra structural level, spatio-temporal *Hox* induction was shown to involve extensive nuclear reorganization including decondensation and extrusion from the chromosome territory[14, 347]. Although distinct mechanisms appear to regulate clusters in different developing systems, progressive looping out from the chromosome territory was proposed to induce sequential transcription activation along clusters in RA treated mouse ES cells. Our results suggest that chromatin loops in silent clusters may represent the underlying structural mechanism of this process. Thus, looping contacts may be required for collinear *Hox* induction.

Conservation of HoxA spatial organization

We compared the long-range cis interactions of the mouse HoxA cluster from P19 cells to that of the human NT2 cell line. The NT2/D1 cell line is also capable of undergoing neuronal differentiation with RA treatment similar to P19 cells and has been extensively used in studying the regulation of *Hox* genes [348, 349]. This allowed us to qualitatively compare the possible conservation of chromatin conformation signatures (CCSs). The transcriptionally silent NT2/D1 *HoxA* cluster was analyzed by 3C technology (Figure A1) and the results were compared to our data. Interestingly, interactions between *HoxA10* (MF58-59) and *HoxA1* and *2* (MF86), *HoxA5*

(MF70), *HoxA7* (MF63) and *HoxA13* (MF53) were conserved between species. A contact between *HoxA11* (MF56) and *HoxA5* (MF70) were also observed in both species (Compare Figure 10 and Figure A1).

The chromosomal architecture of the human and mouse *HoxA* clusters appear to have similar configurations. This may prove to be vital in terms of *Hox* function; however, further studies must be carried out. Considering the degree of conservation of the *Hox* clusters over evolutionary time, with regard to duplication events [350] as well as sequence similarities among species[351], we would like to propose that 3-dimensional conformation of the clusters is also a conserved element of *Hox* gene function. The identification of these conserved interactions across vertebrate species may indicate that the hypothesized collinear control mechanism described earlier is also conserved.

We have also begun preparation for the examination of the effects of enhancer of zeste homolog 2 (EZH2) knock down in order to verify its role in regulating the *HoxA* chromosomal architecture. EZH2 is just one of the many components of the Polycomb repressive complex 2 (PRC2) and is responsible for altering the methylation status of histones though its intrinsic HMTase activity [169, 352, 353]. Lysine 27 of histone H3 (H3K27) is an undisputed methylation site for both the Drosophila and human versions of the complex. This conclusion is supported by both *in vitro* [168, 352] and *in vivo* [168] evidence. This methylation marking is the preferential binding site of the PcG protein and therefore denotes EZH2 as the initiator of the repressive mechanism involved in Hox gene silencing [168]. Cao et al. also determined that the removal of EZH2 through RNA-mediated interference both *in vivo* and in cultured cells, relieved polycomb binding [168]. We believe that knocking down this gene through RNAi mediated techniques, not

only will relieve the K27H3 methyl groups epigenetic mark but will also affect the overall 3-dimensional landscape of the cluster due to the removal of these silencing modifications. The loss and/or gain of contacts resulting from such a knockdown could reveal more information as to the mechanism involved in regulating Hox collinear expression.

The knock down of EZH2 in the mouse P19 cell line was examined by western blot (Figure A2). A 3C library of these cells has been constructed; however, results examining the loss or gain of certain contacts have yet to be determined.

A technology that has recently been developed to analyze large regions of chromatin conformation using high-throughput methods can also be implemented to verify our findings. 5C technology, developed by Dr. Josée Dostie [269] is similar to 3C technology; however, it utilizes microarray detection and therefore has a much larger degree of analysis. This method can not only be used to prove our findings, but it can be implemented in examining the interaction between Hox clusters. Considering the mechanism of collinearity that affects all Hox clusters, it would not be unlikely to find interactions between clusters.

In conclusion, it would appear that the structural characterization of the mouse HoxA gene cluster that we determined directly enables the mechanism of collinear gene expression through the network of conserved non-coding sequences that enable the specifically required chromosomal structure of 4 distinct loops within this cluster. Considering the degree of conservation between animal species (mouse and human), we must also consider the possibility that this degree of control is conserved. Further characterization of individual contacts within this cluster could reveal the role of specific players in this form of regulation and should be considered to corroborate the evidence presented thus far.

Section 4: Materials and methods

Design and preparation of control 3C libraries. Mouse Hox clusters control 3C libraries were generated from bacterial artificial chromosomes (BACs) as previously described[354]. Libraries were generated from BAC clones covering each of the four Hox clusters and one gene desert region (ENCODE region Chr.19 (28282589-28459924) for mouse control library). Mouse BACs were quantified by real-time PCR with a LightCycler (Roche) in the presence of SYBR Green I stain (Molecular Probes®). Equimolar ratio of BAC clones were mixed, digested with a restriction enzyme and randomly ligated. The mouse control 3C library was generated by digesting the following BACs with EcoRI: RP23-20F21, RP23-196F5, RP23-36P11, RP23-101K1, RP23-450L4. BAC clones were obtained from Invitrogen™.

Cell culture. The P19 cell line are mouse pluripotent embryonal carcinoma cells derived from a C3H/He mouse[355]. These cells were kindly provided by Dr. Mark Featherstone (Nanyang Technological University, Singapore) and grown in Alpha Modified Eagle's Medium (AMEM; Wisent cat. no. 310-010-CL) supplemented with 10% FBS (HyClone®). The cell line was grown at 37°C in 5% CO_2 atmosphere in the presence of 1% penicillin-streptomycin. All experiments presented in this study were performed using log-phase cells.

To induce Hox expression in P19, exponentially growing cells were seeded at 1 X 10^6 per 75 cm^2 flasks in 12 ml of complete DMEM containing 10 μM all trans retinoic acid (ATRA; Sigma cat. no. R2625) or no RA control. Cells were treated continuously with RA to achieve maximal induction and passaged to maintain exponential growth. Cells were collected after 3 and 4 days for RNA extraction.

RNA quantification. Total RNA was extracted from undifferentiated control (Silent) and RA-treated (Induced) P19 cells with the GenElute™ Mammalian Total RNA Miniprep Kit as described by the manufacturer (Sigma®). Reverse transcription was performed with oligo$(dT)_{20}$ (Invitrogen™) using the Omniscript Reverse Transcription Kit (Qiagen®). The primer sequences used to measure mouse HoxA genes and actin controls are summarized in Supplemental Table A1. The basal expression levels all 39 mouse Hox genes were characterized by endpoint PCR in undifferentiated P19 cells to verify very low expression levels. In these experiments, total cDNA was amplified under quantitative real-time PCR conditions and PCR products were detected by ethidium bromide staining on agarose gels.

3C analysis. Human and mouse cellular 3C libraries were generated as previously described[354]. Briefly, exponentially growing cells were fixed in 1% formaldehyde, digested with a restriction enzyme and ligated under diluted conditions to promote intermolecular ligation of cross linked restriction fragments. 3C libraries were purified and titrated by PCR with 3C primers detecting neighboring DNA fragments in a gene desert region (mouse Chr.19 as described above). The quality of cellular 3C libraries was verified systematically by generating compaction profiles in gene desert regions as described previously[356] (quality control). PCRs

were performed manually using amplification conditions described elsewhere[354]. PCR products were resolved on agarose gels containing 0.5 μg/ml ethidium bromide and visualized by UV transillumination (302 nm). Gel documentation and quantification was performed using a ChemiDoc™ XRS system featuring a 12-bit digital camera coupled to the Quantity One® computer software (version 4.6.3; BioRad). Mouse 3C primer sequences are shown in Supplemental Table A1.

RNA interference and western blotting. Mouse EZH2 knockdown was performed in P19 cells by reverse transfection in the presence of 5 nM siRNAs (control or EZH2) using HiPerfect transfection reagent as recommended by the manufacturer (Qiagen). Briefly, 6 X 10^5 cells were plated over siRNA/HiPerfect complexes in 35 mm dishes containing a final volume of 2.3 ml of complete DMEM (0 hour transfection). Cells attached onto plates in the presence of siRNAs and were collected 48 h post-transfection for western blotting. Control siRNA (siGENOME Non-Targeting siRNA #2) was purchased from Dharmacon. Mouse EZH2 siRNA (5‘-UUCAAUGAAAGUGCCAUCC-3‘) was purchased from Qiagen.

For western blot analysis, protein samples were prepared by scraping cells directly in 1X SDS sample buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 7.5% glycerol, 5% beta-mercaptoethanol, 0.04% bromophenol blue). Samples were transferred to eppendorfs, sonicated twice for 15 sec and heated at 95ºC for 5 min. 20 μl of each sample was resolved by SDS-PAGE as previously described (gideon’s papers). Gels were transferred onto 0.45 μM nitrocellulose membrane for 45 min at 100 volts. Immunoblotting was performed with anti-human EZH2 mouse mAb AC22 (Cell Signalling; cat. no. 3147) and anti-actin rabbit polyclonal Ab (Cell Signalling; cat. no.

4967) as recommended by the manufacturer. Horseradish peroxidase-conjugated secondary antibodies were purchased from Jackson ImmunoResearch Laboratories. Signals were visualized by chemiluminescence (Perkin Elmer LAS, Inc.; cat. no. NEL105) followed by autoradiography. Films were scanned with the ChemiDoc™ XRS Imaging system and signals were quantified with the Quantity One® software.

Informatics.

The July 2007 mouse (*Mus musculus*) genome sequence assembly (mm9) used for our mouse 3C experimental design was from the NCBI Build 37 and the Mouse Genome Sequence Consortium.

References

1. Szutorisz, H., N. Dillon, and L. Tora, *The role of enhancers as centres for general transcription factor recruitment.* Trends Biochem Sci, 2005. **30**(11): p. 593-9.
2. Sutter, N.B., et al., *Chromatin insulation by a transcriptional activator.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1105-10.
3. Cremer, T., et al., *Chromosome territories--a functional nuclear landscape.* Curr Opin Cell Biol, 2006. **18**(3): p. 307-16.
4. Spector, D.L., *The dynamics of chromosome organization and gene regulation.* Annu Rev Biochem, 2003. **72**: p. 573-608.
5. Tolhuis, B., et al., *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Mol Cell, 2002. **10**(6): p. 1453-65.
6. Spilianakis, C.G. and R.A. Flavell, *Long-range intrachromosomal interactions in the T helper type 2 cytokine locus.* Nat Immunol, 2004. **5**(10): p. 1017-27.
7. Vakoc, C.R., et al., *Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1.* Mol Cell, 2005. **17**(3): p. 453-62.
8. Dekker, J., *Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction.* J Biol Chem, 2008. **283**(50): p. 34532-40.
9. Dekker, J., *Gene regulation in the third dimension.* Science, 2008. **319**(5871): p. 1793-4.
10. Spilianakis, C.G., et al., *Interchromosomal associations between alternatively expressed loci.* Nature, 2005. **435**(7042): p. 637-45.
11. Lomvardas, S., et al., *Interchromosomal interactions and olfactory receptor choice.* Cell, 2006. **126**(2): p. 403-13.
12. Engel, J.D. and K. Tanimoto, *Looping, linking, and chromatin activity: new insights into beta-globin locus regulation.* Cell, 2000. **100**(5): p. 499-502.
13. Palstra, R.J., et al., *The beta-globin nuclear compartment in development and erythroid differentiation.* Nat Genet, 2003. **35**(2): p. 190-4.
14. Chambeyron, S. and W.A. Bickmore, *Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription.* Genes Dev, 2004. **18**(10): p. 1119-30.
15. Scott, M.P., et al., *The molecular organization of the Antennapedia locus of Drosophila.* Cell, 1983. **35**(3 Pt 2): p. 763-76.
16. McGinnis, W., et al., *A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes.* Nature, 1984. **308**(5958): p. 428-33.
17. Chourrout, D., et al., *Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements.* Nature, 2006. **442**(7103): p. 684-7.
18. Greer, J.M., et al., *Maintenance of functional equivalence during paralogous Hox gene evolution.* Nature, 2000. **403**(6770): p. 661-5.
19. Tvrdik, P. and M.R. Capecchi, *Reversal of Hox1 gene subfunctionalization in the mouse.* Dev Cell, 2006. **11**(2): p. 239-50.
20. Malicki, J., K. Schughart, and W. McGinnis, *Mouse Hox-2.2 specifies thoracic segmental identity in Drosophila embryos and larvae.* Cell, 1990. **63**(5): p. 961-7.
21. McGinnis, N., M.A. Kuziora, and W. McGinnis, *Human Hox-4.2 and Drosophila deformed encode similar regulatory specificities in Drosophila embryos and larvae.* Cell, 1990. **63**(5): p. 969-76.
22. Duboule, D., *The vertebrate limb: a model system to study the Hox/HOM gene network during development and evolution.* Bioessays, 1992. **14**(6): p. 375-84.

23. Patterson, L.T., M. Pembaur, and S.S. Potter, *Hoxa11 and Hoxd11 regulate branching morphogenesis of the ureteric bud in the developing kidney.* Development, 2001. **128**(11): p. 2153-61.
24. Rossel, M. and M.R. Capecchi, *Mice mutant for both Hoxa1 and Hoxb1 show extensive remodeling of the hindbrain and defects in craniofacial development.* Development, 1999. **126**(22): p. 5027-40.
25. Duboule, D., *The rise and fall of Hox gene clusters.* Development, 2007. **134**(14): p. 2549-60.
26. Holland, P., *Homeobox genes in vertebrate evolution.* Bioessays, 1992. **14**(4): p. 267-73.
27. Cameron, R.A., et al., *Unusual gene order and organization of the sea urchin hox cluster.* J Exp Zoolog B Mol Dev Evol, 2006. **306**(1): p. 45-58.
28. D'Esposito, M., et al., *EVX2, a human homeobox gene homologous to the even-skipped segmentation gene, is localized at the 5' end of HOX4 locus on chromosome 2.* Genomics, 1991. **10**(1): p. 43-50.
29. Negre, B. and A. Ruiz, *HOM-C evolution in Drosophila: is there a need for Hox gene clustering?* Trends Genet, 2007. **23**(2): p. 55-9.
30. Seo, H.C., et al., *Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica.* Nature, 2004. **431**(7004): p. 67-71.
31. Duboule, D., *Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony.* Dev Suppl, 1994: p. 135-42.
32. Lemons, D. and W. McGinnis, *Genomic evolution of Hox gene clusters.* Science, 2006. **313**(5795): p. 1918-22.
33. Pearson, J.C., D. Lemons, and W. McGinnis, *Modulating Hox gene functions during animal body patterning.* Nat Rev Genet, 2005. **6**(12): p. 893-904.
34. Mann, R.S. and G. Morata, *The developmental and molecular biology of genes that subdivide the body of Drosophila.* Annu Rev Cell Dev Biol, 2000. **16**: p. 243-71.
35. McGinnis, W. and R. Krumlauf, *Homeobox genes and axial patterning.* Cell, 1992. **68**(2): p. 283-302.
36. Estrada, B., F. Casares, and E. Sanchez-Herrero, *Development of the genitalia in Drosophila melanogaster.* Differentiation, 2003. **71**(6): p. 299-310.
37. Mastick, G.S., et al., *Identification of target genes regulated by homeotic proteins in Drosophila melanogaster through genetic selection of Ultrabithorax protein-binding sites in yeast.* Genetics, 1995. **139**(1): p. 349-63.
38. Mann, R.S. and M. Affolter, *Hox proteins meet more partners.* Curr Opin Genet Dev, 1998. **8**(4): p. 423-9.
39. Mahaffey, J.W., *Assisting Hox proteins in controlling body form: are there new lessons from flies (and mammals)?* Curr Opin Genet Dev, 2005. **15**(4): p. 422-9.
40. Gebelein, B., D.J. McKay, and R.S. Mann, *Direct integration of Hox and segmentation gene inputs during Drosophila development.* Nature, 2004. **431**(7009): p. 653-9.
41. Vachon, G., et al., *Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less.* Cell, 1992. **71**(3): p. 437-50.
42. Graba, Y., et al., *Homeotic control in Drosophila; the scabrous gene is an in vivo target of Ultrabithorax proteins.* EMBO J, 1992. **11**(9): p. 3375-84.
43. Strutt, D.I. and R.A. White, *Characterisation of T48, a target of homeotic gene regulation in Drosophila embryogenesis.* Mech Dev, 1994. **46**(1): p. 27-39.

