Coordinated Hoxd genes expression during embryogenesis is dependent on a strict disposition of genes and polarized regulators within the HoxD cluster

THÈSE N^{O} 8604 (2018)

PRÉSENTÉE LE 29 AOÛT 2018

À LA FACULTÉ DES SCIENCES DE LA VIE

UNITÉ DU PROF. DUBOULE

PROGRAMME DOCTORAL EN APPROCHES MOLÉCULAIRES DU VIVANT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Fabrice DARBELLAY

acceptée sur proposition du jury:

Prof. J. Lingner, président du jury
Prof. D. Duboule, directeur de thèse
Prof. A. Reymond, rapporteur
Prof. J. Deschamps, rapporteuse
Prof. D. Trono, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

SUMMARY

HOX transcription factors determine the identity of body regions along the rostro-caudal axis during bilaterian embryogenesis. In vertebrates *Hox* genes distinctively lie organized in dense clusters, each typically composed of a dozen paralogous transcription units spread over 100 kb of genomic DNA. In every *Hox* cluster the *Hox* genes are arranged following a complete 5'-3' transcriptional polarization. Even though *Hox* are found across the whole animal kingdom, this ordered physical arrangement distinctly characterizes vertebrate *Hox* clusters. This organization, and the attendant collinearity phenomena have been proposed to be a crucial feature to *Hox* genes' co-option in several structures organized along secondary body axes, as exemplified by the patterning of various tetrapod-specific structures (e.g. limbs, digits and external genitalia).

The aim of this thesis is to investigate to what extent Hox transcriptional polarization results from constraints imposed by their close linear proximity when arranged in clusters. Therefore, we challenged the murine *HoxD* cluster by engineering inversions inside the cluster using CRISPR/Cas9 mutagenesis approaches, in vivo. We produced two novel alleles bearing inversions of either *Hoxd11* or *Hoxd12*. Complementarily, we also reanalyzed extensively a larger targeted inversion encompassing *Hoxd11-d12* loci in a variety of embryonic organs and tissues. Together, the comparative analysis of these alleles by a combination of approaches (WISH, RNAseq, ChIPseq and 4Cseq) illuminated different types of outcomes resulting from the disruption of *HoxD* transcriptional polarity and the inversion of non-coding regulatory elements.

From our work, we developed the view that engineered inversions of particular Hox genes are tolerated by this genetic system and does not result systematically in major functional disruption of the affected cluster. This result therefore demonstrates that *Hox* transcriptional polarity is not a feature arising from an absolute developmental constraint. This conclusion is nevertheless nuanced by two instances where intra-*HoxD* inversion disrupts the regulation and function of the *HoxD* cluster. In one case, the inversion of *Hoxd11* affects the transcriptional output of its posterior neighbor *Hoxd12*. This perturbation is likely implying interference mechanism related to the transcriptional leakage emitted onto Hoxd12 by the inverted *Hoxd11* locus. In the second case, the inversion of the *Hoxd11-d12* loci in a *Hoxd13lacZ* background results in the specific gain-of-expression of this reporter gene in a variety of embryonic structures. Of interest, we present here diverse pieces of evidence that this dysregulation may not be directly related to the disruption of Hox gene transcriptional polarity but instead caused by the translocation and inversion of a CTCF binding site. This observation supports the hypothesis that CTCF dependent topological organization of the chromatin plays a role in the regulatory insulation of *Hoxd13*.

It appears therefore that spontaneous inversions within polarized *Hox* clusters might affect the function and regulation of the *Hox* patterning mechanism in different functional contexts. This conclusion leads us to conjecture that once disposed in dense clusters, the developmental expression of *Hox* genes is extremely sensitive to genetic perturbations, due notably to the presence of several conserved polarized regulatory elements interspersed between the individual *Hox* genes in cis.

Keywords: *Hox* genes organization, *HoxD* cluster, embryonic development, spatial collinearity, chromatin contacts and TADs, CTCF, posterior prevalence, metanephric kidneys

RÉSUMÉ

Lors de l'embryogenèse, les facteurs de transcription HOX déterminent l'identité des différents segments d'un animal bilatérien le long de l'axe rostro-caudal. Dans le cas particulier des gnathostomes, les gènes *Hox* sont arrangés en clusters, chacun composé d'une dizaine de paralogues disposés sur 100 kb de DNA génomique. Au sein d'un même cluster, les gènes *Hox* sont systématiquement orientés avec la même polarité transcriptionnelle 5'-3'. Bien que les gènes *Hox* soient conservés à travers l'entier du règne animal, cette organisation génétique est strictement exclusive aux gnathostomes. Cette organisation, extrêmement dense, est considérée comme ayant joué un rôle important pour la co-option des gènes *Hox* dans plusieurs structures secondaires spécifiques aux tétrapodes tels que les doigts et les organes génitaux externes.

Au cours de cette thèse, nous avons cherché à déterminer si l'organisation transcriptionnelle polarisée des gènes *Hox* répond à des contraintes inhérentes propres à leur proximité génétique. Pour cela, nous avons utilisé le système CRISPR/Cas9 pour générer deux allèles portant une inversion interne au cluster *HoxD* et comprenant un seul gène *Hoxd*, soit *Hoxd11* soit *Hoxd12*. En parallèle, nous avons également réexaminé une plus grande inversion comprenant l'entier du locus *Hoxd11-d12*. Ces trois inversions sont à l'origine de conséquences sur la régulation et la fonctionnalité du cluster *HoxD*, nous les avons étudiées à travers différents tissus embryonnaires murins.

Les résultats expérimentaux nous ont permis de comprendre que l'inversion d'un gène *Hox* au sein d'un cluster est possible et tolérée par l'ensemble du système génétique, sans occasionner de manière systématique un disfonctionnement. Ainsi, la polarisation transcriptionnelle telle qu'observée dans les clusters *Hox* des gnathostomes ne répond pas à une contrainte génétique absolue. Cependant, cette conclusion est nuancée par deux résultats illustrant les différentes perturbations qui peuvent être générées par de telles reconfigurations génétiques. D'une part, l'inversion du gène *Hoxd11* provoque une importante réduction quantitative de *Hoxd12*, par des mécanismes d'interférence qui restent à explorer mais qui semblent liés à l'existence d'un transcrit émis par le locus *Hoxd11* et couvrant le brin DNA antisens de *Hoxd12*. D'autre part, l'inversion du locus *Hoxd11-d12* dans une souche génétique *Hoxd13lacZ* cause un gain d'expression spécifique de ce reporteur dans différentes structures embryonnaires. Nous présentons des observations qui suggèrent que cette dérégulation n'est probablement pas uniquement causée par l'inversion de la direction de la transcription de ces deux gènes, mais pourrait résulter de la translocation et de l'inversion d'un site reconnu par le facteur CTCF. Cet élément semble en effet important pour organiser les contacts de la chromatine à ce locus et déterminer l'organisation topologique du cluster *HoxD* en isolant le gène *Hoxd13*, aux propriétés homéotiques particulières, des régulations spécifiques aux gènes *Hoxd* centraux.

En définitive, il apparaît donc qu'une inversion spontanée au sein d'un cluster *Hox* transcriptionnellement polarisé peut, à certaines conditions, affecter la fonction et la régulation des gènes qui le composent. Cette conclusion nous amène donc à proposer que les gènes *Hox*, une fois condensés en cluster, sont sensibles à des perturbations génétiques, notamment par la présence d'éléments régulateurs eux-mêmes polarisés.

Mots-clés : organisation et régulation des gènes *Hox*, *HoxD* cluster, développement embryonnaire, contacts de la chromatine et TADs, CTCF, prévalence postérieure, reins

It's a magical world, Hobbes, ol' buddy ... let's go exploring!

Bill Watterson

Un voyage se passe de motifs. Il ne tarde pas à prouver qu'il se suffit à lui-même.
On croit qu'on va faire un voyage, mais bientôt, c'est le voyage qui vous fait, ou vous défait.

Nicolas Bouvier, *L'usage du monde*

1 INTRODUCTION

1.1 Functions of *Hox* genes during development

Hox genes encode a family of transcription factors highly conserved throughout the bilaterian animals. These transcription factors are characterized by a helix-turn-helix DNA-binding domain (homeodomain). This homeodomain allows them to modulate the expression of target genes with co-factors (Merabet et al., 2009). Their names derive etymologically from the term homoeosis coined by Bateson in 1894 to describe genetic mutations affecting the identity and integrity of a segment (metamere) of an animal, often changed into the likeness of another. The spontaneous apparition of remarkable homeotic mutations in the model fly *Drosophila melanogaster* was of crucial importance for *Hox* genes discovery and initial description. Such homeotic mutations were notably studied and reported in this organism by Bridges and Morgan in 1923 and by Hollander in 1937 and were attributed to putative homeotic genes. However, the first *Hox* genes were genetically identified and mapped by Ed Lewis in a seminal work in *D. melanogaster* (Lewis, 1978). Homologous *Hox* genes were later found in the mouse genome, composed of a conserved homeodomain encoded by an homeobox (McGinnis et al., 1984; Scott and Weiner, 1984) and have been since then observed systematically throughout the animal kingdom (Garcia-Fernàndez, 2005).

As originally described in the fly, the main function of these transcriptional regulators is to establish segmental identity along the antero-posterior (AP) axis of the developing animal (Lewis, 1978). This function is conserved in vertebrates and more generally across the bilaterians (Krumlauf, 1994; Lemons and McGinnis, 2006; Mallo et al., 2010). Because of the important genetic networks that these transcriptional factors control and influence during the embryonic development, they have been proposed to be one of the major drivers of morphological variety in the animal kingdom (Pearson et al., 2005; Wagner et al., 2003). In vertebrates, their functions are essential for the proper embryonic development of the animal, their regulations being directly coupled with the elongation of the vertebrate embryo. Indeed, *Hox* genes are initially activated in the early phase of the gastrulation by a series of morphogenic factors, such as Wnt/β-catenin or Oct4 (Denans et al., 2015; Deschamps and Wijgerde, 1993; Mallo, 2017; Neijts and Deschamps, 2017). *Hox* genes are then progressively transcribed in the primitive node region, following a very particular dynamic that will be discussed later in this introduction. Secondary regulations implying important morphogenic pathways and factors, frequently associated with the somitic segmentation process such as e.g. retinoic acid, FGF and Notch, but also *Cdx* and *Gdf11* further modulate and lock the expression patterns of *Hox* genes (Deschamps and van Nes, 2005; Dubrulle and Pourquié, 2004; Neijts et al., 2017; Zákány et al., 2001). This progressive activation of *Hox* genes results in the regionalized patterning of the segments of the embryo generated during its axial elongation and composed of mesoderm (paraxial, intermediate and lateral plate) and neuroectoderm.

Hox genes are found expressed in a large variety of tissues and derivatives, from the neurectoderm (hindbrain and spinal cord) to the somitic (axial and paraxial) mesoderm (Carpenter, 2002; Wellik, 2007; Zakany and Duboule, 2007). In these structures, HOX factors are orchestrating the morphogenesis of the developing animal by controlling, together with co-factors, large series of genetic pathways (Foronda et al., 2009; Pearson et al., 2005). These pathways are diverse and frequently interdependent. Their functionality remaining yet to be fully deciphered and rationalized, notably in the challenging context of embryonic development. However, a global picture of *Hox* function is apparent and resolves around three principal axis: control of cell death, control of cell proliferation and control of cell positional identity (Foronda et al., 2009; Hueber and Lohmann, 2008; Pearson et al., 2005).

This system has been extensively co-opted by vertebrates, especially by the tetrapods, to regulate the morphogenesis of several distinctive secondary structures (Soshnikova et al., 2013; Wagner et al., 2003). The limbs and the digits, the external genitals and the metanephric kidneys are appropriate examples illustrating the diverse HOX functions and roles in the global morphogenesis of the developing tetrapod embryo (Dollé et al., 1991; Wellik et al., 2002; Zakany and Duboule, 2007). Therefore, genetic perturbations of *Hox* genes in murine models, i.e. loss-of-functions or gain-of-functions, result in a series of complex phenotypes illustrating their importance and the pleiotropy of their functions.

A representative example of this statement is the complete deletion of the *Hox11* genes from the mouse genome (Wellik et al., 2002). In this lethal genetic condition, mutant mice display a cumulative variety of defects. Firstly, they suffer from a complete renal agenesis, *Hox11* being crucial to control the branching of the ureteric bud in the developing metanephric kidney as well as the proper induction of this organ (Patterson et al., 2001; Wellik et al., 2002). Secondly, these mutants also display a very damaged skeletal anatomy of the limb zeugopod, i.e. the absence of the ulna and the radius (Davis et al., 1995; Wellik and Capecchi, 2003). Lastly, in contrast to these region-specific malformations, the loss-of-function of *Hox11* genes drives profound homeotic phenotypes along the axial skeleton. *Hox11* genes are indeed normally expressed at the lumbosacral transition, therefore the loss of *Hox11* paralog group results in the homeotic transformation of this region, adopting a lumbar morphology (Rux and Wellik, 2017; Wellik and Capecchi, 2003). Following the same fashion, the loss of *Hox10* genes results in the morphologic transformation of the lumbar and sacral vertebrae to rib-bearing, thoracic-like vertebrae (Rux and Wellik, 2017; Wellik and Capecchi, 2003). The series of genetic mutations produced in murine models to address the functionality of HOX proteins led to the development of two theoretical frameworks to explain how *Hox* genes exert their proper homeotic function. On the one hand, the "*Hox code*" model postulates that the combination of HOX proteins present in each single cell of the developing animal determines its identity by generating a unique transcriptional output (Kessel and Gruss, 1990). In other words, each combination of HOX factors would result in a unique and specific transcriptional response. On the other hand, the "*posterior prevalence*" model proposes that the most posterior *Hox* gene expressed in a given metamere, or cell, suppresses genetically the function of more anterior *Hox* genes (Duboule, 1991; Duboule and Morata, 1994). In view of this last model, one single HOX would therefore control the identity of a cell and not the qualitative or quantitative combinations of HOX factors found within this cell. These two models will be challenged later in this work.

For practical reasons, the functions of *Hox* genes in adult vertebrate animals have been far less studied than during embryogenesis. However, scarce studies reveal that *Hox* genes expression is maintained throughout adulthood in different organs and tissues. *Hox* genes are expressed in precise nested territories of the adult mouse hindbrain, accurate to their initial embryonic pattern (Hutlet et al., 2016). *Hox* genes are also reported to be required to control the proliferation of adult mouse hematopoietic stem and progenitor cells (Lebert-Ghali et al., 2015), and are continuously expressed in mesenchymal stem/stromal cells (MSCs) in the adult skeleton, where they probably participate in correct bone repair and healing (Rux and Wellik, 2017). Finally, during this thesis, we observed a very specific and controlled expression pattern of central *Hoxd* genes in murine adult kidneys, where these genes may participate in maintaining the integrity of the nephrons (Romagnani et al., 2013). Lastly, *Hox* genes are frequently found expressed at various dose in a large variety of human cancers, reflecting their potential role in tumorigenesis (Bhatlekar et al., 2014; Morgan, 2006). All together, we conclude this overview of *Hox* functions by emphasizing the fact that these transcription factors are of the utmost importance. Their homeotic functions, mediated by their ability to modulate various genetic networks, must be therefore precisely controlled in space and time.

1.2 Structure of *Hox* complexes

1.2.1 *Hox* genes are found in clusters

In the *Drosophila melanogaster* genome, eight *Hox* genes are arranged in two complexes (Fig. 1). These complexes are positioned at two independent loci and are named after some famous homeotic transformations which they display when mutated: the *Antennapedia* (ANT-C, size of 340 kb) and the *Bithorax* complexes (BX-C, size of 315 kb) (Kaufman et al., 1980; Lewis, 1978). Surprisingly, *Hox* genes were found arranged in an even tighter and denser organization in higher vertebrate genomes (Fig. 1). In mouse and human genomes, *Hox* genes are indeed disposed in four dense gene clusters, each cluster in the range of 100 kb of total size (Boncinelli et al., 1988; Duboule and Dollé, 1989; Graham et al., 1989; Krumlauf, 1994). These four clusters, *HoxA* to *HoxD*, incorporate the totality of the 39 *Hox* genes found in mammals, and are composed of an incomplete series of the 13 paralogous *Hox* (*Hox1* to *Hox13*), identified by their sequence similarity and relative position within the cluster (Scott, 1992, 1993).

This particular genetic distribution of *Hox* genes in four independent clusters, each composed of an incomplete series of members of the 13 paralogous groups, is resulting from the two rounds of genome duplication that occurred in the early lineage of the vertebrates (Holland et al., 1994; Putnam et al., 2008). Of interest, not a single mammalian *Hox* cluster displays a complete catalogue of paralogous *Hox* genes, from *Hox1* to *Hox13*, thus illustrating the different outcomes for each copy of the duplicated genes or group of genes, following a duplication event and the momentary relief of evolutionary constrains (Holland, 1999; Ohno, 1970). These duplications resulted in an apparent redundancy and interdependence of some *Hox* patterning functions, shared between sister-clusters. A relevant example is the well-studied expression of *Hoxa* and *Hoxd* genes in the developing limbs where the share similar functions and regulations (Andrey et al., 2013; Beccari et al., 2016; Berlivet et al., 2013; Montavon and Duboule, 2012; Spitz et al., 2003; Woltering et al., 2014).

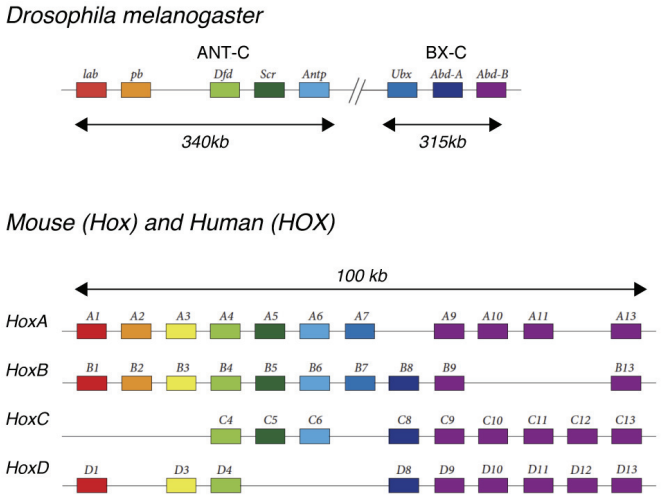

Figure 1: *Hox* genes arrangement in *D. melanogaster* and in mammalian genomes. In *D. melanogaster*, eight *Hox* genes are grouped in two large independent complexes: the Antennapedia (ANT-C) and Bithorax (BX-C) complexes. On the contrary, the 39 mammalian Hox genes are disposed in four tight independent clusters: *HoxA*, *HoxB*, *HoxC* and *HoxD* in the case of *Mus musculus*. Each *Hox* cluster is of approximately 100 kb of size and composed of an incomplete series of the 13 paralogous *Hox* (*Hox1* to *Hox13*). Figure adapted from Taniguchi, 2014.

1.2.2 Evolution of *Hox* gene clusters

Vertebrate *Hox* genes are found arranged in dense and compact gene clusters. These clusters are typically of a size of 100-120 kb (Fig. 2) and are constituted exclusively of *Hox* genes, devoid of any repeat elements (Simons et al., 2006, 2007). Moreover, these *organized* clusters display an additional level of genetic organization. Indeed, in each individual cluster, *Hox* transcription units are found systematically encoded by the same DNA strand, each gene orientated with the same 5'-3' polarity.

The sequencing and reconstruction of genomes belonging to many different groups of animals have shown that *Hox* genes are not always found into such dense clusters (Darbellay and Duboule, 2016; Duboule, 2007; Lemons and McGinnis, 2006). Three different classes of organization can be distinguished, outside of the vertebrate prototypic *organized Hox* cluster already discussed (Fig. 2). Firstly, some bilaterians carry *disorganized Hox* clusters composed of *Hox* genes spanned over large genetic intervals, typically from 400 kb to 1 mb in size. These loci are also frequently entangled with other non-*Hox* genes and repeat elements whereas the global transcriptional directionality is not strictly maintained. This situation is typically found in the amphioxus (cephalochordate) genome (Amemiya et al., 2008; Putnam et al., 2008). Secondly, the *split* organizational configuration is frequently observed across protostomian genomes. In this configuration, *Hox* genes are separated in two sub-groups of unequal size distributed at two independent loci. Each sub-cluster may show different level of organization. The *Hox* genes of *D. melanogaster* which are arranged in two independent and large complexes are relevant illustration of this category (Kaufman et al., 1980; Lewis, 1978). Thirdly, a last category of *Hox* gene organization has been observed in different species of bilaterian groups, such as the urochordates and different protostomians. The *atomized* configuration represents the absolute 'no-cluster' situation, in which *Hox* genes are found scattered over different genomic loci. The genome of the urochordate *Ciona intestinalis* is a representative illustration of such a disposition of *Hox* genes (Ikuta et al., 2004).

The distribution of the four different classes of *Hox* genes structural organization across the bilaterian phylogenic tree reveals a global heterogeneity, not recapitulated by the phylogenic relationships existing between these groups of animals (Fig. 3). First of all, this heterogeneity exemplifies to what extend the vertebrate-specific *organized* structure is indeed not required for the development of many group of animals. This observation infers that for a large and important fraction of bilaterian animals, the primordial function of *Hox* genes, i.e. the patterning of the AP axis does not require an *organized Hox* cluster. A second important point that arise from considering the distribution of *Hox* organizations across bilaterians is the specificity of this *organized* configuration. Such an arrangement is indeed unique to the jawed vertebrates. As of today, not a single genome obtained from any other bilaterian has shown a level of *Hox* cluster compaction similar to the one observed in gnathostome clusters (100-120 kb of size). Conversely, not a single jaw vertebrate genome has revealed a *Hox* cluster resembling the large *disorganized* structures found in echinoderm genomes (600 kb of size, Cameron et al., 2006).

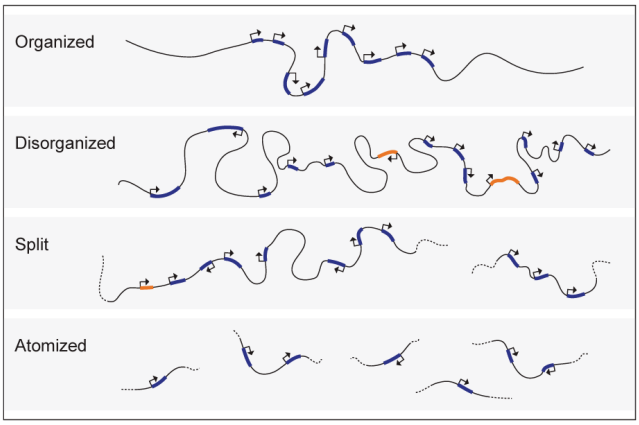

Figure 2: Classification of *Hox* genes structures observed across the animal kingdom (see Duboule, 2007). *Organized Hox* clusters are tight genetic structures where *Hox* genes are encoded by the same DNA strand. *Organized* clusters are devoid of non-*Hox* protein coding genes, they contain few repeats but may include lincRNAs and miRNAs. To this date, vertebrate clusters are the unique known example of this dense organization. In contrast, the *disorganized* clusters are large continuous structures, typically >300 kb of size. The genetic distance between each *Hox* gene is therefore larger and the density of *Hox* genes smaller compared against the *organized* clusters. Moreover, *disorganized* cluster may contain entangled non-*Hox* genes and display heterogeneity in *Hox* genes 5'-3' transcriptional direction. The *Hox* cluster of the echinoderm *Strongylocentrotus purpuratus* illustrates this type of organization (Cameron et al., 2006). *Split* clusters are *disorganized* clusters separated into two distinct and independent complexes. The disposition of *D. melanogaster Hox* genes into the notorious ANT-C/BX-C complexes, separated by 10 mb of chromatin, is a relevant illustration of a *Split* organization. *Atomized* clusters represent the quintessence of *Hox* "no-cluster" situation, where *Hox* genes are found disseminated across the whole genome of the considered animal, without any continuous *cis* organization. This *atomized* disposition of *Hox* genes is observed in the genome of the tunicate *Oikopleura dioica* (Seo et al., 2004). Orange boxes: non-*Hox* coding genes, blue boxes: *Hox* genes, the arrows represent the 5'-3' transcription orientation of each gene. Figure adapted from Darbellay and Duboule, 2016.

This observation raises an important point regarding the evolution of *Hox* structures. There are indeed two possible hypotheses to explain the distinctiveness of gnathostome *organized Hox* clusters. Either gnathostomes are the sole representative of the primordial *Hox* organization, already found in the last common ancestor at the root of the bilaterians lineage, or, alternatively, the gnathostome *Hox* clusters result from an active process of *Hox genes* compaction from a *disorganized* into an *organized* configuration. There are however no valid reasons to consider that gnathostomes are the last and unique representative of the ancestral bilaterian *Hox* cluster organization. Such an hypothesis would indeed require that all the metazoans would have systematically and extensively disorganized, split or atomized their own *organized Hox* clusters, whereas the gnathostomes would have carefully carried an ancestral *organized Hox* cluster directly, inherited directly from *urbilateria* (Duboule, 2007).

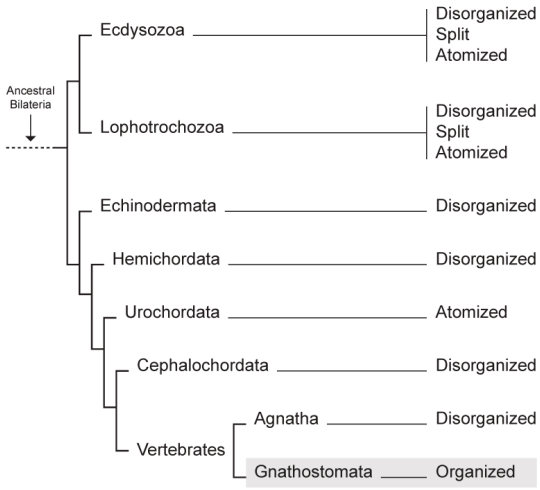

Figure 3: Distribution of the four categories of *Hox* structures among the principal groups of bilaterians (as described in Duboule, 2007). *Disorganized*, *split* and *atomized Hox* clusters are observed throughout different protostomian taxa. In contrast, deuterostomian animals display a lower heterogeneity in *Hox* cluster organization, the predominant category being the large but continuous *disorganized* arrangement. Two important exceptions are the atomized *Hox* genes reported in urochordates (e.g. *O. dioica*) and the tight *organized Hox* cluster exclusively and constantly found in gnathostomes (jaw vertebrates) genomes. This distribution of *Hox* organizations suggest either that the gnathostomes are the only bilaterians to have conserved intact a primordial *organized* cluster, already present in the ancestral *urbilaterian*, or that a process of consolidation occurred specifically during the radiation of the gnathostomes, actively condensing *Hox* clusters from a *disorganized* to an *organized* structure. Figure adapted from Darbellay and Duboule, 2016.

One shall therefore adopt the opposed hypothesis, which is that an ancestral *disorganized* complete cluster, i.e. containing *Hox1* to *Hox13*, existed at the root of the bilaterians, initially constructed by series of tandem duplications (Gehring et al., 2009). This hypothesis might appear counter-intuitive at first, as the idea that an increase of (local) genetic *organization* through time is antagonist to the perception of evolution as a mechanism that leads to an increase of disorganization, reminiscent of the principle of entropy (Darbellay and Duboule, 2016; Duboule, 2007). This ancestral cluster might have been structurally similar to the *disorganized* cluster carried by the cephalochordate *Amphioxus* (Amemiya et al., 2008; Gehring et al., 2009; Putnam et al., 2008). This *disorganized* structure would have then diverged extensively throughout the metazoan expansion and radiations, including rounds of genome duplication, to finally produce the current distribution of *Hox* gene. This scenario would thus imply that *Hox* clusters have been *organized*, further compacted, during the evolution of the gnathostome lineage as well as losing some redundant *Hox* genes following the R1/R2 duplications events (Holland, 1999; Soshnikova et al., 2013). The mechanisms that have been proposed to consolidate and condensate *Hox* genes in dense *organized* clusters are linked with the regulations of these genes.

1.3 *Hox* genes regulation

1.3.1 Spatial collinearity (SC)

The notion of spatial collinearity is tightly associated to the primordial function of HOX factors in the patterning the AP axis of the developing animal. This concept evolved from the original ground-breaking paper of Ed Lewis (Lewis, 1978) who reported an association between domains of *Hox* genes expression along the AP axis of the *Drosophila* larvae and the relative position of the *Hox* genes within BX-C, before being consequently extended to ANT-C (Bender et al., 1983; Harding et al., 1985; Lewis, 1978; Sánchez-Herrero et al., 1985). Surprisingly, a similar observation was then made in murine embryos, where the relative position of *Hox* genes within each cluster was linked with their respective expression patterns along the trunk of the developing embryo (see Akam, 1989; Duboule and Dollé, 1989; Gaunt, 1988; Graham et al., 1989). Vertebrate embryos are patterned by series of partially overlapping domains of *Hox* gene expression. The relative position of those genes along their clusters (from *Hox1* to *Hox13*) corresponds to the position of their transcription domains along the AP axis of the trunk (Deschamps and Duboule, 2017; Deschamps and van Nes, 2005; Gaunt and Gaunt, 2016; Zakany and Duboule, 2007). For instance, the anterior boundary of the expression domain of *Hox3*, which marks the developing cervical vertebrae and diverse anterior embryonic structures is situated more anteriorly along the AP axis than the expression boundary of *Hox8*, which in turn patterns the developing thoracic vertebrae. The same observation holds true with *Hox8* expression pattern which is located more anteriorly than the zone of expression of *Hox11*, which is typically found expressed at the level of the developing lumbar vertebrae (Fig. 4A). This spatial feature is common across the different paralog *Hox* clusters and is observed in all model vertebrates.

Several regulative mechanisms have been proposed to account for vertebrate spatial collinearity. For examples shared enhancers and local regulative relays may act synchronously over several *Hox* genes located in dense *organized* cluster, their activity progressing gradually across the cluster from *Hox1* to *Hox13*. However, this remarkable and aesthetic correspondence between *Hox* genetic positions within their clusters and *Hox* patterns along the AP axis does not seem to be required to obtain a series of nested HOX pattern along the AP axis of animals. Indeed, the spatial *sequential* expression of *Hox* genes is widely observed throughout metazoans, independently of the manner their *Hox* genes are organized (or disorganized) in their genomes (Duboule, 2007; Lemons and McGinnis, 2006). Two examples support this observation. In a first instance, *D. melanogaster Hox* genes are grouped into two complexes, ANT-C and BX-C. These two *Hox* complexes are separated by more than 10 mb of distance, but this configuration is obviously not an obstacle to spatial collinearity. In a latter instance, *Oikopleura dioica* exhibits a spatially synchronized expression of *Hox* genes along its AP axis during its early development. Yet, this tunicate displays a completely *atomized* distribution of *Hox* genes without any apparent order nor continuous organization of its *Hox* genes, spread over several genetically unconnected loci (Seo et al., 2004).

Thus, although the densely *organized Hox* structures found in gnathostomes' genomes present some interesting characters in term of global and shared regulations, it appears that the sequential patterning of HOX along the AP axis does not rely upon this specific disposition of *Hox* genes to operate. This observation is in full agreement with the numerous results obtained these last 30 years by murine transgenic approaches. By extensively evaluating the capability of randomly integrated fraction of *Hox* genes to recapitulate their expression pattern, the conclusion emerged that cluster organization is indeed dispensable for establishing the majority of the expected rostro-caudal expression boundaries, at least within a certain spatial window (Krumlauf, 1994; Tschopp et al., 2012). Therefore, the concept of

collinearity should be considered within the context of time to be further appreciated and to fully reveal the unique genetic specificity of the *organized Hox* cluster.

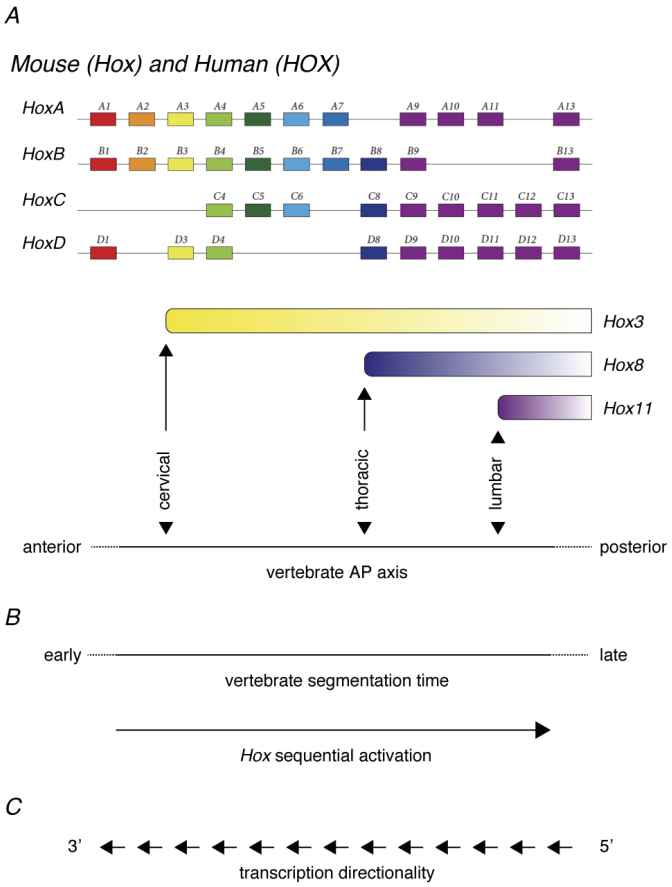

Figure 4: Spatio-temporal collinearity of *Hox* genes in mammals. (*A*) The relative positions of *Hox* genes in *organized* clusters correlate with the limit of their respective domains of expression along the AP axis of the developing embryo. *Hox3* genes are marking the identity of metameres positioned more anteriorly than *Hox8* (i.e. cervical versus thoracic vertebrae). In turn, *Hox8* genes are expressed more anteriorly than *Hox11* genes, which are typically marking the metameres at the level of the lumbar vertebrae of the future adult animal. The association between the respective position of *Hox* genes along the chromatin and the position of their domains of expression along the AP axis is called spatial collinearity. (*B*) The activation of *Hox* genes is progressive and coupled with a temporal dimension. *Hox3* genes are systematically expressed earlier during the development of the embryo than *Hox8* genes, which are in turn activated earlier than *Hox11* genes. This whole process is well synchronized among the *Hox* cluster and is called temporal collinearity. Temporal collinearity leads to the complete patterning of the AP axis of the murine embryo by E10. Of note, neither spatial nor temporal collinearity are exclusive to the *organized Hox* clusters found characteristically in mammals. (*C*) In *organized Hox* clusters, the progressive activation of *Hox* genes from the "anterior" to the "posterior" extremity is directed against *Hox* 5'-3' transcriptional orientation. The 3' extremity of an *Hox* clusters corresponds therefore to its "anterior" extremity. *Organized Hox* clusters are always displaying such a complete and absolute transcriptional polarization, suggesting that a ratchet-like mechanism may be at play during their progressive activation (Soshnikova and Duboule, 2009). Figure adapted from Taniguchi, 2014.

1.3.2 Temporal collinearity (TC)

Spatial collinearity, as described in murine embryos, is encompassed within a fundamental temporal dimension. This notion was described in murine embryo shortly after the observation of *Hox* spatial collinearity in vertebrates (see Dollé et al., 1989; Gaunt and Strachan, 1994; Izpisúa-Belmonte et al., 1991)**.** Temporal collinearity states that the construction of the final spatially collinear nested *Hox* patterns ("*Russian doll"*) is a progressive process which is set by the successive activations of *Hox* genes from the "anterior" extremity of the cluster to the "posterior" end (see references in Deschamps and Duboule, 2017). To reconsider the examples presented in the previous sections of this introduction, temporal collinearity results in the activation of the *Hox3* genes earlier, during the embryonic life, than the central *Hox8* genes, which in turn are activated earlier than the posterior *Hox11* genes (Fig. 4B). This spatio-temporal regulation is common to the different paralog *Hox* clusters known in model vertebrate genomes, independently of their composition in *Hox* genes paralogs.

Hox spatio-temporal collinearity is set within a very determined time frame in the murine embryo, from E7 to E10.5, and occurs in close relationships with the segmentation mechanisms used by the vertebrate embryo. The synchronization of the *Hox* genes activation within this developmental strategy is assured by a pool of mesoderm progenitors (MP) and stem cell-like neuromesodermal progenitors (NMPs) which are found originally in the primitive node region and then, at later embryonic stages, in the tail bud (Deschamps and Duboule, 2017; Henrique et al., 2015). This pool of NMPs/MPs express gradually more posterior *Hox* genes and is directly contributing in the elongating AP axis, feeding notably the pre-somitic mesoderm and is therefore participating in the somitogenesis process. The expression of *Hox* genes in these mesodermal and neuroectodermal precursors are set and memorized at the time of their differentiation, the descendant structures marked by a fixed combination of *Hox* genes (Fig. 5). The precise pace of *Hox* genes activation in these NMPs/MPs stem cell-like progenitors, starting from the anterior *Hox* genes and progressing in a processive manner towards the posterior *Hox* genes, is therefore of the utmost importance to maintain a synchronization with the construction of the AP axis. *Hox* initial activation is mediated by endogenous factors such as Wnt3/Wnt3a, followed by the gradual opening of the central fraction of *Hox* cluster in the NMPs/MPs cells in response to Cdx factors (Neijts et al., 2016, 2017). Subsequent to the Cdx regulation, *Gdf11*, a TGFβ signal released by posterior embryonic tissues after E8.0 is also participating in the progressive activation of *Hox* genes. This activation is principally centered at the level of the *Hox11* genes and involves *cis* enhancers scattered within the *Hox* clusters (Gaunt et al., 2013). Finally, the activation of the posterior *Hox13* is in turn antagonizing the transcription of *Cdx2*, resulting in the progressive repression of the *Hox* locus coupled with the termination of the elongation process (Amin et al., 2016; Young et al., 2009).

As a direct result of the progressive activation that traverses *Hox* clusters, the position of each *Hox* gene within the cluster determines its timing of activation in a distance-dependent manner from the 3' extremity. Therefore, changing *Hox* genes respective positions by deleting or duplicating single or multiple *Hox* results in alterations of their expression patterns along the AP axis, being either anteriorized or posteriorized (Tschopp et al., 2009). Yet, the position of *Hox* genes within the cluster is not the only factor determining *Hox* rostral limits of expression. The presence of *cis* enhancers dispersed within *Hox* clusters and responding to activating factors such as *Gdf11* or *Cdx2* work as additional regulators. These elements can be therefore disconnected from their endogenous clusters and identified in transgenic context as they are able to reproduce the expression patterns of their endogenous neighboring genes although with some temporal shift (Gérard et al., 1993; Hérault et al., 1998; Renucci et al., 1992; Tschopp et al., 2012).

Because *Hox* progressive activation is restricted to NMPs/MPs, the distribution of *Hox* in the descendant tissues is therefore mirroring the status of *Hox* activation at the time the differentiation occurred (Deschamps and Duboule, 2017). The memorization and the robust transmission of this information, i.e. where *Hox* activation boundary lies in a given segment along the AP axis, is assured by various mechanisms that must be stable across time and cell divisions. Specific features of the chromatin at the locus, such as histones marks and chromatin structures participate in fixing a stable boundary between the active and the inactive fractions of the cluster. Some of these modifications are characteristically mediated by the antagonist Polycomb/trithorax-group proteins (reviewed in Steffen and Ringrose, 2014). In this view, the collinear and continuous disposition of *Hox* genes in an organized *cluster* can also be viewed as an advantage to minimize the interphase between active and inactive genes, limiting as much as possible the situation where transcriptional activity might be entangled (Gaunt, 2015).

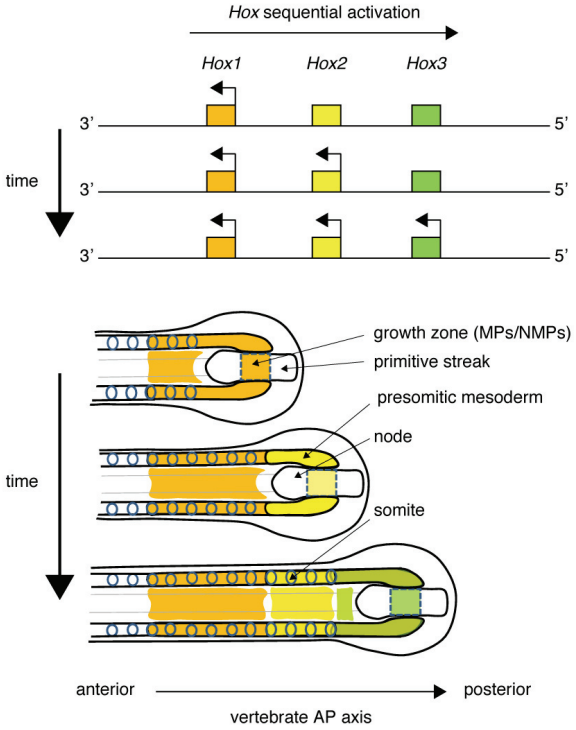

Figure 5: Temporal collinearity insures the synchronization of *Hox* sequential activation with the progressive elongation of the AP axis. Three time-points are represented here to illustrate the temporal collinear expression of three *Hox* genes: *Hox1*, *Hox2* and *Hox3*. In a first instance, *Hox1* is transcribed in the mesoderm progenitors (MP) and neuromesodermal progenitors (MNPs), two populations of stem cell-like present in the anterior region of the streak (growth zone) and feeding the elongating axis of the vertebrate embryo with mesoderm and neurectoderm tissues. In a second time, the activation of the gene *Hox2* in this pool of self-renewing MPs/NMPs marks the metameres with a new set of HOX factors (HOX1 and HOX2). Lastly, following the same logic, the activation of *Hox3* in the MPs/MNPs precursors pattern the next embryonic segment (mesoderm and neuroectoderm) with yet another set of HOX proteins (HOX1, HOX2 and HOX3). The figure is simplified in that presomitic/somitic mesoderm originates from more than one growth zone (Deschamps and Duboule, 2017; Henrique et al., 2015). The offset between the domains of expression of the *Hox* in the neuroectoderm and in the mesoderm derivates is here voluntarily not represented. This figure is adapted from Gaunt, 2015.

At this point however, it is pertinent to note that *Hox* spatio-temporal collinearity is not exclusively observed in vertebrates. It is a process that can be displayed more or less extensively by a variety of metazoans, whenever their segmentation strategy integrate the dimension of time in the progressive elongation of their AP axis, such as the short germband insects or the crustaceans (Duboule, 1992, 1994, 2007). Of note, the particular prerequisite observed in these various metazoan species using *Hox* spatio-temporal collinearity is the disposition of *Hox* genes in continuous structures, either of the *disorganized* or *organized* cluster type (Duboule, 1994, 2007; Garcia-Fernàndez, 2005; Schiemann et al., 2017).

Contrariwise, animals relying on other segmentation strategies, disconnected from the temporal dimension, seem to be free to distribute their *Hox* genes in *atomized*/*split* arrays, organization probably more adapted to their own regulative needs. Such is for instance the situation of *D. melanogaster* where, on the one hand, the identity of the segments of the fly are fixed simultaneously at once and early during the embryogenesis, without neither any AP axis extension nor somitogenesis. On the other hand, *D. melanogaster Hox* genes are split in two independent structures, the famous ANT-C/BX-C complexes (Duboule, 1992, 1994; Schiemann et al., 2017). It thus appears that temporal collinearity is observed independently of the degree of compaction of the *Hox* clusters, requiring simply a continuous genetic disposition of *Hox* to occur. Therefore, this regulative feature cannot explain *per se* the distinctive organization of *Hox* genes in dense clusters in vertebrate genomes.

1.3.3 Regulatory landscapes

From what precedes, we conclude that the disposition of *Hox* genes in a continuous locus is closely connected to the spatio-temporal collinear developmental strategy. The adaptive values of such a continuous genetic disposition are easy to conceptualize; series of shared enhancers could act synergistically and in a combinatorial manner over *Hox* genes, located advantageously in dense clusters (Deschamps and van Nes, 2005; Duboule, 1994; Kmita and Duboule, 2003). Multiplying these long-range enhancers may have allowed for a maximized transcriptional response to different inputs signals (Spitz and Furlong, 2012). Numerous experimental evidences have illustrated the regulatory dependence of murine *Hox* clusters towards series of long-range enhancers located in their genetic neighborhoods (Andrey et al., 2013; Berlivet et al., 2013; Delpretti et al., 2013; Lonfat et al., 2014; Montavon et al., 2011; Spitz et al., 2005). These regulatory landscapes are typically composed by partially redundant long-range enhancers, amplifying together the complex and pleiotropic regulations of these key developmental genes. For example, the contribution of a series of long-range enhancers is predominant in the regulation of *Hoxd13* in the developing limb autopod during the secondary phase of expression of *Hox* genes. In this context, the genetic disconnection of *Hoxd13* from its regulatory landscape leads to a complete loss-of-expression of *Hoxd13* in the developing autopods, resulting in an identical phenotype as the direct loss-of-function of the same gene (Dollé et al., 1993; Montavon et al., 2011; Tschopp and Duboule, 2011).

Nonetheless, in this respect *Hox* clusters do not constitute an exception. Numerous other murine developmental genes are indeed controlled by such large and complex regulatory landscapes; loci as diverse as *Shh* (Amano et al., 2009), *Arx* (Dickel et al., 2018), *Myc* (Uslu et al., 2014) or *Pitx1* (Guillaume Andrey, personal communication) have been reported to be under the control of clusters of long-range enhancers. Similar observations were also made when considering groups of developmental genes expressed in a given embryonic structure, such as the developing limbs (Andrey et al., 2017; Osterwalder et al., 2018). It is relevant to note that the disposition and organization of the majority of regulatory landscapes is frequently encompassed within the global three-dimensional organization of the chromatin in topologically associating domains (TADs, see in Dixon et al., 2012; Nora et al., 2012). This inclusion suggests potential functions for TADs as units of coordinated gene regulation and/or as a resource to

insulate regulations and buffer cross-activations between specific genes and long-range enhancers (Andrey and Mundlos, 2017; Symmons et al., 2014). TADs also correlate with lamina-associated domains (LADs) and DNA replication domains and may thus be considered as complete units of chromosome organization (Gonzalez-Sandoval and Gasser, 2016).

Here again, the case of *Hox* clusters is not different from what is observed at several developmental genes loci, as *Hox* regulatory landscapes are also contained within TADs. The causality of such an organization remains yet to be fully established, i.e. to what degree and by which functional means the specific interactions between enhancers and target genes are constraining TADs organization, but in view of recent evidences it appears clearly that these two landscapes are functionally linked, notably at the *HoxD* cluster (Rodríguez-Carballo et al., 2017). This simultaneous superposition of regulatory landscapes and topological landscapes, i.e. domains of chromatin displaying preferential physical interactions, is also assumed to lead to the appearance of genomic niches. These loci catalyze the development of new pleiotropic regulations, controlled by partially redundant enhancers synchronously regulating sub-group of *Hox* genes (Darbellay and Duboule, 2016). In turn, these regulations, occasionally beneficial and probably very frequently disadvantageous and not conserved, would create opportunities for *Hox* genes co-options in various embryonic structures (Lonfat and Duboule, 2015; Wagner et al., 2003). Such a proposal is supported by several examples of shared regulatory landscapes controlling specific *Hox* genes across different embryonic structure at different time and place; such as e.g. *Hoxa13/Hoxd13* transcription in the autopod and in the genital tubercule or *Hoxd9* in the limb bud and in the mammary bud (Lonfat et al., 2014; Schep et al., 2016).

