Impact of copy number variations (CNVs) on long-range gene regulation at the HoxD locus

Thomas Montavona1,2, Laurie Thevenet3, and Denis Duboule1,2

1National Research Centre Frontiers in Genetics, School of Life Sciences, Ecole Polytechnique Fédérale, 1015 Lausanne, Switzerland; and 2National Research Centre Frontiers in Genetics, Department of Genetics and Evolution, University of Geneva, 1211 Geneva 4, Switzerland

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2012.

Contribution by Denis Duboule, October 11, 2012 (sent for review September 1, 2012)

Copy number variations are genomic structural variants that are frequently associated with human diseases. Among these copy number variations, duplications of DNA segments are often assumed to lead to dosage effects by increasing the copy number of either genes or their regulatory elements. We produced a series of large targeted duplications within a conserved gene desert upstream of the murine HoxD locus. This DNA region, syntenic to human 2q31-32, contains a range of regulatory elements required for HoxD gene transcription, and it is often disrupted and/or reorganized in human genetic conditions collectively known as the 2q31 syndrome. Unexpectedly, one such duplication led to a transcriptional down-regulation in developing digits by impairing physical interactions between the target genes and their upstream regulatory elements, thus phenocopying the effect obtained when these enhancer sequences are deleted. These results illustrate the detrimental consequences of interrupting highly conserved regulatory landscapes and reveal a mechanism where genomic duplications lead to partial loss of function of nearby located genes.

chromatin architecture | enhancer-promoter interaction

Genetic variation is both a source of phenotypic diversity and a cause of many human diseases. Such variations range from single-nucleotide exchanges to large rearrangements, including inversions, translocations, or copy number variations (CNVs), collectively referred to as structural variants (1). CNVs are either a gain (duplication) or a loss (deletion) of DNA sequences that can span from kilobase- to megabase-sized intervals. CNVs may include genes and/or noncoding sequences, and they account for a large part of the genetic diversity in animal populations. Over the last years, increasing evidence has associated them with a number of diseases in humans (2, 3).

However, although the consequences of point mutations are relatively well-understood for many genetic conditions, even when the causative variant is located in noncoding intervals (4, 5), the mechanisms where CNVs can cause diseases or malformations are more elusive. In particular, duplications are generally expected to cause dosage effects because of gene duplications within the DNA segment, leading to a global increase in protein products. However, the comparison of tissue transcriptomes from different mouse inbred strains has suggested that CNVs can also affect the expression of genes located nearby the rearrangements at distances of up to several hundred kilobases (6). Such large-scale effects are in agreement with current models of gene regulation, involving complex sets of control elements that can span large genomic distances, particularly for genes with special roles during embryonic development (7, 8). These genes often display complex expression patterns and accordingly, duplications of either known or putative associated enhancers were reported to cause various malformations in humans (for example, at the SHH, IHH, or BMP2 loci) (9–11). However, except for some rare cases (12), the relevant human material could not be assessed, thus calling for the development of animal models of CNVs-induced pathologies.

We have used the murine HoxD gene cluster as a model locus to investigate the impact of structural variation on gene regulation. In mammals, 39 Hox genes are grouped at four genomic loci, referred to as the HoxA to HoxD gene clusters (13). These genes encode transcription factors, which are critical for proper patterning of the embryonic anterior to posterior axis, as shown by genetic evidence in vivo. In addition to this ancestral function, specific Hox clusters have evolved novel functions along with the emergence of diverse embryonic structures (14). The evolution of these new global specificities were often associated to cluster-wide regulations (i.e., to the presence of strong enhancer sequences controlling several Hox genes at one time). For example, the expression of Hoxd genes was co-opted to organize the development of both the proximal (forearm and lower leg) and distal (hands and feet) limb segments (15), as shown by genetic and biochemical studies in mice.