44. Heuer, J.G., K. Li, and T.C. Kaufman, *The Drosophila homeotic target gene centrosomin (cnn) encodes a novel centrosomal protein with leucine zippers and maps to a genomic region required for midgut morphogenesis.* Development, 1995. **121**(11): p. 3861-76.
45. Wu, K. and D.J. Wolgemuth, *Protein product of the somatic-type transcript of the Hoxa-4 (Hox-1.4) gene binds to homeobox consensus binding sites in its promoter and intron.* J Cell Biochem, 1993. **52**(4): p. 449-62.
46. Sprecher, S.G., et al., *Hox gene cross-regulatory interactions in the embryonic brain of Drosophila.* Mech Dev, 2004. **121**(6): p. 527-36.
47. Hueber, S.D., et al., *Comparative analysis of Hox downstream genes in Drosophila.* Development, 2007. **134**(2): p. 381-92.
48. Walsh, C.M. and S.B. Carroll, *Collaboration between Smads and a Hox protein in target gene repression.* Development, 2007. **134**(20): p. 3585-92.
49. Rauskolb, C., et al., *extradenticle determines segmental identities throughout Drosophila development.* Development, 1995. **121**(11): p. 3663-73.
50. Moens, C.B. and L. Selleri, *Hox cofactors in vertebrate development.* Dev Biol, 2006. **291**(2): p. 193-206.
51. Phelan, M.L., I. Rambaldi, and M.S. Featherstone, *Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif.* Mol Cell Biol, 1995. **15**(8): p. 3989-97.
52. Shen, W.F., et al., *AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins.* Mol Cell Biol, 1997. **17**(11): p. 6448-58.
53. Capellini, T.D., et al., *Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression.* Development, 2006. **133**(11): p. 2263-73.
54. Gebelein, B., et al., *Specificity of Distalless repression and limb primordia development by abdominal Hox proteins.* Dev Cell, 2002. **3**(4): p. 487-98.
55. Dubrulle, J. and O. Pourquie, *Coupling segmentation to axis formation.* Development, 2004. **131**(23): p. 5783-93.
56. Kessel, M. and P. Gruss, *Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid.* Cell, 1991. **67**(1): p. 89-104.
57. de Stanchina, E., et al., *Selection of homeotic proteins for binding to a human DNA replication origin.* J Mol Biol, 2000. **299**(3): p. 667-80.
58. Kumar, S., et al., *Utilization of the same DNA replication origin by human cells of different derivation.* Nucleic Acids Res, 1996. **24**(17): p. 3289-94.
59. Gabellini, D., et al., *Early mitotic degradation of the homeoprotein HOXC10 is potentially linked to cell cycle progression.* EMBO J, 2003. **22**(14): p. 3715-24.
60. Bosco, G., W. Du, and T.L. Orr-Weaver, *DNA replication control through interaction of E2F-RB and the origin recognition complex.* Nat Cell Biol, 2001. **3**(3): p. 289-95.
61. Stagljar, I., U. Hubscher, and A. Barberis, *Activation of DNA replication in yeast by recruitment of the RNA polymerase II transcription complex.* Biol Chem, 1999. **380**(5): p. 525-30.
62. Lewis, E.B., *A gene complex controlling segmentation in Drosophila.* Nature, 1978. **276**(5688): p. 565-70.
63. Gaunt, S.J., *Mouse homeobox gene transcripts occupy different but overlapping domains in embryonic germ layers and organs: a comparison of Hox-3.1 and Hox-1.5.* Development, 1988. **103**(1): p. 135-44.
64. Kmita, M., et al., *Mechanisms of Hox gene colinearity: transposition of the anterior Hoxb1 gene into the posterior HoxD complex.* Genes Dev, 2000. **14**(2): p. 198-211.

65. Deschamps, J. and J. van Nes, *Developmental regulation of the Hox genes during axial morphogenesis in the mouse.* Development, 2005. **132**(13): p. 2931-42.
66. Deschamps, J., et al., *Initiation, establishment and maintenance of Hox gene expression patterns in the mouse.* Int J Dev Biol, 1999. **43**(7): p. 635-50.
67. Kmita, M. and D. Duboule, *Organizing axes in time and space; 25 years of colinear tinkering.* Science, 2003. **301**(5631): p. 331-3.
68. Dubrulle, J., M.J. McGrew, and O. Pourquie, *FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation.* Cell, 2001. **106**(2): p. 219-32.
69. Ikeya, M. and S. Takada, *Wnt-3a is required for somite specification along the anteroposterior axis of the mouse embryo and for regulation of cdx-1 expression.* Mech Dev, 2001. **103**(1-2): p. 27-33.
70. Houle, M., et al., *Retinoic acid regulation of Cdx1: an indirect mechanism for retinoids and vertebral specification.* Mol Cell Biol, 2000. **20**(17): p. 6579-86.
71. Charite, J., et al., *Transducing positional information to the Hox genes: critical interaction of cdx gene products with position-sensitive regulatory elements.* Development, 1998. **125**(22): p. 4349-58.
72. Pownall, M.E., et al., *eFGF, Xcad3 and Hox genes form a molecular pathway that establishes the anteroposterior axis in Xenopus.* Development, 1996. **122**(12): p. 3881-92.
73. Papalopulu, N., et al., *Retinoic acid causes abnormal development and segmental patterning of the anterior hindbrain in Xenopus embryos.* Development, 1991. **113**(4): p. 1145-58.
74. Huang, D., S.W. Chen, and L.J. Gudas, *Analysis of two distinct retinoic acid response elements in the homeobox gene Hoxb1 in transgenic mice.* Dev Dyn, 2002. **223**(3): p. 353-70.
75. Nolte, C., et al., *The role of a retinoic acid response element in establishing the anterior neural expression border of Hoxd4 transgenes.* Mech Dev, 2003. **120**(3): p. 325-35.
76. Huang, D., et al., *A conserved retinoic acid responsive element in the murine Hoxb-1 gene is required for expression in the developing gut.* Development, 1998. **125**(16): p. 3235-46.
77. Packer, A.I., et al., *Expression of the murine Hoxa4 gene requires both autoregulation and a conserved retinoic acid response element.* Development, 1998. **125**(11): p. 1991-8.
78. Zhang, F., et al., *Elements both 5' and 3' to the murine Hoxd4 gene establish anterior borders of expression in mesoderm and neurectoderm.* Mech Dev, 1997. **67**(1): p. 49-58.
79. Langston, A.W., J.R. Thompson, and L.J. Gudas, *Retinoic acid-responsive enhancers located 3' of the Hox A and Hox B homeobox gene clusters. Functional analysis.* J Biol Chem, 1997. **272**(4): p. 2167-75.
80. Langston, A.W. and L.J. Gudas, *Retinoic acid and homeobox gene regulation.* Curr Opin Genet Dev, 1994. **4**(4): p. 550-5.
81. Kastner, P., et al., *Genetic analysis of RXR alpha developmental function: convergence of RXR and RAR signaling pathways in heart and eye morphogenesis.* Cell, 1994. **78**(6): p. 987-1003.
82. Germain, P., et al., *Co-regulator recruitment and the mechanism of retinoic acid receptor synergy.* Nature, 2002. **415**(6868): p. 187-92.
83. Minucci, S. and P.G. Pelicci, *Retinoid receptors in health and disease: co-regulators and the chromatin connection.* Semin Cell Dev Biol, 1999. **10**(2): p. 215-25.
84. Minucci, S., et al., *Oligomerization of RAR and AML1 transcription factors as a novel mechanism of oncogenic activation.* Mol Cell, 2000. **5**(5): p. 811-20.
85. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

86. Kmita, M., et al., *Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs.* Nature, 2002. **420**(6912): p. 145-50.
87. Spitz, F., F. Gonzalez, and D. Duboule, *A global control region defines a chromosomal regulatory landscape containing the HoxD cluster.* Cell, 2003. **113**(3): p. 405-17.
88. Kondo, T. and D. Duboule, *Breaking colinearity in the mouse HoxD complex.* Cell, 1999. **97**(3): p. 407-17.
89. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.
90. Hornstein, E., et al., *The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development.* Nature, 2005. **438**(7068): p. 671-4.
91. Ronshaugen, M., et al., *The Drosophila microRNA iab-4 causes a dominant homeotic transformation of halteres to wings.* Genes Dev, 2005. **19**(24): p. 2947-52.
92. Mainguy, G., et al., *Extensive polycistronism and antisense transcription in the Mammalian Hox clusters.* PLoS ONE, 2007. **2**(4): p. e356.
93. Sessa, L., et al., *Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster.* RNA, 2007. **13**(2): p. 223-39.
94. Milne, T.A., et al., *MLL targets SET domain methyltransferase activity to Hox gene promoters.* Mol Cell, 2002. **10**(5): p. 1107-17.
95. Kawagoe, H., et al., *Expression of HOX genes, HOX cofactors, and MLL in phenotypically and functionally defined subpopulations of leukemic and normal human hematopoietic cells.* Leukemia, 1999. **13**(5): p. 687-98.
96. Golub, T.R., et al., *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.* Science, 1999. **286**(5439): p. 531-7.
97. Dorrance, A.M., et al., *Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations.* J Clin Invest, 2006. **116**(10): p. 2707-16.
98. Horton, S.J., et al., *Continuous MLL-ENL expression is necessary to establish a "Hox Code" and maintain immortalization of hematopoietic progenitor cells.* Cancer Res, 2005. **65**(20): p. 9245-52.
99. Ayton, P.M. and M.L. Cleary, *Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9.* Genes Dev, 2003. **17**(18): p. 2298-307.
100. Milne, T.A., et al., *Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications.* Cancer Res, 2005. **65**(24): p. 11367-74.
101. Nakamura, T., et al., *Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia.* Nat Genet, 1996. **12**(2): p. 154-8.
102. Borrow, J., et al., *The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9.* Nat Genet, 1996. **12**(2): p. 159-67.
103. Moore, M.A., et al., *NUP98 dysregulation in myeloid leukemogenesis.* Ann N Y Acad Sci, 2007. **1106**: p. 114-42.
104. Lawrence, H.J., et al., *Frequent co-expression of the HOXA9 and MEIS1 homeobox genes in human myeloid leukemias.* Leukemia, 1999. **13**(12): p. 1993-9.
105. Thorsteinsdottir, U., et al., *Defining roles for HOX and MEIS1 genes in induction of acute myeloid leukemia.* Mol Cell Biol, 2001. **21**(1): p. 224-34.
106. Cauwelier, B., et al., *Clinical, cytogenetic and molecular characteristics of 14 T-ALL patients carrying the TCRbeta-HOXA rearrangement: a study of the Groupe Francophone de Cytogenetique Hematologique.* Leukemia, 2007. **21**(1): p. 121-8.

107. Soulier, J., et al., *HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL).* Blood, 2005. **106**(1): p. 274-86.
108. Zhang, H., et al., *Global regulation of Hox gene expression in C. elegans by a SAM domain protein.* Dev Cell, 2003. **4**(6): p. 903-15.
109. Abate-Shen, C., *Deregulated homeobox gene expression in cancer: cause or consequence?* Nat Rev Cancer, 2002. **2**(10): p. 777-85.
110. Akarsu, A.N., et al., *A large Turkish kindred with syndactyly type II (synpolydactyly). 2. Homozygous phenotype?* J Med Genet, 1995. **32**(6): p. 435-41.
111. Mortlock, D.P., L.C. Post, and J.W. Innis, *The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads to arrest of digital arch formation.* Nat Genet, 1996. **13**(3): p. 284-9.
112. Mortlock, D.P. and J.W. Innis, *Mutation of HOXA13 in hand-foot-genital syndrome.* Nat Genet, 1997. **15**(2): p. 179-80.
113. Utsch, B., et al., *A novel stable polyalanine [poly(A)] expansion in the HOXA13 gene associated with hand-foot-genital syndrome: proper function of poly(A)-harbouring transcription factors depends on a critical repeat length?* Hum Genet, 2002. **110**(5): p. 488-94.
114. Mollard, R. and M. Dziadek, *Homeobox genes from clusters A and B demonstrate characteristics of temporal colinearity and differential restrictions in spatial expression domains in the branching mouse lung.* Int J Dev Biol, 1997. **41**(5): p. 655-66.
115. Kim, C. and H.C. Nielsen, *Hoxa-5 in mouse developing lung: cell-specific expression and retinoic acid regulation.* Am J Physiol Lung Cell Mol Physiol, 2000. **279**(5): p. L863-71.
116. Volpe, M.V., R.J. Vosatka, and H.C. Nielsen, *Hoxb-5 control of early airway formation during branching morphogenesis in the developing mouse lung.* Biochim Biophys Acta, 2000. **1475**(3): p. 337-45.
117. Golpon, H.A., et al., *HOX genes in human lung: altered expression in primary pulmonary hypertension and emphysema.* Am J Pathol, 2001. **158**(3): p. 955-66.
118. Calvo, R., et al., *Altered HOX and WNT7A expression in human lung cancer.* Proc Natl Acad Sci U S A, 2000. **97**(23): p. 12776-81.
119. Volpe, M.V., et al., *Expression of Hoxb-5 during human lung development and in congenital lung malformations.* Birth Defects Res A Clin Mol Teratol, 2003. **67**(8): p. 550-6.
120. Margueron, R., P. Trojer, and D. Reinberg, *The key to development: interpreting the histone code?* Curr Opin Genet Dev, 2005. **15**(2): p. 163-76.
121. Weintraub, H. and M. Groudine, *Chromosomal subunits in active genes have an altered conformation.* Science, 1976. **193**(4256): p. 848-56.
122. Choo, K.H., *Centromerization.* Trends Cell Biol, 2000. **10**(5): p. 182-8.
123. Howman, E.V., et al., *Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice.* Proc Natl Acad Sci U S A, 2000. **97**(3): p. 1148-53.
124. Shelby, R.D., O. Vafa, and K.F. Sullivan, *Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites.* J Cell Biol, 1997. **136**(3): p. 501-13.
125. Ahmad, K. and S. Henikoff, *The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly.* Mol Cell, 2002. **9**(6): p. 1191-200.
126. Costanzi, C. and J.R. Pehrson, *Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals.* Nature, 1998. **393**(6685): p. 599-601.
127. Pehrson, J.R. and R.N. Fuji, *Evolutionary conservation of histone macroH2A subtypes and domains.* Nucleic Acids Res, 1998. **26**(12): p. 2837-42.
128. Liu, X., B. Li, and GorovskyMa, *Essential and nonessential histone H2A variants in Tetrahymena thermophila.* Mol Cell Biol, 1996. **16**(8): p. 4305-11.