It is important to take into account the two rounds of genome duplications that occurred at the basis of the vertebrate phylum, prior to their radiation (Holland et al., 1994). These duplications may have indeed provided a very stimulating opportunity to start the process of expansion of regulatory landscapes by temporally alleviating the evolutionary pressure over *Hox* genes. In turn, the expansion and development of complex regulatory landscapes triggered *Hox* neo-functionalization, not through variation in their protein sequences but through variations, modulations and novelties in their regulatory landscapes (Duboule, 2007; Holland, 1999; Ohno, 1970; Soshnikova et al., 2013). This view may help to rationalize the development of pleiotropic regulations and the functional co-option of *Hox* genes in a variety of vertebrate specific secondary structures.

Accordingly, the dependence of *Hox* genes regulations towards long-range *trans* enhancers has been proposed to result from a virtuous cycle of organization of *Hox* cluster, i.e. more enhancers within a topological landscape allows for more complex and elaborated regulations within a stronger topological organization. The development of complex and pleiotropic regulations would then in turn requires the arrangement of *Hox* genes in dense loci, to maximize the variety and the functional efficiency of the regulation of combinatorial sub-group of *Hox* genes. This necessity would thus generate a force to further consolidate and densify *Hox* genes in *organized Hox* clusters (Darbellay and Duboule, 2016; Lonfat and Duboule, 2015). We thus assume that *disorganized Hox* clusters display less organized regulatory landscapes and reduced chance of co-option because of the limited quantity of possible meta-*cis* regulations, an issue directly resulting from larger physical distances separating *Hox* genes in such clusters. As a consequence, the level of associated pleiotropy is reduced in these *disorganized* clusters.

1.3.4 Directionality of transcription

Alongside the dense arrangement of *Hox* genes (typically 10 *Hox* genes within 100 kb) and the total absence of any other non-*Hox* protein coding gene, vertebrate *Hox* clusters display yet another genetic feature. In each of these clusters, all the *Hox* transcriptional units are systematically positioned following the same 5'-3' transcriptional polarity, as described originally in the murine *HoxD* cluster (Izpisúa-Belmonte et al., 1991). The most anterior *Hox1* gene, which is activated spatio-temporally first, is found systematically positioned at the 3' extremity of the given cluster, whereas the most posterior *Hox13* gene, activated last at the end of the embryo segmentation, is always situated at the 5' extremity of the cluster (Fig 4C). Therefore, the collinear activation of *Hox* genes is gradually propagating from the 3' to the 5' extremity of *Hox* cluster, following a direction antagonist to the transcription of these genes, in a ratchet-like mechanism (Soshnikova and Duboule, 2009). Thus, a simple hypothesis whereby the progressive *Hox* genes activation might be mediated by their own transcription, i.e. the generation of extending polycistronic transcripts emitted by *Hox1* or the progressive degradation of repressive chromatin marks in response to the gradual increase of transcriptional activity, cannot be taken into account to functionally explain *Hox* progressive activation.

This atypical organization has been observed in every *organized Hox* clusters described so far. Consequently, it suggests that this arrangement may be required for the functionality of this genetic structure, both for the regulation and for the transcriptional outputs of *Hox* clusters. This statement is further supported by two evidences illustrating the drastic consequences observed when experimentally disrupting this genetic organization in the mouse genome. In a first instance, the complete inversion head-to-tail of the *HoxD* cluster, conserving therefore a complete transcriptional polarity across *Hoxd* genes but repositioning *Hoxd13* in the position of *Hoxd1* and *Hoxd1* in the position of *Hoxd13*, results in an embryonic lethal phenotype (Zákány et al., 2004). In this reconfigured condition, the spatio-temporal activation of *Hoxd* genes fails to be maintained; *Hoxd13* expression is considerably anteriorized along the AP axis of the mutant embryo, adopting a pattern similar to *Hoxd1* and causes important homeotic damages in these anterior structures (Young et al., 2009; Zákány et al., 2004). In a second instance, the intra-*HoxD* inversion of *Hoxd11* and *Hoxd12*, breaking the absolute 5'-3' polarization canon of *Hox* clusters, disturbs the proper regulation of the posterior extremity of the *HoxD* cluster (Kmita et al., 2000a). The functional causes of this dysregulation remain to be clarified and will be addressed extensively during this work. Still, recent observations pointed to the important transcription generated by the inversion of *Hoxd11* and *Hoxd12* on the anti-*HoxD* DNA strand as a possible explanation for the observed disruptions (Delpretti et al., 2013).

This interpretation is supported by the very limited quantity of transcriptional activity scored on the anti-*Hox* DNA strand in a wild type configuration. Indeed, few vertebrate *Hox* genes promoters are showing any significant transcriptional bi-directional activity, resulting in robust upstream antisense transcription. This notable strand stringency is evident in embryonic tissues and organs where the transcriptional output of the cluster is maximal, such as in the autopod at E12.5 or in the metanephric kidney at E13.5 (see Fig. 12B and 13B). Some of the rare lincRNA mapped to the anti-*Hox* DNA strand have been described and extensively studied, such as e.g. *Hotair* located in the *HoxC* cluster, or *Haglr* located in the *HoxD* cluster. However, whenever the function of these two lincRNA was properly challenged by genetic approaches, none of these transcripts revealed any evident functionality, further clarifying the relatively simple but fundamental disposition of vertebrate *Hox* genes in a polarized array (Amândio et al., 2016; Zakany et al., 2017).

1.3.5 Mechanisms for TC in vertebrates

The collinear activation of *Hox* genes, as observed in vertebrates, is a puzzling mechanism. Because of its direct integration within the vertebrate elongation strategy, it results in the spatial regulation of the position and shape of HOX proteins expression domains. In turn, these patterns are then directly controlling the morphogenesis, identity and functions of the different metameres of the future body. Accordingly, this whole morphogenetic process relies upon the progressive and gradual activation of the *Hox* clusters. The regulation of this progressive activation has been extensively studied and genetically challenged (see references in Deschamps and Duboule, 2017) and if the exact mechanism regulating this activation is not yet fully understood, its dynamic and some of its constraints are now better described. This understanding has been principally acquired during the last decade, thanks to the extensive development of chromosome conformation capture techniques and DNA sequencing technologies (see references in Denker and de Laat, 2016; de Wit and de Laat, 2012).

Some functional characteristics of this peculiar regulation will be highlighted here. As presented above, temporal collinearity is acting like an activating wave, travelling from the 3' to the 5' extremity of *Hox* clusters. Each *Hox* gene relative position within its cluster thereby determines its timing of activation, in a distance and *Hox*-dependent manner, i.e. how many *Hox* promoters lie in between the 3' extremity of its cluster and the considered *Hox* gene (Tschopp et al., 2009). This mechanism has been extensively studied in the case of the *HoxA*/*HoxD* clusters, where the initiation of this activation is driven by enhancers located next to the anterior extremity of the clusters, in their 3' regulatory landscapes (Neijts et al., 2016, personal communication by Leonardo Beccari and Aurélie Hintermann).

The progression of this activating wave correlates with at least two important features displayed by the chromatin at this locus, their respective causality not having been yet demonstrated nor established. In a first instance, the nature of the histone modifications decorating the *Hox* clusters are progressively exchanged, concomitantly with the transcriptional activation. A gradual replacement is observed from a transcriptionally repressed landscape, characteristically garnished by marks deposited by Polycomb-group repressive complexes, such as e.g. H3K27me3 or H2A119ub, to a transcriptionally active one, typically decorated with H3K4me3 deposited by trithorax (TRX) complexes (Noordermeer et al., 2011; Soshnikova and Duboule, 2009). In a second instance, the propagation of this activation wave is also materialized by significant structural changes in the local topological organization of the chromatin, a distinctive feature visible with both optic and molecular conformational approaches (Fabre et al., 2015; Noordermeer et al., 2011, 2014). Indeed, the transcriptionally inactive, H3K27me3 covered *Hox* genes tend to coalesce together in dense hubs of privileged topological interactions, therefore insulating themselves from the influence of the active *trans* enhancers. On the other hand, the activated *Hox* genes are structurally open and display higher quantity of contacts with the neighboring regulatory landscape (Noordermeer and Duboule, 2013).

Finally, the collinear activation of the *HoxD* cluster is also influenced by the regulatory landscape located on the other side of the cluster, 5' of *HoxD*. This regulation is however principally indirect as no enhancers activity controlling the *Hoxd* genes during their initial activation along the AP axis is located on this side of the cluster (Spitz et al., 2003). Still, the genetic disconnection of the *HoxD* cluster from this posterior regulatory domain by a large targeted inversion affects the dynamic of *HoxD* activation, whereby the posterior *Hoxd* genes are activated too early during AP axis elongation. Such an anteriorization of their expression patterns along the AP axis results in turn in series of homeotic phenotypes (Tschopp and Duboule, 2011). The exact causes of this up-regulation have not been extensively addressed yet in the context of the temporal collinearity along the AP axis but they are related to two

factors, likely entwined. First, the domain of repressive Polycomb-group marks at the *HoxD* locus extends a dozen of kb on the 5' side of the *HoxD* cluster. The structural integrity of this repressive domain is therefore affected in this targeted inversion, as it is split into two discontinuous locus (Schorderet et al., 2013; Soshnikova and Duboule, 2009). Second, the preferential constitutive topological contacts observed between the posterior regulatory landscape and the posterior extremity of the *HoxD* are also affected by this inversion. Altogether, these perturbations affect the insulation capability of the posterior fraction of the *HoxD* cluster to buffer and control the progression of the activation wave (Andrey et al., 2013).

Indeed, the topological organization of the chromatin is a crucial parameter participating in the dynamic of *Hox* clusters activation. Both *HoxA* and *HoxD* clusters are each flanked by two TADs, which are constitutive chromatin domains of favorite topological interactions (Dixon et al., 2012; Nora et al., 2012). These TADs encompass the different regulatory landscapes and collections of enhancer regulating *HoxA* and *HoxD* (Andrey et al., 2013; Berlivet et al., 2013; Montavon et al., 2011; Neijts et al., 2016). The importance of their contributions in the collinear activation of the *Hox* clusters can be further appreciated through the extensive knowledge accumulated at the *HoxA* cluster, thanks to an approach combining *in vivo* and *in vitro* experiments using differentiation assay. In this cluster, the early activation of collinearity is mediated by series of Wnt3/Wnt3a responding enhancers (Cao et al., 2017; Neijts et al., 2016). These enhancers are located in *HoxA* 3' regulatory landscape, at a relative short distance of *HoxA* (100 kb). They are thus positioned next to the anterior *Hoxa* genes, the first genes to be activated (Cao et al., 2017; Neijts et al., 2016). The disposition of these enhancers at this genetic position is intriguing. Indeed, 3' *HoxA* genes are scoring significant chromatin contacts with this particular fraction of the *HoxA* regulatory landscape, series contacts visible in an HiC map as a subTAD (Phillips-Cremins et al., 2013), included within the principal 3' TAD (3' subTAD, Neijts et al., 2016). Thus, the topological proximity between these loci appear to be constitutive to the chromatin and observed prior to the proper activation of the Wnt responding enhancers. Moreover, these chromatin contacts are restricted to the exclusive anterior fraction of the *Hoxa* genes, the central fraction of *HoxA* is excluded from this local hub of interaction. On the other hand, the posterior *Hoxa13* is characteristically engaged in contacts with *HoxA* 5' TAD (Neijts et al., 2016). Altogether, the specificity of the initial response seems therefore to be assured by topological contacts priming series of enhancers to act specifically on their regulative targets, in this case a sub-fraction of *Hoxa* genes.

Overall, we can conclude this overview of the mechanisms regulating the temporal activation of *Hox* genes by highlighting three important parameters. First, the nature and the variety of input signals controlling the *Hox* genes activation are closely coupled with the elongation of the AP axis (*Wnt*, *Cdx*, *Gdf11*, *Hox13*). Second, the specificity displayed by these regulative signals, e.g. Wnt responding enhancers are not activating *Hoxa1* concomitantly with *Hoxa4* even though these genes are genetically close, stresses the importance of the topological segmentation of *Hox* clusters to restrict regulatory inputs in time and space. Third, the whole system requires specific transcriptional repressive marks, efficient at participating in the topological insulation of the inactive *Hox* genes in hubs but labile and susceptible to be gradually substituted by specific activating signals.

1.3.6 Bimodal structure and regulation of *Hox* genes

We have described that the division of the *Hox* genes in a determined three-dimensional structure is actively participating in their global regulation. By priming specific interactions between regulatory landscapes and their target *Hox* genes, these conformational organizations restrict the numerous regulatory inputs in time and space, avoiding cross-activation and unspecific response of subset of *Hox* genes to enhancers (Andrey et al., 2013; Rodríguez-Carballo et al., 2017; Tschopp and Duboule, 2011). It is therefore not surprising that *organized* vertebrate *Hox* clusters have extensively integrated the topological dimension to further refine, and potentially reinforce, their complex and pleiotropic regulations (Andrey et al., 2013; Cao et al., 2017; Delpretti et al., 2013; Lonfat et al., 2014; Montavon et al., 2011; Neijts et al., 2016). In this view, the vertebrate paralog *HoxA* and *HoxD* clusters have developed this three-dimensional segmentation strategy up to the very functional limit.

These two clusters have indeed the particularity to stands at a boundary between two TADs. Therefore, both the *HoxA* and the *HoxD* clusters are topologically split in two halves of unequal size. Each TADs encompasses a fraction of the cluster, i.e. the posterior or the central/anterior *Hox* genes, together with their corresponding regulatory landscapes; respectively the 5' regulatory landscape with the posterior *Hoxa/d* genes and the 3' regulatory landscape with the central/anterior *Hoxa/d* genes (Andrey et al., 2013; Berlivet et al., 2013; Rodríguez-Carballo et al., 2017; Spitz et al., 2005). This bipartite segregation of the clusters helps to partition the regulatory inputs to subsets of *Hox* genes that are transcribed in a variety of developmental contexts, after their initial activation during the elongation of the AP axis. The functional role of this organization has been extensively studied in the *HoxD* cluster in the context of appendicular development. This emblematic illustration of *Hox* genes co-option has exposed the bimodal regulation of the *HoxD* cluster. Initially, during the first phase of limb development, the enhancers located within *HoxD* anterior and telomeric TAD (T-DOM), located 3′ of the *HoxD* cluster, are active. These enhancers control the transcription of *Hoxd3* to *Hoxd11* into the most proximal part of the future limb: the arm and the forearm. In a second phase, in distal limb bud cells, the T-DOM is functionally switched off while the enhancers located within the opposite 5′-located TAD (C-DOM) become active to control the expression of the posterior *Hoxd* genes, i.e. principally *d13* to *d11* and to a lower extent *Hoxd10* (Andrey et al., 2013; Beccari et al., 2016). This second patterning will ultimately result in the development of the digits and of the autopods of the future animal (Delpretti et al., 2012; Zakany and Duboule, 2007).

Therefore, two successive waves of transcription occur controlled by distinct enhancer disposed in two different landscapes. Functionally, HOXD13 and HOXA13 are thought to participate in the permutation of the activity of the two domains between the proximal and the distal limb, i.e. respectively from a state of T-DOM ON/C-DOM OFF to a state of T-DOM OFF/C-DOM ON, in a sort of feed-forward loop where HOXA13/D13 proteins favor their own transcription (Beccari et al., 2016; Sheth et al., 2016). During limb development, the segregation of the productive interactions between the central/anterior *Hoxd* and the enhancers of the T-DOM on one hand and the posterior *Hoxd* and the C-DOM on the other hand is crucial to avoid an uncontrolled expression of the posterior *Hoxd* in the proximal part of the limb. Such a situation is observed when genetically dissociating the posterior fraction of the *HoxD* cluster from the C-DOM by a targeted inversion. This reconfiguration results in the local loss of insulation of the posterior *Hoxd* genes, who fall partially under the regulative influence of the T-DOM (Andrey et al., 2013). In turn, the gain-of-expression of *Hoxd12* and *Hoxd13* in the proximal limb, an area normally exclusively patterned by the central/posterior *Hoxd* genes, alters the proper development of the structure and provokes a mesomelia (Andrey et al., 2013; Tschopp and Duboule, 2011).

A recent study challenged the nature of the boundary segregating *HoxD* into two insulated genetic units in the perspective of limb development by comparing interaction profiles generated in a variety of mutant embryos bearing different deletions of this locus (Rodríguez-Carballo et al., 2017). The main conclusion reached by the authors is that the two TADs composing the C-DOM and the T-DOM show a remarkable structural resistance when confronted to different disruptions of their boundary zone. Indeed, a significant topological merging between these two structures is only observed when the whole *HoxD* cluster is removed together with a significant portion of the C-DOM, in a large 400 kb deletion. But even in such an organizational situation and in absence of a complete set of *Hoxd* target genes to control, the proximal and the distal limb enhancers remained independently active (Rodríguez-Carballo et al., 2017). This bimodal regulation of the *HoxD* cluster has been so far experimentally observed in each developmental situation where *Hoxd* genes have been co-opted. The same regulative strategy is indeed at play in the external genitalia, in the intestinal hernia (caecum) as well as in the developing mammary glands (Delpretti et al., 2013; Lonfat et al., 2014; Schep et al., 2016). In a fashion similar to what was initially observed in the developing limbs, the C- and the T-DOM regulations have never been observed working concomitantly and the segregation of the *HoxD* cluster in two inter-dependent regulatory unit is systematically located at the interphase between *Hoxd13* and *Hoxd10*. In accordance with this bimodal regulation, enhancers driving the expression of the central/anterior *Hoxd* genes are systematically described in the 3' T-DOM, whereas enhancers controlling posterior *Hoxd* genes have been specifically identified within the 5' C-DOM.

Because of the *HoxD* transcriptional profile reported in embryonic metanephric kidneys (Fig. 6A), it is not surprising to see that the bimodal regulation of the *HoxD* cluster is also at work during the development of this organ (Brunskill et al., 2014; Di-Poï et al., 2007; Patterson and Potter, 2004). This embryonic organ will be used extensively in the present work as a functional read-out to challenge the questions underlying this thesis, it constitutes therefore a relevant example to illustrate this unusual genetic regulation as well as to discuss the nature of the topological boundary found over the posterior extremity of the *HoxD* cluster. The development of the kidneys requires the coordinated expression of the central/anterior set of *Hoxd* genes, from *d3* to *d11*. A small fraction of cortical metanephric cells express *Hoxd12* whereas *Hoxd13* remains completely silenced. The active transcriptional control is mediated by a combination of *trans* enhancers and *cis* element located within the *HoxD* cluster and in the regulatory landscape, altogether encompassed within the T-DOM (Di-Poï et al., 2007). The local exclusion of contacts between *Hoxd13* and the active T-DOM, expected from the description of the transcriptional situation, is confirmed when the conformation of the chromatin is revealed by 4Cseq (Fig. 6A). The interaction profile generated from a viewpoint located within the promoter of *Hoxd13* shows a distribution of contacts extensively polarized towards the posterior extremity of the *HoxD* cluster. In accordance with this observation, two 4Cseq profiles generated from viewpoints positioned in the vicinity of the active *Hoxd9* and *Hoxd11* genes show the reverse picture (Fig. 6A).

This boundary effect is even more visible when considering the same set of 4Cseq profiles plotted over a larger genetic distance, in the range of the C- and the T-DOM size (Fig. 7A). This representation reveals the extensive inclusion of the central/anterior fraction of the *HoxD* cluster within the T-DOM, matching remarkably the position of the TAD as described in HiC data obtained from mESC or embryonic limb cells (Dixon et al., 2012; Rodríguez-Carballo et al., 2017). On the one hand, the two transcriptionally active *Hoxd* genes considered here are engaged in series of specific long-range interactions, up to 800 kb away from their genetic positions within the *HoxD* cluster. On the other hand, no significant contacts are scored between these two *Hoxd* genes and the C-DOM, even at a short distance. From the *Hoxd13* viewpoint, the contacts observed are the exact opposite of the picture obtained with *Hoxd9* and *Hoxd11*. The repressed *Hoxd13* gene scores series of specific constitutive long-range interactions with

element located within the centromeric regulatory landscape at distance up to 500 kb and its global topological interactions are fully incorporated within the C-DOM.

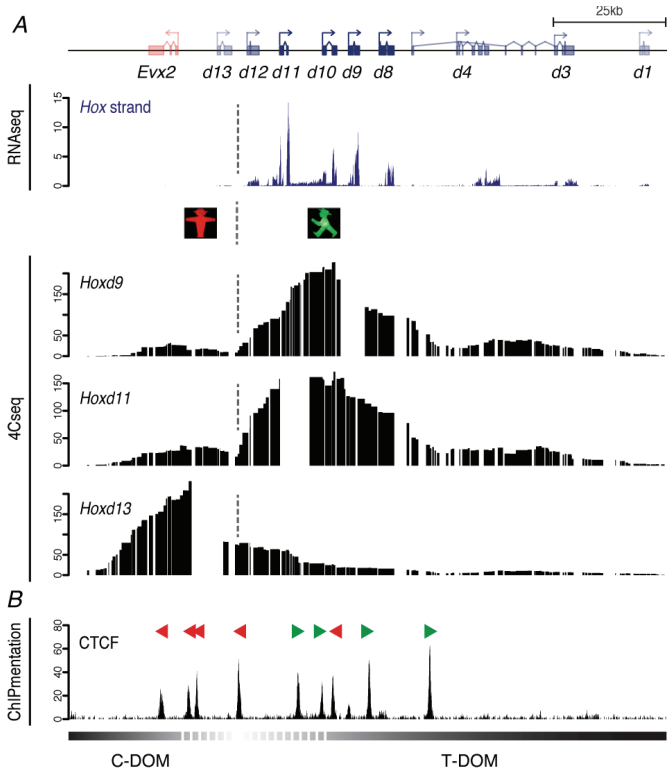

Figure 6: Topological sequestration of the posterior *Hoxd* genes in E13.5 metanephros. (*A*) Embryonic kidneys are patterned by a combination of central and anterior *Hoxd* genes, as illustrated by a strand specific RNAseq. The quantitative peak of transcripts is shared between *Hoxd11* and *Hoxd8*. These transcriptionally active *Hoxd* genes are interacting together as revealed by 4Cseq generated from two viewpoints located within the active fraction of *HoxD* (4Cseq *Hoxd9* and *Hoxd11* viewpoints). These two profiles show concordantly a significant drop of contacts between *Hoxd11* and *Hoxd13*, correlating with the transcriptional status of the locus. This segregation is confirmed by the antagonist contacts observed from the transcriptionally inactive *Hoxd13* (4Cseq *Hoxd13*). This segmentation of the *HoxD* in the metanephros is similar to what has been describe in other tissues patterned by a similar set of *Hoxd* genes, such as the caecum and the proximal part of the limb (Andrey et al., 2013; Delpretti et al., 2013). The low intensity of contacts observed between *Hoxd9*/*Hoxd11* and the anterior extremity of the *HoxD* cluster (*d4*/*d3*) is assumed to be caused by the heterogeneity displayed by developing metanephros, as a fraction of this complex organ (i.e. the ureteric bud) is exclusively patterned by anterior *Hoxd* genes. (*B*) The position of the boundary discriminating the contacts between the active and inactive *Hoxd* genes correlates with a major inversion in the polarity of functional CTCF binding sites, as revealed by CTCF ChIPmentation. Red arrowheads represent CTCF binding sites which contain a CTCF motif orientated towards the posterior extremity of the cluster whereas green arrowheads represent CTCF binding motifs facing the anterior extremity of the *HoxD* cluster. In view of the loop extrusion model (Fudenberg et al., 2016), the inversion of CTCF binding sites polarity suggests a role of chromatin loops in the segregation of the posterior *Hoxd* genes.

Of note, the position of the TADs boundary within the *HoxD* cluster correlates with a series of binding site for CCCTC-binding factor (CTCF) engaged with components of the cohesion complex (Rodríguez-Carballo et al., 2017; Soshnikova et al., 2010), a feature frequently found at TADs boundaries genome wide (Ghirlando and Felsenfeld, 2016). Several indications supports the idea that cohesin and CTCF are indeed two factors actively participating in the formation of chromatin loops, following the so-

called loop-extrusion model (Alipour and Marko, 2012; Fudenberg et al., 2016; Goloborodko et al., 2016; Nasmyth, 2001; Rao et al., 2014; Sanborn et al., 2015). The exact importance of these loops in the context of TADs position and structural integrity is not completely established yet, however several lines of experimental evidences obtained by removing these two elements revealed a global destabilization of TADs (Haarhuis et al., 2017; Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017).

A crucial point in the loop-extrusion model is the proposal that the position of chromatin loops is fixed by convergent orientated CTCF proteins. Chromatin extrusion mediated by the cohesin complex is proposed to stop when this complex encounter a CTCF protein engaged "head first" on the chromatin, whereas the same CTCF protein engaged on a genetically inverted binding site, i.e. "tail first", will not stop the extrusion process (Fudenberg et al., 2016; de Wit et al., 2015). In view of this model, the distribution of the polarized CTCF binding site motif is perceived as an important landmark to orientate structural interactions of the chromatin. Such is the case of the boundary between the *HoxD* C- and the T-DOM where an important inversion in the orientation of the functional CTCF sites is visible. On one hand, the four sites located between *Evx2* and *Hoxd12* face the C-DOM whereas on the other hand four out of the five CTCF sites positioned between *Hoxd11* and *Hoxd4* are facing the T-DOM (Fig. 6B). The disposition and orientation of these binding sites are thus matching remarkably the segregation of contacts observed between the posterior and the central/anterior fraction of the *HoxD* cluster, as discussed above. Moreover, a significant fraction of the defined long-range contacts observed from the 4Cseq viewpoints are falling on CTCF binding sites located, from *Hoxd13*, within the C-DOM and from *d11* and *d9*, within the T-DOM (Fig. 7B). In accordance with the loop-extrusion model, these sites located within *HoxD* regulatory landscapes are each orientated facing the *HoxD* cluster, further supporting a role of this mechanism to participate in *HoxD* bimodal regulation (Rodríguez-Carballo et al., 2017).

While the functional causes of the topological boundary separating both the *HoxA* and the *HoxD* clusters into two genetic units is not fully deciphered yet, the regulative advantages provided by the bipartite segregation of the posterior from the central/anterior *HoxA*/*D* can be appreciated by the fact that such a structural organization is conserved across the vertebrate phylum, from fish to mammals through reptiles and chicken (Guerreiro et al., 2016; Woltering et al., 2014 and Nayuta Yakushiji-Kaminatsui, personal communication). However, this segregation is observed neither in the murine paralogous *HoxB* nor in *HoxC* cluster (Dixon et al., 2012). Interestingly, these two clusters have a *cis*-organization different from the *HoxA*/*HoxD* clusters. Indeed, both *HoxB* and *HoxC* display a significant physical separation of the 3' and 5' *Hox* genes and might therefore not require a specific structure to segregate the *Hox* genes from their respective regulatory landscapes. In the *HoxB* cluster, the posterior *Hoxb13* gene is located very remotely from the *Hoxb1*-*Hoxb9* genes, *de facto* insulated by 80 kb of intergenic chromatin separating these two series of genes. On the other hand, the *HoxC* cluster lacks the entire 3' part of the cluster, being composed only of *Hoxc4* to *Hoxc13*.

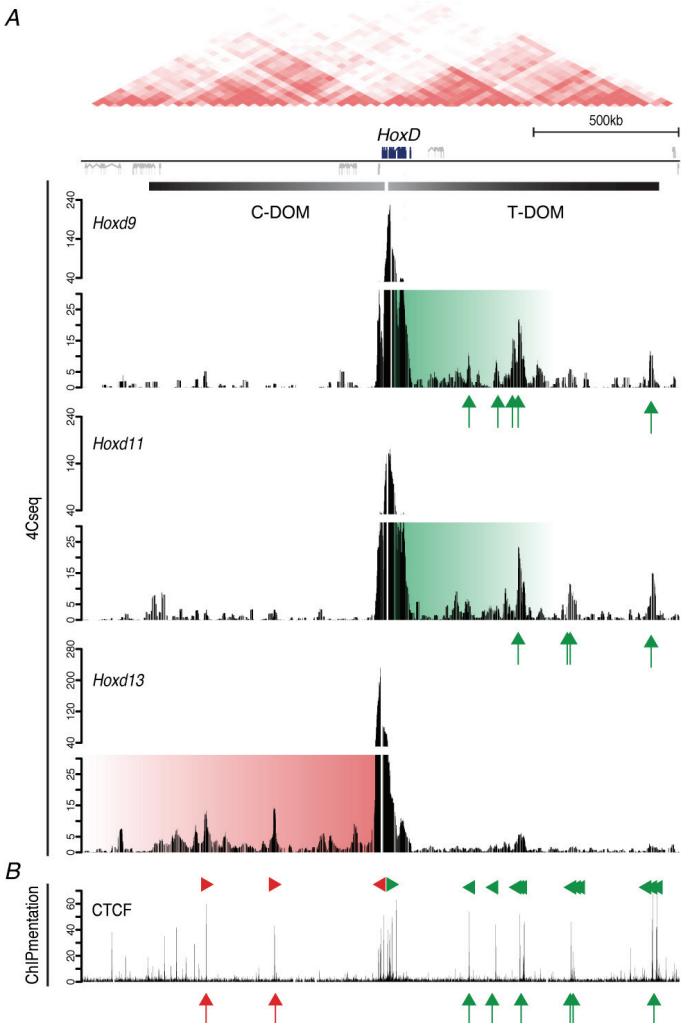

Figure 7: The segregation of topological contacts observed over the posterior extremity of the *HoxD* cluster corresponds to a boundary between two TADs. (*A*) The topological contacts observed by 4Cseq in E13.5 metanephros from the three viewpoints *Hoxd9*, *d11* and *d13* reveal specific long-range contacts between *Hoxd* genes and the two adjacent regulatory landscapes. In accordance with the segregation of contacts observed locally at the *HoxD* locus, transcribed *Hoxd* genes (*Hoxd9*, *d11*) score a majority of contacts with element located within the *HoxD* T-DOM. The interaction profile of the inactive *Hoxd13* gene presents an opposed view, with contacts restricted to the C-DOM. Globally, these contacts are encompassed within the limit of the C- and T-DOM as defined by Andrey et al., 2013 and experimentally confirmed by Rodríguez-Carballo et al., 2017 by HiC generated from embryonic material, confirming the existence of a TAD boundary between *Hoxd11 and Hoxd13*. In E13.5 metanephros, the T-DOM harbors active enhancers engaged in productive long-range contacts with the anterior extremity of the cluster. On the other hand, the C-DOM is inactive but remains engaged in constitutive long-range contacts with the posterior inactive fraction of the *HoxD* cluster. (*B*) An important fraction of long-range contacts observed from the three different *HoxD* 4Cseq viewpoints fall specifically over engaged CTCF sites, within the C- or the T-DOM, orientated facing the *HoxD* cluster. In the first instance, the 4Cseq profile generated from the *Hoxd13* viewpoint, located between four CTCF sites facing the C-DOM (see Fig. 6), is engaged in at least two precise interaction with CTCF sites located within the C-DOM and facing the *HoxD* cluster (red arrows). In a second instance, a similar observation is recorded from the 4Cseq profiles generated from the *Hoxd11* and *Hoxd9* viewpoints, capturing series of long-range interactions between central *Hoxd* genes and convergent CTCF sites spread in the T-DOM. Red arrowheads represent convergent C-DOM CTCF binding sites, whereas green arrowheads represent convergent CTCF located within the T-DOM. The HiC map is extracted from an HiC dataset generated from mESC and published in Dixon et al., 2012.

The recent description of the three-dimensional organization of the single *disorganized Hox* cluster of the cephalochordate *Amphioxus* (550 kb of size) is further supporting the idea that this bimodal regulation is specific to the some *organized Hox* clusters. *Amphioxus* is often considered as a legitimate model to emulate the ancestral *Hox* genetic organization in the chordate last common ancestor. Therefore, the fact that the single *Amphioxus Hox* cluster is organized in what emerges as a single large TAD of 1 mb is a strong indication that *Hox* bimodal regulation is likely a vertebrate specificity (Acemel et al., 2016). *Amphioxus* unique TAD encompasses the whole *Hox* cluster as well as its regulatory landscape located 3' to the *Hox* cluster, further reinforcing the indication that the enhancers controlling the early steps of temporal collinearity are located primordially in this area (Acemel et al., 2016).

In contrast to what is known for the vertebrate *HoxA*/*HoxD*, the 5' posterior *Amphioxus Hox* genes are not topologically detached from the large *Hox* cluster. No exclusive contacts are observed between these genes and the 5' neighborhood of the *Hox* cluster, therefore ruling out a separation of this group of genes into two regulatory units. In parallel, no organized regulatory landscapes are apparent in the 5' vicinity of the *Amphioxus Hox* cluster. This observation is then further raising the causal relationships between, on the one hand, the position of the TADs boundary over the *HoxA*/*D* clusters and, on the other hand, the pleiotropic regulatory landscapes composed of long-range enhancers controlling the posterior *HoxA*/*HoxD* clusters (Acemel et al., 2017; Darbellay and Duboule, 2016; Neijts and Deschamps, 2017).

2 AIM OF THE THESIS AND GENERAL APPROACH

2.1 Scope of the thesis

In vertebrate genomes, *Hox* genes are disposed in tight gene clusters, composed of dense series of *Hox* paralogs arrayed over a short genomic distance, usually in the range of 100 kb. These clusters display an entirely polarized orientation of *Hox* open reading frames; *Hox* genes are transcriptionally orientated following the same 5'-3' direction. This curious organization is specific to vertebrate *Hox* genes and has been proposed to result from the accumulation of inter-dependent regulations, implying both *cis* and *trans* controls, and is thought to reflect the numerous co-options of these genes through vertebrate embryogenesis (Darbellay and Duboule, 2016; Duboule, 1994, 2007).

The aim of this PhD work was to study the functional importance of the transcriptional polarization systematically found in *Hox* clusters. To do so, we generated and studied a series of murine alleles bearing targeted inversions of *HoxD* genes. This approach was meant to let us to assess the exact importance of this particular transcriptional organization within the global context of *HoxD* regulation, in different tissues during embryonic development. In other words, we wanted to assess to what extent to transcriptional polarity of *Hox* genes must be conserved within the murine *HoxD* cluster to allow the system to work synchronously. The results of such approaches and their conclusions are presented below.

2.2 General approach

2.2.1 Selection of the appropriate inversions

We used three different alleles to address the importance of the transcriptional polarization of *Hox* genes: *HoxD* $^{\text{inv}(11)}$, *HoxD* $^{\text{inv}(12)}$ and *HoxD* $^{\text{inv}(11\text{-}12)d13lacZ}$ (see Fig. 8). The two first alleles, inv(*11*) and inv(*12*), present the most direct and simple way to challenge *Hox* transcriptional organization and were motivated by a simple rationale; we wanted to generate *Hox* gene inversions as small as possible, containing a single *Hoxd* gene, therefore encompassing its main promoter to conserve as much as possible its endogenous regulation, its exons and introns as well as a fraction or the totality of its 3'UTR. We intended at conserving the functionality of the *HOX* protein encoded by the inverted gene, reallocated on the anti-*Hox* DNA strand. This was a central concern during the design of these alleles, as the motivation of this experimental work was to observe the consequence of the inversion of a single *Hox* transcriptional unit. Thus, we tried to conserve the global disposition of non-coding regulative elements located within the *HoxD* cluster, as we assumed that the perturbation of too many factors would complicate the interpretation of these alleles.

This strategy was made possible by the development of the CRISPR/Cas9 technology which emerged at the beginning of this thesis. CRISPR/Cas9 allowed us to design and generate a variety of custom alleles directly within the murine zygote, thus bypassing the long and tedious work of targeting loxP sites by homologous recombination in mESC, generating chimeric founders and stabilizing a floxed allele to finally obtain the desired inversion by an exposure to the Cre recombinase. In the second chapter of this work, we used the murine allele inv(*11-12*)*d13lacZ*, an inversion that was already available *sur pattes* in the Duboule laboratory. This allele allowed us to evaluate the consequences of a larger local intra-cluster inversion. Inv(*11-12*)*d13lacZ* encompasses indeed two *Hoxd* genes as well as the intergenic chromatin, containing notably a CTCF site located upstream of *Hoxd12* (Fig. 8). In order to assess the importance of the locus *Hoxd11* in this configuration, we altered this allele by CRISPR/Cas9 mutagenesis, thus finally creating a *HoxD* $^{\text{inv}(11\text{-}12)\text{del}(11)d13lacZ}$ allele. This additional allele allowed us to challenge the responsibility of the *d11* locus in the observed dysregulation of the *Hoxd13lacZ* reporter.

Different reasons led us to decide to use and work with these three inversions, positioned over a similar portion of the *HoxD* cluster. First, the *Hoxd* genes that we modified are activated and transcribed relatively late during the elongation of the developing embryo and their expressions are maintained throughout an important period of the embryonic development. Typically, their respective expression domains within the trunk are still well visible in an E12.5 embryo. This feature facilitates sampling by increasing the quantity of cells micro-dissected from the tissues of each mutant embryo, allowing us to describe and characterize these alleles using up-to-date genomic tools. Second, this series of *Hox* genes, i.e. from *Hoxd10* to *Hoxd13*, permitted us to use a variety of *HoxD*-patterned secondary structures as functional read-out to study these inversions and their global consequences for the *HoxD* cluster regulation. Previous works from the Duboule laboratory have indeed interrogated the regulation of the *HoxD* cluster within the limbs (Andrey et al., 2013; Montavon et al., 2011), the intestinal hernia (Delpretti et al., 2013; Zacchetti et al., 2007) and the metanephric kidneys (Di-Poï et al., 2007), structures that arise and develop between E9 and E13.5. Third, it was of crucial importance that we did not disturb the proper expression of the anterior *HoxD* genes. Indeed, *Hoxd1* and *d3* have been demonstrated to be necessary for embryo development, as different loss of functions were showed to lead to homeotic phenotypes handicapping adult mutant animals (Zákány et al., 2001), a condition that would therefore complicate the management of the mutant stocks. Finally, we were curious to invert a series of *HoxD* genes that are highly expressed and show a transcriptional activity strictly located on the *HoxD* coding strand. This would *de facto* exclude *Hoxd1* and *Hoxd8*. Indeed, the promoters of these two genes are each firing one long intergenic non-coding RNA (lincRNA) on the anti-*HoxD* strand. And, to conclude this short presentation of the rationales that motivated our approaches, we were particularly interested in the genomic region located between *Hoxd11* and *Hoxd13.* This is where lies the topological and functional boundary, at the very root of the bimodal regulation of *HoxD* (see Fig. 6 and 7). Thus, we anticipated that disturbing *HoxD* transcriptional organization at the edge of the two TADs would be an appealing idea to assess to what extent the bimodal regulation of the *HoxD* cluster requires a transcriptionally polarized cluster.

2.2.2 Selection of the model tissues and organs

To evaluate the consequences of the three alleles described above over *Hoxd* genes regulations and expressions, we selected a series of embryonic structures and tissues that can be used as relevant read-outs. These tissues and their respective regulations are summarized below.

We used E12.5 embryos as a global read-out of the different expression domains of *Hoxd* genes. At this developmental stage, *Hoxd* genes are expressed in the lateral mesoderm, along the AP axis, in series of nested domains and these expression patterns are convenient to be visualized by Whole mount In-Situ Hybridization (WISH). *Hoxd* patterns along the AP axis thus provides the experimenter with a first comprehensive picture of the expression of *Hoxd* genes and provide indications of possible disruptions in the different alleles considered. The regulation of the *Hox* genes within the lateral mesoderm along the AP axis at this stage is controlled and maintained by series of relay elements positioned directly in *cis* within the *Hox* clusters (Neijts et al., 2017; Tschopp et al., 2009), even though *Hox* patterns are primordially acquired during the early spatio-temporal collinearity activation of *Hox* clusters (Dollé et al., 1989; Nelson et al., 1996; Tarchini and Duboule, 2006).

In contrast with the situation known in the trunk, the expression of central/anterior *Hoxd* genes (*d1* to *d11*) in the caecum, a secondary structure budding out of the intestinal hernia around E12, relies entirely on long-range *trans* enhancers located within the T-DOM. This regulation follows the same bimodal regulation as observed in the context of limbs development and previously presented (Delpretti

et al., 2013; Zacchetti et al., 2007). True to this general regulatory strategy, *Hoxd13* is inactive and maintained topologically insulated, scoring important topological interactions with the C-DOM (Delpretti et al., 2013). The exact opposite regulative situation is observed in the autopod ("digits") of the developing limb at E12.5. In this tissue, the posterior *Hoxd* genes, from *Hoxd13* to *Hoxd11* are expressed and controlled entirely by long-range enhancers encompassed within the global C-DOM structure (Montavon et al., 2011; Spitz et al., 2003). Meanwhile, the three dimensional structure of the T-DOM is conserved but its enhancers are not active (Andrey et al., 2013).

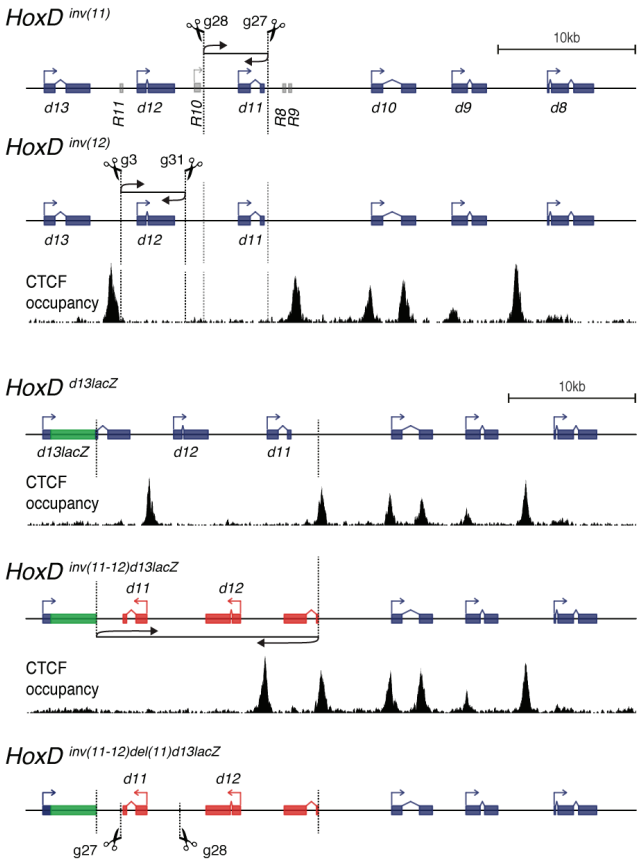

Figure 8: Representation of the different intra-*HoxD* inversions studied in this work. *HoxD* $^{inv(11)}$ and *HoxD* $^{inv(12)}$ were generated using a CRISPR/Cas9 mutagenesis approach whereas the allele *HoxD* $^{inv(11\text{-}12)d13lacZ}$ and the control *HoxD* ^{d13lacZ} were already available in the laboratory and generated through a Cre/Lox approach (Kmita et al., 2000a). Position of the inversions breakpoints are depicted as dashed vertical lines. CTCF occupancy tracks illustrate the position of CTCF at this locus and are extracted from three different CTCF ChIPmentation datasets each generated from E13.5 metanephros with different genotypes: wild type for the alleles inv(*11*) and inv(*12*), homozygous *HoxD* ^{d13lacZ} and *HoxD* $^{inv(11\text{-}12)d13lacZ}$ mutants for the *d13lacZ* and inv(*11-12*)*d13lacZ* alleles (see Results section for details). For the sake of clarity, *Hoxd11* annotation is here voluntarily restricted to its two coding exons, its prolonged 3'UTR encompasses both the gRNA27 and the *d11_d10* loxP site used in the inv(*11-12*)*d13lacZ*.

Finally, the metanephric kidneys have been frequently used during this work. This structure is in a way similar to the caecum as this developing organ also presents a fixed transcriptional pattern of central *Hox* genes. *Hox11* genes has been reported to be necessary to build a functional organ (Davis et al., 1995; Patterson et al., 2001; Wellik et al., 2002) whereas *Hox13* genes are not expressed throughout the development of this organ. The importance to conserve posterior genes silenced is illustrated by the drastic consequences of a gain-of-expression of *Hoxd13* in embryonic metanephros, a condition leading to the complete agenesis of this organ. This phenotype is lethal for the animal shortly after birth (Kmita et al., 2000b). Metanephros sampling by micro-dissection of eviscerated embryos is convenient and robust, therefore reducing the risk of artificially biasing quantitative analysis of the *HoxD* transcriptional and regulative landscapes. At E13.5, central and anterior genes, from *d3* to *d11*, are actively expressed with a quantitative peak of transcripts detected over the central fraction of *HoxD*. This heterogeneous transcriptional landscape reflects the different intertwined roles of *Hoxd* in the assembly of this developing organ. The kidneys are indeed composed of two tissues derived from the intermediate mesoderm: the ureteric bud (UB) and the metanephric mesenchyme (MM) (Shah et al., 2004; Takasato and Little, 2015). *Hoxd11* transcription is restricted specifically to the metanephric mesenchyme, a tissue composed of cells which represent a large fraction of the developing organ. On the other hand, cells expressing a variety of central/anterior *Hoxd*, from *Hoxd3* to *d9*, are specifically found in the epithelial ureteric bud (Brunskill et al., 2014; Mugford et al., 2008; Patterson and Potter, 2004; Yallowitz et al., 2011). *HoxD* regulation in the metanephros also follow the bimodal-regulation. In accord with the transcriptional landscape observed in these structures, the T-DOM is active whereas the C-DOM is inactive. An earlier study revealed this global regulation and notably demonstrated the dependency of the UB signaling to putative enhancers located within the T-DOM, in a fashion similar to the caecum (Di-Poï et al., 2007).

We refined this model during our research and identified a long-range UB enhancer located within the T-DOM. We also showed that an important fraction of the regulative inputs controlling *Hoxd* genes in developing metanephros is not located within the distant T-DOM and thus does not rely only upon long-range enhancers. Interestingly, we found *cis* regulative elements located within *HoxD*, around the locus of *Hoxd8*. On the other hand, the precise regulation of *Hoxd11* in the MM remains not fully understood, even though evidences acquired during this thesis led us to consider the locus of *Hoxd10*, in *cis* within the cluster, to be a probable controller of *Hoxd11* expression. However, these results are beyond the scope of this thesis and will therefore not be further discussed.