In humans, the HoxD cluster is in a several megabase-sized syntenic region, which expectantly contains all sequences identified as important for the regulation of HoxD genes during murine limb development. Interestingly, many human genetic syndromes displaying limb malformations involve structural variants overlapping with either the HOXD cluster itself or the conserved gene deserts flanking this gene cluster (Fig. 1A). For example, the 2q31 microdeletion syndrome is caused by different deletions of various sizes, overlapping with the LNP-ATP5G3 gene desert on the centromeric side of the HOXD cluster. This syndrome is associated with hand malformations resembling mutations into the HOXD13 gene, even in patients where the HOXD gene cluster itself is not deleted (16), suggesting that such deletions affect regulatory elements controlling HoxD gene expression in developing limbs rather than the genes themselves.

Within the syntenic mouse genomic interval, multiple enhancer sequences were recently described, including a global control region (GCR), several regulatory islands (I to V) dispersed within the Lnp-Atp5g3 gene desert, and the Prox element, which is located between Lnp and Evx2 (17–19) (Fig. 1B). These enhancers collectively form a regulatory archipelago spanning over 800 kb on the centromeric side of the gene cluster and controlling the coordinated transcription of Hoxd13 to Hoxd10, Lnp, and Evx2 in developing digits. In these cells, these various elements are brought to the vicinity of the HoxD cluster by the formation of chromatin microarchitecture, such as looping (Fig. 1C). Furthermore, the genetic dissection of this interval has revealed that each of these regulatory islands participate, in a partially redundant fashion, in the transcriptional activation of the target genes (19).

Author contributions: T.M. and D.D. designed research; T.M. performed research; T.M. and L.T. contributed new reagents/analytic tools; T.M. and D.D. analyzed data; and T.M. and D.D. wrote the paper.

The authors declare no conflict of interest.

Freedly available online through the PNAS open access option.

1Present address: Department of Epigenetics, Max-Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany.

2To whom correspondence should be addressed. E-mail: denis.duboule@unige.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1217659109/-/DCSupplemental.
In addition to deletions, other rearrangements, such as inversions, translocations, or duplications involving sequences flanking the gene cluster, are observed in human patients, which are often linked to various limb anomalies (20–23) (Fig. 1A). In these latter cases, however, the variability of the clinical outcomes and the extent of the genomic modifications, which can be sometimes very large, make the determination of a molecular mechanism linking genotype to phenotype problematic. To understand better the molecular origins of the phenotypes displayed by these human syndromes, we engineered a series of inversions and duplications within the regulatory archipelago controlling mouse Hoxd gene transcription in digits. These rearrangements led to diverse morphological outcomes depending on the genomic topology of the modified locus. Surprisingly, a duplication of a subset of the enhancer sequences displayed a phenotype similar to the phenotype associated with a nonoverlapping partial deletion within the gene desert. We show that this duplication induces a reorganization of the spatial conformation of the regulatory interval in developing digits. This reorganization leads to a loss in the functional contribution of some distal enhancer sequences, resulting in a concurrent down-regulation in the expression of Hoxd genes similar to the down-regulation observed upon deletion of the same sequences. We discuss the relevance of such mechanisms to our understanding of the molecular etiologies of CNVs-associated diseases.

Results
Large Inversions Interrupt a Regulatory Landscape. To characterize the extent of the DNA interval necessary for Hoxd gene transcription in developing digits with more precision, we used a set of inversions with one and the same breakpoint within the Itga6 gene (i.e., 3 Mb centromeric to the gene cluster) and a variety of second breakpoints at distinct positions within the regulatory archipelago (Fig. 2 A and B). In the inverted configurations, DNA sequences normally located centromeric to the proximal breakpoints were repositioned at a large distance from the HoxD cluster. We had reported previously that a 3-Mb large inversion with a breakpoint immediately upstream of Hoxd13 led to a complete loss of Hoxd13 expression in developing digits at embryonic day 12.5 (E12.5). Accordingly, animals of this genotype displayed severe shortening of digits at birth, a phenotype similar to the phenotype obtained upon deletion of the gene cluster itself (24) (Fig. 2C). A shorter inversion, with a proximal breakpoint near the telomeric extremity of the gene desert, separated the cluster from regulatory islands I–V, but it maintained the linkage between the GCR, Prox, and Hoxd genes. In this configuration, Hoxd13 expression was limited to a faint domain at the posterior margin of the hand plate (19) (Fig. 2D). This configuration revealed the functional outcome of both the GCR and Prox sequences alone when left in their endogenous context, and thus, it emphasized the importance of the islands I–V in the transcriptional readout of the system.