129. van Daal, A. and S.C. Elgin, *A histone variant, H2AvD, is essential in Drosophila melanogaster.* Mol Biol Cell, 1992. **3**(6): p. 593-602.
130. Allis, C.D., et al., *hv1 is an evolutionarily conserved H2A variant that is preferentially associated with active genes.* J Biol Chem, 1986. **261**(4): p. 1941-8.
131. Leach, T.J., et al., *Histone H2A.Z is widely but nonrandomly distributed in chromosomes of Drosophila melanogaster.* J Biol Chem, 2000. **275**(30): p. 23267-72.
132. Rangasamy, D., et al., *Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development.* EMBO J, 2003. **22**(7): p. 1599-607.
133. Parseghian, M.H., et al., *Characterization of a set of antibodies specific for three human histone H1 subtypes.* Chromosoma, 1994. **103**(3): p. 198-208.
134. Zlatanova, J., P. Caiafa, and K. Van Holde, *Linker histone binding and displacement: versatile mechanism for transcriptional regulation.* FASEB J, 2000. **14**(12): p. 1697-704.
135. Wagner, T.E., et al., *Phosphorylation and dephosphorylation of histone (V (H5): controlled condensation of avian erythrocyte chromatin. Appendix: Phosphorylation and dephosphorylation of histone H5. II. Circular dichroic studies.* Biochemistry, 1977. **16**(2): p. 286-90.
136. Talasz, H., et al., *In vitro binding of H1 histone subtypes to nucleosomal organized mouse mammary tumor virus long terminal repeat promotor.* J Biol Chem, 1998. **273**(48): p. 32236-43.
137. Verona, R.I., M.R. Mann, and M.S. Bartolomei, *Genomic imprinting: intricacies of epigenetic regulation in clusters.* Annu Rev Cell Dev Biol, 2003. **19**: p. 237-59.
138. Thorvaldsen, J.L., K.L. Duran, and M.S. Bartolomei, *Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2.* Genes Dev, 1998. **12**(23): p. 3693-702.
139. Engel, N., et al., *Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations.* Nat Genet, 2004. **36**(8): p. 883-8.
140. Hark, A.T., et al., *CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus.* Nature, 2000. **405**(6785): p. 486-9.
141. Bell, A.C., A.G. West, and G. Felsenfeld, *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.* Cell, 1999. **98**(3): p. 387-96.
142. Kurukuti, S., et al., *CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2.* Proc Natl Acad Sci U S A, 2006. **103**(28): p. 10684-9.
143. Yoon, Y.S., et al., *Analysis of the H19ICR insulator.* Mol Cell Biol, 2007. **27**(9): p. 3499-510.
144. Engel, N., et al., *Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking.* Hum Mol Genet, 2008. **17**(19): p. 3021-9.
145. Plath, K., et al., *Xist RNA and the mechanism of X chromosome inactivation.* Annu Rev Genet, 2002. **36**: p. 233-78.
146. Wang, J., et al., *Imprinted X inactivation maintained by a mouse Polycomb group gene.* Nat Genet, 2001. **28**(4): p. 371-5.
147. Kohlmaier, A., et al., *A chromosomal memory triggered by Xist regulates histone methylation in X inactivation.* PLoS Biol, 2004. **2**(7): p. E171.
148. Khalil, A.M. and D.J. Driscoll, *Histone H3 lysine 4 dimethylation is enriched on the inactive sex chromosomes in male meiosis but absent on the inactive X in female somatic cells.* Cytogenet Genome Res, 2006. **112**(1-2): p. 11-5.
149. Heard, E., P. Clerc, and P. Avner, *X-chromosome inactivation in mammals.* Annu Rev Genet, 1997. **31**: p. 571-610.
150. Ehrich, M., *Biochemical and pathological effects of Clostridium difficile toxins in mice.* Toxicon, 1982. **20**(6): p. 983-9.

151. Bird, A.P., *CpG-rich islands and the function of DNA methylation.* Nature, 1986. **321**(6067): p. 209-13.
152. Tamaru, H. and E.U. Selker, *A histone H3 methyltransferase controls DNA methylation in Neurospora crassa.* Nature, 2001. **414**(6861): p. 277-83.
153. Strahl, B.D., et al., *Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression.* Mol Cell Biol, 2002. **22**(5): p. 1298-306.
154. Marmorstein, R., *Structure of SET domain proteins: a new twist on histone methylation.* Trends Biochem Sci, 2003. **28**(2): p. 59-62.
155. Sims, R.J., 3rd, K. Nishioka, and D. Reinberg, *Histone lysine methylation: a signature for chromatin function.* Trends Genet, 2003. **19**(11): p. 629-39.
156. Santos-Rosa, H., et al., *Active genes are tri-methylated at K4 of histone H3.* Nature, 2002. **419**(6905): p. 407-11.
157. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
158. Jenuwein, T., *Re-SET-ting heterochromatin by histone methyltransferases.* Trends Cell Biol, 2001. **11**(6): p. 266-73.
159. Jenuwein, T., et al., *SET domain proteins modulate chromatin domains in eu- and heterochromatin.* Cell Mol Life Sci, 1998. **54**(1): p. 80-93.
160. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-9.
161. Nakayama, J., et al., *Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly.* Science, 2001. **292**(5514): p. 110-3.
162. Bienz, M. and J. Muller, *Transcriptional silencing of homeotic genes in Drosophila.* Bioessays, 1995. **17**(9): p. 775-84.
163. Ringrose, L. and R. Paro, *Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins.* Annu Rev Genet, 2004. **38**: p. 413-43.
164. Fischle, W., Y. Wang, and C.D. Allis, *Histone and chromatin cross-talk.* Curr Opin Cell Biol, 2003. **15**(2): p. 172-83.
165. Cao, C., et al., *Ubiquitination and degradation of the Arg tyrosine kinase is regulated by oxidative stress.* Oncogene, 2005. **24**(15): p. 2433-40.
166. Schwartz, Y.B., et al., *Genome-wide analysis of Polycomb targets in Drosophila melanogaster.* Nat Genet, 2006. **38**(6): p. 700-5.
167. Rastegar, M., et al., *Sequential histone modifications at Hoxd4 regulatory regions distinguish anterior from posterior embryonic compartments.* Mol Cell Biol, 2004. **24**(18): p. 8090-103.
168. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science, 2002. **298**(5595): p. 1039-43.
169. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.* Genes Dev, 2002. **16**(22): p. 2893-905.
170. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
171. Bickmore, W., *Fluorescence in situ hybridization analysis of chromosome and chromatin structure.* Methods Enzymol, 1999. **304**: p. 650-62.
172. Ng, H.H., et al., *Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association.* Genes Dev, 2002. **16**(12): p. 1518-27.
173. van Leeuwen, F., P.R. Gafken, and D.E. Gottschling, *Dot1p modulates silencing in yeast by methylation of the nucleosome core.* Cell, 2002. **109**(6): p. 745-56.

174. Huyen, Y., et al., *Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks.* Nature, 2004. **432**(7015): p. 406-11.
175. Kleinschmidt, M.A., et al., *The protein arginine methyltransferases CARM1 and PRMT1 cooperate in gene regulation.* Nucleic Acids Res, 2008. **36**(10): p. 3202-13.
176. Wang, Y., et al., *Human PAD4 regulates histone arginine methylation levels via demethylimination.* Science, 2004. **306**(5694): p. 279-83.
177. Shi, Y., et al., *Histone demethylation mediated by the nuclear amine oxidase homolog LSD1.* Cell, 2004. **119**(7): p. 941-53.
178. Metzger, E., et al., *LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription.* Nature, 2005. **437**(7057): p. 436-9.
179. Klose, R.J., et al., *The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36.* Nature, 2006. **442**(7100): p. 312-6.
180. Jenuwein, T., *The epigenetic magic of histone lysine methylation.* FEBS J, 2006. **273**(14): p. 3121-35.
181. Agger, K., et al., *UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development.* Nature, 2007. **449**(7163): p. 731-4.
182. De Santa, F., et al., *The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing.* Cell, 2007. **130**(6): p. 1083-94.
183. Issaeva, I., et al., *Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth.* Mol Cell Biol, 2007. **27**(5): p. 1889-903.
184. Dhalluin, C., et al., *Structure and ligand of a histone acetyltransferase bromodomain.* Nature, 1999. **399**(6735): p. 491-6.
185. Zhang, W., et al., *Essential and redundant functions of histone acetylation revealed by mutation of target lysines and loss of the Gcn5p acetyltransferase.* EMBO J, 1998. **17**(11): p. 3155-67.
186. Imoberdorf, R.M., I. Topalidou, and M. Strubin, *A role for gcn5-mediated global histone acetylation in transcriptional regulation.* Mol Cell Biol, 2006. **26**(5): p. 1610-6.
187. Candau, R., et al., *Identification of human proteins functionally conserved with the yeast putative adaptors ADA2 and GCN5.* Mol Cell Biol, 1996. **16**(2): p. 593-602.
188. Sterner, D.E. and S.L. Berger, *Acetylation of histones and transcription-related factors.* Microbiol Mol Biol Rev, 2000. **64**(2): p. 435-59.
189. Puri, P.L., et al., *p300 is required for MyoD-dependent cell cycle arrest and muscle-specific gene transcription.* EMBO J, 1997. **16**(2): p. 369-83.
190. Blanco, J.C., et al., *The histone acetylase PCAF is a nuclear receptor coactivator.* Genes Dev, 1998. **12**(11): p. 1638-51.
191. Xu, L., et al., *Signal-specific co-activator domain requirements for Pit-1 activation.* Nature, 1998. **395**(6699): p. 301-6.
192. Wesoly, J., et al., *Differential contributions of mammalian Rad54 paralogs to recombination, DNA damage repair, and meiosis.* Mol Cell Biol, 2006. **26**(3): p. 976-89.
193. Verdin, E., F. Dequiedt, and H.G. Kasler, *Class II histone deacetylases: versatile regulators.* Trends Genet, 2003. **19**(5): p. 286-93.
194. Grozinger, C.M. and S.L. Schreiber, *Deacetylase enzymes: biological functions and the use of small-molecule inhibitors.* Chem Biol, 2002. **9**(1): p. 3-16.
195. Vaziri, H., et al., *hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase.* Cell, 2001. **107**(2): p. 149-59.
196. Cheung, P., et al., *Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation.* Mol Cell, 2000. **5**(6): p. 905-15.

197. Thomson, S., et al., *The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase.* EMBO J, 1999. **18**(17): p. 4779-93.
198. De Souza, C.P., et al., *Mitotic histone H3 phosphorylation by the NIMA kinase in Aspergillus nidulans.* Cell, 2000. **102**(3): p. 293-302.
199. Downs, J.A., N.F. Lowndes, and S.P. Jackson, *A role for Saccharomyces cerevisiae histone H2A in DNA repair.* Nature, 2000. **408**(6815): p. 1001-4.
200. Hochstrasser, M., *Protein degradation or regulation: Ub the judge.* Cell, 1996. **84**(6): p. 813-5.
201. Osley, M.A., A.B. Fleming, and C.F. Kao, *Histone ubiquitylation and the regulation of transcription.* Results Probl Cell Differ, 2006. **41**: p. 47-75.
202. Minsky, N. and M. Oren, *The RING domain of Mdm2 mediates histone ubiquitylation and transcriptional repression.* Mol Cell, 2004. **16**(4): p. 631-9.
203. Hwang, W.W., et al., *A conserved RING finger protein required for histone H2B monoubiquitination and cell size control.* Mol Cell, 2003. **11**(1): p. 261-6.
204. Lee, J.S., et al., *Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS.* Cell, 2007. **131**(6): p. 1084-96.
205. Henry, K.W., et al., *Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8.* Genes Dev, 2003. **17**(21): p. 2648-63.
206. Emre, N.C., et al., *Maintenance of low histone ubiquitylation by Ubp10 correlates with telomere-proximal Sir2 association and gene silencing.* Mol Cell, 2005. **17**(4): p. 585-94.
207. Shiio, Y. and R.N. Eisenman, *Histone sumoylation is associated with transcriptional repression.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13225-30.
208. Walter, W., et al., *14-3-3 interaction with histone H3 involves a dual modification pattern of phosphoacetylation.* Mol Cell Biol, 2008. **28**(8): p. 2840-9.
209. Daujat, S., et al., *Crosstalk between CARM1 methylation and CBP acetylation on histone H3.* Curr Biol, 2002. **12**(24): p. 2090-7.
210. Cho, Y.W., et al., *PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex.* J Biol Chem, 2007. **282**(28): p. 20395-406.
211. Nakagawa, T., et al., *Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation.* Genes Dev, 2008. **22**(1): p. 37-49.
212. Martin, C. and Y. Zhang, *The diverse functions of histone lysine methylation.* Nat Rev Mol Cell Biol, 2005. **6**(11): p. 838-49.
213. Duncan, E.M., et al., *Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation.* Cell, 2008. **135**(2): p. 284-94.
214. Herman, J.G., et al., *Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers.* Cancer Res, 1995. **55**(20): p. 4525-30.
215. Esteller, M., et al., *DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis.* Hum Mol Genet, 2001. **10**(26): p. 3001-7.
216. Esteller, M., *Epigenetic lesions causing genetic lesions in human cancer: promoter hypermethylation of DNA repair genes.* Eur J Cancer, 2000. **36**(18): p. 2294-300.
217. Lapidus, R.G., et al., *Methylation of estrogen and progesterone receptor gene 5' CpG islands correlates with lack of estrogen and progesterone receptor gene expression in breast tumors.* Clin Cancer Res, 1996. **2**(5): p. 805-10.
218. Chen, R.Z., et al., *DNA hypomethylation leads to elevated mutation rates.* Nature, 1998. **395**(6697): p. 89-93.