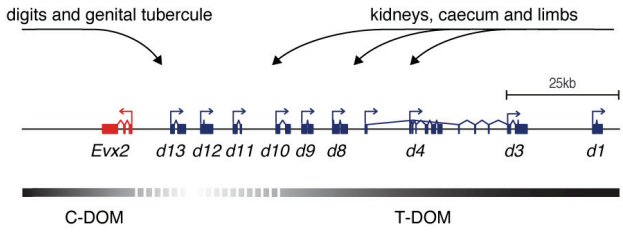

Figure 9: Schematic illustration of *HoxD* bimodal regulation. The *Hoxd* genes are controlled by two different regulatory networks that are integrated within two TADs, the C-DOM and the T-DOM. On one side, the central *Hoxd* genes are under the control of both long-range and local enhancers located within the T-DOM. The caecum, the limbs as well as the metanephric kidneys are embryonic structures typically displaying this regulation, expressing an overall similar set of central *Hoxd* genes. These structures maintain systematically the posterior *Hoxd13* gene silenced and repressed. *Hoxd13* score long-range contacts with the C-DOM and inactive enhancers, therefore participating in its regulative insulation. On the opposite, in the autopods and in the genital tubercule the posterior *Hoxd13* is actively transcribed and is controlled by the active long-range enhancers positioned within the C-DOM. In these tissues, the T-DOM is still present but inactivated (Andrey et al., 2013).

3 RESULTS

For the sake of clarity, the different alleles used during this thesis are presented and discussed separately. A figure representing the three intra-*HoxD* inversions studied is provided in the previous section of this work (Fig. 8).

3.1 Inversion of the locus *Hoxd11*

3.1.1 Design of the inversion

By using a nested inversion of the *Hoxd11* gene, we could address the importance of the polarized transcriptional landscape systematically observed within *organized Hox* clusters. This allele was generated by the simple and customizable CRISPR/Cas9 technology adapted to work in mouse zygotes (Cong et al., 2013; Ran et al., 2013). We designed two guide RNAs (gRNA) to flank *Hoxd11* in order to generate DNA double strand breaks (DSB) mediated by the Cas9 nuclease. We expected that repair of these DSBs would result in the majority of instance in alleles containing a complete deletion of the flanked locus as well as combinations of small inversions/deletions at the two Cas9 breakpoints. Altogether, these alleles would be of no particular relevance to this project. However, a fraction of all the alleles generated by this mutagenic approach would very likely contain an inversion of the target locus, resulting from the different DNA repair pathways used by the zygote (Blasco et al., 2014).

We carefully selected the position of the gRNAs as to generate a functional inversion of the gene *Hoxd11*. We desired to conserve the whole transcriptional unit of *Hoxd11* without altering the internal distribution of putative *cis* regulators and spurious transcription start sites (Fig. 10). Therefore, we positioned the first centromeric guide (gRNA 28) 2.5 kb upstream of *Hoxd11* to conserve intact *Hoxd11* promoter. This guide was located close by but telomeric to the highly conserved region X. The deletion of this element of 230 bp has no effect upon *Hoxd* genes expression, yet this sequence has been annotated as an alternative transcription start site (TSS) upstream of *Hoxd11* (Beckers and Duboule, 1998; Beckers et al., 1996). We decided therefore to exclude it from our design to avoid inverting more than one TSS. We positioned the second gRNA (gRNA 27) in the 3'UTR of *Hoxd11*, 300 bp downstream of *d11* stop codon, in a position relatively proximal within *d11* long annotated 3'UTR. This position was chosen to avoid the relocation within the *HoxD* cluster of two other elements containing *cis*-acting enhancers, the conserved RVIII and RIX located in the middle of *d11* 3'UTR (Gérard et al., 1996; Zákány et al., 1997). A supplementary reason to avoid disrupting the integrity of the RVIII/RIX locus was motivated by the presence of a CTCF binding site in the close vicinity of these two elements (Fig. 10). Because of the possible contribution of CTCF in driving long-range chromatin looping (Fudenberg et al., 2016; de Wit et al., 2015), we took care to position these two gRNAs in between CTCF sites located within the *HoxD* cluster. In total, the genomic distance measured between the two considered gRNAs was just over 4.6 kb.

gRNAs 27/28 were then cloned in pX330 vectors, each vector containing a functional copy of h*SpCas9* (Mashiko et al., 2013) and were injected at an equimolar concentration within the pronuclear of mouse zygotes (PNI). Alternatively, to increase our chance to obtain a full-length inversion of the *Hoxd11* locus, these two gRNA were also *in vitro* transcribed and electroporated directly in zygotes, together with a functional preparation of *Cas9* mRNA (Hashimoto and Takemoto, 2015). Following transfer to foster mothers, animals were screened at weaning for a complete inversion of the *Hoxd11* locus. The genotyping was effectuated by a combination of PCRs made with different pairs of convergent/divergent primers. Using the two CRISPR/Cas9 mutagenesis techniques in parallel, we managed to obtain two independent *HoxD* $^{inv(11)/+}$ F0 out of 160 animals that were born and genotyped. This new allele was then established from one single F0 animals which was extensively characterized and sequenced to confirm the integrity of the breakpoints of the inversion. The *HoxD* $^{inv(11)}$ genome was reconstructed for

future genomic analysis and the allele was then back-crossed with wild type animals for two generations to dilute any off-target mutations caused by the CRISPR/Cas9 mutagenesis.

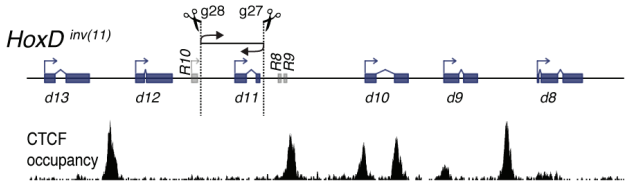

Figure 10: Schematic representation of the allele *HoxD* $^{inv(11)}$ produced by a CRISPR/Cas9 mediated mutagenesis. The blue boxes represent the exons of the depicted *Hoxd* genes. The respective position of the two gRNAs used to generate inv(*11*) are represented. gRNA27 is positioned within *Hoxd11* 3'UTR but centromeric to the conserved region RVIII and RIX (Gérard et al., 1996; Zákány et al., 1997) whereas gRNA28 is located telomeric to the conserved region RX which is annotated as a transcription start located upstream of *Hoxd11* (Beckers and Duboule, 1998; Beckers et al., 1996). For the sake of clarity, *Hoxd11* annotation is here voluntarily restricted to its two coding exons, *Hoxd11* long 3'UTR encompasses both RVIII and RIX. CTCF occupancy track is extracted from a CTCF ChIPmentation dataset produced in E13.5 wild type metanephros.

3.1.2 *HoxD* expression in the inv(*d11*)

3.1.2.1 General observations

This allele did not present any obvious homeotic phenotypes; at least none were observed in heterozygous adult animals. Even though *Hoxd11* patterns important secondary structures such as the forelimbs and the hindlimbs as well as the kidneys, adult heterozygous animals appeared to be unaffected by this genetic reconfiguration. This first observation led us to conclude that the inversion of a single *Hoxd* gene at its endogenous genomic position within the *HoxD* cluster is viable and does not cause sterility, thus allowing us to expand this allele. By extension, we also concluded that the inv(*11*) was not causing a drastic gain of expression of posterior *Hoxd* genes, as illustrated notably in Kmita et al. in the lethal relocation of a *Hoxd9* transgene upstream of *Hoxd13* (Kmita et al., 2000b). Having thus excluded that the inversion of *Hoxd11* is leading to a massive dysregulation of the *HoxD* cluster function, we focused our attention at the *Hoxd* genes located in the neighborhood of *Hoxd11* and examined their respective patterns of expression. We crossed heterozygous adult animals to obtained homozygous mutant and wild type E12.5 littermates embryos. This embryonic stage has many advantages since central and posterior *Hoxd* genes are still expressed and their patterns of expression remain visible in the lateral mesoderm and in the spinal cord, as well as in the developing limbs of the embryo. Consequently, we could perform Whole mount In-Situ Hybridization (WISH) on these samples using antisense probes matching *Hoxd10*, *Hoxd11*, *Hoxd12* and *Hoxd13* mRNA (Fig. 11).

Hoxd11 expression appeared to be fully conserved in the inv(*11*) allele. Its expression in the trunk was similar to the one displayed by the control. We also observed specific signal in a variety of secondary structures known for displaying high level of *Hoxd11* in a wild type situation. The zeugopod and autopod of the forelimb, as well as the developing intestinal hernia showed a strong signal for *d11* in the mutant inv(*11*) (Fig. 11, black arrows). Finally, the intensity of the global signal detected in the inv(*11*) homozygous embryo using a *d11* probe was stronger, although these WISH experiments where performed in parallel. We speculated that this quantitative difference might be a direct consequence of the new sequence of *Hoxd11* 3'UTR in the inv(*11*) as the 3' breakpoint of the inversion, targeted by gRNA27 (Fig. 10), is positioned within *Hoxd11* 3'UTR. This partially artificial 3' UTR might therefore change the half-life and the stability of *Hoxd11* mRNA in the inv(*11*) allele. Importantly, the integrity of the portion of

d11 mRNA detected by the *d11* WISH probe was not affected by the genetic reconfiguration of this locus. Altogether, we concluded that the different regulations of *Hoxd11* were unaffected by the inversion of this locus.

Hoxd10 expression pattern in the trunk was similar in the inversion as in the control embryo, extending from the anterior extremity of the hindlimb up to caudal extremity of the developing animal. This expression pattern clearly differs from *d11* expression, as typically observed in both the inv(*11*) and in the wild type control embryo. In those samples, the anterior extremity of *Hoxd11* pattern starts indeed at the center of the hindlimb (Fig. 11, compare the relative patterns along the rostro-caudal axis against the position of the hindlimb, marked by a dotted black line). This observation is thus excluding any major defects in *Hoxd10* expression along the AP axis of the inv(*11*), as *Hoxd10* remains expressed within its own nested pattern partially excluded from the pattern of *Hoxd11*.

However, if the expression of *Hoxd10* seems qualitatively unmoved along the AP axis, its expression was found to be quantitatively affected. This observation is notable in the limbs of the homozygotes mutant embryos examined (Fig. 11, blue arrowheads). *Hoxd10* signal detected in the autopod appeared to be restricted to the three central digits, compared to the usual four digits pattern displayed by the same gene in the wild type control. A similar reduction of signal was made on the zeugopod of the same embryo, while *d10* expression appeared to have been whipped out of the intestinal hernia (Fig. 11, blue arrowheads). Interestingly, this disruptions seems to concern also *Hoxd12* and to be therefore specific to the two *Hoxd* genes flanking the inv(*11*) mutation. *Hoxd12* is indeed subjected to a similar albeit even more drastic alteration. First, its expression seems to be almost whipped out from the trunk of the inv(*11*) mutant embryo and second, it appears also to be quantitatively reduced in both the zeugopod and the autopod (Fig. 11, red arrows). These two results were totally unforeseen.

Finally, *Hoxd13* expression and regulation is unaffected by the inversion of *Hoxd11*. Specific *Hoxd13* signal is indeed collected in the thumb of the autopod (digit I) in both the mutant and the wild type embryos, whereas this staining is systematically absent from the other WISH (Fig. 11, black arrowheads). Correlatively, *Hoxd13* is not observed in any significant level in the zeugopod. *Hoxd13* expression in the trunk was generally low, restricted to the caudal extremity of the developing embryo, a situation frequently observed in wild type animals at this embryonic stage. Therefore, we concluded from this preliminary exploration of the inv(*11*) allele that the functional inversion of *Hoxd11,* within the *HoxD* cluster, does not induce a global dysregulation of *Hoxd* activation across different regulations and tissues. Instead, this inversion affects negatively *Hoxd10* and *Hoxd12* in a very specific but robust manner.

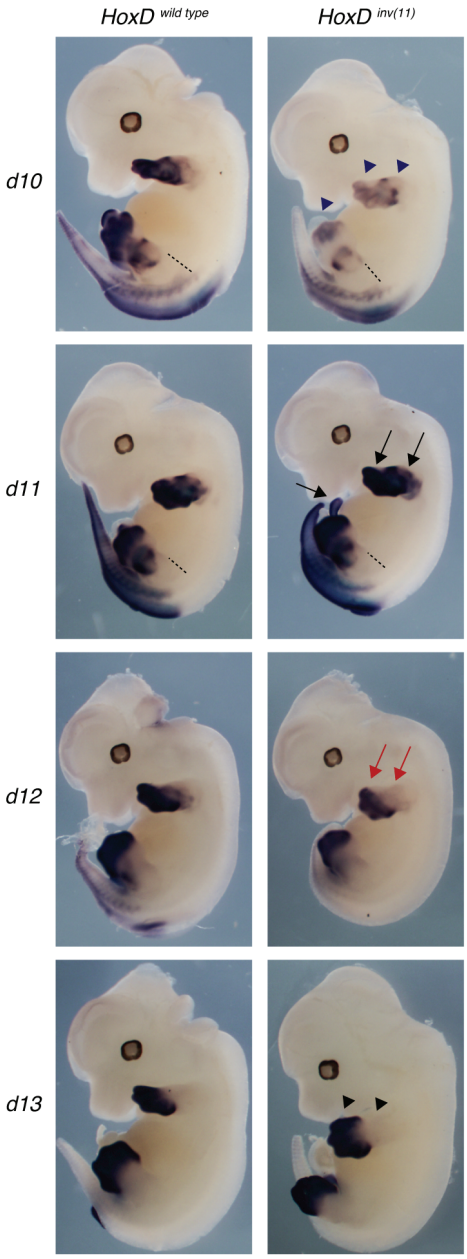

Figure 11: Whole mount In-Situ Hybridization done in E12.5 embryos with four antisense probes revealing the domain of expressions of *Hoxd10*, *Hoxd11*, *Hoxd12* and *Hoxd13*. Wild types and homozygous inv(*11*) littermates embryos were used to produce in parallel these WISH. *Hoxd10* expression in the inv(*11*) allele appears quantitatively reduced both in the trunk along the AP axis and through different secondary structures where it is usually concomitantly expressed with *Hoxd11*, e.g. the limb autopod and the zeugopod. Moreover, *Hoxd10* is not detected in the caecum of the mutant inv(*11*) embryo (see black arrowheads). Still, its nested domain of expression along the AP axis is unmoved, *Hoxd10* remains express in an anterior position compared to *Hoxd11* in the mutant.). In a similar manner to what is observed with *Hoxd10*, the inversion of the *d11* locus causes a pronounced reduction of *Hoxd12* throughout the trunk and in the developing limbs (red arrows). On the contrary, *Hoxd11* expression is unmoved along the AP axis as well as in secondary structures (caecum, autopod and zeugopod, see black arrows). Finally, *Hoxd13* expression is unchanged in the mutant allele and displays the usual autopod-exclusive staining in the limb, marking the five developing digits (black arrowhead) as well as the developing genital tubercule (not visible). Data produced by C. Bochaton and F. Darbellay, unpublished. The black dashed lanes mark the anterior extremity of the hindlimb.

3.1.2.2 Inv(*11*) in the autopod

The apparent connections between the inverted *Hoxd11* locus and its two flanking genes *d12* and *d10* were intriguing. Two preliminary results led us to consider an unexpected feature of *Hoxd11*. In a first instance, a strand specific transcriptome produced from the caecum was re-analyzed (Delpretti et al., 2013) and revealed that a significant signal exists downstream of *Hoxd11*, in the 2.5 kb annotated as *d11* 3'UTR but also extending further downstream towards and over *Hoxd10* promoter. Because no alternative start sites are reported between *d11* and *d10* (Montavon, 2008), we presumed that this transcriptional activity was generated from the main promoter of *Hoxd11*. This observation implies that an important transcriptional leakage occurs over the 3'UTR of *d11*, a sequence that appeared surprisingly inefficient at stopping the transcriptional activity fired by the *d11* promoter. In a latter instance, 3' and 5' RACE screens performed in the Duboule laboratory confirmed the existence of non-canonic *Hoxd* transcripts spliced across different *Hoxd* genes. Among others, we remarked spliced transcripts between RX and the second exons of *Hoxd10* and *Hoxd9*, as well as between the two exons of *Hoxd11* and the second exon of *Hoxd10* mRNA (Mainguy et al., 2007; Montavon, 2008). This observation attest of the presence of polycistronic *Hox* mRNA of unknown function. Altogether, we speculated that the important transcriptional leakage displayed by *Hoxd11* in a wild type situation could also exist in the inv(*11*) allele. Because of the inversion of the locus, we proposed that this transcriptional activity may be relocated directly towards *Hoxd12* on the anti-*HoxD* strand and might therefore interfere with *Hoxd12* transcription.

Hence, we decided to address these transcriptional issues to better understand the perturbations displayed by inv(*11*). We did so by generating a series of strand specific transcriptomes and we worked first with the autopod of inv(*11*) embryos. We chose this structure because of the important quantitative perturbation displayed by *Hoxd12*, as previously reported by WISH. Moreover, *Hoxd* regulation in the autopod is well documented and has been extensively studied in the Duboule laboratory. Advantageously, the complete dependence of the posterior *Hoxd* genes expression (*Hoxd13*, *d12*, *d11* and to a lower extent *d10)* towards long-range enhancers located within the C-DOM was a crucial justification for the choice of this experiment (Montavon et al., 2008; Spitz et al., 2003). We were thus certain to rule-out the relocation of putative autopod enhancers located over inverted locus as a source of this perturbations. We were therefore confident to be in a position to attribute any changes strictly to the transcriptional inversion of *Hoxd11*.

Following these rationales, we generated strand specific RNAseq out of control and mutant autopod dissected from the forelimbs of E12.5 littermates (Fig. 12A). Because of limited stock of mutant animals at the time of this experiment, we decided to work with trans-heterozygous mutant embryos, carrying on one chromosome the desired inv(*11*) allele and a balancer deletion *HoxD* $^{\mathrm{del}(8\text{-}13)\mathrm{RXII}}$ on the other chromosome. The RNAseq generated from the balancer control revealed the expected pattern of expression of *Hoxd* gene in this secondary structure. Quantitatively, a peak of transcripts is scored over *Hoxd13* (Fig. 12B), whereas the expected gradient is observed from *Hoxd13* up to *Hoxd10* (Montavon et al., 2008). Interestingly, we recorded a limited but significant transcriptional activity of *Hoxd11* leaking towards *Hoxd10* (Fig. 12B, black arrow), in agreement to what was previously reported in the caecum (Delpretti et al., 2013). Of note, a similar transcriptional leakage was also apparent in this dataset between *Hoxd12* and *Hoxd11* (Fig. 12B, black arrows). The transcription generated by the anti-*HoxD* strand was null over this fraction of the *HoxD* cluster and found restricted to the *Evx2* gene.

In the mutant trans-heterozygote dataset, the quantitative predominance of *Hoxd13* was still clearly apparent, hence confirming the observation that *d13* transcription is unaffected by the inversion of *Hoxd11*. The normalized signal collected over the inverted *Hoxd11* gene confirmed an augmentation

of the amount of *d11* transcript, in accordance with the observation made in the WISH experiments with homozygous mutants. Yet, we were not able to deduce, from this dataset, whether this quantitative difference is caused by the new artificial sequence composing *Hoxd11* 3'UTR, which may change the stability of the transcript or whether the reposition of the *d11* promoter in the inv(*11*) is causing an upregulation of *Hoxd11* transcription.

This dataset confirmed the abnormal distribution of transcripts over *Hoxd12* and *Hoxd10*. As speculated earlier, we found first that the transcriptional leakage emitted by the inverted *d11* was substantial and progressed over the anti-*HoxD* strand towards *Hoxd12*. The principal part of this signal seemed to be stopped just before it reached *d12* 3'UTR (Fig. 12C, black arrowhead). A lower fraction of this transcriptional activity extends more centromeric, covers *Hoxd12* antisense and stretches almost up to the 3'UTR of *Hoxd13*. From this dataset, it appears that this artificial activity affects drastically the expression of *Hoxd12*, diminishing it by at least a factor of fivefold compared against the wild type control, through mechanisms that remain to be discussed and further explored. On the other hand, it was more complicated to make conclusive observations with *Hoxd10*. Undeniably, the relocation of *d11* on the anti-*HoxD* strand results in a complete loss of leakage running from *Hoxd11* and covering *Hoxd10* promoter. But because of the overall low level of expression of *Hoxd10* in this tissue, it was not pertinent to address here to what precise extent *Hoxd10* was affected by the inversion of *d11*.

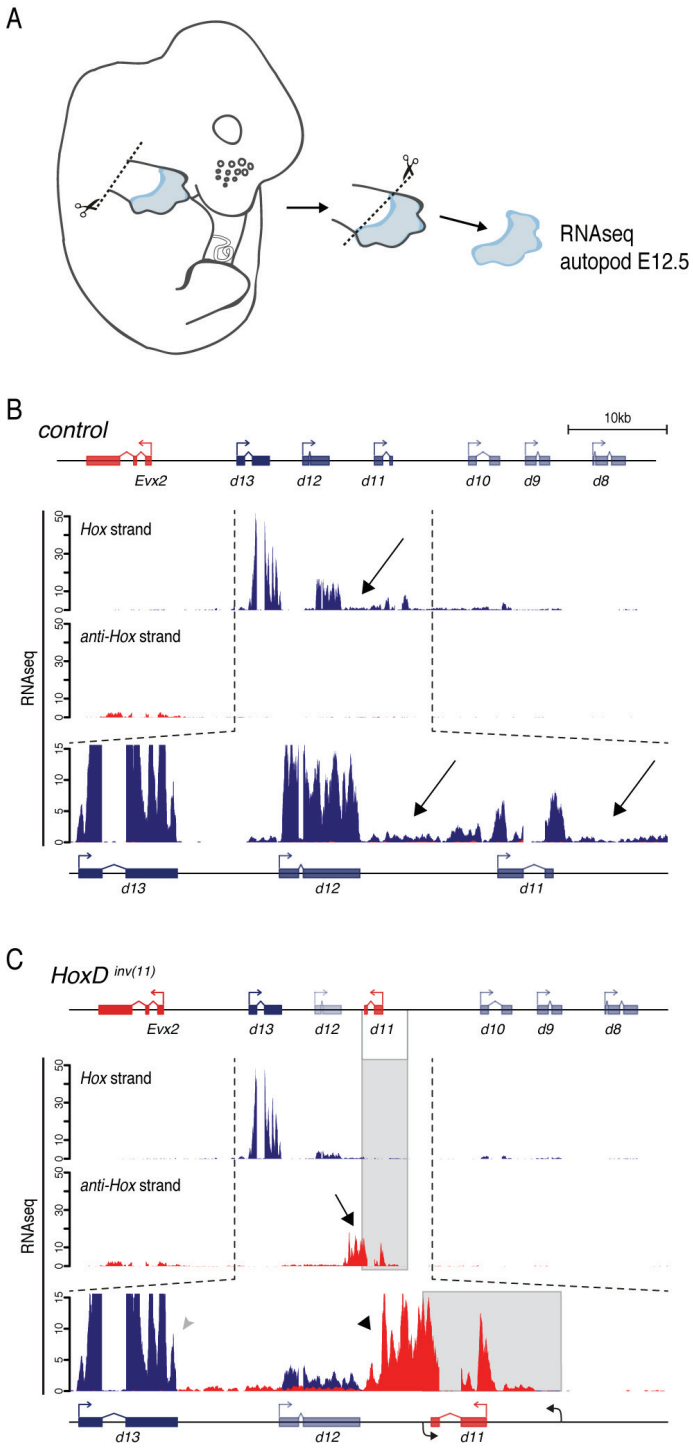

Figure 12: Transcriptional profiles of the posterior *HoxD* cluster in the inv(*11*) allele in E12.5 autopods. (A) Schematic representation of the sampling strategy used with E12.5 embryos to generate these datasets. (B) Normalized strand specific mRNA RNAseq signal generated from the heterozygous control *HoxD* $^{del(8\text{-}13)RXII/+}$ autopods. The posterior *Hoxd* genes display the characteristic reverse collinearity with *Hoxd13* accounting for the majority of the signal at the locus (Montavon, 2008). Both *Hoxd12* and *Hoxd11* display a similar tendency to produce transcriptional leakage stretching over their 3' *Hoxd* neighbor, respectively *Hoxd11* or *Hoxd10* (black arrows). Interestingly, this behavior is not detected between *Hoxd13* and *Hoxd12*, *d13* transcription being efficiently stopped in its 3' UTR. Although masked by the scale, the gene *Evx2* is active and transcribed from the anti-*Hox* strand. (C)

Normalized strand specific mRNA RNAseq signal extracted from the trans-heterozygote mutant *HoxD* del(*8-13*)RXII/inv(*11*). The inversion of the *Hoxd11* locus does not affect the expression of *Hoxd11* but results in the reallocation of the transcriptional leakage emitted by *Hoxd11* on the anti-*Hox* DNA strand leaking towards *Hoxd12* (black arrow). The majority of this artificial transcript is stopped downstream of *Hoxd12* (black arrowhead). A fraction of transcripts continues towards *Hoxd13* and thus covers the antisense DNA strand of the gene *Hoxd12*. This observation correlates with a significant reduction of transcripts matching *Hoxd12* mRNA. This configuration of convergent transcription and its outcome suggest a mechanism of transcriptional interference at the *Hoxd12* locus. *Hoxd13* transcription is unaffected and untouched by the leakage emitted by *Hoxd11* which seems to stop entirely 3' of *Hoxd13* (grey arrowhead). The normalized signal mapped to the *Hox* DNA strand, corresponding to the *Hoxd* genes coding strand, is depicted in blue whereas the normalized signal mapped to the anti-*Hox* DNA strand, in red, corresponds to the *Hoxd* genes antisense DNA strand.

3.1.2.3 Inv(*11*) in the metanephros

We next transported our attention to another embryonic structure which is expressing high dose of *Hoxd11*, the metanephric kidneys. Important points remained to be experimentally addressed with the in(*11*) allele and could be tackled advantageously in this developing organ. First, we wanted to work with a secondary structure expressing higher level of *Hoxd10* than the autopod in order to develop a better understanding of the consequence of the inversion of *d11* over this gene. Second, based upon our previous conclusions, we wondered whether the inverted locus of *Hoxd11* and the transcriptional activity emitted towards *Hoxd12* would robustly result in a reduction of *Hoxd12*. Because *Hoxd11* is expressed in the kidneys in an important fraction of mesenchymal cells that are not expressing *Hoxd12*, we thus questioned whether a part of the transcriptional activity emitted by the inverted *d11*, and running towards *Hoxd12*, might indirectly result in a significant gain-of-expression of *Hoxd12*. Lastly, the preliminary WISH results obtained from the inv(*11*) allele did not reveal any strong perturbations in the patterning of *Hoxd13* in the inv(*11*) allele. Yet, other types of genetic perturbations, e.g. perturbing the long-range contacts between *Hoxd13* and the centromeric regulatory landscape, are known to trigger gain-of-expression of the dominant negative *Hoxd13* gene in *Hoxd11* patterned structures, resulting in viable hypomorphic phenotypes in heterozygous mutant (Tschopp and Duboule, 2011). Therefore, we wanted to study the inv(*11*) allele in a context where the bimodal regulation of *HoxD* is different from the autopods, i.e. *d11* is strongly transcribed whereas *d13* is completely silenced, and observe how *HoxD* genetic regulation can cope with this perturbation.

Following these motivations, we generated strand specific transcriptomes from embryonic kidneys, micro-dissected at E13.5 from control and mutant embryos (Fig. 13A). We followed the same approach for this experiment as for the RNAseq produced in the autopod and presented above. We worked with the same set of trans-heterozygous mutant embryos carrying the inv(*11*) mutation and the del(*8-13*)RXII balancer. RNA was also processed from the balancer allele to be used as a direct comparison against the inv(*11*)/del(*8-13*)RXII transcriptome. This control dataset confirmed the conceptual validity of our approach as we recorded a high transcription of the central fraction of the *Hoxd* genes, from *d11* to *d8*, cumulating over *Hoxd11* (Fig. 13B). We also observed the expected transcriptional leakage issued by *Hoxd11* and covering the intergenic region downstream of *Hoxd11* as well as the promoter of *Hoxd10* (Fig. 13B, black arrow). *Hoxd12* transcripts appeared quantitatively reduced compared to the autopod transcriptome previously discussed. Interestingly, this reduction of *Hoxd12* transcript correlates with the absence of leakage extending from *d12* 3'UTR over *Hoxd11*. At the posterior extremity of the *HoxD* cluster, neither *Hoxd13* nor *Evx2* showed any significant transcription.

The strand specific RNAseq obtained from the inv(*11*) mutant kidneys revealed a couple of noteworthy features. First of all, we found that *Hoxd11* was still expressed in the kidney following the inver-

sion of *d11* locus (Fig. 13C). As anticipated from the autopod RNAseq, the leakage towards the 5' extremity of the *HoxD* cluster was visible on the anti-*HoxD* strand, emanating from the inv(*11*) locus (Fig. 13C, black arrow). In accordance with the previous observations, the amount of normalized signal mapped to the *Hoxd11* gene was higher in the inv(*11*) than in the control transcriptome. Importantly, the leakage generated by the inv(*11*) locus seemed to stop principally at the same locations already observed in the E12.5 autopod transcriptome, i.e. a few kilobases downstream of *Hoxd12* and finally downstream of *Hoxd13* (Fig. 13C, grey and black arrowheads). Yet, this rather limited transcriptional activity correlates again with a pronounced reduction of *Hoxd12* transcripts.

The fact that *Hoxd13* remained completely silenced, its regulation un-challenged by the transcriptional activity generated by the inv(*11*) allele was in agreement with both the initial WISH experiments and the absence of homeotic mutations displayed by heterozygous adult mutants. At the other extremity of the *Hoxd11* locus, a slight change in *Hoxd10* was also visible, particularly noticeable over the first exon of this gene (Fig. 13C, green arrow). Because the whole promoter of *Hoxd10* is cleansed of the leakage generated by *Hoxd11*, we interpreted this result as yet another illustration of the capability of *Hoxd11* to affect, through unknown mechanisms, the transcription of a gene located downstream its own transcriptional unit. These different quantitative observations are confirmed when computing the normalized FPKM values of the genes directly flanking the inverted locus (Fig. 13D).

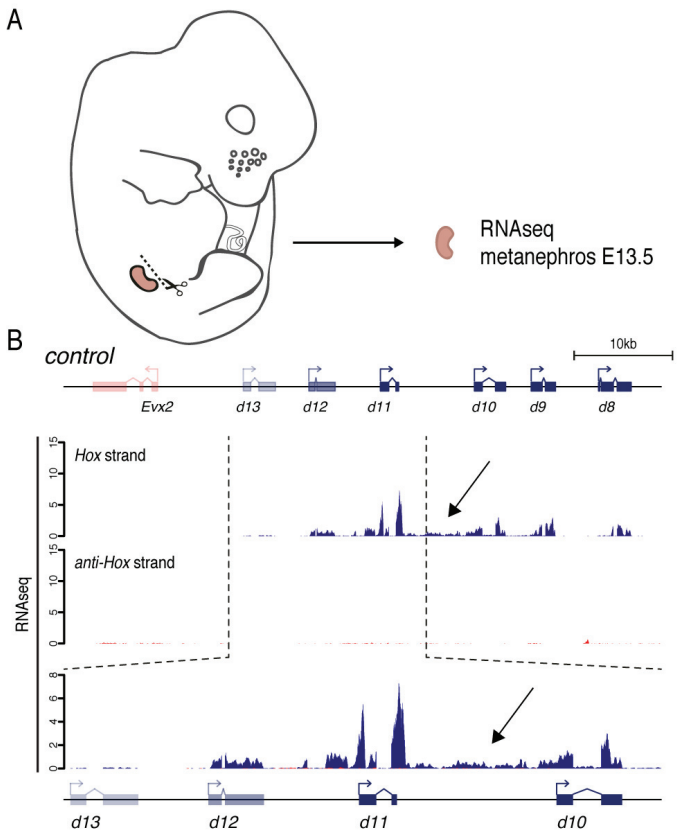

Figure 13: Transcriptional profiles of the posterior *HoxD* cluster in the inv(*11*) allele in E13.5 metanephros. *See captions next page.*

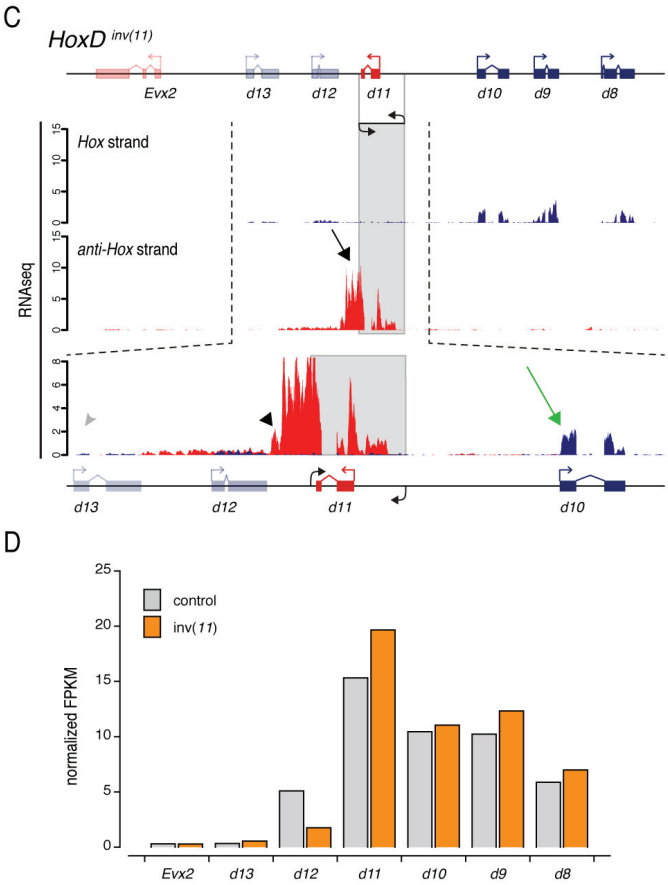

Figure 13: Transcriptional profiles of the posterior *HoxD* cluster in the inv(*11*) allele in E13.5 metanephros. (A) Schematic representation of the sampling strategy used with E13.5 embryos to generate these datasets. (B) Normalized strand specific mRNA RNAseq signal generated from the heterozygous control *HoxD* $^{\text{del(8-13)RXII/+}}$ E13.5 metanephros. The developing kidneys present a transcriptional landscape very different from the E12.5 autopod (Fig. 12). Central *Hoxd* genes, from *Hoxd8* to *Hoxd11*, are highly transcribed whereas the posterior *Hoxd13* and *Evx2* genes are silenced. The reduced amount of *Hoxd12* transcript reflects its restricted distribution within the developing kidney. In contrast with the RNAseq generated from E12.5 autopods, *Hoxd12* does not generate any visible transcriptional leakage from its 3'UTR towards *Hoxd11*, a behavior still observed with *Hoxd11* over *Hoxd10* (black arrow). (C) Normalized strand specific mRNA RNAseq signal extracted from the trans-heterozygote mutant *HoxD* $^{\text{del(8-13)RXII/inv(11)}}$. The inversion of the *Hoxd11* locus results in a limited increase of *Hoxd11* signal probably caused by the artificial 3'UTR of this gene resulting from the inversion of the locus. *Hoxd11* transcriptional leakage on the anti-*HoxD* DNA strand is maintained in this embryonic structure (black arrow). The majority of this transcript is stopped downstream of *Hoxd12* (black arrowhead) whereas a fraction continues towards *Hoxd13* and thus covers the *Hoxd12* antisense, correlating with an important reduction of *d12* mRNA signal. This configuration of convergent transcription and its outcome suggest a mechanism of transcriptional interference at the *Hoxd12* locus. *Hoxd13* repression appears unaffected, the posterior *Hoxd* gene remaining firmly silenced in this mutant (grey arrowhead). The reallocation of *Hoxd11* leakage appears to change qualitatively the signal recorded at *Hoxd10*, notably over the first exon of this gene (green arrow), even though the quantitative variation does not appear significant. (D) Normalized quantification (FPKM) of the posterior *Hoxd* genes computed from the RNAseq dataset presented in B and C. Note the decrease in *Hoxd12* and the increase of *Hoxd11* in the inv(*11*) allele. The central *Hoxd* genes are not significantly impacted. The normalized signal mapped to the *Hox* DNA strand, corresponding to the *Hoxd* genes DNA coding strand, is depicted in blue whereas the normalized signal mapped to the anti-*Hox* DNA strand, in red, corresponds to the *Hoxd* genes antisense DNA strand.

3.1.3 Conclusions

We conclude from this examination of the inv(*11*) allele the following points:

a. The inversion of the *Hoxd11* locus does not result in any significant dysregulation of *Hoxd* activations. This inversion is globally well tolerated by the adult animal. Notably, the spatial collinearity regulation along the AP axis seems maintained, except for *Hoxd12*. The bimodal regulation controlling notably *Hoxd13* in secondary structures is resistant to the transcriptional stresses caused by this reconfiguration.

b. The inversion of the *Hoxd11* locus affects quantitatively its two neighboring *Hoxd*. Both *Hoxd12* and, to a lower extent, *Hoxd10* transcription are affected by unknown mechanism(s), likely to involve transcriptional interference provoked by the important transcriptional leakage displayed by *Hoxd11*.

c. The artificial reorientation of *Hoxd11* transcriptional activity, directed towards the 5' extremity of the cluster and therefore orientated parallel to the collinear activation of the *HoxD* cluster, did not result in any gain-of-expression of the *Hoxd* genes located posterior to the inversion.

3.2 Inversion *HoxD* inv(*11-12*)*d13lacZ*

3.2.1 General observations

The allele *HoxD* *inv(11-12)d13lacZ* consists of an inversion of 15 kb which encompasses both *Hoxd11* and *Hoxd12* (Kmita et al., 2000a), as well as cis-regulating elements that may be located in the intergenic chromatin such as the CTCF site located in between *Hoxd12* and *Hoxd13lacZ* (Fig. 14, Fig. 8). This inversion was obtained by a Cre/Lox recombination. The first *loxP* site which was used is located in between *Hoxd11* and *Hoxd10*, 1.8kb more telomeric than the breakpoint of the inv(*11*) but still within *Hoxd11* 3'UTR, whereas the second loxP site used lies directly downstream of a *lacZ* reporter (Fig. 14). This *lacZ* sequence is knocked-in the first exon of *Hoxd13* and allowed the experimenters to monitor the perturbations of *Hoxd13* expression caused by the intra-cluster inversion. Indirectly, this knock-in results in a loss-of-function of *Hoxd13*, a situation tolerated by heterozygous mutant adult animals (Beccari et al., 2016). *Hoxd13* loss-of-function would prove to be also extremely advantageous to neutralize any gain-of-expression of *d13* as a consequence of the *d11-d12* inversion.

Indeed, this inversion was reported to disrupt the regulation of the posterior *Hoxd13lacZ* reporter in the intestinal hernia (caecum), causing a strong gain-of-expression of this gene and to a lower extent of the neighboring *Evx2* gene (Delpretti et al., 2013; Kmita et al., 2000a). In the original publication, the authors postulated that a polarized silencer element, located in between *Hoxd12* and *Hoxd13lacZ*, fails to insulate this gene when inverted, leading to *d13lacZ* gain-of-expression in the caecum of mutant embryo (Kmita et al., 2000a). This interpretation was then challenged when strand specific RNAseq became available. Transcriptome datasets were generated from caeca micro-dissected from control *HoxD* *d13lacZ* and *HoxD* inv(*11-12*)*d13lacZ* E13.5 embryos. Surprisingly, Delpretti et al. reported the existence of a significant transcriptional activity leaking from the *d11* locus in the inv(*11-12*)*d13lacZ* allele (Delpretti et al., 2013). These transcripts extend towards the 5' extremity of the cluster over the anti-*HoxD* DNA strand, covering *Hoxd13lacZ* antisense and continuing in the direction of *Evx2*. It was thus proposed that the loss of repression/gain-of-expression of the reporter *d13lacZ* and of *Evx2* would be driven by mechanisms linked with this artificial transcriptional activity and not by the reallocation of putative non-coding regulating elements.

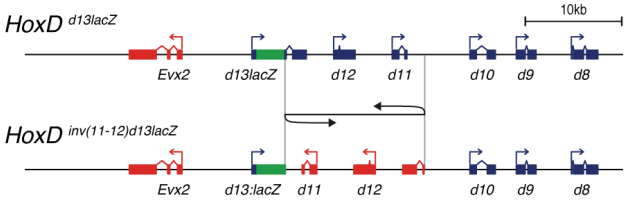

Figure 14: Schematic representation of the allele *HoxD* inv(*11-12*)*Hoxd13lacZ*. Inv(*11-12*)*d13lacZ* was produced by a Cre/Lox-mediated inversion from the *HoxD* *d13lacZ* allele which contains a *lacZ* cassette and a loxP site recombined within the first exon of *Hoxd13*. A second loxP site is located between *Hoxd11* and *Hoxd10* (Kmita et al., 2000a). The inversion is therefore separating *Hoxd13* into two entities, on the one side the promoter, a fraction of *Hoxd13* first exon and the *lacZ* cassette and, on the other side, the rest of the *Hoxd13* gene (second part of the first exon, the intron and the second exon). The *d11_d10* loxP site is positioned within *Hoxd11* 3'UTR, the annotation of this gene being here restricted to its two exons for an easier visualization.

In view of the conclusions reached with the allele inv(*11*), i.e. the functional inversion of *Hoxd11* results in the reduction of *Hoxd12* transcript without affecting the regulation of *Hoxd13,* we were specifically interested in studying this allele. Therefore, we investigated first whether the anteriorization of *Hoxd13lacZ* and *Evx2* in the context of the inversion of the locus *d11-d12* was a general feature of this allele or if this perturbation was specific to the intestinal hernia. Thus, we evaluated the expression patterns of two key genes involved in this reconfiguration, *Hoxd11* and *Hoxd13lacZ*. This point was addressed by a series of WISH experiments done in mutant homozygous embryos at E12.5 using the floxed allele *Hoxd13lacZ* as control (Fig. 15).

We first compared the signal observed in homozygous *d13lacZ* control embryos using two antisense ISH probes revealing respectively *d13lacZ* and *Hoxd11* expression patterns. As expected, *Hoxd11* signal along the AP axis followed the pattern described in the literature at this stage. *Hoxd11* is expressed with a pattern extending from an area located slightly anterior to the center of the proximal hindlimb, at the level of the prevertebra (pv) 27 – the first sacral vertebra – up to the tail bud at the caudal extremity of the embryo (Gérard et al., 1993, 1996; Izpisúa-Belmonte et al., 1991; Zákány et al., 1997). Instead, the *lacZ* signal recapitulated extensively the pattern of expression of *Hoxd13* along the AP axis and could be therefore used as a proxy to visualize *d13lacZ* expression, extending from the pv31 (first caudal vertebra) up to the posterior extremity of the embryo tail (Izpisúa-Belmonte et al., 1991; Tschopp et al., 2009). Interestingly, this clear difference between the position of the patterns of *d11* and *d13lacZ* was strongly reduced in the inv(*11-12*)*d13lacZ* embryo (Fig. 15). Indeed, upon inversion of *d11-d12*, the *lacZ* WISH signal appeared to be shifted and expressed higher up in the trunk, mimicking the signal obtained in the mutant embryo with *d11* WISH probe.

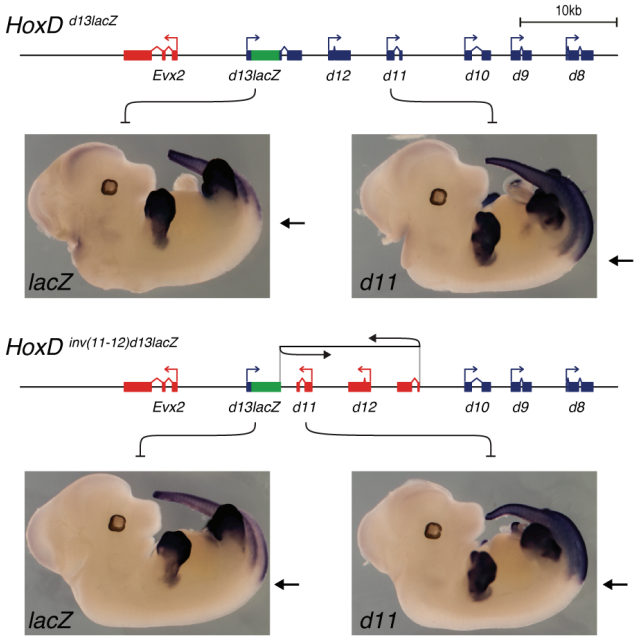

Figure 15: Whole mount In Situ Hybridization of the reporter gene *Hoxd13lacZ* and *Hoxd11* in *HoxD* ^{d13lacZ} and *HoxD* $^{inv(11\text{-}12)d13lacZ}$ E12.5 mutant homozygous. In the control *d13lacZ* allele, the anterior limits of *Hoxd11* and *d13lacZ* expression domains are clearly separated along the AP axis, resulting from the spatial collinear activation of the *HoxD* cluster (black arrows). In the mutant inv(*11-12*)*d13lacZ*, the patterns of *Hoxd11* and *d13lacZ* along the trunk are merged, the expression of *Hoxd13lacZ* is anteriorized, suggesting a simultaneous activation of these two genes during the extension of the AP axis. C. Bochaton and F. Darbellay, unpublished.

The shift of pattern of *Hoxd13lacZ* along the AP axis is also apparent with an X-gal staining processed at the same developmental stage (Fig. 16A) which allows to compare directly the position of expression of the *Hoxd13lacZ* reporter in the two alleles. This anteriorization of *Hoxd13lacZ* appears to occur in both the spinal cord and in the lateral mesoderm of the mutant embryo, an observation that led us to conclude that the inversion of the *d11-d12* locus disrupts the regulation of *Hoxd13lacZ* also along the AP axis and results in the expression of this gene in the same metameres of the embryo as *Hoxd11*. Lastly, we noticed that the inversion of the *d11-d12* locus resulted in yet another effect upon the expression of the *Hoxd* gene concerned. When repositioned *Hoxd11* at the position of a putative *Hoxd12* (moved from the 7th to the 8th genetic position, counting from the 3' anterior extremity), *d11* pattern of expression moved in a subtle but reproducible way and was posteriorized, constrained posteriorly to the center of the hindlimb. WISH effectuated with a probe specific to *Hoxd12* confirmed the exchange of *d11*/*d12* nested expression domains along the AP axis (data not shown).

Overall, these observations motivated us to investigate more closely these perturbations by using recent genomic tools and technologies. Thus, we produced a series of micro-dissections of posterior trunk sampled from mutant embryos at E12.5 in order to obtain biological material to be used for further experiments. We set-up a dissection pattern centered at the level of the sacral vertebrae and corresponding to the area where the gain of expression of *d13lacZ* appeared stronger in the inv(*11-12*)*d13lacZ* mutant, compared to the *d13lacZ* control (Fig. 16A). Using these samples, we generated a series of strand-specific transcriptomes from both control wild type and homozygous inv(*11-12*)*d13lacZ* embryos. Focusing our analysis on the posterior extremity of the *HoxD* cluster, the whole locus appeared to be transcribed in the mutant dataset (Fig. 16B). As expected from both the WISH and the X-gal experiments presented above, we obtained a strong signal corresponding to the reporter *d13lacZ* together with important quantity of transcripts matching both *Hoxd12* and *Hoxd11*, now relocated on the anti-*HoxD* DNA strand. When compared against the wild type control (Fig. 16C), the increase of *d12* transcript was apparent whereas the global level of *Hoxd11* was globally maintained.

