Control of *Hoxd* gene transcription in the mammary bud by hijacking a preexisting regulatory landscape

Ruben Schep[a], Anamaria Necsulea[b], Eddie Rodríguez-Carballo[a], Isabel Guerreiro[a], Guillaume Andrey[b,1], Thi Hanh Nguyen Huynh[a], Virginie Marcet[a], Jozsef Zákány[a], Denis Duboule[a,b,2], and Leonardo Beccari[a]

[a]Department of Genetics and Evolution, University of Geneva,1211 Geneva, Switzerland; and [b]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Contributed by Denis Duboule, October 23, 2016 (sent for review July 31, 2016; reviewed by Marcelo A. Nobrega and Gunter P. Wagner)

Vertebrate *Hox* genes encode transcription factors operating during the development of multiple organs and structures. However, the evolutionary mechanism underlying this remarkable pleiotropy remains to be fully understood. Here, we show that *Hoxd8* and *Hoxd9*, two genes of the *HoxD* complex, are transcribed during mammary bud (MB) development. However, unlike in other developmental contexts, their coexpression does not rely on the same regulatory mechanism. *Hoxd8* is regulated by the combined activity of closely located sequences and the most distant telomeric gene desert. On the other hand, *Hoxd9* is controlled by an enhancer-rich region that is also located within the telomeric gene desert but has no impact on *Hoxd8* transcription, thus constituting an exception to the global regulatory logic systematically observed at this locus. The latter DNA region is also involved in *Hoxd* gene regulation in other contexts and strongly interacts with *Hoxd9* in all tissues analyzed thus far, indicating that its regulatory activity was already operational before the appearance of mammary glands. Within this DNA region and neighboring a strong limb enhancer, we identified a short sequence conserved in therian mammals and capable of enhancer activity in the MBs. We propose that *Hoxd* gene regulation in embryonic MBs evolved by hijacking a preexisting regulatory landscape that was already at work before the emergence of mammals in structures such as the limbs or the intestinal tract.

enhancers | TAD | mammalian development | mammary gland

Animal development is orchestrated by a limited number of signaling molecules and transcription factors that cooperate in complex gene regulatory networks (1). The various elements of these networks are repeatedly used in time and space and were often coopted in the course of evolution to support the appearance of anatomical novelties (2–6). *Hox* genes are a good example of genes that underwent several rounds of neofunctionalization to accompany the emergence of morphological and functional novelties of many kinds, such as tetrapod digits (7) or pregnancy in mammals (8). They encode transcription factors with critical roles in the development of various embryonic structures, including the main body and appendicular axes, as well as virtually all major systems (9–11).

In amniotes, 39 *Hox* genes are generally found located in four different genomic clusters that arose following two rounds of whole-genome duplications (12–14). Although this clustered organization facilitates their coordinated transcriptional activation during the elongation of the major body axis (10, 15), it also favors the evolution of global regulations outside the gene clusters themselves (16). Consequently, specific subgroups of *Hox* genes, at least within the *HoxA* and *HoxD* clusters, are controlled by the same series of global enhancers; hence, they are coregulated in different contexts, which provide the system with both quantitative and qualitative modulations. This regulation has been well documented for the *HoxD* cluster, which contains a range of long-acting enhancers of different specificities within its two large flanking gene deserts (17–20).

The application of chromosome conformation capture approaches (21, 22) on this locus using in vivo material has revealed that these two large regulatory landscapes match in their extent two topologically associating domains (TADs) [i.e., chromosome domains where specific and constitutive physical interactions are privileged (23, 24)]. These two TADs are separated by a tight boundary localized within the *HoxD* cluster itself, isolating those genes controlled by the telomeric TAD (T-DOM) from those genes responding to the centromeric regulation domain (C-DOM) (17) (see Fig. 5*A*). Within the T-DOM, series of enhancers are found, which regulate groups of genes lying in the central part of the cluster, as if the regulatory sequences would contact a chromatin pocket encompassing *Hoxd8* to *Hoxd11* (18). These large and apparently constitutive chromosome domains (TADs) may facilitate the required regulatory switches at important developmental loci, and were also proposed to have triggered the evolution of pleiotropy by providing the regulatory context for evolving novel enhancers (16, 19).

Mammary glands (MGs) are a defining characteristic of mammals. They are highly specialized skin appendages that allow for extensive postnatal care of the progeny by providing a primary source of nutrition and immune protection (25, 26). Their early development depends on complex epithelial–mesenchyme interactions (27, 28), starting with the formation of the mammary lines, a thickening of the ectoderm on the lateral sides of the embryo stretching between the forelimbs and hind limbs. At embryonic day

Significance

During vertebrate evolution, *Hox* gene function was coopted through the emergence of global enhancers outside the *Hox* gene clusters. Here, we analyze the regulatory modalities underlying *Hoxd* gene transcription into the developing mammary glands where *Hox* proteins are necessary. We report the existence of a long-distance acting mammary bud enhancer located near sequences involved in controlling *Hox* genes in the limbs. We argue that the particular constitutive chromatin structure found at this locus facilitated the emergence of this enhancer element in mammals by hijacking a regulatory context at work in other cell types, supporting a model wherein enhancer sequences tend to cluster into large regulatory landscapes due to an increased probability to evolve within a preexisting regulatory structure.

Author contributions: R.S., J.Z., D.D., and L.B. designed research; R.S., E.R.-C., I.G., G.A., T.H.N.H., V.M., J.Z., and L.B. performed research; R.S., A.N., E.R.-C., I.G., G.A., T.H.N.H., V.M., J.Z., D.D., and L.B. analyzed data; and R.S., A.N., D.D., and L.B. wrote the paper.

Reviewers: M.A.N., The University of Chicago; and G.P.W., Yale University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The RNA sequencing data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE84943).

[1]Present address: Max Planck Institute for Molecular Genetics, Development and Disease Research Group, 14195 Berlin, Germany.

[2]To whom correspondence should be addressed. Email: denis.duboule@unige.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617141113/-/DCSupplemental.

(E) 11.5, the murine milk lines split into five symmetrical pairs of mammary placodes (MPs). Each placode is associated with a specialized underlying mammary mesenchyme, which drives mammary ectoderm invagination and formation of the duct system. By E12.5, the placodes invaginate, and around E13.5, they form the mammary buds (MBs), which will then elongate and sprout to create a ductular structure deeper in the dermal mesenchyme. The mammary mesenchyme remains associated with the surface ectoderm, driving the formation of the mammary papilla or nipple, which plays a critical role in lactation (29). Nipples are found in eutherian mammals and marsupials but not in monotremes. The molecular bases of these complex intertissue interactions have recently started to be investigated in some detail (30).

The function of *Hox* genes during the development of skin-derived appendages has been documented in several instances, including hair follicles and feathers (31–35). Likewise, these genes have been implicated in MG development (36–38), with *Hoxc8* being one of the first markers of MP specification, where it contributes to define the number and positions where these structures will form (37). The expression of *Hoxc8* is nevertheless completely abrogated by E12.5. In contrast, *Hox* gene members of the group 9 are transcribed during both embryonic and postnatal development of the MGs, and their combined loss of functions resulted in severe MG hypoplasia and the death of offspring due to milk starvation (36).

We analyzed the transcriptomes of microdissected MBs and observed that *Hoxd8* and *Hoxd9* are the only members of the *HoxD* cluster expressed during MB development. By using a panel of mutant alleles, we show that expression of these two genes in the MB depends on two different and largely independent regulatory mechanisms. Although DNA sequences located near *Hoxd8* are required to drive its expression in the MBs, *Hoxd9* transcription seem to depend mainly on a regulatory input located 450 kb downstream of the cluster, within the T-DOM. We further isolated a 200-bp long DNA segment, which displayed enhancer activity in the MBs. This DNA sequence is conserved among therian mammals and is located near a previously described limb enhancer. Finally, the regulatory potential of this sequence, when isolated from either the platypus or the opossum, was tested to address the evolutionary origin of this element across the mammalian lineage.

Results

Characterization of MB-Enriched Transcripts by RNA-Sequencing. We complemented previous transcriptome studies of the adult MG (39–42) by a transcriptional analysis of microdissected E13.5 embryonic MBs. MB pairs 2 and 3 were thus obtained, and their expression profiles were established by RNA-sequencing (RNA-seq). To distinguish MB-specific expression from the background transcription found in the presumptive dermis and epidermis, we dissected from the same animals an equivalent area of tissue immediately adjacent to the MB (hereafter referred to as "skin"), including the nonmammary epithelium and its underlying mesenchyme (Fig. 1*A*). We identified 409 genes specifically up-regulated (*SI Materials and Methods*) and 204 down-regulated genes in the MBs compared with control skin (false discovery rate < 10%, minimum expression fold change = 1.5; Fig. S1 *A–C* and Datasets S1 and S2). Up-regulated genes included the mammary epithelium markers *Krt8*, *Bmp2*, *Fgf17*, *Gata3*, *Pthlh*, and *Tbx3* (30, 43–47), as well as the mammary mesenchyme-specific genes *Tbx18*, *Esr1*, and *Meox2* (30), thus validating our experimental approach (Fig. 1*B* and Datasets S1 and S2). Among the differentially expressed genes, 28 long noncoding RNAs (lncRNAs) were up-regulated and four were down-regulated in the E13.5 MBs, compared with embryonic skin (Fig. S1 *A* and *B* and Datasets S1 and S2).

