Epigenetic control of human *HOX* clusters

Thesis

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.21954/ou.ro.0000f9c6

Luca Sessa

Epigenetic control of human *HOX* clusters

PhD Thesis submitted in fulfilment of the requirements of the Open University for the degree of Doctor of Philosophy in Molecular and Cellular Biology

April 2004

Director of study:
Dr. V. Orlando

External supervisor
Prof. B. Turner

DIBIT
Department of Biological and Technological Research
San Raffaele Scientific Institute
Milan, Italy

Submission date: 30 March 2004

ABSTRACT

In *Drosophila*, Polycomb group (PcG) and trithorax group (trxG) genes are part of a cell memory system that prevents changes in cell identity by maintaining transcription patterns starting from the developing embryo throughout adult stages. PcG and trxG control epigenetically repressed and active transcriptional states of *Hox* genes and developmentally regulated genes. Both PcGs and trxGs exert their functions by binding to specific DNA element called Polycomb/Trithorax Response Elements (PREs/TREs). Many PREs were identified and characterized in the *BX-C* of *D.melanogaster*. Since no human PREs/TREs have been identified, we decided to map PcG/trxG protein distribution in human *HOX* loci.

Sequence comparisons "*in silico*" revealed that *HOX* clusters present a unique redistribution of the Repetitive Elements (REs) which localize in the flanking regions and are absent from the inner part of the cluster. Mapping of PREs/TREs by chromatin immunoprecipitation (X-ChIP), revealed that PcGs/trxG proteins associate with RE-containing fragments at the boundaries of the human *HOXA* cluster. We discuss the possibility that REs could modulate *Hox* gene expression by working as epigenetic elements able to nucleate the formation of heterochromatin-like structures.

PRE/TRE characterization in *D.melanogaster* showed that PcG and trxG proteins could interact with the same DNA element, implying that these elements could work both as a PRE and as a TRE. Recently it has been hypothesized that the switch from PRE to TRE involves the transcription of these intergenic modules. As for other genetic phenomena (X-chromosome inactivation, dosage compensation, genomic imprinting), these non-coding RNAs play a role in epigenetic control of gene expression, suggesting that non-coding transcripts may undescore the position of *cis*-regulatory elements. I then investigate about 400 genomic loci for the presence of non-coding intergenic RNAs. Among the others, I have identified and characterized these non-coding transcripts in intergenic regions of human and mouse homeotic loci.

DECLARATION

This thesis has been composed by myself and has not been used in any previous application for a degree.
Results presented in this thesis have been obtained by myself. Throughout the text I use both “I” and “We”. Both forms are used interchangeably, and “We” does not mean other than “I”, thus “We” does not necessarily mean that more people contributed to the result or discussion.
All sources of information are acknowledged by means of reference.
Part of this work has been published:
Lavorgna G, Sessa L, Guffanti A, Lassandro L, Casari G. AntiHunter: searching BLAST output for EST antisense transcripts. Bioinformatics. (2004) 20(4): 583-5.

Contents List

Chapter I: INTRODUCTION

Part I: Common aspects of epigenetic phenomena. Pag. 8

"Genetic" and "Epigenetic" regulation of gene expression. Pag. 9

DNA and histone proteins modifications represent common aspects of many epigenetic mechanisms of gene regulation. Pag. 11

Non-coding RNAs: a novel epigenetic mark contributing to gene regulation. Pag. 17

Heterochromatin assembly requires repetitive DNA, ncRNAs and epigenetic modifications. Pag. 22

Part II: Epigenetic regulation of *Hox* gene expression by the Polycomb and trithorax protein families. Pag. 28

Hox genes: an evolutionary conserved family of master control genes. Pag. 29

Maintenance of *Hox* genes expression through development: the *Polycomb* and *trithorax* genes families. Pag. 33

The PcG and trxG cell memory system is maintained through evolution. Pag. 34

PcG and trxG proteins act in large multiprotein complexes. Pag. 38

PcG/trxG multiprotein complexes modify the chromatin structure. Pag. 40

PcG and trxG complexes bind specialized DNA elements. Pag. 45

Promoters share similar sequence motifs with intergenic PREs and are "in vivo" binding sites for PcG complexes both in *D.melanogaster* and mammals. Pag. 48

Transcription through PREs correlates with epigenetic switches. Pag. 49

ncRNA in eukariotic genomes. Pag. 51

Chapter II: RESULTS Pag. 54

Identification of PREs in the human HOXA cluster. Pag. 55

HOXA cluster: from *in silico* cloning to cosmid contig. Pag. 55

Polycomb, trithorax and *HOXA* gene expression in human cell lines. Pag. 58

PcGs and MLL1 distributions in the *HOXA* cluster and pericentromeric repeats by X-ChIP/Southern blots. Pag. 59

X-ChIP and "genome complexity". Pag. 62

PcGs and MLL1 distributions in the *HOXA* locus by X-ChIP/PCRs. Pag. 65

HOX clusters show a peculiar and unique distribution of REs among the human genome. Pag. 68

Alu insertion inside human *HOX* clusters was negatively selected. Pag. 75

Exceptions to the rule and gene regulation. Pag. 76

Comparison between human and mouse clusters. Pag. 77
Non-coding intergenic transcripts coordinate *Hox* gene expression. Pag. 88
AntiHunter: searching BLASTN output for EST antisense (AS) transcripts. Pag. 88
"In silico" identification of intergenic "Opposite Strand" (OS) RNAs in human *HOX* clusters. Pag. 89
"In vivo" identification of intergenic OS RNAs inside the *HOXA* cluster. Pag. 91
Intergenic OS RNAs are transcribed before *HOX* genes. Pag. 98
"In silico" identification of intergenic OS RNAs in the "dispersed" homeotic loci. Pag. 99

Chapter III: DISCUSSION Pag. 105
Large blocks of Repetitive Elements (REs) could contribute to regulation of *Hox* loci by promoting high order chromatin domains. Pag. 106
Polycomb and trithorax proteins associate with REs at the HOXA locus. Pag. 109
PcG/trxG proteins could map at both active and repressed regions. Pag. 109
RE might insulate *Hox* genes from the surrounding environment. Pag. 110
Absence of REs: "spatial", "quantitative" colinearity and regulation of *Hox* genes. Pag. 111
Could PcG/trxG proteins promote sub-nuclear re-localization of *Hox* genes? Pag. 113
Mapping of epigenetic elements by ncRNA. Pag. 114
ncRNAs could promote the expression of *Hox* genes. Pag. 117
Future plans Pag. 122

Chapter IV: MATERIALS AND METHODS Pag. 124
Reagents and standard procedures. Pag. 125
PAC, BAC, cosmid and plasmid clones. Pag. 125
Southern blots. Pag. 126
Western Blot. Pag. 126
RNA purification and RT-PCRs. Pag. 127
Chromatin immunoprecipitation from human cell line. Pag. 127
HOXA genomic microarray. Pag. 131
NT2 cell line treatment with Retinoic Acid. Pag. 131

Chapter V: TABLES. Pag. 132
Table 1.1: PcG protein family. Pag. 133
Table 1.2: trxG protein family. Pag. 134
Table 2.1: *H.sapiens HOXA* genes. Pag. 136

Table 2.2: DNA markers. Pag. 137
Table 2.3: *HOXA* cluster plasmid clones. Pag. 138
Table 2.4: *HOXA5-HOXA6* intergenic region. Pag. 140
Table 2.5: *HOXA* genomic microarray. Pag. 141
Table 2.6: X-ChIP/PCRs. Pag. 146
Table 2.7: X-ChIP/PCR amplified fragments contain different REs. Pag. 148
Table 2.8: *H.sapiens HOXB* cluster genes. Pag. 149
Table 2.9: *H.sapiens HOXC* cluster genes. Pag. 150
Table 2.10: *H.sapiens HOXD* cluster genes. Pag. 151
Table 2.11: "REs-distribution" in the human genome. Pag. 152
Table 2.12: *M.musculus Hoxa* cluster genes. Pag. 153
Table 2.13: *M.musculus Hoxb* cluster genes. Pag. 154
Table 2.14: *M.musculus Hoxc* cluster genes. Pag. 155
Table 2.15: *M.musculus Hoxd* cluster genes. Pag. 156
Table 2.16: Intergenic OS-EST mapping in the human *HOXA* locus. Pag. 157
Table 2.17: Intergenic OS-EST mapping in the human *HOXB* locus. Pag. 158
Table 2.18: Intergenic OS-EST mapping in the human *HOXC* locus. Pag. 159
Table 2.19: Intergenic OS-EST mapping in the human *HOXD* locus. Pag. 160
Table 2.20: RT-PCR conditions. Pag. 161
Table 2.21: OS-ESTs mapping in *H.sapiens* dispersed homeotic loci. Pag. 162
Table 2.22: OS-ESTs mapping in *M.musculus* dispersed homeobox loci. Pag. 167
Table 3.1: OS-ESTs mapping in 90 *H.sapiens* genomic loci. Pag. 171
Table 3.2: OS-ESTs mapping in 90 *M.musculus* genomic loci. Pag. 173

Chapter VI: REFERENCES. Pag. 175

ILLUSTRATIONS.

Fig 1.1: A) sites of post-translational modifications on the histone tails.
B) heterochromatin assembly leading to epigenetic silencing. Pag. 18
Fig 1.2: A) chromosome X inactivation pathway.
B) Igf2r imprinted gene cluster. Pag. 23
Fig 1.3: DNA repeats might work as nucleation center for heterochromatin formation. Pag. 27
Fig 1.4: A) a homeotic mutation.
B) establisment and maintenance of homeotic gene expression in D.melanogaster. Pag. 30
Fig 1.5: homeodomain protein families. Pag. 31
Fig 1.6: *HOM-C/Hox* clusters organisation. Pag. 32
Fig 1.7: axial skeletal transformations in Bmi1$^{-/-}$ mutant mice. Pag. 35
Fig 1.8: an epigenetic model for the action of PcG/trxG complexes. Pag. 45

Fig 2.1: cosmid contig spanning the human HOXA cluster. Pag. 56
Fig 2.2: genomic organisation of the human HOXA cluster. Pag. 57
Fig 2.3: Western blot analysis. Pag. 58
Fig 2.4: X-ChIP/Southern blots. Pag. 60
Fig 2.5: hPC2 distribution in the human *HOXA* locus. Pag. 61
Fig 2.6: X-ChIP/Southern blots. Pag. 61
Fig 2.7: X-ChIP/Southern blots. Pag. 63
Fig 2.8: X-ChIP/PCRs. Pag. 64
Fig 2.9: ChIP to CHIP. Pag. 66
Fig 2.10: X-Chip/PCRs. Pag. 67
Fig 2.11: genomic organisation of the human *HOXB* cluster. Pag. 69
Fig 2.12: genomic organisation of the human *HOXC* cluster. Pag. 71
Fig 2.13: genomic organisation of the human *HOXD* cluster. Pag. 73
Fig 2.14: genomic organisation of the mouse *HoxA* cluster. Pag. 78
Fig 2.15: genomic organisation of the mouse *HoxB* cluster. Pag. 80
Fig 2.16: genomic organisation of the mouse *HoxC* cluster. Pag. 82
Fig 2.17: genomic organisation of the mouse *HoxD* cluster. Pag. 84
Fig 2.18: Hox clusters percentage identity plots. Pag. 86
Fig 2.19: percentage identity plots. Pag. 87
Fig 2.20: adult tissues RT-PCRs. Pag. 92
Fig 2.21: fetal tissues RT-PCRs. Pag. 94
Fig 2.22: OS-EST alternative spliced forms. Pag. 96
Fig 2.23: long intergenic OS transcripts. Pag. 97
Fig 2.24: *HOXA* expression patterns in NT2 cell line treated with RA. Pag. 98
Fig 2.25: expression timing of both HOXA genes and OS transcripts in NT2 cells treated with RA Pag. 100
Fig 2.26: identification of OS transcripts in dispersed homeobox loci. Pag. 101
Fig 2.27: OS transcripts in the human TLX1 locus. Pag. 102
Fig 3.1: *Hox* genes colinearity: the "high order" chromatin model. Pag. 108
Fig 3.2: a mechanism underlying collinearity in mouse limbs. Pag. 112
Fig 3.3: nuclear compartments and gene regulation. Pag. 114
Fig 3.4: OS transcripts colocalized with *cic*-regulatory elements. Pag. 120
Fig 3.5: breaking the silence. Pag. 121

Chapter I:
INTRODUCTION

Part I:

Common aspects of epigenetic phenomena

“Genetic” and “Epigenetic” regulation of gene expression.

Genetics can be defined as the study of the inheritance of specific traits. More specifically, genetics deals with the analysis of genes, the carriers of heredity. For many years, it was thought that genes carry all the information, in the shape of DNA sequences, necessary to define phenotypes (the sum of all the traits we can observe). Many developmental biologists, indeed, thought that different phenotypes (i.e. different cell types in a multi-cellular organism) were just the result of gene loss during development (reviewed by Pennisi E., 2001). They propose, for instance, that muscle cells became muscle cells by loosing unnecessary genes, such as those involved in making liver or epithelial cells. This “genetic” view of development and cellular differentiation was accepted until the middle of last century, when it became clear that all the cells of a multi-cellular organism have the same sets of genes and none of them was lost during development or differentiation. These observations lead to the hypothesis that there should be other mechanisms, in addition to the genetic one, to explain the generation of different phenotypes from the same genotype. This mechanism was called “epigenetic”, that literally, from Greek, means “in addition to genetics”. With this meaning, the term “epigenetics” was coined by Conrad Waddington more than sixty years ago to describe:

> *“the study of the processes by which genotype give rise to phenotype”.*

As the complement of genes remains constant during differentiation, the epigenetic hypothesis propose that genes are “switched on and off” differently to make various cells in the body. In particular, different cellular phenotypes can be ascribed to different patterns of gene expression more than to genes themselves. To go back to the previous example, a muscle cell could be “epigenetically” defined by the expression of muscle-specific genes and by the silencing of liver- and epithelial-specific genes. In 1987, R. Holliday included this concept his definition of epigenetics:

> *“The changes in gene activity during development are generally referred to as epigenetic.”*
>
> (Holliday R., 1987).

He also added two supplementary definitions of epigenetics to include the stable transmission of these expression patterns from one cell to the daughter cells:

> *"study of the changes in gene expression, which occur in organisms with differentiated cells, and the mitotic inheritance of given patterns of gene expression"*
>
> *"transmission of information from one generation to the next, other than the DNA sequence itself... Nuclear inheritance which is not based on differences in DNA sequence"*
>
> (Holliday R., 1994).

C. Wu and J.R. Morris summarized recently all these findings looking at epigenetics as:

> *"the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence"*
>
> (Wu C., 2001).

As argued by the last two definitions, the DNA sequence itself could not represent the sole target for the epigenetic regulation of gene expression. As all the cells posses the same sets of genes, a "labeling system" should exist to mark differently genes that should be turned on, from those that should be silenced. Moreover, these epigenetic modifications, in order to be functional during development and cellular differentiation, should possess two main characteristics. They need to be maintained during cell divisions and to be erased when required. Recent studied revealed that these epigenetic marks are represented by covalent and reversible modification of both DNA and histone proteins associated with DNA. In particular these modifications are due to DNA methylation and histone tail modifications (such as acetylation, methylation, phosphorylation, ubiquitination, ADP-ribosylation, glycosylation (Jenuwein T., 2001).

The first associations between these types of modification and particular heritable states of gene expression derived from the observation that decondensed, transcriptionally active chromatin fractions (termed euchromatin) were enriched in acetylated histones, while highly-condensed and transcriptionally silent chromosomal domains (the so called heterochromatin) were found to be deacetylated (reviewed by Spotwood H.T., 2002). Moreover, it was shown that when an expressed (euchromatic) gene was place near or within heterochromatic domains, this gene could undergo to stochastic silencing that could be stably propagated through multiple cell divisions (reviewed by Karpen G.H., 1997). These evidences suggested that the link between epigenetic modifications and heritable gene expression pattern could be represented by dynamic changes in the chromatin structure. A general model hypothesizes that the same DNA element can be "labeled" in different ways in diverse tissues or during different developmental stages. These labels, maintained through different cell divisions, could be "read" by proteins and multi-protein complexes that promote *in cis* stable modifications in chromatin architecture at target loci. Long-term maintenance of gene expression could be due to the presence of a chromatin fiber more accessible to transcription factors and activators, while heritable gene silencing is thought to be the result of the establishment of a more closed, heterochromatin-like, configuration.

DNA and histone proteins modifications represent common aspects of many epigenetic mechanisms of gene regulation.

The chromosomal DNA associates with proteins and undergoes hierarchical packaging. The minimal structural unit of the chromosome, ubiquitous for all eukaryotes, is the nucleosome. It consists of a protein 'core' formed by two molecules of each of the histone proteins H2A, H2B, H3, H4 (histone octamer), and about 150 bp of DNA being wrapped around the histone octamer (Fig. 1.1). Two adjacent nucleosome core particles are connected by 15-55 bp of "linker" DNA (Luger K., 1997). This nucleosome array forms the "11 nm" chromatin fiber that could be folded, with the addition of another histone protein, H1, into the "30 nm" chromatin fiber. Moreover this fiber structure could be packed into large chromatin domains and chromosome territories (Belmont A.S., 1999). The comparative analyses of epigenetic phenomena (such as dosage compensation in fly, chromosome X inactivation in mammals; genomic imprinting; maintenance of *Hox* gene expression by Polycomb/trithorax proteins families; heterochromatin formation at centromeres and telomeres in yeast) lead to the identification that common epigenetic marks involve the two components of the chromatin fiber, namely DNA and histone proteins.

DNA methylation.

In vertebrate genomes, methyl group usually modifies the 5' carbon of cytosine in CpG dinucleotides and about 70% of these CpG sequences are methylated (Robertson K.D., 2000; Jones P.A., 2001). Interestingly, the CpG dinucleotide distribution in our genome is not random. In fact, CpGs are under-represented throughout the genome, a phenomenon termed "CpG suppression" which is probably due to the conversion of methylated cytosine into thymidine (Robertson K., 2000). CpG dinucleotides are clustered in CpG rich sequences called "CpG islands" mainly associated with promoters and other regulatory DNA as well as the first exon of many genes (Robertson K., 2000; Jones P.A., 2001).

Methylation of DNA is established and maintained by DNA methyltransferases (DNMTs). In mammals three different DNMT families have been identified: DNMT1, DNMT2 and DNMT3a/b (Bird A., 1999). DNMT1, in particular, possesses an important "epigenetic" characteristic. It has a 10/40-fold preference for hemimethylated DNA (Pradhan S., 1999) and, for this reason, is believed to be the enzyme responsible for copying methylation patterns after DNA replication. In proliferating cells this "maintenance" methyltransferase has been localized to replication foci (Leonhardt H., 1992), ensuring methylation of the daughter strand during replication. In addition to this function, DNMT1 as well as the DNMT3 family, possesses also the capability of inducing *de novo* methylation.

Although there are evidences of genes remaining active in methylated regions (Walsh C.P., 1999), DNA methylation is generally associated with silencing. Indeed a large number of tumor suppressor genes that are silenced by aberrant DNA methylations have been described (Tsou J.A., 2002). Generally this abnormal epigenetic modification takes place in CpG islands near or inside the promoter regions (reviewed by Momparler R., 2003). Many mechanisms have been proposed to explain the silencing induced by DNA methylation. The most "direct" mechanism suggests that CpG methylation could interfere with the local binding of transcriptional machinery or of general transcription factors. Indeed many transcription factors cannot bind DNA when their recognition sequences are methylated (Tate P., 1993). Nevertheless, this interference could not easily account for more wide-ranging gene silencing phenomena, such as, the chromosome X-inactivation. A more common feature in methylation-dependent regulation of gene expression is the recruitment of methylation–specific DNA binding proteins (reviewed by Bird A., 1999). The characterization of these methyl-binding proteins has given the key to understanding the connection between DNA methylation and transcriptional silencing. The "founder" of this family is the protein MeCP2, which contains both a methyl-CpG-binding domain (MDB) and a transcriptional repression domain (TRD). Interestingly, the TRD domain could recruit (through the interaction with the co-repressor Sin3A) histone deacetylase (HDAC) activities, which are known, as discussed below, to establish a repressive chromatin environment. Also others MBD-containing proteins present HDAC-dependent repression of gene expression (Bird A., 1999).

Histone modifications and the "histone code".

Another set of epigenetic modifications that regulates gene transcription involves histone proteins. Histones are small basic proteins that consist of a globular domain and a more flexible, charged NH_2-terminus (the "histone tail"; Luger K., 1997). The N-terminal 'tails' contain high amounts of basic amino acids such as lysine and arginine; these terminal tails are external to the nucleosome core structure (Fig. 1.1) and are therefore accessible for protein-protein interaction (reviewed by Sims R., 2003). Histone tails can be subjected to a number of post-translational modifications, such as acetylation, methylation (both discussed below) and phosphorylation, ubiquitination or ADP-ribosylation (reviewed by Spencer V.A., 1999; Fig. 1.1). These epigenetic modifications play an important role in the regulation of gene transcription as they can influence the chromatin architecture and by that DNA accessibility to transcription factors (reviewed by Strahl B., 2000).

Histones can be acetylated at specific lysine residues of histone H3 (K9/K14/K18/K23), of H4 (K5/K8/K12/K16/K20), as well as of H2A and H2B (Fig. 1.1). The association of acetylated histones with transcriptionally active chromatin was proposed soon after the recognition of this modification (Allfrey V., 1964) and is now well documented (Spencer V.A., 1999 and reference therein). Indeed several studies in various organisms show a positive correlation between acetylation of H3 and H4, in promoter regions, and transcriptional activity (reviewed by Grunstein M., 1997; Struhl K., 1998). This is consistent with other studies showing that histones corresponding to heterochromatin, transcriptionally inactive, are hypoacetylated (Spotwood H.T., 2002); in mammals, for example, the inactivated X chromosome is largely free of acetylated histones (Jeppesen P., 1993).

Histone acetylation is a dynamic process and two main factors are responsible for the maintenance of the acetylation status: histone acetyl transferases (HATs) and histone deacetylases (HDACs; reviewed by Kouzarides T., 1999). So far many transcriptional co-activators (i.e.: GCN5, TAFII250, pCAF, p300/CBP), have been isolated in different organism for their HAT activities (Spencer V.A., 1999). It has been suggested that the recruitment of HAT co-activators acetylates histone proteins leading to the destabilization of higher order chromatin structures (Roth S.Y., 2001). This is probably due to a decrease of the net positive charge of histone tails resulting in less condensed chromatin structure and increased accessibility of regulatory factors to DNA (Roth S.Y., 2001). HATs can be grouped into two classes: the "type A" HATs, localized in nuclei, probably acetylate nucleosomal histones in reactions linked to transcriptional activation, while "type B" HATs (represented solely by the yeast Hat1p protein) are purified from the cytoplasm where they are responsible for acetylating newly synthesized histones before chromatin assembly (Roth S.Y., 2001).

HAT proteins are generally found as component of large multiprotein complexes (i.e.: SAGA, NuA4, NuA3, Ada) whose composition is generally conserved among different species (Roth S.Y., 2001). Interestingly, some of the HAT complexes work in concert with chromatin remodeling complexes that promote gene expression by increasing promoter accessibility to the transcriptional machinery (Hassan A.H., 2001). In yeast, for example, histone acetylation could increase the binding of ATP-dependent SWI/SNF chromatin remodeling complexes at promoters (Hassan A.H., 2001).

Also the HDAC is a widespread protein family with various orthologs described in different species (Kou M., 1998). HDAC proteins are generally divided into two classes: the class I is made of HDACs that are components of large multiprotein complexes (such as yeast Rpd3, Hos1/2 or mammals HDAC1/2/3 found in the Sin3A and NuRD complexes), while the class II is made of those HDACs not found in multiprotein complexes (i.e.: the mammalian HDAC4/5/6 protein; Spencer V.A., 1999). Interestingly, both HDAC complexes and single HDAC proteins can be recruited to specific genomic loci by direct interaction with DNA-binding transcription factors; this recruitment leads generally to gene silencing (Davie J.R., 1998). Indeed the targeting of HDAC1/2/3 to a promoter by fusing HDAC to a DNA-binding domain (Gal4 DNA-binding domain) results in transcriptional repression of a Gal4-dependent reporter in transient transfection assays (Davie J.R., 1998). Moreover, as already mentioned, also MeCP2, via Sin3, could recruit the HDAC1/2 complex (Bird A., 1999), suggesting that DNA methylation and histone deacetylation are coupled events in the formation of repressive chromatin structures and gene silencing.

Another epigenetic modification concerns the methylation of histone tails by Histone Methyl-Transferases (HMTs). Histone methylation involves both arginine (R2/R17/R26 of H3, and R3 of H4) and lysine residues (K4/K9/K27/K36/K79 of H3, and K20/K79 of H4; Spencer V.A., 1999; Fig. 1.1). As more that one methyl group can be added to the same residue, mono- or dimethylated arginines and mono-, di-, or trimethylated lysines have been reported (reviewed by Bannister A.J., 2002). While arginine methylation is largely connected with transcriptional activation (Zhang Y., 2001), lysine methylation could correlate with both gene activation and gene silencing. Indeed, methylated K4 H3 was found at transcriptionally active chromatin in yeast (Noma K., 2001; Bernstein B.E., 2002) and chicken (Litt M.D., 2001), while high amounts of methylated K9 H3 were first detected in transcriptionally silenced domains of yeast (Noma K., 2001; Nakayama J., 2001). The trimethylation of K4 H3 defines a transcriptionally active state, whereas methylation of K9 H3 defines a repressed state of a chromosomal region. High amount of methylated K9 H3 have been found in heterochromatic domains of both *D.melanogaster* (Schotta G., 2002) and mammals (Peters A., 2001). These findings suggested that heterochromatin formation could involve the methylation of K9 H3. Indeed, in yeast, it has been show that this epigenetic modification of histone tails is essential for the assembly of heterochromatin at *mat* and *cen1* loci (Nakayama J., 2001). Moreover, K9 H3 methylated histones represent the binding site for the chromatin organization modifier (chromo)domain of the heterochromatic HP1 protein (Lachner M., 2001). This is a well-known protein characterized by its association with heterochromatic foci, such as centromeres, where it plays a fundamental role in gene silencing (Bannister A.J., 2001).

A model proposed to explain heterochromatin formation suggests that HDACs and HMTs (as well as other histone modifying enzymes) could cooperate to establish a silenced chromatin fiber. Recent findings (Nakayama J., 2001) show that HDACs could deacetylate K9/K14 H3 before K9 H3 methylation by HMTs. This modification is then responsible for the binding of HP1 proteins (Nakayama J., 2001). As HP1 could also interact directly with HMTs (Ekwall K., 1996), its binding to methylated K9 H3 could promote a "self-propagating" heterochromatin assembly (Fig.1.1).

The high diversity of histone modifications, as well as the high number of residues that can be modified within histone tails, lead to the hypothesis that specific combinations of histone modifications provide a "histone or epigenetic code", which after "translation" by downstream factors determines specific chromatin functions (Strahl B., 2000; Turner B., 2000; Turner B., 2002). Similarly to the binding of transcriptions factors to specific DNA sequences, different epigenetic modifications of histone tails could regulate gene transcription through the interaction of chromatin-associated proteins with nucleosomes (Strahl B., 2000; Turner B., 2000; Turner B., 2002). Referring to Bryan Turner words:

> *"These modifications are not just a means of reorganizing nucleosome structure, but provide a rich source of epigenetic information. It has been suggested that specific tail modifications, or combinations thereof, constitute a code that defines actual or potential transcriptional states."*
>
> (Turner B., 2002).

> *"This [epigenetic code] may be used both to mediate transient changes in transcription, through modification of promoter-proximal nucleo-*

> *somes, and for the longer-term maintenance and modulation of patterns of gene expression. The latter may be achieved by setting specific patterns of histone acetylation [...] across relatively large chromatin domains. [...] The code may be read (i.e. exert a functional effect) either through non-histone proteins that bind in an acetylation-dependent manner, or through direct effects on chromatin structure."*
>
> (Turner B., 2000)

Non-coding RNAs: a novel epigenetic mark contributing to gene regulation.

In addition to DNA and histone covalent modifications, many epigenetic processes are characterized by the involvement of non-coding RNAs (ncRNA; Andersen A.A., 2003). Phenomena such as chromosome X-inactivation (Plath K., 2002), dosage compensation in *D.melanogaster* (Franke A., 2000) or genomic imprinting in mammals (Reik W., 2001), require ncRNAs to regulate monoallelic gene expression. Other such examples come from *S.pombe*, where the formation of heterochromatin-like structures at *mating type* loci and centromeres is based on RNA-dependent mechanisms (Volpe T.A., 2002; Volpe T.A., 2003). Many evidences (reviewed by Andersen A.A., 2003) suggest that these ncRNAs could be seen as a new element of the "labeling system" that regulates gene expression promoting epigenetic chromatin modifications at target loci. These modifications could lead both to the assembly of heterochromatin-like configurations (as in chromosome X inactivation or genomic imprinting) with the consequent gene silencing, or vice-versa could sustain a more accessible chromatin structure (i.e. in *Drosophila* dosage compensation), thereby promoting gene activation.

X chromosome inactivation and *D.melanogaster* dosage compensation.

In mammals, despite the different X chromosome copy number between males and females, the level of expression of X-linked genes is maintained equal between the sexes as a result of a dosage compensation mechanism. This is achieved by selectively silencing most of the genes present on one of the two X chromosome, a process called X-inactivation (reviewed by Plath K., 2002). The inactivated X (Xi) exhibits a condensed

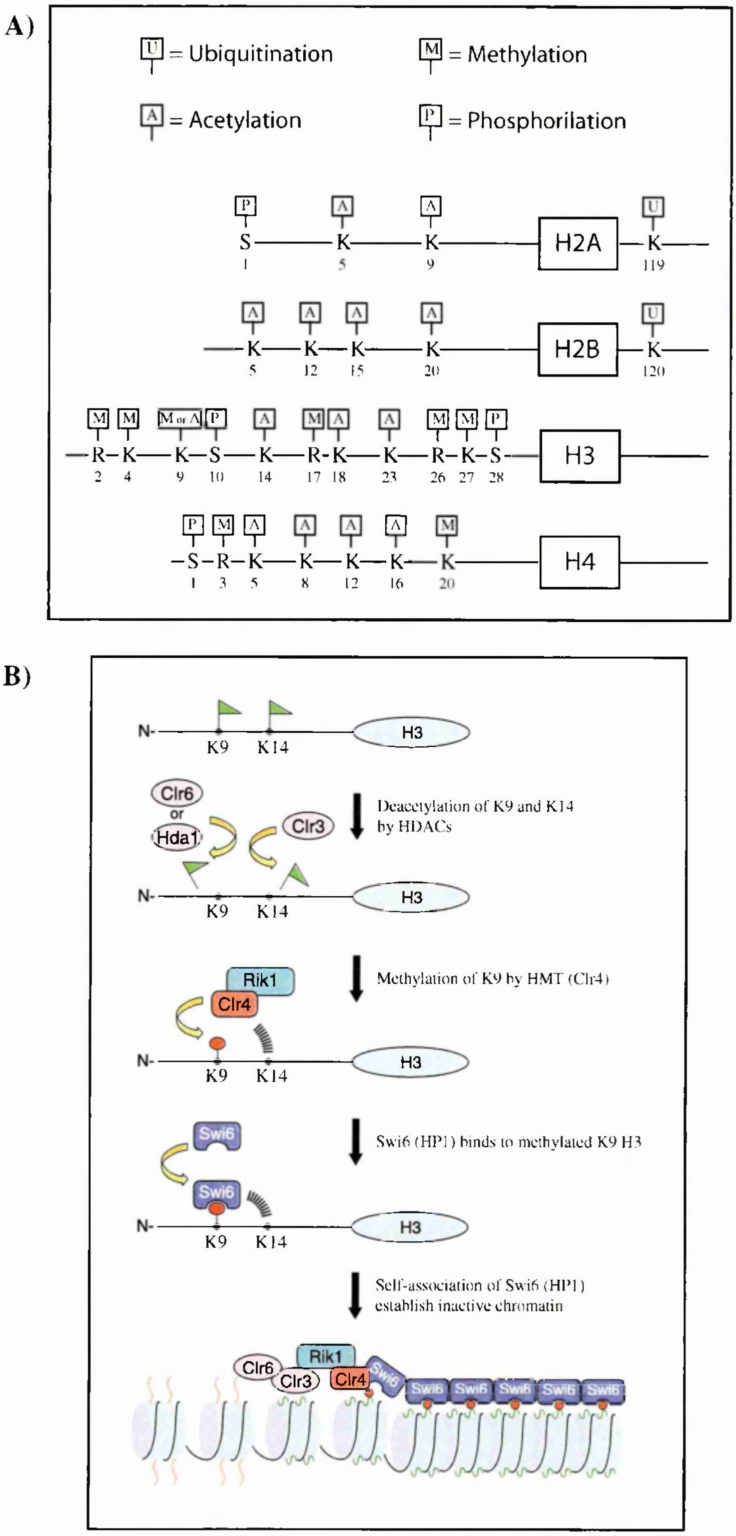


Fig 1.1:

A) Sites of post-translational modifications on the histone tails.

Adapted from Zhang Y., 2001; Richards E.J., 2002. K9 of histone H3 can be either acetylated or methylated. Additionally, within the globular domain of H3 K79 can be methylated (Feng Q., 2002).

B) Heterochromatin assembly leading to epigenetic silencing.

Adapted from Nakayama J., 2001. Green flags and red lollipops represent acetyl and methyl modifications, respectively. A nucleosome is composed of DNA (black line) wrapped around a histone octamer (blue). Orange or green protrusion represent the N-ter tails of histones with or without acetyl modifications, respectively.

heterochromatin structure, characterized by methylated K9 H3 and hypomethylation at K4 H3 (Boggs B.A., 2002; Heard E., 2001) as well as a diffused hypoacethylation of histone proteins (Jeppesen P., 1993; Belyaev N., 1997). Xi is also replicated late in S phase (Takagi N., 1974) and presents a high level of CpG island methylation (Wolf S.F., 1984). Interestingly, the initiation and propagation of the X inactivation is tightly linked to the presence of a specialized DNA elements, called the X inactivation center (Xic), which include the *Xist* gene (Heard E., 1997). *Xist* stays for "X inactivation specific transcripts" as it is transcribed only from the Xic of the Xi. This gene encode a 15-17 kb transcripts, that could be alternatively spliced and polyadenylated (Brown C.J., 1992). The *Xist* RNA posses two other main characteristics: is a non-coding RNA that has the capability to spread *in cis* from its site of synthesis to coat the Xi. Transgenic translocations of the Xic into autosomes revealed that the *cis* spreading of the *Xist* transcript was sufficient alone to induce the silencing of autosomal genes (Wutz A., 2000). This, unequivocally, proves that this ncRNA molecule is capable to induce chromosome-wide silencing. Interestingly, *Xist* is involved only in the initial establishment of X inactivation, that later on becomes *Xist*-independent (Plath K., 2002). This observation suggests also that the "silencing factors" working in concert with *Xist* should be developmentally regulated. Recently, these "silencing factors" have been identified and correspond to the protein complex EED/EZH2 (Silva J., 2003). As it will be discussed extensively in next sections, EED/EZH2 are two proteins both belonging to the Polycomb family. These proteins form a complex that contains HMT activity (specific for K9 and K27 of histone H3 and due to the SET domain of EZH2) and could interact with HDACs. The epigenetic modifications induced by the recruitment of HDAC and HMT activities are then responsible for the early establishing and subsequent spreading *in cis* of heterochromatic structures associated with a chromosome-wide silencing (summarized in Fig. 1.2). Additionally, it seems very likely that maintenance of the association between the Xi and HDAC-EED/EZH2 is *Xist* RNA-dependent (Silva J., 2003).

As we have seen, the transcription of the ncRNA *Xist* from the Xic of the Xi is the key event leading (i) to the recruitment of chromatin modifying complexes; (ii) to X inactivation and gene silencing. Thus we might consider the transcription through the

Xist gene as the first step of the epigenetic process resulting in X-inactivation. But, how can the cell discriminate between the Xa Xic and the Xic on the future Xi? The answer, once again, derives from the different transcription pattern of these DNA elements: a second ncRNA was identified in the Xic regions. The corresponding gene was called *Tsix* as it is transcribed (compared to *Xist*) with an antisense orientation. The expression of *Tsix* is able to promote *in cis* the choice of an X chromosome to become Xa, indeed the main function ascribed to *Tsix* is to lower *Xist* RNA steady state levels (Plath K., 2002) and various molecular mechanisms could be responsible for this *Tsix* regulation of *Xist*. It has been proposed, for example, that antisense transcription could decrease *Xist* gene transcription; alternatively duplex *Tsix/Xist* RNA molecules could interfere with the *Xist* RNA folding or be degraded by RNAi resulting in a deficit of both transcripts (Plath K., 2002). In mice, the asymmetrical persistence of *Tsix* expression is under the control of a cis regulatory element called *Xite* (X-inactivation intergenic transcription element, Ogawa Y., 2003). A deletion of this elements leads to a down regulation of *Tsix* expression, rendering the linked X chromosome more likely to be silenced (Ogawa Y., 2003). Interestingly, the *Xite* element itself can be transcribed, although intact intergenic RNAs are not required for its activity (Ogawa Y., 2003). It has been suggested that *Xite* could function as an enhancer element regulating *Tsix* transcription (Ogawa Y., 2003).

D.melanogaster has developed another epigenetic dosage compensation mechanism to ensure the equal expression of X-linked genes in males and females. In fly males, the transcription rate of all genes in the single X chromosome has been doubled (Franke A., 2000). Noteworthy, also the male dosage-compensated chromosome is “labeled” by the presence of ncRNAs transcribed from genes located on the X chromosome itself (Meller V.H., 1997). The hypertranscription of male X-linked genes results from the association to the X chromosome of a particular ribonucleoprotein complex called dosage compensation complex (DCC) or male specific lethal complex (MSL) consisting of six proteins (MSL1-2-3; MLE; MOF; JIL-1) and two non coding RNAs called *roX1*-*roX2* (reviewed by Andersen A.A., 2003). To promote the hyperactivation of the male X chromosome, DCC complexes coat the X chromosome and introduce epigenetic modifications of the chromatin fiber (Akhtar A., 2000). Indeed MOF contain a HAT activity specific for K16 H4 (Akhtar A., 2000), while JIL-1 is a protein kinase, which displays H3 kinase activity *in vitro* (Jin Y., 2000). These epigenetic modifications are thought to "open" the chromatin structure allowing an easier interaction between general transcription factors and regulatory regions of X-linked genes (Amrein H., 2000) The first association between DCC complexes and the X chromosome take place in correspondence of 30-40 genomic loci called "chromatin entry sites"; interestingly two of these sites correspond to the *roX1* and *roX2* loci themselves (Kelley R.L., 1999). Transgenic assays have shown that *roX1* and *roX2* genes are sufficient to attract DCC complexes to ectopic sites on autosomes (Kelley R.L., 1999; Meller V.H., 2002); these complexes could then spread *in cis* acetylating (K16) H4 in the flanking chromatin regions (Kelley R.L., 1999). This ectopic spreading correlates also with increased transcription in the flanking regions (Park Y., 2000). Recently it has been shown that DCC binding at *roX* chromatin entry sites is dependent on the presence of one of the *roX* ncRNAs (Park Y., 2003). As *roX* transcripts are unstable in the absence of Mls proteins (Meller V.H., 2000), it seems likely that *roX* ncRNAs should be quickly assemble into complexes locally, at their sites of transcription (Park Y., 2003). Moreover, Mls proteins become competent to bind *roX* chromatin entry sites only after their association with *roX* ncRNAs (Park Y., 2003). Although these evidences do not explain how the DCC complexes interact with the other chromatin entry sites, they imply that *roX* transcription, like for the *Xist/Tsix* locus, should be considered the first event leading to the *Drosophila* epigenetic dosage compensation.

Genomic imprinting.

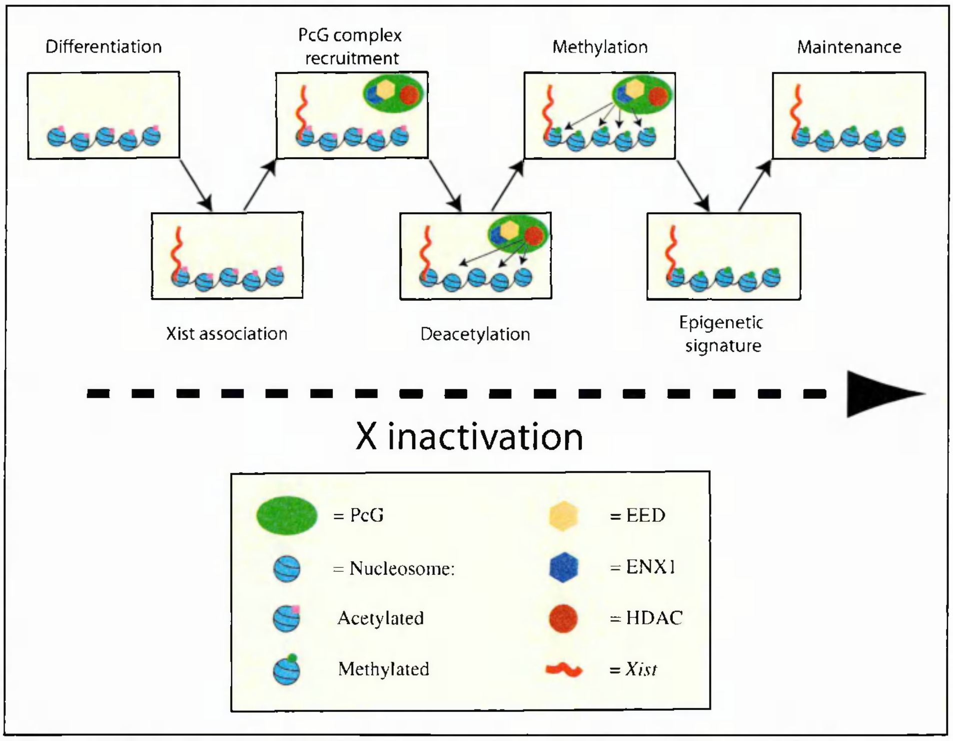
Tsix was the first example of a presumably ncRNA whose transcription is critical to negatively regulate the product of another gene. Recently, transcription of other antisense RNAs has been shown to be required for the regulation of imprinted genes. These genes represent a small group of genes, which are epigenetically "imprinted" so that only one of the parental alleles is expressed. In this case, the epigenetic modification coupled with silencing of imprinted genes is represented by DNA methylation. In fact endogenous imprinted genes are associated with *cis*-acting DNA elements (called imprinting centers, ICs) that are differentially methylated either on the maternal or the paternal allele (Ferguson-Smith A.C., 2001). ICs then could influence epigenetic modification of other *cis*-regulatory elements that regulate the expression of imprinted genes in allele-, tissue- or temporal-specific ways (Ferguson-Smith A.C., 2001).

Many imprinted regions are characterized by the presence of sense and antisense RNAs (Reik W., 2001). Recent findings suggest that these antisense RNA molecules regulate gene expression, as shown for the antisense ncRNA "*Air*" in the *Igf2r* locus. (Sleutels F., 2002). This RNA is a 108 kb, unspliced, repeat-rich ncRNA expressed only from the paternal allele (Fig. 1.2). Its promoter lies within the intron 2 of the Igf2r gene (another imprinted gene, maternally expressed), in a region genetically defined as "imprinting control element" (ICE). Interestingly, the paternal repression of *Igf2r* is due to the methylation of its promoter region. Thus the expression of the ncRNA correlates with the silencing of *Igf2r* on the paternal allele, a situation resembling the *Xist*/*Tsix* one. To test whether imprinted expression of *Air* ncRNA is required for silencing of *Igf2r*, truncated forms (*Air-T*) of the ncRNA were generated by the insertion of a polyadenylation cassette (Sleutels F., 2002). This *Air-T* allele retained a normal imprinted expression (namely, was correctly transcribed on the paternal allele), but the normal paternal repression of *Igf2r* was lost (Sleutels F., 2002). Interestingly, the bi-allelic expression of *Igf2r* correlates with a complete demethylation of his promoter. Thus the proper expression of *Air* is the first step leading to the epigenetic modification of the *Igf2r* promoter. Similarly to *Xist*, it has been suggested that also this ncRNA could mediate the recruitment *in cis* of repressor proteins that could induce a silent chromatin state. Interestingly, the same Polycomb proteins involved in the X inactivation seem to play a role also in the maintenance of the silencing of some imprinted alleles. Indeed, *eed*$^{-/-}$ mice (as discussed below) show the loss of imprinting with the subsequent bi-allelic expression of a subset of imprinted genes (Mager J., 2003).

Heterochromatin assembly requires repetitive DNA, ncRNAs and epigenetic modifications.

In addition to smaller heterochromatin domains found interspersed throughout

A)



B)

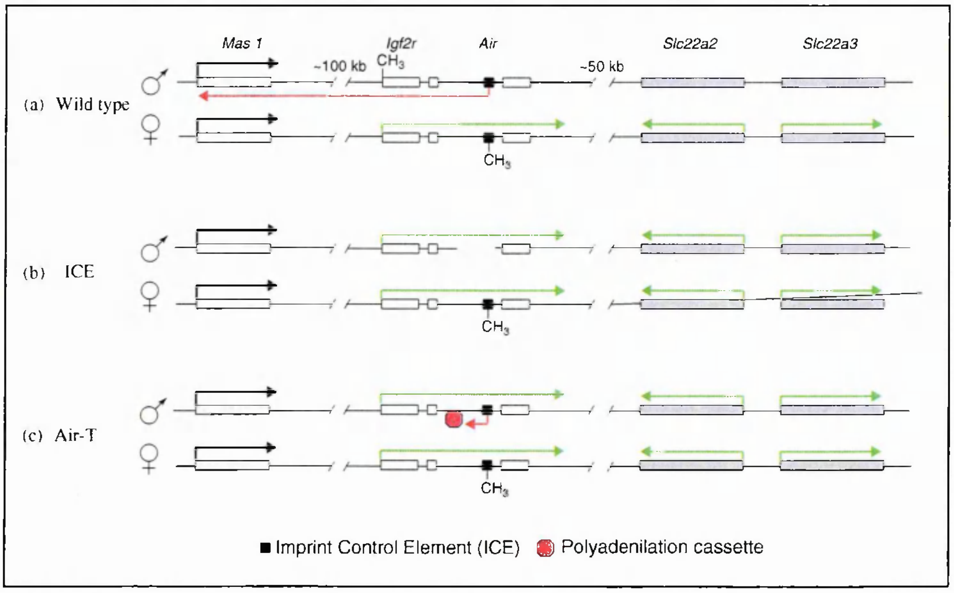


Fig 1.2: A) Chromosome X inactivation pathway. *Xist* expression and cis association with the future Xi recruits the Eed-Enx1-containing PcG complex to the Xi chromatin. A histone deacetylase (HDAC) removes acetyl groups from the lysine residues in the N-terminal tails of nucleosomal histone H3, exposing K9 and/or K27 for methylation by the SET domain of Enx1. The methylated Xi nucleosomes contribute to the signal to adopt a heterochromatic state and consequently underlie gene silencing on the Xi. Subsequently, heterochromatin of the Xi is stably inherited (adapted from Chadwick B.P., 2002).

Fig 1.2: B) Igf2r imprinted gene cluster. In wild-type mice(a), the imprinted genes Igf2r, Slc22a2 and Slc22a3 are expressed exclusively from the maternal allele (green arrows) and Air is expressed exclusively from the paternal allele (red arrow). The Mas 1 gene is upstream of Igf2r and is not imprinted. (b) Mice that inherit a deletion in the ICE on the paternal allele are not able to repress Igf2r, Slc22a2 and Slc22a3 expression on this allele. Due to deletion of the Air promoter, the defect in imprinting might result from the loss of Air expression on the paternal allele. (c) Mice that inherit a paternal Air-T allele express a truncated version of Air owing to the insertion of a polyadenylation cassette. These mice display the same phenotype as mice inheriting a paternal ICE deletion, owing to de-repression of Igf2r, Slc22a2 and Slc22a3 on the paternal allele. -CH3: DNA methylation.

chromosomes, large blocks of heterochromatin are generally located in both centromeric and telomeric regions. As centromeres and telomeres are made up of tandem array of simple DNA repeats, it was supposed that the highly packed heterochromatin fiber could work as a stabilizer of these chromosomal structures, i.e. preventing translocation and fusion of the chromosome ends. Apart from this "structural" role, pericentromeric heterochromatin posseses also the ability to silence gene expression: numerous evidences (reviewed by Karpen G.H., 1997) have shown that genes placed within fission yeast centromeres are transcriptionally inactivated. Interestingly mutants that erase centromere-induced gene silencing also interfere with chromosome segregation, suggesting a critical link between chromatin structure and centromere functions (Karpen G.H., 1997). Recently, it has been shown that the normal chromatin assembly, as well as gene silencing (Volpe T.A., 2002; Volpe T.A., 2003), at centromeres depends on a RNA based mechanism resembling those described for Xi silencing and imprinted loci.

The involvement of RNA molecules in heterochromatin assembly at centromeres was firstly deduced from the analysis of RNAi mutant from yeast (Volpe T.A., 2002; Volpe T.A., 2003). The "RNA interference" (RNAi) is a widespread process by which short double stranded RNAs inhibit the accumulation of homologous transcripts from cognate genes (reviewed by Hannon G.J., 2002). It is a typical example of a posttranscriptional gene silencing (PTGS) mechanism based on the sequence specific pairing of RNA molecules. Even if many aspects of this model are still unclear, analysis of *S.pombe* mutants has revealed that the disruption of the RNAi machinery results (i) in the transcriptional derepression of transgenes integrated in the pericentromeric repeats; (ii) loss of the typical epigenetic imprint for hetrerochromatin: H3-K9 methylation and, consequently, (iii) loss of the heterochromatin-associated Swi6 (*S.pombe* HP1 homolog) from the pericentromeric loci (Volpe T.A., 2002). Moreover, RNAi mutants lead to the abnormal accumulation of complementary ncRNAs transcribed from the DNA elements found at centromeres (Volpe T.A., 2002). Interestingly, double stranded RNAs have been identified also in assembly of mouse centromeric DNA (Rudert F., 1995). Additionally, H3-K9 methylation and HP1 binding to centromeres are abrogated in permeabilized mouse cells after ribonuclease treatment (Maison C., 2002).

As for other epigenetic phenomena, the starting point of this heterochromatinization process is the transcription of both DNA strands in the target locus with the production of complementary RNA molecules. As described in the model presented in Fig. 1.3, these double stranded RNAs (dsRNAs) could recruit the RNAi silencing machinery with the consequent degradation of dsRNAs into short, single strand RNAs that can pair to the target locus in a sequence specific manner. Based on different sets of evidences (reviewed by Jenuwein T., 2002 and Grewal S., 2003) the proposed "pairing hypothesis" might involve chromo domain proteins followed by the recruitment of histone modifying enzymes. Indeed: (i) the chromodomain is a RNA binding domain; (ii) in yeast the chromodomain of Clr4 is required for methylation of H3-K9; (iii) in Tetrahymena mutants of the chromodomain protein pdd1 abrogate H3-K9 methylation. Like for the model proposed in Fig. 1.1, HP1 may bind the methylated histones stabilizing and promoting the spreading of the heterochromatin fiber (Fig. 1.3).

Centromeric DNA is not the only repetitive DNA that is able to direct the assembly of a heterochromatin structure. Also transposable elements (TEs) could work as nucleation centers for the formation and subsequent spread of heterochromatin, in interspersed chromosomal domains. Recent evidence indeed suggests that the full repression of meiotically induced genes, in *S.pombe*, requires the presence of nearby TEs (Schramke V., 2003). Yeast strains mutated in the components of the RNAi machinery present a de-repression of meiotically induced genes, an accumulation of TE transcripts as well as a loss of dimethyl-K9 H3 and Swi6 from these elements (Schramke V., 2003). Moreover, targeted deletions of TEs revealed that they play a direct role in this repression (Schramke V., 2003). All these data lead to the hypothesis that interspersed TEs could direct gene silencing through the recruitment of the RNAi pathway and the subsequent formation of a silent heterochromatin (Fig. 1.3). Other evidences suggest that TEs from higher eukaryotes could utilize a similar silencing mechanism. Indeed, SINE and LINE elements contain promoters that can be transcribed from both RNA pol II and (predominantly) pol III (Weiner A.M., 2002). The mechanism of TE retrotransposition is itself based on the transcription of these interspersed DNA elements (Orstentag E.M., 2001). Interestingly, the close proximity of two TEs having a "tail to tail" orientation, could generate two complementary transcripts from these internal promoters that could give rise to a double stranded RNA molecule (Fig. 1.3). Additionally, the same results can be achieved with the simultaneous transcription of SINEs and LINEs from external promoters (Fig. 1.3). The production of double stranded RNAs from transposons has been described in different organisms and it is thought to promote transposons silencing (Sijen T., 2003; Weiner A.M., 2002; Aravin A.A., 2001; Jensen S., 1999) thus playing a protective or defensive role against TEs spreading. RNAi seems also to be involved in the protection of human cells from the infection of different types of retroviruses (Gitlin L., 2002; Jacque J., 2002).

As for pericentromeric repeats, the double stranded RNAs could be processed by the RNAi machinery inducing the subsequent recruitment of histone and DNA modification enzymes in the TEs-containing regions. Many observations seem to support this hypothesis. Chromatin immunoprecipitation experiments, with antibodies against H3 K9 methylation, revealed that *Alus* and LINEs are the major genomic target for this epigenetic modification (Kondo Y., 2003). The same experiments showed also that *Alu* flanking regions present H3-K9 methylation, suggesting that this modification could spread from these repetitive elements into surrounding regions (Kondo Y., 2003). Additionally, TEs correspond to hypermethylated DNA; this DNA modification has been correlated with the inhibition of their mobility or the silencing of their promoters (Robertson, K.D., 2000; Jones P.A., 1999).

The heterochromatinization of TEs (as for the centromeric and telomeric repeats) was seen as mechanisms, developed by evolution, to protect the genome from the deleterious effects of TEs transposition. Nevertheless, this RNA-based mechanism could have been used also to regulate (silence) genes flanking the TEs (Schramke V., 2003). As this phenomenon involves stable epigenetic modification of the surrounding chromatin, TEs as well as centromeric repeats could be seen as

> *"epigenetic elements" that control the stability of gene expression programs and organize heterochromatic domains [...]"*
>
> (Jenuwein T., 2002).

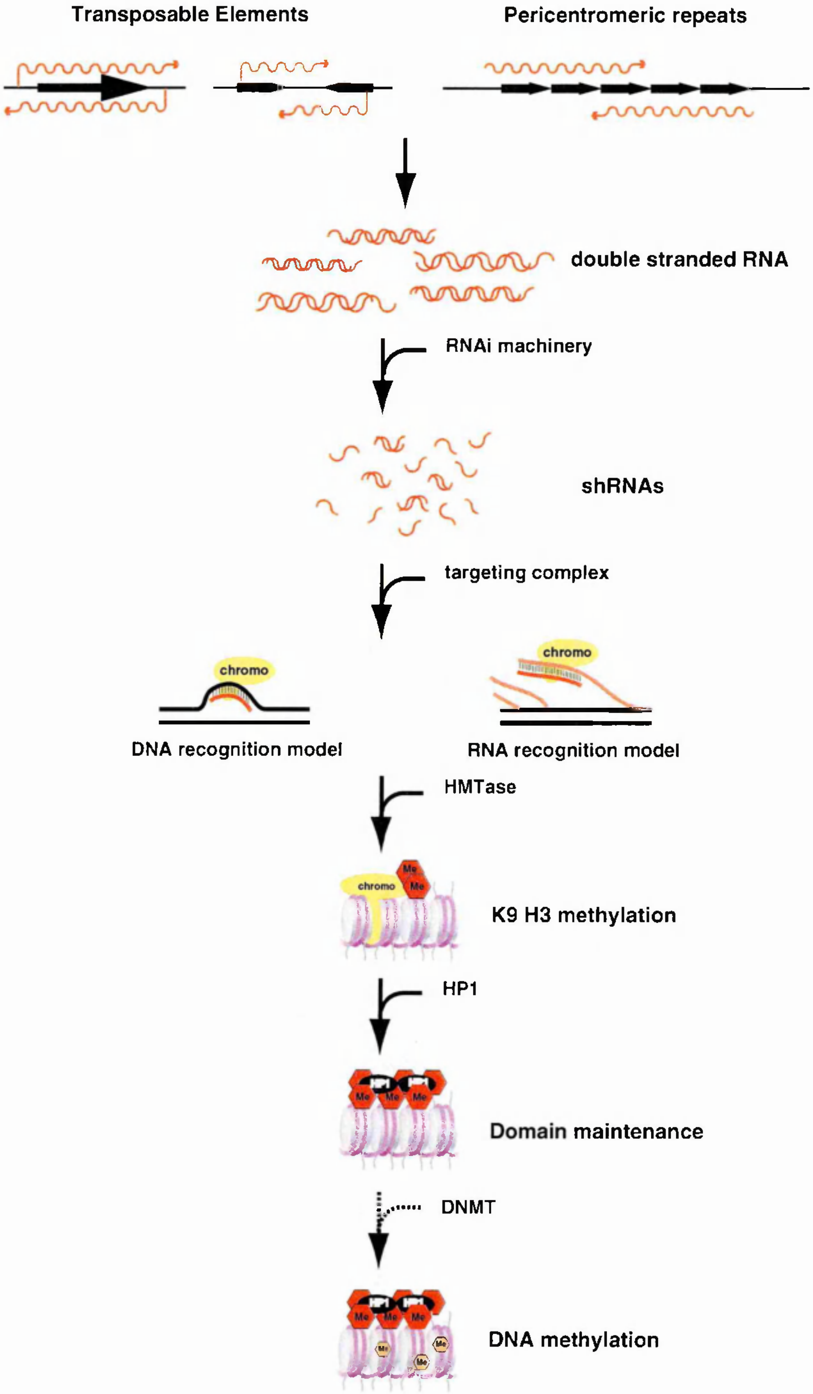


Fig. 1.3: DNA repeats might work as nucleation center for heterochromatin formation.

Many evidences suggest that pericentromeric repeats could promoter the formation of heterochromatin through a mechanism based on their transcription. One of the first steps of this process is the transcription on both strands of the pericentromeric repeats to generate double stranded RNAs (dsRNAs). Similar RNA molecules can also be obtained from the transcription of other repeats, such as Transposable Elements (i.e.: LINEs/SINEs) from external or internal promoters. Short heterochromatic RNAs (shRNAs) are then generated by the RNAi machinery that cleaves dsRNA. These shRNAs could drive the formation of heterochromatin through sequence specific mechanisms involving the recognition of DNA or nascent RNA. These complexes are then able to recruit HMTase activities that, modifying epigenetically the histone tails, induce the binding of heterochromatic proteins (i.e. HP1) and the spreading in *cis* of the heterochromatin fiber. The maintenance of the hetrerochromatin structures could also be enhanced by the subsequent recruitment of DNA methyltransferase (DNMT). (Figure adapted from Jenuwein T., 2002; Grewal S., 2003).

Chapter I: INTRODUCTION

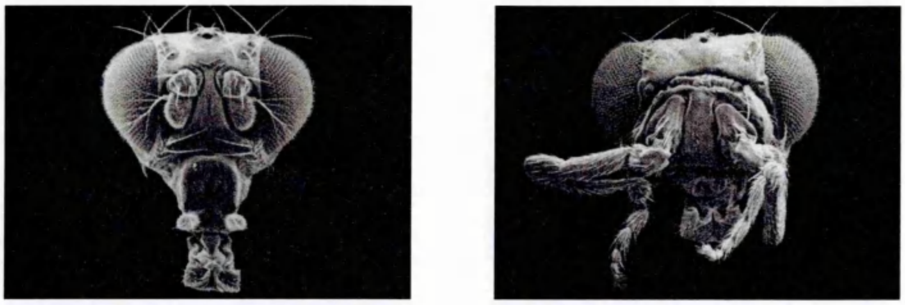
Part II:

Epigenetic regulation of *Hox* gene expression by the Polycomb and trithorax protein families.

Hox genes: an evolutionary conserved family of master control genes.

The metameric organization of the *Drosophila* embryo requires the proper spatial expression of a particular family of genes, the homeotic genes. These genes can be considered as master control genes that regulate the expression of many other genes and program specific developmental pathways. As a general class, the homeotic mutations result in the apparent expression of altered states of determination. For example, in the dominant mutation *Antennapedia* shown in Fig. 1.4, the antennae of the adult fly are transformed into legs that extend from the head of the fly. The correct pattern of homeotic genes expression is the result of a cascade that begins with the presence in the unfertilized egg of the products of the "maternal genes" and subsequently involves other genes families such as "gap" genes, "pair-rule" genes, "segment polarity" genes (Fig. 1.4). The analyses of *D.melanogaster* mutants revealed that homeotic genes are not required to form a specific part of the segmentation pattern but rather they assign a specific identity to these regions. Indeed, mutations in homeotic genes do not eliminate body structures but rather cause these elements to develop with inappropriate identities. Following W. Bateson definition of "homoeosis"- Fig. 1.4- these kind of transformations were named "homeotic phenotypes". Thus, we can compare the homeotic genes to genetic switches that turn on or off different programs of cellular differentiation.

