REVIEW

doi:10.1038/nature12753

Topology of mammalian developmental enhancers and their regulatory landscapes

Wouter de Laat[1] & Denis Duboule[2,3]

How a complex animal can arise from a fertilized egg is one of the oldest and most fascinating questions of biology, the answer to which is encoded in the genome. Body shape and organ development, and their integration into a functional organism all depend on the precise expression of genes in space and time. The orchestration of transcription relies mostly on surrounding control sequences such as enhancers, millions of which form complex regulatory landscapes in the non-coding genome. Recent research shows that high-order chromosome structures make an important contribution to enhancer functionality by triggering their physical interactions with target genes.

Access to animal genome sequences has revealed that the level of complexity of an organism does not relate to its number of genes. Mammals are more complex in morphology and behaviour than roundworms, but their genomes both contain around 20,000 genes. Various parameters can contribute to increased complexity, such as the extent of protein modifications or the diversity of splicing patterns. Pleiotropy is another possible contributor, whereby genes acquire multiple functional tasks at different times and places either during development or in adult life. In this case, gene regulation, rather than function, had to evolve to associate regulatory alternatives to particular genes. Although gene transcription is initiated at promoters, which recruit the basal transcription machinery, these sequences have little impact on transcription control during development and hence this latter task mostly relies on enhancers[1].

Enhancers are sequence modules that contain binding motifs for transcription factors. They are preferentially located in the non-coding part of the genome, at various distances from their target genes[2,3]. In mammals, more than 95% of the genome is non-coding and large gene deserts can sometimes span several megabases. The recent development of high-throughput methods has made it possible to systematically search for enhancers; millions of such regulatory modules have been predicted[4], with 40% of our genome now estimated to carry some regulatory potential[5]. The importance of enhancers for normal development and disease is further underscored by the fact that disease-associated single nucleotide polymorphisms (SNPs) often co-localize with these modules[6]. In addition, congenital diseases and cancers can be induced by chromosomal rearrangements that affect the regulatory neighbourhoods of target genes[7,8].

With so many more potential enhancers than genes, an outstanding task is to functionally connect mammalian regulatory sequences to target genes. In this context, the three-dimensional (3D) configuration of the genome is important because it must accommodate the physical contacts between promoters and distant enhancers. Chromosome conformation studies and genetic analyses of representative loci have recently started to uncover the complex and versatile mechanisms behind target gene selection and enhancer landscape recruitment. In this Review, we discuss a few specific cases involving long-range gene regulation in mammals to illustrate emerging principles whereby remote enhancers can achieve their functions in complex genomic environments.

Evolution of mammalian enhancer landscapes

Vertebrate genomes are unique in that they contain large gene deserts with enhancers acting over distances in the megabase range (see ref. 9 for a review). Invertebrate species studied so far tend to have more local regulatory controls, which can often be recapitulated by short transgenes, such has been shown for the roundworm *Caenorhabditis elegans*. Admittedly, in *Drosophila,* gene regulation during development is complex, with multiple enhancers acting on individual genes[10] and some loci controlled by series of intricate enhancers[11]. However, these enhancer–promoter interactions generally occur over distances shorter than 50 kb (see for example ref. 12) (Fig. 1).

The apparent restriction of megabase-sized regulations to vertebrate genomes is puzzling. It has been argued that the emergence of vertebrates was accompanied by a burst of pleiotropy and gene multi-functionality, such that crucial gene functions were co-opted for a variety of additional tasks (see refs 13 and 14 for references). This was probably achieved by multiplication of enhancers per gene of interest — a process triggered by the two genome duplications that occurred at the root of this taxon. Duplications made task-sharing among paralogous genes possible and may thus have given duplicated genes the licence to evolve additional regulations[15]. As a result, many vertebrate genes that are essential for important developmental pathways and active at different places and times (for example, *Hox*, *Pax*, *Fgf*, *Bmp* and *Hh*), have been kept in several paralogous copies and display complex regulatory landscapes.

Finding regulatory sequences

The complexity of the mammalian regulatory genome is revealed by the analysis of transgenic mice carrying a transposable reporter gene cassette[16] used as an enhancer trap[17]. The staining of hundreds of embryos with an insertion at different genomic locations showed that the minimal promoter was silent at 40% of the integration sites. In nearly 60% of the embryos, however, the reporter gene was active, usually with tissue-restricted expression, showing its capacity to integrate resident regulatory signals. These specific patterns often followed that of the nearest gene. Tissue-specific transcription was also detected near housekeeping genes, suggesting that ubiquitous transcription may result from the integration of various specific cues. Although integration sites located hundreds of kilobases apart could give rise to the same expression patterns, elsewhere in

[1]Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT, Utrecht, the Netherlands. [2]School of Life Sciences, Ecole Polytechnique Fédérale, CH-1015, Lausanne, Switzerland. [3]Department of Genetics and Evolution, University of Geneva, CH-1211, Geneva, Switzerland.

the genome distinct transcription patterns were identified between integration sites separated by only few kilobases, making the mouse genome a regulatory jungle[16] (Fig. 2 and Box 1).

This perception is supported by the systematic mapping of enhancers, carried out either by looking at particular chromatin features associated with enhancer activity, or by multi-species alignments of non-coding DNA sequence in syntenic regions (Table 1). Although the latter approach has identified many enhancers genome wide[18], enhancers may diverge in their sequences between species[19]. Nevertheless, DNA sequence is the prime determinant of transcription factor binding; when a human chromosome was introduced into mice, both the murine transcription factor binding profiles and the resultant gene expression patterns across the human chromosome were nearly identical when compared with human hepatocytes[20]. Therefore, DNA motifs allowing enhancer prediction must exist and, by using experimentally derived enhancer motifs as inputs for machine-learning algorithms, several groups recently made accurate predictions[21–23].

