THE ROLE OF CHROMATIN STRUCTURE IN REGULATING THE EXPRESSION OF CLUSTERED GENES

Duncan Sproul, Nick Gilbert and Wendy A. Bickmore

Abstract | Much of what we know about the chromatin-based mechanisms that regulate gene expression in mammals has come from the study of what are, paradoxically, atypical genes. These are clusters of structurally and/or functionally related genes that are coordinately regulated during development, or between different cell types. Can unravelling the mechanisms of gene regulation at these gene clusters help us to understand how other genes are controlled? Moreover, can it explain why there is clustering of apparently unrelated genes in mammalian genomes?

Chromosomes and Gene Expression Section, Medical Research Council Human Genetics Unit, Crewe Road, Edinburgh EH4 2XU, United Kingdom.
Correspondence to W.A.B.
e-mail: W.Bickmore@hgu.mrc.ac.uk
doi:10.1038/nrg1688
Published online
9th September 2005

The sequencing of whole genomes has introduced significant changes in biology, moving from the study of individual genes and proteins towards the investigation of pathways and systems. The same genome-wide approach is now required to understand the mechanisms that regulate gene expression. An individual gene and its regulatory elements should not be considered in isolation, but in the context of their effect on neighbouring genes, and *vice versa*. Gene order in eukaryotic genomes is not random and the co-expression of clustered genes has been reported for several organisms including yeast, *Drosophila melanogaster* and *Caenorhabditis elegans*[1]. Because gene and genome sizes differ among organisms, and these parameters affect gene clustering, we confine ourselves here to the analysis of mammalian genes and gene clusters.

There have been extensive studies of how chromatin structure can control the expression of individual genes. However, chromatin structure is also an obvious candidate for simultaneous regulation of multiple genes. In this review we consider how long-range regulatory elements, histone modifications and higher-order chromatin structures might each coordinate the expression of genes in clusters. The recent development of large-scale and genome-scale approaches to the detailed study of these chromatin features makes it especially apposite to investigate the influence of chromatin structure across genomic regions. By exploring what we know about chromatin-based mechanisms that regulate the gene clusters which have arisen by tandem gene duplication, we ask whether any of these mechanisms can explain why apparently unrelated genes are clustered in mammalian genomes.

Classes of mammalian gene clusters

It has long been recognized that some genes that are related by primary sequence and/or function can be found adjacent to one another in the mammalian genome. However, it is only with the sequencing of the genome, and the annotation of gene function, that the evolutionary origin and biological function of gene clusters can be understood.

Temporally regulated clusters of structurally related genes. The β-globin and homeobox (Hox) loci are models for the coordinate regulation of clustered genes that encompass 50–250 kb. Both loci arose through tandem duplications of ancestral genes. Selection presumably then kept the paralogous genes together, which implies that there is a functional link between the clustering of these genes in the genome and the mechanisms that regulate their expression.

Box 1 | Structure and function of the MHC

The classical major histocompatibility complex (MHC) covers approximately 3.6 Mb of chromosome 6p21.3; one of the most gene-rich regions of the human genome[13,14]. It contains many clusters of structurally unrelated genes that encode proteins involved in innate and adaptive immunity — for example, human leukocyte antigens, IMMUNO-PROTEOSOME components and peptide transporters — as well as genes that have no obvious immunological function.

MHC class I genes are almost ubiquitously expressed. The most gene-dense region of the MHC is the class III region[13]. Its genes are expressed in various cell types, including B cells. Expression of the class II region is restricted to antigen-presenting cells and thymic epithelial cells (FIG. 1). However, transcription of the MHC class II genes can be strongly induced in other cell types by cytokines such as interferon-γ (IFNG)[42]. Within the MHC there are also genes that have different expression patterns, such as the stem cell transcription factor *OCT4* (also known as *POU5F1*).

Clustering of globin genes on one chromosome is not essential for ensuring their expression in the appropriate cell type. Mammalian α-globin and β-globin genes are located on different chromosomes, but both must be co-expressed at the necessary levels in the same erythroid cells to ensure the correct production of haemoglobin. It is thought that the ancestral globin locus was an α–β locus, and that subsequent genome duplication allowed for separate evolution of α- and β-clusters. There are relics of such an α–β cluster in marsupials[2] and fish[3]. However, with both α- and β-clusters, what has generally remained inviolable is the COLLINEARITY between the gene order and the temporal sequence of gene activation: embryonic, fetal (in the case of β-globin) and adult. Interestingly, the organization of the chicken β-globin locus is different. The first genes (ρ, β^H, β^A) after the upstream DNaseI hypersensitive sites (DHSs) are expressed sequentially in development (early embryonic, hatching and adult), whereas the last gene of the cluster (ε) is only expressed in early embryos. This different regulation might be the result of a novel regulatory enhancer that has been placed between the β^A and ε genes[4].

Likewise, it seems that Hox-gene clustering is more important for the temporal regulation of expression during development than for spatial patterning of expression along the anterior–posterior axis of the embryo. Hox transgenes that are inserted at ectopic locations in the genome often show the correct spatial expression pattern but lose their temporal regulation[5]. These observations are also supported by the analysis of Hox-gene clustering in two UROCHORDATES. The Hox complex is partially dispersed in the *Ciona intestinalis* genome; although the anterior–posterior expression of Hox genes is spatially controlled along the larval neural tube, temporal coordination seems to be lost[6]. Even more strikingly, in *Oikopleura dioica* the Hox genes are completely dispersed around the genome. Whereas the temporal progression of their activation is lost[7,8], the anterior–posterior patterning of Hox expression has been maintained.

For both Hox and globin loci, all the genes are transcribed from the same strand. Although the functional significance of this is unclear, it might point to a role for strand-specific intergenic transcription in the temporal regulation of these clusters[9,10].