We generated inversions using the STRING approach (25), with breakpoints either between islands II and III in the middle of the gene desert [Inv(SB-Itga6)] or on the other side of the desert within the Atf2 gene [Inv(Atf2-Itga6)]. The (SB-Itga6) inversion led to a loss of Hoxd13 expression in the anterior hand plate, including both presumptive digit I and a part of digit II. At birth, affected animals displayed a shorter digit II with a missing phalange (Fig. 2E). In contrast, the Inv(Atf2-Itga6) configuration did not elicit any detectable change in Hoxd13 expression or alterations in the limb phenotype compared with the WT situation (Fig. 2F). From this series, we concluded that (i) sequences farther centromeric from Atf2 are not critical for Hoxd gene expression in developing digits and (ii) the interruption of the regulatory archipelago affects Hoxd gene regulation with a severity that depends on the extent of the DNA segment, which was disconnected from the gene cluster (19).

Duplications Within the Regulatory Interval. To investigate the specificity of these regulatory islands (i.e., to assess whether they could replace one another) as well as to evaluate the impact of their copy number on the transcription of Hoxd genes in digits, we
The Itga6 cluster is shown in trying a targeted deletion from cluster. Each duplication was balanced with a chromosome carrying regulatory elements and their organization relative to the gene. Consequently, they modulated an anterior expression of various modulations compared with controls (Fig. 3A). A duplication of the whole-gene desert from HoxD duplications included either parts or all of the gene desert, the cluster maintains a limited HoxD duplications included either parts or all of the gene desert (24). An inversion leaving the gene desert uninterrupted had no detectable impact on expression to a faint posterior domain (i.e., in the middle of SB-Atf2). A genomic context of HoxD genes in cis with the various modified configurations. 

Fig. 2. A set of nested inversions disrupts the regulatory archipelago controlling HoxD gene transcription in digits. (A) Map of the centromere gene desert along with the positions of the various LoxP sites located within the HoxD regulatory archipelago (red triangles) used for the inversions, deletions, and duplications shown in this study. (B) The WT genomic context of the HoxD cluster is shown in Left, with the location of a remote loxP site within the Itga6 gene used for the set of nested inversions described in C-F. Gray rectangles represent genes, and the HoxD cluster is in white. (Right) The expression of Hoxd13 in an E12.5 limb bud is depicted as well as a WT hand skeleton at birth. (C) A 3-Mb large inversion separates the HoxD cluster from its regulatory elements and thus, abrogates all Hoxd13 expression in developing digits. The resulting phenotype is identical to a full deletion of the HoxD cluster (24). (D) A smaller inversion separating the gene desert from the HoxD cluster maintains a limited Hoxd13 expression to a faint posterior domain. Only the GCR and Prox elements keep their vicinity to Hoxd13. An inversion separating the distal half of the gene desert leads to a decreased anterior expression of Hoxd13 and a shortening of digit II at birth. In this inversion, islands I and II have been removed from the archipelago. An inversion leaving the gene desert uninterrupted had no detectable impact on either Hoxd13 expression or limb morphology. All specimens are homozygous for the various indicated inversions.