To classify the MB-enriched genes functionally, we performed gene ontology (GO) term enrichment analysis (*SI Materials and Methods*) using Gorilla (48) ($P < 10^{-4}$). Enriched GO terms were further clustered using Revigo (49) (similarity threshold = 0.7). Among the 20 most enriched GO term categories, we found several terms related to epithelial structures and MG development, as well as BMP signaling, previously described to be essential for MG development (e.g., refs. 43, 50) (Fig. 1*C* and Dataset S3).

To evaluate spatial expression patterns within the MB further, we analyzed the overlap between our list of MB up-regulated genes and an existing estimate of both ectoderm- and mesenchyme-specific genes derived from a microarray-based analysis of the posterior MB at E12.5 (30). We found that of our 409 identified MB up-regulated genes, 65 had been reported as ectoderm-specific in this previous work, whereas 19 of them were found to be enriched in mammary mesenchyme (Fig. 1*D*). However, the majority of the MB up-regulated genes in our dataset were expressed at similar levels in the E12.5 mammary mesenchyme and ectoderm (Fig. S1*D*). We think that differences in the embryonic stage and tissue [E12.5 posterior MB analyzed by Wansbury et al. (30) versus E13.5 anterior MB analyzed here], as well as the increased sensitivity of RNA-seq compared with microarrays (51), may account for the moderate overlap between the two datasets. In support of the latter, most of our MB-enriched genes were expressed at moderate to low levels in E12.5 mammary primordia (Fig. S1*E*), and were thus likely not detectable as being differentially expressed with microarrays. The distribution of these genes was nevertheless clearly shifted compared with the distribution of genes classified as down-regulated in our analysis, which tended to be expressed at lower levels in the E12.5 mammary primordium (Fig. S1*E*).

The Embryonic MG *Hox*ome. Due to their potential role in MG development (discussed above), we particularly looked at the *Hox* gene family (Fig. S2*A*) and noticed that members of the *HoxB* and *HoxD* clusters were expressed at high levels in the MB compared with the nearby skin control, whereas *HoxA* and *HoxC* paralogs were either not differentially expressed or even down-regulated in these structures (Fig. 2*A* and Fig. S2*A*). Among *HoxB* and *HoxD* genes, *Hoxb3*, *Hoxb6*, *Hoxb9*, and *Hoxd9* were significantly up-regulated in the MBs compared with the skin tissue (Fig. 2*A* and Fig. S2*A*). *Hoxd3* was also up-regulated in the MBs, although to a lower level than the former genes. Although the absolute mRNA levels of *Hoxd10* were also significantly up-regulated in the MB, the levels of *Hoxd10* remained very low (Fig. 2*A* and Fig. S1*F*). We performed whole-mount in situ hybridization (WISH) at different stages of MB development (Fig. 2*B* and Fig. S2*B*) and detected strong *Hoxd9* expression in all five MB pairs at E12.5 and E13.5, whereas neither *Hoxd10* nor *Hoxd11* transcripts were scored. *Hoxd8* expression was also observed, but only in the first to third pairs of MBs at E12.5. Its expression was strongly down-regulated to become virtually nondetectable 1 day later (Fig. 2*B*).

In support of our RNA-seq data, WISH also revealed expression of *Hoxb6* and *Hoxb9* in all MBs, whereas *Hoxa9* transcripts were only scored in the fourth and fifth MB pairs (Fig. S2*B*). We confirmed that *Hoxc9* was not transcribed in the MBs at any of the developmental times analyzed (36), and we could not detect any *Hoxd3* transcripts by WISH, despite a slight, yet significant increase, compared with skin (Fig. S2*A*), likely due to the different sensitivities of the two approaches. To determine the precise localization of the *Hoxd8* and *Hoxd9* transcripts, we cryostat sectioned WISHed embryos, and both *Hoxd8* and *Hoxd9* were detected in the mammary mesenchyme below the mammary epithelium (Fig. 2*C*). Although *Hoxd9* was detected at a high level in all mammary mesenchymal cells, *Hoxd8* transcripts were found in the most lateral portion of this mesenchyme and were barely detectable in its most central part at E12.5. This expression pattern likely paralleled the rapid down-regulation of *Hoxd8* transcription occurring in the MB between E12.5 and

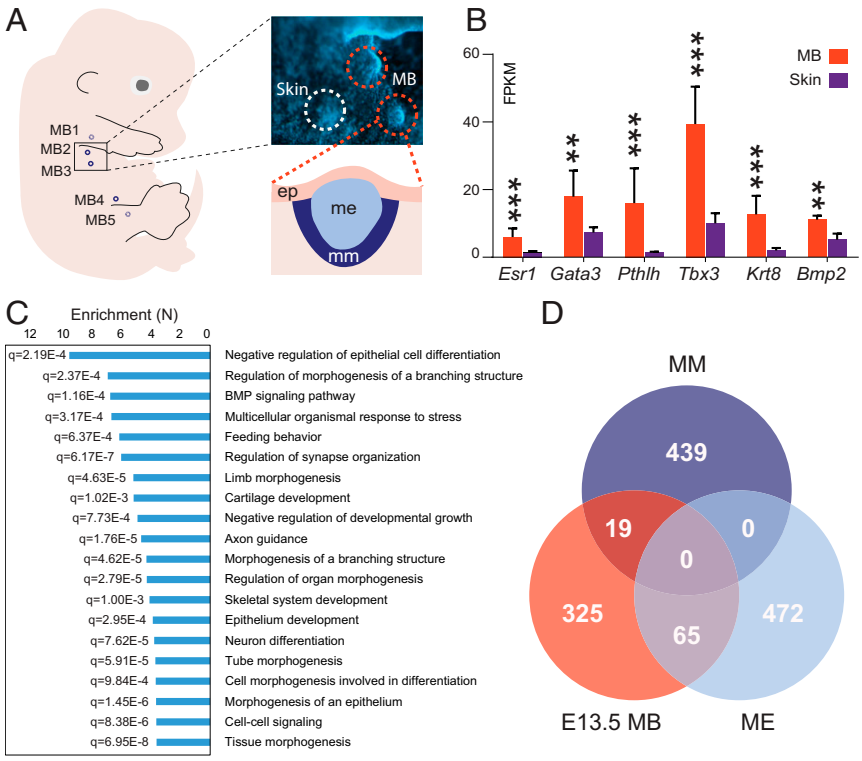

Fig. 1. Transcriptome analysis of MG. (*A*) Schematic view of an E13.5 embryo showing the location of the different MB pairs (MB1–MB5) along the anteroposterior axis (*Left*) and zoomed-in lateral view of a whole-mount DAPI-stained embryo (*Right*) showing the MB2 and MB3 dissected in this study (red dashed circles). A portion of the adjacent skin tissue (including nonmammary ectoderm and mesenchyme) was used as a control (white dashed circle) (Magnification: 30×.). A schematic view of a transversal section of an E13.5 MB is depicted below. ep, epithelium; me, mammary ectoderm; mm, mammary mesenchyme. (*B*) Bar graph plot showing the expression levels of mammary ectodermal (*Gata3*, *Tbx3*, *Pthlh*, *Bmp2*) and mesodermal (*Esr1*, *Lhx9*, *Tbx18*, *Ptx2*) markers measured in fragments per kilobase of exon per million mapped reads (FPKM). $^{**}P < 0.001$, $^{***}P < 0.0001$. (*C*) Bar graph displaying the 20 most enriched GO term categories and their respective false discovery rate (q) values. GO terms significantly enriched among MG up-regulated genes were identified using Gorilla ($P < 10^{-4}$). To reduce the number of redundant GO term categories, we used the Revigo algorithm (similarity threshold = 0.7). (*D*) Venn diagram showing the overlap of genes specifically expressed in the E13.5 MBs with mammary ectodermal (ME)-specific or mesenchymal (MM)-specific genes identified by Wansbury et al. (30). Among the 409 E13.5 MB-specific genes, 19 corresponded to mesenchyme-specific genes, 65 were specifically enriched in the mammary ectoderm, and 19 were up-regulated in the mammary mesenchyme.

E13.5. This dataset, along with previous contributions (36–38), indicates that *Hox* genes are differentially expressed both among the various pairs of MBs and during the development of these structures, suggesting that temporal and spatial *Hox* combinations may contribute to MG development and specification.

***Hoxd8* and *Hoxd9* Transcription Depends on Distinct Regulations.** To search and identify regulatory elements controlling *Hoxd8* and *Hoxd9* expression in MBs, we initially analyzed a BAC transgenic line containing the entire *HoxD* cluster (TgBAC^{HoxD}; Fig. 3*A*). To discriminate transgenic *Hox* genes from their endogenous counterparts, TgBAC^{HoxD} mice were crossed with the *HoxD*$^{Del(1–13)d11Lac}$ line [also known as *Del(1–13)d11lac*] where all *Hoxd* genes are deleted (Fig. 3*A*). Backcrosses generated *Del(1–13)d11lac*$^{-/-}$*;* TgBAC^{HoxD} embryos where only transgenic *Hoxd* genes were present. Neither *Hoxd8* nor *Hoxd9* was expressed in the MB in these embryos (Fig. 3*B*), suggesting that DNA elements located outside the region covered by the BAC^{HoxD} were required.