Superclusters of different co-regulated gene families. There are also larger superclusters of mammalian genes that extend over many megabases. Examples are the major histocompatibility complex (MHC) (BOX 1) and the epidermal differentiation complex (EDC) (BOX 2). Although genes in these clusters are not necessarily all structurally related to each other, they can be coordinately regulated. Apart from transcriptional co-regulation, another selection pressure on MHC clustering is the maintenance of haplotypes in LINKAGE DISEQUILIBRIUM[11]. However, it is difficult to understand how this might be a factor for maintaining the EDC, because it is not involved in immune functions.

The MHC is believed to have evolved from an ancestral region that predates chordates[12], but the splitting of class I and II MHC genes in the genomes of TELEOSTS argues against an absolute requirement for their linkage in *cis*[13]. On the other hand, an extended mammalian MHC has now been defined; it stretches far beyond the classical MHC and includes clusters of olfactory receptor (OR), tRNA and histone-coding genes, the functions of which are not obviously linked to those of the immune system[14] (FIG. 1). Either this region of the genome is especially prone to gene duplication, providing a fertile environment within which

Box 2 | Structure and function of the EDC

The epidermal differentiation complex (EDC) spans ~2 Mb of the human chromosome 1q21.3 and contains genes of at least three families that are involved in the terminal differentiation of the epidermis (FIG. 1). The first group consists of the S100A genes, which flank the EDC and encode calcium-binding regulatory proteins that are involved in signal transduction cascades[56]. The second group encodes precursors of the cornified envelope that are expressed in terminally differentiated keratinocytes. These include small proline-rich (SPRR) proteins, late envelope proteins, loricrin, involucrin and NICE (newly identified cDNA from the EDC)[57,58]. Finally, there are genes that encode so-called 'fused' proteins, which combine structural features of the other two groups. These proteins are associated with keratin intermediate filaments and are partially crosslinked to the cornified envelope[59,60]. There are also at least three other genes that do not seem to have direct roles in the epidermis.

Little is known about the control of EDC expression. Studies on individual gene promoters have shown that widely expressed transcription factors are involved in regulating the EDC genes[61]. In addition, clusters of keratinocyte-specific DHSs have been shown to lie within the SPRR cluster[62]. However, the coordinated upregulation of SPRR genes can occur independently of the neighbouring involucrin and loricrin-coding genes[62]. Other key components of the epidermis, such as keratins, are not encoded within the EDC. Both keratins and EDC genes are co-regulated by the Aire transcription factor[63]. The function of gene clustering within the EDC still remains unclear.

IMMUNO-PROTEASOME
A proteasome complex that degrades proteins into peptides for presentation with MHC class I molecules.

COLLINEARITY
The correspondence between the linear order of genes on the chromosome and the sequential order of their expression.

UROCHORDATES
A subphylum of Chordata that are also known as tunicates. They have a notochord during their early stages of development.

LINKAGE DISEQUILIBRIUM
The non-random association of alleles at adjacent loci along a chromosome.

TELEOSTS
A taxonomic group that comprises most extant bony fishes.

PML NUCLEAR BODIES
Sub-nuclear compartments that are defined by the presence of the PML (promyelocytic leukaemia) protein. They have been associated with diverse nuclear functions including transcription, DNA repair, viral defence, stress, cell-cycle regulation, proteolysis and apoptosis.

LORICRIN
The predominant protein of the cornified envelope in keratinocytes, which is encoded by a gene in the EDC.

FILAGGRIN
A protein that is involved in aggregating keratin during the terminal differentiation of epidermal keratinocytes, and is encoded by a gene in the EDC.