Fig. 3. Morphological effects of inducing duplications within the regulatory archipelago. (A) The various duplications were produced by using the same LoxP sites as for Fig. 2 (red arrowheads), and they are depicted by double thick black lines. Below the set of duplications, the position of the del(SB-Atf2) is indicated by a dashed gray line. (B–G) Schematics of the locus after the various duplications (or deletions) were produced as shown in Left, with a hand skeleton at birth shown in Right. In vivo, each configuration was balanced by a chromosome carrying a deletion of Hoxd13 to Hoxd8 [Del(8–13)] allele, indicated as Δ. For the sake of simplicity, the three segments of the regulatory archipelago, as defined by the positions of LoxP sites, are highlighted using different colors (control in B). In B–G, these colors are used to identify the parts of the archipelago that are duplicated (C–F) or deleted (G). In all cases, LacZ reporter genes were associated with the various configurations. They are indicated on the schemes by a blue rectangle along with the presence of the associated LoxP site (red arrowhead). In E, two such LacZ reporters are present (in text). (B) WT configuration. (C) Duplication of the full archipelago from Evx2 to Atf2. (D) Duplication of islands I–V. (E) A duplication extending from Evx2 until the proximal half of the gene desert is associated with a shortening of digit II at birth (arrowhead). (F) Duplication of the Prox-GCR segment. (G) Deletion of the distal half of the gene desert complementary to the duplication in E. Note the similar shortening of digit II (arrowhead).

applied TAMERE (26) to generate a series of large duplications. We used the same breakpoints as described above as well as a loxP site located immediately centromeric of Evx2 (27). These duplications included either parts or all of the gene desert, the Evx2, Lnp, and Atp5g3 genes and the first exons of Atf2 (Fig. 3A). Consequently, they modified both the number of copies of these regulatory elements and their organization relative to the gene cluster. Each duplication was balanced with a chromosome carrying a targeted deletion from Hoxd13 to Hoxd8 [Del(8–13)] to compare animals with a single copy of Hoxd genes in cis with the various modified configurations.

Animals carrying a duplication of the full regulatory interval from Evx2 to Atf2 did not display any detectable defects at birth compared with controls (Fig. 3B and C). Likewise, animals with a duplication of the whole-gene desert from Rel5 to Atf2 were phenotypically normal (Fig. 3D). In contrast, a duplication extending from upstream Hoxd13 to the SB position (i.e., in the middle of the gene desert) induced a shortening of digit II, with a missing phalange [Dup(Nsi-SB)] (Fig. 3E). Digit I was also affected but with an incomplete penetrance. Such a phenotype was not observed, however, when a shorter duplication was analyzed, including Prox, Lnp, and the GCR but without additional copies of regulatory islands III–V (Fig. 3F). Surprisingly, animals carrying the Dup(Nsi-SB) duplication displayed a phenotype strikingly similar to the phenotype associated with a deletion of the SB to Atf2 DNA segment [Del(SB-Atf2)], a deletion of the distal half of the gene desert that removes regulatory islands I and II with no overlap with the duplicated Nsi-SB fragment (Fig. 3G). In addition, the same shortening of digit II was also observed in Inv(SB-Itga6) animals (i.e., in a configuration where the same DNA segment containing islands I and II was disconnected from the gene cluster) (Fig. 2E). Altogether,
these genetic approaches indicated that the duplication of the \textit{Nsi} to \textit{SB} DNA interval had an impact on the function of distal regulatory elements, including islands I and II.

\textbf{Proximal Duplications Affect the Expression of Remote Target Genes.} To address the origin of these phenotypes, we looked at the effect of the duplications on the transcription of \textit{Hoxd} genes in developing digits. Although the expression profile of \textit{Hoxd13} was not affected in either Dup(\textit{Nsi-Atf2}) or Dup(\textit{Rel5-Atf2}) animals (Fig. 4A–D), \textit{Hoxd13} transcripts were lost from the anterior limbs of Dup(\textit{Nsi-SB}) embryos in a territory precisely corresponding to presumptive digit I and part of digit II (Fig. 4E). A milder reduction was observed in the distal limbs of Dup(\textit{Rel1-Rel5}) embryos (Fig. 4F). Interestingly, a similar loss of expression of \textit{Hoxd13} in the anterior aspect of the developing limb bud was observed in the Del(\textit{SB-Atf2}) specimen (Fig. 4G), pointing again to a convergent effect of both the deletion of two upstream regulatory islands and the duplication of a nonoverlapping piece of DNA located between these two islands and the target promoters. We used quantitative RT-PCR (RT-qPCR) to quantify the steady-state levels of mRNAs in these various configurations and found a 60% reduction in the amount of \textit{Hoxd13} mRNA in Dup(\textit{Nsi-SB}) homozygous digits, whereas \textit{Hoxd12} and \textit{Hoxd10} were reduced to \textasciitilde50\% (Fig. 4H). These values are close to the values that we observed in Del(\textit{SB-Atf2}) digits, where \textit{Hoxd} genes are expressed at approximately half their WT levels (19). In Dup(\textit{Rel1-Rel5}) animals, a milder down-regulation was scored, with posterior \textit{Hoxd} genes expressed at ca. 60–70\% of WT levels. In contrast, neither the Dup(\textit{Nsi-Atf2}) nor the Dup(\textit{Rel5-Atf2}) allele caused any significant change in these steady-state levels. In addition, none of these latter duplications did elicit an elevation of mRNAs copies, suggesting that supernumerary regulatory elements did not work more efficiently to activate gene transcription. Altogether, these results indicated that duplications of the proximal part of the regulatory archipelago led to a partial loss of \textit{Hoxd} gene expression, likely by interfering with the activity of more distally located elements (islands I and II), and their normal activities were not compensated for by other regulatory islands when duplicated.