We further analyzed a series of targeted deletions within the *HoxD* cluster spanning different intervals around the *Hoxd8*-to-*Hoxd9* locus (Fig. 3*C*) and obtained contrasting results. Although *Hoxd9* expression was not affected in any of these deletions (Fig. 3*D*, *Upper*), *Hoxd8* transcripts were strongly down-regulated in all homozygous mutant embryos where the *Hoxd4*-to-*Hoxd8* intergenic region was absent [*HoxD*$^{Del(1–4i)}$ or *HoxD*$^{Del(i)}$; Fig. 3*D*, *Bottom*]. *Hoxd8* transcription was scored in other deletions maintaining this DNA region, however, such as the *HoxD*$^{Del(1–4)}$ or *HoxD*$^{Del(10–13)}$. Although phylogenetic foot-printing analysis of the *Hoxd4-Hoxd8* intergenic region revealed the presence of sequences conserved in mammals (Fig. S3*A*), a transgenic construct carrying this region upstream of a *LacZ* reporter cassette did not display any β-gal activity in the E12.5 MBs (Fig. S3*B*), thus corroborating the results obtained with the BAC^{HoxD} transgenic fetuses. Altogether, these results indicated that although enhancers outside of the *HoxD* cluster drive *Hoxd9* expression in the MBs, *Hoxd8* transcription depends on the combined activity of regulatory elements located in its immediate 3′ vicinity and in the flanking gene deserts.

The selective expression of *Hoxd8* and *Hoxd9* in the MBs was not due to the incapacity of other *Hoxd* genes to be transcribed there, because mutant lines carrying deletions encompassing the *Hoxd8* and/or *Hoxd9* locus ectopically transcribed *Hoxd10*, *Hoxd11*, or *Hoxd12* (Fig. S4). This ectopic expression was at least partially driven by regulatory elements located outside the cluster because it was also scored in embryos homozygous for the *HoxD*$^{Del(1–10)}$ deletion, which removes the putative local *Hoxd8* enhancer(s) [Fig. S4*C*; *Del*$^{(1–10)}$]. In all of these cases of ectopic expression, the concerned *Hoxd* genes were positioned closer to the T-DOM as a result of the deletion, suggesting that the T-DOM may contain embryonic MB enhancers specifically interacting with this central part of the gene cluster (Fig. S4*E*). Ectopic expression of *Hoxd13* was nevertheless never scored in any of the mutant lines analyzed in this study, suggesting that not all promoters are responsive to the MB enhancers (Fig. S4).

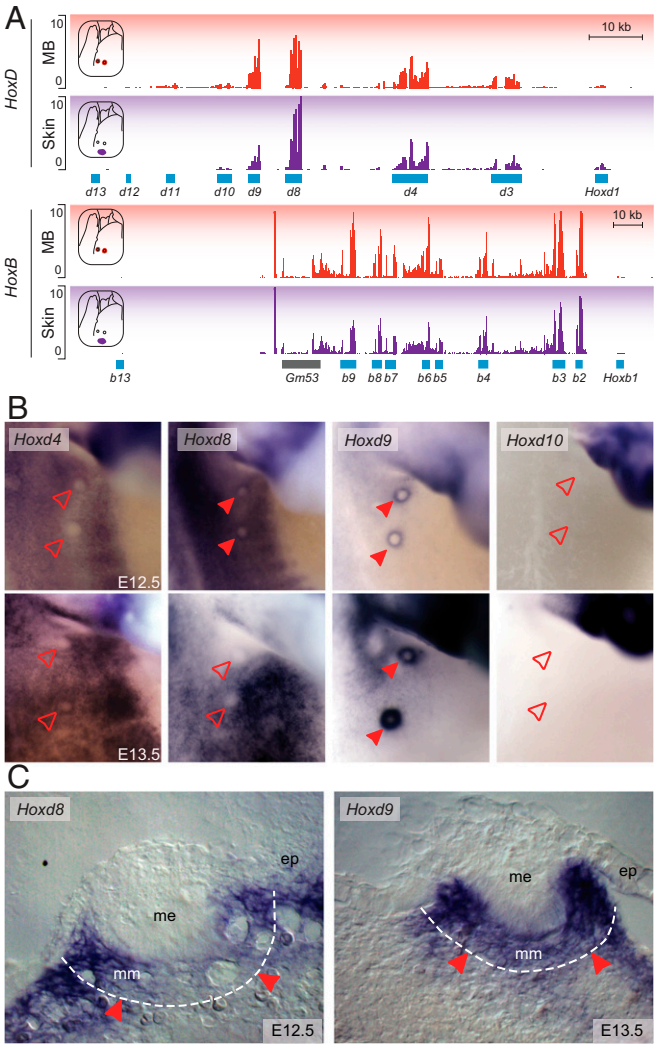

Fig. 2. *Hoxd8* and *Hoxd9* are expressed during MB development. (*A*) RNA-seq profiles of E13.5 MG and neighboring skin showing the expression of either *Hoxd* (*Top*) or *Hoxb* (*Bottom*) genes. (*B*) WISH showing *Hoxd* expression in MB2 to MB3. Filled red arrowheads represent those MBs where *Hoxd* gene expression was detected. Empty red arrowheads represent MBs lacking *Hoxd* expression. (Magnification: 23×.) (*C*) Coronal cross-section of MB2 showing expression of *Hoxd8* (E12.5) and *Hoxd9* (E13.5) in the MBs. The dashed line delimits the mammary mesenchyme from the surrounding nonmammary mesoderm. Red arrowheads point to the expression of *Hoxd8* and *Hoxd9* in the mammary mesenchyme. ep, nonmammary epithelium; me, mammary epithelium. (Magnification: 125×.)

An MB Enhancer Region Is Located in the T-DOM. To verify that distal MB enhancer(s) were located in the T-DOM, we tested two large inversions that separate the *HoxD* cluster from either one or the other gene desert (Fig. 4*A*). In the $HoxD^{Inv(Itga6\text{-}TgHd11LacNsi)}$ inversion, a 3-Mb region containing the entire C-DOM was inverted, repositioning potential enhancers far away from the gene cluster [Fig. 4*A*; *Inv(Itga6-Nsi)*]. On the other hand, the $HoxD^{Inv(Itga6\text{-}Attp)}$ inversion disconnected the *HoxD* cluster from the T-DOM, still keeping its relative distance with potential C-DOM enhancers [Fig. 4*A*; *Inv(Attp-Itga6)* and *SI Materials and Methods*]. Although *Hoxd9* expression in the MBs remained unaffected in homozygous *Inv(Itga6-Nsi)* embryos, it was fully abrogated in the MBs of *Inv(Attp-Itga6)* homozygous mice (Fig. 4*B*, *Upper*). To rule out a potential negative effect of the novel neighboring DNA sequence associated with the *HoxD* cluster following the latter inversion, we also analyzed a large deletion removing almost the entire T-DOM [Fig. S5*A*; *Del(Attp-SB3)*]. Embryonic lethality was prevented by balancing this allele with a deletion of the *HoxD* cluster [*Del(1–13)d11Lac*; *SI Materials and Methods*], and *Hoxd8* and *Hoxd9* expression in the mammary mesenchyme was expectedly lost in $Del(Attp\text{-}TpSB3)^{+/-}/Del(1–13)d11Lac^{+/-}$ embryos, compared with the $Del(1–13)d11Lac^{+/-}$ control littermates (Fig. S5 *B* and *C*).

To locate the MB regulatory region more precisely within the T-DOM, we used targeted deletions covering this area, including the $HoxD^{Del(Attp\text{-}TpSB2)}$, $HoxD^{Del(TpSB2\text{-}TpSB3)}$, $HoxD^{Del(65\text{-}TpSB3)}$, and $HoxD^{Del(TpSB2\text{-}65)}$ alleles (Fig. 4*A* and Fig. S5*A*), which were all analyzed over the *Del(1–13)d11Lac* balancing allele. The expression of *Hoxd8* and *Hoxd9* in the MB was not significantly affected in either *Del(Attp-SB2)* or *Del(CS65-SB3)* embryos (Fig. 4*B* and Fig. S5*B*). In contrast, expression was no longer scored in *Del(SB2-SB3)* or *Del(SB2-CS65)* embryos (Fig. 4*B*, *Bottom*). It is noteworthy that this region was previously shown to contain