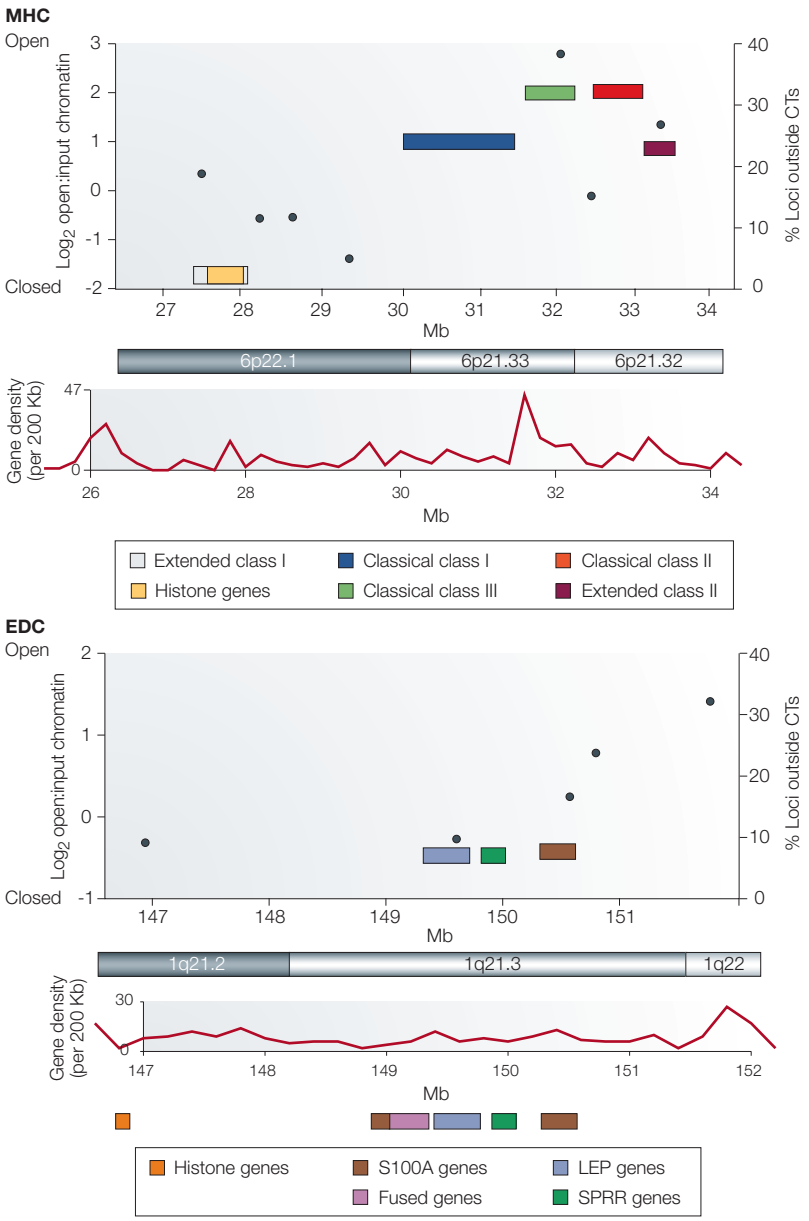

Figure 1 | **Chromatin structure at the MHC and EDC superclusters.** The graphs show the chromatin-fibre structure, expressed as the $\log_2$ hybridization ratio between purified open chromatin and total input chromatin ($\log_2$ open:input chromatin), which is determined on BAC clones (filled circles) from these regions. Open chromatin-fibre structure is designated by values that are >0 on the left-hand *y* axis[48]. The percentages of hybridization signals (right-hand *y* axis) that are located outside chromosome territories (CTs) 6 and 1 are represented by the coloured bars[20,47]. Both data sets were obtained from human lymphoblastoid cells. The map positions (Mb) were determined from the National Center for Biotechnology Information build 35.1. Different groups of genes from each supercluster are indicated by coloured boxes. EDC, epidermal differentiation complex; LEP, late envelope protein; MHC, major histocompatibility complex; SPRR, small proline-rich.

gene clusters could arise, or there is some selective advantage to maintaining this extended chromosomal region of gene clusters. Similar to the classical MHC, tRNA and histone-coding genes represent regions of high transcriptional activity. There is also a spatial proximity of both the MHC and histone clusters to PML NUCLEAR BODIES, but the functional significance of this remains unclear because disruption of PML bodies (by knockdown of the PML protein) has no significant consequence for the transcription of the MHC genes[15]. By contrast, OR genes are transcriptionally silent in most cell types and, even in olfactory cells, only one allele of one gene from the entire OR gene family is expressed in a particular cell[16]. Therefore, it is difficult to postulate how there could be a common transcriptional link between OR genes and the MHC. Intriguingly, the β-globin locus is also embedded within a cluster of OR genes, and class II MHC and OR genes are interspersed in the mouse[17].

The MHC has been duplicated several times during evolution, giving rise to paralogous regions on other human chromosomes[14], including 1q21–1q25 — the location of the EDC[18]. Similar to the extended MHC, this region also contains clusters of histone, OR and tRNA-coding genes (FIG. 1). However, the EDC itself seems to be a recent mammalian addition. Although the flanking regions are conserved in chicken, frog and fish, there is no detectable equivalent to the EDC itself in the currently annotated genomes of these species. However, proteins that are similar to LORICRIN and FILAGGRIN are present in the epidermis of birds[19]. Because the genes of the EDC are not structurally or functionally related to those of the classical MHC, the EDC must have arisen independently through a new round of gene duplication within a chromosomal region that is distantly related to the extended MHC.

Even paralogous, co-expressed genes within the MHC and EDC are not necessarily transcribed from the same strand and, to our knowledge, there is no evidence for a temporal sequence of activation of MHC genes during development. There is, however, some temporal regulation between gene families of the EDC. S100A genes, which encode calcium-binding regulatory proteins, are expressed first — in the basal cell layer of the epidermis. In the SUPRABASAL LAYERS of the skin there is a progressive expression of genes that encode structural components of the CORNIFIED ENVELOPE; including loricrin, INVOLUCRIN, small proline-rich (SPRR) proteins, the so-called 'fused' proteins and finally the late envelope proteins.

General clustering of unrelated genes. Although most mammalian genes are not arranged into the clusters and superclusters of related genes that are described above, they are not randomly distributed along chromosomes. Even before the sequencing of the human genome, it was clear that most 'housekeeping' genes are clustered into specific domains on human chromosomes[21]. This issue has now been rigorously addressed by bioinformatic analysis of genome sequences and through the study of gene-expression patterns[1]. Early suggestions were that there is clustering of genes expressed in a specific tissue or clustering of genes with high transcription rates[22,23]. However, it is now clear that the common feature of the genes that are clustered in the most gene-dense regions of human chromosomes is their expression in a wide range of tissues and cell types[1,24]. Because of their prevalence in expression databases, broadly expressed genes also tend to be categorized as being highly expressed.

SUPRABASAL LAYERS
Layers of progressively differentiating keratinocytes that are found above the basal layer of stem cells.

Could this effect be purely coincidental? Genes might be switched on just because of their proximity to active genes. Such a 'bystander effect' has been shown for the *CD79B* gene, which is located between the human growth hormone cluster and its locus control region (LCR) on chromosome 17 (REF. 25). *CD79B* is expressed in the pituitary, although its function is regarded as B-cell-specific. If this is a generic feature of clustered genes then there is no *a priori* reason why gene clusters would be conserved. However, examination of breaks in SYNTENY between mouse and human indicate that gene-rich clusters of broadly expressed genes are under functional selection[26].

We can therefore conclude that some long-range feature of chromatin or chromosome structure is influencing the expression of clustered genes. So, what could be the mechanistic basis for this regulation?