\textbf{Expression of the Duplicated Genes.} This down-regulation of \textit{Hoxd} gene transcription in the two duplicated configurations could be caused by the presence of additional copies of target promoters within the duplicated intervals. The duplicated copies of \textit{Lnp} and \textit{Eva2}, as well as their promoters, might, indeed, titrate the activity of the various enhancers at the expense of \textit{Hoxd} gene expression (28). We investigated this possibility by quantifying RNA levels of these duplicated genes in the various configurations and observed increased steady-state levels of \textit{Lnp} and \textit{Eva2} mRNAs in Dup(\textit{Nsi-Atf2}) digits (Fig. 5). Surprisingly, however, their expression levels were comparable with the WT situation in the Dup(\textit{Nsi-SB}) configuration, which is also associated with supernumerary copies of these genes. Likewise, the Dup(\textit{Rel1-Rel5}) allele did not lead to an increased expression of the duplicated \textit{Lnp} gene. Therefore, gene copy number was not predictive of expression levels in digits, suggesting that the global organization of the interval was critical in defining the final transcriptional activity.

In the same context, several of the parental \textit{loxP} sites, which were used to produce the duplicated alleles, are associated with a \textit{LacZ} reporter gene. As a consequence, \textit{LacZ} reporter transgenes are present within each duplication allele. In particular, the Dup(\textit{Nsi-SB}) allele, which lead to the strongest transcriptional interference, displayed two neighboring \textit{LacZ} copies in between the duplicated fragments (Fig. 5E). The potential impact of these transgene insertions was evaluated by comparing their expression profiles in the various configurations (Fig. 5F).

In the parental strains, the reporter genes displayed expression profiles specific for their insertion sites within the regulatory landscape. For example, when an \textit{Hoxd11} \textit{LacZ} gene was located immediately upstream of \textit{Hoxd13}, its transcription was detected in both the posterior trunk and proximal limb, corresponding to the future forearm (Fig. 6A), in addition to the strong staining observed in distal limb buds. In contrast, although a \textit{LacZ} transgene integrated in the middle of the gene desert was expectedly expressed in distal limb buds as well, staining also appeared in limited cell populations within the neural tube (Fig. 6B). Likewise, the Dup(\textit{Nsi-SB}) and Dup(\textit{Nsi-Atf2}) alleles were associated with slightly different patterns of transcriptional activity in the neural tube, but \textit{LacZ} staining was always detected in developing digits in the same domain (Fig. 6C and D) and at comparable levels (Fig. 5E).

