

Analysis of long-range gene regulation at the *HoxD* locus

Thèse N° 9375

Présentée le 18 avril 2019

à la Faculté des sciences de la vie

Unité du Prof. Duboule

Programme doctoral en approches moléculaires du vivant

pour l'obtention du grade de Docteur ès Sciences

par

Ana Rita DE CARVALHO AMÂNDIO LHOPITALIER

Acceptée sur proposition du jury

Prof. M. Blokesch, présidente du jury

Prof. D. Duboule, directeur de thèse

Dr A. Taddei, rapporteuse

Prof. A. Reymond, rapporteur

Prof. B. Deplancke, rapporteur

2019

Résumé

Les gènes *Hoxd* sont essentiels pour le développement des différents axes du corps chez les vertébrés d'où l'importance des mécanismes sous-jacents de régulation. Parmi ces différents mécanismes on retrouve les amplificateurs agissant à longue distance qui sont situés dans deux paysages de régulation adjacents. Les analyses de l'architecture de la chromatine au niveau de ce cluster de gènes a mis en évidence l'existence de deux "topologically associating domains" de part et d'autre du cluster et qui incluent ces paysages de régulation. Néanmoins, les dynamiques de ces régions de régulation ainsi que la stabilité et la contribution fonctionnelle des interactions amplificateur-promoteur pendant le développement embryonnaire restent à définir.

Dans ce travail nous avons analysé l'organisation tridimensionnelle de la chromatine et le profil de transcription au locus *Hoxd*, ceci à différentes étapes du développement du tubercule génital (TG). Nous observons que la conformation tridimensionnelle de cette région est antérieure à l'émergence du TG. En parallèle, nous observons une diminution des niveaux de l'ARN qui corrèle avec la diminution des contacts promoteur-amplificateur dans le désert génique adjacent. Cette diminution a lieu tout en maintenant un sous-ensemble de contacts associés au complexe CTCF/Cohésine qui est préservé indépendamment de l'état de transcription du cluster de gènes.

Afin d'explorer la contribution fonctionnelle de ces paysages de régulation nous avons utilisé la technologie CRISPR-Cas9 pour créer des souris avec des délétions partielles de cette région et des délétions ciblées de contacts transitoires (associés aux amplificateurs) et constitutifs (associés au complexe CTCF/Cohésine). Nous observons que des délétions uniques de contacts transitoires ou constitutifs ont peu d'effet sur l'expression de *Hoxd* dans le TG. Au contraire, la délétion unique d'un amplificateur déjà caractérisé de *Hoxd*, l'élément Prox, ou des délétions incluant de nombreux amplificateurs résultent en une diminution de l'expression des gènes *Hoxd*. Nos résultats suggèrent alors que tous les éléments amplificateurs dans les paysages de régulation complexe n'ont pas la même importance fonctionnelle soulignant l'existence d'un système dynamique mais aussi robuste pour réguler l'expression génique de *Hoxd*.

Mots clefs:

Gènes *Hox*, amplificateur, paysages régulateurs, TAD, régulation à longue distance, CTCF, CRISPR-Cas9, 4C-seq, tubercule génital, *Hotair*.

Abstract

Hoxd genes are essential for the development of the various body axes in vertebrates and hence the underlying regulatory mechanisms are of paramount importance. Among these various mechanisms are long-range acting enhancers, which are located in the two adjacent regulatory landscapes. Analyses of chromatin architecture at this gene cluster has revealed the existence of two topologically associating domains (TADs) flanking the cluster and encompassing these regulatory landscapes. However, the dynamics of such regulatory regions as well as the stability and functional contribution of specific enhancer-promoter interactions during development remains to be established.

In this work, we analysed the 3D chromatin organization and transcription profile at the *HoxD* locus, at different time points during genital tubercle (GT) development, and observe that the 3D conformation of this regulatory region predates the embryonic emergence of the GT. Along with this tissue development, we observe a reduction in transcript levels correlating with a decrease in enhancer-promoter chromatin loops within the adjacent gene desert. This decrease occurs while maintaining a subset of CTCF/Cohesin associated contacts, which are preserved independently from the transcriptional status of the gene cluster.

To further explore the functional contribution of this regulatory landscape, we used CRISPR-Cas9 technology to generate mice carrying partial deletions of this region, as well as targeted deletions of both transient (enhancer associated) and constitutive (CTCF/Cohesin associated) contacts. We observe that single deletions of both transient and constitutive contacts displayed little if any effect on *Hoxd* genes expression in the GT. On the contrary, the single deletion of a previous characterized *Hoxd* enhancer, the Prox element, or deletions comprising several enhancers, result in the reduction of *Hoxd* genes expression levels. Overall our results suggest that not all enhancer elements within a complex regulatory landscape have the same functional strength, and highlight the existence of a dynamic yet robust system to tightly regulate *Hoxd* genes expression.

Keywords:

Hox genes, enhancers, regulatory landscapes, TAD, long-range regulation, CTCF, CRISPR-Cas9, 4C-seq, genital tubercle, *Hotair*

Table of Contents

1. Introduction	9
1.1 Hox genes	9
1.1.1 Homeobox genes	9
1.1.2 Evolution and organization of <i>Hox</i> genes	9
1.1.3 Collinearity	11
1.2. Hox genes in the primary body axis	12
1.2.1 Expression and function	12
1.2.1.1 Early <i>Hox</i> expression	12
1.2.1.2 <i>Hox</i> genes role in the patterning of the axial skeleton	12
1.2.1.3 Functional role of <i>Hox</i> paralogous genes	14
1.2.2 Regulation	14
1.2.2.1 Regulation by the Polycomb and Trithorax protein group	14
1.2.2.2 Regulation of <i>Hox</i> genes by long non-coding RNAs	15
1.3. Hox genes in the secondary body axis	17
1.3.1 Regulation and transcription of <i>Hoxd</i> genes in the limb	17
1.3.2 Regulation of <i>Hox</i> genes in the genital tubercle	18
1.4. Genital tubercle development	20
1.4.1 Early development	20
1.4.2 Late development (sex specific GT development)	20
1.4.3 GT developmental regulators (genetic pathways involved in GT development)	21
1.4.4 <i>Hox</i> genes in the GT	23
1.5. Regulatory landscapes	24
1.5.1 Enhancers	24
1.5.2 Long-range gene regulation, regulatory landscapes, and redundancy	25
1.5.3 Topologically associating domains (TADs)	26
1.5.4 CTCF and TAD boundaries	27
2. Scope of the thesis	29
3. Results	31
3.1 Hox genes expression and epigenetic profile during GT development	31
3.2 Characterization of the <i>HoxD</i> regulatory landscapes during GT development	34
3.3 Structural organization of the <i>HoxD</i> cluster and the C-DOM during GT development	36
3.4 <i>In vivo</i> activity of mouse <i>HoxD</i> cis-regulatory elements	42
3.5 Analysis of partial C-DOM deletions <i>in vivo</i>	44
3.6 Functional contribution of different cis-regulatory elements	47
3.7 Chromatin organization in the <i>Del(GT2)</i> allele	50
3.8 The case of <i>Evx2</i>	51
3.9 Deletion or inversion of CTCF	52
4. Discussion	57
4.1 <i>Hoxd</i> genes regulation	57
4.2. Constitutive vs. transient interactions: The spatial organization of enhancer-promoter interactions at the <i>HoxD</i> locus	59
4.3 The C-DOM regulatory potential: Effects of partial C-DOM deletions	62
4.4 Contribution of specific enhancer sequences to the regulation of <i>Hoxd</i> genes	63
4.5 A transcriptional hub: In-light of the phase separation model	64

4.6 Not all enhancers behave the same: The case of <i>Prox</i>	66
4.7 Not all genes behave the same: The case of <i>Evx2</i>	66
4.8 Effects of deleting or inverting island V	67
4.9 Conclusion	69
5. Materials and Methods	71
5.1 Mouse strains and genotyping	71
5.2 CRISPR-Cas9	72
5.3 Transgenic analysis	74
5.4 X-gal staining	75
5.5 Whole-mount <i>in situ</i> hybridization	75
5.6 RT-qPCR	76
5.7 4C-seq	76
5.7.1 Sample preparation	76
5.7.2 Data processing	76
5.8 ChIP-seq	78
5.8.1 Sample preparation	78
5.8.2 Data analysis	78
5.9 ATAC-seq	78
5.9.1 Sample preparation	78
5.9.2 Data analysis	79
5.10 RNA-seq	79
5.10.1 Sample preparation	79
5.10.2 Data analysis	79
6. Bibliography	81
7. Acknowledgements	95
8. Annex 1	97
9. Curriculum Vitae	126

List of figures and tables

Figure 1: Schematic representation of the <i>Hox</i> clusters organization	10
Figure 2: Regulation of <i>Hoxd</i> genes in various tissues	19
Figure 3: Characterization of <i>Hox</i> genes expression in developing GT	33
Figure 4: Characterization of the <i>HoxD</i> regulatory landscapes during GT development	35
Figure 5: Analysis of chromatin architecture at the <i>HoxD</i> locus during GT development	37
Figure 6: Structural organization of the C-DOM during GT development	39
Figure 7: Validation of the C-DOM structural organization	40
Figure 8: Decrease in enhancer-promoter contacts during GT development	41
Figure 9: <i>In vivo</i> activity of mouse <i>HoxD</i> regulatory elements	43
Figure 10: Analysis of partial C-DOM deletions	46
Figure 11: Analysis of single <i>cis</i> -regulatory elements deletion alleles	49
Figure 12: Chromatin architecture in the <i>Del(GT2)</i> allele	51
Figure 13: Regulation of <i>Evx2</i> in the <i>Del(GT2)</i> allele	52
Figure 14: CTCF deletion or inversion	54
Figure 15: Chromatin architecture in CTCF deletion or inversion	56
Figure 16: Schematic representation of the <i>HoxD</i> locus topology during GT development	61
Figure 17: Effect of <i>cis</i> -elements deletions on the <i>HoxD</i> locus topology and gene expression	65
Figure 18: Alleles generated by CRISPR-Cas9	73
Table 1: List of genotyping primers	71
Table 2: List of sgRNAs	72
Table 3: List of fosmids	74
Table 4: List of primers used for recombineering	74
Table 5: List of 4C-seq primers	77

1. Introduction

1.1 *Hox* genes

1.1.1 Homeobox genes

Homeobox genes are master regulators of a wide range of developmental processes, they are found in almost all eukaryotes including animals, plants, and fungi (reviewed in (Holland, 2013)). These genes contain a common DNA sequence element, the homeobox, that encodes a DNA binding domain, the homeodomain. The majority of homeodomain containing proteins act as transcription factors to activate or repress the expression of target genes.

In animals the largest group of homeobox genes is the ANTP-class. It is thought that this class of genes originated from a hypothetical ancestral ANTP-class gene (Proto-ANTP) that underwent extensive tandem gene duplication, leading to an array of related genes positioned along a chromosome (Garcia-Fernández, 2005; Holland, 2015). Subsequently, many of these genes became dispersed in the genome, while others remained aggregated forming three independent gene clusters - *NK*, *Hox*, and *ParaHox*- that retained distinct degrees of conservation in different evolutionary lineages. For example, vertebrates have very organized and conserved *Hox* and *ParaHox* gene clusters but disrupted *NK* gene clusters, whereas dipteran flies retained a tight *NK* gene cluster, but lost some *ParaHox* genes and have a disrupted *Hox* gene cluster (reviewed in Holland, 2015).

1.1.2 Evolution and organization of *Hox* genes

Hox genes are homeobox-containing transcription factors that are present in all bilateria and in cnidaria but have not been found in plants or sponges. It is thought that, through their evolution and expansion, these genes were crucial for the rapid diversification of the body plans of all bilaterians. Usually invertebrates have a single *Hox* gene cluster (although it is split in *Drosophila*), whereas vertebrates possess multiple clusters as a result of whole-genome duplication events. The common ancestor of all chordates is presumed to have had a single cluster. It is thought that two rounds of whole-genome duplication (2R hypothesis) occurred in early vertebrate evolution (Ohno, 1970), resulting in the four *Hox* clusters in jawed vertebrates (*HoxA*, *HoxB*, *HoxC* and *HoxD* clusters) (Duboule, 2007) (Figure 1).

In teleost fishes there was an additional third round (3R) of whole-genome duplication resulting in up to seven or eight *Hox* clusters depending on the species (Amores *et al.*, 1998). Interestingly, the consequences of the teleost fish-specific 3R whole-genome duplication did

not lead to a major difference in total number of *Hox* genes in comparison to other vertebrates (non-teleost). Human and mouse have 39 *Hox* genes, whereas teleost fish have on average 48 *Hox* genes (e.g., zebrafish has 49, sickleback has 48 *Hox* genes), this is due to significant loss of specific *Hox* genes after the 3R event in the latter group (Duboule, 2007; Holland, 2013). Each duplicated ancestral *Hox* gene resulted in several genes that are paralogs of each other. In mammals, the genes in each of the four clusters (*HoxA*, *HoxB*, *HoxC*, and *HoxD*) can be subdivided into 13 paralogous groups based on their sequence similarity and position within the clusters, in each cluster only a subset of paralogous genes was retained (reviewed in Maconochie *et al.*, 1996).

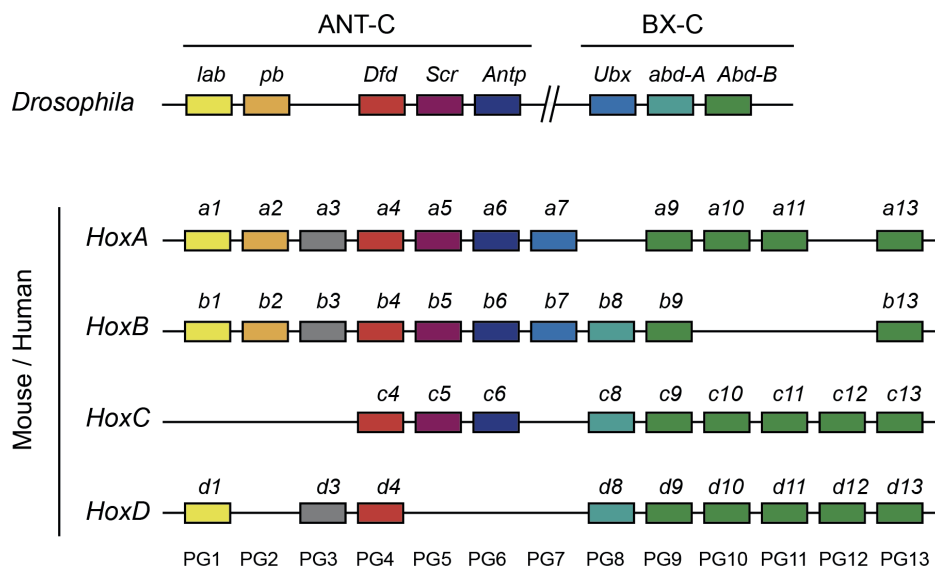


Figure 1: Schematic representation of the *Hox* clusters organization

In *Drosophila* *Hox* genes are grouped into two complex loci, the *Antennapedia* (*ANT-C*) and *Bithorax* (*BX-C*) complexes. In both mouse and human, there are 39 *Hox* genes distributed by four clusters (*HoxA*, *HoxB*, *HoxC*, and *HoxD*) that arose by two rounds of whole genome duplication. Based on sequence similarities, members of the different clusters can be classified into paralogous groups 1 to 13 (PG1 to PG13). Each gene is represented by a colored box. *lab*=labial; *pb*=proboscipedia; *Dfd*=deformed; *Scr*=sex combs reduced; *Antp*=antennapedia; *Ubx*=ultrabithorax; *abd-A*=abdominal-A; *Abd-B*=abdominal-B. (adapted from (Favier and Dollé, 1997; Mallo and Alonso, 2013))

1.1.3 Collinearity

Hox genes were first described in *Drosophila melanogaster* where they are present in a fractured gene cluster, the *Antennapedia* and *Bithorax* complexes (*ANT-C* and *BX-C*, respectively) (Lewis, 1978; Kaufman *et al.*, 1980). E. B. Lewis (1978) work on the *BX-C* reported that mutations on the cluster caused homeotic transformations (i.e., transformation of a body segment into another) and that the gene order reflected the anatomical areas that are under the domain of influence of each single *Hox* genes. As described for *Drosophila melanogaster* (Lewis, 1978), in vertebrates the linear position of the *Hox* genes along the chromosomes reflects their expression and functional domains along the anterior-posterior (AP) body axis during development, in a process named spatial collinearity (Gaunt, Sharpe and Duboule, 1988; Duboule and Dollé, 1989). The collinear pattern of expression of *Hox* genes is observed not only in the various embryonic trunk tissues (Graham *et al.*, 1989) but also along the secondary body axis structures such as the limbs and the external genitalia (Dollé *et al.*, 1991; Haack and Gruss, 1993; Nelson *et al.*, 1996) as well as the digestive track and urogenital system (Dollé *et al.*, 1991; Yokouchi *et al.*, 1995).

In addition to spatial collinearity, vertebrate *Hox* genes are expressed in a time sequence that follows their physical order within the cluster, a process referred to as temporal collinearity (Dolle *et al.*, 1989; Izpisua-Belmonte *et al.*, 1991). Genes at the 3' end of clusters are expressed more anteriorly and earlier, while those at the 5' end of clusters are expressed more posteriorly and later in development. In the primary body axis, 3' *Hox* genes initiate expression during gastrulation (e.g., paralogous groups *Hox1* and *Hox2*) while 5' genes are expressed later in development, in the tail bud after somite formation (e.g., *Hoxd13* is expressed in the most posterior part of the embryo at early embryonic day (E) 9) (Deschamps and Nes, 2005). This temporal activation can also be observed in the secondary body axis (Dolle *et al.*, 1989; Zakany and Duboule, 2007), where anterior genes (3' genes) are activated earlier than posterior genes (5' genes). This way, the precise temporal activation of the *Hox* genes in their exact domain of expression is crucial for establishment of regional identities.

1.2. *Hox* genes in the primary body axis

1.2.1 Expression and function

1.2.1.1 Early *Hox* expression

As mentioned before, *Hox* genes are major players in the patterning of the primary body axis. In vertebrates, the formation of this axis occurs by the sequential assembly of several embryonic tissues in a head to tail fashion as the embryo extends its posterior end. *Hox* genes are present in tissues of ectodermal origin (as the developing central and peripheral nervous system), mesodermal origin (notochord, somites and developing urogenital tract) and endodermal origin (digestive tract) (Nolte and Krumlauf, 2015).

The expression of *Hox* genes starts during the early stages of mouse development and will be essential for the patterning of all three embryonic germ layers. During gastrulation, as the primitive streak extends, the first 3' *Hox* genes are activated in the posterior streak area (at E7.2, in the mouse). Sequentially the domain of expression spreads to the anterior part of the streak that harbors neuro-mesodermal progenitors (reviewed in Neijts and Deschamps 2017). As development progresses, more 5' *Hox* genes start to be expressed and in turn their expression domain spreads anteriorly.

The first region to express *Hox* genes (around the primitive streak) will become extraembryonic mesoderm (Forlani and Deschamps, 2003; Neijts and Deschamps, 2017). Even though these cells will not directly contribute to the future axial and paraxial embryonic structures, early *Hox* expression is relevant for proper trunk development by influencing the timing of *Hox* activation (Neijts and Deschamps, 2017). For example, deletion of an *Hoxc8* early enhancer caused a transient delay in the initial transcription of the gene. This temporal delay is sufficient to produce axial skeleton abnormalities (Juan and Ruddle, 2003). Later during development, the expression of anterior *Hox* genes reaches the node region where the progenitors of the paraxial mesoderm and neuroectoderm reside and will contribute directly to axis elongation.

1.2.1.2 *Hox* genes role in the patterning of the axial skeleton

Both gain and loss-of-function experiments have provided evidence for the pivotal role of *Hox* genes in skeleton patterning, with several adjacent *Hox* genes being linked to vertebrae identity at particular levels of the axial skeleton (reviewed in Mallo *et al.*, 2010). Two main models have been proposed to explain the role of *Hox* genes during patterning the AP axis.

The “*Hox* code” model postulates that the identity of a given body segment is specified by the particular combination of different active *Hox* genes (Kessel and Gruss, 1991). This model was proposed based on the observation that mice embryos exposed to retinoic acid (RA) at E7.5 had posterior homeotic transformation of vertebrae (e.g., the last cervical vertebra gains the identity of thoracic vertebrae) while embryos exposed at E8.5 had anterior homeotic transformation (e.g., lumbar vertebrae gain rib projections becoming thoracic like). In this model, RA would shift the boundaries of *Hox* genes expression modifying the combination of genes expressed in a determinate region of the body plan and in turn lead to homeotic transformations (Kessel and Gruss, 1991). However, several additional studies have shown that upon perturbation of *Hox* expression, the phenotypes are often observed only in the most anterior part of their particular domain of expression (reviewed in Favier and Dollé, 1997). For example, the combined inactivation of *Hox10* paralogous genes in mice, showed phenotypes in the anterior domain of expression (with rib bearing lumbar segments) but no pronounced phenotype in the most posterior domain of expression of the genes (where *Hox11* paralogous genes are co-expressed) (Wellik and Capecchi, 2003). These phenotypes can be explained by the posterior prevalence model.

This model proposes that more posteriorly expressed genes will suppress the activity of more anteriorly expressed genes in adjacent *Hox* domains of expression (Duboule and Morata, 1994). The posterior prevalence model was first proposed to explain the results observed in chicken wing buds upon retinoic acid treatment. When 3' *Hox* genes were activated, in regions where more 5' genes were also expressed, no phenotype was observed. However, when more 5' *Hox* genes, to ones endogenously expressed, were activated there was a duplication of structures (Duboule, 1991). Further evidence in favor of this model came from genetic experiments. For example, loss-of-function of *Hoxa3* only affects the cervical structures even though the expression domain of the gene encompasses a more posterior region (Chisaka and Capecchi, 1991). Conversely, overexpression of *Hoxd4* in regions anterior to its wildtype domain of expression results in a posterior homeotic transformation, whereby the occipital bones gain cervical identity (Lufkin *et al.*, 1992). Of note, neither of the models can fully explain all the observations derived from the genetic analysis of *Hox* function along the AP axis.

1.2.1.3 Functional role of *Hox* paralogous genes

One important aspect to consider when analyzing *Hox* gene function in vertebrates is the existence of functional redundancy between paralogous group genes. Functional redundancy was validated through the generation of compound mutations (reviewed in Mallo *et al.*, 2010; Wellik, 2007). In all cases, paralogous mutants exhibited phenotypes that were more severe than single or double mutant combinations (reviewed in Wellik 2007). For example, whilst single mutants from *Hox4* paralogous group resulted in incompletely penetrant defects in cervical vertebra (C2 and C3), removal of three of the four paralogous *Hox4* genes (*Hoxa4*, *Hoxb4*, and *Hoxd4*) resulted in fully penetrant cervical anterior homeotic transformations (C2 through C5) (Horan *et al.*, 1995). *Hox10* paralogous genes were shown to specify lumbar identity, the functional loss of these genes resulted in the morphological transformation of the lumbosacral region to thoracic-like (vertebrae with small rib projections) (Wellik and Capecchi, 2003). Depletion of *Hox11* paralogous genes lead to a lumbar-like transformation of the sacral region, showing that they are required for sacral vertebrae formation (Wellik and Capecchi, 2003). Further evidence of functional redundancy came from the rescue of *Hoxa11* and *Hoxd11* loss of function phenotype by a *Hoxd11* transgene (Záknáy *et al.*, 1996). Interestingly, *Hox13* paralogous group genes are responsible for terminating the axial elongation program. Accordingly, precocious over-expression of *Hox13* genes resulted in premature arrest of axial growth (Young *et al.*, 2009), conversely deletion of *Hoxb13* caused elongation of the AP axis (Economides and Capecchi, 2003). In snakes, there is a delay in *Hox13* group activation correlating with body plan elongation (Di-poi *et al.*, 2010). These studies, and others not mentioned here, highlight the role of particular *Hox* paralogous group genes in patterning specific axial anatomic domains.

1.2.2 Regulation

1.2.2.1 Regulation by the Polycomb and Trithorax protein group

The Polycomb (PcG) and Trithorax (TrxG) group genes have long been implicated in the regulation of *Hox* genes. Both PcG and TrxG are multi-protein complexes responsible for post-translational modifications of histone tails. Polycomb repressive complexes 1 and 2 (PRC1 and PRC2) are important for gene silencing. EZH2, a PRC2 component, is responsible for the tri-methylation of lysine 27 of the histone H3 tail (H3K27me3), a mark associated with gene repression (Cao *et al.*, 2002). PRC1 contains the RING1B E3 ubiquitin ligase that is

responsible for the ubiquitylation of H2A at lysine 119 (de Napoles *et al.*, 2004; Wang *et al.*, 2004). Inversely, TrxG complexes are associated with transcriptional activation by the trimethylation of histone H3 tail at lysine 4 (H3K4me3) (Ruthenburg and Wysocka, 2007).

In *Drosophila* late development, *Hox* genes expression patterns are maintained by the combined action of these proteins. On the one hand, PcG are required for the repression of *Hox* gene expression, on the other hand, TrxG are responsible for maintaining active *Hox* genes transcription in the correct domains, counteracting the activity of PcG silencing (Schuettengruber *et al.*, 2017). In mouse embryonic stem cells, the *Hox* clusters display both H3K4me3 and H3K27me3 marks. This bivalent chromatin state is thought to reflect a poised chromatin state that would allow for rapid gene activation upon cell differentiation (Bernstein *et al.*, 2006). During mouse early tail bud development, there is a gain of H3K4me3 and a loss of H3K27me3 marks over the *HoxD* cluster that accompanies the collinear activation of these genes (Soshnikova and Duboule, 2009). The transition zone between these two epigenetic states coincides with the transcriptional activation of a specific *Hoxd* gene in time and space (Soshnikova and Duboule, 2009). Furthermore, these chromatin mark dynamics coincide with transitions in the spatial conformation of *Hox* genes (Noordermeer *et al.*, 2011). Repressed *Hox* genes spatially separate from adjacent chromosome regions forming silent domains. Upon *Hox* genes transcriptional activation, the transcribed part of the cluster segregates from the Polycomb-repressed domain (Noordermeer *et al.*, 2011).

While in *Drosophila*, Polycomb response elements (PREs) target PcG to chromatin resulting in gene silencing, in mammals, recruitment of PcG components to target sites appears more complex. Several mechanisms for PcG recruitment have been suggested in mammals including, hypomethylated CpG islands and other CG-rich sequences (Ku *et al.*, 2008; Mendenhall *et al.*, 2010; Riising *et al.*, 2014), interactions between PcG proteins and a number of transcription factors, such as SNAIL, RUNX1, JARID2, AEBP2, and YY1, and long non-coding RNAs (lncRNAs) (reviewed in Schuettengruber *et al.*, 2017). Overall, the available evidence suggests that several mechanisms, that may be complementary, are involved in PcG recruitment to target regions.

1.2.2.2 Regulation of *Hox* genes by long non-coding RNAs *

Long non-coding RNAs (lncRNAs) have been proposed to be involved in the regulation of *Hox* genes. Indeed, there is increasing evidence that long non-coding RNAs can act as gene expression regulators, for example by interacting with transcription factors and chromatin

modifiers to modulate gene expression during development (Lee, 2012). Several lncRNAs associated with the mammalian *Hox* clusters have been identified (Mainguy *et al.*, 2007; Rinn *et al.*, 2007). However, a complete understanding of their roles in *Hox* gene regulation is currently lacking and their *in vivo* functions have yet to be fully assessed.

Perhaps the most prominent example of *Hox* regulation by lncRNAs is the case of *Hotair* (*Hox* transcript antisense intergenic RNA), a lncRNA transcribed from the intergenic region between *Hoxc11* and *Hoxc12* in the *HoxC* cluster. *Hotair* was proposed to repress the expression of 5' *Hoxd* genes in trans by associating with chromatin modification complexes such as PRC2, LSD1, and CoREST/REST (Rinn *et al.*, 2007; Tsai *et al.*, 2010). Indeed, the knockdown of *Hotair* in human fibroblasts led to decreased binding of these repressive complexes at the *HoxD* cluster and to an increase in *Hoxd* genes expression (Rinn *et al.*, 2007). Of note, the role of *Hotair* in direct PRC2 recruitment has been recently disputed (Portoso *et al.*, 2017).

The function of *Hotair* has also been addressed *in vivo*, using genetically engineered mouse models that lack *Hotair* expression. Analyses of a mouse line comprising a genetic deletion of the entire *HoxC* cluster (Suemori and Noguchi, 2000) showed little effect *in vivo*, with no alteration of *Hoxd* genes expression nor of repressive chromatin marks at the *HoxD* locus (Schorderet and Duboule, 2011). However, a targeted deletion of the two exons of *Hotair* was reported to lead to wrist malformation and to homeotic transformations of the spine in mice (Li *et al.*, 2013). These phenotypes were associated with de-repression of *Hox* genes and of imprinted genes by modulation of their chromatin state (Li *et al.*, 2013). Of note, the authors did not observe changes in the neighboring *Hoxc11* and *Hoxc12* genes, supporting the idea that *Hotair* functions in trans (Li *et al.*, 2013). Two additional mouse models were produced, one comprising a partial deletion of the two exons of *Hotair* and the other a partial deletion of the second exon (Lai *et al.*, 2015). In both cases a subtle homeotic transformation of the 4th caudal vertebra was reported, but no wrist malformation nor homeotic transformation of the lumbar vertebrae were observed (Lai *et al.*, 2015). These conflicting results regarding the physiological function of *Hotair* led us to analyze the molecular and phenotypic consequences of deleting the *Hotair* locus *in vivo*.

We addressed the disparities found in the literature by using a targeted deletion of *Hotair* (Li *et al.*, 2013) to assess its role during mouse development. We found that deleting the *Hotair* locus has no detectable effects on wrist and spine morphology in the mouse. Whole transcriptome analyses revealed that this deletion does not affect *Hoxd* genes expression in trans, in any of the embryonic tissues analyzed. However, we observed significant expression

changes of the neighboring *Hoxc11* and *Hoxc12* genes in the mutant mice, most likely due to alterations in the underlying DNA sequence. Taken together, our results suggest that deleting the *Hotair* locus has little to no effect on mouse embryonic development. For further information on the work done on *Hotair* please consult annex 1.

* This section was modified from the introduction of Amândio *et al.*, 2016

1.3. *Hox* genes in the secondary body axis

Hox genes were co-opted for different functions many times during vertebrate evolution, though always respecting a collinear strategy (Duboule, 2007). In addition to their ancestral role in AP patterning of the main body axis, specific genes belonging the *Hox* cluster were co-opted for the development/patterning of the secondary body axes, namely the limbs and external genitalia which are recently evolved structures (Wagner, Amemiya and Ruddle, 2003). Specifically, the *Hoxa* and *Hoxd* genes located at the 5' extremity of their respective clusters, are necessary for patterning both the digits and the genital tubercle (GT) (Dollé *et al.*, 1991; Kondo *et al.*, 1997).

Unlike the primary body axis, where *Hox* gene expression is primarily accomplished by elements situated inside the clusters themselves, the regulation of *Hox* genes in novel body structures (e.g., the GT and the limbs) is thought to have occurred primarily by the recruitment of *cis*-regulatory sequences located outside the gene clusters (Spitz *et al.*, 2001; Spitz *et al.*, 2003; Montavon *et al.*, 2011; Andrey *et al.*, 2013; Lonfat *et al.*, 2014). At the *HoxD* cluster these elements are positioned in two regulatory landscapes located on either side of the cluster.

1.3.1 Regulation and transcription of *Hoxd* genes in the limb

The vertebrate limb develops from the lateral plate mesoderm mesenchymal cells. *Hox* genes are essential for the proper development of this structure. There are two phases of gene activation, in the initial phase, there is a sequential transcription, in both time and space, of 3' and central *Hoxd* genes (from *Hoxd1* to *Hoxd11*), controlled by enhancers located in the regulatory landscape 3' to the cluster. These cells will give rise to the future arm and forearm (Zakany and Duboule, 2007; Andrey *et al.*, 2013). The second phase of transcription occurs in the distal part of the limb bud that will eventually give rise to the digits. In this phase, *Hoxd* genes from *Hoxd8* to *Hoxd13* are transcribed. There is strong transcription of *Hoxd13* and sequentially lower levels of *Hoxd12* to *Hoxd8* transcripts (Montavon *et al.*, 2008). Enhancers

located in the regulatory landscape 5' to the cluster control this phase (Zakany and Duboule, 2007; Montavon *et al.*, 2011; Andrey *et al.*, 2013). Both these regulatory landscapes are encompassed within two topologically associating domains (TADs), with the *HoxD* cluster being positioned at the border of these TADs (Figure 2). TADs represent an intermediate level of chromatin organization (Dixon *et al.*, 2012; Nora *et al.*, 2012) and their role in privileging enhancer-promoter interactions will be explored further in section 1.5.3 of the introduction.

1.3.2 Regulation of *Hox* genes in the genital tubercle

Genetic experiments have highlighted the role of the regulatory sequences positioned outside the *HoxD* cluster in controlling *Hoxd* genes expression in GT and other evolutionary novel structures. Upon replacing the entire *HoxD* cluster with a *Hoxd11lacZ* transgene, the reporter gene showed expression in the GT (as well as in the limbs) (Spitz *et al.*, 2001). Moreover, a PAC containing the human *HoxD* cluster randomly integrated in the mouse genome could not recapitulate *Hoxd* genes expression in the GT nor in any other region outside the trunk (Spitz *et al.*, 2001). Together, these results suggest that in the GT sequences located outside the gene cluster are able to drive *Hoxd* genes expression.

The *cis*-elements controlling 5' *Hoxd* genes in the GT (and digits) have been shown to be located 5' to the cluster (Spitz *et al.*, 2003; Spitz *et al.*, 2005), in the centromeric regulatory landscape (C-DOM). A large inversion that splits the *HoxD* cluster in two independent sub-clusters, resulted in the separation of 5' *Hoxd* genes (*Hoxd11* to *Hoxd13*) and its associated 5' regulatory landscape (C-DOM) away from the remaining *Hoxd* genes and the 3' regulatory landscape. In this allelic conformation *Hoxd11* was expressed in the GT and in the digits but no longer in the proximal limb nor in the cecum, in turn *Hoxd10* was no longer expressed in the GT nor the digits. These results further revealed that the enhancers controlling 5' *Hoxd* genes in the GT and digits were positioned 5' to the cluster (in the C-DOM), while the proximal limb, and cecum enhancers were positioned 3' to the cluster (Spitz *et al.*, 2005) (Figure 2).

Further evidence came from experiments done on mice with reallocations of the centromeric regulatory landscape. Upon repositioning of the full C-DOM approximately 3Mb away from the *HoxD* cluster, along with a *Hoxd11lacZ* reporter cassette, the expression of 5' *Hoxd* genes is greatly reduced in the GT (Lonfat, PhD thesis, Tschopp and Duboule, 2011). Furthermore, the inverted reporter cassette was able to drive *lacZ* expression in the GT in a pattern identical to the wildtype (Lonfat, PhD thesis, Tschopp and Duboule, 2011). Sequential deletions spanning the C-DOM further demonstrated the necessity of this region for the

transcription of 5' *Hoxd* genes in both the GT and the limbs (Montavon *et al.*, 2011; Lonfat *et al.*, 2014).

Several enhancer elements positioned over the C-DOM have been identified. Initially, the use of transgenic analysis identified a Global Control Region (GCR), located approximately 200kb 5' to the *HoxD* cluster, and Prox, an element located between the *Lnp* and the *Evx2* genes, with enhancer activity in both the GT and the digits (Spitz *et al.*, 2003; Gonzalez *et al.*, 2007). More recently came the observation that, several enhancer sequences were located further away from the gene cluster, in a regulatory landscape covering approximately 700kb (Figure 2) (Montavon *et al.*, 2011; Lonfat *et al.*, 2014). These elements were able to interact with 5' *Hoxd* genes via long-range chromatin interaction forming a regulatory archipelago controlling gene expression (Montavon *et al.*, 2011; Lonfat *et al.*, 2014). This regulatory landscape (C-DOM), encompassed within a TAD, is shared between digits and GT, however some tissue-specific enhancers-promoter interactions were observed (Lonfat *et al.*, 2014). The use of chromosome conformation capture based techniques, to study the frequency of chromatin interactions at the locus, revealed a DNA region (i.e., GT2) that interacts with *Hoxd13* in GT but not in digits. Accordingly, this region had a strong GT-specific enhancer activity pattern. Conversely, an enhancer-promoter interaction (i.e., island II – *Hoxd13*) was observed in digits but not in GT. Similar differences in the interactions profiles between GT and digits were also observed at the *HoxA* locus (Lonfat *et al.*, 2014). These tissue specific modifications of the chromatin structure highlight the various possibilities of interactions within a pre-fixed TAD architecture to elicit similar transcriptional responses (Lonfat *et al.*, 2014).

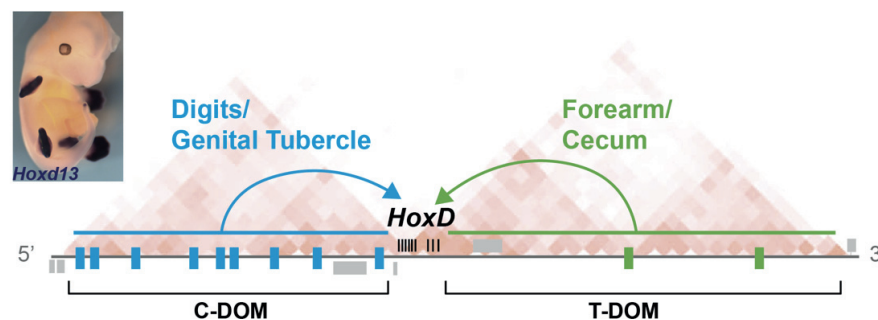


Figure 2: Regulation of *Hoxd* genes in various tissues

Schematic representation of the two regulatory landscapes flanking the *HoxD* cluster and encompassed within two TADs. The *HoxD* cluster is at the border between these two TADs. Enhancers located 3' to the cluster in the telomeric regulatory landscape (T-DOM) are responsible for the regulation of *Hoxd* genes expression in the proximal limb and the cecum (in green). Enhancers located 5' to the cluster in the centromeric regulatory landscape (C-DOM) are responsible for the regulation of *Hoxd* genes

expression in the digits and in the genital tubercle (in blue). WISH of a E12.5 mouse embryo illustrating the pattern of expression of *Hoxd13* in the digits and in the GT (figure adapted from (Andrey *et al.*, 2013; Montavon and Duboule, 2013; Lonfat and Duboule, 2015)).

1.4. Genital tubercle development

1.4.1 Early development

The GT is composed by both surface ectoderm, mesoderm, and cloacal endoderm. The latter will contribute to the formation of the urethral plate epithelium (Seifert *et al.*, 2008). As development progresses, this embryonic tissue will originate the penis in males and the clitoris in females (Cohn, 2011). The initial phase of GT development starts with the appearance of a pair of mesenchymal cell agglomerates, called the genital swellings. These cell agglomerates are positioned on each side of the cloaca membrane, caudal to the posterior hindlimb buds and at the level of the anterior tail (Perriton *et al.*, 2002; Tschopp *et al.*, 2014). At E11.5, the buds will merge to form the GT with the urethral plate epithelium growing out between them (Perriton *et al.*, 2002). At E13.5, the mesenchymal preputial swellings begin to form, in the proximal lateral region of the tubercle, these secondary outgrowths will continue to develop and will form the prepuce, within which the preputial glands are situated (Perriton *et al.*, 2002; Georgas *et al.*, 2015). From this developmental stage on, the GT can be anatomically subdivided into a proximal and a distal region. At E15, the proximal region of the GT is surrounded by the preputial swellings, in the distal region the glans is exposed. These anatomical differences highlight the proximal-distal divisibility of the GT (reviewed in Georgas *et al.*, 2015). Of note, no apparent morphological differences between males and females are observed until this stage (Seifert *et al.*, 2008; Georgas *et al.*, 2015).

1.4.2 Late development (sex specific GT development)

Mouse GT development can be roughly divided in two parts correlating with the action of steroid hormones. The first part, described above, is thought to be independent of steroid hormones. After this initial process, at around E16, the development of the genital tubercle becomes sex-specific (hormonal controlled), eventually generating sexual dimorphic external genitalia (Cohn, 2011). In males, the urethra is internalized by the fusion of the urethral folds. Simultaneously the merging of the preputial swellings originates the prepuce. As development progresses the prepuce envelops the glans. The urethra is closed in the proximal region of the organ and opened in the distal glans (E17.5) (Yamada *et al.*, 2003; Georgas *et al.*, 2015). In

female mice, the preputial swellings envelop the clitoris but without fusing. The urethra is not surrounded by the mesenchyme and as such is not internalized remaining ventral to the clitoris (open at the base of the clitoris). In both sexes, the full development and maturation of the external genitalia is not complete until postnatal stages. Eventually the penis and the clitoris, corresponding to the initial mesenchymal tissue, will develop bone (Georgas *et al.*, 2015).

Several studies have highlighted the critical role of androgens and estrogens in determining sexual differentiation of the genitalia. Disruption of androgen signaling resulted in the de-masculinization of the GT in males. Conversely, exposure to androgen masculinizes female external genitalia (Suzuki *et al.*, 2002). Furthermore, absence of estrogen receptor in mice leads to females with elongated clitoris (Yang *et al.*, 2010). Altogether, a proper balance between estrogen and androgen is required for a correct sexual differentiation of the external genitalia.

1.4.3 GT developmental regulators (genetic pathways involved in GT development)

The evolution of the of two major body appendages, the GT and the limbs, represented a major step in vertebrate evolution by facilitating the colonization of the terrestrial environment. There are similarities between the development of the vertebrate limbs and the GT, with several genetic pathways that are involved in the development of the limbs also having a pivotal role in ensuring proper genital development.

As mentioned above, both endoderm, mesoderm and ectoderm contribute to external genital development, with several pathways involved in mediating the interactions between all three germ layers to ensure proper GT patterning. Classical experiments have demonstrated that removing the distal urethral epithelium (DUE) hampers GT development (reviewed Cohn 2011). Consequently, it was proposed that the DUE would act as a primary signaling center for GT development, similar to the role of the apical ectodermal ridge (AER) during limb development. However, this role of the DUE remains controversial (see below).