Based both on chromosomal clustering and on HD sequence comparisons, the homeobox genes can be divided into three hierarchical levels: superclass, class and family (reviewed by Burglin T.R., 1994; Gehring W.J., 1994). In particular, referring to their chromosomal locations, the homeobox genes can be divided into two superclasses: the *Dispersed Superclass* and the *Complex Superclass*. The first *Superclass* collects all homeobox genes dispersed throughout the *D.melanogaster* genome, while the *Complex Superclass* comprises homeobox genes clustered in homeotic gene complexes (Fig. 1.5). In *D.melanogaster*, these genes are organized into two complexes, the *Antennapedia* complex (*ANT-C*) and the *Bithorax* complex (*BX-C*); together these two complexes form the so-called homeotic gene complex or *HOM-C*. Interestingly, gene complexes evolutionarily related to the *HOM-C* have been found in all the species analysed, including *H.sapiens* where four complexes (called *HOXA-D* complexes) have been identified



"The case of the modification of the antenna of an insect into a foot, of the eye of a crustacean into an antenna, of a petal into a stamen, and like, are examples of the same kind.
It is desirable and indeed necessary that such variation, which consist in the assumption by one member of a meristic series, of the form or characters proper to other members of the series, should be recognized as constituting a distinct group of phenomena...
I therefore propose... the term Homoeosis... the essential phenomenon is not that there has merely a change, but that something has been changed into the likeness of something else."

Bateson W., Material for the study of Variation, *1894.*

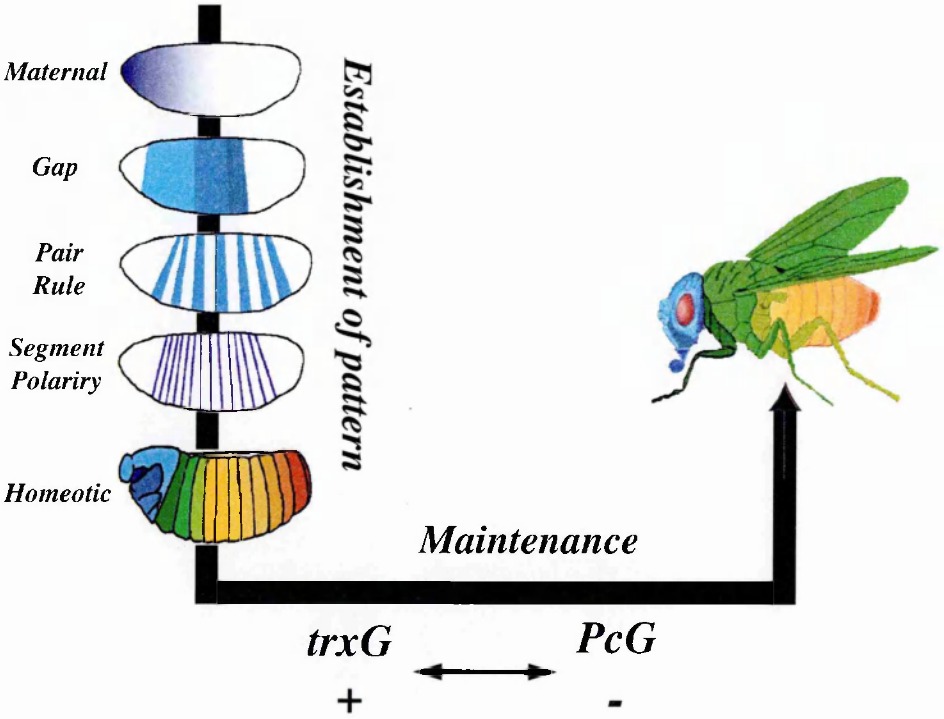


Fig. 1.4:
A) A homeotic mutation.
An *Antennapedia* (right) mutant head is compared to a wild type *D.melanogaster* (left). In the mutant, antennae are converted into leg structures by a mutation in the *Antennapedia* gene that causes its expression in the head. Fly images are from the "Interactive Fly" (http://sdb.bio.purdue.edu/fly).

B) Establishment and maintenance of homeotic gene expression in D.melanogaster.
See text for details.

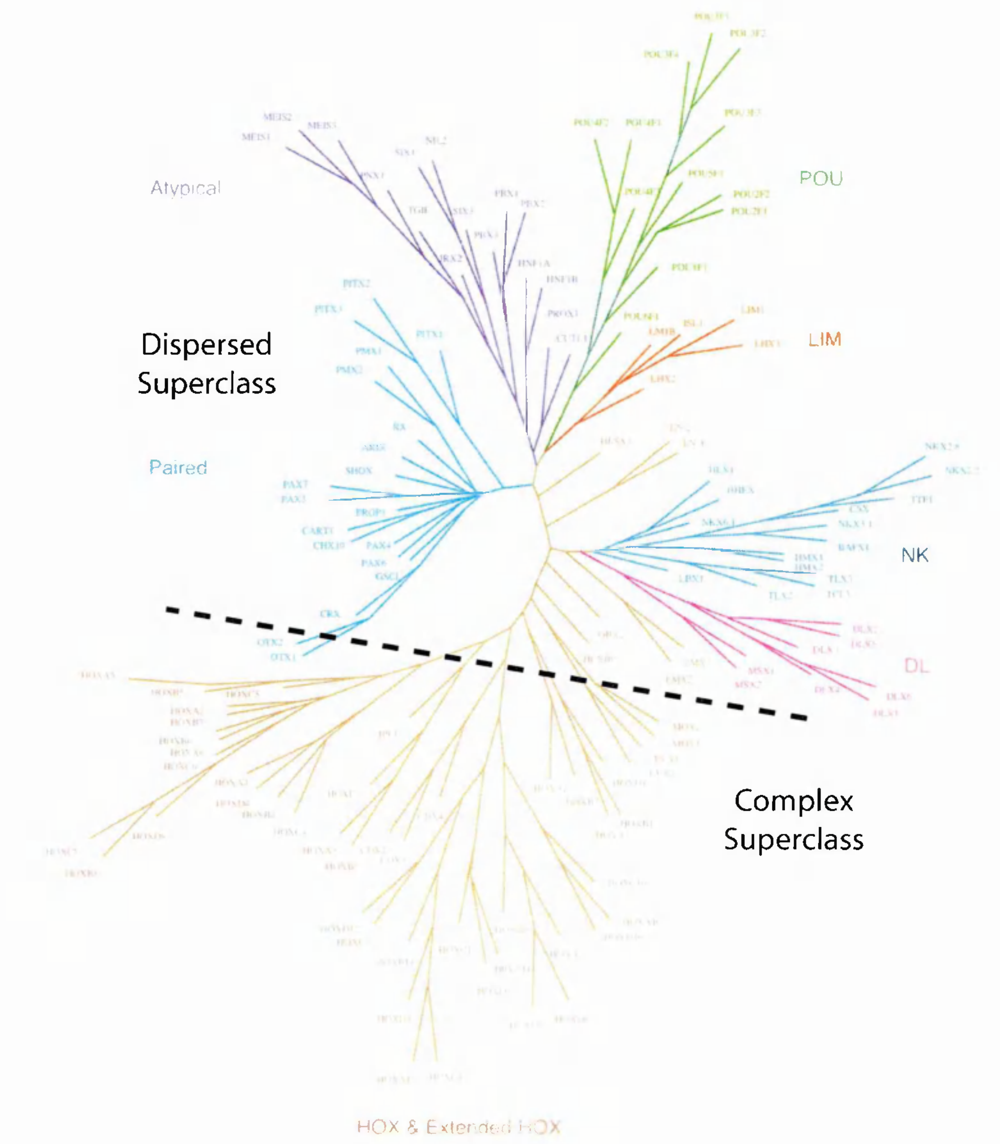


Fig.1.5: homeodomain protein families.

Phylogenetic relationships between members of the homeodomain protein families. A dashed line schematically divides homeodomain families in *Complex* and *Dispersed Superclasses* (adapted from Banerjee-Basu S., 2001).

(Fig. 1.6; Boncinelli E., 1991). The genes in the four vertebrate clusters can be aligned and subdivided into 13 groups on the basis of their HD sequences. These groups were named paralog groups (PG1 to PG13, Fig. 1.6).

One remarkable feature of the homeotic genes organized into clusters is that the order in which they are expressed along the antero-posterior axis of the *D.melanogaster* embryo reflects their physical disposition along the *HOM-C* complex (Fig. 1.6). Indeed, *Hox* genes located at the 3' extremity (corresponding to the first PG groups) of the cluster are activated in anterior embryonic domains whereas 5' located genes (the latest PG

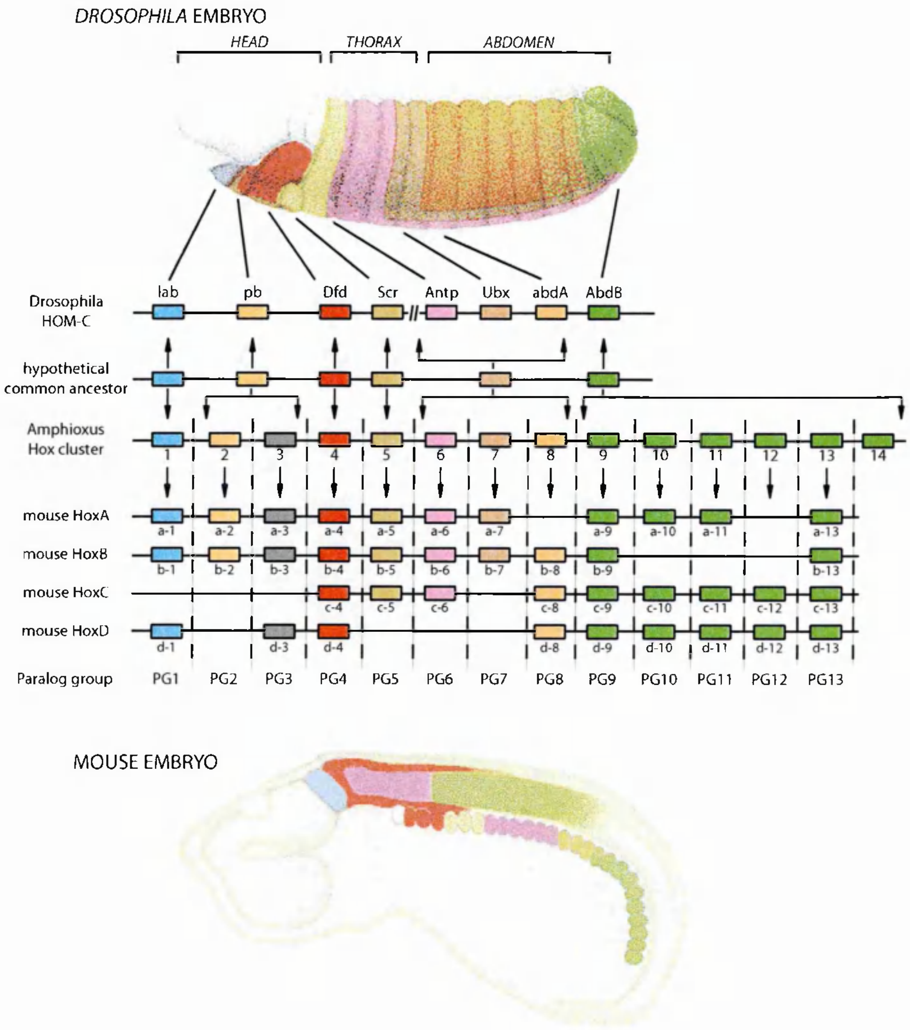


Fig. 1.6: *HOM-C/Hox* clusters organisation.

The evolutionary relationship between *Drosophila*, amphioxus, mouse *Hox* clusters, and the deduced complement of *Hox* genes in the presumed common ancestor of arthropods and chordates. The *Drosophila HOM-C*, the four murine *Hox* gene clusters (*HoxA -HoxD*) and the Amphioxus *Hox* complex are shown with their inferred phylogenetic relationship. A coloured box represents each *Drosophila* homeotic gene with its corresponding vertebrate paralog.group (PG1-PG13) Colours also indicate the expression domains in *Drosophila* as well as in mouse embryos. The anterior-posterior domains of *Drosophila* and vertebrate *Hox* gene expression correspond to the order of the genes within the complex.
HOM gene abbreviations: *Lab=labial*; *Pb=proboscipedia*; *Dfd=Deformed*; *Scr=Sex combs reduced*; *Antp=Antennapedia*; *Ubx=Ultrabithorax*; *abd-A=Abdominal-A*; *Abd-B=Abdominal-B*
(Adapted from http://www.press.uchicago.edu/books/gee/carroll1.html).

groups) are transcribed subsequently and in more caudal areas. This phenomenon, called "spatial colinearity", was originally described in *Drosophila* (Lewis E., 1978) and further extended to all animals exhibiting an anterior-posterior axial polarity (reviewed by Duboule D., 1998), including mammals (Gaunt S.J., 1988; Duboule D., 1988; Graham A., 1989). In vertebrates, there is also a "temporal colinearity", such that the most 3' genes have the earliest onsets of expression with a sequential activation of adjacent more 5' *Hox* genes. Interestingly, in mouse, *Hox* genes expression follows the colinearity rule not only in the trunk axis but also in limb development (Kmita M., 2002). In these secondary axes, genes located at the 5' end of the *HoxA* and *HoxD* clusters are expressed in distal regions (the future hands and feet), whereas genes located in the middle of these clusters (i.e. *Hoxd9*) are expressed in more proximal areas. The developing limb is also characterized by a "quantitative colinearity", namely a decrease in transcription efficiency going from the *Hoxd13* to the *Hoxd9* genes (Kmita M., 2002).

Maintenance of *Hox* genes expression through development: the *Polycomb* and *trithorax* genes families.

Once the expression patterns of the homeotic genes have been established, they should be maintained throughout the late developmental stages and entire life of the organism. As the segmentation genes, involved in the early establishment of homeotic gene expression, are expressed transiently in the developing embryo, the maintenance mechanisms should rely on the products of other genes. Indeed, in *Drosophila,* mutants have been described showing an ectopic expression in anterior segments of homeotic genes normally expressed only in more posterior segments (Lewis E., 1978). This misexpression results in a change of identity of an anterior segment into a more posterior one. Molecular analysis revealed that the initial pattern of homeotic gene expression was normal in these mutants and only late did this pattern degenerate to yield anterior expression (Kuziora M.A, 1988). These mutants led to the identification of a large protein family involved in the maintenance of homeotic gene repression: the Polycomb group (PcG). Its name derives from one of his members, the *Polycomb –Pc-* gene. One of the features of *Pc* null mutations is to transform the second and third leg of adult flies into the first one. In males the first legs present the sex combs, thus this transformation gave rise to the name of the gene. Later on all the mutants with a phenotype resembling the *Polycomb* one, were grouped together into the Polycomb group. PcG proteins are required to keep the homeotic genes silenced outside their normal expression domains. Once established, silencing by the PcG proteins is mitotically stable and maintained over many cell divisions. Whereas maintenance of repression involves the PcG proteins, another group of proteins is required to ensure continued expression of homeotic genes. This is the so-called trithorax group (trxG) since the *trithorax* gene is one of his members. *trxG* genes have been identified by several approaches. Some *trxG* mutants were recognized as they showed an anterior transformation of the abdominal segments, while most of the *trxG* genes were identified in genetic screens for dominant suppressor of *Polycomb* or *Antennapedia* mutations (Kennison J.A., 1988; Kennison J.A., 1995). *PcG* and *trxG* code for proteins that form multiprotein complexes and play an antagonistic role on the maintenance of *Hox* genes expression. The PcG proteins are required to keep silenced homeotic genes in those domain where they should be turned off; vice-versa the trxG maintains the "on state" of the homeotic genes. For these features, the PcG and trxG proteins are considered an integral part of the cellular memory system: they are able to transmit epigenetically repressed or active transcriptional state in a stable and hereditable manner.

Since these first mutant screenings, many *PcG* and *trxG* genes have been identified in *Drosophila.* Using data from two public databases ("the interactive fly" and the "flybase") I have collected the known *PcG* genes in Table 1.1 and the *trxG* genes in Table 1.2. However, the subdivision in PcG/trxG families is an oversimplification as there are genes having both PcG and trxG phenotypes (Brock H.W., 2001; LaJeunesse D., 1996; Milne T.A., 1999; Bajusz I., 2001). Due to these characteristics, these genes are suggested to form a third group of maintenance genes, called *Enhancer of trithorax and Polycomb* (*ETP*; Gildea J.J., 2000).

The PcG and trxG cell memory system is maintained through evolution.

The identification of *PcG* and *trxG* orthologs in as many different organisms as *C.elegans* (Ross J.M., 2003), vertebrates (summarized in Table 1.1 and Table 1.2;) and plants (Schumacher A., 1997; Alvarez-Venegas R., 2003), reveals that this memory system has been conserved through evolution. Mutations in *C.elegans PcG* orthologs present a deregulation of *Hox* genes expression leading to anteroposterior transformations (Ross J.M., 2003). Also *PcG* and *trxG* knock out mice show deregulation of *Hox* gene expression patterns with the consequent onset of homeotic phenotypes (van Lohuizen M., 1998; Hanson R.D., 1999).

Bmi1 and *Mel18* are two mouse gene othologs to *Psc/Su(z)2*. *Bmi1*$^{-/-}$ mice exhibit postnatal lethality due to neurological abnormalities and transformation along the entire axial skeleton (Hanson R.D., 1999). In these mice, many *Hox* genes expression boundaries were shifted anteriorly, inducing posterior transformations of vertebral identities (Fig. 1.7; Hanson R.D., 1999). Vertebral transformations were also found in *Mel18*$^{-/-}$ mice. Interestingly, *Bmi1*$^{-/-}$; M*el18*$^{+/-}$ and *Bmi1*$^{+/-}$; M*el18*$^{-/-}$ showed increased in posterior transformations of axial skeleton and rib cage compared to single knock-outs, with double mutants *Bmi1*$^{-/-}$; M*el18*$^{-/-}$ presenting even more severe transformations. These evidences suggest dosage compensation effects of *Bmi1* and M*el18* on survival of embryos, regulation of *Hox* gene repression and skeletal development (van Lohuizen M., 1998). Additionally, in these mutants, the early expression pattern of *Hox* genes was comparable to the wild type, while it was derepressed only in late development (van Lohuizen M., 1998). Thus, also in vertebrate, PcG proteins play a role in the maintenance and not in the initial establishment of *Hox* gene repression.

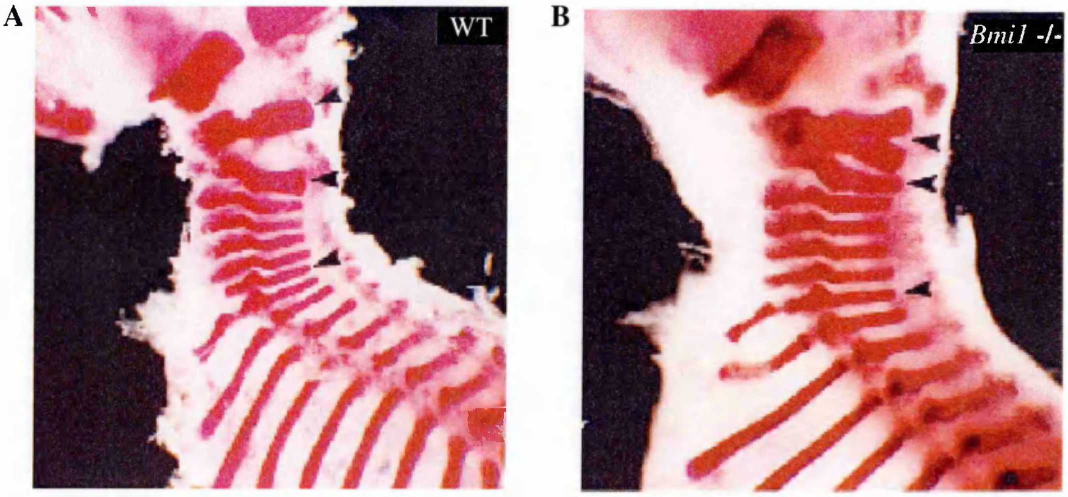


Fig. 1.7: Axial skeletal transformations in Bmi1$^{-/-}$ mutant mice.

The picture shows transformations in the cervical and thoracic vertebrae of *Bmi1* mutant mice. E18 fetuses of the indicated genotypes were stained with alizarin red; arrowheads mark the first, second, and seventh cervical vertebrae in wild type (**A**) and *Bmi1-/-* (**B**) mice. In *Bmi1* mutant mice, the second cervical vertebra is transformed into the third one and the last cervical into the first thoracic vertebra. Adapted from Hanson R.D., 1999.

M33 is considered a mouse ortholog of *Pc*, as it is able to partially rescue the *Pc* mutant phenotype (Muller J., 1995). *M33*$^{-/-}$ mice die soon after birth. Also these knock-outs present severe alteration in the axial skeleton, including exoccipital/atlas malformations, vertebral fusions and posteriorisation of thoracic vertebrae (Coré N., 1997). RNA *in situ* hybridizations revealed an anterior shift in *Hoxa3* and *Hoxc8* expression domains, while the expression of many other *Hox* genes (*Hoxd4*, *Hoxa5*, *Hoxc5*, *Hoxa6*, and *Hoxc6*)) was unaffected. The de-repression of specific subset of *Hox* genes (reviewed by Bel S., 1998) lead to the hypothesis that distinct PcG proteins (or different PcG multi-protein complexes) could regulate restricted groups of *Hox* genes. Double *PcG* knock-outs confirmed this hypothesis. *Bmi1*$^{-/-}$ M*33*$^{-/-}$ mice present an anterior limit of expression for *Hoxc8* and *Hoxc9* more severely affected than the single mutants, while other *Hox* genes were not affected (Bel S., 1998).

Embryonic ectoderm development (*eed)* is a PcG member (ortholog to *Drosophila esc*) with very peculiar functions. Its null phenotype is the more severe than any other murine *PcG* mutant analyzed. *Eed*$^{-/-}$ mice died at mid-gastrulation failing to develop a node, notochord and somites and without neural induction (Schumacher A., 1997). As this phenotype is manifested before *Hox* gene expression, PcG proteins may be involved in the control of other developmental processes that do not require *Hox* gene expression. Indeed recent data show that *eed* has a role in X inactivation both in extra-embryonic and in embryonic lineages (Wang J., 2001; Silva J., 2003). Moreover, *eed* has been identified as a trans-acting factor regulating autosomal imprinted genes during early development (Mager J., 2003: Ferguson-Smith A., 2003). Regulation of X-chromosome as well as autosomal genes silencing could represent new function for the PcG proteins acquired during evolution; their deregulation may also explain the lethal gastrulation defects presented by *Eed*$^{-/-}$ mice.

Mammalian orthologs of *trxG* genes have been also identified. *Mll1* (also called *All1*, *Hrx*) resembles *trx* of *D.melanogaster* and was identified originally as a gene commonly involved in chromosomal translocations leading to acute leukemias. *Mll1* gene targeting results in abnormal development of the skeletron and of hematopoietic cells (Yu B.D., 1998). Segment abnormalities were present in *Mll1*$^{+/-}$ mice with both anteriorization and posteriorization shifts of cervical, thoracic and lumbar regions identities (Yu B.D., 1995). *Hoxa7* and *Hoxc9* analyses revealed a caudal shift of their expression domains (Yu B.D., 1995). In particular, *Hoxa7* early expression pattern was not different between wild type and M*ll*$^{-/-}$ mice, while no substantial *Hoxa7* expression was detected in mutant mice later (Yu B.D., 1998). Thus the loss of late *Hox* gene expression in M*ll*$^{-/-}$ mice, indicates that M*ll* works as a maintenance factor.

In *D.melanogaster trxG* and *PcG* genes play an antagonistic role, indeed double mutants of *trxG* and *PcG* were described to restore a wild type-like phenotype (Ingham P., 1983). This feature seems to be conserved also in mammals, as axial skeleton transformation and altered *Hox* genes expression of both M*ll*$^{+/-}$ and *Bmi1*$^{-/-}$ mice were normalized when both genes were simultaneously deleted (Hanson, R.D., 1999). The normalization of axial defects in the double mutant was not complete, suggesting that these proteins do not universally co-regulate the same homeotic genes at all segmental level. The same study demonstrates that Mll1 and Bmi1 reciprocally regulate the expressions of many (but not all) *Hox* genes.

In addition to skeletal malformations and *Hox* genes deregulation, many *PcG* and *trxG* knockout mice develop different types of tumors, especially in the haematopoietic cell lineage (reviewed by Jacobs J.J.L., 2002).

Bmi1, *Mel18*, *M33* mutant mice display a strong reduction in overall T and B cell numbers as well as hypoplasia of spleen and thymus. These reductions suggest that there is a decrease in proliferation capabilities of T and B cell in these mice, rather than a block in their differentiation (Lessard J., 2003). This hypothesis is confirmed by the impaired proliferative response of these cells to several mitogens (Jacobs J.J.L., 2002). Notably, Bmi1 plays a fundamental role not only in maintenance of proliferation capacities of hematopoietic stem cells but also for the self-renewal of stem cells in the peripheral and central nervous system (Molofsky A.V., 2003). Overexpression of Bmi1 in hematopoietic cells results in a high predisposition to develop B and T cell lymphomas, thus confirming the role of this PcG protein as a positive regulator of hematopoietic cell proliferation (Haupt Y., 1993). Indeed, *Bmi1* was originally identified as an oncogene activated in B cell lymphomas (Alkema M.J., 1997). Vice versa, *Eed* $^{+/-}$ mutant mice exhibit hypoproliferation of bone marrow cells and an increased susceptibility to develop hematopoietic tumors (Lessard J., 1999). These phenotypes suggest that *eed* could work as a tumor suppressor gene.

The accumulating data connecting PcG proteins to the control of cellular identity and proliferation stimulated the search for alterations of PcG expression in human malignancies. In four cases of mantle cell lymphomas an overexpression of BMI1, due to *BMI1* gene amplification, was observed (Beà S., 2001). The overexpression of this gene correlates also with human high-grade osteosarcoma and resectable non-small cell lung cancer (Jacobs J.J.L., 2002). Recently, EZH2 overexpression was observed in hormone-refractory, metastatic prostate cancer (Varambally S., 2002). Both mRNA and EZH2 protein levels are increased in metastatic prostate tumors relative to localized or benign cancers suggesting that a deregulation in EZH2 expression may be involved in cancer progression. The authors report that when EZH2 is activated in tumor cells, a substantial number of other genes are shut down. If some of these genes code for tumor suppressors, their repression by EZH2 could accelerate the cancer progression towards metastasis.

The strongest connection between PcG/trxG members and human malignancies concerns the *mixed-lineage leukemia* gene (*MLL1/ALL1/HRX/Htrx*). More than 30 different translocations involving *MLL1* were described both in acute lymphoblastic (ALL) and acute myeloid leukemias (AML; Cimino G., 1998). As the fusion partners are unrelated, the current model proposes that the translocation products could interfere in a dominant manner with the *MLL1* normal function. Tandem duplication of *Mll1* gene portions have been also reported in AML patients with normal karyotype, suggesting that also MLL1 self-fusion could results in leukemogenesis (Caslini C., 2000).

PcG and trxG proteins act in large multiprotein complexes.

Two distinct observations led to the hypothesis that PcG and trxG proteins could exert their functions forming large multiprotein complexes. First, the synergistic effects of double *PcG* mutants on the homeotic phenotype suggest the existence of dosage interaction between the PcG proteins (as for the double mutant *Bmi1*$^{-/-}$ M*33*$^{-/-}$ discussed above). Second, different studies (Franke A., 1992; Rastelli L, 1993) revealed that many *Drosophila* PcG (such as Pc, Ph and Psc) proteins bind polytenic chromosomes in an overlapping pattern. Interestingly, only Psc was found at some of these loci, while at other loci only Ph and Pc were present (Rastelli L, 1993). These data suggests also that various complexes may exist with different PcG protein compositions. Further biochemical evidences and genetic (yeast two hybrid) screenings demonstrated the association of PcG and trxG into multiprotein complexes (reviewed by Francis N.J., 2001; Otte A.P., 2003). Mammalian PcG proteins BMI1, HPC2, HPC3, RING1, HPH1, HPH2 are reciprocally co-immunoprecipitated (Satijn D., 1997; Gunster M.J., 1997; Bardos J., 2000), colocalize in discrete nuclear bodies (Saurin A.J., 1998; Satijn D., 1997) and interact in a two hybrid system (Satijn D., 1999), suggesting that they are part of a large multiprotein complex. This complex was recently purified from HeLa cell extracts and called "Polycomb Repressive Complex from HeLa cells" (hPRC-H, Levine S.S., 2002). The same kinds of protein-protein interactions were also described in *D.melanogaster*, where an equivalent complex was similarly identified and named "Polycomb Repressive Complex 1" (PRC1, Shao Z., 1999). Biochemical analyses revealed that the "core" of this large (2-6 MDa) multiprotein complex, is made of Ph, Psc, Pc and dRing1 proteins (Francis N.J., 2001). In addition to PRC1/hPCR-H a second PcG complex has been identified both in flies and humans: PRC2. The PcG proteins forming the "core" PCR2 are conserved in both species: EED, EZH2, SU(z)12 (for the human complex); as extensively proven by co-immunoprecitation, colocalization and two hydrid experiments (Satijn D., 1999; Satijn D., 2001).

Genetic and phenotypic analysis of PcG mutants in *Drosophila* suggests different functions for these complexes. In particular, the complex consisting of Esc and E(z) (PRC2) appear to be crucial only early in development, soon after the *Hox* genes silencing has been established by gap and segment polarity genes. The second complex containing Pc, Ph, Psc (PRC1) vice versa is required in later stages to ensure that heritable silencing is maintained (van Lohuizen M., 1999). For these reasons, the two complexes were named PcGi (for PcG *initiating*) and PcGm (for PcG *maintaining*). Nevertheless, this schematic subdivision of PcG proteins into two complexes seems to be clearly oversimplified. The PcG protein Pho/YY1, for example, is found as a constituent of both PRC1 and PRC2 complexes in flies and vertebrates. Moreover, Pho/YY1 is part of the PRC2 complex from both young and older *Drosophila* embryo extracts, while is associated with the PRC1 only in extract from young embryos (Poux S., 2001). Thus the protein composition of the complex could vary in different developmental stages. Another source of variability is the tissue specific expression of many mammalian PcG proteins. When human fetal kidney was stained with antibodies against PcG factors, separate cell layers express different members of the PRC1 complex (Gunster M.J., 2001). Also human cell lines express completely different sets of PcG proteins (Gunster M.J., 2001). The fact that there are cell type-specific PcG target genes and cell type-specific PcG complexes led to the hypothesis that a specific Polycomb complex could regulate distinct target loci (Satijn D., 1999).

Genetic data suggest that also trxG proteins physically interact. Indeed, at least three trxG multiprotein complexes were identified from *Drosophila* embryonic extracts. The analysis of one of these complexes, called BRM, revealed the presence of the trxG protein brahma (BRM; Dingwall A.K., 1995). Another trxG protein, Moira (Mor) was found in this complex (Crosby M.A., 1999) while the majority of BRM associated proteins are not encoded by trxG genes (Papoulas O., 1998). The two further complexes were characterized for the presence of other trxG proteins, such as ASH1 and ASH2 (Papoulas O., 1998). As expected, related complexes (BRG1 and hBRM) were also identified in humans (Wang W., 1996).

PcG/trxG multiprotein complexes modify chromatin structure.

How can different PcG/trxG multiprotein complexes maintain the repressed/activated state of developmental regulated genes? The identification of enzymatic activities associated with components of PRC1/PRC2/BRM complexes start to give (still incomplete) answers to this question.

When the first *trxG* gene was analyzed, its sequences revealed an elevated similarity to the yeast SWI2/SNF2 (Tamkun J.W., 1992), that functions as the ATPase subunit of a 2 MDa chromatin remodeling complex, the SWI/SNF complex. Genetic and biochemical studies suggest that SWI/SNF activates transcription by remodeling nucleosome positions (reviewed by Sudarsanam P., 2000). These complexes can slide nucleosomes along DNA; transfer histone octamers to naked DNA and alter the topology of closed circular nucleosomal templates (Langst G., 2001). As a consequence of this chromatin remodeling, activators and transcription factors are probably helped in binding their target sites by the work of SWI/SNF complexes. This also suggested that the maintenance of *Hox* and other developmentally regulated genes mediated by trxG complexes could involve the formation of an "open" chromatin structure, more accessible to transcription machineries. Concordantly, many other studies have highlighted the interaction of trxG proteins with Histone Acetyl Transferases (HATs). Histone acetylation is an epigenetic mark associated with the transcriptionally active euchromatin, while inactive heterochromatin domains are consistently hypoacetylated (Jenuwein T., 2001). A "trithorax acetylation complex" (TAC1), containing trx and the HAT CBP/p300 proteins has been recently identified (Petruk S., 2001). TAC1 characterizations revealed that this complex is able to acetylate histone tails and is required for the maintenance (but not for the proper initiation) of the homeotic *Ubx* gene expression (Petruk S., 2001). Indeed, mutations either in *trx* or *dCBP* genes reduce the expression of *Ubx,* indicating a direct role for TAC1 (and consequently, histone acetylation) in the *Hox* genes expression. Additionally, the trx and CBP/p300 have been shown to interact with the cis regulatory regions of *Ubx* (Petruk S., 2001). These data lead to the hypothesis that trx complexes might maintain transcriptional activation promoting an epigenetic modification (acetylation of histone tails) in correspondence of specific *cis* regulatory elements.

On the other hand, accumulating evidence suggests that PcG complexes may promote the formation of highly packed, transcriptionally repressed chromatin fiber. A first indication derived from the sequence analysis of the Pc protein. Its amino terminus presents a chromo domain homologous to HP1 (Paro R., 1991). This protein is encoded by *Su(var)2-5*, a gene identified as a modifier of heterochromatin-induced position effect variegation (PEV). Mutations in the HP1 chromo domain result in a reduction of its silencing ability; while for Pc it seems to mediate protein-protein interaction (with other members of the multiprotein complex; Cavalli G., 1998a) as well as binding to methylated chromatin (Cao R., 2002; Czermin B., 2002; Müller J., 2002). Due to this similarity, it was proposed that Pc, together with other proteins, could use heterochromatin-like structures to keep genes repressed (Paro R., 1991). The same conclusions can be deduced from different studies showing that histone deacetylases (HDACs) could play a role in PcG function. As already mentioned, deacetylation of lysines, in the N-terminal tails of histone H3 and H4, is an important indicator of chromatin repression (Jenuwein T., 2001). EED and EZH2 indeed co-immunoprecipitate with HDAC1 and HDAC2 proteins from human cell line extracts (van der Vlag J., 1999). Additionally, EED-mediated repression was dependent on HDAC2 activity (van der Vlag J., 1999). The same kind of interaction was seen in *D.melanogaster*, where HDAC RPD3 is found in the PRC2 complex (Tie F., 2003). In human cancer line, EZH2-mediated transcriptional repression is dependent on HDAC activity, as the commonly used HDAC inhibitor (tricostatin A) could abrogate the silencing effects of EZH2 (Varambally S., 2002).

Interestingly, as shown for single *trxG* and *PcG* genes, the large multiprotein trxG and PcG complexes maintain antagonistic behaviors: PRC1 and SWI/SNF can compete each other for interaction with the nucleosomal template. PRC1 is capable of stabilizing *in vitro* a nucleosomal array to the effect of ATP-dependent remodeling by SWI/SNF complex (Shao Z., 1999). Despite the changes in the primary sequence and composition of the complex, also the hPRC-H is able to inhibit SWI/SNF-dependent remodeling of nucleosomal templates (Levine S.S., 2002).

Several evidences tightly link PcG/trxG complexes to regulation of the *Hox* gene expression based on epigenetic mechanisms. Among them, the similarity found between PcG/trxG and protein involved in the remodeling of the chromatin structure and, most important, the association of enzymatic activities, like histone acetylation/deacetylation and histone methylation, with PcG/trxG complexes. These enzymatic activities, as we have seen, are due to the interaction between PcG/trxG proteins with HDACs/HATs but may be also carried by PcG proteins themselves. For example, E(Z) posses a Histone Methyl Transferase (HMT) activity (Czermin B., 2002; Müller J., 2002). It has been shown that the E(Z)/ESC (PRC2) complex possesses HMT activity for both lysines K9 and (preferred) K27 on H3 (Cao R., 2002; Kuzmichev A.,2002). The E(Z) SET domain tri-methylates K9 and (probably) K27 (Czermin B., 2002; Müller J., 2002; Kuzmichev A.,2002). With the exception of DOT1p (van Leeuwen F., 2002), all reported HMTs that methylate lysine residues contain a SET domain, which includes the amino acids important for MTase function. This is true also for E(Z), as SET domain mutations impair HMT activity *in vitro* and *in vivo* (Müller J., 2002). In addition, PRC2 complexes containing these mutant forms of E(Z) fail to maintain the repressed state of *Hox* genes. (Müller J., 2002). Other evidences linking MTase activity to the SET domain derived from metastatic prostate cancer (Varambally S., 2002). The overexpression of EZH2 resulted in the repression of several genes as detected by DNA microarray analysis, and this repression was dependent on an intact SET domain (Varambally S., 2002). Interestingly, Pc, a member of the PRC1 complex, has a preferential binding for H3 methylated at K27, probably through its chromo domain (Cao R., 2002; Czermin B., 2002; Müller J., 2002). It is postulated that the E(Z)/ESC complex makes a methyl mark on the K27 of histone H3 and this mark is recognized by PRC1. This fits very well with the PcGi-PcGm complexes model where the E(Z)/ESC complex initiates the silencing and a subsequent interaction between the two complexes is needed for PRC1 to be able to maintain the silenced state (Cao R., 2002; Czermin B., 2002; Müller J., 2002).

Histone methylation and acetylation also correlate with trx-G mediated activation. TRX, found in the TAC1 complex (Petruk S., 2001), has HMTase activity specific for K4 of histone H3 in *D.melanogaster* (Czermin B., 2002). The SET domain of the human ortholog, MLL1, retains the same methylation activity over K4 H3 *in vitro* (Milne T.A., 2002). *In vivo* this correlation was described for the *Hoxc8* gene, a *Hox* gene directly regulated by MLL1. In wild type mice, the expression of the *Hoxc8* correlates with a hypermethylation of K4 H3 at its locus, while in *Mll1*$^{-/-}$ mice the Hoxc8 locus is hypomethylated and the gene unexpressed (Milne T.A., 2002). In addition, *Mll1*$^{-/-}$ cells transfected with a vector expressing MLL1 showed an enrichment of K4 H3 at the *Hoxc8* locus compared with cells expressing MLL1 lacking the SET domain (Milne T.A., 2002). Interestingly, also histone acetylation seems to play an important role in regulating *Hoxc8* expression. Indeed, *Mll1*$^{-/-}$ mice show a hypoacetylated *Hoxc8* locus, while wild type mice expressing *Hoxc8* present a hyperacetylated locus. MLL1 transfected in

Mll1$^{-/-}$ cells is able to re-establish the acetylation pattern at different positions of the *Hoxc8* locus (Milne T.A., 2002). A similar situation was described also for the *Hoxa9* locus, another gene positively regulated by MLL1. In particular MLL1, as a member of a big multiprotein complex, binds the *Hoxa9* promoter and induces K4 H3 methylation, histone H3 and histone H4 acetylation at the promoter region (Nakamura T., 2002).

ASH1 is another trxG protein that possesses HMTase activity specific for K4 H3, K9 H3 and K20 H4 (Beisel C., 2002). Again, this methylase activity is due to the presence in this protein of a SET domain (Beisel C., 2002). Chromatin immunoprecipitation (X-ChIP) experiments indicate that ASH1 is able to mediate *Ubx* gene (a natural target gene of ASH1) activation promoting methylation of K4 and K9 in H3 and K20 in H4. This trivalent methylation pattern was also able to prevent the binding of HP1, a heterochromatin-associated protein that generally binds methylated K9 H3, as well as of PcG proteins (Beisel C., 2002). Vice-versa the trivalent methylation patter mediates the recruitment of SWI/SNF-like chromatin remodeling complex through the binding of BRM and MOR, both trxG proteins (Beisel C., 2002).

Taken together all these results exemplify how the histone code could be read: ASH1 labels the chromatin methylating H3 at K4/K9 and H4 at K20. This methylation pattern facilitates binding of trxG complexes and inhibits binding of epigenetic repressor such as Pc and HP1. This implies that, while methylated K9 H3 alone could mediate the recruitment of the heterochromatic protein HP1 leading to transcriptional repression, the simultaneous presence also of methylated K4 H3 and K20 H4 brings the recruitment of remodeling complexes that maintain a transcriptional active state.

The integration of genetic and epigenetic data, suggests that the products of gap, pair-rule, segment polarity genes could mediate the interaction of early developmental multiprotein complexes (ESC-E(Z), ASH1) with their target loci. These complexes establish a specific pattern of histone tail modifications that is recognized by other multiprotein complexes (PRC1, SWI/SNF, TAC1) that promote the maintenance of repressed or activated states of target gene (Fig. 1.8).

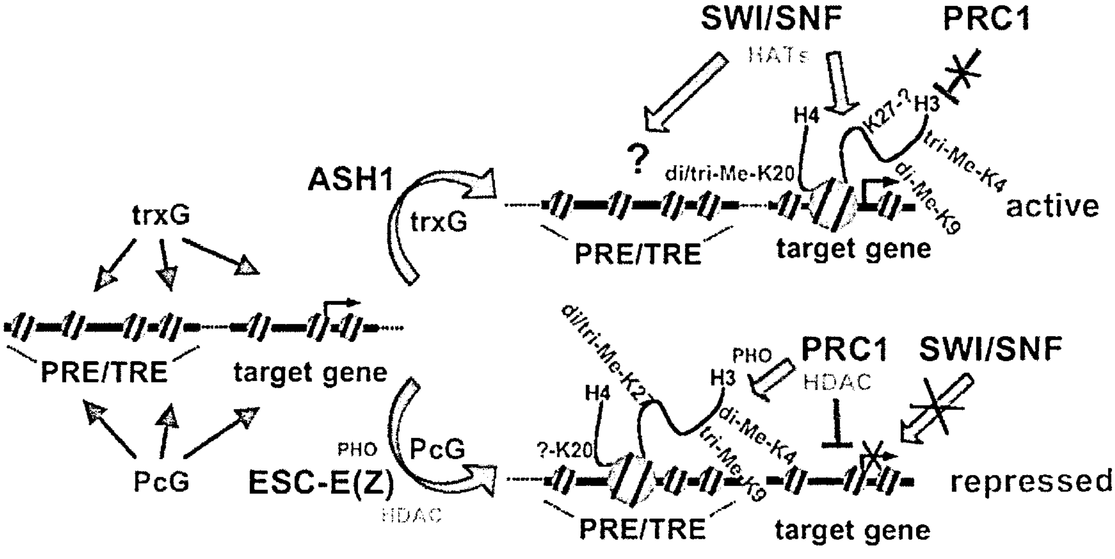


Fig 1.8: An epigenetic model for the action of PcG/trxG complexes.

On the left, a PcG/trxG target locus is schematically depicted. During development transient transcription factors, such as the products of the segmentation genes, interact with target genes promoting their activation or repression. In the first case, ASH1 marks the transcriptionally active promoter by methylating K20 of histone H4 plus K4 and K9 of H3. This, in turn, led to the recruitment of remodeling machines and histone acetyl transferase that are part of trxG complexes. These modifications prevent further interaction of the promoter with gene silencing activities (i.e. the PRC1 complex). In the second case, the ESC-E(Z) complex marks the transcriptionally repressed state by methylating K9 and K27 of H3 in the PRE, thus stably recruiting PcG complexes, such as PRC1. The association of both PRC1 and ESC-E(Z) could be concomitant with the recruitment of HDACs. Adapted from Breiling A., 2002.

PcG and trxG complexes bind specialized DNA elements.

PcG and trxG complexes exert their epigenetic regulation on gene expression by binding to the chromatin fiber. Different DNA elements were described as the *in vivo* binding sites for PcG and trxG complexes and for this reason were named Polycomb Response Element (PRE) and Trithorax Response Element (TRE).

Using immunostaining techniques more than 100 PcG/trxG binding site have been described on polytene chromosomes from salivary glands. This distribution implies that homeotic complexes represent only a subset of *PcG* target genes. An important aspect deduced from these profiles, is that many trx and Pc binding sites overlap (Chinwalla V., 1995). These were the first evidences suggesting that PcG and trxG binding sites could coincide at the same DNA element. Consequently, it was proposed to change the name of PREs and TREs into Maintenance Elements (MEs; Brock H.W., 2001) or Cellular Memory Modules (CMM; Cavalli G., 1998b). I will use generically the term PRE or CMM for both TREs and PREs.

PREs were identified and characterized at five genomic loci in the *D.melanogaster* (*BX-C*, *ANTP-C*, *engrailed*, *polyhomeotic* and *hedgehog*) by means of functional assays (i.e. transgene analysis) and chromatin immunoprecipitation. These analyses (reviewed by Pirrotta V., 1997; Brock H.W., 2001) revealed that (i) PREs are *cis* regulatory elements, whose length can range from several hundred base pair to some kilobases; (ii) the distance between the PREs and the target genes can vary up to 20 kb or even more; (iii) they can mediate the repression of a reporter gene in transgenic flies; (iv) transposon-mediated insertion of a PRE randomly in the genome is able to create an ectopic PRE that recruits PcG proteins; (v) PREs exhibit "homing" and "pairing" effects, namely transposons containing a PRE integrate near endogenous PRE sites.

The PREs belonging to the *BX-C* complex have been mapped by X-ChIP (Orlando V., 1993; Strutt H., 1997; Orlando V., 1998). A first analysis, done in a *Drosophila* cell line (Orlando V., 1993), revealed that Pc covered most of the *BX-C* including known PREs and promoters. Interestingly, no Pc binding sites were detected in the region corresponding to *Abd-B*, the only gene of the *BX-C* expressed in this cell line. The distribution of Pc over large *BX-C* fragments suggested that PcG protein could promote their own binding according to cooperative mechanisms that facilitate the spreading of these proteins along the chromatin fiber (Orlando V., 1993). An improved X-ChIP approach (Strutt H., 1997) confirmed the absence of Pc from the *Abd-B* region and revealed, by defining better the Pc distribution, that Pc binding sites correspond mainly to intergenic sequences previously identified as having PRE activities. In addition to intergenic sites, PcG and TrxG proteins were also mapped at core promoters, in embryos (Orlando V., 1998). An important consequence of this picture is that promoters themselves become key target regions for the mechanism(s) by which PcG complexes regulate transcription. PcG complexes/core promoter association was further analyzed in our lab (Breiling A., 2001). Surprisingly, many general transcription factors (GTF) were detected at the promoters of silenced genes regulated by PcG complexes (Breiling A., 2001). A confirmation of PcG complexes/GTFs interaction derived both from co-immunoprecipitation experiments (Breiling A., 2001) and biochemical purification of PcG complexes (Saurin A., 2001; Nakamura T., 2002). Thus, it was proposed that PcG complexes could maintain gene silencing through development by inhibiting GTF-mediated activation of transcription.

An important aspect of X-ChIP analyses was the confirmation that PcGs and trxG proteins could interact with the same DNA element (Strut H., 1997; Orlando V., 1998). Thus PcG and trxG proteins could act through common chromosomal element to direct the adjacent chromatin structure into either a heritable closed or open conformation. This leads to the possibility that the same DNA fragment could present a "double personality" possessing both a PRE and a TRE soul. Such a feature was demonstrated for the *Frontabdominal 7* (*Fab-7*) element, a DNA fragment containing both a boundary element and a PRE that are in vivo binding sites both for Pc and GAGA factor (Cavalli G., 1998b). The authors used transgenic fly lines with a heat shock–inducible *GAL4* gene as a driver construct and a reporter construct with *Fab-7* linked to a GAL4 UAS-inducible *lacZ* gene and *white* as a transformation marker. In the absence of heat shock, silencing imposed by *Fab-7* on the flanking reporter genes was dependent on PcG proteins, as heterozygous mutant *PcG* genes showed a release of white gene repression. A heat shock–induced burst of GAL4 strongly derepressed *lacZ* in all embryonic tissues. Strikingly, when embryos that had received a GAL4 burst were returned to normal temperature and allowed to develop, more than 70% of the resulting adult flies had red eyes. Thus the GAL4 pulse during embryogenesis can impose a mitotically stable reprogramming of the *Fab-7* CMM, from a silenced to an open chromatin state. This "switch" can be also meiotically transmitted (Cavalli G., 1998b). To interpret these sets of data, the authors propose that the transcription could drive the CMM from a silenced state (due to the presence of PcG factors) to a trxG-dependent open chromatin structure. The same authors later on demonstrated that different PcG proteins were tightly bound to the transgene even in the presence of trxG-mediated transcription (Cavalli G., 1999). The idea that PcG and trxG complexes might compete in a mutually exclusive way for the binding to the same DNA element seems to be oversimplified and probably the maintenance of the transgene transcription relies also on some other epigenetic marks. Indeed mitotically stable hyperacetylation of histone H4 was found at the activated *Fab-7* transgene (Cavalli G., 1999).

Promoters share similar sequence motifs with intergenic PREs and are "in vivo" binding sites for PcG complexes both in *D.melanogaster* and mammals.

The X-ChIP analyses of the PcG/trxG binding sites had the remarkable advantage of improving the resolution of PRE mapping from several hundreds of kb (immunolocalization on polytenic chromosomes) to several hundreds of bp (X-ChIP). This, in turn, favored the alignment of PRE sequences by looking for the presence of a common "binding consensus". Such a search was complicated by the fact that all, except three, PcG and trxG proteins don't have any known DNA binding domain. The exceptional proteins are PHO/YY1 (Pleiohomeotic, PcG), GAGA factor (GAF, PcG/trxG) and Z (Zeste, trxG). All three have sequence specific DNA binding motifs (reviewed by Ringrose L., 2003), that are present at least once in all the known PRE sequences. Unfortunately these consensuses are very short (GAF) or highly degenerate (PHO, Z) and cannot be used singularly as PRE/TRE-specific consensus. Nevertheless, these consensuses were used to develop a bioinformatic tool able to predict PRE elements in the *Drosophila* genome. The program, named "PREdictor" (http://www.techfak.uni-bielefeld.de/marc/pre; Ringrose L., 2003) is based on the evidence that known PREs contain multiple, clustered copies of GAF-PHO-Z binding motives. When used to scan the entire *Drosophila* genome, PREdictor identified 167 putative PREs (Ringrose L., 2003). The cytological positions of the predicted PREs are in excellent agreement (50%) with immunological mapping of PcG and trxG binding sites. Putative PREs were found in the genomic loci of different types of genes, such as transcription factors (including homeobox genes), genes linked to embryonic patterning or cell fate specification and genes having tumor suppressor activities. A major point emerged from this study is that all genes examined are characterized by the presence of a putative PRE in their promoter regions. Thus the presence of common DNA motifs both at intergenic PREs and promoters could explain the binding of PcG and trxG proteins at these regions and suggests that PcG/trxG complexes could directly bind promoters. Finally, *in vivo* binding of PcG and trxG proteins to putative PREs was confirmed with X-ChIP assay (Ringrose L., 2003).

Core promoters seem to represent PcG/trxG binding sites also in mammalian genomes (Milne T., 2002; Nakamura T., 2002). Genetic analyses revealed that the expression patters of many *Hox* genes were severely affected in *PcG* and *trxG* mutant mice, suggesting that these genes represent direct targets for the PcG/trxG activities (as discussed in previous sections). Indeed X-ChIP experiments showed that MLL1 binds directly the proximal promoter of *Hoxc8* (Milne T., 2002). This promoter and known 5' and 3' *Hoxc8* enhancers show MLL1-dependent acetylation of both histone H3 and H4 (Milne T., 2002). The same genomic regions were found to be K4 H3 methylated in wild type but not in *Mll*$^{-/-}$ mice (Milne T., 2002). Another *Hox* gene that is positively regulated by MLL1 is *Hoxa9* (Hanson R.D., 1999). X-ChIP experiments in HeLa cells expressing HOXA9 clearly demonstrated the association of MLL1 as well as other proteins found in its complexes (such as BRM, Mi2, TBP, TAFII80) with the core promoter of *HOXA9* (Nakamura T., 2002). Abolishment of MLL1 expression by small interfering RNA resulted in a loose of these proteins from the promoter, demonstrating that MLL1 is necessary for the association of the complex to this region (Nakamura T., 2002). MLL1 is also able to induce K4 H3 dimethylation and H3 acetylation at the *HOXA9* promoter (Nakamura T., 2002).

Apart from these two *Hox* gene promoters, PcG complexes were also found associated with pericentromeric heterochromatin (Saurin J., 1998). Immuno-FISH experiments demonstrated that HPC2, RING1 and BMI1 form discrete nuclear structures (termed "PcG bodies") stably associated with human pericentromeric regions of autosomal chromosomes (Saurin J., 1998). In particular, PcG bodies were found at the pericentromeric heterochromatin of chromosome 1 in a variety of cell lines (Saurin J., 1998).

Transcription through PREs correlates with epigenetic switches.

Many evidences suggest that PREs could work as epigenetic elements promoting either a silenced or activated chromatin state throughout development. As we have seen, this feature was largely investigated for the *Fab-7* PRE (Cavalli G., 1998b; Cavalli G., 1999). Nevertheless, other epigenetic elements from the *BX-C* (namely, *Mcp* and *bxd* PREs) present identical characteristics. Like for the *Fab-7* recombinant constructs, *Mcp* and *bxd* PREs were cloned upstream to GAL4 UAS controlling the expression of both *lacZ* and *white* genes (Rank G., 2002). Due to the association of PcG proteins to these

PREs, transgenic flies express neither *LacZ* nor *white* reporter genes (Rank G., 2002). Upon GAL4 induction during embryogenesis, LacZ was homogenously expressed in embryos and 90% of the corresponding adult flies have red eyes (Rank G., 2002). Together these results prove that (i) different CMMs posses both PRE/TRE activities; (ii) transcription through the target gene(s) could mediate a switch from the PRE to the TRE activity of the CMM; (iii) once this switch takes place, it will be stably maintained and transmitted through mitoses. Surprisingly, RNA *in situ* hybridizations revealed that the GAL4 pulse was able to induce not only the transcription of the two reporter genes but also of the PRE element itself (Rank G., 2002). The authors then looked for the presence of such transcripts also in transgenic flies having a truncated *Fab-7* element and that were not able to switch on the reporter genes upon GAL4 activation. "CMM transcripts" were not detected in such flies, suggesting the idea that transcription through these epigenetic elements is involved in switching a silent CMM into an activated one. If this last hypothesis is correct, we should expect to find CMM transcripts in those regions where CMM should be in an open chromatin conformation to promote the maintenance of homeotic genes transcription. Indeed RNA *in situ* hybridizations confirmed that transcription of CMMs coincides in *Drosophila* embryos with the transcription of homeotic genes they control (Rank G., 2002).

Other papers were concomitantly published describing the identification of intergenic non-coding RNAs in the *BX-C* (Bae E., 2002; Drewell R., 2002). These transcripts colocalized with peculiar *cis* regulatory elements, called *infraabdominal* (*iab*). At least seven *iab* elements have been mapped in the 100 kb intergenic region between *abd-A* and *Abd-B* in the *BX-C* and each *iab* region controls the expression pattern of the corresponding *abdominal* gene in a specific segment. Systematic RNA *in situ* hybridizations highlighted the presence of non-coding transcripts in each of these elements, including regions corresponding to known PREs (Bae E., 2002). Interestingly, the expression domains of *iab* non-coding RNAs overlap precisely with the segmental domains that are affected by mutations in each particular *iab* region. The intergenic transcription seems also to precede the transcription of those two homeotic genes in early embryogenesis (Bae E., 2002). It was then proposed that this early intergenic transcription could define the domains of activity for *cis* regulatory elements within each *iab* element. This hypothesis is reinforced from the analysis of *Mcp* mutants (Drewell R., 2002). *Mcp* is a dominant mutation due to a 3 kb deletion in the intergenic region between *iab4* and *iab5*. The corresponding phenotype (a posterior transformation of the fourth abdominal segment into the fifth) is due to the expression of *Abd-B* driven by *iab5* in the fourth abdominal segment, anterior to its normal expression domain. In the wild type embryo, the regulatory region *iab4* drives the correct expression of *abd-A* in the fourth abdominal segment. Deletion of the *Mcp* element results in the absence of intergenic *iab4* but not of the intergenic *iab5* and *Abd-B* transcription (Drewell R., 2002). These evidences suggest that *Mcp* could contain part of the promoter or initiation elements required for the intergenic transcription of *iab4*. If this transcription is required for the correct activation of *abd-A* *cis* regulatory elements present in the *iab4* region, then the loss of intergenic transcription may result in a failure to produce the fourth abdominal segment (Drewell R., 2002).

A general model can be drawn from these experimental evidences. The inactive state of a hypothetical gene could correspond to a situation where the gene itself and its cis-regulatory elements (i.e. enhancers) are embedded in a highly packed chromatin structure. This "close" structure prevents the interaction of trans-activator proteins with the enhancer. Thus we can suppose that intergenic transcription could represent a mechanism to "open" the chromatin structure allowing the interaction between *cis* regulatory DNA elements and regulatory proteins. In addition, we can even suppose that transcription through these DNA elements could modify epigenetically the chromatin structure.

ncRNA in eukariotic genomes.

Genome sequencing projects have revealed an unexpected feature of eukaryotes evolution: the developmental complexity of higher eukaryotes, more than the result of an increase of the total amount of protein coding genes, could be mainly due to variation in gene expression patterns (Mattick J., 2001a-b). Humans, for example, have only twice as many protein coding genes as *D.melanogaster* or *C.elegans* (International Human Genome Sequencing Consortium, 2001; Venter J.C., 2001) and share 99% of their protein coding genes with mice (Mattick J., 2001a-b). These analyses revealed also a second striking feature: the transcription of ncRNA is significantly increased during the evolution of higher eukaryotes (Mattick J., 2001a-b). About 98% of all the transcriptional output (or "transcriptome") in humans is represented by ncRNAs and half to three quarters of the transcriptome consists of antisense and intergenic transcripts (Mattick J., 2001; Yelin R., 2003). A similar situation was found also in the mouse transcriptome, where about 50% of RIKEN full-length cDNAs appear to be non-coding (Okazaki Y., 2002). Moreover, whole-chromosome analysis using oligonucleotide microarrays in humans has revealed that the level of transcription from both chromosome 21 and 22 is an order of magnitude higher than expected (Kapranov P., 2002). Many of these transcripts correspond to ncRNAs that could be polyadenylated, alternatively spliced and transported to the cytoplasm (Kapranov P., 2002; Cawley S., 2004). A high resolution co-mapping of these non-coding transcripts and transcription factor binding sites, revealed that the same transcriptional regulatory machinery could drive the transcription of both non-coding and protein-coding genes (Cawley S., 2004).

These findings suggest that RNA-mediated gene regulation could be a more common event than had been thought (Mattick J., 2001 a-b). In addition to ncRNA molecules such as *Xist/Tsix*, *roX1/roX2*, *Air*, a new class of ncRNAs, called microRNAs (miRNAs), supports this hypothesis. Firstly discovered in *C.elegans* (Lee R.C., 1993; Wightman B., 1993; Reinhart B.J., 2000), miRNAs are 21-24 nucleotide-long RNAs cut from longer (60-80 nucleotides) hairpin-shaped RNAs that bind to specific mRNA targets, blocking *"in trans"* their translation into proteins (reviewed by Bartel D.P., 2004). Interestingly, the miRNAs-based gene silencing is due to the recruitment of the RNAi machinery that could either promote the degradation of target mRNAs or the block of their translation (Bartel D.P., 2004). Computational and cloning approaches have led to the identification of orthologous miRNAs in different eukaryotic genomes, suggesting that this class of ncRNA has been conserved during the evolution of eukaryotes (Lim L.P., 2003). Noteworthy, miRNA genes should correspond to ~1% of the predicted human genes, a fraction similar to that seen for genes coding for transcription factors (Lim L.P., 2003).

The presence and the functional characterization in our transcriptome of such a large amount of ncRNAs lead to the hypothesis that ncRNAs could represent an unexplored layer of gene regulation. Nowadays ncRNAs seem to be involved in the onset of networks regulating gene expression with almost infinite potential complexity (Mattick J., 2001a-b). Referring to John Mattick's words:

> *"Complexity is hidden in the non-coding output of the genome"*
>
> (Dennis C., 2002).

Chapter II:
RESULTS

Identification of PREs in the human HOXA cluster.