219. Jin, B., et al., *DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function.* Hum Mol Genet, 2008. **17**(5): p. 690-709.
220. Alves, G., A. Tatro, and T. Fanning, *Differential methylation of human LINE-1 retrotransposons in malignant cells.* Gene, 1996. **176**(1-2): p. 39-44.
221. Feinberg, A.P., *Imprinting of a genomic domain of 11p15 and loss of imprinting in cancer: an introduction.* Cancer Res, 1999. **59**(7 Suppl): p. 1743s-1746s.
222. Yoo, C.B. and P.A. Jones, *Epigenetic therapy of cancer: past, present and future.* Nat Rev Drug Discov, 2006. **5**(1): p. 37-50.
223. Jones, P.A. and S.M. Taylor, *Cellular differentiation, cytidine analogs and DNA methylation.* Cell, 1980. **20**(1): p. 85-93.
224. Ginder, G.D., M.J. Whitters, and J.K. Pohlman, *Activation of a chicken embryonic globin gene in adult erythroid cells by 5-azacytidine and sodium butyrate.* Proc Natl Acad Sci U S A, 1984. **81**(13): p. 3954-8.
225. Plumb, J.A., et al., *Reversal of drug resistance in human tumor xenografts by 2'-deoxy-5-azacytidine-induced demethylation of the hMLH1 gene promoter.* Cancer Res, 2000. **60**(21): p. 6039-44.
226. Christman, J.K., *5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy.* Oncogene, 2002. **21**(35): p. 5483-95.
227. Yoo, C.B., J.C. Cheng, and P.A. Jones, *Zebularine: a new drug for epigenetic therapy.* Biochem Soc Trans, 2004. **32**(Pt 6): p. 910-2.
228. Cheng, J.C., et al., *Inhibition of DNA methylation and reactivation of silenced genes by zebularine.* J Natl Cancer Inst, 2003. **95**(5): p. 399-409.
229. Brueckner, B., et al., *Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases.* Cancer Res, 2005. **65**(14): p. 6305-11.
230. Cornacchia, E., et al., *Hydralazine and procainamide inhibit T cell DNA methylation and induce autoreactivity.* J Immunol, 1988. **140**(7): p. 2197-200.
231. Scheinbart, L.S., et al., *Procainamide inhibits DNA methyltransferase in a human T cell line.* J Rheumatol, 1991. **18**(4): p. 530-4.
232. Zacharias, J.F., *The new genetics.* J Obstet Gynecol Neonatal Nurs, 1990. **19**(2): p. 122-8.
233. Bigelow, R.L. and J.A. Cardelli, *The green tea catechins, (-)-Epigallocatechin-3-gallate (EGCG) and (-)-Epicatechin-3-gallate (ECG), inhibit HGF/Met signaling in immortalized and tumorigenic breast epithelial cells.* Oncogene, 2006. **25**(13): p. 1922-30.
234. Davis, A.J., et al., *Phase I and pharmacologic study of the human DNA methyltransferase antisense oligodeoxynucleotide MG98 given as a 21-day continuous infusion every 4 weeks.* Invest New Drugs, 2003. **21**(1): p. 85-97.
235. Herman, J.G., et al., *Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands.* Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9821-6.
236. Taipale, M., et al., *hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells.* Mol Cell Biol, 2005. **25**(15): p. 6798-810.
237. Eckner, R., *p300 and CBP as transcriptional regulators and targets of oncogenic events.* Biol Chem, 1996. **377**(11): p. 685-8.
238. Janknecht, R., *The versatile functions of the transcriptional coactivators p300 and CBP and their roles in disease.* Histol Histopathol, 2002. **17**(2): p. 657-68.
239. Livengood, J.A., et al., *p53 Transcriptional activity is mediated through the SRC1-interacting domain of CBP/p300.* J Biol Chem, 2002. **277**(11): p. 9054-61.

240. Jin, Y., et al., *MDM2 inhibits PCAF (p300/CREB-binding protein-associated factor)-mediated p53 acetylation.* J Biol Chem, 2002. **277**(34): p. 30838-43.
241. Legube, G., et al., *Role of the histone acetyl transferase Tip60 in the p53 pathway.* J Biol Chem, 2004. **279**(43): p. 44825-33.
242. Kitabayashi, I., et al., *Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein.* EMBO J, 2001. **20**(24): p. 7184-96.
243. Sobulo, O.M., et al., *MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3).* Proc Natl Acad Sci U S A, 1997. **94**(16): p. 8732-7.
244. Armstrong, S.A., et al., *MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia.* Nat Genet, 2002. **30**(1): p. 41-7.
245. Cebrat, M., et al., *Synthesis and analysis of potential prodrugs of coenzyme A analogues for the inhibition of the histone acetyltransferase p300.* Bioorg Med Chem, 2003. **11**(15): p. 3307-13.
246. Cui, L., et al., *Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro Plasmodium falciparum development.* Eukaryot Cell, 2008. **7**(7): p. 1200-10.
247. Balasubramanyam, K., et al., *Small molecule modulators of histone acetyltransferase p300.* J Biol Chem, 2003. **278**(21): p. 19134-40.
248. Balasubramanyam, K., et al., *Polyisoprenylated benzophenone, garcinol, a natural histone acetyltransferase inhibitor, represses chromatin transcription and alters global gene expression.* J Biol Chem, 2004. **279**(32): p. 33716-26.
249. Chen, Y., et al., *Curcumin, both histone deacetylase and p300/CBP-specific inhibitor, represses the activity of nuclear factor kappa B and Notch 1 in Raji cells.* Basic Clin Pharmacol Toxicol, 2007. **101**(6): p. 427-33.
250. Kamat, A.M., G. Sethi, and B.B. Aggarwal, *Curcumin potentiates the apoptotic effects of chemotherapeutic agents and cytokines through down-regulation of nuclear factor-kappaB and nuclear factor-kappaB-regulated gene products in IFN-alpha-sensitive and IFN-alpha-resistant human bladder cancer cells.* Mol Cancer Ther, 2007. **6**(3): p. 1022-30.
251. Biel, M., et al., *Design, synthesis, and biological evaluation of a small-molecule inhibitor of the histone acetyltransferase Gcn5.* Angew Chem Int Ed Engl, 2004. **43**(30): p. 3974-6.
252. Costi, R., et al., *Cinnamoyl compounds as simple molecules that inhibit p300 histone acetyltransferase.* J Med Chem, 2007. **50**(8): p. 1973-7.
253. Boyle, G.M., A.C. Martyn, and P.G. Parsons, *Histone deacetylase inhibitors and malignant melanoma.* Pigment Cell Res, 2005. **18**(3): p. 160-6.
254. Bi, G. and G. Jiang, *The molecular mechanism of HDAC inhibitors in anticancer effects.* Cell Mol Immunol, 2006. **3**(4): p. 285-90.
255. Fraga, M.F., et al., *Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer.* Nat Genet, 2005. **37**(4): p. 391-400.
256. Faretta, M., L. Di Croce, and P.G. Pelicci, *Effects of the acute myeloid leukemia--associated fusion proteins on nuclear architecture.* Semin Hematol, 2001. **38**(1): p. 42-53.
257. Finnin, M.S., et al., *Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors.* Nature, 1999. **401**(6749): p. 188-93.
258. Yoshida, M., et al., *Structural specificity for biological activity of trichostatin A, a specific inhibitor of mammalian cell cycle with potent differentiation-inducing activity in Friend leukemia cells.* J Antibiot (Tokyo), 1990. **43**(9): p. 1101-6.
259. Cameron, E.E., et al., *Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer.* Nat Genet, 1999. **21**(1): p. 103-7.

260. Kelly, W.K., et al., *Phase I study of an oral histone deacetylase inhibitor, suberoylanilide hydroxamic acid, in patients with advanced cancer.* J Clin Oncol, 2005. **23**(17): p. 3923-31.
261. Qiu, L., et al., *Histone deacetylase inhibitors trigger a G2 checkpoint in normal cells that is defective in tumor cells.* Mol Biol Cell, 2000. **11**(6): p. 2069-83.
262. Cremer, M., et al., *Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells.* Chromosome Res, 2001. **9**(7): p. 541-67.
263. Pienta, K.J. and D.S. Coffey, *A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosome.* J Cell Sci Suppl, 1984. **1**: p. 123-35.
264. Trask, B.J., et al., *Mapping of human chromosome Xq28 by two-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei.* Am J Hum Genet, 1991. **48**(1): p. 1-15.
265. Warrington, J.A. and U. Bengtsson, *High-resolution physical mapping of human 5q31-q33 using three methods: radiation hybrid mapping, interphase fluorescence in situ hybridization, and pulsed-field gel electrophoresis.* Genomics, 1994. **24**(2): p. 395-8.
266. Sachs, R.K., et al., *A random-walk/giant-loop model for interphase chromosomes.* Proc Natl Acad Sci U S A, 1995. **92**(7): p. 2710-4.
267. Banerji, J., S. Rusconi, and W. Schaffner, *Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences.* Cell, 1981. **27**(2 Pt 1): p. 299-308.
268. Dekker, J., et al., *Capturing chromosome conformation.* Science, 2002. **295**(5558): p. 1306-11.
269. Dostie, J., et al., *Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements.* Genome Res, 2006. **16**(10): p. 1299-309.
270. Simonis, M., J. Kooren, and W. de Laat, *An evaluation of 3C-based methods to capture DNA interactions.* Nat Methods, 2007. **4**(11): p. 895-901.
271. Stamatoyannopoulos, J.A., et al., *NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region.* EMBO J, 1995. **14**(1): p. 106-16.
272. Keys, J.R., et al., *A mechanism for Ikaros regulation of human globin gene switching.* Br J Haematol, 2008. **141**(3): p. 398-406.
273. Zhou, G.L., et al., *Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes.* Mol Cell Biol, 2006. **26**(13): p. 5096-105.
274. Lee, G.R., et al., *Regulation of the Th2 cytokine locus by a locus control region.* Immunity, 2003. **19**(1): p. 145-53.
275. Bacher, C.P., et al., *Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation.* Nat Cell Biol, 2006. **8**(3): p. 293-9.
276. Spilianakis, C.G., G.R. Lee, and R.A. Flavell, *Twisting the Th1/Th2 immune response via the retinoid X receptor: lessons from a genetic approach.* Eur J Immunol, 2005. **35**(12): p. 3400-4.
277. Bantignies, F., et al., *Inheritance of Polycomb-dependent chromosomal interactions in Drosophila.* Genes Dev, 2003. **17**(19): p. 2406-20.
278. Hakim, O., et al., *Glucocorticoid receptor activation of the CIZ1-LCN2 locus by long range interactions.* J Biol Chem, 2009.
279. Xu, M. and P.R. Cook, *Similar active genes cluster in specialized transcription factories.* J Cell Biol, 2008. **181**(4): p. 615-23.
280. Landegren, U., et al., *A ligase-mediated gene detection technique.* Science, 1988. **241**(4869): p. 1077-80.
281. Frith, M.C., M. Pheasant, and J.S. Mattick, *The amazing complexity of the human transcriptome.* Eur J Hum Genet, 2005. **13**(8): p. 894-7.

282. Bejerano, G., et al., *Ultraconserved elements in the human genome.* Science, 2004. **304**(5675): p. 1321-5.
283. Glazko, G.V., et al., *A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions.* Trends Genet, 2003. **19**(3): p. 119-24.
284. Plessy, C., et al., *Enhancer sequence conservation between vertebrates is favoured in developmental regulator genes.* Trends Genet, 2005. **21**(4): p. 207-10.
285. Gumucio, D.L., et al., *Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the beta-like globin genes.* Mol Phylogenet Evol, 1996. **5**(1): p. 18-32.
286. Loots, G.G. and I. Ovcharenko, *Mulan: multiple-sequence alignment to predict functional elements in genomic sequences.* Methods Mol Biol, 2007. **395**: p. 237-54.
287. Davidson, E.H., D.R. McClay, and L. Hood, *Regulatory gene networks and the properties of the developmental process.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1475-80.
288. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
289. Aparicio, S., et al., *Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.* Science, 2002. **297**(5585): p. 1301-10.
290. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.
291. Wasserman, W.W. and A. Sandelin, *Applied bioinformatics for the identification of regulatory elements.* Nat Rev Genet, 2004. **5**(4): p. 276-87.
292. Hardison, R.C., J. Oeltjen, and W. Miller, *Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome.* Genome Res, 1997. **7**(10): p. 959-66.
293. Dermitzakis, E.T., et al., *Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs).* Science, 2003. **302**(5647): p. 1033-5.
294. Tautz, D., *Evolution of transcriptional regulation.* Curr Opin Genet Dev, 2000. **10**(5): p. 575-9.
295. Thomas, J.W., et al., *Comparative analyses of multi-species sequences from targeted genomic regions.* Nature, 2003. **424**(6950): p. 788-93.
296. Marshall, H., et al., *A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1.* Nature, 1994. **370**(6490): p. 567-71.
297. Rowitch, D.H., et al., *Identification of an evolutionarily conserved 110 base-pair cis-acting regulatory sequence that governs Wnt-1 expression in the murine neural plate.* Development, 1998. **125**(14): p. 2735-46.
298. Bagheri-Fam, S., et al., *Comparative genomics of the SOX9 region in human and Fugu rubripes: conservation of short regulatory sequence elements within large intergenic regions.* Genomics, 2001. **78**(1-2): p. 73-82.
299. Kumar, S. and S.B. Hedges, *A molecular timescale for vertebrate evolution.* Nature, 1998. **392**(6679): p. 917-20.
300. Santini, S., J.L. Boore, and A. Meyer, *Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters.* Genome Res, 2003. **13**(6A): p. 1111-22.
301. Haerry, T.E. and W.J. Gehring, *Intron of the mouse Hoxa-7 gene contains conserved homeodomain binding sites that can function as an enhancer element in Drosophila.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13884-9.
302. Knittel, T., et al., *A conserved enhancer of the human and murine Hoxa-7 gene specifies the anterior boundary of expression during embryonal development.* Development, 1995. **121**(4): p. 1077-88.

303. Odenwald, W.F., et al., *The Hox-1.3 homeo box protein is a sequence-specific DNA-binding phosphoprotein.* Genes Dev, 1989. **3**(2): p. 158-72.
304. Doerksen, L.F., et al., *Functional interaction between a RARE and an AP-2 binding site in the regulation of the human HOX A4 gene promoter.* Nucleic Acids Res, 1996. **24**(14): p. 2849-56.
305. Rossi, P., et al., *A nuclear factor 1 binding site mediates the transcriptional activation of a type I collagen promoter by transforming growth factor-beta.* Cell, 1988. **52**(3): p. 405-14.
306. Ekker, S.C., et al., *The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins.* EMBO J, 1994. **13**(15): p. 3551-60.
307. Margalit, Y., et al., *Isolation and characterization of target sequences of the chicken CdxA homeobox gene.* Nucleic Acids Res, 1993. **21**(21): p. 4915-22.
308. Catron, K.M., N. Iler, and C. Abate, *Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins.* Mol Cell Biol, 1993. **13**(4): p. 2354-65.
309. Nonchev, S., et al., *Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20.* Development, 1996. **122**(2): p. 543-54.
310. Vesque, C., et al., *Hoxb-2 transcriptional activation in rhombomeres 3 and 5 requires an evolutionarily conserved cis-acting element in addition to the Krox-20 binding site.* EMBO J, 1996. **15**(19): p. 5383-96.
311. Maconochie, M., et al., *Paralogous Hox genes: function and regulation.* Annu Rev Genet, 1996. **30**: p. 529-56.
312. Frasch, M., X. Chen, and T. Lufkin, *Evolutionary-conserved enhancers direct region-specific expression of the murine Hoxa-1 and Hoxa-2 loci in both mice and Drosophila.* Development, 1995. **121**(4): p. 957-74.
313. Kim, C.B., et al., *Hox cluster genomics in the horn shark, Heterodontus francisci.* Proc Natl Acad Sci U S A, 2000. **97**(4): p. 1655-60.
314. Nobrega, M.A., et al., *Scanning human gene deserts for long-range enhancers.* Science, 2003. **302**(5644): p. 413.
315. Dickmeis, T., et al., *Expression profiling and comparative genomics identify a conserved regulatory region controlling midline expression in the zebrafish embryo.* Genome Res, 2004. **14**(2): p. 228-38.
316. Masters, J.R., *HeLa cells 50 years on: the good, the bad and the ugly.* Nat Rev Cancer, 2002. **2**(4): p. 315-9.
317. Martin, G.R. and M.J. Evans, *The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture.* Cell, 1974. **2**(3): p. 163-72.
318. Andrews, P.W., *Teratocarcinomas and human embryology: pluripotent human EC cell lines. Review article.* APMIS, 1998. **106**(1): p. 158-67; discussion 167-8.
319. McBurney, M.W., *P19 embryonal carcinoma cells.* Int J Dev Biol, 1993. **37**(1): p. 135-40.
320. McBurney, M.W. and B.J. Rogers, *Isolation of male embryonal carcinoma cells and their chromosome replication patterns.* Dev Biol, 1982. **89**(2): p. 503-8.
321. Rossant, J. and M.W. McBurney, *The developmental potential of a euploid male teratocarcinoma cell line after blastocyst injection.* J Embryol Exp Morphol, 1982. **70**: p. 99-112.
322. Martin, G.R. and M.J. Evans, *Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro.* Proc Natl Acad Sci U S A, 1975. **72**(4): p. 1441-5.
323. Jones-Villeneuve, E.M., et al., *Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells.* J Cell Biol, 1982. **94**(2): p. 253-62.
324. Morassutti, D.J., et al., *Murine embryonal carcinoma-derived neurons survive and mature following transplantation into adult rat striatum.* Neuroscience, 1994. **58**(4): p. 753-63.

