

# Why the activity of a gene depends on its neighbors

Alexander Feuerborn and Peter R. Cook

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

**Sixty years ago, the position of a gene on a chromosome was seen to be a major determinant of gene activity; however, position effects are rarely central to current discussions of gene expression. We describe a comprehensive and simplifying view of how position in 1D sequence and 3D nuclear space underlies expression. We suggest that apparently-different regulatory motifs including enhancers, silencers, insulators, barriers, and boundaries act similarly – they are active promoters that tether target genes close to, or distant from, appropriate transcription sites or ‘factories’. We also suggest that any active transcription unit regulates the firing of its neighbors – and thus can be categorized as one or other type of motif; this is consistent with expression quantitative trait loci (eQTLs) being widely dispersed.**

## Position effects and gene regulation

In 1950 (three years before the description of the double helix), this is how Nobel laureate E.B. Lewis began a review [1]:

‘That the effect of a gene may be dependent upon its position with respect to neighboring genes is now a well-established fact. . . This phenomenon of position effect. . . should throw light on the organization of the chromosomes as well as on the primary reactions of specific genes.’ (Note: ‘primary reactions’ are now known as ‘transcription’).

In 2015, position effects (those due to position in 1D sequence space on a chromosome; see [Glossary](#)) are known to be commonplace; for example, expression levels of a reporter gene can vary  $\sim 10^4$ -fold when integrated at thousands of different sites around the mouse genome [2]. However, they are usually not central to current discussions of genome organization and gene regulation, where the underlying molecular mechanisms remain obscure. Nevertheless, all agree these mechanisms are complex, with  $\sim 10^6$  sequences regulating only  $\sim 23\,000$  human genes [3]. Moreover, regulators are diverse. We build on Lewis’ ‘fact’ that ‘position’ is the key, and describe a comprehensive and simplifying view of how position (in 1D sequence and 3D nuclear space) determines gene expression (and vice versa). Our purpose is to specify more precisely what the underlying molecular mechanisms might be.

Corresponding author: Cook, P.R. ([peter.cook@path.ox.ac.uk](mailto:peter.cook@path.ox.ac.uk)).

Keywords: boundary; domain; enhancer; expression quantitative trait loci; position effect; silencer.

0168-9525/

© 2015 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tig.2015.07.001>

## Some forces shaping structure

As pathologists know, genome architecture varies from cell to cell – even in clonal populations. At the molecular level, high-throughput chromosome conformation capture (Hi-C) applied to single mouse cells reveals that none share exactly the same genic contacts; however, some contacts are seen more often than others, and therefore the organization is non-random [4]. Time-lapse imaging of living human cells also shows the organization changes from moment to moment; a locus tagged with GFP might diffuse through a local volume (diameter 0.5–1  $\mu\text{m}$ ) for a minute or more (to contact briefly many other sequences), ‘jump’ to a neighboring volume the next (to contact others), and then become transiently immobilized [5,6].

Which proteins might stabilize specific contacts? We begin with transcription factors because they provide the necessary specificity. Many factors (either acting alone, or complexed with others) are ‘bivalent’ in the sense that they (or the complex) can bind to two different segments of DNA to form a loop. [Box 1](#) illustrates three different ways they can stabilize loops, but only the first two require such bivalency. Any loops that are formed will persist for the order of seconds – the average residence time of a typical factor on DNA (again shown by GFP tagging [5]). However, engaged RNA polymerases can remain bound for longer (polymerase II takes  $\sim 10$  min to transcribe a typical human gene of 30 kb), and this tight binding is specific in that it occurs throughout the transcription unit but not elsewhere. If two engaged polymerases are associated with other bivalent factors or complexes, then the same three ways can drive polymerases together. Since one-third of

## Glossary

**3C:** chromosome conformation capture, a technique for assessing the proximity between two sequences on a chromosome in 3D nuclear space.

**CTCF:** CCCTC-binding factor, originally defined as a transcription factor.

**eRNAs:** transcripts encoded by enhancers.

**eQTLs:** expression quantitative trait loci.

**DRB:** 5,6-dichloro-1- $\beta$ -D-ribo-furanosyl-benzimidazole, a transcriptional inhibitor.

**GWAS:** genome-wide association studies – the examination of many genetic variants in different individuals to see if any one variant is associated with a given phenotypic trait.

**ENCODE:** The Encyclopedia of DNA Elements.

**Hi-C:** a high-throughput variant of 3C.

**HMR:** hidden mating-type locus right, a locus controlling yeast mating type.

**HUVECs:** human umbilical vein endothelial cells.

**NF- $\kappa$ B:** nuclear factor  $\kappa$ B, a transcription factor.

**TNF $\alpha$ :** tumor necrosis factor  $\alpha$ , a cytokine.

**Position effect:** effects on expression of changing the location of a gene on a chromosome.

**YY1:** Yin Yang 1, usually considered a transcriptional repressor.

### Box 1. Equilibria favoring looping

A typical transcription factor is present at  $\sim 1$  nM, and many bind to others with equilibrium constants of  $\sim 10^{-7}$  M; then, these numbers mean that  $<1\%$  are instantaneously in protein:protein complexes (Figure IAi). However, if DNA with two cognate binding sites 10 kb apart is present,  $\sim 67\%$  are in protein:protein complexes bound to DNA; DNA binding increases the local protein concentration, and thus interaction frequencies. Because such protein:protein complexes are 'bivalent', they can loop DNA [67] (Figure IAii).