A major goal of this study was the identification of PREs in the human genome. As many PREs have been mapped in the *BX-C* of *D.melanogaster* (Orlando V., 1993; Orlando V., 1998), I looked for the distribution of Polycomb and trithorax proteins in the human loci that are orthologous to the *BX-C*, namely the *HOX* clusters. In particular, I focused my attention on the *HOXA* cluster that was the only fully sequenced among the four human *HOX* clusters at that time. Following the experimental approaches used for the *D.melanogaster*, we decide to investigate PRE distribution by means of Chromatin Immunoprecipitation assay, using the immunoprecipitated DNA (I-DNA) as a probe for Southern analyses. Therefore, I started defining the genomic organisation and assembling a cosmid contig of this target region.

HOXA cluster: from *in silico* cloning to cosmid contig.

The *HOXA* cluster is made of 11 genes that localize in the short arm of chromosome 7 (7p15; Apiou F., 1996). We deduced its genomic organization *in silico* via both the Non Redundant (nr) and the Unfinished High Throughput Genomic Sequences (htgs) BLAST databases. These databases were screened with cDNA sequences from different human *HOXA* genes (see Table 2.1 for their accession numbers) and three PAC/BAC and eight cosmid clones spanning the entire cluster were identified (Kim M.H., 1998). Different DNA markers (Table 2.2) belonging to this region were used to define the cosmid contig by PCRs, restriction analyses and Southern blots. This cosmid contig is shown schematically in Fig. 2.1. The DNA sequence of the entire *HOXA* locus (more than 300 kb) was inferred from the sequences of the PAC/BAC clones. The EcoRI restriction pattern from the cosmid contig was compared with the restriction map obtained *in silico*, confirming that the entire *HOXA* cluster was cloned without any gaps.

Repeatmasker Web Server (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) was used to identify and localize all the Repetitive Elements (REs) in the *HOXA* genomic sequence. The analysis includes repetitive DNA families such as LINEs, SINEs, LTRs, MERs and satellite DNAs (i.e. simple repeats and low complexity DNAs). CpG islands were mapped using the PipMaker (http://bio.cse.psu.edu/pipmaker) and NIX programs (http://www.hgmp.mrc.ac.uk). The detailed genomic organization of the human *HOXA* locus is shown in Fig. 2.2.

A)

Marker/Gene	PAC DJO170O19	PAC DJO167F23	Cos 248G5	Cos 248C6	Cos 230G4	Cos 64D8	Cos 81G6	Cos 11E9	Cos 83F7
SWSS646	●		●	●					
HOXA13	●				●				
D7S1903	●				●				
HOXA11	●				●				
HOXA10	●				●	●			
D7S2774	●				●	●			
HOXA7	●					●	●		
SWSS2109	●					●	●		
HOXA6	●					●	●		
HOXA5	●					●	●		
HOXA4	●						●		
D7S2834	●						●	●	
HOXA3		●						●	
HOXA2		●						●	
HOXA1		●						●	●
SWSS2757		●							●
SWSS2243		●							●
SWSS3140		●							

B)

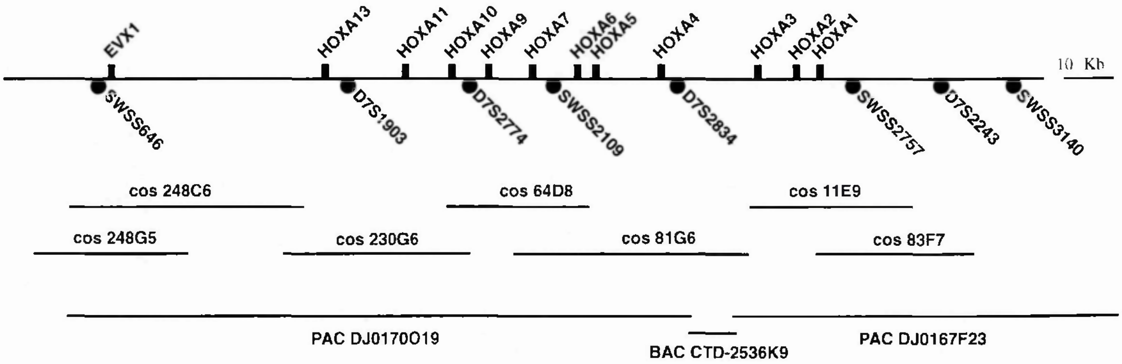


Fig. 2.1: Cosmid contig spanning the human HOXA cluster.

A) Schematic representation of DNA markers and HOXA genes used to assemble the cosmid contig. The order of genes and markers reflects their relative positions along the HOXA locus. Table 2.2 and Table 2.20 summarises PCR conditions specific for each DNA marker/HOXA gene, respectively. Genes/DNA markers belonging to different clones are shown as black dots.

B) Physical map of the human HOXA cluster. On the genomic line, genes are shown as black rectangles, while DNA markers as black circles. PAC, BAC and cosmid clones are shown as black lines below the genomic line.

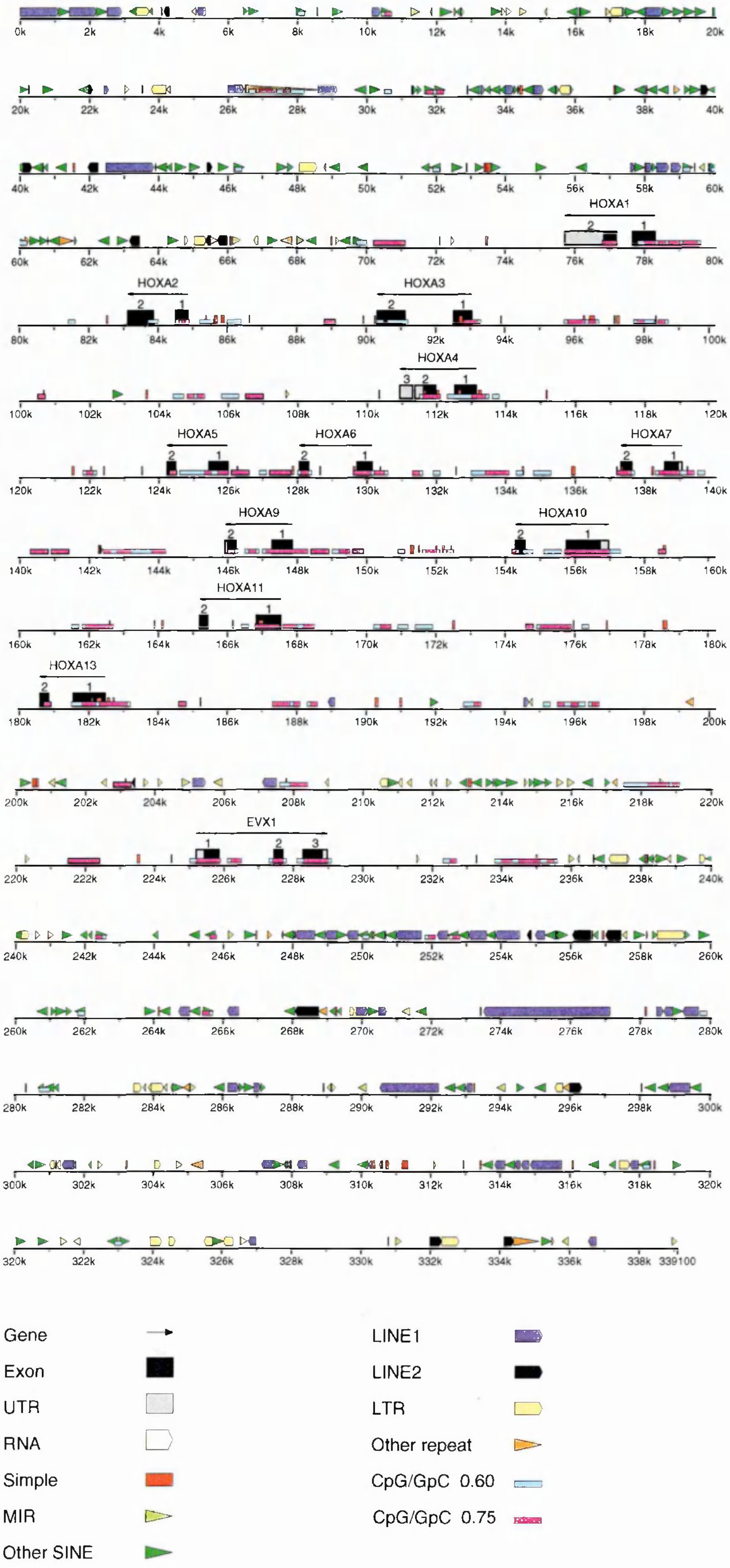


Fig. 2.2: genomic organisation of the human HOXA cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Kilobase (k) markings refer to our genomic contig.

Polycomb, trithorax and *HOXA* gene expression in human cell lines.

After having adapted the X-ChIP analysis to human cell lines (as described in Chapter IV), we started the analysis of the expression of different PcG and trxG proteins in HeLa and U937 cells. The expression pattern was investigated by Western Blot analysis using affinity-purified α-hPC2, α-BMI1, α-RING1, α-EZH2, α-HPH1, α-MLL1, α-YY1 and α-TBP. All proteins are expressed in both human cell lines (Fig. 2.3); MLL1 is also expressed in HeLa cells (C. Caslini, personal communication).

I investigated the expression patterns of *HOXA* genes in HeLa cells by RT-PCR. Table 20 summarised both primers and PCR conditions. From this analysis, *HOXA1*, *HOXA2*, *HOXA3*, *HOXA4* and *HOXA13* seem to be not expressed, vice versa *HOXA5*, HOXA6, *HOXA7*, *HOXA9*, *HOXA10* and *HOXA11* are expressed in HeLa cells (data not shown).

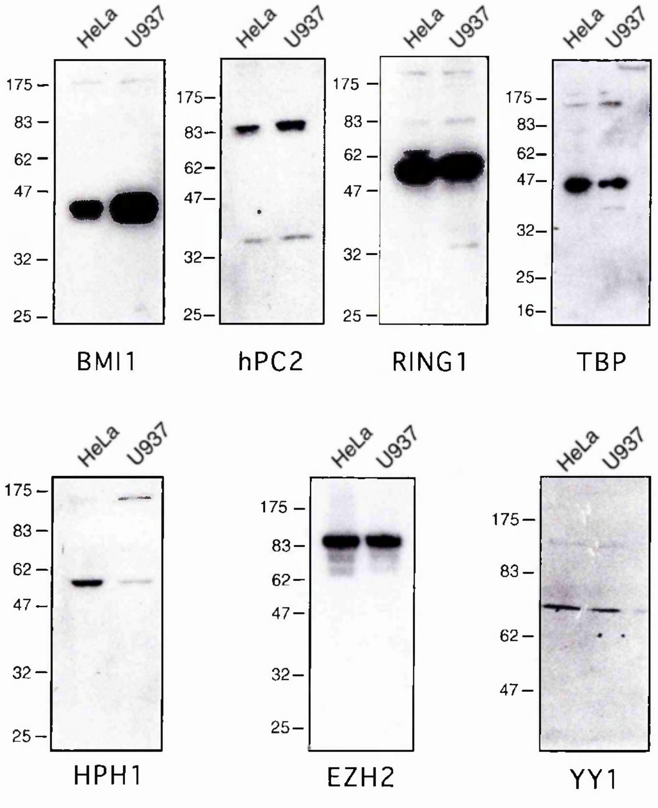


Fig. 2.3: Western blot analysis.

Western blot experiments were performed with total nuclear extracts from HeLa and U937 cell lines. Protein showed approximately the expected molecular weights: BMI1 (44-47 kDa); hPC2 (82 kDa); RING1 (54 kDa); HPH1 (124 kDa); EZH2 (90 kDa); TBP (36kDa); YY1 (65 kDa).

PcGs and MLL1 distributions in the *HOXA* cluster and pericentromeric repeats by X-ChIP/Southern blots.

The cosmid contig covering the entire *HOXA* cluster was digested with EcoRI (Fig. 2.4A), blotted and hybridized with DNA immunoprecipitated with anti-hPC2 antibodies (Fig. 2.4B). DNA aspecifically immunoprecipitated (mock, Fig. 2.4C) was used as control. With this approach, putative PcGs binding site could be identified quantifying the hybridisation signals in the two Southern blots. The restriction fragments showing a higher signal in the hybridisation with I-DNA, versus the mock, might be considered as the *in vivo* binding site for the investigated proteins. Thus, we have quantified all the hybridisation signals with the "Quantity One Quantification Software", Bio-Rad (Fig. 2.4D and E), identifying fourteen enriched EcoRI fragments in the α-hPC2 Southern blot. Knowing the EcoRI restriction pattern of the *HOXA* locus, it was possible to precisely localize these enriched fragments. This mapping revealed that they are localized mainly in the surrounding regions of the *HOXA* cluster, where only REs have been mapped (compare Fig. 2.5 and Fig. 2.2). The presence of REs and low complexity DNA was also confirmed hybridising the same cosmid contig with labelled CoT1 DNA (Roche, Fig. 2.4D). Indeed, CoT1 DNA is a fraction of the human genome composed essentially by LINEs, SINEs and other repeats, generally used to compete REs in Southern blots.

These results represent the first attempt to map the distribution of PcG/TrxG proteins in large human genomic regions using X-ChIP approach. Other groups have tried to analyse the genomic distribution of these proteins by mean of immunohistochemistry (Saurin A.J., 1998). Their analyses revealed that Polycomb complexes or PcG bodies were found in the pericentromeric regions, in particular of chromosome 1 (Saurin A.J., 1998). To investigate better these associations, we repeated the X-ChIP/Southern blot assay with three different families of human centromeric repeats, looking for the binding of hPC2, BMI1 (two PcGs proteins) and MLL1. Although the latter is a TrxG member, it has been recently shown to interact both with hPC2 and BMI1 (Xia Z.B., 2003). As target regions, we choose the Sau3A, the EcoRI (or alphoid) and the Long Sau3A satellite families known, by FISH analysis, to map to the pericentromeric region of chro-

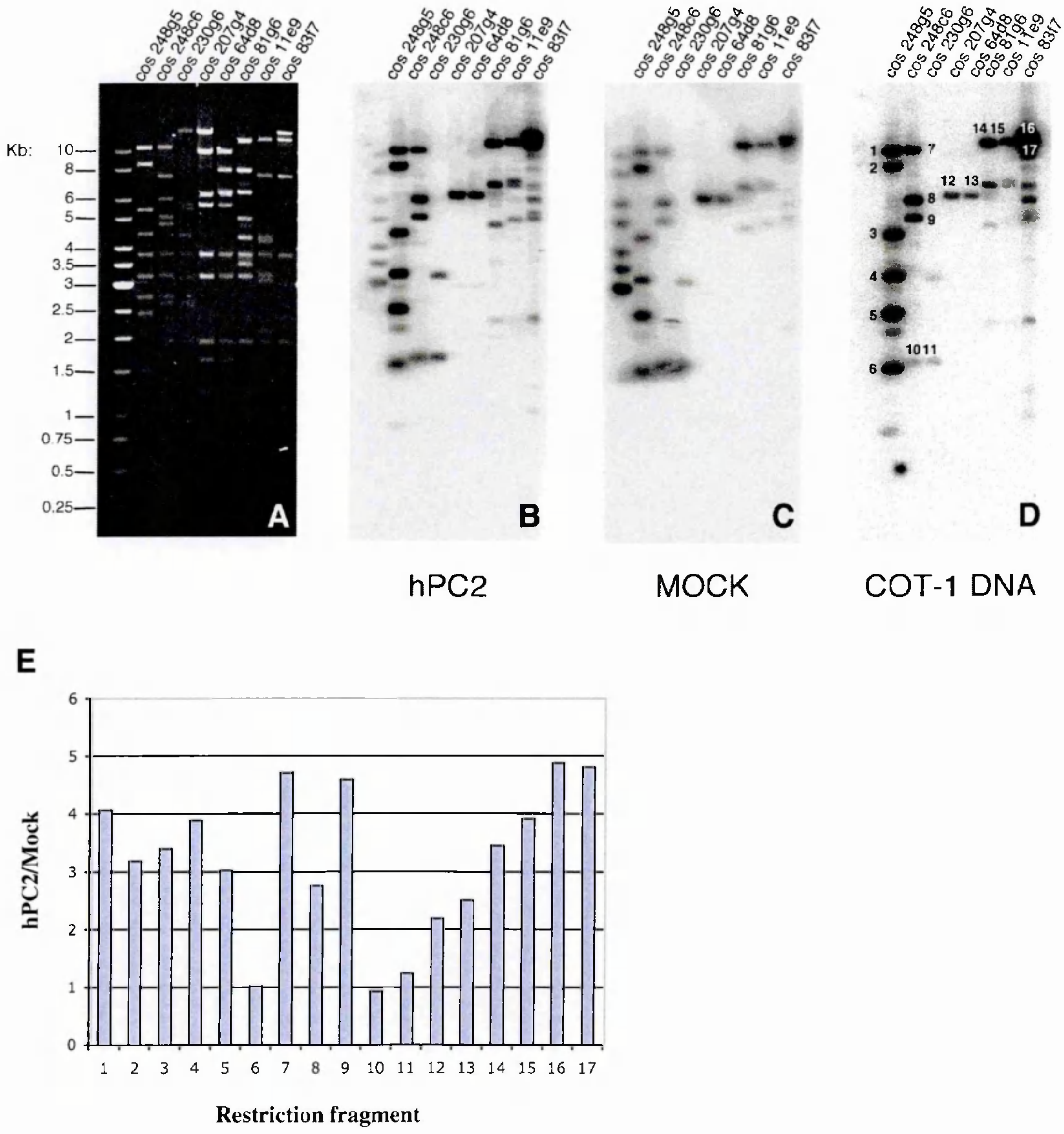


Fig. 2.4: X-ChIP/Southern blots.

Human *HOXA* cosmid clones were digested with EcoRI (**A**). Gels were blotted and hybridised with α-hPC2 I-DNA (**B**), Mock DNA (**C**) or CoT1-DNA (**D**). Restriction fragments representing enriched bands are numbered in **D** (1 to 5; 7 to 9; 12 to 17). Fragments 6, 10, 11 were taken as internal controls. **E**) Quantitative analysis. Hybridization signals from **B** and **C** were quantified and plotted with Quantity One Quantification Software" (Bio-Rad). Intensity values were normalized to an arbitrary 1 kb fragment after the subtraction of the background. The x-axis represents the restriction fragments analyzed; numbers as in **D**. The y-axis represents the ratio hPC2/Mock after normalization and subtraction of the background.

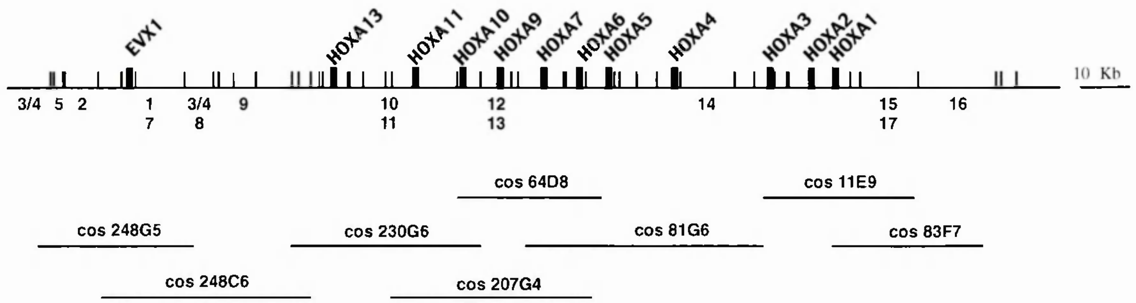


Fig. 2.5: hPC2 distribution in the human *HOXA* locus.

Small black vertical lines represent EcoRI restriction sites. The cosmid contig is drawn below the genomic line. Black rectangles indicate *HOXA* genes and of the flanking *EVX1* gene positions. Numbers below the genomic line identify enriched bands as in Fig. 2.4 and represent putative binding sites for hPC2. Fragments 3 and 4 correspond to the terminal fragments of the cos 248G5 and their relative position was not determined.

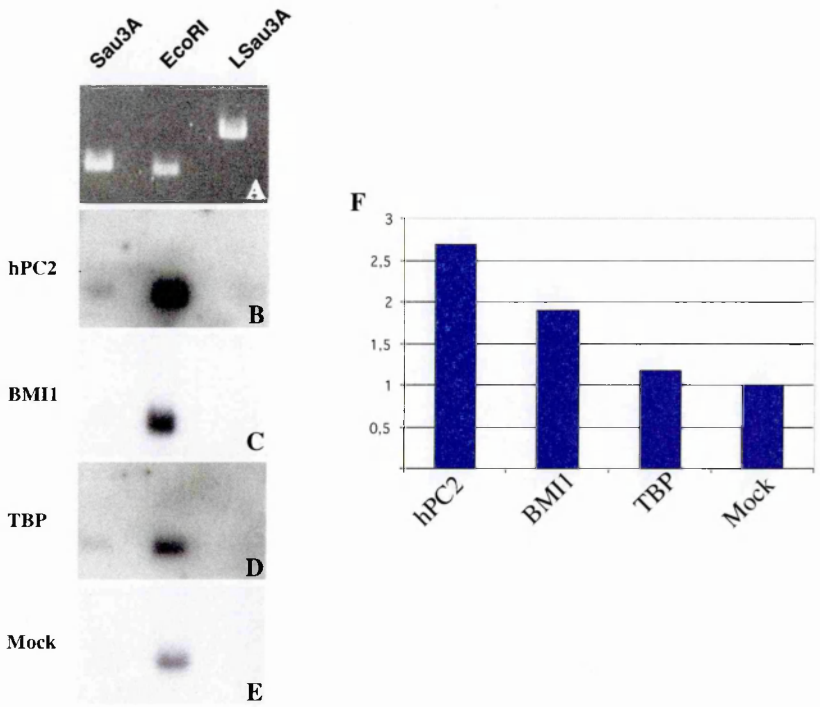


Fig. 2.6: X-ChIP/Southern blots.

DNA fragments corresponding to Sau3A, EcoRI (alphoid) and Long Sau3A (Lsau3A) families of pericentromeric repeats were run on an agarose gel (**A**). Gels were blotted and hybridized with α-hPC2 I-DNA (**B**), α-BMI1 I-DNA (**C**), α-TBP I-DNA (**D**) and Mock DNA (**E**). Quantitative analysis of hybridization signals for the alphoid monomer is shown (**F**). Signals were quantified and plotted with Quantity One Quantification Software" (Bio-Rad) after the subtraction of the background. The x-axis represents the X-ChIP/southern blot analyzed; the y-axis represents the ratio I-DNAs/Mock after subtraction of the background.

mosome 1 (Agresti A 1987; Agresti A., 1989; Meneveri R., 1985). Our results (Fig. 2.6) confirmed the association of the PcG proteins to the pericentromeric region of chromosome 1. Additionally we clearly demonstrate that these proteins bind preferentially the alphoid repeats while no significant interaction was observed with both the Sa3uA and Long Sau3A families.

X-ChIP and "genome complexity".

Repeating the X-ChIP/Southern blot approach with other α-PcGs immunoprecipitated DNAs, we always get the same result: the "inner core" of the *HOXA* cluster seems to be devoid of PcGs proteins that vice-versa seem to be bound to flanking regions containing REs.

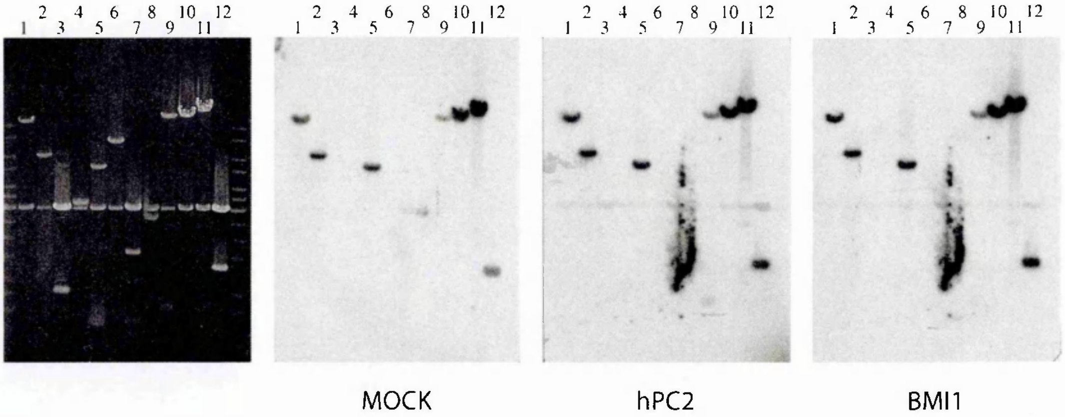
These distributions could represent only partially the *in vivo* PcGs distribution. The analysis of these hybridisation results should be done keeping in mind some aspects of the hybridisation kinetics specific for the human genome. When we started our analysis we followed the X-ChIP/Southern blot approach applied for *Drosophila* cells and embryos (Orlando V., 1997). Nevertheless *D.melanogaster* is an organism having a lower "genome complexity" than the human one (the complexity of a genome can be defined as the total length of all different sequences present in it) and I-DNA from *D.melanogaster* and *H.sapiens* genome could have hybridisation kinetics completely different. While "single sequence" DNAs, in the *D.melanogaster* I-DNA, could hybridise their complementary sequence on the filter, this could not be true for human I-DNA. The main feature of the human genome responsible of this different behaviour is the fact that 50% of our genome is made of highly repeated DNA. This DNA will correspond to the major fraction in a human I-DNA, while "unique" DNAs will represent the minor portion. This means that a specific "single sequence" DNA present in the human I-DNA could be insufficiently labelled and therefore will give an undetectable hybridisation signal on the filter. Thus, in our X-ChIP/Southern blot experiments we could have detected only the PcG binding sites located in regions containing REs. We could have lost the binding sites in the inner core of the *HOXA* cluster, composed exclusively of unique sequences. To test this hypothesis, we carried out three different kinds of hybridisations. First we used total genomic DNA as probe on the cosmid contig obtaining hybridisation signals perfectly identical to the one shown in Fig. 2.4C-D. These experiments confirmed that fragments corresponding to "single sequence" DNAs could not be detected even using human genomic DNA. We then cloned the entire human *HOXA* clusters into plasmid vectors (summarised in Table 2.3), separating the "unique sequence" clones (corresponding to the *HOXA1-HOXA13* region) from clones containing repetitive DNA

(*HOXA* surrounding regions). We prepared two different sets of filters using these clones and repeated the X-ChIP/Southern hybridisations with I-DNAs against different PcG proteins. As shown in Fig. 2.7B-C, I-DNA hybridises mostly with clones containing REs but only with three "single sequence" clones. Sequence analysis of these three clones

A) *HOXA* plasmid clones:

1) *248.1*
2) *248.2*
3) *248.116*
4) *248.62*
5) *248.35*
6) *248.32*
7) *1589*
8) *230.93*
9) *11.4*
10) *83.57*
11) *83.24*
12) *1254.1*
13) 248.16
14) 248.101
15) 230.91
16) 230.89
17) 230.2
18) 230.81
19) 2631
20) 230.12
21) 230.9
22) 230.I14
23) 64.22
24) 6180
25) 1635.11
26) 986.41
27) 64.34
28) 3572.31
29) 5000.11
30) 11.31
31) 4353.21
32) 83.23

B) X-ChIP/Southern blots: "Repetitive DNA" clones.



C) X-ChIP/Southern blots: "Single Sequence DNA" clones.

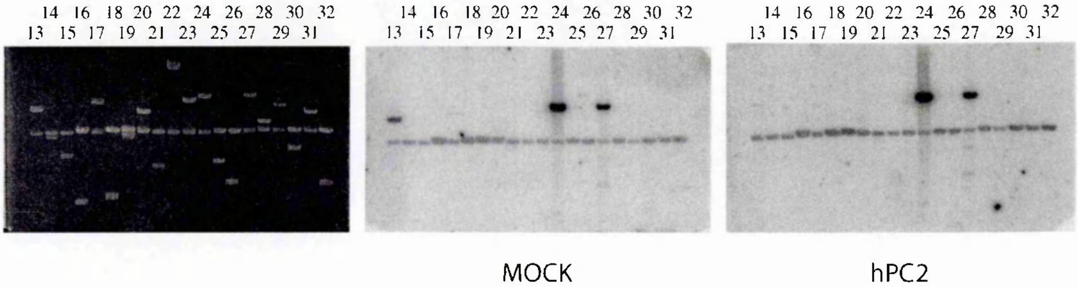


Fig 2.7: X-ChIP/Southern blots.

Plasmid clones corresponding to the entire HOXA cluster are schematically shown in (**A**) and summarized in Table 3; clones containing REs are in italics. (**B**) Plasmid clones containing REs were digested with EcoRI, run in agarose gels and blotted. Filters were hybridized with Mock DNA, α-hPC2 I-DNA and α-BMI1 I-DNA. (**C**) Plasmid clones containing single sequence DNAs were digested with EcoRI, run in agarose gels and blotted. Filters were hybridized with Mock DNA, and α-hPC2 I-DNA. See text for details.

revealed that they contain stretches of low complexity DNA, probably responsible for the hybridisation signals. Moreover, we obtained the same hybridisation patterns with human genomic DNA.

These results indicate that the X-ChIP/ Southern approach could be useful to investigate the binding sites of proteins that associate with repetitive DNAs but could not be sensitive enough for the analysis of "single sequence" regions. The distribution of protein binding sites in such regions should be performed analysing the I-DNA with more sensitive techniques, such as PCR or microarray assay. To investigate this hypothesis, we focused our attention on the *HOXA5-HOXA6* intergenic region. The EcoRI fragments spanning this portion of the cluster were never detected in the X-ChIP/Southern analysis. This region, arbitrary chosen, was subdivided into fourteen amplifiable PCR fragments (Table 2.4). I-DNA with antibodies against different PcG and TrxG proteins was utilized as template for PCR reactions (Fig. 2.8). Different primer pairs specifically

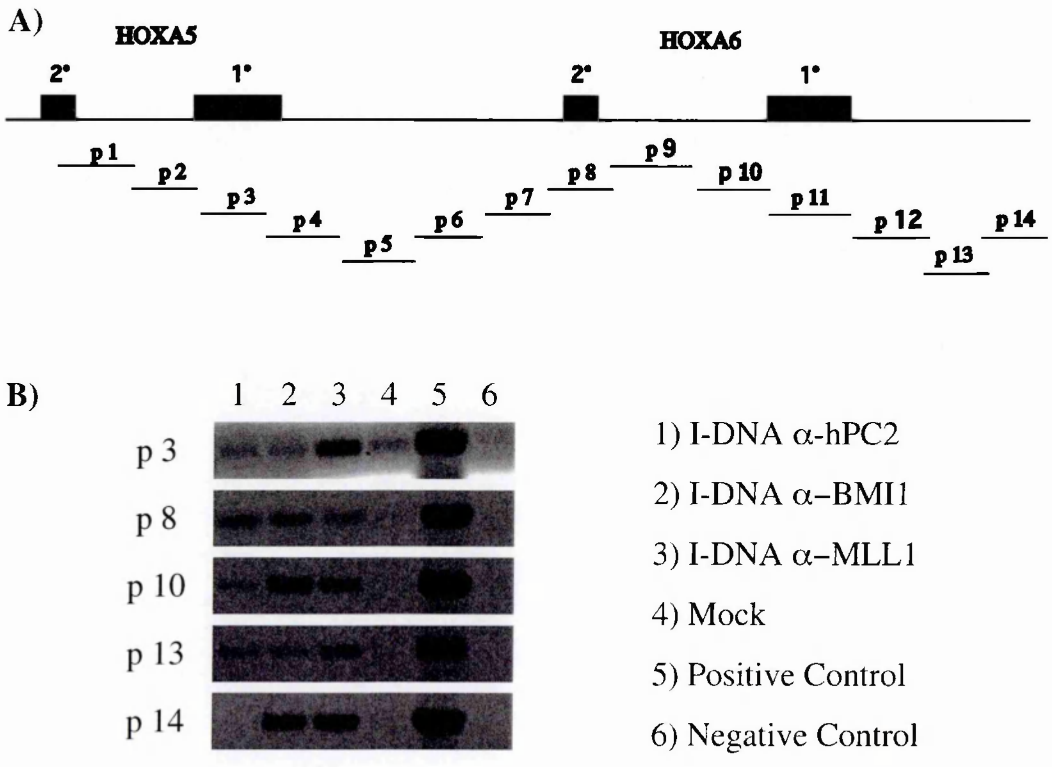


Fig. 2.8: X-ChIP/PCRs.

The *HOXA5-HOXA6* locus is schematically represented in (**A**). Black rectangles and numbers indicate coding exons and exon order, respectively. The position of amplified fragments (p1-p14) are indicated below the genomic line. (**B**) X-ChIP/PCR analyses. I-DNAs with antibodies against hPC2, BMI1, MLL1 and a Mock control, were amplified following PCR conditions given in Table 2.4. Human genomic DNA and water were amplified as positive and negative controls, respectively.

amplified only I-DNAs, while no amplifications were detectable in the mock. The distribution of these binding sites revealed that PcGs and MLL1 associate both with regions containing *HOXA5-HOXA6* exons (primers p3 and p8) and with promoters (primers p13 and p14). Such results clearly show that the X-ChIP/PCR is more sensitive than the X-ChIP/Southern blot analysis and allows a more detailed investigation of human I-DNAs. Nevertheless, the survey of large genomic regions (such as the *HOX* cluster loci) would require hundreds of primer pairs and could be hardly investigated by this approach. To this aim, the X-ChIP should be associated with another sensitive method that allows rapid analyses of big genomic loci. Recent publications (Weinmann A.S., 2002) have shown that the microarray technology could be coupled to the X-ChIP to investigate the genomic distribution of target proteins, an approach called "ChIP to CHIP". We then decided to study the association of PcG and TrxG proteins to the human *HOXA* cluster by means of the "ChIP to CHIP" strategy. We created a microarray containing the entire human *HOXA* cluster, one of the first examples of "locus specific genomic microarray". Taking advantage both of the plasmid clones and the full sequence of the *HOXA* cluster, the entire human *HOXA* locus was subcloned into 137 PCR fragments (0.5-2.3 kb, Table 2.5) and the corresponding DNAs were quantified, checked by restriction analysis and at the end spotted on a glass slide. The resulting "*HOXA* genomic microarray" was then hybridized with human genomic DNA. Using this approach, the sensitivity of the hybridization on the "HOX microarray" is clearly increased: both the "unique sequences" clones and the ones containing REs can be efficiently detected even by a genomic DNA probe (Fig. 2.9). In the new future, we will hybridize the "HOX microarray" with I-DNAs against different PcG and TrxG proteins, following the scheme shown in Fig. 2.9.

PcGs and MLL1 distributions in the *HOXA* locus by X-ChIP/PCRs.

To confirm the PcG proteins distribution in the surrounding regions of the HOXA cluster, we decided to utilize the X-ChIP/PCR approach. To this end, the fourteen EcoRI fragments containing putative PcG proteins binding sites were divided into amplifiable segments (Table 2.6 summarised both primer sequences and PCR conditions). The chro-

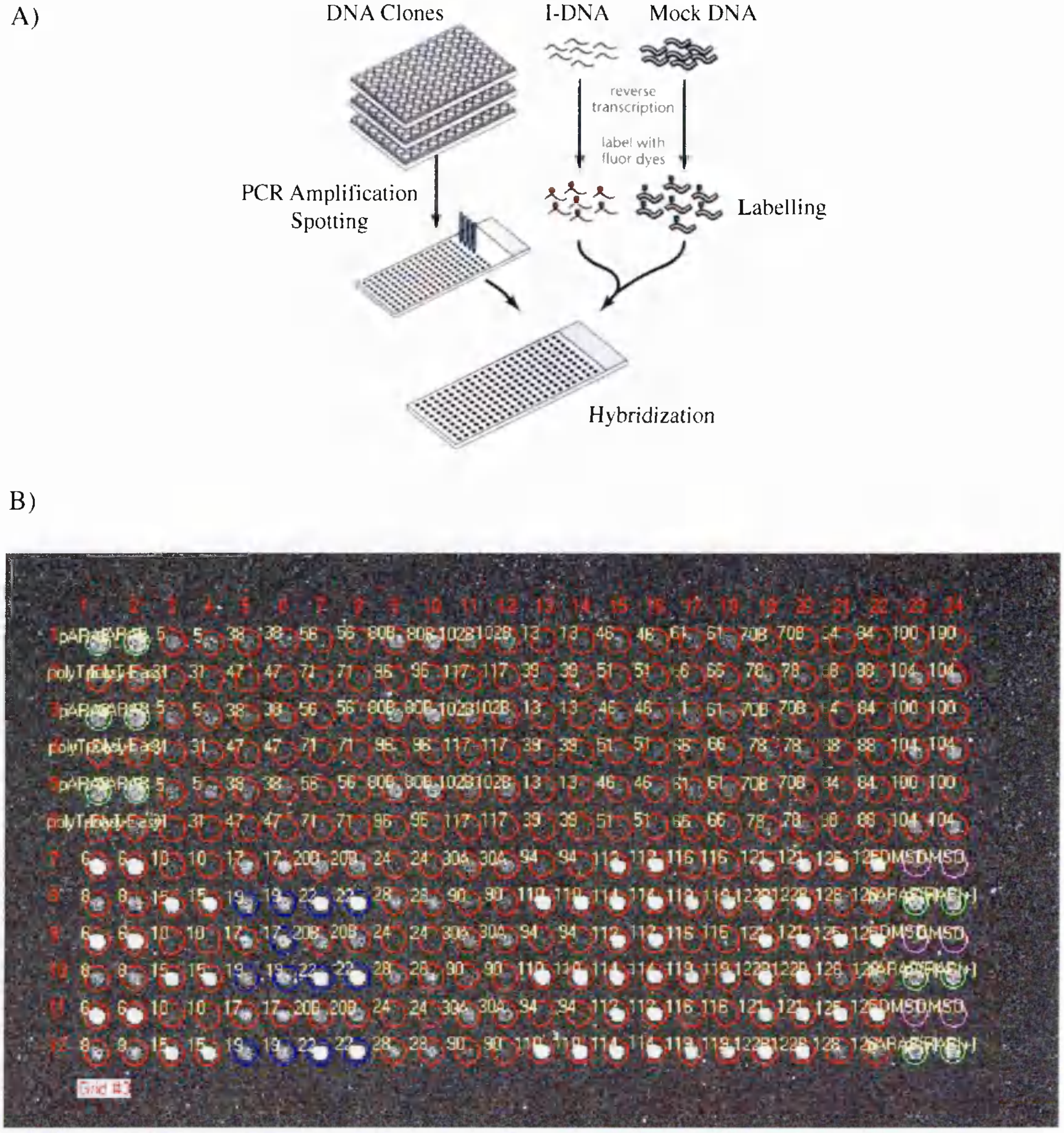


Fig. 2.9: ChIP to CHIP.

(**A**) The Chip to CHIP approach is schematically presented.
(**B**) Total genomic DNA was used as a probe on a microarray containing a portion of the *HOXA* locus. The first 12 columns spots correspond to unique sequence DNA fragments, while the columns 13-24 correspond to DNA fragments containing REs. Many of the unique sequence DNA were positively hybridized.

matin immunoprecipitation was repeated and the I-DNA used as template for PCR reactions with these sets of primers (Fig 2.10). These X-ChIP/PCR experiments confirmed the association between PcG proteins to regions containing REs both upstream and downstream the human *HOXA* cluster (Fig. 2.10). We included in this analysis also the protein MLL1, observing a similar distribution between MLL1 and the PcG proteins, in particular with BMI1 (Fig. 2.10).

The *in silico* analysis of the human sequence, reveals that many different LINE and SINE families localize in the flanking regions of the *HOXA* cluster. A more detailed

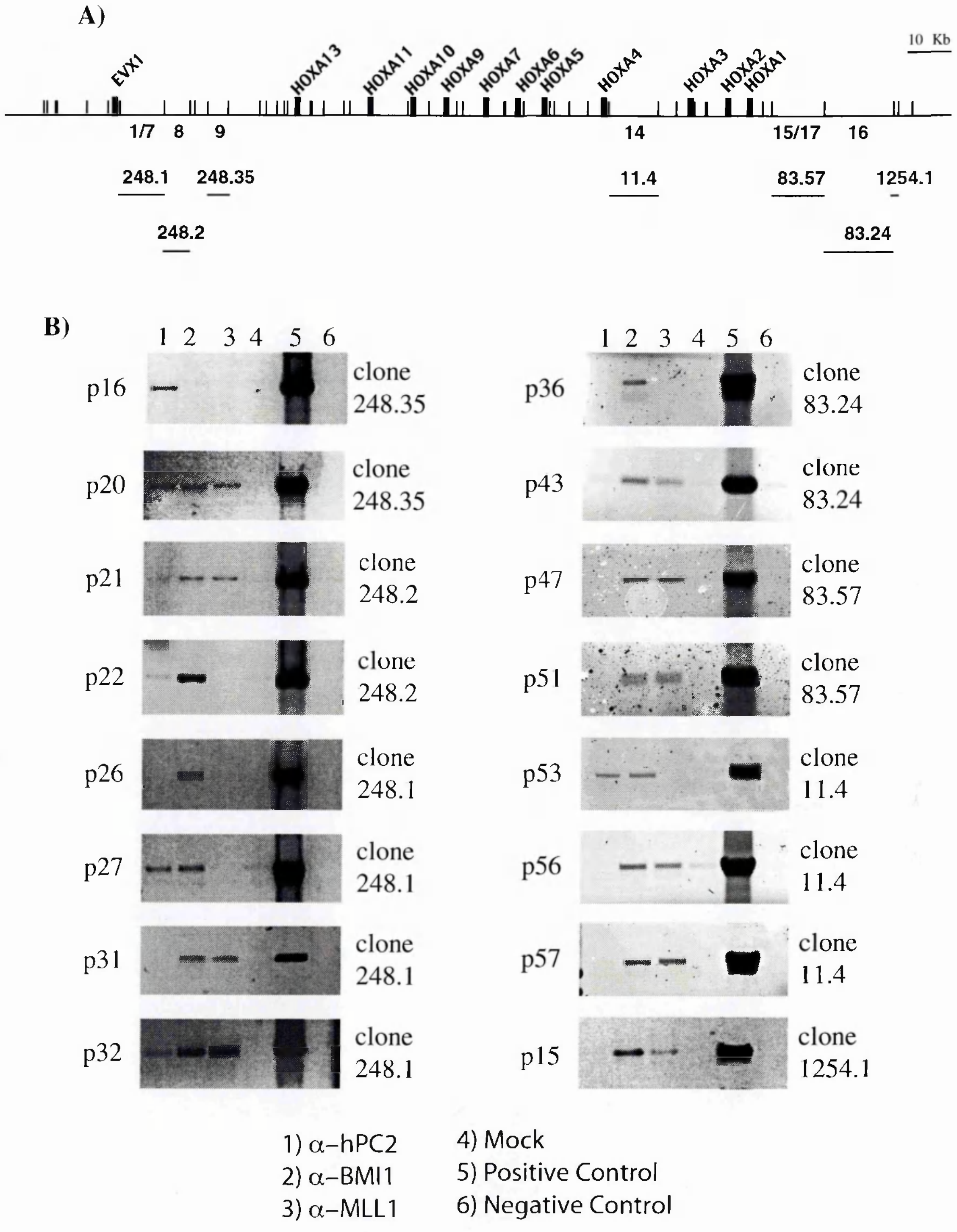


Fig. 2.10: X-Chip/PCRs.

A) Schematic organization of the *HOXA* locus. Small black vertical lines represent EcoRI restriction sites; black rectangles indicate *HOXA* and *EVX1* genes. Numbers below the genomic line identify putative PcG binding sites as in Fig. 2.4. The corresponding plasmid clones are drawn below the genomic line.

B) X-ChIP/PCRs analyses. EcoRI restriction fragments corresponding to putative PcG binding sites were subdivided into 60 amplifiable fragments (see Table 6). I-DNA with antibodies against hPC2, BMI1, MLL1 and a Mock control were amplified with this set of primers. Human genomic DNA and water were used as positive and negative controls, respectively. See text for discussion of informative results.

mapping of the REs present in the X-ChIP/PCR amplified fragments indicates that these segments contain different LINE and SINE elements, as depicted in Table 2.7. Furthermore, some of these binding sites do not contain any REs. Thus we could exclude that PcG and TrxG proteins associate preferentially with specific REs subfamilies.

HOX clusters show a peculiar and unique distribution of REs among the human genome.

The striking co-localization of the PcGs and trxGs binding sites with the in the flanking regions of the *HOXA* cluster prompted us to define the genomic organization of the other three human HOX clusters. In particular, we want to address the question about the distribution of REs in the *HOXB-D* cluster.

The genomic organization of the human *HOXB-D* clusters was deduced applying the same *in silico* approach described for the human *HOXA* cluster. Non-Redundant (nr) and the Unfinished High Throughput Genomic Sequences (htgs) BLAST databases were screened with cDNA sequences from human *HOXB-D* genes (cDNA accession numbers corresponding to the *HOXB*, *HOXC* and *HOXD* genes are listed in Table 2.8, Table 2.9 and Table 2.10, respectively). In this way we identified several PAC/BAC clones, whose sequences were assembled into contigs covering the entire human *HOXB-C-D* loci. As for the human *HOXA* cluster, Repeatmasker Web Server (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) was used to identify and localize all the REs in these genomic sequences. The genomic organization of these three loci is shown in Fig. 2.11, Fig. 2.12. Fig. 2.13. This analysis confirmed also in the other human *HOX* clusters the peculiar localization of REs. The intergenic regions are devoid of repetitive elements (with some exception discussed below), while they are grouped in broad domains upstream and downstream the coding part of the clusters. The internal part of each cluster is particularly rich of CpG islands, simple repeats and low complexity DNAs.

According to the data provided by the human genome project, around 50% of our genome contains repetitive sequences, while the coding sequences comprise less than 5%. Thus, it is quite surprising that the *HOX* loci are virtually devoid of REs. The den-

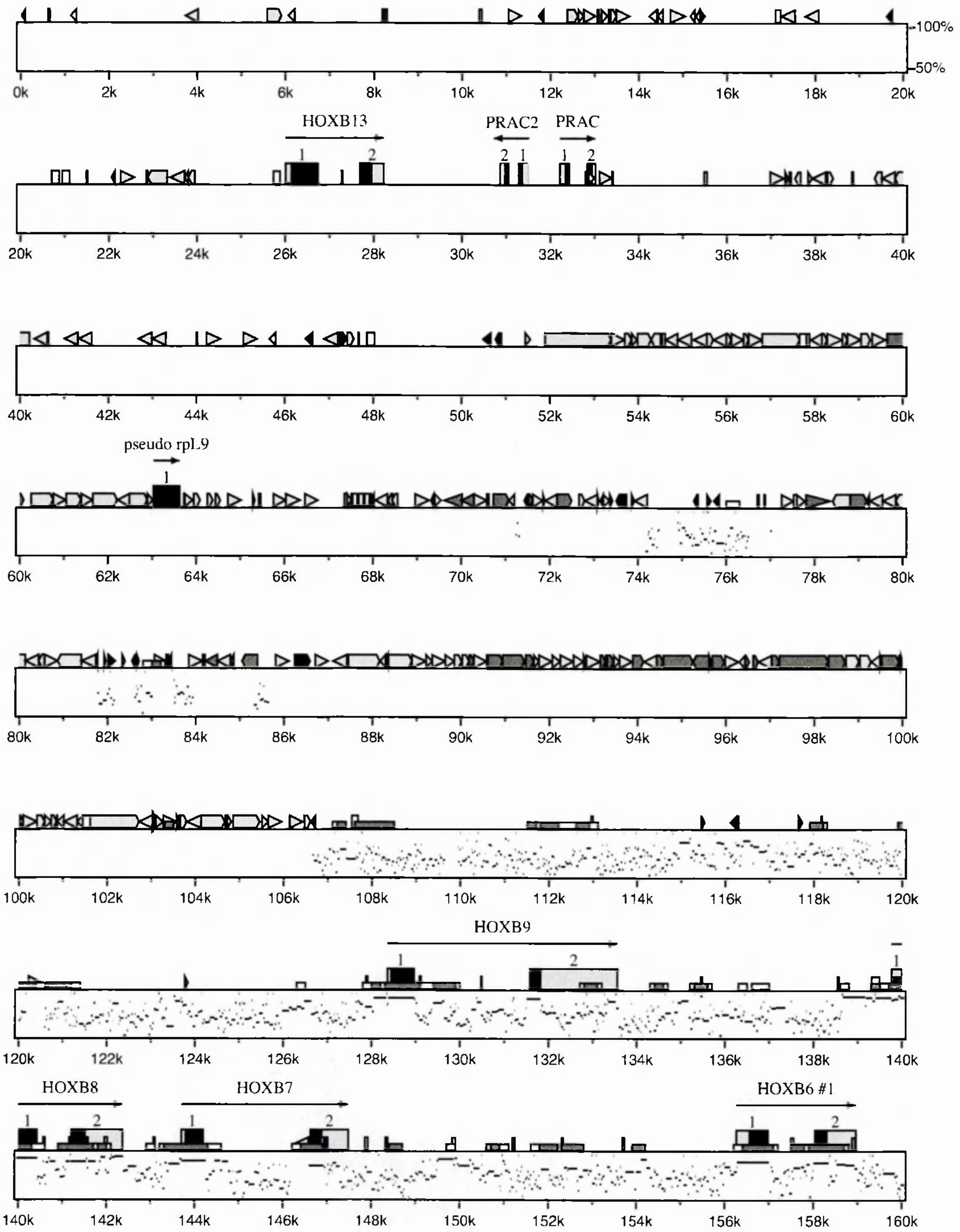


Fig. 2.11: genomic organisation of the human *HOXB* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the next page. Black arrows indicate the direction of the transcription of the corresponding genes. Genes presenting different alternative splice forms (see Table 8) are indicated with "#". Only one (#1) alternative spliced form is drawn. Comparisons of nucleotide sequence between human *HOXB* and mouse *HoxB* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the human *HOXB* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

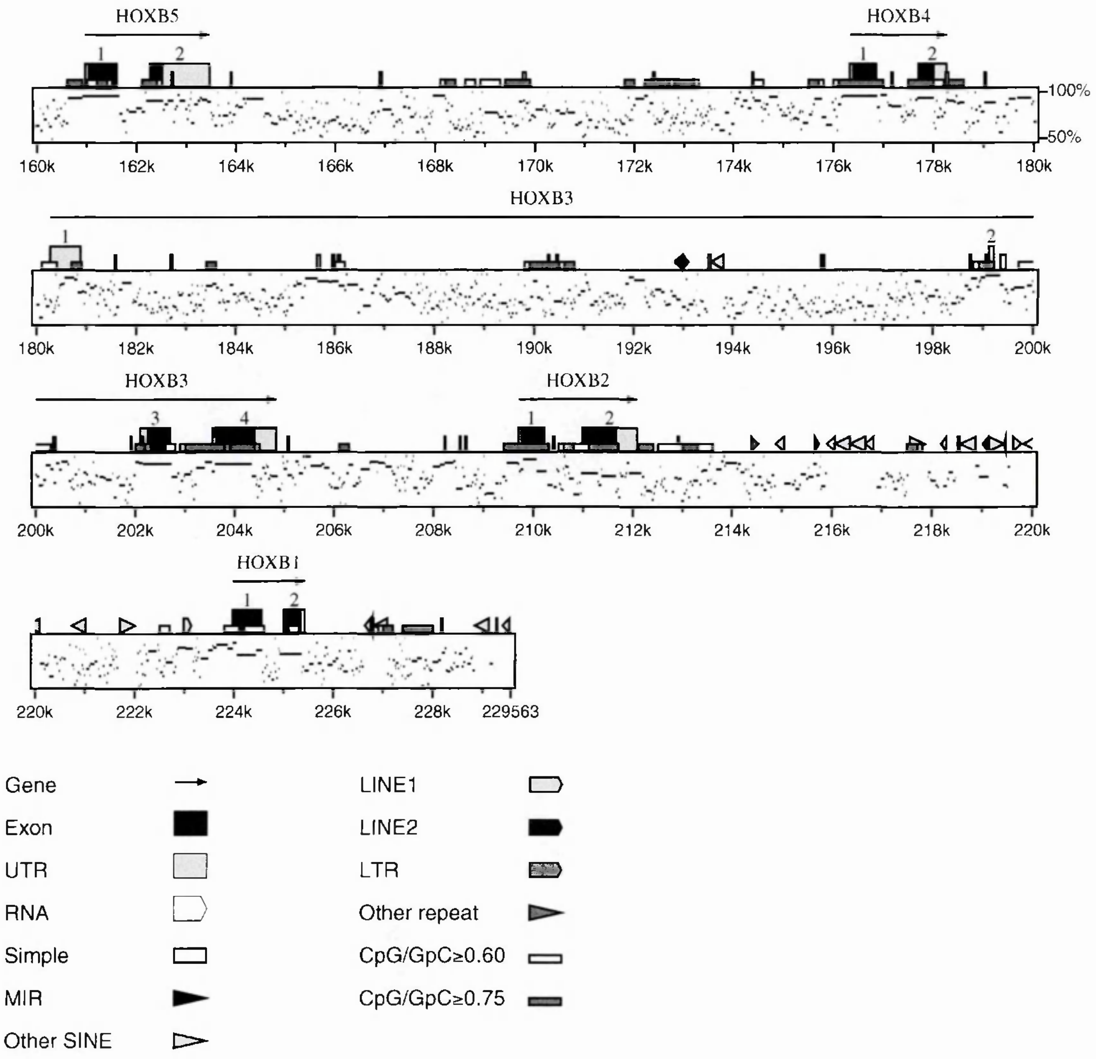


Fig. 2.11: genomic organisation of the human *HOXB* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Genes presenting different alternative splice forms (see Table 8) are indicated with "#". Only one (#1) alternative spliced form is drawn. Comparisons of nucleotide sequence between human *HOXB* and mouse *HoxB* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the human *HOXB* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

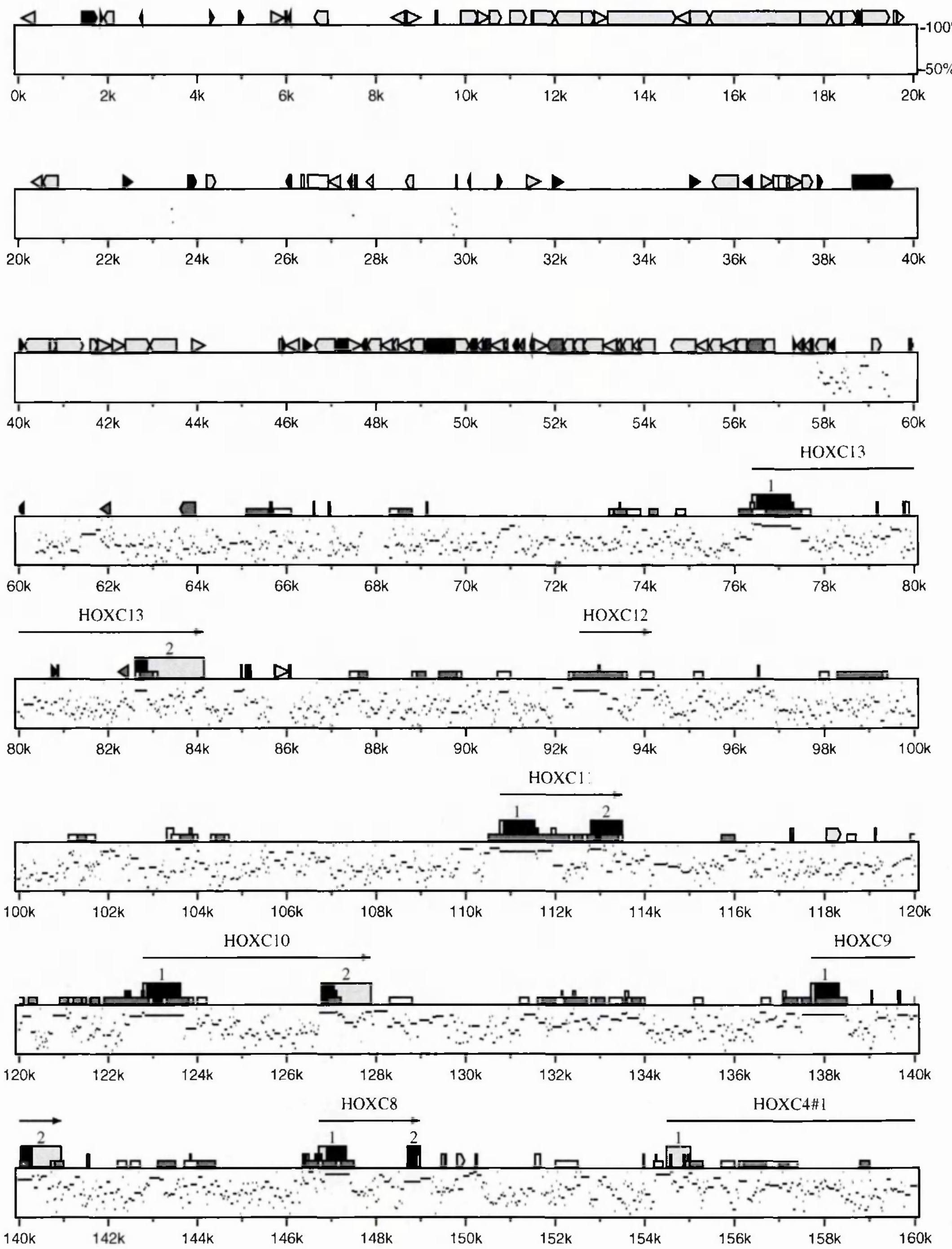


Fig. 2.12: genomic organisation of the human *HOXC* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the next page. Black arrows indicate the direction of the transcription of the corresponding genes. Genes presenting different alternative splice forms (see Table 9) are indicated with "#". Only one (#1) alternative spliced form is drawn. Comparisons of nucleotide sequence between human *HOXC* and mouse *HoxC* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the human *HOXC* cluster as a reference; mouse *Hoxc* sequence ends between *HOXC5* and *HOXC4*. Kilobase (k) markings refer to our genomic contig.

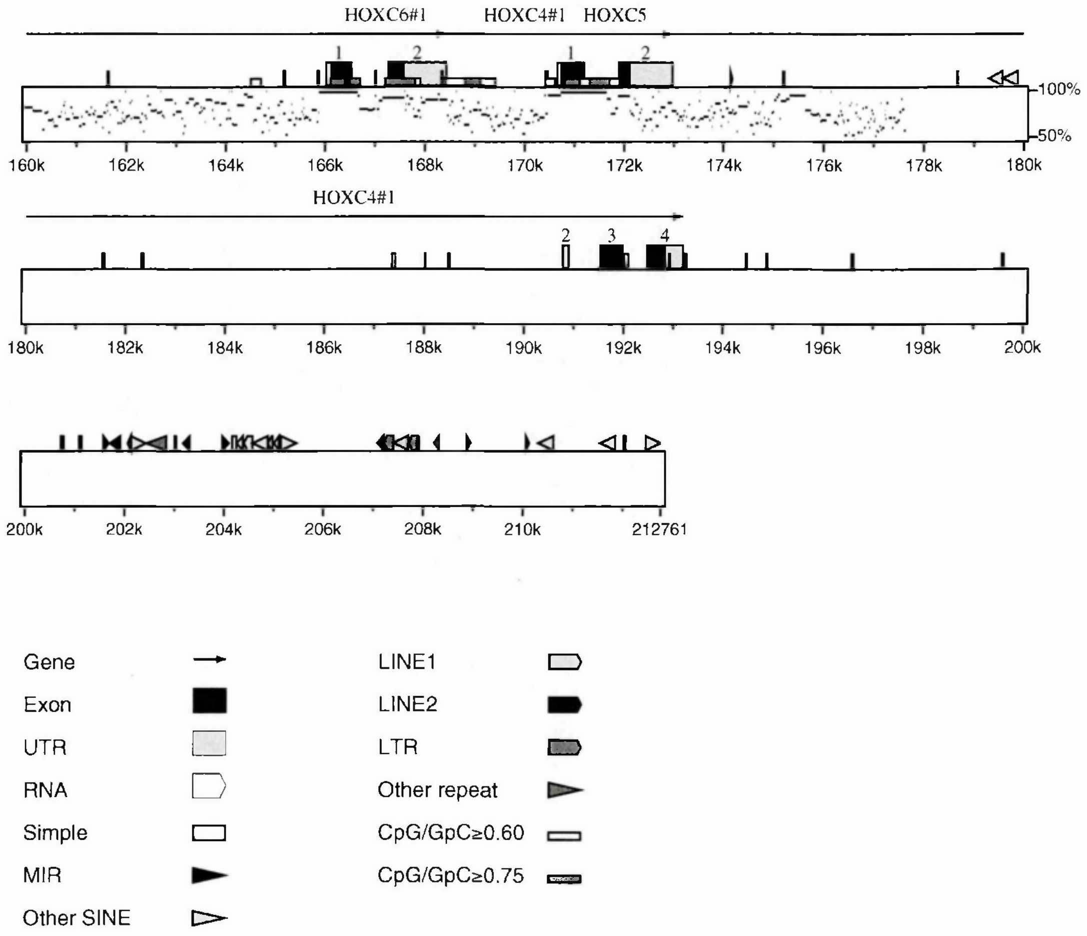


Fig. 2.12: genomic organisation of the human *HOXC* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Genes presenting different alternative splice forms (see Table 9) are indicated with "#". Only one (#1) alternative spliced form is drawn. Comparisons of nucleotide sequence between human *HOXC* and mouse *HoxC* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the human *HOXC* cluster as a reference; mouse *Hoxc* sequence ends between *HOXC5* and *HOXC4*. Kilobase (k) markings refer to our genomic contig.