Mechanisms of coordinate gene regulation

Sharing regulatory elements. One reason to keep genes clustered together on a chromosome is if they share a common regulatory element. Efficient β-globin transcription requires an LCR, located ~10 kb upstream of the first gene in the cluster — the embryonic ε-globin DHSs that are located in the LCR, and more to the 5′ and 3′ end of the locus, are physically close to each other in the nucleus[27] and can be crosslinked to one another[28]. This is thought to form the base of a looped domain to which globin genes are then sequentially recruited when they are expressed. This chromatin architecture is erythroid-specific and depends on erythroid transcription factors[29,30]. Most importantly, it is established during development before the transcription of the β-globin genes. In murine progenitor erythrocytes that are committed to, but not yet initiating, globin expression there is spatial association of some of the DHSs and part of the LCR, but not of the globin genes themselves. The genes are then sequentially recruited to the loop hub as they become expressed — the embryonic genes in primitive embryonic erythrocytes, and the adult β-globin genes in definitive erythroid cells from the fetal liver[31].

Does a similar mechanism operate at Hox loci? There are localized control elements and enhancers within the Hox clusters that function over a few neighbouring genes, but only in the case of the homeobox D cluster, *Hoxd*, is there evidence for global regulatory elements that are located outside the clusters. A global control region (GCR) that lies to the 5′ end of the *Hoxd* cluster regulates the expression of 5′ *Hoxd* genes in the distal limb bud[32], but its mode of action is still unknown. Unlike the β-globin LCR and DHSs, it also regulates genes (even-skipped homeobox homologue 2, *Evx2*, and lunapark, limb and neural patterns, *Lnp*) that are unrelated to Hox genes. It has been proposed that this GCR is a recent evolutionary innovation that was co-opted to the *Hoxd* locus at the same time as the appearance of limbs[32]. Therefore there is no *a priori* reason to expect that a similar element is necessary for collinear Hox expression along the main anterior–posterior embryonic axis, or for early temporal collinearity of expression. Recently it has been proposed that there is an early limb control region (ELCR) at the 3′ end of the *Hoxd* cluster and that it controls anterior–posterior collinear expression[33].

Long-range regulatory elements that exert their influence over superclusters, or even clusters of gene families within superclusters, have not been found. DHSs that stretch up to 9 kb upstream of the MHC, class II, DRα (*HLA-DRA*) promoter have been classified as an LCR[34]. However there is no evidence that these elements exert any effect on other class II genes. Class II genes do share regulatory motifs that are located close to each of the individual gene promoters. These S–Y elements are the binding sites for the regulatory factor X complex (RFX) and for the MHC class II transactivator (CIITA, also known as MHC2TA)[35]. CIITA is a master regulator of class II expression. Induction of its expression by interferon-γ (IFNG) precedes that of MHC class II genes. However, the S–Y regulatory elements themselves cannot explain the clustering of class II genes in the MHC, because the MHC class-II-associated invariant chain gene *Ii* (*CD74*) is also regulated by the same element and transcription factors, although it is located on a different chromosome[35].

Histone modifications. The role of chromatin structure in modulating gene expression has been extensively studied at the level of histones and their post-translational modification. Among these modifications, histone acetylation is generally considered to be an indication of transcriptional activity. There is hyperacetylation of histones at the β-globin LCR, but the levels of histone acetylation at the globin genes themselves then depend on the stage of development. In the embryonic YOLK SAC, acetylated histones are found at the promoters of both active embryonic genes and the promoters of the inactive adult genes. Later, in the fetal liver, this modification is retained at the promoters of the adult genes, but is lost from the now inactive embryonic genes[36]. This is remarkably similar to the situation at the *Hoxb* cluster. Induction of the temporal programme of *Hoxb* expression during the differentiation of embryonic stem cells is accompanied by histone acetylation at early (3′) and late (5′) genes simultaneously. This acetylation is then lost from the earliest gene, homeobox B1 (*Hoxb1*), as its transcription ceases, but is retained at homeobox B9 (*Hoxb9*), which is expressed later during differentiation[37]. For both β-globin and *Hoxb* loci, histone acetylation could therefore be equated with a locus-wide potentiation for transcription. Subsequent gene-specific deacetylation might then be important for the sequential silencing of expression. Nonetheless, artificial induction of histone acetylation using histone deacetylase (HDAC) inhibitors is not sufficient to either activate transcription or open chromatin structure[36,37]. Methylation of histone H3 at lysine 4 (H3-K4) is also associated with the activation of gene expression, and the acquisition and subsequent loss of this mark at the *Hoxb* locus largely parallels histone acetylation[37].

CORNIFIED ENVELOPE
A tough protein–lipid structure that is formed under the plasma membrane of keratinocytes during their terminal differentiation.

INVOLUCRIN
A component of the cornified envelope.

SYNTENY
The preserved order of genes along a chromosome in related organisms.

YOLK SAC
The first site of blood formation in the mammalian embryo.

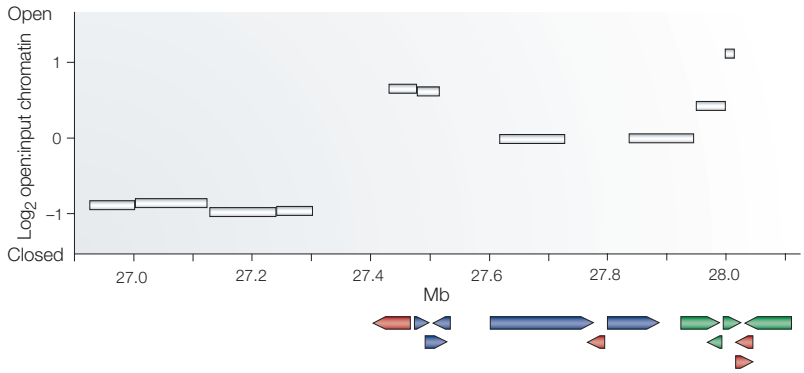

Figure 2 | **Relationship of chromatin-fibre structure to gene density and gene expression.** The graph shows the chromatin-fibre structure, expressed as the $\log_2$ hybridization ratio between purified open chromatin and total input chromatin ($\log_2$ open:input chromatin), for a region of human chromosome 22q in lymphoblastoid cells. Open chromatin-fibre structure is designated by values that are >0. Data are taken from REF. 48. The positions of known genes in this region and the directions of their transcription are shown below the graph. The expression status of these genes was determined in the same cells. Green represents genes that are expressed, red represents genes that are not expressed and blue represents genes for which expression is not determined.