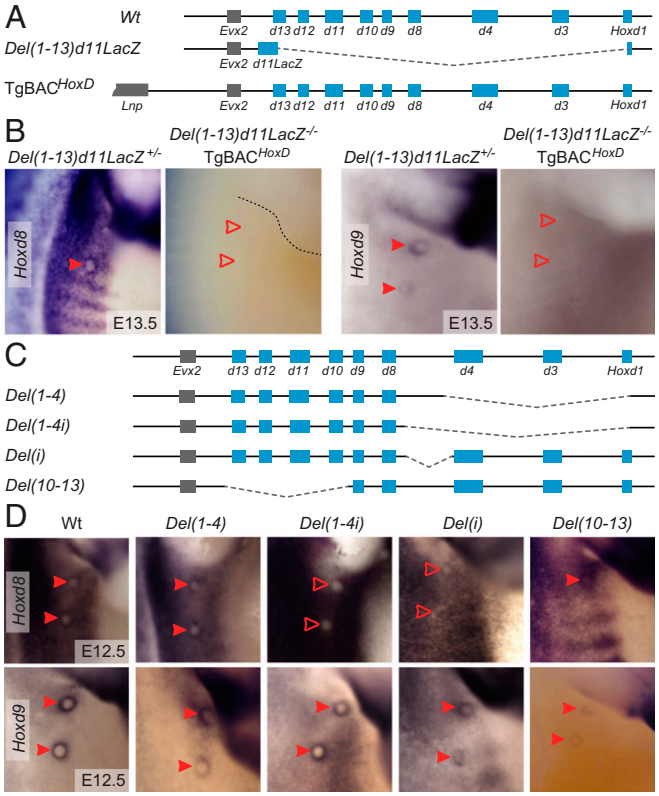

Fig. 3. *Hoxd8* and *Hoxd9* expression in embryonic MBs depends on different regulatory mechanisms. (*A* and *B*) WISH analysis showing expression of *Hoxd8* and *Hoxd9* in a mouse carrying a transgenic BAC coding for the *HoxD* cluster randomly integrated into the genome. The BAC insertion was analyzed in $Del(1–13)d11LacZ^{-/-}$ animals, which lack the endogenous *HoxD* cluster. $Del(1–13)d11LacZ^{+/-}$ animals were used as controls. The different alleles used in this study are schematized in *A*. *Hoxd* genes are represented by light blue boxes, whereas gray rectangles represent other, non–*Hox*-coding genes. The deleted region is depicted with dashed lines. (*C* and *D*) Mice carrying deletions of different portions of the *HoxD* cluster were analyzed to identify potential regulatory elements controlling *Hoxd8* and/or *Hoxd9* expression in the MBs. (*C*) Different *Hoxd* genes and *Evx2* are represented with light blue and gray boxes, respectively. Dashed lines represent the deleted DNA segments. (*D*) WISH analysis showing *Hoxd8* and *Hoxd9* expression in E12.5 MBs in the homozygous embryos of the different lines shown in *C*, compared with wild-type (Wt) embryos. Expression of *Hoxd9* remained unchanged in all of the deletions analyzed. *Hoxd8* expression in the MB, however, was lost in all lines carrying the deletion of region *i*. In *B* and *D*, filled red arrowheads represent MBs where expression of *Hoxd8* and/or *Hoxd9* was detected. Empty red arrowheads represent MBs lacking expression of these genes. (Magnification: *B*, 20×; *D*, 18×.)

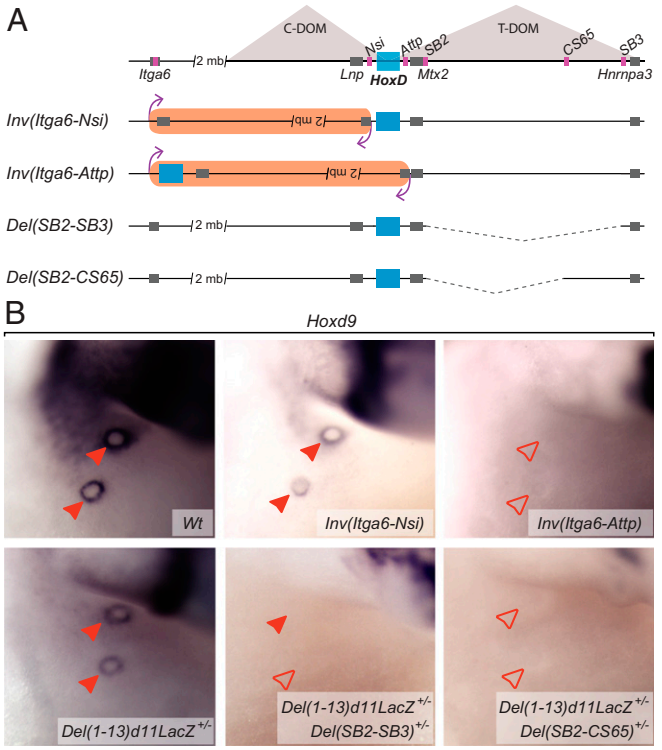

Fig. 4. *HoxD* telomeric gene desert is required for *Hoxd9* expression in the MBs. (*A*) Schemes representing the alleles used in this study and carrying different inversions or deletions spanning the *HoxD* centromeric or telomeric deserts. The *HoxD* cluster is represented by a light blue rectangle. The T-DOM and C-DOM are represented by light gray triangles. Gray boxes illustrate other, non-*Hox* genes. Breakpoints for the different inversions/deletions are shown with pink boxes. The inverted DNA regions are highlighted in pale orange, whereas dashed lines represent the deleted DNA segments. (*B*) WISH analysis showing *Hoxd9* expression in E13.5 MBs in the different alleles depicted in *A*, compared with control embryos. Embryos homozygous for the *Inv(Itga6-Attp)* and *Inv(Itga6-Nsi)* alleles were compared with *Wt* littermates. Deletions spanning the *HoxD* telomeric gene desert were crossed with mice carrying the balancer mutation *Del(1–13)d11lacZ*, as in Fig. 3, to avoid the embryonic lethality observed in homozygous embryos. In this case, $Del(1–13)d11LacZ^{+/-}$ littermates were used as controls. Filled red arrowheads represent MBs where *Hoxd9* expression was detected. Empty red arrowheads represent MBs lacking *Hoxd9* expression. (Magnification: 38×.)

regulatory elements controlling *Hoxd* gene expression both in the proximal limb buds and the intestinal cecum (17, 18). This DNA segment also contains the transcriptional start sites of *Hog* and *Tog*, two lncRNAs coregulated with selected *Hoxd* genes in the developing cecum (18), and whose transcription was also slightly enriched in E13.5 MBs compared with the surrounding control tissue (Fig. 5*B*).

This particular DNA region maps at the boundary between two sub-TADs within the T-DOM (17) (Fig. 5*A*), and chromosome conformation capture approaches have revealed its strong contacts with the *HoxD* gene cluster in either transcriptionally active or inactive cellular contexts (17, 18, 23, 52). Of note, the interactions observed between this DNA region and *Hoxd9* are usually stronger than with the other *Hoxd* genes (Fig. 5*C* and Fig. S6 *A* and *B*). This privileged interaction involving *Hoxd9* was not scored with any other sequence within the T-DOM, except for the other limb-specific enhancer CS65 (17), ruling out a technical bias associated with the *Hoxd9* probes used for chromosome conformation capture (4C) studies. The same preferential interaction was observed when *Hoxd9* and *Hoxd11* 4C profiles were compared from either E10.5 brain cells (where *Hoxd* genes are not expressed) or anterior and posterior trunk cells (expressing different combinations of *Hoxd* genes) (53) (Fig. S6 *B* and *C*), supporting the idea that *Hoxd9* specifically contacts this region, regardless of its transcriptional activity.

A high density of DNA sequences conserved among mammals were found in the 60-kb surrounding this *Hoxd9* interacting region, with six of them specifically conserved in all eutherian mammals and marsupials but not in monotremes (Fig. 5*D*, blue arrows). A 20-kb-large deletion was engineered, which removed four of the six eutherian-specific elements, as well as the previously reported conserved sequences CS38 to CS40 (17). Embryos homozygous for this deletion [$HoxD^{Del(CS38–CS40)}$, also known as *Del(CS38–CS40)*] were viable and displayed no major alteration in *Hoxd* gene expression in either the proximal limb or the cecum at E13.5. Likewise, the expression of *Hoxd8* in the E12.5 MBs remained unchanged (Fig. S5*D*). In contrast, *Hoxd9* transcripts were no longer observed in these structures (Fig. 5*E*), confirming that expression of *Hoxd8* and *Hoxd9* in the MBs depends on different regulatory mechanisms. Despite the loss of *Hoxd9* expression in the MB of $Del(CS38–CS40)^{-/-}$ embryos, no major alterations were observed either in the milking behavior or in the survival of the offspring of knockout females, in agreement with the apparent lack of phenotype observed in the $Hoxd9^{-/-}$ mice (36).

To assess whether additional regulatory elements controlling *Hoxd8*/*Hoxd9* in the developing MB were located outside of this deleted region, we tested two stable transgenic lines carrying random insertions of BAC clones flanking the CS38–CS40 region (Fig. 5*B*; BAC^{T1} and BAC^{T2}) carrying an integration of a *LacZ* reporter under the control of the β-globin minimal promoter. Neither of the lines displayed any β-gal activity in the MB at any stage analyzed (Fig. 5*E*). Therefore, although the CS38–CS40 region is required for *Hoxd9* transcription, the control of *Hoxd8* transcription depends on the combined activities of local and distal regulatory elements.

A Eutherian-Specific MB Element. This 20-kb-large DNA region necessary to drive *Hoxd9* expression in the developing MB interacts strongly with the *HoxD* cluster in all tissues analyzed thus far (17, 18, 53). A close examination of 4C interaction profiles revealed that the strongest contacts coincide with the CS38 and CS39 conserved elements (Fig. S6*D*). The latter had been previously described as an enhancer driving *Hoxd* expression in developing limb buds (17, 52). Immediately 5′ of CS39, we identified a *ca.* 200-bp sequence conserved among eutherian mammals. A stable transgenic line carrying this eutherian-specific element, the CS39 sequence, and a *LacZ* reporter cassette revealed β-gal activity in the developing MBs at E12.5 and E13.5, but not in the E11.5 MPs (Fig. 6*A*), thus matching the expression of endogenous *Hoxd9*. Instead, the CS38 region did not display any enhancer activity when tested in a classical transgenic assay.