Transcription-factor-binding to DNA creates chromatin signatures that can also be used to experimentally identify enhancers. Such signatures include the local opening up of chromatin (uncovered by DNaseI sensitivity[24,25]), the presence of transcription factors and co-factors such as p300 and the deposition of histone marks such as monomethylation of histone H3 lysine 4 (H3K4) and acetylation of H3K27, as assayed by chromatin immunoprecipitation (ChIP). Systematic mapping by these methods across cell types have uncovered millions of sites with regulatory potential, most of them recognizable only in a tissue-specific manner[26].

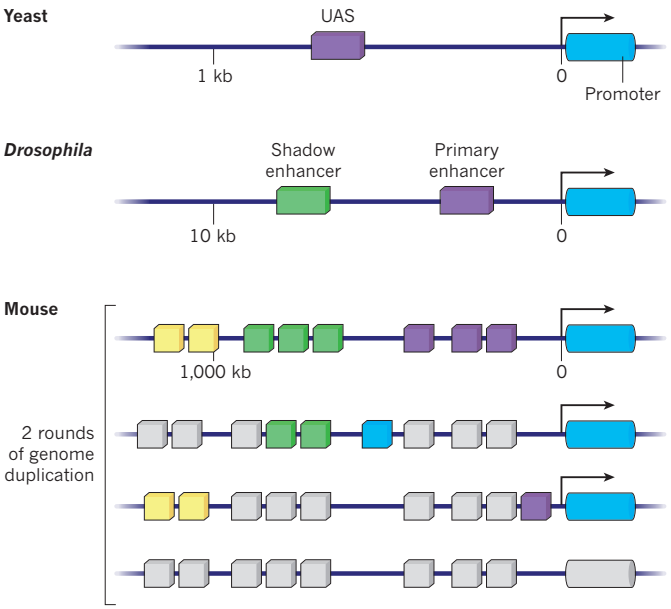

Figure 1 | Variations in long-range gene regulation. Differences exist between the enhancer–promoter distances of yeast, *Drosophila* and mice. Hypothetical and representative gene loci are shown with yeast locus in which very few enhancers, or upstream activating sequences (UAS), are usually found; those that are, exist within 1 kb (up to maximum of a few kilobases) of the promoter. In *Drosophila*, often multiple enhancers exist and they are usually located within 10 kb of the promoter. Occasionally, they are found at distances of up to 100 kb, and some complex developmental regulatory landscapes have been reported, for example the BX-C locus, which stretches over distances of around 300 kb, but such intricate regulations do not seem to be the rule in *Drosophila*. In mammals, shown here in mice, regulatory landscapes found around developmental genes often extend over several hundred kilobases to more than 1 Mb. The two rounds of genome duplication that accompanied the emergence of vertebrates may have allowed for additional regulatory complexity to develop, owing to the release of constraints associated with the target gene, thus triggering the *de novo* evolution of enhancers and the diversification of their use. As a result, different paralogous landscapes display various enhancer combinations (coloured squares). Concomitantly, large gene deserts may have evolved to prevent bystander effects and regulatory interferences.

Functional screens for enhancers

Predicted enhancers should be validated in functional essays. Transgenic animals are typically used for this purpose and can validate up to 90% of the inserted DNA sequences. Recently, self-transcribing active regulatory region sequencing (STARR-seq) has facilitated screens based on enhancer activity, whereby candidate sequences stimulate their own transcription[27]. Enhancer strength in one particular cellular context is thus reflected by the abundance of the corresponding transcripts. When applied to *Drosophila* S2 cells in which 11 million random fragments were tested for activity, a number of interesting observations were made. For example, 5% of the 5,499 regions displaying enhancer activity were bona fide transcription start sites, showing that, to some extent, promoters may also enhance transcription. Many strong enhancers were mapped near housekeeping genes, and often multiple enhancers (five or more) seemed to target the same gene, including housekeeping genes, further suggesting that ubiquitous expression may also be controlled by complex networks of tissue-specific enhancers rather than being a promoter-intrinsic feature.

When compared with DNaseI and various ChIP profiles obtained from S2 cells, a third of the enhancer sequences scored by STARR-seq were within a repressive chromatin configuration, often near silent, important developmental genes. These sequences lacked H3K27ac but carried H3K27me3 (reflecting Polycomb-mediated repression) and H3K4me1 (a modification found at enhancers). These sequences were thus recognized as functional enhancers, but they were actively silenced[27]. This emphasizes a limitation associated with all enhancer screening methods; although a given DNA sequence may display regulatory capacity, it may not exert this property in its physiological context, raising the need not only to evaluate enhancer functionalities by genetic approaches in their natural environment, but also to study in some detail the nature of these environments. Chromosome topology has become an important parameter in this context, because it accommodates the wiring between enhancers and their target endogenous genes. Therefore, enhancer action must be considered in the 3D structure of the genome.

Enhancer action through DNA looping

The development of chromosome conformation capture (3C) technologies[28] has advanced our understanding of regulatory interactions in the 3D genome. 3C considers the amount of ligation products between cross-linked DNA segments as a function of their contact frequencies *in vivo*. Originally designed for the analysis of one-to-one contacts between selected pairs of genomic sequences, high-throughput variants now allow the assessment of one-to-all (4C), many-to-many (5C) or all-to-all (Hi-C) contacts, thus providing distinct details about chromatin topology[29]. Using the β-globin (also known as *Hbb*) locus, 3C studies showed that globin genes form tissue- and differentiation-specific contacts with the distant locus control region (LCR), thus illustrating that spatial proximity to enhancers can increase promoter activity[30,31]. Presumably, this proximity increases the local concentration of DNA binding modules, causing a nearby accumulation of cognate transcription factors at the promoter that strengthens its transcriptional output.