Fgf8 is expressed in the DUE in a pattern that resembles its expression domain in the limb bud AER. Initially, at E10.5 *Fgf8* is expressed in the distal-posterior cloacal endoderm. As development progresses from E11.5 to E14.5, this gene is expressed in the DUE. For this reason, *Fgf8* has been proposed as a candidate for initial regulation of GT development. In fact, early experiments have shown that ectopic application of *Fgf8* beads to the mouse GT resulted in mesenchymal gene expression, and consequentially induced outgrowth of the GT suggesting the role of *Fgf8* as an initial activator of GT development (Haraguchi *et al.*, 2000). Surprisingly,

conditional genetic deletion of *Fgf8* from the cloacal endoderm (prior to GT outgrowth) resulted in normal external genital development (Seifert *et al.*, 2009). Furthermore, analysis of *Fgf8* target genes and the distribution of *Fgf8* protein showed that *Fgf8* is not translated in DUE (Seifert *et al.*, 2009). This lack of phenotypic defects in *Fgf8*-null mice, albeit high gene expression in the DUE, challenged the view of the DUE as an organizing center of GT development, arguing divergence between the mechanisms that control limb and GT outgrowth. However, a more recent study reported that ectopic expression of *Fgf8* in the urethra epithelium led to cell over-proliferation and aberrant GT morphology (Lin *et al.*, 2013). Accordingly, conditional deletion from the urethral epithelium of FGF receptors (*Fgfr1* and *Fgfr2*) resulted in smaller GT and abnormal maturation of the urethral epithelium (Lin *et al.*, 2013). These later results are in agreement with a similar role of *Fgf8* in limbs and GT development. As such, the role of the DUE as the GT organizing center is still under debate. Other members of the *Fgf* gene family are also involved in regulating GT development. Deletion of *Fgf10* does not affect initial outgrowth of the GT. However, later in development, the *Fgf10* knockout mice show abnormal opening along the ventral surface of the GT suggesting a role of *Fgf10* in later GT development (Haraguchi *et al.*, 2000).

Sonic hedgehog (Shh) has also been implicated in GT development. This gene is expressed in the whole cloaca epithelium and urethral plate epithelium (Seifert *et al.*, 2009). *Shh* null mice lack external genitalia. Although the initial outgrowth of the GT is started it is not maintained (Haraguchi *et al.*, 2001; Perriton *et al.*, 2002; Seifert *et al.*, 2009). In the absence of *Shh* signaling, *Wnt5a*, *Bmp2*, *Bmp4*, *Fgf8*, and *Fgf10*, are downregulated and apoptosis increases (Haraguchi *et al.*, 2001; Perriton *et al.*, 2002). Using conditional knockout mice for *Shh* it was shown that its signal is not only required during the initial GT budding but also during the androgen-independent GT morphogenesis, highlighting that the primary target tissue of *Shh* signaling is the GT mesenchyme, rather than the urethral plate epithelium (Lin *et al.*, 2009). In GT mesenchyme cells, the expression of *Hoxd13*, *Fgf10*, *Ptch1*, *Gli1*, and *Bmp4* is regulated by *Shh* signaling from the urethral plate epithelium (Haraguchi *et al.*, 2001; Perriton *et al.*, 2002; Lin *et al.*, 2009; Miyagawa, Moon, *et al.*, 2009; Seifert *et al.*, 2009). GT outgrowth is rescued in *Shh* null embryos by β -catenin gain-of-function mutation or exogenous upregulation (Lin *et al.*, 2009; Miyagawa *et al.*, 2009). These data indicates that *Wnt*/ β -catenin signaling in the distal urethral epithelium acts downstream of *Shh* signaling during GT outgrowth.

Wnt/β-catenin signaling plays an important role in early androgen-independent GT development by both maintaining homeostasis of the urethra, and promoting mesenchymal cell proliferation and tissue integrity. Knockout animals for *β-catenin* develop severe hypospadias (i.e., deficiencies in the placement of the urethral opening) in both sexes (Lin *et al.*, 2008). Besides the *Wnt/β-catenin* pathway role in early GT development it was shown that this pathway is also essential in late, androgen dependent, GT development. Accordingly, overexpression of *β-catenin* in embryonic mice resulted in masculinization of female external genitalia (Miyagawa *et al.*, 2009).

The expression of *Bmp* (bone morphogenetic protein) genes has also been implicated in regulating GT outgrowth. *Bmp7* loss led to a halt in cloacal septation, and defects in urethra morphogenesis (Wu *et al.*, 2009).

1.4.4 *Hox* genes in the GT

Hox genes have been shown to be indispensable for the formation of the genital tubercle. During mouse GT development, 5' *Hoxa* and *Hoxd* genes (Dollé *et al.*, 1991; Montavon *et al.*, 2008; Lonfat *et al.*, 2014), and *Hoxc10* and *Hoxc11* are expressed (Peterson *et al.*, 1994; Hostikka and Capecchi, 1998; Hostikka *et al.*, 2009). Both the remaining *Hoxc* genes and *Hoxb* genes are not expressed in this tissue (Lonfat, PhD thesis; Zeltser *et al.*, 1996). Functional studies of the *HoxC* cluster have not reported a role for this group of genes during GT development (Papenbrock *et al.*, 2000; Suemori and Noguchi, 2000; Hostikka *et al.*, 2009).

Both 5' *Hoxa* and *Hoxd* genes have been functionally linked to genitalia patterning. Inactivation of *Hoxd13* alone resulted in a small morphological alteration in the penile bone (baculum), with the proximal segment of this structure being shorter and deformed (Dolle *et al.*, 1993). The simultaneous deletion of *Hoxd11*, *Hoxd12* and *Hoxd13* induced a stronger reduction of this structure. Almost complete agenesis of the baculum was observed upon compound mutations for *Hoxd11*, *Hoxd12*, *Hoxd13*, and *Hoxa13* demonstrating the contribution of *Hoxa13* for the morphology of these structure (Zákány *et al.*, 1997). This data suggests functional cooperation between these four genes (*Hoxd11*, *Hoxd12*, *Hoxd13* and *Hoxa13*) for proper GT development. Of note, *Hoxa13* loss of function alone results in hypospadias as a consequence of a combined loss of expression of *Bmp7* and *Fgf8* in the urethral plate epithelium, leading to loss of apoptosis in this tissue, and a decrease in mesenchymal cell proliferation (Morgan *et al.*, 2003).

The disruption *Hoxa13* and *Hoxd13* genes affects the patterning of the GT in a dose dependent manner. The analysis of compound mutations for *Hoxa13* and *Hoxd13* genes has revealed that while *Hoxa13*^{+/-}/*Hoxd-13*^{-/-} genotype results in abnormal male and female external genitalia, homozygous-null mutants for both *Hoxa13* and *Hoxd13* have complete genital tubercle agenesis lacking external genitalia (Dolle *et al.*, 1993; Kondo *et al.*, 1997; Warot *et al.*, 1997). In humans, mutations in the *Hoxa13* gene are responsible for hand-foot-genital syndrome affecting the distal aspects of the limb and external genitalia. Mutations in *Hoxd13* are responsible for synpolydactyly type II characterized by distal limb malformation and more rarely hypospadias (Quinonez and Innis, 2014).

1.5. Regulatory landscapes

1.5.1 Enhancers

Enhancers were first identified in the simian virus 40 (SV40) over 30 years ago. Canonically they are defined as short noncoding DNA sequences (between 100-1000bp) that are able to regulate transcription of target genes in a manner that is independent of their orientation, location, and distance (reviewed in Long *et al.*, 2016). Since their first discovery, many enhancers have been characterized, and their biochemical and functional properties have been extensively studied. In the last decade, large-scale studies have identified thousands of enhancers that drive gene expression patterns with spatiotemporal specificity (Visel *et al.*, 2006). Furthermore, techniques to assess opening of chromatin (Song and Crawford, 2010; Buenrostro *et al.*, 2013), post-translational histone modifications (reviewed in Shlyueva *et al.*, 2014), or self-transcribing active regulatory region sequencing (STARR-seq) (Arnold *et al.*, 2013), have also been used to identify putative enhancers. Overall, enhancers have been established as a pervasive feature of the genome, vastly out-numbering the ~20,000 protein coding genes (ENCODE Project Consortium, 2012). For example, in the mouse (in a dataset of 19 tissues and cell types), the *cis*-regulatory sequences annotated add up to 11% of the genome (nearly 300,000 murine *cis*-regulatory sequences) (Shen *et al.*, 2012).

Genes can be regulated by multiple distal enhancers. In the case of mammals these enhancers can be positioned at distances that can exceed a megabase. They can act cooperatively or redundantly, with different spatiotemporal activities, permitting a large array of combinatorial gene expressions. Enhancers contribute to a specific and robust regulation of gene expression. It is thought that alterations in these *cis*-elements during evolution played an important role in the acquisition and/or modification of gene expression patterns allowing for

pleiotropic gene function and to the appearance of morphological innovations (reviewed in Long et al. 2016). Each functional enhancer is composed of several transcription factors binding sites. In general, enhancer activation is dependent on the binding of multiples transcription factors. This cooperative behavior is thought to be involved in the nucleosome eviction process, facilitate downstream effector binding and in turn the activation of enhancers (Spitz and Furlong, 2012). Transcription factors recruit co-activators and co-repressors. It is the combinatory signals of all bound factors that determine enhancer activity (reviewed in Long et al., 2016).

1.5.2 Long-range gene regulation, regulatory landscapes, and redundancy

Cis-regulatory landscapes play an important role in regulating developmental gene expression. They contain enhancers, insulators and other architectural elements that can be positioned close to the gene or spread over large genomic distances (reviewed in Spitz, 2016). Several studies suggest that the majority of developmental genes are controlled by multiple enhancers with both overlapping and distinct spatiotemporal activities. Nonetheless, single enhancer can be responsible for controlling the expression pattern of a gene in a specific tissue as exemplified by the case of the *Shh* locus. At this locus, a limb-specific enhancer (ZRS) located approximately 1 megabase away from the *Shh* target promoter, is necessary for the expression of this gene in the limbs and in turn for proper limb development (Sagai et al., 2005).

Although single enhancers can be responsible for controlling expression patterns in a specific tissue, and their deletion can have phenotypic effects, this type of regulation seems to be the exception. Increasing evidence suggests that the expression of most developmental genes is controlled by multiple regulatory elements. Several mechanisms of action have been proposed to classify the cooperation between enhancers (extensively reviewed in Long et al., 2016). For example, the analysis of the enhancer elements controlling α -globin locus (collectively termed super-enhancer) showed that each element acts in an additive and independent fashion (Hay et al., 2016). Another example of additive enhancer behavior comes from a study of the *Indian hedgehog (Ihh)* locus. Here several enhancers with individual tissue specificities work in an additive manner to ensure proper gene regulation (Jerković et al., 2017). These studies highlight some of the many examples of modular/additive behavior of enhancers, however there is increasing examples of non-additive enhancer activity.

The notion of redundancy or “shadow enhancers” within a regulatory region has been used to describe the activity of enhancers whose function partially or completely overlaps

(Hong *et al.*, 2008). A recent study has suggested that enhancer redundancy is a widespread feature of mammalian genomes (Osterwalder *et al.*, 2018). The authors have used CRISPR–Cas9 genome editing technology to individually delete several embryonic enhancers, with no noticeable phenotypic effect. However, upon removal of pairs of limb enhancers regulating the same gene they observed phenotypic differences (Osterwalder *et al.*, 2018). Also, a study in *Drosophila* has shown that redundant or “shadow enhancer” are present in most regulatory landscapes (Cannavò *et al.*, 2015). In some cases, however, shadow enhancers may show redundant/overlapping activity patterns but may not be necessarily functionally identical, as demonstrated in *Drosophila* for the *shavenbaby* (*svb*) locus (Frankel *et al.*, 2010). In this locus enhancers that seem to act redundantly under normal environmental conditions were shown to be essential under genetic or environmental stress, conferring robustness against environmental or genetic variability (Frankel *et al.*, 2010). Furthermore, a meticulous dissection of the *Sox9* gonadal regulatory elements found that, although several enhancers were redundant, one, *Enh13*, was essential to initiate mouse testis development. Alone, deleting this enhancer was sufficient to phenocopy the loss of *Sox9* in the gonads (Gonen *et al.*, 2018). This result is an illustration that different enhancers, within a regulatory landscape, don’t all act in the same way.

In the case of *Hox* genes, both *HoxA* and *HoxD* clusters digit regulation is controlled by several enhancers with complementary specificities (Montavon *et al.*, 2011; Berlivet *et al.*, 2013). Partial deletions of the *HoxD* regulatory landscape resulted in alterations in the expression of these genes. The combined action of all enhancers present in this regulatory landscape is responsible for the final transcription pattern of the genes (Montavon *et al.*, 2011). Conversely, *HoxD* transcription regulation in the GT also relies upon several enhancers positioned in the same regulatory landscape, however tissue-specific enhancer-promoter interactions were scored (Lonfat *et al.*, 2014). The studies mentioned above illustrate the complexity of the *cis*-regulatory interactions, and describe several ways enhancers can act to regulate gene expression.

1.5.3 Topologically associating domains (TADs)

An important characteristic of enhancers is their capacity to regulate transcription over long genomic distances. The development of chromosome conformation capture (3C) based technologies, that have been used to determine the physical proximity between DNA sequences based on their probability to be crosslinked together, have provided key insights supporting the

idea that direct physical contact or proximity is an essential aspect of enhancer function (reviewed in de Laat and Duboule, 2013). Enhancer-promoter interactions have now been identified in many loci, however they are not always associated with transcription, suggesting that some loci display a preformed enhancer-promoter spatial conformation which is poised for transcription (reviewed in de Laat & Duboule 2013). 3C based techniques have shown that, in human fibroblasts cell-line, 54% of the active promoters were interacting with on average 5 distal enhancers, while the other 46% of the genes did not show any specific distal interactions (Jin *et al.*, 2013).

Most promoter-enhancer interactions occur within topologically associating domains (TADs) (Dixon *et al.* 2012; Nora *et al.* 2012), which were described as domains of privileged DNA interactions (i.e., regions within the same TAD interact with each other at a higher frequency than with regions located elsewhere in the genome). TADs are thought to be conserved between cell types and across species (Dixon *et al.*, 2012; Rao *et al.*, 2014), however variability is observed in intra-TAD domains (Phillips-Cremins *et al.*, 2013). Importantly, this compartmentalization of the genome into TADs seems to limit the activity of *cis*-regulatory elements to genes that are in the same TAD creating a hub where enhancer-promoter interactions are favored (de Laat and Duboule, 2013). Furthermore, when TAD boundaries are disrupted leading to the fusion of neighboring TADs, genes are exposed to enhancers from which they were insulated before. These topological abnormalities may lead to ectopic or aberrant gene expression and consequentially to congenital disorders as well as cancer (reviewed in Spielmann *et al.*, 2018).

As mentioned above, the *HoxD* cluster is at the border between two TADs, each containing distinct enhancer-rich regulatory landscapes. Different genes in the cluster respond to distinct enhancers located in each of the flanking TADs. Furthermore, both of these TADs include enhancers of unrelated functions, such as those regulating *Hoxd* genes in the cecum and the proximal forelimb (located in the TAD 3' to the cluster) or in the GT and the digits (located in the TAD 5' to the cluster). This suggests that TADs at the *HoxD* cluster, as in other *loci*, act as hubs where novel enhancer sequences have evolved (Lonfat and Duboule, 2015).

1.5.4 CTCF and TAD boundaries

A great proportion of TAD boundaries are enriched for both CTCF, a DNA-binding protein containing an 11 zinc-finger domain, and Cohesin, a ring-like shaped multiprotein complex composed of SMC1-SMC3 heterodimer, RAD21, and SA1/2 subunits, that is able to

topologically imprison DNA. Both these proteins have also been largely implicated in chromatin loop formation within TADs (reviewed in Bonev and Cavalli, 2016). Loops are formed primarily between convergent CTCF sites (i.e., sites with consensus CTCF motifs oriented towards each other) (Rao *et al.*, 2014), CRISPR-Cas9 genome editing studies have confirmed that this specific orientation is required for CTCF mediated DNA looping (de Wit *et al.*, 2015; Guo *et al.*, 2015). The use of the auxin-mediated protein degradation system, to induce rapid degradation of CTCF (Nora *et al.*, 2017) and Cohesin (Rao *et al.*, 2017; Schwarzer *et al.*, 2017), resulted in a loss of TAD structure. However, gene expression was only partially affected by these losses indicating that other mechanisms should be at play to ensure proper gene expression.

Deletion of specific CTCF-binding sites within the *HoxA* and *HoxC* clusters resulted in ectopic transcriptional activation of these genes (Narendra *et al.*, 2015), ultimately causing phenotype alterations (Narendra *et al.*, 2016). At the *HoxD* cluster the TAD boundary resides between *Hoxd11* and *Hoxd12* in the distal forelimb. The robustness of this TAD boundary was assessed by sequential deletions of border elements. Interestingly, merging of the two TADs was only observed upon deletion of a 400kb DNA sequence that includes the entire gene cluster, highlighting the robust nature of the boundary at the *HoxD* locus (Rodríguez-carballo *et al.*, 2017).

2. Scope of the thesis

The accurate and precise regulation of *Hox* genes expression during development is of paramount importance to ensure proper embryogenesis and ultimately establish tissue structure and function. In evolutionary novel structures (e.g., the GT and the limbs) the regulation of *Hoxd* genes is primarily accomplished by *cis*-elements positioned in two regulatory landscapes flanking the cluster and encompassed within two TADs (Spitz *et al.*, 2001; Spitz *et al.*, 2003; Montavon *et al.*, 2011; Andrey *et al.*, 2013; Lonfat *et al.*, 2014). Extensive work, has cemented the importance of these regulatory landscapes in controlling *Hoxd* genes expression. However, both the dynamic behavior of such regulatory landscapes, and the functional contribution of specific *cis*-elements during development remained to be established.

The aim of this work was to understand the complex functional organization of these regulatory landscapes to control transcription of *Hoxd* genes during development. To this end we used GT development as a model system. In GT, 5' *Hoxd* genes regulation is primarily accomplished by elements located 5' to the gene cluster, in a regulatory landscape (C-DOM) extending approximately 700kb. We used a combination of molecular and genetic approaches to dissect the regulatory potential of the C-DOM during development. This work will contribute to a larger understanding of long-range gene regulation in complex organisms.

3. Results

3.1 *Hox* genes expression and epigenetic profile during GT development

Hox genes are essential for proper development of the genital tubercle (GT) (Dolle *et al.*, 1993; Kondo *et al.*, 1997). Since GT differentiation is sex-specific in later stages of development (see introduction), we decided to assess if there were any differences in the expression of *Hox* genes between females and males in this tissue. To this aim, we micro-dissected GT from two females and two males at embryonic day (E) 16.5 (in the transition to the sex-specific stage) and at E18.5 (hormonal control period of sexual differentiation). We extracted RNA from those tissues and performed RNA-seq. We conducted differential gene expression analyses between females and males at the two aforementioned stages of GT development to assess potential differences in *Hox* genes expression. We considered as significant an absolute expression fold change greater than 1.5 and set the false discovery rate (FDR) threshold at 5%. By using these parameters, we observed a total of 182 protein-coding genes differentially expressed (114 genes male-biased; 68 genes female-biased) between females and males at E16.5 (Figure 3A). At later stage of GT development (E18.5) a total of 765 protein-coding genes were differentially expressed (332 genes male-biased; 433 genes female-biased) between females and males (Figure 3B). These analyses revealed no changes in *Hox* genes expression levels between females and males for any of the developmental stages analyzed (Figure 3A and 3B).

For subsequent analysis we used E17.5 as a proxy for late GT development (see below) and performed RT-qPCR to assess the transcript levels of a subset of *Hoxd* genes (*Hoxd13*, *Hoxd12*, *Hoxd11*) in both females and males at this developmental stage. Using this method, no difference between sexes were detected regarding the expression of 5' *Hoxd* genes (Figure 3C). Consequently, we included both females and males in the following experiments.

To obtain a global idea of the expression profile of *Hox* genes during GT development we quantified their expression levels using RNA-seq. We analyzed data from three different stages of GT embryonic development, E12.5 (data from Amandio *et al.*, 2016), E16.5 and E18.5, corresponding to an early stage (independent of steroid hormones), intermediate stage (in the transition to the sex specific stage) and a late stage (corresponding to a hormonal control period of sexual differentiation) respectively.

RNA-seq data revealed that mainly 5' *Hoxa* (*Hoxa7* to *Hoxa13*) and 5' *Hoxd* (*Hoxd8* to *Hoxd13*) genes are expressed in all analyzed stages of GT development (Figure 3D-G). Furthermore, with the exception of *Hoxc11* and *Hoxc10*, only minimal transcript levels of the

HoxC and *HoxB* clusters were detected (Figure 3E-F), consistent with previous findings (Dollé *et al.*, 1991; Hostikka and Capecchi, 1998; Montavon *et al.*, 2008; Hostikka *et al.*, 2009). Overall, we observed a global decrease in 5' *Hox* transcript levels during GT development, in particular for *Hoxd12* and *Hoxd13* between E12.5 and E16.5 (Figure 3G).

To further characterize the pattern of *Hoxd* transcripts during GT development, we included additional developmental stages. We micro-dissected cloaca at E10.5 (an embryonic tissue that will contribute to the formation of the GT) as well as GT at E12.3, E13.5, E15.5, E16.5, E17.5 and E18.5, and performed RT-qPCR for *Hoxd13* (Figure 3H). This dataset revealed that *Hoxd13* is already expressed in the cloaca at E10.5. As development progresses, we observed a significant increase of *Hoxd13* transcripts between the cloaca at E10.5 and the GT at E12.5 ($p < 0.0001$). The levels of these transcripts remained constant in GT between E12.5 and E13.5, while they were significantly reduced between E13.5 and E15.5 ($p < 0.0001$). After E15.5, transcript levels continue to decrease but to a lesser extent (significant between E16.5 and E18.5; $p = 0.038$) confirming the RNA-seq results for *Hoxd13* (Figure 3G).

We compared chromatin accessibility and histone modifications in selected developmental stages to correlate with transcript levels. We used the cloaca region at E10.5 (prior to GT formation), GT at E13.5 (an early stage of GT development), and GT at E17.5 (late stage of GT development) and performed ATAC-seq and ChIP-seq for both H3K27ac and H3K27me3 chromatin marks. At the *HoxD* locus we observed that prior to GT formation, at E10.5, all genes in the cluster and *Evx2* were accessible, as seen by ATAC-seq (Figure 3I, track 1). H3K27ac marks accumulate in both promoters and gene bodies of *Hoxd9* to *Evx2*, and at the promoters of *Hoxd3* and *Hoxd1* (Figure 3F, track 2). However, H3K27me3 repressive marks were still detected, in particular over 5' *Hoxd* genes and *Evx2* (Figure 3I, track 3, gray area), suggesting the presence of a subpopulation of non-expressing cells in this tissue. As development progresses, at E13.5, 5' *Hoxd* genes and *Evx2* lost H3K27me3 repressive marks and gained H3K27ac positive marks (up to the level of *Hoxd11*) (Figure 3I track 5). In contrast, 3' genes lost H3K27ac marks and gained H3K27me3 (Figure 3I, track 6). Accordingly, chromatin accessibility at 3' *Hoxd* genes was reduced, as seen by the loss of ATAC-seq peaks (Figure 3I, track 4). At this developmental stage, there was a clear separation of the cluster into two distinct epigenetic domains, with an active domain at 5' and an inactive domain at 3', reminiscent of what was observed in distal forelimb buds (Andrey *et al.*, 2013). Altogether, these results suggest that at E13.5 the majority of cells analyzed expressed primarily 5' *Hoxd* genes. At E17.5, 5' *Hoxd* genes started to gain H3K27me3 marks up to the level of *Evx2* (Figure 3I, gray box in bottom track), indicating that at least some of the cells in this tissue no

longer expressed these genes. This data confirms the reduction of 5' *Hoxd* genes mRNA levels in later stages of development observed by both RNA-seq and RT-qPCR.

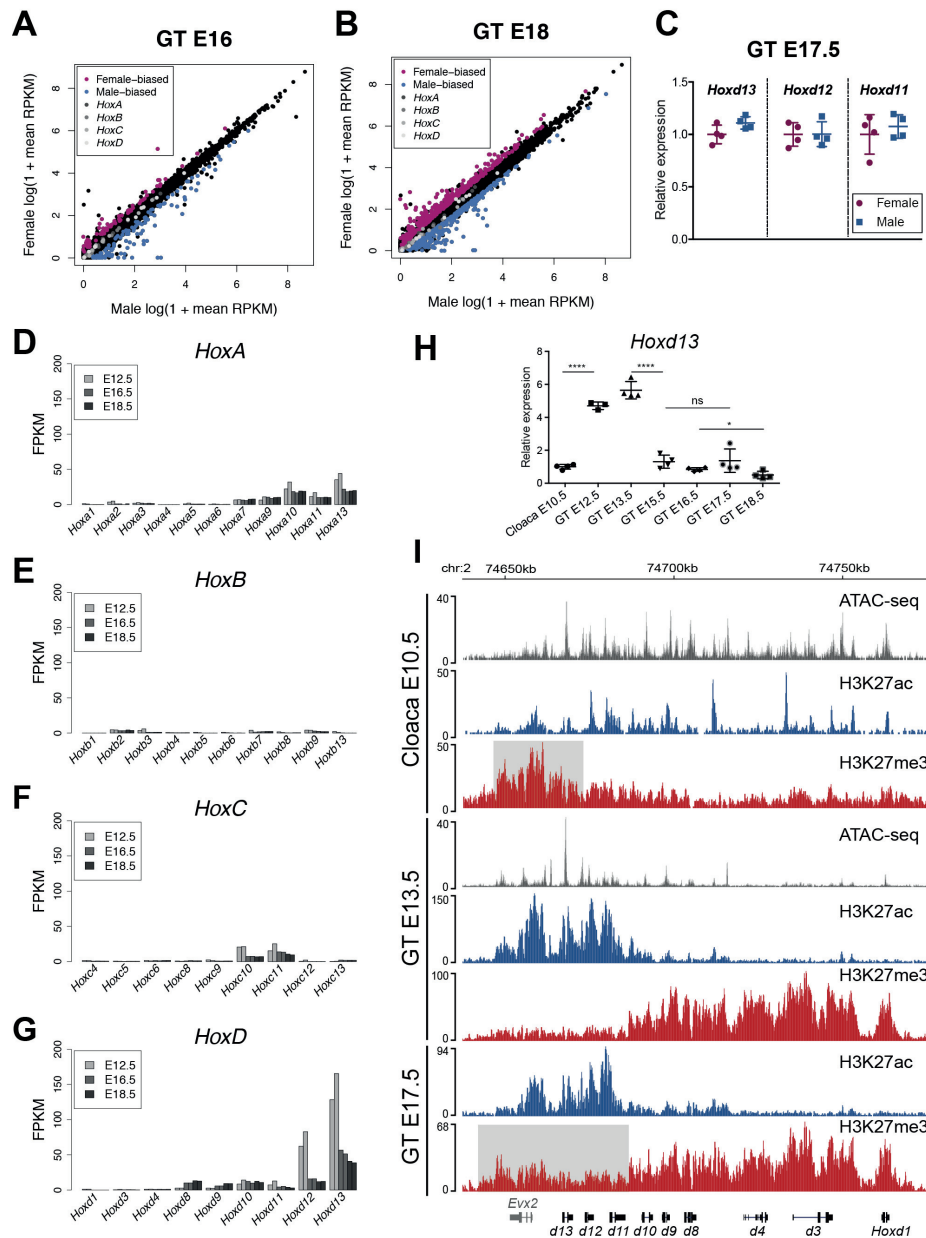


Figure 3: Characterization of *Hox* genes expression in developing GT

A, B) Scatterplots show the correlation of log(1+RPKM) values of all expressed protein coding genes between females and males in GT at E16 and in GT at E18. Female biased genes (absolute fold change > 1.5 and FDR < 0.05) are plotted in pink and male biased genes (absolute fold change > 1.5 and FDR < 0.05) are plotted in blue. *Hox* genes are plotted in gray. **C**) Dot-plot shows relative expression obtained by RT-qPCR of the indicated genes. Values plotted indicate the ratio of expression using female as reference for each gene. Bars indicate mean with SD, 4 individual biological replicates of female (pink dots) and male (blue dots) GT at E17.5 were used. No significant difference between sexes was observed for any of the genes analyzed (p-value > 0.05 for t-test). **D-G**) Quantification of *Hox* genes

expression by RNA-seq (FPKM values) in GT at E12.4, E16.5 and E18.5. **H)** RT-qPCR of *Hoxd13* in different stages of GT development, values plotted indicate the ratio of expression using cloaca as a reference (n>3 biological replicates for each sample). We used an unpaired two-tailed t-test to evaluate the putative significant changes in *Hoxd13* expression. Bars indicate mean with SD, *p<0.05, ****p<0.0001. **I)** ATAC-seq (gray) and ChIP-seq profiles for H3K27ac (blue) and H3K27me3 (red) show the coverage at the *HoxD* locus in wildtype cloaca E10.5, GT E13.5 and GT E17.5. Coordinates (mm10): chr2:74637433-74775728. Gray box in 3rd track indicate enrichment of H3K27me3 at 5' *Hoxd* genes in the cloaca, gray box in bottom track indicate gain of H3K27me3 at 5' *Hoxd* genes in GT E17.5 when compared to E13.5.

3.2 Characterization of the *HoxD* regulatory landscapes during GT development

It was previously proposed that the functional recruitment of *Hox* genes during evolution of structures, such as the limbs and the GT, occurred preferentially through the use of multiple enhancer sequences located outside the gene cluster and distributed into two TADs (Figure 4A) (Spitz *et al.*, 2001; Spitz *et al.*, 2003; Montavon *et al.*, 2011; Andrey *et al.*, 2013; Lonfat *et al.*, 2014). In the GT, long-range transcriptional regulation of 5' *Hoxd* genes is primarily accomplished by the centromeric regulatory landscape (hereafter C-DOM) (Spitz *et al.*, 2001, Spitz *et al.*, 2005; Spitz *et al.*, 2003; Gonzalez *et al.*, 2007; Lonfat *et al.*, 2014). To further explore the functional dynamics of this regulatory landscape during GT development, we examined in detail our ATAC-seq and ChIP-seq datasets (see above) in the two regulatory landscapes flanking the *HoxD* cluster (C-DOM and T-DOM).

In the cloaca, we observed that both these regions displayed several open chromatin sites (as seen by ATAC-seq), some of which corresponded to previously identified GT enhancers (e.g., GT1 and GT2) (Lonfat *et al.*, 2014) (Figure 4B, red arrowheads in track 1). Interestingly, Prox a known GT and limb enhancer was not yet accessible (Figure 4B, red arrow in track 1). Nonetheless, we were unable to detect any clear H3K27ac enrichment, a mark which correlates with active regulatory regions (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011) suggesting that long-range enhancers were not yet active at this early stage (Figure 4B, track 3). As mentioned before, 5' *Hoxd* genes were covered by the Polycomb-associated mark H3K27me3, however this mark was not detected in the two flanking regulatory landscapes (Figure 4B, track 6).

At E13.5, the C-DOM became progressively active, both chromatin accessibility and H3K27ac marks were scored over previously characterized enhancers within this region (e.g., Prox and GT2) (Gonzalez *et al.*, 2007; Montavon *et al.*, 2011; Lonfat *et al.*, 2014) (Figure 4B, red arrowheads in track 2 and black arrows in track 4). In contrast, the telomeric regulatory landscape (T-DOM) remained largely depleted of active H3K27ac marks, while enriched for H3K27me3 repressive marks (Figure 4B, tracks 4 and 7). This enrichment was most prominent

between *Mtx2* and the conserved non-coding sequence CS39 (Andrey *et al.*, 2013) (Figure 4B, gray box in track 7). At E17.5, the C-DOM became inactive losing H3K27ac marks and accumulating some H3K27me3 marks, preferentially at the level of island I and island V (Figure 4B, arrowheads in bottom track). As expected from the absence of 3' *Hoxd* genes transcripts in GT (Figure 3G), the repression of T-DOM activity was maintained as seen by the preservation of H3K27me3 marks over this region (Figure 4B, bottom track).

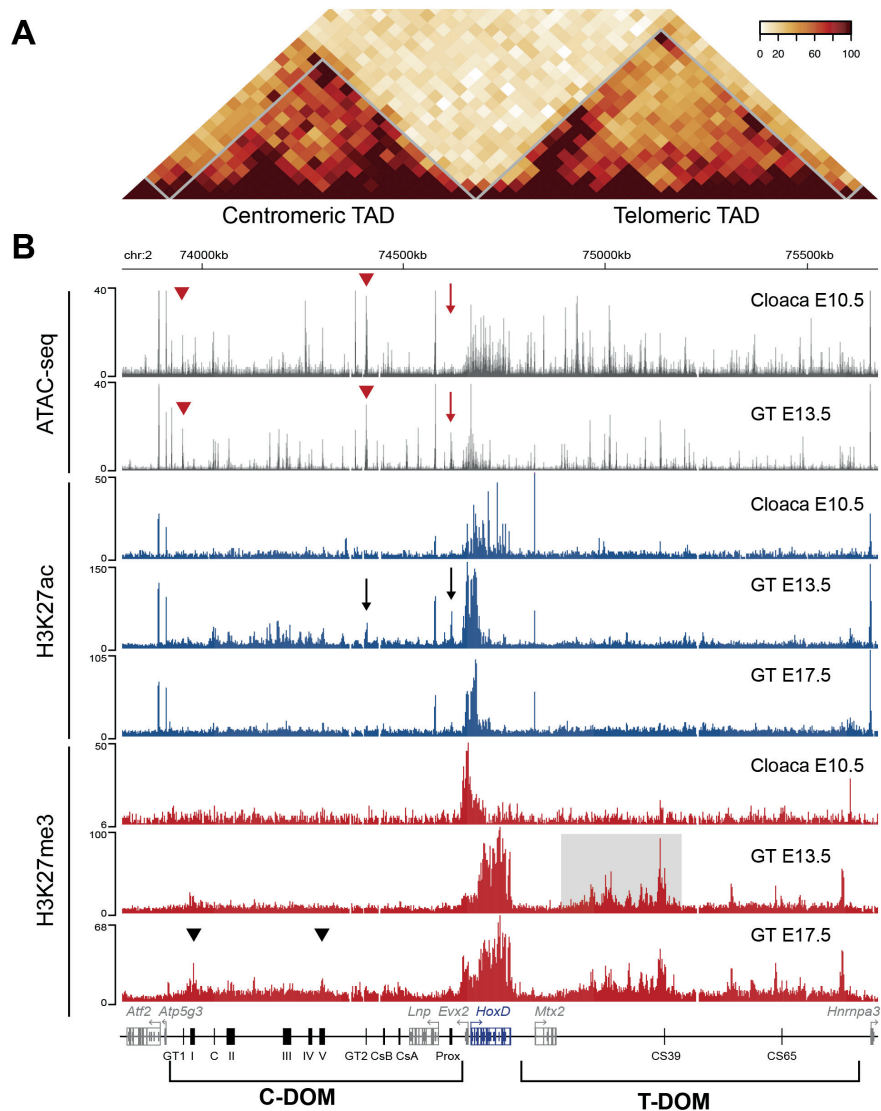


Figure 4: Characterization of the *HoxD* regulatory landscapes during GT development
A) Hi-C heat map of E12.5 mouse distal limb cells (Rodríguez-carballo *et al.*, 2017) centered at the *HoxD* locus and covering the two regulatory gene deserts (C-DOM and T-DOM). Coordinates (mm10): chr2:73800000-75680000. **B)** ATAC-seq profiles (gray) of the *HoxD* locus and its regulatory gene deserts in cloaca at E10.5 and GT at E13.5. The cloaca track (upper gray track) represents the average of two biological replicates and the GT E13.5 track (lower gray track) represents the average of three biological replicates. Red arrowheads indicate GT1 and GT2 enhancers, red arrows indicate the Prox

enhancer element. CHIP-seq profiles for H3K27ac (blue) and H3K27me3 (red) show the coverage at the *HoxD* locus and its regulatory gene deserts in cloaca at E10.5, GT at E13.5, and GT at E17.5. Black arrows indicate H3K27ac enrichment in the Prox and GT2 enhancer elements in GT at E13.5. Gray box indicates H3K27me3 enrichment in the telomeric gene desert in GT at E13.5. Black arrowheads indicate H3K27me3 enrichment in the centromeric gene desert at the level of island I and island V in GT at E17.5. Coordinates (mm10) for all plots: chr2:73801753-75678206.

3.3 Structural organization of the *HoxD* cluster and the C-DOM during GT development

We sought to understand the dynamics of the physical interactions between 5' *Hoxd* genes and their regulatory landscapes during the development of this embryonic structure. We used circularized chromosome conformation capture combined with high-throughput sequencing (4C-seq) to study the frequency of chromatin interactions and how their pattern is established during development. We micro-dissected cloaca at E10.5 and GT at E12.5, E13.5, E15.5 and E17.5 and focused on *Hoxd13* as a viewpoint, as it is the *Hoxd* gene with the highest mRNA levels in these tissues and because its disruption lead to morphological alterations in external genitals (Dolle *et al.*, 1993). As a control for these experiments we generated 4C contact maps for the forebrain at E12.5, an embryonic tissue that does not express any *Hoxd* genes (Figure 5). As a starting point of our developmental temporal series, we used a mouse embryonic stem cells (mESC) dataset (Noordermeer *et al.*, 2014), reasoning that these cells would reflect the ground state 3D architecture of the *HoxD* cluster (Noordermeer *et al.*, 2014). As expected, in all samples analyzed, we observed that *Hoxd13* contacted preferentially the C-DOM in comparison to the T-DOM, independently of its transcriptional status (see percentages in Figure 5). The majority of contacts observed in mESC were scored within the gene cluster itself (Figure 5, track 1) (Noordermeer *et al.*, 2014). Additional contacts were observed in the flanking gene deserts, in particular with island V and island II in the C-DOM and with CS39 in T-DOM (Figure 5, track 1). The CS39 region contains the transcription start sites of the two opposite lncRNAs *Hotdog* and *Twin of Hotdog* (Delpretti *et al.*, 2013). This 3D architecture was globally maintained in the forebrain (Figure 5, track 1 and bottom track), despite some reported differences (Noordermeer *et al.*, 2014). Overall, it appears that in the absence of transcription, the chromatin structure over the *Hoxd* gene cluster was generally conserved, with discrete contacts established between *Hoxd13* and island V and II.

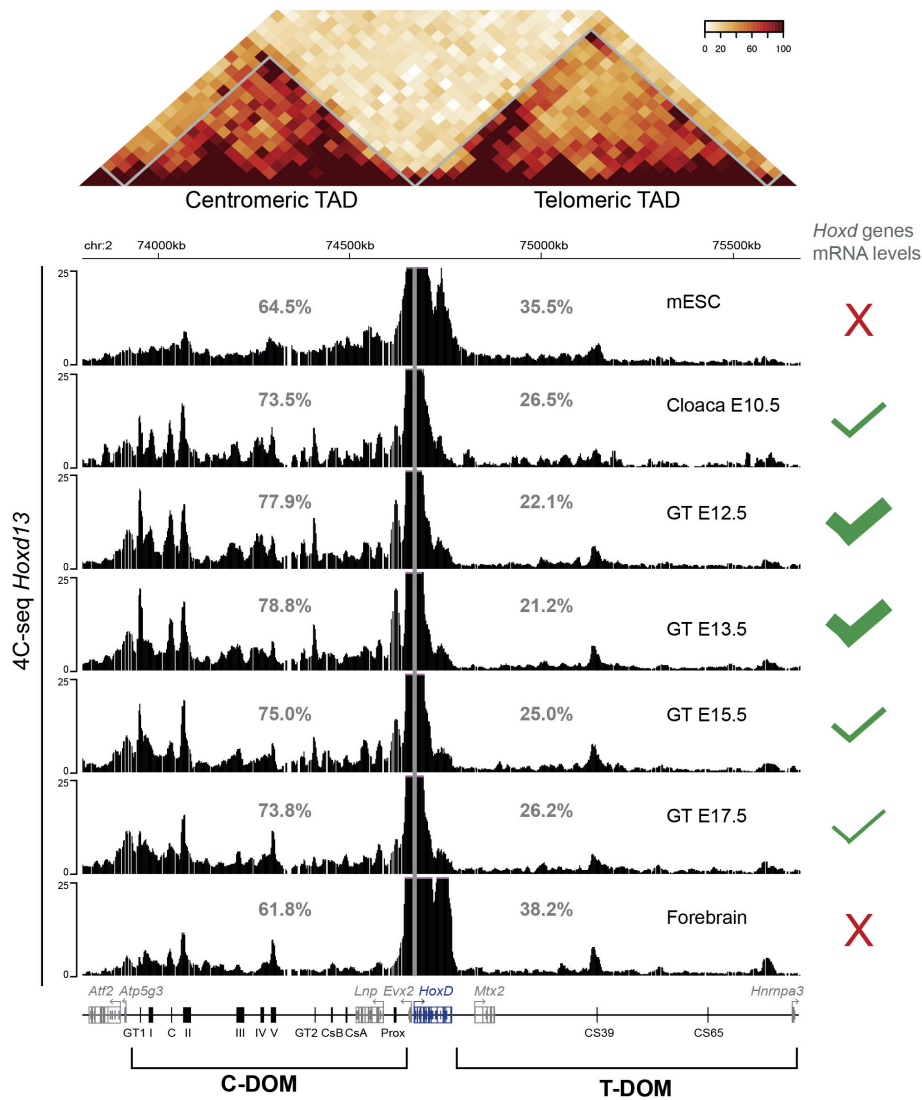


Figure 5: Analysis of chromatin architecture at the *HoxD* locus during GT development
 Hi-C heat map of E12.5 mouse distal limb cells (Rodríguez-carballo *et al.*, 2017) centered at the *HoxD* cluster and covering the two regulatory gene deserts (coordinates (mm10): chr2: 73800000-75680000). 4C-seq contact profiles from the *Hoxd13* viewpoint (gray line) at the *HoxD* cluster and its regulatory gene deserts. 4C-seq profiles from mESC (Noordermeer *et al.*, 2014) (upper track), Cloaca, GT at E12.5, GT at E13.5, GT at E15.5, GT at E17.5 and forebrain (bottom track) are represented. Coordinates (mm10): chr2:73801753-75678206. The percentages in gray represent the ratios of contacts scored in either regulatory gene desert (i.e., C-DOM and T-DOM). A schematic representation of the expression of *Hoxd* genes in each tissue is represented in green (if expressed) or red (if not expressed) on the side of the corresponding 4C track.

Upon transcriptional activation, the frequencies of contacts with the C-DOM increased and interactions between *Hoxd13* and previously characterized enhancers (e.g., Prox, GT2) (Gonzalez *et al.*, 2007; Lonfat *et al.*, 2014) were observed (Figure 5 and 6A). By quantifying the percentage of interactions between *Hoxd13* and the whole C-DOM, we observed a 22%

increase in overall contacts in this regulatory region between mESC and the cloaca at E10.5 (Figure 6A, track 2). Our 4C-seq cloaca dataset showed that the chromatin architecture of the C-DOM predated the embryonic emergence of the GT (Figure 6A, track 2). In the cloaca at E10.5, we observed contacts between the enhancer regions and *Hoxd13*, however their frequency was lower than what was observed in the early GT developmental stages (E12.5 or E13.5). We observed an additional 22% of contacts in the C-DOM of GT at E12.5 when compared to cloaca (Figure 6A, tracks 2 and 3), and a difference of 35% when comparing cloaca and GT at E13.5 (Figure 6A, track 4). As development progresses, the overall contacts established with the C-DOM were progressively lost (Figure 6A, tracks 5 and 6). From E13.5 to E17.5, there was a 44% decrease in all contacts within the C-DOM. Qualitatively, the GT centromeric landscape at E17.5 became similar to either the forebrain or the mESC landscape, with loss in contacts with specific enhancers (e.g., GT1 and GT2, see below) (Figure 6A).