Because chromatin immunoprecipitation on small numbers of cells is technically difficult, there has been only limited analysis of histone modifications at Hox loci along the anterior–posterior embryonic axis. Histone acetylation at the enhancer, but not the promoter, of homeobox D4 (*Hoxd4*) precedes the activation of transcription during embryogenesis. Histone acetylation is then observed at the gene itself in regions of the spinal cord that are posterior to the boundary of expression[38]. In a genome-wide study of histone modifications, an unusually wide-spread domain of H3-K4 methylation was observed at 3′ Hox genes in fibroblasts that had been derived from anterior regions of the animal[39].

CIITA associates with histone acetyltransferases (HATs) and has intrinsic HAT activity[40,41]. The domain of histone acetylation that is induced by CIITA binding to the MHC class II region is broad (up to ~16 kb) (REF. 42), but there is no evidence for locus-wide histone modifications that are similar to those observed at Hox loci[39].

It has been suggested that the 'bystander activation' of *CD79B* is a consequence of the spreading of histone acetylation and H3-K4 methylation out from the growth hormone LCR to the pituitary-specific *GH-N* growth hormone (also known as *GH1*) and thereby encompassing CD79B (REF. 25). There is no activation of *CD79B* in the placenta, where histone modifications are limited to the placental growth hormone genes themselves and do not occur in the intervening region[25,43]. This raises the possibility that the spread of histone modifications might be responsible for generalized gene activation across clusters of broadly expressed genes. However, broad domains of histone acetylation seem to be the exception rather than the rule in mammalian genomes. Chromosome-wide mapping of histone acetylation and H3-K4 methylation indicates that, with the exception of unusual regions such as Hox clusters, these marks of 'active chromatin' are generally confined to the promoters and regulatory elements of mammalian genes, and are not distributed over broad domains[39,44].

Large-scale alterations in higher-order chromatin structure. With no evidence for an overarching LCR-like element that coordinates the regulation of Hox loci, other than *Hoxd*, and with the inability of histone modification patterns to explain the temporal regulation of Hox expression, what other mechanisms could control collinearity? One mechanism that has been proposed, at least for the initial sequential activation of Hox loci, is a progressive 'opening' of chromatin structure that is propagated from the 3′ to 5′ end, or a progressive escape from a repressive chromatin environment[5]. Evidence that is consistent with this has been obtained at the *Hoxb* locus by interphase FLUORESCENCE *IN SITU* HYBRIDIZATION (FISH). On induction of transcription, the *Hoxb* locus decondenses and the *Hoxb* genes move out of the confines of their chromosome territory in a temporal sequence that is in synchrony with their expression[37,45].

Looping out from the chromosome territory has also been seen at the β-globin locus in erythroid cells, but there is no evidence that this is coordinated with the sequential activation of the embryonic and adult genes during development. This change in large-scale chromatin structure is suggested to correspond to a poised state, rather than transcriptional activation *per se*[46]. The LCR is implicated in this change in chromosome-territory architecture, but how this relates to the DHS-mediated β-globin looped domain discussed earlier remains to be investigated.

The study of the nuclear organization of the human MHC first brought to light the looping of specific genomic regions outside chromosome territories. In B-LYMPHOBLASTOID CELLS the classical MHC cluster shows a high frequency of looping out from its chromosome territory, most markedly at the class II and III regions[47]. This phenomenon is not seen for the extended MHC (FIG. 1). The frequency of extra-territory loops is low in non-ANTIGEN-PRESENTING CELLS (for example, fibroblasts), but can be rapidly induced by IFNG, which induces CIITA and class II expression. The EDC is also extruded from its chromosome territory in expressing cells, such as keratinocytes, but not in non-expressing lymphoblasts[20] (FIG. 1).

The extrusion of the MHC into loops outside its chromosome territory was suggested to represent a cytological manifestation of long-range chromatin decondensation[47]. Similarly, during the induction of *Hoxb* expression, the nuclear reorganization of *Hoxb* genes outside chromosome territories is also accompanied by cytologically detectable levels of chromatin decondensation[37,45]. We remain ignorant about how chromatin is folded up at the complex levels of chromatin organization beyond the nucleosome,

FLUORESCENCE *IN SITU* HYBRIDIZATION
This is a cytological technique that is used to detect and localize DNA sequences on chromosomes, or in nuclei, using fluorescent probes.

B-LYMPHOBLASTOID CELLS
Peripheral blood mononuclear cells that are transformed with the Epstein–Barr virus.

ANTIGEN-PRESENTING CELLS
Dendritic cells, macrophages and B cells. These cells express MHC class II genes, display foreign antigens that form complexes with MHC on their surfaces, and can activate T cells.