325. Sharma, S. and M.F. Notter, *Characterization of neurotransmitter phenotype during neuronal differentiation of embryonal carcinoma cells.* Dev Biol, 1988. **125**(2): p. 246-54.
326. Itoh, F., T. Nakane, and S. Chiba, *Gene expression of MASH-1, MATH-1, neuroD and NSCL-2, basic helix-loop-helix proteins, during neural differentiation in P19 embryonal carcinoma cells.* Tohoku J Exp Med, 1997. **182**(4): p. 327-36.
327. Reese, D.H. and M. Ramos-Valle, *A high-throughput method for monitoring changes in homeobox gene expression.* Biochem Biophys Res Commun, 2007. **357**(4): p. 882-8.
328. Jonk, L.J., et al., *Isolation and developmental expression of retinoic-acid-induced genes.* Dev Biol, 1994. **161**(2): p. 604-14.
329. Pratt, M.A., et al., *Retinoic acid fails to induce expression of Hox genes in differentiation-defective murine embryonal carcinoma cells carrying a mutant gene for alpha retinoic acid receptor.* Differentiation, 1993. **53**(2): p. 105-13.
330. Baron, A., et al., *Hox-1.6: a mouse homeo-box-containing gene member of the Hox-1 complex.* EMBO J, 1987. **6**(10): p. 2977-86.
331. Parikh, H., et al., *Organization, sequence and regulation of expression of the murine Hoxa-7 gene.* Gene, 1995. **154**(2): p. 237-42.
332. Galliot, B., et al., *The mouse Hox-1.4 gene: primary structure, evidence for promoter activity and expression during development.* Development, 1989. **107**(2): p. 343-59.
333. Gillespie, R.F. and L.J. Gudas, *Retinoic acid receptor isotype specificity in F9 teratocarcinoma stem cells results from the differential recruitment of coregulators to retinoic response elements.* J Biol Chem, 2007. **282**(46): p. 33421-34.
334. Gillespie, R.F. and L.J. Gudas, *Retinoid regulated association of transcriptional co-regulators and the polycomb group protein SUZ12 with the retinoic acid response elements of Hoxa1, RARbeta(2), and Cyp26A1 in F9 embryonal carcinoma cells.* J Mol Biol, 2007. **372**(2): p. 298-316.
335. Tan, D.P., et al., *Murine Hox-1.11 homeobox gene structure and expression.* Proc Natl Acad Sci U S A, 1992. **89**(14): p. 6280-4.
336. Wang, L., et al., *Hierarchical recruitment of polycomb group silencing complexes.* Mol Cell, 2004. **14**(5): p. 637-46.
337. Kondo, T., N. Takahashi, and M. Muramatsu, *The regulation of the murine Hox-2.5 gene expression during cell differentiation.* Nucleic Acids Res, 1992. **20**(21): p. 5729-35.
338. Zwartkruis, F., et al., *The murine Hox-2.4 promoter contains a functional octamer motif.* Nucleic Acids Res, 1992. **20**(7): p. 1599-606.
339. Chakalova, L., et al., *Developmental regulation of the beta-globin gene locus.* Prog Mol Subcell Biol, 2005. **38**: p. 183-206.
340. Bartlett, J., et al., *Specialized transcription factories.* Biochem Soc Symp, 2006(73): p. 67-75.
341. Lanzuolo, C., et al., *Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex.* Nat Cell Biol, 2007. **9**(10): p. 1167-74.
342. Simeone, A., et al., *Differential regulation by retinoic acid of the homeobox genes of the four HOX loci in human embryonal carcinoma cells.* Mech Dev, 1991. **33**(3): p. 215-27.
343. Boncinelli, E., et al., *HOX gene activation by retinoic acid.* Trends Genet, 1991. **7**(10): p. 329-34.
344. Zeiser, S., et al., *Number of active transcription factor binding sites is essential for the Hes7 oscillator.* Theor Biol Med Model, 2006. **3**: p. 11.
345. Dupe, V., et al., *In vivo functional analysis of the Hoxa-1 3' retinoic acid response element (3'RARE).* Development, 1997. **124**(2): p. 399-410.
346. Oosterveen, T., et al., *The direct context of a hox retinoic acid response element is crucial for its activity.* J Biol Chem, 2003. **278**(26): p. 24103-7.

347. Morey, C., et al., *Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation.* Development, 2007.
348. Faiella, A., et al., *Inhibition of retinoic acid-induced activation of 3' human HOXB genes by antisense oligonucleotides affects sequential activation of genes located upstream in the four HOX clusters.* Proc Natl Acad Sci U S A, 1994. **91**(12): p. 5335-9.
349. Moroni, M.C., M.A. Vigano, and F. Mavilio, *Regulation of the human HOXD4 gene by retinoids.* Mech Dev, 1993. **44**(2-3): p. 139-54.
350. Hurley, I., M.E. Hale, and V.E. Prince, *Duplication events and the evolution of segmental identity.* Evol Dev, 2005. **7**(6): p. 556-67.
351. Gaunt, S.J., *Conservation in the Hox code during morphological evolution.* Int J Dev Biol, 1994. **38**(3): p. 549-52.
352. Czermin, B., et al., *Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites.* Cell, 2002. **111**(2): p. 185-96.
353. Muller, J., et al., *Histone methyltransferase activity of a Drosophila Polycomb group repressor complex.* Cell, 2002. **111**(2): p. 197-208.
354. Dostie, J. and J. Dekker, *Mapping networks of physical interactions between genomic elements using 5C technology.* Nat Protoc, 2007. **2**(4): p. 988-1002.
355. McBurney, M.W., et al., *Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line.* Nature, 1982. **299**(5879): p. 165-7.
356. Miele, A., et al., *Mapping chromatin interactions by Chromosome Conformation Capture (3C)*, in *Current Protocols in Molecular Biology*, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, Editor. 2006, John Wiley & Sons: Hoboken, N.J. p. 21.11.1-21.11-20.

Appendix

Table A1. Primer List

Mouse HoxA 3C Primers

MF	Primer sequence
1	GAGAGATGATACCAACCCTGAGATTTGC
2	CTGGGCTTTCTTCAGCACTTAACAGTTCG
4	CCAAACCAGACTAAAAGTGCTAACCTTGG
5	CCAAGCCTCATAGTCTGAGTTCAAACACC
6	AGCCAAGAAGAGGATGTTGTCTTGAAAGG
7	GTTCTTCACATAATCTAGGCCAGGGAGAGG
8	GATCTGGTAGGAAGGACCTTGAAGACTAGC
9	GTTCAGTCCTTGGAAACTTGTAGTGGAAGG
10	CTGGAAACCTTGAGAGGTAGATCTAGTTGG
11	CCTGTGAGAGATGAAGAACGTGTGTTTAGG
12	TTTTGCTCTCTTTGGGTCAAGGGTCTGAGG
13	CATTTAAGGTGAGCAGCAATTGAGGAAGG
19	CTTTAACCTTGTGGGGAGGTTTAATATGC
20	CTTGTATGTTTCAGCGTCTCCCAGAAAGC
21	GTAGGACCAAGTACCCAAGGGTAGTACAGC
24	TAATCTTTTGGATACCATACCCCCAACC
25	ACCTGTGCAGCTATTAGGAGGTAGTGTCC
27	CTATCTCCTGCACTAGTAACGCAAGAACC
28	GGGGTTTATAGTTGGGATTTGATTCTCC
30	CTAGGCAACCCAAACCTTCCTAGCTTCC
31	CTAAGGCATACTGGGTCTAAAGGCTTCC
32	AGAAACGTCTGCCACAGTGAAGATGAAAGG
33	GCTTATCTCCGGGCTATTTCAGGTATGTCG
35	AACCCTTCGAAAGCAAAGAGAGGTGACAGC
36	CCAAGGTTTGGATAGGTCAGGAGACATAGG
37	AGTGAGGCACAAAGACATTAGCCCACTTGC
38	GGTGAGATCTGAGTTGTCATCTCTGTGTAGC

MF	Primer sequence
39	CTATTCTGTGCTCCTCGCTGTGACTTTTCC
41	AGAGTGCTTGTGAAATATGCATGAGACC
42	GACAGGTTCTACTGTTGCCTGATTCTGC
44	GGGTTATAAAGAGAGACGTGACAATAACAACC
46	CCAGGACCCATACTAGGTATCTTATTACC
47	TTTCCAGCACTTCCAGCTGGCACATTATCC
48	AAGTGGCTGAGAGCAAGTGATTGTTAGC
50	CAGTGTGTACTTGAAGGGGTCTTGAGTTTCC
52	GCCCTTGATACAAGCCTAGTAGAAAATCC
53	CATGGTGGATTAACCAACACTGTCAAAGG
54	TTGGTGAGACTCTTAATGCCCTTCTACC
55	CTCTAAGGCTCCAGCCTACTGGAATTGG
56	GTTGGCATTTTCCAACAAAATCCCCTAGCC
57	GAGAATGTGTGTTATCACCAGGAAGAAAACC
58	GTCATGCCTTTTTGGAGGGATTTCTTGC
59	GCCTTCATTACTTCAGTTGAAAGCATCTCC
60	GGAAGCTGGTCAGGTCCGTGTTTTCTGG
61	CTGGCTGTATAGTTTTTAGTCTCTGTAGC
62	GTTGTTGCAGACAGAAAGGACTTCAAAGC
63	GGAAAGTCACAGGAAGGTCTTGTGGATGG
65	GCCTTCCTCTTTGTAGATGGTTTTTGTCTGG
67	TACTAATTCAGAGTTCTCCGATGGTTGC
68	ACTTTTCTCCCGAGCAGCAGTACAAACC
69	GGTTGTGATGTGAGAAGCTGGAGTGAAGC
70	GACAATATAGGTGGCCCAGAAGGCAAAAGG
71	TAAAACAAAATCCACCCCATCCGCTCTGC
72	CTGAGGGAATAGCAGCAGGTAAATAGGC

Mouse HoxA 3C Primers cont.

MF	Primer sequence
73	CTACTTCTTCCTCTTTGGCCTTCTCTTTCG
74	AGACAGGGTGAGCTGGGGTCACATTCTAAGC
75	GTAAAGTGAACTTTCCCACCACACTCTCC
76	CACTTATCCGATTTAGTCACAGAGCCAAGC
77	AGGCCATAAAGAAGTCTGCTGACTGATCC
78	GCTTAGTTTGAGCTCCCACCAGTGTTTTTGG
79	ACCTGGAGGTTATTGTCCAGGAACAGAGC
80	TAGGAAGGTGATACGGTTACTGTGAGAGG
85	AAAGCCCAAATTGTCATTGGGCAGAAGC
86	GGGATTTCTTACCCAAAGCACATTCTTAGC
87	ATAGGGGTTTATTAAGCTTCCCTGATGG
90	CCTAAACTTAAGAGATTCTGGGGCGCAACC
92	AGGAATTTTCTCACTGGGCCATAAAGTCAAGC
94	GTCTATAATTTTTCCTTAGGGCCGATAACAGG
96	CCTCTTTTGTTTCAAAGACCCAGGTGAGAGG
98	GTCAGAGCTGGTCCATTGTGTTCAGAGG
100	GCTGCCATTATTTGCACAAACAACCTGTCC
101	AGAGCTAACCTACAAGCTACCCTCTGACC
102	GTAAGATGCTTAGTAGGCCAACCACACACTGG
103	GGTTTCTCTTCTGCTCTCACTGATGTGG
104	GCTCCATTCCATTAAGAACAGGGAGAGTGC
105	TATAGCTCCAGATGCCAGGTAACAACAGC
107	ATAGAGCAGACGATCAGGTTTCCTTTCTCC
110	CCCTGCATTCCCTAACACTTGTACTTGAACC
111	CCTTTCTTGTCTCACTCCTATCTCTCACC
115	GGAAGGAATAGGTATGTGCTTGGAGAGC
116	TGAACTTGGCTCCCCACGTCTTAGTTGC
117	GAATTGCAAACAAGATGTCCCTTTGAGTGG

MF	Primer sequence
118	CACTCTCTTCTAGACTTAATTGCATGTGTGG
119	CTGGGAATTTACTAAAGATGGGCCTTCC
120	TCAGGAGGAGTGAGGAATGGGAAAGAACACG
121	CAGAAATGCTTGCCGTGTAATCTTGACAGC
122	AAAAGATAGGCTTCTGATCCCAGCTCATCG
123	AGACTGATGAGTGACAGAAGGTGACAAGG
126	CCCTACACAGTGAGTGCTGAGATTAAAGG
127	AGGGATAATCCTTCCTAACACAGACTGC
128	CTTTCTCCTTTAGGCAGATTCAGGTCAGC
129	GGTATGTGTTTAGACCTTCAAGATGAGAGTGC
130	CTGTTTTAGAACTTGTTCAGGTGTGTGG
131	GATGCCAAAGGATGTACATGTCCCTTTACC
133	GGCAGGCACTAGCAGTATCTGAATGTTGATGG
134	CACAGTGCAATGTTTGTGCACTGGTGTAAGC
135	GCACTGAGATTTGTCTGTAAGCTATTCC
137	CAAGATCTGGCCACTTAAAGTTCATAACC
138	AGAGGACCAGCATACTTTGTGTGAGTGACC
139	ACAGCAGACAGACATTTCCCAGCATCATCC
140	GGGGAGAAGAGCTAAATCAGATGTCACTGG
142	ACTACATACATGCCTGGTGCTGAAGTCAGAGG
143	GAGACCTGCCAATTTATAGCTGAAGAGC
144	CTCGCAAGGTTTGTTGAGTGTATTGTAGCC
145	CATATGTAGCAGAGGGATGCATGTCTGG
146	CATTAGGACATGCCAGTGAGAGATAGACC
147	AAAAACCTAGTGAATGTGTTAGCAGGAGTGC
150	GAGAGTGGTTGCCAAGGTATAGAGAAAGG
151	AGCTCAAAGGCGTAGTGTTCCTGTAGAAGG
152	CCTACCTATCATTTCTCCTTTCAGTTATCC