We focused our analysis on specific regions and quantified the percentage of fragments in each island by using mESC as a reference (Figure 6B). We observed that the frequency of contacts between island II and island V remained fairly constant in all samples analyzed. In contrast, the contacts between *Hoxd13* and either Prox, GT2 or GT1 increased from mESC to GT at E13.5. From GT at E13.5 to GT at E17.5, we observed a decrease in the frequency of contacts, correlating with a decrease in *Hoxd13* transcript levels. Forebrain cells (where *Hox* genes are not expressed) showed the lowest values of interactions with Prox and GT2 (Figure 6B). This data suggests that, as transcription decreased, some contacts established with the centromeric landscape were lost whereas others were maintained (i.e., island V and island II), regardless of the transcriptional activity of *Hoxd13* (Figure 6B). This suggests that at the time when transcription is switched off, the locus goes back to the pre-organized chromatin backbone that was previously described for the tissues that do not express *Hox* genes (Noordermeer *et al.*, 2014). Of interest, these “constitutive” interactions overlapped with strong binding of both CTCF and RAD21 to the same sites (see below, Figure 7A). Both CTCF, an insulator element, and the Cohesin complex of which RAD21 is a component, are involved in facilitating enhancer-promoter contacts by DNA-looping (see Ong and Corces, 2014).

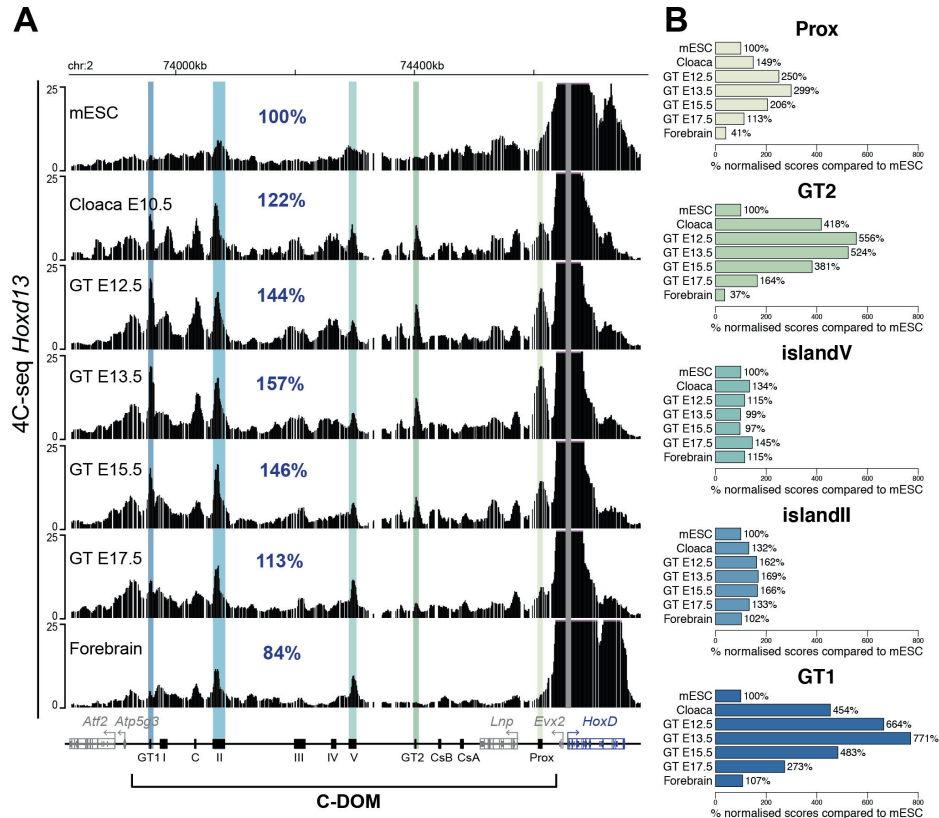


Figure 6: Structural organization of the C-DOM during GT development

A) 4C-seq contact profiles from the *Hoxd13* viewpoint (gray line) at the *HoxD* cluster and C-DOM. 4C-seq profiles from mESC (Noordermeer *et al.*, 2014)(upper track), Cloaca, GT E12.5, GT E13.5, GT E15.5, GT E17.5 and forebrain (bottom track) are represented. Coordinates (mm10): chr2:73815520-74792376. The percentages of the sum of scores in the C-DOM using mESC as a reference are represented in blue. **B)** Bar plots show the percentage of the sum of scores in selected regulatory regions using mESC as a reference.

We validated the interactions identified with *Hoxd13* by using the contacted DNA itself as a viewpoint for the 4C-seq experiments. These experiments confirmed and further detailed the previously identified interactions. We used GT at E12.5 as a proxy for an “active” landscape and prepared a multiplexed library containing viewpoints corresponding to seven previously identified *Hoxd13* interacting DNA regions (Figure 7, island II, III, IV, and V identified in Montavon *et al.*, 2011), (Figure 7, GT1, C, and GT2 identified in Lonfat *et al.*, 2014). Using this approach, we observed that all islands were able to contact *Hoxd13* to some degree (Figure 7A). We used a peak calling algorithm (PeakC) (de Wit *et al.*, 2015) to systematically identify chromatin loops in our 4C-seq datasets. In all datasets, the algorithm was able to call peaks in the *Hoxd13* region (Figure 7A, red dashed line). Of note when using island III as a viewpoint, the algorithm was able to call a peak close to *Hoxd13* but the frequency of contacts in this region was not as prominent as it was when using other islands as viewpoints. We also observed

enhancer-enhancer (e.g., GT2-C), enhancer-CTCF (e.g., GT2-island II) and CTCF-CTCF (e.g., island V-island II) interactions, as seen by the ability of each island to contact other islands (Figure 7A). Upon careful analysis, we observed that all islands contacted with higher frequency the *Evx2* region and its associated CTCF sites. These contacts were stronger than those established with the *Hoxd13* gene body (Figure 7B, gray box).

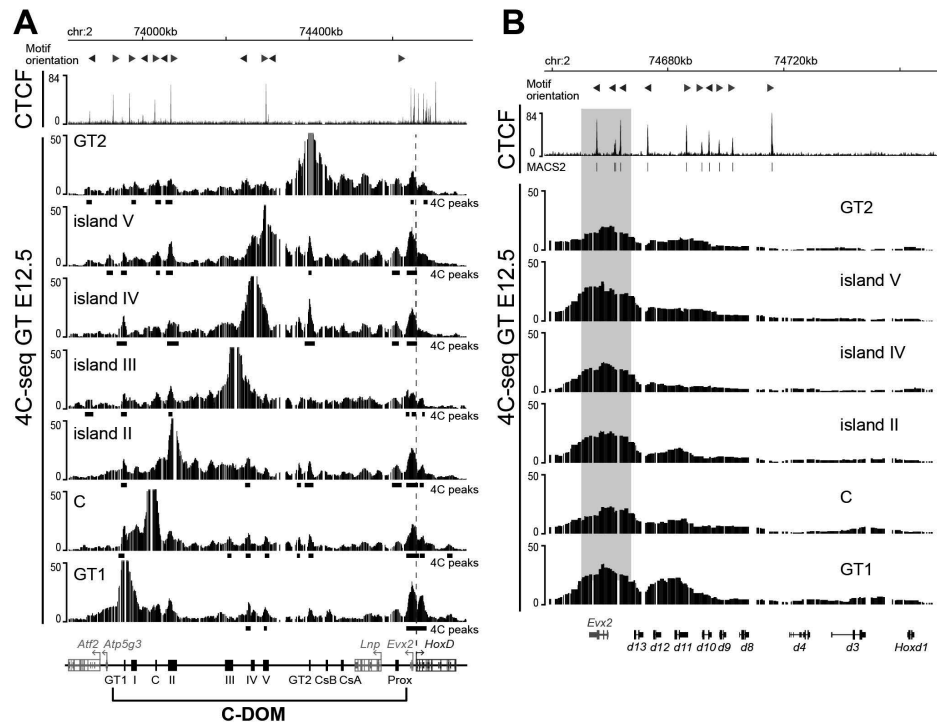


Figure 7: Validation of the C-DOM structural organization

A) Upper panel shows CTCF ChIP-seq profile of GT E13.5 at the *HoxD* cluster and C-DOM. Red and blue arrowheads show forward and reverse CTCF motif orientation respectively. Lower panels show 4C-seq contact profiles at the *HoxD* locus and C-DOM of GT at E12.5, viewpoints are indicated in each track. Below each 4C-seq track, peaks identified by the peak calling algorithm (PeakC) are shown by black boxes (“4C peaks”). Coordinates (mm10): chr2:73815520-74792376. Red dashed line indicates the *Hoxd13* region. **B)** Upper panel shows a zoom in CTCF ChIP-seq profile of GT at E13.5 at the *HoxD* cluster. Coordinates (mm10): chr2:74637433-74775728. Red and blue arrowheads show forward and reverse CTCF motif orientation respectively. Lower panels show 4C-seq contact profiles of GT E12.5 at the *HoxD* locus, viewpoints are indicated in the right top corner of each respective track. Gray box indicates the region of interactions with *Evx2*.

To further explore the organization of the *HoxD* regulatory landscape during GT development, we selected a subset of *Hoxd13* interacting regions, which displayed changes in contact frequencies with *Hoxd13* during GT development (GT2, island IV, GT1), as well as a “constitutive” contact maintained even in the absence of gene transcription (island V). We prepared 4C-seq libraries for GT at developmental stages E12.5, E13.5, E15.5, and E17.5, and for forebrain at E12.5 as a negative control. We quantified the percentages of fragments in the

Evx2 region, where all islands seemed to preferential contact (see above, Figure 7B), using GT at E12.5 as a reference (Figure 8B). By using these viewpoints, we confirmed that, with the exception of island V, where contact frequency was roughly maintained, all viewpoints showed a decrease in contacts with the region close to *Hoxd13* from E12.5 to E17.5 (Figure 8A,B), confirming our previous analysis from the *Hoxd13* viewpoint itself (Figure 6B). Overall, as GT development progressed, a decrease in contacts between *Hoxd13* and its previously characterized enhancers was scored, independently of which viewpoint was considered (Figure 6, 8A,B).

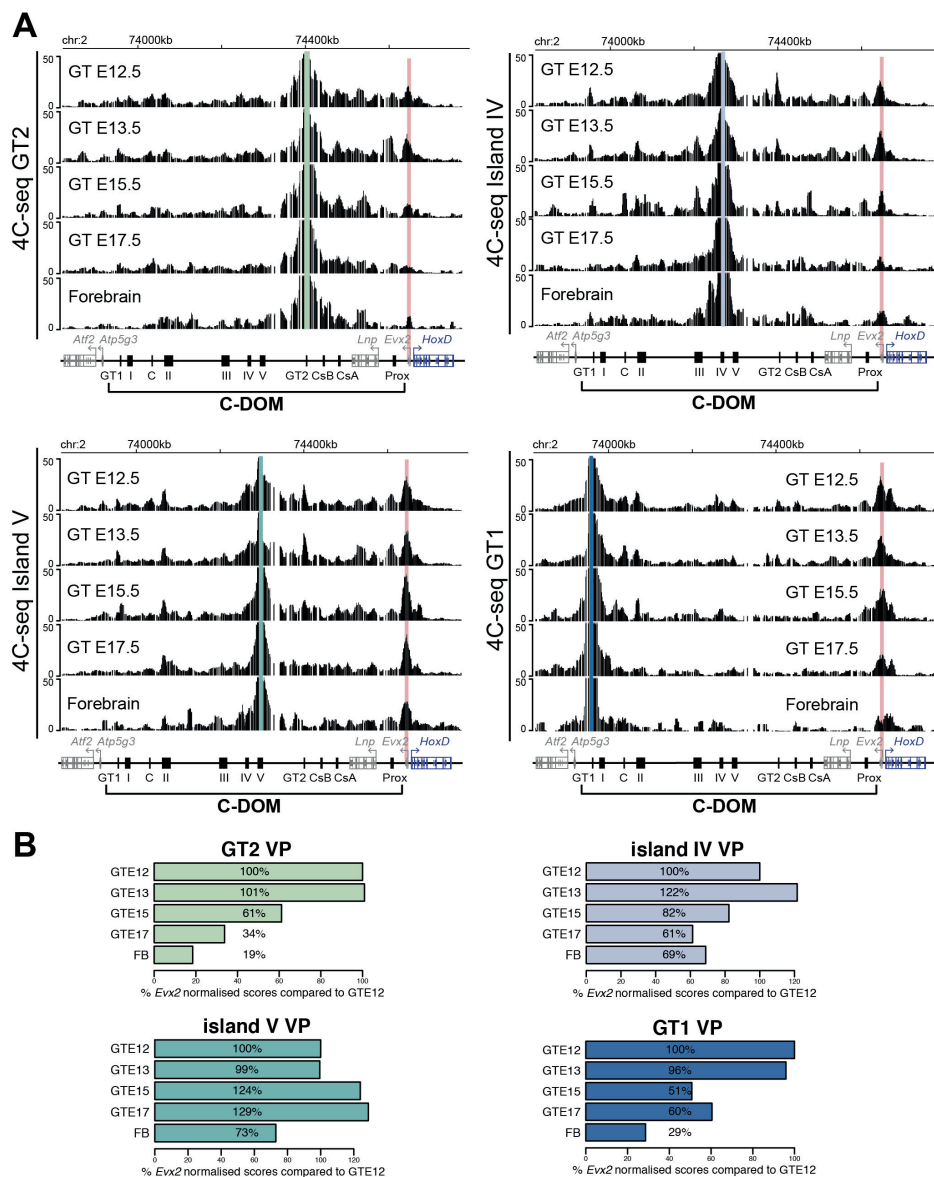


Figure 8: Decrease in enhancer-promoter contacts during GT development

A) 4C-seq profiles at the *HoxD* cluster and C-DOM of GT at E12.5, GT at E13.5, GT at E15.5, GT at E17.5, and forebrain are represented. Coordinates (mm10): chr2:73815520-74792376. Viewpoints are

identified by green to blue colored lines. Clockwise, first panel represent GT2 viewpoint, then island IV, GT1 and island V viewpoints. Red line identifies the *Evx2* gene region used for the quantifications. **B)** Bar plots show the percentage of the sum of scores in the *Evx2* region using GT E12.5 as a reference, for each viewpoint dataset. VP=viewpoint.

3.4 *In vivo* activity of mouse *HoxD* cis-regulatory elements

Previous work from the laboratory identified several regulatory elements located within the C-DOM and involved in the regulation of posterior *Hoxd* genes (Figure 9A) (Spitz *et al.*, 2003; Gonzalez *et al.*, 2007; Montavon *et al.*, 2011; Lonfat *et al.*, 2014). We set out to further explore the regulatory potential of these region during GT development. To this aim, we focused on the genomic segment positioned between the SB and *rel5* sequences, since this region accounted for 76% of *Hoxd13* expression in the GT (see below, Figure 10). We selected five regions of approximately 30kb in size based on ATAC-seq, H3K27ac, 4C-seq datasets and on DNA sequence conservation, and tested them for enhancer activity (Figure 9B). Each candidate regulatory region was cloned upstream of a *LacZ* reporter gene driven by a minimal *beta-globin* promoter and integrated at random positions in the mouse genome to test their enhancer potential *in vivo*.

Examination of transgenic embryos at E13.5, after X-gal staining, revealed that the IIIIE and IVE sequences contained enhancer activity in both GT and limb buds (Figure 9C) in a pattern that matched at least partially the endogenous expression domains of *Hoxd13* in these tissues. In the developing GT, the two sequences showed complementary patterns of expression, with IIIIE being active in the dorsal part of the GT (6 embryos), whereas the IVE sequence was more specific for the basal region of the GT (4 embryos) (Figure 9C). These two transgenes contained islands III and island IV, respectively, two previously characterized limb enhancers (Montavon *et al.*, 2011) for which activity in the GT was not reported. Therefore, these results suggest that the regions responsible for GT enhancer specificity correspond to the ATAC-seq and H3K27ac peaks identified outside island III and island IV and included in the transgenes (see arrows on Figure 9B). Interestingly, in the IIIIE construct, the pattern of expression in the distal forelimb was almost identical to the domain of expression of *Hoxd13*, whereas the island III transgene generated a signal only in a subset of the proximal limb (Montavon *et al.*, 2011). Three embryos injected with the E1 sequence showed a weak basal signal on the GT and no reproducible staining in the limb buds (Figure 9C). No reproducible enhancer activity was detected neither in the VE sequence (out of 9 PCR positive embryos), nor in E2 (out of 4 PCR positive embryos) (Figure 9C). Therefore, despite their chromatin signatures, the VE and E2 elements were not able to drive any specific transcription, at least in this transgenic assay. Of

note, the VE region includes a CTCF binding site, an insulator element involved in facilitating enhancer-promoter contact by DNA-looping (see Ong and Corces, 2014).

Overall, of the five elements tested only E1, IIIE, and IVE showed activity in GT. We also re-investigated the activity of GT2 by analyzing embryos transgenic for this region (Lonfat *et al.*, 2014), as expected the GT2 sequence displayed strong activity in GT cells exclusively (Figure 9C). This set of experiments highlighted the inherent regulatory complexity of this cluster, where almost every individual enhancer element displayed a unique pattern of activity. In some instances, these activities appeared in complementary domains (e.g., IIIE and IVE), while in others, largely overlapping activities were scored (e.g., GT2).

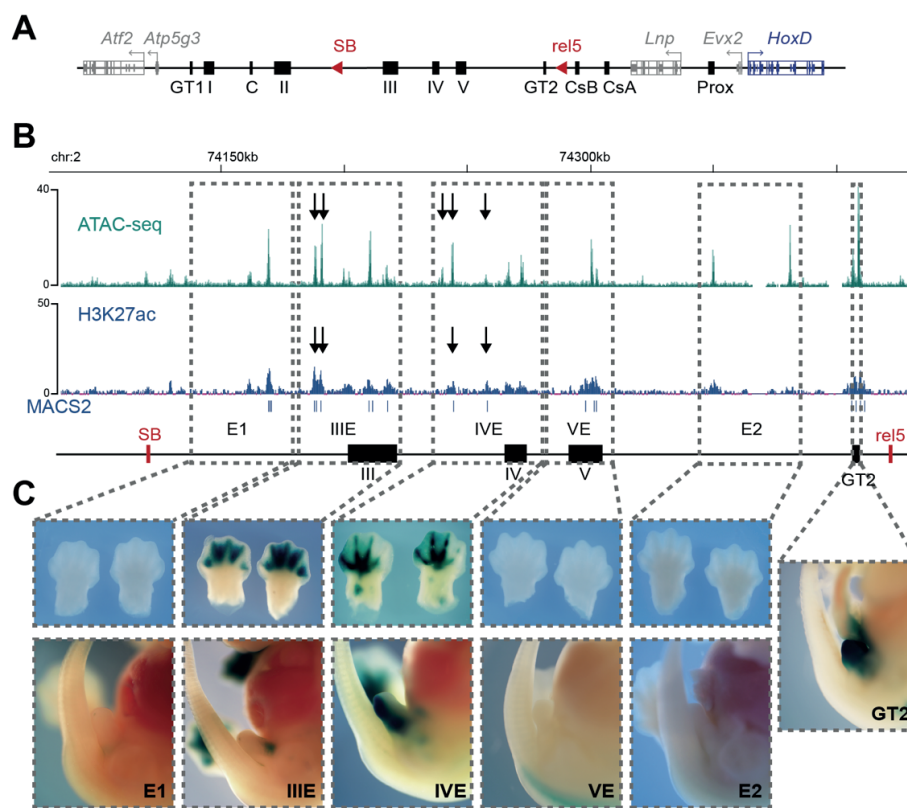


Figure 9: *In vivo* activity of mouse *HoxD* regulatory elements

A) Schematic representation of the *HoxD* cluster and its centromeric gene desert. Previously characterized enhancers are shown as black boxes, red arrowheads represent the SB and *rel5* sequences. **B)** ATAC-seq profile (average of three biological replicates) (upper track), and ChIP-seq for H3K27ac (bottom track) of GT E13.5 in the region between *rel5* and SB (Coordinates (mm10): chr2:74084880-74432824). Input was used for MACS2 peak caller tool, on the H3K27ac ChIP-seq blue boxes under the bottom track. This information in combination with 4C-seq and conservation (not shown) was used to predict the regions to test for enhancer activity. **C)** Enhancer activity of individual regulatory elements from the *rel5*-SB interval, gray dash-line boxes highlight the tested regions and the GT2 enhancer (Lonfat *et al.*, 2014). For each regulatory element a representative staining is shown.

3.5 Analysis of partial C-DOM deletions *in vivo*

To evaluate the functional importance of the C-DOM for 5' *Hoxd* genes during the development of the GT, we assayed some deletions involving parts of this region, in particular the *Del(rel1-rel5)*, *Del(rel5-SB)* and the *Del(SB-Atf2)* alleles (Montavon *et al.*, 2011) (Figure 10A). We also generated a novel allele corresponding to the deletion of the region located between island IV and the SB sequence (*Del(IV-SB)*), by using the CRISPR/Cas9 genome editing technology (Figure 10A). This new 154kb large deficiency, removed half of the regulatory region between the rel5 and SB sequences and contained three GT enhancers regions, E1, IIIE and IVE (see above, Figure 9B, C). We analyzed the effect of each of these deletions on the expression of 5' *Hoxd* genes and *Evx2* by RT-qPCR using RNA extracted either from control GT or from GT dissected from homozygous deleted mutant at E12.5.

The *Del(rel1-rel5)* allele removes one-third of the C-DOM, including two digit and/or GT enhancers (GCR and Prox) (Spitz *et al.*, 2001; Gonzalez *et al.*, 2007). In these mice, a 47% reduction in *Hoxd13* mRNA levels was scored in the GT (Figure 10B) ($p=0.0005$). The level of *Evx2* transcripts was reduced by 61% ($p<0.0001$), whereas, *Hoxd12*, *Hoxd11*, and *Hoxd10* were less affected (Figure 10B). Therefore, in the GT at E12.5, the expression of these genes was altered by the absence of the GCR and Prox enhancers in agreement with what Lonfat *et al.*, 2014 had observed at E15.5.

The *Del(rel5-SB)* allele removes 300kb of the C-DOM, including the GT2, island V, IV and III sequences. Mice carrying this deletion displayed a greater effect on gene expression. The level of *Hoxd13* mRNAs was reduced by 76% ($p<0.0001$), while the expression of *Evx2* was almost abolished ($p<0.0001$) (Figure 10C). *Hoxd12*, *Hoxd11* and *Hoxd10* were also affected, yet to a lesser degree (Figure 10C).

The *Del(SB-Atf2)* allele deleted the most distal part of the regulatory landscape. In these mutant mice, we observed a slight but significant upregulation of both *Evx2* and *Hoxd13* mRNA levels ($p=0.0009$ and $p=0.0005$ respectively) in the GT, whereas other 5' *Hoxd* genes were not affected (Figure 10D). All the above results are generally in agreement with what was observed for *Hoxd13* in GT at E15.5 (Lonfat *et al.*, 2014). Of note, the observed effects seem to be more pronounced at E12.5 than at E15.5.

We next analyzed the *Del(IV-SB)* allele in both GT and distal forelimb buds (DFL) at E12.5 by both RT-qPCR and whole mount *in situ* hybridization (WISH). This allele removes island IV and III but leaves the GT2 and island V sequences in place (Figure 10A). In the GT, we observed a 38% decrease in the amount of *Hoxd13* mRNAs ($p=0.002$) and a 57% decrease

for *Evx2* ($p=0.0008$) (Figure 10E), yet no significant effect was detected for any other 5' *Hoxd* genes. In the DFL, a minor reduction of transcript levels for both *Evx2* and *Hoxd13* was observed by RT-qPCR (Figure 10F). The WISH expression pattern of *Hoxd13* revealed a loss of staining in digit one (Figure 10G, arrows). Noteworthy, we observed a greater reduction of *Hoxd13* mRNA levels in the GT than in the DFL in the absence of this region. This may be due to a difference in the percentage of cells responding to the various GT or DFL enhancers and/or to the strengths of these enhancers. Interestingly, when we compared the RT-qPCR results for the GT by using either the *Del(rel5-SB)* or the *Del(IV-SB)* alleles, we observed that for both *Hoxd13* and *Evx2*, the GT2 enhancer alone could account for almost half of the expression of these genes, since it was the only active enhancer present in the region between island IV and *rel5* (as seen by the transgenic assay-Figure 9C).

We further used the *Del(IV-SB)* allele (Figure 10A) to carry out ATAC-seq in both the developing GT and DFL at E13.5 to look at all sequences associated with chromatin accessibility. In DFL, with the exception of the deleted region, we detected no significant difference in ATAC-seq profiles between the various alleles (Figure 10H, tracks 1 and 2 and MACS2 tracks). In the GT, we observed a loss of a ATAC-seq peak located between GT2 and CsB in the *Del(IV-SB)* allele (Figure 10H, arrows). This peak corresponds to island E, a region that showed some activity in a sub-population of cells in the GT when tested using a lacZ lentivirus-based reporter transgenesis (Lonfat, PhD thesis). Outside the island E region, any minor differences observed were due to individual variation (Figure 10H, MACS2 tracks). Therefore, with the exception of island E, which seemed to be sensitive to the deletion of the region between island IV and SB in the GT, other accessible DNA sequences appeared to be independent from the presence or absence of the (*IV-SB*) DNA interval. Taken together, these results indicate that several non-overlapping regions located on the C-DOM are required for target gene activation in the developing GT.

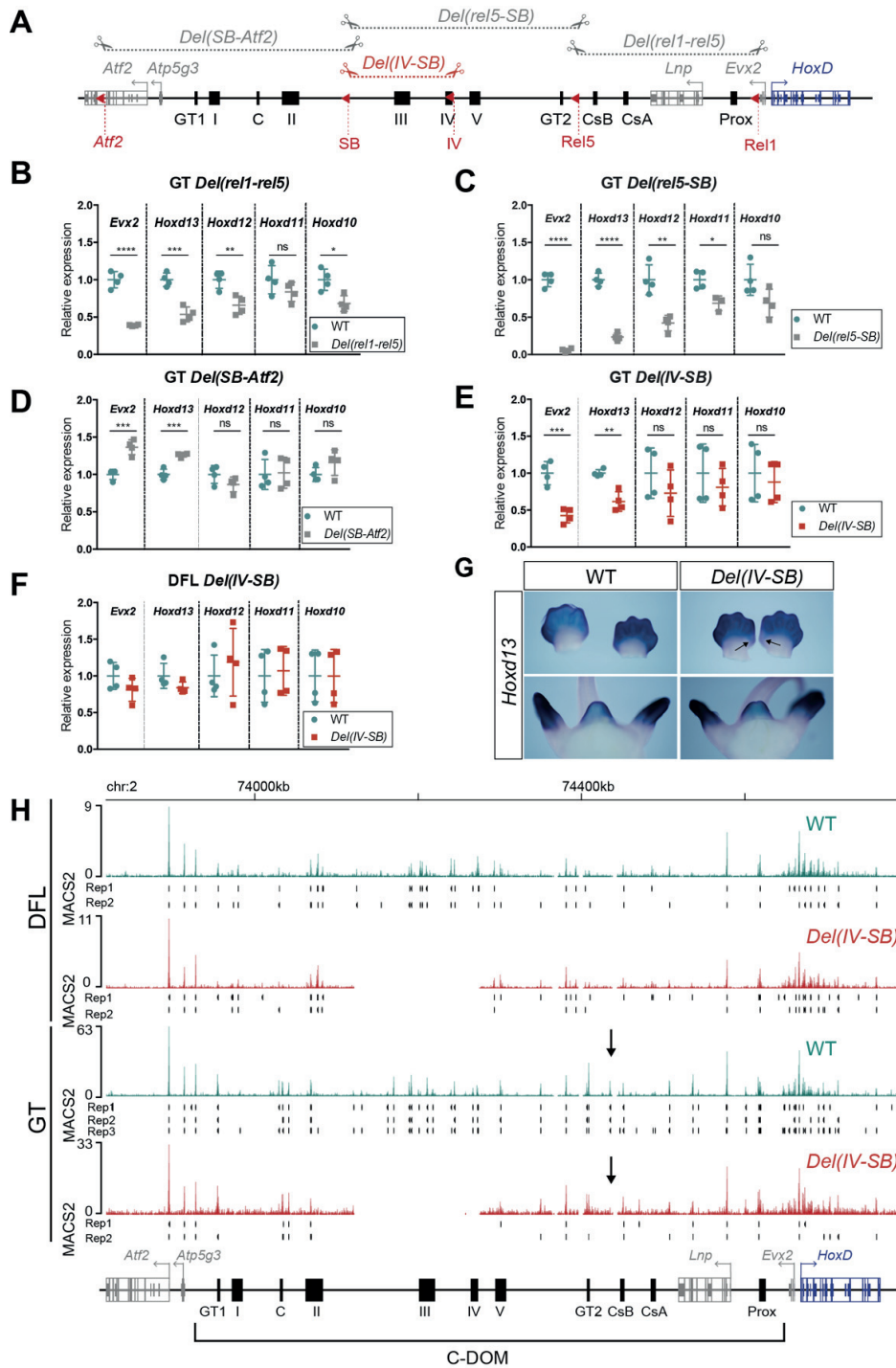


Figure 10: Analysis of partial C-DOM deletions

A) Schematic representation of the alleles analyzed. **B-F)** Relative expression obtained by RT-qPCR of wildtype and the different homozygous null alleles of E12.5 embryos. The title of each plot indicates the tissue and the allele used for the RT-qPCR. The mRNA levels of 5' *Hoxd* genes and *Evx2* were analyzed, values plotted indicate the ratio of expression using wildtype as a reference (green dots) for each gene (n=4 biologically independent WT and mutant GT or distal forelimb (DFL) pairs). We used an unpaired two-tailed t-test to evaluate the statistical significance of changes in gene expression. Bars indicate mean with SD, * p<0.01; ** p<0.001; *** p<0.0001, ****p<0.00001; ns= non-significant

($p > 0.01$). **G**) WISH of *Hoxd13* in wildtype (left) and *Del(IV-SB)* homozygous null (right) littermates E12.5 embryos. The arrowheads indicate a loss of expression in presumptive digit I in the mutant genotype. **H**) ATAC-seq profiles of wildtype (green) and *Del(IV-SB)* (orange) DFL (upper tracks) and GT at E13.5 (bottom tracks) at the *HoxD* cluster and C-DOM regions. Coordinates (mm10): chr2:73815520-74792376). Each track represents the average of the corresponding biological replicates. Peaks called using MACS2 are displayed under the corresponding tracks for each replicate (black rectangles under each track). The black arrow highlights the island E peak which disappears in the *Del(IV-SB)* GT track.

3.6 Functional contribution of different *cis*-regulatory elements

The complex expression patterns of *Hox* genes as well as of many other developmental genes, are usually controlled by the action of multiple enhancers, either regulating subsets of the global expression pattern, or acting together in a partially redundant manner, as observed by transgenesis and genetic experiments (Spitz *et al.*, 2003; Montavon *et al.*, 2011; Lonfat *et al.*, 2014). To determine the function of single enhancer element in the regulation of *Hoxd* genes in the GT, we used CRISPR/Cas9 genome editing in fertilized eggs to generate a series of alleles where these regulatory elements were either deleted or inverted (Figure 11A, Figure 18). We generated deletion alleles for all known enhancers located between the rel5 and SB sequences (*Del(GT2)*, *Del(IV)*, and *Del(IIIIE)*), i.e. in the DNA interval that when removed, had the greater impact upon *Hoxd13* expression (see above, Figure 10E). We also generated a deletion allele of the Prox sequence, *Del(Prox)*, a known limb and GT enhancer (Gonzalez *et al.*, 2007). In addition, to assess the functional importance of the CTCF binding sites located within island V and its polarity, we both deleted and inverted this region and its associated CTCF sites (*Del(V)* and *Inv(V)* respectively).

We examined whether the loss of individual elements would influence the levels of expression of either *Hoxd* genes or *Evx2*, by RT-qPCR (Figure 11B). Unexpectedly, with the exception of a minor reduction in the mRNAs level of *Evx2* in the GT after deletion of the GT2 enhancer ($p = 0.0079$) (Figure 11B), we did not detect any significant difference in transcript levels after the deletion of any individual regulatory elements, neither in the GT, nor in DFL at E12.5 for all regulatory elements located between the rel5 and SB sequences (i.e., *Del(GT2)*, *Del(IV)*, *Del(IIIIE)*, *Del(V)* and *Inv(V)*)(Figure 11B). We performed whole mount *in situ* hybridization for *Hoxd13* on littermates of both wildtype and single regulatory elements deletions to evaluate any potential variation in the expression domain (Figure 11C). There again, we did not observe any significant changes, neither in the intensity of the signal, nor in the spatial transcript pattern in any of the alleles. These results suggest that no single enhancer sequence, located between the rel5 and SB, seems to be essential for proper expression of 5'

Hoxd genes in both GT and DFL, at least at the developmental stages and tissues analyzed. These results were particularly surprising for the *Del(GT2)* allele that from previous experiments seemed to be relevant for expression of *Hoxd13* in the GT.

Interestingly, upon analyzing the *Del(Prox)* allele we observed that homozygous animals carrying this deletion displayed a 36% decrease in the expression of *Hoxd13* in the GT ($p=0.0013$), a similar effect was observed for *Evx2* and *Hoxd12* (Figure 11B). In the DFL we also detected a significant decrease in expression *Hoxd13*, *Evx2*, and *Hoxd12* transcript levels (16% ($p=0.0008$), 22% ($p=0.006$), and 25% ($p=0.016$) respectively), but to a lesser degree when comparing to the GT (Figure 11C). These results suggest that deleting *Prox* has an impact on the regulation of 5' *Hoxd* genes in both the GT and DFL. Of note we were unable to detect changes in the expression pattern of *Hoxd13*, by WISH (Figure 11C).

All together these results highlight that distinct regulatory sequences work in a different fashion to regulate gene expression, suggesting a possible temporal and/or functional hierarchy between the different *cis*-regulatory elements present in the C-DOM. In this case, *Prox* seems to be both sufficient and necessary for *Hoxd* expression in the GT whilst the individual *cis*-regulatory elements located between the *rel5* and *SB* sequences appear to have a more modest role.

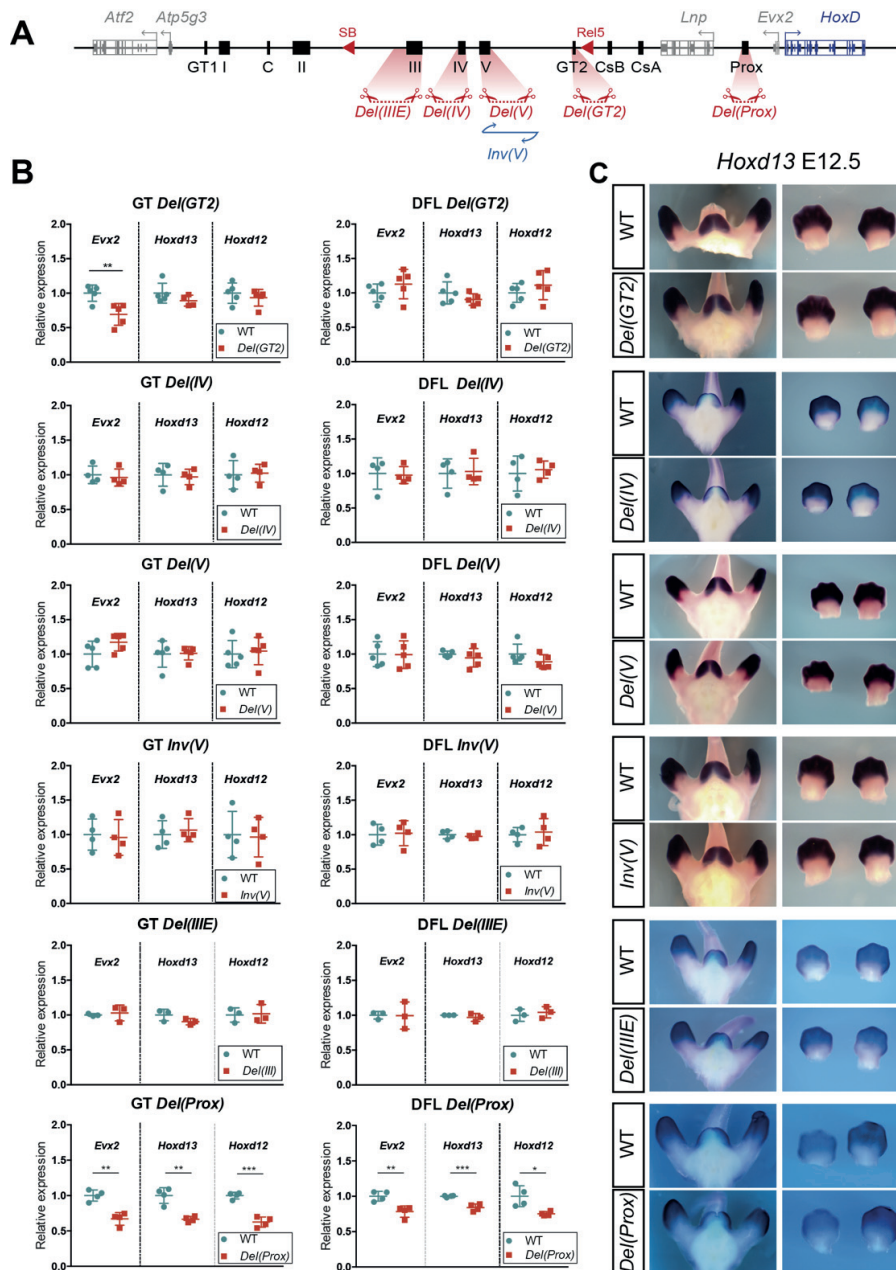


Figure 11: Analysis of single *cis*-regulatory elements deletion alleles

A) Schematic representation of the alleles generated by CRISPR-Cas9 genome editing technology. **B)** Relative expression obtained by RT-qPCR of wildtype and the different homozygous null or inverted alleles of GT and DFL of E12.5 embryos. The title of each plot indicates the tissue and the allele used for the RT-qPCR. The mRNA levels of 5' *Hoxd* genes and *Evx2* were analyzed, values plotted indicate the ratio of expression using wildtype as a reference (green dots) for each gene ($n \geq 4$ biologically independent WT and mutant GT or DFL pairs). We used an unpaired two-tailed t-test to evaluate the statistical significance of changes in gene expression. Bars indicate mean with SD, * $p \leq 0.01$; ** $p \leq 0.001$; *** $p \leq 0.0001$. **C)** WISH using the *Hoxd13* probe in both wildtype and the correspondent homozygous null or inverted littermates E12.5 embryos. The spatial expression of *Hoxd13* remains globally unchanged.

3.7 Chromatin organization in the *Del(GT2)* allele

We assessed the impact of the deletion of a single regulatory element upon the functional reorganization of the regulatory landscape. To this aim, we selected the *Del(GT2)* allele. This enhancer sequence showed a ubiquitous and specific pattern of expression in the GT by transgenic assay and results obtained by using large deletions pointed to its critical importance (see above, Figure 10). We performed 4C-seq and ATAC-seq in both control and *Del(GT2)* homozygous GT at E13.5 to try and detect any potential alterations in the regulatory landscape (Figure 12).

The analysis of the ATAC-seq datasets revealed a loss of accessibility (peak) associated with island E in the *Del(GT2)* mutant allele, similar to what had been observed for the *Del(IV-SB)* allele in the GT (Figure 12A, black arrows). Overall, the removal of the GT2 sequence had only a small impact on C-DOM structure. This result supports the existence of a regulatory redundancy, whereby other elements in the C-DOM are capable of compensating for the loss of the GT2 enhancer and thus ensure proper *Hoxd13* gene expression. Therefore, it is likely that the wildtype levels of *Hoxd13* transcripts observed in the *Del(GT2)* mice were compensated for by the remaining intact *cis*-regulatory elements and not by the recruitment of novel elements.

We also examined the potential effect of removing this enhancer upon the interaction profile of the regulatory landscape and gene cluster. We observed only minor differences in chromatin organization after the deletion of GT2 (Figure 12B). When we used *Hoxd13* as a viewpoint, both in wildtype and in *Del(GT2)* mutant GT, this gene was able to contact all other previously characterized enhancers in this regulatory region with apparently comparable frequencies (Figure 12B, tracks 1 and 2). To evaluate the potential effect of the GT2 deletion upon the interactions between other regulatory regions, we used a view-point positioned on island IV and produced a 4C-seq profile. With this set up, minor differences in the frequency of contacts were observed between the two genotypes (Figure 12B, tracks 3 and 4). The statistical significance of these differences was nevertheless not fully established and hence they may be due to individual variation. In addition, they were not corroborated by corresponding alterations in accessible sites (Figure 12A). Overall the absence of GT2 did not seem to importantly affect either the interactions between *Hoxd13* and other enhancers, or the contacts between the enhancers themselves, indicating that the removal of one enhancer in isolation has little effect on the global architecture at the locus.