Crucially, we found that the inversion was breaking the long transcript produced by *Hoxd11,* already discussed in the context of the inv(*11*) allele and reported in the caecum transcriptome obtained from the same mutant (Delpretti et al., 2013). In a wild type configuration, this transcript was observed extending from *d11* 3'UTR towards and over *Hoxd10* promoter (Fig. 16C, blue arrow). Following the inversion of the locus *d11-d12*, *d11* generates transcriptional leakage across its 3'UTR, on the anti-*HoxD* strand, covering *d13lacZ* antisense DNA strand and progressing towards *Evx2* (Fig. 16B, red arrow). This important artificial activity correlates with the activation of the posterior extremity of the *HoxD* cluster (Delpretti et al., 2013). Interestingly, we also found that the relocated *Hoxd12* was displaying a similar tendency to leak towards its 3' neighbor *Hoxd11* in this genetic configuration (Fig. 6B, black arrow), the signature of an increased transcriptional activity. However, the degree of leakage was observed here was lower than the one emitted from *Hoxd11*.

It seems also that the relocation of *d11* transcriptional activity towards the 5' extremity of the cluster resulted in a noticeable increase of *Hoxd10* transcript (Fig. 16B). This observation was confirmed by WISH in the caecum of E13.5 mutant embryos (Kmita et al., 2000a). Interestingly, *Hoxd9* seemed to follow the same tendency in this dataset, its normalized signal being also slightly more important in the inv(*11-12*)*d13lacZ* mutant. However, it was then not possible to clearly attribute these relative quantitative changes solely to the inversion. We were indeed first not in possession of a transcriptome obtained from a floxed *HoxD* ^{d13lacZ} allele, a dataset which would provide an effective control for the stability

and dynamic of *Hoxd13lacZ* mRNA. Second, the anterior boundary of our sampling was not fixed as stringently as the posterior limit and could therefore bias the distribution of the signal specific to central *Hoxd* genes.

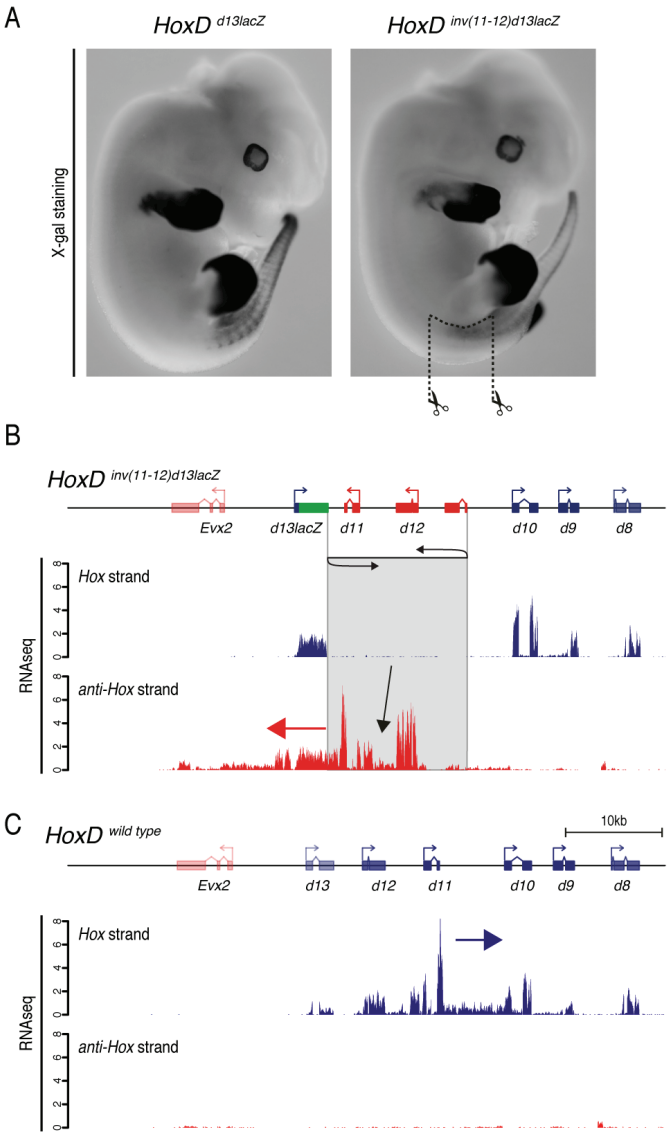

Figure 16: Transcriptional profiles of the posterior *HoxD* cluster in the posterior trunk of E12.5 inv(*11-12*)*d13lacZ*. (A) X-gal staining of E12.5 heterozygous *HoxD* ^{d13lacZ} and *HoxD* $^{inv(11\text{-}12)d13laZ}$ embryos. The anteriorization of the expression of *d13lacZ* in the inv(*11-12*)*d13lacZ* mutant is clearly visible, extending up to the base of the hindlimb. Dashed lanes represent the micro-dissected sampling, centered at the level of the sacral vertebras, used to generate the RNAseq presented in the next panel. (B) Stranded RNAseq generated from homozygous inv(*11-12*)*d13lacZ* mutant E12.5 posterior trunk (rRNA-depleted). The transcriptome reveals the strong transcriptional leakage released by the inverted *d11-d12* locus and covering the *HoxD* antisense DNA strand over the posterior extremity of the cluster, up to the 3' of *Evx2* (red arrow). The expression of *Hoxd13lacZ* is clearly visible, despite the convergent transcription emitted by *Hoxd11*. *Hoxd12* displays a similar behavior, albeit reduced in intensity, of leaking over *Hoxd11* (black arrow). (C) Control stranded RNAseq generated from wild type E12.5 posterior trunk (rRNA-depleted). The transcriptional leakage of *Hoxd11* over *Hoxd10* is visible (blue arrow). A limited but significative transcription of *Hoxd13* is also apparent, probably caused by a contamination of *Hoxd13* positive cells present in the micro-dissected tissues. This signal prompted us to move to another embryonic structure displaying

a more robustly fixed set of central *Hoxd* genes to better apprehend the mechanisms causing the upregulation of the posterior *HoxD* cluster. The normalized signal mapped to the *Hox* DNA strand, corresponding to the *Hoxd* genes coding strand, is depicted in blue whereas the normalized signal mapped to the anti-*Hox* DNA strand, in red, corresponds to the *Hoxd* genes antisense DNA strand.

However, the control wild type transcriptome revealed a limited but significant signal recorded over *Hoxd13* (Fig. 16C). This observation indicated the presence of *d13* positive cells within our sampled tissues. These contaminating cells might originate from several sources, not mutually exclusive, such as the presence of *d13* positive cells projecting in the caudal extremity of the spinal cord (Dollé et al., 1994; Tschopp et al., 2012) or the presence of traces of uro-genital or cloacal tissues attached to the micro-dissected fragment of posterior trunk. These dissections were moreover complicated by the fact that minimal variations in the individual embryonic development stage at the time of sampling, something frequently observed within a murine litter, might shift significantly positions and relative patterns of HOX domains, as the embryo undergoes a very dynamic three-dimensional development. These limited but significant population of *Hoxd13* positive cells presents in the control weakens any further conclusions made with this allele. Indeed, such a sampling plan along the AP axis would make impossible to detect and attribute *Hoxd13lacZ* regulative changes exclusively to the transcriptional reconfiguration of *d11* and *d12*, because any specific perturbations would be diluted by the signal coming from *d13lacZ* positive cells responding to their intrinsic regulations.

Hence, to reinforce our sampling and get rid of spatio-temporal variations, we looked for another model structure displaying a more stable transcriptional *Hoxd* pattern throughout embryonic time and a robust shape to facilitate the sampling. Preferably, such a structure would express high level of central *Hoxd11* to *Hoxd8* genes and would show a strong and complete repression of posterior *Hoxd13*. For these reasons, metanephric kidneys were selected, as group *Hox11* genes has been reported to be necessary to pattern and build a functional organ (Davis et al., 1995; Patterson et al., 2001; Wellik et al., 2002) whereas an important gain of *Hoxd13* in embryonic metanephros results in a complete and lethal agenesis of this organ in newborn animal (Kmita et al., 2000b). Finally, we also decided to use systematically homozygote *d13lacZ* floxed samples to control for any possible artefacts triggered by the *lacZ* sequence inserted within the *Hoxd13* gene. This synthetic sequence derived from a prokaryotic gene and we considered whether it could potentially impact the stability of the hybrid *d13lacZ* mRNA or partially disrupt the endogenous regulation of the locus.

3.2.2 Inv(*11-12*)*d13lacZ* in the metanephros

We first performed an X-gal staining in E13.5 metanephros dissected from mutant and control embryos to assess *lacZ* expression in embryonic kidneys (Fig. 17A). In this simple experiment, homozygous inv(*11-12*)*d13lacZ* metanephros displayed a very strong X-gal staining revealing a gain-of-expression of the *d13lacZ* reporter. On the other hand, the *d13lacZ* floxed control allele showed a complete absence of staining, in accordance with the absence of *Hoxd13* expression observed in wild type situation (see Fig. 13B, 17A). This observation defines that the regulation of *d13lacZ* is affected by the inversion of the locus *d11-d12* in this embryonic tissue, in addition to the caecum and the AP axis.

We then produced a series of strand specific RNAseq datasets using these samples. In the control *HoxD* ^{d13lacZ} transcriptome, the peak of transcriptional activity is centered over *Hoxd8* to *Hoxd11* (Fig. 17B). We did not observe significant amount of *d13lacZ* transcript in this control dataset, reflecting the repression of the posterior *Hoxd13* gene in wild type metanephros. The characteristic leakage of *Hoxd11* throughout *Hoxd10* was also visible whereas *Hoxd12* is discreetly expressed (Fig. 17B). The transcriptome generated from the mutant inv(*11-12*)*d13lacZ* metanephros revealed a very different transcriptional panorama (Fig. 17C). We found the equivalent transcript leaking from the inverted *d11* towards *Evx2* (Fig. 17C, red arrow), correlating with an important gain of *d13lacZ* transcript (Fig. 17C, black arrow). Interestingly, the gain of *Evx2* appeared to be rather weak and limited in this mutant allele. The signal collected at this locus seemed indeed diluted within the transcriptional noise emitted from *Hoxd11*, inciting us to wonder to what extent this gene was actively transcribed from its own promoter in this configuration. Normalized FPKM values computed from these two datasets summed-up these quantitative observations (Fig. 17D) and confirmed that the gain of *Hoxd12* was concomitant with the reduction of *Hoxd11* transcripts. These values also revealed a high enrichment of *Hoxd13lacZ* transcripts, in the range of tenfold in the inv(*11-12*)*d13lacZ* mutant compared to the control. This strong enrichment was largely driven by the minimal quantity of *Hoxd13lacZ* transcript found in the control *HoxD* ^{d13lacZ} metanephros, thus confirming our decision to shift our experimental approach to this embryonic structure. Lastly, these quantified data also revealed an unexpected significant decrease of *Hoxd10* and *Hoxd9* transcripts, to a lower extent of *d8* and more generally of the whole anterior *Hoxd* genes.

To complete the description of these transcriptional profiles, we produced two series of ChIPseq targeted against H3K27me3 and H3K4me3 in these two alleles. We supposed that the analysis of the distribution of these two histones marks would allow us to address clearly the regulative changes caused by the inversion of the locus *d11-d12* over the posterior extremity of the *HoxD* cluster. In accordance with the RNAseq dataset presented above, we observed an enrichment of H3K27me3 between *Hoxd13lacZ* and a locus downstream (3') of *Evx2* in the control allele (Fig. 8A, orange arrowhead). The distal position of this H3K27me3 domain, located centromeric to *Evx2,* is consistent with the known distribution of PRC2 elements (Suz12, Ring1B) and H3K27me3 marks in situations where the *HoxD* cluster is either inactive or partially active (Schorderet et al., 2013; Soshnikova and Duboule, 2009). The sequence of the *lacZ* reporter was enriched for H3K27me3 and, expectedly, was also deprived of H3K4me3 (Fig. 18A, black arrow). This observation was in accordance with the silenced status of the reporter *d13lacZ* in this configuration. The profile of H3K4me3 revealed by this ChIPseq correlated well with the global transcriptional landscape of the *HoxD* cluster, discriminating the actively transcribed central *Hoxd* gene, from *Hoxd8* to *Hoxd11*, from the inactive posterior extremity of the cluster (*Hoxd13lacZ* and *Evx2*, Fig. 18A, green arrowhead). The balance of marks located at the *Hoxd12* locus further reveals the regulative heterogeneity of this gene, expressed in a minority of cells restricted to cortical mesenchyme of the developing kidneys (Patterson and Potter, 2004).

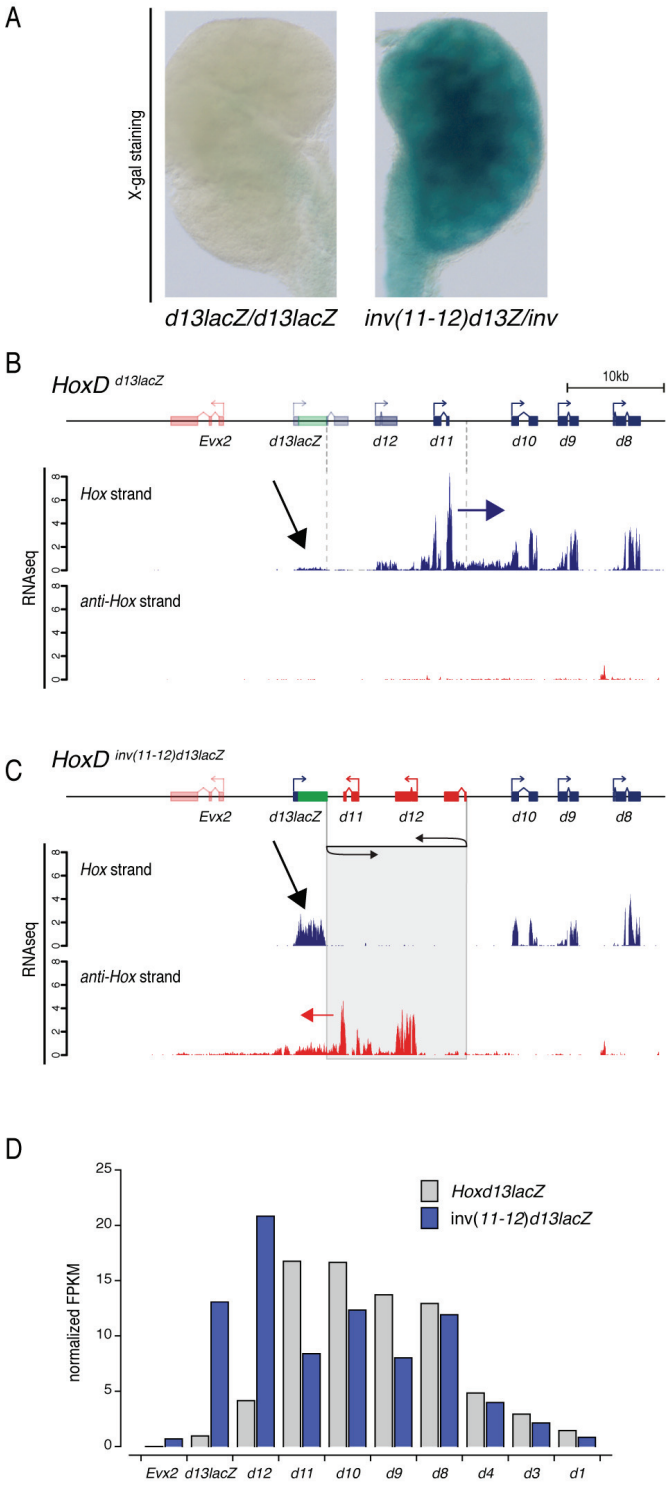

Figure 17: *Hoxd13lacZ* gain-of-expression in the inv(*11-12*)*d13lacZ* metanephros. (A) X-gal staining of E13.5 metanephros dissected from homozygous *HoxD* ^{d13lacZ} and *HoxD* $^{inv(11-12)d13lacZ}$ mutant embryos. No significative staining is observed in the control *d13lacZ*, therefore confirming the relevance of the choice of this embryonic structure to address the regulative interferences provoked by the inversion of the *d11-d12* locus. (B) Stranded RNAseq generated from control homozygous *d13lacZ* E13.5 metanephros (rRNA-depleted). The transcriptional activity observed here is characteristic of the developing kidneys with the majority of transcripts produced by the

central *Hoxd* genes (*Hoxd11* to *Hoxd8*) and specific to the *HoxD* DNA coding strand, depicted here in blue. Transcripts emitted by the anti-*HoxD* DNA strand, depicted in red, are globally nil except for a single lincRNA fired by *Hoxd8* TSS. As expected from the bimodal regulation of the *HoxD* cluster, *Hoxd13lacZ* is not transcribed (black arrow) whereas *Hoxd12* shows a reduced level of activity. The transcriptional leakage generated by *Hoxd11* and extending towards *Hoxd10* already detected in the posterior trunk RNAseq is here also apparent (blue arrow). (C) Stranded RNAseq generated from mutant homozygous inv(*11-12*)*d13lacZ* E13.5 metanephros (rRNA-depleted). The inversion of the *d11-d12* locus reallocates the strong transcriptional activity generated by *Hoxd11* on the anti-*Hox* DNA strand and leaks towards *Evx2* (red arrow), explaining the detection of this gene by WISH in the same structures as *Hoxd13lacZ* (Delpretti et al., 2013). This transcriptional activity correlates with the transcription of the reporter *Hoxd13lacZ*. (D) Normalized quantification (FPKM) of the posterior *Hoxd* genes computed from the RNAseq dataset presented in B and C. The gain of signal of *Hoxd13lacZ* and *Hoxd12*, translocated in the position of *Hoxd11*, are evident. Inv(*11-12*)*d13lacZ* appears also to quantitatively affect the transcription of the central/anterior *Hoxd* genes, from *Hoxd10* to *Hoxd3*.

The combination of H3K27me3 and H3K4me3 enrichment in the mutant inv(*11-12*)*d13lacZ* provided a valuable complement to the transcriptional profile described previously. On the one hand, H3K27me3 was absent from the activated *d13lacZ* reporter and the inverted *Hoxd12*, relocated in the position of *Hoxd11*. Both genes showed only minor enrichment of H3K27me3, quantitatively similar to the highly transcribed *Hoxd9*, in contrast to the significant enrichment recorded at this locus in the control datasets. On the other hand, both loci displayed high enrichment of H3K4me3 (Fig. 18B, black arrows and green arrowhead). After a careful calibration of the different ChIPseq signals based upon signal collected at biologically relevant locus outside of the *HoxD* cluster, we concluded that the global level of the H3K27me3 over *Evx2* did not significantly change (Fig. 18B, orange arrowhead).

This impression was further confirmed when considering the results obtained from a MACS2 analysis effectuated with these ChIP datasets to statistically test the distribution of the H3K27me3. Crucially, the distribution of these mark at the *Evx2* locus appeared to be conserved in the mutant allele. This result was finally supported by the fact that this whole locus, located 5' of *Hoxd13lacZ*, did not show any enrichment of H3K4me3, even though a limited transcriptional activity was detected by RNAseq. We thus concluded that this transcriptional activity was produced from either the activation of the *d13lacZ* locus or from the leakage emitted by the inverted *Hoxd11* but did not result from the activation of *Evx2 per se*. Furthermore, the H3K4me3 ChIPseq obtained from the mutant metanephros showed an unexpected enrichment in the central fraction of the mutant *HoxD* cluster. The whole series of *Hoxd* genes, from *d10* to *d8*, displayed an enhanced coverage of H3K4me3 (Fig. 18B), even though the transcriptional output of these genes was quantitatively reduced (Fig. 17D).

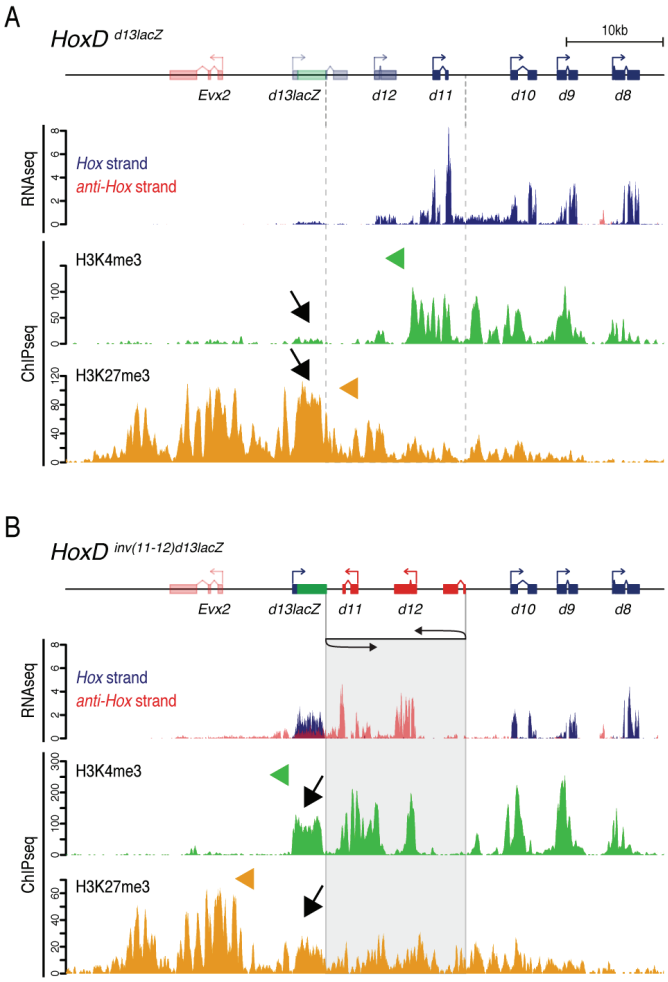

Figure 18: The inversion of the *d11-d12* locus disrupts the endogenous regulation of *d13lacZ* in E13.5 metanephros. (A) RNAseq and ChIPseq signals specific to the histone modifications H3K4me3 (in green) and H3K27me3 (in orange) in control homozygous *HoxD* ^{d13lacZ} E13.5 metanephros. H3K4me3 covers the active *Hoxd* genes (green arrowhead) whereas H3K27me3, a read-out of the repressing activity of PcG, marks the posterior silenced extremity of the *HoxD* cluster (orange arrowhead). *Hoxd13lacZ* is extensively covered by H3K27me3 and deprived of H3K4me3. (B) RNAseq and ChIPseq signals specific to the histone modifications H3K4me3 and H3K27me3 in mutant homozygous *HoxD* $^{(inv11-12)d13lacZ}$ E13.5 metanephros. *Hoxd13lacZ* is activated and enriched in H3K4me3, showing a reduced level of H3K27me3 (black arrows). Interestingly, the weak transcriptional activity observed between *Hoxd13lacZ* and *Evx2* is not correlating with the activation of *Evx2* and does not affect H3K27me3 coverage (position of the green and orange arrowheads). The quantitative gain-of-expression of *Hoxd12* is confirmed and correlates with an important gain of H3K4me3. Superimposed strand-specific RNAseq are from Fig. 17.

At this point, we concluded from this initial investigation of the inv(*11-12*)*d13lacZ* allele the following points:

a. The particular regulation of *d13lacZ* is not maintained in the *HoxD* $^{\text{inv}(11\text{-}12)d13lacZ}$ configuration. The transcription of *d13lacZ* is activated more anteriorly along the AP axis as well as in different secondary structure (metanephros, caecum) where it is usually repressed and covered by H3K27me3.

b. The gain-of-expression of *Hoxd13lacZ* correlates with an important transcriptional activity observed on the anti-*HoxD* DNA strand, generated by the inverted *Hoxd11* gene. This transcript runs towards the posterior extremity of the *HoxD* cluster and encompasses notably *Hoxd13lacZ* antisense and *Evx2*.

c. The regulative disruption is local and restrained to *Hoxd13lacZ*. Even though the transcriptional activity traverses the locus of *Evx2*, the intrinsic regulation of this locus is conserved and its promoter is resistant to this transcriptional activity.

d. The relocation of *Hoxd11* in the position of *Hoxd12* results in its posteriorization in the trunk along the AP axis and in a significant quantitative reduction in the metanephros. Concomitantly, a strong gain-of-expression of *Hoxd12*, relocated in the position of *Hoxd11*, is visible by RNAseq in different structures (trunk and metanephros).

e. In the metanephros, the transcriptional output of the series of central *Hoxd* genes located anteriorly to the inverted locus, i.e. *Hoxd1* to *Hoxd10*, is reduced even though these genes show extremely high level of H3K4me3.

3.2.3 Targeting *d11* in the inv(*11-12*)*d13lacZ*

3.2.3.1 Design of the deletion

In the light of those intermediate conclusions, it seems that the transcriptional leakage emitted by the inverted *Hoxd11* correlates with two completely opposite outcomes. In the context of the inv(*11-12*)*d13lacZ* allele, we showed that the transcriptional activity covering the antisense of *Hoxd13lacZ* correlates with a clear gain-of-expression of this gene. Meanwhile, in the inv(*11*) mutant, an apparently similar transcriptional activity on the anti-*HoxD* DNA strand correlates with a robust reduction of *Hoxd12* transcript. In an effort to challenge these two conflicting observations, we seek one more time for the CRISPR/Cas9 technology. We took advantage of the mutagenic efficiency of this technique to delete the *Hoxd11* locus out of the inv(*11-12*)*d13lacZ* allele, hence generating the allele *HoxD* $^{\text{inv}(11\text{-}12)\text{del}(11)d13lacZ}$ (Fig. 19A, Fig. 8). We speculated that this genetic manipulation may reveal to what extent the transcription generated by the inverted *d11* is contributing to the gain-of-expression of *Hoxd13lacZ* in the inv(*11-12*)*d13lacZ*. To engineer this allele, we used the same two gRNAs described previously in this work, gRNA27 and gRNA28.

This experiment coincided with the development of the electroporation of gRNAs and *Cas9* mRNA in murine zygote, an approach which bypasses the tedious PNI technology and increases the throughput of the whole experiment (Hashimoto and Takemoto, 2015). This technical advance was crucial for the establishment of this allele, as the mutagenesis of the *d11* locus in the inv(*11-12*)*d13lacZ* context turned out to be particularly challenging. In a first instance, the *lacZ* reporter sequence inserted within *Hoxd13*

appeared to negatively affect the efficiency of any CRISPR/Cas9 targeting in the genetic vicinity of *d13lacZ*. We observed indeed a drastically reduced number of DSB events 3' of *Hoxd11*, next to the *lacZ* reporter (Fig. 19) compared to a similar targeting made over a wild type chromosome. The repression displayed by the *lacZ* resulted in the frequent appearance of F0 animals of the genotype inv(*11-12*)*d13lacZ*/del(*11*).

This local inhibition was even stronger when using another gRNA, located 3' of the *d11* locus but closer to the *lacZ* than g27 (data not shown). In a second instance, the loss-of-function of *Hoxd13*, resulting from the knock-in of the *lacZ* reporter in this gene, did not permit us to work with a stock of homozygous mutant male breeders to produce homozygous mutant zygotes to be processed. Indeed, homozygous inv(*11-12*)*d13lacZ* males suffer from a *Hoxd13* $^{-/-}$ phenotype and thus show structural default in their digits and in their external reproductive organs, a condition leading to a complete infertility (Dollé et al., 1993). Therefore, we had to continue working with heterozygous inv(*11-12*)*d13lacZ* breeders to generate zygotes to be exposed to CRISPR/Cas9. This prerequisite reduced the quantity of putative inv(*11-12*)*d13lacZ* allele that can be mutagenized by CRSIRP/Cas9 by a mendelian factor of two. Eventually, out of the 63 pups that were treated with this technique and tested, a single F0 animal of the desired inv(*11-12*)del(*11*)*d13lacZ* genotype was obtained. This founder was then extensively characterized and his genome reconstructed for future genomic experiments. The lane was then further backcrossed with wild type animals stock for two generations to dilute any off-target mutations caused by the CRISPR/Cas9 mutagenesis.

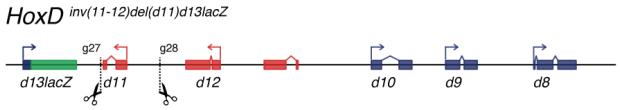

Figure 19: Schematic representation of the allele *HoxD* $^{inv(11\text{-}12)del(11)Hoxd13lacZ}$ produced by CRISPR/Cas9 mediated mutagenesis. The locus *Hoxd11* was deleted in *cis* out of the inv(*11-12*)*d13lacZ* allele using the pair of guides gRNA27/28 already used to generate the inv(*11*) allele.

3.2.3.2 Inv(*11-12*)del(*11*)*d13lacZ*

Once in possession of this allele, we assessed whether the gain-of-expression of *d13lacZ* was still observed or if the normal regulation of *d13lacZ* was recovered. This last possibility would infer that *d13lacZ* expression should disappear from the developing kidneys, the caecum as well as retrieve its regular pattern along the AP axis. Following these rationales, we sampled a series of E12.5 heterozygous mutant embryo of relevant genotypes to be exanimated by WISH. We used the same *lacZ* antisense probe as previously presented (Fig. 15) to reveal the expression pattern of the reporter *d13lacZ*. We found that the pattern of expression of *d13lacZ* along the AP axis was still subjected to a significant anteriorization in the double mutant inv(*11-12*)del(*11*)*d13lacZ* (Fig. 20). *LacZ* WISH signal appeared to be comparable with the signal observed in inv(*11-12*)*d13lacZ*, extending from a position located in the posterior root of the hindlimb up to the caudal extremity of the embryo (Fig. 20, black arrow). In addition, a signal in the budding caecum was also visible and therefore indicated that *d13lacZ* gain-of-expression was similarly conserved in this structure (Fig. 20, black arrowhead).

Figure 20: Whole mount In Situ Hybridization of the reporter gene *Hoxd13lacZ* in the alleles *HoxD* ^{d13lacZ}, *HoxD* $^{inv(11\text{-}12)d13lacZ}$ and *HoxD* $^{inv(11\text{-}12)del(11)d13lacZ}$ in E12.5 heterozygotes embryos. The deletion of the locus *Hoxd11* from the inv(*11-12*)*d13lacZ* allele does not prevent the anteriorization of the reporter *Hoxd13lacZ* which conserves the domain of expression observed in the inv(*11-12*)*d13lacZ* mutant, extending up to the base of the hindlimb (black arrow) along the AP axis. The gain of *d13lacZ* expression in the caecum reported by both Delpretti et al. and Kmita et al. is also conserved in the inv(*11-12*)del(*11*)*d13lacZ* embryo (black arrowhead). The white dashed lanes mark the posterior extremity of the hindlimb.

We next verified whether the gain-of-expression of *d13lacZ* was also visible in the metanephros. We processed a series of heterozygous mutant metanephros through X-gal staining and recorded an intense staining in the metanephric mesenchyme of the developing kidneys of the double mutant inv(*11-12*)del(*11*)*d13lacZ* (Fig. 21A). This staining was significantly stronger and appeared to follow a broader pattern compared to the one observed in the heterozygous inv(*11-12*)*d13lacZ* control stained at the same time (Fig. 21A). This observation led us to conclude that the leakage emitted by the locus *d11* in the inv(*11-12*)*d13lacZ* allele was not necessary to trigger the activation of *Hoxd13lacZ* in the metanephros, a result in accordance with the observation made previously in the caecum and along the trunk.

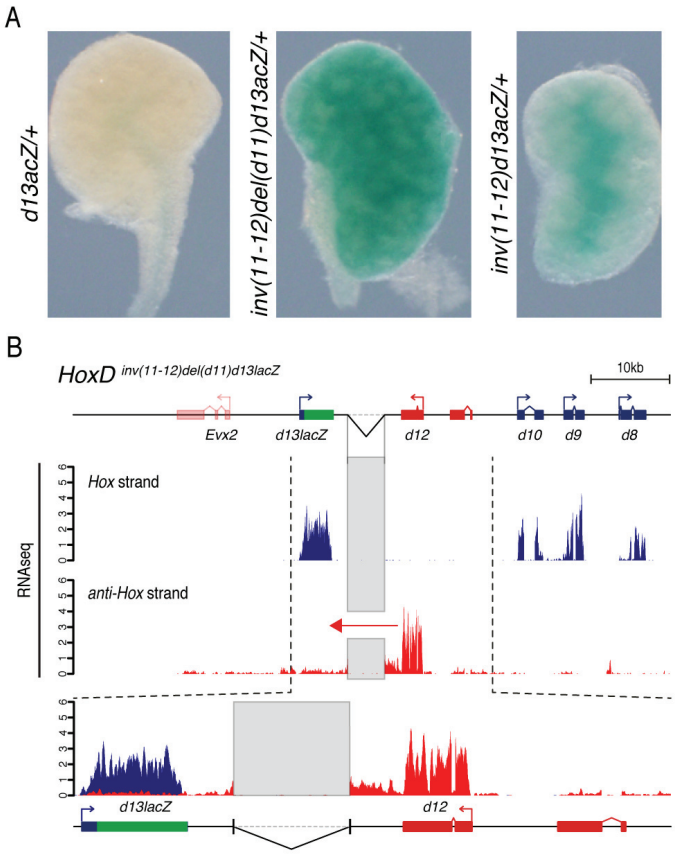

Figure 21: Gain-of-expression of the posterior *d13lacZ* gene in the *HoxD* $^{inv(11\text{-}12)del(11)d13lacZ}$ allele in E13.5 metanephros. (A) Series of X-gal staining of E13.5 metanephros micro-dissected from heterozygous mutant *HoxD* ^{d13lacZ}, *HoxD* $^{inv(11\text{-}12)del(11)d13lacZ}$ and *HoxD* $^{inv(11\text{-}12)d13lacZ}$ embryos. The inv(*11-12*)del(*11*)*d13lacZ* metanephros displays an important X-gal staining, quantitatively and qualitatively more important than the X-gal signal obtained in the control inv(*11-12*)*d13lacZ* sample. (B) Stranded mRNA RNAseq generated from E13.5 metanephros micro-dissected from a mutant *HoxD* $^{inv(11\text{-}12)del(11)d13lacZ/del(8\text{-}13)RXII}$. Transcripts generated by the *HoxD* coding DNA stand are in blue, transcripts mapped to the anti-*HoxD* DNA strand are depicted in red. The deletion of the *Hoxd11* locus does not avert the appearance of an anti-*HoxD* transcriptional activity released by the *Hoxd12* locus (red arrow) at the posterior extremity of the *HoxD* cluster and correlates with the activation of the reporter *d13lacZ*.

In a counterintuitive manner, we speculated whether the up-regulation of *Hoxd13lacZ* would be actually even more important in the inv(*11-12*)del(*11*)*d13lacZ* double mutant when *Hoxd13lacZ* is free of the transcriptional leakage emitted by the inverted *Hoxd11* locus. The transcriptional landscape of the double mutant was thus assessed more precisely and we produced a strand specific RNAseq (Fig. 21B). This dataset confirmed in a first instance the strong transcription of the *d13lacZ* and revealed in a latter instance that the deletion of the *d11* locus leads to the appearance of a new synthetic transcript on the anti-*HoxD* DNA strand (Fig. 21B, red arrow). This synthetic transcript seems generated from the inverted *d12* locus and displays a similar genetic identity as the transcript emitted from *d11* in the inv(*11-12*)*d13lacZ* allele. It covers the posterior extremity of the *HoxD* cluster, including the antisense of the *d13lacZ* reporter and progress up to the 3' extremity of *Evx2.* The exact quantification of this transcript from this dataset is not robustly determined but it seems that the transcript is less abundant that the related one detected in the inv(*11-12*)*d13lacZ* allele generated by the inverted *Hoxd11* locus.

Overall, the experiments done using the inv(*11-12*)del(*11*)*d13lacZ* allele led us to the following conclusions:

a. The *Hoxd13lacZ* gain-of-expression is maintained even in absence of the *d11* locus in different tissues (AP axis, caecum, metanephros) and appears to be even quantitatively increased.

b. The deletion of the *d11* locus did not abolish totally the transcriptional activity running on the posterior extremity of the *HoxD* cluster on the anti-*HoxD* strand over, the *Hoxd12* locus appears to be capable in this configuration to spawn transcriptional leaking over its 3'UTR.

c. The transcriptional activity thus generated on the anti-*HoxD* DNA strand shares some genetic similarities with the one displayed by *d11* in the inv(*11-12*)*d13lacZ*.

3.2.4 Topological perturbations

We finally investigated to what extent the gain-of-expression of *Hoxd13lacZ* could be functionally caused by the *cis* reallocation of non-coding regulators located within the inverted *d11-d12* locus. We speculated that the genetic inversion, i.e. the physically reorientation of DNA motifs, and the reposition of such elements might participate in the loss of repression of *d13lacZ*. This hypothesis was consequential to the remarkable interpretation proposed by the authors of the original publication describing this allele (Kmita et al., 2000a) and also motivated by a set of new rationales. In a first instance, a series of CCCTC-binding factor (CTCF) binding sites are distributed over the *HoxD* cluster and notably in between *Hoxd12* and *Hoxd13* (Rodríguez-Carballo et al., 2017; Soshnikova et al., 2010). Several recent observations highlighted the putative role of CTCF and cohesin in constraining particular chromatin loops, in the range of 100 kb of distance following the so-called loop-extrusion model (Alipour and Marko, 2012; Goloborodko et al., 2016; Nasmyth, 2001; Rao et al., 2014; Sanborn et al., 2015).

The exact contributions of these particular contacts in the context of global chromosome organization, at the scale of TADs, remains yet to be fully validated (Schwarzer et al., 2017). However, a key feature of this model is that long-range DNA loops are mediated between convergent orientated CTCF sites. This characteristic implies that the polarity of CTCF sites determine the specificity of an important fraction of the topological contacts between key loci (de Wit et al., 2015). The genetic inversions of a series of CTCF sites further illustrates the importance of this polarity to establish precise long-range chromatin contacts between distant enhancers and gene promoters (Guo et al., 2015). Besides, the deletion of a single CTCF site in the paralog *HoxA* cluster results in a similar outcome, a gain-of-expression of *Hoxa7* observed in a cellular differentiation assay and correlating with the local disruption of the topological organization (Narendra et al., 2015, 2016). As presented in the introduction section, the distribution and orientation of CTCF sites positioned inside the *HoxD* cluster display a clear inversion in their directionality between the sites located downstream and upstream of *Hoxd11* (Rodríguez-Carballo et al., 2017). This break in CTCF sites orientation correlates extremely well with the topological boundary observed between *d11* and *d13* and are moreover extensively incorporated within the global bimodal regulation model (Andrey et al., 2013; Dixon et al., 2012; Rodríguez-Carballo et al., 2017).

In view of these rationales, the single CTCF site comprised within the inverted *d11-d12* locus in the inv(*11-12*)*d13lacZ* mutant was of pertinent interest for this work. In a wild type situation this site has indeed been reported to be orientated facing the centromeric gene desert, scoring long-range contacts with the C-DOM (Rodríguez-Carballo et al., 2017). Importantly, this site is the more external site located at the extremity of the C-DOM, its orientation facing the centromeric gene desert just before the T-DOM is called and the inversion of CTCF polarity is observed (Fig. 6, green arrowheads). Hence=, this

CTCF site may be a key element to participate in the insulation of the posterior extremity of the cluster through putative loop-extrusion events. Therefore, we postulated that its inversion and relocalisation next in line with the series of CTCF sites positioned over the central fraction of *HoxD* and included within the T-DOM, sites orientated majorly facing the distal T-DOM, could result in a local loss of topological insulation at the posterior *HoxD*. Such a synthetic topological no-man's-land would contain *d13lacZ*, the *d11* and *d12* locus and might lead to an amalgamation of their respective regulations, ultimately activating *d13lacZ* irrespectively of any anti-*HoxD* transcription. In this model, we propose that *Evx2* regulation would not be affected, as the topological insulation of this gene would be still assured by the two conserved CTCF sites located between *Evx2* and *Hoxd13lacZ*.

Therefore, we looked at the distribution of the CTCF factor in inv(*11-12*)*d13lacZ* metanephros. We wanted first to see whether the CTCF site of particular interest was still functional and bound by this protein, as the engineered construction of this allele could have affected the integrity of the motif. If this site was used and bound by CTCF, we wanted next to appreciate to what extent the cohesin complex was found engaged there. This signature (CTCF+/Rad21+) would indicate, in the light of the loop-extrusion model (Goloborodko et al., 2016; Rao et al., 2014; Sanborn et al., 2015), the functional integration of this site in chromatin interactions directed towards the T-DOM. We thus questioned whether the integration of this site within an abnormal topological organization could potentially cause the gain-of-expression of *Hoxd13lacZ* observed in different tissues and structures.

Because metanephros are composed of a limited amount of cells (typically $1.5x10^5$ cells in a pair of E13.5 metanephros) we adapted and optimized a recently published technical variation of ChIPseq to address successfully these points starting from a limited amount of biological material (Rodríguez-Carballo et al., 2017; Schmidl et al., 2015). ChIPmentation combines the benefits of the stringency and the specificity of chromatin immuno-precipitation with the high efficiency of an engineered transposase to construct a sequencable DNA library. The Tn5-derived transposase used here integrate partial sequencing adapters directly on top of the immuno-precipitated chromatin fragments. By optimizing the concentration of the enzyme and the time of reaction, transposase activity can be controlled and adapted to different amount of immuno-precipitated chromatin. This step allows the experimenter to build a library ready to be sequenced by running a nested PCR on top of the purified transposed DNA. This approach turned out to be extremely efficient to obtain sequencable ChIP libraries from limited amount of chromatin, because it efficiently bypasses the numerous enzymatic steps used to construct a regular ChIPseq library (i.e. DNA repair followed by A-tailing, ligation of sequencing adapters and PCR amplification).

We took advantage of this technique to generate ChIPmentation profiles CTCF and for Rad21/Scc1, a canonical component of the cohesin complex used here as a proxy for cohesin binding, in metanephros of the mutant inv(*11-12*)*d13lacZ* and control *d13lacZ*. The control dataset showed the expected set of eight principal CTCF sites between *Hoxd8* and *Evx2* (Fig. 9A). In accordance with published data, we also observed the presence of an additional weaker site found over the promoter of *Hoxd9* (Rodríguez-Carballo et al., 2017) that will not be further considered here. The orientation of these sites (Fig. 22A, black arrowheads) was determined by the direction of the CTCF binding motif extracted from these ChIPmentation datasets (Ziebarth et al., 2013). The Rad21 profile generated in this control allele revealed an heterogenous presence of cohesin at CTCF sites, as Rad21 was not observed systematically at each *HoxD* CTCF site. This result was in accordance with a series of both published and unpublished observation made in the Duboule laboratory (Rodríguez-Carballo et al., 2017). Notably, the CTCF site located between *Hoxd12* and *Hox13lacZ* is bound by both CTCF and Rad21 (Fig. 22A, black arrow). This observation is an important indication of the role of this element in the topological organization of this

locus. The CTCF ChIPmentation profile generated from inv(*11-12*)*d13lacZ* metanephros revealed that this site, now translocated next to *Hoxd10* and its motif inverted, was still functional and bound by CTCF at level quantitatively similar to one observed in the control dataset (Fig. 22B, black arrows). Remarkably, this inverted CTCF site showed also an important amount of Rad21 engaged (Fig. 22B, black arrows), whereas the global distribution of CTCF and Rad21 appeared overall well conserved between the two alleles.

This result led us to conclude that this particular CTCF site remains functional, in view of the alleged loop-extrusion model, in the mutant inv(*11-12*)*d13lacZ*. It might be engaged in topological contacts with the metanephros transcriptionally active fraction of the *HoxD* cluster and the whole T-DOM. As developed above, such a rewiring of contacts, either in *cis* or at a longer-range, in *trans*, could result in local loss of topological insulation of *Hoxd13lacZ*. Therefore, we next examined whether the reposition of this CTCF site correlates with important changes in the position and distribution of topological contacts in this mutant allele. We took advantage of the *lacZ* reporter sequence inserted within *Hoxd13* to generate a 4Cseq profile from this viewpoint in mutant metanephros. This sequence was indeed conveniently located, its genetic position conserved in both the control and mutant allele relative to loci such as *Evx* and *Hoxd8*.

The 4Cseq profile generated from the *lacZ* viewpoint in the *d13lacZ* control allele revealed a tight and dense local hub of interactions between *d13lacZ* and *Evx2* (Fig. 22A, *lacZ* viewpoint red headpin). These contacts scored from a transcriptionally inactive locus are following the distribution of H3K27me3, a feature also visible at long-range distances between fraction of the *HoxD* cluster and other loci sharing a similar epigenetic identity (Vieux-Rochas et al., 2015). This hub of contact is also included within the territory delimited by the four similarly orientated CTCF sites found between *Evx2* and *d12*, systematically facing the C-DOM. A pronounced exclusion of the central fraction of the cluster, i.e. *d10*, *d9* and *d8*, is clearly apparent, but not absolute. It reflects the cellular heterogeneity displayed by this model organ where at least two different regulatory circuits are controlling *Hoxd* expression (Di-Poï et al., 2007). Upon the inversion of the *d11-d12* locus and the relocation of the CTCF site, the 4Cseq contacts observed from the *lacZ* viewpoint shifted strongly and presents a very different picture (Fig. 22B, *lacZ* viewpoint red headpin). The dissociation of the local hub of contacts composed previously by *d13lacZ* and *Evx2* is evident, in accordance with the important reduction of H3K27me3 observed by ChIPseq over *d13lacZ*. Instead, a large proportion of these contacts are redirected towards the inverted *d11-d12* locus which display a similar regulative status. A certain fraction of the contacts is also gained over the transcriptionally active central portion of *HoxD* whereas the inverted CTCF site and its surrounding environment is strongly deprived in contacts in this reconfiguration (Fig. 22B, black arrow).