Mouse HoxB 3C Primers

MF	Primer Sequence
2	CATAGTAGAACCCAGTGTCTGCATGCTTGC
3	ACAAACAACGGTTTCCAGGTAGTGGTGGCG
5	ACCACTAGCTCTTACTAGGCACAACACTGC
9	ATTGCCAAGAAGAAACTCGCAGAGGTGAGC
10	GTACTGACATTTCGCCAGGCAGAGGTGG
12	GTTATAGTCTGTTGTAAGCTGCCTCCTTGG
13	AAAATTTGGTCATCTGGGGCCAGTGAGG
14	GTGAGGCTACCTGGTTACTTAGTGACCC
15	CCATGTAACACACACTCAAGTACAAGTAGC
17	AAGACAACTACTGAAACTGCCGTGTAGG
18	GAGACAGGAGGTTAAAGAACATCAACTTGG
19	TATTGGGAGTGGGGAGAACTCTAAAACG
20	CCTTGAGTCAGTATTAACACTTCTAGGAACC
21	GAGACAGAGATAGGAAAGTCACTGCAAGG
22	TGATTTCCAGAGCCTGTAACGTCATGGC
24	TCTTAACTGCTGAGGCCTCTTTCCAAACC
27	CCAACCTCTATTGTCCCTAATCTTCTTACGC
29	CCCTCCTGAGCACAGTGGAATCTTAGACC
30	GTCTTCCACCCTACCTGCAAGGTCAAGG
31	TGGTGTTAGGGGTGTGCTGAAAGAAAGC
32	CTCAAATGCAGACACCCTTCTGTCTCTACC
33	ACTGTGGGTAATCGGAGCACTCCAAGACG
34	GTCTTTCTGCACAAACCTAGGCAAAAGG
35	AAGAAAGAAAGAAGCAAGCTGGGCGTGG
36	CCCAGCTCCTTATCTTCCTCTTAACTCC
37	GCCTGGACTACTTGAAATCCTGTACTTTGC
39	TGGAATAACCCCGTGTCTCCATGTATGC

MF	Primer Sequence
40	GTGTGAGTATGTGTGATTACCCGCTTGC
41	TGTCACACAAGAGCAGACAGATTTTGAC
42	AATATACCTAGGGACATGCAAACGCGCC
43	TGGAAATGAAGAGCTACCAGGTACGTCC
44	ATTTTTACTGGCTGTCCTGGAACTCACC
45	CAGGAGCATCCAGATTTCACAGTCATTG
46	GTGCATGTGTTTGTGTGTATGTGTAGTG
47	TAGAGATAGGAGAGTCGAGAGGGAGAAAGG
48	AGAGGAAGAGATTTCTGAGAGGCTTCGC
49	CCTCCTTTTTCTCCGTCAGTCCTGATTAGC
50	CCCATGGCTTTAGATCTCTTCCAGGTTAGC
51	CCTTACCGTTTCTGTCTACGAACCTAGGCC
53	GCTGACTGACAATGCCTTTCCAGAATACC
54	CTATGTATTTTAGCCAGATGTGCTGGCG
55	GGAGTTAAGCAACTTGAAGCAGGAGTGAGG
56	AAGTTCGTGGTAATCCTGCCTTAGCTTCC
57	CGTACAGTACAGTAAATAACCGTGCTTGG
59	CTGAGACCCAATGGACTGTAAAACTATGC
60	ATGGACAACACAAGATCAAGGGGAGTGG
61	TGTCTGGCGGGTATGTAAAGGAGCCTGC
62	CCGGTAGACATTTGCCATTTCAGGAAACC
63	AGGGATTCTGCCTCACCCCAGAGTTAGG
64	GTGCGTCTTTCCCAGAGATCTGCTGAGAGC
66	GCAGAAGTCTGAAATGAGCATTGAGTTCC
67	TTGGTCTTAACTGGCTGTTCAGGTGTAGGC
68	AAGATGACAGAACCGTACACACAGAACC
69	ACTCCACTGGTACCTGTACTCAGCATGCGC

Mouse HoxB 3C Primers cont.

MF	Primer Sequence
70	AGCTTCATGTCATAGACCACTGGCAACTGG
74	GGAAGAAAGAAACTGTGAAATCCGGAGAGG
75	TGGCATGAGCTTCGATCACCGACTAAGG
76	GCTATATGTGAGTGCGGGTTTTTGTAGAAGG
77	ACACCAGTTGTACTGTGTTCCTCCTCAACG
78	GGAGGCAATTAACTGGCTAATTCCAAGACC
79	GAGACAAGGATGTACTCAGACTAGAGGAAGC
80	TACTGCTTCTTGCTTAGCTGGGGGAGTTCG
82	CCCAGCTTGAAAAACAGAACAACTCTTATGG
83	CTCTCAGAACATTCCCTTGTGGAGGACTGG
84	ACAGGACAAGAGCGTGTTCGGAGAGACC
85	AGTGCCCCCTCCTTTAACCGGAATAACGTTGC
86	CCCTTAGTCCCAAGGAAAGAAACTTTATGG
87	TCACTACCCATCAGCCACTTGAAGGTCACG
88	TTAGAGACGCTAGAGAGCGACCGCTCAAGG
89	AGATTGCTGAAGGCTGCAGTGTGCAAAGG
90	ATTTGTAGCTAGGGACCTTGTAGACTGTGG
91	AGGGTGGAGAAGGAGCTACCTAACTTGG
93	CCCGGTTTCTGTGTAGAGGAATGAAGCG
94	GACTTGGAACTCTAGGCAAACCCACTGC
95	GTATGTTTTTGCGTGCATCCATTTTGTGC
96	AAAGGACGTACACTGTTCTCCTATCTCC
97	ACAGATCTGCTTGAGGTCTCGTGATAGG
98	GACAGAGGGAATATAAAGAGAGGCTGAGG
99	GGGAACACTCAACTCATCAGAGGTCGAAGG
100	CATCTATGTACTCCCACAACTTCCCTTACC
101	ACATTCGTTCCACATACCCCTCCCATTTCC
103	GCTTCAGTTTTACGTAGAAGCACACACC

MF	Primer Sequence
105	TCCAGAGAAGGGATGACATGTTCTCAGAGG
107	AAATCATCTCCCTGGCCCAGATTTTCAAGG
108	CCTGAGCTCCCAGCACTTGATTTCATTTGC
109	GATCCTGTCTGAATGGTGGGAACTAATTGC
111	ATAGGTTCTGATTCTGGGAAATTCCTTGG
113	CCTTTCACTTCCTGCCTCATGTTCTTTCC
115	CAGAGGCTTCTCTTAGTGCTTCTGTACC
116	AACTATTACCCCAGCTCCTTTAACAAAGCC
118	GTGTAACACATATCTGGAACTCTCAGCGGC
119	GGAAGATTTTCATTTGTAACGGAGACAGG
120	CTCAGCTCCCAAAAGACGGAGTCACAGG
121	ATTAATCGCACTAGGATTGGTTCCTGGC
122	CGTCATTCTTGGCCCTCAGTGTTAGTATCC
124	TCCTCCGGTTATTGCCAGAAATGATCAGGG
125	GAGTAATTGAGGAAGACGCTTGATGTGAGC
127	CGTGATGTAGAAATGGAATCGGGGGATGCC
129	CAGTTCTTCTAGAGCACTGTGAGACACTGG
130	GAAATCACCAGGCATCTAACGAGATTTAGC
131	GCTTACTTTTAAACTTCAGCAGTCGCCC
132	CTCGGTTTCATTTCCTCTTAGGTCTTCAGC
135	GCTATTTAGGCATCCCACTATGTTTCTCACG
136	GTCTATCACTTGGTTGGTTACCTCTCTGC
137	TTCCACCTACTCAGCTGGAGTGAATGTAGC
139	ACATAGCGCTTAATGGGGAAGAGCCTAGC
140	CTCTTAGAGAGCATCTCCAGGCTTAGGG
141	TACTCACGTGGCATGACCGTTGTATCTACC
142	CCTGTGTGCATATCTATGTATCACTGTGC
143	CTCATTTGTGAATCCTGCTCTAACCACAAGG
144	CTCATGAAATGAGGTATCGCTGACAGAGC

Mouse HoxC 3C Primers

MF	Primer Sequence
1	ATTTAACATGAGCTCTTCTTGGCTCCTTGG
2	TGACAAATAGCTACTCCCCAACCTTCTAGC
4	AGATGCCTGCTTGGTTAAGAATACAGTGG
5	AGGACTTTCACAGTGATCTGAGATGTTGC
8	AGTTACAAGCAAGCATAACTCCACTAGGC
9	TTATGTCCCTAGCCCCTGGTCTACTCTTCC
10	GTCCCAGTCATACCCCTTTTGTAACTGG
11	ATGCAATATGACTCTTGAAACCCCTTGACC
12	AGGAACATTCCCGGATCCAATGATGAAAGG
13	AGTGGGAGCATCTTCCTCTAGTCTCAGG
14	AACTAGGAGGATGTGTAAGCGAGGGCTTGG
15	CTCCTGTCTTCCGTGTCTCCAATCTTGC
16	GTTATTACCATCAGGGTAGTGGCGACAGC
17	TCAGGACTTGATAGCACACCACACTACC
18	TCAGACCTAGGTATTTAGGTGAAAGTGC
20	CCCCGGCCTATCTATATCTTTATTCAGG
22	ATCCTTTCCTTTCCCCTGAGGTCAGTATCC
23	TCAAGACCTTTGGAATGGGGTATAGGAAGG
24	CTTCTCTTGTCCACAAAGTCACTGGTGTCC
25	GTGAAGGAACTTTGAGTGCATAGTAGTGG
27	CAGTACTTCCTAAACGAACGGACATACTCC
28	CTCAGTGGCTGCTGAGGATGAGTGATGC
29	CACATTTAGAGCGATTTTGTGTGACTGG
30	CGATACCTCACATAGAACTGGGTACATGG
31	CTCCAGTTTCTTGAGATTCTGCCTCTGC
33	ACTGACTGGATCACTTACCTTGAACAGG
34	GAAGACTTGGTAGCTTGCAGGAAGACATGG
35	CCCACACCTTTAATCCCAGCTTTTTACAGG

MF	Primer Sequence
38	CATGGTGGCATACAGGTTGTAATCCTAGC
39	ACCCTCTTATAATCCTGGGGCAGTTTCTCC
40	GTTGTTACTCAGTGGTGGAGCAGTTACC
41	GAGGTGGACTTTGGTGAAACTAACATGC
43	GTGCTAGGGACTGTAATCAAGAAGAACTGC
44	TTGAGCAAGTTCCTGGCGGGATCTCCACTGC
45	CCCTTACCAATATGGAGGGACCAGAAAAGG
46	TACCTGAAGTCCACTGATGCCATAGAAAGC
47	CTCTAAACCCCTGGCCTTTGTGACTTGTCC
48	TGTCTTTCTGGTTCTCTTTGCCTGATGG
49	ATGCATAGCACTGTCATCTCTGTAAAGG
50	AGGAAGGCTCGAGCTGAGGAAGGAGAACC
52	CATTCCTAATTCCCAACTCCCTTGTAGAGC
53	GGAATGTAGGAACACCGGACTCACTCAGC
54	TTTCAGCCTTTCTCTCTGGAAGGACTTAGC
55	CTCAAAGTGCCTGGGTCCTTTGTGTAGG
56	AGAGGCTGAATTTTAAGACCTCCTTTTCC
58	GATGCGATTTTTAGGGATTAGCAAAGAGG
60	TCCAGGACAATATTCCCTGAAGTGGGTAGC
61	TGGGGTGTGTGTGGTCTGGTTTCAATTCC
63	TATACCTGAAGTATGTGGCCAGGAAAGG
64	GAAAACCTAGGCTGTATGCACAACAAACC
65	GTCCACAGGAGAGAAGGAGTTGTAAATGG
67	AAAAAGAAAGGTAGTGGGATGGTCAGCTAGG
68	TGGAAAAGCTGGCCTTGTCCATTAGCTAGG
69	AAAATCCCAACAACTGAGACTGCCTAGC
70	TAAAGGTAGCTGCATAGTCAGGAGAGTTCG
71	GTAATTTATACTGGCCTCCCCACACTGC

Mouse HoxC 3C Primers cont.

MF	Primer Sequence
72	CCATCCTAGAAGCTGAAGTCCACTCTCC
73	TTGGAGCCGTCCCTATAACCATCTAGTTCC
75	GGAATTTATGTTCCAGAGACGGATGGCAAGC
76	GTATACAAAGGAGACGCTTGTGTTTGTGG
77	GTTTCCTCCTAAATTACAGGTTGTGTGTGG
78	GGCAGTGACCTATTGGAACAAGTCAAACACC
79	CCTCATACTTAAGGTTCAGGTCCAAATGAGC
80	GAGATACATAGTTTTCTCGAGGATGGGAAGG
81	GGGAGAGAGAGTTACAGAGTCCAATAAAAGG
82	GCCCTAGACTGAACATGGCCTTTCTAACC
83	GGACTCAGCTTAGTGTTTGACTTGTCTCC
84	ATACGAATGTACGAGCAACTGGAGTCTTGC
85	GTAATTGAGGCAGACCTTGACTCTCAATCC
86	CAAGAGGGAGTTCTGTAAATGTTCTGAGAGG
87	GTGTTTGAGTGTTACTCAGTGGATCATGC
88	AGGTCACAGGTGGTAGATGATGCGATTTGG
89	GCTCTGGGATACAGTGTAACAACTAAGAAGC
90	GGCGATTTGAAGTAATAGAGACTAACTGAGG
91	CCAGAGACAGGAGGATAGAGGTATAGATGG
92	GCAAAAGAGAGACTCCAAGACCACACATTGG
95	GTGTCTTCTTCTACGATGCTCAATAGGC
96	GCTGTAGCTCACTGAACTTGATGTCTAGG
97	GCTCCGCTTTATTTGCAGCATCTTTCATGC
98	AGAAGTAGGAGACAAGGGCAATATGACG
99	TCTTGGAGAAGGGTAGTAAGACTGGGTTGG
100	GTTACAAAACATTTGGCAAGCTGGGCATGG
101	CTCGTTCTGACCTGAAGAGTGGATTTGACC
102	TGGAATCCTAACCTATAGTTACCTCCTCTGG
103	GGCCAAGCTAGAAAGCCAGACATAAAATCC