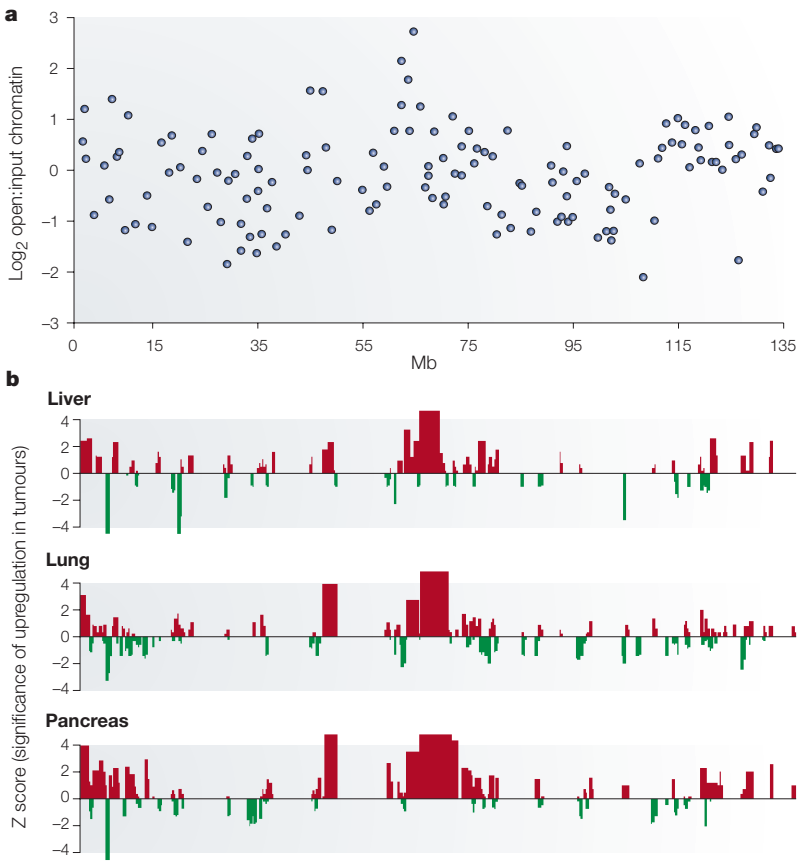

Figure 3 | **Correlation between chromatin-fibre structure and the transcriptional misregulation of genes in cancer. a** | Chromatin-fibre structure, expressed as the $\log_2$ hybridization ratio between purified open chromatin and total input chromatin ($\log_2$ open:input chromatin) for human chromosome 11 in lymphoblastoid cells. Open chromatin-fibre structure is designated by values that are >0 (REF. 48). **b** | The distribution of genes on human chromosome 11 that tend to be either upregulated (red), or downregulated (green) in liver, lung and pancreatic cancers. Adapted, with permission, from REF. 51 © (2003) American Association for Cancer Research.

but the biophysical compaction of 30-nm chromatin fibres, purified from B-lymphoblastoid cells, has been measured across the human genome by SUCROSE GRADIENT SEDIMENTATION[48]. The data from the analysis of the MHC and EDC are consistent with the extent of extra-territory looping: the MHC seems to have a mostly 'open' chromatin-fibre structure in lymphoblasts, whereas the chromatin structure of the EDC is 'compact' (FIG. 1).

These open chromatin fibres are not randomly dispersed across the rest of the human genome; they are clustered. Moreover, these domains of open chromatin structure correspond to the gene-dense clusters of broadly expressed genes[48,49]. FISH analysis also indicates that these domains have a decondensed higher-order structure that is often looped outside the corresponding chromosome territory[48]. Therefore the higher-order chromatin structure of these gene clusters looks remarkably similar to the activated state of *Hoxb* and the MHC supercluster.

Expression analysis shows that within the domains of open chromatin-fibre structure, individual genes can be inactive as well as active (FIG. 2). Therefore there is no 'bystander effect' of open chromatin-fibre structure on gene expression. Instead, we propose that constitutively open chromatin-fibre domains create an environment that is permissive, or poised, for gene expression. The actual expression of individual genes within these regions is then determined by the presence or absence of specific transcription factors. For example, the expression of α-globin remains mainly erythroid-specific, although it is in an open gene-rich cluster[48] that loops out from its chromosome territory in various tissues[50].

Although not all the genes in domains of open structure need to be expressed in any one cell type, we think that there will be a critical threshold for the proportion of active genes in a region that is needed to maintain the open chromatin structure and to propagate it across a region. By virtue of the fact that genes in gene-rich clusters have broad expression patterns[1,24], this ensures that in most cells these domains will remain 'open'. Conversely, the requirement for these regions to be constitutively open will function as a selection pressure to maintain the clustering of broadly expressed genes in the genome.

A consequence of this model is that the expression of genes within open chromatin domains should be readily inducible if the relevant transcription factors are supplied, not only at an appropriate time (for example, during development), but also inappropriately. Such a situation might arise in disease states, such as cancer, where gene expression is generally misregulated. Preliminary analysis indicates that indeed there is a tendency for the genes that are generally upregulated in various cancers to originate from the gene-rich clusters of open chromatin-fibre structure in the human genome[51] (FIG. 3).

Outlook and future directions

We have argued that there might be unexpected similarities between the chromatin-based mechanisms that regulate the coordinated expression of genes in specialized clusters (such as globin, Hox and MHC clusters) and those that regulate unrelated genes that are clustered together in the genome. Specifically, we suggest that an 'opening-up' of domains of secondary chromatin (fibre) structure might place genes in a transcriptionally permissive environment, where their expression can then be triggered by transcription factors.

There are two experimental directions that could be taken to explore this possibility further. First, we need to expand our knowledge of chromatin structures beyond the analysis of individual promoters or genes to a more global scale that encompasses large gene clusters or even whole genomes. Several studies have begun to make progress in this direction[39,44,48,52–55]. Second, the experimental approach of rearranging gene clusters, which has been important for the evaluation of the mechanisms that regulate Hox and globin expression, might also be a profitable route towards investigating the requirement for clustering of other genes in the human genome.

SUCROSE GRADIENT SEDIMENTATION
An ultracentrifugation technique that separates macromolecules on the basis of their mass and their size or shape (frictional coefficient).