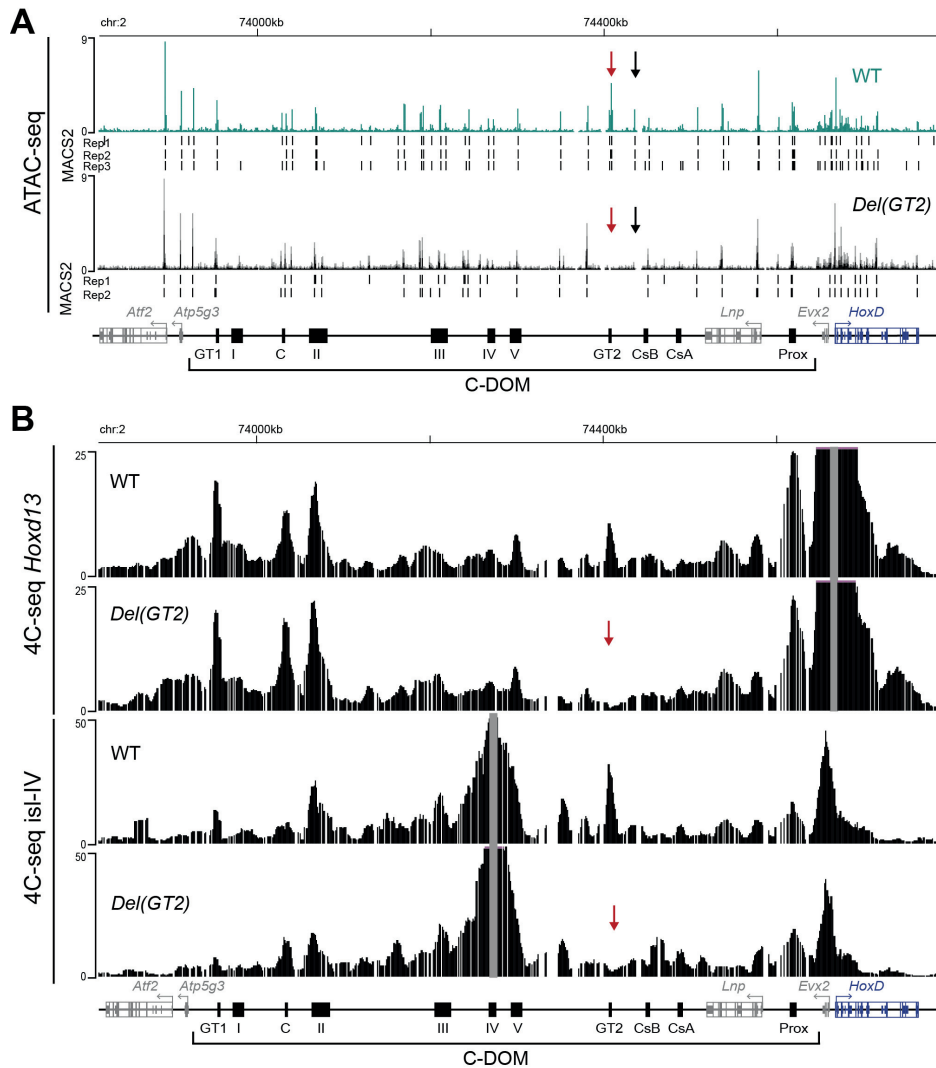


Figure 12: Chromatin architecture in the *Del(GT2)* allele

A) ATAC-seq profiles of the *HoxD* cluster and C-DOM in wildtype and *Del(GT2)* homozygous null GT at E13.5. Coordinates (mm10): chr2:73815520-74792376. The wildtype track (upper track) represents the average of three biological replicates and the *Del(GT2)* track (lower track) represents the average of two biological replicates. Peaks called using MACS2 are displayed under the corresponding tracks for each individual replicate (black rectangles under each track). Red arrows highlight the deleted region, black arrows indicate a lost peak in *Del(GT2)* (island E). **B)** 4C-seq profiles of wildtype and *Del(GT2)* homozygous null GT at E13.5. The two upper tracks show profiles from the *Hoxd13* viewpoint (average of two biological replicates). The two lower tracks correspond to 4C-seq profiles from the island IV viewpoint of wildtype (average of two biological replicates) and *Del(GT2)* homozygous (average of three biological replicates) null GT at E13.5 Coordinates (mm10): chr2:73815520-74792376. Viewpoints are highlighted by a gray line. Red arrows show the deleted region.

3.8 The case of *Evx2*

The homeobox-containing gene *Evx2* is located immediately 5' to *Hoxd13*. As a consequence, *Evx2* shares several regulatory elements with 5' *Hoxd* genes, for example in the

distal limb buds and in the GT (see above and (Kmita *et al.*, 2002; Montavon *et al.*, 2011)). Because a mild reduction of *Evx2* transcript levels was observed in mice deleted for the GT2 enhancer (see above, Figure 11B), we performed whole mount *in situ* hybridization with littermates of both control and *Del(GT2)* embryos at E11.5 and E12.5. We observed a slight decrease in signal intensity in *Del(GT2)* embryos in both E11.5 and E12.5 (Figure 13A, arrows), confirming the RT-qPCR results. We thus examined the potential consequences of the GT2 deletion on long-range interactions involving *Evx2*. We carried out 4C-seq using *Evx2* as a viewpoint and all contacts with GT2 were expectedly lost. However, we also observed that deleting GT2 reduced the frequency of interactions between GT1, island IV, and *Evx2* (Figure 13B, arrows), suggesting that GT1 and island IV may have also contributed to the decrease in *Evx2* mRNA levels observed in mutant *Del(GT2)* GT (see above, Figure 11B). Therefore, the deletion of GT2 appeared to have a stronger impact over *Evx2* than over *Hoxd13*.

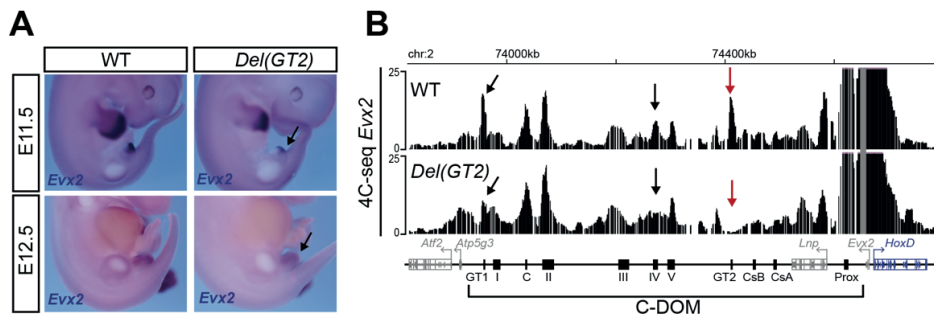


Figure 13: Regulation of *Evx2* in the *Del(GT2)* allele

A) WISH using the *Evx2* probe in both wildtype and *Del(GT2)* homozygous null littermates at both E11.5 (upper panel) and E12.5 (lower panel). Arrows indicate a decrease in intensity of signal in the GT of *Del(GT2)* embryos. **B)** 4C-seq profiles (average of two biological replicates) of wildtype and *Del(GT2)* homozygous null GT at E13.5. Coordinates (mm10): chr2:73815520-74792376. *Evx2* was used as a viewpoint (gray line). Red arrows show the deleted region, black arrows indicate changes in interactions with GT1 and island IV between genotypes.

3.9 Deletion or inversion of CTCF

The enrichment of bound CTCF and Cohesin at TAD boundaries (Dixon *et al.*, 2012; Rao *et al.*, 2014) and their functional importance in maintaining these structures was recently documented (Nora *et al.*, 2017; Rao *et al.*, 2017; Schwarzer *et al.*, 2017). However, less is known about the function of the CTCF and Cohesin sites that are present within TADs. These sites are thought to contribute to the formation of smaller sub-structures, which appear more dynamic and cell-type specific (Phillips-Cremins *et al.*, 2013; Rao *et al.*, 2014). In the case of the *HoxD* locus, several CTCF sites are spread throughout the two regulatory landscapes

(Rodríguez-carballo *et al.*, 2017). We used ChIP-seq to analyze the occupancy of CTCF and RAD21 (a Cohesin sub-unit) in the GT. In the C-DOM, we observed several binding sites for both CTCF and RAD21, accumulating primarily between island II and *Atf2* (Figure 14A). Of note island V was the only region between *Evx2* and island II where binding of both these proteins was detected (Figure 14A). At the level of the gene cluster itself, binding occurred mainly in the 5' and central parts of the cluster (Figure 14B). With the exception of a more prominent CTCF binding on the promoter of *Hoxd9*, the distribution of CTCF and RAD21 in the GT was similar to what was observed in limb buds (Rodríguez-carballo *et al.*, 2017).

As it was shown that the majority of CTCF mediated chromatin loops are established between sites positioned in convergent orientations (i.e., sites with consensus CTCF motifs pointing toward each other) (Rao *et al.*, 2014; de Wit *et al.*, 2015; Vietri Rudan *et al.*, 2015), we determined the orientation of CTCF sites in the cluster and its centromeric regulatory region using the CTCFBSDB database (<http://insulatordb.uthsc.edu>). Accordingly, we defined the CTCF sites located between *Evx2* and *Hoxd11* as being positioned in the reverse orientation, in contrast to the majority of the other sites within the cluster, which are in the forward orientation (Figure 14B). In the C-DOM, the sites with the highest enrichments for CTCF and RAD21 are in the forward orientation (towards the cluster) (e.g., island C, island II and island V). A close examination of island V revealed one prominent CTCF/RAD21 co-binding site and a less prominent site close by, which was bound by CTCF only. Interestingly this CTCF/RAD21 site on island V correspond to a “stable” chromatin interaction, that is maintained independently of the transcriptional status of the gene cluster.

To investigate the potential effect of removing CTCF/Cohesin on sub-domain formation within the *HoxD* centromeric regulatory landscape, we used the *Del(V)* and *Inv(V)* alleles (see above, Figure 11A), which either delete or invert the two CTCF sites, respectively. We first analyzed the occupancy of CTCF and Cohesin (RAD21) by ChIP-seq in *Del(V)* GT at E13.5. While we observed a complete loss of CTCF and RAD21 associated with island V, this deletion did not seem to affect the binding of CTCF and RAD21 either to other sites in the centromeric regulatory landscape, or to the gene cluster (Figure 14A and 14B).

To assess the impact of either removing or inverting these CTCF sites on the remaining regulatory elements, we performed ATAC-seq in both wildtype, *Del(V)*, and *Inv(V)* homozygous GT at E13.5. In the *Del(V)* GT, with the exception of the deleted region, we did not observe any change in the peaks profile (Figure 14C) when compared to the control GT. Minor changes were not reproduced in the replicates and are likely due to individual variation (Figure 14C, MACS2 on replicates). In the *Inv(V)* GT, we observed again a loss of the ATAC-

seq peak associated with island E (Figure 14C, black arrow in bottom track), similar to what was scored in the *Del(GT2)* and *Del(IV-SB)* alleles. These observations support a view whereby the deletion or inversion of a CTCF site within a regulatory landscape has little impact on the implementation of the remaining regulatory elements, thus corroborating the RT-qPCR results where no change in expression of either 5' *Hoxd* genes or of *Evx2* were observed (Figure 11B).

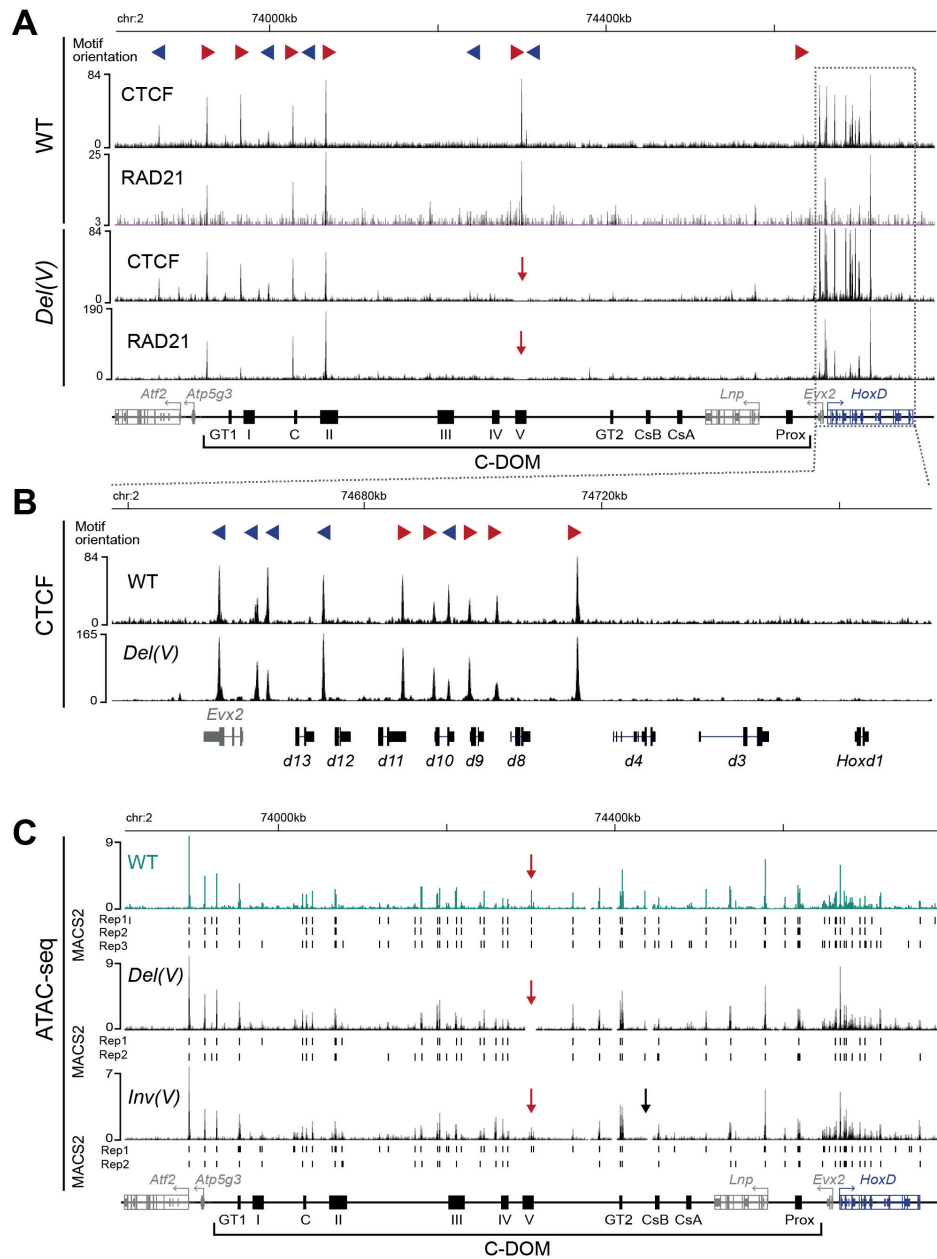


Figure 14: CTCF deletion or inversion

A) CTCF and RAD21 ChIP-seq profiles of wildtype and *Del(V)* homozygous null GT at E13.5. Red and blue arrowheads show forward and reverse CTCF motif orientation respectively. Red arrows indicate deleted CTCF and RAD21 sites on island V. **B)** CTCF ChIP-seq profiles of wildtype and *Del(V)*

homozygous null GT at E13.5 at the *HoxD* locus (coordinates (mm10): chr2: 74637433-74775728). C) ATAC-seq profiles of the *HoxD* cluster and C-DOM regions in wildtype, *Del(V)*, and *Inv(V)* GT at E13.5. Coordinates (mm10): chr2:73815520-74792376. The wildtype track (green track) represents the average of at three biological replicates and the *Del(V)* and *Inv(V)* tracks represents the average of two biological replicates. Peaks called using MACS2 are displayed under the corresponding tracks for each individual replicate (black rectangles under each track). Red arrows highlight the deleted or inverted region. Black arrows indicate a lost peak (island E) in the *Inv(V)* allele.

To investigate the mechanisms by which CTCF/Cohesin mediated contacts may influence the interactions between *cis*-regulatory elements and promoters and in turn their activity, we performed 4C-seq in wildtype, *Del(V)* and *Inv(V)* GT at E13.5. We first analyzed the *Del(V)* allele by using *Hoxd13* as a viewpoint. As expected, contacts with island V were not observed. This contact loss was confirmed by the peak caller. Surprisingly, abolishing CTCF binding in island V had a reduced impact on the interactions between *Hoxd13* and the other enhancers within the C-DOM (Figure 15A). We confirmed this result by using a viewpoint positioned on island IV (the closest enhancer to island V in linear distance) and we observed a slight reduction in the frequency of contacts between island IV and *Hoxd13* (Figure 15A, bottom track). Therefore, it seems that island V and its associated CTCF sites have a marginal importance, if any, in maintaining the chromatin structure of this regulatory landscape.

We next analyzed the impact of inverting island V along with its two CTCF sites. Qualitative analysis of the interaction profile generated using *Hoxd13* as a viewpoint revealed a specific disruption in the contacts with island V, which was confirmed by the peak caller (Figure 15B). Inverting the CTCF binding sites at island V reduced, but did not abolish the interactions between this sequence and *Hoxd13*. We validated this result by doing the reverse experiment and placing a viewpoint on island V. In this set up, we observed a reduction in the frequency of interactions between island V and *Hoxd13* thus confirming the previous result (Figure 15B arrow). Noteworthy, we observed an increase in contacts with both island II and island C, i.e. with the next convergent oriented CTCF sites (Figure 15B, arrowheads). When we used island IV as a viewpoint, we also observed a reduction in contacts with *Hoxd13* (Figure 15B, arrow). Taken together, these results suggest that either the loss or the inversion of island V and its associated CTCF sites is only partially important for the maintenance of the chromatin structure in this regulatory landscape. Overall, we observed that slight modifications in chromatin structure within a TAD are not sufficient to disturb its global functional organization in a way that may lead to modifications in gene expression.

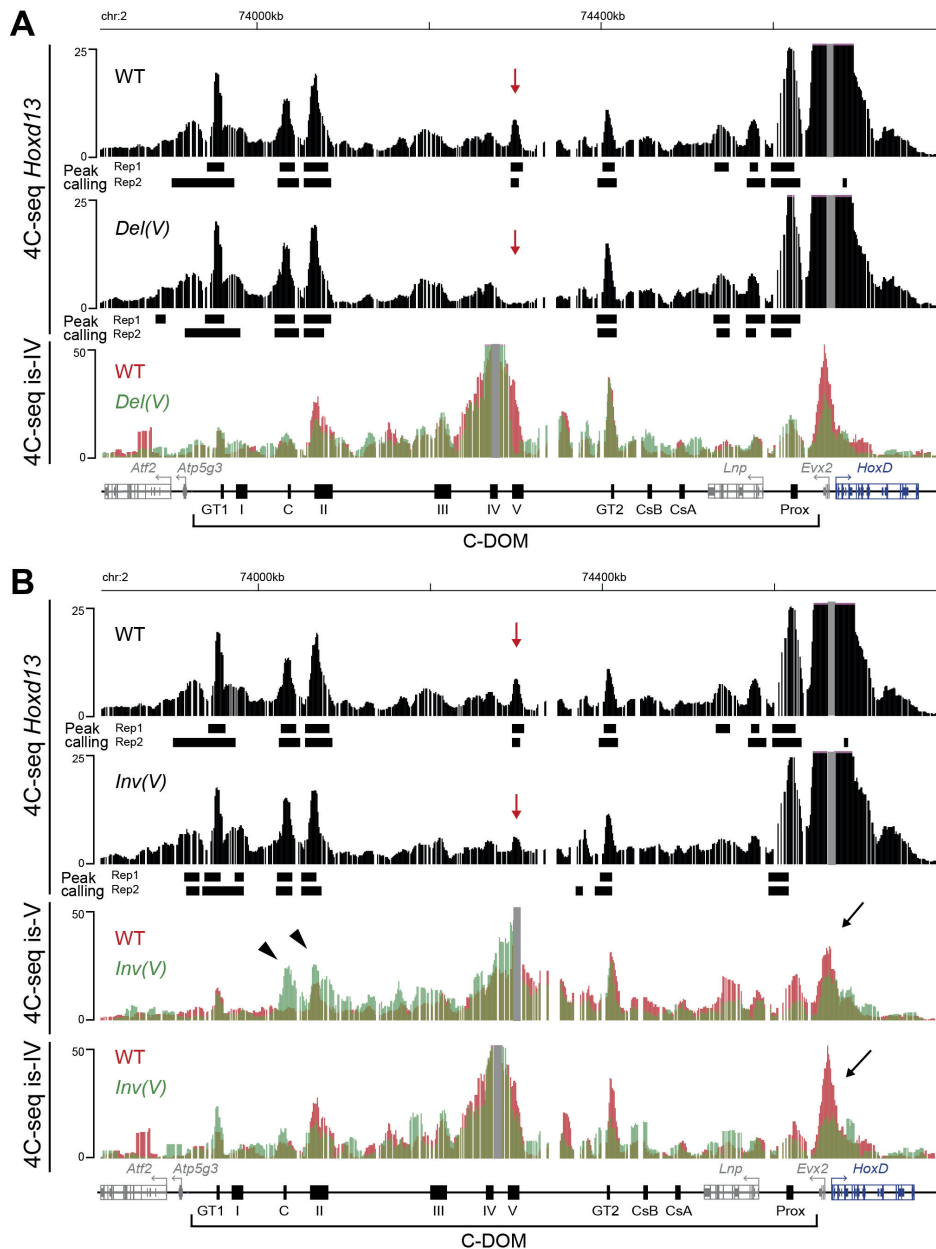


Figure 15: Chromatin architecture in CTCF deletion or inversion

A) 4C-seq profiles (average of two biological replicates) of wildtype and *Del(V)* homozygous-null GT at E13.5. Coordinates (mm10): chr2:73815520-74792376. *Hoxd13* (upper tracks) and island IV (lower track) viewpoints are highlighted by a gray line. Black boxes under the *Hoxd13* viewpoint 4C-seq tracks show peaks identified by the peak calling algorithm (PeakC) for each individual replicate (Peak calling). Red arrows show the deleted region. The profiles from island IV viewpoint are shown as overlays of contacts formed by wildtype (red) and *Del(V)*(green). **B)** 4C-seq profiles (average of two biological replicates) of wildtype and *Inv(V)* homozygous GT at E13.5. Coordinates (mm10): chr2:73815520-74792376. Viewpoints are highlighted by a gray line. On *Hoxd13* viewpoint, below each 4C-seq track, peaks identified by the peak calling algorithm (PeakC) are shown by black boxes for each individual replicate (Peak calling). Red arrows show the inverted region. The profiles from island IV and island V viewpoint are shown as overlays of contacts formed by wildtype (red) and *Del(V)*(green). Black arrows indicate loss of contacts with the *Hoxd13* region in *Inv(V)*, arrowheads highlight gain of contacts with island II and island C (the next CTCF sites) in *Inv(V)*.

4. Discussion

In this work we studied the chromatin dynamics at the *HoxD* locus as well as the functional contribution of specific *cis*-elements to the regulation of *Hoxd* genes during GT development. Upon analysis of the chromatin conformation and transcription profile of this locus we observed that the overall chromatin structure of the C-DOM predates the embryonic appearance of the GT. As GT development progresses, we observed a reduction in transcript levels correlating with a decrease in enhancer-promoter chromatin loops within the C-DOM. This decrease occurred while maintaining a subset of CTCF/Cohesin associated contacts, which were preserved independently from the transcriptional status of the gene cluster.

Functional analysis of the *HoxD* regulatory landscape revealed that while deletion of the Prox element or deletions comprising several enhancers affected *Hoxd* genes expression levels, other single deletions of both transient (enhancer associated) and constitutive contacts (CTCF/Cohesin associated) displayed little if any effect on gene expression in the GT. Overall our results suggest enhancer elements within this complex regulatory landscape have different functional strengths, and highlight the existence of a dynamic yet robust system to tightly regulate *Hoxd* genes expression.

4.1 *Hoxd* genes regulation

We compared different timepoints during GT development and characterized the transcriptional profile, epigenetic status, and chromatin conformation at the *HoxD* locus. Our results show that the chromatin conformation of the C-DOM predates the emergence of the GT and that as development progresses there is a decrease in overall *Hoxd* transcripts that is accompanied by the corresponding alterations in histone marks and chromatin accessibility sites.

In the cloaca, a tissue that was proposed to be an organizing center for GT formation (Tschopp *et al.*, 2014), all *Hoxd* genes are accessible as seen by ATAC-seq and positively marked for H3K27ac. Our datasets are in accordance with RNA-seq data from a public dataset (GEO accession number - GSE88764), showing that both 3' and 5' *Hoxd* genes are expressed in the cloaca at E10.5. Interestingly already in the cloaca *Hoxa13* and *Hoxd13* are the highest expressed *Hox* genes of each respective cluster. Regarding the regulation of *Hoxd13* in the cloaca, we observed that the C-DOM, the region primarily responsible for the regulation of this gene in the GT (Lonfat *et al.*, 2014), displayed several enhancer-promoter contacts with *Hoxd13* (as seen by 4C-seq) and ATAC-seq peaks, some of them corresponding to previously

identified GT enhancers (e.g., GT1 and GT2). Indeed, analysis of the GT2 transgene (Lonfat *et al.*, 2014) at E10.5 showed enhancer activity of this element in the cloaca (data not shown). Of note, we were unable to detect enrichment for H3K27Ac mark for this element in our ChIP-seq dataset. 4C-seq analysis from the *Hoxd13* viewpoint showed the presence of novel interactions with the C-DOM upon *Hoxd* genes activation in the cloaca. This is contrary to what was observed in the trunk, a tissue where the regulation is primarily accomplished by elements positioned within the cluster (Spitz *et al.*, 2001). In the latter case, upon gene activation, no differences in the interaction profiles between *Hox* target genes and their neighboring gene deserts were observed (Noordermeer *et al.*, 2014). Although further analyses are necessary, these data suggest that the C-DOM contributes, at least in part, to the long-range regulation of *Hoxd13* in the cloaca.

Although the C-DOM may contribute to *Hoxd13* regulation in the cloaca, it is possible that elements present within the cluster are also involved in this regulation. Indeed, WISH analysis of *Hoxd13* in mice lacking the endogenous *HoxD* cluster but carrying a transgenic BAC containing the murine *HoxD* cluster, randomly integrated into the genome, showed staining for this gene in the cloaca region (Beccari unpublished data). This observation supports the idea that the cluster alone is able to drive expression of *Hoxd13* in the cloaca in the absence of the adjacent regulatory regions. Furthermore, a large inversion that relocates the C-DOM approximately 3Mb away but leaves the *HoxD* cluster intact, resulted in expansion of the expression domain of *Hoxd13* and increased intensity of WISH staining for this gene the cloaca region at E10 (Tschopp and Duboule, 2011). Although a full quantification of the effect on gene expression of this inversion is needed, it is clear, as observed by WISH, that upon re-allocation of the C-DOM the expression of *Hoxd13* is not reduced, if anything it increases. This suggests that the expression of *Hoxd13* in the cloaca can be sustained without the C-DOM, contrary to what was observed for the DFL at E12.5 (Tschopp and Duboule, 2011) and in the GT (Lonfat, PhD thesis). While the underlying regulatory mechanisms controlling *Hoxd* gene regulation in the cloaca remains to be fully understood, it is likely that both sequences located inside and outside of the *HoxD* cluster are involved in controlling *Hoxd* genes expression in this tissue.

In the GT at E12.5, there is a loss of accessibility sites and H3K27ac marks at 3' *Hoxd* genes and a gain of H3K27me3 marks, as expected these genes are no longer expressed. In turn, 5' *Hoxd* genes are expressed and H3K27ac positive. At this stage, the C-DOM is fully active with known GT enhancers being positively marked by H3K27ac. Here the contribution of the C-DOM to the transcription of *Hoxd13*, and other 5' *Hoxd* genes to a lesser degree, is

evident not only by the activation of enhancers but also by the results of the transgenic assays and the C-DOM deletions series, which affect 5' *Hoxd* genes expression. In late stages of GT development, we observed a reduction in the amount of 5' *Hoxd* genes transcript levels correlating with loss of enhancer-promoter interactions and H3K27ac marks in the C-DOM and with gain in H3K27me3 marks at both 5' *Hoxd* genes and at discrete regions of the C-DOM (island I and island V), at least in a subpopulation of cells. These results suggest that the loss of 5' *Hoxd* genes transcript levels is accompanied by a loss of enhancer activity in the C-DOM. Taken together, our results are in accordance with previous experiments suggesting that in the GT the C-DOM is the primary contributor to the regulation of *Hoxd13* (see introduction) (Spitz *et al.*, 2001; Lonfat *et al.*, 2014).

4.2. Constitutive vs. transient interactions: The spatial organization of enhancer-promoter interactions at the *HoxD* locus

During embryonic development, three-dimensional chromatin organization of the genome is critical for the accurate control of gene expression. Extensive study of the *HoxD* cluster has provided great insight into gene regulation at this locus. As mentioned before, in the case of the GT, the regulation of *Hoxd* genes is primarily accomplished by the C-DOM (Lonfat *et al.*, 2014). However, how the pattern of interactions between the genes and their regulatory domains is established during development was still unknown.

We have examined chromatin dynamics at the *HoxD* locus during GT development. Both in mESC and in forebrain (i.e., the basal state of our 4C-seq analysis and the negative control, respectively), in the absence of transcription, the entire *HoxD* cluster forms a dense domain (Noordermeer *et al.*, 2011, 2014). Some discrete contacts are established with the surrounding regulatory regions, island V and island II in C-DOM and CS39 in T-DOM, correlating with co-binding of CTCF and Cohesin at those genomic locations. Overall, it is clear that in the absence of transcription the chromatin structure surrounding the *HoxD* cluster is roughly maintained when using *Hoxd13* as a viewpoint. This is consistent with what was previously observed for *Hoxd13* and other viewpoints (see Noordermeer *et al.*, 2014). Upon activation of target genes transcription, in the cloaca, new contacts with the surrounding regulatory regions were observed (e.g., *Hoxd13*-Prox and *Hoxd13*-GT2). This observation suggests that the overall chromatin structure of the C-DOM predates the emergence of the GT. It is interesting to notice that some of the interactions observed in the cloaca correspond to ATAC-seq peaks (e.g., GT2) while others do not (e.g., Prox), suggesting that in the case of

Prox the initial interaction with *Hoxd13* may be independent of the enhancer activity. Indeed, a study in *Drosophila* showed that in some cases, enhancer-promoter interactions are already established prior to gene activation, and are frequently associated with paused RNA polymerase (Ghavi-Helm *et al.*, 2014). Further confirmation of the Prox activity (or lack of) in the cloaca is however needed, for example by an *in vivo* enhancer assay.

As development progresses, in GT at E12.5 and E13.5, the C-DOM is fully active with the highest frequency of contacts being scored. These contacts occurred within regions enriched for H3K27ac marks, which are generally associated with active genes and enhancers (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011), as exemplified by robust interactions with GT2. These interactions in a transcriptionally active context were shown to be facilitated by the Mediator complex that helps to recruit factors, such as the Cohesin complex, to enable promoter-enhancer chromatin interactions (Kagey *et al.*, 2010; Phillips-Cremins *et al.*, 2013). While the majority of these interactions were already scored in the cloaca, although with lower frequency, the interaction with a sub-domain of island III only appeared in the GT when this enhancer became positively marked by H3K27ac, suggesting a correlation between the activation of this element during GT development and the appearance of this particular interaction peak. This is in contrast with what was observed for Prox, where the interactions were pre-formed even before activation of the regulatory element, as evaluated by the absence of the Prox ATAC-seq peak in the cloaca.

Upon continuous decrease in *Hoxd* transcript levels, all contacts associated to active regulatory regions within the C-DOM seem to disappear. In GT at E17.5 the C-DOM acquires a mESC/forebrain-like chromatin architecture, keeping a framework of constitutive interactions associated to CTCF/Cohesin co-binding sites (i.e., island II and island V). This result may be explained by a dilution effect. In this scenario, the decrease in both *Hoxd* transcript levels and contacts associated with active regulatory regions would reflect only having a small subset of cells in the tissue that express *Hoxd* genes.

Overall it appears clear that there are changes in the regulatory landscape associated with the transcriptional status of the cluster. This seems to be a pervasive feature during development. Indeed, recent work has also illustrated changes in intra-TAD enhancers-promoter interactions both in different cell types or during embryonic development (Phillips-Cremins *et al.*, 2013; Andrey *et al.*, 2016; Freire-Pritchett *et al.*, 2017). Similar to what was observed in our case for the *HoxD* locus, the analysis of hundreds of viewpoints during limb development also identified two types of interactions (Andrey *et al.*, 2016). The first type is stable across tissues and stages of development, and correspond to CTCF/RAD21 co-binding

sites, similar to what we observe for island V and island II. The second type are tissue and cell type specific, and are thought to belong to regulatory regions, based on their chromatin marks profiles (Andrey *et al.*, 2017). The latter bare resemblance to some of the enhancer-promoter interactions we observe (e.g., GT2, Prox).

It is interesting to notice that contrary to the case of the trunk (Noordermeer *et al.*, 2014) and similar to the limb (Montavon *et al.*, 2011; Andrey *et al.*, 2013), in the GT new contacts with the C-DOM appear upon gene activation. These new contacts may not occur all in the same cell at the same time. It is probable that, in each allelic conformation *Hoxd13* is interacting with a subset of enhancers at a time. As GT development progresses and the total *Hoxd* transcript levels in the tissue are reduced, the 3D-chromatin conformation of the C-DOM becomes similar to that of mESC or forebrain (Figure 16). This way, while some contacts are constitutive, others disappear when the enhancers seemingly stop to work, as indicated by the loss of H3K27ac and gain of H3K27me3 in discrete domains. Furthermore, multiple regulatory elements interact with each other and the vicinity of *Hoxd13* pointing to a complex regulatory system to fine tune gene expression. Our results indicate a relationship between transcription and chromatin conformation, illustrating the various possibilities of interactions within a pre-existing chromatin structure.

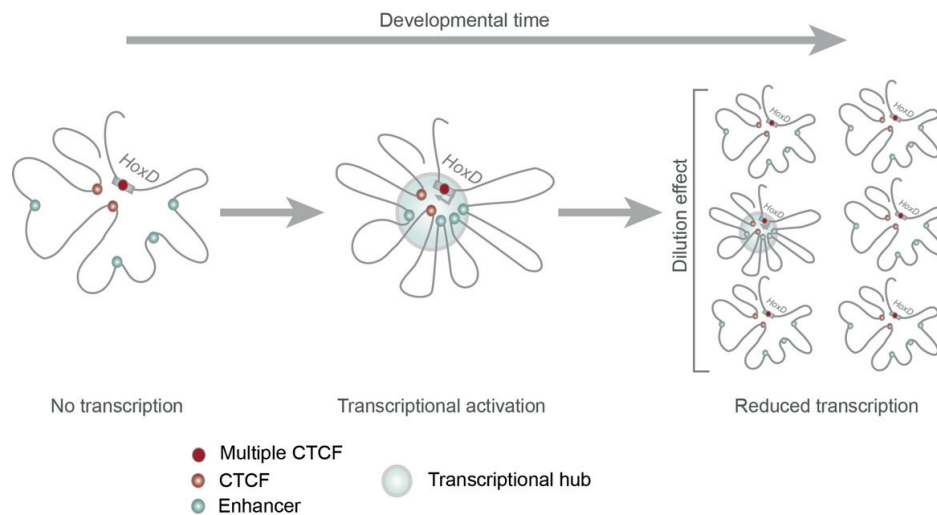


Figure 16: Schematic representation of the *HoxD* locus topology during GT development

In tissues or cells where *Hoxd* genes are not expressed there are constitutive interactions established between the cluster and some regulatory islands associated to CTCF/RAD21 co-binding sites. In GT, upon transcriptional activation, new contacts, associated with active enhancers, are formed. Of note, although in this schematic representation all integrations are depicted in the same transcriptional hub, this may not be the case. It is likely that at a given allele *Hoxd13* is interacting with only a subset of enhancers at a time. As development progresses, the total amount of *Hoxd* transcripts in the tissue are reduced and the spatial chromatin conformation at the *HoxD* locus becomes similar to that of non *Hoxd*

expressing tissues. This later observation is likely a consequence of only having a sub-population of cells with an “active conformation” (dilution effect).

4.3 The C-DOM regulatory potential: Effects of partial C-DOM deletions

As discussed above, in the GT, 5' *Hoxd* genes transcription regulation relies primarily on the presence of elements located in the C-DOM. Indeed, our analysis of the allelic series that corresponds to partial deletions of the C-DOM reflects the presence of several elements required for 5' *Hoxd* genes expression in the developing GT. With the exception of the *Del(SB-Atf2)* allele, all other deletions resulted with different degrees, in a down regulation of the expression of 5' *Hoxd* genes that are not completely cumulative.

The deletion between SB and *Atf2* resulted in a slight upregulation of *Evx2* and *Hoxd13* expression. Several, not mutually exclusive, explanations can account for this latter result. Firstly, some weak enhancers in this region may contact and sequester *Hoxd13*, thus preventing it to fully access more potent enhancers located at more proximal positions. Secondly, parts of this CTCF-rich region may have insulator properties. Finally, this effect may be a consequence of a global reorganization of the genomic landscape, as this deletion removes a TAD boundary. Interestingly, removing this region (between SB and *Atf2*) in the distal limb buds resulted in a down-regulation of *Hoxd* genes expression and phenotypic alteration of the digits (Montavon *et al.*, 2011) highlighting that even though both tissues share the same regulatory landscape (the C-DOM) some differences in regulation exist (Lonfat *et al.*, 2014).

Removing the region between the *rel5* and SB sequences resulted in the greatest reduction of expression suggesting that the sequences present in this region are of great relevance for the regulation of 5' *Hoxd* genes and *Evx2*. We further dissected the regulatory potential of this region by generating an allelic conformation that removed both island IV and III (*Del(SB-IV)*), while leaving island V and GT2 sequences intact. Using this allele, we observed a reduction in *Hoxd13* transcript levels that corresponded to half of what was observed with the *Del(rel5-SB)* allele. These results, in combination with our *in vivo* enhancer activity assays, suggested that both island IV and III and GT2 sequences are important for the regulation of *Hoxd13* in the GT, at least in combination.

The *Del(rel1-rel5)* allele also resulted in a decrease in 5' *Hoxd* genes and *Evx2* transcript levels, suggesting that elements in the region between the *rel1* and *rel5* sequences are important for the regulation of these genes. In this region we observed several ATAC-seq peaks, however with the exception of Prox and the promoter of *Lnp*, they did not correspond to 4C-seq or H3K27ac peaks in GT at E13.5. Furthermore, *in vivo* enhancer reporter assays for both the CsB

and CsA sequences showed no enhancer activity of these elements in the GT (Gonzalez *et al.*, 2007). Nonetheless, the ATAC-seq peak corresponding to island E showed enhancer activity in a sub-population of cells in the GT (Lonfat, PhD thesis). Although further analysis of full regulatory potential of the region between the *rel1* and *rel5* sequences is necessary, Prox seems to be the primary active enhancer in this region, as observed by both the presence of 4C-seq, ATAC-seq, and H3K27ac peaks as well as by the ability of this sequences to drive reporter gene transcription in the GT (Gonzalez *et al.*, 2007), and the single deletion of the Prox sequence.

Of interest, the different partial C-DOM deletions showed that *Hoxd12*, *Hoxd11*, and *Hoxd10* expression in GT is less affected than that of *Hoxd13* and *Evx2* by the sequences present in the C-DOM, suggesting that the enhancers found in this region have a stronger effect on *Hoxd13* and *Evx2* expression. Overall, in accordance with Lonfat *et al.*, 2014, we conclude that combined action of a collection of enhancers present in the C-DOM is important for the regulation of *Hoxd* genes in the GT.

4.4 Contribution of specific enhancer sequences to the regulation of *Hoxd* genes

Hox genes, as well as many other developmental genes, are regulated by the action of multiple enhancers (reviewed in (Spitz and Furlong, 2012; Long *et al.*, 2016)). To assess the importance of single *cis*-acting elements within the C-DOM to the regulation of 5' *Hoxd* genes and *Evx2* expression we have used two approaches. First, we evaluated the capacity of candidate sequences to drive transcription by *in vivo* reporter assays. Second, we evaluated the necessity of candidate regions for endogenous gene expression by generating mice, using CRISPR-Cas9 genome editing technology, with deletions of the target regions. Using these approaches, we tested both the sufficiency and necessity, respectively, of the candidate regulatory regions.

Our results suggest that no single enhancer sequence located between the *rel5* and SB sequences, seem to be essential in isolation, for proper expression of 5' *Hoxd* genes in both GT and DFL, at least at the developmental stages and tissues analyzed. On the contrary, the deletion of Prox resulted in a decrease in the transcript levels of 5' *Hoxd* genes in both the GT and DFL, suggesting that not all enhancers have the same functional strength within the C-DOM. A possible explanation for the lack of an obvious transcriptional deregulation upon deleting individual enhancers located between the *rel5* and SB sequences, is the presence of multiple enhancers associated with 5' *Hoxd* genes regulation that may have spatiotemporally

redundancy, rather than unique activity in both the GT and the distal forelimb buds. Regarding the GT, our transgenic assays in combination with previous results reveal that in some cases there is at least partial overlap in the functional domain of expression (e.g., GT2, GT1, Prox) (Lonfat, PhD thesis; Gonzalez *et al.*, 2007)), whereas in others there is a complementary domain of expression (e.g., IIIE and IVE). In conjunction with our single deletion results, these findings illustrate a degree of functional overlap between at least some enhancers regulating 5' *Hoxd* genes. In this view, the loss of one enhancer may be compensated by the presence of the others. Interestingly, it was recently shown that at several *loci* important for limb development, deletions of single enhancers caused no noticeable changes in limb morphology, suggesting that this apparent redundancy is a widespread feature of mammalian genomes (Osterwalder *et al.*, 2018).

The notion of redundancy or “shadow enhancers” within a regulatory region has been used to describe the activity of enhancers whose function partially or completely overlaps (Hong *et al.*, 2008). It is thus possible that the evolution of multiple enhancers in the C-DOM provided robustness to *Hoxd* genes expression. Indeed, with the exception of Prox, only upon removal of several regulatory elements (large deletions) did we observe changes in gene expression. Alternatively, our experimental approach is lacking the resolution to discern mild alterations in gene expression only occurring in a subpopulation of cells. Another possibility is that the single enhancer deletions analyzed have a more substantial role at different stages of development, or tissues that were not analyzed in this work.

4.5 A transcriptional hub: In-light of the phase separation model

Recent studies have provided evidence for the phase-separation model for transcriptional control (Hnisz *et al.*, 2017). At the super-enhancer level, the formation of RNA-Polymerase II (RNA Pol II), transcription factors (TF), and Mediator phase-separated condensates were shown to ensure robust gene transcription (Boija *et al.*, 2018; Sabari *et al.*, 2018). It is thus possible that the same principles are applied to long-range enhancer interactions to control gene expression.

Many enhancers seem to contribute to the expression of *Hoxd13*, and there is evidence suggesting that *Hoxd13* can interact with multiple elements (e.g., enhancers) in the same cell at the same time (Fabre *et al.*, 2017). In this view condensates would provide a permissive hub for transcription of a gene, where deleting “weak” enhancers in isolation would not be sufficient to elicit alterations in the physical properties necessary for condensate formation (e.g.,

Mediator; TFs) while deletion of bigger regions or a “stronger” enhancer (i.e., Prox) would (Figure 17). Furthermore, the slight alteration in the architecture of the *HoxD* locus observed upon the deletion of single regulatory regions seems not to be enough to perturb the transcriptional hub at least concerning *Hoxd* genes (Figure 17).