Phylogenetic footprinting of the sequence contained in the TgNCS39 transgenic construct (Fig. 6*B*) revealed four regions of high conservation (Peak1–Peak4). Whereas Peak1 is specific to therian mammals, Peak2–Peak4 matched the vertebrate conserved CS39 region (17). We generated transgenic constructs carrying different combinations of these conserved regions (Fig. 6*C*) to define which sequence displayed the MB enhancer capacity. All constructs carrying the Peak1 sequence activated *LacZ* transcription in the MBs (Fig. 6 *D* and *E*), in agreement with the conservation of Peak1 in therian mammals. In contrast, the TgPeak3-Peak4/*LacZ* reporter displayed β-gal staining in the limb buds but not in the E12.5 MB (Fig. 6 *D* and *E*). We next generated lentiviral reporter constructs carrying either Peak1 or Peak2 individually, and only the TgPeak1/*LacZ* construct was able to drive *LacZ* reporter expression in the developing MBs, in addition to a high background activity in the rest of the embryo (Fig. 6 *D* and *E*). Both constructs were consistently active in the limb bud. However, the TgPeak1/*LacZ* construct showed significantly less expression intensity and robustness than observed

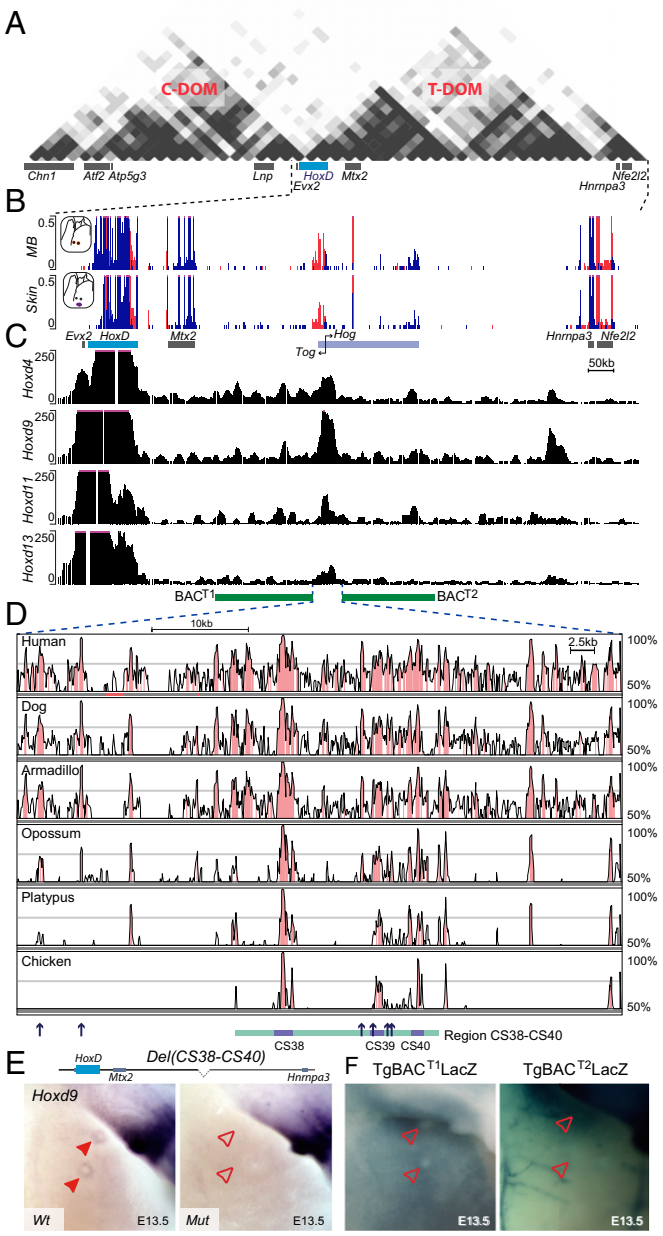

Fig. 5. A 20-kb-large region of the *HoxD* telomeric gene desert specifically interacts with *Hoxd9* and is required for its expression in MBs. (*A*) Hi-C data adapted from Dixon et al. (23) showing the C-DOM and T-DOM flanking the *HoxD* cluster. The light blue and gray boxes, respectively, represent the *HoxD* cluster and the other gene loci contained within the region analyzed. (*B*) RNA-seq analysis comparing the mouse E13.5 MBs with the adjacent epidermal tissue (skin). The *HoxD* cluster is depicted by a blue rectangle, whereas other, non–*Hox*-coding genes are represented by gray boxes. Light blue rectangles point to the *Hog* and *Tog* lncRNAs, and black arrows represent the transcription start site of these two genes. (*C*) Chromosome conformation capture analysis of different *Hox* genes in E11.5 mouse embryos aligned with the RNA-seq from *B*. All 4C samples were normalized for the respective number of mapped reads for each sample. *Hoxd9* displays increased interactions with a short region, including the CS38-CS39 regions and containing the transcription start site of the *Hog* and *Tog* lncRNAs. Data are from Guerreiro et al. (75). Green rectangles represent the region of the *HoxD* telomeric gene desert contained within the BAC^{T1} and BAC^{T2} transgenic lines tested in this study. (*D*) Vista alignment showing the noncoding conserved elements across different vertebrate species within a 60-kb window around the CS38–CS40 region. Regions of conservation (defined by at least 70% sequence identity over 100 bp) are shown in pink. Several eutherian-specific sequences shown by blue arrows at the bottom were identified in this area. The turquoise rectangle represents the region spanned by the *Del(CS38–CS40)*, which removes most of these eutherian-specific sequences. (*E*) WISH analysis comparing *Hoxd9* expression in E13.5 MBs from either *Wt* or embryos homozygous for the *Del(CS38–CS40)* deletion (*Mut*). A schematic representation of the *Del(CS38–CS40)* allele is shown above the panels. The dashed line represents the deleted region. (*F*) *LacZ* staining of embryos carrying random integrations of BACs coding for the *HoxD* telomeric regions on both sides of the CS38–CS40 elements. In *E* and *F*, filled red arrowheads represent MBs where *Hoxd9/LacZ* expression was detected. Empty red arrowheads represent MBs lacking expression of these genes. (Magnification: 18×.)

with the TgPeak1-Peak2/*LacZ* reporter. In addition, β-gal activity was detected in the vibrissae and in the incipient hair follicles, as well as in the presumptive skin dermis, although at low levels. These results suggest that, albeit Peak1 is the minimal region required for MB enhancer activity, other sequences might contribute to drive robust and specific transcription into these structures.

The Peak1 sequence [hereafter referred to as mammary bud regulatory element (MBRE)] was not detected in the monotreme platypus or in other nonmammalian vertebrates (Fig. 6*B*). To evaluate the regulatory potential of this region across eutherian and noneutherian mammals, we cloned a DNA segment orthologous to the murine Peak1-Peak2 sequence from both the platypus and opossum, and tested it in lentiviral transgenic assays. In agreement with its lack of conservation, the platypus transgene failed to drive *LacZ* activity in the developing MBs (one of nine), although it had similar expression in the developing limb (five of nine) compared with the murine sequence. The opossum sequence also displayed reproducible *LacZ* staining in the proximal limb (five of seven) but showed only sporadic expression in the developing MB (two of seven). Nevertheless, it could control expression in the hair follicle placodes, thus partially mimicking the reporter activity of the murine TgPeak1/*LacZ* construct.

Discussion

***Hox* Genes in the MGs.** Our comparative transcriptome analysis identified 409 genes enriched in the E13.5 MBs, among which were several *Hox* genes. The mammary mesenchyme is instrumental in the development of the MGs and can induce ectopic expression of mammary ectoderm markers in the nonmammary skin epithelium of different tetrapod species (54, 55). Also, the mesenchyme underlying the ectodermal precursor of different skin appendages determines the type of structure to be formed (56, 57). Therefore, the molecular identities of these different mesenchymes must be distinctly identified. In a study by Wansbury et al. (30), the authors dissected E12.5 posterior MBs and separated the ectoderm from the underlying mammary mesenchyme, making it difficult to identify which genes are specifically expressed in this structure. Indeed, an important number of the genes identified in that study are also expressed in the adjacent epithelium and mesenchyme (30). Our data, thus, represent a substantial complement in the description of the gene network operating during MB development. We confirmed the strong expression of the *Hoxa9*, *Hoxb9*, and *Hoxd9* paralogous genes (36) in the MBs and report the detectable transcription of additional *Hox* genes, such as *Hoxd8*, *Hoxb3*, and *Hoxb6*.

Hoxd8 transcription is rapidly down-regulated, however, whereas other *Hox* gene mRNAs remain at high steady-state levels. Also, *Hoxa9* and *Hoxd8* expression is restricted to a subset of MBs, with the latter expressed only in MB1 to MB3 and the former in the posteriorly located MB4 and MB5. These observations, along with the reported role of *Hoxc8* in the early formation of the MPs from the milk line (36), suggest that spatial and temporal *Hox* combinations contribute to the patterning and development of the different pairs of MGs (Fig. 7*A*). Hairs and feathers are other skin-derived appendages whose early development shares similarities