1. Hurst, L. D., Pal, C. & Lercher, M. J. The evolutionary dynamics of eukaryotic gene order. *Nature Rev. Genet.* **5**, 299–310 (2004).
2. De Leo, A. A. *et al.* Sequencing and mapping hemoglobin gene clusters in the Australian model dasyurid marsupial *Sminthopsis macroura*. *Cytogenet. Genome Res.* **108**, 333–341 (2005).
3. Gillemans, N. *et al.* Functional and comparative analysis of globin loci in pufferfish and humans. *Blood* **101**, 2842–2849 (2003).
4. Mason, M. M., Lee, E., Westphal, H. & Reitman, M. Expression of the chicken β-globin gene cluster in mice: correct developmental expression and distributed control. *Mol. Cell. Biol.* **15**, 407–414 (1995).
5. Kmita, M. & Duboule, D. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**, 331–333 (2003).
6. Ikuta, T., Yoshida, N., Satoh, N. & Saiga, H. *Ciona intestinalis* Hox gene cluster: its dispersed structure and residual colinear expression in development. *Proc. Natl Acad. Sci. USA* **101**, 15118–15123 (2004).
7. Seo, H. C. *et al. Hox* cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431**, 67–71 (2004).
8. Patel, N. H. Evolutionary biology: time, space and genomes. *Nature* **431**, 28–29 (2004).
9. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. & Fraser, P. Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-globin locus. *Mol. Cell* **5**, 377–386 (2000).
10. Plant, K. E., Routledge, S. J. & Proudfoot, N. J. Intergenic transcription in the human β-globin gene cluster. *Mol. Cell. Biol.* **21**, 6507–6514 (2001).
11. Trowsdale, J. The gentle art of gene arrangement: the meaning of gene clusters. *Genome Biol.* **3**, COMMENT2002 (2002).
12. Danchin, E. G. & Pontarotti, P. Towards the reconstruction of the bilaterian ancestral pre-MHC region. *Trends Genet.* **20**, 587–591 (2004).
13. Kumanovics, A., Takada, T. & Lindahl, K. F. Genomic organization of the mammalian MHC. *Annu. Rev. Immunol.* **21**, 629–657 (2003).
14. Horton, R. *et al.* Gene map of the extended human MHC. *Nature Rev. Genet.* **5**, 889–899 (2004).
15. Wang, J. *et al.* Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions. *J. Cell Biol.* **164**, 515–526 (2004).
16. Shykind, B. M. Regulation of odorant receptors: one allele at a time. *Hum. Mol. Genet.* **14** (Suppl. 1), R33–R39 (2005).
17. Amadou, C. *et al.* Co-duplication of olfactory receptor and MHC class I genes in the mouse major histocompatibility complex. *Hum. Mol. Genet.* **12**, 3025–3040 (2003).
18. Shiina, T. *et al.* Genomic anatomy of a premier major histocompatibility complex paralogous region on chromosome 1q21–q22. *Genome Res.* **11**, 789–802 (2001).
19. Alibardi, L. & Toni, M. Localization and characterization of specific cornification proteins in avian epidermis. *Cells Tissues Organs* **178**, 204–215 (2004).
20. Williams, R. R., Broad, S., Sheer, D. & Ragoussis, J. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp. Cell Res.* **272**, 163–175 (2002).
21. Craig, J. M. & Bickmore, W. A. The distribution of CpG islands in mammalian chromosomes. *Nature Genet.* **7**, 376–382 (1994).
22. Caron, H. *et al.* The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289–1292 (2001).
23. Versteeg, R. *et al.* The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.* **13**, 1998–2004 (2003).
24. Lercher, M. J., Urrutia, A. O. & Hurst, L. D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genet.* **31**, 180–183 (2002).
This paper proposes that genes that are broadly expressed, rather than those that are highly expressed, cluster in the human genome.
25. 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).
26. Singer, G. A., Lloyd, A. T., Huminiecki, L. B. & Wolfe, K. H. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Mol. Biol. Evol.* **22**, 767–775 (2005).
27. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. & Fraser, P. Long-range chromatin regulatory interactions *in vivo*. *Nature Genet.* **32**, 1–4 (2002).
28. 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).
This study used the 3C (chromosome conformation capture) technique to show that regulatory and coding regions of the β-globin locus are brought into close proximity with each other.
29. Drissen, R. *et al.* The active spatial organization of the β-globin locus requires the transcription factor EKLF. *Genes Dev.* **18**, 2485–2490 (2004).
30. 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 (2005).
31. Palstra, R. J. *et al.* The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genet.* **35**, 190–194 (2003).
32. Spitz, F., Gonzalez, F. & Duboule, D. A global control region defines a chromosomal regulatory landscape containing the *HoxD* cluster. *Cell* **113**, 405–417 (2003).
33. Zakany, J., Kmita, M. & Duboule, D. A dual role for Hox genes in limb anterior–posterior asymmetry. *Science* **304**, 1669–1672 (2004).
34. Carson, S. & Wiles, M. V. Far upstream regions of class II MHC Ea are necessary for position-independent, copy-dependent expression of Ea transgene. *Nucleic Acids Res.* **21**, 2065–2072 (1993).
35. Krawczyk, M. *et al.* Long distance control of MHC class II expression by multiple distal enhancers regulated by regulatory factor X complex and CIITA. *J. Immunol.* **173**, 6200–6210 (2004).
36. Forsberg, E. C. *et al.* Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc. Natl Acad. Sci. USA* **97**, 14494–14499 (2000).
37. Chambeyron, S. & Bickmore, W. A. Chromatin decondensation and nuclear reorganization of the *HoxB* locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