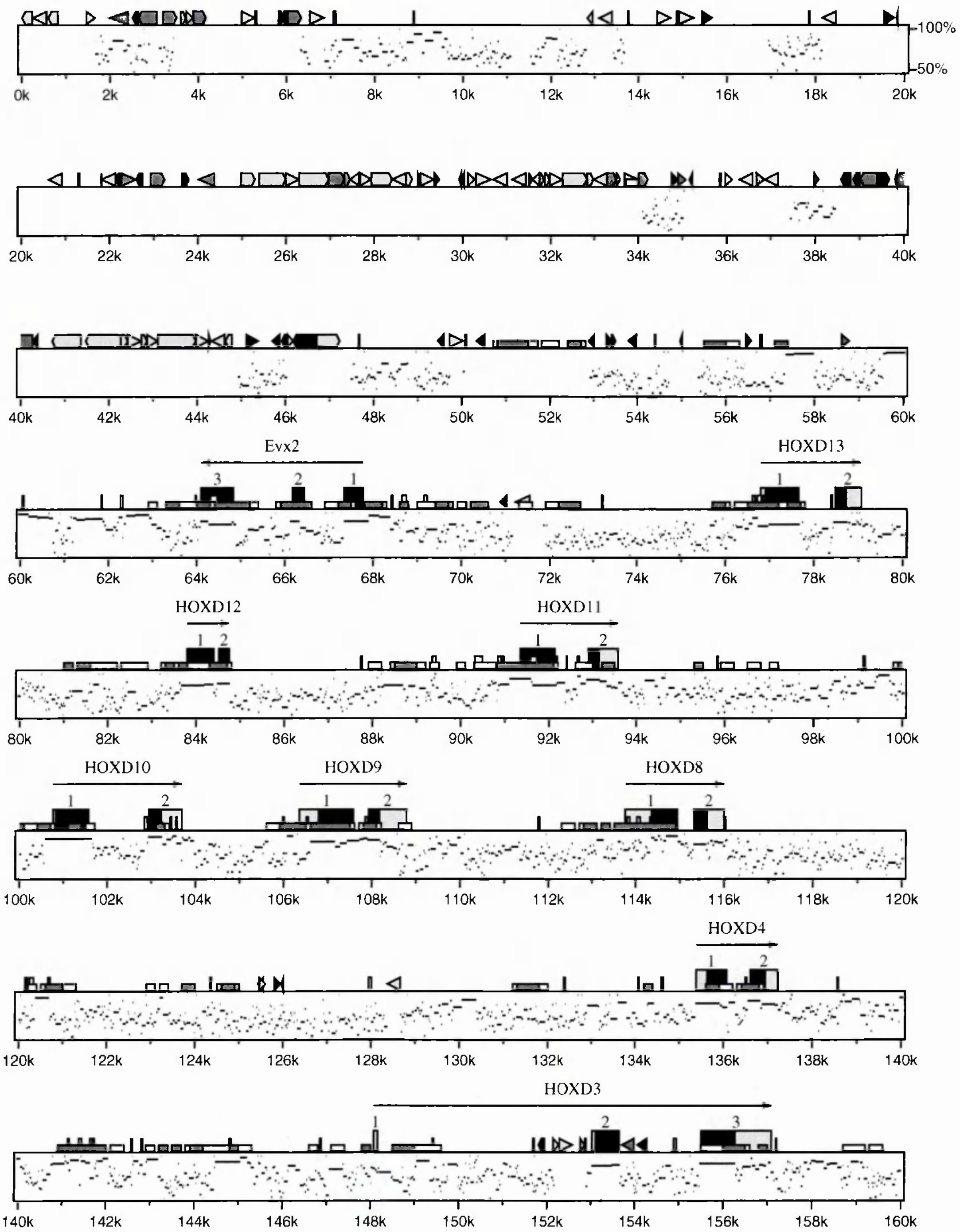


Fig. 2.13: genomic organisation of the human *HOXD* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Comparisons of nucleotide sequence between human *HOXD* and mouse *HoxD* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the human *HOXD* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

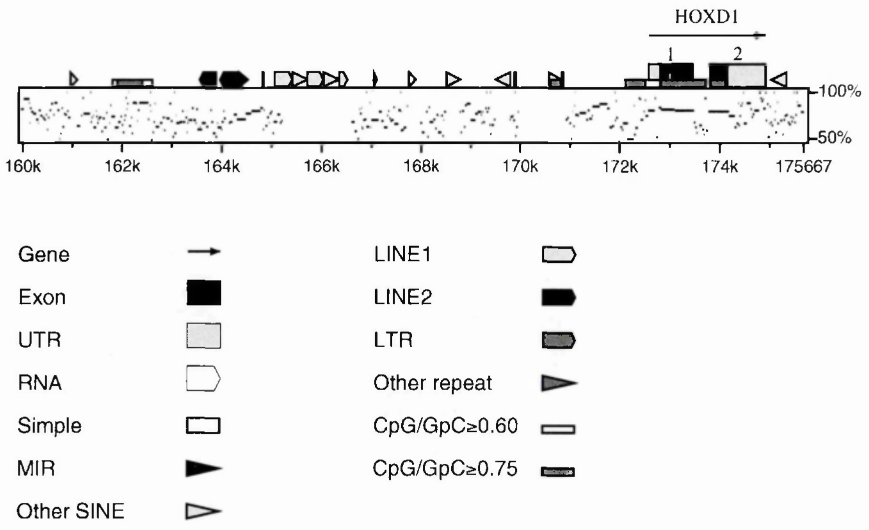


Fig. 2.13: genomic organisation of the human *HOXD* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Comparisons of nucleotide sequence between human *HOXD* and mouse *HoxD* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the human *HOXD* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

sity of different REs families varies a lot among the human genome and it has been possible to calculate the percentage of REs present in each human isochore (Smit A.F., 1999). From the output of RepeatMasker we deduce that the percentage of REs in the intergenic portions of the human clusters is less than 1-2%. This value is dramatically lower than expected 41-42%, based on the GC content of these regions. Conversely, the percentage of REs in the flanking regions approaches the expected values. These data underline that the absence of REs in the intergenic portions of *HOX* cluster should have some functional meanings and could not be due just to the base composition of the clusters themselves. As controls we have analyzed the genomic organization of other gene clusters (such as the beta-globin cluster, the major histocompatibility complex (MHC) class I and class II clusters, histone gene cluster) at the UCSC Genome Browser (www.genome.ucsc.edu). These clusters show a rather uniform intergenic distribution of REs with no large blocks of REs in the flanking portions (Schwartz S., 1991; Shiina T.1999). Thus, the distribution of REs in the four *HOX* loci seems to be peculiar and specific, as other gene clusters do not share it. In collaboration with Diego Di Bernardo, we investigated how this "RE-distribution" represents a unique situation in the human genome. We developed a bio-informatic algorithm able to scan the entire human genome looking for any region containing less than 5 REs and at least 3 transcribed cDNA. Surprising this programme identifies only other 6 genomic regions satisfying starting parameters (Table 2.11). Nevertheless, none of these regions is as long as the *HOX* clusters and does not contain as many genes. These data clearly demonstrate that the peculiar REs distribution of the *HOX* clusters represents also a unique characteristic of these loci among the entire human genome.

Alu insertion inside human *HOX* clusters was negatively selected.

We hypothesized that the absence of REs was due to negative selective pressure, to investigate this hypothesis, we analyzed the *Alu* sequences distribution in the human *HOX* clusters.

Different *Alu* subfamilies can be identified on the basis of their putative age, *AluY* being the most recent dispersed subfamily, *AluS* the intermediate and *AluJ* the oldest (Mighell A.J, 1997). Analysis of random sequences from Genbank showed a global (*AluS* + *AluY*)/*AluJ* ratio =3 (Jurka J., 1988). This value represents a random integration of the *Alu* subfamilies in the absence of any selection and it has been used to analyse the accessibility of genomic regions for *Alu* insertions (Beck S., 1999). Only one *AluS* element was identified in the intergenic regions of both *HOXA* (between *HOXA3* and *HOXA4*, Fig. 2.2) and *HOXC* (between *HOXC12* and *HOXC13*) clusters (Fig. 2.12). The complete absence of "old" *AluJ* elements and the presence of a "young" *Alu* mean that the insertion of REs may have not been tolerated during *HOX* cluster evolution. Vice versa both the *HOXB* and *HOXD* clusters contain *Alu* elements belonging to different subfamilies. As shown in Fig. 2.11 and Fig. 2.13, these *Alu* elements are mainly grouped into specific portions of the clusters, rather than being uniformly distributed. In particular they are localized in the regions between *HOXB13-HOXB9*, *HOXB2-HOXB1* and *HOXD3-HOXD1*, where the *Alu* ratio is 1.48, 1.28 and 3, respectively. Except for the latest region, these ratios indicate these regions initially accepted *Alu* elements but became later sensitive to further insertions. In the rest of the *HOXB* and *HOXD* clusters, only two *Alu* elements (*AluS*, *AluY*) were identified in each of the remaining *HOXB9-HOXB2* and *HOXD13-HOXD3* portions. This scenario is quite similar to the one observed for the

HOXA and *HOXC* clusters, suggesting again that the *Alu* insertion was negatively selected. As controls we have analyzed the human *beta-globin* and *MHC classII* clusters. *Beta-globin* shows a very high ratio (*AluS* + *AluY*)/*AluJ* = 7) and seems to have tolerated *Alu* insertion during evolution. On the other hand, *MHCII* has a 1.42 ratio, perhaps initially tolerating *Alu* sequences and later becoming refractory to further insertions (Beck S., 1999).

We have further completed the *Alu* insertion analysis of *HOXA*, *HOXC* and *HOXD* clusters, spanning 60-70 kb of upstream and/or downstream regions. Regions with REs surrounding upstream and downstream the *HOXA* cluster show a ratio of 2.70 and 2.95 respectively, which is comparable to the expected value. A similar situation was observed for the upstream region of the *HOXD* cluster, where the ratio value was 2.2. Conversely, the ratio obtained for the *HOXC* flanking region was much lower (=1.5). This could indicate that, at least for *HOXC*, some negative selection prevented further insertions of *Alu* subfamilies in the upstream flanking region.

Exceptions to the rule and gene regulation.

As already underlined in the previous paragraph, we noticed some exceptions featuring the presence of REs in intergenic regions. In the human *HOXB* cluster (Fig. 2.11), *HOXB13* is separated from the previous paralog gene by more than 100 kb made uniquely of repetitive DNA and a pseudogene of ribosomal protein L9 (Fig. 2.11). Other REs were identified between the *HOXB1* and *HOXB2* genes. The same situation was detected in the *HOXD* clusters, in the region separating *HOXD1* and *HOXD3* (Fig. 2.13). Moreover, in the *HOXA* cluster a region containing REs separates *EVX1* from *HOXA13* (Fig. 2.2) whereas its paralog *EVX2* in the *HOXD* cluster remains linked to the *HOXD13* gene (Fig. 2.13). Notably, the four mouse *Hox* loci present the same genomic organization.

Strikingly, all these exceptions correspond to cases in which the gene placed next to the REs presents specific expression profiles, unshared with neighbouring *Hox* genes (see discussion).

Comparison between human and mouse clusters.

The precise and co-ordinate expression of *Hox* genes requires an intricate net of cis-regulatory elements. A way to identify regulatory elements, conserved during evolution, is to compare genomic DNA from orthologous loci (Hardison R.C., 1997). For this reason we deduced from the public databases the sequences corresponding to the four mouse *Hox* clusters. All mouse *Hox* clusters show the same genomic organization as the human ones, confirming the striking distribution of the REs. Fig. 2.14 - Fig. 2.17 represent the genomic maps of the mouse *HoxA-D* clusters. The accessions numbers of cDNA sequences used for *HoxA-D* genes mapping are listed from Table 2.12 to Table 2.15, respectively.

Using PipMaker, the sequences from human and mouse were compared. As control, we performed the same analysis using human and mouse *beta-globin* loci (accession numbers: human U01317.1, mouse X14061). As summarized in Fig. 2.18, this approach revealed a strikingly high sequence identity in the *Hox* clusters maintained between the orthologous loci both in coding and intergenic regions (Fig. 2.18A-D). Cluster segments with the highest identities (more than 80-90%) correspond both to exons and to several other intergenic fragments, particularly CpG islands. Vice-versa when the sequences of two paralogous *HOX* loci were compared (both in human and mouse), no sequence similarities were identified except for the genomic portions coding for homeodomains (Fig. 2.19A). In *beta-globin* loci high identity values are restricted to coding regions (Fig. 2.19B; Bulger M., 1999). This high sequence similarity between human and mouse Hox loci didn't help us in identifying putative *cis* regulatory elements. On the other hand, this approach revealed that all human and mouse *Hox* clusters present the same genomic organization with the REs confined in regions surrounding *Hox* clusters. This seems to be a new major common characteristic shared by all *Hox* clusters, in addition to the spatial and temporal colinearity.

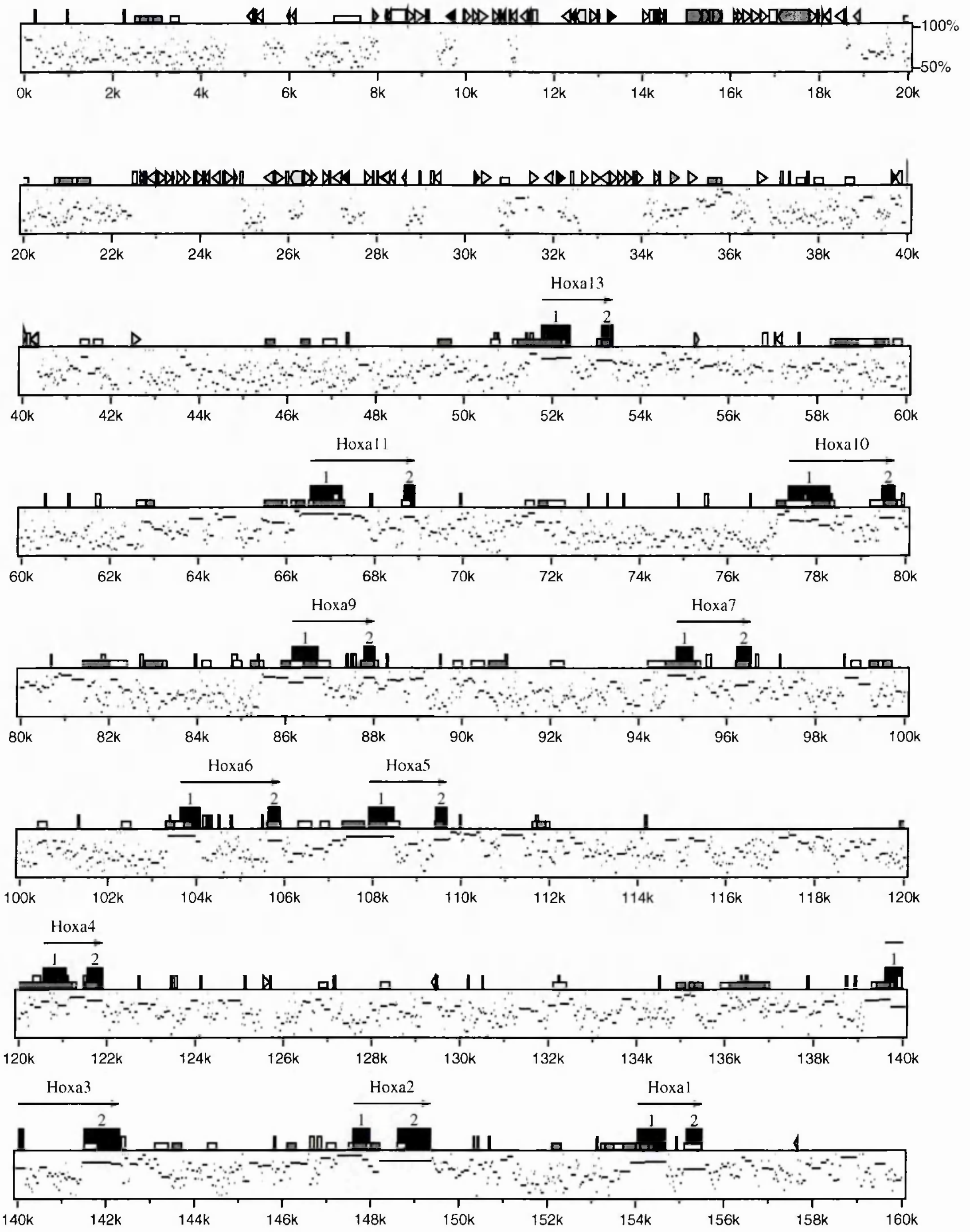


Fig. 2.14: genomic organisation of the mouse *HoxA* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the next page. Black arrows indicate the direction of the transcription of the corresponding genes. Comparisons of nucleotide sequence between mouse *HoxA* and human *HOXA* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the mouse *HoxA* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

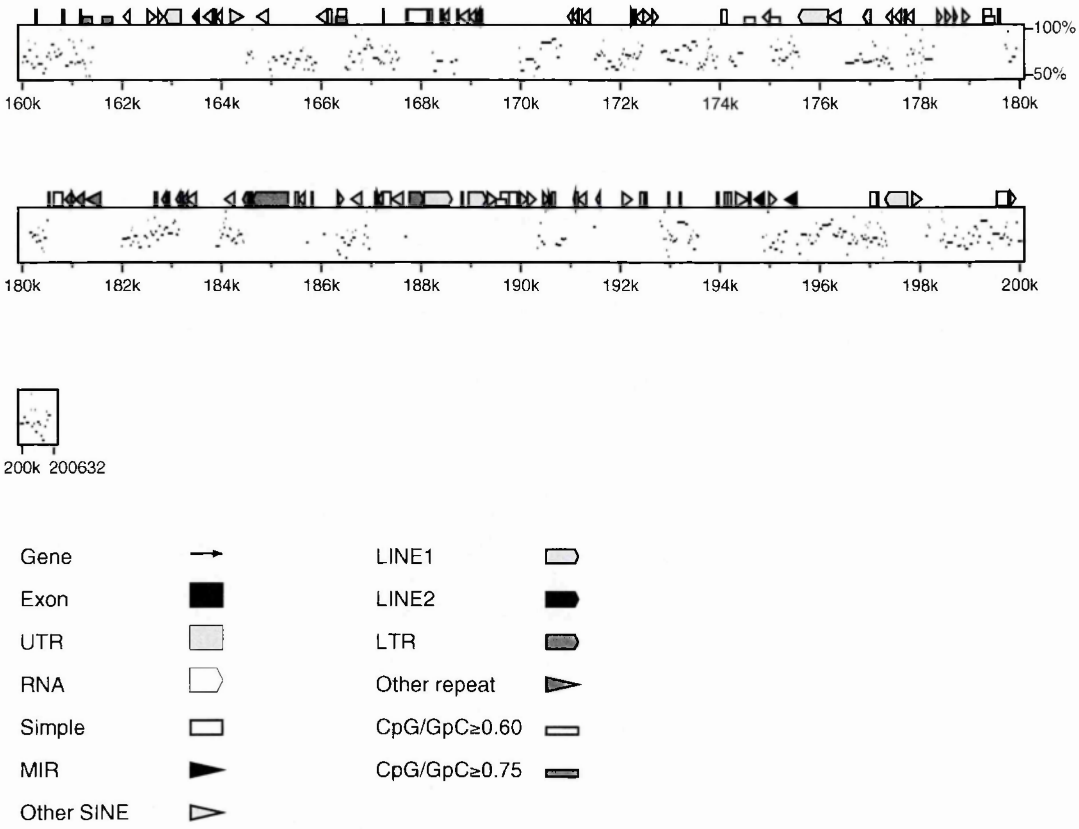


Fig. 2.14: genomic organisation of the mouse *HoxA* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Comparisons of nucleotide sequence between mouse *HoxA* and human *HOXA* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the mouse *HoxA* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

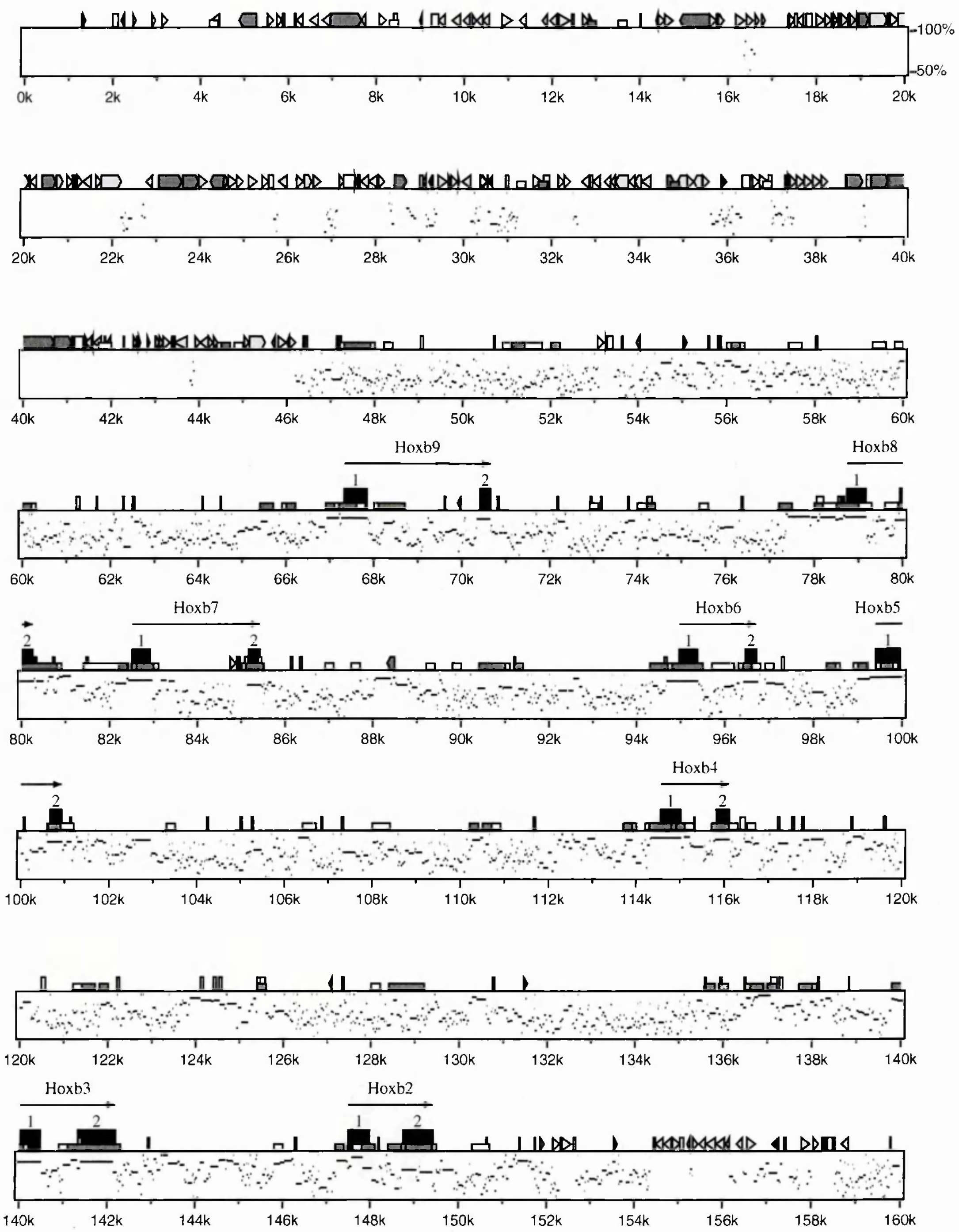


Fig. 2.15: genomic organisation of the mouse *HoxB* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the next page. Black arrows indicate the direction of the transcription of the corresponding genes. Comparisons of nucleotide sequence between mouse *HoxB* and human *HOXB* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the mouse *HoxB* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

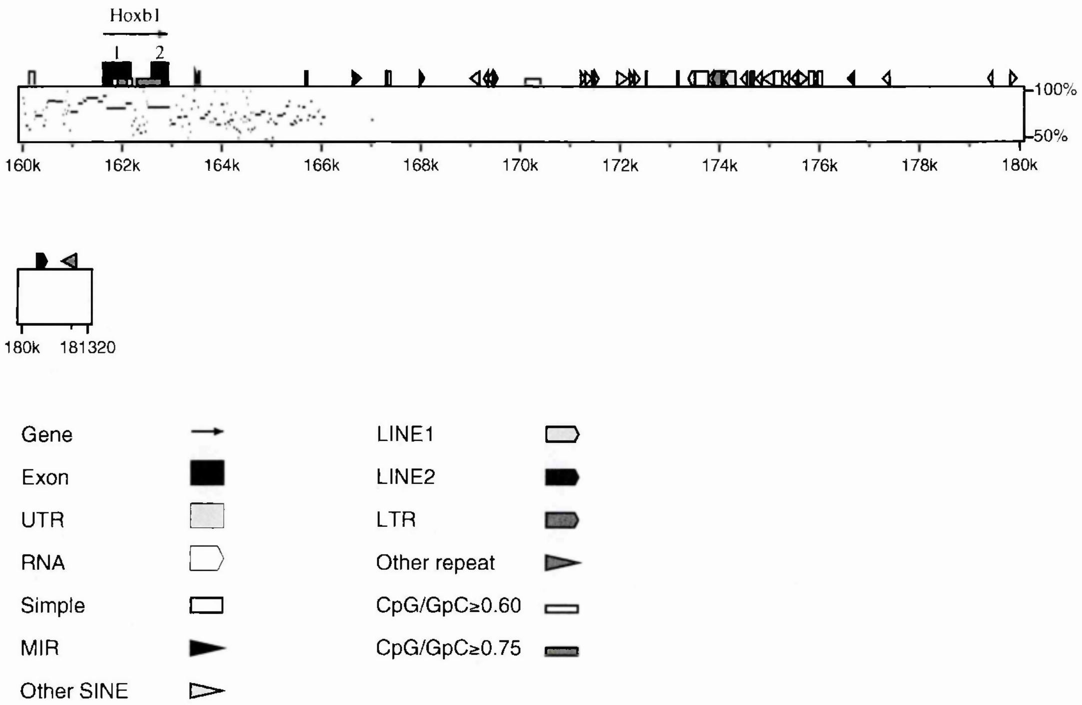


Fig. 2.15: genomic organisation of the mouse *HoxB* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Comparisons of nucleotide sequence between mouse *HoxB* and human *HOXB* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the mouse *HoxB* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

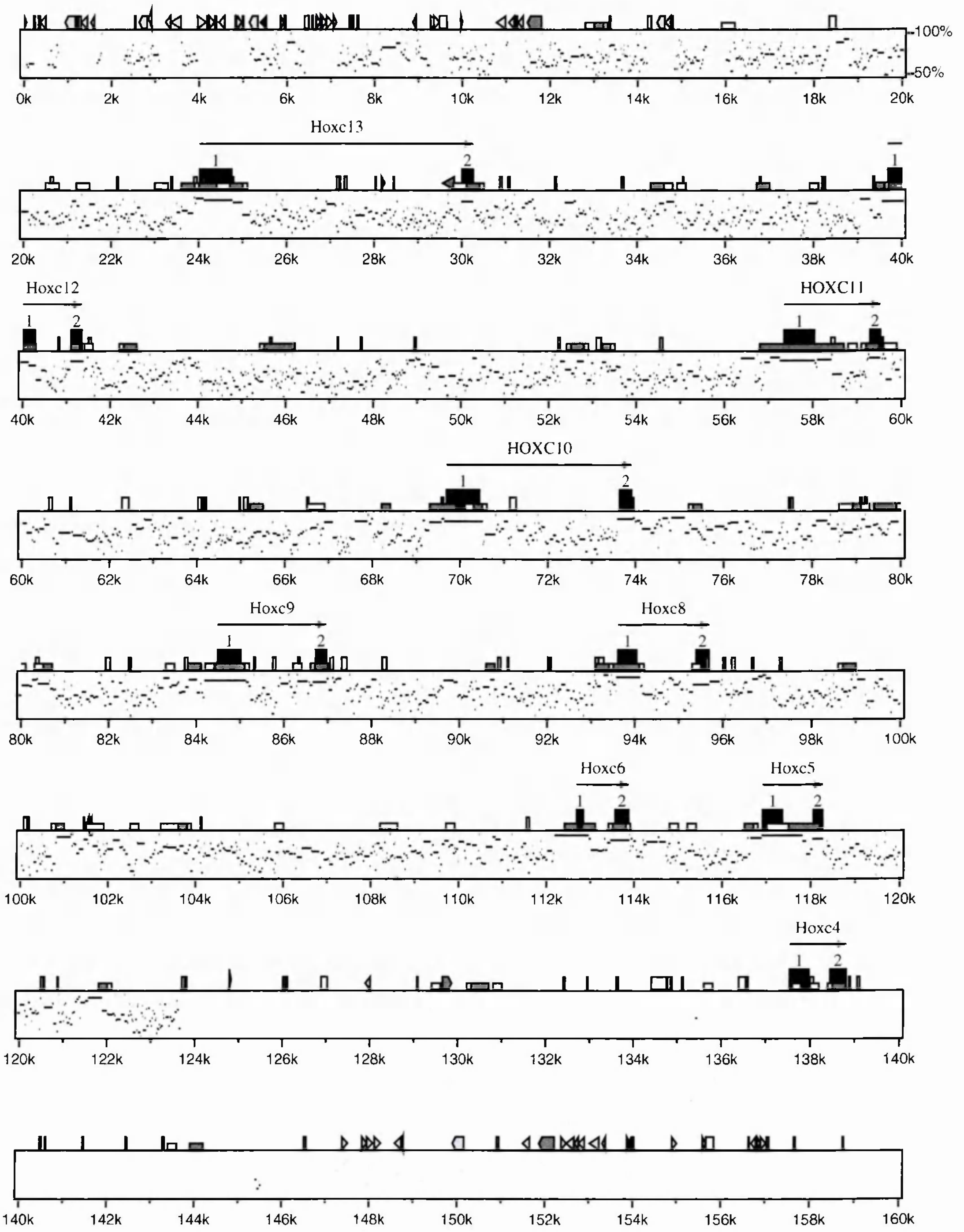

100%
50%
0k 2k 4k 6k 8k 10k 12k 14k 16k 18k 20k
Hoxc13
1
2
1
20k 22k 24k 26k 28k 30k 32k 34k 36k 38k 40k
Hoxc12
1 2
HOXC11
1 2
40k 42k 44k 46k 48k 50k 52k 54k 56k 58k 60k
HOXC10
1 2
60k 62k 64k 66k 68k 70k 72k 74k 76k 78k 80k
Hoxc9
1 2
Hoxc8
1 2
1
80k 82k 84k 86k 88k 90k 92k 94k 96k 98k 100k
Hoxc6
1 2
Hoxc5
1 2
100k 102k 104k 106k 108k 110k 112k 114k 116k 118k 120k
Hoxc4
1 2
120k 122k 124k 126k 128k 130k 132k 134k 136k 138k 140k
140k 142k 144k 146k 148k 150k 152k 154k 156k 158k 160k

Fig. 2.16: genomic organisation of the mouse *HoxC* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the next page. Black arrows indicate the direction of the transcription of the corresponding genes. Human *HOXC11* and *HOXC10* genes were used to map the position of mouse *Hoxc11* and *Hoxc10*. Comparisons of nucleotide sequence between mouse *HoxC* and human *HOXC* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the mouse *HoxC* cluster as a reference; human sequence ends between *Hoxc5* and *Hoxc4*. Kilobase (k) markings refer to our genomic contig.

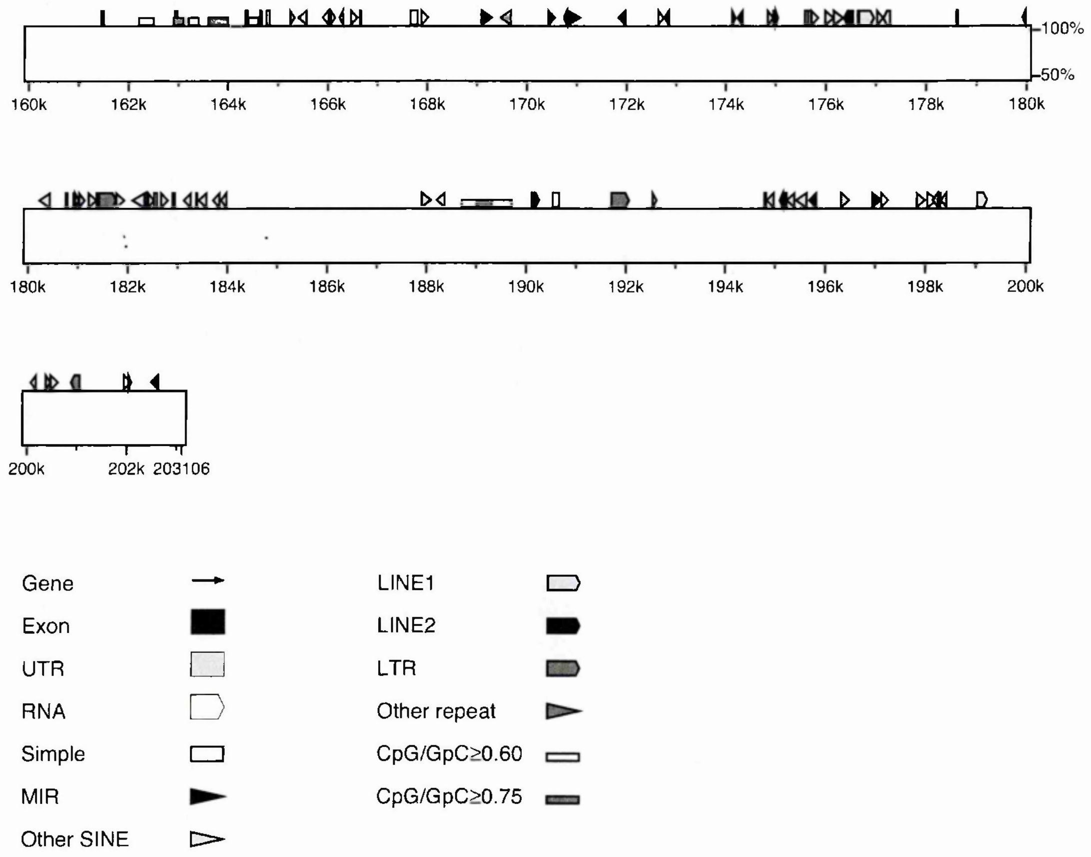


Fig. 2.16: genomic organisation of the mouse *HoxC* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Human *HOXC11* and *HOXC10* genes were used to map the position of mouse *Hoxc11* and *Hoxc10*. Comparisons of nucleotide sequence between mouse *HoxC* and human *HOXC* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the mouse *HoxC* cluster as a reference; human sequence ends between *Hoxc5* and *Hoxc4*. Kilobase (k) markings refer to our genomic contig.

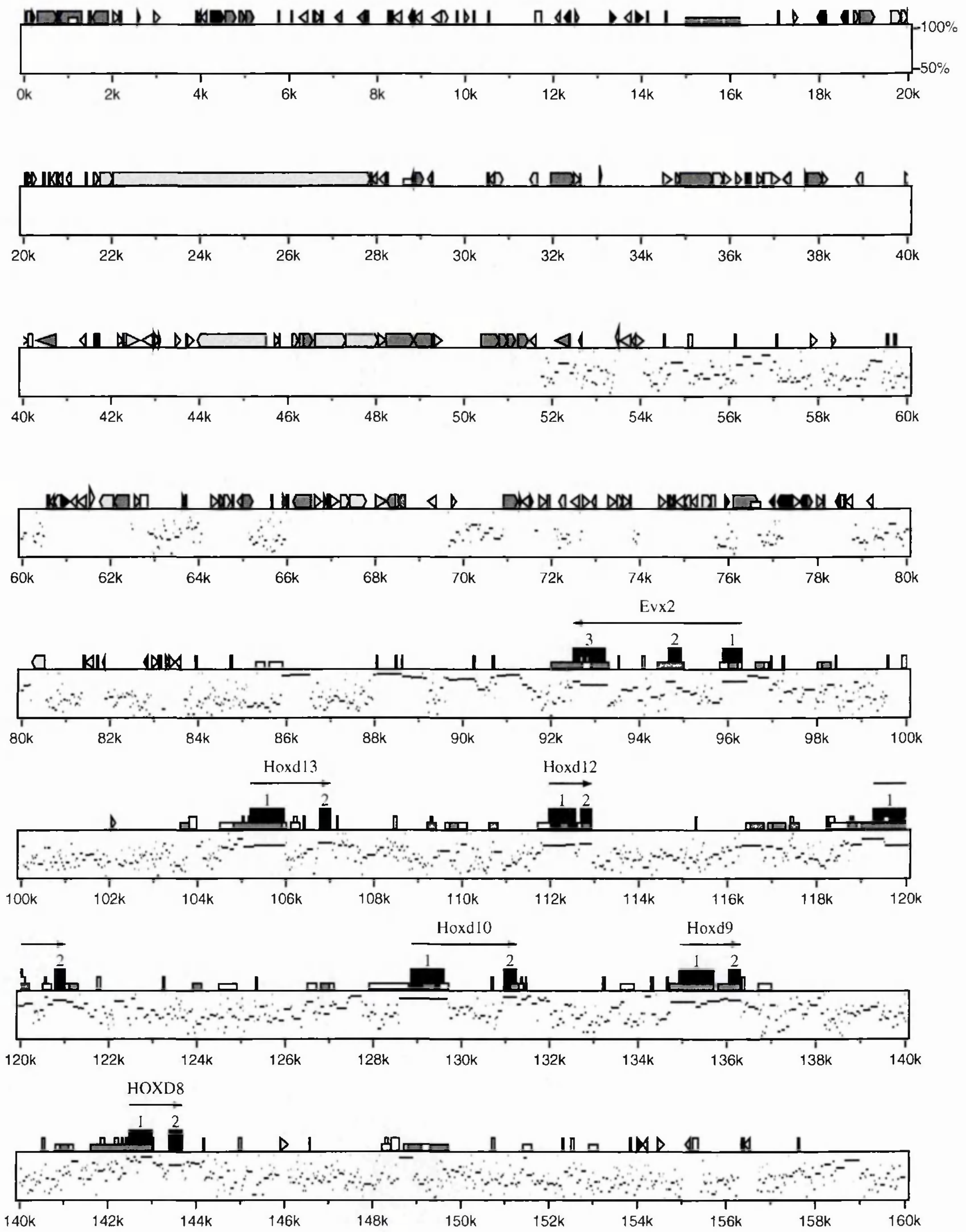


Fig. 2.17: genomic organisation of the mouse *HoxD* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the next page. Black arrows indicate the direction of the transcription of the corresponding genes. Human *HOXD8* gene was used to map the position of mouse *Hoxd8*. Comparisons of nucleotide sequence between mouse *HoxD* and human *HOXD* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the mouse *HoxD* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

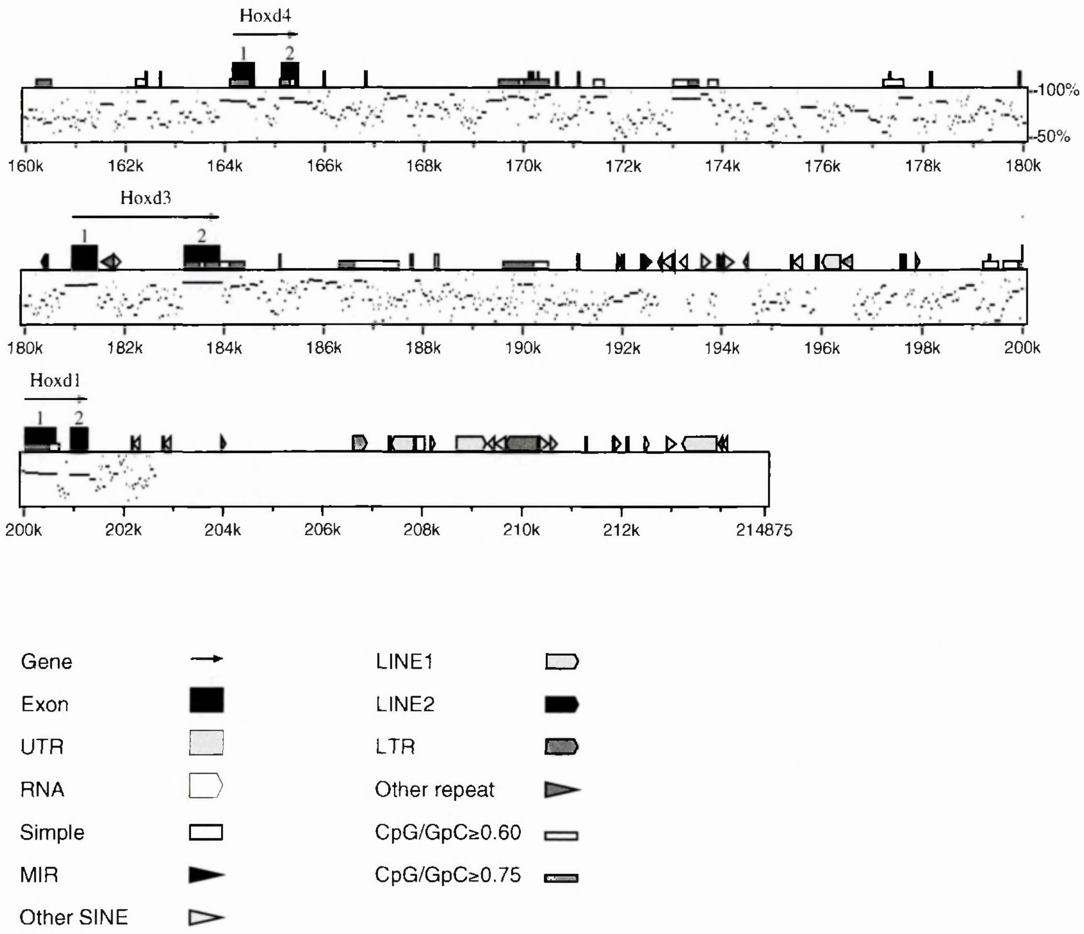


Fig. 2.17: genomic organisation of the mouse *HoxD* cluster.

Repetitive elements, genes and CpG islands are represented by different symbols and colours as described in the lower part of the figure. Black arrows indicate the direction of the transcription of the corresponding genes. Human *HOXD8* gene was used to map the position of mouse *Hoxd8*. Comparisons of nucleotide sequence between mouse *HoxD* and human *HOXD* clusters are shown in rectangles below the genomic line. Sequence comparisons are based on the mouse *HoxD* cluster as a reference. Kilobase (k) markings refer to our genomic contig.

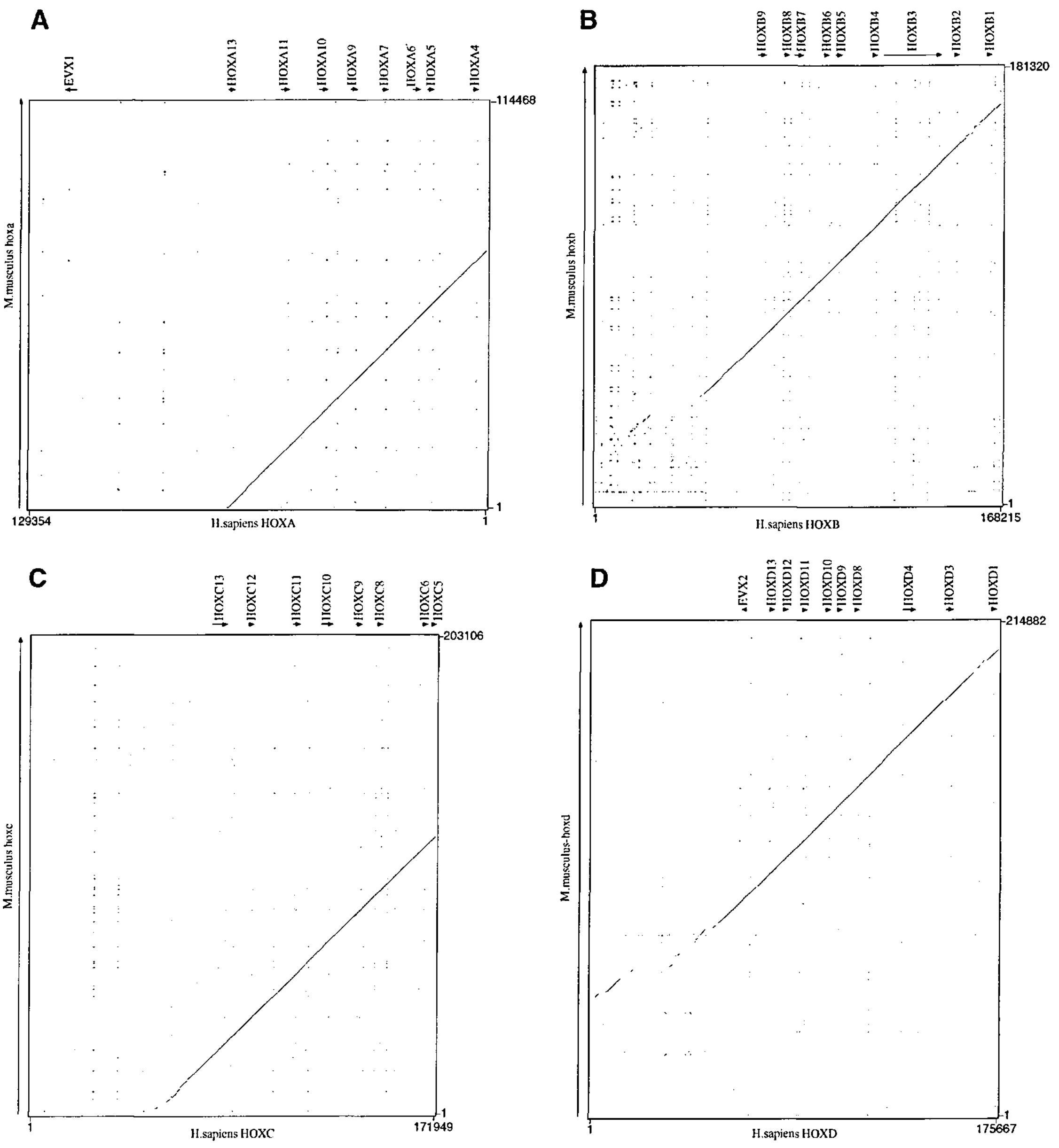


Fig. 2.18: *Hox* clusters percentage identity plots.

Sequence comparisons between human and mouse *Hox* clusters. In each plot, the x-axis corresponds to the human sequence, while the mouse orhologous locus is on the y-axis. Numbers represents sequences length. Small arrows indicate human gene positions. See text for discussion.

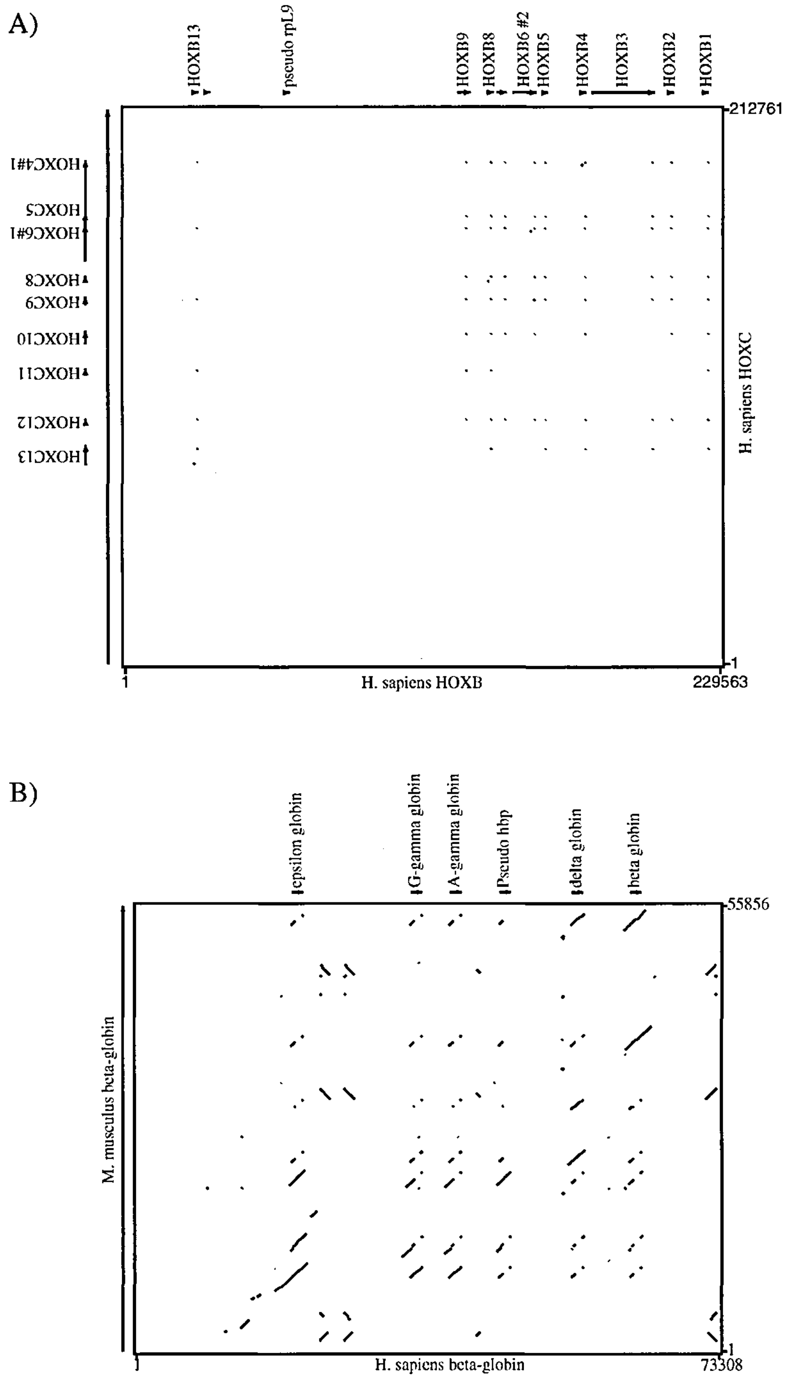


Fig. 2.19: Percentage Identity Plots.

A) Sequence comparison between human HOXB (x-axis) and HOXC (y-axis) clusters. Numbers represents sequences length; small arrows indicate genes. See text for discussion.

B) Sequence comparisons between human and mouse *beta-globin* gene clusters. The x-axis correspond to the human sequence, while the mouse orhologous locus is on the y-axis. Numbers represents sequences length. Genes are represented by small arrows. See text for discussion.

Non-coding intergenic transcripts coordinate *Hox* gene expression.

More than fifteen years ago non-coding RNAs were identified in the *BX-C* of the *D.melanogaster* (Lipshitz H.D., 1987); recently others intergenic transcripts have been extensively characterized (Bae E., 2002; Drewell R.A.; 2002; Rank G., 2003). These RNAs result from the transcription of large portions of regulative regions controlling *Hox* gene expressions. Interestingly, *BX-C* intergenic transcription precedes the transcription of the neighbouring *Hox* gene and seems to be necessary for appropriate *Hox* expression. As for *Hox* genes in the *BX-C*, the expression of these non-coding RNAs is also regulated according to their relative position along the cluster. Namely, both the coding and non-coding transcripts located at the 3' end of the cluster are transcribed before and in more anterior embryonic territories than their 5' counterparts. These observations lead to the hypothesis that the transcription of the intergenic regions may play an essential role in coordinating *Hox* genes expression. Moreover, non-coding RNAs appear to be implicated in several unexpected aspects of chromatin reprogramming and control of gene expression (i.e., X chromosome inactivation, dosage compensation, co-suppression and position effect variegation), suggesting that non-coding transcription may underscore the position of "non canonical" regulatory cis elements. Thus, we performed a systematic search for this kind of transcripts firstly inside the four human *HOX* clusters.

AntiHunter: searching BLASTN output for EST antisense (AS) transcripts.

The genomic sequences corresponding to the four human *HOX* loci represented the starting point for the identification of intergenic RNAs by a bionformatic approach. To this aim, in collaboration with Giovanni Lavorgna, I developed the "AntiHunter" software tool (available at http://bio.ifom.firc.it/ANTIHUNTER, Lavorgna G., 2003) capable of identify *in-silico* potential AS EST transcripts within a given genomic region of interest from BLASTN output.

AntiHunter takes as input a genomic sequence and a list of annotated transcripts of the genomic region. This list includes transcript names, their beginning and ending positions plus their strand occurrence. Then, it will perform the following tasks:

1) Run the RepeatMasker (http://repeatmasker.genome.washington.edu/cgi-

bin/RepeatMasker) program on the genomic sequence in order to filter out repeated sequences.

2) Perform a BLASTN search of the resulting sequence vs. the EST database.

3) Scan the BLASTN output looking for AS EST respect to the annotated genes.

4) Report the results to the user by Email.

The program gains independent information on EST strand source by looking (i) at the splice junctions of the genomic region matching a spliced EST and (ii) to the presence of a PolyA tail in "3'" annotated ESTs. Only EST showing at least one of these independent evidences for strand source are further considered for potential sense-antisense pairing. Moreover, since Oligo(dT)-priming can also take place on internal PolyA stretches within an unspliced transcript, the algorithm identifies such genomic PolyA stretches and disregards the relative PolyA information obtained from the EST sequence. AntiHunter can be used, in principle, to analyze genomic regions from any species for which there are EST and genomic data available. Moreover, it was also able to identify, despite the fact that it was using as query a human sequence, the presence of EST antisense transcripts from other species than human, unravelling the possible evolutionary conservation of the phenomenon. Also, AntiHunter can tolerate a variable number of bases between an annotated gene and an antisense transcript. This can be useful in detecting AS transcripts to genes with only partially characterized 5' and/or 3' ends. It can also facilitate the detection of transcribed gene regulatory regions that originate from intergenic regions and that contribute to regulation of their neighbour genes. The accuracy of AntiHunter was tested using genes that possess AS transcripts and many examples can be found at the web page: http://bio.ifom.firc.it/ANTIHUNTER.

"In silico" identification of intergenic "Opposite Strand" (OS) RNAs in human *HOX* clusters.

As all human *HOX* genes are transcribed on the same strand ("sense strand"), we used "AntiHunter" to identify intergenic OS-ESTs transcribed from the opposite DNA strand ("antisense strand"). Indeed we found this kind of transcripts inside the four human *HOX* clusters. Their mapping revealed an overlap between some AS-ESTs with

HOX sense transcripts, as well as an intergenic localization, with no overlap with the *HOX* genes, for other AS-ESTs. As the latter transcripts cannot be formally considered as “true” antisenses, we decided to refer generically to all the identified ESTs as “Opposite Strand (OS) ESTs”.

All the spliced intergenic OS-EST clones identified in the human *HOX* cluster loci are schematically grouped in Tables 2.16-2.19. We decided to focus our analysis only on the spliced OS-ESTs because the sequences corresponding to the acceptor and donor splicing sites are highly conserved (99.24% of introns have GT-AG at their 5’-3’ ends, respectively) and can be used to confirm the DNA strand that has been transcribed. Those spliced OS-EST clones that do not follow the GT-AG rule, probably result from misleading annotations of the EST sequences and were not further investigated. As shown in Tables 2.16-2.19, “opposite-strand” transcripts derived from many different tissues (both normal and malignant), from adult and foetal samples and from different cell lines. Thus their expression is not confined to particular tissues or specific developmental stages. OS-ESTs mapping in the human *HOX* loci, revealed that some of them present very large introns. In the *HOXB* cluster, for example, in the OS-ESTs AI685673 and AI125255 there are introns bigger than 13 or 50 kb, respectively. In the human *HOXA* cluster, we identified a spliced OS transcript having an intron larger than 25 kb (see below).

We look for the presence of ORFs in OS-ESTs belonging to the human *HOXA* cluster. The vast majority of these OS-ESTs contain only very small ORFs (14-80 aminoacids). In some other cases, computational analysis using gene prediction methods identified hypothetical proteins with no similarity to any known protein (i.e. OS-EST AK022839, hypothetical protein XP_168220; OS-EST BC035889, hypothetical protein XP_212093; OS-EST AK091933, hypothetical protein XP_212080). Therefore we concluded that these OS-ESTs most probably represent non-coding transcripts. Nevertheless, we should keep in mind that these ESTs could correspond to fragments of cDNAs and could not represent full-length transcripts. Thus, the complete ORFs analysis should be repeated with full-length cDNAs.

"In vivo" identification of intergenic OS RNAs inside the *HOXA* cluster.

Moving from the *"in silico"* to the *"in vivo"* approach, the existence of OS transcripts was investigated in different human tissues. The analysis was done in the *HOXA* cluster by RT-PCR using both foetal and adult tissue samples. Table 2.20 presents the primer pairs used to amplify both *HOXA* genes and OS-EST transcripts. These OS-ESTs are distributed along the entire *HOXA* cluster and localized between *HOXA1-A2* (OS-ESTs: BG325728, AA489505); between *HOXA3-A4* (OS-ESTs: BI823151, BE8733499); between *HOXA6-A7* (OS-EST: BF510786); between *HOXA11-A13* (OS-EST: BE305073). First of all, specific "single strand RT-PCRs" were performed to prove that OS-ESTs derive from the transcription of the "antisense strand". We confirmed the strand specificity of all OS-ESTs investigated. Then, RT-PCRs were repeated with cDNAs from seven adult and seven foetal human tissues. These cDNAs were obtained from total RNAs of different tissues retrotranscribed with an oligo(dT). As shown in Fig. 2.20 and Fig 2.21, the vast majority of OS-ESTs transcripts were detectable in many different tissues. One striking feature is represented by the expression of OS-ESTs that seems to correlate with the expression of the neighbour *HOX* genes. Due to this characteristic, it seems more likely that OS-ESTs transcription plays a role in promoting or sustaining *HOX* gene expression, rather than counteracting it.

Analysing the AntiHunter output, we realised that different OS-ESTs could correspond to alternative spliced forms of the same transcripts. In particular, we focus our attention in the region between *HOXA1* and *HOXA2*. We investigated the existence of these alternative splice OS transcripts by mean of RT-PCRs, both in adult and foetal human tissues. Using a specific primer pair, we identified four different PCR products (Fig. 2.22). These amplified fragments were cloned and their sequences confirmed that they are alternative spliced OS transcripts belonging to this intergenic region. All these antisenses are correctly spliced, none of them present a significative ORF. Interestingly, this analysis revealed that some of these alternative forms are tissue specific, especially in foetal samples (Fig. 2.22).