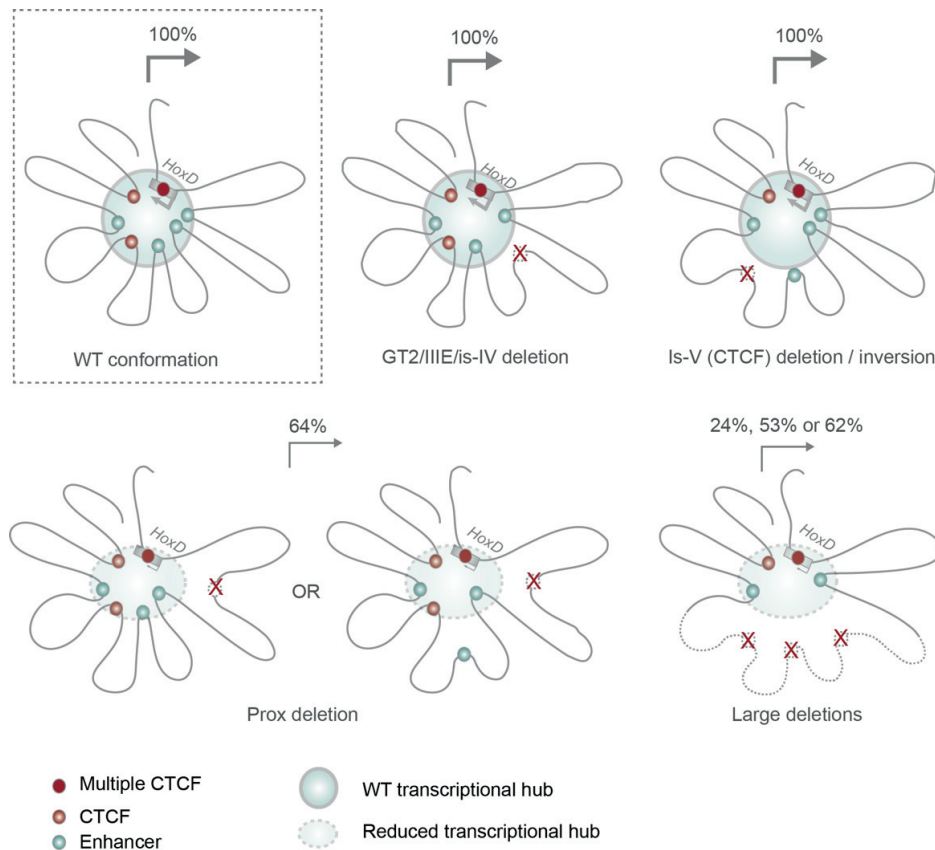


Figure 17: Effect of *cis*-elements deletions on the *HoxD* locus topology and gene expression
 Schematic representation of chromatin topology and transcriptional output in different genetic configurations at the *HoxD* locus. The wildtype (WT) topology and transcript levels are depicted on the top-left corner (dashed-line box). Upon deletion of either island IV, IIIIE, or GT2 no changes in transcription were observed. Deleting or inverting island V and its associated CTCF sites resulted in minor alterations in chromatin conformation, nonetheless they were not sufficient to perturb transcription. Deletion of Prox resulted in a reduction of transcription with potential alterations in chromatin interaction within the *HoxD* locus. Deletions comprising several regulatory elements resulted in reduction of *Hoxd* genes transcription. Of note, this schematic representation does not take into account the dynamic behavior of each individual element. Transcriptional hubs are depicted by blue circles or faded blue ellipses, representing dynamically shared transcriptional resources (e.g., RNA Pol II; TF). Arrows and percentages represent the levels of *Hoxd13* transcripts in each allelic conformation.

4.6 Not all enhancers behave the same: The case of Prox

While all the single regulatory element deletions between the rel5 and SB sequences showed no effect on the transcription of *Hoxd13*, the deletion of Prox resulted in a decrease in the transcript levels of this gene in both the GT and DFL. This observation suggests that not all enhancers have the same functional strength within the C-DOM. Prox is located closer to the cluster in linear distance and its interaction frequency, by 4C-seq, is higher than that of the enhancers located in-between the rel5 and SB sequences. One could think of a model where enhancer behavior depends on the frequency of the interaction with the promoter. In this view, Prox would interact more often and as such would contribute more frequently to the transcriptional hub, while the enhancers located between the rel5 and SB sequences (i.e., GT2, IV and III) would interact less frequently and in turn would contribute less often. The absence, in isolation, of any of the latter group of enhancers would be compensated by the remaining elements. Contrarily, in the absence of Prox, because it is more frequently in the transcriptional hub, in the wildtype situation, the total amount of transcription factors in the hub would be diminished and in turn impact the transcriptional outcome. It is also possible that deleting Prox has an effect on the interaction between the other enhancers and *Hoxd13* and as a consequence impact target gene transcription. Analysis of chromatin conformation would be of great help to resolve these questions.

4.7 Not all genes behave the same: The case of *Evx2*

Evx2, a mouse gene orthologue to the *Drosophila* pair-rule gene even-skipped (*eve*), is located 8kb upstream of *Hoxd13*. This physical proximity to the *HoxD* cluster is accompanied by a partial overlap in the expression pattern of both genes (Dollé *et al.*, 1994). Both *Hoxd13* and *Evx2* are expressed in a similar pattern during DFL and GT development (Dollé *et al.*, 1994; Herault *et al.*, 1996), and share a common regulatory landscape (the C-DOM) (Kmita *et al.*, 2002; Spitz *et al.*, 2003; Montavon *et al.*, 2011). Both of these genes are located in the same TAD and as such exposed to similar regulatory elements, suggesting at least some degree of overlap between the enhancer sequences controlling these two genes. Indeed, large deletions of the C-DOM affect the expression of these genes (Montavon *et al.*, 2011). Of note, while inactivation of *Evx2* in the mouse leads to very minor alterations in digits morphology, no changes in phenotype were reported for the GT (Herault *et al.*, 1996).

We observed a mild decrease in *Evx2* mRNA levels in the GT upon deleting GT2. As expected, because GT2 is not active in DFL, no change was observed in this tissue. The same

deletion did not affect the expression of *Hoxd13* in the GT. One possible explanation for this difference comes from our 4C-seq results using the *Evx2* viewpoint. In this context we observed changes in interactions with GT1 and island IV suggesting that in the case of *Evx2*, but not *Hoxd13*, GT2 is important for the interactions between the gene and other enhancers. In this conformation the absence of GT2 seems to impair the interactions between GT1 and *Evx2* and, island IV and *Evx2* possibly contributing to the decrease in *Evx2* transcript levels. When doing the converse experiment and placing a view point on island IV or GT1 (not shown) we observed a mild decrease in interactions, in the *Del(GT2)* allele, with both these enhancers and the CTCF site located in the *Evx2* gene body, supporting the previous results. However, upon using a viewpoint in *Hoxd13* we did not observe differences in the interaction frequencies with GT1 and island IV in the *Del(GT2)* allele suggesting that *Evx2* and *Hoxd13* are not in the same transcriptional hub. One could envision one allelic and/or cell conformation where *Evx2* will interact with multiple enhancers in a GT2 dependent way, where in the other conformation *Hoxd13* would interact with multiple enhancers in a GT2 independent way. As our analysis is done at a cell population level we are unable to distinguish between transcriptional hubs.

4.8 Effects of deleting or inverting island V

The work of our lab and others have suggested that intra-TAD CTCF/Cohesin constitutive interactions are pervasive feature during development (Phillips-Cremmins *et al.*, 2013; Noordermeer *et al.*, 2014; Andrey *et al.*, 2016). We wanted to evaluate the impact of deleting or inverting island V, a region shown to interact with *Hoxd13* in all the tissues and developmental stages analyzed thus far, on the genomic 3D-architecture of the *HoxD* locus. We have hypothesized that island V, and its associated CTCF sites, would be important to ensure proper chromatin architecture of the locus and in turn correct gene expression, as it is one of the two constitutive contacts maintained in the absence of transcription (island II being the other), and the only CTCF/Cohesin co-binding site between the cluster and island II (i.e., within a 550kb region of the C-DOM).

The deletion of island V had a negligible effect on the architecture of the regulatory landscape. This result is a plausible explanation for the lack of alterations in transcription observed upon deleting this island. Indeed, all enhancers in C-DOM specifically the ones between rel5 and SB, that when jointly deleted resulted in a 76% decrease in mRNA levels of *Hoxd13*, are still able to contact *Hoxd13* suggesting that other factors may be sufficient to account for those interactions. Possibly, these interactions reflect the activity of transcription

factors bound to these enhancers. The interactions between *Hoxd13* and island II, the next CTCF/Cohesin co-binding site convergent to the cluster, remained unaltered, as seen both by the *Hoxd13* viewpoint and the island II viewpoint (data not shown). It is thus likely that the deletion of island V has little impact in the formation and/or maintenance of the interactions between the CTCFs in the cluster and the remaining convergent CTCFs in the C-DOM.

The inversion of island V and its associated CTCF sites resulted in a decrease of interactions with *Hoxd13* and increase in interactions with both island II and island C. In this new allelic conformation, these islands are placed in a convergent position to the CTCF/Cohesin co-binding site on the inverted island V. This result is in accordance with work showing that the majority of CTCF mediated chromatin loops are established between sites positioned in opposed and convergent orientations (Rao *et al.*, 2014; de Wit *et al.*, 2015; Vietri Rudan *et al.*, 2015). Nonetheless, we need to formally show, by ChIP-seq, that the inversion of island V does not impair binding of CTCF and Cohesin. Our ATAC-seq data from the *Inv(V)* allele shows that the chromatin is open at the inverted island V suggesting the presence of CTCF and/or Cohesin in this site. However, whether the binding pattern is similar to the wildtype allele still needs to be evaluated.

Inverting island V resulted in slight modifications in intra-TAD structure, specifically it reduced the interaction frequency between island IV and *Hoxd13* (as seen from the island IV viewpoint). However it was not sufficient to have an impact on gene expression, in accordance with the RT-qPCR results for the *Del(IV)* allele. Furthermore, interactions between *Hoxd13* and other enhancers (e.g., GT2, Prox) do not seem to be greatly affected by the inversion of island V. Again, as in the case of the deletion of island V, some enhancer-promoter contacts seem to be formed independently of the orientation of this island. Overall, altering the CTCF-Cohesin mediated loop associated to island V, has negligible impact on long-range regulation of *Hoxd* genes within the C-DOM. Alternatively, we are missing the time resolution to observe the impact of removing these sites, and as a consequence at the timepoint analyzed the system was able to recover and ensure proper transcriptional output of the genes.

4.9 Conclusion

In this work we have analyzed the function of complex regulatory landscapes in controlling transcription of *Hoxd* genes during development. Using a combination of molecular and genetic approaches we have found a complex and robust system of regulation where several regulatory elements encompassed within a TAD play different roles in controlling gene expression. While perturbation of structural elements resulted in slight modifications in intra-TAD structure, they were not sufficient to disturb gene expression. Furthermore, though deletion of some single long-range acting enhancers had no detectable effect on transcription, the Prox element is necessary to ensure proper transcription of *Hoxd* genes, suggesting that not all elements within a complex regulatory landscape have the same functional strength, at least at a determined timepoint in development.

Similar concepts are likely to apply to other regulatory contexts as long-range regulation within complex regulatory landscapes is not specific to the *Hox* clusters. In recent years it has become evident that the non-coding part of the genome is important for the regulation of gene expression during development and, consequently, how failures in this regulation can cause disease. Further understanding of the contribution of complex regulatory landscape to gene regulation would be of great value to the understanding of both development and disease processes.

5. Materials and Methods

5.1 Mouse strains and genotyping

Genotyping of all alleles was done by PCR. Mouse tissue biopsies were lysed for 15' at 95°C, 800rpm, in lysis buffer (50mM NaOH, 0.2mM EDTA). For all genotyping reactions PCR was performed with a standardized cycling protocol (1x(94°3'), 2x(94°1'.62°1',72°1'), 30x(94°30''.62°30'',72°30''), 1x(72°10')). The primers used to genotype the *Del(rel1-rel5)*, *Del(rel5-SB)*, and *Del(SB-Atf2)* alleles are described in (Montavon *et al.*, 2011). Primers used to genotype the remaining alleles are listed in Table 1.

Table 1: List of genotyping primers

Allele	Primer name	Sequence	Product Size (bp)
<i>Del(Prox)</i>	590	GACTGTGTTTTGAGGAGGACAGTG	PCR 590-591: WT: 488bp PCR 594-597: Del:369bp
	591	CTGTGGCAGTAAGAGTGTGCTGAG	
	594	CCCAGCTGCAGCAGAAGACC	
	597	GTTTCAAAGGCAAGTCCATGACTCTCTG	
<i>Del(GT2)</i>	435	TGGGTACAGTTTGGCTCCAT	PCR 435-436: WT: 575bp PCR 433-436: Del: 500bp
	436	GCCCTTGGTGGCATGTTTAG	
	433	CTGCCACCTACCTTCTCCTC	
<i>Del(V) and Inv(V)</i>	RA7	GCTTGTGTCTTGCTGTGTCA	PCR RA7-RA8: WT: 419bp PCR RA7-RA10: Del:440bp PCR RA7- RA9: Inv: approx 300bp
	RA8	AAGGAAAGTGTGTGTGCTGG	
	RA9	CCGAATCCCTAGCTGTGCGAG	
	RA10	CATCTGTAGGTTCTGTGCTTATG	
<i>Del(IV)</i>	502	CAGTTCTTTCCACGGTGAGGAAGC	PCR 502-503: WT: 504bp PCR 501-507: Del: approx 450bp
	503	CACAAAGATGTGTCAAAGTTGGGGCTG	
	501	CCTCTTGAACCACGTGCATCGC	
	507	GCCAGCAGGAATAGCTATACTAACAGG	
<i>Del(III)</i>	545	GGAGAGCTCTGGGTGTGATTGC	PCR 545-542: Wt: 397bp PCR 537-544: Del: 600bp
	542	TGGGAAGTGTGGAGTCTTCTGCC	
	537	CCCTCCCCCTATCACTGTATCTCC	
	544	CAGACCTTTGCAGTAGGGTCATGG	
<i>Del(IV-SB)</i>	RA21	TCTGCCTCCGTTCTCACAAT	PCR RA21-RA22: Wt: 480bp PCR RA21-RA24: Del:300bp
	RA22	GGACCATCAAGAAGCATCCG	
	RA24	GACTAATCAAAGCCAGCA	

5.2 CRISPR-Cas9

With the exception of *Del(rel1-rel5)*, *Del(rel5-SB)*, and *Del(SB-Atf2)* alleles (Montavon *et al.*, 2011), all mouse strains carrying deletions or inversions of the different regulatory regions were generated using CRISPR–Cas9 genome editing technology.

Single guide RNAs (sgRNAs) were designed flanking the genomic regions of interest (5' and 3' to the regions of interest) using the crispr.mit.edu web tool (Zang lab) for the *Del(V)*, *Inv(V)*, and *Del(GT2)* alleles, or CCTop (Stemmer *et al.*, 2015) for the *Del(III)*, *Del(Pox)*, and *Del(IV-SB)* alleles (Table 2). All sgRNAs were cloned, as recommended in (Cong *et al.*, 2013), into the BbsI site of the pX330:hSpCas9 (Addgene ID 42230) vector.

The mouse strains *Del(V)*, *Inv(V)*, and *Del(GT2)* were produced by pronuclear injection (Mashiko *et al.*, 2013) of a mix of the two appropriated sgRNA cloned into the pX330:hSpCas9 vector (sgRNA:pX330:hSpCas9) (25 ng/μl each). The mouse strains *Del(IV-SB)*, *Del(IV)*, *Del(III)*, and *Del(Prox)* were produced by electroporation (Hashimoto and Takemoto, 2015) using a mix containing Cas9 mRNA (final concentration of 400ng/μl) and two sgRNAs (300ng/μl each) in Opti-MEM 1x injection buffer.

PCR based genotyping was carried out with primers designed on both sides of sgRNAs targets, with an approximate distance of 150-300bp from the cutting site (Table 1). Sanger sequence of positive PCR bands was used to identify and confirm the deletion or inversion breakpoints of the F0 funder animals (Figure 18).

Table 2: List of sgRNAs

Allele	5' sgRNA sequence	3' sgRNA sequence
<i>Del(Prox)</i>	ATCTATCACTATTTGCTCCC	ACGATACTTAAGACCTAGTC
<i>Del(GT2)</i>	CAGAGATCTGCTTGGCACGT	GCAGTGCGTGCTGGTTCGCC
<i>Del(V)/Inv(V)</i>	TCAAATCGTACAGCGCTGCC	CACGGGGTGGAGTTATCTAC
<i>Del(IV)</i>	CGCAAAGGCAGATGCGACTT	CTTGCCGCTTGACCTGCTAA
<i>Del(III)</i>	TAGAATCATTGGTAGTACGG	TGGTTCTAAGATTTCCGTAA
<i>Del(IV-SB)</i>	CTCTTGTTCTTGATCGACAT	CTTGCCGCTTGACCTGCTAA

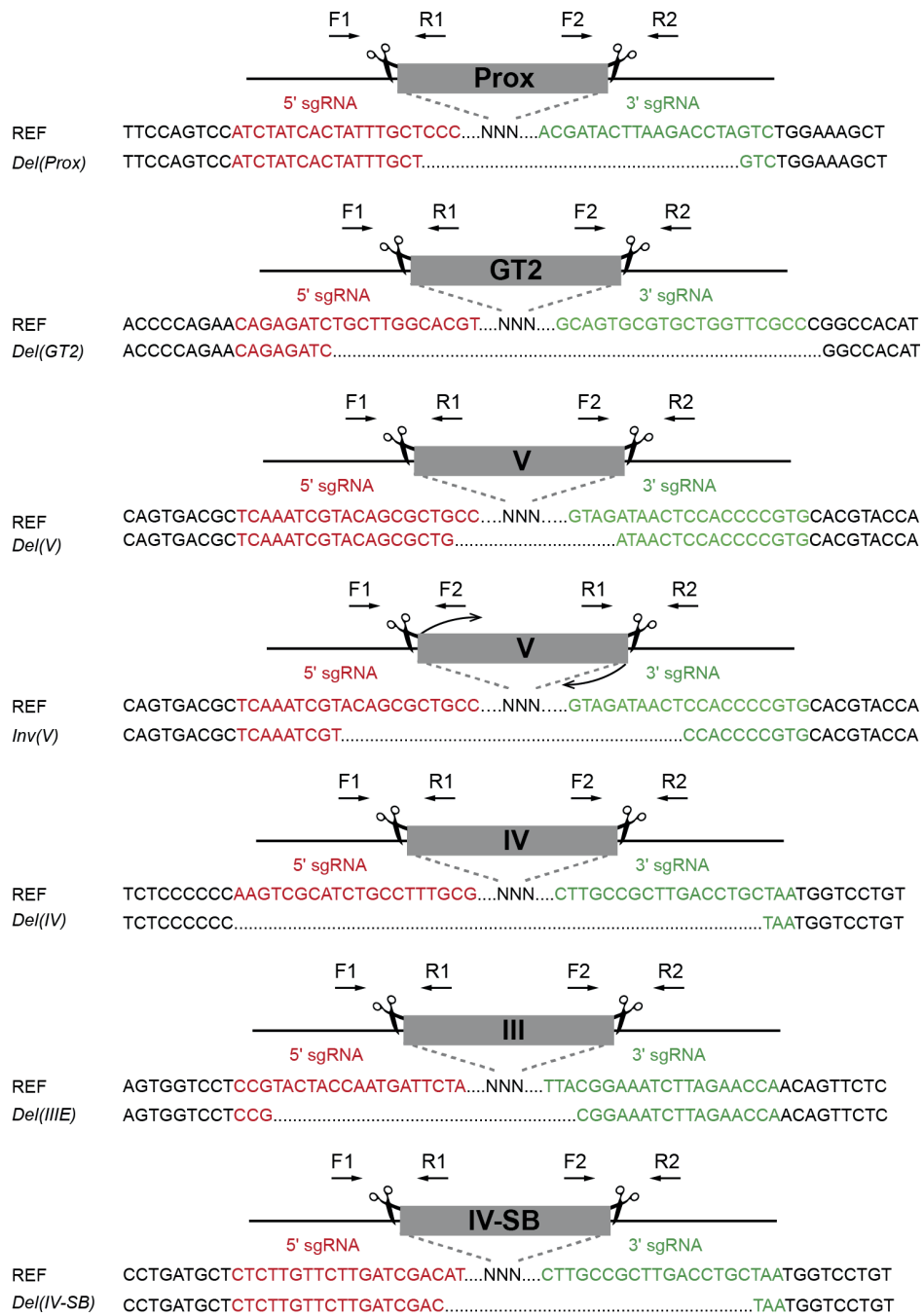


Figure 18: Alleles generated by CRISPR-Cas9

Sanger sequencing results of F0 animals for all alleles generated. Scissors indicate CRISPR-Cas9 mediated breakpoints flanking each regulatory region. SgRNAs are marked in red or in green. PCR based genotyping was carried out with primers designed on both sides of sgRNAs targets, deletions were screened with primers F1/R2, inversions with primers F1/F2 and R2/R1, and WT were amplified with primers F1/R1 and F2/R2.

5.3 Transgenic analysis

All mouse fosmid clones were obtained from BACPAC Resources Center (<https://bacpacresources.org>) (Table 3). Their integrity was verified by Sanger sequence and restriction enzyme fingerprinting. The fosmids were introduced in EL250 cells (Lee *et al.*, 2001) and targeted, by ET-recombineering, with a construct containing a *PI-SceI* restriction site, a *βglobin::LacZ* reporter gene with a FRT-flanked kanamycin selection marker, and flanked by 50 bp-long homology arms. The targeting constructs were produced by PCR amplification using the primers indicated in Table 4 to introduce the homology arms. The W11-D5 was shortened to remove the sequences that corresponded to island-IV. The targeted fosmids were selected at 30°C on LB plates containing chloramphenicol and kanamycin. The integrity of each modified fosmid was verified by restriction enzyme fingerprinting, and the correct integration of the *βglobin::LacZ* reporter gene was confirmed by PCR and Sanger sequence. All fosmids were linearized with *PI-SceI* and micro-injected into mouse oocytes. Embryos were harvested at E13.5 and stained for β-galactosidase activity (according to standards procedures). A minimum of three transgenic animals with consistent staining were obtained per construct. The transgenic mouse embryos for either the Prox or GT2 were obtained as described in (Gonzalez *et al.* 2007; Lonfat *et al.*, 2014).

Table 3: List of fosmids

Clone Name	Coordinates	Length
W11-1879J12 (island IV)	chr2:74235922-74281101	45180 bps
W11-2556D5 (island III)	chr2:74180517-74223072	42556 bps
W11-1741G6 (island V)	chr2:74270669-74309212	38544 bps
W11-1129A16 (between is-III/SB)	chr2:74141402-74180933	39532 bps
W11-109B16 (between is-V/GT2)	chr2:74346499-74384278	37780 bps

Table 4: List of primers used for recombineering

Primer name	Sequence 5' > 3'
W11-2556D5 (island III) Fw	CAGTTCTCAGTCTTCAACTTGCTGAGTCAAAAATCTGTTGTTCTGTATTG CTCGAGGTCGACGGTATCG
W11-2556D5 (island III) Rev	CTTTTCTGTGCTATTCTGAGGAGTGTGGTGTGTTACTTCTGGGAAGTG ATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGG CGTCCGCCATCTCCAGCAGC
W11-1879J12 (island IV) Fw	AATAAATGAACAGCGCTGCTCAGCTGCCCCCTCCCCGTGGCTAGGCATT GCTCGAGGTCGACGGTATCG
W11-1879J12 (island IV) Rev	TCTTAACAGAGTAGCTTCCTATGTGGAATGTCTGTGGTGGAAAGCGAGG AATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGG CGTCCGCCATCTCCAGCAGC

W11-1741G6 (island V) Fw	<u>ACTGTTCCCTAATTAATACTAAGCCATATGCAAAAACATTCTTGAATACC</u> <u>ATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGGGCTCGAGGTGACGG</u> <u>TATCG</u>
W11-1741G6 (island V) Rev	<u>TCTTAACAGAGTAGCTTCCTATGTGGAATGTCTGTGGTGGAAAGCGAGG</u> <u>A CGTCCGCCATCTCCAGCAGC</u>
W11-1129A16 (between is- III/SB) Fw	<u>TTTTGATCCTGAGGCAGCGGAAGACTGTGCCACACTAGATGTAACCTGAG</u> <u>GCTCGAGGTGACGGTATCG</u>
W11-1129A16 (between is- III/SB) Rev	<u>TGATTCTTACCAGACTCTAGTTGTCTGCATCCCAAGTTCCTACTAGGAC</u> <u>ATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGG</u> <u>CGTCCGCCATCTCCAGCAGC</u>
W11-109B16 (between is- V/GT2) Fw	<u>CACCCAGTTTTTTAATCCTGCCGGATCTTGCATTGACGGTGGCTGTAA</u> <u>GCTCGAGGTGACGGTATCG</u>
W11-109B16 (between is- V/GT2) Rev	<u>CAAGCATCTCCTCTTTGGTGTCTCAGAAGCAGTTGGCTAAATGAGCTCCT</u> <u>ATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGG</u> <u>CGTCCGCCATCTCCAGCAGC</u>

Homology Arm - restriction site (PISceI) - **Primer**

5.4 X-gal staining

Embryos were stained using standard procedures. Briefly, whole embryos (E13.5) were fixed in 4% paraformaldehyde at 4°C for 35 min, stained in a solution containing 1 mg/ml X-gal at 37°C overnight, washed in PBS, imaged, and stored in 4% paraformaldehyde.

5.5 Whole-mount *in situ* hybridization

Whole-mount *in situ* hybridization (WISH) was performed according to (Woltering *et al.*, 2014). Briefly, embryos were dissected in PBS and fixed overnight in 4% paraformaldehyde (PFA), washed in PBS, dehydrated, and stored in 100% methanol at -20°C. Rehydration was performed by a series of methanol/TBS-T washes, followed by a short digestion of Proteinase K, and re-fixation in 4% PFA. Pre- prehybridization, hybridization, and post-hybridization steps were carried out at 67°C. For all genotypes, both mutant and control wildtype (E12.5) littermate embryos were processed in parallel to maintain identical conditions throughout the WISH procedure. DIG-labeled probes for *in situ* hybridizations were produced by *in vitro* transcription (Promega) and detection was carried out using an alkaline phosphatase conjugated anti-digoxigenin antibody (Roche). WISH probes templates were previously described in: *Hoxd13* (Dollé *et al.*, 1991), *Evx2* (Herault *et al.*, 1996). For detection the chromogenic substrates NBT/BCIP or BM-purple were used.

5.6 RT-qPCR

Before processing, all tissues were stored at -80°C in *RNAlater* stabilization reagent (Invitrogen). RNA was extracted from single micro-dissected GT (E12.5) or one pair of distal forelimbs (E12.5), using Qiagen Tissue Lyser and RNeasy Plus kit (Qiagen), according to the manufacturer's instructions. RNA was reversed transcribed using Superscript III (Invitrogen) or Superscript IV (Invitrogen) and random hexamers. qPCR was performed on a CFX96 real-time system (BioRad) using GoTaq qPCR Master Mix (Promega). Primers were previously described in (Montavon *et al.*, 2008). Three technical replicates were used per biological replicate. Relative gene expression levels were calculated by the $2^{-\Delta\text{Ct}}$ method using a reference gene. *Tub β* was chosen as internal control and the mean of wildtype control samples was set as reference to calculate the ratio between the different samples. Graphical representation and statistical analysis were performed with GraphPad Prism 7.

5.7 4C-seq

5.7.1 Sample preparation

Circular chromosome conformation capture (4C-seq) was performed as described in (Noordermeer *et al.*, 2011). Briefly, tissues (20-40 GT or 40 cloaca) were isolated in PBS supplemented with 10% Fetal Calf Serum and dissociated to single cell by collagenase treatment. Samples were fixed in 2% formaldehyde, lysed, and stored at -80°C . Pools of between 20-40 GT or 40 Cloaca were primarily digested with *NlaIII* (NEB, R0125L) followed by ligation under diluted conditions. After decrosslinking and DNA purification *DpnII* (NEB, R0543M) was used for the second restriction. All ligation steps were performed using highly concentrated T4 DNA ligase (Promega, M1794). For each viewpoint approximately $1\mu\text{g}$ of DNA was amplified by using 12 individual PCR reactions. Libraries were constructed with primers for different viewpoints (see Table 5) containing Illumina Solexa adapter sequences and sequenced on an Illumina HiSeq 2500 sequencer, as single-end reads (read length 100 base pairs). In some samples 4-bp barcodes were added between the adapter and each specific viewpoint to allow sample multiplexing.

5.7.2 Data processing

4C-seq reads were demultiplexed, mapped on GRCm38/mm10 mouse assembly, and analyzed using the 4C-seq pipeline of the Bioinformatics and Biostatistics Core Facility (BBCF)

HTSstation (<http://htsstation.epfl.ch>) (David *et al.*, 2014). Profiles were normalized to a 5Mb region surrounding the *HoxD* cluster and smoothed using a window size of 11 fragments. 4C-seq peaks were called using the peak calling algorithm (PeakC) (de Wit *et al.*, 2015). C-DOM and T-DOM quantifications in Figure 5 were performed as in (Yakushiji-kaminatsui *et al.*, 2018). C-DOM quantifications on Figure 6 were done by dividing the sum of the scores in the C-DOM (chr2: 73,921,943–74,648,943) by the sum of the scores that fall in a non-interacting region of the T-DOM (chr2:75,166,258-75,571,741) (background local normalization). Quantifications of the interactions established with the cis-regulatory elements (in Figure 6 and 8) were calculated as a percentage of the sum of the scores of each element using the mESC sample as a reference.

Table 5: List of 4C-seq primers

Viewpoint		Sequence (5' > 3')
<i>Hoxd13</i>	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXAAAATCCTAGACCTGGTCATG
	iR	CAAGCAGAAGACGGCATAACGAGGCCGATGGTGCTGTATAGG
<i>Evx2</i>	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXTTGGAAACCCTGTGAGCCTAC
	iR	CAAGCAGAAGACGGCATAACGAGGGAAGAAACCTACCACGACAC
GT2	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXTTCTCTCTTTTAGTGACCTTGAACA
	iR	CAAGCAGAAGACGGCATAACGAGAAATATCCAAAGGTAAAAATCAAGAA
island V	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXGCTACAAGACTCATTCGTTAA
	iR	CAAGCAGAAGACGGCATAACGAACTAACTTAAGTCCCCTCG
island IV	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXTACAGCCTAGTCTTTTCTCATCACAT
	iR	CAAGCAGAAGACGGCATAACGATGTAATTATTTCAGGGTTGGAGTAGAATCA
island III	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXCTGCTTGAGTGAATACCTGT
	iR	CAAGCAGAAGACGGCATAACGATCACAGGCTGAGAAAGTCT
island II	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXGCATTCATCAAGCTGTGATTAGCA
	iR	CAAGCAGAAGACGGCATAACGAAATCCCATAAATGTAGACTGTAGTGTTGC
island C	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXATGATTTGAGAGGACTATTATGC
	iR	CAAGCAGAAGACGGCATAACGAAAGTTGGGCTTGAGATAACA
GT1	iF	AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTTCCGATCT XXXXTTCTAAAATTAACAACAAAGTTCATG
	iR	CAAGCAGAAGACGGCATAACGAGGCGTCTGTGCATTATTAA

XXXX – Corresponds to possible barcode sequences

5.8 ChIP-seq

5.8.1 Sample preparation

Micro-dissected 20-40 GT or 70 Cloaca were crosslinked in 1% formaldehyde/PBS for 20 min and stored at -80°C until further processing. Chromatin was sheared using a water-bath sonicator (Covaris E220 evolution ultra-sonicator). Immunoprecipitation was done using the following antibodies, anti-CTCF (Active Motif, 61311), anti-RAD21 (Abcam, ab992), anti-H3K27ac (Abcam, ab4729), and H3K27me3 (Merck Millipore, 07-449). Libraries were prepared using the TruSeq protocol, and sequenced on the Illumina HiSeq system (100bp single-end reads) according to manufactures instructions.

5.8.2 Data analysis

ChIP-seq reads processing was done on the Duboule lab local Galaxy server (Guerler *et al.*, 2016). Adapters and bad-quality bases were removed with Cutadapt version 1.16 (Martin, 2011) options -m 15 -q 30 -a GATCGGAAGAGCACACGTCTGAACTCCAGT CAC. Reads were mapped to the mouse genome (mm10) using Bowtie2 (v2.3.4.1) (Langmead and Salzberg, 2012), with standard settings. The coverage was obtained as the output of MACS2 (v2.1.1.20160309) (Zhang *et al.*, 2008). CTCF motif orientation was assessed using the CTCFBSDB 2.0 database (Ziebarth *et al.*, 2013), with EMBL_M1 identified motifs.

5.9 ATAC-seq

5.9.1 Sample preparation

ATAC-seq was performed as described in (Buenrostro *et al.*, 2013). Briefly, micro-dissected tissues (a poll of 2 GT or 2-3 cloaca) were isolated in PBS supplemented with 10% Fetal Calf Serum and dissociated to single cell by collagenase treatment. After isolation, 50,000 cells were lysed in 50 µl of lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂ and 0.1% IGEPAL CA-630), nuclei were carefully resuspended in 50µl transposition reaction mix (25µl TD buffer, 2.5µl Tn5 transposase and 22.5µl nuclease-free water) and incubated at 37 °C for 30min. DNA was isolated with a MinElute DNA Purification Kit (Qiagen). Library amplification was performed by PCR (10 to 12 cycles) using NEBNext High-Fidelity 2x PCR Master Mix (NEB, M0541S). Library quality was checked on a

fragment analyzer, and paired-end sequencing was performed on an Illumina NextSeq 500 instrument (read length 2×37 base pairs).

5.9.2 Data analysis

ATAC-seq reads processing was done on the Duboule lab local Galaxy server (Guerler *et al.*, 2016). Reads were mapped to the mouse genome (mm10) using Bowtie2 (v2.3.4.1) (Langmead and Salzberg, 2012), (-I 0 -X 2000 --fr --dovetail --very-sensitive-local). Reads with mapping quality below 30, mapping to mitochondria, or not properly paired were removed from the analysis. PCR duplicates were filtered using Picard (v1.56.0). Peak calling was done using MACS2 (v2.1.0.20151222) call peak (--cutoff qvalue 0.05 --nomodel --shift -100 --extsize 200 --call-summits). The coverage was done using the center of the Tn5 insertion and extended on both sides by 20bp (script developed by Lucille Delisle). When indicated, some coverage profiles represent an average of the replicates, this was done by dividing each replicate by the number of million reads that fall within peaks in each sample (for normalization) and calculating the average coverage.

5.10 RNA-seq

5.10.1 Sample preparation

Micro-dissected GT from different embryonic stages were individual stored at -80°C in RNAlater stabilization reagent (Ambion) before further sample processing. Total RNA was extracted from tissues using Qiagen RNeasy Plus Micro Kit (Qiagen) after disruption and homogenization. RNA quality was assessed using an Agilent 2100 Bioanalyser. Only samples with high RNA integrity number were used. Sequencing libraries were prepared according to TruSeq Stranded mRNA Illumina protocol, with polyA selection. RNA-seq libraries were sequenced on an Illumina HiSeq 2500 sequencer, as single-end reads (read length 100 base pairs).

5.10.2 Data analysis

Raw RNA-seq reads were aligned on the mouse mm10 genome assembly using TopHat 2.0.9 (Yates *et al.*, 2016). Gene expression computations were performed using uniquely mapping reads extracted from TopHat alignments and genomic annotations from Ensembl release 82 (Kim *et al.*, 2013) FPKM (fragments per kilo- base per million mapped fragments)

expression levels for each gene were calculated using Cufflinks (Roberts *et al.*, 2011). Differential gene expression analysis was done using DESeq2 (Love *et al.*, 2014) in R.

6. Bibliography

- Amores, A., Force, A., Yan, Y.-L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y.-L., Westerfield, M., Ekker, M. and Postlethwait, J. H. (1998) 'Zebrafish hox Clusters and Vertebrate Genome Evolution', *Science*, 282(5394), p. 1711 LP-1714. doi: 10.1126/science.282.5394.1711.
- 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 (New York, NY)*, 340(June), p. 1234167. doi: 10.1126/science.1234167.
- Andrey, G., Schöpflin, R., Jerković, I., Heinrich, V., Ibrahim, D. M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M. and Mundlos, S. (2016) 'Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding.', *Genome research*, p. gr.213066.116. doi: 10.1101/gr.213066.116.
- Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M. and Stark, A. (2013) 'Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq', *Science*, 339(6123), p. 1074 LP-1077. doi: 10.1126/science.1232542.
- 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 Genetics*, 9(12). doi: 10.1371/journal.pgen.1004018.
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L. and Lander, E. S. (2006) 'A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells', *Cell*, pp. 315–326. doi: 10.1016/j.cell.2006.02.041.
- Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnesse, A., Coffey, E. L., Zamudio, A. V., Li, C. H., Shrinivas, K., Manteiga, J. C., Hannett, N. M., Abraham, B. J., Afeyan, L. K., Guo, Y. E., Rimel, J. K., Fant, C. B., Schuijers, J., Lee, T. I., Taatjes, D. J. and Young, R. A. (2018) 'Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains', *Cell*, pp. 1–14. doi: 10.1016/j.cell.2018.10.042.
- Bonev, B. and Cavalli, G. (2016) 'Organization and function of the 3D genome', *Nature Review Genetics*, 17, pp. 661–678. doi: 10.1038/nrg.2016.112.
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.', *Nature methods*, 10(12), pp. 1213–8. doi: 10.1038/nmeth.2688.
- Cannavò, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., Ciglar, L., Korb, J. O. and Furlong, E. E. M. (2015) 'Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks', *Current Biology*, pp. 38–51. doi: 10.1016/j.cub.2015.11.034.
- Cao, R., Wang, L., Wang, H. and Xia, L. (2002) 'Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing', *Science*, 298(November), pp. 1039–1044.
- Chisaka, O. and Capecchi, M. R. (1991) 'Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene hox-1.5', *Nature*, 350, p. 473. Available at: <https://doi.org/10.1038/350473a0>.
- Cohn, M. J. (2011) 'Development of the external genitalia: conserved and divergent

- mechanisms of appendage patterning.’, *Developmental Dynamics*, 240(5), pp. 1108–15. doi: 10.1002/dvdy.22631.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Hsu, P. D., Wu, X., Jiang, W. and Marraffini, L. a (2013) ‘Multiplex Genome Engineering Using CRISPR/Cas Systems’, *Science*, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.Multiplex.
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. a, Frampton, G. M., Sharp, P. a, Boyer, L. a, Young, R. a and Jaenisch, R. (2010) ‘Histone H3K27ac separates active from poised enhancers and predicts developmental state.’, *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp. 21931–21936. doi: 10.1073/pnas.1016071107.
- 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(1). doi: 10.1371/journal.pone.0085879.
- 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 Reports*, 5(1), pp. 137–150. doi: 10.1016/j.celrep.2013.09.002.
- Deschamps, J. and Nes, J. Van (2005) ‘Developmental regulation of the Hox genes during axial morphogenesis in the mouse’, *Development*, pp. 2931–2942. doi: 10.1242/dev.01897.
- Di-poi, N., Montoya-burgos, J. I., Miller, H., Pourquie, O., Milinkovitch, M. C. and Duboule, D. (2010) ‘Changes in Hox genes ’ structure and function during the evolution of the squamate body plan’, *Nature*, 464(March), pp. 99–103. doi: 10.1038/nature08789.
- 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(7398), pp. 376–80. doi: 10.1038/nature11082.
- Dolle, 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(3), pp. 431–441. doi: 10.1016/0092-8674(93)90378-4.
- 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’, *Development*, pp. 143–153.
- Dolle, P., Izpisua-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(6251), pp. 767–772. Available at: <http://dx.doi.org/10.1038/342767a0>.
- Dollé, P., Izpisúa-Belmonte, J. C., Brown, J. M., Tickle, C. and Duboule, D. (1991) ‘Hox-4 genes and the morphogenesis of mammalian genitalia’, *Genes and Development*, 5(10), pp. 1767–1776. doi: 10.1101/gad.5.10.1767.
- Duboule, D. (1991) ‘Patterning in the vertebrate limb’, *Current Opinion in Genetics & Development*.
- Duboule, D. (2007) ‘The rise and fall of Hox gene clusters.’, *Development*, 134(14), pp. 2549–60. doi: 10.1242/dev.001065.
- Duboule, D. and Dollé, P. (1989) ‘The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes.’, *The EMBO journal*, 8(5), pp. 1497–1505.