MF	Primer Sequence
104	CTTCTGAAAGAATAAAGCATCGCCAATCC
105	CTTTACTACACTCGTTTCCATCAGTTACCC
106	CTTAATTAAAATCCATGAGGAAGCCCCTTGC
107	CGTGCGTTTTGGTAAGTAGACTTTTTCAAGG
108	CACCCAGATGACAGCTTCCAAAGTATAAGG
109	TCCGTAGTCATGTAAAAGCCTCTGGTTCTCC
110	TCCTGCACTCATTTACCTAGGCTTTCAGC
111	TGAGGGCTAGTCAGAAGCAGAAATGTAGG
112	CTGCTTTTTGCAATACTCTACATAGTCCTAGC
113	AACCAGCAGCTATTGAGTGGCACTTTGACC
114	AAGGAAGGGCTAGCAGGATGTCTCATCAGC
115	GGCTGTATGTGGCTAGTGATTGTTGTAATGG
117	GGGTAAAAATCCACTACAGTCAGAGCATAGG
118	AACTTGCTTTCTTGTTTCCCTCGGGTTTCC
119	TTCTAGCTGGATTCACCTCCCTCTCAGC
120	GGTGTGATATGTGTGGCAATCATCTTGG
121	AATATCCTTCTGCAGCATGGAGCTGTGC
122	AAGCAGACAAGTTTGGCTGAAACAGAACC
123	AGGACATACTGATGGCTTCTCCTACTACC
124	TTAGACAACGACCAAAGTCAAGCTTCATGC
125	TTGAAGGTCCAGTCTCTACAAAAGAAGG
126	TGACCCTCAGTGATGGAAGCTTTACTTAGC
127	TAGGTACATGCCTAAAACCCACCAATCAGG
129	GTATCTTTTCTCTTTCACGCGCTATGAGACC
130	ATCTGCTTCTTCCTCACAAGTGTAGTGTGC
131	GACTTGACTGGTCCTCTTCCCATTACATCC
133	CTAGGGTAGCCTGCACCATAAAGTAAACTCC
134	GTCAGATCTCCAGAGTCATCGGTTATTTTCC
136	CTATGGCCTTTCTGTATTTCACGTCTCTTGC

Mouse HoxD 3C Primers

MF	Primer Sequence
1	GGAGACAAGAAAAGGATCCAATGTTAGC
2	GTGAATGGTGCACACTACATTCATTCAGG
3	CCTGCTAAGGCTTATATCCCACTACTAGAGC
4	AGAGAGCAGGTAGGGGACTGAAGAAAAGG
5	CTCTTCCAAGATCTGACTCATGTTTTCTCC
6	GATGCAACACCATTAACCCACAGTATCG
7	GAATTTATGGAGAAGTTAGGTTGACAGTCC
8	GCACACACACTGATCCTATGATAATCTCC
9	AAAGGAGAGCACACAGATAGTAAAGAAAGC
10	CTGGTACGTTATCCACATGAAAAATGAGC
11	CAAGGTATTTTTCAGATAGCCAAGTCATCC
12	GGACTCAAAAGAGTTCACACAGACTGAAGC
13	CAGTTGGAAGTTTTCTACTTCACCTTGTCC
14	GCAATGCTTAACCAACTGGTGCTAAACG
15	TAAGCAGAGGTAAAAGGCTAAACCCAGAGG
16	GGGCTCTAACACAATCTAGGCTCTAACC
18	ACAGTTCTTAATGTCCAGATGCTAGAGAGG
19	AGTTTGAGTCTTAGTACTAGGGGGTTGAGG
20	GAGGCCTTCCAGGACCCTTGTAAAGAAACC
22	GGCTGGGAAATACAGATCAGACTAAAGG
23	ATATTTGGTATTGCTGGTTTGCTGAGTCC
25	ATGGCTCTATTGGGAGGTATCAAAAAGTCG
26	ATCTCAATATGTACCCTGGTTGGTCTGG
27	AACTGGAACCAGTGAAAACAGCAAATCTCC
28	GTACTGCAGTTTAACAGACAGACTCACC
29	TCTATAGGACCTGCCAAAGATAATGTGG
30	CTACATTTCCACTGAGCCAATGATATTCC

MF	Primer Sequence
31	AAGCAGCACATATACTTAAACCGGAACG
32	GGAAGAGATTACTTCCACACTAAGCAGAGC
33	ACCAGTTACATCAGAATCACTGTGTGTAGG
34	GATCCATCTCTATAGGACCCCACCTACC
35	CCTATTTGAGGTGGGCATAAGATACAAGG
36	GTGTTGTCCCTTACCTGACATCCTGAGC
38	AACCGAACATCTGTGGACTTCCTAGTTGC
39	AGGCCTGTAACTACCTAGGAATCTCTGTGC
40	GGCAGTTGATCTGAGCGAGCTGACATGG
41	TAGCTAGGGTGTCTCTTCTCAGAGACTGG
42	AAAACAAAACATCTCTCTGCCAAGTTGTGG
43	AAAGCAATTTGCCACCCTGCTAAATAAACG
44	AGCAACTACTACGTGGACTCGCTCATAGG
45	CTCTTCCTAGACAACCCTGCAGCGAACTGG
46	AGGACCAGTTGTAAATGTTACCGCTTCC
47	AAAATCAAGTTTATGCTAAGGCCTGCATCC
48	AAAACTCCAACAGCATTCAGAGAATCAACC
49	CAATCTCTTTACCCACTTCTATGTGTATCG
50	CTGGGAGTGCAGTCTCTGGAAAGTCTAGG
51	TAGCAATGTCCCCCTCTTTGTTAATTTAGG
52	AACTTTTGTACTCTTCTGTGCTGCTGTCG
53	TGGCAGAAAAGCATAGTGGTTACTCATTGC
54	GAACTCTAAGTACGTGGACCCCAAGTTTCC
55	CTTTCCTGGAGACAAATTCCCGCAAAGC
56	GTTCCTCCAGCGTTTATTGGTAGTTGAACC
57	AGACTGGTCCACCATTAAAGACAAACTTGG
58	AATGCTCCGTCCCCTCCTGTAAAGATTGC

Mouse HoxD 3C Primers cont.

MF	Primer Sequence
59	ATCAAGGTTTCTTATTGACCCCAAGTCTGC
60	CGATCTCTTAGACTTGAAGTACAGCAAACG
61	CCCACAGAAACTGTATGCTATTATGTTTAGGC
62	ACATGCAAAGAGTAGCCACTAAAACAACC
63	TTAGGGGACATCAAACTGTTGAGTACATCC
64	AGGAGAAGCCTGGAAGTAATAAAGATGACC
65	AGCCCCACTAAGCTGTAGAGTCTGTGAGC
66	GTACTCACATGTCGTGTTACAACTGTCTGG
67	GTTACTGATGGCTGCTTCTAGGTAGTAGGG
68	AGTTCAAAAAGGATAGTAAGCGCTTTAGGG
69	TACTTACCTCCTGGCTAGTATGCACATTCC
70	TAGTGGTTCTCGGGGTCTCTAAGTGTCTGG
71	CAATACATTTGCCGAAGGTGGGACATGG
72	TATCGTATGAAAGAAAGGCTTACAGTGTCC
73	CTGCATACAAGTACAGGTGTCCTCAGAGC
74	AAGTTGAAAACAAAGGACCTTGGAGTGAGG
75	TCAGCTTGACTGAAACCTAGAGTTACCC
76	GCAGTTCATTTTGTCTCTCGTGAAATAGG
77	AGATCTTCTAGCAGTACTGAGGGCTGTTTCC
79	GATGGCAATACTTCTGTATCAATCTTCTCG
80	CTTTAGGTAGCAGGGCATCTCTGCTGTGG
81	AGATTTAGGGAGGGTATATCTTTGCTAGGC
82	TCCTGTGACTTTAGATAAGACATCACTACC
83	CTCACAAGCTGTGTAAGCCTCCTTCTGG
84	GACCTGTCTGTCAGTTCTCTTTAAATACCC
85	CAGAAGAAAGTACCTGTAAATTAGCCATGC

MF	Primer Sequence
86	ACCTGGTGTAGTTGCCTTATCTTCTTTGC
87	CTGCATTTGACTGCTTTATACTTCTACTGC
88	CCTTATTTCTAGCACCCAACGTTAAAGTCC
89	TATCCCCTTGGTCTTTAGCTTTCTAATTGC
91	GGAGTTTTACTTCTCTTTGAACTTGAGTGACC
92	TTACCTTAAAAACAACACTCTGGGTTCAGC
93	AAGATAATTCCTTGATGGGCATGGTTCTCC
94	GGGACTTGTTATCTCTGAAGCAGGAAGC
95	AACTGTTTCATGGAGATGAACACTTCTCG
96	CGGTTGTTAGTCTCCTGTACCTCCTTCC
97	GCAAGGAGCTTTCTCTGTAGTACATCTTACC
98	CTTTCTCTCAAGAACAGTCCAAAAGAAGC
99	TAAGGGTCTCTATCTCCAAGAGACCAGTCC
100	AGATTAGTTGCAGATGCCCTCCCCACACC
101	GAATTGAAAAAGAAAATCGGTGCCTCTGG
102	GTTGTAAGTCTGTGCTCAATCAGCTTTCC
103	TCACTTGCCAGATACTCAAACAGAAATACG
104	ATTGAGTATGGTGGCAGTTGCAGAATCG
106	AAGTTCAGGGTTACCAGCGAGCTGAGAAGG
107	ACAATCTGAATGTTATTAAGGGGGAAGTGG
108	CCTACTCTTCGAGGTACAGGAAGTTGTCC
109	ACTGGCAGTGTCCAAATGTATTCTAAATGG
110	TTTAAGATGTCACTGGTACCAAGGTTGAGG
111	CTGCACATATACATTCTTAAGTCAGCTTGG
112	AATCTGTCTAGCCATTTGGGTATGATGTGG
113	ATGAGACTGTAGCGAGTATCACCATCTTCC
114	TGTAATCCAAGTATTGGGGAGTCTGAGG

Mouse Gene Desert 3C Primers

MF	Primer sequence
1	GGCTAATGAGAAGACACACAAATTCTGG
2	AGCAGATGTGAAGACCCTGATGAAAACACC
3	CTTCATTACTCTACAGGTCCAGAGGCAAGG
5	CCACCCTGGTGGCATCTACTTAGAATAGG
6	ACAACCAACTGGTGGAGGGTTTGTTTTGC
7	TTCTGGCTGCTTGTGTTGTGACTGAAGAGG
8	GTAGACTAGAGTAGAGAGTTCAGAGATGAGC
10	CTCATCCAGTCATCATCAGAGCATTTACC
11	TTGTCTCTCTAGTGTGCAGTCTGCTGTAGG
12	GGTCTCCTGTGCTATTTGTTTGAGTACTAGG
14	ACAGAGACTATGTCTGGATGATGTCTTGC
15	GGTTAGCAGCTTTCTTTGCTTCCCTTTAGC
16	CCCACAGACATATGGAAACTGAAAAATGC
18	ACAAACCCTGAAAGTATGGGAAGATTAGGG
20	CAAAATTTGCAGGCTACTATGGCCTGAGC
21	ACTTAGTAGTTTGCTTAAGGTGGGCTTCC
22	CATTCTGACTGGAGTGAAGTGGAATCTCAGG
23	CAGATTTGGGGGCATATCTGACACTATGG
24	ACTACTTTGCATTAAAGAGGCTGGAACC
25	CACATGATCATCTCGTTGGATGCAGAAAAAGC
26	GAAGGTGGACTTCAGAAAATCAAACAACC
27	CCCTCTCTCCACCTTTGGTGATTTTGTTCC
28	GCTGAATCTGCAAAAATGCACAGTAGATGG
29	GCAGATGTTTGGCCATTTCTTTTGAGTCTTCC
30	CAGAAACAACAGAAAGCCTACAGACTCATGG
31	AAAGAACCTCGGAAGATGTCAGAAGATGG
33	GACCCATGACTACAGCCGTATATGTAGC
34	GAGCTCTCCTATTTAGTTTACCCCGAGAGG

MF	Primer Sequence
35	GAAAGAGGATTTGAAGTTGGGTGAGTAGGG
37	TTGTTTGTTTTTGGTGGGGGAGAGTAGG
38	GATTAAACACACACCACCCGTAGTACCC
39	ATAGATATTGCCTTGCAGCCTCTTTCTCC
40	GTGAAAGACAAAACACAGCCAATAACTGC
41	GGTTTGTCTTGGTGGCTAACATAAGAGC
42	CCCTCATTAAACCCACATAATTGTAGCC
44	GATAAGGAGAATACAAGGGTGAGTTCTTGC
45	GAAATGACAGTGCCTGTTGGAAAACTTCG
46	CAGTTCTCGCAGTTTTGGTGCTACTTGC
47	TATCAATGATGGAAGGTGAGATGGGATAGGG
49	ATTCTGACTGGTGTGAGGTGGAATCTCAGG
50	CAGCCTTGTCTAGTCCCTGATTTTAGTGG
51	GGTGGAATCTCAGGGTTGTTTTGATTTGC
52	CTAGTCCCTGATTTTAGTGGGATTGCTTCC
56	CCTTCTGTCTCTTACTCTTGCCCCATGAGC
57	GGTTCCTGGGAGGATCCTAGTAAAATAGCC
58	GGATCTGCAAACAGAGGAACAGATGCTAGG
59	GCTTCTTAGTGCTTGAGATATGACTGAGG
60	GAGATGGAATCTCAGGGTTAATTGATTTGC
62	GCAGCCTTTTCTAGTCCCTGATTTTAGAGG
63	AGCCCGTGATACATGAAGGTTAACAGAGC
64	CTTATGCTTCTTGTTCAAGTCCTAACATCC
66	CTTCTATAGCAAAGTCTACCCAAATGAAGC
67	GTGTGATAAACACAACAGGCAAGCAAAGG
68	GCAGTCTTGTCTTGTCCCTGATTTTAGTGG
70	TATGCCCAGGAGTGGTATTGCTGGATCTTCC
72	TTATCTTCTCTCCCCATGACTTTCTACACC

Mouse Gene Desert 3C Primers cont.