We then started the characterization of long intergenic OS transcripts in the *HOXA* cluster (Fig. 2.23). Although not predicted by AntiHunter, we supposed that dif-

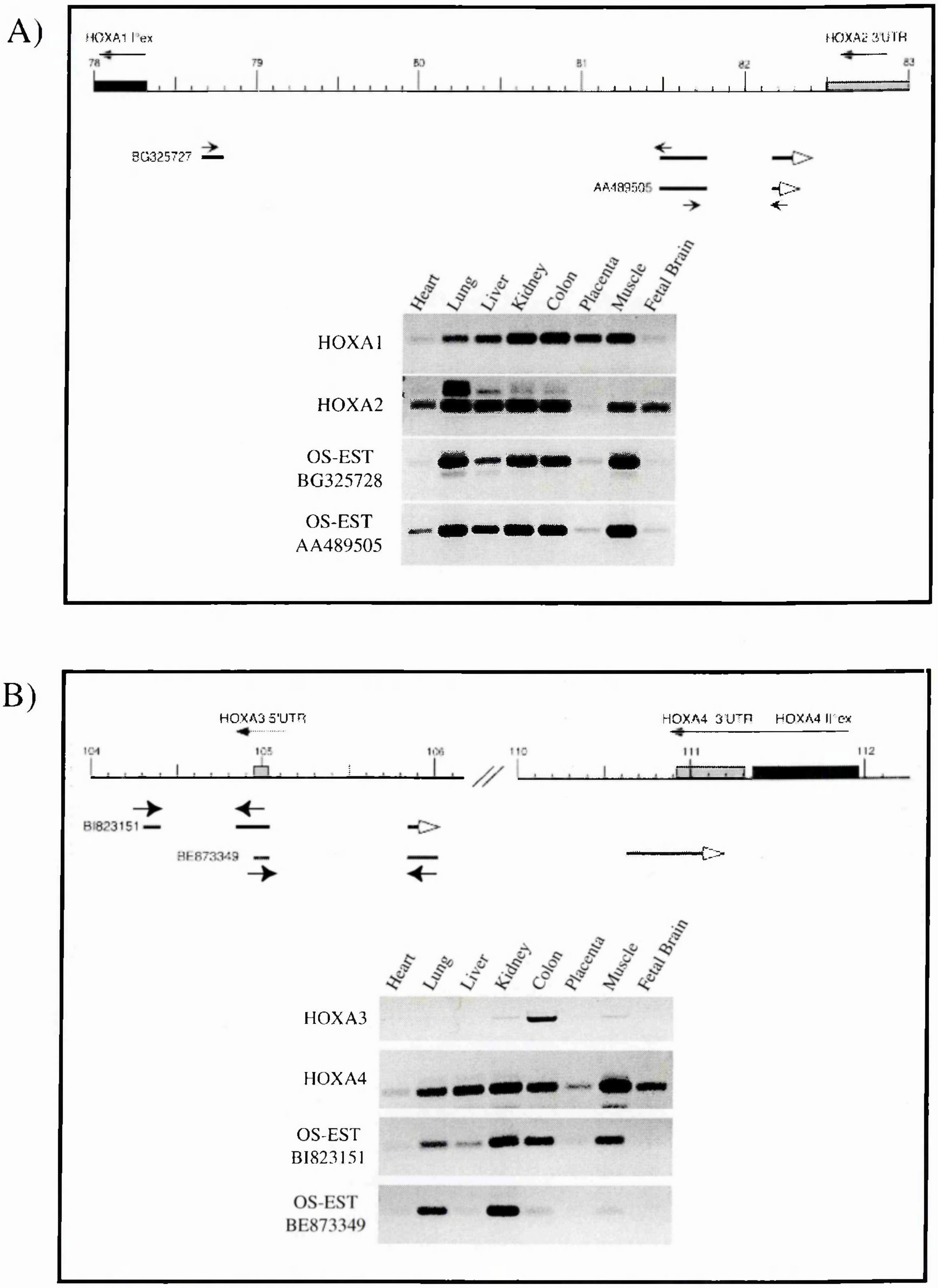


Fig. 2.20: Adult tissues RT-PCRs.

The figure is continued in the next page.

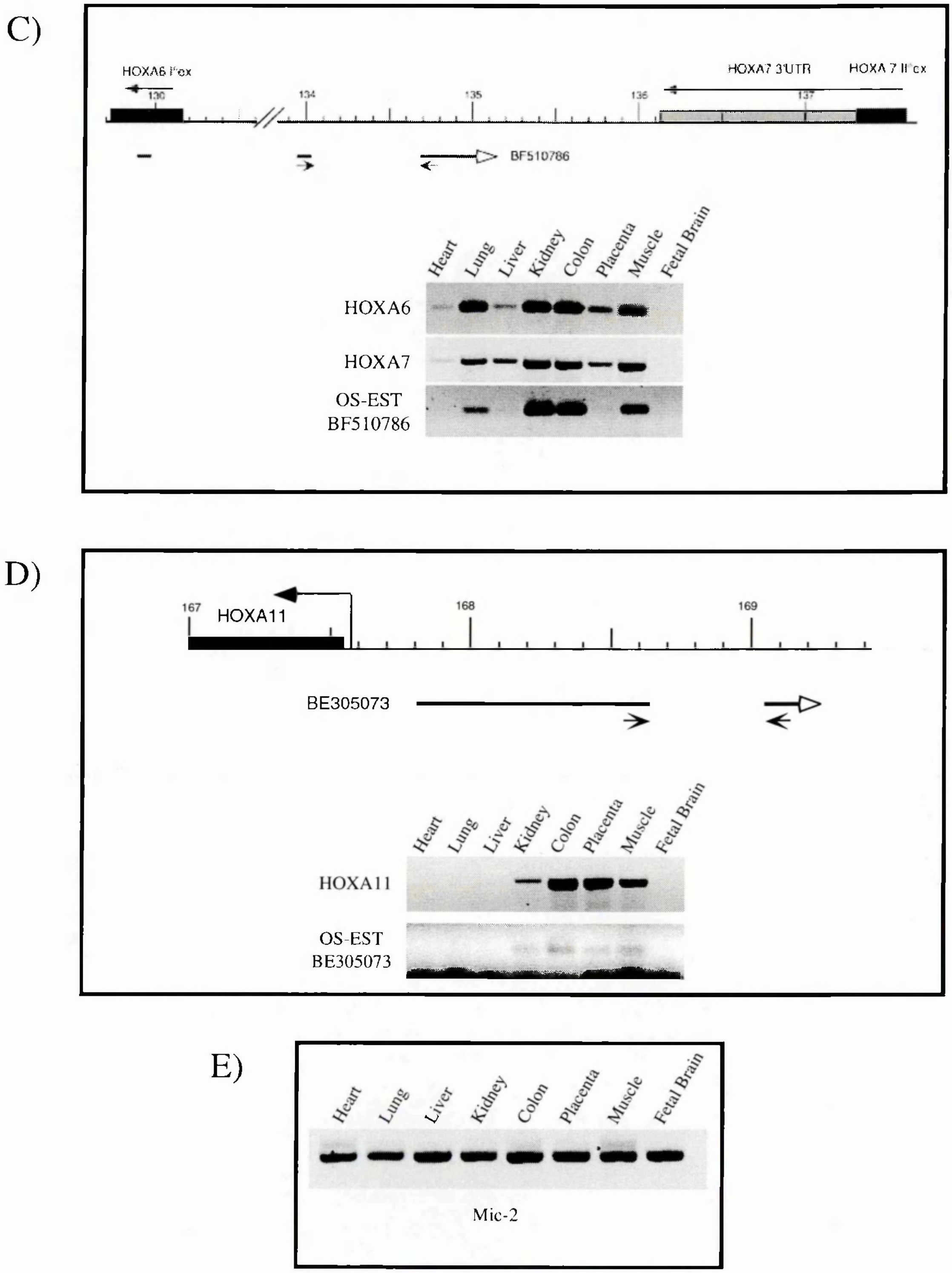


Fig. 2.20: Adult tissues RT-PCRs.

Intergenic regions of the *HOXA* locus are schematically shown (**A-D**). Kilobase markings refer to our genomic contig. Black rectangles correspond to coding exons while gray rectangles correspond to 5' and 3' UTR *HOXA* regions. Long black arrows indicate the transcriptional orientation of *HOXA* genes; long white arrows indicate both exon positions and transcriptional orientation of OS-ESTs; short black arrows specify position and orientation of primer pairs used for OS-EST RT-PCRs. RT-PCRs for both *HOXA* genes and OS-ESTs are shown below the genomic maps (**A-D**); RT-PCR normalization with a house-keeping gene is shown in (**E**). Tissues are indicated above the RT-PCR imagines. See Table 2.20 for PCR details.

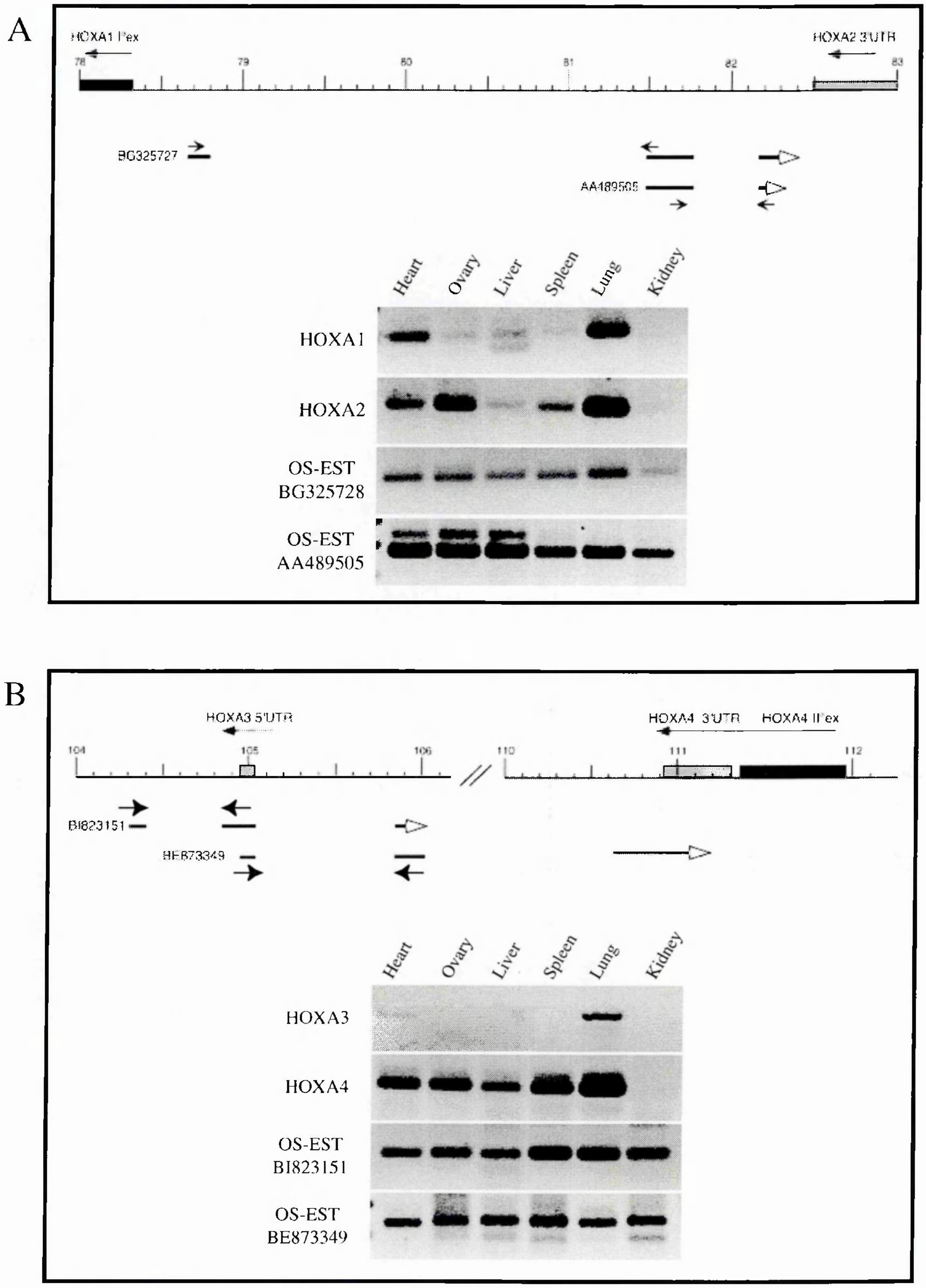


Fig. 2.21: Foetal tissues RT-PCRs.

The figure is continued in the next page.

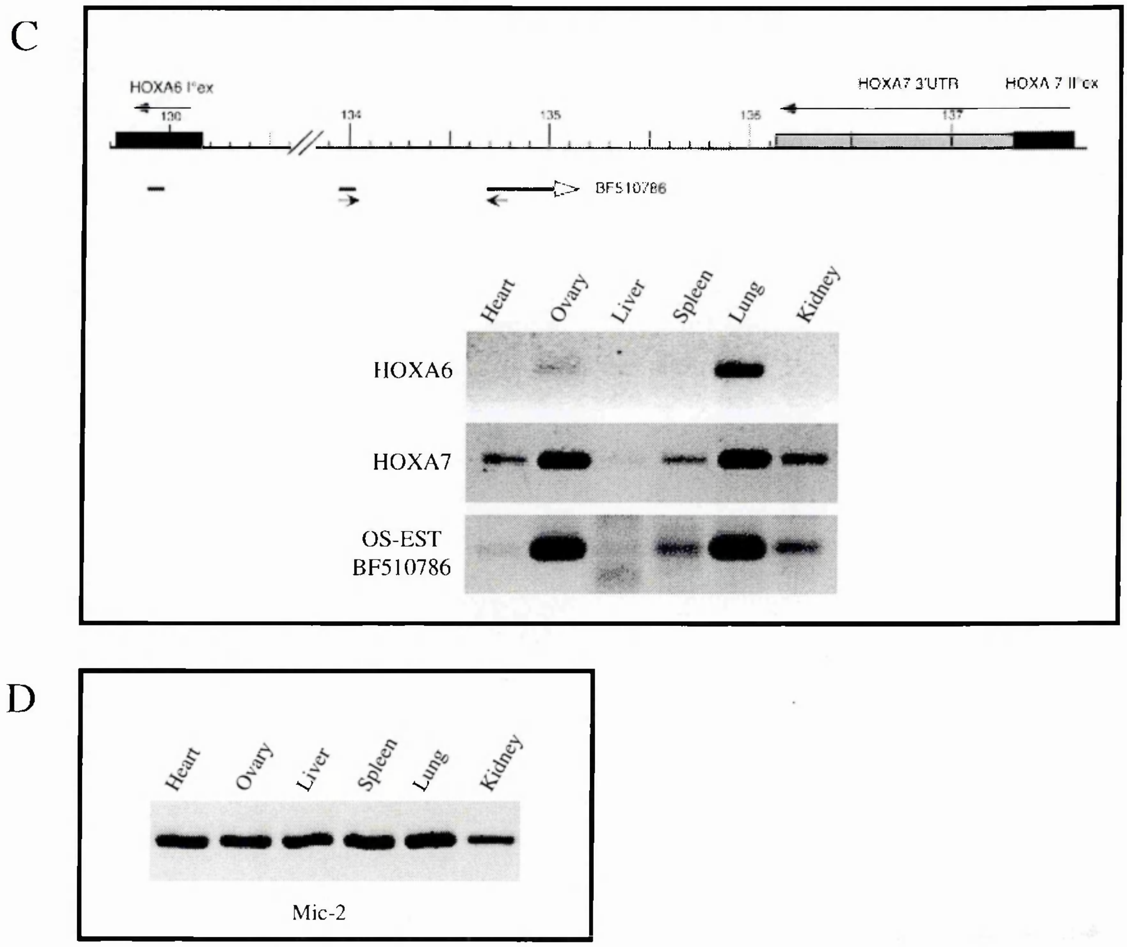


Fig. 2.21: Foetal tissues RT-PCRs.

Intergenic regions of the *HOXA* locus are schematically shown (**A-C**). Kilobase markings refer to our genomic contig. Black rectangles correspond to coding exons while gray rectangles correspond to 5' and 3' UTR *HOXA* regions. Long black arrows indicate the transcriptional orientation of *HOXA* genes; long white arrows indicate both exon positions and transcriptional orientation of OS-ESTs; short black arrows specify position and orientation of primer pairs used for OS-EST RT-PCRs.

RT-PCRs for both *HOXA* genes and OS-ESTs are shown below the genomic maps (**A-C**); RT-PCR normalization with a housekeeping gene is shown in (**D**). Tissues are indicated above the RT-PCR imagines. See Table 2.20 for PCR details.

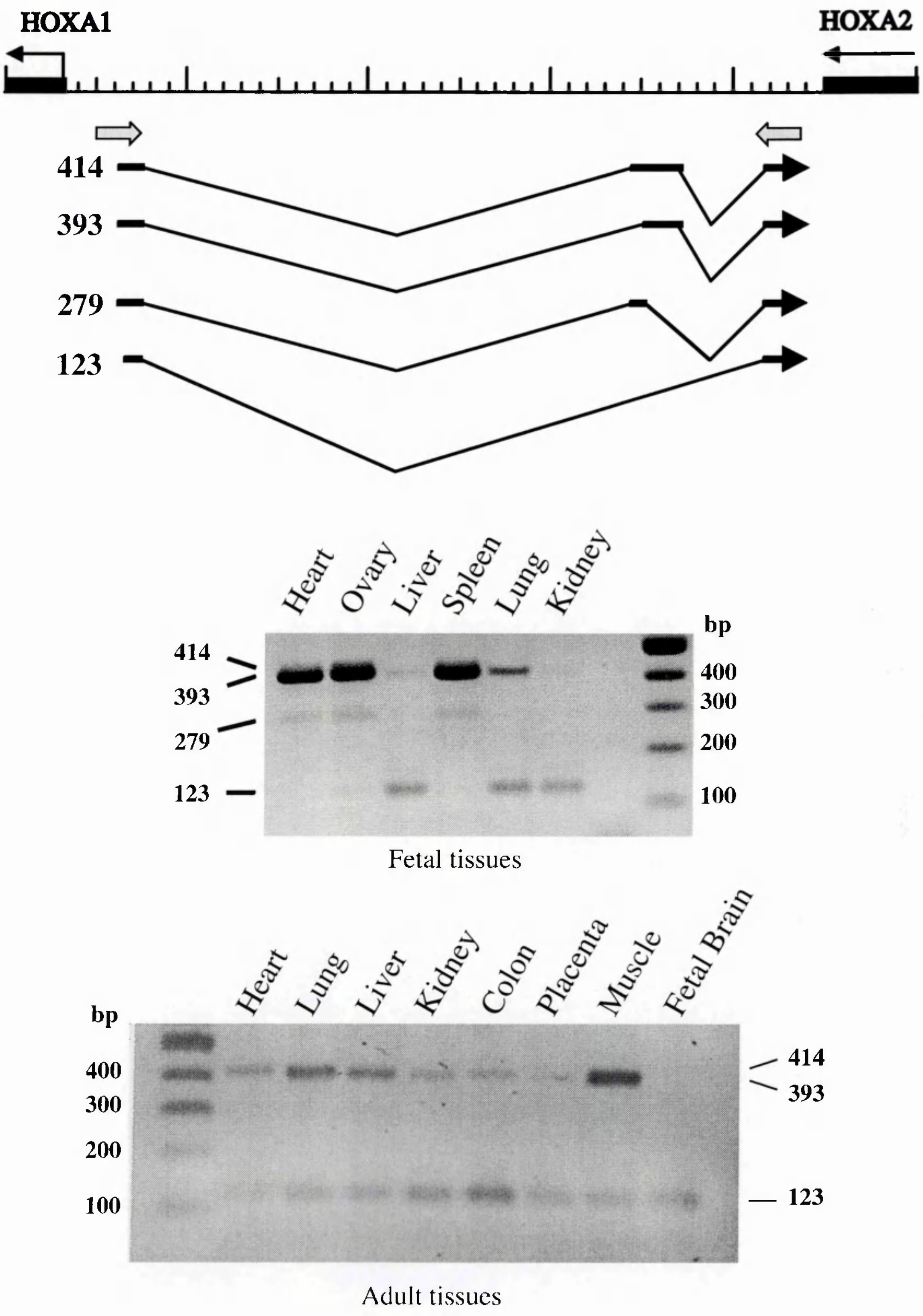


Fig. 2.22: OS-EST alternative spliced forms.

Human *HOXA1-HOXA2* intergenic region is shown in the upper part of the figure. Kilobase markings refer to our genomic contig. Black rectangles correspond to *HOXA* exons. Thin black arrows indicate the transcriptional orientation of *HOXA* genes; long black arrows indicate both exon positions and transcriptional orientation of alternatively spliced OS-ESTs. Short gray arrows represent PCR primers (BG325728.U and AA489505.L; see Table 2.20 for PCR details). RT-PCR product lengths (in bp) are indicated on the left. Amplification patters obtain with human fetal and adult tissues are shown in the lower part of the figure.

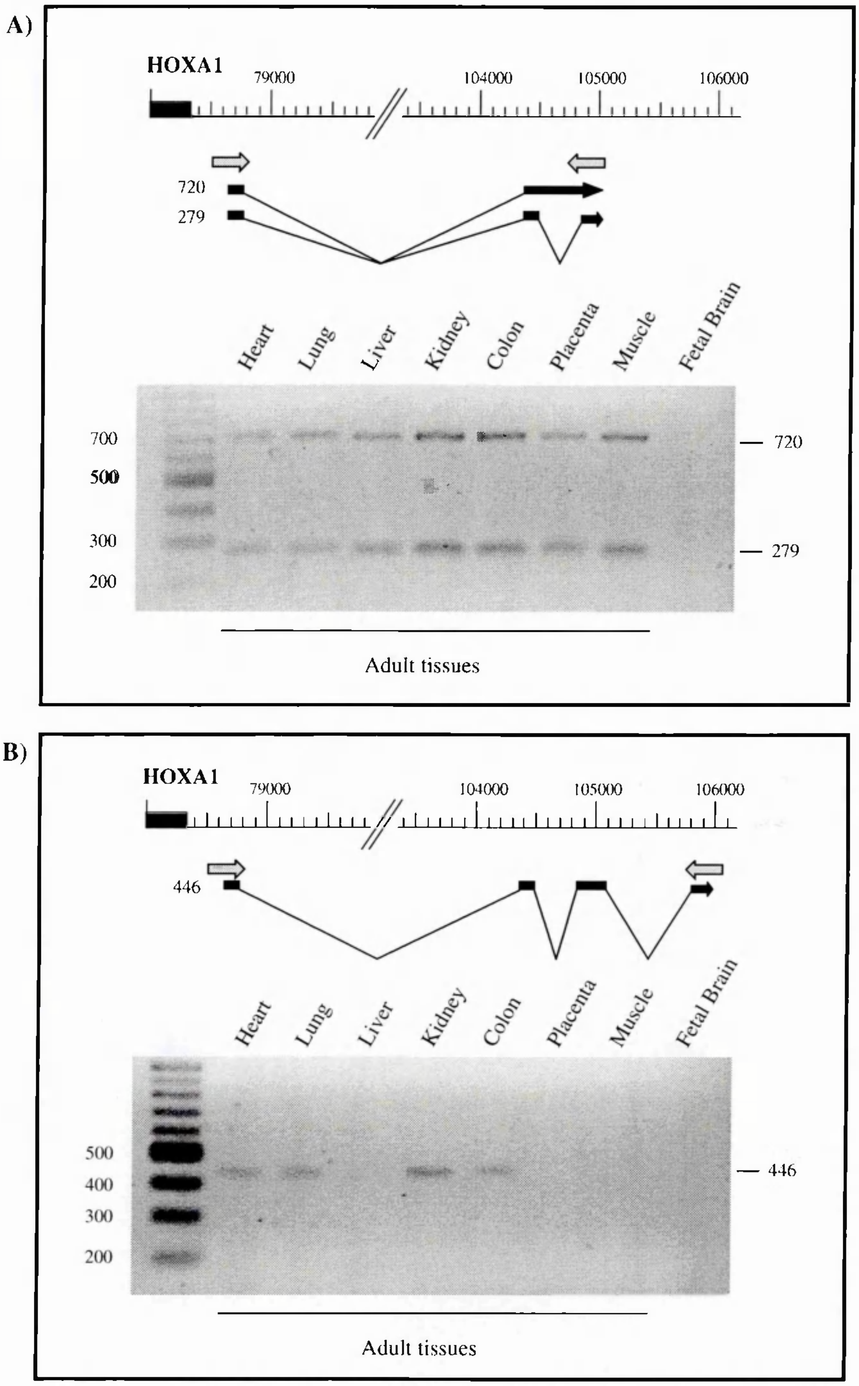


Fig. 2.23: long intergenic OS transcripts.

A) Long intergenic OS transcripts detected with BG325728.U and BI823151.L primers (see Table 2.20 for PCR details). Kilobase markings refer to our genomic contig. Black rectangle corresponds to *HOXA* first exon. Long black arrows indicate both exon positions and transcriptional orientation of alternatively spliced OS transcripts. Short gray arrows schematically represent PCR primers. RT-PCR product lengths (in bp) are indicated on the left. Amplification patters obtain with human adult and fetal tissues are shown in the lower part of the figure.

B) Long intergenic OS transcripts detected with BG325728.U and BE873349.L primers (see Table 2.20 for PCR details). Kilobase markings refer to our genomic contig. Black rectangle corresponds to *HOXA* first exon. Long black arrows indicate both exon positions and transcriptional orientation of OS transcripts. Short gray arrows schematically represent PCR primers. RT-PCR product lengths (in bp) are indicated on the left. Amplification patters obtain with human adult and fetal tissues are shown in the lower part of the figure.

ferent OS-EST could be part of the same transcript and analysed their existence by RT-PCRs. In this way we were able to identify three different long intergenic OS transcripts (Fig. 2.23). From their sequence analyses, we confirmed that these transcripts map in the *HOXA* cluster and are correctly spliced. We identified alternative spliced and tissue specific forms also in this kind of transcripts.

Intergenic OS RNAs are transcribed before *HOX* genes.

The intergenic non-coding RNAs in the *D.melanogatser* present two major features. In early embryos, they are expressed before and in the same territories as the flanking sense genes. Our characterisation clearly demonstrated that the expression of both intergenic OS- ESTs and *HOXA* genes takes place in the same tissues. To investigate the expression timing of these OS transcripts, we used the NT2 cell line. This teratocarcinoma cell line has been extensively used to study the collinear expression profile of the *HOX* genes after treatment with retinoic acid (RA, Fig. 2.24 and Mavilio F., 1993). The RA adbministration induces the expression of *HOX* genes from the 3' to the 5' regions of the cluster (in the *HOXA* cluster, *HOXA7* is generally the most 5' gene activated with RA). Thus we decide to repeat the RA treatment analysing, by RT-PCR, the expression of both of *HOXA* genes and OS-ESTs. We choose different OS-EST transcripts distrib-

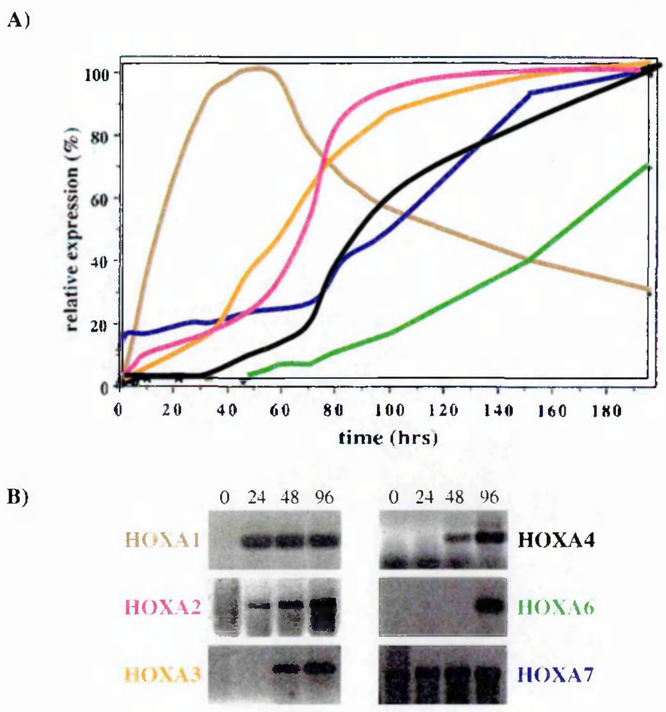


Fig. 2.24: *HOXA* expression patterns in NT2 cell line treated with RA.

A) Profiles of *HOXA* gene expression in NT2 cell line treated with RA as published in the PRO HOX database (http://www.evol.nw.ru/labs/lab38/spirov/hox_pro/hox-pro00). *HOXA* gene colors are as in B). **B**) *HOXA* RT-PCR amplifications. NT2 cells were treated with RA (10^{-7} M) for 24, 48, 96 hours. Total RNA was retrotrancribed with Oligo(T) and amplified with specific *HOXA* primers as described in Table 2.20. There is a good correlation between these *HOXA* expression profiles (A) and the expected ones (B).

uted along the entire *HOXA* cluster and localized between *HOXA1-A2*, between *HOXA3-A4*, between *HOXA6-A7*, between *HOXA11-A13*. Total RNA was extracted from NT2 cells after 0, 1, 2, 4, days of RA treatment and retrotranscribed with an oligo dT. Our analysis revealed that OS ESTs are transcribed contemporary (OS EST BG325728; OS EST AA489505; OS EST BI823151) or even before (OS EST BE873349; OS EST BF510786) the corresponding *HOX* senses (Fig. 2.25). Interestingly, also the OS-EST expression follows the spatial and the temporal colinearity, the most 3' OS-ESTs being transcribed before the 5' ones. Looking at the 5' regions of the *HOXA* cluster, neither *HOXA11* nor the OS transcripts localized between *HOXA11-A13* were expressed after RA treatment.

"In silico" identification of intergenic OS RNAs in "dispersed" homeotic loci.

As discussed in the introduction, the homeotic genes can be divided into two big *Superclasses*: the *Complex* and the *Dispersed Superclass*. The *Complex Superclass* is made of those homeobox genes that are organized into gene clusters, while the *Dispersed* one groups all the homeobox gene classes that are dispersed throughout the genome. By mean of AntiHunter, we were able to identify intergenic non-coding OS-ESTs in both human and mouse *Hox* loci belonging to the *Complex Superclass*. We then used both AntiHunter and the Santa Cruz Genome Browser (UCSC Genome Bionformatics: http://genome.ucsc.edu) to verify the existence of these kinds of OS transcripts also in human and mouse loci containing the "dispersed" homeobox genes.

As starting point for this *in silico* analysis, we used a phylogenetic tree recently published (Banerjee-Basu S., 2001) containing 91 homeobox genes belonging to the *Dispersed Superclass*. Other four known homeobox genes were added to this set, for a total of 95 dispersed loci. Using this approach we found OS-EST transcripts in the corresponding loci both in human and mouse. In particular we identified OS-ESTs in 34 human and 38 mouse dispersed loci (Fig. 2.26; Tables 2.21 and 2.22). Interestingly, both human and mouse show a common organization of OS-ESTs in 19 orthologous dispersed loci (Fig. 2.26). As for the *HOX* clusters, I focused my research only on correctly spliced OS-ESTs (Table 2.21 and Table 2.22). I also look for the presence of the com-

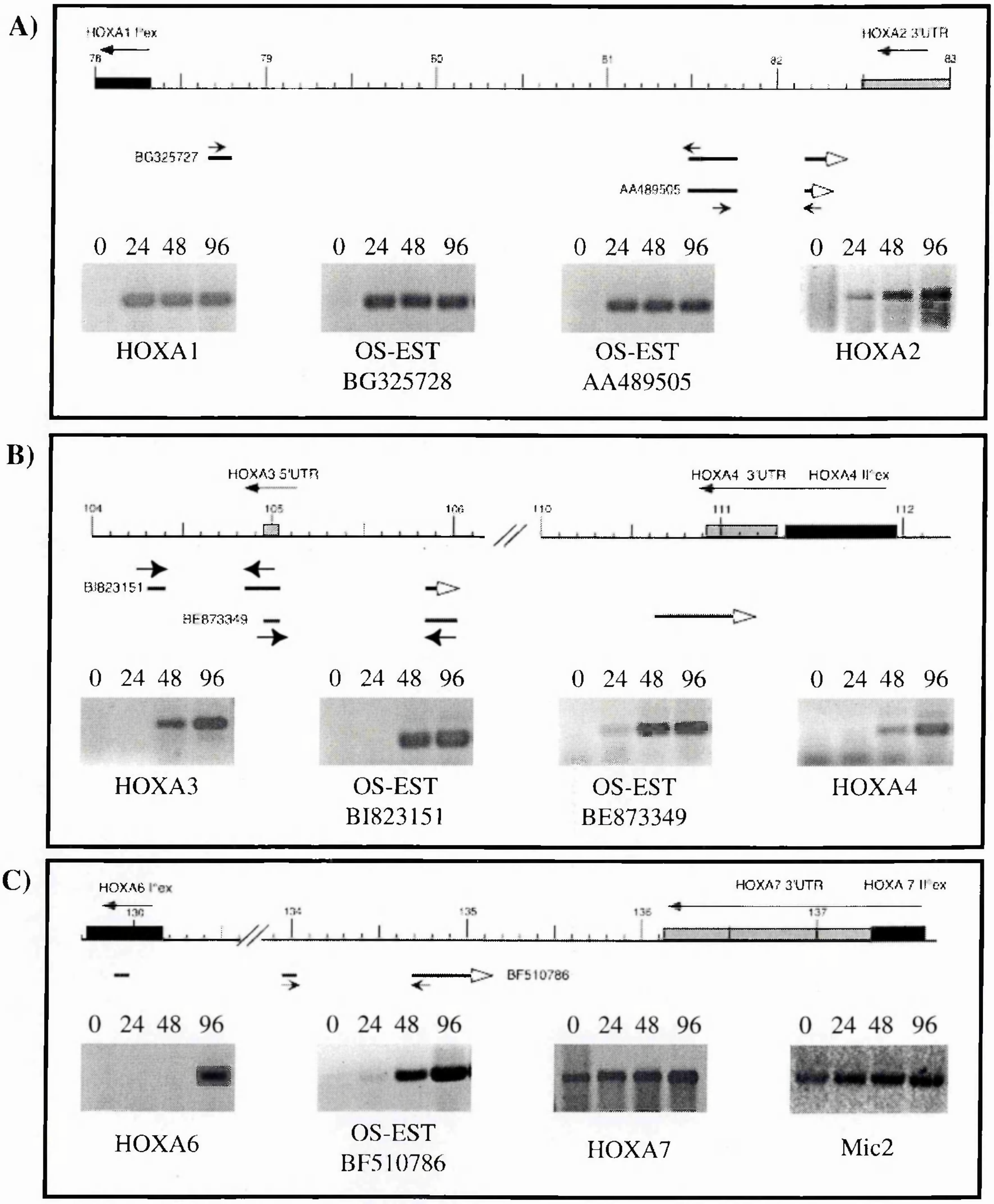


Fig. 2.25: Expression timing of both *HOXA* genes and OS transcripts in NT2 cells treated with RA.

Intergenic regions of the *HOXA* locus are schematically shown (**A-C**). Kilobase markings refer to our genomic contig. Black rectangles correspond to coding exons while gray rectangles correspond to 5' and 3' UTR *HOXA* regions. Long black arrows indicate the transcriptional orientation of *HOXA* genes; long white arrows indicate both exon positions and transcriptional orientation of OS-ESTs; short black arrows specify position and orientation of primer pairs used for OS-EST RT-PCRs. (**A-C**) RT-PCRs for both *HOXA* genes and OS-ESTs after RA treatment are shown below the genomic maps. Total RNA from NT2 cells, treated with RA for 0, 24, 48 and 96 hours, was retrotranscribed with oligo(T).cDNAs were amplified with primer pairs specific (see Table 2.20) for *HOXA* genes and OS-EST. RT-PCR normalization with the housekeeping gene *Mic-2* is shown in (**C**).

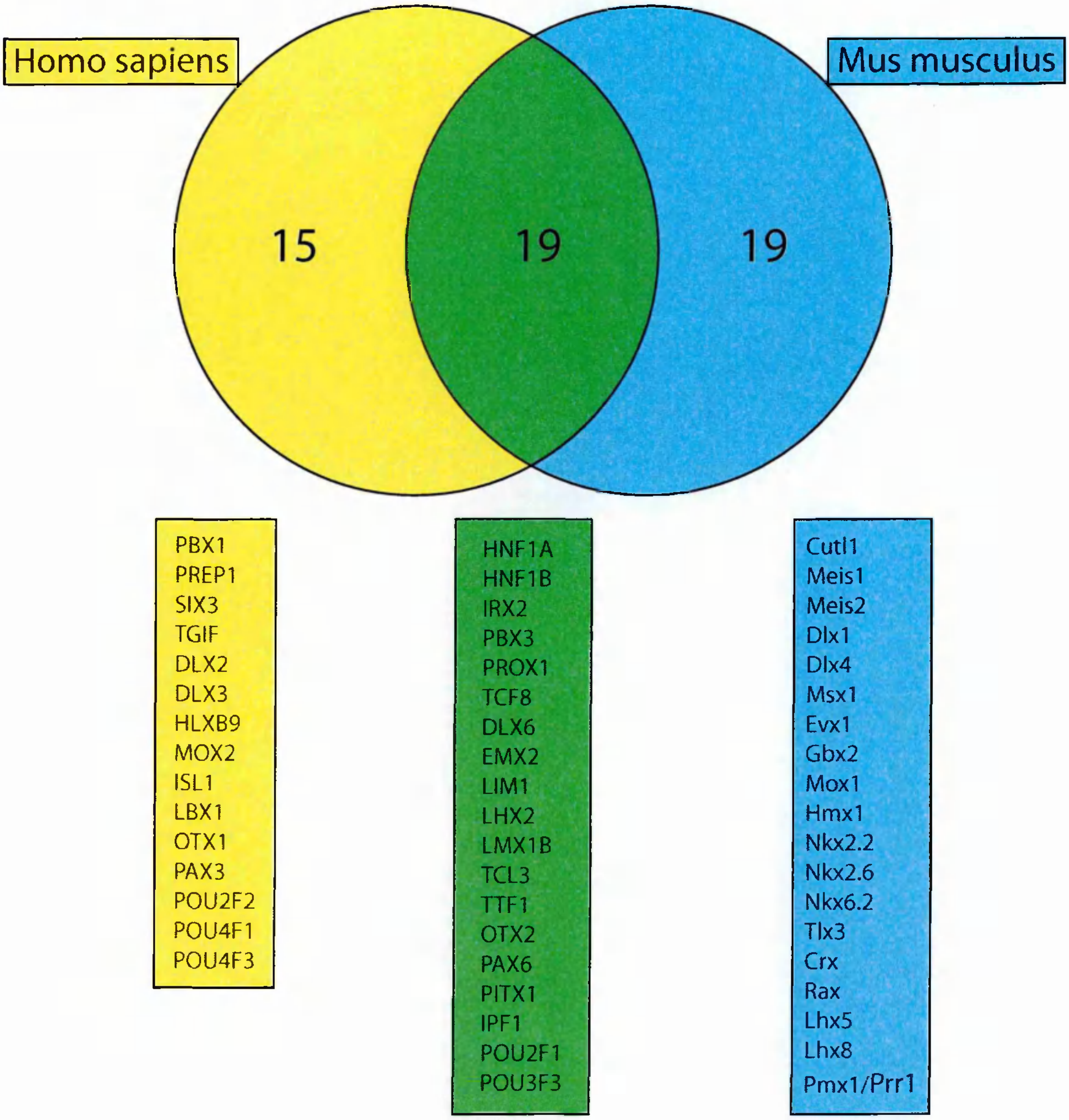


Fig. 2.26: Identification of OS transcripts in dispersed homeobox loci.

Human (yellow) and mouse (blue) dispersed homeobox loci presenting OS-ESTs. Loci characterized by the presence of OS transcripts in both species (green) are listed in the central part of the figure.

mon polyadenylation signal (AATAAA) in the OS-ESTs sequences that, indeed, was identified at the 3' end of many OS-ESTs (Table 2.21 and Table 2.22).

Keeping in mind that the *in silico* identified OS-ESTs could correspond to fragments of the full-length transcript, we analysed in detail the distribution of OS transcripts in the positive loci. In nearly all cases (97% in human and 71% in mouse), the dispersed loci present OS transcripts in the 5' regions of dispersed homeobox genes. The distance between the first exon of the OS-EST transcript and the homeobox gene can vary. The *EMX2* and the *TGIF1* loci (Table 2.21) represent two opposite situations. The first exon of *EMX2* overlaps with the first exon of an OS-EST transcript (AY117413; Noonan F.C., 2003), while the first exon of *TGIF1* is more than 10 kb away from the *in silico* mapped first exon of the OS-EST BG490809. Also the relative position between the first exon of OS EST transcripts and the homeobox gene can be different. There are five human (*EMX2*, *PAX3*, *HNF1A*, *PITX1*, *IPF1*) and eight mouse (*Hnf1a*, *Tcf8*, *Pax6*, *Pitx1*, *Ipf1*, *Dlx4*, *Evx1*, *Rax*) loci where the first exon of OS transcripts maps in intronic regions of the homeobox genes, while the other(s) exon(s) map in the 5' upstream regions. Other OS ESTs map completely in intronic regions (*MEOX2* locus in human; *Pbx3*, *Prox1*, *Lhx2*, *Pmx1*, *Cutl1*, *Meis1*, *Meis2*, *Meox1* in mouse). OS EST transcripts having their first exons in the 3' regions of the dispersed homeobox genes represent the last situation (*POU2F2*, *POU4F1* in human; *Hnf1a*, *Tlx1*, *Pou2f1*, *Dlx1*, *Msx1*, *Nkx6.2* in mouse). Among them, in human *POU4F1*, *POU4F3*, *DLX6* and *TCL3* loci and mouse *Dlx6*, *Dlx1*, *Nkx2.2* loci, there are OS-ESTs that "surround" completely the homeobox gene having the first exon in its 3' region and the other exons in the 5' upstream region of the homeobox gene. The human TLX1 locus is used to exemplify many of these different OS transcripts (Fig 2.27). Thus, as revealed by this *in silico* analysis, the presence of OS transcripts seems to be a common characteristic also for the homeobox genes belonging

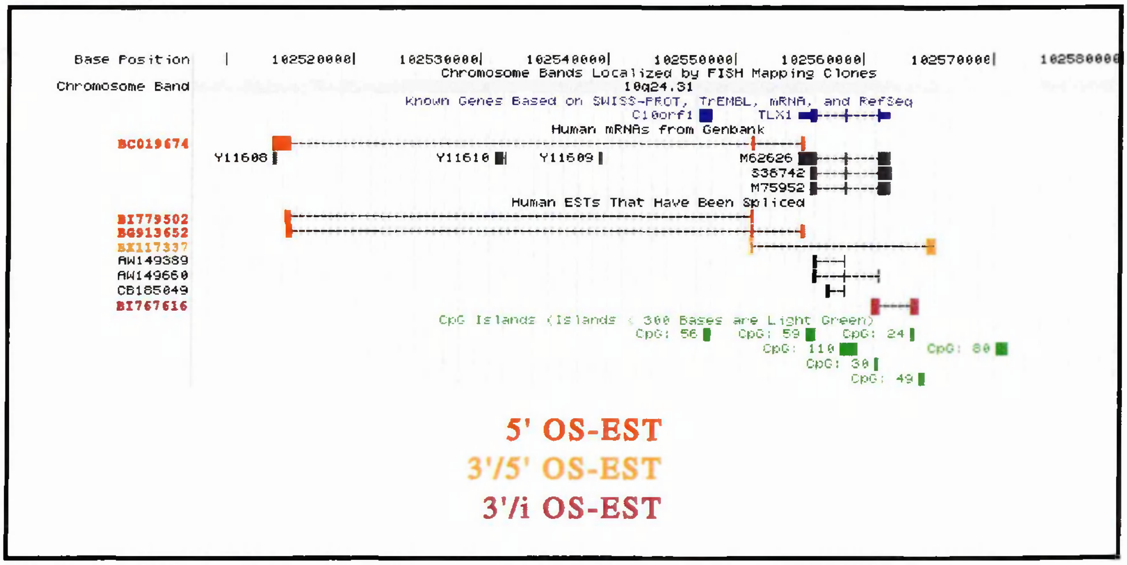


Fig. 2. 27: OS transcripts in the human TLX1 locus.

Different kinds of OS-ESTs characterize the TLX1 locus, as deduced from the UCSC Genome Browser. The *TLX1* gene (in blue) maps at chromosome 10q24.31. Red ESTs: OS spliced transcripts that map completely in the 5' region of the *TLX1* gene. Orange EST: OS spliced transcript having the first exon in the 3' region and the second exon in the 5' region of *TLX1*. Brown EST: OS spliced transcript having the first exon in the 3' region and the second exon in the intronic region of *TLX1*. Black ESTs: sense transcripts. Arrows indicate transcriptional orientations of EST clones. CpG islands are shown in green. Base positions of the *TLX1* locus refer to the UCSC Genome Contig.

to the *Dispersed Superclass.*

We decided to further investigate the genomic organization of those 19 pairs of orthologous loci (Fig. 2.26) presenting OS-ESTs both in human and mouse. These loci can be extremely useful in order to define common features of the OS transcripts. The mapping of OS-ESTs and homeobox genes revealed that 16 pairs of human/mouse loci present OS transcripts in similar positions; we will use the term "related" for these OS-ESTs. For example, related OS-ESTs have been mapped at the 5' regions of both human and mouse *OTX2/Otx2* loci. Using BLAST (http://www.ncbi.nlm.nih.gov/blast/) we then looked for sequence similarities between related OS-ESTs. Surprisingly, no significant similarities were found between OS-ESTs that map in similar position in orthologous loci. Even if the sequences analysed could represent partial portions of the full-length OS transcripts, this scenario strongly suggests that OS-ESTs correspond to non-coding transcripts. A virtual translation of related OS-ESTs confirmed the absence of significative ORFs.

Recently, the presence of OS-ESTs in a *dispersed* locus has been described *in vivo.* Indeed OS transcripts have been identified in the 5' upstream region of *Emx2/EMX2* loci (Noonan F.C., 2003). The mouse transcript, named *Emx2os* for "*Emx2 opposite strand*", presents four alternative transcripts that overlap with their 5' ends the *Emx2* 5'UTR by 271 nucleotides. Also the human *EMX2OS* presents at least two alternative spliced forms, *EMX2OS_v1* and *EMX2OS_v2.* Only *EMX2OS_v1* overlaps with the 5'UTR end of *EMX2*, while the 5' end of *EMX2OS_v2* is localized in the first intron of the *EMX2* gene. As for many other related OS-ESTs, *EMX2OS* and *Emx2os* show no detectable homology other than those regions that overlap the *EMX2/Emx2* transcripts. Virtually translated, both these transcripts present several ORFs but none of the predicted peptides are conserved between the two species nor they share similarity with proteins in the databases. The overall picture described *in vivo* for the *EMX2/Emx2* loci resembles the transcriptional map deduced by our *in silico* approach for many other dispersed loci. In particular, the localization of divergent transcripts in the 5' regions of different dispersed homeotic loci, underlines the possibility that these loci are characterized by the presence of bi-directional transcripts. A recent study revealed that a common feature of bi-directional genes is the presence of a CpG island that overlaps with the first exon of both genes (Adachi N., 2002). Interestingly, the analysis of both *EMX2/Emx2* loci revealed the presence of such a CpG island (Table 2.21 and Table 2.22) and the *in vivo* mapping of the 5' end of both sense and OS transcripts showed that some OS-ESTs could originate from this common region. Additionally, using the program "First Exon Finder" (http://rulai.cshl.org/tools/FirstEF/) I predicted the presence of two overlapping and divergent promoters in this common CpG island driving, respectively, OS and sense transcription. I decided then to look for the presence of such "common" CpG islands in the dispersed homeobox genes. In particular I focus my attention at those loci where the first exon of the OS-EST map in close proximity (less than 3 kb) with the first exon of the homeobox gene. Even if we cannot exclude the presence of others 5' exons in the OS transcripts, this approach revealed that 11 (*HLXB9*, *ISL1*, *LBX1*, *PAX3*, *TCF2/HNF1B*, *TCF8*, *EMX2*, *LIM1/LHX1*, *LMX1B*, *TITF1*, *POU3F3*) human and 9 (*Emx2*, *Lim1/Lhx1*, *Lmx1b*, *Pou3f3*, *Evx1*, *Gbx2*, *Hmx1*, *Tlx3*, *Lhx5*) mouse dispersed loci present a CpG island overlapping the putative first exon of both OS and homeobox gene transcripts (Table 2.21 and Table 2.22). In many of these regions "First Exon Finder" predicted the existence of divergent promoters, suggesting that some dispersed homeobox loci could be characterized by the presence of bi-directional, CpG-related, promoters. On the other hand, the mapping of OS-ESTs in intronic and 3' regions of the homeobox genes implies that other OS transcripts could also originate from different CpG islands and promoters.

Chapter III: DISCUSSION

Large blocks of Repetitive Elements (REs) could contribute to regulation of *Hox* loci by promoting high order chromatin domains.

According to the data provided by the human genome project, only 2–5% of our genome encodes for exons, with the remainder corresponding predominantly to intergenic regions and introns. Interspersed REs occupy up to 50% (or even more) of our genome while the other half is made of "unique" DNA. Since their identification, REs functions were the subject of a heated and controversial debate. Many researchers consider these elements as unnecessary and compare them to parasites, selfish DNAs exploiting eukaryotic genomes (Hickey D.A., 1982), while, many other biologists look at them as genomic symbionts that could interact with the surrounding genomic environments thus increasing the host genome plasticity and its capacity to evolve (for a review: Deininger P.L., 2002).

REs could have had a main role in genome evolution by promoting chromosome reshuffling through unequal homologous recombination. Due to their sequence similarity, REs enable pairing and exchange between unrelated fragments of chromatin, leading to deletions, duplications and inversion (Hughes J., 2001). Additionally, as different type of transposons carry promoter and enhancer motifs, they could even influence directly host "native" genes by interfering with or modifying their expression profiles (Brosius J., 1999a). An updated list of vertebrate regulatory elements and genes (or single exons) generated by retrosequences can be found at http://www-ifi.uni-muenster.de/exapted-retrogenes/tables.html (Brosius J., 1999b).

REs have been also identified as the genetic elements responsible for the onset of many genetic diseases. Many studies suggest that repetitive DNAs contribute to diseases through insertional mutagenesis associated with retrotransposition or through recombination processes involving unequal crossing-overs (Kazazian H.H. Jr, 1998; Kidwell M.G., 2001). Consequently, it has been proposed that organisms should have developed a "genome defense" mechanism in order to counteract the spreading of transposons and to suppress recombination between REs (Henikoff S., 2000). As the silencing of many duplicated or repetitive sequences, as well as of many viruses and transposons, was linked to DNA methylation, it was proposed that the host-defense system would involve heterochromatin formation at these loci (reviewed by Yoder J.A., 1997; Henikoff S., 2000). Since then, further evidence suggested that REs could be embedded in a chromatin structure characterized by those epigenetic modifications typical of the silent heterochromatin (Kondo Y., 2003). Moreover, in a recent paper (Schramke V., 2003) transposons have been shown to directly drive the heterochromatinization process of a yeast locus and by that gene silencing *in cis*. These findings suggest that during host evolution the heterochromatin assembling at REs might have evolved also as an epigenetic mechanism controlling gene expression.

We then decided to include in our investigation the distribution of REs in the mouse and human *Hox* loci, considering them as potential "epigenetic elements" playing a role in *Hox* gene regulation. Interestingly, our bioinformatic analysis has revealed a distinctive distribution of REs in the *Hox* loci, namely the presence of large blocks (up to 100 kb) of SINEs, LINEs and other types of repeats in the regions surrounding the clusters, while "the core" of mouse and human clusters (with some exceptions discussed below) are devoid of REs. In agreement with a previous model designed to explain both "spatial" and "temporal" colinearity of mouse *Hox* genes (Kondo T., 1998; Kondo T., 1999), we hypothesize that these blocks of REs flanking the *Hox* clusters might represent nucleation centers for the assembly of a heterochromatin-like structure that, by spreading *in cis*, is responsible for *Hox* clusters silencing. This model, called "higher-order regulatory mechanism" (Fig. 3.1), is based on many transgenic analyses (Kondo T., 1998; Kondo T., 1999) and suggests a multi-step mechanism for the activation of *Hox* genes.

The model foresees that early in development, the entire cluster would be in a transcriptionally inactive configuration and all *Hox* genes would be repressed as the result of a closed chromatin conformation. Starting from the 3' end of each cluster, *Hox* genes and their *cis*-acting elements are sequentially released from this "heterochromatin-like" environment, becoming accessible to transcriptional activating factors (Kondo T., 1999; Kmita M., 2000). Based on their unique distribution at the *Hox* loci, we suggest that the REs could play a fundamental role in *Hox* gene silencing by their ability to nucleate heterochromatin structures.

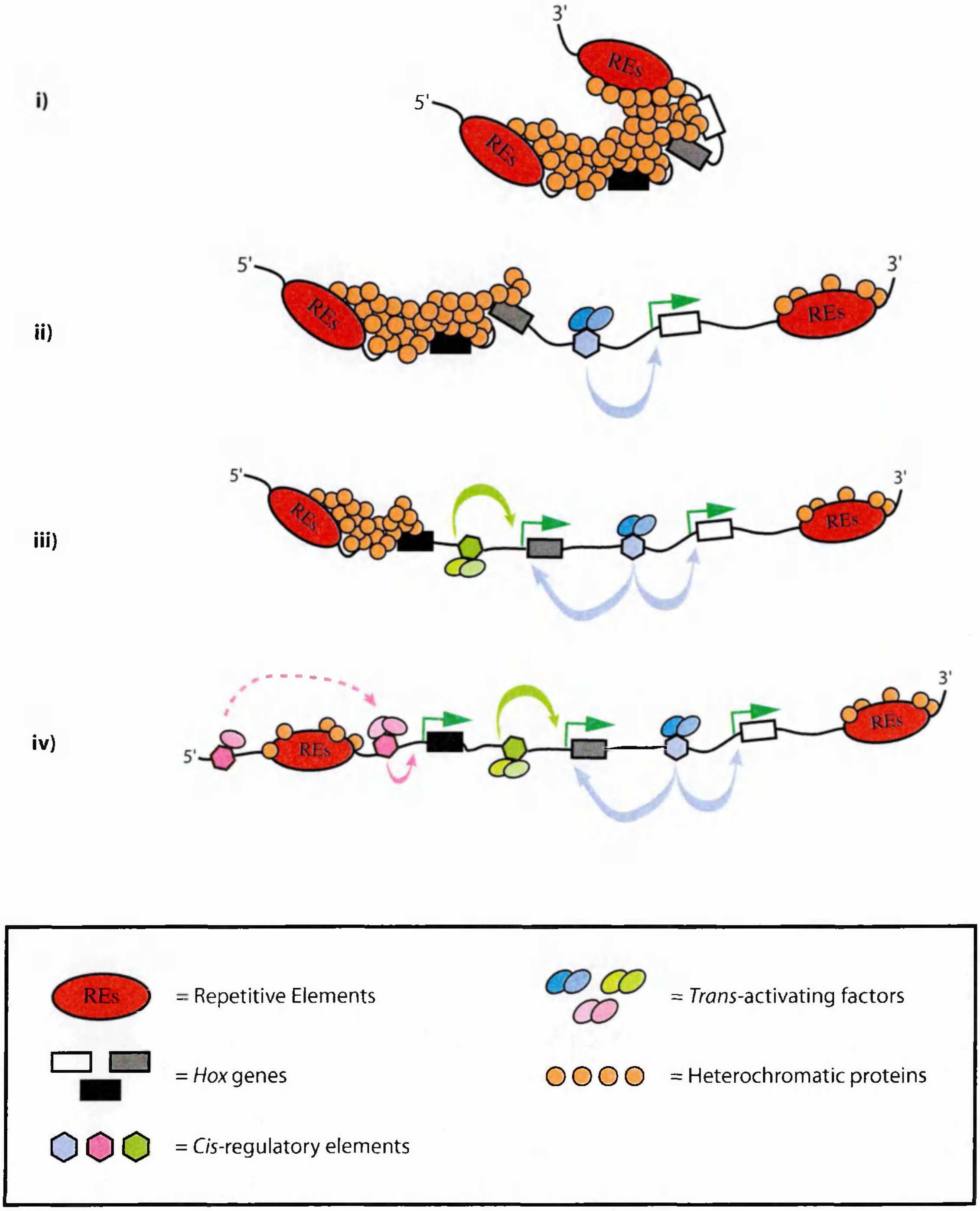

i)
3'
5'
REs
REs
ii)
5'
3'
REs
REs
iii)
5'
3'
REs
REs
iv)
3'
5'
REs
REs
REs = Repetitive Elements
= *Trans*-activating factors
= *Hox* genes
= Heterochromatic proteins
= *Cis*-regulatory elements

Fig. 3.1: *Hox* genes colinearity: the "high order" chromatin model.

The molecular mechanism driving collinear activation of three *Hox* genes (white, gray and black rectangles) derive from the gradual accessibility of genes to the transcription machinery. Early in development, the entire cluster is embedded in a heterochromatin-like structure (i) and *Hox* genes are silenced. Starting from the 3' end, a progressive remodeling of the chromatin fiber (ii-iv) allows the interaction between *trans*-activating factors (colored ovals) and *cis*-regulatory elements (colored hexagons). These elements could be both gene-specific (green hexagon) or shared among different *Hox* genes (blue hexagon). Trans-activating complexes then could recruit the transcriptional machinery promoting the proper gene expression (green arrows).

Moreover, when this "wave" has passed through the entire cluster, also DNA elements located outside (i.e.: global enhancer; pink hexagonal) could play a role in the activation of *Hox* genes expression, as for the vertebrate limb development (Kmita M., 2002).

Polycomb and trithorax proteins associate with REs at the HOXA locus.

As discussed in the first chapter, Polycomb and trithorax protein families regulate epigenetically *Hox* genes expression. In *D.melanogaster*, in order to exert this function, PcG and trxG proteins bind both particular *cis* regulatory elements (the so called PREs/TREs/CMMs), as well as core promoters (Breiling A., 2001). Although a recent bioinformatic approach has shown that in flies these binding sites share similar sequence motifs (Ringrose L., 2003), in mammals no PREs have been identified yet. The only characterized binding sites for human and mouse PcG/trxG proteins were the *Hoxc8* (Milne T., 2002) and *HOXA9* (Nakamura T., 2002) promoters, in addition to pericentromeric heterochromatin (Saurin J., 1998).

In agreement with the hypothesized role of REs as epigenetic elements involved in *Hox* gene silencing, X-ChIP experiments revealed that PcG and trxG proteins bind to regions surrounding the *HOXA* clusters and that contain repetitive DNAs (Fig 2.10). The comparative analyses of the sequences of X-ChIP/PCR amplified fragments revealed that these segments contain REs belonging to different families (as summarized in Table 2.7), thus we exclude a sequence-specific binding of PcG and TrxG proteins to REs. Vice versa, we suggest that this association could reflect the organization of a heterochromatin structure in correspondence of REs and its subsequent spread *in cis*. In agreement with this hypothesis, at both ends of the *HOXA* cluster the distribution of PcG/trxG proteins co-map with silenced *HOXA* genes. Indeed, PcG/trxG proteins have been localized both downstream *HOXA1* and in the intergenic region between *HOXA3-A4*, thus surrounding the *HOXA1-A4* genes that are all silenced in HeLa cells. A similar situation was found also in correspondence of the repressed *HOXA13* gene.

PcG/trxG proteins could map at both active and repressed regions.

We have presented the Polycomb and the trithorax groups as antagonistic protein families that regulate the expression of *Hox* genes in an opposite way. The Pc and trx families are responsible for the maintenance of a repressed and activated state of *Hox* gene expression, respectively. Thus, our X-ChIP mapping of the PcG and trxG proteins revealed an apparent contradiction: the MLL1 protein (a trxG member) and HPC2 and

BMI1 proteins (PcG members) have been co-mapped both in regions containing REs as well as active HOXA5-A6 genes (see Fig 2.10 and Fig. 2.6, respectively).

The simultaneous presence of PcG and trxG proteins at both silenced and active regions is a well-documented phenomenon (reviewed by Orlando V., 2003). Like our findings (Fig 2.6), in *D.melanogaster* Pc repressor has been mapped at active promoters (Breiling A., 2001) and PcG proteins were found at transgenes even in the presence of trxG-mediated transcription (Cavalli G., 1999). Although no direct proofs have been collected to explain this associations, a recent study seems to give an explanation for the presence of trxG proteins at silenced loci. Indeed it has been shown that MLL1 could interact directly with both HPC2 and BMI1 proteins (Xia Z.B., 2003). Pull-down and co-immunoprecipitation assays clearly demonstrated a direct binding of HPC2 and BMI1 protein to a specific portion of the MLL1 repression domain. Using a reporter gene system, this domain has been shown to possess repression activity (Zeleznik-Le N., 1994) due to its interaction with HDAC1 and HDAC2 (Xia Z.B., 2003). In agreement with these biochemical interactions, our X-ChIP/PCR co-mapping of MLL1/HPC2/BMI1 suggest that these protein could be part of a multiprotein complex linked with gene repression more that with gene activation. Nevertheless, the analysis of double *Mll*$^{-/-}$ *Bmi1*$^{-/-}$ knockouts revealed that these proteins regulate in an opposed way the expressions of many *Hox* genes. This suggests also the presence of MLL1 and BMI1 proteins in separated multiprotein complexes with distinct activities on gene expression.

RE might insulate *Hox* genes from the surrounding environment.

Our data show that practically no REs are found in the inner part of all eight *Hox* gene clusters. Remarkably, the only “exceptions” map in the vicinity of *Hox* genes that seem to be not coordinated with the rest of the cluster, in particular in the developing brain (Keynes R., 1994). This situation involves *HOXB1*, *HOXB13*, *HOXD1* that are surrounded by LINEs and SINEs, both in human and mouse. *Hoxd1* is the only *Hox* gene not expressed in the central nervous system. *Hoxb1* is expressed more posteriorly than the paralogous group 2 genes (Keynes R., 1994). Another example is *Hoxb13*. This gene is situated more than 70 kb away from the other members of the *HoxB* cluster. As shown in Fig. 2.15, this spacing region is almost made up of repetitive DNA. Despite the far location, *Hoxb13* expression seems to follow the colinearity pattern (Zeltser L., 1996). Nevertheless, in mice where all the other members of the *HoxB* cluster were deleted, no differences in the *Hoxb13* expression pattern were found compared to the wild type situation (Medina-Mendez O., 2000). These data suggest that the regulation of *Hoxb13* expression is independent from the rest of the cluster and due to *cis* elements not included in the deleted *Hoxb1-Hoxb9* region.