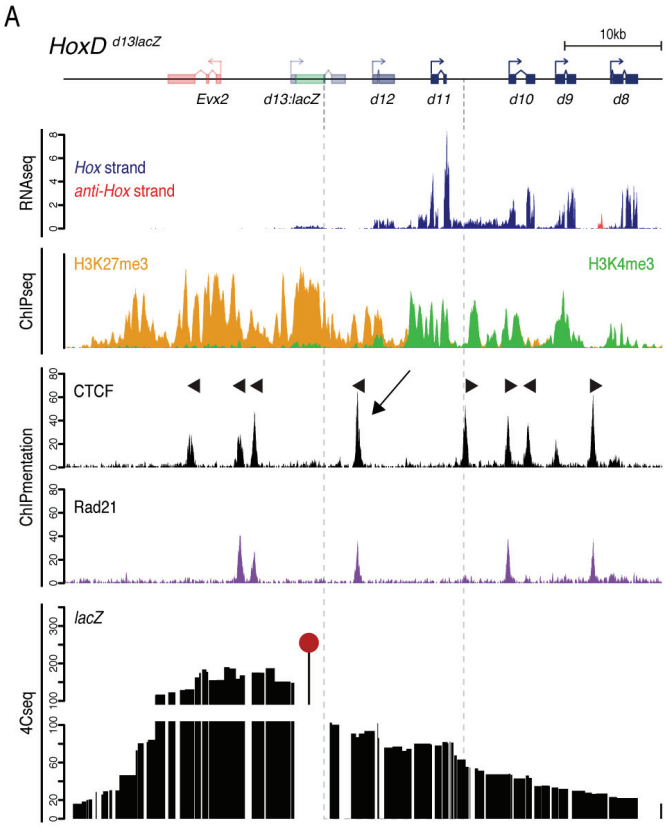

Figure 22: The inv(*11-12*)*d13lacZ* changes the position and the orientation of a CTCF site. (A) ChIPmentation of CTCF (black) and Rad21 (purple), a sub-unit of the cohesin complex, in control homozygous *HoxD* ^{d13lacZ} E13.5 metanephros. The orientation of each CTCF site is indicated by a black arrowhead, either facing the T-DOM (e.g. the d11_d10 CTCF) or the C-DOM (e.g. the d13lacZ_d12 CTCF, black arrow). The chromatin contacts observed from a *lacZ* 4Cseq probe (red pin) show a distinct separation of the posterior extremity of the *HoxD* cluster, containing *Hoxd13lacZ*, from the central faction of *HoxD*. These contacts correlate with the transcriptional status of the locus. Superimposed strand-specific RNAseq and H3K4m3/H3K27me3 ChIPseq are from Fig. 17 and 18.

A broader approach was also attempted to assess to what degree these local changes were integrated within the C-DOM and the T-DOM, using a series of 4Cseq viewpoints distributed over the *HoxD* cluster. However, these efforts failed to reveal meaningful variations in the position and global distribution of contacts across the locus studied. This result prompted us to start working with a more quantitative and robust chromatin conformational method, UMI-4Cseq (Olivares-Chauvet et al., 2016; Schwartzman et al., 2016). However, at the time of writing, this alternative approach has not produced results yet. Altogether, we interpreted these results as an indication of the functional integration of this site, when inverted and relocated, with DNA interactions located within the T-DOM. We postulate that this reorganization shifts locally the position of the boundary separating the C-DOM from the T-DOM at the *HoxD* cluster. In turn, this shift results in the local loss of topological insulation of *Hoxd13lacZ*, leading to the activation of this genes in structures controlled by regulating elements located within the T-DOM, such as the metanephros or the caecum (Delpretti et al., 2013; Kmita et al., 2000a)

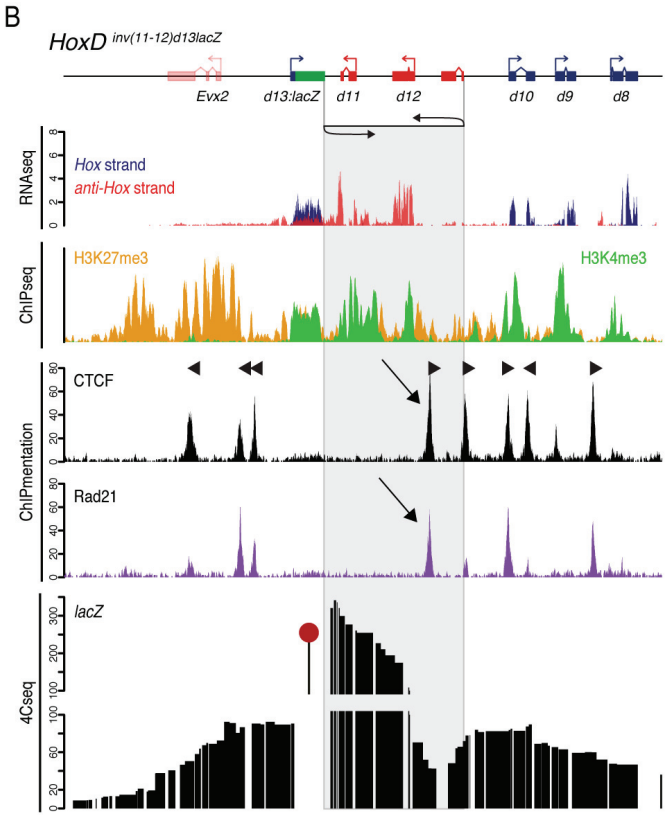

Figure 22: The inv(*11-12*)*d13lacZ* changes the position and the orientation of a CTCF site. (B) ChIPmentation of CTCF (black) and Rad21 (purple) in mutant homozygous *HoxD* $^{inv(11\text{-}12)d13lacZ}$ E13.5 metanephros. The orientation of each CTCF site is indicated by a black arrowhead, either facing the T-DOM or the C-DOM. The inversion of the *d11-d12* locus relocate and invert the d13lacZ_d12 CTCF (black arrow). This site is still functional and actively engaged by Rad21, suggesting its integration within the T-DOM topological organization. 4Cseq contacts generated from the *d13lacZ* (red pin) reporter indicates the appearance of a local hub of contacts between the actively transcribed *d13lacZ*, *d11* and *d12* genes. The contacts between *Hoxd13lacZ* and *Evx2* are greatly reduced.

3.2.5 Conclusions

In view of the data presented above, we conclude that the inversion of the *d11-d12* locus in the inv(*11-12*)*d13lacZ* allele results in the following defects in the regulations of *Hoxd* genes:

a. The relocation of *Hoxd11* in the position of *Hoxd12* leads to its posteriorization along the AP axis and to a significant quantitative reduction in the different structures investigated (AP, metanephros). Concomitantly, a strong gain-of-expression of *Hoxd12*, relocated in the position of *Hoxd11*, is quantitatively visible by RNAseq along the AP axis as well as in secondary structures (metanephros and caecum).

b. The regulation of *d13lacZ* is not maintained in this genetic configuration; the transcription of *d13lacZ* is found located more anteriorly along the AP axis and is activated in secondary structure (metanephros, caecum) where it is usually repressed and covered by H3K27me3 marks (bimodal regulation). The activation of the posterior fraction of the *HoxD* cluster results, in the metanephros, in a global quantitative reduction of the central/anterior *Hoxd*.

c. These changes in regulations correlate with the presence of a significant transcriptional activity on the anti-*HoxD* DNA strand, stretching over the 5' extremity of the cluster up to *Evx2*. This activity seems to be generated by the inverted *d11* locus.

d. This anti-*HoxD* transcriptional activity is not able to cause a transcriptional activation of *Evx2*; the whole posterior extremity of the *HoxD* locus, centered over *Evx2*, remains extensively covered by H3K27me3. Following the deletion of the *d11* locus in this allele, this anti-*HoxD* transcriptional activity appears to be reduced but is nevertheless still present, probably generated by the inverted *d12* locus. On the other hand, the global up-regulation of *Hoxd13lacZ* is conserved.

e. The *d11-d12* locus contains one CTCF binding site, engaged by both CTCF and Rad21. This site is located at a very particular position in the *HoxD* locus, being located next to the TAD boundary which topologically segregating *HoxD* into two regulatory domains. In non-inverted condition, this site is the last of a series showing a shared centromeric orientation, included within the C-DOM. Upon inversion of the *d11-d12* locus, this CTCF site is relocated next to *Hoxd10*, its motif returned, now facing the T-DOM and aligned with the series of CTCF distributed over the central fraction of *HoxD* and engaged in contacts with the T-DOM. The reoriented/repositioned CTCF site remains actively bound by CTCF and Rad21.

f. Local topological contacts recorded from the *d13lacZ* reporter are shifting importantly and are reassigned over the inverted *d11-d12* locus and, to a lower extent, towards the central portion of the *HoxD* cluster.

3.3 Inversion of the locus *Hoxd12*

3.3.1 Design of the inversion

The results obtained with the two intra-*HoxD* inversions presented in the previous sections of this work are contrasted and their interpretations are somewhat conflicting. In a first instance, the inversion of the *Hoxd11* locus decreases the transcriptional output of *Hoxd12* whereas, in the second allele studied, the larger inversion of the *Hoxd11*/*Hoxd12* locus correlates with the gain-of-expression of *Hoxd13lacZ*. In an attempt to clarify the consequences of the loss of transcriptional polarity within the *HoxD* cluster, we constructed an allele bearing the inversion of the *Hoxd12* locus. The choice of this locus was motivated by several rationales, each resolving around the central question of discriminating between the effects exclusively caused by the reallocation of the transcriptional activity on the anti-*HoxD* DNA strand and the rearrangement of non-coding regulatory elements in *cis* which may perturb *Hoxd* genes regulations.

The allele *HoxD* $^{inv(12)}$ was generated in a similar fashion as the allele inv(*11*). Here again, we used the CRISPR/Cas9 technology adapted to work in murine zygote (Cong et al., 2013; Ran et al., 2013). Briefly, two gRNAs (gRNA3 and gRNA31) were designed to flank *Hoxd12* locus (Fig. 23) and cloned in the pX330 backbone (Mashiko et al., 2013). The gRNA positioned 3' to *Hoxd12* was designed on purpose 800 bp downstream of the annotated *Hoxd12* 3'UTR, in an effort to preserve the regular transcriptional stop site(s) of *Hoxd12*. The region X, a block of 230 bp conserved throughout vertebrate genomes and containing an alternative transcription start site for *Hoxd11* (Beckers et al., 1996; Hérault et al., 1998), was excluded from the inversion for reasons already discussed with the inv(*11*) allele. The position of the gRNA3, upstream of *Hoxd12*, was more challenging to determine. We sought the inv(*12*) to include a significant portion of *d12* promoter in order to conserve the endogenous regulation of this gene. But few options were available for the design of a functional gRNA with a sequence specific enough to target efficiently this locus. Moreover, we had to take into consideration an important constrain: the presence of d13_d12 CTCF site located upstream of *Hoxd12*. This site is inverted and reallocated in the inv(*11-12*)*d13lacZ* allele and is likely participating in the disruption of the regulation of the posterior *HoxD* locus by affecting the topological organization of the locus. Therefore, we positioned the gRNA31 telomeric to this CTCF site to avoid affecting its integrity and specific orientation in the inv(*12*) allele (Fig. 23). The size of the inversion of the *Hoxd12* locus was just below 4.7 kb, a size comparable with the inv(*11*) allele.

Because of a limited choice of possible targetable CRISPR/Cas9 sites, we had to design gRNA31 directly within the conserved region XI (211 bp, Hérault et al., 1998). This locus contains at least one *cis* enhancer which has been demonstrated to control *Hoxd12* transcription, maintaining notably the expression of this gene in the trunk along the AP axis (Hérault et al., 1998). However, this enhancer is not required to control *Hoxd* expression in the autopod, as the deletion of the locus RXI did not affect neither *Hoxd12* nor the specific expression of the neighboring genes *Hoxd11* and *Hoxd13*. Similarly, RXI does not appear to contain an enhancer active in the metanephric kidneys since the expression of both *Hoxd11* and *Hoxd12* is completely lost in this structure when the *HoxD* cluster is genetically dissociated into two independent entities separated by a large inversion of 3 mb, i.e. *Hoxd13*, *d12* and *d11* on the one hand, *Hoxd10* to *Hoxd1* on the other hand (Di-Poï et al., 2007; Spitz et al., 2005). We therefore assumed that if the inversion of the *Hoxd12* locus interrupts the function of RXI, potentially affecting *Hoxd12* expression along the trunk, the metanephros and the autopod of the inv(*12*) allele should remain valid read-outs available to record the consequences of this inversion.

The two pX330 vectors encoding the gRNAs 3 and 31 with the Cas9 mRNA were injected at an equimolar concentration within the pronucleus of mouse zygotes (PNI). The generation of this allele was effectuated in collaboration with the TCF platform at EPFL. Following transfer to foster mothers, animals were screened at weaning for the complete inversion of the *Hoxd12* locus. The genotyping was done by using a combination of PCRs from series of convergent/divergent primers. A single F0 founder was isolated out of about 60 animals obtained and genotyped for the inversion of the locus *Hoxd12*. The *HoxD* $^{inv(12)}$ allele was then back-crossed with wild type animals for two generations to dilute any off-target mutations caused by the CRISPR/Cas9 mutagenesis and the inverted locus was sequenced to confirm the integrity of the breakpoints and to reconstruct the genome of this new allele.

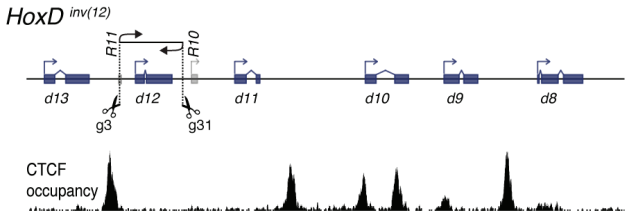

Figure 23: Schematic representation of the inv(*12*) allele produced by CRISPR/Cas9 mediated mutagenesis. Positions of the guide gRNA3/31 used to target the *Hoxd12* locus are depicted. gRNA3 is situated within the conserved region RXI (Hérault et al., 1998) but telomeric to the d13_d12 CTCF to avoid disrupting the presumed regulative function of this site (see text). gRNA31 is positioned downstream of *Hoxd12* 3' UTR and centromeric to RX (Beckers et al., 1996; Hérault et al., 1998). CTCF occupancy track is extracted from a CTCF ChIPmentation dataset produced in E13.5 wild type metanephros.

3.3.2 Inv(*12*) in the metanephros

Once in possession of *HoxD* $^{inv(12)}$, we addressed whether the relocation of *Hoxd12* over the anti-*HoxD* strand could affect the normal regulation and expression of the posterior *Hoxd13* gene. We postulated that such a disruption might be mediated by transcriptional leakage emitted by the inverted *Hoxd12* locus and potentially leading to either a loss-of-expression or a gain-of-expression of *Hoxd13*. Alternatively, we anticipated that *Hoxd12* transcriptional stop site(s) might be able to contain this leakage because these sites are still linked to the inverted locus in the inv(*12*) allele, a configuration which was not present in both the inv(*11*) and inv(*11-12*)*d13lacZ* alleles presented above.

In order to obtain a precise picture of the transcriptional situation at play in this allele, we generate a strand specific RNAseq from metanephros dissected in E13.5 mutant embryos. We justify the choice of this structure by two principal rationales. First, the posterior *Hoxd13* is not expressed in E13.5 metanephros in wild type condition, therefore making the metanephros an adapted model to observe any possible gain-of-expression of this posterior gene following the inversion of the *Hoxd12* locus. This approach is similar to what has been successfully used with the inv(*11-12*)*d13lacZ* allele. Second, *Hoxd12* is only expressed in a limited number of cells composing the developing kidney whereas its anterior neighbor *Hoxd11* marks a large portion of this organ, i.e. the metanephric mesenchyme. Therefore, an appropriate question to address was to verify whether *Hoxd12* was upregulated following its reallocation closer to *Hoxd11* promoter in the inv(*12*) allele. To challenge these points, we employed the same experimental strategy as discussed previously in the case of the inv(*11*); we worked with the del(*8-13*)(RXII) allele as a balancer in order to study the inv(*12*) in a trans-heterozygote genetic context.

The transcriptome obtained from mutant metanephros revealed that the reallocation of the transcriptional activity over the anti-*HoxD* DNA strand is restricted at the *Hoxd12* locus (Fig. 24A, red arrow).

In this embryonic structure and genetic configuration, *Hoxd12* seems to release a limited transcriptional activity leaking towards *Hoxd13* but without covering it. The posterior extremity of the cluster remains indeed free of any significative anti-*HoxD* transcription, *Hoxd13* and *Evx2* remain silenced (Fig. 24A, black arrow). This observation is similar to the situation displayed by the control dataset generated from heterozygous del(*8-13*)RXII metanephros (Fig. 24B). In contrast to the two inversions previously studied in this embryonic organ, *HoxD* transcriptional landscape appears therefore globally unaffected by the inversion of *Hoxd12*.

Finally, the sole variation apparent here is the increase of normalized signal computed over *Hoxd12* (Fig. 24A and C). In our opinion the exact interpretation of this observation is debatable. At the time of writing, we are not in possession of biological duplicate datasets to confirm this tendency. Altogether, if the inversion of *Hoxd12* results in a significative up-regulation of this gene, the magnitude of this augmentation is in any case not reaching quantitatively the level of *Hoxd11* therefore arguing against a complete functional merging of these two promoters. On the other hand, the quantitative increase of *Hoxd12* transcripts is supported by the observation of the transcriptional leakage emitted from *Hoxd12* towards *Hoxd13*. This transcriptional behavior is indeed not visible in the control dataset where *Hoxd12* is not leaking over *Hoxd11* (Fig. 24B, black arrow), whereas on the contrary a similar observation is recorded in E12.5 autopod RNAseq, in a structure where the transcriptional activity of *Hoxd12* is higher (Fig. 12B). In this view, the visualization by WISH of the distribution of the expression domains of the three posterior *Hoxd* genes along the AP axis and in other secondary structures (i.e. caecum and zeugopod) of the inv(*12*) embryo should help us to clarify our understanding of this allele in the near future.

3.3.4 Conclusions

In the metanephros, the inversion of the locus *Hoxd12* does not result in a gain-of-expression of the posterior *Hoxd13* gene. In this embryonic organ, the inverted *d12* locus emits a limited transcriptional leakage on the anti-*HoxD* DNA strand but does not cover *Hoxd13* antisense. This observation suggests us that no significative loss-of-function of *Hoxd13* should be observed in tissues where both *Hoxd12* and *Hoxd13* genes are expressed concomitantly (i.e. in the limb autopod and along the trunk in the *Hoxd13* expression domain). We are not yet in a position to address this point as a stock of homozygous inv(*12*) animals is currently breed. However, preliminary observation of the digits of inv(*12*) heterozygote animals did not reveal any apparent *Hoxd13* $^{+/-}$ phenotype in these structures.

More investigations are obviously required, notably to assess to what extent the reallocation and the partition of RXI is affecting the initial *Hoxd12* expression along the AP axis. But overall these observations support the impression that then engineered inversion of a single *Hox* gene in a tight *organized* cluster can be tolerated and not affect drastically *Hox* genetic system. In view of the two inversions considered in the previous sections of this work (i.e. inv(*11*) and inv(*11-12*)*d13lacZ*), this observation can be considered as a promising paradigm to illustrate the constrains to overcome to break *Hox* organization once the clusters are structurally consolidated.

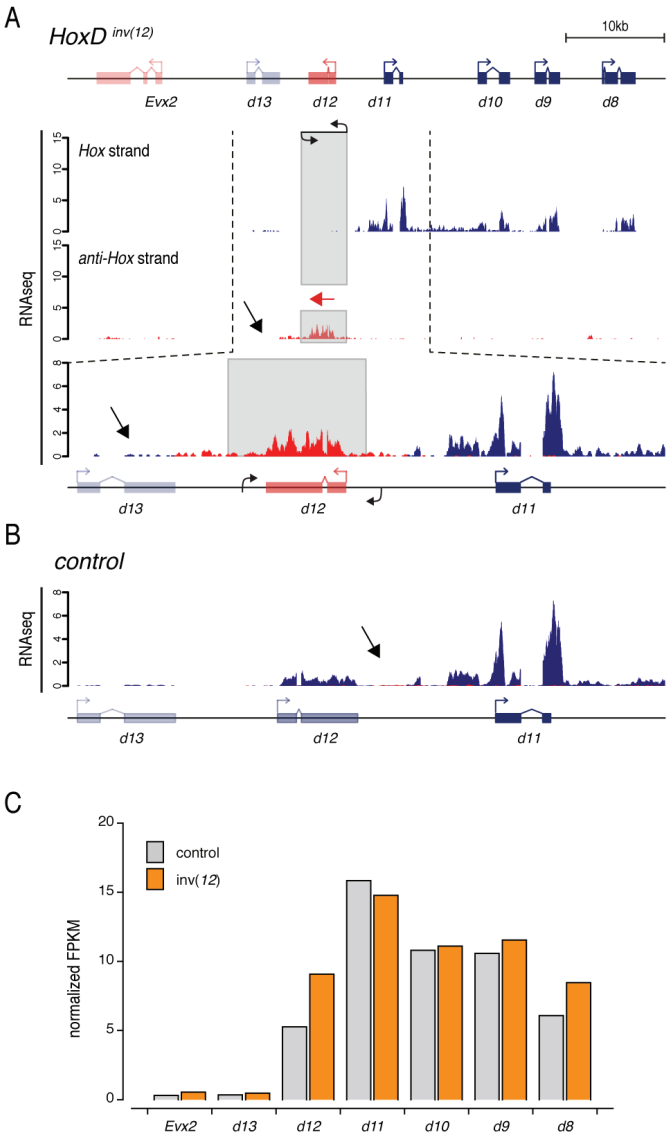

Figure 24: Transcriptional profile of the posterior *HoxD* cluster in the inv(*12*) allele in E13.5 metanephros. (A) Normalized strand specific mRNA RNAseq signal generated from the mutant *HoxD* $^{\text{del}(8\text{-}13)\text{RXII/inv}(12)}$. The inversion of the locus *Hoxd12* results in a limited transcriptional leakage covering the anti-*HoxD* DNA strand, progressing towards *Hoxd13* (red arrow). However, this transcriptional activity does not reach *Hoxd13*, which remains transcriptionally silenced (black arrow). The quantity of *Hoxd12* transcripts appears slightly increased, compared to the control transcriptome. Transcripts mapped to the anti-*HoxD* DNA strand are depicted in red, transcripts specific to the *HoxD* DNA strand are represented in blue. (B) Control strand specific mRNA RNAseq obtained from *HoxD* $^{\text{del}(8\text{-}13)\text{RXII/+}}$ (see Fig. 13B). In a wild type non-inverted configuration, *Hoxd12* does not generate any transcriptional leakage towards *Hoxd10* (black arrow). (C) Normalized quantification (FPKM) of the posterior *Hoxd* genes computed from the RNAseq dataset presented in B and C. Note the increase of *Hoxd12* in the inv(*11*) allele. The quantification of the central *Hoxd* genes are otherwise not significantly impacted, except for a limited gain of *Hoxd8*.

3.4 Consequences of a functional gain-of-expression of *Hoxd13*

3.4.1 General observations

The homeotic consequences of the gain-of-expression of *d13lacZ* in the inv(*11-12*)*d13lacZ* allele are impossible to evaluate because of the knock-in of the *lacZ* reporter in the first exon of *Hoxd13*. This recombination therefore results in a general *Hoxd13* loss-of-function phenotype, discussed above in this work. We could not easily attempt to reconstruct a similar inv(*11-12*) allele in the context of a functional *Hoxd13* gene because of the important lethality of mutant displaying gain-of-expression of *Hoxd13*. It was indeed demonstrated that the anteriorization of *Hoxd13* expression along the AP axis, together with its gain-of-expression in the metanephros was not tolerated by mutant animals, resulting in the complete agenesis of the kidneys but also in the premature arrest of the AP axis extension (Kmita et al., 2000b; Young et al., 2009). These phenotypes would obviously complicate the maintenance of such an allele, requiring floxed mutant animal stocks and a very particular breeding strategy.

Therefore, in the final chapter of this thesis, we took advantage of an engineered murine allele that was available in the laboratory to illustrate the consequences of a gain-of-expression of *Hoxd13* in the developing kidneys. We used the inversion *HoxD* $^{(TgHd11lacNsi\text{-}Itga6)}$, a 3 mb inversion obtained by TAMERE (Tschopp and Duboule, 2011). This inversion was mediated by the Cre recombinase acting over two *loxP* sites, one located within the gene *Itga6* and a second site situated within a transgene containing a reporter composed of the *Hoxd11* gene fused to a *lacZ* sequence and integrated in between *Evx2* and *Hoxd13* (Gimond et al., 1998; van der Hoeven et al., 1996). Crucially, the *HoxD* cluster and the whole T-DOM remain intact in this allele (Fig. 25A). The main consequence of this genetic reconfiguration is the disconnection of the *HoxD* cluster from the centromeric regulatory landscape; the detachment of the posterior *Hoxd* genes from the C-DOM.

This reconfiguration leads to specific loss of *Hoxd13* expression in tissues where *d13* is under control of centromeric long-range enhancers, e.g. in the digits and in the genital tubercule (Lonfat et al., 2014; Tschopp and Duboule, 2011). Furthermore, this allele also presents the particularity to display gain-of-expression of *Hoxd13* in several tissues, such as e.g. in the trunk along the AP axis and in the developing proximal limb, where it provokes series of hypomorphic phenotypes, tolerated by heterozygous adult animal (Andrey et al., 2013; Tschopp and Duboule, 2011). This genetic behavior was functionally interpreted as resulting from the rewiring of the locus of *Hoxd13* over the T-DOM regulative landscape, triggered by the disconnection of *Hoxd13* from the C-DOM environment (Andrey et al., 2013). Since then, several genes have been described to behave similarly, whenever their topological organization are genetically challenged (see notably Lupiáñez et al., 2015).

3.4.2 Homeotic mutation in the kidney

To determine whether this allele was displaying a gain-of-expression of the posterior *Hoxd13* in embryonic kidneys, we first generated a strand specific transcriptome from homozygous inv(*Nsi-Itga6*) metanephros micro-dissected in E13.5 embryos. We also generated concomitantly a control RNAseq from wild type metanephros to be directly compared against the mutant sample (Fig. 25B). In accordance with our suppositions, the RNAseq profile obtained from the mutant samples revealed important transcriptional changes over the two genes located in the posterior fraction of the *HoxD* cluster, a marked up-regulation of *Hoxd12* and the activation *Hoxd13*. The normalized FPKM values computed from these datasets illustrate this quantitative redistribution of the transcriptional activity (Fig. 25C). In the mutant metanephros, transcription is reallocated principally towards *Hoxd12* and *Hoxd13*, at the expense of the *Hoxd* genes located anteriorly. This global transcriptional situation is reminiscent of the situation found

in the metanephros of the inv(*11-12*)*d13lacZ* where the up-regulation of *d12* and the activation of *d13lacZ* in E13.5 metanephros resulted in a similar reduction of the central/anterior *Hoxd* genes.

We then isolated and assayed series of heterozygous inv(*Nsi-Itga6*) and wild type adult animals to estimate their daily consumption of water. We assumed here that renal homeotic mutations, potentially resulting from changes in *Hoxd* patterning could affect water homeostasis in the mutant allele. This measure was effectuated by weighing the water bottles of the relevant mouse cages over several days, an extremely non-invasive manipulation. The quantification of water consumption revealed that heterozygous inv(*Nsi-Itga6*) mutant display a pronounced polydipsia (Fig 25D, blue arrow). We observed that mutant animals drink daily at least two times more water than control animals of similar sex, age and genetic background. This symptom was complemented by polyuria, a condition easily detectable in the cage of the considered mutant animals.

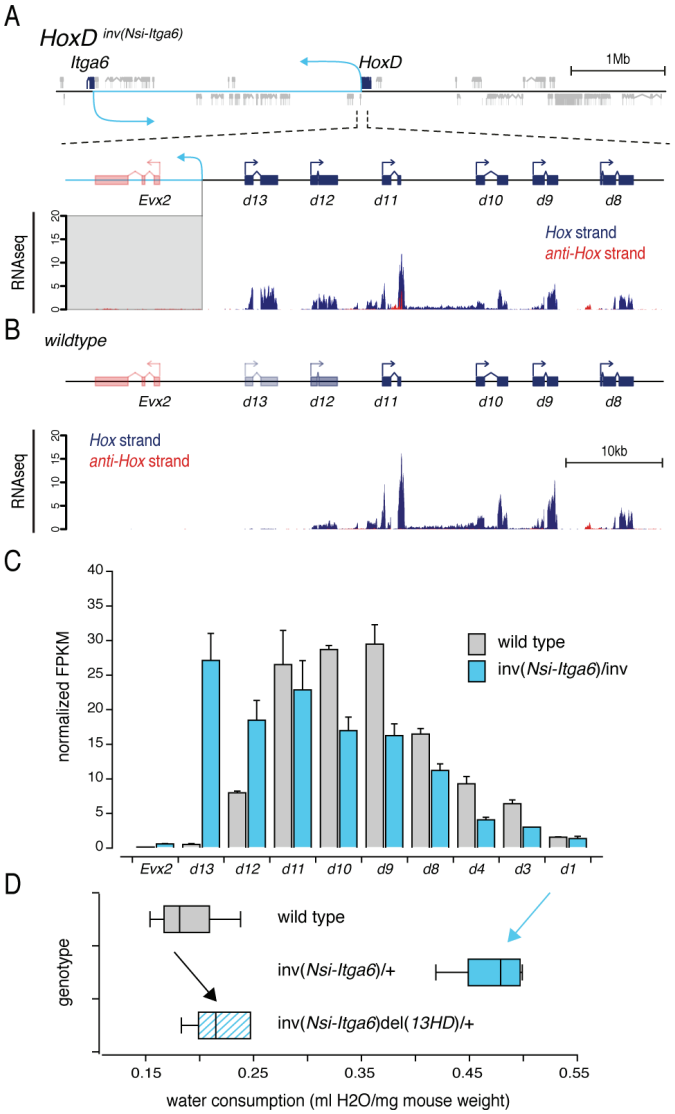

Figure 25: Inversion of *HoxD* centromeric regulatory landscape causes a gain-of-expression of *Hoxd13* in developing kidneys resulting in a pronounced polydipsia. *See captions next page.*

Figure 25: Inversion of *HoxD* centromeric regulatory landscape causes a gain-of-expression of *Hoxd13* in developing kidneys resulting in a pronounced polydipsia. (A) The inversion inv(*Nsi-Itag6*) separates *Hoxd13* from its endogenous regulatory landscape, effectively disrupting the integrity of the C-DOM (Andrey et al., 2013; Tschopp and Duboule, 2011). This inversion results in a pronounced gain-of-expression of the posterior *Hoxd13* gene in 13.5 metanephros as pictured by a strand specific mRNA RNAseq generated in homozygous *HoxD* $^{inv(Nsi\text{-}Itag6)/inv(nsi\text{-}Itag6)}$. Transcripts mapped to the *HoxD* coding DNA strand are displayed in blue, transcripts mapped to the anti-*HoxD* DNA strand are depicted in red. (B) Control strand specific mRNA RNAseq obtained from wild type E13.5 metanephros. (C) Normalized quantification (FPKM) of the posterior *Hoxd* genes computed from the RNAseq presented in A and B. The gain-of-expression of the posterior *Hoxd13* and *Hoxd12* genes in the inv(*Nsi-Itag6*) mutant correlates with the reduction of the transcription of the central *Hoxd* genes. This transcriptional situation is reminiscent of the situation observed in the inv(*11-12*)*d13lacZ* (see Fig. 17D). (D) Quantification of the daily water intake in 3 months old animals of different genotypes: wild type (grey), heterozygote inv(*Nsi-Itga6*) (blue) and double mutant heterozygote inv(*Nsi-Itag6*)del(*13HD*) (blue stripes). Inv(*Nsi-Itag6*) heterozygous animals show a twofold increased consumption of water (blue arrow). The deletion of the homeodomain of *Hoxd13* in *cis* with the inv(*Nsi-Itag6*), engendering a loss-of-function of *Hoxd13*, rescues this phenotype (black arrow) (see Fig. 26 and text for explanations).

3.4.3 Posterior prevalence

We then verified that this phenotype is caused by the sole gain-of-expression of the posterior *Hoxd13* gene in developing metanephros. This point was important to address, as several interpretations could be legitimately raised regarding the causes of this renal perturbation. Firstly, the inversion changes the whole quantitative balance of central and anterior *Hoxd* genes. Secondly, following the Cre-Lox inversion, the last exon of the gene *Itag6* is genetically disconnected from the main *Itga6* transcriptional unit. Although this gene has not been reported to play any role in kidney development or renal homeostasis, *Itga6* is nevertheless expressed in E13.5 metanephros. Moreover, a related gene, *Itga8*, has been demonstrated to be critical for the epithelial-mesenchymal transition during kidney morphogenesis (Müller et al., 1997). Lastly, this large genetic inversion can result in unexpected genetic perturbation outside of the *HoxD* locus in unknown loci, potentially topologically stressed by the inversion and their regulation thereby affected (Franke et al., 2016; Lupiáñez et al., 2015).

Thus, we employed one more time the CRISPR/Cas9 system to generate a small deletion within *Hoxd13*, in *cis* over the inv(*Nsi-Itga6*) allele (Fig. 26A). We designed and cloned two gRNA, gRNA10 and gRNA11, to target the second exon of *Hoxd13* and to delete 120 bp of the homeobox of this gene. The CRISPR/Cas9-mediated mutagenesis was then effectuated by PNI injection of the two corresponding pX330 vectors (Mashiko et al., 2013) in zygotes obtained from crosses between a stock of heterozygous mutant male breeders and a stock of wild type female. Following transfer in foster mothers, 49 injected animals were obtained and genotyped at weaning. Two F0 animals of the inv(*Nsi-Itga6*)del(*13HD*)/del(*13HD*) genotype were isolated and characterized, one male and one female. The regular breeding of the male turned out to be impossible because of the *Hoxd13* $^{-/-}$ phenotype. The lane was therefore established from the female F0 after two generations of back-crossing with a wild type stock of animals to dilute any off-target mutations caused by the CRISPR/Cas9 mutagenesis. We further confirmed by Sanger sequencing that the homeodomain of the HOXD13 protein targeted by the del(*13HD*) mutation was combining the desired 120 bp deletion and a frameshift mutation within the *Hoxd13* open reading frame, leading therefore to a loss-of-function of this gene in *cis* with the inv(*Nsi-Itga6*) background (data not shown).

We next quantified the daily water consumption of this new allele, expecting that the loss-of-function of HOXD13 in the context of its gain-of-expression would restore the proper development and function of the kidneys. This prediction was confirmed and the deletion of HOXD13 homeodomain in the inv(*Nsi-Itag6*) rescues entirely the polydipsia/polyuria symptoms observed with the inv(*Nsi-Itga6*) (Fig.

25D, black arrow). We finally generated a strand specific transcriptome of E13.5 metanephros dissected from homozygous inv(*Nsi-Itg6*)del(*13HD*) embryos to control that the del(*13HD*) mutation did not disturb transcriptionally the activation of *Hoxd13* in this allele. The RNAseq confirmed that the *d13HD* gene remain transcriptionally active and that the global *HoxD* transcriptional landscape in this double mutant allele is unchanged (Fig. 26A). Lastly, we performed a series of histological sections in adult kidneys dissected from adult heterozygous animals of relevant genotype. This final experiment was done to visualize two points: the expected histological defects of the kidneys of the inv(*Nsi-Itga6*) mutant and their disappearance from the kidney of the double mutant inv(*Nsi-Itga6*)del(*13HD*), following the rescue by the loss-of-function of *Hoxd13*. H&E-stained sections revealed a disorganized structure, with an extremely reduced medulla in the single mutant whereas sections obtained from the double mutant where anatomically identical to the wild type control (Fig. 26B).

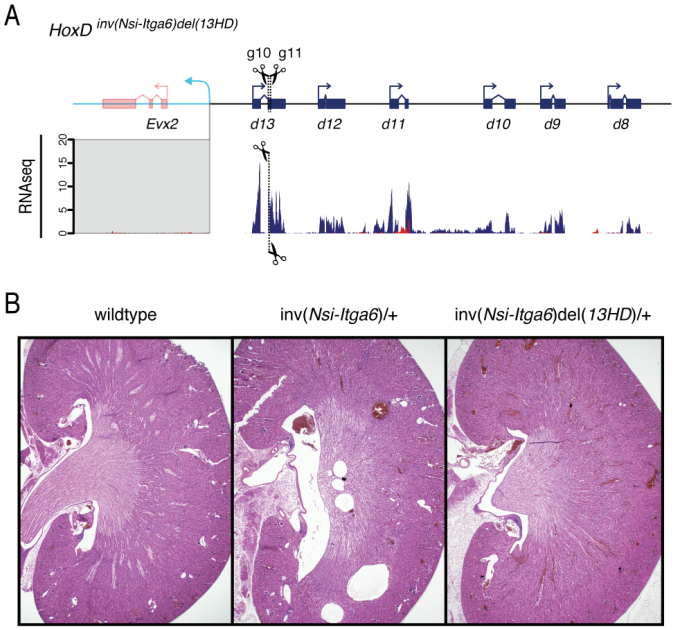

Figure 26: Loss-of-function of *Hoxd13* rescues the renal phenotype induced by the gain-of-expression of this gene in the inv(*Nsi-Itga6*) allele. (A) The loss-of-function of *Hoxd13* is produced by a CRISPR/Cas9 mediated deletion of 120 bp of the homeodomain of *Hoxd13*, positioned in the second exon of this gene. This deletion was obtained by the combined action of the guides gRNA10/11, in *cis* with the inv(*Nsi-Itga6*). Strand specific mRNA RNAseq obtained from E13.5 metanephros micro-dissected from homozygous *HoxD* $^{inv(Nsi\text{-}Itga6)del(13HD)/inv(Nsi\text{-}Itga6)del(13HD)}$ embryos shows that the truncated *Hoxd13* is still up-regulated in this allele, even though the renal phenotype is rescued (see Fig. 25D). (B) Representative H&E sections of kidney dissected from adult animal of different genotypes: wild type, heterozygote inv(*Nsi-Itga6*) and double mutant heterozygote inv(*Nsi-Itag6*)del(*13HD*). The defects visible in the mutant inv(*Nsi-Itag6*) are not present anymore in the double mutant inv(*Nsi-Itga6*)del(*13HD*).

3.4.4 Conclusions

We thus demonstrated with the hypomorphic inv(*Nsi-Itga6*) allele that the activation of the posterior *Hoxd13* gene in developing metanephros results in polydipsia in the adult animal. This symptom is likely caused by the sub-optimal organization of the kidney, leading to a less efficient retention of water. We proved that this phenotype is caused by the sole gain-of-expression of *Hoxd13*, in a necessary and suffi-

cient manner. These results illustrate the dominant negative genetic character of HOXD13, causing hypomorphic homeotic mutations when activated in structures normally patterned by HOXD11. Lastly, this result confirms the validity of the posterior prevalence displayed by *Hox* genes in vertebrate embryogenesis (Duboule, 1991; Duboule and Morata, 1994)

4 DISCUSSION

4.1 *Hox* clusters organizational constrains

This work shows that different inversions designed within the *HoxD* cluster result in a variety of perturbations in this genetic unit. Altogether, we demonstrated the difficulty to generate an inversion breaking the 5'-3' polarity within the murine *HoxD* cluster without affecting the complex regulation of these genes. As discussed above, the genetic proximity of *Hox* genes observed in gnathostome *organized Hox* clusters likely results from the consolidation of a primordial *disorganized* cluster during the radiation of the vertebrates. On the one hand, this process involves the development of networks of pleiotropic *trans* enhancers located in regulatory landscapes; the *organized* clusters providing by itself a genetic toolkit to co-opt *Hox* genes in a large variety of vertebrate-specific anatomical structures and organs (Darbellay and Duboule, 2016; Duboule, 2007; Soshnikova et al., 2013). On the other hand, this genetic densification coerces *organized Hox* clusters to conserve a determined transcriptional polarity and organization throughout the mechanisms that will be discussed here. Therefore, these constraints might be appreciated as the cost for the close proximity displayed by *Hox* genes when disposed in a tight cluster, illustrating the difficulties for a spontaneous mutation to invert the directionality of one *Hox* gene within an *organized* cluster. Altogether, the results discussed here provide elements to explain why large *disorganized* or *split Hox* structures, where the genetic distance between *Hox* genes is more important, can abide local inversion of *Hox* genes (e.g. the *Dfd* and *Ftz* genes in *D. melanogaster* ANT-C).

The mechanisms constraining this peculiar organization seem to be diverse and their effects can have different magnitudes, often detrimental for the global regulation of the *organized Hox* clusters. Interestingly, the results presented in this work illustrate those different possibilities, each triggered by an intra-cluster inversion. Two principal categories of perturbations can be defined based on their functional dependence or independence of transcription. These effects are not exclusive from one another and altogether they provide the framework where our results will be discussed. We shall first review them in details through the examination of relevant theoretical examples presented in the Fig. 27.

4.1.1 Transcription dependent perturbations

The disruption of *Hox* genes expressions and regulations might be directly provoked by the transcriptional activity relocated on the anti-*Hox* DNA strand. This activity would be generated by the inverted locus and leaks towards the 5' posterior extremity of the *Hox* cluster. This transcriptional behavior is by itself not artificial, as several *Hox* genes leak over their downstream *Hox* neighbor in a wild type configuration. This peculiar comportment results in polycistronic *Hox* mRNA transcripts (Mainguy et al., 2007; Montavon, 2008). This phenomenon can be further favorized if the inversion of an *Hox* locus genetically disconnects the principal poly(A) site (PAS from the 3' UTR of the considered *Hox* gene. For instance, this situation is observed with both the inv(*11*) and the inv(*11-12*)*d13lacZ* alleles. In this category, we will consider two series of different mechanisms, each resulting in antagonist effects over the considered *Hox* genes: promoter clearing and transcriptional interference.

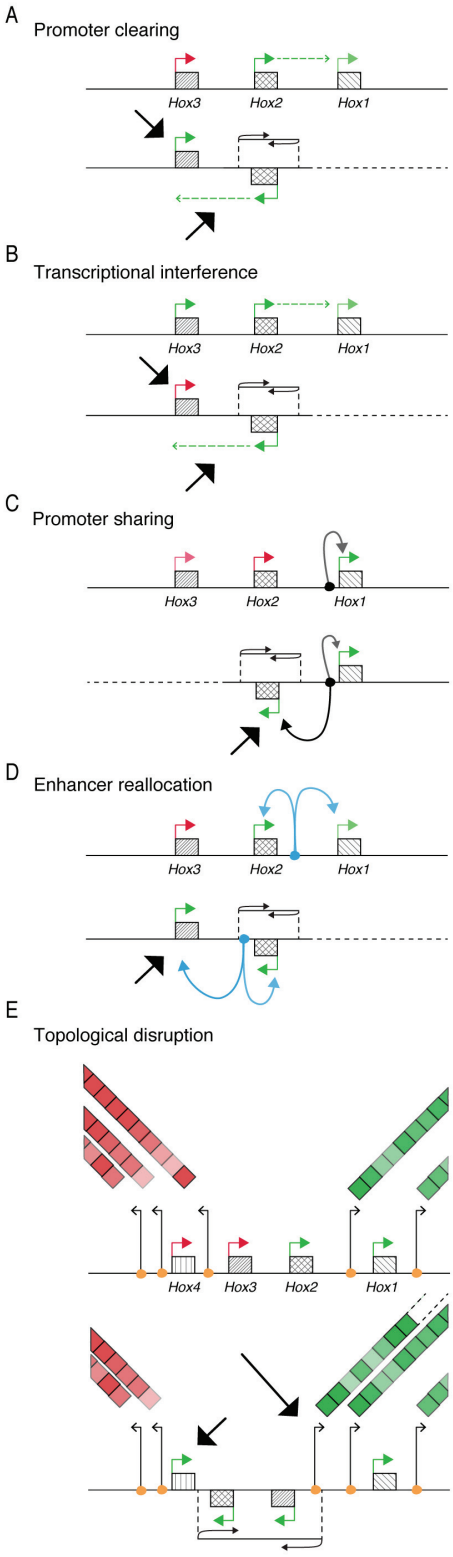

Figure 27: Schematic representation of the potential disruptions caused by an intra-Hox inversion. The first two classes of perturbations are based on hypothetic mechanisms directly related to the transcriptional directionality of *Hox* genes and their tendency to leak out of their 3'UTR (A and B). This situation is typically observed

with the gene *Hoxd11* in various situations *in vivo*. The *promoter clearing* (A) hypothesis proposes that the inversion of the gene *Hox2* results in the reallocation of the transcriptional activity, normally leaking out of *Hox2* over *Hox1*, on the anti-Hox DNA strand towards the posterior *Hox3* gene. This transcriptional activity might affect the regulation of the gene *Hox3* by disrupting its repression (black arrows), thus resulting in *Hox3* gain-of-expression in tissues and organs where this gene is not expressed. On the other hand, the *transcriptional interference* (B) model states that the same transcriptional leakage generated by the inverted *Hox2* locus will affect the transcriptional output of the posterior *Hox3*, reducing the dose of *Hox3* mRNA in a knock-down manner. The three other class of perturbations are based on mechanisms not involving transcription (C, D and E). In a first instance, *promoter sharing* (C) suggests that the inversion of a single *Hox* gene can results in its anteriorization and gain-of-expression in tissue where this *Hox* gene is normally silenced (black arrow, C). This interference in *Hox2* regulation could be caused by the reallocation and increased proximity between the two *Hox* promoters, *Hox1* and *Hox2*, leading to the cross-activation of their respective promoters (black arrow, C). In a second instance, *enhancer reallocation* (D) states that an inversion can translocate short-range *cis* enhancer within the *Hox* cluster, therefore affecting the regulation of the *Hox* genes. In this example, the inversion of the *Hox2* locus reallocates an enhancer normally active over *Hox1* and *Hox2* (blue, D). The repositioning of this enhancer might cause new regulations over neighboring *Hox* gene, causing for example *Hox3* gain-of-expression in tissues where this gene is usually not transcribed (black arrow, D). Finally, an intra-cluster inversion might also reallocate and invert a motif recognized by architectural elements potentially implied in the topological segmentation of an *Hox* cluster, such as a CTCF site (E). In this hypothetic situation, inspired by the inv(*11-12*)*d13lacZ* allele, the inversion of the locus containing *Hox2* and *Hox3* is also translocating a CTCF site (black arrow, E). In tissues where the posterior extremity of the *Hox* cluster is normally insulated from activating inputs by series of specific long-range contacts (red diagonals, E), the translocation and inversion of the CTCF site located at the frontier between the active and the inactive genes might participate in new topological contacts with distant active elements (green diagonals, E). Furthermore, the reallocation of this site could reduce *Hox4* insulation and expose this gene to different local activating influences, resulting in its activation (black arrow, E). In such a model, we propose that because several CTCF sites remain untouched upstream of *Hox4*, this gene shall maintain its own specific active long-range regulation, not represented here. Together, these mechanisms are not exclusive and their representations are voluntarily simplified.

The promoter clearing model proposes that anti-*Hox* transcriptional activity generated by the inverted *Hox* gene disrupts the PcG-dependent repression of the neighboring posterior *Hox*, resulting in its gain-of-expression in tissues where the inverted *Hox* is expressed (Fig. 27A, *promoter clearing*). At the time of writing, this hypothetic mechanism is not supported by abundant indications in the literature. But at least two observations should be taken into account and presented here. First, the different repressive modifications of the chromatin induced by PcG and covering the *Hox* clusters prior to their activations are labile. Such is notably the case of the well-studied H3K27me3 marks deposited by PRC2, one of the two canonical complexes of PcG. Indeed, a progressive turn-over of this mark is observed during the initial collinear activation of the *HoxD* cluster, replaced by the antagonist H3K4me3/H3K27ac marks (Soshnikova and Duboule, 2009). Second, the recruitment and maintenance of the *Hox* PcG domains are not extensively understood and possible functional relations with transcription remain ambiguous. The model developed in mESC some years ago to explain the recruitment of PRC2 at the *HoxD* locus is based on the presence of series of PcG response element (PRE) positioned within the cluster and an high density of unmethylated CpG dinucleotides (Schorderet et al., 2013). Recent studies have highlighted the capacity of different PRC2 components to bind nascent RNA transcripts, resulting in some case in the eviction of PRC2 out of the chromatin and modulating the methyltransferase enzymatic activity of PRC2 core component, EZH2 (Cifuentes-Rojas et al., 2014; Davidovich et al., 2013; Herzog et al., 2014; Kaneko et al., 2014).