\textbf{Impaired Long-Range Interactions.} To challenge this hypothesis, we applied chromosome conformation capture (4C) (29, 30) to establish the long-range interaction profile of \textit{Hoxd13} in both WT and Dup(\textit{Nsi-SB}) developing digits. In WT digit cells, \textit{Hoxd13} is...
brought to the vicinity of the various distal regulatory elements, likely through chromatin looping (19) (Fig. 7A), suggesting that a tight physical proximity is established between the promoter of this gene and the various regulatory islands. In the duplicated configuration, this profile of interactions was substantially altered. The detected interactions between Hoxd13 and sequences located within the duplicated segment were globally stronger, particularly with DNA segments displaying relatively weak interactions in the WT situation, such as the vicinity of the centromeric breakpoint (Fig. 7A, gray arrowhead). In contrast, sequences located farther centromeric of the duplicated segment (i.e., whose genomic distance to Hoxd13 had been increased) displayed decreased interaction frequencies, including regulatory islands I and II (Fig. 7A, black arrowheads).

This result suggests that the phenotypes observed in the duplications derive from reduced interactions between Hoxd target promoters and islands I and II because of either an increase in the absolute distance in between, or an increase in the total number of islands. We controlled this experiment by using island I instead of Hoxd13 as a viewpoint in a 4C approach. In WT developing digits, island I interacted with both other regulatory elements and the centromeric part of the HoxD cluster. In Dup(Nsi-SB) digits, these contacts were stronger with sequences located within the duplicated segment (Fig. 7B). As for the Hoxd13 interaction profile, the interactions observed for island I seemed generally less specific for the regulatory elements than in the WT situation. Instead, interactions were reinforced within most of the duplicated interval. In agreement with what we observed when using Hoxd13 as a viewpoint, the interactions of island I with the centromeric part of the cluster, including sequences located in the Hoxd13 to Hoxd10 interval, were reduced (Fig. 7C). We concluded that the duplication led to an altered conformation of the regulatory landscape, impairing the association of Hoxd genes with distal regulatory elements and thus, leading to a localized loss of Hoxd gene expression in the digit-forming area with concurrent morphological defects (Fig. 7D and E).

Discussion

Copy Numbers and Genomic Topology. The transcription of Hoxd genes in developing digits relies on the activity of multiple regulatory elements dispersed over an interval of ca. 800 kb, overlapping with a conserved gene desert that extends from Lnp to Atg5p3 (19). A duplication covering the proximal part of this regulatory archipelago led to an unexpected and localized loss of Hoxd gene transcription, with the associated morphological defects in digits. This effect was strikingly similar to the effect elicited by a deletion of the most distal third of the landscape, suggesting that both types of rearrangements had a common impact on Hox genes transcription.

In contrast, a duplication of the entire regulatory interval had no detectable effect on Hox genes regulation, indicating that it is the genomic organization of the regulatory elements relative to their target genes rather than their absolute number of copies that is of importance for the transcriptional output of the system. Accordingly, we did not observe any correlation between the impact of duplications on Hoxd gene regulation, on the one hand, and the expression levels of either duplicated target genes (Lnp, Evx2) or exogenous transcription units within the landscape, on the other hand, arguing against an interference caused by enhancer/promoter competition, which was described in other contexts (31). This effect is also distinct from known cases of copy number-induced silencing, in which multicopy transgenes or repeat elements are themselves repressed rather than their neighbors (32). Only duplications increasing the distance between the HoxD cluster and distal enhancers induced a decreased transcription in distal limbs; these regulatory interferences were a function of the size of the duplicated segment, because a short proximal duplication, including both Prox and the GCR but not regulatory islands III–V, caused a down-regulation of Hoxd genes milder than the down-regulation observed with a larger duplication, insufficient to elicit a detectable morphological alteration.

This observed decrease in the transcriptional readout whenever the distal islands I and II are moved away from their target further highlights the specific requirement for all of the various regulatory elements to establish the genuine expression profile of Hoxd genes in WT condition. Additional copies of a subset of these enhancers, indeed, cannot compensate for the relocation of others at a distance. This observation indicates that multiple regulatory elements are not merely required to provide a sufficient number of binding sites for a similar set of trans-acting factors but instead, that various islands may recruit (at least partially) distinct molecular complexes leading to subtle qualitative and quantitative differences.