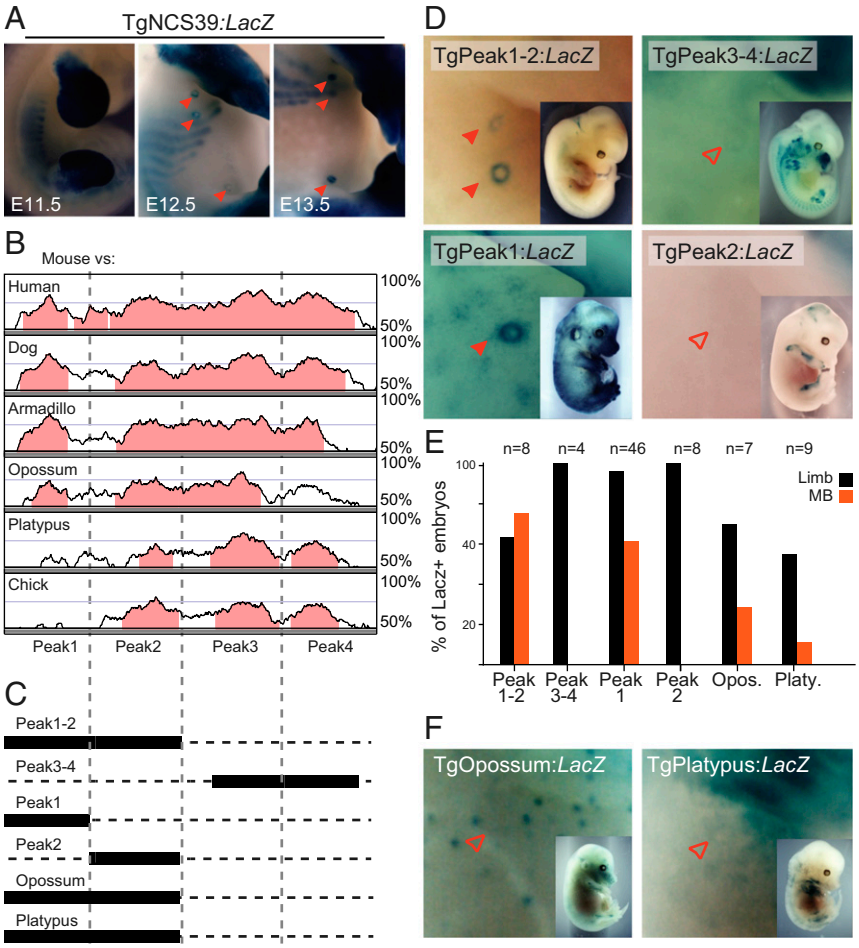

Fig. 6. Eutherian-specific sequence displays MB enhancer activity. (*A*) *LacZ* staining of mouse transgenic embryos carrying a 1.2-kb construct, including the CS39 regulatory element and its 5′ adjacent therian-specific conserved element. Robust expression of the *LacZ* reporter is scored in the developing MBs (red arrowheads). (*B*) Vista alignment of the sequence encoded by the TgNCS39 transgene, showing four peaks of conservation (defined by at least 70% sequence identity over 100 bp). The most 5′-located sequence (Peak1) is conserved in therian mammals, yet not in the monotreme platypus. (*C*) Scheme showing the different transgenic constructs used in this study, including either different portions of the TgNCS39 transgene or the opossum and platypus regions paralogous to the Peak1-Peak2 sequence (black rectangles). (*D* and *F*) *LacZ* staining of E12.5 mouse embryos transgenic for the different lentivirus constructs described in *C*. Filled red arrowheads represent MBs where *LacZ* expression was detected. Empty red arrowheads represent MBs lacking expression of the reporter. (Magnification: 18×.) (*E*) Bar graphs representing the percentage of *LacZ*-stained embryos showing specific *LacZ* expression either in the MB or the limb for each of the analyzed lentiviral constructs over the total number of *LacZ*-stained embryos.

with MP induction and invagination (58–61). Of note, they also express different *Hox* gene combinations (32–34, 62). Furthermore, within the same skin structure, a considerable level of interspecies heterogeneity in the expression of these genes can be observed (31), suggesting that various *Hox* codes were differentially coopted during evolution for the specification of epidermal appendages. In this context of multiple *Hox* gene expression, it is not surprising that our deletion of the entire CS38–CS40 region, including the *Hoxd9* MBRE element, did not elicit a strong abnormal phenotype. Functional redundancy between group 9 HOX proteins in these and other developing structures was indeed previously reported (9, 36, 63).

Distinct Regulations in the MB. We show that the telomeric gene desert is required for the expression of both *Hoxd8* and *Hoxd9* in the embryonic MB. These two genes are nevertheless regulated by different mechanisms. *Hoxd9* transcription depends on a 20-kb-large DNA segment (region CS38–CS40) containing the eutherian conserved element MBRE, whereas sequences inside the *HoxD* cluster have little impact upon its regulation. On the other hand, the MBRE-containing region only weakly, if at all, contributes to *Hoxd8* expression in the MBs, which instead requires a 13-kb-large region located 3′ of the *Hoxd8* locus within the cluster (Fig. 7*B*). Surprisingly, although the *Del(SB2-CS65)* deficiency abrogated *Hoxd8* expression in the MBs, none of the BAC clones covering this region displayed enhancer activity in the MBs. One explanation is that regulatory elements driving *Hoxd8* expression may be located within the SB2–SB3 region, yet outside of the transgenic BAC tested in this study. Also, the concomitant activity of various *Hoxd8* enhancers located both in the 3′ vicinity of the gene and in the telomeric gene desert may be required such that neither one of these sequences alone would be sufficient. Therefore, although the precise mechanisms behind these regulations remain to be elucidated, our results reveal that *Hoxd8* and *Hoxd9* are differentially controlled in the developing MBs. This observation is intriguing, because long-range regulations involving the *HoxD* flanking gene deserts were previously reported to involve series of contiguous genes systematically, thus allowing for a coordinated function.

A Regulatory Hub. The particular topology of chromatin domains at *Hox* loci could have triggered the emergence of novel enhancers, through preexisting structural and regulatory contacts (16).

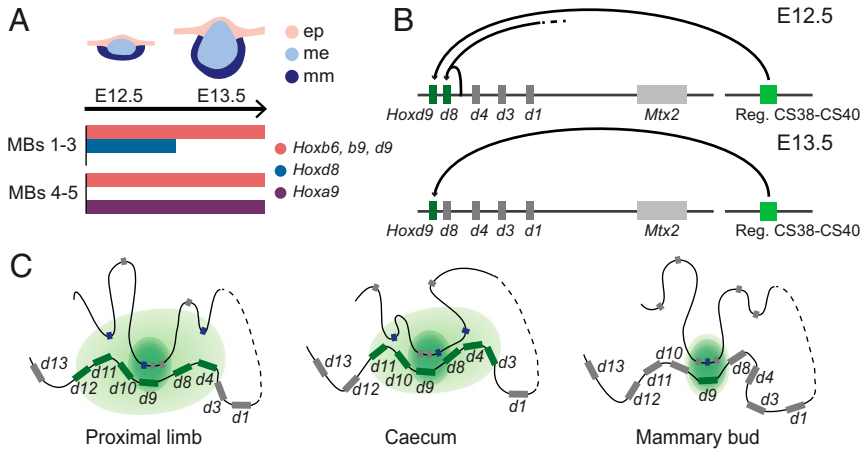

Fig. 7. Model for *Hoxd* gene regulation in the MB. (*A*) Schematic representation of *Hox* gene expression in E12.5 and E13.5 MBs. The different *Hox* genes analyzed in this study are represented with bars of different colors spanning the two developmental stages. ep, epithelium; me, mammary epithelium; mm, mammary mesenchyme. (*B*) Scheme representing the regulation of *Hoxd* gene expression in the MBs. *Hoxd* genes specifically expressed in the MBs are represented with green rectangles, whereas others are shown in gray. The regulatory (Reg.) region CS38–CS40 identified in this study is shown in light green. The activation controlled by either proximal or distal regulatory elements is represented by black arrows. The activity of the telomeric gene desert is required for the expression of both genes in these structures. However, *Hoxd9* expression depends strictly on the regulatory activity of the CS38–CS40 region containing the therian-specific enhancer MBRE, whereas *Hoxd8* transcription requires the combined activity of the *HoxD* telomeric gene desert and a local regulatory element located within the cluster, at the 3′ end of the *Hoxd8* locus. (*C*) Schematic of the CS38–CS40 regulatory hub. Strong constitutive contacts (depicted here as a dark green sphere) are fixed between the CS38–CS40 region and the central part of the *HoxD* cluster, particularly with *Hoxd9*. These contacts drag specific enhancers toward this part of the gene cluster (76–85). The number of genes responding to these enhancers, as well as the number of additional active enhancers located nearby, varies from tissue to tissue (shown by the light green cloud). As a result of this preformed chromatin architecture, novel enhancers can evolve by hijacking this potent regulatory module, as may have been the case for the MB-specific regulatory sequence. Transcriptionally active genes are represented by green boxes, whereas inactive *Hox* genes are depicted by gray rectangles.

Noteworthy, region CS38–CS40 is located at the boundary between the two sub-TADs of the T-DOM and was reported to establish robust contacts with the *HoxD* cluster in all tissues analyzed thus far, regardless of the transcriptional state of the gene cluster (17, 53, 64). These constitutive contacts may be associated with the presence of several CTCF binding sites clustered in this region. Here, we show that among *Hoxd* genes, the strongest interaction with this region, and particularly with the CS39 enhancer, is established by *Hoxd9*, suggesting that this locus may help secure a robust constitutive interaction with this region, which might be used by various regulatory regions located around the CS3–CS40 region to interact with their sets of target genes in the vicinity of *Hoxd9* (Fig. 7*C*). In this view, these constitutive contacts may serve as a regulatory hub by dragging various tissue-specific enhancers always at the same position within the *HoxD* cluster, centered on *Hoxd9*, as previously observed for the proximal limbs and the cecum (17, 18) (Fig. 7*C*). Also, this strong constitutive interaction with *Hoxd9* could have been instrumental in the emergence of MB enhancers in mammals by providing the necessary chromatin architecture, and thus optimal conditions, for a novel regulatory sequence to evolve, a mechanism proposed for the de novo appearance of enhancers located in the centromeric gene desert (19).