- Duboule, D. and Morata, G. (1994) 'Colinearity and functional hierarchy among genes of the homeotic complexes', *Trends in Genetics*, 10(10), pp. 358–364. doi: 10.1016/0168-9525(94)90132-5.
- Economides, K. D., Zeltser, L. and Capecchi, M. R. (2003) 'Hoxb13 mutations cause overgrowth of caudal spinal cord and tail vertebrae', *Developmental Biology*, 256(2), pp. 317–330. doi: [https://doi.org/10.1016/S0012-1606\(02\)00137-9](https://doi.org/10.1016/S0012-1606(02)00137-9).
- ENCODE Project Consortium (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*. doi: 10.1038/nature11247.
- Fabre, P. J., Leleu, M., Mormann, B. H., Lopez-Delisle, L., Noordermeer, D., Beccari, L. and Duboule, D. (2017) 'Large scale genomic reorganization of topological domains at the HoxD locus', *Genome Biology*, 18(1), pp. 1–15. doi: 10.1186/s13059-017-1278-z.
- Favier, B. and Dollé, P. (1997) 'Developmental functions of mammalian Hox genes.', *Molecular human reproduction*, 3(2), pp. 115–31. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9239717>.
- Forlani, S., Lawson, K. A. and Deschamps, J. (2003) 'Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo', *Development*, pp. 3807–3819. doi: 10.1242/dev.00573.
- Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F. and Stern, D. L. (2010) 'Phenotypic robustness conferred by apparently redundant transcriptional enhancers.', *Nature*, 466(7305), pp. 490–3. doi: 10.1038/nature09158.
- Freire-Pritchett, P., Schoenfelder, S., Várnai, C., Wingett, S. W., Cairns, J., Collier, A. J., García-Vílchez, R., Furlan-Magaril, M., Osborne, C. S., Fraser, P., Rugg-Gunn, P. J. and Spivakov, M. (2017) 'Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells', *eLife*, 6, pp. 1–26. doi: 10.7554/eLife.21926.
- García-Fernández, J. (2005) 'Hox, ParaHox, ProtoHox: facts and guesses.', *Heredity*, 94(2), pp. 145–152. doi: 10.1038/sj.hdy.6800621.
- Gaunt, S. J., Sharpe, P. T. and Duboule, D. (1988) 'Spatially restricted domains of homeo-gene transcripts in mouse embryos : relation to a segmented body plan', *Development*, 179, pp. 169–179.
- Georgas, K. M., Armstrong, J., Keast, J. R., Larkins, C. E., McHugh, K. M., Southard-Smith, E. M., Cohn, M. J., Batourina, E., Dan, H., Schneider, K., Buehler, D. P., Wiese, C. B., Brennan, J., Davies, J. a., Harding, S. D., Baldock, R. a., Little, M. H., Vezina, C. M. and Mendelsohn, C. (2015) 'An illustrated anatomical ontology of the developing mouse lower urogenital tract', *Development*, 142(10), pp. 1893–1908. doi: 10.1242/dev.117903.
- Georgas, K. M., Armstrong, J., Keast, J. R., Larkins, C. E., McHugh, K. M., Southard-Smith, E. M., Cohn, M. J., Batourina, E., Dan, H., Schneider, K., Buehler, D. P., Wiese, C. B., Brennan, J., Davies, J. A., Harding, S. D., Baldock, R. A., Little, M. H., Vezina, C. M. and Mendelsohn, C. (2015) 'An illustrated anatomical ontology of the developing mouse lower urogenital tract', *Development*, 142(10), pp. 1893–1908. doi: 10.1242/dev.117903.
- Ghavi-Helm, Y., Klein, F. a., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W. and Furlong, E. E. M. (2014) 'Enhancer loops appear stable during development and are associated with paused polymerase', *Nature*, 512(7512), pp. 96–100. doi: 10.1038/nature13417.
- Gonen, N., Futtner, C. R., Wood, S., Garcia-moreno, S. A., Salamone, I. M., Samson, C., Sekido, R., Poulat, F., Maatouk, D. M. and Lovell-badge, R. (2018) 'Sex reversal following

- deletion of a single distal enhancer of Sox9', *Science*, 9408(June), pp. 1–10.
- Gonzalez, F., Duboule, D. and Spitz, F. (2007) 'Transgenic analysis of Hoxd gene regulation during digit development', *Developmental Biology*, 306(2), pp. 847–859. doi: 10.1016/j.ydbio.2007.03.020.
- Graham, A., Papalopulu, N., Krumlauf, R., Ridgeway, T. and Hill, M. (1989) 'The Murine and Drosophila Homeobox Complexes Have Common Features of Organization and Expression', *Cell*, 57, pp. 367–378.
- Guerler, A., Eberhard, C., Baker, D., Clements, D., Afgan, E., Turaga, N., Taylor, J., van den Beek, M., Blankenberg, D., Bouvier, D., Hillman-Jackson, J., Chilton, J., Čech, M., Coraor, N., Nekrutenko, A., Grüning, B., Von Kuster, G., Rasche, E., Soranzo, N. and Goecks, J. (2016) 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update', *Nucleic Acids Research*, 44(W1), pp. W3–W10. doi: 10.1093/nar/gkw343.
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., Ren, B., Krainer, A. R., Maniatis, T. and Wu, Q. (2015) 'CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function.', *Cell*, 162(4), pp. 900–10. doi: 10.1016/j.cell.2015.07.038.
- Haack, H. and Gruss, P. (1993) 'The Establishment of Murine Hox-1 Expression Domains during Patterning of the Limb', *Developmental Biology*, 157(2), pp. 410–422. doi: <https://doi.org/10.1006/dbio.1993.1145>.
- Haraguchi, R., Mo, R., Hui, C., Motoyama, J., Makino, S., Shiroishi, T., Gaffield, W. and Yamada, G. (2001) 'Unique functions of Sonic hedgehog signaling during external genitalia development', *Development*, 128(21), pp. 4241–4250.
- Haraguchi, R., Suzuki, K., Murakami, R., Sakai, M., Kamikawa, M., Kengaku, M., Sekine, K., Kawano, H., Kato, S., Ueno, N. and Yamada, G. (2000) 'Molecular analysis of external genitalia formation: the role of fibroblast growth factor (Fgf) genes during genital tubercle formation.', *Development*, 127(11), pp. 2471–2479.
- Hashimoto, M. and Takemoto, T. (2015) 'Electroporation enables the efficient mRNA delivery into the mouse zygotes and facilitates CRISPR / Cas9-based genome editing', *Scientific reports*, (May), pp. 1–3. doi: 10.1038/srep11315.
- Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Sharpe, J. A., Suci, M. C., Telenius, J., Williams, R., Rode, C., Li, P., Pennacchio, L. A., Sloane-stanley, J. A., Ayyub, H., Butler, S., Sauka-spengler, T., Gibbons, R. J., Smith, A. J. H., Wood, W. G. and Higgs, D. R. (2016) 'Genetic dissection of the -globin super-enhancer in vivo', *Nature genetics*, 48(8). doi: 10.1038/ng.3605.
- Herault, Y., Hraba-renevey, S., Hoeven, F. Van Der and Duboulet, D. (1996) 'Function of the Evx-2 gene in the morphogenesis of vertebrate limbs', *The EMBO journal*, 15(23), pp. 6727–6738.
- Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. and Sharp, P. A. (2017) 'A Phase Separation Model for Transcriptional Control', *Cell*, 169(1), pp. 13–23. doi: 10.1016/j.cell.2017.02.007.
- Holland, P. W. H. (2013) 'Evolution of homeobox genes', *WIREs Dev Biol*, 2(February). doi: 10.1002/wdev.78.
- Holland, P. W. H. (2015) 'Did homeobox gene duplications contribute to the Cambrian

- explosion?', *Zoological letters*, pp. 1–8. doi: 10.1186/s40851-014-0004-x.
- Hong, J.-W., Hendrix, D. A. and Levine, M. S. (2008) 'Shadow enhancers as a source of evolutionary novelty', *Science*, 321(5894), p. 1314. doi: 10.1126/science.1160631.
- Horan, G. S., Ramírez-Solis, R., Featherstone, M. S., Wolgemuth, D. J., Bradley, A. and Behringer, R. R. (1995) 'Compound mutants for the paralogous *hoxa-4*, *hoxb-4*, and *hoxd-4* genes show more complete homeotic transformations and a dose-dependent increase in the number of vertebrae transformed.', *Genes & Development*, 9(13), pp. 1667–1677. doi: 10.1101/gad.9.13.1667.
- Hostikka, S. L. and Capecchi, M. R. (1998) 'The mouse *Hoxc11* gene: Genomic structure and expression pattern', *Mechanisms of Development*, 70(1–2), pp. 133–145. doi: 10.1016/S0925-4773(97)00182-2.
- Hostikka, S. L., Gong, J. and Carpenter, E. M. (2009) 'Axial and appendicular skeletal transformations, ligament alterations, and motor neuron loss in *Hoxc10* mutants', *International Journal of Biological Sciences*, 5(5), pp. 397–410.
- 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.', *The EMBO journal*, 10(8), pp. 2279–2289.
- Jerković, I., Ibrahim, D. M., Andrey, G., Haas, S., Hansen, P., Janetzki, C., González Navarrete, I., Robinson, P. N., Hecht, J. and Mundlos, S. (2017) 'Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCF', *PLOS Genetics*, 13(1), p. e1006567. doi: 10.1371/journal.pgen.1006567.
- Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C., Schmitt, A. D., Espinoza, C. A. and Ren, B. (2013) 'A high-resolution map of the three-dimensional chromatin interactome in human cells', *Nature*, 503(7475), pp. 290–294. doi: 10.1038/nature12644.
- Juan, A. H. and Ruddle, F. H. (2003) 'Enhancer timing of Hox gene expression: deletion of the endogenous *Hoxc8* early enhancer', *Development*, 130(20), p. 4823 LP-4834. doi: 10.1242/dev.00672.
- Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., Berkum, N. L. Van, Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J. and Young, R. A. (2010) 'Mediator and cohesin connect gene expression and chromatin architecture', *Nature*, 467(7314), pp. 430–435. doi: 10.1038/nature09380.
- Kaufman, T. C., Lewis, R. and Wakimoto, B. (1980) 'Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: The homeotic gene complex in polytene chromosome interval 84A-B', *Genetics*, 94(1), pp. 115–133.
- Kessel, M. and Gruss, P. (1991) 'Homeotic Transformations of Murine Vertebrae and Concomitant Alteration of Hox Codes Induced by Retinoic Acid', *Cell*, 67(1894).
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L. (2013) 'TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.', *Genome biology*. BioMed Central Ltd, 14(4), p. R36. doi: 10.1186/gb-2013-14-4-r36.
- Kmita, M., Tarchini, B., Duboule, D. and Hérault, Y. (2002) 'Evolutionary conserved sequences are required for the insulation of the vertebrate *Hoxd* complex in neural cells', *Development*, pp. 5521–5528. doi: 10.1242/dev.00151.
- Kondo, T., Zákány, J., Innis, J. W. and Duboule, D. (1997) 'Of fingers, toes and penises.', *Nature*, 390(6655), p. 29. doi: 10.1038/36234.

- Ku, M., Koche, R. P., Rheinbay, E. and Mendenhall, E. M. (2008) ‘Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains’, *PLoS Genetics*, 4(10). doi: 10.1371/journal.pgen.1000242.
- de Laat, W. and Duboule, D. (2013) ‘Topology of mammalian developmental enhancers and their regulatory landscapes.’, *Nature*, 502(7472), pp. 499–506. doi: 10.1038/nature12753.
- Lai, K.-M. V., Gong, G., Atanasio, A., Rojas, J., Quispe, J., Posca, J., White, D., Huang, M., Fedorova, D., Grant, C., Miloscio, L., Droguett, G., Poueymirou, W. T., Auerbach, W., Yancopoulos, G. D., Friendewey, D., Rinn, J. and Valenzuela, D. M. (2015) ‘Diverse Phenotypes and Specific Transcription Patterns in Twenty Mouse Lines with Ablated LincRNAs’, *Plos One*, 10(4), p. e0125522. doi: 10.1371/journal.pone.0125522.
- Langmead, B. and Salzberg, S. L. (2012) ‘Fast gapped-read alignment with Bowtie 2’, *Nat Methods*, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.
- Lee, E., Yu, D., Velasco, J. M. De, Tessarollo, L., Swing, D. A., Court, D. L., Jenkins, N. A., Copeland, N. G. and The, J. C. (2001) ‘A Highly Efficient Escherichia coli -Based Chromosome Engineering System Adapted for Recombinogenic Targeting and Subcloning of BAC DNA’, *Genomics*, 65, pp. 56–65. doi: 10.1006/geno.2000.6451.
- Lee, J. T. (2012) ‘Epigenetic Regulation by Long Noncoding RNAs’, *Science*, 338(6113), pp. 1435–1439. doi: 10.1126/science.1231776.
- Lewis, E. B. (1978) ‘A gene complex controlling segmentation in Drosophila’, *Nature*, 276, pp. 565–570.
- Li, L., Liu, B., Wapinski, O. L., Tsai, M. C., Qu, K., Zhang, J., Carlson, J. C., Lin, M., Fang, F., Gupta, R. a., Helms, J. a. and Chang, H. Y. (2013) ‘Targeted Disruption of Hotair Leads to Homeotic Transformation and Gene Derepression’, *Cell Reports*. The Authors, 5(1), pp. 3–12. doi: 10.1016/j.celrep.2013.09.003.
- Lin, C., Yin, Y., Bell, S. M., Veith, G. M., Chen, H., Huh, S. H., Ornitz, D. M. and Ma, L. (2013) ‘Delineating a Conserved Genetic Cassette Promoting Outgrowth of Body Appendages’, *PLoS Genetics*, 9(1), pp. 1–12. doi: 10.1371/journal.pgen.1003231.
- Lin, C., Yin, Y., Long, F. and Ma, L. (2008) ‘Tissue-specific requirements of beta-catenin in external genitalia development’, *Development*, 135(16), pp. 2815–2825. doi: 10.1242/dev.020586.
- Lin, C., Yin, Y., Veith, G. M., Fisher, A. V, Long, F. and Ma, L. (2009) ‘Temporal and spatial dissection of Shh signaling in genital tubercle development’, *Development*, 136(23), pp. 3959–3967. doi: 10.1242/dev.039768.
- Lonfat, N. and Duboule, D. (2015) ‘Structure, Function and Evolution of Topologically Associating Domains (TADs) at Hox loci’, *FEBS Letters*. Federation of European Biochemical Societies, 589(20), pp. 2869–2876. doi: 10.1016/j.febslet.2015.04.024.
- 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(6212), pp. 1004–1006. doi: 10.1126/science.1257493.
- Long, H. K., Prescott, S. L. and Wsocka, J. (2016) ‘Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution’, *Cell*, 167(5), pp. 1170–1187. doi: 10.1016/j.cell.2016.09.018.
- Love, M. I., Huber, W. and Anders, S. (2014) ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’, *Genome Biology*, 15(12), pp. 1–34. doi: Artn

550\DOI 10.1186/S13059-014-0550-8.

Lufkin, T., Mark, M., Hart, C. P., Dollé, P., LeMeur, M. and Chambon, P. (1992) 'Homeotic transformation of the occipital bones of the skull by ectopic expression of a homeobox gene', *Nature*, 359, p. 835. Available at: <https://doi.org/10.1038/359835a0>.

Maconochie, M., Nonchev, S., Morrison, A. and Krumlauf, R. (1996) 'PARALOGOUS HOX GENES : Function and Regulation', *Annual review of genetics*, 30, pp. 529–56.

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(4), pp. 1–7. doi: 10.1371/journal.pone.0000356.

Mallo, M. and Alonso, C. R. (2013) 'The regulation of Hox gene expression during animal development.', *Development*, 140(19), pp. 3951–63. doi: 10.1242/dev.068346.

Mallo, M., Wellik, D. M. and Deschamps, J. (2010) 'Hox genes and regional patterning of the vertebrate body plan', *Developmental Biology*, 141(4), pp. 520–529. doi: 10.1016/j.surg.2006.10.010.Use.

Martin, M. (2011) 'Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads', *EMBnet*. doi: <https://doi.org/10.14806/ej.17.1.200>.

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.', *Scientific reports*, 3, p. 3355. doi: 10.1038/srep03355.

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W. and Issac, B. (2010) 'GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells', *PLoS Genetics*, 6(12), pp. 1–10. doi: 10.1371/journal.pgen.1001244.

Miyagawa, S., Moon, A., Haraguchi, R., Inoue, C., Harada, M., Nakahara, C., Suzuki, K., Matsumaru, D., Kaneko, T., Matsuo, I., Yang, L., Taketo, M. M., Iguchi, T., Evans, S. M. and Yamada, G. (2009) 'Dosage-dependent hedgehog signals integrated with Wnt/beta-catenin signaling regulate external genitalia formation as an appendicular program.', *Development*, 136(23), pp. 3969–78. doi: 10.1242/dev.039438.

Miyagawa, S., Satoh, Y., Haraguchi, R., Suzuki, K., Iguchi, T., Taketo, M. M., Nakagata, N., Matsumoto, T., Takeyama, K., Kato, S. and Yamada, G. (2009) 'Genetic interactions of the androgen and Wnt/beta-catenin pathways for the masculinization of external genitalia.', *Molecular endocrinology (Baltimore, Md.)*, 23(June 2009), pp. 871–880. doi: 10.1210/me.2008-0478.

Montavon, T., Duboule, D. and B, P. T. R. S. (2013) 'Chromatin organization and global regulation of Hox gene clusters Chromatin organization and global regulation of Hox gene clusters'.

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 & development*, 22(3), pp. 346–59. doi: 10.1101/gad.1631708.

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(5), pp. 1132–45. doi: 10.1016/j.cell.2011.10.023.

Morgan, E. A., Nguyen, S. B., Scott, V. and Stadler, H. S. (2003) 'Loss of Bmp7 and Fgf8 signaling in Hoxa13-mutant mice causes hypospadias', *Development*, pp. 3095–3109. doi: 10.1242/dev.00530.

- de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., Koseki, H. and Brockdorff, N. (2004) 'Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation', *Developmental Cell*, 7(5), pp. 663–676. doi: <https://doi.org/10.1016/j.devcel.2004.10.005>.
- Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E. O. and Reinberg, D. (2016) 'CTCF-mediated topological boundaries during development foster appropriate gene regulation', *Genes & Development*, 30(24), pp. 2657–2662. doi: 10.1101/gad.288324.116.
- 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(6225), pp. 1017–1021. doi: 10.1126/science.1262088.
- 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', *Developmental Biology*, 428(2), pp. 293–299. doi: 10.1016/j.ydbio.2017.02.009.
- 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', 1466, pp. 1449–1466.
- Nolte, C., Alexander, T. B. and Krumlauf, R. (2015) 'Mammalian Embryo : Hox', *eLS. John Wiley & Sons, Ltd: Chichester*. doi: 10.1002/9780470015902.a0000740.pub3.
- 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, p. e02557. doi: 10.7554/eLife.02557.
- Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W. and Duboule, D. (2011) 'The dynamic architecture of Hox gene clusters.', *Science (New York, N.Y.)*, 334(6053), pp. 222–5. doi: 10.1126/science.1207194.
- 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(5), p. 930–944.e22. doi: 10.1016/j.cell.2017.05.004.
- 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., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J. and Heard, E. (2012) 'Spatial partitioning of the regulatory landscape of the X-inactivation centre.', *Nature*, 485(7398), pp. 381–5. doi: 10.1038/nature11049.
- Ohno, S. (1970) 'Evolution by genes duplication'. doi: <https://doi.org/10.1007/978-3-642-86659-3>.
- Ong, C.-T. and Corces, V. G. (2014) 'CTCF: an architectural protein bridging genome topology and function.', *Nature reviews. Genetics*, 15(4), pp. 234–46. doi: 10.1038/nrg3663.
- Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-frick, I., Pickle, C. S., Kato, M., Garvin, T. H., Pham, Q. T., Harrington, A. N., Akiyama, J. A., Afzal, V., Lopez-rios, J., Dickel, D. E. and Visel, A. (2018) 'Enhancer redundancy provides phenotypicrobustness in mammalian development', *Nature*, 554(7691), pp. 239–243. doi: 10.1038/nature25461.
- Papenbrock, T., Visconti, R. P. and Awgulewitsch, A. (2000) 'Loss of fibula in mice overexpressing Hoxc11', *Mechanisms of Development*, 92(2), pp. 113–123. doi:

10.1016/S0925-4773(99)00344-5.

Perriton, C. L., Powles, N., Chiang, C., Maconochie, M. K. and Cohn, M. J. (2002) 'Sonic hedgehog signaling from the urethral epithelium controls external genital development.', *Developmental biology*, 247(1), pp. 26–46. doi: 10.1006/dbio.2002.0668.

Peterson, R. L., Papenbrock, T., Davda, M. M. and Awgulewitsch, A. (1994) 'The murine Hoxc cluster contains five neighboring AbdB-related Hox genes that show unique spatially coordinated expression in posterior embryonic subregions', *Mechanisms of Development*, 47(3), pp. 253–260. doi: 10.1016/0925-4773(94)90043-4.

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., Bland, M. J., Wagstaff, W., Dalton, S., McDevitt, T. C., Sen, R., Dekker, J., Taylor, J. and Corces, V. G. (2013) 'Architectural protein subclasses shape 3D organization of genomes during lineage commitment', *Cell*, 153(6), pp. 1281–1295. doi: 10.1016/j.cell.2013.04.053.

Portoso, M., Ragazzini, R., Moiani, A., Michaud, A., Wassef, M., Servant, N., Sargueil, B. and Margueron, R. (2017) 'PRC 2 is dispensable for HOTAIR-mediated transcriptional repression', *The EMBO journal*, 36(8), pp. 981–994. doi: 10.15252/embj.201695335.

Quinonez, S. C. and Innis, J. W. (2014) 'Human HOX gene disorders', *Molecular Genetics and Metabolism*, 111(1), pp. 4–15. doi: 10.1016/j.ymgme.2013.10.012.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Bruggmann, S. A., Flynn, R. A. and Wysocka, J. (2011) 'A unique chromatin signature uncovers early developmental enhancers in humans', *Nature*. 2010/12/15, 470(7333), pp. 279–283. doi: 10.1038/nature09692.

Rao, S. S. P., Huang, S., Glenn, B., Hilaire, S., Engreitz, J. M., Perez, E. M., Sanborn, A. L., Sarah, E., Bascom, G. D., Bochkov, I. D., Huang, X., Muhammad, S., Glenn, B., Hilaire, S., Casellas, R., Lander, E. S. and Aiden, E. L. (2017) 'Cohesin Loss Eliminates All Loop Domains', *Cell*, 171(2), p. 305–309.e24. doi: 10.1016/j.cell.2017.09.026.

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. and Aiden, E. L. (2014) 'A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping', *Cell*, 159(7), pp. 1665–1680. doi: 10.1016/j.cell.2014.11.021.

Riising, E. M., Comet, I., Leblanc, B., Wu, X., Johansen, J. V. and Helin, K. (2014) 'Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide', *Molecular Cell*, 55(3), pp. 347–360. doi: 10.1016/j.molcel.2014.06.005.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Bruggmann, S. a, Goodnough, L. H., Helms, J. a, Farnham, P. J., Segal, E. and Chang, H. Y. (2007) 'Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.', *Cell*, 129(7), pp. 1311–23. doi: 10.1016/j.cell.2007.05.022.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. and Pachter, L. (2011) 'Improving RNA-Seq expression estimates by correcting for fragment bias.', *Genome biology*, 12(3), p. R22. doi: 10.1186/gb-2011-12-3-r22.

Rodríguez-carballo, E., Lopez-delisle, L., Zhan, Y., Fabre, P. J., Beccari, L., El-idrissi, I., Hanh, T., Huynh, 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 & Development*, pp. 2264–2281. doi: 10.1101/gad.307769.117.

Ruthenburg, A. J., Allis, C. D. and Wysocka, J. (2007) 'Methylation of Lysine 4 on Histone

- H3: Intricacy of Writing and Reading a Single Epigenetic Mark', *Molecular Cell*, 25(1), pp. 15–30. doi: <https://doi.org/10.1016/j.molcel.2006.12.014>.
- Sabari, B. R., Dall'Agnesse, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., Roeder, R. G., Sharp, P. A., Chakraborty, A. K. and Young, R. A. (2018) 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, 361(6400). doi: 10.1126/science.aap9195.
- Sagai, T., Hosoya, M., Mizushima, Y., Tamura, M. and Shiroishi, T. (2005) 'Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb', *Development*, 132(4), p. 797 LP-803. doi: 10.1242/dev.01613.
- Schorderet, P. and Duboule, D. (2011) 'Structural and functional differences in the long non-coding RNA hotair in mouse and human.', *PLoS genetics*, 7(5), p. e1002071. doi: 10.1371/journal.pgen.1002071.
- Schuettengruber, B., Bourbon, H., Croce, L. Di and Cavalli, G. (2017) 'Review Genome Regulation by Polycomb and Trithorax : 70 Years and Counting', *Cell*, 171(1), pp. 34–57. doi: 10.1016/j.cell.2017.08.002.
- Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G. and Loe-mie, Y. (2017) 'Two independent modes of chromatin organization revealed by cohesin removal', *Nature*, 551(7678), pp. 51–56. doi: 10.1038/nature24281.
- Seifert, A. W., Bouldin, C. M., Choi, K. S., Harfe, B. D. and Cohn, M. J. (2009) 'Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development', *Development*, 136(23), pp. 3949–3957. doi: 10.1242/dev.042291.
- Seifert, A. W., Harfe, B. D. and Cohn, M. J. (2008) 'Cell lineage analysis demonstrates an endodermal origin of the distal urethra and perineum', *Developmental Biology*, 318(1), pp. 143–152. doi: 10.1016/j.ydbio.2008.03.017.
- Seifert, A. W., Yamaguchi, T. and Cohn, M. J. (2009) 'Functional and phylogenetic analysis shows that Fgf8 is a marker of genital induction in mammals but is not required for external genital development.', *Development*, 136(15), pp. 2643–2651. doi: 10.1242/dev.036830.
- Sexton, T. and Cavalli, G. (2015) 'Review The Role of Chromosome Domains in Shaping the Functional Genome', *Cell*, 160(6), pp. 1049–1059. doi: 10.1016/j.cell.2015.02.040.
- Shen, Y., Yue, F., Mcclary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenko, V. V and Ren, B. (2012) 'A map of the cis-regulatory sequences in the mouse genome', *Nature*. Nature Publishing Group, 488(7409), pp. 116–120. doi: 10.1038/nature11243.
- Shlyueva, D., Stampfel, G. and Stark, A. (2014) 'Transcriptional enhancers: from properties to genome-wide predictions', *Nature Reviews Genetics*, 15(4), pp. 272–286. doi: 10.1038/nrg3682.
- Song, L. and Crawford, G. E. (2010) 'DNase-seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells', *Cold Spring Harbor Protocols*, 2010(2), p. pdb.prot5384. doi: 10.1101/pdb.prot5384.
- Soshnikova, N. and Duboule, D. (2009) 'Epigenetic regulation of vertebrate', *Epigenetics*, pp. 537–540.
- Spielmann, M., Lupiáñez, D. G. and Mundlos, S. (2018) 'Structural variation in the 3D

genome', *Nature Reviews Genetics*. Springer US, 19(July). doi: 10.1038/s41576-018-0007-0.

Spitz, F. (2016) 'Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles', *Seminars in Cell and Developmental Biology*, 57, pp. 57–67. doi: 10.1016/j.semcdb.2016.06.017.

Spitz, F. and Furlong, E. E. M. (2012) 'Transcription factors : from enhancer binding to developmental control', *Nature*, 13(9), pp. 613–626. doi: 10.1038/nrg3207.

Spitz, F., Gonzalez, F. and Duboule, D. (2003) 'A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster', *Cell*, 113(3), pp. 405–417. doi: 10.1016/S0092-8674(03)00310-6.

Spitz, F., Gonzalez, F., Peichel, C., Vogt, T. F., Duboule, D. and Zákány, J. (2001) 'Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations', *Genes & Development*, 6, pp. 2209–2214. doi: 10.1101/gad.205701.where.

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.', *Nature genetics*, 37(8), pp. 889–93. doi: 10.1038/ng1597.

Stemmer, M., Thumberger, T., Keyer, S., Wittbrodt, J. and Mateo, J. L. (2015) 'CCTop : An Intuitive , Flexible and Reliable CRISPR / Cas9 Target Prediction Tool', 9, pp. 1–11. doi: 10.1371/journal.pone.0124633.

Suemori, H. and Noguchi, S. (2000) 'Hox C cluster genes are dispensable for overall body plan of mouse embryonic development.', *Developmental biology*, 220(2), pp. 333–42. doi: 10.1006/dbio.2000.9651.

Suzuki, K., Ogino, Y., Murakami, R., Satoh, Y. and Bachiller, D. (2002) 'Embryonic development of mouse external genitalia : insights into a unique mode of organogenesis', *Evolution & development*, 141, pp. 133–141.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E. and Chang, H. Y. (2010) 'Long noncoding RNA as modular scaffold of histone modification complexes.', *Science*, 329(5992), pp. 689–93. doi: 10.1126/science.1192002.

Tschopp, P. and Duboule, D. (2011) 'A regulatory "landscape effect" over the HoxD cluster.', *Developmental biology*, 351(2), pp. 288–96. doi: 10.1016/j.ydbio.2010.12.034.

Tschopp, P., Sherratt, E., Sanger, T. J., Groner, A. C., Aspiras, A. C., Hu, J. K., Pourquié, O., Gros, J. and Tabin, C. J. (2014) 'A relative shift in cloacal location repositions external genitalia in amniote evolution', *Nature*. doi: 10.1038/nature13819.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A. and Hadjur, S. (2015) 'Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture', *Cell reports*, 10(8), pp. 1297–1309. doi: 10.1016/j.celrep.2015.02.004.

Visel, A., Minovitsky, S., Dubchak, I. and Pennacchio, L. A. (2006) 'VISTA Enhancer Browser—a database of tissue-specific human enhancers', *Nucleic Acids Research*, 35(suppl_1), pp. D88–D92. doi: 10.1093/nar/gkl822.

Wagner, G. P., Amemiya, C. and Ruddle, F. (2003) 'evolutionary novelties', 100(25), pp. 14603–14606.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. and Zhang, Y. (2004) 'Role of histone H2A ubiquitination in Polycomb silencing', *Nature*. Macmillan Magazines Ltd., 431, p. 873. Available at: <https://doi.org/10.1038/nature02985>.

- Warot, X., Fromental-ramain, C., Fraulob, V., Chambon, P. and Dollé, P. (1997) 'Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts', 4791, pp. 4781–4791.
- Wellik, D. M. (2007) 'Hox patterning of the vertebrate axial skeleton', *Developmental Dynamics*, 236(9), pp. 2454–2463. doi: 10.1002/dvdy.21286.
- Wellik, D. M. and Capecchi, M. R. (2003) 'Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton.', *Science*, 301(5631), pp. 363–367. doi: 10.1126/science.1085672.
- 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', *Molecular Cell*, 60(4), pp. 676–684. doi: 10.1016/j.molcel.2015.09.023.
- 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 biology*, 12(1), p. e1001773. doi: 10.1371/journal.pbio.1001773.
- Wu, X., Ferrara, C., Shapiro, E. and Grishina, I. (2009) 'Bmp7 expression and null phenotype in the urogenital system suggest a role in re-organization of the urethral epithelium', *Gene expression patterns : GEP*. 2008/12/29, 9(4), pp. 224–230. doi: 10.1016/j.gep.2008.12.005.
- Yakushiji-kaminatsui, N., Lopez-delisle, L., Bolt, C. C., Andrey, G., Beccari, L. and Duboule, D. (2018) 'Similarities and differences in the regulation of HoxD genes during chick and mouse limb development', *PLoS biology*, pp. 1–33. Available at: <https://doi.org/10.1371/journal.pbio.3000004>.
- Yamada, G., Satoh, Y., Baskin, L. S. and Cunha, G. R. (2003) 'Cellular and molecular mechanisms of development of the external genitalia.', *Differentiation; research in biological diversity*. International Society of Differentiation, 71(8), pp. 445–60. doi: 10.1046/j.1432-0436.2003.7108001.x.
- Yang, J. H., Menshenina, J., Cunha, G. R., Place, N. and Baskin, L. S. (2010) 'Morphology of Mouse External Genitalia: Implications for a Role of Estrogen in Sexual Dimorphism of the Mouse Genital Tubercle', 184, pp. 1604–1609. doi: 10.1016/j.juro.2010.03.079. Morphology.
- Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S. J., Cunningham, F., Aken, B. L., Zerbino, D. R. and Flicek, P. (2016) 'Ensembl 2016', *Nucleic acids research*, 44(D1), pp. D710–D716. doi: 10.1093/nar/gkv1157.
- Yokouchi, Y., Sakiyama, J.-I. and Kuroiwa, A. (1995) 'Coordinated Expression of Abd-B Subfamily Genes of the HoxA Cluster in the Developing Digestive Tract of Chick Embryo', *Developmental Biology*, 169(1), pp. 76–89. doi: <https://doi.org/10.1006/dbio.1995.1128>.
- 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., Beck, F., Mallo, M. and Deschamps, J. (2009) 'Cdx and Hox Genes Differentially Regulate Posterior Axial Growth in Mammalian Embryos', *Developmental Cell*, 17(4), pp. 516–526. doi: <https://doi.org/10.1016/j.devcel.2009.08.010>.
- Zakany, J. and Duboule, D. (2007) 'The role of Hox genes during vertebrate limb development',

Current Opinion in Genetics and Development, 17(4), pp. 359–366. doi: 10.1016/j.gde.2007.05.011.

Zákány, J., Fromental-Ramain, C., Warot, X. and Duboule, D. (1997) 'Regulation of number and size of digits by posterior Hox genes: a dose-dependent mechanism with potential evolutionary implications.', *Proceedings of the National Academy of Sciences of the United States of America*, 94(25), pp. 13695–13700. doi: 10.1073/pnas.94.25.13695.

Zákány, J., Favier, B., Potter, S. S. and Duboule, D. (1996) 'Functional Equivalence and Rescue Among Group 11 Hox Gene Products in Vertebral Patterning', *Developmental Biology*, 328(0137), pp. 325–328.

Zeltser, L., Desplan, C. and Heintz, N. (1996) 'Hoxb-13 : a new Hox gene in a distant region of the HOXB cluster maintains colinearity', *Development*, 2484, pp. 2475–2484.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. and Liu, X. S. (2008) 'Open Access Model-based Analysis of ChIP-Seq (MACS)', *Genome biology*, (9). doi: 10.1186/gb-2008-9-9-r137.

Ziebarth, J. D., Bhattacharya, A. and Cui, Y. (2013) 'CTCFBSDB 2 . 0 : a database for CTCF-binding sites and genome organization', *Nucleic Acids Research*, 41(November 2012), pp. 188–194. doi: 10.1093/nar/gks1165.

7. Acknowledgements

First, I would like to thank Denis Duboule, a truly great and humane mentor, for all the knowledgeable advice during my PhD.

I must thank present and past members of the Duboule Lab, Anouk, Lucille, Chase, Nayuta, Fabrice, Hocine, Célia, Anne-Catherine, Gordana, Leo, Isabel, Eddie, Aurélie, and Joska for all the interesting scientific discussions and help during this time, it is a pleasure to work in such great environment. Specifically, I would like to thank Lucille and Chase for reviewing my thesis.

Thank you to the various platforms for their help during this work. Particularly, everyone at the iGE3 and GECF genomics platforms for their sequencing expertise, and to the EPFL Transgenic Core Facility and UNIGE Transgenic core facility for their help with transgenesis.

I would like to thank Bénédicte, Hanh, Sandra, and Julien in Geneva for all their help in generating some of the mouse lines used in this work as well as taking care of all the animal stocks.

I must thank Angela Taddei, Alexandre Reymond, Bart Deplancke, and Melanie Blokesch for accepting to be in my thesis committee.

I would especially like to thank my family for all the encouragement and love during these years. My husband Loïc, who's unconditional support made it all possible, and our children Mathilde and Arthur who brought a lot of happiness and love to our lives.

8. Annex 1

***Hotair* is dispensible for mouse development**

Ana Rita Amândio, Anamaria Necsulea, Elisabeth Joye, Bénédicte Mascrez,
Denis Duboule

Published as: A. R. Amândio, A. Necsulea, E. Joye, B. Mascrez, and D. Duboule, “Hotair Is Dispensible for Mouse Development,” pp. 1–27, 2016, DOI: 10.1371/journal.pgen.1006232

RESEARCH ARTICLE

Hotair Is Dispensible for Mouse Development

Ana Rita Amândio¹, Anamaria Necseulea¹, Elisabeth Joye¹, Bénédicte Mascrez², Denis Duboule^{1,2*}

¹ School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ² Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland

* denis.duboule@epfl.ch, denis.duboule@unige.ch



CrossMark
click for updates

 OPEN ACCESS

Citation: Amândio AR, Necseulea A, Joye E, Mascrez B, Duboule D (2016) *Hotair* Is Dispensable for Mouse Development. *PLoS Genet* 12(12): e1006232. doi:10.1371/journal.pgen.1006232

Editor: Gregory S. Barsh, Stanford University School of Medicine, UNITED STATES

Received: April 18, 2016

Accepted: July 8, 2016

Published: December 15, 2016

Copyright: © 2016 Amândio et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw and processed RNA-seq data are available in GEO, accession number GSE79028.

Funding: This work was supported by institutional funding from the EPFL (<http://www.epfl.ch>) and the University of Geneva (<http://www.unige.ch>) as well as by grants from the Swiss National Research Fund (www.snf.ch) (No. 310030B_138662) and the European Research Council (<https://erc.europa.eu>) (ERC SystemsHox, No: 232790 and ERC RegulHox, No: 667156) (to DD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Despite the crucial importance of *Hox* genes functions during animal development, the mechanisms that control their transcription in time and space are not yet fully understood. In this context, it was proposed that *Hotair*, a lncRNA transcribed from within the *HoxC* cluster regulates *Hoxd* gene expression in *trans*, through the targeting of Polycomb and consecutive transcriptional repression. This activity was recently supported by the skeletal phenotype of mice lacking *Hotair* function. However, other loss of function alleles at this locus did not elicit the same effects. Here, we re-analyze the molecular and phenotypic consequences of deleting the *Hotair* locus *in vivo*. In contrast with previous findings, we show that deleting *Hotair* has no detectable effect on *Hoxd* genes expression *in vivo*. In addition, we were unable to observe any significant morphological alteration in mice lacking the *Hotair* transcript. However, we find a subtle impact of deleting the *Hotair* locus upon the expression of the neighboring *Hoxc11* and *Hoxc12* genes *in cis*. Our results do not support any substantial role for *Hotair* during mammalian development *in vivo*. Instead, they argue in favor of a DNA-dependent effect of the *Hotair* deletion upon the transcriptional landscape *in cis*.

Author Summary

During mammalian embryonic development, *Hox* genes must be tightly regulated. It was proposed earlier that part of this regulation relies upon *Hotair*, a long non-coding RNA that recruits repressive protein complexes onto the *HoxD* gene cluster to keep these genes silent before they become activated. A genetic deletion of *Hotair* in mice induced homeotic transformations, thus supporting this hypothesis. However, other alleles involving this locus gave controversial results and hence we re-assessed the effect of the full deletion of *Hotair in vivo*. In our genetic background and using our analytical conditions, we could not confirm the reported morphological alterations, nor could we detect any mis-regulation of *Hoxd* genes in those fetal tissues where *Hotair* is detected in control animals. However, the genomic deletion induces the mis-regulation *in-cis* of the neighboring *Hoxc11* and *Hoxc12* genes, a side-effect which may underlie a weakly penetrant alteration observed in the shape of some tail vertebrae.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Hox genes encode transcription factors with crucial roles in the specification of regional identities along the body axes during development. Mutations affecting specific *Hox* genes typically lead to homeotic transformations, whereby a particular body part is transformed into the identity of another one [1–4]. In mammals, following the two rounds of genome duplication that occurred at the basis of the vertebrate lineage (see [5]), four distinct clusters of *Hox* genes are found (*HoxA* to *HoxD*) (ref. in [6]). During development, *Hox* genes are transcriptionally activated in a precise temporal and spatial sequence, which follows their chromosomal order [7,8]. These collinear patterns of transcription are regulated at multiple levels and studies focusing on the *HoxA*, *HoxB* and *HoxD* loci have revealed the importance of intricate combinations of local and long-range *cis*-regulatory elements. Also, studies using micro-dissected embryonic material have shown that the transcriptional activation of these genes, in different ontogenetic contexts, is accompanied by major changes in both the epigenetic modifications of the surrounding chromatin and its 3D spatial organization [9–11].

Long non-coding RNAs (lncRNAs) have been proposed to represent yet another layer of regulatory control at these important developmental loci (e.g. [12–15]). Increasing evidence indeed suggests that lncRNAs can act as regulators of gene expression, for example by interacting with transcription factors and chromatin modifiers to modulate transcription during development [16]. Several lncRNAs associated with the mammalian *Hox* clusters have been identified, amongst which *Hotair* (*Hox* transcript antisense intergenic RNA), a lncRNA transcribed from the intergenic region between *Hoxc11* and *Hoxc12* within the *HoxC* cluster and the founding member of this new class of RNAs. *Hotair* was proposed to help repress some 5'-located (posterior) *Hoxd* genes *in trans*, through its association with chromatin modification complexes such as PRC2, LSD1 and CoREST/REST [13,17]. Accordingly, *Hotair* would recruit or enrich this part of the *HoxD* cluster with *Polycomb* (*Pc*) complex, thus contributing to its repressed state before transcription starts. This proposal was substantiated by the knockdown of *Hotair* in human fibroblasts, which led to a decreased binding of *Pc* repressive complexes in the *HoxD* cluster and to a concurrent increase in *Hoxd* genes expression [13].

This important function for a lncRNA in cultured human fibroblasts was however not supported by the analysis of a mouse line carrying a targeted deletion of the entire *HoxC* cluster [18], i.e. including the mouse *Hotair* lncRNA. This deletion showed little effect *in vivo*, with no alteration of *Hoxd* genes expression. Also, the presence and enrichment of H3K27me3 repressive chromatin marks at the *HoxD* locus was not dramatically modified [19]. This lack of effect was tentatively explained by the concomitant *in-cis* deletion of all *Hoxc* genes, which may have masked or compensated a potential alteration caused by the absence of *Hotair* alone [20]. To alleviate this problem, three alleles were recently produced where the *Hotair* transcript was specifically targeted (Fig 1). The first allele is a targeted deletion of the two major exons of *Hotair*. Mice carrying this deletion were reported to display a malformation of the wrist and homeotic transformations of the spine, either from six lumbar vertebrae (L6) to a L5 vertebral formula, or within the post-sacral region [20]. These phenotypes were associated with a derepression of *Hoxd* genes and of a set of imprinted genes by modulation of their chromatin state [20]. These effects were scored in the absence of any change in the transcription of the neighboring *Hoxc11* and *Hoxc12* genes, supporting a function of *Hotair* *in trans* [20].

Two additional *Hotair* deletion mutant alleles combined with *LacZ* reporter knock-in were recently reported by Lai and colleagues [21]. The first allele deleted nearly the entire *Hotair* sequence and the second one comprised a smaller deletion starting in the second exon [21]. In both cases, while a subtle alteration of the 4th caudal vertebra was scored, the wrist and the spine appeared normally formed, without any sign of the lumbar homeotic transformation

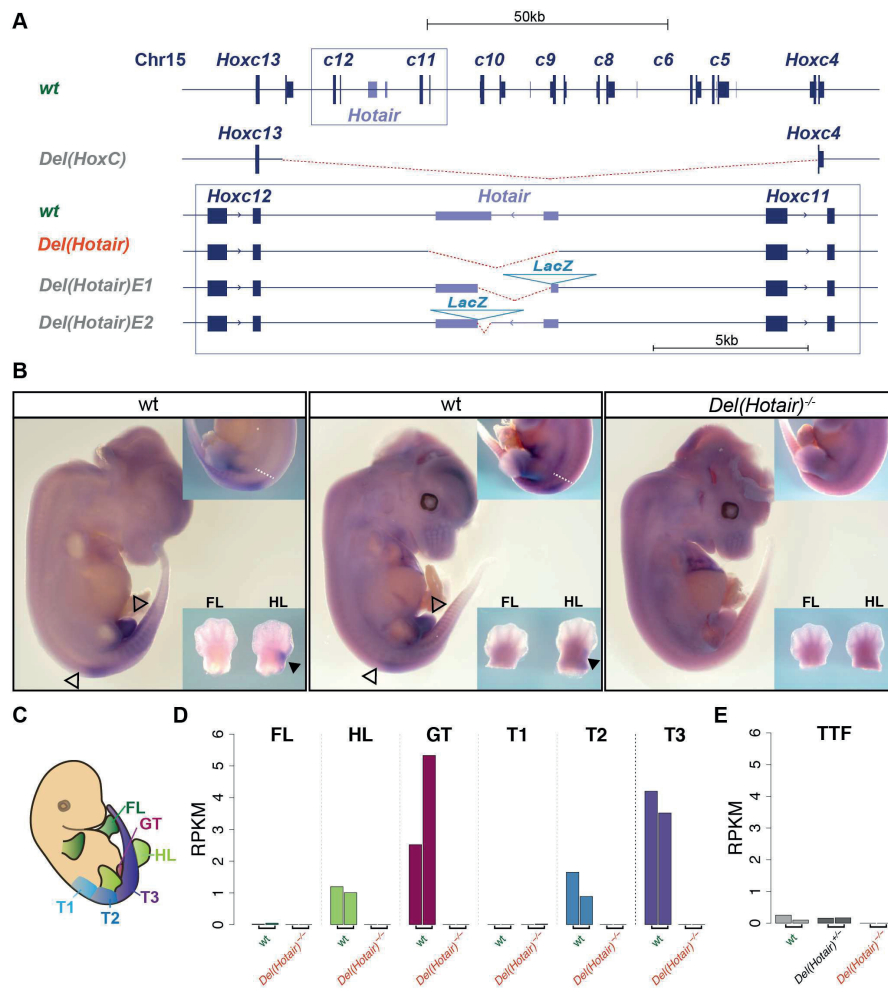


Fig 1. *Hotair* expression in vivo. (A) Schematic representation of the wild type *Hotair* locus and the various *Hotair* deletion alleles. The deleted DNA is in red. The *HoxC* allele is from [18] and the shorter deletions in the box from [20] and [21]. (B) Whole mount in situ hybridization (WISH) of *Hotair* RNAs on E12.5 wild type CD1 (left) and CBA/C57/B6 (center) control embryos and of a *Del(Hotair)*^{-/-} mouse embryo (right, n = 3). No signal was detected in *Del(Hotair)*^{-/-} embryos, demonstrating the specificity of the probe. *Hotair* is expressed with a posterior restriction (white dashed line), resembling the transcript distribution of either a *Hox11* or a *Hox12* gene. Black arrowheads indicate expression domain of *Hotair* in the hindlimbs, hollow arrowheads indicate the limit of *Hotair* expression in the trunk and in the genital tubercle. The common artifact signal in the cerebral vesicles results from incomplete opening of these vesicles and subsequent probe trapping. (C) Schematic representation of the dissection patterns for RNA-seq. These dissections involved forelimbs (FL, dark green), hindlimbs (HL, green) and the genital tubercle (GT, magenta), as well as three trunk sections corresponding to the lumbar/sacral (T1, light blue); sacro/caudal (T2, blue) and caudal (T3, purple)

regions. (D) Quantification of *Hotair* expression by RNA-seq (normalized RPKM values). (E) Quantification of *Hotair* expression (normalized RPKM values) in tail tip fibroblasts (TTF), using data from [20].

doi:10.1371/journal.pgen.1006232.g001

and wrist alterations previously reported for the deletion of both exons [20]. Due to our long-lasting interest in the transcriptional regulation of *Hoxd* genes during development (e.g. [22]), we addressed these apparently conflicting results by re-assessing the effects of deleting the *Hotair* lncRNA during early mouse development, using the largest deletion allele previously described [20]. In agreement with earlier and more recent results [19,21], we find that the deletion of *Hotair* has no substantial effect, neither on wrist morphology, nor on the vertebral formula at the lumbo-sacral level. In addition, transcriptome analyses reveal that the absence of *Hotair* does not impact upon *Hoxd* genes expression *in trans*, in any of the embryonic tissues analyzed. In contrast, we observe subtle yet significant changes in the expression of the neighboring *Hoxc11* and *Hoxc12* genes in the mutant mice, indicating an *in-cis* impact of modifying the genomic locus. Taken together, our results strongly suggest that the *Hotair* lncRNA has little effect—if any—on mouse embryonic development.