Modified Architecture of the Regulatory Landscape. The down-regulation in Hoxd gene transcription scored with the large proximal duplication was associated with a modified spatial organization of the genomic regulatory interval. The contacts established both by Hoxd13 and island I, which are located telomeric and centromeric of the duplicated segment, respectively, were strengthened with those sequences lying within the duplicated segment, whereas they were clearly weakened with more distally located sequences (Fig. 7D and E). Because our 4C approach cannot discriminate between contacts experienced by each of the two copies of any duplicated DNA sequences, we do not know whether the increased levels of interactions observed with these sequences reflect an association of the viewpoint with both copies of the same island or alternatively, a reinforced interaction with one of them only.
Likewise, decreased interactions between the viewpoints and sequences distal to the duplicated segment (such as islands I and II when assessed from Hoxd13) could be because of either a larger genomic distance between the target genes and the enhancer sequences or alternatively, a competition between the various elements for the formation of long-range interactions, such that additional copies of islands within the duplicated sequences would compete out the contacts between Hoxd genes and distal sites. In the former case, the intercalation of any similar-sized piece of DNA would lead to the same effect. However, the concomitant increase in the quantity of contacts with duplicated enhancers suggests that these sequences participate in the spatial organization of the landscape in the duplicated mutant, as if the reiterated segment would now actively take part to this regulatory architecture. As a consequence, distal islands I and II would be somehow left out of the structure (Fig. 7 D and E).

**Genome Evolution and Human Disease.** Interestingly, Hoxd gene regulation seemed unaffected in all of the mutant configurations where at least one complete regulatory archipelago was maintained upstream of the gene cluster. In contrast, any condition interrupting this interval, by either an inversion or a duplication intercalating some DNA sequences within the regulatory landscape, resulted in a partial loss of expression. Therefore, it is critical for the proper transcriptional control of these genes that an integral regulatory block be preserved, including the gene desert. Such a regulatory constraint likely provided a selective pressure to maintain this highly syntenic region uninterrupted, because the Lpr-Atp5g3 gene desert is present upstream of the HoxD cluster in all vertebrate genomes that have been sequenced so far. Similar constraints might participate in the stability of other gene deserts, since such DNA intervals have been associated with long-range regulation in several instances (33–40).

These observations also suggest a mechanism where duplications overlapping regulatory regions may lead to a decreased expression of critical target gene(s) by disturbing the intricate organization of complex regulatory landscapes. Such a mechanism could underlie the molecular etiology of some CNV-associated diseases in humans. This possibility is rarely discussed, because an increase in copy number of putative regulatory elements is usually expected to result in an increased expression of their target genes. More precise and exhaustive analyses may, thus, reveal a higher complexity in the organization of regulatory landscapes, and hence, the effects of CNVs affecting such regions may have to be integrated into a global topographic context rather than using mere quantitative parameters.

The complexity of CNV-associated diseases is highlighted by three overlapping duplications, including the human HOXD cluster and flanking sequences. Two such duplications are associated with mesomelic dysplasia (a shortening of the forearm and lower leg), whereas a third and larger duplication causes syndactyly (fused digits) (21–23). Such distinct clinical outcomes as well as the present report illustrate the difficulty in elucidating those molecular mechanisms underlying the pathological consequences of complex genetic conditions in humans in the absence of a proper and adapted animal model.

**Materials and Methods**

**Mouse Strains.** The Del(8–13), Inv(Nsi-itga6), Del(SB-Atf2), and Inv(Rei5-itga6) alleles were described previously (19, 24, 41). The inversions (SB-itga6) and (Atf2-itga6) where generated by STRING (25) using an loxP site inserted into the itga6 gene (42) and a second loxP site located either at the SB position within the gene desert (43) or in the Atf2 gene (44). Recombinant offspring with both loxP sites in cis were crossed with Hprt-Cre mice (45). Duplications were produced by TAMERE (26) using the loxP sites in Nsi (46), Rei1 (27), Rei5 (19), SB, and Atf2. Genotyping of mice and embryos was performed by PCR analysis (SI Materials and Methods).
LacZ Staining, in Situ Hybridization, and Skeletal Preparation. Detection of LacZ reporter activity and in situ hybridization were performed according to standard protocols. The Hoxd13 probe was previously described (47). For skeletal preparation, newborns were stained with standard Alcian blue/Alizarin red protocols.