We propose that REs may act as insulators or boundary elements separating specific *Hox* genes from the other transcription units in the cluster, perhaps restricting the activity of *cis*-regulatory elements. This idea is reinforced by the recent identification in human (no mouse orthologs have been identified) of *PRAC* and *PRAC2* genes (Liu X.F., 2001; Olsson P., 2003). These genes code for small nuclear proteins specifically expressed in prostate, colon and rectum (Liu X.F., 2001; Olsson P., 2003). As shown in Fig. 2.11, *PRAC* and *PRAC2* are localized upstream the large block of REs separating *HOXB13* from the other members of the cluster. These are the only examples of non-homeobox gene located between two *Hox* genes in human clusters. Similarly to *PRAC* and *PRAC2*, also *HOXB13/Hoxb13* are expressed in the prostate, colon and rectum (Liu X.F., 2001; Sreenath T., 1999). Both the location of the *PRAC* genes and *HOXB13* upstream to the REs and their expression profiles, suggest that these genes might share some specific *cis*-regulatory elements. A stronger case can be made for the *Evx1* and *Evx2* genes, linked respectively to the *HoxA* and *HoxD* clusters. As shown in Fig. 2.2 and Fig. 2.13, several REs were found between *HOXA13* and *EVX1* but none between *HOXD13* and *EVX2*. The analysis of these genes has shown that, while *Evx2* follows the colinearity of the *HoxD* cluster, *Evx1* is expressed independently from *HoxA* genes (Herault Y., 1996; Dolle P., 1994).

Absence of REs: "spatial", "quantitative" colinearity and regulation of *Hox* genes.

The bioinformatic analysis on the *Alus* subfamilies distribution in the *HOX* clusters let us suppose that absence of REs is probably the result of a negative selection playing against REs insertion in these loci. As discussed in the Chapter I phenomena such as

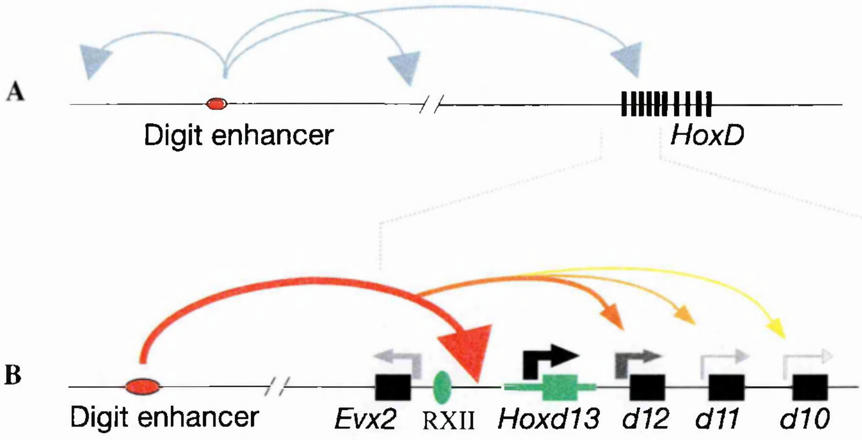


Fig 3.2: A mechanism underlying collinearity in mouse limbs.

A) Red oval represents the digit enhancer able to drive (blue arrows) the expression of different transcription units including some HoxD genes.
B) Once a contact has been established with the HoxD cluster, a sequence-specific mechanism, driven by other regulatory elements (RXII green oval) and sequences of the Hoxd13 locus (green bar), targets most of the enhancer activity to the 5' extremity of the complex (thick red arrow), leading to a preferential activation of the most 5' gene (Hoxd13; black arrow). Accordingly, the enhancer becomes gradually less efficient (orange to yellow arrows) in controlling promoters located further apart in the 3' direction (grey arrows), thus accounting for quantitative collinearity. Figure adapted from Kmita M., 2002.

"spatial" and "quantitative" colinearity rely on the precise expression, both in time and space, of *Hox* genes during embryo development and adulthood. Indeed, many homeotic phenotypes could be due just to the misexpression of one of these master control genes (i.e. single mutants of the *HoxA* and *HoxD* PG9-13 genes presents many different limb and digit malformations; Zakany J., 1999). Although the mechanisms involved in these coordinated expressions are still mysterious, the clustered organization of *Hox* loci may play a main rule in explaining colinearity phenomena. Many transgenic approaches in mouse have clearly shown that adjacent *Hox* genes can share and/or compete for *cis-regulatory* elements (Duboule D., 1998). In limbs, for example, the expression of the *Hoxd13-d9* genes is under the control of a "digit enhancer", located more than 200 kb upstream the *Hoxd13* gene (Kmita M., 2002). Systematic deletion and duplication in the *Hoxd* genes have clearly shown that the activity of the digit enhancer is mostly related to the position of the target *Hoxd* promoter than on their identity. The presence of SINEs and LINEs could have had deleterious effects in many aspects of this *Hox* gene regulation. First, the presence of extra enhancers or promoters (i.e. those carried by REs) could have interfered with "digit enhancer-*Hox* promoter" associations. Additionally, the development of these interactions can be obtained only with the co-evolution of an "uninterrupted" *HoxD* cluster (Fig. 3.2; Kmita M., 2002). The presence of REs could have also dangerous effects on the physical integrity of *Hox* clusters. RE-induced non-homologous recombinations could result in *Hox* gene duplications, inversions or deletions. The limb homeotic phenotypes shown by transgenic mice carrying these kinds of genomic alterations undoubtedly demonstrate that appropriate *Hox* gene expressions require an intact and correctly organized *Hox* cluster (Zakany J., 1999; Kmita M., 2000).

Could PcG/trxG proteins promote sub-nuclear re-localization of *Hox* genes?

Sub-nuclear re-localization events represent a strategy used by various cell types to "turn off" many different genes. This kind of transcriptional regulation foresees that the silencing of a genomic locus could be achieved through its dynamic recruitment into constitutive heterochromatic compartments of the nucleus. We have already discussed the capability of the pericentromeric regions to nucleate a repressed chromatin structure "*in cis*"; several examples have demonstrated that constitutive heterochromatin could also modify gene expression "*in trans*", by moving target loci in close spatial proximity to pericentromeric loci (Francastel C., 2000).

Many pieces of evidence suggest that the sequence-specific transcription factor Ikaros plays a fundamental role in this type of gene silencing. First, immuno-FISH approaches demonstrated that Ikaros complexes co-localized in B cell nuclei with constitutive heterochromatin, while many different silenced loci showed a selective association with Ikaros foci (Brown K.E., 1997). This kind of nuclear re-localization concerns not only single transcription units (i.e.: *CD2*, *CD4*, *CD8a*, *CD19*, *CD45*; Brown K.E., 1997), but also may involve gene clusters, as the β globin cluster (Brown K.E., 2001) or the mouse *HoxB* cluster (Brown K.E., 1999). Indeed, following mitogenic stimulation of B cell nuclei, one or both alleles of the silenced *Hoxb4-Hoxb9* region were found associated with centromeric mouse γ-satellite domains. It is not know if the *HoxB* nuclear re-localization is mediated by Ikaros complexes or by other factors. Yet, the peculiar association of PcG/trxG proteins with both human pericentromeric α-satellite repeats and REs at the *Hox* loci, as well as other characteristics (i.e. the homing phenotype) of the PcG proteins, lead us to suppose an involvement of PcG proteins in this nuclear re-local-

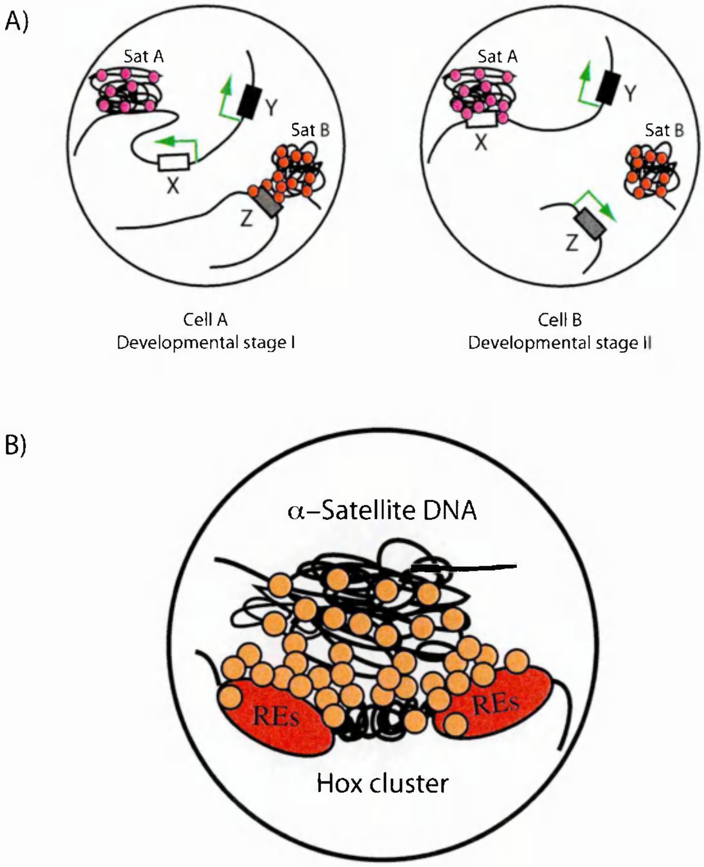


Fig. 3.3: Nuclear compartments and gene regulation.

A) Sub-nuclear re-localization of three genes (white, gray and black rectangles) is schematically shown in two hypothetical cell types/developmental stages. The association of different genes (X, Z) with pericentromeric heterochromatin (Sat A/B) induces their silencing in a cell/developmental stage –specific manner. This nuclear re-localization could be due to protein-protein interactions (red and purple dots). Green arrows represented expressed genes.

B) The sub-nuclear re-localization of *Hox* cluster could be due to the interaction of PcG (orange dots) proteins with both pericentromeric DNA repeats (a-satellite) and repetitive elements (REs, red ovals) flanking the cluster.

ization. Following the model proposed in Fig. 3.3, we speculate that the interaction between the PcG complexes, both at the *Hox* loci and at the pericentromeric regions, could drive the nuclear relocation of the target loci.

Mapping of epigenetic elements by ncRNA.

Nevertheless, the observed association of PcG/trxG proteins with REs could represent only partially the *in vivo* PcG/trxG distribution in the *HOXA* cluster. In the second chapter, we argued that in our X-ChIP/Southern blot experiments we could have failed to identify PREs/CMMs in the inner core of the *HOXA* cluster, due to the human genome complexity and to hybridization kinetics of the X-ChIP/Southern approach. Indeed the analysis of I-DNA with more sensitive techniques (Fig 2.8) revealed the presence of PcG/trxG binding sites in *HOXA* intergenic regions, that were undetected with previous X-ChIP/Southern blots. To investigate the presence of epigenetic elements also in the core portion of the *Hox* clusters, we then decided to set up two strategies.

The first one consists of combining the X-ChIP assay with a more powerful detection method based on microarray hybridizations. We generated a "locus specific genomic microarray" containing the entire human *HOXA* cluster and the flanking regions. Following an approach called "ChIP to CHIP", I-DNAs will be hybridized to this "*HOXA* specific microarray"(Fig 2.9). As shown by our first experiments with human total genomic DNA, the use of very small amounts of hybridization buffers (50-100 μl) modifies the hybridization kinetics on the microarray (as compared to a normal Southern blot), thus favoring the pairing and the detection also of single copy DNA (Fig 2.9).

The second approach we developed to identify putative epigenetic elements is based on the recent findings that these kinds of elements could be transcribed. We have already discussed evidence suggesting that transcription through centromeric satellites, LTRs and (retro)transposons could represent the first event leading to the assembling of a heterochromatin structure embedding these elements. We have also shown that non-coding transcripts could correlate with an epigenetic maintenance of the expression of homeotic genes. Experimental evidence suggested that, in the *BX-C*, many non-coding transcripts take place in correspondence with particular *cis*-regulatory elements (i.e. *iab* regions; Bae E., 2002; Drewell R., 2002) or known epigenetic elements, such as CMMs (Rank G., 2002). Moreover many other different epigenetic phenomena (i.e. X-chromosome inactivation, genomic imprinting, dosage compensation in flies) are all characterized by the presence of ncRNAs. Thus we decided to systematically map ncRNAs in the human and mouse *Hox* clusters as putative landmarks for the presence of epigenetic elements. To this end we specifically developed "AntiHunter", a bioinformatic program capable to identify antisense EST transcripts in any given genomic locus. We focused our attention on antisense RNAs for to two main reasons: i) antisense RNA molecules play a fundamental role in many different epigenetic phenomena (RNAi-based gene silencing, X-inactivation, genomic imprinting); ii) as *Hox* genes possess different alternative splice forms that are all transcribed from the same DNA strand (the "sense strand"), it is easier to detect antisense transcripts coming from the transcription of the "opposite strand". By the use of this approach we have identified many different "opposite strand" ESTs (OS-ESTs) in all four human and mouse intergenic regions of *Hox* clusters. The OS-ESTs mapping revealed two major characteristics shared by these RNA molecules: they are scattered through all human and mouse clusters and overlap both with intergenic regions and *Hox* coding exons. This latter evidence leads to the possibility that *Hox* gene regulation could involve epigenetic mechanisms based on the RNAi machinery. Although this possibility could not be excluded and should be investigated in more detail, our RT-PCRs from both adult and fetal human tissues seems not to argue in favor of this hypothesis. In RNAi-based gene silencing, the accumulation of both sense and antisense molecules takes place only in the presence of RNAi mutants, while our RT-PCRs revealed the simultaneously presence of both antisense and sense transcripts in wild type tissues. Moreover, I have identified spliced OS-ESTs that correspond to primary transcripts longer that 25 kb and that overlap more than one *Hox* sense transcript. If these transcripts were to drive the silencing of *Hox* genes by RNAi, then these long transcripts should be rapidly degraded and it should be quite unlikely to identify their spliced forms.

Further evidence suggests that the presence of OS-ESTs is not a peculiar characteristic only of the *Hox* genes, but could also be found at other loci. As described in the introduction the homeotic genes can be divided into two main superclasses: the *Complex* and the *Dispersed Superclass*. The first superclass is made of homeotic genes organized into clusters, while the second groups all the homeotic genes found dispersed through the genome. Our analysis of 95 dispersed homeotic loci revealed that 34 human and 38 mouse dispersed loci are characterized by the presence of OS-ESTs, suggesting that these ncRNAs could play a role also in the regulation of dispersed homeotic loci.

As a control, I further investigated 90 genomic loci, randomly chosen both in mouse and human, for the presence of correctly spliced OS-ESTs (Table 3.1 and Table 3.2). This bioinformatic analysis revealed that the presence of OS-EST is a common feature also of many other non-homeotic loci. Indeed 40 human and 36 mouse genomic loci present sense-antisense transcripts and among them, 23 orthologous loci present "related" OS-ESTs. Like for the homeotic loci, the OS-ESTs could map at the 5', 3' or intronic regions of the sense transcripts (Table 3.1 and Table 3.2). These bioinformatic analyses agree perfectly with recent studies (Cawley S., 2004; Yelin R., 2003) showing that large portions of human and mouse genomes are characterized by the presence of sense-antisense transcripts. Although ncRNAs at various loci could have different biological functions (Storz G., 2002), our *in silico* and *in vivo* characterization of the homeotic OS-ncRNA lead us to suppose that they could play a fundamental role in the transcriptional activation of homeotic genes.

ncRNAs could promote the expression of *Hox* genes.

Like for the intergenic transcripts mapped at *cis*-regulatory elements of the *BX-C*, we speculate that the transcription of OS-ESTs could represent one of the first steps leading to the stable expression of *Hox* genes. The experimental evidence obtained from the *D.melanogaster* proposes that the inactive state of a homeotic gene could correspond to a situation where the gene itself and its *cis*-regulatory elements (i.e. its enhancers or CMMs) are embedded in a highly packed chromatin structure. This "close" structure prevents the interaction of trans-activator proteins with the *cis*-regulatory elements and stably maintained the silencing of target genes. Thus it was supposed that intergenic transcription could represent a mechanism to "open" the chromatin structure allowing the interaction between *cis*-regulatory DNA elements and regulatory proteins. In addition, we can even suppose that transcription could increase the accessibility of these DNA elements modifying epigenetically their chromatin structure. The intergenic transcripts that could mediate this opening should possess three main characteristics: i) they should be produced in the same tissue or embryonic territories where the *cis* elements regulate the expression of the target gene; ii) they should be produced in correspondence with *cis*-regulatory elements; iii) they should be transcribed before target genes.

Although the direct involvement of OS RNAs in the epigenetic regulation of *HOX* expression requires further investigations, our analyses revealed that OS transcripts present all features characterizing the *BX-C* intergenic ncRNAs. Indeed, they are mainly co-expressed in the same adult and fetal tissues expressing also neighbor HOXA genes. Moreover, the simultaneous mapping of known *cis*-regulatory elements (as deduced by a PubMed search) and OS-ESTs, revealed that the many enhancer and RAREs map in introns of the OS-ESTs. Yet, the primary (unspliced) OS transcripts include also these elements. Nevertheless, I found three direct links between the mapping of *cis*-regulatory elements and the position of OS-EST exons. First, in the intergenic region between *Hoxa1* and *Hoxa2* a rhombomere 2-specific enhancer has been mapped, driving the expression of *Hoxa2* (Frasch M., 1995). As shown in the Fig. 3.4, this enhancer co-localizes with one exon of 7 different OS-ESTs; concordantly all the OS-ESTs derive from mouse brain cDNA libraries. I detected a similar situation also in the intergenic regions of the dispersed *Dlx1-Dlx2* and *Dlx5-Dlx6* genes. Two enhancers driving *Dlx* genes expression in different regions of the developing brain were identified and characterized in vivo in each of the two intergenic regions (Fig 3.4; Ghanem N., 2003). Mapping of the enhancers and OS-ESTs revealed that the OS transcription passes through these enhancers. Also in these cases, OS-ESTs derived from tissues expressing the four *Dlx* genes. Finally we investigated the expression of both OS-ESTs and sense transcripts in NT2 cells treated with retinoic acids (RA). This cell line has been particularly useful to study the kinetics of the *Hox* gene activation. Without any RA treatment all *HOX* genes are silenced and upon addition of RA they start to be expressed following their physical order from the 3' end the *HOX* locus, in a way that resembles the embryonic colinearity. Our RT-PCR analysis revealed that OS-EST transcription precedes the transcription of the surrounding *HOXA* genes. Moreover, also the transcription of OS-ESTs follows the colinearity rule, being 3' OS-ESTs transcribed before the 5' ones.

Taking all this evidence together, we propose a model (shown in Fig 3.5) where the "antisense transcription" might induce an opening of the chromatin structure of homeotic loci, allowing the subsequent interaction between *cis*-regulatory elements (i.e.: enhancers, RAREs, CMMs) and specific transcription factors (i.e.: segment polarity proteins, RARs). The transcriptional machinery, due to its interaction with histone modification enzymes (mainly HATs), could induce specific histone tail modifications in the regions corresponding to *cis*-regulatory (or epigenetic) elements, increasing their local accessibility to transcriptional factors (Fig 3.5). Our observations reveal that this "antisense transcription" goes from the 3' to the 5' end of the *HOXA* cluster (Fig 2.25). The identification of long OS transcripts has shown that different OS RNAs present the same 5' exon (Fig 2.23), suggesting that they may share the same transcriptional starting point. Moreover, the *in silico* identification of 5'-alternatively spliced OS-ESTs may imply the existence of independent transcriptional starting points, as shown for the human *EMX2* locus (Noonan F.C., 2003). The characterization of the full-length OS transcripts will probably help us to discriminate between these two possibilities. Although the model presents the "opening" by antisense transcription of an hypothetical *Hox* cluster, the identification of ncRNAs in dispersed loci suggests that a similar mechanism could be responsible also for the transcriptional activation of many other genes.

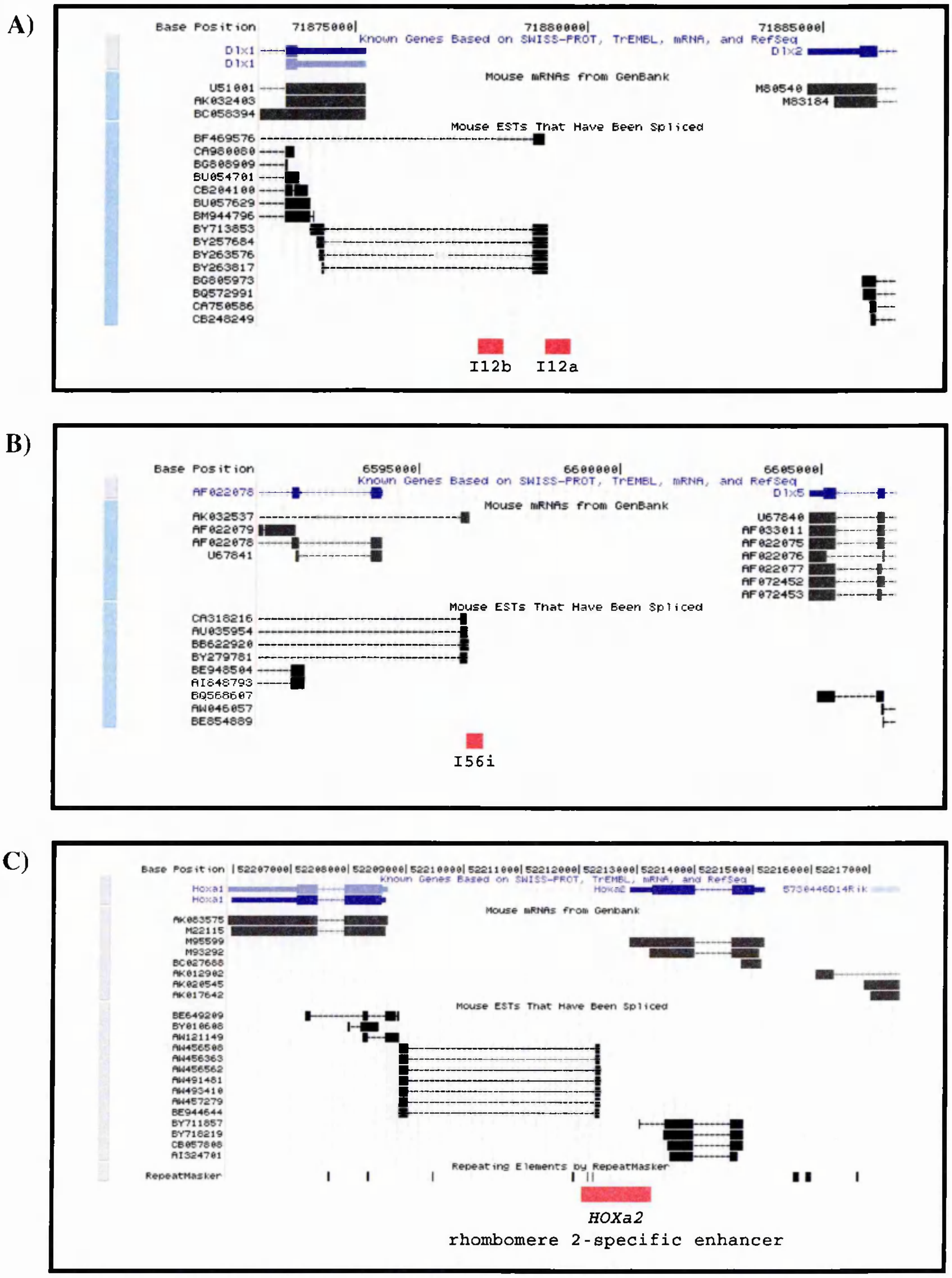


Fig. 3.4: OS transcripts colocalized with *cis*-regulatory elements.

The genomic organization of three mouse homeotic loci is represented as in the UCSC Genone Browser. Genes are in blue and EST clones are in black. Big rectangles indicate exon positions and small rectangles the 5'/3'UTR regions. Arrows indicate the transcriptional direction of genes and ESTs. Known *cis*-regulatory elements in the intergenic region between *Dlx1-Dlx2* (**A**), *Dlx5-Dlx6* (**B**) and *Hoxa1-Hoxa2* (**C**) genes, are shown schematically as red rectangles. Both primary as well as spliced OS transcripts colocalize with these regulatory elements.

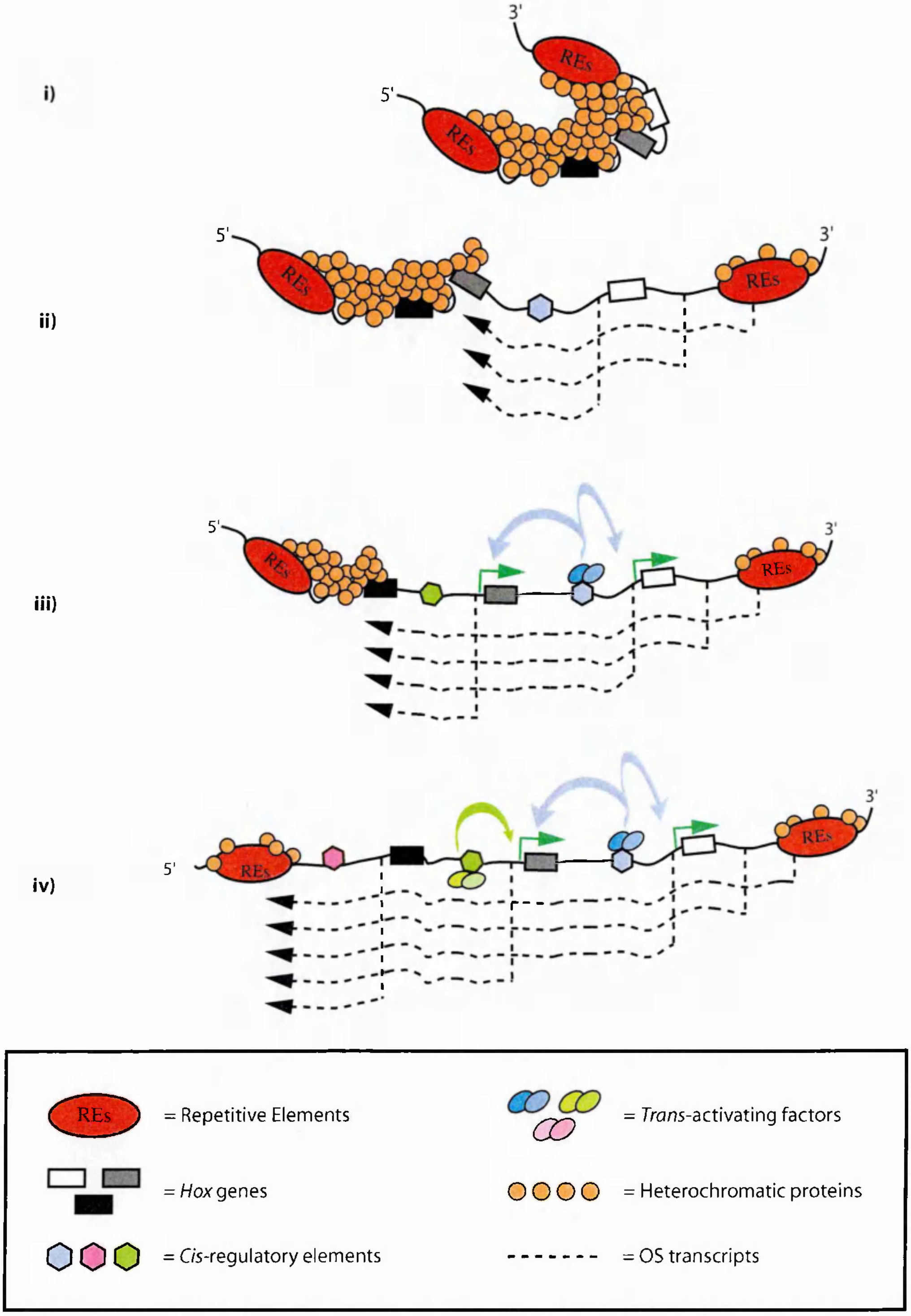


Fig. 3.5: Breaking the silence.

The collinear activation (i-iv) of three different *Hox* genes (colored rectangles) is schematically shown. Early in development the entire cluster is embedded in a hetrerochromatin-like structure (i) and *Hox* genes are silenced. Starting from the 3' end, a progressive remodeling of the chromatin fiber (ii-iv) takes place. The first step of this process is based on the antisense (or opposite strand, OS) transcription of the 3' portions of the cluster (dashed line). The OS transcripts could have different starting point both inside and/or outside the cluster (ii). As far as the OS transcription goes in the 5' portion of the cluster, *cis*-regulatory elements (colored hexagonals) and promoters are released from the heterochromatic structure and *trans*-activating factors (colored ovals) could promote the proper *Hox* genes expression (green arrows). Alternatively, the OS transcripts could mediate directly the recruitment of *trans*-activating factors on the *cis*-elements. As discussed in the text, the OS transcription could also induce an epigenetic modification of histone tail in the correspondence of *cis*-regulatory elements, promoting a stable maintenance of the *Hox* expression pattern. Only unspliced OS transcripts are shown.

Future plans.

The results discussed in this Thesis could represent the starting point for future experiments aimed at further investigation of epigenetic mechanisms controlling *HOX* gene expression.

In particular, the human "*HOXA* microarray" will be used for "ChIP on CHIP" analyses. This locus-specific genomic microarray will be hybridized with different immunoprecipitated DNAs against PcG/trxG proteins. This approach (described in Fig. 2.9) should enhance the mapping resolution of PcG/trxG proteins binding sites in the human *HOXA* cluster and complete the identification of mammalian PREs or Cellular Memory Modules (CMM). As the hybridization sensitivity of the "ChIP on CHIP" approach is clearly increased compare to the classical " ChIP/Southern blot" (Fig. 2.9), "ChIP on CHIP" experiments will be used to investigate PcG/trxG proteins distribution both in the "core" as well as in the flanking regions of the human *HOXA* cluster. This association will be analyzed in the NT2D1 cell line upon induction of *HOX* gene expression with Retinoic Acid (RA).

It will be interesting to see if intergenic transcription is accompanied by chromatin modifications. "ChIP on CHIP" experiments will be used also to investigate changes in histone tail modifications and in particular the putative deposition of histone H3.3 variant along with transcription opening of the cluster. Moreover, epigenetic marks such as H3 K9/K27 or H3 K4 methylation will be analyzed in NT2D1 cells treated with RA.

The in silico identification of OS transcripts in mouse homeotic loci belonging both to the Dispersed as well as Complex Superclasses represents the starting point for further analysis of these transcripts in vivo. The expression of OS transcripts as compared to the *Hox* sense transcripts will be investigated by RNA in situ hybridizations on mouse embryo sections. The in situ analysis will be useful to validate in the context of the developing embryo, the findings obtained in human adult and fetal tissues (Fig. 2.20; Fig. 2.21). Indeed, as for the NT2D1 cell line treated with RA (Fig. 2.25), I expect that the expression of OS transcripts will precede the expression of homeotic genes. Finally, to test the model presented in Fig. 3.5, loss of function experiments will be performed to knock-down (RNAi) OS transcripts in NT2D1 cells. This set of experiments should tell if the OS RNA has any direct role in *HOXA* gene activation.

Chapter IV:
MATERIAL AND METHODS

Reagents and standard procedures.

All basic DNA standard methods as well as bacterial transformation and cultures, media and buffer preparations were performed according to a Laboratory Manual (Maniatis T., 1982). DNA fragments and PCR amplifications were purified from agarose gel by the Qiaquick DNA purification system (Qiagen). DNAs were cloned into pBluescript or pGem-TEasy (Promega) vectors using T4 DNA Ligase from Promega. Ligations were performed according to manufacture's protocols. Ligation products were transformed in the E.*coli* DH5a or XL-1 Blue strains. Small-scale plasmid preparations (mini-preps) were made by the alkaline lysis method (Maniatis T., 1982). Large scale and midi-scale preparations (maxi and midi-preps) were done by purification with Qiagen MAXI-MIDI kits. Restriction and modification enzymes were obtained from Roche or Promega and used according to manufacture's outlines.

PCRs were performed with Promega or Qiagen PCR reaction kits, following manufacture's indications. Specific PCR conditions are summarized in different Tables (see text for details). PCR primers were designed with the OLIGO 4 program for Macintosh and synthesized by Roche, Primm or MWG Companies.

PAC, BAC, cosmid and plasmid clones.

PAC (DJ0170O19; DJ0167F23) and BAC (CTD-2536K9) clones were identified *in silico* (see Charter II), and ordered to the UK HMGP Resource Centre, Cambridge. PAC/BAC DNA preparations were carried out with Qiagen MIDI kits. PAC/BAC clones were grown overnight (ON), shaking, in 10 ml of LB medium + 25 mg/ml Kanamycin at 37°C. Next morning, 2.5 ml of bacterial clones were add to 97.5 ml of fresh LB + 25 mg/ml Kanamycin. After 1.5 hours at 37°C, IPTG was added to the final concentration of 0.5 mM; cells were collected by centrifugation after further 6-8 hours at 37°C.

Cosmid contig spanning the HOXA cluster (Kim M.H., 1998) were request to the MRC Genome Resource Facility, Dept of genetics, The Hospital for Sick Children, Toronto, Ontario. Cosmid clones were grown in 50 ml of LB medium + Kanamycin at 37°C ON. Cells were collected by centrifugation and DNA extracted with Qiagen MIDI kits.

The PAC/BAC/Cosmid contig was assembled by PCRs, restriction analyses and Southern blots using different DNA markers (Table 2.2 and Fig.2.1).

Plasmid clones were produced by sub-cloning EcoRI restriction fragments from cosmid clones into pBluescript vectors. Large EcoRI fragments were subdivided into amplifiable PCR fragments and subsequently cloned into pGem-TEasy (Promega) vectors.

Plasmid clone corresponding to three different families of pericentromeric repeats were kindly provided by A, Agresti, DIBIT Milan. These clones were digested with EcoRI (Sau3A and EcoRI families) or EcoRI + NotI (Long Sau3A family) and Southern blots were performed as described below.

Southern blots.

About 5-10 µg of total human DNA, 1-3 µg of PAC/BAC DNA, 1-2 µg of cosmid DNA were completely digested with different restriction enzymes (Promega).
Different amount of plasmid clones were digested with various restriction enzyme to get about 1µg of insert DNA. Samples were run on 0.6-1 % agarose gels in 1X TBE generally ON at 4°C. Gels were treated with 0.25 N HCl for 10' (shaking), 0.4 N NaOH for 30' (shaking). DNA was transferred to nylon membranes (GeneScreen ^{TM} & GeneScreen Plus®, Du Pont). Membranes were pre-wet with water and equilibrate in NaOH 0.4 N for 10'-15'. A capillary blot was set up using 0.4 N NaOH as transfer buffer. After transfer, the membrane was washed with 2X SSC for 1'-2' and DNA was fixed further on by baking at 80°C for 1 hour. Pre-hybridisation and hybridisation were done at 65°C in 7% SDS, 1 mM Na-EDTA, 1% BSA, 0.25M Na_2HPO_4 pH 7.2. Filters were washed at 65°C once for 10' with 5% SDS, 1mM Na-EDTA, 0.5% BSA, 20 mM Na_2HPO_4 pH 7.2; five times at 65°C for 5'each with 1% SDS, 1mM Na-EDTA, 0.5% BSA, 20 mM Na_2HPO_4 pH 7.2.

Western blots.

Protein extraction from human cell lines, electrophoresis and blotting were done according to a Laboratory Manual (Maniatis T., 1982). Rabbit affinity-purified antibodies against hPC2, BMI1, RING1, EZH2, HPH1 were kindly provided by A. Otte, Amsterdam. C. Caslini, Milan, kindly provided rabbit affinity-purified antibodies against MLL1. Affinity-purified antibodies against YY1 and TBP were ordered from Santa Cruz. Antibodies were diluted as follows: BMI-1: 1/10000; hPC-2: 1/4000; RING-1: 1/500; HPH-1: 1/1000; EZH-2: 1/1000; TBP: 1/5000, YY1: 1/2000.

RNA purification and RT-PCRs.

Total RNA was extracted from human cell lines by means of TRIzol" reagent (Life technologies), following manufacture's instructions. Cells were collected by centrifugation and TRIzol" reagent was added directly to the pellet (1-2 ml TRIzol" / 10^6 cells). Total RNA was dissolved in RNase-free water and store at –20/-80 °C. RNA quantification was done measuring the A_{260}/A_{280} ratio with a spectrophotometer. About 10 µg of total RNA was digested with 2U of DNase (Promega) at 37°C for 30'-60'. DNase was then inactivated at 70°C for 10'.

"Specific strand RT-PCRs". To define the DNA strand transcribed, we used the "One-Step" RT-PCR system by Promega. About 200 ng -1 µg of total RNA (DNase treated) was retrotranscribed with specific primers for 45'-60' at 48°C. The RT enzyme (AMV) wad inactivated by heating 2' at 94°C. The second primer was added and PCRs performed as described in Table 20.

"Oligo-dT RT-PCRs". About 3-4 mg of total RNA (DNase treated) was retrotranscribed with oligo-$(dT)_{18}$ RevertAid^{tm} H Minus First Strand cDNA Synthesis Kit (Fermentas). PCRs were performed using 200 ng of cDNA as described in Table 20.

A. Banfi and M.Cocchia (Tigem, Naples) kindly provided total RNA from human adult tissues. Total RNA from foetal tissues was request to L. Nitsch (Telethon, Naples).

Chromatin immunoprecipitation from human cell line.

In order to prepare fixed chromatin from HeLa cells, I have adapted the protocol originally used for the *D.melanogaster* (Orlando V., 1997). Schematically:

1). Grow 6 x 10^6 HeLa cells at 37°C in DMEM medium/100 U/ml penicillin/100 mg/ml streptomycin) in cell culture dishes.

2). Add fixing solution directly into the dishes and mix. The final formaldehyde concentration should be 1%. Leave the dishes at 37°C for 20'.

3). To stop HCHO fixation, add solid glycine to 125 mM and mix well. Leave at room temperature for 10'. Discard the fixing solution and wash twice the dishes with cold 1% PBS. Discard the last wash and add 2-4 ml of 1% cold PBS. Using a cell scraper, remove and collect the cells in two 50 ml Falcon tubes. Pellet cells at 1000 rpm, 5' (Hereaus Minifuge or equivalent) and wash once with cold PBS. Pellet cells at 1000 rpm, 5'.

4). Take up cell pellet into 15 ml of wash solution A and shake/rotate slowly for 10 min. at RT. Spin down cells and take up pellet into 15 ml wash solution B and shake/rotate slowly for 10 min. at RT. Spin down again and resuspend cells into 3-4 ml TEE (solution C). Transfer the solution into a 15 ml Falcon tube that has been cut off at the 11 ml mark.

5). Add ca. 1 ml glass beads, acid washed. Store on ice.

6). Sonicate each aliquot, 6x 30 sec. in 30 sec. intervals using a high power sonicator. Each tube is cooled in a beaker with an ice/water mix. Avoid foaming.

7). Transfer the suspension in a new 15 ml Falcon tube and add sodium lauryl sarcosine (Sarkosyl) to 0.5% from a 10% stock. Rotate at RT for 10 min. on a wheel.

8). Distribute the suspension into 1.5 ml microfuge tubes and spin for 5 min. at maxspeed in a tabletop centrifuge at 4°C. Unite supernatants into a 15 ml Falcon tube and add 2.84g of CsCl to each tube. Mix gently until the salt has dissolved and fill up with TEE/0.5% Sarkosyl to 5 ml (final density should be 1.42 g/cm^3). Transfer the solution into a 5 ml polyallomere Beckman tube (for Beckman SW55 rotor). Spin at 40000 rpm (Beckman L7-65 ultracentrifuge, SW55 rotor, 20 °C) for 24-48 hours.

9). Elute 10 x 500 µl fractions per gradient with a peristaltic pump or a similar device. Check the density profile of fractions with a refractometer. The peak-fraction of crosslinked chromatin should have a density of 1.39 g/cm^3. Routinely, chromatin is found at gradient fractions spanning density values between 1.350 and 1.450 g/ cm^3, transfer the fractions into dialysis bags.

10). Dialyse fractions in dialysis bags at least 2 h against 300 volumes of dialysis buffer. After 2 hours, in order to estimate the average size of the DNA and trace it along the gradient, we remove about 1/10 volume (50 µl) of each fraction to a microfuge tube. Change the buffer and continue dialysis overnight with remaining chromatin.

11). Add to the 50µl of chromatin one volume of TE and incubate tubes ON at 65∞C. Add Proteinase K to 500 mg/µl and SDS to 0.5%. Incubate at 50∞C for ca. 3h. Extract once with phenol-chloroform, once with chloroform. Precipitate with ethanol in the presence 0.3 M Na-acetate.

12). Spin 20 min, wash in 70% ethanol.

13). Take up pellets in 10 µl of TE. Add RNase A to 10 µg/ml and incubate 30 min at 37°C. Load and run the samples in 0.8% agarose gel in 1x TBE.

For chromatin-IPs, fractions that contain the cross-linked chromatin are pooled and stored in 500 µl aliquots at -80 °C or used directly for IP. For this purpose one aliquot of chromatin suspension is adjusted to RIPA buffer by sequentially adding appropriate amounts of NaCl, Triton-X100, SDS and deoxycholate and used for the preclearing.

Chromatin immunoprecipitation.

14). For each IP and the mock-control take 400 µl of chromatin and add the same volume of RIPA buffer. Add 20 µl of Protein A/G agarose beads (Santa Cruz Biotechnology). Incubate for 1-2 hour at 4°C for pre-clearing and spin for 10' in a tabletop centrifuge (14000 rpm, 4°C).

15). Transfer supernatant to a new tube and add the appropriate amount of antibody (usually dilutions of 1:100 - 500 µl). The same amount of precleared chromatin is used as negative control, without the addition of antibody (mock-control). Incubate the samples overnight at 4°C on a wheel. Spin samples 10' in a tabletop centrifuge (14000 rpm, 4°C). Transfer the IPs to new tubes. Then add 20 µl of Protein A/G agarose beads. After incubation for further 2-4 hours, pellet and wash the beads 5 times with 600 µl of RIPA buffer, 1x with 600 µl LiCl-buffer and 1x with 600 µl TE (pH 8), always pelleting the beads with short spins (15" at max speed) with a tabletop centrifuge. At the end, take up beads in 100 µl of TE.

16). Add 10 µg/ml of RNAse (DNAse-free) and incubate samples 30' at 37 °C. Adjust samples to 0,5% SDS and 0.5 mg/ml Proteinase K and incubate ON at 37°C. Next day, spin down the beads and transfer the supernatant to a new tube. Incubate for 6 hours at

65°C to complete the reversal of cross-link. Phenol-Chloroform extract the samples. Back-extract the phenol phase by adding an equal volume of TE (pH 8). Combine the aqueous phases and chloroform extract. Precipitate DNA by adding Glycogen to 100 µg/ml as carrier, 1/10 volume of 3M sodium acetate pH 5.2 and 3 volumes of 100% ethanol. Incubate at -20 °C for some hours up to ON.

X-ChIP/PCR.

17a). Spin down DNA and wash pellet in 70% ethanol, air dry briefly and resuspend the precipitated DNA in 30-60 µl of TE. I have used 1 µl of the I-DNA for each PCR reaction.

X-ChIP-Southern blot.

17b). Spin down DNA and wash pellet in 70% ethanol, air dry briefly and resuspend the precipitated DNA in 21 µl of TE.

18). A ligation reaction is set up with 7 µl of I-DNA, 1 µM of linker (Strutt H., 1997), 4U of T4 DNA Ligase (Promega) in a total volume of10 µl of ligase 1X buffer (Promega). Ligation was carried out ON at 4°C.

19). 5 µl of the ligated I-DNA was used as a template for a PCR reaction with linker-specific primers (Strutt H., 1997). PCR conditions: 94°C x 2'; 30-35 cycles: 94°C x 1', 55°C x 1', 72°C x 3'; 94°C x 1', 55°C x 1', 72°C 10'.

20) Amplified DNA was purified with Qiaquick PCR Purification System (Qiagen), quantified with a spectrophotometer; 1/10 of the amplified DNA was checked on an agarose gel.

21) About 50-100 ng of amplified DNA was labelled with Megaprime^{TM} DNA labelling systems (Amersham). Labelled DNA, at the concentration of 10^6 counts / ml of hybridisation buffer, was used as a probe in Southern hybridisation (Orlando V., 1997). Signals quantification was done with "Quantity One Quantification Software" (Bio-Rad).

22) X-ChIP solutions:

Fixing solution: 11% HCHO (from a 37% stock equilibrated with methanol), 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 50 mM Hepes pH 8.

Solution A: 10 mM Tris-HCl pH 8, 10 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.25%

Triton X100.

Solution B: 10 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.01% Triton X100.

Solution C (TEE): 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8.0, 0.5 mM EGTA, pH 8.0.

Dialysis buffer: 10mM Tris pH 8, 1mM EDTA, 0.1mM EGTA, 5% glycerol.

TE: (10mM Tris pH 8, 1mM EDTA)

RIPA Buffer: (10 mM Tris-HCl pH 8, 1 mM EDTA pH 8.0, 0.5 mM EGTA, 1% Triton X-100, 0.1% Na-Deoxycholate, 0.1% SDS, 140mM NaCl, 1mM PMSF)

LiCl-buffer: (0.25M LiCl, 0.5% NP 40, 0.5% Na-Deoxycholate, 1mM Na-EDTA, 10mM Tris-HCl, pH 8).

HOXA genomic microarray.

The entire human HOXA locus was subcloned into 137 PCR fragments (0.5-2.3 kb, Table 2.5). After amplification, the corresponding DNAs were purified, quantified and checked by proper restriction analysis.

Spotting, hybridization and quantification of signals were done by M.Cocchia at the Telethon Institute for Genetics and Medicine (Tigem, Naples).

NT2 cell line treatment with Retinoic Acid.

The human embryonic carcinoma cell line NT2/D1 was maintained in Dulbecco's modified minimal essential medium supplemented with 10% foetal bovine serum, 20 mM Hepes Buffer, 2 mM Glutamine, 200 U/ml Penicillin and 200 mg/ml Streptomycin (all from Invitrogen) in a humidified atmosphere of 5% CO2 in air. NT2/D1 cells were induced to differentiate with 10 μM all-trans Retinoic Acid (RA, Sigma). Cells were seeded at a density of 10^6 cells / 75 cm^2 tissue culture flask and re-fed every 24 h with fresh medium containing RA.

Chapter V
TABLES

TABLE 1.1: PcG protein family.

D.melanogaster	Vertebrates	Protein characteristic
additional sex combs -*ads/Asx*		glutamine-rich regions; Alanine-rich regions
cramped -*crm*	CRAMP1L	PEST sequences; Nuclear Localization Signals; Alanine-rich regions
enhancer of zeste -*E(z)*	Ezh1/EZH1; Ezh2/EZH2	SET domain; Cystein-rich regions; Histone-metyl-transferase
Enhancer of Polycomb -*E(Pc)*	Epc1/EPC1; Epc2/EPC2	Nuclear Localization Signals putative leucine zipper
extra sex comb -*esc*	eed/Eed	WD domains
pleiohomeotic -*pho*	mYY1/hYY1; Mph1	DNA binding zinc finger; acidic domains ; glycine-alanine-rich domain
Polycomb -*Pc*	M33/hPC1(CBX2); MPc2/hPC2; hPC3	chromo domain; Nuclear Localization Signals; C-terminal domain
Polycomblike -*Pcl*	M96	Cystein-rich zinc finger
polyhomeotic -*ph*	mph1(rae28)/hPH1; mph2/hPH2; hPH3	glutamine-rich regions; putative Cys2-Cys2 zinc fingers
Posterior sex comb -*psc* Suppressor of zeste 2 -*Su(z)2* Suppressor of zeste 2D -*Su(z)2D*	bmi1/BMI1; mel18/MEL18	cysteine-rich sequence; RING-finger motif; C3HC4 zinc finger
Sex comb on midleg -*Scm*	SCML1; SCML2	Nuclear Localization Signal; Alanine-rich regions; SPM domain; SAM domain
Ring	RING1/RING2	RING (zinc) finger motif
dMi-2 -*dMi2*	hZFH (hMi-2)	HMG box-like region PHD fingers chromodomains SNF2-type ATPase domain
Sex comb extra -*Sce*		
Sex comb reduced -*Scr*		
super sex comb -*sxc*		

TABLE 1.2: trxG protein family. Part I.

D.melanogaster	Vertebrates	Protein characteristic
absent, small or homeotic discs 1 and 2 *-ashl, ash2*	ASXL1, ASXL2	SET domain PHD finger
Brahma *-brm*	BRG1; hbrm	DNA-dependent ATPase; bromodomain; ATP-dependent nucleosome remodelling
eyelid *-osa*	Bright	putative DNA-binding domain
ISWI	hSNF2L	DNA-dependent ATPase domain
modifier of mdg4 *mod(mdg4)*	?	BTB domain DSD domain
moira *-mor*	BAF170; BAF155; SRG3; SWI3	SANT domain leucine zipper motif
trithorax *-trx*	MLL1 (ALL1; HRX; Htrx)	C4HC3 zinc finger motif; cys-rich zinc finger (PHD); SET domain; atypical bromodomain; DNA binding domain
trithoraxlike *-trl* GAGA		zinc finger domain POZ domain DNA-binding domain glutamine (Q) domain
zeste *-z*		DNA binding domain leucine zipper
Snf5-related1 *-snr1*	hSNF5/INI1	member of the SWI/SNF chromatin ATP-dependent remodeling complex,
Bekka *-Bka*		
grappa *-gpp*		
kismet *-kis*	Chd1; chromodomain helicase DNA binding protein 2	Chromo domain DNA binding domain, DEAD/DEAH Helicase C-terminal domain
kohtalo *-kto*		
leg arista wing complex *-lawc*		
skuld *-skd*		
sallimus *-sls*		
Suppressor ofPolycomb at 37D *-Su(Pc)37D*		
taranis *-tara*	TRIP-Br1/p34(SEI-1); TRIP-Br2/Y127; RBT1	
tonalli *-tna*		
urdur *-urd*		
verthandi *-vtd*		

TABLE 1.2: trxG protein family. Part II.

D.melanogaster	Vertebrates	Protein characteristic
little immaginal disc -*lid*	RBP2	PH domains leucine zipper Nuclear localization Signal

TABLE 2.1: ***H.sapiens HOXA*** **genes.**

Gene	mRNA Accession Number	cDNA Position (bp)	Coding (bp)
HOXA1 #1	NM_005522	75388-77190; 77656-78370	76835-77190; 77656-78306
HOXA1 #2	NM_153620	75388-77193; 77656-77750; 77954-78370	77691-77750; 77954-78292
HOXA2	NM_006735	82497-83860; 84505-85170	83121-83860; 84505-84895
HOXA3 #1	NM_030661	88585-91115; 92510-93155; 101911-101990	90310-91115; 92510-93035
HOXA3 #2	NM_153631	88585-91115; 92510-93155; 104954-105037; 109281-109415	90310-91115; 92510-93035
HOXA3 #3	NM_153632	88585-91115; 92510-92672; 96314-96392	90310-91115; 92510-92561
HOXA4	NM_002141	110786-111302; 111363-111968; 112513-113175	111620-111968; 112513-113128
HOXA5	NM_019102	123772-124480; 125441-126063	124230-124480; 125441-126002
HOXA6	NM_024014	127892-128312; 129703-130144	128053-128312; 129703-130144
HOXA7	NM_006896	137066-137616; 139072-138562	137304-137616; 138562-138940
HOXA9 #1	NM_152739	144833-146236; 147273-147925	145998-146236; 147273-147852
HOXA9 #2	NM_002142	144833-146236; 147273-147368; 147536-147925	147359-147368; 147536-147820
HOXA10 #1	NM_018951	152986-154568; 155744-156698	154294-154569; 155743-156649
HOXA10 #2	NM_153715	152986-154568; 162039-162618	154294-154568; 162039-162049
HOXA11	AF039307	165190-165442; 166830-167538	165190-165442; 166830-167538
HOXA13	NM_000522	179274-180836; 181550-182471	180592-180836; 181550-182471
EVX1	NM_001989	225196-225851; 227442-227698; 228280-228967	225425-225851; 227442-227698; 228280-228819

TABLE 2.1:
Alternative spliced forms of the human *HOXA* genes are indicated with "#" followed by a number. Nucleotide positions of all genes refer to our DNA contig.

TABLE 2.2: DNA markers.

DNA Marker	Primer Sequences (5'-3')	Amplification size (bp)	Tm (°C)
D7S2243	D7S2243.1: GCCACCCTCAAATTGCTTCC D7S2243.2: GAATAGAGCTGCTGACTTCC	391	62
SWSS3140	SWSS3140.1: ACTTTGTAAGCACACTGG SWSS3140.2: TTGAAAGAGAAACAGAGG	280	54
SWSS2757	SWSS2757.1: CTTAAAGACTGGAATCTCTG SWSS2757.2: AGAGAAGGCTGAACTTTG	326	54
D7S2834	D7S2834.1: TGCCTTAATGGGGGAAGAG D7S2834.2: ACCAGTTTCTATCCCTTACCTGC	394	63
SWSS2109	SWSS2109.1: GCGGGATAATTGATGGGCTC SWSS2109.2: TTCAAGGGGTTTTACAAG	358	55
D7S2774	D7S2774.1: AGACACCTCAGCGGCCAAC D7S2774.2: GGAACTCGCACCTGTGCT	339	60
D7S1903	D7S1903.1:GAATGGGTGGCTTCAGCTC D7S1903.2: TGACAGAAGACATTTCAG	499	55
SWSS646	SWSS646.1: CAGTGAAAGGACATGGACTG SWSS646.2: GGGAAGTAAATGCCAACTGC	267	60

TABLE 2.2

DNA markers used to assemble the *HOXA* cluster cosmid contig.

PCR conditions:

1 cycle:	94°C x 2’
30 cycles:	94°C x 1’
	tm°C x 1’
	72°C x 1’
1 cycle:	72°c x 5’

TABLE 2.3: *HOXA* cluster plasmid clones. Part I.

Plasmid Clone	Insert lenght (bp)	PAC Clone	Map position	Restriction Pattern	Vector
248.16	4800	DJ0170O19	118590-123946	NotI	pBluescript
248.101	2694	DJ0170O19	115896-118590	HincII	pBluescript
248.1	10330	DJ0170O19	105566-115896	NotI	pBluescript
248.2	5966	DJ0170O19	99600-105566	SacI	pBluescript
248.116	803	DJ0170O19	98797-99600	SacII; RsaI	pBluescript
248.62	3223	DJ0170O19	95575-98797	SacI	pBluescript
248.35	5054	DJ0170O19	90520-95575	NotI	pBluescript
248.32	7330	DJ0170O19	83190-90520	NotI	pBluescript
1589.1	1589	DJ0170O19	81601-83190	BamHI	pTEasy
230.93	2607	DJ0170O19	78992-81601	BglII	pBluescript
341.1	341	DJ0170O19	78651-78992	SacI	pBluescript
230.91	1899	DJ0170O19	76752-78651	HincII	pBluescript
230.89	640	DJ0170O19	76112-76752	SacI	pBluescript
230.2	5364	DJ0170O19	70748-76112	BamHI	pBluescript
230.81	760	DJ0170O19	69988-70748	BglII	pBluescript
2631.1	2631	DJ0170O19	67357-69988	EcoRV/BamHI	pBluescript
230.12	4325	DJ0170O19	63032-67357	NotI	pBluescript
230.9	1523	DJ0170O19	61509-63032	HincII	pBluescript
230.14	13322	DJ0170O19	48187-63032	NotI	pBluescript
64.22	5576	DJ0170O19	42611-48187	EcoRV; BamHI	pBluescript
6180.1	6180	DJ0170O19	36431-42611	NotI	pBluescript
1635.11	1635	DJ0170O19	34796-36431	HindIII;PvuII	pTEasy
P1-P2	1552	DJ0170O19	34796-	ApaI	pTEasy
P3-P4	2009	DJ0170O19		EcoRI	pTEasy
P5-P6	2098	DJ0170O19		EcoRI	pTEasy
P7-P8	1517	DJ0170O19		KpnI	pTEasy
P8rev-P19	1156	DJ0170O19		PstI	pTEasy
P20-P9rev	1172	DJ0170O19		PstI	pTEasy
P9-P10	1923	DJ0170O19		SmaI	pTEasy
P11-P12	1752	DJ0170O19		PvuII	pTEasy
P13-P14	715	DJ0170O19		EcoRI	pTEasy
P15-P16	1133	DJ0170O19		PvuII	pTEasy
P17-P18	905	DJ0170O19	-24994	EcoRI	pTEasy
1865.17	1865	DJ0170O19	23079-24944	EcoRI;PstI	pTEasy
P3/4	3172	DJ0170O19	19907-23079	EcoRI	pTEasy
P1/2	3172	DJ0170O19	19907-23079	ApaI	pTEasy
3172.30	3172	DJ0170O19	19907-23079	EcoRI;ApaI	pTEasy
3172.30	218	DJ0170O19	19689-19907	EcoRI;ApaI	pTEasy
64.34	6347	DJ0170O19	13342-19689	BamHI; HindIII	pBluescript
986.41	986	DJ0170O19	12356-13342	BamHI	pBluescript
3572.31	3527	DJ0170O19	8829-12356	BamHI; KpnI	pBluescript
4355.21	4355	DJ0170O19	4474-8829	BssHII; HpaI	pBluescript
5000.11	4474	DJ0170O19	1-4474	HpaI; NotI	pBluescript
11.4	6017	DJ0167F23	96701-102718	BamHI	pBluescript
11.25/11.26	4300	DJ0167F23	92401-96701	BamHI; PstI	pBluescript
11.31	2062	DJ0167F23	90339-92401	PstI	pBluescript
1085.11	643	DJ0167F23	89696-90339	EcoRI	pTEasy
1085.11	345	DJ0167F23	89351-89696	EcoRI	pTEasy
1085.11	97	DJ0167F23	89254-89351	EcoRI	pTEasy
11.21/11.22	4172	DJ0167F23	85082-89254	BamHI; PstI	pBluescript
83.23	1024	DJ0167F23	84058-85082	NotI	pBluescript
11.11	7488	DJ0167F23	76570-84058	BamHI	pBluescript

TABLE 2.3: *HOXA* cluster plasmid clones. Part II.

Plasmid Clone	Insert lenght (bp)	PAC Clone	Map position	Restriction Pattern	Vector
1915.6	1915	DJ0167F23	74655-76570	BamHI	pTEasy
83.57	11976	DJ0167F23	62679-74655	NotI	pBluescript
83.24;83.30	16079	DJ0167F23	46600-62679	EcoRV;BamHI	pBluescript
1254.1	1254	DJ0167F23	45346-46600	SacI	pTEasy

TABLE 2.3.

Cosmid clones were digested with EcoRI and the restriction fragments were cloned in the indicated vectors. Map position of EcoRI sites refer to PAC sequences. The EcoRI fragment 34796-24994 (PACDJ0170O19) was subdivided into amplifiable PCR fragments (P1-P18) that were cloned independently. The identity of all clones was checked by restriction analyses with the indicated enzymes.

TABLE 2.4: *HOXA5-HOXA6* intergenic region.

PCR Fragment	Oligo	Sequence (5'-3')	Tm (°C)	Amplification product (bp)
pl	1U	TTGAAGTGGAACTCCTTCTCCAGC	66	535
	1L	AAACAGGCTCCCCAACCCTGC		
p2	2U	GCCGCCTCCCGTTTCCAGCC	66	579
	2L	TGCGCAAGCTGCACATAAGTCATG		
p3	3U	CATGACTTATGTGCAGCTTGC	60	474
	3L	TGAGCGAGAATTCAGGGACTC		
p4	4U	GAGTCCCTGAATTGCTCGCTC	60	510
	4L	TTTGCCATAATGGGTGTAACC		
p5	5U	GCCCGCCGCCAGTTGCCG	60	553
	5L	GTGTATGGAATTTGACCTGCGC		
p6	6U	GGCGAGGTCAAATTCCATACAC	66	668
	6L	CCAACAGAGCCCAGTCTCTCG		
p7	7U	CGAGAGACTGGGCTCTGTTGG	66	530
	7L	ACTTTTCTCCCCGCCTGCTCC		
p8	8U	GGAGCAGGCGGGGAGAAAAG	60	546
	8L	TAGGTGGAAAACATCTCGCTTG		
p9	9U	ACCAAGCGAGATGTTTTCCACC	60	627
	9L	GAGCTTCATACACCTGTCTTG		
pl0	10U	CAAGCAAGGTGTATGAAGCTC	60	619
	1OL	GCGGGTAAGACATATCCCAAGAC		
pl1	11U	GTCTTGGGATATGTCTTACCCGC	60	511
	11L	GCCAGATGTACTAATACACAAC		
pl2	12U	GTTGTGTATTAGTACATCTGGC	60	377
	12L	AAAAGGGGAGGAGGAAGGAGG		
pl3	13U	CCTCCTTCCTCCTCCCCAAAAA	60	548
	13L	TCGGTCCTTTCTTGTTGAAACC		
pl4	14U	GGTTTCAACAAGAAGGACCG	60	668
	14L	CAGAAATCCATCCAAATCACG		

TABLE 2.4

The intergenic region between the *HOXA5-HOXA6* genes has been divided into 14 amplifiable fragments (pl to pl4).