Evolution of an MB Enhancer. MGs likely originated from an apocrine gland already present before the divergence of the sauropsid and synapsid lineages, which further evolved in the synapsid lineage (the precursors of mammals). It is believed that the ancestral MG was a unit composed of a hair follicle, a sebaceous gland, and an apocrine gland, thus termed the apo-pilo-sebaceus unit (APSU) (65, 66). Monotremes have ∼200 repetitions of APSUs per MG, which secrete directly to the body surface; hence, they lack mammary papilla (or nipples). In contrast, marsupial and eutherian MGs release their secretion into mammary ducts that converge toward a larger duct opening into nipple-like structures (65). Therian MGs differ in their number and positions, as well as in the shape and function of the nipples, which have adapted according to the nursing behavior and number of the progenies (29, 67). It is therefore possible that changes in the transcription profiles of the mammary mesenchyme, which is required for the development of both structures, accompanied the evolution and diversification of MG morphology and functions across the mammalian lineages.

Our investigations on the evolutionary origin of the enhancer activity driving *Hoxd9* expression in the mouse embryonic mammary mesenchyme identified a sequence (MBRE) capable of directing reporter gene expression in the mesenchyme of the developing MBs. The MBRE, however, requires the presence of sequences located nearby (within Peak2 conserved throughout vertebrates) to ensure specific activity into the MBs. We were not able to identify the monotreme MBRE, and the orthologous sequence from platypus did not display enhancer activity in the mammary mesenchyme. Of note, the marsupial orthologous sequence (from opossum) did trigger *LacZ* reporter expression in the mesenchyme underlying the forming hair follicle placodes, whereas expression in the developing MBs was only sporadic.

The lack of evidence about *Hox* gene expression in the MBs of noneutherian mammals makes it difficult to interpret such changes in the regulatory activity of the MBRE in these lineages. Also, interspecies divergence within enhancer sequences does not necessarily mean that their regulatory output would be fundamentally modified (68), and enhancers associated with genes under positive selection in mammals were shown to evolve rapidly (69). As a possible scenario however, this sequence may have acquired a nonspecific enhancer activity in APSU-associated mesenchyme in the common ancestor of metatherians and eutherians, thus preluding the actual murine MB regulatory element. Alternatively, this sequence exerted its MB activity in the therian common ancestor and further diversified toward a hair follicle-specific activity in the marsupial lineage. Our observation that a transgenic construct carrying only MBRE retains its broad APSU enhancer activity, thus resembling the opossum sequence,

nevertheless suggests that this activity was an ancestral property. In this view, the MB specificity of this sequence evolved through the acquisition of responsiveness to upstream mammary mesenchyme determinants. The nature of these upstream factors is currently under study.

It was recently shown that transposable elements can acquire a regulatory potential during evolution, and thus have an impact on gene expression (e.g., refs. 70–72). On the other hand, vertebrate *Hox* clusters are largely devoid of transposable elements, unlike their flanking regulatory regions. Interestingly, the CS38–CS40 region is also relatively poor in transposable element insertions compared with the surrounding *HoxD* telomeric gene desert. This low number of transposons could be due either to the high density of regulatory elements found within this region or to the importance of this DNA segment in the establishment of a regulatory conformation at the *HoxD* locus.

Materials and Methods

Mouse Strains. Mice were handled according to the Swiss law on animal protection (LPA), with the requested authorization (GE/81/14 to D.D.). Mice were raised and killed according to good laboratory practice standards. Genetically modified mice were maintained and crossed in heterozygosis. The mouse mutant lines used in this work and the primers used to genotype them are described in Table S1. The description of the *Inv(Attp-Itga6)* and *Del(CS38–CS40)* alleles can be found in *SI Materials and Methods*.

RNA Extraction and RNA-Seq. Pairs of embryonic MBs 2 and 3 from E13.5 embryos were dissected in cold PBS and immediately processed for RNA extraction. To evaluate MB-specific gene expression, we dissected an equivalent portion of ectoderm and its underlying mesoderm tissue immediately adjacent to the MB. RNA extraction was performed using the RNAeasy Micro Kit (Qiagen) following the manufacturer's instructions. The RNA-seq libraries were prepared from 100 ng of pure total RNA using the TruSeq Stranded mRNA protocol from Illumina with polyA selection. Libraries were sequenced on a HiSeq 2500 machine as single-end, 100-bp reads. The RNA-seq datasets produced in this study are publically available in the Gene Expression Omnibus (GEO) database (accession no. GSE84943).

Cloning. The probe sequences of *Hoxa9*, *Hoxb6*, *Hoxb9*, and *Hoxc9* were amplified using Toptaq (Qiagen) and specific primers (Table S2), and cloned into pGEMT-easy vector following the manufacturer's instructions. For the cloning of lentiviral constructs, the target DNA was amplified by PCR using the Expand High Fidelity PCR system (Sigma) and specific primers (Table S3), and ligated in the PCR8/GW/TOPO (Thermo Fisher Scientific). Evolutionary conserved sequences were identified using the Vista alignment algorithm (see *SI Materials and Methods*). Coordinates used for the alignments are listed in Table S4. The opossum and platypus Peak1 and Peak2 regions were identified based on sequence conservation corresponding to the mouse orthologous CS39 Peak2. The cloned sequences correspond to the coordinates chr4:188,167,036–188,167,486 (monDom5) and Ultra514:15,095,67–15,096,077 (ornAna1) of the opossum and platypus genomes, respectively. These sequences were synthesized in vitro and inserted into the pENTR vector (GeneArt; Thermo Fisher Scientific). All of the PCR8/GW/TOPO or pENTR clones were verified by standard Sanger sequencing using the T7 primer located in the vector backbone, and the clones carrying the correct insert were recombined into the pRRLbLacGW (18) by LR reaction (Thermo Fisher Scientific). The final recombined constructs were verified by Sanger sequencing.

Lentiviral Transgenesis and β-Gal Staining. Lentiviral transgenesis was performed as in the study by Friedli et al. (73), and E12.5 embryos were dissected and stained for β-gal activity following a standard protocol.

WISH. Probes used in WISH were synthesized using T3, T7, or SP6 RNA polymerase (Table S2). They were subsequently purified using the RNA Easy Mini Kit (Qiagen). The WISH protocol was the protocol used by Woltering et al. (74).

Chromosome Conformation Capture (4C) Sequencing Dataset Analysis. The 4C-sequencing data from E11.5 whole embryo (75) were downloaded from the GEO database (accession no. GSE79048), whereas the 4C-sequencing data from E10.5 brain and anterior and posterior trunk were obtained from Noordermeer et al. (53). The pipeline used to analyze the 4C-seq data is described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Dr. B. Howard from the Breakthrough Breast Cancer Research Centre (London) for sharing information about the MB microarray data reported by Wansbury et al. (30), as well as members of the Duboule laboratories for sharing ideas and discussions. We also thank Dr. B. Mascrez, S. Gitto, and J. Couderey for help with mutant stocks. We thank the Geneva genomics platform (University of Geneva), the transgenesis unit of the Geneva Medical Centre, and the transgenic core facility at the Ecole Polytechnique Fédérale de Lausanne (EPFL) for their assistance. This work was supported by funds from the EPFL, the University of Geneva, Swiss National Research Fund Grant 310030B_138662, European Research Council System*Hox* Grant 232790 and Regul*Hox* Grant 588029, and the Claraz Foundation (D.D.).