Results

Hotair expression *in vivo*

We extended the analysis of a mouse strain that includes a deletion of the two major *Hotair* exons (Fig 1A) [20]. Even though we concluded that this mutation is primarily an allele of the *HoxC* cluster (see below) and hence that it should be referred to as *HoxC^{Del(Hotair)}*, we shall refer to it as *Del(Hotair)* throughout this study for sake of simplicity. We first confirmed the expression pattern of *Hotair* *in vivo* by whole mount *in situ* hybridization (WISH) using wild type mice of two distinct genetic backgrounds (CD1 and CBA/C57/B6) as well as *Del(Hotair)^{-/-}* mouse embryos at embryonic day 12.5 (E12.5) (Fig 1B). Staining of *Del(Hotair)^{-/-}* embryos confirmed the specificity of the *Hotair* probe as no signal was detected in these embryos (Fig 1B). In contrast, wild type embryos of both genetic backgrounds showed the presence of *Hotair* transcripts in the genital tubercle, the proximal part of the hindlimbs and in the caudal part of the embryo (Fig 1B), confirming previously published data [19] and consistent with the *LacZ* staining reported for the *Hotair* knocked-in allele [21]. In both cases, staining was observed just posterior to the lumbar region and was not scored in developing forelimbs [21].

We quantified the expression levels of *Hotair* with high coverage RNA-sequencing (RNA-seq) (Materials and Methods, S1 Table). Based on the spatial expression pattern of *Hotair* as determined with WISH and on the skeletal phenotypes reported in mice by Li et al. [20], we micro-dissected both *Del(Hotair)^{-/-}* and wild type E12.5 embryos into six distinct parts for comparative RNA-seq analyses (Fig 1C). We thus separately collected the forelimbs (FL), the hindlimbs (HL), the genital tubercle (GT), a piece of trunk corresponding to the lumbo-sacral region (T1); a piece of trunk corresponding to the sacro-caudal region (T2) and finally, a piece of trunk corresponding to the developing caudal region (T3, Fig 1C). As expected from the WISH experiments, *Hotair* transcripts were scored in the hindlimbs, the genital tubercle and the trunk samples T2 and T3. The highest steady-state levels of *Hotair* RNAs were detected in the GT and the T3 embryonic tissues (Fig 1D, S1 Dataset). As a control, *Hotair* transcripts were not detected in any tissues derived from homozygous *Del(Hotair)^{-/-}* mutant embryos (Fig 1D). To better compare this dataset with published results, we analyzed in parallel the RNA-seq data obtained from primary tail tip fibroblast (TTF), derived from both wild type and *Del(Hotair)^{-/-}* mice [20]. This analysis revealed that the expression level of *Hotair* in control TTF

was very low when compared to its expression levels in the GT or the posterior T3 trunk sample (Fig 1E, S2 Dataset).

Phenotype of mice lacking the *Hotair* lncRNA

Hotair was reported to be important for both the proper establishment of the mouse vertebral column and for the formation of the forelimb mesopodial articulation: the wrist [20,21]. To confirm this phenotypic effect, we inter-crossed *Del(Hotair)^{+/-}* heterozygous mice and examined the skeletons of F1 animals at postnatal day 22 (P22). We investigated in particular the three reported sites of observed alterations in mutant *Del(Hotair)^{-/-}* mice [20,21]. We first searched for potential differences in vertebral formulae, as it was reported that 58% of *Del(Hotair)^{-/-}* mice had five lumbar vertebrae, while 100% of wild type CBA/CS7/BL6 mice had six lumbar vertebrae [20].

All our mutant alleles at *Hox* loci (see e.g. [23]) are systematically backcrossed onto mixed (B6xCBA)F1 animals to maintain heterogeneous but similar backgrounds when comparing experimental crosses. After bringing the *Hotair* mutant mice [20] onto this genetic background for some generations, we found that 80% of *Del(Hotair)^{-/-}* mice displayed five lumbar vertebrae, similar to wild type littermates (Chi-square test, *p*-value 0.97, Fig 2A, Table 1). In both wild type and homozygous mutant animals, the L6 formula was sporadically scored, as well as the mixed L6/S1 vertebral type, often observed in our stocks. Despite a limited number of specimens observed, but together with the fact that we were unable to detect specific transcripts in the embryonic trunk at this vertebral level by two independent methods, we conclude that this lncRNA is very unlikely to have a function in the organization of this very flexible morphological boundary (see discussion).

We next analyzed the morphology of caudal vertebrae in the post-sacral region. Previous analyses had concluded that mice with *Hotair* deletions had longer lateral processes on the fourth vertebra when compared to wild type animals, with full penetrance. In our case, we observed that three out of ten *Del(Hotair)^{-/-}* mutant mice had longer processes on the fifth caudal vertebra, compared to wild type (Fig 2B). This may indeed correspond to a very subtle

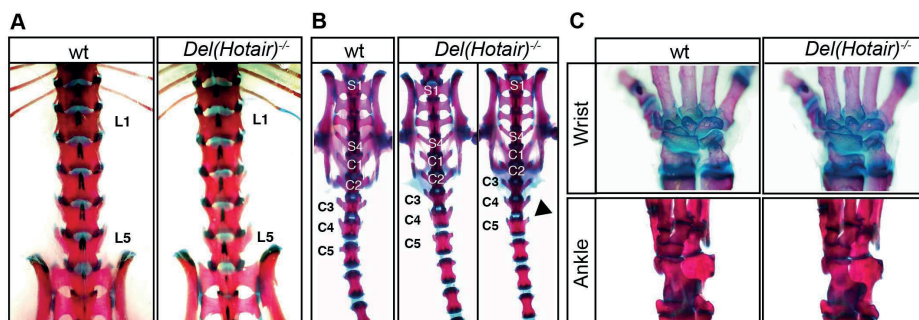


Fig 2. *Hotair* deletion has little impact, if any, on skeletal morphology. Alizarin Red and Alcian Blue skeletal staining of wild type and *Del(Hotair)^{-/-}* mice. (A) Lumbar region of wild type (left) and *Del(Hotair)^{-/-}* (right). In our (B6xCBA) background, both control and mutant animals have five lumbar vertebrae (L5), with an equally low incidence of L6 (see Table 1). (B) The sacro-caudal region of wild type (left) and *Del(Hotair)^{-/-}* (right) animals, with the black arrowhead pointing to a moderate gain of lateral protrusion in mutant caudal vertebra 5 (C5), usually not observed in control animals. (C) Normal wrist and ankle bones in both wild type (left) and *Del(Hotair)^{-/-}* (right) animals. The number and organization of mesopodial bones remained unchanged in the mutant condition.

doi:10.1371/journal.pgen.1006232.g002

Table 1. Phenotypic analysis of wild type and *Del(Hotair)*^{-/-} skeletons at post-natal day 22 (P22).

	wt (n = 11)	<i>Del(Hotair)</i> ^{-/-} (n = 11)	<i>Del(Hotair)</i> ^{-/-} (n = 10)
Lumbar vertebrae			
L6	1	1	1
L6/S1*	2	1	1
L5	8	9	8
Wrist			
normal	11	11	10
ill-formed	0	0	0
Ankle			
normal	11	11	10
ill-formed	0	0	0

* Mixed identity

Various control and mutant specimen were scored for having either five lumbar vertebrae (L5), six (L6) or a mixed L5-6/S1 vertebra. The mesopodial bones (wrist and ankle) were found normal in shape and number in all cases.

doi:10.1371/journal.pgen.1006232.t001

morphological alteration in this region of the caudal spine, although the penetrance of this light phenotype is not 100%. Unlike the lumbo-sacral and wrist alterations [20], this particular tail vertebral morphology was also scored by Lai and colleagues when analyzing another mutant allele of *Hotair* [21].

Finally, and even though we were unable to detect any *Hotair* transcripts in the forelimbs of E12.5 mice embryos, unlike for hindlimbs (Fig 1B), we carefully examined both forelimb and hindlimb skeletons of wild type and *Del(Hotair)*^{-/-} mutant mice. We did not detect any alteration in limb morphology (Fig 2C and Table 1), in particular in the anatomy of the wrist, where malformations due to the loss of *Hotair* had been previously reported (Fig 2C). The same conclusion was reached concerning the hindlimbs, even though *Hotair* transcripts could clearly be scored in the proximal part. Altogether, we could not reproduce the reported phenotypic effects of *Hotair* deletion at two sites, the wrist and the lumbo-sacral region, where we were also unable to detect any *Hotair* transcripts. Regarding tail vertebrae, a slight effect could indeed be observed, poorly penetrant and likely dependent on the genetic background (see below).

Transcription profiles of *Del(Hotair)*^{-/-} mutant embryonic tissues

To more globally evaluate the effect of *Hotair* deletion upon developmental gene regulation, we performed a principal component analysis (PCA) using the expression levels of all autosomal protein-coding genes detected in our RNA-seq experiments (Materials and Methods). We observed a good separation of the data according to tissue type, although the T1 and T2 samples clustered together (Fig 3). Principal component 1 (PC1), which explained 61.6% of the total gene expression variance, separated the trunk segments (T1, T2 and T3) from the other embryonic tissues. Likewise, the differences between GT, HL and FL were resolved along PC2, which accounted for 12.5% of the total variance (Fig 3). Part of the variance was also explained by the genotypes and we observed that wild type and *Del(Hotair)*^{-/-} samples were separated along PC2 (Fig 3). Of note, the same separation between wild type and *Del(Hotair)*^{-/-} on PC2 was observed in all tissues, even in T1 and FL where *Hotair* is not expressed (Fig 3). In agreement with the results from the clustering between samples, we observed high expression level

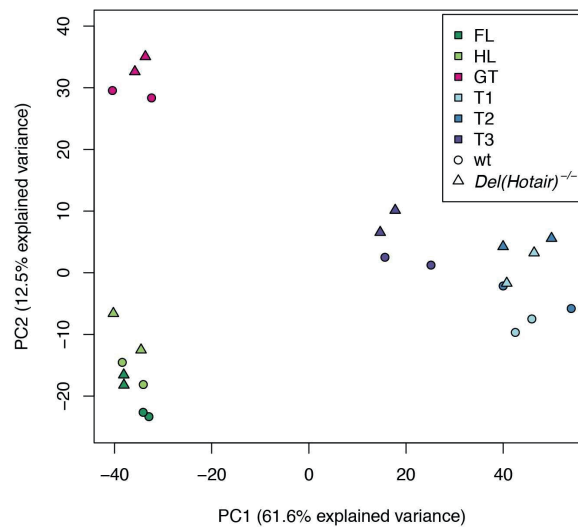


Fig 3. Overview of gene expression patterns in wild type and *Del(Hotair)^{-/-}* embryonic tissues. First factorial map for the principal component analysis (PCA) of gene expression levels. Tissues are color-coded (upper right corner) and the genotypes are indicated by either a circle (wild type) or a triangle (*Del(Hotair)^{-/-}*). The numbers in parentheses indicate the proportion of the variance explained by PC1 or by PC2.

doi:10.1371/journal.pgen.1006232.g003

correlations among biological replicates for all tissues (S1 Fig). Furthermore, gene expression clustering based on pairwise Euclidean distances between samples (see [Materials and Methods](#)) showed a clear separation between four different groups: the limbs (FL and HL), the GT, the T3 trunk segment and the remaining T1 and T2 trunk segments (S2 Fig). Using this method, we observed a separation between genotypes only when the T3 sample was considered.

Altogether, these results point to the reproducibility of replicates and illustrate the good separation between tissues, with the exception of the T1 and T2 trunk samples. The high similarity in global gene expression between T1 and T2 likely reflects the spatial proximity of these two tissues, even though we cannot exclude some variation in the positioning of the precise T1/T2 boundary during dissection, which is a challenging task in such young embryos.

Expression analysis of wild type versus *Del(Hotair)^{-/-}* embryonic tissues

Since *Hotair* was proposed to act as a repressor of gene expression in cultured fibroblasts [13,20], we conducted tissue-specific differential gene expression analyses between wild type and *Del(Hotair)^{-/-}* samples to assess such a potential function under physiological conditions. We only considered as significant an absolute expression fold change greater than 1.5 and we set the false discovery rate (FDR) threshold at 5%. By using these parameters, we observed between 64 and 588 protein-coding genes differentially expressed in the various tissues analyzed (S3 and S4 Datasets). We first compared all tissues that express *Hotair* in the wild type

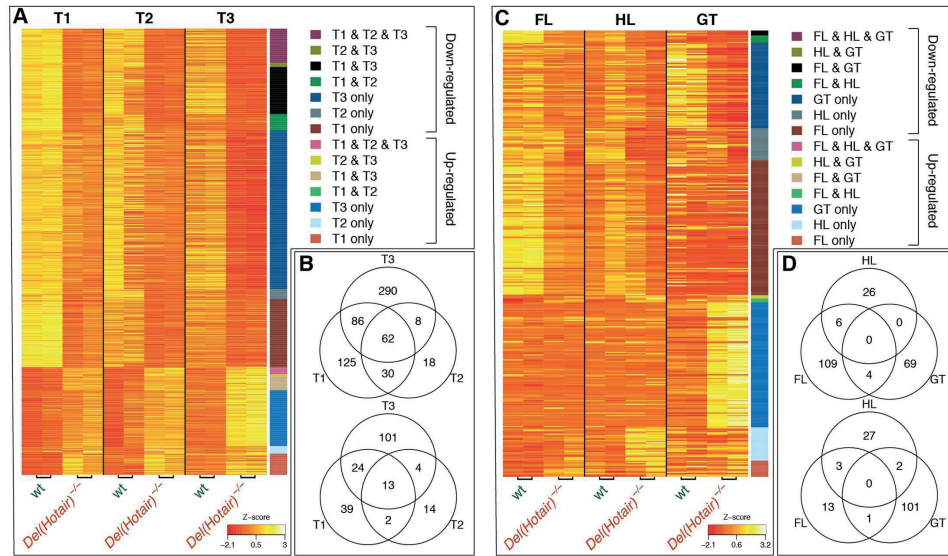


Fig 4. Differential expression between wild type and *Del(Hotair)*^{-/-} dissected samples. A-B) Differential gene expression analysis between wild type and *Del(Hotair)*^{-/-} trunk tissues (T1, T2, T3). The absolute fold change is > 1.5 and FDR < 0.05. The different columns correspond to sample type and rows correspond to differentially expressed genes. (A) Heat map of centered and scaled gene expression levels (Z-score log₂ RPKM). Genes are color coded vertically, according to tissue and expression changes between genotypes. (B) Venn diagram showing the number of down-regulated (top) and up-regulated (bottom) genes. C-D) Differential gene expression analysis between wild type and *Del(Hotair)*^{-/-} forelimbs (FL), hindlimbs (HL) and genital tubercle (GT). The absolute fold change is > 1.5 and FDR < 0.05. (C) Heat map of centered and scaled gene expression levels (Z-score log₂ RPKM). Genes are color coded vertically according to the tissue and orientation of expression between genotypes. (D) Venn diagram showing the number of down-regulated (top) and up-regulated (bottom) genes.

doi:10.1371/journal.pgen.1006232.g004

condition, i.e. the T2, T3, GT and HL samples, reasoning that potential differentially expressed *Hotair* targets should be identified in these contexts. However, we were not able to identify any common genes with altered expression between wild type and *Del(Hotair)*^{-/-} samples (S3 Fig), suggesting that the *Hotair* deletion does not affect the same set of genes in all tissues analyzed.

We thus divided the differential expression analysis based on the global expression clustering results. First, we compared the trunk samples T1 (lacking *Hotair* expression), T2 and T3. We identified 62 down-regulated genes and 13 up-regulated genes between wild type and *Del(Hotair)*^{-/-} samples, which are shared in all trunk sections (Fig 4A and 4B). Of note, we observed a common trend in gene expression differences in all trunk samples, even though only some of them passed the established thresholds (Fig 4A). Gene ontology (GO) analysis for either common down-regulated or common up-regulated genes showed no enrichment of functional terms and the majority of differentially expressed genes were down-regulated, which was unexpected given the previously proposed role of *Hotair* as a repressor [20] (Fig 4A and 4B). GO analysis of down-regulated genes in distinct *Del(Hotair)*^{-/-} trunk tissues revealed significant enrichment (FDR < 10%) in functional terms related to organ development and multicellular organismal process for most tissues (S5 Dataset). Up-regulated genes in T3 were

enriched for functional terms related to metabolic and biosynthetic processes and a weak enrichment for neuron differentiation genes was observed for T2 (S5 Dataset). Differential expression analysis for FL, HL and GT showed no common genes with altered expression (Fig 4C and 4D). GO analyses for differentially expressed genes in these tissues showed enrichment for functional terms related to development (S5 Dataset).

We next asked whether *Polycomb* target genes were preferentially up or down-regulated upon *Hotair* deletion. We defined putative target genes using H3K27me3 ChIP-seq data from wild type tail tip fibroblasts [20], selecting genes with a minimum 5-fold enrichment between H3K27me3 ChIP and input DNA in the gene promoter region (Materials and Methods) and thus obtained 861 putative target genes (S6 Dataset). We analyzed their pattern of differential expression in the T3 trunk sample, which includes the fetal tail and thus likely has the cell type composition *in vivo* most related to tail fibroblasts. Out of the 485 putative *Pc* targets that were expressed in the T3 segment (RPKM >1 in at least one wild type or *Del(Hotair)*^{-/-} sample), 60 genes were significantly differentially expressed (absolute fold change > 1.5 and FDR <10%), including 50 down-regulated and 10 up-regulated in the *Del(Hotair)*^{-/-} samples (S4 Fig). This indicates that only 17% of all differentially expressed *Pc* targets were up-regulated in *Del(Hotair)*^{-/-} samples, which is slightly lower than the proportion of up-regulated genes among non-targets (25% up-regulated genes out of 564 differentially expressed non-target genes, Chi-square test p-value 0.18). Thus, we could not detect any enrichment for up-regulation of putative *Pc* target genes when compared to all other expressed protein-coding genes. Therefore, under these physiological conditions, we could not find evidence supporting a role for *Hotair* in setting up, maintaining or re-enforcing the repression of this set of *Polycomb* target genes. While physiologically relevant, our analysis is however difficult to directly compare with the situation in tail fibroblasts, as *Polycomb* occupancy naturally depends on both the tissue-type and the developmental stage.

Expression of imprinted genes in *Del(Hotair)*^{-/-} embryonic tissues

Interestingly, a subset of imprinted genes including *H19* and *Meg3* was shown to be up-regulated upon deletion of *Hotair* in TTF [20]. We thus analyzed the expression status of known imprinted genes transcribed (RPKM >1) in at least one sample (S7 Dataset). To ensure maximum sensitivity, we lowered our FDR threshold to 10% while maintaining an absolute expression fold change greater than 1.5. With these parameters, we observed a total of 21 imprinted genes differentially expressed in our samples (S5 Fig, S7 Dataset). We found that 71% of differentially expressed imprinted genes were down-regulated, while only 29% were up-regulated. Notably, *H19* and *Meg3* were down-regulated in our samples, in contrast to what was observed in TTF.

In conclusion, these global transcriptome analyses comparing *Hotair* deletion mutant and wild type micro-dissected tissues revealed changes in gene expression upon deletion of the *Hotair* locus. Noteworthy, we observed numerous expression changes not only in tissues that normally express *Hotair* at detectable levels, i.e. the T2, T3, GT and HL samples, but also in tissues like the anterior trunk (T1) or the forelimb (FL), where *Hotair* lncRNAs were not detected. This suggests that such observed differences in gene expression cannot be explained by a mere direct effect of the *Hotair* RNA. Potential explanations to these unexpected observations are discussed below.

Hox genes expression in wild type and *Del(Hotair)*^{-/-} mutant mice

The original observation, which led *Hotair* to become the paradigm of lncRNAs acting in *trans*, was its capacity to regulate several genes members of the *HoxD* cluster by interacting

with components of the *Polycomb* repressive complex 2 (PRC2)[13]. In contrast, no effect was initially reported upon *Hoxc* genes expression levels [13,20], despite the fact that *Hotair* is encoded from within the *HoxC* locus in both humans and mice [13,19]. We re-assessed this issue by analyzing the expression of all *Hox* genes across our various tissue samples (Fig 5). The global expression patterns of all four *Hox* clusters expectedly corresponded to previously described expression patterns for such embryonic stage and body levels (e.g. [24,25]).

In order to detect even subtle effects of *Hotair* upon *Hox* gene regulation, we lowered our FDR threshold to 10% for differential expression analyses. Under these conditions, we detected significant down-regulation of some anterior *Hoxa* genes (*Hoxa3*, *Hoxa5* and *Hoxa6*) in the three trunk samples (Fig 5 and S6 Fig). Interestingly, these differences were present in all trunk samples, including in T1 where *Hotair* is not expressed. We also observed a slight up-regulation of *Hoxb9* in HL and GT (Fig 5 and S6 Fig). Notably, in some of the tissues analyzed, we detected significant expression changes for both *Hoxc11* and *Hoxc12*, i.e. the two genes in the *HoxC* cluster that flank the *Hotair* locus (see below).

However, in contrast with previous reports using tail fibroblasts [13,20], we did not detect any significant change in the steady-state levels of *Hoxd* genes RNAs, in any of the analyzed tissues (Figs 5 and 6A). To clarify this contradictory observation, we re-analyzed the previously published RNA-seq data from both wild type and *Del(Hotair)^{-/-}* TTF [20]. By implementing our analytical pipeline, we could not detect any significant difference in expression for any of the *Hoxd* genes (S7 Fig, S4 Dataset). Noteworthy, the expression levels of posterior *Hoxd* genes in this TTF dataset are either barely detectable or not detected at all, as for the *Hoxd12* gene, for example, suggesting that previous conclusions were raised based on particularly low transcript levels.

Visualization of *Hoxd* genes expression in *Del(Hotair)^{-/-}* embryos

The deletion of *Hotair* was claimed to alter both the expression levels and the spatial transcript distribution of the *Hoxd10* and *Hoxd11* genes in the trunk [20]. We performed whole mount *in situ* hybridization (WISH) on both wild type and *Del(Hotair)^{-/-}* littermates to appreciate potential variations in the expression domains of these two genes. By using our well established protocol, we found that *Hoxd10* and *Hoxd11* transcripts showed wild type distributions in *Del(Hotair)^{-/-}* mutant specimen (Fig 6B). To more precisely evaluate any potential difference in these expression domains between homozygous mutant and control littermates, we carried out double WISH for the *Hox* gene of interest in combination with a probe specific for the *MyoD* gene, which allowed for unambiguous somite visualization [26]. In both wild type and *Del(Hotair)^{-/-}* embryos, *Hoxd10* was expressed in the future spine up to the level of somite 26, whereas *Hoxd11* was scored from somite level 29 and caudally, as previously reported [27]. Neither *Hoxd10*, nor *Hoxd11* showed any detectable increase in the intensity of the signal or in their spatial expression pattern, confirming the RNA-seq results (Fig 6A and 6B). Taken together, these observations suggest that *Hotair* has no effect on the regulation of *Hoxd* genes, at least in the developmental context and at the stage where the function of *Hox* genes is critical for morphological development.

In-cis effect of *Hotair* deletion upon *Hoxc* genes transcription

Unlike for *Hoxd* genes, our differential expression analyses between *Del(Hotair)^{-/-}* and wild type samples revealed modest but significant changes for both *Hoxc11* and *Hoxc12*, the two genes neighboring the *Hotair* locus and thus flanking the deletion breakpoint (Fig 5). To further verify this new observation, we carefully analyzed the expression levels of both *Hoxc11* and *Hoxc12* in all tissue samples from where RNA-seq datasets had been obtained. As expected

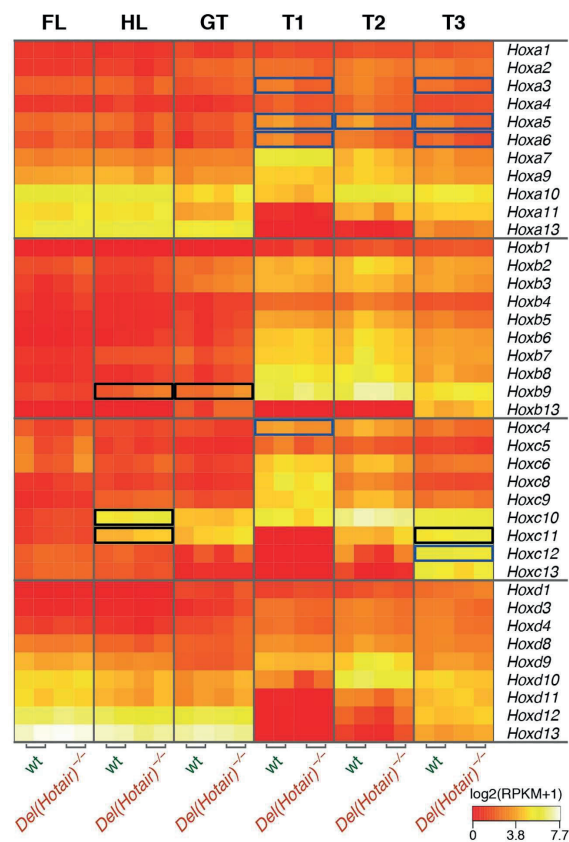


Fig 5. Expression of *Hox* genes in the various wild type and *Del(Hotair)*^{-/-} embryonic tissues. Heat map of log₂-transformed RPKM expression levels for all *Hox* genes. The columns correspond to sample type (indicated on top) and the rows correspond to *Hox* genes (indicated on the right). The blue boxes point to down-regulated genes, whereas the black boxes indicate up-regulated genes (FDR < 10%, no minimal fold change threshold).

doi:10.1371/journal.pgen.1006232.g005

from their collinear transcription [28], *Hoxc12* transcripts were mainly detected in the most posterior T3 trunk sample. In this sample, a significant reduction in the level of *Hoxc12* RNAs was scored in *Del(Hotair)*^{-/-} specimen (Fig 7A). On the other hand, *Hoxc11* transcripts were detected in the hindlimbs (HL), the genital tubercle (GT) and the T2 and T3 trunk samples. In these tissues, we observed an up-regulation of *Hoxc11* RNAs upon deletion of the *Hotair* locus, which was statistically significant for both HL and T3 (Fig 7B). In addition, a strong positive

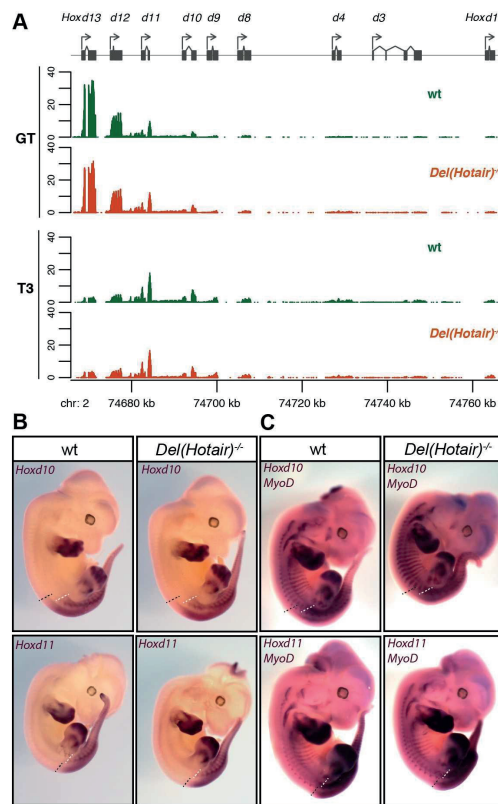


Fig 6. The deletion of *Hotair* does not alter *Hoxd* genes expression in embryo. (A) RNA-seq expression profiles of *Hoxd* genes in both the GT and T3 tissues of wild type (green) and *Del(Hotair)*^{-/-} (orange) E12.5 embryos. The Y-axis represents the per-base read coverage, normalized by dividing by the total number of million mapped reads in the corresponding samples. The two biological replicates were pooled for this representation and only uniquely mapping reads were used. (B) WISH of *Hoxd10* and *Hoxd11* on E12.5 wild type (left) and *Del(Hotair)*^{-/-} (right) embryos. The dashed lines indicate the rostral limits of the expression domains in the trunk, neural tube (black) and paraxial mesoderm (white). Adult vertebrae derive from the latter tissue. (C) Double WISH for the *MyoD* RNAs (for somite visualization) and either the *Hoxd10* (upper panel) or *Hoxd11* (lower panel) on E12.5 wild type (left) and *Del(Hotair)*^{-/-} (right) embryos. There was no detectable difference in the anterior limit of expression for any *Hoxd* gene analyzed.

doi:10.1371/journal.pgen.1006232.g006

correlation between the expression levels of *Hotair* and *Hoxc11* was detected in wild type samples (S8 Fig). The correlation was weaker between *Hotair* and *Hoxc12* expression (S8 Fig).

We asked whether these changes in expression of *Hoxc* genes in some *Del(Hotair)*^{-/-} samples were accompanied by alterations in their spatial expression patterns. We analyzed the expression of both *Hoxc12* and *Hoxc11* by WISH in *Del(Hotair)*^{-/-} E12.5 embryos and wild type

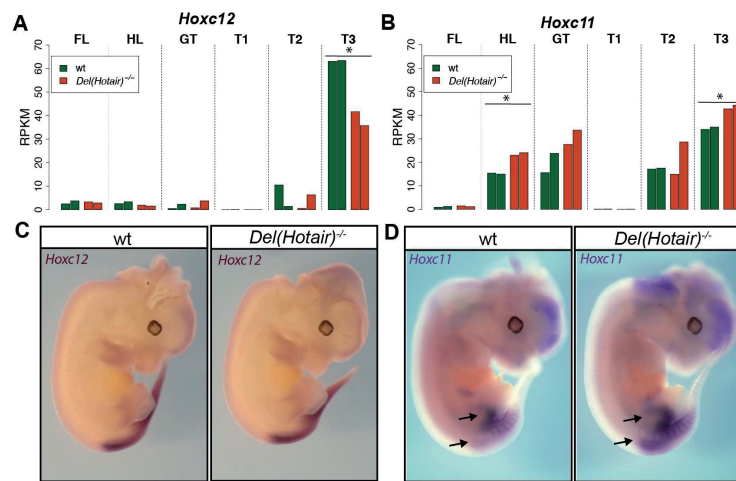


Fig 7. The deletion of *Hotair* affects the expression of the neighboring *Hoxc11* and *Hoxc12* genes. *Hoxc12* (A) and *Hoxc11* (B) expression (normalized RPKM values) in the various dissected tissue samples for wild type (green) and *Del(Hotair)^{-/-}* (orange) E12.5 embryos. The asterisk* indicates those samples where significant differences in transcript levels between genotypes were scored (FDR < 10%). (C) WISH using the *Hoxc12* probe in both wild type (left) and *Del(Hotair)^{-/-}* (right) E12.5 embryos. The spatial expression of *Hoxc12* remains globally unchanged. (D) WISH of *Hoxc11* in wild type (left) and *Del(Hotair)^{-/-}* (right) E12.5 embryos. The arrows indicate the slight anterior shift in the expression profile and the increase in signal intensity in the mutant genotype.

doi:10.1371/journal.pgen.1006232.g007

littermates. We did not observe any change for *Hoxc12* expression, neither in the transcript domain, nor in the intensity of RNA signal (Fig 7C). In contrast, *Del(Hotair)^{-/-}* embryos showed a clear rostral expansion of the *Hoxc11* transcript domain in the trunk, as well as an increased signal intensity in both the hindlimb buds and the trunk, in agreement with our RNA-seq data (Fig 7D).

The local impact of deleting the *Hotair* locus

To understand more precisely the reason why the deletion of the *Hotair* locus impacted the transcription of the flanking *Hoxc* genes, we analyzed in details the RNA-seq profiles of the region comprising *Hoxc12*, *Hotair* and *Hoxc11*. We first asked if all transcript isoforms derived from the *Hotair* locus were abrogated in the *Del(Hotair)^{-/-}* allele and observed that the deleted region almost perfectly coincides with the annotated boundaries of the locus in the mouse. However, the inspection of the RNA-seq profiles in tissues that transcribe *Hotair* RNA revealed the existence of larger transcripts, extending over at least 2.4 kb upstream of the annotated promoter (Fig 8A). Although we cannot determine the precise location of *Hotair* transcription start site(s), the presence of continuous transcription upstream of the annotated gene boundaries indicates that at least one, and probably two of the possible *Hotair* promoters were not deleted. Indeed in *Del(Hotair)^{-/-}* tissues, we detected transcripts initiating upstream of the annotated *Hotair* promoter, for instance within the *Hoxc11* intron, and spanning over the deleted region (Fig 8A).

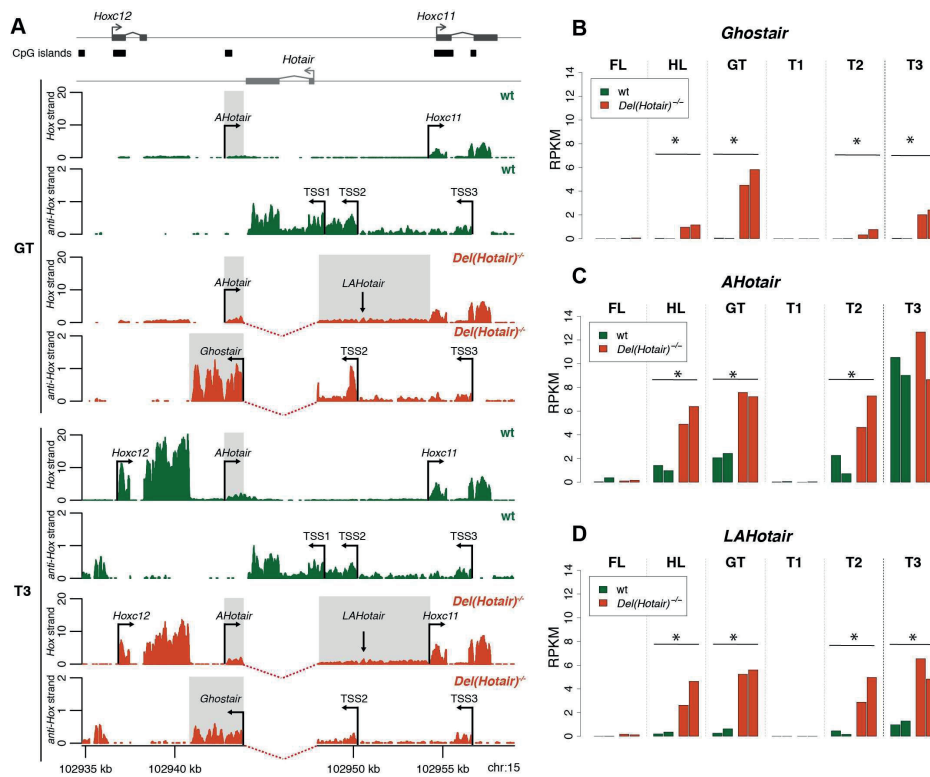


Fig 8. In-cis effects of the *Hotair* deletion on the local transcriptional activity. A) RNA-seq expression profiles of the genomic region neighboring *Hotair* in both the developing genitalia (GT, four profiles at top) and the most posterior trunk tissue sample (T3, four profiles at the bottom) from either wild type (green) and *Del(Hotair)^{-/-}* (orange) E12.5 embryos. In the wild type GT, only the *Hoxc11* gene is expressed along with *Hotair* on the opposite strand, which shows at least three putative start sites (arrows, TSS1 to TSS3). In the mutant GT, a long form of a new lncRNA (*AntiHotair*) now extends (grey box) on the *Hox* DNA strand, going over the deleted region up to the *Hoxc11* promoter. On the opposite DNA strand, the *Hotair* TSS2 and TSS3 are still functional and produce *Ghost of Hotair* (*Ghostair*), yet another new species of lncRNA, specific for the *Del(Hotair)^{-/-}* allele (in orange) and absent from the control allele (in green). A similar situation is observed in the T3 trunk sample, except that *Hoxc12* and *AHotair* are also expressed there. In the native locus, *anti-Hotair* is produced and meets with the end of the *Hotair* transcript. In the deleted allele, *Ghostair* is produced by the remaining *Hotair* TSS and terminates close to the 3' end of the *Hoxc12* transcript (bottom two profiles). The gray boxes indicate the genomic regions used for the expression quantifications of *AHotair*, *LAHotair* and *Ghostair*. The Y-axis represents the per-base RNA-seq read coverage, normalized by dividing by the total number of million mapped reads in the corresponding samples. The two biological replicates were pooled for this representation and only uniquely mapping reads were used. B-D) Expression values (normalized RPKM) for *AHotair* (B), *LAHotair* (C) and *Ghost of Hotair* (D) in all tissue samples. Genotypes are color-coded with wild type in green and *Del(Hotair)^{-/-}* in orange. The asterisk* indicates those samples where significant differences in expression were scored between the two genotypes (FDR < 10%).

doi:10.1371/journal.pgen.1006232.g008

In contrast to its multiple start site(s), *Hotair* displayed a very sharp transcription termination site. In wild type tissues, transcription of *Hotair* terminated at the annotated site, with virtually no RNA-seq reads mapped downstream of this position (Fig 8A). However, in the *Del(Hotair)^{-/-}* samples, we observed transcription downstream of the deleted locus terminating within 100bp of the *Hoxc12* termination site (Fig 8A). The presence of this extended transcript, which likely derives from one of the native *Hotair* promoters (as predicted with a *de novo* transcript assembly procedure, S9 Fig), likely resulted from the deletion of the wild type *Hotair* termination signals (Fig 8A). To quantify this gain of transcription, we counted RNA-seq reads mapping on the region between the annotated *Hotair* termination site and the *Hoxc12* termination site, on the *Hotair* strand. We referred to this transcript, which only appears upon deletion of *Hotair*, as *Ghost of Hotair* (*Ghostair*). We observed significant gains of *Ghostair* transcription in *Del(Hotair)^{-/-}* samples, in all tissues that expressed *Hotair* in the wild type condition, i.e. the hindlimb buds, the genital bud and the two trunk samples T2 and T3 (Fig 8B).

Ghost of Hotair and Anti-Hotair

Subsequent analyses of the RNA-seq profiles revealed an additional un-annotated promoter sequence, yet located on the *Hox* DNA strand. This promoter lies between *Hoxc12* and *Hoxc11* and overlaps with a CpG island (Fig 8A). In the wild type situation, it generates a relatively short, poorly abundant and un-spliced transcript, ca 1.8kb in size. The estimated termination site for this transcript was found within the region deleted in *Del(Hotair)*, close to the termination of *Hotair* itself on the other strand. Accordingly, we refer to this short transcript as *Anti-Hotair* (*AHotair*). In *Del(Hotair)^{-/-}* samples, this CpG island promoter was still active, giving rise to a much longer *AHotair* transcript (Fig 8A, long *AHotair* or *LAHotair*), consistent with the deletion of its termination site. We did not observe any clear boundaries between this extended transcript and *Hoxc11*, suggesting that this *AHotair* RNA could leak onto the *Hoxc11* transcription unit. This was confirmed by a *de novo* transcript assembly procedure (see [Materials and Methods, S9 Fig](#)).

To quantify this gain of transcription from the *Hox* strand, we further defined two transcribed regions; The first one largely corresponded to the short *Anti-Hotair* transcript detected in wild type samples, starting at the CpG island promoter and ending at the boundary of the deleted region (Fig 8A). The second one, long *AHotair*, corresponded to the longer transcript observed in *Del(Hotair)^{-/-}* samples, starting at the deleted region boundary and ending at the annotated *Hoxc11* transcription start site (Fig 8A). In agreement with our observations based on the RNA-seq profiles, we detected significant increases in expression for both *AHotair* and *LAHotair* in *Del(Hotair)^{-/-}* mutant tissues (Fig 8C and 8D). Therefore, the deletion of the two annotated exons of *Hotair* [20] had a previously ignored important impact in *cis* by generating two new transcripts, which may potentially interact with the transcription of neighboring *Hoxc* genes.

Discussion

Homeotic versus Homeopathic phenotypes

In this study, we have re-investigated the phenotypic and molecular effects of deleting the *Hotair* lncRNA on mouse development *in vivo*, as reported in [20]. In this previous study, three phenotypic differences between wild type and *Del(Hotair)^{-/-}* mice were reported, namely wrist malformation, a posterior homeotic transformation from lumbar vertebra L6 to sacral vertebra S1 identity and a mild anterior homeotic transformation of the 4th caudal vertebra. We did not detect any wrist malformation, nor did we see any substantial homeotic phenotype in the lumbar region of mutant animals, thus contradicting two of the three reported

phenotypic effects of the *Hotair* deletion. In *Mus musculus*, the lumbo-sacral transition shows great variability between L5 and L6 depending on the inbred strain considered and the total number of pre-sacral vertebrae. In fact, this number not only varies between inbred strains but also within the same strain and can even be biased by the sex of the animal [29]. Therefore, this region must be considered with great care before concluding on the presence of a homeotic transformation.