38. Rastegar, M., Kobrossy, L., Kovacs, E. N., Rambaldi, I. & Featherstone, M. Sequential histone modifications at *Hoxd4* regulatory regions distinguish anterior from posterior embryonic compartments. *Mol. Cell. Biol.* **24**, 8090–8103 (2004).
39. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
This large-scale analysis of histone modifications in mammalian cells reveals punctate domains of modification for most genomic regions, but broad domains of active histone modifications at Hox gene loci.
40. Raval, A. *et al.* Transcriptional coactivator, CIITA, is an acetyltransferase that bypasses a promoter requirement for TAF$_{II}$250. *Mol. Cell* **7**, 105–115 (2001).
41. Zika, E. & Ting, J. P. Epigenetic control of *MHC-II*: interplay between CIITA and histone-modifying enzymes. *Curr. Opin. Immunol.* **17**, 58–64 (2005).
42. Masternak, K., Peyraud, N., Krawczyk, M., Barras, E. & Reith, W. Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nature Immunol.* **4**, 132–137 (2003).
43. Kimura, A. P., Liebhaber, S. A. & Cooke, N. E. Epigenetic modifications at the human growth hormone locus predict distinct roles for histone acetylation and methylation in placental gene activation. *Mol. Endocrinol.* **18**, 1018–1032 (2004).
44. Roh, T. Y., Cuddapah, S. & Zhao, K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19**, 542–552 (2005).
This study used a novel combination of immuno-precipitation and serial analysis of gene expression (SAGE) for mapping histone modifications across the genome.
45. Chambeyron, S., Da Silva, N. R., Lawson, K. A. & Bickmore, W. A. Nuclear re-organisation of the *Hoxb* complex during mouse embryonic development. *Development* **132**, 2215–2223 (2005).
46. Ragoczy, T., Telling, A., Sawado, T., Groudine, M. & Kosak, S. T. A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome. Res.* **11**, 513–525 (2003).
47. Volpi, E. V. *et al.* Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* **113**, 1565–1576 (2000).
This was the first study to show that a region of the genome can be extruded from a chromosome territory in a regulated fashion.
48. Gilbert, N. *et al.* Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibres. *Cell* **118**, 555–566 (2004).
This paper highlights the first global survey of the biophysical properties of chromatin fibres across the human genome. It also proposes that an open chromatin structure is present in regions of high gene density.
49. Lercher, M. J., Urrutia, A. O., Pavlicek, A. & Hurst, L. D. A unification of mosaic structures in the human genome. *Hum. Mol. Genet.* **12**, 2411–2415 (2003).
50. Mahy, N. L., Perry, P. E. & Bickmore, W. A. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol.* **159**, 753–763 (2002).
51. Zhou, Y. *et al.* Genome-wide identification of chromosomal regions of increased tumor expression by transcriptome analysis. *Cancer Res.* **63**, 5781–5784 (2003).
52. Murrell, A., Rakyan, V. K. & Beck, S. From genome to epigenome. *Hum. Mol. Genet.* **14** (Suppl. 1), R3–R10 (2005).
53. Crawford, G. E. *et al.* Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc. Natl Acad. Sci. USA* **101**, 992–997 (2004).
54. Sabo, P. J. *et al.* Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc. Natl Acad. Sci. USA* **101**, 4537–4542 (2004).
55. Weil, M. R., Widlak, P., Minna, J. D. & Garner, H. R. Global survey of chromatin accessibility using DNA microarrays. *Genome Res.* **14**, 1374–1381 (2004).
56. Marenholz, I., Heizmann, C. W. & Fritz, G. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem. Biophys. Res. Commun.* **322**, 1111–1122 (2004).
57. Marenholz, I. *et al.* Identification of human epidermal differentiation complex (EDC)-encoded genes by subtractive hybridization of entire YACs to a gridded keratinocyte cDNA library. *Genome Res.* **11**, 341–355 (2001).
58. Marshall, D., Hardman, M. J., Nield, K. M. & Byrne, C. Differentially expressed late constituents of the epidermal cornified envelope. *Proc. Natl Acad. Sci. USA* **98**, 13031–13036 (2001).
59. Huber, M. *et al.* Isolation and characterization of human repetin, a member of the fused gene family of the epidermal differentiation complex. *J. Invest. Dermatol.* **124**, 998–1007 (2005).
60. Contzler, R., Favre, B., Huber, M. & Hohl, D. Cornulin, a new member of the 'fused gene' family, is expressed during epidermal differentiation. *J. Invest. Dermatol.* **124**, 990–997 (2005).
61. Jang, S. I. & Steinert, P. M. Loricrin expression in cultured human keratinocytes is controlled by a complex interplay between transcription factors of the Sp1, CREB, AP1, and AP2 families. *J. Biol. Chem.* **277**, 42268–42279 (2002).
62. Martin, N., Patel, S. & Segre, J. A. Long-range comparison of human and mouse *Sprr* loci to identify conserved noncoding sequences involved in coordinate regulation. *Genome Res.* **14**, 2430–2438 (2004).
63. Johnnidis, J. B. *et al.* Chromosomal clustering of genes controlled by the aire transcription factor. *Proc. Natl Acad. Sci. USA* **102**, 7233–7238 (2005).

Acknowledgements

D.S. is a Medical Research Council (MRC) pre-doctoral training fellow and W.A.B. is a Centennial fellow of the James S. McDonnell foundation. This work was funded by the MRC UK, and in part by the EU FP6 Network of Excellence Epigenome.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene

CD74 | *CD79B* | *Evx2* | *GH1* | *HLA-DRA* | *Hoxb1* | *Hoxb9* | *Hoxd* | *Hoxd4* | IFNG | *Lnp* | MHC2TA

Nuclear Protein Database: http://npd.hgu.mrc.ac.uk

PML bodies

Access to this interactive links box is free online.