In view of these evidences, we propose that anti-*Hox* transcripts emitted by the inverted *Hox2* towards the posterior *Hox3* might disrupt the PcG repression found at this locus by reducing PRC2 enzymatic activity or separating it from the chromatin. In turn, this reduction of PRC2 presence or activity might allow *Hox3* promoter to response to the same regulatory inputs as *Hox2*, thus resulting in the gain-of-expression of *Hox3* in cells expressing *Hox2* (Fig. 27A). In a second instance, the outcome of the

hypothetic mechanism considered here would result in an opposite effect for the expression and regulation of the *Hox* gene posterior to the inversion, although the initial cause would remain the same. Because the anti-*Hox* transcriptional activity generated by the inverted *Hox2* gene leaks over *Hox3* antisense, it might hence result in a physical interference with the proper transcription of *Hox3* (Fig. 27B, *transcriptional interference*). This mechanism would thus result in the reduction of *Hox3* transcripts whenever *Hox2* and *Hox3* genes are co-expressed. In this situation, we propose that *Hox2* by itself might not be affected as long as a transcription stop site exists between *Hox2* and *Hox3*, therefore assuring the production of a functional and stable *Hox2* mRNA.

Based on studies addressing convergent transcription in yeast, a genetic disposition frequently observed in the *S. cerevisiae* genome, the principal source of transcriptional interference is assumed to be caused by the frontal collision between two transcriptional machineries traveling concomitantly on opposed DNA strand (Hao et al., 2017; Palmer et al., 2011; Pannunzio and Lieber, 2016). These head-to-head collisions result in mutual arrest of the two complexes, leading to their dissociation from the DNA through polyubiquitylation of the RNA polymerase II (RNAPs) and the degradation of the aborted RNA transcript (Hobson et al., 2012; Osato et al., 2007; Prescott and Proudfoot, 2002; Somesh et al., 2005). It is however not known to what extend this degradation is concerning systematically the two stalled RNAPs or whether some factors, such as e.g. the relative firing rate of the two convergent promoters or characteristics of the chromatin can impact more specifically one RNAP (Boldogköi, 2012; Hao et al., 2017). Other theoretical mechanisms of similar outcome might also be considered, even if experimental evidences are less numerous. Because the anti-*Hox* transcriptional activity covers *Hox3* TSS and its proximal promoter, the initiation of *Hox3* transcription might therefore be disturbed, as the elongating RNAPs leaking from *Hox2* could remove transcription factors or promoter-bound RNA polymerase, or obstruct their binding at the *Hox3* locus (Hao et al., 2017).

Lastly, the convergent transcription observed between *Hox2* and *Hox3* may potentially activate RNA interference (RNAi) mechanisms through the production of small interfering RNA (siRNA, see a complete review in Castel and Martienssen, 2013). These short RNAs would result from the processing by endoribonuclease of the *Dicer* family of double stranded RNA. In accordance with our model, siRNAs targeting *Hox3* would potentially be generated from the sense and antisense transcripts of this gene in the nuclei of the cells expressing concomitantly the two *Hox* genes considered. Evidences obtained in human cell culture models as well as zebrafish transgenic experiments suggest that the generation of such siRNA through the artificial transcription of the same gene from two convergent promoters is able to efficiently knock-down the targeted gene, both in *cis* and in *trans* (Andrews et al., 2014; Calero-Nieto et al., 2010; Gullerova and Proudfoot, 2012; White et al., 2014).

This transcriptional gene silencing (TGS) is mediated by mechanisms that are not yet extensively understood. However, they seem to involve the modifications of histones and the deposition of stable marks such as H3K9me3. In turn, these marks lead to a reduction of the transcriptional output of the targeted locus (Andrews et al., 2014; Gullerova and Proudfoot, 2012; White et al., 2014), a process evocative of the RNA-induced transcriptional silencing (RITS) complex extensively studied in the yeast *Schizosaccharomyces pombe* (Castel and Martienssen, 2013; Moazed, 2009). The export of these functional siRNA in the cytoplasm can also theoretically activate post-transcriptional gene silencing (PTGS) through the cleavage and the destruction of the mRNA targeted by these siRNA, in our example the gene *Hox3*, by RNA-induced silencing complex (RISC). This mechanism would be analogous to a classical knock-down mediated by RNAi and short hairpin RNA. Overall, the outcomes of these hypothetic RNAi perturbations would be a reduction of *Hox3* transcripts whenever both *Hox2* and *Hox3* are transcribed simultaneously.

4.1.2 Transcription independent perturbations

An inversion within an *Hox* cluster might disrupt the regulations of *Hox* genes by reshuffling different types of non-coding elements, such as promoters and enhancers. These hypotheses are legitimate interpretations to consider as a significant portion of *Hox* genes regulation lies in *cis* within the clusters, where numerous enhancers and relays fix *Hox* expression across time and space. These regulative elements are likely inherited from the primordial regulation of *Hox* genes set in distended *disorganized* structures, then consolidated in tight *organized* clusters during the radiation of the gnathostomes.

In a first instance, the hypothesis of promoter reshuffling is supported by the fact that each *Hox* TSS is typically separated from the next one by less than 10 kb of distance. Therefore, the inversion of a single *Hox* gene increases the proximity between the inverted *Hox2* promoter and the promoter of the *Hox1* gene (Fig. 27C, *promoter sharing*). The reduction of the linear distance between these two *Hox* genes promoters may result in the amalgamation of their regulation because of possible cross-activations between their respective promoters and the proximity of the transcriptional machinery. Following this model, *Hox2* and *Hox1* would acquire a similar expression dynamic, resulting practically in the gain-of-expression of *Hox2* in tissues ordinarily patterned exclusively by *Hox1*. In a second instance, an intra-cluster inversion could result in the redisposition of a local enhancer positioned in *cis* within the cluster, a reallocation which might affect the regulation of the neighboring *Hox* genes (Fig. 27D, *enhancer reallocation*). The outcomes of such a reconfiguration are various, as the activity of the alleged enhancer can be modulated by the topologic structure of its local regulatory environment. For the sake of clarity, we will consider here a simple situation where the relocated enhancer, active initially on *Hox1* and *Hox2*, acquires the ability to control the gene *Hox3* following the inversion of the *Hox2* locus. Such a scenario would therefore lead to the gain-of-expression of *Hox3* in tissues normally patterned by *Hox1* and *Hox2*. Interestingly, this reallocation might also impact the expression of *Hox1*.

Finally, an intra-*Hox* inversion can *de facto* translocate binding sites engaged by factors implicated in the organization of the clusters. In the view of the loop-extrusion model (Sanborn et al., 2015; Schwarzer et al., 2017; de Wit et al., 2015), convergent CTCF sites are thought to block chromatin extrusion mediated by cohesin. Because the position and the orientation of such CTCF sites correlate with specific long-range contacts scored between adjacent regulatory landscapes and *Hox* genes (Rodríguez-Carballo et al., 2017), we propose that the reallocation and the inversion of a CTCF binding site might disrupt the topological organization of the locus (Fig. 27E, *topological disruption*). In the case where the inverted CTCF site is located at the boundary between two series of divergent CTCF sites, each series facing convergent CTCF sites located in their respective regulatory landscapes, the chromatin loop established at this locus would be inverted and redirected towards the opposed regulatory landscape.

The consequences of such a topologic rearrangement on the *Hox* genes could be important. The insulation of the posterior *Hox4* might be disrupted, thus potentially permitting cross-activation of this gene with the neighboring *Hox3*/*Hox2* locus or any local enhancer located in the vicinity. Moreover, the novel chromatin loops that could result from the translocation and the inversion of the CTCF site polarity might also trigger productive long-range interactions with distant enhancers, normally sequestered from *Hox4* and *Hox3*. In such a model, the endogenous regulation of *Hox4* could be conserved, as a global polarized structure is maintained on the side of the gene *Hox4*, therefore assuring that a fraction of the privileged contacts remains between *Hox4* promoter and its original regulatory landscape. Such a topological disruption of *Hox* genes regulation have been described in the context of the *HoxA* cluster following the deletion of two CTCF sites (Narendra et al., 2016).

4.2 Interpretation of the results

In the second section of this chapter, we will discuss the results previously presented in the light of the five different models proposed above.

4.2.1 Inversion of *Hoxd11*

The major consequence of the inversion of the *Hoxd11* locus is the significant reduction of *Hoxd12* mRNA. This knock-down is observed in a variety of embryonic structures whenever these two genes are co-expressed; we reported this observation along the AP axis in the trunk in the nested domain of expression of *Hoxd12*, in the distal autopod as well as in metanephric kidneys. However, outside of this disruption, the regulation of the *HoxD* cluster appears qualitatively unaltered by the inversion of this locus. We did not uncover substantial disruptions in the global collinear expression of *Hoxd* genes along the AP axis, neither significant global changes in expression of central *HoxD* genes in the kidneys nor the posterior *Hoxd13* in the limb autopod. Overall, the diminution of *Hoxd12* transcripts does not trigger ostensible homeotic phenotypes, this inversion appears therefore well tolerated by the animal. By itself, this result can hardly be considered as a surprise. Previous studies have indeed showed that the homozygous loss-of-function of *Hoxd12* is resulting in a very modest and limited malformation of the distal part of limb as well as subtle changes in the axial skeleton (Davis and Capecchi, 1996; Hérault et al., 1998; Kondo et al., 1996), emphasizing the apparently dispensable function of this gene.

In order to appreciate the functional causes resulting in the disruption of *Hoxd12* in this allele, we will examine this result under the light of the mechanisms presented in the previous section of this chapter. In a first instance, we shall consider first the two types of transcription dependent mechanisms presented: *promoter clearing* and *transcriptional interference*. This class of perturbation is particularly relevant to discuss here because of the transcriptional leakage released by the *Hoxd11* locus. Indeed, the telomeric breakpoint of *HoxD* $^{inv(11)}$ is positioned within *Hoxd11* 3'UTR, upstream of the annotated transcription stop site of *Hoxd11*. Therefore, the inversion of *Hoxd11* disconnects this locus from its normal 3'UTR and results in a significant transcriptional activity leaking from *Hoxd11* and extending towards and over *Hoxd12*.

On the one hand, no evidence currently supports a hypothetic mechanism of *promoter clearing*. The inversion of *Hoxd11* correlates indeed with a prominent decrease of *Hoxd12* transcript in the developing kidneys as well as along the trunk and in the autopods, suggesting a consistent loss-of-expression of this gene. This observation is thus in opposition with the predicted gain-of-expression of the *Hox* gene posterior to the inversion, as predicted by this model (Fig. 27A). In this aspect, the fact that in wild type metanephros the number of cells expressing *Hoxd12* is lesser than the number of cells expressing *Hoxd11* is an important observation to exclude a massive up-regulation of *Hoxd12* in the inv(*11*) allele (Patterson and Potter, 2004).

On the other hand, the results obtained by strand specific RNAseq from both the autopod and the kidneys suggest the existence of an active mechanism of transcriptional interference between the inverted *Hoxd11* and *Hoxd12*. Interestingly, this interference seems to affect essentially *Hoxd12*, by disrupting either the transcriptional output of this gene or the stability of this mRNA whereas *Hoxd11* transcripts are not severely affected. These observations are in accordance with the mechanisms designated under the term of *transcriptional interference* (Fig. 27B). An important point to examine here is the substantial fraction of transcripts generated by the inverted *Hoxd11* transcriptional unit and stopped between *Hoxd11* and *Hoxd12*, about 800 bp downstream of the latter gene (Fig. 12C and 13C, black arrowheads). We think that this reduction of RNAseq signal corresponds with a transcriptional termination site efficient enough to halt an important portion of RNAPs transcribing the inverted *Hoxd11* gene,

therefore producing functional and stable mRNA of *Hoxd11* detectable by stranded mRNA RNAseq. This efficient but not absolute termination site insures that not every transcript generated by the inverted *Hoxd11* locus interferes with *Hoxd12*, an important feature to avoid the systematic collision between the two transcriptional units. We think that such a configuration would indeed results in a pronounced reduction of both transcripts, a result not observed in our experiments.

Interestingly, we did not detect any canonical (AAUAAA) motifs typical of RNA polymerase II termination at the position where the RNAseq signal appears to diminish strongly, downstream of *Hoxd12* (Cheng et al., 2006; Richard and Manley, 2009). However, we found a combination of alternative stop sites, sharing several characteristic compatible with the generation of mature *Hoxd11* mRNA (Cheng et al., 2006; Proudfoot, 2016). Besides, another putative transcription termination site lies on the anti-*HoxD* DNA strand between *Hoxd12* and *Hoxd11*, a stem-loop (hairpin) structure coupled with a polyA stretch positioned close to *Hoxd12* 3'UTR (200 bp downstream). This arrangement is reminiscent of a Rho-independent terminator sequence, typically found in *Escherichia coli* (Lesnik et al., 2001; Naville et al., 2011) and may provide a clarification for the production of a stable mRNA encoding the inverted *Hoxd11* gene.

It is nonetheless relevant to note that such a transcriptional behavior is not artificial and does not result from the disruption of *d11* 3'UTR in the inv(*11*). *Hoxd11* shows indeed a similar feature in a wild type situation, in presence of an intact 3'UTR. Different alternative transcripts spliced either from the main promoter of *d11* or from RX appear to leak over the promoter and the gene body of the anterior neighbor *Hoxd10* (Montavon, 2008). These transcripts are therefore detected and assigned indiscriminately to the gene *Hoxd10* by ISH and RNAseq approaches. We proposed here that after the inversion of *Hoxd11*, the quantity of *Hoxd10* transcripts is reduced because of the relocation on the opposite DNA strand of the transcription of *Hoxd11*. An example supporting this interpretation is the transcriptional situation of *Hoxd10* in the limb autopod at E12.5. In a wild type configuration, *Hoxd10* is expressed with an expression pattern similar to *Hoxd11*, suggesting an association in their respective regulations (Zakany and Duboule, 2007). Of note, *Hoxd10* signal is reduced to a minimal level in the inv(*11*) allele when the transcriptional contribution from *Hoxd11* is reallocated towards *Hoxd12*, even though the promoters of *Hoxd11* and *Hoxd10* are physically positioned closer (Fig. 11 and Fig. 12). This result suggests that in a wild type configuration the promoter of *Hoxd10* is transcriptionally weakly active in this embryonic structure, an observation in accordance with the fact that the loss-of-function of *Hoxd10* does not elicit a phenotype in the autopod when all other *Hoxd* genes remain functional (Carpenter et al., 1997).

Out of the three mechanisms that are not directly associated with transcription, one configuration can be directly excluded: CTCF sites have indeed been voluntarily excluded from the design of the inv(*11*). This approach allows us to exclude legitimately the *topological disruption* model as a cause for the perturbations observed within this allele. From the two mechanisms that remain to be considered here, the *enhancer reallocation* model (Fig. 27D) is not supported by any of the results obtained with this allele. First of all, no gain-of-expression of *Hoxd12* is detectable, in opposition to what is proposed by this model. Complementarily, *Hoxd10* expression domain is not qualitatively affected in the inv(*11*), at least not in the embryonic stage studied by WISH (E12.5, see Fig. 11). Moreover, thanks to the specificity and precision of the CRISPR/Cas9 mutagenesis, we excluded the conserved regions RVIII, RIX and RX from the inv(*11*) allele.

This is an important point to consider, as at least one of those elements have been demonstrated to harbor *cis* enhancers responding to signaling factors controlling *HoxD* transcription along the AP axis (i.e. Gdf11/Smad in the RVIII, Gaunt et al., 2013). Lastly, the quantitative reduction of *Hoxd12* observed

in the autopod and in the metanephros is independent of any enhancers relocation because long-range elements controlling *Hoxd* genes transcription in these two structures are located respectively in the center of the C-DOM (autopods) and in the T-DOM, either within the central fraction of the *HoxD* or in the adjacent regulatory landscape (kidneys). Therefore, the reallocation of putative *cis* enhancers within the *HoxD* cluster cannot explain the disruption of *Hoxd12* expression as observed in the inv(*11*). Lastly, we did not obtain any experimental results supporting the alleged mechanism of *promoter sharing* (Fig. 27C) in this allele and are here able to rule out confidently this hypothesis. Indeed, the respective domains of expression of both *Hoxd11* and *Hoxd10* do not appear to merge and remain partially exclusive. This observation is particularly apparent along the AP axis of homozygous inv(*11*) embryos, where the metameres patterned exclusively by *Hoxd10* conserve their identity in the mutant allele (Fig. 11) whereas no gain-of-expression of *Hoxd10* is observed in the autopod even though the inverted *Hoxd11* is highly transcribed in this tissue (Fig. 12).

In conclusion, from the five mechanisms examined here, the *transcriptional interference* scenario is the only one which matches extensively the experimental data presented. Further investigations are still required to uncover the exact factors responsible for the knock-down of *Hoxd12*. In our opinion, two series of experimentations can help to discriminate between different possibilities. In a first instance, the presence of siRNA corresponding to the sequence of *Hoxd12* would be a clear indication of active RNA interference pathways at play at this locus in this allele. This information can be nowadays easily obtained thanks to the adaptation of NGS techniques to sequence classes of short RNAs (<100 bp) isolated from relevant tissues, such as the autopod in this case. In a second instance, a ChIPseq of RNA polymerase II exhibiting the phosphorylation of the Ser5 might be an instructive dataset to determine whether the amount of RNAP released from *Hoxd12* promoter is affected (Harlen and Churchman, 2017).

4.2.2 Inversion *Hoxd11-d12*

The inversion of the locus *Hoxd11-d12* in the inv(*11-12*)*d13lacZ* allele leads to a pronounced gain-of-expression of the posterior *Hoxd13lacZ* in a variety of tissues and organs, normally patterned by the central *Hoxd11* gene and controlled by the T-DOM. In these structures, the posterior *Hoxd13lacZ* gene is normally not expressed and the C-DOM is typically inactive. In addition to the intestinal hernia (Delpretti et al., 2013; Kmita et al., 2000a), we demonstrated that this gain of expression is also detectable in the metanephric kidneys as well as along the AP axis of the mutant embryos, typically following the pattern of expression of *Hoxd11* (Fig. 15 and Fig. 17). In these two structures, we found a common characteristic: an important transcriptional activity released by the inverted *Hoxd11* locus, apparently similar to the one observed in the inv(*11*) and covering the posterior extremity of the *HoxD* cluster. This transcriptional feature appears to be inherent to the inv(*11-12*)*d13lacZ* allele, as a similar observation was made in mutant caecum (Delpretti et al., 2013). These results thus legitimately raises the question of a potentially causative association between the anti-*Hox* transcription and the gain-of-expression of *Hoxd13lacZ* (Delpretti et al., 2013). Of note, such a possibility would be in stark contrast with the results obtained with the inv(*11*) allele and discussed above. In an attempt to distinguish between the different mechanisms responsible for the gain-of-expression of *Hoxd13lacZ*, we shall review the results obtained with this allele using the theoretical framework already presented.

The experimental evidences supporting the *promoter clearing* model (Fig. 27A) to explain the dysregulation of *Hoxd13lacZ* are neither definite not absolute, making therefore the validation of this hypothetic mechanism a complicated task. On the one hand, the antisense transcripts leaking from *Hoxd11* over the 5' extremity of the *HoxD* cluster correlates systematically with the gain-of-expression of *d13lacZ*. This activation is reported exclusively in tissues expressing *Hoxd11* and the *Hoxd13lacZ* locus

is subjected to a complete exchange of histones marks, shifting from a transcriptionally silenced landscapes (high quantity of H3K27me3, absence of H3K4me3) to an active one (limited amount of H3K27me3, high coverage of H3K4me3, see Fig. 18). In an effort to address rigorously the relationships between the antisense transcription generated by the *Hoxd11* locus and the gain-of-expression of *d13lacZ*, we deleted the *Hoxd11* locus from the inv(*11-12*)*d13lacZ* allele. Interestingly, this reconfiguration failed utterly to suppress the gain-of-expression of the reporter in the caecum, along the AP axis and in the metanephros of the double mutant embryo (Fig. 20 and 21). However, in this latter instance, we reported that *Hoxd12* shows in turn an aptitude to transcriptionally leak over the antisense of *Hoxd13lacZ*, albeit in a lower extent than *Hoxd11*. Considered together, these elements are supportive of the *promoter clearing* model.

But, on the other hand, the combination of RNAseq and ChIPseq data generated with the inv(*11-12*)*d13lacZ* allele showed us that low transcription activity is not sufficient to antagonize PcG-dependent repression, such as the *promoter clearing* model would imply. An important result supporting this interpretation is the preservation of the posterior extremity of the PcG domain encompassing the *HoxD* cluster, even though part of the locus is traversed by a transcriptional activity (Fig. 18). More precisely, the promoter of *Evx2* located within the C-DOM next to *Hoxd13lacZ* remains largely repressed and transcriptionally inactive in the metanephros of the inv(*11-12*)*d13lacZ* mutant embryos. The distribution of histones marks remains undisturbed (H3K27me3 rich, H3K4me3 poor) across the locus even though a clear transcriptional activity is visible between *Hoxd13lacZ* and the 3' extremity of *Evx2* on the anti-*HoxD* DNA strand (Fig. 18). The source of this transcriptional activity remains to be established but is reminiscent of the capacity of *Hoxd13* to display a bidirectional transcriptional activity, a feature observed in wild type embryos in different tissues where these two genes are co-expressed.

These transcripts cover *Evx2* without affecting the regulation of this gene. They can be nevertheless detected by WISH with the help of an anti-*Evx2* probe. This observation, i.e. the gain of *Hoxd13lacZ* and *Evx2* in the caecum of inv(*11-12*)*d13lacZ* embryos prompted Delpretti et al., to suggest that the posterior domain of the *HoxD* cluster was globally activated in response to the inversion of *Hoxd11-d12* in the caecum (Delpretti et al., 2013). The evidences presented here and acquired in the metanephros support a different interpretation; a very local and limited disruption of the repressive machinery centered at the *Hoxd13lacZ* locus. Hence, in view of the elements reviewed above, we developed the opinion that the transcriptional leakage generated by the inverted *Hoxd11* may eventually contribute in the gain-of-expression *Hoxd13lacZ*, as we were not in a position not refute this hypothesis. However, we are confident that the exclusive presence of the transcriptional machinery is not sufficient to thoroughly destabilize and remove PcG repression, thus triggering the active transcription of a given gene, an interpretation deduced from the absence of H3K4me3 at the locus of *Evx2* in the mutant metanephros.

Because of the outcome of the inversion of the *Hoxd11-d12* locus, the existence of *transcriptional interference* (Fig. 27B) mechanisms over the locus of *Hoxd13lacZ* could be at first hand disregarded. However, we cannot rule out that the transcriptional convergence observed between *Hoxd11* and *Hoxd13lacZ* is not resulting in some sort of negative obstruction. A relevant point to note is the *lacZ* staining in the inv(*11-12*)del(*11*)*d13lacZ* allele which appears stronger and quantitatively increased. This observation might imply that the transcriptional output of *d13lacZ* is actually increased when the antisense transcription covering this locus is reduced, as *Hoxd12* seems to release a reduced amount of leakage over the antisense of *Hoxd13lacZ* (Fig. 21B). In any case, it is evident that the consequences of any hypothetical *transcriptional interference* are overwhelmed by activating mechanisms and signals resulting in a specific gain-of-expression of *Hoxd13lacZ* in a variety of embryonic tissues. The situation is particularly flagrant in the autopod, where the convergent transcription of *Hoxd11* and *Hoxd13lacZ*

does not seem to affect appreciably neither the regulation nor the transcriptional outputs of both genes in any appreciable way (Kmita et al., 2000a). Therefore, *transcriptional interference* cannot be ruled out at this time but should be considered as an issue of secondary importance for the understanding of this peculiar allele.

We have not yet generated robust experimental data that permit us to examine the *promoter sharing* (Fig. 27C) model in the inv(*11-12*)*d13lacZ*. However, the results of WISH experiments performed in E12.5 mutant autopods and published in the initial article describing this allele provides a valid set of results which permits us to rule out this hypothesis, at least in the autopods (Kmita et al., 2000a). The inversion of the *Hoxd11-d12* locus results in the reallocation of *Hoxd12* upstream of *Hoxd10*, at a distance of 11 kb (Fig. 14). But the regulation of *Hoxd12* and *Hoxd10* remains clearly distinct in this tissue, *Hoxd12* conserving an important level of transcription whereas *Hoxd10* transcription is completely lost (Kmita et al., 2000a) in a fashion reminiscent of the situation discussed with the allele inv(*11*) in the same tissue. A similar separation of the patterns of expression between *Hoxd12* and *Hoxd10* is also clearly visible along the AP axis of the mutant embryos (data not shown).

Enhancer reallocation (Fig. 27D) is a mechanism that is complicated to address thoroughly in the context of this allele. The inverted *Hoxd11-d12* locus includes a couple of highly conserved sequences disposed between *Hoxd13* and *Hoxd10*, i.e. RVIII, RIX and RX already presented in this work. These sequences are therefore translocated in the inv(*11-12*)*d13lacZ* allele. Some of these elements harbor local short-range enhancers, contributing to adjust and maintain the different *HoxD* expression domains during embryogenesis. As presented in the context of the inv(*11*), RVIII is notably known to harbor a binding site for the Gdf11/Smad factors active on *HoxD* transcription along the AP axis (Gaunt et al., 2013; Gérard et al., 1996). It is therefore legitimate to consider that repositioning these elements might participate in the gain-of-expression of *Hoxd13lacZ* in selected tissues and structures.

But the relations between *enhancer reallocation* and the dysregulation of *Hoxd13lacZ* are not exclusively causative. This hypothetic mechanism is indeed not providing a valid explanation to explain *Hoxd13lacZ* gain-of-expression in two embryonic structures extensively studied here. First, *Hoxd* genes regulation in the caecum strictly depends on long-range enhancers located within the telomeric regulatory landscape (Delpretti et al., 2013). Second, central *Hoxd* genes regulation in the developing metanephros depends on enhancers located within the T-DOM, telomeric to the inverted locus (Di-Poï et al., 2007). In these two examples, the genetic position of important *HoxD* enhancers remains hence unchanged in the mutant inv(*11-12*)*d13lacZ* allele. Therefore, the *enhancer reallocation* model cannot be taken into account to explain to gain-of-expression *d13lacZ* observed in at least these two secondary structures and is thus dispensable to activate the *d13lacZ* reporter in structures where *HoxD* rely on long-range regulative inputs. But it is not possible to exclude that enhancers reallocation might contribute in the disruption of *Hoxd13lacZ* regulation in some other situations, notably in the anteriorization of *d13lacZ* reported in the trunk along the AP axis.

Finally, we presented evidences obtained in the metanephros to address the possibility that a CTCF site, reallocated in the inversion, might provoke the gain-of-expression of *d13lacZ* by disrupting locally the topological organization of the *HoxD* cluster. In our view, this hypothesis, i.e. *topological disruption* (Fig. 27E), provides a pertinent framework to explain several particularities of the inv(*11-12*)*d13lacZ* allele. It is appropriate to note that this interpretation is deriving from the initial analysis proposed by Kmita et al. in the original description of this allele (Kmita et al., 2000a). Back then, the authors postulated the existence of an undetermined polar silencer positioned between *Hoxd13lacZ* and *Hoxd12*. Following the translocation of this element between *Hoxd12* and *Hoxd10* and the inversion of

its sequence, the posterior *Hoxd13lacZ* reporter gene is assumed to be released from its exclusive regulation and topological insulation, hence capable to response to regulatory inputs located within the T-DOM and normally constrained at the central *Hoxd* genes (Kmita et al., 2000a). In view of the alleged loop extrusion model (Fudenberg et al., 2016; Sanborn et al., 2015) and numerous experimental evidences demonstrating the existence of a resilient topological boundary separating *Hoxd13* from the central *Hoxd* genes and their regulations in the T-DOM (Andrey et al., 2013; Rodríguez-Carballo et al., 2017), we postulate that the polarized silencer discussed by Kmita et al. is actually the CTCF binding site located between *Hoxd12* and *Hoxd13*. This site is of great interest as it stands at a very specific position. It is indeed the most outward CTCF site of the C-DOM, the first out of four sites found at the posterior extremity of *HoxD* and systematically orientated facing the C-DOM (Fig. 6 and 7). It is also the only site located downstream of *Hoxd13*. For these reasons, we think that changing the genetic position and orientation of this site can affect locally the topological organization of the locus and may in turn reduce the regulatory insulation of the posterior *Hoxd13lacZ*.

We showed that this site, i.e. designated as d13_d12 CTCF, is occupied by both CTCF and Rad21 in a control configuration (Fig. 22). These results suggest that d13_d12 CTCF is likely engaged in long-range interactions with convergent CTCF sites located within the C-DOM. In combination with the three other CTCF sites located between *Evx2* and *Hoxd13,* we think that this site is setting the boundary between the T-DOM and the C-DOM and contributes in the tight and exclusive topological contacts observed between the *Hoxd13* locus and the posterior extremity of *HoxD*. This segregation of the posterior *HoxD* is observed by 4Cseq in metanephros in both wild type and control *HoxD* ^{d13lacZ} allele. Following the inversion of the *Hoxd11-d12* locus, d13_d12 CTCF site is translocated upstream of *Hoxd10*, its sequence is inverted, and it faces the transcriptionally active T-DOM and its numerous convergent CTCF sites (Fig. 6 and 7). This reallocation is not affecting the functionality of the site and we scored a significant enrichment of Rad21 at this site, suggesting its probable integration within the T-DOM (Fig. 22). 4Cseq profile generated from the *d13lacZ* locus reveals a significant change in the conformation of the chromatin and the establishment, in the mutant allele, of a local hub of short range interactions encompassing the active *Hoxd13lacZ*, *Hoxd11* and *Hoxd12* genes (Fig. 22).

At the present time, there is no experimental evidence indicating that the global structure of either the C-DOM or the T-DOM are disrupted in the inv(*11-12*)*d13lacZ* allele. But in our opinion, there are no valid reasons to expect massive changes in these two structures, as the inversion of the single d13_d12 CTCF site, standing on the edge of the C-DOM, would very likely result in a local shift of the TAD boundary without affecting the internal structure of either the C-DOM or the T-DOM. An argument in favor of this interpretation is the pattern of expression of the reporter *Hoxd13lacZ*, which conserves the unambiguous *Hoxd13* identity in the developing autopod and marks the five presumptive digits, whereas the expression domains of the inverted *Hoxd11* and *Hoxd12* are restricted to a fraction of the autopod (Kmita et al., 2000a). This observation leads us to consider that the C-DOM and its enhancers are still able to contact and activate specifically *Hoxd13lacZ*, a feature likely assured by the three CTCF sites located upstream of *Hoxd13lacZ* and unmoved by the inv(*11-12*)*d13lacZ* mutation. Moreover, we did not observe a significant dysregulation of *Evx2* in this mutant, having demonstrated that the signal detected by WISH and RNAseq in a variety of tissues is not specific to the activity of *Evx2* promoter but is indirectly resulting from the activation of the *Hoxd13lacZ* promoter. Similarly, the integrity and the functionality of the T-DOM is not particularly affected, as the global expression of the *Hoxd* genes, up to *Hoxd10*, are qualitatively conserved and unmoved. Therefore, we propose that the translocation and inversion of d13_d12 CTCF site destabilizes locally the organization of the boundary between the C-DOM and the T-DOM by creating an artificial topological no-man's land. This zone encompasses the

three posterior *Hoxd* genes and might alleviate the insulation of *Hoxd13lacZ*, which in turn is capable to response to regulatory inputs otherwise restricted to the T-DOM.

The determination of the exact contribution of the d13_d12 CTCF site in separating the *HoxD* cluster into two domains is a relevant question to tackle. It appears indeed that this segmentation lies at the base of *HoxA/D* bimodal regulation. It is also relevant to note that such a CTCF functional site is observed across the four murine *Hox* clusters, as well as in the chicken *Hox* clusters (Nayuta Yakushiji-Kaminatsui, personal communication). In each of these examples, a paralogous 13_12 CTCF site is found orientated facing the posterior 5' extremity of the considered *Hox* cluster, therefore supporting a specific topological organization already presents in the primordial *disorganized* cluster. This point is currently addressed by a CRISPR/Cas9 mutagenesis approach directed against the d13_d12 CTCF site. This experiment is ongoing and is currently performed in a wild type context, using a single gRNA to generate a precise mutation within the central motif of the d13_d12 CTCF. This mutation is expected to abrogate CTCF binding, which should in turn abolish the enrichment of Rad21/cohesin at this locus and potentially affect the distribution of long-range contacts and the position of the boundary between the C-DOM and the T-DOM. Even though this approach will not produce the exact same disposition of CTCF sites as the one observed in the inv(*11-12*)*d13lacZ*, i.e. one less "C-DOM" CTCF site compared to one less "C-DOM site" and one more "T-DOM" site, this allele should nevertheless shed light into the importance of this site in insulating *Hoxd13*. We speculate that any gain-of-expression of *Hoxd13* shall be hypomorphic and viable, similarly to the inv(*Nsi-Itga6*) allele discussed later in this work, a situation that could not be guaranteed in an inv(*11-12*) allele reconstructed in the presence of a functional *Hoxd13* gene. Moreover, this allele would present the advantage to be fully compatible with a CaptureHiC protocol, using a set of RNA probes spanning this fraction of the chromosome, a protocol developed these last months in the Duboule lab (unpublished data from Chase Bolt and Lucille Delisle).

Our understanding of the *HoxD* $^{\text{inv}(11\text{-}12)d13lacZ}$ allele is not conclusive yet as numerous factors are affected by this inversion: the respective position of two *Hoxd* genes, the directionality of their transcription and the production of an artificial transcriptional activity over the posterior extremity of the *HoxD* cluster, the position and the orientation of a CTCF binding site and the translocation of presumed *cis* enhancers. Moreover, this allele bears a loss-of-function of *Hoxd13* caused by the knock-in of a *lacZ* cassette within this gene. Together, these factors might amalgamate to produce the regulatory disorder of *d13lacZ* and are complex to separate and consider individually. However, this allele illustrates well the variety of challenges that would represent a spontaneous inversion happening within a consolidated *organized Hox* cluster, a situation overall well tolerated by the animal here thanks to the loss-of-function of the posterior *Hoxd13* which protect the animal from any damaging gain-of-expression of this posterior gene.

4.2.3 Inversion *Hoxd12*

In contrast against the two alleles discussed above, *HoxD* $^{\text{inv}(12)}$ seems to be well tolerated by this genetic system and does not yield severe changes in the *HoxD* endogenous regulation. Heterozygous animals do not show any deleterious phenotypes and according to data currently available the global regulation of the *HoxD* cluster seems overall unaffected by this inversion. Among the five theoretical models used to discuss these allele, *topological disruption* (Fig. 27E) cannot be considered here. The original design of this allele was indeed aimed at producing an inversion not affecting the position nor the orientation of any of the CTCF sites found at the *HoxD* cluster, similarly to the inv(*11*) allele. Likewise, the *enhancer reallocation* (Fig. 27D) hypothesis will not be discussed here in the context of the E13.5 metanephros RNAseq (Fig. 24), as the inverted locus is deprived of short-range *cis* enhancers active in the developing kidneys (Di-Poï et al., 2007).

On the other hand, *promoter sharing* (Fig. 27C) can be at least partially addressed, the metanephros being a relevant read-out to assess this point. As presented above, we observed a quantitative gain of *Hoxd12* transcripts, which seems significant and is likely reflecting an upregulation of *Hoxd12* promoter. This observation is supported by two arguments. First, the 3'UTR of *Hoxd12* is untouched by the inversion of the *Hoxd12* locus, the breakpoint targeted by the gRNA31 being positioned downstream of *d12* 3'UTR. Therefore, any post-transcriptional mechanisms controlling *Hoxd12* mRNA stability are conserved in the inv(*12*) allele (Fig. 23). Second, the limited but detectable anti-*HoxD* transcriptional leakage emitted by the inverted *Hoxd12* locus is not observed in the non-inverted control transcriptome but remind us of the transcriptional leakage observed in the E12.5 autopod RNAseq, tissues where *Hoxd12* is quantitatively more expressed (Fig 12B). However, *Hoxd12* is clearly not up-regulated to the magnitude of transcription displayed by *Hoxd11* in E13.5 metanephros, suggesting that these two genes might conserve a separated regulation and do not extensively cross-activate their respective promoters. This interpretation needs to be confirmed by additional experiments, based on the visualization of the respective expression domains of *Hoxd11* and *Hoxd12* across different tissues and structures.

Of the two transcription-depend mechanisms, *promoter clearing* is not observed here and can therefore be excluded. Indeed, the inversion of *Hoxd12* locus provokes a limited transcriptional leakage on the anti-*HoxD* DNA strand progressing towards *Hoxd13* but it fails to reach this gene which remains silenced, its regulation apparently unaffected at least in the embryonic kidneys. This observation reminds us of what is seen in the inv(*11*) allele, where a similar transcriptional noise is detected upstream of *Hoxd12* on the anti-*HoxD* DNA strand, without neither reaching *Hoxd13* nor affecting its endogenous regulation and transcription. On the other hand, *transcriptional interference* at the *Hoxd13* locus is impossible to address here as we are not in possession of data collected in tissues expressing concomitantly *Hoxd13* and *Hoxd12*, such as in the distal autopod. However, based upon the strong transcriptional insulation observed between *Hoxd12* and *Hoxd13* and the presence of several sequences favorizing transcription to stop 3' of *Hoxd13*, both on the *HoxD* and the anti-*HoxD* DNA strand, this hypothesis appears unlikely. It will nevertheless be tested by the production of an RNAseq dataset in inv(*12*) autopods.

In conclusion, further work is required to confirm that the targeted inversion of *Hoxd12* locus represents a sustainable intra-cluster inversion. But based on the preliminary results, this last allele seems to have a neutral impact on the *HoxD* cluster. At the present time, inv(*12*) exemplifies the fact that the polarized disposition of *Hox* genes, within *organized Hox* clusters, may not be a systemic requirement for *organized Hox* clusters function. However, the chance to obtain randomly such a precise and specific reconfiguration are obviously nil.

4.2.4 Demonstration of the dominant negative nature of *Hoxd13*

Out of the three intra-cluster inversions that we studied and reviewed here, none showed a major phenotype. This observation is explained by two parameters; on the one hand, the apparently limited homeotic function of *Hoxd12* (Davis and Capecchi, 1996; Hérault et al., 1998; Kondo et al., 1996) and on the other hand the knock-in of the *lacZ* reporter within the posterior *Hoxd13*. In the first case considered, the quantitative and qualitative alterations of *Hoxd12* expressions (i.e. anteriorization of *d12* in the *HoxD* $^{\text{inv}(11\text{-}12)d13lacZ}$ and the knock-down of *d12* in the allele *HoxD* $^{\text{inv}(11)}$) is apparently well tolerated by the mutant mice. In the latter case, the anteriorization of *d13lacZ* in the inv(*11-12*)*d13lacZ* is compensated by the loss-of-function of this gene, therefore masking the dire results of a gain-of-expression of this gene for the development of the embryo.

We took advantage of the allele *HoxD* $^{\text{inv}(Nsi\text{-}Itga6)}$ to illustrate the effects resulting from the interruption in the collinear activation of the *HoxD* cluster and apprehend the consequences of the activation

of *Hoxd13* in tissues where it is usually maintained repressed. This large inversion shows a robust tendency to express *Hoxd13* in places usually exclusively patterned by *Hoxd11* and central *Hoxd* genes, e.g. in the proximal limb. This gain-of-expression of posterior *Hoxd* genes affects the identity and shape of this anatomical structure in a classical illustration of the homeotic function of *Hox* genes (Tschopp and Duboule, 2011). It has been demonstrated that this transcriptional behavior is triggered by the genetic disaggregation of the C-DOM and the disappearance of the organized series of long-range contacts observed notably between convergent CTCF sites, resulting in a reduced topological insulation of the posterior extremity of the *HoxD* cluster, *Hoxd13* and *d12* falling under the control of the T-DOM regulation (Andrey et al., 2013).

In the case of the developing metanephros, the disconnection of *Hoxd13-d12* from their regulatory landscape results in the up-regulation of the posterior *Hoxd* genes in a similar manner as previously described in the zeugopod (Fig. 25A and C). Of interest, the up-regulation of both *Hoxd12* and *Hoxd13* in the inv(*Nsi-Itga6*) is also accompanied by a concomitant quantitative reduction of transcripts attributed to the central *Hoxd10* and *Hoxd9*. In an intriguing coincidence, a very similar observation was also made in the metanephros of the inv(*11-12*)*d13lacZ* mutant (Fig. 17D). Together, these results suggest that the *HoxD* cluster abides for a factor limiting the transcriptional rate of these genes, where the activation of the posterior *Hoxd* genes leads to a reduction of the transcripts of the central *Hoxd* genes. We speculate that this transcriptional reduction, which is observed across several datasets, is actually resulting from unknown mechanistic constrains specific to the chromatin at the given locus.

We demonstrated that the renal phenotype observed in inv(*Nsi-Itag6*) adult heterozygous mutant was the sole consequence of the gain-of-expression of *Hoxd13* during the embryonic development of this organ (Fig. 26). The changes in central *Hoxd* genes expression (i.e. the variation in the qualitative or quantitative *Hox* code) do not contribute in the renal inv(*Nsi-Itga6*) phenotype, an alternative hypothesis that could have been proposed based on recent studies demonstrating the role of *Hox9/10/11* genes in maintaining a fixed cellular identity within the developing organ (Drake et al., 2018; Magella et al., 2018). Likewise, the rescue of the renal phenotype by the deletion of *Hoxd13* homeodomain in the inv(*Nsi-Itga6*) excludes several alternative causes that could have been anticipated, such as e.g. the loss-of-function of *Itga6* or other disruptive genetic variations that could have been caused by the nature of this large inversion.

These result are therefore confirming the relevance and the validity of the *posterior prevalence* model (Duboule, 1991; Duboule and Morata, 1994), particularly when considering *Hoxd13* and its homeotic dominant negative function which has been already observed in several different developmental models (Beccari et al., 2016; Tschopp and Duboule, 2011; Young et al., 2009). The study of this allele can thus be considered as another example of the importance to keep *Hoxd13* and its regulation tightly contained and insulated within the C-DOM, as assured by the integration of the bimodal regulation of the *HoxD* genes within domains of preferred topological interactions. We infer that any spontaneous mutation that would disrupt this organization would likely generate a set of heavy homeotic phenotypes in the mutant animal of relatively limited viability.

4.3 Concluding remarks and perspectives

The aim of this thesis was to assess the importance of the transcriptional polarization found in mammalian *organized Hox* clusters. To address this question, we used the murine *HoxD* cluster as a model to produce targeted inversions of posterior *Hoxd* genes. We then studied the structural and functional perturbations caused by these reconfigurations in a set of embryonic tissues known to be patterned by *Hoxd* genes. We reported here that specific inversions differently impacted *HoxD* cluster regulation and function.

First, we showed that inversion of the *Hoxd12* gene resulted in a minimal disruption of *HoxD* function. Therefore, inv(*12*) represented an example supporting the idea that within an *organized* cluster a single *Hox* gene can be stably inverted. This remarkable result is unexpected and counter-intuitive, in view of the evolutionary conservation of *Hox* transcriptional polarity in all characterized gnathostome *Hox* clusters. The systematic transcriptional polarization observed in the *organized Hox* clusters indeed suggested *a priori* a specific role for the spatio-temporal collinear activation of these genes. An extensive breeding program would need to be performed in order to explore potential reproductive inefficacy of inv(*12*) homozygous individuals, or aging related alterations.

Two other alleles presented here attest to the singularity of the above finding. In a first instance, the inversion of the *Hoxd11* gene reallocated a significant transcriptional activity on the anti-*HoxD* DNA strand and promoted the accumulation of *Hoxd12* antisense transcripts. Coincidentally, accumulation of the native *Hoxd12* transcripts was substantially reduced. The molecular mechanism of this interference remains yet to be explored. In the second instance, we revisited the consequences of the simultaneous inversion of two *Hoxd* genes, *Hoxd11* and *Hoxd12*, activating specifically the reporter *Hoxd13lacZ*. We showed that in addition of the inversion of these two genes, a CTCF site was translocated and inverted in this allele. This reconfiguration seemed to perturb the the topological organization of the *HoxD* cluster and changed locally the distribution of long-range contacts, disrupting the regulation of these genes. Together, the latter two examples indicated that spontaneous inversions in an *organized Hox* cluster could change drastically the regulation of the affected *Hox* cluster. In addition to exchanging *Hox* gene order and thus potentially affecting *Hox* gene open reading frame integrity, inversions would be therefore of uncertain evolutive value. To conclude, once set in an *organized* cluster, *Hox* genes are locked in a complex and compacted structure, not allowing major changes in *cis*. This restriction transfers *de facto* the appearance of new evolutionary regulatory mechanisms to the adjacent regulatory landscapes, rich in pleiotropic *trans* enhancers.

In our opinion, a more fruitful window of opportunity to invert successfully a single *Hox* gene could have been the primordial *disorganized Hox* cluster, an organization observed phylogenetically prior to the *Hox* condensation and the appearance of tightly *organized* clusters during the radiation of vertebrates. A question that shall remain open, being impossible to address experimentally, is to know to what extent *Hox* consolidation could have been successful starting from a *disorganized Hox* cluster showing a distribution of *Hox* genes incompletely transcriptionally polarized. We speculate that the consolidation may have been as successful and efficient as the one observed at the present time, on condition that *Hox* gene transcriptional leakage would remain minimal and controlled, therefore limiting interference mechanisms between convergent *Hox* genes.

5 MATERIALS AND METHODS

5.1 Mouse models

5.1.1 CRISPRed alleles

Three murine alleles were produced during this thesis: *HoxD* $^{inv(11)}$, *HoxD* $^{inv(11\text{-}12)del(11)d13lacZ}$ and *HoxD* $^{inv(Nsi\text{-}Itga6)del(13HD)}$. These alleles were obtained using the CRISPR/Cas9 technology in mouse zygote. Two specific guide RNA were cloned and used to generate each allele, designed to flank the locus of interest. Targeted inversion/deletion obtained from the resolution of the CRISPR/Cas9 induced DNA double strand breaks were then identified and extensively described in F0 adult animals. The gRNA used and their respective sequences are listed below. Each guide was cloned in a *BbsI* site in the pX330:hSpCas9 (Addgene ID 42230) vector (Cong et al., 2013). The efficiency and specificity of these guide sequence designs was evaluated *in silico* using different tools: CCtop (Stemmer et al., 2015), CHOPCHOP (Montague et al., 2014) and crispr.mit.edu (Zhang lab).