PCR conditions:

1 cycle:	94°C x 2’
35 cycles:	94°C x 1’
	Tm°C x 1’
	72°C x 1’
1 cycle:	72°C x 5’

TABLE 2.5: *HOXA* genomic microarray. Part I.

Fragment	Plasmid Clone	Oligo Sequence (5'-3')	PCR Programme/ MgCl2(mM)	Amplification Product (bp)	Restriction Pattern	DNA Content
1	248.16	248.16lowI: ATGCTGTGTAGTCGAGGC G248.16up: ATTGCGCGGCTGGAGAAG	G/1.5	1865	SmaI	U
2	248.16	248.16 upI: TGGATGGCAACATACCAG 248.16 lowII: TGGCATTAAGGCAACATGC	F/1.5	1874	HindIII	U
3	248.16	248.16upII: TCCTGAGTGCCAAATCCTG G248.16low: ATGACACCCTGAAGAGTGG	L/1.5	1950	StuI	R
4	248.101	N248.101 low: AGTCCAGGACTTGCATCC T7: TAATACGACTCACTATAGGG	L/1	1255	SmaI	C
5	248.101	248.101 up: TGATCTTGCCCTCGCCTC T3: ATTAACCCTCACTAAAGGGA	F/1.5	1625	SmaI	U
6	248.1	10366.1U: GTCTGAGACTGATTCTTGTC 10363.1L: AGATAGATTTCATTTCCAAAGG	E/1	1216	ApaLI	R
7	248.1	10366.2U: AGATCCACGGTAGTCACAGG 10366.2L: TGCTCCTTCCACCTTATCAC	E/1	1035	PstI	R
8	248.1	10366.3U: GGTGATAAGGTGGAAAGGAGC 10366.3L: CGGGAGCCAGGTGAGGTTG	E/1	1500	BamHI	R
9	248.1	10366.4U: CAACCTCACCTGGCTCCCG 10366.4L: AGGCAGTTTTAACCTCTTAGG	E/1	1507	ApaI	R
10	248.1	10366.5U: TGGTGAGTCCTGCTC_TTTCC 10366.5L: CGCTTCCCCCTTCTGCCTTG	E/1	1655	PstI	R
11	248.1	10366.6U: CAAGGCAGAAGGGGGAAGCG 10366.6L: GGAGTGAGAGATCAAGGAGTG	E/1	1351	XhoI	U
12	248.1	10366.7U: CTCCTTACTCGGCTTTCATCTC 10366.7L: AGTGGATGGCGTTTGAGGTAG	E/1	1160	ApaI	U
13	248.1	10366.8U: CGACTTCACTCTCTTCTACCTC 10366.L: GCAGGGGCAGCGGGACTC	E/1	915	XbaI	C
14	248.2	5966.1U: ACTCCAGTCTTCTTCAATTCCC 5966.1L: GGTGGTAATTTCATACAGAGTG	E/1	1250	SphI	C
15	248.2	5966.2U: CCACCCCATTTTAATAAAGTCC 5966.2L: GAATGAAGCAACAACTCCTTAC	E/1	1307	StuI	R
16	248.2	5966.3U: GCTTCATTCCTCTTCCTTTCAG 5966.3L: ACTGTCCTTCCCTCATAACCC	E/1	1433	StuI	R
17	248.2	5966.4U: TCATAGAGAAAGTGGATTCATG 5966.4L: GGAGTTTACCTTACTATTCTTAG	E/1	1092	BglII	R
18	248.2	5966.5U: CTAAGAATAGTAAGGTAAACTC 5966.5L: GCCTGGAAATACTGATAGATGC	H/1	1132	SmaI	R
19	248.16	T3: ATTAACCCTCACTAAAGGGA T7: TAATACGACTCACTATAGGG	A/2	973	PstI	R
20A	248.62	248.62low: GAGCTTGGCAACATGGTC T7: TAATACGACTCACTATAGGG	O/1.5	1782	HindIII	R
20B	248.62	T3: ATTAACCCTCACTAAAGGGA 248.62up: GTATGAACTGAGCTCACG	O/2	1691	HindIII	R
21	248.35	5054.1U: GGGCTGTTCTTGATCTCCTTAC 5054.1L: TAGGGGTCACATAACTTAACC	E/1	1235	StuI	R
22	248.35	5054.2U: TAGAACACAAATGAAGGATGG 5054.2L: GTGGTGACAGGTTGATGACTC	E/1	1204	StuI	R
23	248.35	5054.3U: GGAACAGAAATCAGTGGCAGC 5054.3L: AAGAGGGTGGATCACGGTTTG	E/1	1449	StuI	R
24	248.35	5054.4U: AAACCGTGATCCACCCTCTTC 5054.4L: GAGAACTTTTAACCCCCTTGG	E/1	1391	StuI	R
25	248.32	248.32lowI: AATCCAAGGCTGGTATGATG T7: TAATACGACTCACTATAGGG	F/1	1971	XbaI	U
26	248.32	248.32upI: GTATGTATTAGCTATTGAGG 248.32lowII: AACCTGGAGCAGCTTCGG	F/1	1899	StuI	R
27	248.32	248.32upII: ATTGTTCGAAGTATCTAT 248.32lowIII: AACACTGGATTGCTGTGAG	F/1	2147	XhoI	U
28	248.32	248.32upIII: TGTAGATATCCAGGATCAG T3: ATTAACCCTCACTAAAGGGA	F/1	2044	XbaI	R
29	1589.1	1589.1: CCTTCTGCCCATAGAGAGTCC 1589.2: CCCACCAGCACCTCCTAACC	C/1	1662	BamHI	R
30A	230.93	T3: ATTAACCCTCACTAAAGGGA 230.93low: GACACACTAAGGACTGAGG	O/1.5	1469	SacI	R
30B	230.93	230.93up: TACCTCAGTCCTTAGTGTG T7: TAATACGACTCACTATAGGG	L/1	1329	SacI	R
31	341.1	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	F/2	689	ApaI	U
32	230.93	T3: ATTAACCCTCACTAAAGGGA T7: TAATACGACTCACTATAGGG	C/1	2065	SacI	U
33	230.89	T3: ATTAACCCTCACTAAAGGGA T7: TAATACGACTCACTATAGGG	B/1	806	SphI	C
34	230.2	N230.2lowI: TTCTTCTCCGCTCGCACC T3: ATTAACCCTCACTAAAGGGA	L/1	1814	PstI	U
35	230.2	N230.2upI: AGAAGCTGGAGCAGAGCC I230.2low: ATGACAGCCTCCGTGCTC	I/1	975	PstI	C

TABLE 2.5: *HOXA* genomic microarray. Part II.

Fragment	Plasmid Clone	Oligo Sequence (5'-3')	PCR Programme/ MgCl2(mM)	Amplification Product (bp)	Restriction Pattern	DNA Content
36	230.2	I230.2up: CACGGAGGCTGTCATAGC N230.2lowII: TACCAGGTCAGAGCAAGG	G/1	901	PstI	C
37	230.2	230.2upII: TTGCTCTGACCTGGTAAC T7: TAATACGACTCACTATAGGG	F/2	1876	BamHI	U
38	230.81	T3: ATTAACCCTCACTAAAGGGA T7: TAATACGACTCACTATAGGG	B/1	926	PstI	U
39	2631.1	I2631.1low: AATGGTCATTGATAGTGTCG SP6: GATTTAGGTGACACTATA	N/1	818	PstI	C
40	2631.1	I2631.1up: GATCGACACTATCAATGACC N2631.1low: AACCTCTCCTGAGGACTC	N/1	599	XhoI	U
41	2631.1	2631.1up: AAGAAGTTGTGAGTCCTCAG T7: TAATACGACTCACTATAGGG	C/1	1506	XbaI	C
42	230.12	230.12low: TAGGAACATAGGGCCGACTTG T3: ATTAACCCTCACTAAAGGGA	C/1	2310	HindIII	C
43	230.12	N230.12up: AAGTCGGCCGTATGTTCC I230.12low: TCTTGGCAGCGTGGAAGG	L/1.5	1091	ApaI	U
44	230.12	I230.12up: AACCACCTTCCACGCTGC T7: TAATACGACTCACTATAGGG	O/1.5	1134	PstI	C
45	230.9	T3: ATTAACCCTCACTAAAGGGA T7: TAATACGACTCACTATAGGG	C/1	1689	XhoI	U
46	230.14-1	230.14lowI: ACGGTTCATTATCCTTGAG T3: ATTAACCCTCACTAAAGGGA	C/1	1917	HindIII	C
47	230.14-1	230.14upI: GCTTTGCTCTTCAGCACTG 230.14lowII: TCCAGTTCTAAGGCCAGAG	C/1	2004	PstI	U
48	230.14-1	230.14upII: TGAATCTATCAGGTGCAATC 230.14lowIII: CAGACACGCAGACATGAAC	C/1	1984	NotI	C
49	230.14-1	230.14upIII: CAGTCCTCCTGGCTAGAC 230.14lowIV: ACACAGAGTGGACTGTTCC	I/1.5	1855	XbaI	C
50	230.14-1	230.14upIV: AACTGCCACACTCCACAG 230.14lowV: CTTCTGCCCAACTGCATG	C/1	1890	ApaI	C
51	230.14-1	230.14upV: TAGGCCTTCGCTGAGCAC 230.14lowVI: TGAGAGCTCAACTCTCGTC	N/1.5	2153	PstI	C
52	230.14-1	230.14upVI: TGAGAGCTCAACTCTCGTC 230.14lowVII: TGCCTTGTTGGTCACTGC	C/1	1925	StuI	U
53	230.14-1	230.14upVII: ACTGTGGCTTGTCCAGAG T7: TAATACGACTCACTATAGGG	C/1	873	XbaI	U
54	64.22	I64.22low1: CCACAGCGCATCTCTAGC 6180-64.22up: AGGCTCAGGATGGAAGCG	C/1	1026	NotI	C
55	64.22	I64.22up1: GCTAGAGATGCGCTGTGG N64.22lowI: ACTGGTAACATGTGACAAG	N/1	852	HindIII	U
56	64.22	64.22upI: AATTGTGGTGTGCTTGTCAC 64.22lowII: ATAGCGCGGATGTTTGTAAG	F/2	1948	StuI	U
57	64.22	N64.22upII: ACATCCGCGCTATCTGCG I64.22low2: GCAAGCGCAATGAGGCAG	C/1.5	1096	PstI	C
58	64.22	I64.22up2: TGCCTCATTGCGCTTGCC T3: ATTAACCCTCACTAAAGGGA	P/1.5	871	ApaI	C
59	6180.1	6180.1lowI: TACATGCGCTCCTGGCTG T3: ATTAACCCTCACTAAAGGGA	C/1	2164	HindIII	U
60	6180.1	6180.1upI: ACAGCGGTTCAGGTTTAATG 6180.1lowII: TACTGGAACTGCTGGTGAG	F/2	2175	ApaI	U
61	6180.1	6180.1upII: TGTGCCTGCAACAGTGTGT T7: TAATACGACTCACTATAGGG	F/2	2219	HindIII	C
62	1635.11	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	1795	HindIII	U
63	P1-P2	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	1729	PstI	C
64	P3-P4	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	2187	PstI	C
65	P5-P6	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	2276	SacI	C
66	P7-P8	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	1695	PstI	C
67	P8R-P19	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	B/1	1334	HindIII	C
68	P2O-P9R	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	B/1	1370	PstI	C
69	P9-P10	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	2101	BamHI	U
70A	P11-P12	P11: TTCCAGGCTTCCTTCTCCCC NP12: TCGCAGATCCGTGGATGC	H/2	805	SmaI	C
70B	P11-P12	NP11: CTGCATCCACGGATCTGC P12: CCCCCGAAACCTGCGACCTG	Q/1.5	966	SmaI	C
71	P13-P14	P13: GAGCGGTGAGGGTCGGGC P14: CTTCCCCCACTCCACGAC	D/1	712	SmaI	U

TABLE 2.5: *HOXA* genomic microarra. Part III.

Fragment	Plasmid Clone	Oligo Sequence (5'-3')	PCR Programme/ MgCl2(mM)	Amplification Product (bp)	Restriction Pattern	DNA Content
72	P15-P16	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	B/1	1311	XhoI	U
73	P17-P18	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	B/1	1086	StuI	U
74	1865.17	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	2025	PstI	C
75	3172.30	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	1691	PstI	C
76	64.34	N64.34lowI: TACACGCGCTACCAGACC T7: TAATACGACTCACTATAGGG	M/1.5	2148	SmaI	C
77	64.34	64.34upI: GTCTGGTAGCGCGTGTAG 64.34lowII: TTGTAGCTGCTGCTGCTG	F/1.5	2282	XhoI	C
78	64.34	64.34upII: GAGGATGCAGAGGATTGG T3: ATTAACCCTCACTAAAGGGA	F/2	2327	XhoI	C
79	986.41	986.41up: TTGGGCACTGAAGACACC 986.41low: TTCAATAACTCAGGCTGCC	L/1	1152	BamHI	C
80A	3572.31	T3: ATTAACCCTCACTAAAGGGA N3572.31low: GGTCACACTTACTTAGTGG	O/1.5	1962	HindIII	U
80B	3572.31	3572.31upI: GGCAAGCCATGGACAGAC T7: TAATACGACTCACTATAGGG	O/1.5	2049	HindIII	U
81	4355.21	4355.21low: TCTTGGATATTGTGTAAAGGC T7: TAATACGACTCACTATAGGG	C/2	2313	HindIII	C
82	4355.21	4355.21up: TCTGAACCTTCCTCAATCTC T3: ATTAACCCTCACTAAAGGGA	C/2	2315	HindIII	U
83	5000.11	5000.11low: TTGAGATGGCTACAGTCAG T3: ATTAACCCTCACTAAAGGGA	C/1	2388	PstI	C
84	5000.11	N5000.11up: AGCTAGCTGACTGTAGCC I5000.11low: GTGAGATGTGGGTATGTG	N/1	1077	XbaI	C
85	5000.11	I5000.11up: ACATACCCACATCTCACCG I5000.11low2: CACTGCCTCCTACTACGC	G/1.5	558	PstI	C
86	5000.11	I5000.11up2: GTAGTAGGAGGCAGTGGG 4355-5000low: CTGACACGCAAGGAGCAG	O/1.5	871	EcoRI	C
87	11.4	11.4-1U: GAAACTTCGTATTCCTCTGCC 11.4-1L: ACGCTGGACCTTCATCTTGAG	D/1	1531	XbaI	C
88	11.4	11.4-2U: CAAGATGAAGGTCCAGCGTCC 11.4-2L: CTAACATTCGGGAGAGAGAGG	E/1	1573	HindIII	C
89	11.4	11.4-3U: CCGAATGTTAGAATAAGGAGC 11.4-3L: ATACAATGGCAGGCTGTGAGG	E/1	1448	XbaI	C
90	11.4	11.4-4U: GAGTTTGCTAACACCCACACC 11.4-4L: GGCTTTAAGAGTTTTCATTTGC	E/1	1413	PstI	R
91	11.4	11.4-5U: GAGCAAATGAAAACTCTTAAAGC 11.4-5L: ATCATCGGTTCTTCCTACGG	E/1	1747	StuI	R
92	11.4	11.4-6U: AGCCCTGATGAAAGAAGGAAG 11.4-6L: GCTCGGTGACCATTTGCTGG	E/1	1498	PstI	U
93	11.4	11.4-7U: CTTTCTGGGAGTGGGAGATG 11.4-7L: GCTCAGACAGGTGGACTGGC	E/1	1503	BamHI	U
94	11.4	11.4-8U: CTATGACTAGGGTGGCTTGG 11.4-8L: TAGATAGATGGGGGAGGAGGG	E/1	1039	XbaI	R
95	11.25	11.25low: AGAGGGACTAAGATTCTGAGTGTG T7: TAATACGACTCACTATAGGG	F/1.5	2283	ApaI	C
96	11.25	N11.25up: ACACTCAGAATCTTAGTCC I11.25low: TCCAGAGAGACAATAGGGC	N/1	1066	BamHI	U
97	11.25	I11.25up: TCTCTCTGGAAGATGTGCC I11.25low2: TCTCGGCTGATGCTGAGG	N/1	504	PstI	U
98	11.25	I11.25up2: AGCATCAGCCGAGATGGC 11.4-1Urevlow: AGAGGAATACGAAGTTTCTC	N/1	598	SacI	C
99	11.31	T3: ATTAACCCTCACTAAAGGGA T7: TAATACGACTCACTATAGGG	C/1	2228	BglI	U
100	1085.11	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	1468	EcoRI	C
101A	11.21	11.21low1: GGGAGCACTCGCTGTATC T7: TAATACGACTCACTATAGGG	Q/1.5	1398	PstI	C
101B	11.21	11.21low: AAAGCTGCCATCCCAGGGTTC 11.21upI: TACAGCGAGTGCTCCCTG	H/2	910	PstI	C
102A	11.21	11.21low2: AGAAGCAGGCACTGGAGC 11.21up: AGAGGAGAGCAGGAAATGCAAG	G/2	1024	XhoI	U
102B	11.21	11.21up2: AATGGCTCCAGTGCCTGC T3: ATTAACCCTCACTAAAGGGA	O/1.5	1169	XhoI	U
103	83.23	T3: ATTAACCCTCACTAAAGGGA T7: TAATACGACTCACTATAGGG	A/1.5	1190	BamHI	C
104	11.11	11.11lowI: AACGGTGATCCATCACTGC T7: TAATACGACTCACTATAGGG	C/1	2031	PstI	C
105	11.11	11.11upI: ACGGTGTCAGCCAATGGC 11.11lowII: ACTTCGGTACGTGGACAAC	C/1	1755	XbaI	U

TABLE 2.5: *HOXA* genomic microarray. Part IV.

Fragment	Plasmid Clone	Oligo Sequence (5'-3')	PCR Programme/ MgCl2(mM)	Amplification Product (bp)	Restriction Pattern	DNA Content
106	11.11	11.11upII: GTGTACAATGTTGGAATCAG 11.11lowIII: TCTGGTTCAGTCACTCTAG	C/1	2062	PstI	U
107	11.11	11.11upIII: TGCTAACCTGACTTCGAAG T3: ATTAACCCTCACTAAAGGGA	C/1	2180	PstI	C
108	1915.6	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	C/1	2165	BamHI	U
109	pD	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	N/1	1591	SmaI	R
110	pE	11976.2U2: CCCTCCCTTGCCCCAAAATG 11976.2L2: TATAGTATAGATGAAACATTAGTC	N/1	842	Sau3AI	R
111	pH	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	N/1	1736	HindIII	R
112	83.57	11976.4U: GTGTTTTCTTGGTGGGTAAAG 11976.4L: ACATTTATATTTTAGAAGCCTGG	E/1	1405	HindIII	R
113	pX	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	N/2	1896	PstI	R
114	pY	11976.6U: CCCAAGTCCTTGTTCACTTATG 11976.6L: CCCCAACCCCCGAAATATCAGC	E/1.5	1492	PstI	R
115	pS	11976.7U: CGTGTTGCTGATATTTCGGGG 11976.7L: GGATCAGAAAAAGGCACCCAG	E/1	1454	XbaI	R
116	83.57	11976.8U: CTGGGTGCCTTTTTCTGATCC 11976.8L: GGGTCACTTTAGAATGAGTCC	E/1	1172	HindIII	R
117	pAB	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	N/1	1115	ScaI	U
118	pU	16079.1U: GAAAACAATCATGTTGAAATCCC 160791.L: GCTACCTCTACCAAGAAGATAC	E/1	1426	XbaI	R
119	pV	16079.2U: CACAGTATCTTCTTGGTAGAGG 16079.2L: GTCTTCTTCCATCTTCTCTTCC	E/1	1521	SphI	R
120	pZ	16079.3U: AGAAATAAATGGGAGTCAGTGG 16079.3L: ATCTGTGGCTAAGGGAAGAGC	E/1	1574	XbaI	R
121	83.24	16079.4U: CCACACTTTTCAGAGAAGAAC 16079.4L: GGGAATAGACATTTAGAGGAAC	E/1	1360	SmaI	R
122A	83.24	J-16079.5U3:CCAGTTGGAAAGATGTGAGAAAG J-16079.5L3: TGACCTAAGAATGACCCTGAGC	T/1.5	1282	SmaI	R
122B	83.24	J-16079.5L2: TTTCTCACATCTTTCCAACTGG 16079.5U: GTTCCTCTAAATGTCTATTCCC	S/1.5	561	SmaI	R
123	83.24	16079.6U: CGTTTGTATCGCTGTTTATCAG 16079.6L: TCTCATCCCTCCTGTTCATCTG	E/1	1555	BamHI	R
124	83.24	16079.7U: GATGAGAGGACTATGGTAACAG 16079.7L: TCTCATCCCTCCTGTTCATCTG	E/1	1704	ScaI	R
125	83.24	16079.8U: GATGAGAGGACTATGGTAACAG 16079.8L: CCTGTATTTGATGTGTATCTCG	E/1	1475	PstI	R
126	83.23	16079.9U: GAGATACACATCAAATACAGGT 16079.9L: TCGCTTGAACTGGAATTGAAC	E/1	997	SmaI	R
127A	P127A	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	N/1	1234	SmaI	R
127B	83.24	16079.10U: TCATTTGATAATAAAGAAGTAGC Blow: GATTCTCCTGCCTCAGCC	R/1	925	SmaI	R
128	83.24	16079.11U: GGCATACCATCTACCCCTTCC 16079.11L: ATATTGCTTGAATTAACATACTGC	E/1	880	PstI	R
129	1254.1	T7: TAATACGACTCACTATAGGG SP6: GATTTAGGTGACACTATA	B/1	1536	SacI	R

TABLE 2.5

The entire *HOXA* locus was subcloned into 137 amplifiable fragments that were spotted on a glass slide.

The sequences of each oligo pairs as well as the PCR condition, including MgCl2 concentration, are specified.

The identities of all amplified fragments were verified by restriction analyses with the indicated enzymes.

"DNA content" refers to the presence of repetitive elements (R), low complexity DNA (C) or unique DNA (U) in the corresponding amplified fragments.

Repetitive elements and low complexity DNA were identified with Repeatmasker (http://repeatmasker.genome.washington.edu/cgin/RepeatMasker).

TABLE 2.5: *HOXA* genomic microarray. Part V.

PCR programmes

A			**B**			**C**			**D**		
1x:	94°C	5’	1x:	94°C	5’	1x:	94°C	5’	1x:	94°C	5’
30x:	94°C	30”	35x:	94°C	30”	35x:	94°C	45”	35x:	94°C	45”
	50°C	30”		55°C	30”		55°C	45”		60°C	45”
	72°C	1’		72°C	1’40”		72°C	2’		72°C	2’
1x:	72°C	5’	1x:	72°C	5’	1x:	72°C	5’	1x:	72°C	5’

E			**F**			**G**			**H**		
1x:	94°C	5’	1x:	94°C	5’	1x:	94°C	5’	1x:	94°C	5’
35x:	94°C	1’	35x:	94°C	45”	35x:	94°C	1’	35x:	94°C	45”
	60°C	1’		50°C	45”		56°C	1’		58°C	45”
	72°C	2’		72°C	2’		72°C	2’		72°C	2’
1x:	72°C	5’	1x:	72°C	5’	1x:	72°C	5’	1x:	72°C	5’

I			**L**			**M**			**N**		
1x:	94°C	5’	1x:	94°C	5’	1x:	94°C	5’	1x:	94°C	5’
35x:	94°C	1’	35x:	94°C	1’	35x:	94°C	1’	35x:	94°C	1’
	58°C	1’		55°C	45”		54°C	1’		55°C	1’
	72°C	2’		72°C	2’		72°C	2’		72°C	2’
1x:	72°C	5’	1x:	72°C	5’	1x:	72°C	5’	1x:	72°C	5’

O			**P**			**Q**			**R**		
1x:	94°C	5’	1x:	94°C	5’	1x:	94°C	5’	1x:	94°C	5’
35x:	94°C	1’	35x:	94°C	45”	35x:	94°C	1’	35x:	94°C	1’
	53°C	1’		52°C	45”		57°C	1’		55°C	1’
	72°C	2’		72°C	1’40”		72°C	1’		72°C	1’
1x:	72°C	5’	1x:	72°C	5’	1x:	72°C	5’	1x:	72°C	5’

S			**T**		
1x:	94°C	2’	1x:	94°C	2’
35x:	94°C	1’	35x:	94°C	1’
	65°C	1’		60°C	1’
	72°C	1’20”		72°C	1’20”
1x:	72°C	5’	1x:	72°C	5’

TABLE 2.6: X-ChIP/PCRs. Part I.

PCR Fragment	Oligo	Sequence 5'-3'	Amplification product (bp)	Plasmid Clone
p15	1254.1U	CATTTTCCTCTTATGGCACTT	1180	1254.1
	1254.1L	TGAGTTTCAATACAGGCAACC		
p16	5054.1U	GGGCTGTCTTGATCTCCTTAC	1235	248.35
	5054.1L	TAGGGGTCACATAACTTAACC		
p17	5054.2U	TAGAACACAAATGAAGGATGG	1204	248.35
	5054.2L	GTGGTGACAGGTTGATGACTC		
p18	5054.3U	GGAACAGAAATCAGTGGCAGC	1449	248.35
	5054.3L	AAGAGGGTGGATCACGGTTTG		
p19	5054.4U	AAACCGTGATCCACCCTCTTC	1391	248.35
	5054.4L	GAGAACTTTTAACCCCCTTGG		
p20	5966.1U	ACTCCAGTCTTCTCAATTCCC	1250	248.2
	5966.1L	GGTGGTAATTTCATACAGAGTG		
p21	5966.2U	CCACCCCATTTTAATAAAGTCC	1307	248.2
	5966.2L	GAATGAAGCAACAACTCCTTAC		
p22	5966.3U	GCTTCATTCCTCTTCCTTTCAG	1433	248.2
	5966.3L	ACTGTCCTTCCCTCATAACCC		
p23	5966.4U	TCATAGAGAAAGTGGATTCATG	1092	248.2
	5966.4L	GGAGTTTACCTTACTATTCTTAG		
p24	5966.5U	CTAAGAATAGTAAGGTAAACTC	1132	248.2
	5966.5L	GCCTGGAAATACTGATAGATGC		
p25	10366.1U	GTCTGAGACTGATTCTTGTC	1216	248.1
	10366.1L	AGATAGATTTCATTTCCAAAGG		
p26	10366.2U	AGATCACACGGTAGTCACAGG	1035	248.1
	10366.2L	TGCTCCTTCCACCTTATCAC		
p27	10366.3U	GGTGATAAGGTGGAAGGAGC	1500	248.1
	10366.3L	CGGGAGCCAGGTGAGGTTG		
p28	10366.4U	CAACCTCACCTGGCTCCCG	1507	248.1
	10366.4L	AGGCAGTTTTAACCTCTTAGG		
p29	10366.5U	TGGTGAGTCCTGCTCCTTTCC	1655	248.1
	10366.5L	CGCTTCCCCCTTCTGCCTTG		
p30	10366.6U	CAAGGCAGAAGGGGGAAGCG	1351	248.1
	10366.6L	GGAGTGAGAGATCAAGGAGTG		
p31	10366.7U	CTCCTACTCGGCTTTCATCTC	1160	248.1
	10366.7L	AGTGGATGGCGTTTGAGGTAG		
p32	10366.8U	CGACTTCACTCTCTTCTACCTC	915	248.1
	10366.8L	GCAGGGGCAGCGGGACTC		
p33	16079.1U	GAAAACAATCATGTTGAAATCCC	1426	83.24
	16079.1L	GCTACCTCTACCAAGAAGATAC		
p34	16079.2U	CACAGTATCTTCTTGGTAGAGG	1521	83.24
	16079.2L	GTCTTCTTCCATCTTCTCTTCC		
p35	16079.3U	AGAAATAAATGGGAGTCAGTGG	1574	83.24
	16079.3L	ATCTGTGGCTAAGGGAAGAGC		
p36	16079.4U	CCACACTTTTCAGAGAAGAAC	1360	83.24
	16079.4L	GGGAATAGACATTTAGAGGAAC		
p37	16079.5U	GTTCCTCTAAATGTCTATTCCC	1634	83.24
	16079.5L	AAATAAGCAGAGTTGTGAGTGG		
p38	16079.6U	CGTTTGTATCGCTGTTTATCAC	1555	83.24
	16079.6L	TCTCATCCCTCCTGTTCATCTG		
p39	16079.7U	GATGAGAGGACTATGGTAACAG	1704	83.24
	16079.7L	GGATTGGTTGGTGAGGACTGC		

TABLE 2.6: X-ChIP/PCRs. Part II.

PCR Fragment	Oligo	Sequence 5'-3'	Amplification product (bp)	Plasmid Clone
p40	16079.8U	CAGTCCTCACCAACCAATCCC	1475	83.24
	16079.8L	CCTGTATTTGATGTGTATCTCG		
p41	16079.9U	GAGATACACATCAAATACAGGT	998	83.24
	16079.9L	TCGCTTGAACTGGAATTGAAC		
p42	16079.10U	TCATTTGATAATAAAGAAGTAGC	1965	83.24
	16079.10L	AATAGAGGCTGTCATAATCTG		
p43	16079.11U	GGCATACCATCTACCCCTTCC	861	83.24
	16079.11L	ATATTGCTTGAATTAACATACTGC		
p44	11976.1U	CAACTACACTTTGTATTCCTAGTC	1431	83.57
	11976.1L	ACATTTTGGGGCAAGGGAGGG		
p45	11976.2U	TCTTTAACCCTCCCTTGCCC	849	83.57
	11976.2L	TATAGTATAGATGAAACATTAGTC		
p46	11976.3U	TCTTGACTAATGTTTCATCTATAC	1558	83.57
	11976.3L	CCCTTTTCTTTACCCACCAAG		
p47	11976.4U	GTGTTTTCTTGGTGGGTAAAG	1405	83.57
	11976.4L	ACATTTATATTTTAGAAGCCTGG		
p48	11976.5L	AACAACCTAACACCAGGCTTC	1720	83.57
	11976.5U	CATAAGTGAACAAGGACTTGGG		
p49	11976.6U	CCCAAGTCCTTGTTCACTTATG	1492	83.57
	11976.6L	CCCCAACCCCCGAAATATCAGC		
p50	11976.7U	CGTGTTGCTGATATTTCGGGG	1406	83.57
	11976.7L	GGATCAGAAAAAGGCACCCAG		
p51	11976.8U	CTGGGTGCCTTTTTCTGATCC	1172	83.57
	11976.8L	GGGTCACTTTAGAATGAGTCC		
p52	11976.9U	TCATCTCTTCTTGAACTTTCGG	937	83.57
	11976.9L	TGCTGGGTTCTGAGGTTGCC		
p53	11.4-1U	GAAACTTCGTATTCCTCTGCC	1531	11.4
	11.4-1L	ACGCTGGACCTTCATCTTGAG		
p54	11.4-2U	CAAGATGAAGGTCCAGCGTCC	1573	11.4
	11.4-2L	CTAACATTCGGGAGACAGAGG		
p55	11.4-3U	CCGAATGTTAGAATAAGGAGC	1448	11.4
	11.4-3L	ATACAATGGCAGGCTGTGAGG		
p56	11.4-4U	GAGTTTGCTAACACCCACACC	1413	11.4
	11.4-4L	GGCTTTAAGAGTTTTCATTTGC		
p57	11.4-5U	GAGCAAATGAAAACTCTTAAAGC	1747	11.4
	11.4-5L	ATCATCGGTTCTTCCTACGG		
p58	11.4-6U	AGCCCTGATGAAAGAAGGAAG	1498	11.4
	11.4-6L	GCTCGGTGACCATTTGCTGG		
p59	11.4-7U	CTTTCTGGGAGTGGGAGATG	1503	11.4
	11.4-7L	GCTCAGACAGGTGGACTGGC		
p60	11.4-8U	CTATGACTAGGGTGGCTTGG	1039	11.4
	11.4-8L	TAGATAGATGGGGGAGGAGGG		

TABLE 2.6: X-ChIP/PCRs.

The EcoRI enriched fragments were divided into 60 amplifiable subfragments. Plasmid clones used as positive controls in the PCR reactions are shown.

All primer pairs work at the same PCR conditions:

1 cycle: 94°C x 2'
35 cycles: 94°C x 1'
60°C x 1'
72 °C x 1' 30"
1 cycle: 72°C x 1'

TABLE 2.7: the X-ChIP/PCR amplified fragments contain different REs.

PCR Fragment	LINE	SINE	LTR Element	DNA Element	Low Complexity	Simple repeat
p15	LINE2	AluSq				
		AluSg				
p16		AluSg			CT-rich	(TCCC)n
p20		MIR				
		MIR				
p21		AluSx	MLT1J			
			MLT1J			
p22		MIR				
		MIR				
		MIR				
p26		MIR				
		AluSx				
p28		MIR				
p31					GA-rich	
p32					AT-rich	
p36		AluJ				
		AluY				
p43		AluJb				
p47		AluSq	LTR67	MER46B		
p51					A-rich	
p53						(CA)n
						(TA)n
p56						
p57		AluSg				(TG)n

TABLE 2.7

The X-ChIP/PCR positive fragments were analyzed for the presence of REs, low complexity DNA and simple repeats with Repeatmasker. Fragments p31, p32, p51, p53 contain only low complexity DNA or simple repeats. Neither REs nor low complexity DNA and simple repeats were identified in the fragment p56.

TABLE 2.8: *H.sapiens HOXB* cluster genes.

Gene	mRNA Accession number	cDNA position (bp)	Coding (bp)
HOXB1	NM_002144	223968-224550; 225003-225426	223974-224550; 225003-225331
HOXB2	NM_002145	209697-210207; 210980-212072	209757-210207; 210980-211659
HOXB3	NM_002146	180269-180880; 199107-199197 202093-202699; 203544-204840	202252-202699; 203544-204392
HOXB4	NM_024015	176326-176854; 177696-178259	176398-176854; 177696-177994
HOXB5	NM_002147	160960-161595; 162260-163458	161034-161595; 162260-162507
HOXB6 #1	NM_156037	156261-156978; 158042-158963	156564-156978; 158042-158301
HOXB6 #2	NM_018952	149741-149851; 1500098-150233 156486-156978; 158042-158963	156564-156978; 158042-158301
HOXB6 #3	NM_156036	156261-158963	156564-156986
HOXB7	NM_004502	143694-144192; 146618-147481	143793-144192; 146618-146871
HOXB8	NM_024016	139760-140418; 141202-142365	139995-140418; 141202-141509
HOXB9	NM_024017	128355-128957; 131575-133554	128441-128957; 131575-131809
HOXB13	NM_006361	26007-26737; 27683-28224	26137-26737; 27683-27940
PRAC	NM_032391	32210-32412; 32831-33008	32338-32412;32831-32929
Pseudo rpL9		63023-63618	

TABLE 2.8.

Alternative spliced forms of the human *HOXB* genes are indicated with "#" followed by a number. Nucleotide positions of all genes refer to our DNA contig.

TABLE 2.9: *H.sapiens HOXC* cluster genes.

Gene	mRNA Accession Number	cDNA Position (bp)	Coding (bp)
HOXC13	NM_017410	76396-77246; 82601-84148	76511-77246; 82601-82860
HOXC12	NM_173860	92534-93143; 93932-94170	
HOC11	NM_014212	110749-111527; 112785-113489	110846-111527; 112785-113489
HOXC10	NM_017409	122774-123614; 126773-127883	122864-123614; 126773-127048
HOXC9	NM_006897	137697-138330; 140034- 140941	137793-138330; 140034-140278
HOXC8	NM_022658	146706-147324; 148693-148985	146889-147324; 148693-148985
HOXC6 #1	NM_004503	166014-166525; 167259-168418	166126-166525; 167259-167566
HOXC6 #2	NM_153693	154656-155018; 166180-166525 167259-168418	166372-166525; 167259-167566
HOXC5	NM_018953	170652-171180; 171882-172965	170727-171180; 171882-172096
HOXC4#1	NM_014620	154493-155018; 190767-190883 191524-191968; 192457-193186	191530-191968; 192457-192812
HOXC4 #2	NM_153633	191484-191986; 192457- 193186	191530-191968; 192457-192812

TABLE 2.9

Alternative spliced forms of the human *HOXC* genes are indicated with “#” followed by a number. Nucleotide positions of all genes refer to our DNA contig.

TABLE 2.10: ***H.sapiens HOXD* cluster genes.**

Gene	mRNA Accession Number	cDNA Position (bp)	Coding (bp)
HOXD1	NM_024501	172568-173442; 173797-174896	172792-173442; 173797-174131
HOXD3	NM_006898	148066-148157; 153020-153644 155506-157087	153104-153644; 155506-156263
HOXD4	NM_014621	135374-136066; 136597-137212	135623-136066; 136597-136931
HOXD8	NM_019558	113738-114930; 115305-115989	114356-114930; 115305-115601
HOXD9	NM_014213	106349-107574; 107923-108785	106788-107574; 107923-108164
HOXD10	NM_002148	100761-101567; 102943-103687	100823-101567: 102943-103220
HOXD11	NM_021192	91345-92125; 92896-93577	91345-92125; 92896-93131
HOXD12	NM_021193	83791-84392; 84511-84749	83791-84392; 84511-84749
HOXD13	NM_000523	76793-77660; 78469-79061	76904-77660; 78469-78719
Evx2	M93128	64096-64827; 66167-66438 67338-67765	64096-64827; 66167-66438 67338-67765

TABLE 2.10

Alternative spliced forms of the human *HOXD* genes are indicated with "#" followed by a number. Nucleotide positions of all genes refer to our DNA contig.

TABLE 2.11: “REs-distribution” in the human genome.

Chromosome	Position (bp)	Lenght (pb)	Genes	Cluster (Y/N)
1	146631104-146634561	3458	HIST2H2BE-HIST2H2AC-HIST2H2AB	Y
2	176928553-176981971	53418	HOXD cluster	Y
6	27837795-27845806	8011	HIST1H2BL-HIST1H2AI-HISTH3H-HISTH2AJ-HISTH2BM	Y
7	26840887-26947968	107081	HOXA cluster	Y
11	877937-891072	13135	IFITM3-IFITM1-IFITM2	Y
11	2113242-2141711	28469	IGF2-IGF2AS-INS	N
12	10854479-10862941	8463	TAS2R7-TAS2R8-TAS2R9	Y
12	54049322-54166110	116788	HOXC cluster	Y
17	46962099-47045796	83697	HOXB2-HOXB9	Y
17	80517099-80531381	14283	*MGC230806* -RAC3-DCXR	N

TABLE 2.11

Ten human genomic loci are characterized by the presence of less than 5 REs and at least 3 transcribed cDNA. Genes and genomic coordinates of these loci derive from the Ensembl Genome Browser (www.ensembl.org). Putative genes are in italics. Eight of these loci (Y) correspond to gene clusters.

TABLE 2.12: *M.musculus Hoxa* cluster genes.

Gene	mRNA Accession Number	Coding (bp)
Hoxa1	NM_010449	154033-154673; 155150-155505
Hoxa2	NM_010451	147595-147973; 148614-149353
Hoxa3	Y11717	139590-140118; 141497-142299
Hoxa4	S70444	120550-121060; 121550-121896
Hoxa5	NM_010453	107902-108463; 109421-109671
Hoxa6	AF247663	103628-104068; 105628-105887
Hoxa7	NM_010455	94847-95222; 96222-96535
Hoxa9	NM_010456	86148-86724; 87770-88008
Hoxa10	NM_008263	77369-78294; 79472-79746
Hoxa11	NM_010450	66533-67241; 68660-68892
Hoxa13	NM_008264	51745-52388; 53104-53348

TABLE 2.12
Nucleotide positions of all genes refer to our DNA contig.

TABLE 2.13: *M.musculus Hoxb* cluster genes.

Gene	mRNA Accession Number	Coding (bp)
Hoxb1	NM_008266	161610-162174; 162592-162922
Hoxb2	NM_002145	147494-147976; 148733-149412
Hoxb3	NM_010458	140032-140479; 141329-142182
Hoxb4	NM_010459	114552-115004; 115811-116109
Hoxb5	NM_008268	99396-99957; 100672-100919
Hoxb6	NM_008269	94959-95375; 96451-96710
Hoxb7	NM_010460	82506-82905; 85157-85412
Hoxb8	NM_010461	78741-79164; 79941-80248
Hoxb9	XM_147563	67320-67836; 70402-70637
Hoxb13	NM_008267	

TABLE 2.13

Nucleotide positions of all genes refer to our DNA contig.

TABLE 2.14: *M.musculus Hoxc* cluster genes.

Gene	mRNA Accession Number	Coding (bp)
Hoxc4	NM_013553	137540-137978; 138453-138809
Hoxc5	NM_008271	116905-117358; 118053-118265
Hoxc6	XM_181367	112675-112829; 113556-113864
Hoxc8	NM_010466	93606-94038; 95386-95678
Hoxc9	NM_008272	84476-85013; 86718-86962
HOXC10	NM_017409	69684-70434; 73623-73901
HOXC11	NM_014212	57352-58033; 59295-59527
Hoxc12	XM_111599	39679-40282; 41105-41343
Hoxc13	XM_203647	24015-24744; 29996-30252

TABLE 2.14.

Nucleotide positions of all genes refer to our DNA contig. *Hoxc10* and *Hoxc11* coding regions were mapped using the corresponding mRNA from *H.sapiens*.

TABLE 2.15: *M.musculus Hoxd* cluster genes.

Gene	mRMA Accession Number	Coding (bp)
Hoxd1	NM_010467	199978-200629; 200931-201265
Hoxd3	NM_010468	180932-181427; 183194-183897
Hoxd4	NM_010469	164151-164577; 165134-165459
HOXD8	NM_019558	142475-142989; 143385-143668
Hoxd9	NM_013555	134927-135704; 136051-136292
Hoxd10	NM_013554	128851-129595; 130962-131239
Hoxd11	NM_008273	119264-120038; 120774-121009
Hoxd12	NM_008274	111958-112554; 112685-112923
hoxd13	NM_008275	105181-105949; 106757-107007
Evx2	NM_007967	92495-93214; 94646-94926 95871-96298

TABLE 2.15.

Nucleotide positions of all genes refer to our DNA contig.
Hoxd8 coding regions were mapped using the *H.sapiens* mRNA.

TABLE 2.16: Intergenic OS-EST mapping in the human *HOXA* locus. Nucleotide positions refer to our DNA contig.

Intergenic region	Accession number	Nucleotide position	Tissue
HOXA1-HOXA2	BG325728	78628-78783; 81469-81761; 82174-82361	Renal cell adenocarcinoma
HOXA1-HOXA2	AA489505	81477-81761; 82174-82361	HeLa S3 cell line
HOXA1-HOXA2	BF593636	78744-78783; 82174-82645	Brain glioblastoma
HOXA1-HOXA2	BE857357	78773-78783; 82174-82653	Brain glioblastoma
HOXA1-HOXA2	AW469060	81726-81761; 82174-82534	Stomach adenocarcinoma
HOXA1-HOXA2	AA531291	78599-78783; 82174-82361	Prostate
HOXA3-HOXA4	BF431347	103522-103588; 104314-104568	Kidney
HOXA3-HOXA4	AW771616	103360-103588; 104314-104559	Kidney
HOXA3-HOXA4	BE857368	96352-96388; 104314-104585	Brain glioblastoma
HOXA3-HOXA4	BF114958	90172-90264; 92430-92590; 104314-104568	Kidney
HOXA3-HOXA4	BI823151	103572-103588; 104314-104421; 104862-105061; 105855-106018	Pooled brain, lung, testis
HOXA3-HOXA4	BE873349	104971-105061; 105855-106033; 110595-111083	Colon adenocarcinoma
HOXA3-HOXA4	AW449791	96331-96388; 104314-104594	Lung
HOXA3-HOXA4	AI795971	97800-97966; 104314-104568	Kidney
HOXA6-HOXA7	AI887912	129940-129981; 133970-134076; 134689-135095	Lung carcinoid
HOXA6-HOXA7	BF510786	129878-129981; 133970-134076; 134689-135098	Lung
HOXA6-HOXA7	AW205417	129897-129981; 133970-134076; 134689-135098	Lung
HOXA6-HOXA7	AW470294	134038-134076; 134689-135098	Lung carcinoid
HOXA6-HOXA7	AK091933	129558-129807; 133970-134076; 134689-136680	Kidney
HOXA9-HOXA10	BE856108	151294-151388; 151702-152096	Prostate
HOXA9-HOXA10	BE672845	151312-151388; 151702-152096	Prostate
HOXA9-HOXA10	AI768431	151014-151388; 153702-153768	Kidney
HOXA9-HOXA10	BM549446	151323-151388; 151702-152159; 153703-154158	Brain astrocytoma
HOXA11-HOXA13	BE293400	167974-168657; 169077-169193	Colon adenocarcinoma
HOXA11-HOXA13	BE293794	168135-168657; 169077-169171	Colon adenocarcinoma
HOXA11-HOXA13	AW137133	168451-168657; 169077-169251	Colon
HOXA11-HOXA13	AW769105	168437-168657; 169077-169241	Colon adenocarcinoma
HOXA11-HOXA13	BG746816	170213-170414; 170995-171614	Colon adenocarcinoma
HOXA11-HOXA13	BM788987	168589-168657; 170995-171360	SNU-16 cell line
HOXA11-HOXA13	AF071171	170424-170649; 170995-171688	

TABLE 2.17: Intergenic OS-EST mapping in the human *HOXB* locus. Nucleotide positions refer to our DNA contig.

Intergenic region	Accession number	Nucleotide position	Tissue
H0XB2-HOXB3	BE676309	203478-230702; 204324-204386; 204780-204836; 206266-206354;	B-cell, chronic lymphotic leukemia
H0XB2-HOXB3	AI312851	203489-203704; 204324-204386; 204780-204836; 206266-206313	Fetal lung
HOXB5-HOXB7	BM994156	148311-148581; 150667-150786; 164127-164161	Lung metatastic chondrosarcoma
HOXB5-HOXB7	AI685673	148311-148581; 150667-150786; 164127-164199	Prostate
HOXB5-HOXB7	AI765957	148317-148581; 150667-150786; 164127-164201	Kidney
HOXB5-HOXB7	AW242766	148318-148581; 150667-150786; 164127-164197	Kidney
HOXB5-HOXB7	BE552174	148320-148581; 150667-150786; 164127-164239	Kidney
HOXB5-HOXB7	AI743312	148321-148581; 150667-150786; 164127-164228	Pooled
HOXB5-HOXB7	AW268484	148304-148581; 150667-150786; 162352-162391	Pooled
HOXB5-HOXB7	BF431268	148319-148581; 150667-150786; 162352-162396	Kidney
HOXB5-HOXB7	AI125836	148319-148581; 150667-150786; 158137-158194	Testis
HOXB5-HOXB7	AI684852	148319-148581; 150667-150786; 152481-152575	Pooled feal lung, testis, B-cell
HOXB5-HOXB7	AI220200	148319-148581; 150667-150700	Pooled feal lung, testis, B-cell
HOXB5-HOXB7	AI335246	152376-152638; 154170-154268; 162352-162391	2 kidney pooled tumors
HOXB5-HOXB7	AI866933	152378-152638; 154170-154268; 162352-162415	Stomach adenocarcinoma
HOXB5-HOXB7	AI623194	152378-152368; 153050-153172	Pooled germ cell tumors
HOXB5-HOXB7	AW291017	152355-152638; 153050-153184	Colon
HOXB5-HOXB7	AI949711	152375-152638; 153050-153233	Kidney
HOXB5-HOXB7	AW025831	152378-152638; 153050-153142	Pooled germ cell tumors
HOXB5-HOXB7	AA936427	152378-152638; 154170-154269	Pooled
HOXB5-HOXB7	AI125255	152381-152638; 154170-154268; 204780-204836; 205077-205095	Testis
HOXB9-HOXB13	BF476543	29972-30402; 31417-31466	Prostate
HOXB9-HOXB13	BF437711	29972-30402; 31417-31472	Prostate
HOXB9-HOXB13	AI972893	29972-30402; 31417-31447	Prostate
HOXB9-HOXB13	AI400849	29972-30402; 31417-31432	Prostate
HOXB9-HOXB13	AI817776	29972-30402; 31417-31477	Prostate
HOXB9-HOXB13	AI762586	29973-30402; 30959-30984	Colon tumor

TABLE 2.18: Intergenic OS-EST mapping in the human *HOXC* locus. Nucleotide positions refer to our DNA contig.

Intergenic region	Accession numbers	Nucleotide positions	Tissue
HOXC10-HOXC11	AI830498	121555-121758; 122568-122687; 123080-123123	Kidney
HOXC9-HOXC10	BF195843	132812-133122; 133437-133513; 134274-134389	Kidney
HOXC9-HOXC10	AI935658	132812-133122; 133437-133513; 134274-134362	Kidney
HOXC5-HOXC6	AW070489	167286-167578; 168067-168225	Pooled fetal lung, testis and B-cell
HOXC5-HOXC6	AW241966	167303-167578; 168067-168365	Pooled fetal lung, testis and B-cell

TABLE 2.19: Intergenic OS-EST mapping in the human *HOXD* locus. Nucleotide positions refer to our DNA contig.

Intergenic region	Accession number	Nucleotide position	Tissue
HOXD1-HOXD3	BQ028175	160469-160681; 161321-161483; 161700-161815; 163018-163087;	Bladder carcinoma
HOXD1-HOXD3	BG324915	160137-160681; 161321-161483; 172408-172511	Renal cell adenocarcinoma
HOXD1-HOXD3	BG421760	160247-160681; 161321-161483; 172408-172511	Renal cell adenocarcinoma
HOXD1-HOXD3	AL120749	160148-160681; 161321-161483; 172408-172431	Melanoma cell line
HOXD1-HOXD3	AA405218	161395-161483; 161700-161821; 161930-162103	Testis
HOXD1-HOXD3	BE378592	160476-160681; 161321-161483; 161700-161815; 163018-163049	Endometrium, adenocarcinoma cell line
HOXD1-HOXD3	AA406052	160469-160681; 161321-161483; 161700-161776	Testis
HOXD1-HOXD3	AI672018	160469-160681; 161321-161487	Kidney
HOXD1-HOXD3	AA479660	160470-160681; 161321-161483; 161700-161808	Ovarian tumor
HOXD1-HOXD3	AW082733	160477-160681; 161321-161486	Pooled fetal lung, testis and B-cell
HOXD1-HOXD3	AI024707	160498-160681; 161700-161889	Testis
HOXD1-HOXD3	AA778683	160498-160681; 161321-161483; 161700-161930	Testis
HOXD1-HOXD3	AA974759	160499-160681; 161321-161459	Pooled fetal lung, testis and B-cell
HOXD4-HOXD9	BF432371	106349-106433; 119308-119411; 120610-120857	Brain glioblastoma
HOXD4-HOXD9	AW450846	106350-106433; 119308-119411; 120610-120871	Prostate
HOXD4-HOXD9	AI131064	106367-106433; 119308-119411; 120610-120918	Fetal heart
HOXD4-HOXD9	AI244997	106393-106433; 119308-119411; 120610-120889	Kidney
HOXD4-HOXD9	AI650322	118595-118939; 119308-119411; 120610-120661	Kidney
HOXD4-HOXD9	AI521882	118831-118939; 119308-119411; 120610-120690	Kidney
HOXD4-HOXD9	BQ015065	118833-118939; 119308-119411; 120610-121002	Left pubic bone chondrosarcoma
HOXD4-HOXD9	AI347448	118833-118939; 119308-119411; 120610-120872	2 pooled kidney tumors
HOXD4-HOXD9	AW302186	118834-118939; 119308-119411; 120610-120819	Kidney
HOXD4-HOXD9	AW269157	118834-118939; 119308-119341	Kidney
HOXD4-HOXD9	BE999977	118835-118939; 119308-119411; 120610-120857	Colon tumor
HOXD4-HOXD9	AI949937	118835-118939; 119308-119411; 120610-120855	2 pooled kidney tumors
HOXD4-HOXD9	AW070867	118835-118939; 119308-119411; 120610-120820	Pooled fetal lung, testis and B-cell
HOXD4-HOXD9	AI253251	118835-118939; 119308-119411; 120610-120813	Kidney
HOXD4-HOXD9	AW207561	118835-118939; 119308-119411; 120610-120844	Kidney
HOXD4-HOXD9	AI949770	118835-118939; 119308-119411; 120610-120760	2 pooled kidney tumors
HOXD4-HOXD9	AI139468	118835-118939; 119308-119411; 120610-120814	Fetal heart
HOXD4-HOXD9	AI393873	118836-118939; 119308-119411; 120610-120823	Pooled human melanocyte, fetal heart, uterus

TABLE 2.20: RT-PCR conditions.

Transcript	Oligo	Sequence (5'-3')	cDNA amplification size (bp)	Genomic amplification size (bp)	Tm (°C)
OS-EST BG325728	BG325728.U	CCTCCCCACCGTTCAATG	211	2897	60
	BG325728.L	TGCTCCCTACCTTCCCTC			
OS-EST AA489505	AA489505.U	AAACCAGCCATAGTCCCCACACTC	281	691	60
	AA489505.L	TTCAACCCCCTCCCCCATAAATCC			
OS-EST BI823151	BI823151.U	GAAACGCCAAGACATAGAAAAC	236	675	60
	BI823151.L	GTTGGAGACAATTCCTGGTTC			
OS-EST BE873349	BE873349.U	GGCGAAGAGTTTAGGAGCAG	240	1030	60
	BE873349.L	GCAGAACTTCTCCAAGCCAG			
OS-EST BF510786	BF510786.U	GCTTTGCAACTCGCAACC	222	816	60
	BF510786.L	CCACGCACCTATTCCCCC			
OS-EST BE305073	BE305073.U	TTCTTTCGCTTCCTCCAAAC	126	544	60
	BE305073.L	TCTTCTGGAGCTGAGGACTC			
HOXA1	HOXA1U	GCCGTACTCTCCAACTTTC	221	673	60
	HOXA1L	CTCGCCTCAATACATTCACC			
HOXA2	HOXA2U	CACTTGTCTCTCAGTCAAATCC	252	892	60
	HOXA2L	AGAAAACCGCACTTCTGCC			
HOXA3	HOXA3RA3.1	GCGATCTACGGTGGCTACC	770	2100	60
	HOXA3RA3.2	GCTGCGACTTGGAGACTGG			
HOXA4	HOXA4U	CATGCGAGCCACGTCCTG	375	920	60
	HOXA4L	AGACAAACAGAGCGTGTGGG			
HOXA5	HOXA5A	TTGAAGTGGAACTCCTTCTCCAGC	112	1072	60
	HOXA5S	TGCGCAAGCTGCACATAAGTCATG			
HOXA6	HOXA6U	TGAAGTGGAACTCCTTCTCCAG	199	1589	60
	HOXA6L	TACAAACCCGACAGCAGCAG			
HOXA7	HOXA7U	GCCAATTTCCGCATCTACC	351	1296	60
	HOXA7L	TCATTCCTCCTCGTCTTCC			
HOXA9	HOXA9A	CACTCGTCTTTTGCTCGGTC	229	229	55
	HOXA9S	CAGCCAACTGGCTTCATGCG			
HOXA10	HOXA10A	GGACGCTGCGGCTAATCTCTAGGCG	198	1318	65
	HOXA10S	AGAGCAGCAAAGCCTCGCCGGAGAAG			
HOXA11	HOXA11U	AACCAGATTTTGACTTGACGATCAGTG	444	1900	60
	HOXA11L	ATTTCTATAGCACCGTGGGCAG			
HOXA13	HOXA13A	CGTCGTGGCTGATATCCG	114	114	55
	HOXA13S	GGGAGAAAGAAGCGCGTG			

TABLE 2.20

RT reactions: see Chapter IV.
PCR conditions:

1 cycle:	94°C x 2’
35 cycles:	94°C x 1’
	Tm°C x 1’
	72°C x 1’
1 cycle:	72°C x 5’

TABLE 2.21: OS-ESTs mapping in the *H.sapiens* dispersed homeotic loci. Part I.

Gene	OS-EST	Spliced	PolyA signal	Position	CpG	Distance
DLX2	AA625909	y	n	5'	-	d
DLX3	W60102	y	n	5'	S	c
PBX1	AK124128	y	n	5'	-	e
PREP1	AA621062	y	y	5'	-	d
	AA812731	y	y	5'	-	d
	AI138243	y	y	5'	-	d
	AW269528	y	y	5'	-	d
	BX095448	y	y	5'	-	d
SIX3	BM663835	y	y	5'	-	a
	BM681783	y	y	5'	-	a
	BM690547	y	y	5'	-	a
	BM692614	y	y	5'	-	a
	BM727566	y	y	5'	-	a
	BQ188468	y	y	5'	-	a
	BU726060	y	y	5'	-	a
	BX116389	y	y	5'	-	a
TGIF	BG490809	y	y	5'	-	e
	AA469958	y	n	5'	-	e
	BX282843	y	n	5'	-	e
HLXB9	AI761375	y	n	i	S	
	BC022831	y	n	5'	C	b
	BC033117	y	n	5'	C	b
	BU506764	y	n	5'	C	b
	BG470747	y	n	5'	C	b
	BE257756	y	n	5'	C	b
MEOX2	BX538274	y	n	i	-	
	BG184437	y	n	i	-	
	AI417984	y	n	i	-	
ISL1	BE792152	y	n	5'	C	a
	AL524737	y	n	5'	C	a
	BE798355	y	n	5'	C	a
	BE798361	y	n	5'	C	a
	BX394889	y	n	5'	C	a
	BX345045	y	n	5'	C	a
	BX490971	y	n	5'	C	a
LBX1	AK123344	y	n	5'	C	b
	AK096698	y	y	5'	S	d
	BG208726	y	n	5'	-	d
	BX115274	y	y	5'	-	d
	AI016804	y	y	5'	-	d
OTX1	BC047612	y	y	5'	C	d
	R25812	y	y	5'	-	d
	BM925861	y	y	5'	C	d
PAX3	AK057009	y	y	i/5'	C	a
	BC008048	y	n	5'	S	e
	BC033986	y	n	5'	S	e
POU2F2	AK027895	y	y	3'	S	
	BC001184	y	y	3'	S	
	BG285118	y	n	3'	S	
	BI459577	y	n	3'	S	
	BG751692	y	n	3'	S	
	BI820816	y	n	3'	S	
	CB043897	y	n	3'	S	
	AL602303	y	n	3'	S	
	BI227004	y	n	3'	S	
	AL524763	y	n	3'	S	
	AL560857	y	n	3'	S	
	AL542363	y	n	3'	S	
	BU153121	y	n	3'	S	
	BX387879	y	n	3'	S	
	BE797935	y	n	3'	S	
	BG715073	y	n	3'	S	
	BU155902	y	n	3'	S	

TABLE 2.21: OS-ESTs mapping in the *H.sapiens* dispersed homeotic loci. Part II.

Gene	OS-EST	Spliced	PolyA signal	Position	CpG	Distance
	CB961738	y	n	3'	S	
	BM128274	y	n	3'	S	
	BM128670	y	n	3'	S	
	BG749041	y	n	3'	S	
	BQ896500	y	n	3'	S	
	BF689676	y	n	3'	S	
	BG764613	y	n	3'	S	
	BG764265	y	n	3'	S	
	BG760880	y	n	3'	S	
	BG766980	y	n	3'	S	
	BG762733	y	n	3'	S	
	BG764249	y	n	3'	S	
	BG761130	y	n	3'	S	
	BE794366	y	n	3'	S	
	BE513293	y	n	3'	S	
	BE513290	y	n	3'	S	
	BU173019	y	n	3'	S	
	BE559733	y	n	3'	S	
	AL535748	y	n	3'	S	
	BM549479	y	n	3'	S	
	CD557618	y	n	3'	S	
	BE297663	y	n	3'	S	
	BE887124	y	n	3'	S	
	BE294828	y	n	3'	S	
	BE729379	y	n	3'	S	
	BE296231	y	n	3'	S	
	AL558191	y	n	3'	S	
	BF315779	y	n	3'	S	
	BG771866	y	n	3'	S	
	BC015776	y	y	i	-	
	BC042152	y	y	i	-	
	AA310711	y	n	i	-	
	BG397451	y	y	i	-	
	BG759321	y	y	i	-	
	AA829777	y	y	i	-	
	AI802162	y	y	i	-	
	AU120278	y	n	5'	-	d
	AK024119	y	y	5'	-	d
POU4F1	BX647243	y	y	3'	-	
	AK095779	y	y	3'/5'	-	
	AV723850	y	n	3'/5'	-	
	AV723416	y	n	3'/5'	-	
	AL706733	y	n	3'/5'	-	
POU4F3	AA843559	y	n	3'/5'	-	
	BX115732	y	n	3'/5'	-	
	AI125723	y	n	3'/5'	-	
TCF1-HNF1A	AW451008	y	y	i/5'	-	
	BE293343	y	n	i/5'	-	
	BG433970	y	n	i/5'	-	
TCF2-HNF1B	BG419303	y	n	5'	S	a
IRX2A-IRX5	AF275804	y	y	5'	C	d
	AI245732	y	y	5'	-	d
	AW117234	y	y	5'	-	d
	BM696001	y	y	5'	-	d
	BM974647	y	y	5'	-	d
	BU729242	y	y	5'	-	d
	BX096724	y	y	5'	-	d
	AA954994	y	y	5'	-	e
	BU569024	y	y	5'	C	d
	AI342725	y	y	5'	-	e
	AI004004	y	y	5'	-	d
	AW023444	y	y	5'	-	d
	BG720116	y	y	5'	-	d

TABLE 2.21: OS-ESTs mapping in the *H.sapiens* dispersed homeotic loci. Part III.