MF	Primer sequence
74	CTTGTCTAGTCCTGGATTTTAGTGGGATTGC
80	CACATTCATAGAAGTGGGCAATCTTTACC
81	GTAATGGGCAAACCTGTCATATTCTAGGC
82	ACCAGGTACCAAAGCTAAATCAAGATCAGG
84	GGGAAAGGATCTTTACCTATCCCAAATCG
85	AAACACACTTCTACTTTCCGTTGACACC
87	GTGTTAGCAGGGTATCTCTAATGTCAAAAGC
89	AGCACCATGGGTATTAGAAGAGTCTCTAGG
90	ATCTCTCTTTATTTGTCCCTCCCTCACC
91	GAATTGAGTTGGAATTTTGATGGGGATTGC
92	CACTCAAGAATACATAGCTAGCAGATACAGC
93	TTCTCTCTTCCCCTCCATGGTGACTGACTCC
95	GGGGTGCTTAATGTGTATTAGTGTCAGAGG
96	GGCCATCTGTTTCTTAATGGTGTTTTGACTCC
97	CTTAATCTCTAGTCTCTGCCTGTCACTGC
99	TCTAGGAAGCTGGCTTTTGACTACCACACG
100	ATGCCTAAGGAGTTTGACTGTATGTTGG
101	AGAGAAGCCAGATTCAGTCAGTCAGCTTGG
102	CACACTGCTTGCAATTTTATGTTCTGAGC
103	TGGGCTGTTTTGGAGTCAGCTTTTATCC
104	TGCAGACTTTACTTACTATCCTGTGAGGAAGG
105	GTTCCTCAGACAAGTGACTTCTTGCTACC
106	TCTCCACTTTCTGCTCTTGCTGACCATACC
109	GGGTACAAGGAAGAAACAAATGGCTGAGG
110	GCCACTCCCTGGTCCAAGAATATACAAACC
112	GCTTCTGTCCAGACATAGTGATACAGACC
113	GAAAGATTGGAACTGGAGGATCAGGTGAGG
114	AGAAAAAGCTGTCAAGTCCTCCGTCAGAGC
116	CCTACCACCTCCACTGTCAGGATTATTGG

MF	Primer Sequence
117	TCTATTCCCGCATTAGTCTGGTGTATGG
120	CACAACTGCCCATCTTTCTGAATATCAACTGG
122	GCCACTTACCAAAGTGAAATCAAGATCAGG
124	CAAAGCTAGAGACAATTCTTGACAGAGG
126	AAAGGACATCCAGTGATTCCTACAGAGG
127	GCAGAAAAAGCATTTGACAAGATCCAACACC
129	CCTATAGCCAAACAGACAGATAGACAGAAGG
132	GAGGAATTTGAATCCACCCACTTCACTGC
133	CCCTCTCAAATATGTAGGGTTGGAAAGG
134	GGAAGTGCTGGAGCTGAGTAAAAGATGG
135	AGACTTACTTCCACGCTTCTCTCAAAGC
136	GGTTGAGGCAGAACAATTGTAAGTTCAAAGC
137	CAACCCAGATATCCCAAAATACAAGAATGG
138	AAAAGAACTTCTGCCTGAATCACCATCC
139	TCTTATATTCAGCTGGTCAGCCAGCATTCC
140	GGCCATGTTAGGTTGTCTTGTTTGGTTGG
141	GTTGGGCTTATACTGAACAATGAAAGTGG
142	TCTGTGGAGTGGGGATGTTGAATCCTTTGG
144	TGAGTGTAGATGAGTTTCATGTGTCATCG
145	TTCCAATTCCGATTTAGGGAGATAGTCC
146	GGAGGCCATTATTCTTAAACAAATGTGTGTGC
147	GTTCAGCAGTAAACATCACATCTGCTCTCC
148	CCTATGGAAGAGTTAGGGGACACAGAGC
151	TCCCTATAAACACACCAACTGAAGAGACC
152	GGGCAGTATAGGGAACTTTCAGGGTAGC
155	GCAGACAAAATGGCAGCCTACAGAATGG
156	GAGGACAGAGATGGCTGTTATTGGGTACG
157	GAATGGAAGTCACCAGGAGAACAAGTGC
158	TAGGCTCAGACATCTGCTTTAGAAATCC
159	CTCAAGAAGCTAGCCACCAAAAGAAACC

Gene expression primers

Primer name	Primer sequence
HoxA1 Fwd	GTGGTCCAGAGCAGAGTTTACG
HoxA1 Rev	GCTGCAAGCTTCATGACAGAGG
HoxA2 Fwd	CCTACTCTCTCAAATGTTGAGG
HoxA2 Rev	TATGGTCACTTTCTTGCAGGCC
HoxA3 Fwd	ATACTTTATGGCTCCCGAGAGG
HoxA3 Rev	GTTCACAAGCACACACACACGG
HoxA4 Fwd	TTTGAGTCAGCCAGACAGCACG
HoxA4 Rev	CAAGGAGAGGAAACTACATGCC
HoxA5 Fwd	TGCATGTACGTGGAAGTGTTCC
HoxA5 Rev	AAAGGGTCCTACAAAGGCACGC
HoxA6 Fwd	TACCTGCACTTTTCTCCCGAGC
HoxA6 Rev	CTTCTCAAGCTCCAGTGTCTTGG
HoxA7 Fwd	ACCAAGGCTTCACTGACAAGGC
HoxA7 Rev	AGCTATTCCAGAGGATGGCAGC
HoxA9 Fwd	CAACTGGCTACATGCTCGCTCC
HoxA9 Rev	TCACTCGTCTTTTGCTCGGTCC
HoxA10 Fwd	GAGAAGGAAACCTCTCTTCCCG
HoxA10 Rev	AATAGGGAGAATTGTGGTGTGC
HoxA11 Fwd	TCTGGCAATCTTCTAAGAGCCG
HoxA11 Rev	TTAGCAGTGAGCCGAGTTTAACC
HoxA13 Fwd	CAGTTGAAAGAACTCGAACGGG
HoxA14 Rev	CCTCTGCTCCACCTTTTAATCC
HoxB1 Fwd	GGAACACTCAACTCATCAGAGG
HoxB1 Rev	GACCTAGAAAGAAGAGTTGGGC
HoxB2 Fwd	CCACTCGTGGATACATTTATGC
HoxB2 Rev	CTTTGTAAGAAAACCCTCTCCC
HoxB3 Fwd	TACGGCCTCAATCACCTTTCCC
HoxB3 Rev	CATCACAGGTGTGTCAGTTTGG
HoxB4 Fwd	CAGGTCCTGGAGTTGGAGAAGG
HoxB4 Rev	CTTGGTGTTGGGCAACTTGTGG
HoxB5 Fwd	TGGACATTGAGAAGAGGAGAGG
HoxB5 Rev	GGGTGAGAGAGGAAGTGATTGG
HoxB6 Fwd	ATCGGACTCACTTGATGTCTCC
HoxB6 Rev	TAGTATGTGCTCCTTCCAGTGG
HoxB7 Fwd	AAGCTGTTTGTGAACTGTGG
HoxB7 Rev	CAGAAAGAGGCTCGTGAATAGG
HoxB8 Fwd	AGAGAGACAGGTCAAAATCTGG
HoxB8 Rev	TACTTCTTGTCACCCTTCTGCG
HoxB9 Fwd	GTTCCTAGAAAAGTGATGAGCC
HoxB9 Rev	CTCCAAACTCTATGTCTGTGGG
HoxB13 Fwd	CTTCTGAGTTCTGGACTGTTCC
HoxB13 Rev	ATCAGCTCAACTCATGAAAGCG
HoxC4 Fwd	AGTGGAGTCTCATCCTCTCTCC
HoxC4 Rev	AGGCTCTCCAGATGCATTTAGC
HoxC5 Fwd	GAGCTGTATAGATCAGTGTGGG
HoxC5 Rev	TCTCCTGTAAAGTACTTCCTCC
HoxC6 Fwd	GGCATATTTCGAACTGTGACCG
HoxC6 Rev	TCTCCTACTGGCTAAACAAACG
HoxC8 Fwd	AACTTACAGCCGGTATCAGACC
HoxC8 Rev	TTCCACCTTCTCCTCATCTCGG
HoxC9 Fwd	TACACCAAGTACCAGACGCTGG
HoxC9 Rev	GGCTTAGGATTGTTCCTTGTCG
HoxC10 Fwd	TTGTGTGTGCCCTCATAGATGG
HoxC10 Rev	GAAGGAGCAATGCTAACTGTGC
HoxC11 Fwd	CCTTCCACAGTCACTGAGATCC
HoxC11 Rev	GAAGAAGCGGTCGAAAGCTTGG
HoxC12 Fwd	AGTTCGTCCCTACTCAACGAGG
HoxC12 Rev	GAACTCGTTGACCAGAAACTCG
HoxC13 Fwd	CATAAGCAGTAGCAGATAGTGG
HoxC13 Rev	TTGTAATGAGCCAGTGGGAAGG
HoxD1 Fwd	TCCTGAGATGATCAAAGCTAGC
HoxD1 Rev	CTAATCCTTACCTACCCTGTGG
HoxD3 Fwd	ACCTGACAGTGCCATAAACTGC
HoxD3 Rev	AGTCCGATGCAATCTTTACAGG
HoxD4 Fwd	CCTTGGCGAGTCATTAAACTCC
HoxD4 Rev	TCCTGCTTGCACTGAAAGCACC
HoxD8 Fwd	CAAGCTACCTTAATGTCACTGC
HoxD8 Rev	GTTGTTGAGGCAAACCACCAGC
HoxD9 Fwd	TCTGAAGGAAGACAGAGTGTGG
HoxD9 Rev	TGCTCTGAGGACTAAGCCTATCC
HoxD10 Fwd	TGCAGGGTAACTATTACTGCGC
HoxD10 Rev	GCCATGAGTTCATGGGCTAACC
HoxD11 Fwd	CAGACGTGCTCTTCAAGGCTCC
HoxD11 Rev	GCCTCGTAGAACTGATCAAAGC
HoxD12 Fwd	TGGTGGTCAGCTCTTGTAAAGC
HoxD12 Rev	GTGCATCTTGCTAAAGGTTGGG
HoxD13 Fwd	AGCCAGGTGTACTGTGCCAAGG
HoxD13 Rev	TGAGCTGCAGTTTGGTGTAAGG
Actin Fwd	CTCAATTGCCTTTCTGACTAGG
Actin Rev	GCTGTGTTCTTGCACTCCTTGC

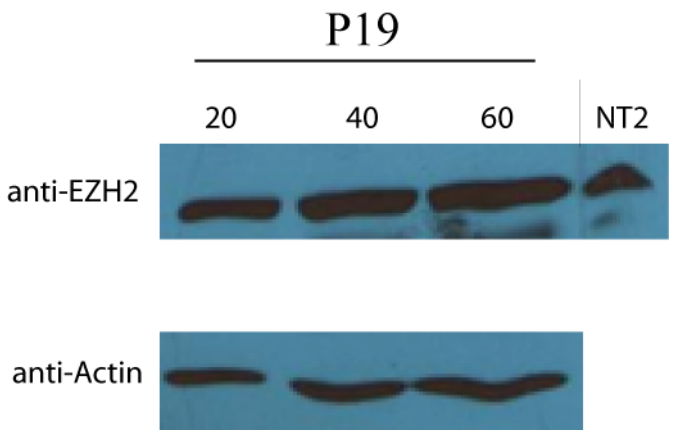


Figure A1. Western Blot of Mouse EZH2

Protein analysis for the mouse EZH2 enzyme from P19 cells and the Human control from NT2 cells. Volumes of cellular extract from 20µl to 60µl were used to assay for efficiency of antibody. Anti-Actin was used as a positive control

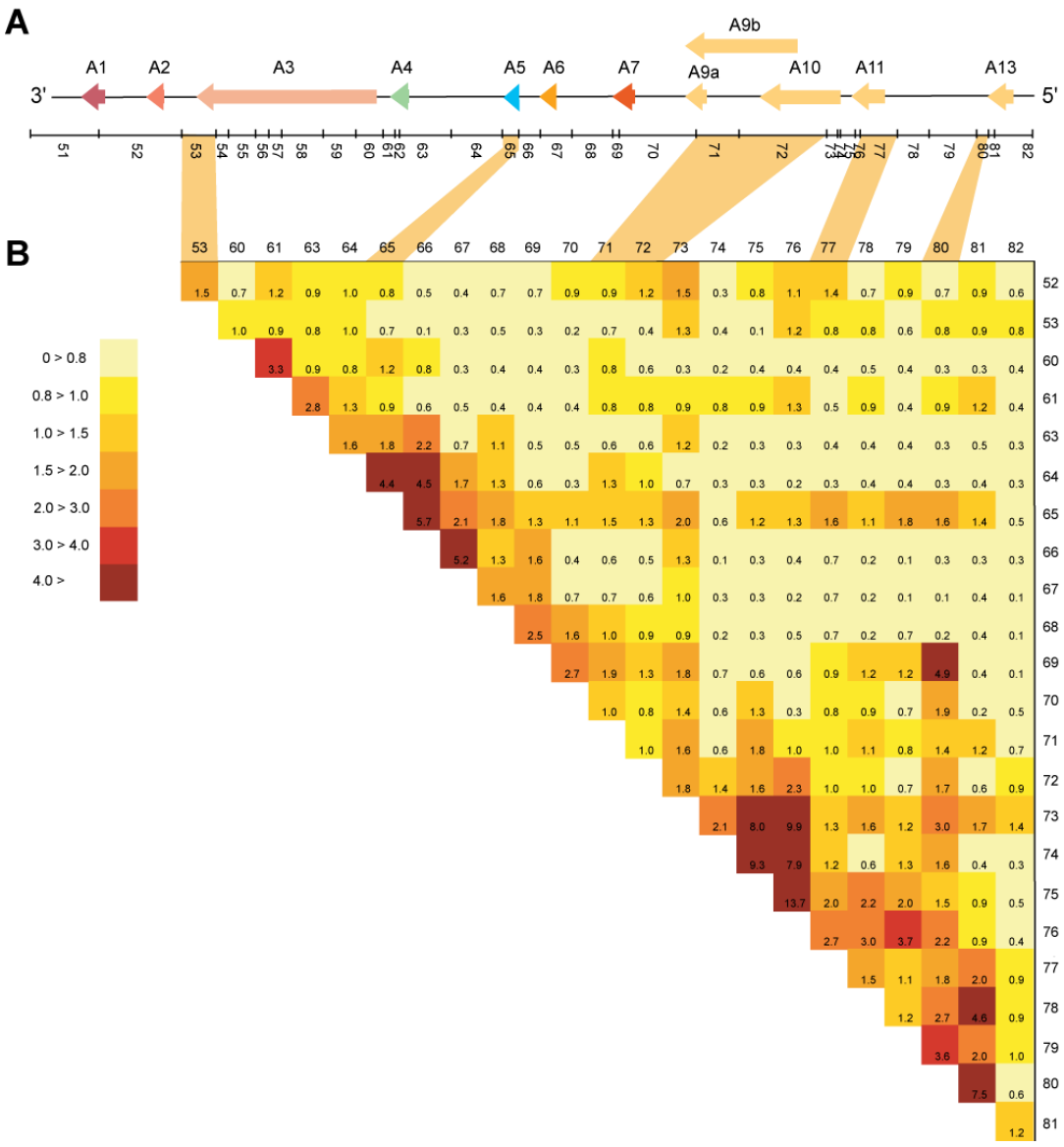

A
A9b
A1
A2
A3
A4
A5
A6
A7
A9a
A10
A11
A13
3'
5'
B
0 > 0.8
0.8 > 1.0
1.0 > 1.5
1.5 > 2.0
2.0 > 3.0
3.0 > 4.0
4.0 >

Figure A2. Analysis of 3C experiments in the Human NT2 HoxA cluster

A) A schematic representation of the annotated Human HoxA gene cluster including the location of all fragments used for analysis. The schematic is drawn to scale. B) A heat map depicting the culmination of all fixed point experiments used in examining the conformation of the Human HoxA gene cluster. Interaction frequencies (IFs) and their corresponding representative colors are indicated in the figure legend. Values for interaction frequency ranges were selected based on average IF over the entire cluster.

Provided by Dr. Maria Ferraiuolo