HoxD $^{inv(11)}$ and *HoxD* $^{inv(Nsi\text{-}Itga6)del(13HD)}$ were obtained by pro-nuclear injection (PNI) of an equimolar solution of the two pX330:hSp:Cas9 vectors containing the appropriate guides (Mashiko et al., 2013). These manipulations were done by Jozsef Zakany in Duboule lab, at UniGe. *HoxD* $^{inv(12)}$ was generated using a similar PNI approach by the EPFL transgenesis platform (TCF). *HoxD* $^{inv(Nsi\text{-}Itga6)del(13HD)}$ was obtained by targeting *Hoxd13* homeobox in zygote obtained from a cross between *HoxD* $^{inv(TgHd11lacNsi\text{-}Itga6)/+}$ males and wild type females. Lastly, *HoxD* $^{inv(11\text{-}12)del(11)d13lacZ}$ was obtained by the electroporation in zygote of a solution containing *Cas9* mRNA and the two desired guide, in-vitro transcribed (Hashimoto and Takemoto, 2015). The zygotes used were obtained from a cross between *HoxD* $^{inv(11\text{-}12)d13lacZ/+}$ males and wild type females. This manipulation was realized by Bénédicte Mascrez in the Duboule lab at UniGe.

Tissue biopsies obtained from weaned pups were extensively genotyped, identified F0 animals were separated and characterized by DNA Sanger sequencing at the breakpoints of the desired mutations. This information was then used to reconstruct the different genomes used in this work. These animals were then back-crossed over wild type (B6CBA)F1 mice for at least two generations and single heterozygous mutant animals were used to expand each alleles. Transgenic integration of the pX330:hSp:Cas9 vectors was regularly observed in animals pro-nuclear injected, across different founders and alleles. The segregation of these transgenes was monitored by PCR and only pX330 negative animals were used to expand the new alleles. Genotyping primers were designed to amplify specifically the different inversion/deletions breakpoints, working at routine PCR conditions established in Duboule laboratory (see 5.1.3). These primers are listed below.

The position of the CRISPR/Cas9 targeted mutations in the alleles *HoxD* $^{inv(11\text{-}12)del(11)d13lacZ}$ and *HoxD* $^{inv(Nsi\text{-}Itga6)del(13HD)}$ in *cis* over the original inv(*11-12*)*d13lacZ* and inv(*Nsi-Itga6*) allele was confirmed, when needed, by long-range PCR. Otherwise, the genotyping of the progeny of back-crossed F0 animals allowed us to discriminate between *cis* and *trans* mutations. F0, F1 and F2 animals obtained from CRISP/Cas9 mutagenesis directed by DNA injection were also genotyped for the transgenic integration of the pX330 vector with the primers referenced below.

project	sequence gRNA	ID	plasmid	used to generate
deletion, 2nd exon *Hoxd13*	CGGACATGTGCGTCTACCGA	g10	pFD10	*HoxD* *inv(Nsi-Itga6)del(13HD)*
	TTGTCTCTCCGAAAGGTTCG	g11	pFD11	
inversion/deletion, locus *Hoxd11*	CTACAGATCTCGCCGCGTTG	g27	pFD27	*HoxD* *inv(11)* *HoxD* *inv(11-12)del(11)d13lacZ*
	GGTTCTCTAGCGGAACGCGC	g28	pFD28	
Inversion, locus *Hoxd12*	TCCCTTCCTGAACAGTAAGA	g31	pFD31	*HoxD* *inv(12)*
	CGACCAATTGGGCGGCTCTC	g3	pFD32	

primer	sequence	combination	allele
FD126	TGGGTTCAATGACGCCTCTT	FD126/137 wt=635bp FD126/138 inv(11)=around 470bp	*HoxD* *inv(11)*
FD137	GAAAAGACAACCAGCACTGC		
FD138	TTTACGAGAGCTGCCAGACA		
g515	GTATGGAGTCCCAATTCCTGAAAAGGAG	g515/g516 wt=278bp g519/g520 inv(12)=539bp	*HoxD* *inv(12)*
g516	GCGCTAGCCTAAAGCAGGACC		
g519	CTTCGGTTAGGGAAGGAGACTCC		
g520	GCCAAGCTGGAACTGGTTAGGC		
g56	CCGTCCAATGTGCGTGTTTTCC	g56/134inv(Ns-Itga6)=around 650bp g56/202 wt=around 450bp	*HoxD* *inv(Nsi-Itga6)del(13HD)*
g134	GAGTTTCTCTTTGCTGTAATGAAGAGCTG		
g202	GCAAGCCACTTGGAAACAACTGTTAATGG		
FD59	AGGCTAATGTTTCCCCTGCT	FD59/60wt=320bp, del(13HD)=200bp	
FD60	CGCATTTAAACTGTCTGTGGC		
FD105	CATGGAGATGTATGCCTCGTG	FD105/106 inv(11-12)d13lacZ=117bp	*HoxD* *inv(11-12)del(11)d13lacZ*
FD106	AGGTAAACCAATGCCCAGAAT		
FD137	same as FD137	FD137/138 del(d11)=around 650bp	
FD138	same as FD138		
FD97	AATTACCTGGAGCACCTGCC	FD97/98 pX330=around 1045bp	Tg pX330
FD98	AGAAACAGGTCGGCGTACTG		

5.1.2 Other alleles

Routine genotyping was done by PCR amplification. Tissue biopsies were lysed overnight at 65°C in lysis buffer (10mM KCl, 20mM TrisHCl pH 8.0, 10mM (NH4)2SO4, 1mM EDTA, 0.1% Triton,) complemented with 0.5% proteinase K (20mg/ml). After pK heat-inactivation at 98°C for 10 minutes, PCR were performed with a standardized cycling protocol established in Duboule laboratory ((1x 94°3'), 2x(94°1', 62°1', 72°1'), 30x(94°45'', 62°45'', 72°45''), 1x(72°3')). The following alleles were used in this work and have been already described elsewhere: *d13lacZ* and inv(*11-12*)*d13lacZ* in Kmita et al., 2000, del(*8-13*)RXII in Tarchini et al., 2005, inv(*Nsi-Itga6*) in Tschopp and Duboule, 2011, TgN*d8* and TgN*d9* in Tschopp et al., 2012. All experiments were performed in agreement with the swiss law on animal protection under license number GE 81/14 attributed to D. Duboule.

5.2 *In situ* hybridization and X-gal staining

Standard protocols were used for whole mount *in situ* hybridization and X-gal staining. Briefly, WISH were performed using DIG labelled RNA probes (Woltering et al., 2009) available in the Duboule lab (*Hoxd13*, *d12*, *d11*, *d10* and *lacZ*) and already described respectively in Dollé et al., 1991, Izpisúa-Belmonte et al., 1991 and Gérard et al., 1996. *LacZ* detection in metanephros at E13.5 was done on eviscerated embryos, fixed in 4% PFA for 20 minutes at 4°C. Staining was performed overnight at 37°C, following the usual X-gal staining procedure, except that the staining solution was complemented with TrisHCl pH=7.4 (20uM final concentration) to avoid background signal in the metanephros.

5.3 Histology

Kidneys were dissected from sacrificed adult heterozygous mice, rinsed abundantly in 1x PBS and fixed overnight in 4% PFA before dehydration and paraffin embedding. Sections were produced and H&E stained at HCF (EPFL).

5.4 Water consumption

Water consumption was quantified during a series of 5 consecutive days by weighing each noon the water bottles from cages containing 2 to 4 adult mice of equivalent age and sex.

5.5 Genomic data

5.5.1 Generalities

Each dataset presented in this thesis has been mapped on GRCm38/mm10. Mutant chromosome 2 of *HoxD* $^{inv(11)}$, *HoxD* ^{d13lacZ} and *HoxD* $^{inv(11\text{-}12)d13lacZ}$ were reconstructed from Sanger sequencing of the breakpoints and the *lacZ* reporter of each allele, using a series of R scripts developed by Lucille Delisle with the help of Anamaria Necsulea. GTF annotations used in this work derive from ENSEMBL GRCm38.90 and are filtered against read-through/overlapping transcripts, keeping only transcripts annotated as 'protein-coding', thus discarding transcripts flagged as 'retained_intron', 'nonsense-mediated decay' etc. to conserve only non-ambiguous exons and avoid quantitative bias during data analysis by STAR/Cufflinks/HTSeq-count (Amândio et al., 2016). ChIPseq and RNAseq libraries were constructed and sequenced by UniGe iGE3 Genomics platform. These libraries, together with 4Cseq, were then sequenced on HiSeq2500 or HiSeq4000, single-end 100bp.

ChIPmentation libraries were sequenced at GECF (EPFL) on a NextSeq500, pair-end, with respectively 75bp and 37bp read length. Bioinformatic analysis were done using an internal Galaxy server held at the Duboule lab UniGe, regularly updated and enriched with workflows developed in collaboration with Lucille Delisle and Anamaria Necsulea. The 4Cseq datasets were analyzed with the HTSstation interface (BBCF, EPFL). Visualization of sequenced datasets on wild type and mutant genomes was done

on IGV v.2.3.93 and figures presented in this work were build using a genome browser developed initially by Anamaria Necsulea and assembled as a R script by Lucille Delisle.

5.5.2 RNAseq

RNA was extracted from micro-dissected tissues, stored in RNAlater at -80°C, using RNeasy plus micro kit (Qiagen) following the manufacturer recommendations. At least 1.5ug of total RNA (RIN>9) was submitted to the iGE3 Genomics platform and prepared for stranded RNA sequencing using TruSeq kits (Illumina). Samples were either poly(A)-enriched or ribosomal RNA depleted. Following sequencing, Illumina TruSeq adaptors were cleaved from the raw fastq files. Reads were then mapped to their respective genome using STAR with the ENCODE settings.

Strand specific coverage was computed from single-mapped reads and normalized to this number, for each RNAseq dataset. Normalized FPKM were computed by determining first coefficients extrapolated from a set of 1000 house-keeping stably expressed genes, in between the series of compared RNAseq datasets. These coefficients were then applied to the respective FPKM values, determined by Cufflinks. Biological duplicates were produced for inv(*TgHd11lacNsi-Itga6*) RNAseq. This global quantification workflow, containing also a DESeq2 analysis tool, was built with a series of R script developed by Lucille Delisle and Anamaria Necsulea.

5.4.3 ChIPseq and ChIPmentation

ChIP/ChIPmentation datasets were produced using the following commercial antibodies: anti-H3K4me3 (17-614, Millipore), anti-H3K27me3 (07-449, Millipore), anti-H3K27ac (ab4729, Abcam), anti-CTCF (61311, Active Motif) and anti-Rad21 (ab992, Abcam). Histones ChIP were produced from 5 pairs of metanephros micro-dissected from E13.5 embryos (approx. $1x10^6$ cells). Pairs of metanephros were individually fixed in 1% Formaldehyde for 20 minutes at room temperature and stored at -80°C. Genotyping was done on brain or tail biopsies. Sonication was done using a Bioruptor (Diagenod) in 1% SDS. ChIP was done by over-night incubation of protA/protG magnetic beads conjugated with the desired antibody, in 1.2ml dilution buffer, rotating at 30rpm 4°C. Washes were done using 1ml of washing buffers, incubated at 4°C for 2 minutes: 2x RIPA/2x RIPA-500mM NaCl/2x LiCl/2x TE. Beads were eluted and chromatin fixation was reversed during an over-night incubation at 65°C in presence of protK. The eluate was treated with RNase A, PCI purified, dehydrated and precipitated using NaOAc/Ethanol. DNA was then quantified and shipped to the iGE3 platform for TruSeq library construction.

The ChIPmentation protocol used in this work was largely adapted from Schmidl et al., 2015. Briefly, ChIPmentation datasets were produced from 4 pairs of metanephros (approx. $7.5x10^5$ cells) micro-dissected from E13.5 embryos. Metanephros were dissociated for 10 minutes at 37°C, 700 rpm, in 400ul of a solution containing 1x PBS complemented with 10% FCS and 5ul Collagenase 10mg/ml (sterile filetered**,** C1764 Sigma). Fixation was done in a final concentration of 1% FA (w/o methanol, 28908 Thermo) for 10 minutes at room temperature and cells were then stored at -80°C. Genotyping was done on brain or tail biopsies. Chromatin shearing was done using a Covaris E220, generously shared by the Lingner lab at EPFL, in 120ul sonication buffer (0.25% SDS) with the following settings: duty cycle at 2%, peak incident power at 105W, water level set at 6 with the amplifier engaged and a treatment of 8 minutes. The volume for the over-night IP (4°C, 25 rpm) was adjusted at 300ul using a dilution buffer lo lower the SDS concentration and buffer the pH (0.1% SDS, 20 mM HEPES). Following the last LiCl wash, magnetic beads were washed 2x in 1ml of 10uM TrisHCl pH=7.4 and finally resuspended in 24ul prewarmed tagmentation buffer (FC-121-1030, Illumina). 1ul of TDE1 enzyme was added to the mix and

the beads were incubated for 120" at 37°C. DNA purification, quantification and Nextera library construction were done following the published ChIPmentation protocol (Schmidl et al., 2015), using indexed Illumina adapters (FC-121-1011, Illumina). ChIP and ChIPmentation datasets were analyzed with a workflow developed by Lucille Delisle and running on the Duboule laboratory Galaxy server. Briefly, Illumina adaptors were removed from the reads, the single-end or pair-end reads were then mapped to the corresponding genome using Bowtie2. The final coverage was obtained after duplicates removal, estimation of the average fragments size and extension of the reads by MACS2.

5.4.4 4Cseq

Pairs of metanephros were individually micro-dissected from E13.5 embryos. Metanephros were then dissociated for 10 minutes at 37°C, 700 rpm, in 250ul of a solution containing 1x PBS complemented with 10% FCS and 5ul Collagenase 10mg/ml (sterile filtered**,** C1764 Sigma). Fixation was done in a final concentration of 2% PFA for 10 minutes at room temperature. Following glycine quenching and two 1x PBS washes to stop the fixation, the cells were then lysed and nuclei stored at -80°C. Genotyping of the corresponding samples was done on brain or tail biopsies processed in parallel. Pools of 20-25 pairs of metanephros of similar genotypes were then used to produce a double-digest libraries following the protocol described in Noordermeer et al., 2011, using *NlaIII* and *DpnII*. For each 4Cseq viewpoint, 11 individual PCR were run and pooled together, using between 100-150ng of the double-digested initial library per individual PCR. Series of viewpoints were then multiplexed together and sequenced with an HiSeq2500. The reads were then demultiplexed, mapped over wild type or mutant genomes uploaded in the interface, normalized and analyzed using the workflow provided by HTSstation (David et al., 2014), maintained at the time of this work by the BBCF facility at EPFL. The profiles were smoothened using a window size of 11 fragments, and the final coverage was normalized considering the total signal spanned 5Mb on the two sides of each viewpoint.

The complete list of 4Cseq primers used during this thesis is listed below. The four XXXX base pairs depicted in red represent optional barcodes that were added between the Illumina Solexa adapter sequences and the specific viewpoint the inverse forward primer to allow multiplexing of the same viewpoint throughout a single sequencing lane. The pair of *ilacZ* primers used in this study were kindly provided by Leonardo Beccari whereas the other primers were available in the laboratory and already published in different publications from the Duboule laboratory.

viewpoint	**primer**	**sequence**
Hoxd13	iHoxd13 fwd	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC-GCTCTTCCGATCTXXXXAAAAATCCTAGACCTGGTCATG
	iHoxd13 rev	CAAGCAGAAGACGGCATACGAGGCCGATGGTGCTGTATAGG
Hoxd11	iHoxd11 fwd	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC-GCTCTTCCGATCTXXXXAAGCATACTTCCTCAGAAGAGGCA
	iHoxd11 rev	CAAGCAGAAGACGGCATACGACTAGGAAAATTCCTAATTTCAGG
Hoxd9	iHoxd9 fwd	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC-GCTCTTCCGATCTXXXXCGAACACCTCGTCGCCCT

	iHoxd9 rev	CAAGCAGAAGACGGCATACGACCCTCAGCTTGCAGCGAT
Hoxd4	iHoxd4 fwd	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC-GCTCTTCCGATCTXXXXACAGGACAATAAAGCATCCATAGGCG
	iHoxd4 rev	CAAGCAGAAGACGGCATACGATCCAGTGGAATTGGGTGGGAT
Hoxd1	iHoxd1 fwd	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC-GCTCTTCCGATCTXXXXACTATTTACCTCGGGCTCGCT
	iHoxd1 rev	CAAGCAGAAGACGGCATACGACATCCTGGGGACTGGTCAGA
CS39	iCS(39) fwd	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC-GCTCTTCCGATCTXXXXTTCCAAGGAGAAAGGTGTTGGTC
	iCNS(39) rev	CAAGCAGAAGACGGCATACGACAGGGCGTTGGGTCACTCT
IslandII	iIsl-II fwd	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC-GCTCTTCCGATCTXXXXGCATTCATCAAGCTGTGATTAGCA
	iIsl-II rev	CAAGCAGAAGACGGCATACGATCCCATAATATGTAGACTGTAGTGTTGC
lacZ	ilacZ fwd	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGAC-GCTCTTCCGATCTXXXXTAGTGCAACCGAACGCGAC
	ilacZ rev	CAAGCAGAAGACGGCATACGAGGCAAAGACCAGACCGTTCATAC

6 ACKNOWLEDGMENTS

I would like first to thank the NCCR program "*Frontiers in Genetics*" for funding the first year of this thesis as well as giving me the opportunity to work in different laboratories across Switzerland during my rotations. These different experiences have broadened my scientific knowledge and interests and have generated enriching relationships. I am particularly thankful to David Shore for offering me the possibility to apply to this PhD program some years ago.

On a more personal point of view, I would like to thank my whole family for instigating me with curiosity and a strong desire to observe and challenge nature, as well as materially supporting my daily aspirations during several years. Each of you supported and encouraged this work and provided me with a comforting distance out of the research world in time of stresses. My partner, Mathilde, has been of central importance to go through the final legs of this work. I am extremely grateful and lucky to have her by my side. I am also very thankful to the extended P/F crew who were necessary and sufficient to conserve exquisite connections with the wild world. These are precious moments to keep a balance in life. My sole regret is the difficulty to share with you all the delicate, intimate and swift happiness attained when resolving a humble question, before moving towards a new one.

This work would not have been possible to realize without the technical support of the whole mouse team of the Duboule lab. I am thus particularly in debt towards Hanh, Sandra and Bénédicte for the management of the mouse colony and production of transgenic alleles. I was also extremely lucky to profit from the expertise of Jozsef, a steady and invaluable help throughout all these years. I am also extremely thankful to the whole iGe3 and EPFL sequencing platforms as well as the TGF platform for their collaboration and expertise.

I benefited a lot from the international and stimulating environment of the Duboule lab. I am therefore indebted to present and past laboratory members for their support, expertise, curiosity and affection for genetic and experimental biology. For different reasons, I would like to thank particularly Guillaume, Joska, Léonardo, Aurélie, Pierre, Eddie, Lucille, Chase, Hocine, Rita, Célia and Nayuta. I am also extremely thankful to Andrew Seeber for frequent discussions and reading this manuscript at different times.

I am extremely obliged to my mentor Denis Duboule who let me fulfil my thesis in his lab. He guided me through this adventure with care, a great humanity, a variety of stimulations and his passion for science, genetics and knowledge in general.

Finally, I am honored that Jacqueline Deschamps, Alexandre Reymond, Didier Trono and Joachim Lingner have agreed to sit on the jury of my thesis.

7 REFERENCES

Acemel, R.D., Tena, J.J., Irastorza-Azcarate, I., Marlétaz, F., Gómez-Marín, C., de la Calle-Mustienes, E., Bertrand, S., Diaz, S.G., Aldea, D., Aury, J.-M., et al. (2016). A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. Nat. Genet. *48*, 336–341.

Acemel, R.D., Maeso, I., and Gómez-Skarmeta, J.L. (2017). Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. Wiley Interdiscip. Rev. Dev. Biol. *6*, 1–19.

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

Alipour, E., and Marko, J.F. (2012). Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. *40*, 11202–11212.

Amândio, A.R., Necsulea, A., Joye, E., Mascrez, B., and Duboule, D. (2016). Hotair Is Dispensible for Mouse Development. PLoS Genet. *12*, e1006232.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. Dev. Cell *16*, 47–57.

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

Amin, S., Neijts, R., Simmini, S., van Rooijen, C., Tan, S.C., Kester, L., van Oudenaarden, A., Creyghton, M.P., and Deschamps, J. (2016). Cdx and T Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche. Cell Rep. *17*, 3165–3177.

Andrews, O.E., Cha, D.J., Wei, C., and Patton, J.G. (2014). RNAi-mediated gene silencing in zebrafish triggered by convergent transcription. Sci. Rep. *4*, 5222.

Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Development *144*, 3646–3658.

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

Andrey, G., Schöpflin, R., Jerković, I., Heinrich, V., Ibrahim, D.M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M., et al. (2017). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. Genome Res. *27*, 223–233.

Beccari, L., Yakushiji-Kaminatsui, N., Woltering, J.M., Necsulea, A., Lonfat, N., Rodríguez-Carballo, E., Mascrez, B., Yamamoto, S., Kuroiwa, A., and Duboule, D. (2016). A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus. Genes Dev. 1–15.

Beckers, J., and Duboule, D. (1998). Genetic analysis of a conserved sequence in the HoxD complex: regulatory redundancy or limitations of the transgenic approach? Dev. Dyn. *213*, 1–11.

Beckers, J., Gérard, M., and Duboule, D. (1996). Transgenic analysis of a potential Hoxd-11 limb regulatory element present in tetrapods and fish. Dev. Biol. *180*, 543–553.

Bender, W., Akam, M., Karch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B., and Hogness, D.S. (1983). Molecular Genetics of the Bithorax Complex in Drosophila melanogaster. Science *221*, 23–29.

Berlivet, S., Paquette, D., Dumouchel, A., Langlais, D., Dostie, J., and Kmita, M. (2013). Clustering of Tissue-Specific Sub-TADs Accompanies the Regulation of HoxA Genes in Developing Limbs. PLoS Genet. *9*, e1004018.

Bhatlekar, S., Fields, J.Z., and Boman, B.M. (2014). HOX genes and their role in the development of human cancers. J. Mol. Med. (Berl). *92*, 811–823.

Blasco, R.B., Karaca, E., Ambrogio, C., Cheong, T.C., Karayol, E., Minero, V.G., Voena, C., and Chiarle, R. (2014). Simple and Rapid InVivo Generation of Chromosomal Rearrangements using CRISPR/Cas9 Technology. Cell Rep. *9*, 1219–1227.

Boldogköi, Z. (2012). Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. Front. Genet. *3*, 1–17.

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

Brunskill, E.W., Park, J.-S., Chung, E., Chen, F., Magella, B., and Potter, S.S. (2014). Single cell dissection of early kidney development: multilineage priming. Development *141*, 3093–3101.

Calero-Nieto, F.J., Bert, A.G., and Cockerill, P.N. (2010). Transcription-dependent silencing of inducible convergent transgenes in transgenic mice. Epigenetics Chromatin *3*, 3.

Cameron, R.A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J.P., Berney, K., Arenas-Mena, C., Martinez, P., Lucas, S., Richardson, P.M., et al. (2006). Unusual gene order and organization of the sea urchin hox cluster. J. Exp. Zool. B. Mol. Dev. Evol. *306*, 45–58.

Cao, K., Collings, C.K., Marshall, S.A., Morgan, M.A., Rendleman, E.J., Wang, L., Sze, C.C., Sun, T., Bartom, E.T., and Shilatifard, A. (2017). SET1A / COMPASS and shadow enhancers in the regulation of homeotic gene expression. Genes Dev. 1–15.

Carpenter, E.M. (2002). Hox genes and spinal cord development. Dev. Neurosci. *24*, 24–34.

Carpenter, E.M., Goddard, J.M., Davis, A.P., Nguyen, T.P., and Capecchi, M.R. (1997). Targeted disruption of Hoxd-10 affects mouse hindlimb development. Development *124*, 4505–4514.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat. Rev. Genet. *14*, 100–112.

Cheng, Y., Miura, R.M., and Tian, B. (2006). Prediction of mRNA polyadenylation sites by support vector machine. Bioinformatics *22*, 2320–2325.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory Interactions between RNA and Polycomb Repressive Complex 2. Mol. Cell *55*, 171–185.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Darbellay, F., and Duboule, D. (2016). Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at Hox Loci. In Current Topics in Developmental Biology, (Elsevier Inc.), pp. 299–314.

David, F.P.A., Delafontaine, J., Carat, S., Ross, F.J., Lefebvre, G., Jarosz, Y., Sinclair, L., Noordermeer, D., Rougemont, J., and Leleu, M. (2014). HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. PLoS One *9*, e85879.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol. *20*, 1250–1257.

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

Davis, A.P., Witte, D.P., Hsieh-Li, H.M., Potter, S.S., and Capecchi, M.R. (1995). Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11. Nature *375*, 791–795.

Delpretti, S., Zakany, J., and Duboule, D. (2012). A function for all posterior Hoxd genes during digit development? Dev. Dyn. *241*, 792–802.

Delpretti, S., Montavon, T., Leleu, M., Joye, E., Tzika, A., Milinkovitch, M., and Duboule, D. (2013). Multiple Enhancers Regulate Hoxd Genes and the Hotdog LncRNA during Cecum Budding. Cell Rep. *5*, 137–150.

Denans, N., Iimura, T., and Pourquié, O. (2015). Hox genes control vertebrate body elongation by collinear Wnt repression. Elife *4*, 1–33.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. *30*, 1357–1382.

Deschamps, J., and Duboule, D. (2017). Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. Genes Dev. *31*, 1406–1416.

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

Deschamps, J., and Wijgerde, M. (1993). Two phases in the establishment of HOX expression domains. Dev. Biol. *156*, 473–480.

Di-Poï, N., Zákány, J., and Duboule, D. (2007). Distinct roles and regulations for HoxD genes in metanephric kidney development. PLoS Genet. *3*, e232.

Dickel, D.E., Ypsilanti, A.R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B.J., Khin, Y.S., Fukuda-Yuzawa, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Ultraconserved Enhancers Are Required for Normal Development. Cell *172*, 491–499.e15.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dollé, P., Izpisúa-Belmonte, J.C., Falkenstein, H., Renucci, A., and Duboule, D. (1989). Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. Nature *342*, 767–772.

Dollé, P., Izpisua-Belmonte, J.C., Brown, J.M., Tickle, C., and Duboule, D. (1991). HOX-4 genes and the morphogenesis of mammalian genitalia. Genes Dev. *5*, 1767–1776.

Dollé, P., Dierich, A., LeMeur, M., Schimmang, T., Schuhbaur, B., Chambon, P., and Duboule, D. (1993). Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neotenic limbs. Cell *75*, 431–441.

Dollé, P., Fraulob, V., and Duboule, D. (1994). Developmental expression of the mouse Evx-2 gene: relationship with the evolution of the HOM/Hox complex. Dev. Suppl. *1994*, 143–153.

Drake, K.A., Adam, M., Mahoney, R., and Potter, S.S. (2018). Disruption of Hox9,10,11 function results in cellular level lineage infidelity in the kidney. Sci. Rep. *8*, 6306.

Duboule, D. (1991). Patterning in the vertebrate limb. Curr. Opin. Genet. Dev. *1*, 211–216.

Duboule, D. (1992). The vertebrate limb: a model system to study the Hox/HOM gene network during development and evolution. Bioessays *14*, 375–384.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev. Suppl. *42*, 135–142.

Duboule, D. (2007). The rise and fall of Hox gene clusters. Development *134*, 2549–2560.

Duboule, D., and Dollé, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. EMBO J. *8*, 1497–1505.

Duboule, D., and Morata, G. (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. Trends Genet. *10*, 358–364.

Dubrulle, J., and Pourquié, O. (2004). Coupling segmentation to axis formation. Development *131*, 5783–5793.

Fabre, P.J., Benke, A., Joye, E., Nguyen Huynh, T.H., Manley, S., and Duboule, D. (2015). Nanoscale spatial organization of the HoxD gene cluster in distinct transcriptional states. Proc. Natl. Acad. Sci. U. S. A. *112*, 13964–13969.

Foronda, D., De Navas, L.F., Garaulet, D.L., and Sánchez-Herrero, E. (2009). Function and specificity of Hox genes. Int. J. Dev. Biol. *53*, 1409–1419.

Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265–269.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049.

Garcia-Fernàndez, J. (2005). Hox, ParaHox, ProtoHox: facts and guesses. Heredity (Edinb). *94*, 145–152.

Gaunt, S.J. (1988). 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 *103*, 135–144.

Gaunt, S.J. (2015). The significance of Hox gene collinearity. Int. J. Dev. Biol. *59*, 159–170.

Gaunt, S.J., and Gaunt, A.L. (2016). Possible rules for the ancestral origin of Hox gene collinearity. J. Theor. Biol. *410*, 1–8.

Gaunt, S.J., and Strachan, L. (1994). Forward spreading in the establishment of a vertebrate Hox expression boundary: the expression domain separates into anterior and posterior zones, and the spread occurs across implanted glass barriers. Dev. Dyn. *199*, 229–240.

Gaunt, S.J., George, M., and Paul, Y.-L. (2013). Direct activation of a mouse Hoxd11 axial expression enhancer by Gdf11/Smad signalling. Dev. Biol. *383*, 52–60.

Gehring, W.J., Kloter, U., and Suga, H. (2009). Evolution of the Hox gene complex from an evolutionary ground state. Curr. Top. Dev. Biol. *88*, 35–61.

Gérard, M., Duboule, D., and Zákány, J. (1993). Structure and activity of regulatory elements involved in the activation of the Hoxd-11 gene during late gastrulation. EMBO J. *12*, 3539–3550.

Gérard, M., Chen, J.Y., Gronemeyer, H., Chambon, P., Duboule, D., and Zákány, J. (1996). In vivo targeted mutagenesis of a regulatory element required for positioning the Hoxd-11 and Hoxd-10 expression boundaries. Genes Dev. *10*, 2326–2334.

Ghirlando, R., and Felsenfeld, G. (2016). CTCF: Making the right connections. Genes Dev. *30*, 881–891.

Gimond, C., Baudoin, C., van der Neut, R., Kramer, D., Calafat, J., and Sonnenberg, A. (1998). Cre-loxP-mediated inactivation of the alpha6A integrin splice variant in vivo: evidence for a specific functional role of alpha6A in lymphocyte migration but not in heart development. J. Cell Biol. *143*, 253–266.

Goloborodko, A., Marko, J.F., and Mirny, L.A. (2016). Chromosome Compaction by Active Loop Extrusion. Biophys. J. *110*, 2162–2168.

Gonzalez-Sandoval, A., and Gasser, S.M. (2016). On TADs and LADs: Spatial Control Over Gene Expression. Trends Genet. *32*, 485–495.

Graham, A., Papalopulu, N., and Krumlauf, R. (1989). The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell *57*, 367–378.

Guerreiro, I., Gitto, S., Novoa, A., Codourey, J., Nguyen Huynh, T.H., Gonzalez, F., Milinkovitch, M.C., Mallo, M., and Duboule, D. (2016). Reorganisation of Hoxd regulatory landscapes during the evolution of a snake-like body plan. Elife *5*.

Gullerova, M., and Proudfoot, N.J. (2012). Convergent transcription induces transcriptional gene silencing in fission yeast and mammalian cells. Nat. Struct. Mol. Biol. *19*, 1193–1201.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

Haarhuis, J.H.I., Van Der Weide, R.H., Blomen, V.A., Brummelkamp, T.R., De Wit, E., Rowland Correspondence, B.D., Nl, E.D.W.D.W., and Nl, R.R. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. 693–707.

Hao, N., Palmer, A.C., Dodd, I.B., and Shearwin, K.E. (2017). Directing traffic on DNA—How transcription factors relieve or induce transcriptional interference. Transcription *8*, 120–125.

Harding, K., Wedeen, C., McGinnis, W., and Levine, M. (1985). Spatially regulated expression of homeotic genes in Drosophila. Science *229*, 1236–1242.

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat. Rev. Mol. Cell Biol. *18*, 263–273.

Hashimoto, M., and Takemoto, T. (2015). Electroporation enables the efficient mRNA delivery into the mouse zygotes and facilitates CRISPR/Cas<inf>9</inf>-based genome editing. Sci. Rep. *5*.

Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors and the making of the spinal cord. Development *142*, 2864–2875.

Hérault, Y., Beckers, J., Kondo, T., Fraudeau, N., and Duboule, D. (1998). Genetic analysis of a Hoxd-12 regulatory element reveals global versus local modes of controls in the HoxD complex. Development *125*, 1669–1677.

Herzog, V.A., Lempradl, A., Trupke, J., Okulski, H., Altmutter, C., Ruge, F., Boidol, B., Kubicek, S., Schmauss, G., Aumayr, K., et al. (2014). A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. Nat. Genet. *46*, 973–981.

Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. Mol. Cell *48*, 365–374.

van der Hoeven, F., Zákány, J., and Duboule, D. (1996). Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls. Cell *85*, 1025–1035.

Holland, P.W. (1999). Gene duplication: past, present and future. Semin. Cell Dev. Biol. *10*, 541–547.

Holland, P.W., Garcia-Fernàndez, J., Williams, N.A., and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. Dev. Suppl. 125–133.

Hueber, S.D., and Lohmann, I. (2008). Shaping segments: Hox gene function in the genomic age. BioEssays *30*, 965–979.

Hutlet, B., Theys, N., Coste, C., Ahn, M.-T., Doshishti-Agolli, K., Lizen, B., and Gofflot, F. (2016). Systematic expression analysis of Hox genes at adulthood reveals novel patterns in the central nervous system. Brain Struct. Funct. *221*, 1223–1243.

Ikuta, T., Yoshida, N., Satoh, N., and Saiga, H. (2004). Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development. Proc. Natl. Acad. Sci. U. S. A. *101*, 15118–15123.

Izpisúa-Belmonte, J.C., Falkenstein, H., Dollé, P., Renucci, A., and Duboule, D. (1991). Murine genes related to the Drosophila AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. EMBO J. *10*, 2279–2289.

Kaneko, S., Son, J., Bonasio, R., Shen, S.S., and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. Genes Dev. *28*, 1983–1988.

Kaufman, T.C., Lewis, R., and Wakimoto, B. (1980). Cytogenetic Analysis of Chromosome 3 in DROSOPHILA MELANOGASTER: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. Genetics *94*, 115–133.

Kessel, M., and Gruss, P. (1990). Murine developmental control genes. Science *249*, 374–379.

Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. Science (80-.). *301*, 331–333.

Kmita, M., Kondo, T., and Duboule, D. (2000a). Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. Nat. Genet. *26*, 451–454.

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

Kondo, T., Dollé, P., Zákány, J., and Duboule, D. (1996). Function of posterior HoxD genes in the morphogenesis of the anal sphincter. Development *122*, 2651–2659.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191–201.

Lebert-Ghali, C.-É., Fournier, M., Kettyle, L., Thompson, A., Sauvageau, G., and Bijl, J.J. (2015). Hoxa cluster genes determine the proliferative activity of adult mouse hematopoietic stem and progenitor cells. Blood *127*, blood-2015-02-626390.

Lemons, D., and McGinnis, W. (2006). Genomic evolution of Hox gene clusters. Science *313*, 1918–1922.

Lesnik, E.A., Sampath, R., Levene, H.B., Henderson, T.J., McNeil, J.A., and Ecker, D.J. (2001). Prediction of rho-independent transcriptional terminators in Escherichia coli. Nucleic Acids Res. *29*, 3583–3594.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Lonfat, N., and Duboule, D. (2015). Structure, function and evolution of topologically associating domains (TADs) at HOX loci. FEBS Lett. *589*, 2869–2876.

Lonfat, N., Montavon, T., Darbellay, F., Gitto, S., and Duboule, D. (2014). Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. Science (80-.). *346*, 1004–1006.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Magella, B., Mahoney, R., Adam, M., and Potter, S.S. (2018). Reduced Abd-B Hox function during kidney development results in lineage infidelity. Dev. Biol. 0–1.

Mainguy, G., Koster, J., Woltering, J., Jansen, H., and Durtson, A. (2007). Extensive polycistronism and antisense transcription in the mammalian Hox cluster. PLoS One *2*, 1–7.

Mallo, M. (2017). Reassessing the Role of Hox Genes during Vertebrate Development and Evolution. Trends Genet. *34*, 209–217.

Mallo, M., Wellik, D.M., and Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate body plan. Dev. Biol. *344*, 7–15.

Mashiko, D., Fujihara, Y., Satouh, Y., Miyata, H., Isotani, A., and Ikawa, M. (2013). Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. Sci. Rep. *3*, 3355.

McGinnis, W., Hart, C.P., Gehring, W.J., and Ruddle, F.H. (1984). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila. Cell *38*, 675–680.

Merabet, S., Hudry, B., Saadaoui, M., and Graba, Y. (2009). Classification of sequence signatures: a guide to Hox protein function. Bioessays *31*, 500–511.

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. Nature *457*, 413–420.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. *42*, W401–W407.

Montavon, T. (2008). Mechanisms of Hoxd Genes Collinearity in Developing Digits and External Genitalia. Université de Genève.

Montavon, T., and Duboule, D. (2012). Landscapes and archipelagos: spatial organization of gene regulation in vertebrates. Trends Cell Biol. *22*, 347–354.

Montavon, T., Le Garrec, J.-F., Kerszberg, M., and Duboule, D. (2008). Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. Genes Dev. *22*, 346–359.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. Cell *147*, 1132–1145.

Morgan, R. (2006). Hox genes: a continuation of embryonic patterning? Trends Genet. *22*, 67–69.

Mugford, J.W., Sipilä, P., Kobayashi, A., Behringer, R.R., and McMahon, A.P. (2008). Hoxd11 specifies a program of metanephric kidney development within the intermediate mesoderm of the mouse embryo. Dev. Biol. *319*, 396–405.

Müller, U., Wang, D., Denda, S., Meneses, J.J., Pedersen, R.A., and Reichardt, L.F. (1997). Integrin α8β1 is critically important for epithelial-mesenchymal interactions during kidney morphogenesis. Cell *88*, 603–613.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017–1021.

Narendra, V., Dekker, J., Mazzoni, E.O., and Reinberg, D. (2016). CTCF-mediated topological boundaries during development foster appropriate gene regulation. Gene Dev 2657–2662.

Nasmyth, K. (2001). Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis. Annu. Rev. Genet. *35*, 673–745.

Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A., and Gautheret, D. (2011). ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol. *8*, 11–13.

Neijts, R., and Deschamps, J. (2017). At the base of colinear Hox gene expression: cis-features and trans-factors orchestrating the initial phase of Hox cluster activation. Dev. Biol. *428*, 293–299.

Neijts, R., Amin, S., Van Rooijen, C., Tan, S., Creyghton, M.P., De Laat, W., and Deschamps, J. (2016). Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. Genes Dev. *30*, 1937–1942.

Neijts, R., Amin, S., van Rooijen, C., and Deschamps, J. (2017). Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk segment in the Hox cluster topology. Dev. Biol. *422*, 146–154.

Nelson, C.E., Morgan, B.A., Burke, A.C., Laufer, E., DiMambro, E., Murtaugh, L.C., Gonzales, E., Tessarollo, L., Parada, L.F., and Tabin, C. (1996). Analysis of Hox gene expression in the chick limb bud. Development *122*, 1449–1466.

Noordermeer, D., and Duboule, D. (2013). Chromatin architectures and Hox gene collinearity. In Current Topics in Developmental Biology, pp. 113–148.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011). The dynamic architecture of Hox gene clusters. Science *334*, 222–225.

Noordermeer, D., Leleu, M., Schorderet, P., Joye, E., Chabaud, F., and Duboule, D. (2014). Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. Elife *3*, e02557.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22.

Ohno, S. (1970). Evolution by Gene Duplication (Berlin, Heidelberg: Springer Berlin Heidelberg).

Olivares-Chauvet, P., Mukamel, Z., Lifshitz, A., Schwartzman, O., Elkayam, N.O., Lubling, Y., Deikus, G., Sebra, R.P., and Tanay, A. (2016). Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. Nature *540*, 296–300.

Osato, N., Suzuki, Y., Ikeo, K., and Gojobori, T. (2007). Transcriptional interferences in cis natural antisense transcripts of humans and mice. Genetics *176*, 1299–1306.

Osterwalder, M., Barozzi, I., Tissiéres, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243.

Palmer, A.C., Barry Egan, J., and Shearwin, K.E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. Transcription *2*, 9–14.

Pannunzio, N.R., and Lieber, M.R. (2016). RNA Polymerase Collision versus DNA Structural Distortion: Twists and Turns Can Cause Break Failure. Mol Cell *62*, 327–334.

Patterson, L.T., and Potter, S.S. (2004). Atlas of Hox gene expression in the developing kidney. Dev. Dyn. *229*, 771–779.

Patterson, L.T., Pembaur, M., and Potter, S.S. (2001). Hoxa11 and Hoxd11 regulate branching morphogenesis of the ureteric bud in the developing kidney. Development *128*, 2153–2161.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. *6*, 893–904.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell *153*, 1281–1295.

Prescott, E.M., and Proudfoot, N.J. (2002). Transcriptional collision between convergent genes in budding yeast. Proc. Natl. Acad. Sci. U. S. A. *99*, 8796–8801.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, K., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature *453*, 1064–1071.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305–320.e24.

Renucci, A., Zappavigna, V., Zàkàny, J., Izpisúa-Belmonte, J.C., Bürki, K., and Duboule, D. (1992). Comparison of mouse and human HOX-4 complexes defines conserved sequences involved in the regulation of Hox-4.4. EMBO J. *11*, 1459–1468.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes Dev. *23*, 1247–1269.

Rodríguez-Carballo, E., Lopez-Delisle, L., Zhan, Y., Fabre, P.J., Beccari, L., El-Idrissi, I., Huynh, T.H.N., Ozadam, H., Dekker, J., and Duboule, D. (2017). The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. Genes Dev. 1–44.

Romagnani, P., Lasagni, L., and Remuzzi, G. (2013). Renal progenitors: an evolutionary conserved strategy for kidney regeneration. Nat. Rev. Nephrol. *9*, 137–146.

Rux, D.R., and Wellik, D.M. (2017). Hox genes in the adult skeleton: Novel functions beyond embryonic development. Dev. Dyn. *246*, 310–317.

Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. *112*, E6456–E6465.

Sánchez-Herrero, E., Vernós, I., Marco, R., and Morata, G. (1985). Genetic organization of Drosophila bithorax complex. Nature *313*, 108–113.

Schep, R., Necsulea, A., Rodríguez-Carballo, E., Guerreiro, I., Andrey, G., Nguyen Huynh, T.H., Marcet, V., Zákány, J., Duboule, D., and Beccari, L. (2016). Control of Hoxd gene transcription in the mammary bud by hijacking a preexisting regulatory landscape. Proc. Natl. Acad. Sci. U. S. A. *113*, E7720–E7729.

Schiemann, S.M., Martín-Durán, J.M., Børve, A., Vellutini, B.C., Passamaneck, Y.J., and Hejnol, A. (2017). Clustered brachiopod Hox genes are not expressed collinearly and are associated with lophotrochozoan novelties. Proc. Natl. Acad. Sci. U. S. A. *114*, E1913–E1922.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods *12*, 963–965.

Schorderet, P., Lonfat, N., Darbellay, F., Tschopp, P., Gitto, S., Soshnikova, N., and Duboule, D. (2013). A genetic approach to the recruitment of PRC2 at the HoxD locus. PLoS Genet. *9*, e1003951.

Schwartzman, O., Mukamel, Z., Oded-Elkayam, N., Olivares-Chauvet, P., Lubling, Y., Landan, G., Izraeli, S., and Tanay, A. (2016). UMI-4C for quantitative and targeted chromosomal contact profiling. Nat. Methods *13*, 685–691.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51–56.

Scott, M.P. (1992). Vertebrate homeobox gene nomenclature. Cell *71*, 551–553.

Scott, M.P. (1993). A rational nomenclature for vertebrate homeobox (HOX) genes. Nucleic Acids Res. *21*, 1687–1688.

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

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

Shah, M.M., Sampogna, R. V, Sakurai, H., Bush, K.T., and Nigam, S.K. (2004). Branching morphogenesis and kidney disease. Development *131*, 1449–1462.

Sheth, R., Barozzi, I., Langlais, D., Osterwalder, M., Nemec, S., Carlson, H.L., Stadler, H.S., Visel, A., Drouin, J., and Kmita, M. (2016). Distal Limb Patterning Requires Modulation of cis-Regulatory Activities by HOX13. Cell Rep. *17*, 2913–2926.

Simons, C., Pheasant, M., Makunin, I. V, and Mattick, J.S. (2006). Transposon-free regions in mammalian genomes. Genome Res. *16*, 164–172.

Simons, C., Makunin, I. V, Pheasant, M., and Mattick, J.S. (2007). Maintenance of transposon-free regions throughout vertebrate evolution. BMC Genomics *8*, 470.

Somesh, B.P., Reid, J., Liu, W.F., Søgaard, T.M.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. Cell *121*, 913–923.

Soshnikova, N., and Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. Science *324*, 1320–1323.

Soshnikova, N., Montavon, T., Leleu, M., Galjart, N., and Duboule, D. (2010). Functional analysis of CTCF during mammalian limb development. Dev. Cell *19*, 819–830.

Soshnikova, N., Dewaele, R., Janvier, P., Krumlauf, R., and Duboule, D. (2013). Duplications of hox gene clusters and the emergence of vertebrates. Dev. Biol. *378*, 194–199.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell *113*, 405–417.

Spitz, F., Herkenne, C., Morris, M. a, and Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. Nat. Genet. *37*, 889–893.

Steffen, P. a, and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat. Rev. Mol. Cell Biol. *15*, 340–356.

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS One *10*, e0124633.

Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. Genome Res. *24*, 390–400.

Takasato, M., and Little, M.H. (2015). The origin of the mammalian kidney: implications for recreating the kidney in vitro. Development *142*, 1937–1947.

Taniguchi, Y. (2014). Hox transcription factors: modulators of cell-cell and cell-extracellular matrix adhesion. Biomed Res. Int. *2014*, 591374.

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

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

Tschopp, P., and Duboule, D. (2011). A regulatory "landscape effect" over the HoxD cluster. Dev. Biol. *351*, 288–296.

Tschopp, P., Tarchini, B., Spitz, F., Zakany, J., and Duboule, D. (2009). Uncoupling time and space in the collinear regulation of Hox genes. PLoS Genet. *5*, e1000398.

Tschopp, P., Christen, A.J., and Duboule, D. (2012). Bimodal control of Hoxd gene transcription in the spinal cord defines two regulatory subclusters. Development *139*, 929–939.

Uslu, V.V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N.A., Marioni, J.C., and Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. Nat. Genet. *46*, 753–758.

Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian *Hox* genes with other H3K27me3 targets within an active nuclear domain. Proc. Natl. Acad. Sci. *112*, 4672–4677.