1. Nowick K, Stubbs L (2010) Lineage-specific transcription factors and the evolution of gene regulatory networks. *Brief Funct Genomics* 9(1):65–78.
2. Glassford WJ, et al. (2015) Co-option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty. *Dev Cell* 34(5):520–531.
3. Hinman VF, Cheatle Jarvela AM (2014) Developmental gene regulatory network evolution: Insights from comparative studies in echinoderms. *Genesis* 52(3):193–207.
4. Monteiro A (2012) Gene regulatory networks reused to build novel traits: Co-option of an eye-related gene regulatory network in eye-like organs and red wing patches on insect wings is suggested by optix expression. *BioEssays* 34(3):181–186.
5. Kirschner M, Gerhart J (1998) Evolvability. *Proc Natl Acad Sci USA* 95(15):8420–8427.
6. Duboule D, Wilkins AS (1998) The evolution of 'bricolage'. *Trends Genet* 14(2):54–59.
7. Woltering JM, Duboule D (2010) The origin of digits: Expression patterns versus regulatory mechanisms. *Dev Cell* 18(4):526–532.
8. Lynch VJ, et al. (2008) Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proc Natl Acad Sci USA* 105(39): 14928–14933.
9. Mallo M, Wellik DM, Deschamps J (2010) Hox genes and regional patterning of the vertebrate body plan. *Dev Biol* 344(1):7–15.
10. Deschamps J, van Nes J (2005) Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* 132(13):2931–2942.
11. Zakany J, Duboule D (2007) The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* 17(4):359–366.
12. Pascual-Anaya J, D'Aniello S, Kuratani S, Garcia-Fernàndez J (2013) Evolution of Hox gene clusters in deuterostomes. *BMC Dev Biol* 13:26.
13. Mendivil Ramos O, Barker D, Ferrier DE (2012) Ghost loci imply Hox and ParaHox existence in the last common ancestor of animals. *Curr Biol* 22(20):1951–1956.
14. Duboule D (2007) The rise and fall of Hox gene clusters. *Development* 134(14): 2549–2560.
15. Kmita M, Duboule D (2003) Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301(5631):331–333.
16. Darbellay F, Duboule D (2016) Topological domains, metagenes, and the emergence of pleiotropic regulations at Hox loci. *Curr Top Dev Biol* 116:299–314.
17. Andrey G, et al. (2013) A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 340(6137):1234167.
18. Delpretti S, et al. (2013) Multiple enhancers regulate Hoxd genes and the Hotdog LncRNA during cecum budding. *Cell Reports* 5(1):137–150.
19. Lonfat N, Montavon T, Darbellay F, Gitto S, Duboule D (2014) Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* 346(6212): 1004–1006.
20. Montavon T, et al. (2011) A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147(5):1132–1145.
21. de Wit E, de Laat W (2012) A decade of 3C technologies: Insights into nuclear organization. *Genes Dev* 26(1):11–24.
22. Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. *Mol Cell* 49(5):773–782.
23. Dixon JR, et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398):376–380.
24. Nora EP, et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485(7398):381–385.
25. Lefèvre CM, Sharp JA, Nicholas KR (2010) Evolution of lactation: Ancient origin and extreme adaptations of the lactation system. *Annu Rev Genomics Hum Genet* 11: 219–238.
26. Vorbach C, Capecchi MR, Penninger JM (2006) Evolution of the mammary gland from the innate immune system? *BioEssays* 28(6):606–616.
27. Parmar H, Cunha GR (2004) Epithelial-stromal interactions in the mouse and human mammary gland in vivo. *Endocr Relat Cancer* 11(3):437–458.
28. Ahn Y (2015) Signaling in tooth, hair, and mammary placodes. *Curr Top Dev Biol* 111: 421–459.
29. Koyama S, Wu HJ, Easwaran T, Thopady S, Foley J (2013) The nipple: A simple intersection of mammary gland and integument, but focal point of organ function. *J Mammary Gland Biol Neoplasia* 18(2):121–131.
30. Wansbury O, et al. (2011) Transcriptome analysis of embryonic mammary cells reveals insights into mammary lineage establishment. *Breast Cancer Res* 13(4):R79.
31. Awgulewitsch A (2003) Hox in hair growth and development. *Naturwissenschaften* 90(5):193–211.

32. Kanzler B, et al. (1994) Differential expression of two different homeobox gene families during mouse tegument morphogenesis. *Int J Dev Biol* 38(4):633–640.
33. Chuong CM, et al. (1990) Gradients of homeoproteins in developing feather buds. *Development* 110(4):1021–1030.
34. Stelnicki EJ, et al. (1998) HOX homeobox genes exhibit spatial and temporal changes in expression during human skin development. *J Invest Dermatol* 110(2):110–115.
35. Godwin AR, Capecchi MR (1998) Hoxc13 mutant mice lack external hair. *Genes Dev* 12(1):11–20.
36. Chen F, Capecchi MR (1999) Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy. *Proc Natl Acad Sci USA* 96(2):541–546.
37. Carroll LS, Capecchi MR (2015) Hoxc8 initiates an ectopic mammary program by regulating Fgf10 and Tbx3 expression and Wnt/β-catenin signaling. *Development* 142(23):4056–4067.
38. Garin E, Lemieux M, Coulombe Y, Robinson GW, Jeannotte L (2006) Stromal Hoxa5 function controls the growth and differentiation of mammary alveolar epithelium. *Dev Dyn* 235(7):1858–1871.
39. Andrechek ER, Mori S, Rempel RE, Chang JT, Nevins JR (2008) Patterns of cell signaling pathway activation that characterize mammary development. *Development* 135(14): 2403–2413.
40. Blanchard A, et al. (2007) Gene expression profiling of early involuting mammary gland reveals novel genes potentially relevant to human breast cancer. *Front Biosci* 12:2221–2232.
41. Master SR, et al. (2002) Functional microarray analysis of mammary organogenesis reveals a developmental role in adaptive thermogenesis. *Mol Endocrinol* 16(6): 1185–1203.
42. Stein T, et al. (2004) Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14 and STAT3. *Breast Cancer Res* 6(2):R75–R91.
43. Phippard DJ, et al. (1996) Regulation of Msx-1, Msx-2, Bmp-2 and Bmp-4 during foetal and postnatal mammary gland development. *Development* 122(9):2729–2737.
44. Eblaghie MC, et al. (2004) Interactions between FGF and Wnt signals and Tbx3 gene expression in mammary gland initiation in mouse embryos. *J Anat* 205(1):1–13.
45. Hiremath M, Wysolmerski J (2013) Parathyroid hormone-related protein specifies the mammary mesenchyme and regulates embryonic mammary development. *J Mammary Gland Biol Neoplasia* 18(2):171–177.
46. Asselin-Labat ML, et al. (2007) Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* 9(2):201–209.
47. Douglas NC, Papaioannou VE (2013) The T-box transcription factors TBX2 and TBX3 in mammary gland development and breast cancer. *J Mammary Gland Biol Neoplasia* 18(2):143–147.
48. 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.
49. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7):e21800.
50. Cho KW, et al. (2012) Retinoic acid signaling and the initiation of mammary gland development. *Dev Biol* 365(1):259–266.
51. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63.
52. Beccari L, et al. (2016) A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus. *Genes Dev* 30(10):1172–1186.
53. Noordermeer D, et al. (2011) The dynamic architecture of Hox gene clusters. *Science* 334(6053):222–225.
54. Cunha GR, et al. (1995) Mammary phenotypic expression induced in epidermal cells by embryonic mammary mesenchyme. *Acta Anat (Basel)* 152(3):195–204.
55. Propper A, Gomot L (1973) Control of chick epidermis differentiation by rabbit mammary mesenchyme. *Experientia* 29(12):1543–1544.
56. Dhouailly D, Rogers GE, Sengel P (1978) The specification of feather and scale protein synthesis in epidermal-dermal recombinations. *Dev Biol* 65(1):58–68.
57. Kollar EJ, Fisher C (1980) Tooth induction in chick epithelium: Expression of quiescent genes for enamel synthesis. *Science* 207(4434):993–995.
58. Chen CF, et al. (2015) Development, regeneration, and evolution of feathers. *Annu Rev Anim Biosci* 3:169–195.
59. Dhouailly D (2009) A new scenario for the evolutionary origin of hair, feather, and avian scales. *J Anat* 214(4):587–606.
60. Duverger O, Morasso MI (2008) Role of homeobox genes in the patterning, specification, and differentiation of ectodermal appendages in mammals. *J Cell Physiol* 216(2):337–346.
61. Mikkola ML, Millar SE (2006) The mammary bud as a skin appendage: Unique and shared aspects of development. *J Mammary Gland Biol Neoplasia* 11(3-4):187–203.
62. Kanzler B, Prin F, Thelu J, Dhouailly D (1997) CHOXC-8 and CHOXD-13 expression in embryonic chick skin and cutaneous appendage specification. *Dev Dyn* 210(3): 274–287.
63. Xu B, Wellik DM (2011) Axial Hox9 activity establishes the posterior field in the developing forelimb. *Proc Natl Acad Sci USA* 108(12):4888–4891.
64. Noordermeer D, et al. (2014) Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *eLife* 3:e02557.
65. Oftedal OT (2002) The mammary gland and its origin during synapsid evolution. *J Mammary Gland Biol Neoplasia* 7(3):225–252.
66. Oftedal OT, Dhouailly D (2013) Evo-devo of the mammary gland. *J Mammary Gland Biol Neoplasia* 18(2):105–120.
67. Gilbert AN (1986) Mammary number and litter size in Rodentia: The "one-half rule". *Proc Natl Acad Sci USA* 83(13):4828–4830.
68. Sakabe NJ, Savic D, Nobrega MA (2012) Transcriptional enhancers in development and disease. *Genome Biol* 13(1):238.
69. Villar D, et al. (2015) Enhancer evolution across 20 mammalian species. *Cell* 160(3): 554–566.
70. Notwell JH, Chung T, Heavner W, Bejerano G (2015) A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat Commun* 6:6644.
71. Chuong EB, Elde NC, Feschotte C (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351(6277):1083–1087.
72. Lynch VJ, Leclerc RD, May G, Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 43(11):1154–1159.
73. Friedli M, et al. (2010) A systematic enhancer screen using lentivector transgenesis identifies conserved and non-conserved functional elements at the Olig1 and Olig2 locus. *PLoS One* 5(12):e15741.
74. Woltering JM, Noordermeer D, Leleu M, Duboule D (2014) Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* 12(1):e1001773.
75. Guerreiro I, et al. (August 1, 2016) Reorganisation of Hoxd regulatory landscapes during the evolution of a snake-like body plan. *eLife*, 10.7554/eLife.16087.
76. Spitz F, Herkenne C, Morris MA, Duboule D (2005) Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nat Genet* 37(8):889–893.
77. Gimond C, et al. (1998) Cre-loxP-mediated inactivation of the alpha6A integrin splice variant in vivo: Evidence for a specific functional role of alpha6A in lymphocyte migration but not in heart development. *J Cell Biol* 143(1):253–266.
78. Hérault Y, Rassoulzadegan M, Cuzin F, Duboule D (1998) Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nat Genet* 20(4):381–384.
79. Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
80. Giardine B, et al. (2005) Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* 15(10):1451–1455.
81. Yates A, et al. (2016) Ensembl 2016. *Nucleic Acids Res* 44(D1):D710–D716.
82. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
83. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12(3):R22.
84. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
85. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: Computational tools for comparative genomics. *Nucleic Acids Res* 32(Web Server issue):W273–W279.