Transcription itself is not required for the formation of enhancer–promotor loops[32]; however, the loops are required for transcription, as confirmed by enhancer–promoter loops engineered at the β-globin locus *in vivo*[33]. In mouse cells that lack the erythroid-specific transcription factor Gata1, chromatin loops are absent and globin genes are silent. Although the Gata1-associated protein Ldb1

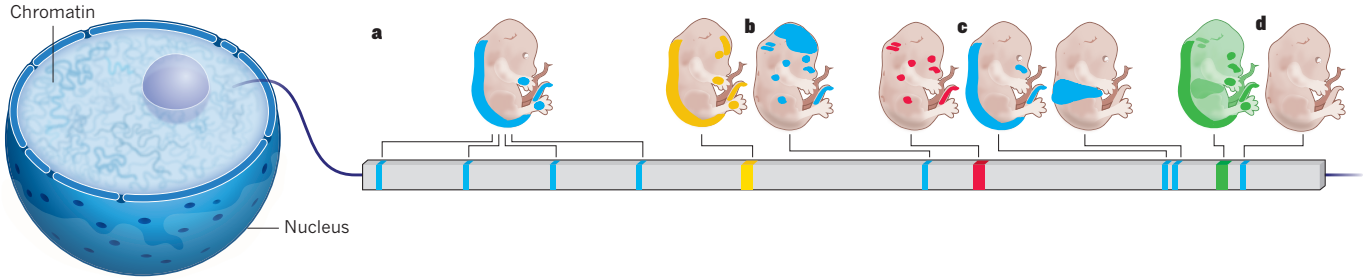

Figure 2 | The mammalian regulatory jungle. A model of three hypothetical genes (yellow, red and green) and their hypothetical expression pattern at a given stage of embryonic development are shown. Embryos coloured blue show the activity of a given reporter gene integrated at different chromosomal locations (adapted from ref. 16). They illustrate that genomic context critically determines expression patterns. Thus, **a**, various insertion sites may display comparable expression patterns despite being spread over a large chromosomal interval. Note that these reporter genes incorporate most of the regulatory activities acting on the downstream gene shown in yellow. **b**, Often, the reporter gene incorporates the enhancer activities that control the expression of one the nearest genes (red gene). **c**, Tissue-specific reporter gene expression can sometimes be seen at sites close to housekeeping genes (green gene). In addition, two closely linked integration sites may show very distinct expression patterns that reveal highly localized regulatory circuits. **d**, At some chromosomal sites, the reporter gene is inactive and apparently not capable of capturing enhancer activity.

is no longer recruited to the β-globin gene promoter, it still binds to the LCR through other transcription factors. When artificial zinc fingers were used to force recruitment of Ldb1 to the β-globin promoter, looping with the LCR was induced in Gata1-null cells leading to robust transcription activation[33]. Likewise, an isolated human LCR introduced into transgenic mice at an ectopic site was capable of *trans*-activating an endogenous β-globin gene located on another chromosome only in those cells in which an inter-chromosomal contact was established[34]. Both studies show that contacts are necessary for enhancers to activate transcription.

Pre-established compared with *de novo* formed loops

Enhancer–promoter loops have now been identified at many loci. However, the presence of such loops is not necessarily associated with an active transcriptional outcome, suggesting that some loci display a spatial configuration, which is poised for transcription. This permissive situation is in contrast to loops that are initiated at the time of transcriptional activation because of an instructive process (Fig. 3). In such a permissive situation, a regulatory landscape exists in a preformed 3D conformation that can be used in any cell type by tissue-specific transcription factors for efficient transcription activation[35]. In the instructive model, it is the *de novo* establishment of a chromatin configuration triggered by specific transcription factors, for example by looping, which will cause transcriptional activation. Studies on a restricted number of loci have highlighted these various alternatives.

Enhancer–promoter loops at the α-globin (also known as *Hba*) and β-globin gene loci are formed exclusively in erythroid cells, with interaction frequencies increasing during erythroid maturation[36,37]. In addition, loop formation depends on the presence of erythroid-specific transcription factors such as Klf1 and Gata1 (refs 38, 39). Likewise, the contacts between the *SatB1* gene and its enhancer landscape found in a 800 kb large flanking gene deserts, are formed *de novo* in thymocytes in which *SatB1* is transcribed at high levels[40] (Fig. 3b). Therefore, in both cases, tissue-specific factors are necessary to trigger long-range contacts, specifically in tissues that require high transcriptional output.

By contrast, evidence for preformed configurations has been found both at the *Hox* (discussed later) and the *Shh* gene loci. *Shh* expression in posterior limb buds is necessary for correct limb development[41] and is controlled by an enhancer (ZRS) located 1 Mb away, within an intron of the *Lmbr1* gene. Mutations in the ZRS cause polydactylies in both humans and mice[42,43]. Despite this large distance, the ZRS loops and contacts the *Shh* locus[44]. This loop, however, is not specific for posterior limb cells because privileged contacts are already observed in embryonic stem cells in which *Shh* is inactive[45]. Also, contacts between the enhancer region and the promoter occur even in the absence of the enhancer itself[44] and the ZRS activity can be expanded to ectopic sites in the limb bud whenever mutations recruit new sets of transcription factors[46]. This suggests the existence of a preformed topology that organizes the physical proximity between the ZRS and its target *Shh* gene. In this view, the locus may be in a 'permissive' configuration and tissue-specific transcription factors acting through the ZRS would merely select and consolidate an existing structure. The transcription factors p53 and FOXO3 seem to act similarly through pre-existing chromatin loops[47,48]. These two proteins are not developmental regulators, however, but factors required for cell proliferation and survival.

Preformed, permissive structures potentially offer some regulatory benefits. They may help to target enhancers to a gene of interest, thereby preventing bystander activation of unrelated neighbouring genes[49]. Transcriptional activation may also be simpler to implement, because it may only involve slight and discrete variations in internal contacts within a largely conserved structure, in a way that is related to allosteric transitions of single molecules[35]. Finally, the existence of preformed and as yet inactive regulatory landscapes could have been a rich playground for the emergence of new enhancers, because both the basic structural context and the transcriptional outcome would be available and ready to be hijacked by factors with distinct tissue specificities.