Wagner, G.P., Amemiya, C., and Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. Proc. Natl. Acad. Sci. U. S. A. *100*, 14603–14606.

Wellik, D.M. (2007). Hox patterning of the vertebrate axial skeleton. Dev. Dyn. *236*, 2454–2463.

Wellik, D.M., and Capecchi, M.R. (2003). Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. Science *301*, 363–367.

Wellik, D.M., Hawkes, P.J., and Capecchi, M.R. (2002). Hox11 paralogous genes are essential for metanephric kidney induction. Genes Dev. *16*, 1423–1432.

White, E., Schlackow, M., Kamieniarz-Gdula, K., Proudfoot, N.J., and Gullerova, M. (2014). Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. Nat. Struct. Mol. Biol. *21*, 552–559.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev. *26*, 11–24.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell *60*, 676–684.

Woltering, J.M., Vonk, F.J., Müller, H., Bardine, N., Tuduce, I.L., de Bakker, M.A.G., Knöchel, W., Sirbu, I.O., Durston, A.J., and Richardson, M.K. (2009). Axial patterning in snakes and caecilians: Evidence for an alternative interpretation of the Hox code. Dev. Biol. *332*, 82–89.

Woltering, J.M., Noordermeer, D., Leleu, M., and Duboule, D. (2014). Conservation and Divergence of Regulatory Strategies at Hox Loci and the Origin of Tetrapod Digits. PLoS Biol. *12*, e1001773.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. *36*, e201798004.

Yallowitz, A.R., Hrycaj, S.M., Short, K.M., Smyth, I.M., and Wellik, D.M. (2011). Hox10 genes function in kidney development in the differentiation and integration of the cortical stroma. PLoS One *6*, e23410.

Young, T., Rowland, J.E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.-N., et al. (2009). Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. Dev. Cell *17*, 516–526.

Zacchetti, G., Duboule, D., and Zakany, J. (2007). Hox gene function in vertebrate gut morphogenesis: the case of the caecum. Development *134*, 3967–3973.

Zakany, J., and Duboule, D. (2007). The role of Hox genes during vertebrate limb development. Curr. Opin. Genet. Dev. *17*, 359–366.

Zakany, J., Darbellay, F., Mascrez, B., Necsulea, A., and Duboule, D. (2017). Control of growth and gut maturation by HoxD genes and the associated lncRNA Haglr. Proc. Natl. Acad. Sci. U. S. A. *114*, 9290–9299.

Zákány, J., Gérard, M., Favier, B., and Duboule, D. (1997). Deletion of a HoxD enhancer induces transcriptional heterochrony leading to transposition of the sacrum. EMBO J. *16*, 4393–4402.

Zákány, J., Kmita, M., Alarcon, P., de la Pompa, J.L., and Duboule, D. (2001). Localized and transient transcription of Hox genes suggests a link between patterning and the segmentation clock. Cell *106*, 207–217.

Zákány, J., Kmita, M., and Duboule, D. (2004). A dual role for Hox genes in limb anterior-posterior asymmetry. Science *304*, 1669–1672.

Ziebarth, J.D., Bhattacharya, A., and Cui, Y. (2013). CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. Nucleic Acids Res. *41*, D188-94.

8 CV

Fabrice Darbellay
fabrice.darbellay@gmail.com
19, chemin du Jerlon · 1223 Cologny · + 41 (0) 76 616 9467

Present position

PhD student 10.2012 – present
School of Life Sciences, Federal Institute of Technology Lausanne, CH
Advisor: Prof. Denis Duboule
Functional organization of Hox genes clusters.

Education

University of Geneva 10.2011 – 10.2012
NCCR - Frontiers in Genetics, PhD rotation program Lausanne/Basel, CH

University of Geneva
Master degree in Biology 09.2009 – 08.2011
Bachelor degree in Biology 10.2006 – 07.2009

Research experience

PhD rotation program *NCCR - Frontiers in Genetics*
3rd rotation, advisor: Prof. Susan M. Gasser 05.2012 – 08.2012
Friedrich Miescher Institute for Biomedical Research Basel, CH
Development of a high-throughput synthetic lethal screen in C. elegans.

2nd rotation, advisor: Prof. Denis Duboule 02.2012 – 04.2012
School of Life Sciences, Federal Institute of Technology Lausanne, CH
PRC2 recruitment at the HoxD locus in murine iPS cells.

1st rotation, advisor: Prof. Winship Herr 10.2011 – 01.2012
Center for Integrative Genomics, University of Lausanne Lausanne, CH
Investigation of HCF-1 subunits interactions in HEK293 cells.

MSc internship, advisor: Prof. David Shore 09.2009 – 09.2011
University of Geneva Genève, CH
Investigation of the CURI complex at the crossroad of ribosome biogenesis in S. cerevisiae.

Skills and techniques

- Expertise in molecular biology (RNA/DNA manipulations).
- Genetic manipulations in budding yeast and murine iPS/ES cells.
- Construction, establishment and maintenance of mouse mutant alleles (CRISPR/Cas9 targeting).
- Microdissection of mouse, rat and chicken embryonic tissues at different developmental stages.
- Characterization of genes expressions patterns in mouse embryo (WISH) and developmental enhancers activity by transgenic X-gal reporter.
- Production and analysis of genomic datasets:
  Chromatin immunoprecipitation and accessibility assays (ChIP, ChIPmentation, ATAC). Chromosome conformation capture assays (3C-derived techniques: 4Cseq, 4Cseq-UMI, HiC). Transcriptome analysis (stranded RNAseq)

Peer-reviewed publications

Darbellay, F., Bochaton, C., Zakany, J., Deslile, L., Mascrez, B., Tschopp, P., and Duboule, D. Arrangement of murine *Hox* genes in dense clusters requires a strict genetic organization. *In preparation.*

Darbellay, F., and Necsulea, A. Comparative transcriptomics analyses across species, organs and developmental stages reveal rapid evolution of lncRNA expression patterns. *In preparation.*

Zakany, J., **Darbellay, F.**, Mascrez, B., Necsulea, A., and Duboule, D. (2017) Control of growth and gut maturation by *HoxD* genes and the associated lncRNA *Haglr*. Proceedings of the National Academy of Sciences U. S. A. 114, 9290-9299.

Lonfat, N., Montavon, T., **Darbellay, F.**, Gitto, S., and Duboule, D. (2014) Convergent evolution of complex regulatory landscapes and pleiotropy at *Hox* loci. Science 346, 1004–1006.

Schorderet, P., Lonfat, N., **Darbellay, F.**, Tschopp, P., Gitto, S., Soshnikova, N., and Duboule, D. (2013) A Genetic Approach to the Recruitment of PRC2 at the *HoxD* Locus. PLoS Genetics 9, e1003951.

Review articles and book chapters

Darbellay, F., and Duboule, D. (2016) Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at *Hox* Loci. Current Topics in Developmental Biology 116, 299–314.

Leadership experiences

School of Life Sciences, Federal Institute of Technology, laboratory of Prof. Denis Duboule

Experience in teaching general biology courses to undergraduate students (TA). Practical training of undergraduate internship students in the laboratory. Direct advising of a MSc student including extensive training in the field of functional genomics.

Presentation and attendance to conferences

Louis-Jeantet Satellite Symposium Workshop, October 2016, Geneva (CH), selected short talk The cost of proximity in the *HoxD* cluster.

12th EMBL Transcription and chromatin conference, August 2016, Heidelberg (D), poster presentation *Hoxd* genes regulation during metanephric kidney development.

4th FP7 IDEAL-Ageing annual meeting, February 2015, Bologna (I), selected short talk Maintaining *Hoxd* genes silent through development and adulthood.

11th EMBL Transcription and chromatin conference, August 2014, Heidelberg (D), poster presentation TAD-sequestering of the *Hoxd13* gene prevents its dominant negative deleterious effects during development.

References

Prof. Denis Duboule - School of Life Sciences
Federal Institute of Technology Lausanne (EPFL), Switzerland - denis.duboule@epfl.ch

Prof. David Shore - Department of Molecular Biology
University of Geneva, Switzerland - david.shore@unige.ch

Prof. Susan M. Gasser - Friedrich Miescher Institute for Biomedical Research
Basel, Switzerland - susan.gasser@fmi.ch

Dr. Guillaume Andrey - Max Planck Institute for Molecular Genetics
Berlin, Germany - andrey@molgen.mpg.de

9 ANNEXE

CHAPTER SIXTEEN

Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at *Hox* Loci

Fabrice Darbellay*, Denis Duboule*,†,1
*School of Life Sciences, Federal Institute of Technology, Lausanne, Switzerland
†Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland
[1]Corresponding author: e-mail addresses: denis.duboule@unige.ch; denis.duboule@epfl.ch

Contents

Abstract

Hox gene clusters of jaw vertebrates display a tight genomic organization, which has no equivalent in any other bilateria genomes sequenced thus far. It was previously argued that such a topological consolidation toward a condensed, metagenic structure was due to the accumulation of long-range regulations flanking Hox loci, a phenomenon made possible by the successive genome duplications that occurred at the root of the vertebrate lineage, similar to gene neofunctionalization but applied to a coordinated multigenic system. Here, we propose that the emergence of such large vertebrate regulatory landscapes containing a range of global enhancers was greatly facilitated by the presence of topologically associating domains (TADs). These chromatin domains, mostly constitutive, may have been used as genomic niches where novel regulations could evolve due to both the preexistence of a structural backbone poised to integrate novel regulatory inputs, and a highly adaptive transcriptional readout. We propose a scenario for the coevolution of such TADs and the emergence of pleiotropy at ancestral vertebrate Hox loci.

Current Topics in Developmental Biology, Volume 116
ISSN 0070-2153
http://dx.doi.org/10.1016/bs.ctdb.2015.11.022

Ever since large *Hox* gene clusters were reported to exist in both protostomes (Bender et al., 1983; Harding, Wedeen, McGinnis, & Levine, 1985; Sanchez-Herrero, Vernos, Marco, & Morata, 1985) and deuterostomes (Boncinelli et al., 1988; Duboule, Baron, Mahl, & Galliot, 1986; Hart, Fainsod, & Ruddle, 1987), thus indicating unexpected commonalities in the developmental strategies of all bilateria (Akam, 1989; Duboule & Dolle, 1989; Graham, Papalopulu, & Krumlauf, 1989), this genetic system, originally described in *Drosophila* (Lewis, 1978) has revealed its remarkable epistemic value. However, the sequencing of genomes belonging to many different groups of animals, such as e.g. cephalopods (Albertin et al., 2015), nematodes (Aboobaker & Blaxter, 2003), platyhelminthes (Pierce et al., 2005), and urochordata (Ikuta, Yoshida, Satoh, & Saiga, 2004; Seo et al., 2004), has shown that *Hox* genes are not always found in a clustered organization, indicating that this particular in-*cis* topology is indeed not necessary for the development of many groups of animals. While some genomes carry compact or disorganized clusters, others have split this series of contiguous genes at one or in multiple positions (see Fig. 1). Nevertheless, at least one

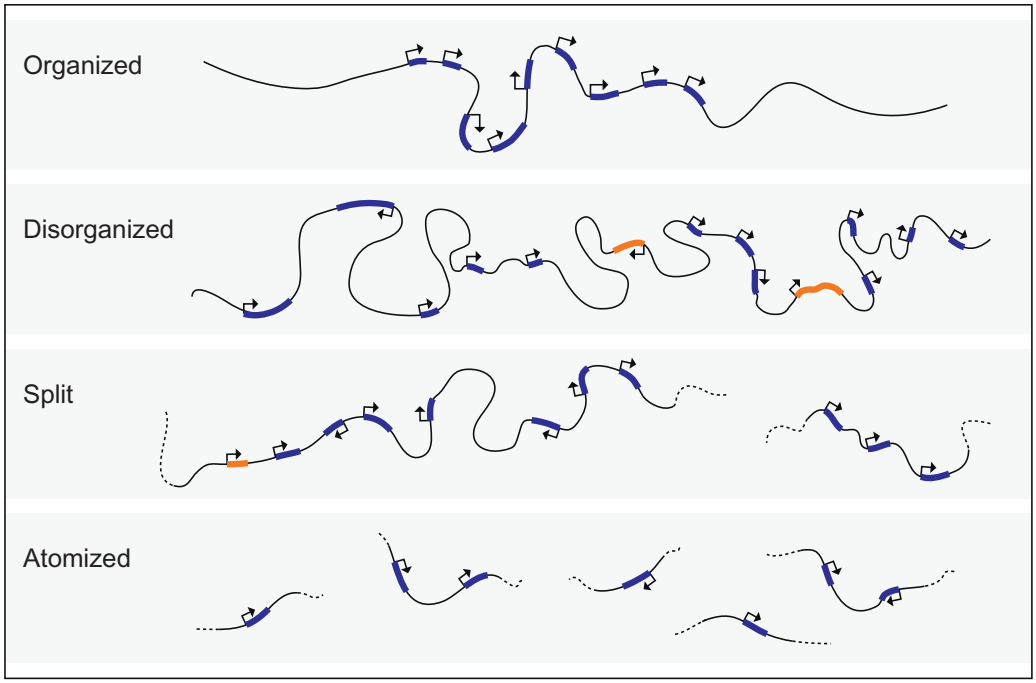

Figure 1 The genomic organization of *Hox* genes varies in different bilaterian animals. *Hox* genes can be found either in tightly organized gene clusters, ca. 100 kb in length, containing few repetitive elements and no foreign genes (organized) or in much larger linkage group, ca. 1 mb in size, including repeats and sometimes foreign genes (disorganized). Such genomic clusters can also be split in various subclusters (split) or even in single unit as in atomized *Hox* (non)clusters (atomized). *Adapted from Duboule (2007).*

entire *Hox* gene cluster is present in all metazoans developing their anterior to posterior axis by following a temporal sequence in the production of segments, such as short germ band insects or vertebrates, suggesting that this clustered organization is closely associated with these particular developmental strategies ("the *Hox* clock"; see Duboule, 1994; Kmita & Duboule, 2003).

1. THE RAISE OF *HOX* METAGENES

Of note, among all animal genomes sequenced thus far and containing one or several full *Hox* gene cluster(s), the level of genomic organization and compaction can vary substantially, from rather large gene complexes, close to one megabase in size (e.g., Mehta et al., 2013), to much shorter and compact structures. Surprisingly, jaw vertebrate (gnathostomes) display the highest level of compaction in their *Hox* clusters, with ca. 10 *Hox* genes within 100–120 kilobases (kb) of DNA, very few repetitive elements intermingled, no foreign genes interspersed, and all transcription units encoded by the same DNA strand (see Duboule, 2007). Even in squamata, where *Hox* clusters are larger than those found in fishes, birds, or in mammals (Di-Poï, Montoya-Burgos, & Duboule, 2009; Di-Poi et al., 2010; Vonk et al., 2013), their general aspect is clearly related to that of other jaw vertebrate clusters and their sizes merely 10–20% larger.

As of today, not a single non-jaw vertebrate genome has shown a comparable level of compaction. Conversely, not a single vertebrate genome has revealed a *Hox* cluster resembling the large structures found, for example, in sea urchins or in *Anopheles* (Cameron et al., 2006; Holt et al., 2002) (Fig. 2). Interestingly, the cephalochordate amphioxus *Hox* cluster, while being rather well organized still expands over half a megabase of DNA (see Minguillon et al., 2005). In jaw vertebrates, since neighboring *Hox* genes often share global regulations (see below), can be included into the same transcriptional unit and display redundant functions, this particular clustered organization was referred to as a "metagene" (Duboule, 1994).

2. A GENOMIC SYNAPOMORPHY

In this particular context, the case of vertebrate agnatha (jawless fishes) has remained difficult to evaluate. Jawless vertebrates indeed contain *Hox* complements whose phylogenetic relations with gnathostomes are difficult to establish. In the Japanese lamprey, for instance, the number of *Hox*

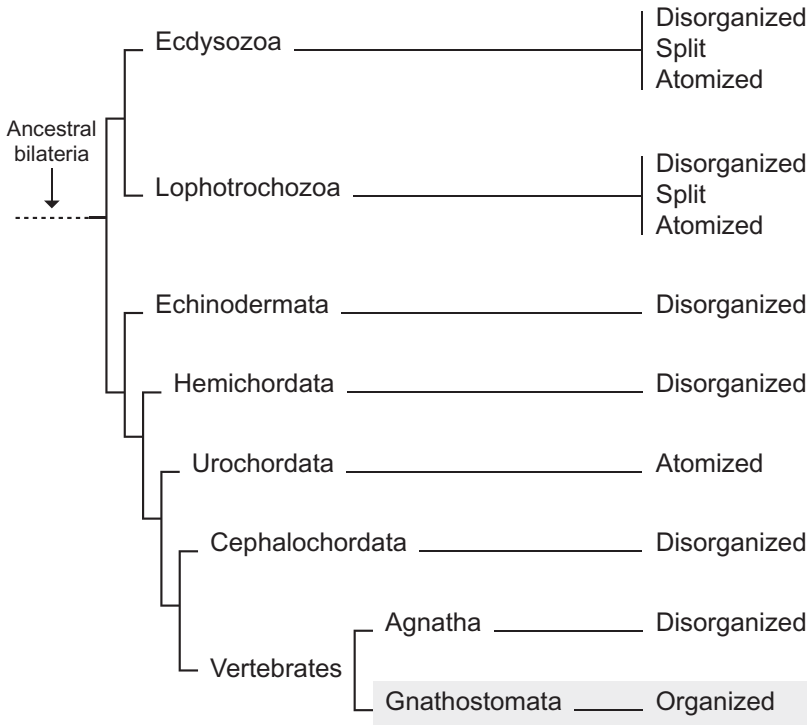

Figure 2 Distribution of various types of *Hox* gene organization among the major groups of bilateria. While invertebrates (Ecdysozoa and Lophotrochozoa) contain taxons with disorganized, split, or atomized *Hox* gene clusters, echinoderms and chordates seems to generally contain disorganized gene clusters, with the exception of urochordates, which do not have distinctive clusters and gnathostomes (jaw vertebrates), which all display tightly organized *Hox* gene clusters. *Adapted from Duboule (2007).*

clusters suggest that at least three events of duplication occurred (Mehta et al., 2013). However, while the R1 and R2 genome duplications (see Putnam et al., 2008) likely predated the separation between agnatha and gnathostomata (Kuraku, Meyer, & Kuratani, 2009), the comparative analysis of potential conserved regulatory sequences raised the possibility that agnatha separated from jaw vertebrates at an earlier stage (discussed in Mehta et al., 2013). In any case, however, the average size of the Japanese lamprey *Hox* clusters, while somewhat variable, is clearly closer to the invertebrate situation than to the gnathostome standard. For example, the *Hoxdelta* cluster (likely related to the tetrapod *HoxD* cluster) is 700 kb large and contains a range of repeats (Mehta et al., 2013). Therefore, by using this particular criterion, lampreys seem to be an interesting exception among vertebrates, which will be further discussed below.

Altogether, compact and well-organized *Hox* clusters thus appear to be a "genomic synapomorphy" of gnathostomes (Fig. 2). As previously argued (Duboule, 2007), there are essentially two exclusive phylogenetic

explanations to account for this observation. In the first view, a bilateria ancestor carried a compacted *Hox* gene cluster, which became subsequently disorganized, split, or atomized in every single branch derived from this ancestor, except in jaw vertebrates. Accordingly, gnathostomes would be the only extant direct representatives (*Hox*-wise) of our bilateria ancestor, an alternative that cannot reasonably be considered outside a theological context.

The alternative scenario involves the presence of a large and disorganized *Hox* cluster in the bilateria ancestor, as a result of gene duplications in *cis*. This type of cluster was conserved in many animal groups, due to the existence of in-*cis* constraints in the temporal activation of these genes (Duboule, 1992). As a consequence, all taxa escaping this major developmental constraint via a distinctive ontogenetic strategy such as a cell lineage-dependent developmental process could have split or atomized their gene clusters without damages (discussed in Duboule, 1994, 2007). This scenario, however, implies that *Hox* gene clusters became organized and compacted during the evolution of gnathostomes. At first, this proposal appears slightly counterintuitive, as we tend to associate some entropy to evolutionary mechanisms. In this case indeed, the *Hox* clusters must have evolved into a higher level of genomic organization, while being located within genomes, which are themselves globally less compact than many invertebrate genomes, at least in terms of gene density.

3. REGULATORY PRIMING

Both the proximal cause of such a "consolidation" of *Hox* clusters in gnathostomes and the nature of its underlying mechanism(s) have remained speculative and poorly understood. Based on genetic studies using the mouse *HoxD* cluster as a paradigm, it was proposed that the evolution and progressive accumulation of long-range regulations at the vicinity of the gene cluster may have triggered the cluster to optimize its transcriptional response by becoming better organized. For example, vertebrate *Hox* genes are all encoded by the same DNA strand and large repeats seem to be excluded from these genomic loci, even though some taxa display more than others (see Di-Poï et al., 2009; Duboule, 2007; Simons, Makunin, Pheasant, & Mattick, 2007). It was also proposed that the presence of a strong global enhancer outside the gene cluster itself may "prime" the region for the *de novo* emergence of novel enhancers, as if the former would facilitate or prepare the evolution of the latter, a process which may have been critical for

the evolution of jaw vertebrates. This concept of "regulatory priming" (Gonzalez, Duboule, & Spitz, 2007), however, has remained devoid of any mechanistic basis.

Recent advances in the study of chromatin structure and long-range gene regulation now provide us with an emerging conceptual framework that helps reconsidering these issues. In particular, the past 10 years have seen important changes in the way we contemplate developmental gene regulation in vertebrates, whereby the complexity of the transcription unit itself has been progressively replaced by the complexity of its regulatory modalities (e.g., Spitz & Furlong, 2012). The first element to consider is that many transcription factors or signaling molecules of key importance for a given developmental mechanism are generally highly pleiotropic and thus require multiple regulatory inputs. Second—and likely related to this former observation—the various enhancer sequences in charge of these distinct controls in space and time are found over very large distances from the target genes. Such a concentration of long-range transcriptional controls (Lettice et al., 2003), referred to as regulatory landscapes (Spitz, Gonzalez, & Duboule, 2003) have now been described around many such genes (e.g., Montavon & Duboule, 2012; Symmons et al., 2014), usually extending over a genomic interval much larger than the size of the target gene itself (e.g., Amano et al., 2009; Berlivet et al., 2013; Ruf et al., 2011).

4. SO NEAR AND YET SO FAR

Gene regulation via long-range acting enhancers, however, raises a number of questions related both to the mechanism of enhancer–promoter contact (de Laat & Grosveld, 2003; Tolhuis, Palstra, Splinter, Grosveld, & De Laat, 2002) and to the choice and selectivity either of the target gene(s) or of the appropriate enhancer whenever complex regulatory landscapes are involved (reviewed in Spitz & Furlong, 2012). The recent observation that such regulatory landscapes often display constitutive chromatin organizations in 3D, i.e., even in the absence of transcription, allows to address these various issue in a slightly different context. By using chromosome conformation capture (de Laat & Dekker, 2012), a large regulatory landscape containing series of digit enhancers was initially described upstream the *HoxD* gene cluster (Montavon et al., 2011). This particular genetic environment also displayed constitutive contacts, i.e., interactions present in cell types devoid of any *HoxD* transcripts. Because this ca.

1 mb large constitutive chromatin domain is delimited by clear boundaries at both ends, which are shared between different, active, or inactive cell types, the existence of a poised regulatory structure was proposed, whose spatial and functional organization would slightly change in digit cells transcribing the requested target *HoxD* genes.

The presence of related chromatin domains along the entire genome was subsequently observed by using Hi-C and 5C technologies (Dixon et al., 2012; Nora et al., 2012) and referred to as topologically associating domains (TADs; Dixon et al., 2012). TADs tend to be structurally conserved during cell differentiation, even though interactions occurring either inside or in-between TADs may vary (Berlivet et al., 2013; Dixon et al., 2015; Giorgetti et al., 2014). While the potential functions of these chromatin domains, for example, as units of coordinated gene regulation or as a means to prevent regulatory interferences, have been discussed (Nora, Dekker, & Heard, 2013; Symmons et al., 2014), the causality of such processes often remains to be clearly established (see de Laat & Duboule, 2013).

5. WITH A LITTLE HELP FROM MY TAD

A comparison between the ES cells dataset from the Ren laboratory (Dixon et al., 2012) and our *in embryo* study of the *HoxD* loci confirmed that the two large regulatory landscapes flanking the *HoxD* cluster were also scored by Hi-C using embryonic cells and hence they correspond to genuine TADs (Andrey et al., 2013). In addition, functional analysis revealed that these two TADs were active at different times without any temporal overlap and contained various sets of enhancers (Andrey et al., 2013; Delpretti et al., 2013; Lonfat, Montavon, Darbellay, Gitto, & Duboule, 2014). Considering these new elements, how can TADs help us think about potential mechanisms leading to the evolutionary consolidation of *Hox* gene clusters in early vertebrates?

If *Hox* clusters became progressively better organized in response to the accumulation of global enhancers located outside, as proposed above, then the force driving such an accumulation of regulations, i.e., the mechanism underlying this "regulatory priming" ought to be the cause of consolidation. In fact, every additional evolution of a given regulatory sequence within the landscape would help tighten the organization of the gene cluster such as to improve a subset of the global adaptive value of the system. We would like to argue that TADs facilitated this process through their constitutive aspect,

i.e., the fact that a poised chromatin structure existed even in the absence of a transcriptional read out. In this view, TADs were used as particular chromatin niches where novel enhancer sequences have evolved (Lonfat & Duboule, 2015) (Fig. 3).

6. TAD HIJACKING

Why should this be the case? At many developmental loci, TADs are often considered as global and independent regulatory units including their target genes (e.g., see Symmons et al., 2014). Accordingly, the *de novo* appearance of a transcription factor specific binding sequence within a given TAD may trigger cells to experience a new differentiation pathway, which may or may not be further elaborated depending on its adaptive value for the organism. In this view, TADs may be seen as regulatory toolkits, as chromatin playgrounds where various factors may have experienced some successful or (mostly) unsuccessful encounters (Lonfat & Duboule, 2015; Lonfat et al., 2014). Understandably, such a trial and error process may function more efficiently should a poised regulatory structure such as TADs already exists, thus merely reducing its cooptation to the progressive evolution of a robust DNA binding sequence inside. Such a structure may favorably "attract" and "sequester" factors and facilitate the latter process, due to a general chromatin environment (a niche) already implementing related types of long-range controls in distinct cellular contexts.

This hypothesis, however, does not explain how this initial protein–DNA recognition can occur. It nevertheless helps thinking about the other necessary factors for a productive interaction to develop, such as the contact with proximal promoter elements and the presence of the required cofactors. Also, the emergence of a novel enhancer within a given TAD likely reenforced the global structure, for example, by evolving novel constitutive contacts in parallel, supporting the new transcription task. This in turn may have triggered new regulatory sequences to emerge, thus leading to a virtuous circle and the observed concentration of enhancers of all kinds within some TADs at developmental loci with high pleiotropic functions (Fig. 3).

7. TAKE IT OR LEAVE IT

Interestingly, *HoxD* cluster is located at the border of two TADs, each chromatin environment regulating a well defined and rather invariant subset of target genes. For instance, the centromeric TAD contains enhancers

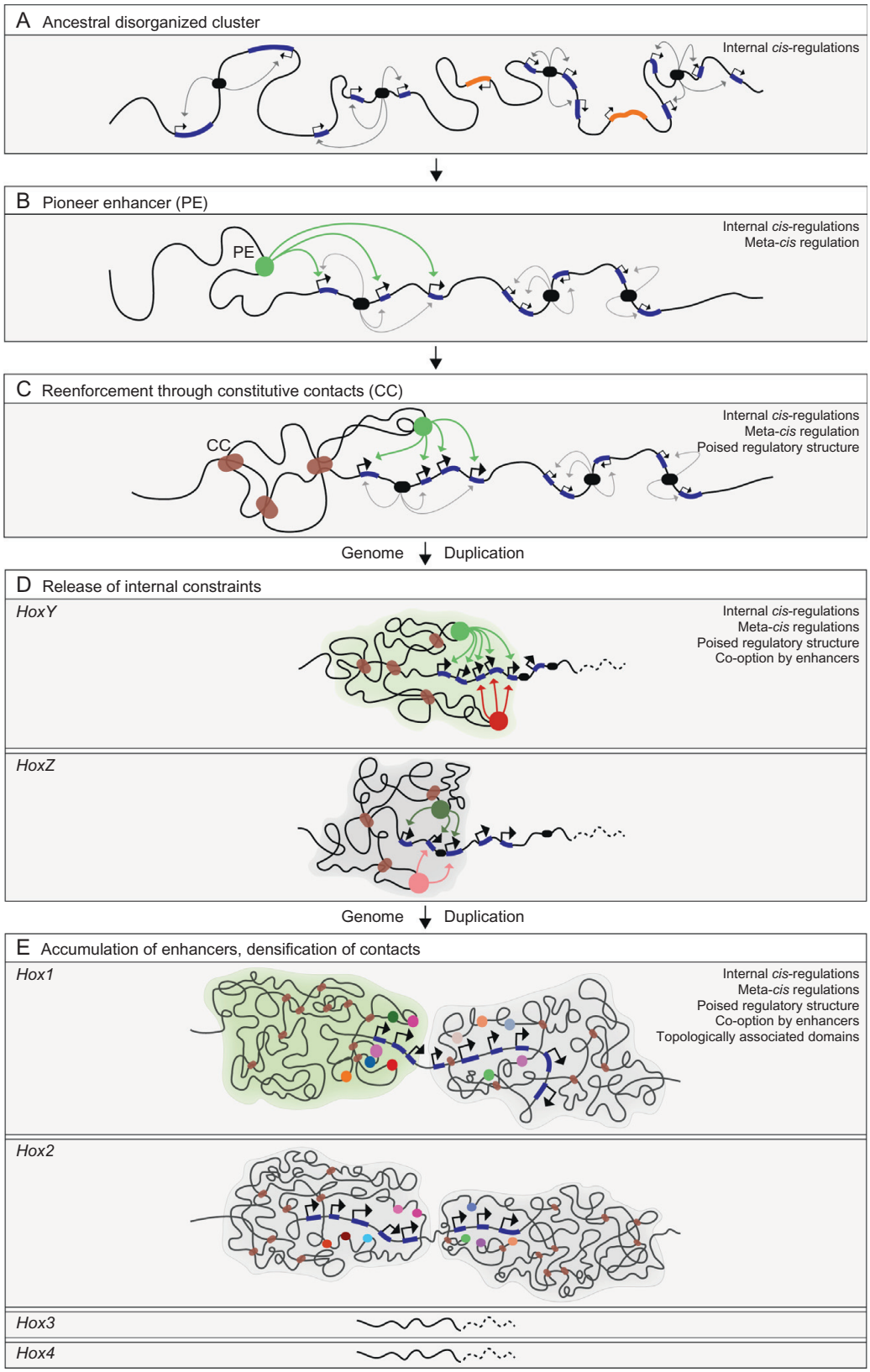

Figure 3 See legend on next page.

active in the developing digits and genitals. In both cases, the *Hoxd13* to *Hoxd9* group of genes is the main target of this regulating landscape (Lonfat et al., 2014; Montavon et al., 2011), even though the function of *Hoxd13* alone would likely be sufficient (Delpretti, Zakany, & Duboule, 2012; Kmita et al., 2005). The same holds true for the telomeric TAD, where the *Hoxd11* to *Hoxd8* subgroup seems to be coordinately regulated, for example, in the developing forearm and intestinal caecum (Andrey et al., 2013; Delpretti et al., 2013). Therefore, it appears that the system itself experiences some internal constraints and that, consequently,

Figure 3 TAD-dependent emergence of multiple global regulations at *Hox* loci and concomitant consolidation of the *Hox* gene clusters in jaw vertebrates. (A) Initially, an ancestral disorganized *Hox* genes (blue) cluster implements its collinear regulation along the rostral to caudal axis through series of internal regulatory sequences (*cis*-regulation; black). The cluster may also contain unrelated genes (orange). In (B), some genes of the cluster become coordinately regulated by the evolution of a strong enhancer located outside the cluster itself, such as not to interfere with the internal regulatory organization. As a consequence, the transcriptional response (the adaptive value) becomes slowly optimized via the progressive genomic rearrangement of the target genes (start of consolidation), in order to optimize their response to this new regulation. This initial meta-*cis* regulation through a "pioneer enhancer" (PE) triggers the formation of a local 3D architecture. (C) Should this coordinated transcriptional control be of some adaptive value, the 3D structure may start to be stabilized or strengthened by the evolution of constitutive interactions (brown ovals), which will help and improve the interactions between the pioneer enhancer and its targets. This global (meta-*cis*) regulation may however interfere with some of the constraints imposed by the internal collinear mechanism. (D) This antagonism between *cis* and meta-*cis* regulations is partly released after genome duplications, which allows some functional compensation to occur between the duplicated clusters and thus restores more flexibility in evolving global regulations. As a consequence of genome duplication, this 3D structure will thus facilitate the evolution of yet other global enhancers on each duplicate (*HoxY* and *HoxZ*), due to the preexistence of a poised transcriptional read out of high adaptive value. This process of "regulatory priming" (Gonzalez et al., 2007) will continue as a virtuous circle and progressively lead to the maximal compaction of the target *Hox* gene cluster, with all genes in the same transcriptional orientation, very few repeats and no longer any foreign genes interspersed. (E) At the same time, large and constitutive regulatory domains will be observed on both sides, reflecting the successful cooption of many global enhancers that accompanied the emergence of most jaw vertebrate innovations (see the text). These domains correspond to topologically associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012). In (B)–(D), only one of the flanking regions of the cluster is shown for sake of simplicity, even though the same mechanism may apply to the other. In (E), only two clusters are shown instead of the four derived from both rounds of genome duplications.

not all possibilities exist regarding the evolution of novel regulations. In this view, whichever new tissue specificity may evolve within one of these TADs, the transcription readout will be predetermined according to the internal properties of the system. This fixed and coordinated response of groups of target genes results from the compaction of the gene cluster, which in our hypothesis is itself due to the accumulation of enhancers, illustrating how the evolving system may generate its own set of constraints.

In this oulipian scheme, it is conceptually difficult to imagine the emergence of a subtle, fine-tuned type of regulation, affecting a single gene in a precise manner. Instead, the evolution of global enhancers within either one of these TADs likely resulted in the massive activation of a prefixed battery of neighboring target genes, leading to major modification of the resulting animals. In this context, it is interesting to note that all enhancers located nearby the *HoxD* clusters are involved in structures of major importance for the vertebrate lineage (Spitz, Herkenne, Morris, & Duboule, 2005). Therefore, while TADs may have represented powerful means to evolve pleiotropy at important developmental loci, it may have equally prevented other regulatory input to emerge, due to the constraints imposed by the system itself and the existence of a metagenic structure at its target end.

8. PIONEER ENHANCERS

It is as yet difficult to evaluate whether the constitutive contacts within a given TADs (i.e., contacts not involved in themselves into a transcriptional read out) predated the appearance of specific regulations or instead, if an initial contact between a promoter and an enhancer became strengthened via the subsequent addition of a constitutive architecture. This conundrum may be addressed in the future by comparing TADs either between different species, or between paralogous loci within the same group. Because a constitutive interaction in itself may not carry an adaptive value as strong as a first specific regulatory contact, we favor the latter possibility as a theoretical starting point (see Fig. 3). Accordingly, an initial productive enhancer–promoter interaction had to be established in order to initiate the process by securing a basic constitutive regulatory landscape leading to a tissue-specific enhanced transcription of a group of *Hox* target genes (Fig. 3A and B).

In this view, a pioneer enhancer sequence must have initiated the system and thus facilitated the evolution of subsequent enhancers within the same TAD. This view is somewhat analogous to the concept of pioneer factors

(Zaret & Carroll, 2011), yet it ought to be considered in an evolutionary—rather than developmental—framework; while pioneer transcription factors facilitate the recruitment of additional factors at a given locus by locally reorganizing chromatin during development or reprogramming (Iwafuchi-Doi & Zaret, 2014), pioneer enhancers would have facilitated the emergence of novel regulatory sequences over evolutionary times, by providing an appropriate chromatin structure and transcriptional readout where trial and error processes can occur.

9. REGULATORY PALEONTOLOGY

As a consequence of the previous proposal, TADs containing series of enhancers associated with highly pleiotropic loci should contain traces of their evolutionary histories. A detailed structure–function comparison between orthologous TADs in various species, associated with the presence or absence of the corresponding morphological traits, should allow for a reconstitution of their progressive evolutionary trajectories, concerning both their specific enhancer sequences and their constitutive architectural contacts, up to the original pioneer enhancer sequence, which may not have been necessarily conserved through time, and corresponding binding factors. While such reconstitutions or parts thereof—may be easy whenever regulations will concern characters separated by large evolutionary distances (for example, the emergence of limbs and mammary glands), they may turn out to be more difficult for characters sharing close evolutionary timing and adaptive value, such as digits and genitals (see Lonfat et al., 2014). In this latter case, the question as to which enhancer was first may remain unsolved.

In summary, we propose that initial TAD-like structures, formed by pioneer enhancers, may have helped accumulating global regulations upstream the *Hox* clusters. The concomitant optimization of the transcriptional response triggered a progressive structural consolidation of the gene clusters. Such an accumulation of regulations, however, was possible only after a genome duplication event had occurred such that internal constraints applied to the gene cluster itself (and to its early critical function during trunk development) were partially released as a consequence of functional redundancy (Fig. 3C and D). Some issues, however, remain difficult to integrate into this conceptual framework. For instance, agnatha such as the Japanese lamprey likely experienced several genome duplications and yet the *Hox* clusters have not been consolidated (Mehta et al., 2013) as in other

vertebrates. This suggests that the emergence of global regulation may not be a trivial event and that genome duplications were not sufficient to trigger this important step, at least at these loci. Concerning lampreys, it will be interesting in the future to evaluate if at least one pioneer global enhancer exists flanking either one of the *Hox* clusters. In any case, it is striking to see that lampreys do not exhibit any of the morphologies that were associated with global regulations at the *HoxD* cluster, such as for example the proximal and distal limbs, external genitals, intestinal caecum, and the metanephric kidneys.

Equally difficult to understand is how all four *Hox* gene clusters were almost equally consolidated along the gnathostome lineage (as schematized in Fig. 3). A parsimonious answer is that consolidation had started before genome duplications had occurred, under the form of a common pioneer global enhancer, which may have started to control several genes coordinately in one particularly adaptive structure of an ancestral chordate (Fig. 3). The accumulation of regulations triggered by this enhancer, however, may have initiated only after duplications of the locus. In this scenario, the presence of a pioneer enhancer would not only have allowed for pleiotropy and consolidation to develop, but would have made it the rule for jaw vertebrates *Hox* clusters. If true, all four *Hox* clusters should *ad minima* be the target of some global regulations. While this is clearly the case for *HoxA* (Berlivet et al., 2013; Woltering, Noordermeer, Leleu, & Duboule, 2014) and *HoxD* (see, e.g., Andrey et al., 2013; Montavon et al., 2011; Spitz et al., 2003), the situation is less clear either for *HoxB* or for *HoxC*. While the *HoxB* cluster seems to reside at the boundary between two TADs, similar to both *HoxA* and *HoxD* (unpublished observations) suggesting the existence of flanking global regulations, the *HoxC* cluster however does not display this particular organization, at least when an ES cells dataset is used (Dixon et al., 2012). Further studies on the regulations of the *HoxB* and *HoxC* clusters may be informative in this respect.

ACKNOWLEDGMENTS

We thank all the past and present members of the Duboule Laboratories for their numerous contributions to these various issues through their data and discussions and in particular C. Bolt and A. Hintermann for their comments on the chapter. This work is supported by funds from the Swiss SNF, the University of Geneva, The Ecole Polytechnique Federale de Lausanne, and the European Research Council (ERC). F.D. was supported by the EU FP7 project IDEAL.

REFERENCES

Aboobaker, A., & Blaxter, M. (2003). Hox gene evolution in nematodes: Novelty conserved. *Current Opinion in Genetics & Development*, *13*, 593–598.

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

Albertin, C. B., Simakov, O., Mitros, T., Wang, Z. Y., Pungor, J. R., Edsinger-Gonzales, E., et al. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, *524*, 220–224.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., & Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: Limb bud-specific differential regulation of competence and active transcription. *Developmental Cell*, *16*, 47–57.

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

Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., et al. (1983). Molecular genetics of the bithorax complex in Drosophila melanogaster. *Science*, *221*, 23–29.

Berlivet, S., Paquette, D., Dumouchel, A., Langlais, D., Dostie, J., & Kmita, M. (2013). Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *PLoS Genetics*, *9*, e1004018.

Boncinelli, E., Somma, R., Acampora, D., Pannese, M., D'Esposito, M., Faiella, A., et al. (1988). Organization of human homeobox genes. *Human Reproduction*, *3*, 880–886.

Cameron, R. A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J. P., Berney, K., et al. (2006). Unusual gene order and organization of the sea urchin hox cluster. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, *306*, 45–58.

de Laat, W., & Dekker, J. (2012). 3C-based technologies to study the shape of the genome. *Methods*, *58*, 189–191.

de Laat, W., & Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, *502*, 499–506.

de Laat, W., & Grosveld, F. (2003). Spatial organization of gene expression: The active chromatin hub. *Chromosome Research*, *11*, 447–459.

Delpretti, S., Montavon, T., Leleu, M., Joye, E., Tzika, A., Milinkovitch, M., et al. (2013). Multiple enhancers regulate Hoxd genes and the Hotdog LncRNA during cecum budding. *Cell Reports*, *5*, 137–150.

Delpretti, S., Zakany, J., & Duboule, D. (2012). A function for all posterior Hoxd genes during digit development? *Developmental Dynamics*, *241*, 792–802.

Di-Poï, N., Montoya-Burgos, J., & Duboule, D. (2009). Atypical relaxation of structural constraints in Hox gene clusters of the green anole lizard. *Genome Research*, *19*(4), 602–610.

Di-Poi, N., Montoya-Burgos, J. I., Miller, H., Pourquie, O., Milinkovitch, M. C., & Duboule, D. (2010). Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature*, *464*, 99–103.

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, *518*, 331–336.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*, 376–380.

Duboule, D. (1992). The vertebrate limb: A model system to study the Hox/HOM gene network during development and evolution. *Bioessays*, *14*, 375–384.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: A basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development* (Suppl.), 135–142.

Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development*, *134*, 2549–2560.
Duboule, D., Baron, A., Mahl, P., & Galliot, B. (1986). A new homeo-box is present in overlapping cosmid clones which define the mouse Hox-1 locus. *The EMBO Journal*, *5*, 1973–1980.
Duboule, D., & Dolle, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *The EMBO Journal*, *8*, 1497–1505.
Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., et al. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, *157*, 950–963.
Gonzalez, F., Duboule, D., & Spitz, F. (2007). Transgenic analysis of Hoxd gene regulation during digit development. *Developmental Biology*, *306*, 847–859.
Graham, A., Papalopulu, N., & Krumlauf, R. (1989). The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell*, *57*, 367–378.
Harding, K., Wedeen, C., McGinnis, W., & Levine, M. (1985). Spatially regulated expression of homeotic genes in Drosophila. *Science*, *229*, 1236–1242.
Hart, C. P., Fainsod, A., & Ruddle, F. H. (1987). Sequence analysis of the murine Hox-2.2, -2.3, and -2.4 homeo boxes: Evolutionary and structural comparisons. *Genomics*, *1*, 182–195.
Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., et al. (2002). The genome sequence of the malaria mosquito Anopheles gambiae. *Science*, *298*, 129–149.
Ikuta, T., Yoshida, N., Satoh, N., & Saiga, H. (2004). Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 15118–15123.
Iwafuchi-Doi, M., & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes & Development*, *28*, 2679–2692.
Kmita, M., & Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. *Science*, *301*, 331–333.
Kmita, M., Tarchini, B., Zakany, J., Logan, M., Tabin, C. J., & Duboule, D. (2005). Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. *Nature*, *435*, 1113–1116.
Kuraku, S., Meyer, A., & Kuratani, S. (2009). Timing of genome duplications relative to the origin of the vertebrates: Did cyclostomes diverge before or after? *Molecular Biology and Evolution*, *26*, 47–59.
Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., et al. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*, 1725–1735.
Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, *276*, 565–570.
Lonfat, N., & Duboule, D. (2015). Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS Letters*, *589*(20 Pt. A), 2869–2876 (review). http://dx.doi.org/10.1016/j.febslet.2015.04.024.
Lonfat, N., Montavon, T., Darbellay, F., Gitto, S., & Duboule, D. (2014). Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science*, *346*, 1004–1006.
Mehta, T. K., Ravi, V., Yamasaki, S., Lee, A. P., Lian, M. M., Tay, B. H., et al. (2013). Evidence for at least six Hox clusters in the Japanese lamprey (Lethenteron japonicum). *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 16044–16049.

Minguillon, C., Gardenyes, J., Serra, E., Castro, L. F., Hill-Force, A., Holland, P. W., et al. (2005). No more than 14: The end of the amphioxus Hox cluster. *International Journal of Biological Sciences*, *1*, 19–23.

Montavon, T., & Duboule, D. (2012). Landscapes and archipelagos: Spatial organization of gene regulation in vertebrates. *Trends in Cell Biology*, *22*(7), 347–354.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., et al. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell*, *147*, 1132–1145.

Nora, E. P., Dekker, J., & Heard, E. (2013). Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? *Bioessays*, *35*, 818–828.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*, 381–385.

Pierce, R. J., Wu, W., Hirai, H., Ivens, A., Murphy, L. D., Noel, C., et al. (2005). Evidence for a dispersed Hox gene cluster in the platyhelminth parasite Schistosoma mansoni. *Molecular Biology and Evolution*, *22*, 2491–2503.

Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature*, *453*, 1064–1071.

Ruf, S., Symmons, O., Uslu, V. V., Dolle, D., Hot, C., Ettwiller, L., et al. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nature Genetics*, *43*, 379–386.

Sanchez-Herrero, E., Vernos, I., Marco, R., & Morata, G. (1985). Genetic organization of Drosophila bithorax complex. *Nature*, *313*, 108–113.

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

Simons, C., Makunin, I. V., Pheasant, M., & Mattick, J. S. (2007). Maintenance of transposon-free regions throughout vertebrate evolution. *BMC Genomics*, *8*, 470.

Spitz, F., & Furlong, E. E. (2012). Transcription factors: From enhancer binding to developmental control. *Nature Reviews. Genetics*, *13*, 613–626.

Spitz, F., Gonzalez, F., & Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell*, *113*, 405–417.

Spitz, F., Herkenne, C., Morris, M. A., & Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nature Genetics*, *37*, 889–893.

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., et al. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Research*, *24*, 390–400.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell*, *10*, 1453–1465.

Vonk, F. J., Casewell, N. R., Henkel, C. V., Heimberg, A. M., Jansen, H. J., McCleary, R. J., et al. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 20651–20656.

Woltering, J. M., Noordermeer, D., Leleu, M., & Duboule, D. (2014). Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biology*, *12*, e1001773.

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: Establishing competence for gene expression. *Genes & Development*, *25*, 2227–2241.