We note that another study involving two distinct deletion alleles of *Hotair*—though of smaller extents—also failed to confirm these latter two phenotypic effects [21]. The lack of effect of *Hotair* deletion upon wrist morphology is consistent with the absence of any detectable *Hotair* transcripts in mouse embryonic forelimbs (Fig 1) also reported previously [19] and by [21] using a sensitive *lacZ* reporter transgene system. Regarding the reported L5 to L6 transition, it is noteworthy that in wild type animals, detectable expression of *Hotair* in the paraxial mesoderm, i.e. in the mesodermal tissue that will generate the vertebrae, barely reaches the level of the lumbo-sacral transition, a transition labeled by its neighboring *Hoxc11* gene [30,31]. This makes a more anterior (at the L5 to L6 transition) *Hotair* loss-of-function dependent gain of function phenotype due to *Hoxd* genes difficult to understand.

Our analyses did nevertheless reveal a subtle difference between wild type and *Del(Hotair)^{-/-}* mice in the morphology of the post-sacral caudal vertebrae. Although this observation is in agreement with one of the previously described morphological alterations [20,21], we note that the penetrance of the mutant phenotype is much lower than the 100% reported by [20]. Also, such an anterior transformation should reflect a loss of function rather than the effect of de-repressed *Hox* genes [1,3], as already scored in some instances, for example when abrogating the function of the nearby located *Hoxc13* gene [32]. Moreover, we also observed variations in these vertebral morphologies amongst wild type animals. A potential explanation for the observed difference in phenotypic penetrance may reside in the genetic background of the animals. In this work, we used a mixed CBAxBL/6 strain, while previous studies used a BL/6 background. This relatively mild difference in genetic backgrounds may account, at least in part, for the discrepancy regarding the penetrance of this weak and physiologically poorly significant morphological transformation. Should this be the case, we would still have to conclude that the previously reported phenotypic effects of *Hotair* deletion are not only very mild but also inbred strain-specific, definitely arguing against a general role—even minor—of *Hotair* during mouse development. Accordingly, we would refer to these phenotypic alterations as homeopathic rather than homeotic [20].

The effects of the *Hotair* deletion in *trans*

Using sensitive RNA-seq measurements, we showed that the expression of hundreds of genes changed significantly upon deletion of the *Hotair* locus *in vivo*. However, none of these changes in gene expression could be reconciled with the suggested role for *Hotair* in silencing gene expression *in vivo* [20]. In particular, the initial proposal that *Hotair* RNA acts in *trans* to repress the expression of posterior *Hoxd* genes and of a subset of imprinted genes *via* the recruitment of the PRC2 complex [13,20] was not supported by our results. Indeed we did not note any significant change either in the levels, or in the spatial distribution of *Hoxd* transcripts, in any of the tissues analyzed. Also, when a larger set of putative *Pc* target genes was considered, the same conclusion was reached (S4 Fig). Finally, the majority (71%) of the differentially expressed imprinted genes including the reported *Hotair* targets *H19* and *Meg3*, were down- rather than up-regulated in *Del(Hotair)^{-/-}* mutant samples, again in contradiction with previous results.

Therefore, our results are at odds not only with the phenotypic outcome of the *Hotair* deletion, but also with its effects upon gene expression [20]. One potential explanation to these

serious discrepancies may be that the regulatory effect of *Hotair* is highly specific for tail tip or foreskin fibroblasts, which were previously used for functional investigations [13,20], whereas not at work *in vivo*, precisely in those embryonic tissues where *Hotair* is expressed at the highest levels. Indeed both *Hotair* and *Hoxd* genes transcripts are rather abundant in our tissue samples, while they are very weakly present in murine tail tip fibroblasts (Fig 1, S7 Fig). Unless *Hotair* would function more efficiently at low concentrations, we conclude that our *in vivo* system is better suited to reveal the role of *Hotair*, if any.

Another possibility is that the function of *Hotair* might not be exerted at the developmental stage analyzed (E12.5) but instead, at other time points. This explanation is nevertheless not compatible with the absence of phenotypic effects on skeletal morphology at P22, which should still be scored, should the deletion of *Hotair* deregulate target genes at other developmental stages. Also, the various genetic backgrounds may influence the penetrance of the phenotype (see above) and, by genotyping through the *Hotair* deleted locus, one may select for one particular haplotype associated with the mutant allele, which may result in some differential gene expression. Finally, it remains possible that a few hours difference in the developmental timing may lead to substantial relative variations in amounts of transcripts for many genes, in particular at an embryological stage where many important differentiation events occur.

In this context, it must be noted that the settings used for our transcriptome analyses overlap in sensitivity with the biological variations of the system itself, as seen for example with the variation in the level of *Hotair* in the GT replicate samples (Fig 1). Such differences can be due to intrinsic variations, yet most likely to slight variations in the micro-dissection plans or in the developmental stage of littermate embryos, or both. For example, a slight variation in the thickness of the piece in the trunk would elicit quantitative differences in *Hox* gene expression, whereas the depth of the piece (trunk) or the proximal level of the section (limbs, genitalia) may involve another presumptive tissue type, leading to large qualitative differences in transcripts. In fact, many of the strongest differentially expressed genes are clearly unrelated to those developmental processes involved in the potential morphological or molecular phenotypes under scrutiny (S10 Fig). This would also explain that differences are seen even in those samples where neither *Hotair*, nor *Hoxc11* are expressed. Accordingly, we do not interpret these results as reflecting changes in biological processes but, instead, as a sign of the sensitivity and intrinsic variations of our *in embryo* approach.

The effects of the *Hotair* deletion in *cis*

When investigating the roles of lncRNAs by genetic approaches *in vivo*, it is often problematic to separate the lncRNA-dependent effects from those generated by the manipulation of the corresponding genomic locus [33]. *Hotair* is transcribed from within the *HoxC* cluster, a tightly packed and gene-dense locus, and its deletion was reported to have no consequence on the transcription of the neighboring *Hoxc* genes at the developmental stage and cell types examined [20]. Here again, our results *in embryo* contradict this view and showed that the expression levels of both *Hoxc11* and *Hoxc12* changed upon deleting the *Hotair* locus. We observed an extension in the spatial distribution of *Hoxc11* transcripts in both the trunk and the hindlimbs of *Del(Hotair)^{-/-}* mutant specimens. Upon examination of the datasets of [20], we also found differences for *Hoxc10* and *Hoxc12* between wild type and *Del(Hotair)^{-/-}* tail tip fibroblasts (S7 Fig). Therefore, the deletion of the *Hotair* locus had a significant impact in *cis* on *Hoxc* gene expression, in both *in vivo* and *in vitro* systems.

This was confirmed by the observation of *Ghost of Hotair* (*Ghostair*), a novel RNA produced by the anti-*Hox* strand in the deletion mutant allele. This transcript initiates at one of the alternative *Hotair* promoters, which was not included into the deletion, and terminates

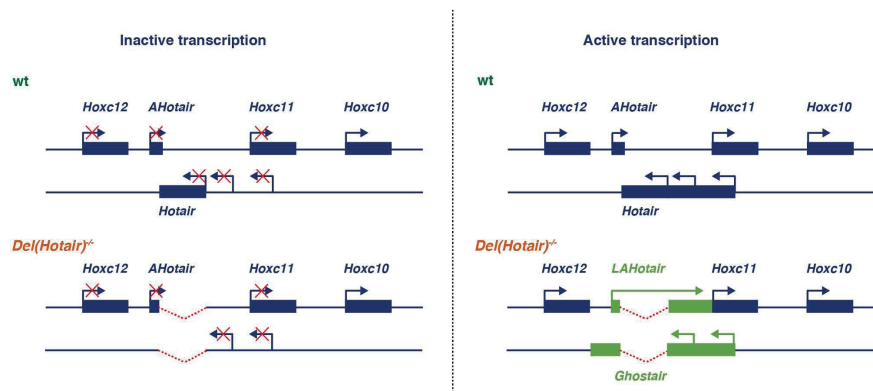


Fig 9. Schematic summarizing the data shown in Fig 8. In a wild type situation and at a body level anterior to *Hoxc11* expression (from example in the upper lumbar region such as L3 or L4), *Hoxc10* is active whereas the whole posterior part of the *HoxC* cluster is repressed (top left). Once *Hoxc11* and *Hoxc12* become activated, in more posterior regions of the body, both the *Hotair* and the *AntiHotair* RNAs are produced, from the anti-*Hox* and *Hox* DNA strands, respectively (top, right). Upon deletion of the *Hotair* locus, the posterior *HoxC* cluster remains closed for transcription in the anterior parts of the body (bottom, left). In contrast, the activation of the *Hoxc11* region (bottom right) triggers the transcription of *Ghostair* on the opposite DNA strand, which meets the 3' end of the *Hoxc12* transcription unit, perhaps causing the light decrease in *Hoxc12* mRNAs. On the *Hox* strand, the anti-*Hotair* RNA can now cross over the deleted region and contribute to the transcription of *Hoxc11*, perhaps inducing the observed light gain of expression of the latter gene.

doi:10.1371/journal.pgen.1006232.g009

close to the 3' end of the *Hoxc12* transcript on the opposite strand. Our analysis also revealed the existence of *AntiHotair*, a previously un-annotated transcript on the *Hox* strand, derived from a CpG island promoter located close to the 3' end of *Hotair*. While in the wild type situation this transcript remains relatively short and ends within the region targeted by the *Hotair* deletion, a longer *AntiHotair* transcript was produced in *Del(Hotair)^{-/-}* samples with no clear separation with *Hoxc11*. As a consequence, this transcript could leak onto *Hoxc11*, acting as an alternative 5' un-translated region, which gives a mechanistic basis for the light gain of *Hoxc11* expression in *Del(Hotair)^{-/-}* tissues (Fig 9). Such in-cis effects on the local transcription landscape by deleting transcription termination signals on both strands are likely independent from any possible *Hotair* function.

Hotair back into context

The *Hotair* lncRNA is transcribed from within the *HoxC* gene cluster [13], i.e. one of the most gene-dense and GC-rich regions of mammalian genomes [34]. Due to the particular regulatory strategy at work on the four *Hox* gene clusters [8], any endogenous or exogenous promoter present within such a gene cluster will be transcribed at the place and time where the neighboring *Hox* genes will be activated. The transcription of *Hotair* is no exception to this rule, for transcripts are found posteriorly, roughly matching the expression domains of *Hoxc11* or *Hoxc12*. While it is indeed possible that *Hotair* exerts a genuine function during development, for example by micro-tuning the transcription of *Hoxc* genes in cis, the question as to whether or not this RNA could be a mere by-product of the complex regulation occurring in the gene cluster remains open, in our opinion.

In any case, a potential mis-regulation of *Hoxc* genes should be carefully considered when investigating *Hotair* functions. This is desirable not only when studying developmental phenotypes [18,19,35], where it may represent a confounding factor due to the known roles of *Hoxc* genes there, but also when studying the roles of *Hotair* in other biological processes including human diseases. For instance, it was reported that *Hotair* is overexpressed in breast cancer and that this RNA regulates metastasis by reprogramming chromatin via *Polycomb* complexes [36]. Our analysis of expression data obtained from a cohort of cancer samples [37] revealed a strong positive correlation between *Hotair* and *Hoxc11* expression (S8 Fig), also observed in our mouse wild type samples (S8 Fig). Therefore, while *Hotair* may indeed be involved in a variety of cancer conditions, it is likely that its over-expression in cancer cells is accompanied by *Hoxc11* over-expression, which may again confound the observed phenotypes.

Conclusion

Thus far, four different alleles have been studied, which partially or entirely removed the *Hotair* lncRNA and no consensus has been found regarding a potential function of this RNA during mouse development [19–21]. In our hands, *Hotair* has no function during mouse development, *a fortiori* when the regulation of *Hoxd* genes in *trans* is concerned. The deletion of the locus engineered by [20] induces modifications in the transcription of some *Hoxc* genes, through complex re-allocations of promoter and termination sites leading to novel RNA species. This mis-regulation of *Hoxc* gene transcription may have a slight effect upon some vertebral morphologies, yet this impact—if any—would be poorly penetrant and inbred strain-specific, i.e. of little interest for our understanding of developmental processes at large. Yet another allele would be necessary to solve these discrepancies, whereby the CRISPR-cas9 technology would help abrogate the *Hotair* transcription without substantially modifying the *in-cis* environment. At this point however, we do not see the urgency of increasing the number of mutant alleles at this locus, as confounds due to genetic background differences may always blur the resolution of such subtle effects.

Materials and Methods

Mouse strains

The *Del(Hotair)* mouse strain was described in [20] and kindly provided by Dr. H. Chang. Heterozygous mice were crossed back onto a mixed CBAXC57/B6 background (Charles River). Wild type, heterozygous and homozygous mutant embryos were obtained by intercrossing heterozygous mice. Genotyping was performed by PCR analysis on individual yolk sac lysates using the following primers:

Wild type (F)–CCTTATTCTCCCGGAGCCTAGC

Wild type (R)–CTGCCTCTGGCTCCACTCC

Del(Hotair)^{-/-} (F)–CCTTGCCAACGTGTGGCTTCC

Del(Hotair)^{-/-} (R)–CCAAGTCTACCGCTACTGCG

Ethics statement

Maintenance of, and experiments on animals were approved by the Geneva Canton ethical regulation authority (authorization GE/81/14 to D.D.) and performed according to Swiss law.

Whole-mount *in situ* hybridization

Whole-mount *in situ* hybridizations (WISH) were performed according to standard protocols. Embryos were dissected in PBS and fixed overnight in 4% paraformaldehyde (PFA), washed in PBS, dehydrated and stored in 100% methanol at -20°C . Both *Del(Hotair)^{-/-}* and control wild type E12.5 littermates embryos were processed in parallel to maintain identical conditions throughout the WISH procedure. DIG-labeled probes for *in situ* hybridizations were produced by *in vitro* transcription (Promega) and detection was carried out using an alkaline phosphatase conjugated anti-digoxigenin antibody (Roche). WISH probes templates were previously described in: *Hotair* [19]; *Hoxd10* and *Hoxd11* [27]; *Hoxc11* [35]; *Hoxc12* [34] and *MyoD* [26].

Skeletal Preparation

Whole mount skeletal preparation of P22 animals was done with standard Alcian blue/Alizarin red staining protocols.

RNA extraction and RNA-seq library preparation

Embryonic tissues were stored at -80°C in RNeasy lysis reagent (Ambion) before genotyping. After genotyping and embryo sorting, total RNA was extracted from tissues using QIAGEN RNeasy Plus Micro Kit after disruption and homogenization. RNA quality was assessed using an Agilent 2100 Bioanalyser. Only samples with high RNA integrity number were used. Sequencing libraries were prepared according to TruSeq Stranded mRNA Illumina protocol, with polyA selection. RNA-seq libraries were sequenced on an Illumina HiSeq 2500 sequencer, as single-end reads (read length 100 base pairs). We obtained between 36 and 54 millions of raw RNA-seq reads for each sample (S1 Table).

RNA-seq processing and gene expression estimation

Raw RNA-seq reads were aligned on the mouse mm10 genome assembly using TopHat 2.0.9 [38]. Gene expression computations were performed using uniquely mapping reads extracted from TopHat alignments and genomic annotations from Ensembl release 82 [39]. We filtered the annotated transcript isoforms for protein-coding genes, keeping only transcripts annotated as 'protein-coding', thus discarding transcripts flagged as 'retained_intron', 'nonsense-mediated decay' etc. For *Hox* genes, we manually inspected annotated transcripts and retained only the canonical isoform for each gene, discarding read-through transcripts and retained introns. For non-coding genes, all annotated isoforms were kept. We then constructed 'flattened' gene models by combining the exon coordinates from all retained isoforms and counted the number of unique reads that aligned on these exons. We discarded reads that aligned on two or more overlapping genes on the same strand, as well as reads containing more than 2 mismatches or small insertions or deletions. We computed RPKM (Read per Kilobase of Exon per Million mapped reads) expression levels for each gene based on the unique read counts. The total number of mapped reads was computed on the entire nuclear genome, discarding reads that mapped on the mitochondria. RPKM expression levels were then further normalized across samples with a median scaling procedure, using as a standard the 100 genes with the least expression rank variation across samples, found in the 25%-75% range of expression levels [40]. As a control, we also computed expression levels using all TopHat mapped reads and the multi-read and fragment bias correction procedures implemented in Cufflinks [41]. The same procedure was applied for previously published tail tip fibroblast RNA-seq samples [20]. The RNA-seq data presented in this previous publication were also strand-specific and

generated with a dUTP protocol that sequences the antisense mRNA strand like the TruSeq Stranded mRNA protocol.

Statistical analyses and graphical representations

The principal component analysis (PCA) was performed using the `dudi.pca` function in the `ade4` package in R [42]. The input table for the PCA consisted of log₂-transformed RPKM expression levels, for all protein-coding genes that had RPKM >1 in at least one of our samples. The data was centered (meaning that the mean expression levels were brought to a value of 0 for each gene, removing between-gene variations in expression levels) but not scaled prior to the PCA analysis. Euclidean distances between samples were computed with the standard `dist` function in R and clustered using the hierarchical clustering method (`hclust`). All statistics and graphical representations were done in R.

Differential expression analyses

We tested for differential gene expression using DESeq2 [43] in R. Specifically, we contrasted a generalized linear model that explains the variation in read counts for each gene, as a function of the genotype (wild type or *Del(Hotair)^{-/-}*) with a null model that assumes no effect of the genotype. The analyses presented in the manuscript were performed with the likelihood ratio test (LRT); the Wald test was performed as a control and results are provided in the supplementary datasets. The tests were performed separately for each tissue. The p-values were corrected for multiple testing with the Benjamini-Hochberg approach, for all six tissues at the same time. The same procedure was applied for previously published tail tip fibroblast RNA-seq samples, which included wild type, heterozygous *Del(Hotair)^{+/-}* and homozygous *Del(Hotair)^{-/-}* samples. In this case, we performed three separate pairwise comparisons between the three genotypes.

Gene ontology enrichments

We tested for gene ontology (GO) enrichment in the sets of differentially expressed genes using the GOrilla webserver [44]. Each enrichment analysis compared two lists of genes, the focal list containing differentially expressed protein-coding genes (up-regulated and down-regulated genes analyzed separately) and the background list containing all protein-coding genes expressed in the corresponding samples. To construct the background list, we computed the minimum number of reads observed for differentially expressed protein-coding genes, summed across all relevant samples and selected all genes that had equivalent or higher read counts.

Prediction of *Polycomb* target genes

To obtain a list of putative *Polycomb* target genes, we analyzed chromatin immuno-precipitation followed by sequencing (ChIP-seq) data for H3K27me₃ and corresponding input data, from wild type tail tip fibroblasts [20]. We mapped the ChIP-seq data on the mm10 mouse genome using Bowtie 2 [45]. We removed identical ChIP-seq reads to avoid biases stemming from PCR duplication and we kept unambiguously mapped reads with at most two mismatches. We computed the average H3K27me₃ and input read coverage in the promoter region (defined as 2kb upstream the annotated transcription start site) for each Ensembl-annotated transcript. The same conclusions were reached when defining promoter regions as 4kb regions centered on the TSS (S6 Dataset). The read coverage was normalized by dividing by the total number of million mapped reads for each sample. We defined putative *Polycomb* targets as those genes for which the ratio between the H3K27me₃ and input was at least 5, and

for which the absolute H3K27me3 normalized coverage was at least 0.1. We discarded genes that had satellite repeats in the promoter regions as we observed that these repeats are enriched in H3K27me3 marks (likely as an artifact).

Imprinted genes

The list of known mouse imprinted genes was extracted from <http://www.geneimprint.com>.

De novo transcript assembly at the *HoxC* locus

We used RNA-seq data from our *Del(Hotair)^{-/-}* samples, combined across all six tissues, to predict transcript sequences for *Ghostair* and *AntiHotair*. To do this, we first split each RNA-seq reads into three segments and aligned them with Bowtie 2 on the DNA sequence delimited by *Hoxc12* and *Hoxc11*. We then extracted all RNA-seq reads that mapped at least partially onto this region and assembled transcripts *de novo* using Trinity [46] setting `SS_lib_type = R` since our data was strand-specific. We kept Trinity contigs with a minimum length of 1000bp and aligned them on the mouse chromosome 15 using BLAT [47]. We manually excluded small, repetitive BLAT hits. See also [S8 Dataset](#).

Expression data in cancer samples

To study the correlation between *Hotair* expression and the expression of neighboring genes *Hoxc11* and *Hoxc12*, we analyzed gene expression levels (RPM = reads per total million mapped reads) for a cohort of cancer samples [37].

Supporting Information

S1 Fig. Correlations between biological replicates. The scatterplots show the correlation of log₂-transformed RPKM expression levels of all expressed protein coding genes between biological replicates (Pearson and Spearman correlation coefficients). FL, Forelimbs; HL, hindlimbs; GT, genital tubercle; T1, T2, T3; trunk samples corresponding to either the lumbosacral, the sacro-caudal region or the caudal region, respectively.
(TIF)

S2 Fig. Pairwise Euclidean distances between samples. Hierarchical clustering and heat map of pairwise Euclidean distances between samples, computed on log₂-transformed RPKM expression levels of all protein coding genes. The distances are color-coded, with blue representing small distances and yellow large distances.
(TIF)

S3 Fig. Numbers of differentially expressed genes, in common among tissues. Venn diagrams of differential expression analysis results (fold change > 1.5 and FDR < 0.05) for all tissue samples that express *Hotair* in the wild type condition (T2, T3, GT and HL). (A) Venn diagram showing the down-regulated genes. (B) Venn diagram showing up-regulated genes.
(TIF)

S4 Fig. Differential expression for candidate *Polycomb* target genes in the T3 trunk sample. (A) Volcano plot representing the log₂ fold change and the false discovery rate (log₁₀-transformed) for candidate *Polycomb* target genes (see [Materials and Methods](#)) in T3. The direction of expression changes is color-coded, with red showing down-regulated genes and green up-regulated genes. Non-significant genes are in blue. (B) Barplot representing the percentage of up-regulated and down-regulated genes for candidate *Polycomb* targets (left) and other genes (right).
(TIF)

S5 Fig. Expression of known imprinted genes in wild type and *Del(Hotair)*^{-/-} embryonic tissues. Heat map of log₂-transformed RPKM expression levels of all imprinted genes (extracted from <http://www.genemimprint.com>) expressed in our samples (RPKM > 1, S7 Dataset). The columns correspond to samples and the rows correspond to imprinted genes. Blue boxes indicate down-regulated genes whereas black boxes indicate up-regulated genes (fold change > 1.5 and FDR < 10%).

(TIF)

S6 Fig. Expression of *Hoxa* and *Hoxb* genes in wild type and *Del(Hotair)*^{-/-} dissected embryonic tissues. (A) RNA-seq expression profiles of the *HoxA* genomic region in the trunk T1 (top), T2 (middle) and T3 (bottom) samples of either wild type (green) or *Del(Hotair)*^{-/-} (orange) E12.5 embryos. Very subtle differences are scored (arrows), which are nevertheless considered as significant using our analytical parameters (see also Fig 5). The Y-axis represents the per-base RNA-seq read coverage, normalized by dividing by the total number of million mapped reads in the corresponding samples. The two biological replicates were pooled for this representation and only uniquely mapping reads were used. (B) RNA-seq expression profiles of the *HoxB* genomic region in the hindlimbs (HL, top) and genital tubercle (GT, bottom) of either wild type (green) or *Del(Hotair)*^{-/-} (red) E12.5 embryos. There again, the difference observed for *Hoxb9* (Fig 5) is weak yet considered as significant in our conditions (FDR < 10%, no fold change threshold).

(TIF)

S7 Fig. Expression of *Hox* genes in tail tip fibroblasts (TTF). Bar plots showing the quantification of all *Hox* genes expression by RNA-seq (normalized RPKM values) in TTF. Datasets are from [20]. The FDR of the differential expression test (likelihood ratio test in DESeq2) between wild type and *Del(Hotair)*^{-/-} samples is indicated in red above each gene.

(TIF)

S8 Fig. Correlations between the expression levels of *Hotair* and of the neighboring *Hoxc* genes. (A) Scatterplot of log₂-transformed RPKM expression levels for *Hotair* and *Hoxc11* shows excellent correlation (R = 0.898). (B) Scatterplot of log₂-transformed RPKM expression levels for *Hotair* and *Hoxc12* with a lower correlation coefficient (R = 0.592). The various tissues samples are represented by a color code and the genotypes are indicated by either a circle (wild type), or a triangle (*Del(Hotair)*^{-/-}) (C) Scatterplot of log₂-transformed RPM expression levels for *Hotair* and the *Hoxc11* gene in a cohort of 376 cancer samples [37], showing a high correlation coefficient (R = 0.92). (D) Scatterplot of log₂-transformed RPM expression levels for *Hotair* and the *Hoxc12* gene in the same cohort as in C) [37], showing a lower correlation coefficient (R = 0.66). The Pearson correlation coefficients are shown above the plot.

(TIF)

S9 Fig. *De novo* transcript assembly in the *HoxC* cluster for a pool of the various samples in *Del(Hotair)*^{-/-} mutant specimens. (A) Existing annotations for *Hoxc11*, *Hoxc12* and *Hotair*, as extracted from the Ensembl database. (B) Variation in the RNA-seq reads coverage at the vicinity of *Hotair*. The *Hox* and anti-*Hox* strands are depicted separately. The Y-axis represents the per-base RNA-seq read coverage, normalized by dividing by the total number of million mapped reads in the corresponding samples. All our *Del(Hotair)*^{-/-} samples were pooled for this representation and only uniquely mapping reads were used. (C) Genomic coordinates of *de novo* assembled transcripts, as predicted by Trinity on the basis of *Del(Hotair)*^{-/-} RNA-seq data and mapped on the genome with Blat (excluding repetitive hits). The different isoforms assigned to a single gene were combined for this representation. Note that transcripts may be fragmented, in particular at repeats and low complexity regions. We observe transcripts

crossing the deleted region on both the *Hox* strand (Trinity identifier *TR5603|c0_g3*) and the anti-*Hox* strand (Trinity identifier *TR5645|c0_g1*). (D) Positions of repeated elements in the vicinity of *Hotair*.

(TIF)

S10 Fig. Top differentially expressed genes in embryonic tissue samples. The volcano plots show the log₂ fold change and the false discovery rate (log₁₀-transformed) for differentially expressed genes (fold change > 1.5 and FDR < 0.05) in the various tissue samples. The top differentially expressed genes are highlighted in red (fold change > 4 and FDR < 0.00001). Dashed lines mark absolute fold changes of 2 and 4. FL, Forelimbs; HL, hindlimbs; GT, genital tubercle; T1, T2, T3; trunk samples corresponding to either the lumbo-sacral, the sacro-caudal region or the caudal region, respectively.

(TIF)

S1 Dataset. Gene expression data for wild type and *Del(Hotair)*^{-/-} embryonic mouse tissue samples. Data are from forelimb buds (FL), hindlimb buds (HL), genital tubercle (GT) and the trunk T1, T2 and T3 samples. Numbers of uniquely mapped reads and RPKM/FPKM values are provided in separate files.

(TGZ)

S2 Dataset. Gene expression for either wild type, *Del(Hotair)*^{+/-} or *Del(Hotair)*^{-/-} tail tip fibroblasts (TTF), after re-analysis of the RNA-seq datasets from Li and colleagues [20]. Numbers of uniquely mapped reads and RPKM/FPKM values are provided in separate files.

(TGZ)

S3 Dataset. Differential expression analyses comparing wild type and *Del(Hotair)*^{-/-} mouse tissue samples. Data are from the FL, HL, GT, T1, T2 and T3 and are provided in separate files.

(TGZ)

S4 Dataset. Differential expression analyses comparing wild type, *Del(Hotair)*^{+/-} and *Del(Hotair)*^{-/-} tail tip fibroblasts (TTF), after our re-analysis of the RNA-seq data from [20].

(TGZ)

S5 Dataset. Gene ontology (GO) enrichment analysis, contrasting genes that are differentially expressed between wild type and *Del(Hotair)*^{-/-} mouse tissues from a background set of genes expressed at comparable levels in the same tissues. The enrichment analysis was done for the GO category 'biological process', using the GORILLA webserver [44].

(TGZ)

S6 Dataset. Analysis of putative *Polycomb* target genes, predicted on the basis of H3K27me3 ChIPSeq data from Li and colleagues [20]. The table provides the H3K27me3 and input read coverage for the predicted *Polycomb* targets, as well as differential expression results for these genes in the T3 trunk segment.

(TGZ)

S7 Dataset. Results of the differential expression analyses for imprinted genes. The list of imprinted genes was extracted from www.geneimprint.com.

(TGZ)

S8 Dataset. Results of the *de novo* transcript assembly procedure, which we used to confirm that *Ghostair* and *AntiHotair* cross the deleted region.

(TGZ)

S1 Table. Basic statistics for the RNA-seq dataset, including the number of raw and mapped RNA-seq reads and the number of detected genes for each sample.
(XLSX)

Acknowledgments

We thank H. Chang for providing the *Hotair* mutant stock. We also thank M. Ros as well as L. Deslile and other lab members for their expertise, comments and materials, T.-H. Nguyen Huyn for mouse breeding, A. Awgulewitsch for WISH probes and J. Zakany for comments on the manuscript. We thank the Geneva Genomics Platform (University of Geneva) for sequencing.

Author Contributions

Conceived and designed the experiments: ARA AN DD.

Performed the experiments: ARA AN EJ BM.

Analyzed the data: ARA AN DD.

Wrote the paper: ARA AN DD.

References

1. Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276: 565–570. PMID: [103000](#)
2. Krumlauf R (1994) Hox genes in vertebrate development. *Cell* 78: 191–201. PMID: [7913880](#)
3. Duboule D, Morata G (1994) Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet* 10: 358–364. PMID: [7985240](#)
4. Wellik DM (2009) Hox Genes and Vertebrate Axial Pattern. *Hox Genes* 88: 257–+.
5. Holland PW, Garcia-Fernandez J (1996) Hox genes and chordate evolution. *Dev Biol* 173: 382–395. doi: [10.1006/dbio.1996.0034](#) PMID: [8605999](#)
6. Duboule D (2007) The rise and fall of Hox gene clusters. *Development* 134: 2549–2560. doi: [10.1242/dev.001065](#) PMID: [17553908](#)
7. Deschamps J, van Nes J (2005) Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* 132: 2931–2942. doi: [10.1242/dev.01697](#) PMID: [15944185](#)
8. Kmita M, Duboule D (2003) Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301: 331–333. doi: [10.1126/science.1085753](#) PMID: [12869751](#)
9. Berlivet S, Paquette D, Dumouchel A, Langlais D, Dostie J, et al. (2013) Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *PLoS Genet* 9: e1004018. doi: [10.1371/journal.pgen.1004018](#) PMID: [24385922](#)
10. Soshnikova N, Duboule D (2009) Epigenetic temporal control of mouse Hox genes in vivo. *Science* 324: 1320–1323. doi: [10.1126/science.1171468](#) PMID: [19498168](#)
11. Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, et al. (2011) The dynamic architecture of Hox gene clusters. *Science* 334: 222–225. doi: [10.1126/science.1207194](#) PMID: [21998387](#)
12. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, et al. (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472: 120–124. doi: [10.1038/nature09819](#) PMID: [21423168](#)
13. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, et al. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129: 1311–1323. doi: [10.1016/j.cell.2007.05.022](#) PMID: [17604720](#)
14. Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, et al. (2013) Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* 2: e01749. doi: [10.7554/eLife.01749](#) PMID: [24381249](#)

15. Mainguy G, Koster J, Woltering J, Jansen H, Durston A (2007) Extensive polycistronism and antisense transcription in the mammalian Hox clusters. *PLoS One* 2: e356. doi: [10.1371/journal.pone.0000356](https://doi.org/10.1371/journal.pone.0000356) PMID: [17406680](https://pubmed.ncbi.nlm.nih.gov/17406680/)
16. Lee JT (2012) Epigenetic regulation by long noncoding RNAs. *Science* 338: 1435–1439. doi: [10.1126/science.1231776](https://doi.org/10.1126/science.1231776) PMID: [23239728](https://pubmed.ncbi.nlm.nih.gov/23239728/)
17. Tsai MC, Manor O, Wan Y, Mosammamaparast N, Wang JK, et al. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329: 689–693. doi: [10.1126/science.1192002](https://doi.org/10.1126/science.1192002) PMID: [20616235](https://pubmed.ncbi.nlm.nih.gov/20616235/)
18. Suemori H, Noguchi S (2000) Hox C cluster genes are dispensable for overall body plan of mouse embryonic development. *Dev Biol* 220: 333–342. doi: [10.1006/dbio.2000.9651](https://doi.org/10.1006/dbio.2000.9651) PMID: [10753520](https://pubmed.ncbi.nlm.nih.gov/10753520/)
19. Schorderet P, Duboule D (2011) Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet* 7: e1002071. doi: [10.1371/journal.pgen.1002071](https://doi.org/10.1371/journal.pgen.1002071) PMID: [21637793](https://pubmed.ncbi.nlm.nih.gov/21637793/)
20. Li L, Liu B, Wapinski OL, Tsai MC, Qu K, et al. (2013) Targeted disruption of Hotair leads to homeotic transformation and gene derepression. *Cell Rep* 5: 3–12. doi: [10.1016/j.celrep.2013.09.003](https://doi.org/10.1016/j.celrep.2013.09.003) PMID: [24075995](https://pubmed.ncbi.nlm.nih.gov/24075995/)
21. Lai KM, Gong G, Atanasio A, Rojas J, Quispe J, et al. (2015) Diverse Phenotypes and Specific Transcription Patterns in Twenty Mouse Lines with Ablated LincRNAs. *PLoS One* 10: e0125522. doi: [10.1371/journal.pone.0125522](https://doi.org/10.1371/journal.pone.0125522) PMID: [25909911](https://pubmed.ncbi.nlm.nih.gov/25909911/)
22. Tschopp P, Duboule D (2011) A genetic approach to the transcriptional regulation of hox gene clusters. *Annu Rev Genet* 45: 145–166. doi: [10.1146/annurev-genet-102209-163429](https://doi.org/10.1146/annurev-genet-102209-163429) PMID: [22060042](https://pubmed.ncbi.nlm.nih.gov/22060042/)
23. Tschopp P, Duboule D (2014) The Genetics of Murine Hox Loci: TAMERE, STRING, and PANTHERE to Engineer Chromosome Variants. *Methods Mol Biol* 1196: 89–102. doi: [10.1007/978-1-4939-1242-1_6](https://doi.org/10.1007/978-1-4939-1242-1_6) PMID: [25151159](https://pubmed.ncbi.nlm.nih.gov/25151159/)
24. Wellik DM (2007) Hox patterning of the vertebrate axial skeleton. *Dev Dyn* 236: 2454–2463. doi: [10.1002/dvdy.21286](https://doi.org/10.1002/dvdy.21286) PMID: [17685480](https://pubmed.ncbi.nlm.nih.gov/17685480/)
25. Zakany J, Duboule D (2007) The role of Hox genes during vertebrate limb development. *Current Opinion in Genetics & Development* 17: 359–366.
26. Hebrok M, Fuchtbauer A, Fuchtbauer EM (1997) Repression of muscle-specific gene activation by the murine Twist protein. *Exp Cell Res* 232: 295–303. doi: [10.1006/excr.1997.3541](https://doi.org/10.1006/excr.1997.3541) PMID: [9168805](https://pubmed.ncbi.nlm.nih.gov/9168805/)
27. Gerard M, Chen JY, Gronemeyer H, Chambon P, Duboule D, et al. (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. PMID: [8824591](https://pubmed.ncbi.nlm.nih.gov/8824591/)
28. Izpisua-Belmonte JC, Falkenstein H, Dolle P, Renucci A, 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. PMID: [1676674](https://pubmed.ncbi.nlm.nih.gov/1676674/)
29. Green EL (1941) Genetic and Non-Genetic Factors Which Influence the Type of the Skeleton in an Inbred Strain of Mice. *Genetics* 26: 192–222. PMID: [17247002](https://pubmed.ncbi.nlm.nih.gov/17247002/)
30. Wellik DM, Capecchi MR (2003) Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science* 301: 363–367. doi: [10.1126/science.1085672](https://doi.org/10.1126/science.1085672) PMID: [12869760](https://pubmed.ncbi.nlm.nih.gov/12869760/)
31. Hostikka SL, Capecchi MR (1998) The mouse Hoxc11 gene: genomic structure and expression pattern. *Mech Dev* 70: 133–145. PMID: [9510030](https://pubmed.ncbi.nlm.nih.gov/9510030/)
32. Godwin AR, Capecchi MR (1998) Hoxc13 mutant mice lack external hair. *Genes Dev* 12: 11–20. PMID: [9420327](https://pubmed.ncbi.nlm.nih.gov/9420327/)
33. Bassett AM, Akhtar A, Barlow DP, Bird AP, Brockdorff N, et al. (2014) Considerations when investigating lincRNA function in vivo. *Elife* 3: e03058. doi: [10.7554/eLife.03058](https://doi.org/10.7554/eLife.03058) PMID: [25124674](https://pubmed.ncbi.nlm.nih.gov/25124674/)
34. Peterson RL, Papenbrock T, Davda MM, Awgulewitsch A (1994) The murine Hoxc cluster contains five neighboring AbdB-related Hox genes that show unique spatially coordinated expression in posterior embryonic subregions. *Mech Dev* 47: 253–260. PMID: [7848872](https://pubmed.ncbi.nlm.nih.gov/7848872/)
35. Papenbrock T, Visconti RP, Awgulewitsch A (2000) Loss of fibula in mice overexpressing Hoxc11. *Mech Dev* 92: 113–123. PMID: [10727851](https://pubmed.ncbi.nlm.nih.gov/10727851/)
36. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, et al. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464: 1071–1076. doi: [10.1038/nature08975](https://doi.org/10.1038/nature08975) PMID: [20393566](https://pubmed.ncbi.nlm.nih.gov/20393566/)
37. Balbin OA, Malik R, Dhanasekaran SM, Prensner JR, Cao X, et al. (2015) The landscape of antisense gene expression in human cancers. *Genome Res* 25: 1068–1079. doi: [10.1101/gr.180596.114](https://doi.org/10.1101/gr.180596.114) PMID: [26063736](https://pubmed.ncbi.nlm.nih.gov/26063736/)

38. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcripts in the presence of insertions, deletions and gene fusions. *Genome Biol* 14: R36. doi: [10.1186/gb-2013-14-4-r36](https://doi.org/10.1186/gb-2013-14-4-r36) PMID: [23618408](https://pubmed.ncbi.nlm.nih.gov/23618408/)
39. Yates A, Akanni W, Amode MR, Barrell D, Billis K, et al. (2016) Ensembl 2016. *Nucleic Acids Res* 44: D710–716. doi: [10.1093/nar/gkv1157](https://doi.org/10.1093/nar/gkv1157) PMID: [26687719](https://pubmed.ncbi.nlm.nih.gov/26687719/)
40. Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, et al. (2011) The evolution of gene expression levels in mammalian organs. *Nature* 478: 343–348. doi: [10.1038/nature10532](https://doi.org/10.1038/nature10532) PMID: [22012392](https://pubmed.ncbi.nlm.nih.gov/22012392/)
41. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12: R22. doi: [10.1186/gb-2011-12-3-r22](https://doi.org/10.1186/gb-2011-12-3-r22) PMID: [21410973](https://pubmed.ncbi.nlm.nih.gov/21410973/)
42. Dray S, Dufour AB (2007) The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* 22: 1–20.
43. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550. doi: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8) PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)
44. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48. doi: [10.1186/1471-2105-10-48](https://doi.org/10.1186/1471-2105-10-48) PMID: [19192299](https://pubmed.ncbi.nlm.nih.gov/19192299/)
45. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–359. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
46. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644–652. doi: [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883) PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/)
47. Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12: 656–664. doi: [10.1101/gr.229202](https://doi.org/10.1101/gr.229202) PMID: [11932250](https://pubmed.ncbi.nlm.nih.gov/11932250/)

9. Curriculum Vitae

Ana Rita de Carvalho Amândio Lhopitalier

Rue do centre 158

1025 St-Sulpice

+41 (0) 787973382

ritaamandio@gmail.com

EDUCATION

2014 – Present	Doctoral Program in Molecular Life Sciences EDMS, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
2009 – 2010	Advanced Study Program in Territory, Environment and Sustainable Development, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCTUNL)
2007 – 2008	MSc in Developmental and Evolutionary Biology, Faculdade de Ciências da Universidade de Lisboa (FCUL)
2002 – 2007	Licenciatura - Cell Biology and Biotechnology, Faculdade de Ciências da Universidade de Lisboa (FCUL)

EMPLOYMENT HISTORY

2014-present	PhD student, Laboratory of Developmental Genomics (Prof. Denis Duboule), Swiss Institute for Experimental Cancer Research (ISREC), EPFL (Switzerland). <i>Analysis of long-range gene regulation at the HoxD locus</i>
2011 – 2013	Research assistant, Actin Dynamics Laboratory, Instituto Gulbenkian de Ciência (IGC) (Portugal)
2010 – 2011	Research Intern, Neuropharmacology Department, Elan Pharmaceuticals inc. CA 94080 (USA)
2009 – 2010	Research assistant, Actin Dynamics Laboratory, Instituto Gulbenkian de Ciência (IGC) (Portugal)
2007 – 2009	MSc student, Laboratory of Development and Evolutionary Morphogenesis, Centre for ecology, evolution, and environmental changes (CE3c) (Portugal). <i>Morphodynamics of somite epithelialisation in chick embryos</i>

PUBLICATIONS

Articles in referred journals

Ana Rita Amândio, Anamaria Necșulea, Elisabeth Joye, Bénédicte Mascrez, and Denis Duboule. 2016. "Hotair Is Dispensable for Mouse Development," *PLoS Genetics* 1–27. doi:10.1371/journal.pgen.1006232.

Ana Rita Amândio, Pedro Gaspar, Jessica L Whited, and Florence Janody. 2014. "Subunits of the Drosophila Actin-Capping Protein Heterodimer Regulate Each Other at Multiple Levels." *PLoS ONE* 9 (5). doi:10.1371/journal.pone.0096326.

Barbara Jezowska, Beatriz Garcia Fernandez, Ana Rita Amândio, Paulo Duarte, Claudia Mendes, Catarina Bras-Pereira, and Florence Janody. 2011. "A Dual Function of Drosophila Capping Protein on DE-Cadherin Maintains Epithelial Integrity and Prevents JNK-Mediated Apoptosis." *Developmental Biology* 360 (1). doi:10.1016/j.ydbio.2011.09.016.

Gabriel G. Martins, Pedro Rifes, Ana Rita Amândio, Gabriela Rodrigues, Isabel Palmeirim, and Sólveig Thorsteinsdóttir. 2009. "Dynamic 3D Cell Rearrangements Guided by a Fibronectin Matrix Underlie Somiteogenesis." *PLoS ONE* 4 (10). doi:10.1371/journal.pone.0007429.

Book chapters

Gabriel G. Martins, Pedro Rifes, Ana Rita Amândio, Pedro Campinho, Isabel Palmeirim, and Sólveig Thorsteinsdóttir, 2007. 3D visualization and analysis of cell-matrix transformations in whole-mount and live embryos using confocal and multi-photon microscopy. pp. 426-433. In A. Méndez-Vilas, J. Díaz (Eds.) *Modern Research and Educational Topics in Microscopy*. Microscopy Series nº 3, Vol. 1. Formatex, Badajoz, Spain. (ISBN-13:978-84-611-9419-3).