Promoter contact networks

High-throughput variants of 3C technology allow the simultaneous analysis of contacts made by multiple promoters. Comprehensive studies of such long-range promoter interactions have either used 5C to interrogate contacts across 1% of the genome[50] or the ChIA-PET approach[51]. ChIA-PET combines ChIP with a 3C strategy to uncover the chromatin loops formed by genomic sites that are bound by a protein of interest (in this case, RNA polymerase II)[51]. Both 5C and ChIA-PET studies have shown that most promoters are engaged in chromatin loops, often in cell-type-specific manners. Using 5C, contacts were found with sites showing enhancer-type chromatin signatures and with sites bound by CCCTC-binding factor (CTCF), a chromatin architecture protein known to be involved in loop formation[52] and chromatin organization. However, most contacts were with 'unclassified' sites without any recognizable chromatin or sequence mark[50].

Inter-promoter contacts were also discovered by ChIA-PET and preferentially occurred between genes displaying coordinated expression. *In vitro* reporter assays further suggested that promoters can enhance each other's activity[51]. Although it is often assumed that enhancers target the nearest gene, 5C data suggest that this is true in only 7% of cases. In addition, nearly 80% of long-range DNA contacts remained unaffected when intervening sequences were bound

BOX 1

Regulatory landscapes in mammalian genomes

- Around 20,000 genes
- More than 10^6 enhancers (potential regulatory sequences)
- About four enhancers contact an active gene on average per cell type[50]
- Average enhancer–promoter loop size[50] is 120 kb
- Largest enhancer–promoter distance so far (*SOX9*, Pierre Robin disease)[74] is 1,300 kb
- 545 gene deserts (>640 kb, that is, top 3% largest deserts)[75]
- Largest gene desert[75] is 5.1 Mb

by CTCF, challenging the idea that the major role of this protein is to block enhancer–promoter contacts. Finally, at least 10% of distal sites were engaged in contacts with multiple genes and, likewise, many genes formed contacts with more than one distant site[50].

The functional relevance of these complex interactions is admittedly difficult to evaluate. Enhancers defined by genetic approaches generally contact their target genes in chromosome conformation studies[31,44] and bona fide enhancers have been isolated using 4C analyses[35,40,53]. However, a physical contact does not necessarily reflect a functional interaction[37,54]. Therefore, although 3C-based strategies can physically connect genes to potential enhancer sequences, genetic approaches in the appropriate *in vivo* systems are required to reveal whether and how such potential enhancers are functionally connected to a specific target gene.

Target selection by promiscuous enhancers

Mammalian promoters not only consist of a core sequence but often also contain immediate upstream binding sites for general and tissue-specific transcription factors, which serve as a first regulatory layer to confer some tissue specificity. These additional modules can integrate regulatory activities from remote enhancers, as illustrated by the effect of a transgenic 'orphan' LCR (without any of the globin family of genes). When positioned into an unrelated locus, an LCR elicited a tissue-specific upregulation of many of its surrounding endogenous genes, which normally do not encounter this enhancer[55]. In another study, the regulation of *Fgf8* was examined genetically. *Fgf8* encodes a protein of a pleiotropic signalling pathway and hence its transcription during development must be tightly controlled. In the 200 kb region surrounding *Fgf8*, several unrelated genes are found, as well as nearly 50 regulatory modules that bear tissue-specific information. Chromosomal rearrangements of the *Fgf8* locus *in vivo*, whereby unrelated genes were placed at the *Fgf8* position, recapitulated the highly specific developmental expression of *Fgf8* (ref. 56).

Such a prevalence of enhancer strength over promoter selectivity must lead to situations in which functionally unrelated neighbour genes share tissue-specific expression patterns. Such bystander effects do exist[57–59], and are best seen when genes are upregulated after genetic rearrangements. For example, the deletion of the two α-globin genes re-directs their enhancer to the *NME4* gene, 300 kb away, causing an eightfold increase in its expression[60]. In addition, enhancers that control *Hoxd* genes in digits can recruit a new set of target genes after a Robertsonian translocation[61]. By contrast, a promoter-creating mutation in between the α-globin genes and their cognate enhancers causes the blood disorder α-thalassemia, presumably by re-allocating the enhancer away from α-globin genes[62]. In evolutionary terms, bystander effects can be prevented by restricting the number of genes within a regulatory landscape. This may explain why key developmental vertebrate genes are frequently flanked by evolutionary conserved gene deserts (or gene-poor regions) that are rich in enhancer sequences[9]. In contrast, housekeeping genes often cluster in the genome, which may help them to collectively compete for and titrate out enhancer activity, thereby buffering potential fluctuations in expression.

Table 1 | Methodology for screening for functionally relevant enhancers

Data type that identification is based on	Sensitivity	Specificity
Comparative genomics		
Screens for evolutionarily conserved sequence blocks (*in silico*)	Will not score evolutionarily diverged enhancers[19]	Also identifies regulatory sites with no enhancer activity (promoters, insulators and architectural sites) and redundant enhancers, for example without target genes (orphan enhancers)[76]
ChIP-seq of transcription factors, p300, H3K27ac and H3K4me1		
Genomic screens for sites associated with protein factors and histone modifications often found at enhancers	Will not score enhancers lacking detectable signature	Also identifies non-enhancers with similar signatures and redundant enhancers, for example those without target genes (orphan enhancers)
DNaseI profiling		
Screens for genomic sites with locally opened up chromatin	Expected to identify (nearly) all sites with regulatory potential	Also identifies other regulatory sites (promoters, insulators and architectural sites) and redundant enhancers, for example those without target genes (orphan enhancers)
Enhancer trap		
Transgenic reporter genes as a read-out for local enhancer activity at different genomic locations	A medium-throughput technique: misses enhancers owing to the limited number of integration sites analysed; and enhancers incompatible with reporter gene promoter	Also identifies redundant enhancers, for example without target genes (orphan enhancers)
STARR-Seq		
Functional screens for genomic sequences across the genome for their capacity to enhance their own transcription	Will not score enhancers incompatible with reporter gene promoter	Also identifies occluded enhancers that are actively repressed in the cell of interest and redundant enhancers, for example those without target genes (orphan enhancers)
3C-based methods (promoter centred)		
Screens for chromosomal sites that physically contact a promoter of interest	Will not score infrequently contacted enhancers, enhancers located close to (<10 kb) the promoter and enhancers acting independently from promoter contact	Also identifies bystander contacts with non-enhancers and redundant enhancers