RT-qPCR Analyses. Presumptive digits were dissected from E12.5 embryos and stored in RNAlater reagent (Qiagen) before genotyping. RNA was isolated from individual embryos using the RNeasy microkit (Qiagen); 500 ng RNA were reverse-transcribed using random primers and SuperScript III RT (Invitrogen). cDNA was PCR-amplified using SYBR green containing qPCR master mix (Invitrogen) with a CFX96 Real-Time System (Bio-Rad). A mean quantity was calculated from triplicate reactions for each sample. Expression changes were normalized to Rps9. Primers used were as described (48).

4C Analysis. Presumptive digits were dissected from E12.5 embryos, dissociated by collagenase, and fixed in 2% formaldehyde for 10 min at room temperature. Nuclei were stored at −80 °C until genotyped. 4C libraries were produced as described (49) using NlaIII and DpnII (New England Biolabs) as primary and secondary restriction enzymes, respectively. Digits samples from 10 embryos were pooled for each library. Religated sequences were amplified by inverse PCR with AmpliTaq DNA polymerase (Applied Biosystems) using 200 ng 4C library per reaction and the following primers: Hoxd13-F 5′-AAAATCCTAGACCTGGTCATG-3′; Hoxd13-R 5′-GGCCGATGGTGCTGTATAGG-3′; island I-F 5′-AAGTAGCAAAGCAACCACAGTAAAG-3′; and island I-R 5′-GGCAGAAATGTGGAAAGGTCA-3′. For each condition, 16 reactions were pooled and purified using the Qiagen PCR Clean-Up Kit, fragmented and labeled using the GeneChip WT Double-Stranded DNA Terminal Labeling Kit (Affymetrix), and hybridized to custom-made tiling arrays (50). Arrays were processed according to the manufacturer's instructions. For each genotype and fragment of interest, two independent samples were analyzed.

Tiling Array Data Analyses. Array data were quantile-normalized within 4C-amplified/input replicate groups and scaled to medial feature intensity of 100 using TAS software (Affymetrix). For each genomic position, a dataset was generated consisting of all (PM-MM) pairs mapping within a sliding window of 2 kb (broad view) or 500 bp (close view). Average ratios were plotted along the genomic DNA sequence using Integrated Genome Browser software (Affymetrix) (SI Materials and Methods).
ACKNOWLEDGMENTS. We thank B. Mascrez for help with mice, E. Joyce for technical assistance, and G. Andrey, N. Soshnikova, and members of the D. D. laboratories for discussions and reagents. This work was supported by a fellowship from the Human Frontier Science Program Organization (HFSPO) (to L.T.) and funds from the Ecole Polytechnique Fédérale de Montavon et al. PNAS

Supporting Information

Montavon et al. 10.1073/pnas.1217659109

SI Materials and Methods


Primer sequences are listed below:


In the case of duplicated alleles, we discriminated between homozygous and heterozygous embryos by quantifying copy numbers of the mutant alleles by quantitative PCR using a primer pair mapping within the LacZ coding sequence in the reporter gene associated with the duplication (5′-ATCAGGATAGTTGGGATGTA-3 and 5′-TGATTTGTGTAGTCGGTTATGCA-3).

Chromosome Conformation Capture Data Analysis. Quantitative differences in chromosome conformation capture interaction profiles were calculated using TAS software. Each interaction profile was normalized with input DNA and mismatch probe values. For each genotype and viewpoint, two replicate samples were processed. For the comparison of WT and Dup(Nsi-SB) intensities, datasets were normalized in parallel and scaled to equal median intensities. Differences in signal intensities between Dup(Nsi-SB) and WT profiles were considered significant for intervals displaying a P value lower than 0.001 over a minimal window of 100 bp (Wilcoxon signed-rank test). Arrowheads in Fig. 7 highlight a subset of these intervals.