Gene	OS-EST	Spliced	PolyA signal	Position	CpG	Distance
	BU590179	y	y	5'	-	e
	BU160166	y	y	5'	C	d
	AA825162	y	y	5'	-	e
	AA829347	y	y	5'	-	e
	AI242819	y	y	5'	-	e
	AI923595	y	y	5'	-	d
	AI804090	y	y	5'	-	d
	BU165662	y	y	5'	C	d
	BU165627	y	y	5'	C	d
	BF672570	y	n	5'	C	d
	CB854553	y	n	5'	-	d
	W35547	y	n	5'	-	d
	BE504515	y	n	5'	-	d
	BF692451	y	n	5'	C	d
	AI204177	y	n	5'	-	d
	BQ011371	y	n	5'	-	d
	AA928654	y	n	5'	-	d
	BU180741	y	n	5'	-	d
	BU625145	y	n	5'	-	e
	AI827680	y	n	5'	-	d
	BM150438	y	n	5'	-	e
	BM150430	y	n	5'	-	e
	BI561324	y	n	5'	C	d
	BM150433	y	n	5'	-	e
	BQ431041	y	n	5'	-	d
	BQ638202	y	n	5'	C	d
PBX3	AA421133	y	n	5'	S	c
	BX095101	y	n	5'	S	c
PROX1	BE179719	y	n	5'	S	b
	AV681804	y	n	5'	-	c
	AK092251	y	n	5'	-	e
TCF8	AK094743	y	n	i	-	
	CD674797	y	n	5'	C	a
	BM703926	y	n	5'	C	a
	AI474070	y	n	5'	C	a
	BX446976	y	n	5'	C	a
	BI518871	y	n	5'	C	a
	AI808408	y	n	5'	C	a
DLX6	AK091367	y	n	5'	S	c
	AX746939	y	n	5'	S	c
	AK095619	y	n	3'/5'	-	
	BE263812	y	y	3'/5'	-	
	AW245528	y	n	3'/5'	-	
	CA865377	y	n	3'/5'	-	
	BX643287	y	n	3'/5'	-	
EMX2	AY117413	y	y	5'	C	a
	AY117034	y	y	i/5'	C	
	BG699095	y	n	5'	C	a
	BM661903	y	n	5'	C	a
	BM687813	y	n	5'	C	a
	BF939629	y	n	5'	C	a
	AI471863	y	n	5'	C	a
	AI912812	y	n	5'	C	a
	AI493115	y	n	5'	C	a
	AW510874	y	n	5'	-	e
	AW470861	y	n	5'	C	b
	AW135979	y	n	5'	C	b
	AI636020	y	n	5'	C	a
	BF592111	y	n	5'	C	a
	AI435896	y	n	5'	C	b
	CB049839	y	n	5'	C	a
	BX113843	y	n	5'	C	a
LIM1-LHX1	AI492288	y	y	5'	C	b

TABLE 2.21: OS-ESTs mapping in the *H.sapiens* dispersed homeotic loci. Part IV.

Gene	OS-EST	Spliced	PolyA signal	Position	CpG	Distance
	BX112232	y	y	5'	C	b
	BX115740	y	n	5'	C	b
	CB047405	y	n	5'	C	b
	BU552680	y	n	5'	C	b
	AW243436	y	y	5'	-	e
LHX2	BM726651	y	n	5'	S	c
	BM681295	y	n	5'	S	c
LMX1B	AA936470	y	n	5'	C	b
	AI694775	y	n	5'	C	c
	AW451323	y	n	5'	C	b
	BE502769	y	n	5'	C	b
	BE670219	y	n	5'	C	b
TCL3-TLX1	BI767616	y	n	3'/i	S	
	BX117337	y	n	3'/5'	-	
	BC019674	y	n	5'	S	a
	BG913652	y	n	5'	S	a
TITF1	BX161496	y	n	5'	C	a
	BX383579	y	n	5'	C	a
	BX367709	y	n	5'	C	a
OTX2	BC041486	y	y	5'	-	c
	BU739906	y	y	5'	-	c
	BU740358	y	y	5'	-	c
PAX6	BX648962	y	y	5'	C	d
PITX1	AK026965	y	n	i/5'	C	
IPF1	BX383058	y	n	i/5'	C	
POU2F1	BU173125	y	n	5'	-	e
	BE251123	y	n	5'	-	e
	BG396198	y	n	5'	-	e
	BU182356	y	n	5'	-	e
POU3F3	AK096498	y	n	5'	C	c
	AI769363	y	y	5'	-	e
	AI433864	y	y	5'	-	e
	AI652629	y	y	5'	-	d
	AW237801	y	y	5'	-	e
	AI934525	y	y	5'	-	e
	AL036798	y	y	5'	-	d
	AI879561	y	y	5'	-	e
	AI418969	y	y	5'	-	e
	AI769730	y	y	5'	-	e
	AI935664	y	y	5'	-	e
	CA412541	y	y	5'	C	c
	AA436623	y	y	5'	-	e
	AI522160	y	y	5'	-	e
	AI763190	y	y	5'	C	c
	BF195976	y	y	5'	-	e
	AW162117	y	y	5'	-	e
	CK004193	y	y	5'	C	c
	BF700938	y	n	5'	-	d
	AI89178	y	n	5'	-	d
	AL036758	y	n	5'	-	d
	AW160500	y	n	5'	-	d

TABLE 2.21: OS-ESTs mapping in the *H.sapiens* dispersed homeotic loci. Part V.

Gene. Investigated locus
OS-EST. Accession numbers of OS transcripts mapped in the corresponding locus.
Slpiced. "y" indicates correct splicing of the corresponding OS-EST.
PolyA signal. "y" indicates the presence of the polyadenilation consensus AATAAA in the 3' region of OS-ESTs; "n" denotes absence of this consensus sequence.
Position. The position of the OS transcript refers to the transcriptional orientation of the homeobox gene.
5': the OS-EST maps in the 5' region of the homeobox gene.
i/5': the first exon of the OS-EST maps in intronic region of the homeobox gene, while the other(s) exon(s) in its 5' region.
i: the OS-EST maps in intronic regions of the homeobox gene.
3'/i: the first exon of OS-EST maps in the 3' regions of the homeobox gene, while the other(s) exon(s) in its intronic regions.
3': the OS-EST maps in the 3' region of the homeobox gene.
CpG. We investigate the presence of a CpG island containing both the first exon of OS-EST and the first exon of the homeobox gene.
"-": the first exon of the OS-EST is not in a CpG island.
"C": the first exons of the OS transcript and the homeobox gene are both in the same CpG island.
"S": the first exons of the OS transcript and the homeobox gene are in separated CpG island
Distance. I calculate the distance between the first exon of the OS-EST and the first exon of the homeobox gene.
"a": < 0.3 kb
"b": 0.3-1 kb
"c": 1-3 kb
"d": 3-10 kb
"e": >10 kb

TABLE 2.22: OS-ESTs mapping in the *M.musculus* dispersed homeobox loci. Part I

Gene	OS-EST	Spliced	PolyA signal	Position	CpG	Distance
Tcf1/Hnf1a	BC031162	y	y	3'	S	
	AK008740	y	n	3'	S	
	AA254815	y	n	i/5'	-	
	AI464806	y	n	i/5'	-	
	AW226772	y	n	i/5'	-	
	CB950971	y	n	i/5'	-	
	CF620880	y	y	i/5'	-	
Tcf2/Hnf1b	BB650367	y	n	5'	-	c
	AK082716	y	n	5'	-	c
	BG963365	y	n	5'	-	b
Irx2	BF162870	y	n	5'	-	d
Pbx3	BB633614	y	n	i	-	
	BY724378	y	n	i	-	
Prox1	BF469000	y	n	i	-	
	BF471703	y	n	5'	-	a
	BY730214	y	n	5'	-	a
Tcf8/Zfhx1a	AK046967	y	n	i/5'	C	
	BY718683	y	n	i/5'	C	
	BY723840	y	n	i/5'	C	
	BB632161	y	n	i/5'	C	
	BB646373	y	n	i/5'	C	
	BY727378	y	n	i/5'	C	
	BE646958	y	n	i/5'	-	
Dlx6	AK038694	y	n	5'	-	c
	AK032537	y	y	3'/5'	-	
	AK032180	y	y	5'	-	c
	AK044034	y	y	5'	-	c
	BB641078	y	n	5'	-	c
	CA318216	y	n	3'/5'	-	
	CA451082	y	n	5'	-	c
	BB631966	y	n	5'	-	c
	AU035954	y	n	3'/5'	-	
	BB622920	y	n	3'/5'	-	
	BY003744	y	n	5'	-	c
	BY279781	y	n	3'/5'	-	
Emx2	AY117414	y	y	5'	C	a
	AK033011	y	y	5'	C	a
	BY002746	y	n	5'	C	a
Lim1/Lhx1	AK018772	y	y	5'	C	a
	W62870	y	n	5'	C	a
	BU560004	y	y	5'	C	a
	CA318627	y	y	5'	C	a
	BB860096	y	n	5'	C	a
	BY705450	y	y	5'	C	a
	BE653895	y	n	5'	C	a
	W98734	y	n	5'	C	a
	AI848463	y	y	5'	C	a
Lim2/Lhx2	BU937255	y	n	i	-	
Lmx1b	AK081497	y	y	5'	C	b
	BY135800	y	n	5'	C	b
	BY740925	y	n	5'	C	b
	BY346398	y	n	5'	C	b
	BY346690	y	n	5'	C	b
	BY345755	y	n	5'	C	b
	BY346636	y	n	5'	C	b
	BY344987	y	n	5'	C	b
	BY345311	y	n	5'	C	b
	BY728389	y	n	5'	C	b
	BY141711	y	n	5'	C	b
	BY317324	y	n	5'	C	b
	BY151591	y	n	5'	C	b
Tlx1*	BB573627	y	n	3'/i	S	
Titf1	BY117165	y	n	5'	-	b

TABLE 2.22: OS-ESTs mapping in the *M.musculus* dispersed homeobox loci. Part II.

Gene	OS-EST	Spliced	PolyA signal	Position	CpG	Distance
	BY120663	y	n	5'	-	b
Otx2	AK042665	y	n	5'	-	c
	BB639822	y	n	5'	-	c
	BB246370	y	n	5'	-	c
	BU504680	y	n	5'	-	c
	BE860198	y	n	5'	S	d
Pax6	AK029183	y	y	i/5'	-	
	AK044354	y	n	i/5'	-	
	BB642285	y	n	i/5'	-	
	BB613704	y	n	i/5'	-	
	BG802265	y	n	i/5'	-	
Pitx1*	AK043531	y	y	i/5'	-	
	BY299882	y	n	i/5'	-	
Pmx1/Prrx1	BB648488	y	n	i	-	
Ipf1	AK008755	y	y	i/5'	-	
	BY708657	y	n	i/5'	-	
	BG085011	y	n	i/5'	-	
Pou2f1	AK016078	y	n	3'/i	-	
	BY715679	y	n	3'/i	-	
Pou3f3	AK011437	y	y	5'	S	c
	AK019154	y	y	5'	S	c
	BC049671	y	y	5'	-	e
	BC058417	y	y	5'	-	d
	BC059892	y	y	5'	-	d
	CA319029	y	y	5'	-	d
	CF750962	y	y	5'	S	c
	AW045679	y	y	5'	-	e
	BQ747456	y	y	5'	-	d
	BU562288	y	y	5'	S	c
	BB650367	y	y	5'	-	d
	BQ748271	y	y	5'	-	e
	BX515333	y	y	5'	-	d
	CA516666	y	y	5'	-	d
	AA028688	y	y	5'	-	d
	BM114333	y	y	5'	-	d
	BQ554853	y	y	5'	-	d
	BU842137	y	y	5'	-	e
	CB205150	y	y	5'	-	d
	BU961251	y	y	5'	-	e
	AV149696	y	y	5'	-	e
	BE864528	y	y	5'	-	e
	AW244777	y	y	5'	S	c
	BQ043769	y	y	5'	-	d
	CF216586	y	y	5'	-	e
	BQ554854	y	y	5'	-	d
	AA041813	y	y	5'	-	d
	BY003997	y	y	5'	S	c
	BY255277	y	n	5'	-	d
	AK048630	y	n	5'	S	c
	BY289704	y	n	5'	-	d
Cutl1	AK016175	y	n	i	S	
	BY715779	y	n	i	S	
Meis1	CF739522	y	n	i	-	
	BI693346	y	n	i	-	
	BB869286	y	n	i	-	
	BX634601	y	n	i	-	
	AA182163	y	n	i	-	
	AA866970	y	n	i	-	
	BX513192	y	n	i	-	
Meis2	BB873229	y	n	i	-	c
	BB872324	y	n	i	-	c
	BB869881	y	n	i	-	c
	BY733759	y	n	i	-	c

TABLE 2.22: OS-ESTs mapping in the *M.musculus* dispersed homeobox loci. Part III.

Gene	OS-EST	Spliced	PolyA signal	Position	CpG	Distance
	BB802079	y	n	i	-	e
Dlx1	BB642046	y	n	5'	-	c
	BB647803	y	n	5'	-	c
	BF469576	y	n	3'/5'	-	
	BY713853	y	n	3'	-	
	BY257684	y	n	3'	-	
	BY263576	y	n	3'	-	
	BY263817	y	n	3'	-	
Dlx4	AK080562	y	n	i/5'	C	
	AK076510	y	n	i/5'	C	
	BB640871	y	n	i/5'	C	
	BB614242	y	n	i/5'	C	
	AI098507	y	n	i/5'	C	
Msx1	AA032878	y	n	3'	-	
Evx1	AK031498	y	n	i/5'	S	
	AK017671	y	n	5'	C	a
	BY717450	y	n	5'	C	a
	BB618942	y	n	5'	C	a
	BB620958	y	n	i/5'	S	
	BY009058	y	n	5'	C	a
	BB873689	y	n	5'	C	a
Gbx2	AK036422	y	n	5'	C	a
	BB629310	y	n	5'	C	a
Meox1	AA184243	y	n	i	-	
	AI591966	y	n	i	-	
	AI662525	y	n	i	-	
Nkx 2.6	AKOO7038	y	n	5'	-	a
	BY707106	y	n	5'	-	a
Nkx 6.2	BC056341	y	y	3'	-	
Nkx 2.2	AK045921	y	y	3'/5'	C	
	AK020097	y	n	5'	-	c
	BY718829	y	n	5'	-	c
Hmx1	AK053452	y	n	5'	C	b
	BB665013	y	n	5'	C	b
Tlx3	AK029182	y	n	5'	C	b
	BB613702	y	n	5'	C	b
Crx	AU045294	y	n	5'	-	e
	AU021231	y	n	5'	-	e
	AU018615	y	n	5'	-	e
	AU043259	y	n	5'	-	e
	BG071513	y	n	5'	-	e
	AU019481	y	n	5'	-	e
	AU044467	y	n	5'	-	e
	CA319029	y	n	5'	-	e
	C88456	y	n	5'	-	e
	AA549101	y	n	5'	-	e
	AA574835	y	n	5'	-	e
	AA638834	y	n	5'	-	e
	AA666904	y	n	5'	-	e
	AA590396	y	n	5'	-	e
	AA446756	y	n	5'	-	e
	AA606869	y	n	5'	-	e
	AA607846	y	n	5'	-	e
	AA608038	y	n	5'	-	e
	AA608222	y	n	5'	-	e
	AA671433	y	n	5'	-	e
	AA683749	y	n	5'	-	e
	AA474743	y	n	5'	-	e
	AA575061	y	n	5'	-	e
	AA473194	y	n	5'	-	e
	BY736445	y	n	5'	-	e
	AA571427	y	n	5'	-	e
	BX526306	y	n	5'	-	e

TABLE 2.22: OS-ESTs mapping in the *M.musculus* dispersed homeobox loci. Part IV.

Gene	OS-EST	Spliced	PolyA signal	Position	CpG	Distance
	AA636324	y	n	5'	-	e
	AA647722	y	n	5'	-	e
	BY732031	y	n	5'	-	e
	BY742111	y	n	5'	-	e
	C89532	y	n	5'	-	e
	BB642159	y	n	5'	-	e
	AA591078	y	n	5'	-	e
	BY032255	y	n	5'	-	e
	BF469102	y	n	5'	-	e
	CB845701	y	n	5'	-	e
Rax	AK033044	y	y	i/5'	-	
	BB623983	y	n	i/5'	-	
Lhx5	AI430373	y	n	5'	C	a
	W89644	y	n	5'	C	a
	BX518548	y	n	5'	C	a
Lhx8	BY730156	y	n	5'	-	c

TABLE 2.22

Gene. Investigated locus

OS-EST. Accession numbers of OS transcripts mapped in the corresponding locus.

Slpiced. “y” indicates correct splicing of the corresponding OS-EST.

PolyA signal. “y” indicates the presence of the polyadenilation consensus AATAAA in the 3’ region of OS-ESTs; “n” denotes absence of this consensus sequence.

Position. The position of the OS transcript refers to the transcriptional orientation of the homeobox gene.

5’: the OS-EST maps in the 5’ region of the homeobox gene.

i/5’: the first exon of the OS-EST maps in intronic region of the homeobox gene, while the other(s) exon(s) in its 5’ region.

i: the OS-EST maps in intronic regions of the homeobox gene.

3’/i: the first exon of OS-EST maps in the 3’ regions of the homeobox gene, while the other(s) exon(s) in its intronic regions.

3’: the OS-EST maps in the 3’ region of the homeobox gene.

CpG. We investigate the presence of a CpG island containing both the first exon of OS-EST and the first exon of the homeobox gene.

“-“: the first exon of the OS-EST is not in a CpG island.

“C”: the first exons of the OS transcript and the homeobox gene are both in the same CpG island.

“S”: the first exons of the OS transcript and the homeobox gene are in separated CpG island

Distance. I calculate the distance between the first exon of the OS-EST and the first exon of the homeobox gene.

“a”: < 0.3 kb

“b”: 0.3-1 kb

“c”: 1-3 kb

“d”: 3-10 kb

“e”: >10 kb

TABLE 3.1: OS-ESTs mapping in 90 *H.sapiens* genomic loci. Part I.

Gene	OS EST	Spliced	Position	Known Gene
aggrecan 1	-	-	-	-
aminolevulinic acid synthase 1	-	-	-	-
amisyn	-	-	-	-
angiogenin	-	-	-	-
aryl-hydrocarbon receptor	y	-	-	-
Bardet-Biedl syndrome 4 homolog	y	y	5'	-
beta-glucuronidase	-	-	-	-
BING4 protein	y	y	5'	H2-KE2
bone morphogenetic protein 5	-	-	-	-
Bcl6 interacting corepressor	-	-	-	-
cadherin 11	-	-	-	-
calcineurin binding protein 1	y	y	5'	-
calpain 3	-	-	-	-
calponin 2	y	y	5'	-
choroidermia	-	-	-	-
claudin 12	-	-	-	-
cullin 3	y	y	3'	-
decorin	-	-	-	-
dipeptidase 1 (renal)	-	-	-	-
dipeptidylpeptidase 6	y	y	i	-
dual specificity phosphatase 1	y	y	3'/5'	-
endothelial differentiation, sphingolipid G-protein-coupled receptor, 3	-	-	-	-
endothelial differentiation-related factor 1	-	-	-	-
Eph receptor A3	-	-	-	-
epsin 2	y	y	i	-
fatty acid desaturase 3	-	-	-	-
fem-1 homolog c (C.elegans)	-	-	-	-
Fgfr1 oncogene partner	y	y	5'	-
fibroblast growth factor 7	y	y	3'/5'	-
fibromodulin	-	-	-	-
G protein coupled receptor 24	-	-	-	-
G protein-coupled receptor 64	y	y	3'/5'	-
glucose phosphate isomerase 1	-	-	-	-
guanine deaminase	-	-	-	-
guanylate cyclase 1, soluble, alpha 3	-	-	-	-
histone deacetylase 10	-	-	-	-
HMG-box protein SOX21	y	y	5'	-
HRD1 protein	y	y	3'	MRPL49
HS1 binding protein	-	-	-	-
insulin induced gene 1	y	y	5'	-
integrin alpha 4	y	y	3'	-
interleukin 1 receptor-like 2	y	y	5'	-
intestinal cell kinase	y	y	5'	FBX09
jagged 1	y	y	3'	-
karyopherin (importin) alpha 1	y	y	3'	-
Kruppel-like factor 4	-	-	-	-
latrophilin 1	y	y	3/i; i	-
legumain	-	-	-	-
leprecan 1	y	y	5'	-
leptin receptor	-	-	-	-
L-fucose kinase	y	y	3'	COG4
ligatin	y	y	3'	RASSF5
lipoic acid synthetase	y	y	5'	RPL9
Ly6/neurotoxin 1	-	-	-	-
lysosomal apyrase-like 1	y	y	5'	-
mab-21-like 1 (C. elegans)	y	y	3'/5'	-

TABLE 3.1: OS-ESTs mapping in 90 *H.sapiens* genomic loci. Part II.

Gene	OS EST	Spliced	Position	Known Gene
mahogunin, ring finger 1	y	y	5';i; 3'	-
Mak3p homolog (S. cerevisiae)	y	y	5'	ATP6V1A
makorin, ring finger protein, 1	-	-	-	-
mannosidase 1, beta	-	-	-	-
Mblk1-related protein-1	y	y	3'	HCAP-G
metallothionein 2	-	-	-	-
methyl CpG binding protein 2	-	-	-	-
microfibrillar-associated protein 1	y	y	3'; 5'	HYPK
mitochondrial ribosomal protein L19	y	y	3'	-
monoamine oxidase A	-	-	-	-
mucolipin 2	y	y	5'	NYD-SP29
multimerin	-	-	-	-
myomesin 2	-	-	-	-
myosin X	-	-	-	-
neurexin II	y	y	i	-
neuritin 1	y	y	3'	-
neuropilin	-	-	-	-
ninjurin 1	-	-	-	-
pallidin	-	-	-	-
parvin, beta	y	y	5'	-
pellino 2	-	-	-	-
pericentrin 2	y	y	5'	-
peroxiredoxin 2	-	-	-	-
phospholipase C, gamma 1	-	-	-	-
plexin B1	y	y	3'	-
polyamine oxidase	-	-	-	-
reticulon 4	y	y	3'	-
retinitis pigmentosa 2 homolog (human)	y	y	3'	-
scinderin	y	y	5';i	-
Sec15A	-	-	-	-
SEC22L1	-	-	-	-
sequestosome 1	y	y	5':3'	MGAT4B
syntaxin 12	-	-	-	-
talin 2	y	y	5'	-

TABLE 3.1

Gene. Investigated locus

OS-EST. "y" indicates the presence of OS-ESTs; "-" indicates the absence of OS-ESTs

Slpiced. "y" indicates correct splicing of the corresponding OS-EST.

Position. The position of the OS transcript refers to the transcriptional orientation of the sense gene.

5': the OS-EST maps in the 5' region of the target gene.

i/5': the first exon of the OS-EST maps in intronic region of the target gene, while other exons in its 5' region.

i: the OS-EST maps in intronic regions of the target gene.

3'/i: the first exon of OS-EST maps in the 3' regions of the target gene, while other exons in its intronic regions.

3': the OS-EST maps in the 3' region of the target gene.

Know gene. Known genes corresponding to OS transcripts.

TABLE 3.2: OS-ESTs mapping in 90 *M.musculus* genomic loci. Part I.

Gene	OS EST	Spliced	Position	Known gene
aggrecan 1	-	-	-	-
aminolevulinic acid synthase 1	y	y	-	-
amisyn	-	-	-	-
angiogenin	-	-	-	-
aryl-hydrocarbon receptor	-	-	-	-
Bardet-Biedl syndrome 4 homolog	y	y	3'	-
beta-glucuronidase	-	-	-	-
BING4 protein	y	y	5'	H2-Ke2
bone morphogenetic protein 5	-	-	-	-
Bcl6 interacting corepressor	-	-	-	-
cadherin 11	-	-	-	-
calcineurin binding protein 1	-	-	-	-
calpain 3	y	y	5'	-
calponin 2	-	-	-	-
choroidermia	-	-	-	-
claudin 12	-	-	-	-
cullin 3	-	-	-	-
decorin	-	-	-	-
dipeptidase 1 (renal)	-	-	-	-
dipeptidylpeptidase 6	-	-	-	-
dual specificity phosphatase 1	-	-	-	-
endothelial differentiation, sphingolipid G-protein-coupled receptor, 3	y	y	i/5'	-
endothelial differentiation-related factor 1	-	-	-	-
Eph receptor A3	-	-	-	-
epsin 2	-	-	-	-
fatty acid desaturase 3	-	-	-	-
fem-1 homolog c (C.elegans)	-	-	-	-
Fgfr1 oncogene partner	-	-	-	-
fibroblast growth factor 7	y	y	3'/5'	-
fibromodulin	-	-	-	-
G protein coupled receptor 24	-	-	-	-
G protein-coupled receptor 64	y	y	3'	-
glucose phosphate isomerase 1	-	-	-	-
guanine deaminase	y	y	5'	-
guanylate cyclase 1, soluble, alpha 3	-	-	-	-
histone deacetylase 10	y	y	5'	-
HMG-box protein SOX21	y	y	5'	-
HRD1 protein	-	-	-	-
HS1 binding protein	y	y	5'	-
insulin induced gene 1	-	-	-	-
integrin alpha 4	-	-	-	-
interleukin 1 receptor-like 2	-	-	-	-
intestinal cell kinase	y	y	5'	Fbxo9
jagged 1	-	-	-	-
karyopherin (importin) alpha 1	y	y	5'	-
Kruppel-like factor 4	-	-	-	-
latrophilin 1	y	y	3'/i	-
legumain	-	-	-	-
leprecan 1	y	y	5';3'	-
leptin receptor	-	-	-	-
L-fucose kinase	y	y	3'	Cog4
ligatin	y	y	3'	Rassf5
lipoic acid synthetase	y	y	5'	Rpl9
Ly6/neurotoxin 1	-	-	-	-
lysosomal apyrase-like 1	y	y	5';3'	-
mab-21-like 1 (C. elegans)	y	y	3'/5'	-

TABLE 3.2: OS-ESTs mapping in 90 *M.musculus* genomic loci. Part II.

Gene	OS EST	Spliced	Position	Known gene
mahogunin, ring finger 1	y	y	5';3'	-
Mak3p homolog (S. cerevisiae)	y	y	5'	Atp6v1a
makorin, ring finger protein, 1	y	y	5'	-
mannosidase 1, beta	-	-	-	-
Mblk1-related protein-1	y	y	3'	-
metallothionein 2	-	-	-	-
methyl CpG binding protein 2	-	-	-	-
microfibrillar-associated protein 1	-	-	-	-
mitochondrial ribosomal protein L19	y	y	3'	-
monoamine oxidase A	-	-	-	-
mucolipin 2	-	-	-	-
multimerin	-	-	-	-
myomesin 2	y	y	i	-
myosin X	y	y	5'	-
neurexin II	y	y	3'	-
neuritin 1	y	y	3'	-
neuropilin	y	y	i	-
ninjurin 1	-	-	-	-
pallidin	y	y	5'	-
parvin, beta	-	-	-	-
pellino 2	-	-	-	-
pericentrin 2	y	y	5'	-
peroxiredoxin 2	y	y	3'	-
phospholipase C, gamma 1	-	-	-	-
plexin B1	-	-	-	-
polyamine oxidase	-	-	-	-
reticulon 4	y	y	3'	-
retinitis pigmentosa 2 homolog (human)	-	-	-	-
scinderin	-	-	-	-
Sec15A	-	-	-	-
SEC22L1	y	y	5'	-
sequestosome 1	y	y	5';3'	-
syntaxin 12	-	-	-	-
talin 2	y	y	i	-

TABLE 3.2

Gene. Investigated locus
OS-EST. "y" indicates the presence of OS-ESTs; "-" indicates the absence of OS-ESTs
Slpiced. "y" indicates correct splicing of the corresponding OS-EST.
Position. The position of the OS transcript refers to the transcriptional orientation of the sense gene.
5': the OS-EST maps in the 5' region of the target gene.
i/5': the first exon of the OS-EST maps in intronic region of the target gene, while other exons in its 5' region.
i: the OS-EST maps in intronic regions of the target gene.
3'/i: the first exon of OS-EST maps in the 3' regions of the target gene, while other exons in its intronic regions.
3': the OS-EST maps in the 3' region of the target gene.
Know gene. Known genes corresponding to OS transcripts.

Chapter VI: REFERENCES

Adachi N, Lieber MR. Bidirectional gene organization: a common architectural feature of the human genome. Cell (2002) 109: 807-9.

Agresti A, Meneveri R, Siccardi AG, Marozzi A, Corneo G, Gaudi S, Ginelli E. Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequences (HaeIII family). J. Mol. Biol. (1989) 205: 625-31.

Agresti A, Rainaldi G, Lobbiani A, Magnani I, Di Lernia R, Meneveri R, Siccardi AG, Ginelli E. Chromosomal location by in situ hybridization of the human Sau3A family of DNA repeats. Hum. Genet. (1987) 75: 326-32.

Akhtar A, Zink D, Becker PB. Chromodomains are protein-RNA interaction modules. Nature (2000) 407: 405-9.

Allfrey V, Faulkner RM, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA (1964) 51:786-94.

Alkema MJ, Jacobs H, van Lohuizen M, Berns A. Pertubation of B and T cell development and predisposition to lymphomagenesis in Emu Bmi1 transgenic mice require the Bmi1 RING finger. Oncogene (1997) 15: 899-910.

Alvarez-Venegas R, Pien S, Sadder M, Witmer X, Grossniklaus U, Avramova Z. ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. Curr. Biol. (2003) 13: 627-37.

Amrein H. Multiple RNA-protein interactions in Drosophila dosage compensation. Genome Biol. (2000) 1: 1030.

Andersen AA, Panning B. Epigenetic gene regulation by noncoding RNAs. Curr. Opin. Genet. Dev. (1998) 8: 173-8.

Apiou F, Flagiello D, Cillo C, Malfoy B, Poupon MF, Dutrillaux B. Fine mapping of human HOX gene clusters. Cytogenet. Cell. Genet. (1996) 73: 114-5.

Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr. Biol. (2001) 11: 1017-27.

Bae E, Calhoun VC, Levine M, Lewis EB, Drewell RA. Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the Drosophila bithorax complex. Proc. Natl. Acad. Sci. USA (2002) 99: 16847-52.

Bajusz I, Sipos L, Gyorgypal Z, Carrington EA, Jones RS, Gausz J, Gyurkovics H. The Trithorax-mimic allele of Enhancer of zeste renders active domains of target genes accessible to polycomb-group-dependent silencing in Drosophila melanogaster. Genetics (2001) 159: 1135-50.

Banerjee-Basu S, Baxevanis AD. Molecular evolution of the homeodomain family of transcription factors. Nucleic Acids Res. (2001) 29: 3258-69.

Bannister AJ, Schneider R, Kouzarides T. Histone methylation: Dynamic or static? Cell (2002) 109:801-6.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromodomain. Nature (2001) 410:120-4.

Bardos JI, Saurin AJ, Tissot C, Duprez E, Freemont PS. HPC3 is a new human polycomb orthologue that interacts and associates with RING1 and Bmi1 and has transcriptional repression properties. J. Biol. Chem. (2000) 275: 28785-92.

Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell (2004) 116: 281-97.

Bea S, Tort F, Pinyol M, Puig X, Hernandez L, Hernandez S, Fernandez PL, van Lohuizen M, Colomer D, Campo E. BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas. Cancer Res. (2001) 61: 2409-12.

Beck S., Trowsdale J. Sequence organisation of the class II region of the human MHC. Immunol. Rev. (1999) 167: 201-210.

Beisel C, Imhof A, Greene J, Kremmer E, Sauer F. Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. Nature (2002) 419: 857-62.

Bel S, Core N, Djabali M, Kieboom K, Van der Lugt N, Alkema MJ, Van Lohuizen M. Genetic interactions and dosage effects of Polycomb group genes in mice. Development (1998) 125: 3543-51.

Belmont AS, Dietzel S, Nye AC, Strukov YG, Tumbar T. Large-scale chromatin structure and function. Curr. Opin. Cell. Biol. (1999) 11: 307-11.

Belyaev ND, Houben A, Baranczewski P, Schubert I. Histone H4 acetylation in plant heterochromatin is altered during the cell cycle. Chromosoma (1997) 106:193-7

Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T, Schreiber SL. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc. Natl. Acad. Sci. USA (2002) 99: 8695-700.

Bird AP, Wolffe AP. Methylation-induced repression—belts, braces, and chromatin. Cell (1999) 99: 451-4.

Boggs BA, Cheung P, Heard E, Spector DL, Chinault AC, Allis CD. Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat. Genet. (2002) 30: 73-6.

Boncinelli E, Simeone A, Acampora D, Mavilio F. HOX gene activation by retinoic acid. Trends Genet. (1991) 7: 329-34.

Breiling A, Orlando V. SET domain proteins reSET gene expression. Nat. Struct. Biol. (2002) 9: 894-6.

Breiling A, Turner BM, Bianchi ME, Orlando V. General transcription factors bind promoters repressed by Polycomb group proteins. Nature (2001) 412: 651-5.

Brock HW, van Lohuizen M. The Polycomb group—no longer an exclusive club? Curr. Opin. Genet. Dev. (2001) 11: 175-81.

Brosius J. Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica (1999) 107: 209-38.

Brosius J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene (1999) 238: 115-34.

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell (1992) 71: 527-42.

Brown KE, Baxter J, Graf D, Merkenschlager M, Fisher AG. Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. Mol. Cell. (1999) 3: 207-17.

Brown KE, Guest SS, Smale ST, Hahm K, Merkenschlager M, Fisher AG. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. Cell (1997) 91: 845-54.

Bulger M., van Doorninck J.H., Saitoh N., Telling A., Farrell C., Bender M.A., Felsenfeld G., Axel R., Groudine M. Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc. Natl. Acad. Sci. USA (1999) 96: 5129-34.

Burglin TR. A comprehensive classification of homeobox genes. (1994) In Guidebook to homeobox genes, D. Duboule, ed. Oxford, UK. Oxford University Press: 27-41.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science (2002) 298: 1039-43.

Caslini C, Alarcon AS, Hess JL, Tanaka R, Murti KG, Biondi A. The amino terminus targets the mixed lineage leukemia (MLL) protein to the nucleolus, nuclear matrix and mitotic chromosomal scaffolds. Leukemia (2000) 14: 1898-908.

Cavalli G, Paro R. Chromo-domain proteins: linking chromatin structure to epigenetic regulation. Curr. Opin. Cell. Biol. (1998a) 10: 354-60.

Cavalli G, Paro R. Epigenetic inheritance of active chromatin after removal of the main transactivator. Science (1999) 286: 955-8.

Cavalli G, Paro R. The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell (1998b) 93: 505-18.

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, Wheeler R, Wong B, Drenkow J, Yamanaka M, Patel S, Brubaker S, Tammana H, Helt G, Struhl K, Gingeras TR. Unbiased Mapping of Transcription Factor Binding Sites along Human Chromosomes 21 and 22 Points to Widespread Regulation of Noncoding RNAs. Cell (2004) 116:499-509.

Chadwick BP, Willard HF. SETting the stage. Eed-Enx1 leaves an epigenetic signature on the inactive X chromosome. Dev Cell. (2003) 4: 445-7.

Chinwalla V, Jane EP, Harte PJ. The Drosophila trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. EMBO J. (1995) 14: 2056-65.

Cimino G, Rapanotti MC, Sprovieri T, Elia L. ALL1 gene alterations in acute leukemia: biological and clinical aspects. Haematologica (1998) 83: 350-7.

Coré N, Bel S, Gaunt SJ, Aurrand-Lions M, Pearce J, Fisher A, Djabali M.
Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. Development (1997) 124: 721-9.

Crosby MA, Miller C, Alon T, Watson KL, Verrijzer CP, Goldman-Levi R, Zak NB. The trithorax group gene moira encodes a brahma-associated putative chromatin-remodeling factor in Drosophila melanogaster. Mol. Cell. Biol. (1999) 19: 1159-70.

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell (2002) 111: 185-96.

Davie JR, Chadee DN. Regulation and regulatory parameters of histone modifications. J. Cell. Biochem. Suppl. (1998) 30-31: 203-13.

Davie JR. Covalent modifications of histones: expression from chromatin templates. Deininger PL, Batzer MA. Mammalian retroelements. Genome Res. (2002) 12: 1455-65.

Dennis C. The brave new world of RNA. Nature (2002) 418: 122-4.

Dingwall AK, Beek SJ, McCallum CM, Tamkun JW, Kalpana GV, Goff SP, Scott MP. The Drosophila snr1 and brm proteins are related to yeast SWI/SNF proteins and are components of a large protein complex. Mol. Biol. Cell. (1995) 6: 777-91.

Dolle P, Fraulob V, Duboule D. Developmental expression of the mouse Evx-2 gene: relationship with the evolution of the HOM/Hox complex. Dev. Suppl. (1994) 143-53.

Drewell RA, Bae E, Burr J, Lewis EB. Transcription defines the embryonic domains of cis-regulatory activity at the Drosophila bithorax complex. Proc. Natl. Acad. Sci. USA (2002) 99: 16853-8.

Duboule D. Vertebrate hox gene regulation: clustering and/or colinearity? Curr. Opin. Genet. Dev. (1998) 8: 514-8.

Ekwall K, Nimmo ER, Javerzat JP, Borgstrom B, Egel R, Cranston G, Allshire R. Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. J. Cell. Sci. (1996) 109: 2637-48.

Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y. Methylation of H3-lysine 79 is mediated by a new HMTases without a SET domain. Curr. Biol. (2002) 12:1052-1058

Ferguson-Smith AC, Reik W. The need for Eed. Nat. Genet. (2003) 33: 433-4.

Ferguson-Smith AC, Surani MA. Imprinting and the epigenetic asymmetry between parental genomes. Science (2001) 293: 1086-9.

Francastel C, Schubeler D, Martin DI, Groudine M. Nuclear compartmentalization and gene activity. Nat. Rev. Mol. Cell. Biol. (2000) 1: 137-43.

Francis NJ, Saurin AJ, Shao Z, Kingston RE. Reconstitution of a functional core polycomb repressive complex. Mol. Cell (2001) 8: 545-56.

Franke A, Baker BS. Dosage compensation rox! Curr. Opin. Cell. Biol. (2000) 12: 351-4.

Franke A, DeCamillis M, Zink D, Cheng N, Brock HW, Paro R. Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. EMBO J. (1992) 118):2941-50.

Frasch M, Chen X, Lufkin T. Evolutionary-conserved enhancers direct region-specific expression of the murine Hoxa-1 and Hoxa-2 loci in both mice and Drosophila. Development (1995) 121: 957-74.

Gaunt SJ. 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: 135-44.

Gehring WJ, Affolter M, Burglin T. Homeodomain proteins. Annu. Rev. Biochem. (1994) 63: 487-526.

Ghanem N, Jarinova O, Amores A, Long Q, Hatch G, Park BK, Rubenstein JL, Ekker M. Regulatory roles of conserved intergenic domains in vertebrate Dlx bigene clusters. Genome Res. (2003) 13: 533-43.

Gildea JJ, Lopez R, Shearn A. A screen for new trithorax group genes identified little imaginal discs, the Drosophila melanogaster homologue of human retinoblastoma binding protein 2. Genetics (2000) 156: 645-63.

Gitlin L, Karelsky S, Andino R. Short interfering RNA confers intracellular antiviral immunity in human cells. Nature (2002) 418: 430-4.

Graham A, Papalopulu N, Krumlauf R. The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell (1989) 57: 367-78.

Grewal SI, Moazed D. Heterochromatin and epigenetic control of gene expression. Science (2003) 301: 798-802.

Grunstein M. Histone acetylation in chromatin structure and transcription. Nature (1997) 389: 349-52.

Gunster MJ, Raaphorst FM, Hamer KM, den Blaauwen JL, Fieret E, Meijer CJ, Otte AP. Differential expression of human Polycomb group proteins in various tissues and cell types. J. Cell. Biochem. (2001) 81: 129-143.

Hannon GJ. RNA interference. Nature (2002) 418: 244-51.

Hanson RD, Hess JL, Yu BD, Ernst P, van Lohuizen M, Berns A, van der Lugt NM, Shashikant CS, Ruddle FH, Seto M, Korsmeyer SJ. Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. Proc. Natl. Acad. Sci. USA (1999) 96: 14372-7.

Hardison R.C., Oeltjen J., Miller W. Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. Genome Res. (1997) 7: 959-66.

Hassan AH, Neely KE, Workman JL. Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. Cell (2001) 104: 817-27.

Haupt Y, Bath ML, Harris AW, Adams JM. bmi-1 transgene induces lymphomas and collaborates with myc in tumorigenesis. Oncogene (1993) 8: 3161-4.

Heard E, Clerc P, Avner P. X-chromosome inactivation in mammals. Annu. Rev. Genet. (1997) 31: 571-610.

Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell (2001) 107: 727-38.

Henikoff S. Heterochromatin function in complex genomes. Biochim. Biophys. Acta. (2000) 1470:1-8.

Herault Y, Hraba-Renevey S, van der Hoeven F, Duboule D. Function of the Evx-2 gene in the morphogenesis of vertebrate limbs. EMBO J. (1996) 15: 6727-38.

Hickey DA. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics (1982) 101: 519-31.

Holliday R. Epigenetics: an overview. Dev. Genet. (1994) 15: 453-7.

Holliday R. The inheritance of epigenetic defects. Science (1987) 238: 163-70.

Hughes JF, Coffin JM. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. Nat. Genet. (2001) 29: 487-9.

Ingham P W. Nature (1983) 306: 591–593.

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature (2001) 409: 860-921.

Jacobs JJ, van Lohuizen M. Polycomb repression: from cellular memory to cellular proliferation and cancer. Biochim. Biophys. Acta. (2002) 1602: 151-61

Jacque JM, Triques K, Stevenson M. Modulation of HIV-1 replication by RNA interference. Nature (2002) 418: 435-8.

Jensen S, Gassama MP, Heidmann T. Taming of transposable elements by homology-dependent gene silencing. Nat. Genet. (1999) 21: 209-12.

Jenuwein T, Allis CD. Translating the histone code. Science (2001) 293: 1074-80.
Jenuwein T. Molecular biology. An RNA-guided pathway for the epigenome. Science (2002) 297: 2215-8.

Jeppesen P, Turner BM. The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell (1993) 74:281-89.

Jin Y, Wang Y, Johansen J, Johansen KM. JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. J. Cell. Biol. (2000) 149: 1005-10.

Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. Science (2001) 293:1068-70.

Jones PA. The DNA methylation paradox. Trends Genet. (1999) 15: 34-7.

Jurka J., Smith T. A fundamental division in the Alu family of repeated sequences. Proc. Natl. Acad. Sci. USA (1988) 85: 4775-78.

Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR. Large-scale transcriptional activity in chromosomes 21 and 22. Science (2002) 296: 916-9.

Karpen GH, Allshire RC. The case for epigenetic effects on centromere identity and function. Trends Genet. (1997) 13: 489-96.

Kazazian HH Jr. Mobile elements and disease. Curr. Opin. Genet. Dev. (1998) 8:343-50.

Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI. Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell (1999) 98: 513-22.

Kennison JA, Tamkun JW. Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc. Natl. Acad. Sci. USA (1988) 85: 8136-40.

Kennison JA. The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Annu. Rev. Genet. (1995) 29: 289-303.

Keynes R, Krumlauf R. Hox genes and regionalization of the nervous system. Annu. Rev. Neurosci. (1994) 17: 109-32.

Kidwell MG, Holyoake AJ. Transposon-induced hotspots for genomic instability. Genome Res. (2001) 11:1321-2.

Kim MH, Park JH, Park HW, Chung IH, Park KA. PAC and cosmid contig spanning the HOXA cluster on human chromosome 7p15. Somat. Cell. Mol. Genet. (1998) 24: 245-8.

Kmita M, Fraudeau N, Herault Y, Duboule D. Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs. Nature (2002) 420: 145-50.

Kmita M, van Der Hoeven F, Zakany J, Krumlauf R, Duboule D. Mechanisms of Hox gene colinearity: transposition of the anterior Hoxb1 gene into the posterior HoxD complex. Genes Dev. (2000) 14: 198-211.

Kondo T, Duboule D. Breaking colinearity in the mouse HoxD complex. Cell (1999) 97: 407-17.

Kondo T, Zakany J, Duboule D. Control of colinearity in AbdB genes of the mouse HoxD complex. Mol. Cell (1998) 1: 289-300.

Kondo Y, Issa JP. Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. J. Biol. Chem. (2003) 278: 27658-62.

Kouzarides T. Histone acetylases and deacetylases in cell proliferation. Curr. Opin. Genet. Dev. (1999) 9: 40-8.

Kuo MH, Allis CD. Roles of histone acetyltransferases and deacetylases in gene regulation. Bioessays (1998) 20: 615-26.

Kuziora MA, McGinnis W. Different transcripts of the Drosophila Abd-B gene correlate with distinct genetic sub-functions. EMBO J. (1988) 7: 3233-44.

Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. (2002) 16: 2893-905.

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature (2001) 410:116-20.

LaJeunesse D, Shearn A. E(z): a polycomb group gene or a trithorax group gene? Development (1996) 122: 2189-97.

Langst G, Becker PB. ISWI induces nucleosome sliding on nicked DNA. Mol. Cell (2001) 8: 1085-92.

Lavorgna G, Sessa L, Guffanti A, Lassandro L, Casari G. AntiHunter: searching BLAST output for EST antisense transcripts. Bioinformatics (2004) 20: 583-5.

Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell (1993) 75: 843-54.

Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell (1992) 71: 865-73.

Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature (2003) 423: 255-60.

Lessard J, Schumacher A, Thorsteinsdottir U, van Lohuizen M, Magnuson T, Sauvageau

G. Functional antagonism of the Polycomb-Group genes eed and Bmi1 in hemopoietic cell proliferation. Genes Dev. (1999) 13: 2691-703.

Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P, Kingston RE. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. Mol. Cell. Biol. (2002) 22: 6070-8.

Lewis EB. A gene complex controlling segmentation in Drosophila. Nature (1978) 276: 565-70.

Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP. Vertebrate microRNA genes. Science (2003) 299: 1540.

Lipshitz HD, Peattie DA, Hogness DS. Novel transcripts from the Ultrabithorax domain of the bithorax complex. Genes Dev. (1987) 1: 307-22.

Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G. Correlation between histone lysine methylation and developmental changes at the chicken b–globin locus. Science (2001) 293:2453-55.

Liu XF, Olsson P, Wolfgang CD, Bera TK, Duray P, Lee B, Pastan I. PRAC: A novel small nuclear protein that is specifically expressed in human prostate and colon. Prostate (2001) 47: 125-31.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature (1997) 389: 251-60.

Mager J, Montgomery ND, de Villena FP, Magnuson T. Genome imprinting regulated by the mouse Polycomb group protein Eed. Nat. Genet. (2003) 33: 502-7.

Maison C, Bailly D, Peters AH, Quivy JP, Roche D, Taddei A, Lachner M, Jenuwein T, Almouzni G. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat. Genet. (2002) 30: 329-34.

Maniatis T, Fritsch EF, Sambrook J. Molecular Cloning: A Laboratory Manual (1982). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Mattick J. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep. (2001a) 2: 986-91.

Mattick JS, Gagen MJ. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol Biol Evol. (2001b) 18: 1611-30.

Mavilio F. Regulation of vertebrate homeobox-containing genes by morphogens. Eur J. Biochem. (1993) 212: 273-88.

Medina-Martinez O, Bradley A, Ramirez-Solis R. A large targeted deletion of Hoxb1-Hoxb9 produces a series of single-segment anterior homeotic transformations. Dev. Biol. (2000) 222: 71-83.

Meller VH, Gordadze PR, Park Y, Chu X, Stuckenholz C, Kelley RL, Kuroda MI. Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila. Curr. Biol. (2000) 10: 136-43

Meller VH, Rattner BP. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J. (2002) 21: 1084-91.

Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell

(1997) 88 :445-57.

Meneveri R, Agresti A, Della Valle G, Talarico D, Siccardi AG, Ginelli E. Identification of a human clustered G + C-rich DNA family of repeats (Sau3A family). J. Mol. Biol. (1985) 186: 483-9.

Mighell A.J., Markham A.F., Robinson P.A. Alu sequences. FEBS Lett. (1997) 417: 1-5.

Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, Hess JL. MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol. Cell (2002) 10: 1107-17.

Milne TA, Sinclair DA, Brock HW. The Additional sex combs gene of Drosophila is required for activation and repression of homeotic loci, and interacts specifically with Polycomb and super sex combs. Mol. Gen. Genet. (1999) 261: 753-61.

Molofsky AV, Pardal R, Iwashita T, Park IK, Clarke MF, Morrison SJ. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature (2003) 425: 962-7.

Momparler RL. Cancer epigenetics. Oncogene (2003) 22: 6479-83.

Muller J, Gaunt S, Lawrence PA. Function of the Polycomb protein is conserved in mice and flies. Development (1995) 121: 2847-52.

Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell (2002) 111: 197-208

Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, Dubois G, Mazo A, Croce CM, Canaani E. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol. Cell (2002) 10: 1119-28.

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SIS. Role of histone H3 lysine methylation in epigenetic control of heterochromatin assembly. Science (2001) 292:110-13.

Noma K, Allis CD, Grewal SIS. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science (2001) 293:1150-55.

Noonan FC, Goodfellow PJ, Staloch LJ, Mutch DG, Simon TC. Antisense transcripts at the EMX2 locus in human and mouse. Genomics. (2003) 81: 58-66.

Ogawa Y, Lee JT. Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol Cell. (2003) 11: 731-43.

Okazaki Y,et al. FANTOM Consortium; RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature (2002) 420: 563-73.

Olsson P, Motegi A, Bera TK, Lee B, Pastan I. PRAC2: a new gene expressed in human prostate and prostate cancer. Prostate (2003) 56:123-30.

Orlando V, Jane EP, Chinwalla V, Harte PJ, Paro R. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. EMBO J. (1998) 17: 5141-50.

Orlando V, Paro R. Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. Cell (1993) 75: 1187-98.

Orlando V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends Biochem. Sci. (2000) 25: 99-104.

Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. (2001) 35: 501-38.

Otte AP, Kwaks TH. Gene repression by Polycomb group protein complexes: a distinct complex for every occasion? Curr. Opin. Genet. Dev. (2003) 13: 448-54.

Papoulas O, Beek SJ, Moseley SL, McCallum CM, Sarte M, Shearn A, Tamkun JW. The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes. Development (1998) 125: 3955-66.

Park Y, Mengus G, Bai X, Kageyama Y, Meller VH, Becker PB, Kuroda MI. Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex. Mol Cell. (2003) 11: 977-86.

Paro R, Hogness DS. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc. Natl. Acad. Sci. USA (1991) 88: 263-7.

Pennisi E. Behind the scenes of gene expression. Science (2001) 293: 1064-7.

Peters AHFM, O'Carrol D, Scherthan H, Mechtler K, Sauer S, Schöfer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia M, Jenuwein T. Loss of the *Suv39h* histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell (2001) 107:323-37.

Petruk S, Sedkov Y, Smith S, Tillib S, Kraevski V, Nakamura T, Canaani E, Croce CM, Mazo A. Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. Science (2001) 294: 1331-4.

Pirrotta V. PcG complexes and chromatin silencing. Curr. Opin. Genet. Dev. (1997) 7: 249-58.

Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. (2002) 36: 233-78.

Poux S, Melfi R, Pirrotta V. Establishment of Polycomb silencing requires a transient interaction between PC and ESC. Genes Dev. (2001) 15: 2509-14.

Pradhan S, Bacolla A, Wells RD, Roberts RJ. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem. (1999) 274: 33002-10.

Rank G, Prestel M, Paro R. Transcription through intergenic chromosomal memory elements of the Drosophila bithorax complex correlates with an epigenetic switch. Mol. Cell. Biol. (2002) 22: 8026-34.

Rastelli L, Chan CS, Pirrotta V. Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on Enhancer of zeste function. EMBO J. (1993) 12: 1513-22.

Reik W, Walter J. Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. (2001) 2: 21-32.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature (2000) 403: 901-6.

Richards EJ, Elgin SCR. Epigentic codes for heterochromatin formation and silencing: Rounding up the usual suspects. Cell (2002) 108: 489-500.

Ringrose L, Rehmsmeier M, Dura JM, Paro R. Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. Dev. Cell. (2003) 5: 759-71.

Robertson KD, Wolffe AP. DNA methylation in health and disease. Nat. Rev. Genet. (2000) 1: 11-9.

Ross JM, Zarkower D. Polycomb group regulation of Hox gene expression in C. elegans. Dev. Cell (2003) 4: 891-901.

Roth SY, Denu JM, Allis CD. Histone acetyltransferases. Annu Rev Biochem. (2001) 70: 81-120.

Rudert F, Bronner S, Garnier JM, Dolle P. Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm. Genome (1995) 6: 76-83.

Satijn DP, Hamer KM, den Blaauwen J, Otte AP. The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in Xenopus embryos. Mol. Cell. Biol. (2001) 21: 1360-9.

Satijn DP, Olson DJ, van der Vlag J, Hamer KM, Lambrechts C, Masselink H, Gunster MJ, Sewalt RG, van Driel R, Otte AP. Interference with the expression of a novel human polycomb protein, hPc2, results in cellular transformation and apoptosis. Mol. Cell. Biol. (1997) 17: 6076-86.

Satijn DP, Otte AP. Polycomb group protein complexes: do different complexes regulate distinct target genes? Biochim. Biophys. Acta (1999) 1447: 1-16.

Saurin AJ, Shiels C, Williamson J, Satijn DP, Otte AP, Sheer D, Freemont PS. The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. J. Cell. Biol. (1998) 142: 887-98.

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J. (2002) 21:1121-31.

Schramke V, Allshire R. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science (2003) 301: 1069-74.

Schumacher A, Magnuson T. Murine Polycomb- and trithorax-group genes regulate homeotic pathways and beyond. Trends Genet. (1997) 13: 167-70.

Schwartz, S., Miller, W., Yang, C.M., Hardison, R.C. Software tools for analyzing pairwise alignments of long sequences. (1991) Nucleic Acids Res. 19: 4663-67.

Shiina, T., Tamiya, G., Oka, A., Takishima, N., Inoko, H. Genome sequencing analysis of the 1.8 Mb entire human MHC class I region. Immunol. Rev. (1999) 167: 193-9.

Sijen T, Plasterk RH. Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature (2003) 426: 310-4.

Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, Webster Z, Peters AH, Jenuwein T, Otte AP, Brockdorff N. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev. Cell. (2003) 4: 481-95.

Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature (2002) 415: 810-3.

Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. (1999) 9: 657-63.

Spencer VA, Davie JR. Role of covalent modifications of histones in regulating gene expression. Gene (1999) 240:1-12.

Spotswood HT, Turner BM. An increasingly complex code. J. Clin. Invest. (2002) 110: 577-82.

Sreenath T, Orosz A, Fujita K, Bieberich CJ. Androgen-independent expression of hoxb-13 in the mouse prostate. Prostate (1999) 41: 203-7.

Storz G. An expanding universe of noncoding RNAs. Science (2002) 296: 1260-3.

Strahl BD, Allis CD. The language of covalent histone modifications. Nature (2000) 403: 41-5.

Struhl K. Histone acetylation and transcriptional regulatory mechanisms. Genes Dev. (1998) 12:599-606.

Strutt H, Cavalli G, Paro R. Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. EMBO J. (1997) 16: 3621-32.

Sudarsanam P, Winston F. The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. Trends Genet.(2000) 16: 345-51.

Takagi N. Differentiation of X chromosomes in early female mouse embryos. Exp Cell. Res. (1974) 86: 127-35.

Tamkun JW, Deuring R, Scott MP, Kissinger M, Pattatucci AM, Kaufman TC, Kennison JA. brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell (1992) 68: 561-72.

Tate PH, Bird AP. Effects of DNA methylation on DNA-binding proteins and gene expression. Curr. Opin. Genet. Dev. (1993) 3: 226-31.

Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA. DNA methylation analysis: a powerful new tool for lung cancer diagnosis. Oncogene (2002) 21: 5450-61.

Turner BM. Cellular memory and the histone code. Cell (2002) 111:285-91.

Turner BM. Histone acetylation and an epigenetic code. BioEssays (2000) 22:936-45.

van der Vlag J, Otte AP. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. Nat. Genet. (1999) 23: 474-8.

van Leeuwen F, Gafken PR, Gottschling DE. Dotlp modulates silencing in yeast by methylation of the nucleosome. Cell (2002) 109:745-56.

van Lohuizen M. Functional analysis of mouse Polycomb group genes. Cell. Mol. Life Sci. (1998) 54: 71-9.

van Lohuizen M. The trithorax-group and polycomb-group chromatin modifiers: implications for disease. Curr. Opin. Genet. Dev. (1999) 9: 355-61.

Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, Rubin MA, Chinnaiyan AM. The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature (2002) 419: 624-9.

Venter JC, et al. The sequence of the human genome. Science (2001) 291: 1304-51. Erratum in: Science (2001) 292: 1838.

Volpe T, Schramke V, Hamilton GL, White SA, Teng G, Martienssen RA, Allshire RC. RNA interference is required for normal centromere function in fission yeast. Chromosome Res. (2003) 11: 137-46.

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen R. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science (2002) 297:1833-37.

Walsh CP, Bestor TH. Cytosine methylation and mammalian development. Genes Dev. (1999) 13: 26-34.

Wang J, Mager J, Chen Y, Schneider E, Cross JC, Nagy A, Magnuson T. Imprinted X inactivation maintained by a mouse Polycomb group gene. Nat. Genet. (2001) 28: 371-5.

Wang W, Cote J, Xue Y, Zhou S, Khavari PA, Biggar SR, Muchardt C, Kalpana GV, Goff SP, Yaniv M, Workman JL, Crabtree GR. Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. EMBO J. (1996) 15: 5370-82.

Weiner AM. SINEs and LINEs: the art of biting the hand that feeds you. Curr. Opin. Cell. Biol. (2002) 14: 343-50.

Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell (1993) 75: 855-62.

Wolf SF, Jolly DJ, Lunnen KD, Friedmann T, Migeon BR. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc Natl Acad Sci USA. (1984) 81: 2806-10.

Wu C, Morris JR. Genes, genetics, and epigenetics: a correspondence. Science (2001) 293: 1103-5.

Wutz A, Jaenisch R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell. (2000) 5: 695-705.

Xia ZB, Anderson M, Diaz MO, Zeleznik-Le NJ. MLL repression domain interacts with histone deacetylases, the polycomb group proteins HPC2 and BMI-1, and the corepressor C-terminal-binding protein. Proc. Natl. Acad. Sci. USA (2003) 100: 8342-7.

Yelin R, Dahary D, Sorek R, Levanon EY, Goldstein O, Shoshan A, Diber A, Biton S, Tamir Y, Khosravi R, Nemzer S, Pinner E, Walach S, Bernstein J, Savitsky K, Rotman G. Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. (2003) 21:379-86.

Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. (1997) 13:335-40.

Yu BD, Hanson RD, Hess JL, Horning SE, Korsmeyer SJ. MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc. Natl. Acad. Sci. USA (1998) 95: 10632-6.

Yu BD, Hanson RD, Hess JL, Horning SE, Korsmeyer SJ. MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc Natl Acad Sci USA. (1998) 95: 10632-6.

Yu BD, Hess JL, Horning SE, Brown GA, Korsmeyer SJ. Altered Hox expression and segmental identity in Mll-mutant mice. Nature (1995) 378: 505-8.

Zakany J, Duboule D. Hox genes in digit development and evolution. Cell. Tissue Res. (1999) 296: 19-25.

Zeleznik-Le NJ, Harden AM, Rowley JD. 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. Proc Natl Acad Sci USA. (1994) 91:10610-4.

Zeltser L, Desplan C, Heintz N. Hoxb-13: a new Hox gene in a distant region of the HOXB cluster maintains colinearity. Development (1996) 122: 2475-84.

Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev. (2001) 15: 2343-2360.