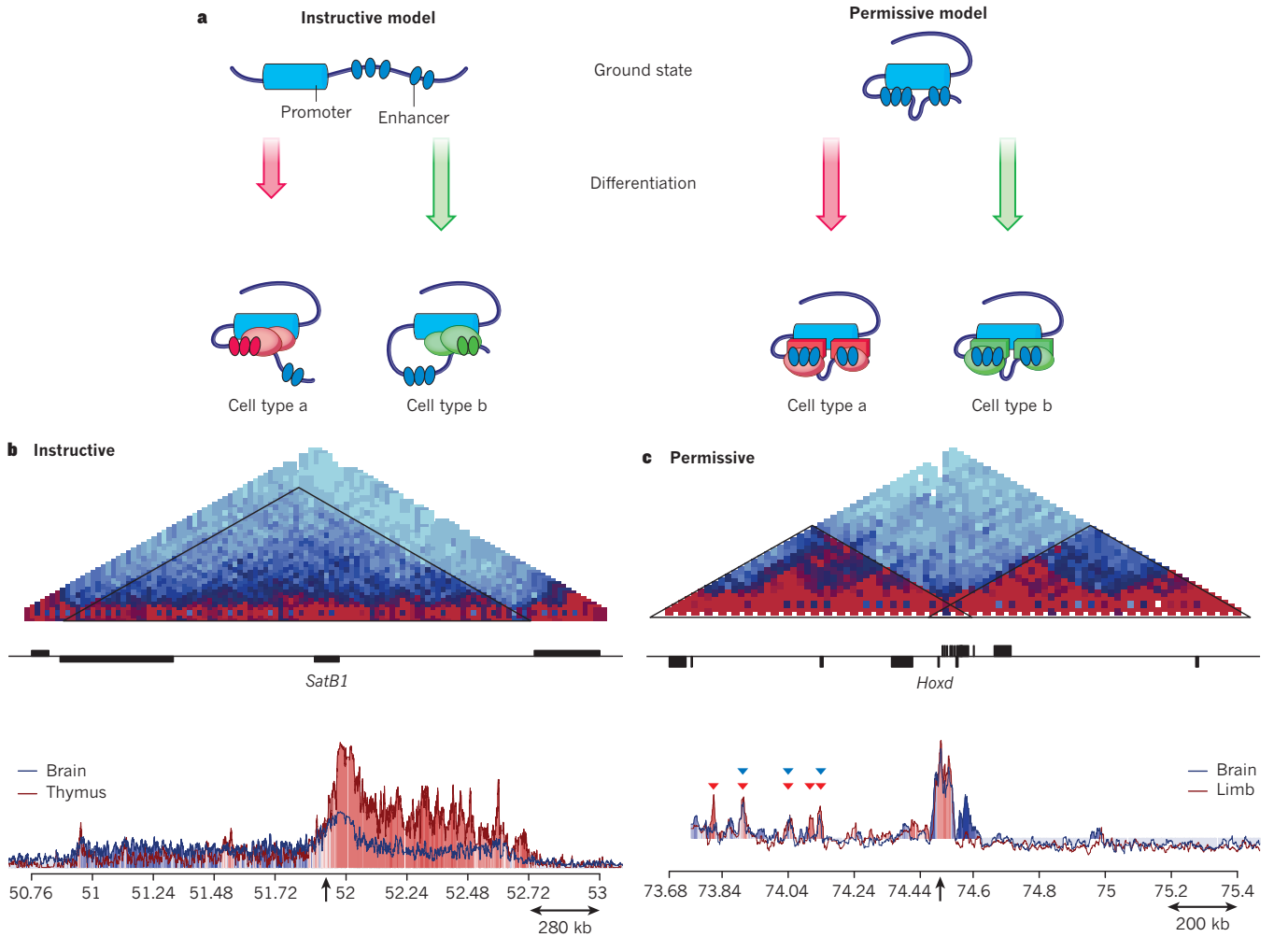

Figure 3 | Comparison of instructive and permissive model for three-dimensional controlled gene expression during differentiation. **a,** In the instructive model, tissue-specific enhancer–promoter contacts are formed *de novo* during differentiation, depending on the available transcription factors acting on the locus. In this view, no particular three-dimensional (3D) structure is formed in those cells in which the gene is inactive. In the permissive model, a preformed (ground-state) structure already exists in progenitor cells, formed either by selective promoter–enhancer interactions or by intrinsic properties of the chromatin domain. On differentiation, transcriptional activation requires merely the additional binding of tissue-specific transcription factors, which will take advantage of the configuration for immediate and robust gene activation. **b**, The *SatB1* locus is an example of an instructive regulatory operation. TADs, identified by Hi-C (top) and site-specific contact profiles (4C data) for the *SatB1* gene promoter (arrow) (bottom). 4C reveals that robust contacts between the gene and sites across the flanking 800 kb TAD are exclusively established in thymocytes (red) that highly express SatB1, and not in brain cells (blue) that poorly express the gene. **c**, *Hoxd* genes are an example of a permissive regulatory operation. In limb cells (red), Hoxd13 exclusively contacts the gene desert inside the TAD on the left, with particular interactions that involve five regulatory islands (red arrowheads). The same region, and three out of five, islands are also contacted in brain cells (blue arrowheads) in which the gene is not expressed, indicating a preformed domain that only requires subtle structural changes to support transcription in the limb (adapted from refs 35, 40, 45, 53).

Recruiting regulatory landscapes

Systematic deletions *in vivo* of potential enhancer sequences from a regulatory landscape have been done in less than a handful of studies. Although the loss of the 20 kb LCR reduced β-globin gene transcription 25- to 100-fold, the deletion of any of its five individual enhancer modules only reduced expression by a factor of 1.03–1.7 (ref. 63). These individual sites thus seem to collaborate, probably by aggregating into a single active chromatin hub[31], which would interact with one target gene at a time[64]. At the α-globin gene cluster a single enhancer, located 35 kb away in the intron of a housekeeping gene, accounts for 95% of the elevated α-globin transcript levels. Three other enhancers do form physical contacts but seem to be genetically redundant. Only when they are deleted along with the main enhancer do they abolish the remaining 5% of transcription[37].

Multiple enhancers have also been described in gene deserts adjacent to the *HoxD* gene cluster. *Hox* genes encode transcription factors that have key roles during the patterning of the various body axes in bilateral animals. In vertebrates, long-range enhancers have evolved within the flanking gene deserts to accompany the emergence of vertebrate-specific features such as the appendicular skeleton. Analyses of deletions *in vivo* and 4C experiments have revealed that the digit enhancers in the centromeric desert form a regulatory archipelago — a set of islands with contacts between themselves and with the target genes. These regulatory islands complement each other to reach the final transcriptional outcome, both in the quantity of transcripts and in their spatial distribution. Some — but not all — of these contacts are maintained when the enhancers are inactive and hence they form a permissive background configuration[35,53] (Fig. 3c). A comparable situation was observed at the opposite telomeric gene desert, in which several enhancers scattered throughout 1 Mb of DNA regulate their target *Hoxd* genes in the developing forearm. Therefore, *Hoxd* genes physically recruit different and preformed regulatory landscapes at different times and in different cells, initially to allow forearm construction and subsequently, to help form digits[35].

TADs as moulds for enhancer–promoter contacts

How are regulatory contacts coordinated in the 3D genome and what are the length scales over which enhancers can find their targets? Recent reports indicate that the regulatory landscapes mapped around the *Hoxd* and *Shh* loci match well with so-called topologically associating domains (TADs). TADs[45,65,66] were defined as chromosomal regions within which sequences preferentially contact each other, based on genome-wide interaction maps generated by Hi-C[67]. These domains, which are conserved among mammalian species but not found in yeast, are about 1 Mb and are separated by boundary regions that often contain CTCF-binding sites, housekeeping genes, transfer RNA genes or short interspersed elements[45,65,66].

The *Shh* gene for instance falls into a TAD, which precisely spans the region extending from the gene to its most distal ZRS enhancer. ZRS–promoter contacts are therefore contained within, or secured by, the overall 3D shape of the chromosomal domain. Likewise, both regulatory landscapes flanking the *Hoxd* gene cluster exactly match two topological domains. In this case, the *Hox* gene cluster itself lies between these two TADs[45], and the genes located right at the boundary are capable of switching their contacts from one TAD to the other[35], suggesting that enhancers acting over a given gene or set of genes may not always be restricted to a single TAD. In this view, TADs may sometimes also be used as large units of tissue-specific transcription.

Are TADs a cause or an effect?

TADs are, for the most part, already formed in embryonic stem cells and therefore seem to exist regardless of the transcription status[45,49]. However, recent topology maps at higher resolution revealed extensive structural reorganization at the sub-megabase scale during differentiation[68], with distinct hierarchical roles for different architectural proteins. At the 0.1 to 1 Mb scale, CTCF and cohesin may anchor constitutive interactions around development-specific genes. Below 100 kb, mediator, a protein complex conventionally associated with enhancer activity, and cohesin might cooperate to bridge tissue-specific enhancer–promoter interactions[68].

Topological domains may reflect an "inherent property of the mammalian genome"[45] that could help to limit the distance over which enhancers operate, providing a mould for enhancer–promoter interactions to occur and thus limiting undesired bystander effects[49]. Alternatively, TADs might merely result from pervasive enhancer–promoter interactions and thus illustrate the existence of a pre-regulatory genome, comprised of poised structures and lacking the final tissue-specific factor to become active. Genetic approaches to preturb TADs by means of chromosomal rearrangements[35,65] will help us both to understand the mechanisms underlying TAD formation and discriminate between these hypotheses.

Very long-range regulation

So far, little evidence exists for mammalian enhancers to act beyond the few megabase scale (that is, more than the TAD organizational level). Although circumstantial evidence supports inter-chromosomal gene regulation[69], it has rarely been verified genetically and whenever such observations were complemented by the deletion of enhancers *in vivo*, only neighbouring genes located on the same chromosome were affected[70]. Mammalian *trans*-activation was, however, seen in transgenic mice carrying an orphan LCR, which was able to activate its natural β-globin target gene on another chromosome. This nevertheless happened through fortuitous inter-chromosomal contacts, made in only a subset of cells, which consequently displayed elevated levels of globin expression[34]. Although artificial, this experiment shows what could probably be expected from a genome that, beyond the level of TADs, is structured in a probabilistic manner with the overall shape and relative location of chromosomes being different from cell to cell: productive inter-chromosomal enhancer–promoter interactions may exist but are likely to result in variegated expression. For pan-cellular expression control, enhancers therefore seem best positioned in *cis*, within domains of preferred chromosomal contacts, such as TADs.

Perspectives

High-throughput technologies for the identification of regulatory sequences have provided us with a wealth of information about the regulatory potential of our genome. The difficult task ahead will be to functionally connect genes to regulatory sequences and establish the relevance of their interactions. The systematic application of functional screening methods such as STARR-Seq[27] to a wide range of mammalian cell types and tissues may thus help us to distinguish between the millions of sites that have been identified so far by sequence- and chromatin-based enhancer screening methods. The enhancer classifications currently used (for example, 'poised' when displaying H3K4me1 only, or 'active' when showing both H3K4me1 and H3K27ac marks) are admittedly incomplete and will have to be refined based on their regulatory capacities. Are enhancers strictly tissue-specific or are they sequences carrying tissue-invariant enhancer activity? What causes the occlusion[71], in a natural context, of regulatory sites that harbour STARR activity in the same cell type?

In addition, although a given DNA sequence may be either suspected (by DNaseI profiling or ChIP-seq), or shown (by STARR-Seq), to have a regulatory capacity, its physiological relevance will have to be established in the appropriate developmental context. As a prerequisite, high resolution contact maps based on ultra-deep sequencing of Hi-C data can be expected for all relevant developmental cell types, which will allow us to physically connect potential regulatory sequences to target genes. Such detailed contact maps will help us to uncover the hierarchical folding principles of our chromosomes and clarify the degree of developmental conservation of structural domains like TADs and sub-TADs across the entire genome. It may also provide a topological framework for long-range enhancer–promoter contacts and might even inform us about those constraints underlying inter-species syntenic DNA segment conservation (see ref. 49). Although current Hi-C and 4C strategies are well equipped to perform these tasks, improvements will be necessary both to computational analysis tools and to downscaling the experimental material, whenever scarce cellular populations will be considered in the small developing mammalian embryo.

Once a particular DNA sequence is found in an open, nucleosome-depleted configuration, bound by tissue-specific transcription factors, displaying STARR-Seq activity and capable of establishing physical contacts with an endogenous gene, the question still remains as to whether this site controls the developmental expression of the target gene. Genetic approaches whereby all the parameters listed above can be perturbed in an ontogenic context will be necessary. Only a few complex regulatory landscapes have been dissected *in situ* so far and they have shown different aspects of enhancers with different recruitment strategies, different means of target gene selection and different ways of cooperating with other regulatory sites. The application of new strategies for site-directed genome editing[72,73] may help to clarify how regulatory sites and genes functionally orchestrate developmental gene expression programs and, consequently, how failures in regulatory wiring can cause diseases. However, genetics analysis on its own will not easily answer these questions because many components of developmental pathways are notoriously redundant. In addition, the developing mammalian embryo is very efficient at implementing compensatory mechanisms, whenever sequence modifications are induced. As a result, current analytical tools will have to be streamlined such that the impact of a genetic manipulation can be investigated as exhaustively and objectively as possible. ■

Received 2 June; accepted 2 September 2013.

1. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27,** 299–308 (1981).
2. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144,** 327–339 (2011).
3. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nature Rev. Genet.* **13,** 613–626 (2012).
4. Dunham, I. *et al*. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
5. Stamatoyannopoulos, J. A. What does our genome encode? *Genome Res.* **22,** 1602–1611 (2012).
6. Maurano, M. T. *et al*. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337,** 1190–1195 (2012).
7. Kleinjan, D. A. & van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76,** 8–32 (2005).
8. Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature Rev. Genet.* **14,** 125–138 (2013).
9. Montavon, T. & Duboule, D. Landscapes and archipelagos: spatial organization of gene regulation in vertebrates. *Trends Cell Biol.* **22,** 347–354 (2012).
10. Lagha, M., Bothma, J. P. & Levine, M. Mechanisms of transcriptional precision in animal development. *Trends Genet.* **28,** 409–416 (2012).
11. Maeda, R. K. & Karch, F. Gene expression in time and space: additive *vs* hierarchical organization of *cis*-regulatory regions. *Curr. Opin. Genet. Dev.* **21,** 187–193 (2011).
12. Levine, M. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20,** R754–R763 (2010).
13. Duboule, D. & Wilkins, A. S. The evolution of 'bricolage'. *Trends Genet.* **14,** 54–59 (1998).
14. Kirschner, M. & Gerhart, J. Evolvability. *Proc. Natl Acad. Sci. USA* **95,** 8420–8427 (1998).
15. Ohno, S. *Evolution by Gene Duplication*. (Springer, 1970).
16. Ruf, S. *et al*. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nature Genet.* **43,** 379–386 (2011).
In this paper, the authors use a transposable reporter gene system in mice to probe for enhancer activity *in vivo* and show widely varying reporter expression patterns at hundreds of genomic integration sites.
17. O'Kane, C. J. & Gehring, W. J. Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl Acad. Sci. USA* **84,** 9123–9127 (1987).
18. Pennacchio, L. A. *et al*. *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444,** 499–502 (2006).
19. Odom, D. T. *et al*. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genet.* **39,** 730–732 (2007).
20. Wilson, M. D. *et al*. Species-specific transcription in mice carrying human chromosome 21. *Science* **322,** 434–438 (2008).
21. Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* **462,** 65–70 (2009).
22. Lee, D., Karchin, R. & Beer, M. A. Discriminative prediction of mammalian enhancers from DNA sequence. *Genome Res.* **21,** 2167–2180 (2011).
23. Gorkin, D. U. *et al*. Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Res.* **22,** 2290–2301 (2012).
24. Crawford, G. E. *et al*. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature Methods* **3,** 503–509 (2006).
25. Thurman, R. E. *et al*. The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 1 (2012).
26. Hardison, R. C. & Taylor, J. Genomic approaches towards finding *cis*-regulatory modules in animals. *Nature Rev. Genet.* **13,** 469–483 (2012).
27. Arnold, C. D. *et al*. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339,** 1074–1077 (2013).
In this paper, the authors report a genome-wide screening method that quantifies the self-transcription of random DNA segments to identify sequence modules with enhancer activity.
28. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).
29. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26,** 11–24 (2012).
30. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. & Fraser, P. Long-range chromatin regulatory interactions *in vivo*. *Nature Genet.* **32,** 623–626 (2002).
31. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* **10,** 1453–1465 (2002).
32. Palstra, R. J. *et al*. Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PloS ONE* **3,** e1661 (2008).
33. Deng, W. *et al*. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149,** 1233–1244 (2012).
In this study, it is shown that the artificial physical tethering of a distant enhancer to a gene promoter leads to highly activated transcription, illustrating that enhancer contacts causally underlie transcriptional output.
34. Noordermeer, D. *et al*. Variegated gene expression caused by cell-specific long-range DNA interactions. *Nature Cell Biol.* **13,** 944–951 (2011).
This article shows that inter-chromosomal gene activation by an ectopic enhancer leads to variegated gene expression in subsets of mouse cells.
35. Andrey, G. *et al*. A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science* **340,** 1234167 (2013).
In this paper, the authors show that different regulatory landscapes are recruited during development by the same genes, which switch contacts from one to the other.
36. Palstra, R. J. *et al*. The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genet.* **35,** 190–194 (2003).
37. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* **26,** 2041–2051 (2007).
38. Drissen, R. *et al.* The active spatial organization of the β-globin locus requires the transcription factor EKLF. *Genes Dev.* **18,** 2485–2490 (2004).
39. Vakoc, C. R. *et al*. Proximity among distant regulatory elements at the β-globin locus requires GATA-1 and FOG-1. *Mol. Cell* **17,** 453–462 (2004.12.028 (2005).
40. van de Werken, H. J. *et al*. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods* **9,** 969–972 (2012).
41. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75,** 1401–1416 (1993).
42. Lettice, L. A. *et al*. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12,** 1725–1735 (2003).
43. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range *cis*-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132,** 797–803 (2005).
44. Amano, T. *et al*. Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16,** 47–57 (2009).
45. Dixon, J. R. *et al*. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
46. Lettice, L. A. *et al*. Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev. Cell* **22,** 459–467 (2012).
47. Eijkelenboom, A. *et al*. Genome-wide analysis of FOXO3 mediated transcription regulation through RNA polymerase II profiling. *Mol. Syst. Biol.* **9,** 638 (2013).
48. Melo, C. A. *et al*. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* **49,** 524–535 (2012).
49. Nora, E. P., Dekker, J. & Heard, E. Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? *BioEssays* **5,** 201300040 (2013).
50. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489,** 109–113 (2012).
This report is the first systematic analysis, based on 5C, of promoter-centred DNA contacts, revealing complex and dynamic contact networks, which change during development.
51. Li, G. *et al*. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
52. Splinter, E. *et al*. CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev.* **20,** 2349–2354 (2006).
53. Montavon, T. *et al*. A regulatory archipelago controls *Hox* genes transcription in digits. *Cell* **147,** 1132–1145 (2011).
In this paper, the authors report the existence of a regulatory archipelago that controls transcription in developing digits and contains several islands with various quantitative and qualitative contributions.
54. Bender, M. A. *et al.* Flanking HS-62.5 and 3-HS1, and regions upstream of the LCR, are not required for β-globin transcription. *Blood* **108,** 1395–1401 (2006).
55. Noordermeer, D. *et al.* Transcription and chromatin organization of a housekeeping gene cluster containing an integrated β-globin locus control region. *PLoS Genet.* **4,** e1000016 (2008).
56. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24,** 530–542 (2013).
57. Cajiao, I., Zhang, A., Yoo, E. J., Cooke, N. E. & Liebhaber, S. A. Bystander gene activation by a locus control region. *EMBO J.* **23,** 3854–3863 (2004).
58. Spitz, F., Gonzalez, F. & Duboule, D. A global control region defines a chromosomal regulatory landscape containing the *HoxD* cluster. *Cell* **113,** 405–417 (2003).
59. Zuniga, A. *et al*. Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev.* **18,** 1553–1564 (2004).
60. Lower, K. M. *et al*. Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proc. Natl Acad. Sci. USA* **106,** 21771–21776 (2009).
61. Tschopp, P., Fraudeau, N., Bena, F. & Duboule, D. Reshuffling genomic landscapes to study the regulatory evolution of Hox gene clusters. *Proc. Natl Acad. Sci. USA* **108,** 10632–10637 (2011).
62. De Gobbi, M. *et al*. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science* **312,** 1215–1217 (2006).
63. Bender, M. A. *et al.* Targeted deletion of 5-HS1 and 5-HS4 of the β-globin locus control region reveals additive activity of the DNaseI hypersensitive sites. *Blood* **98,** 2022–2027 (2001).
64. Wijgerde, M., Grosveld, F. & Fraser, P. Transcription complex stability and chromatin dynamics *in vivo*. *Nature* **377,** 209–213 (1995).
65. Nora, E. P. *et al*. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
66. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148,** 458–472 (2012).
Refs 45, 65 and 66 report that mammalian and fly chromosomes are structurally subdivided into topological domains with flanking boundaries that hamper DNA contacts across.

67. Lieberman-Aiden, E. *et al*. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
68. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153,** 1281–1295 (2013).
69. Williams, A., Spilianakis, C. G. & Flavell, R. A. Interchromosomal association and gene regulation in trans. *Trends Genet.* **26,** 188–197 (2010).
70. Fuss, S. H., Omura, M. & Mombaerts, P. Local and *cis* effects of the H element on expression of odorant receptor genes in mouse. *Cell* **130,** 373–384 (2007).
71. Gaetz, J. *et al*. Evidence for a critical role of gene occlusion in cell fate restriction. *Cell Res.* **22,** 848–858 (2012).
72. Miller, J. C. *et al*. A TALE nuclease architecture for efficient genome editing. *Nature Biotechnol*. **29,** 143–148 (2011).
73. Jinek, M. *et al*. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
74. Kleinjan, D. J. & Coutinho, P. *Cis*-ruption mechanisms: disruption of *cis*-regulatory control as a cause of human genetic disease. *Brief. Funct. Genomic Proteomic* **8,** 317–332 (2009).
75. Ovcharenko, I. *et al*. Evolution and functional classification of vertebrate gene deserts. *Genome Res.* **15,** 137–145 (2005).
76. Ahituv, N. *et al*. Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* **5,** e234 (2007).

Acknowledgements The authors thank Y. Oz for preparing the figures, and G. Geeven, M. Janssen, G. Andrey, L. Beccari, T. Montavon and M. Leleu for their help. The authors' laboratories are supported by funds from the Dutch Scientific Organization (NWO) (714012002 and 724012003 (VICI)), EU grant 2010-259743 (MODHEP), a KWF Dutch Cancer Foundation grant (2009-4459) and a NanoNextNL grant (to W.dL) and the Ecole Polytechnique Fédérale de Lausanne, the University of Geneva, the Swiss National Research Foundation, the ERC grant SystemsHox.ch and the FP7 program IDEAL (to D.D.).

Author Information Reprints and permissions information is available at www.nature.com/reprint. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/cwhy4j. Correspondence should be addressed to W.dL (w.delaat@hubrecht.eu) or D.D. (denis.duboule@epfl.ch).