Such clustering/looping is probably reinforced by a 'bridging-induced attraction' uncovered using (molecular dynamics) simulations of bivalent 'factors' binding to – and dissociating from – 'chromatin fibers' [68] (Figure IB). In the absence of explicit interactions between one factor and another, transiently bound factors nevertheless cluster spontaneously. Once a cluster happens to appear, it tends to persist; dissociating 'bridges' rebind nearby because the local concentration of binding sites is high (i.e., near other 'bridges').

Transcription factors are often found in larger complexes, and then 'depletion attraction' provides another force driving clustering [69] (Figure IC). In a crowded cell, small proteins (diameter  $<5$  nm) continually bombard larger complexes (diameter 5–25 nm) from all sides. When two larger complexes come into contact, the small proteins are sterically prevented from entering the green volume between the two and thus cannot knock them apart; as a result, the small proteins exert a force equivalent to the osmotic pressure on opposite sides of the two larger complexes to keep them together (Figure ICi). If the larger complexes have DNA-binding sites, this 'attraction' can again stabilize loops (Figure ICii).

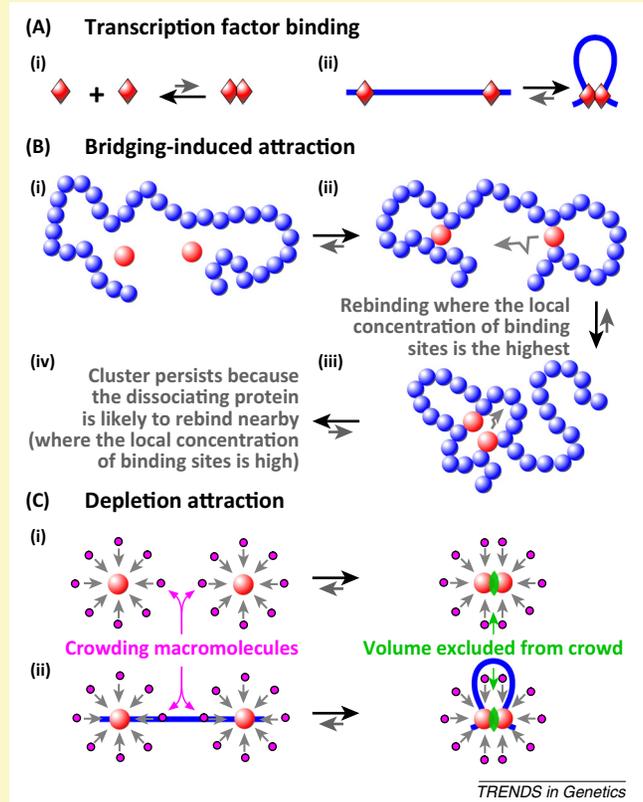


Figure I. Three equilibria favoring looping.

engaged polymerases are also 'paused'/'stalled' [7], these aggregates could persist for longer [7]. Hence, the system must either spend energy to prevent the clustering or – as seems likely – it goes with the flow and uses other familiar forces (charge interactions, H bonds, van der Waals, and hydrophobic forces) to organize the resulting structures.

### Clusters of active polymerases – 'factories'

The forces described above fit comfortably with a model for genome organization in which a central architectural feature is a cluster of active polymerases – a 'transcription factory' – surrounded by loops [8,9] (Figure 1). We define such a factory as a site containing at least two polymerases (plus associated factors) active on at least two templates (to distinguish it from the case where two polymerases are

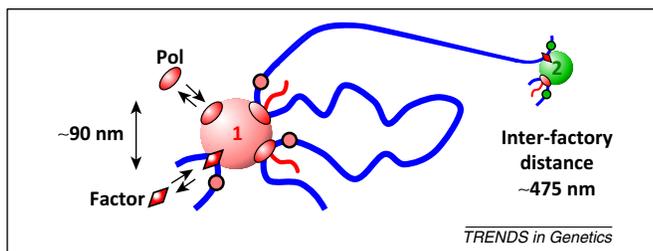


Figure 1. Transcription factories. Chromatin is tethered through clusters of polymerases/factors to two nucleoplasmic factories (1,2) which are rich in different factors. A typical factory is associated with  $\sim 16$  loops ( $\sim 8$  tethered through active polymerases and  $\sim 8$  through factors [9,22,23]). In this and subsequent figures, only a few attached loops are shown, polymorphic factories are represented as uniform spheres, and promoters (colored circles) tend to initiate in factories of the same color.

active on one template). These factories contain high local concentrations that act through the law of mass action to drive production. For example, mammalian nuclei contain a  $1 \mu\text{M}$  pool of polymerase II, but essentially all transcripts are made in factories where concentrations are  $\sim 1000$ -fold higher.

The first evidence for factories came when permeabilized human cells were incubated in bromouridine triphosphate (BrUTP) plus the other triphosphates required for transcription; after immuno-labeling, nascent BrRNA was seen in discrete sites [10]. These sites are so closely spaced they are difficult to resolve one from another by conventional microscopy, but clusters of polymerases [11] and appropriately tagged factors [12] have now been imaged in living cells using modern techniques; even so, the exact relationship of these clusters to active sites of transcription remains obscure. Factories have also been purified and their proteomes and transcriptomes analyzed – they contain the relevant polymerases, factors, RNA-binding proteins [13], and transcripts [14]. Significantly, the most frequently found contacts detected by chromosome conformation capture (3C)-based methods involve active transcription units [15–19] – the expected result if active units are tethered to factories.

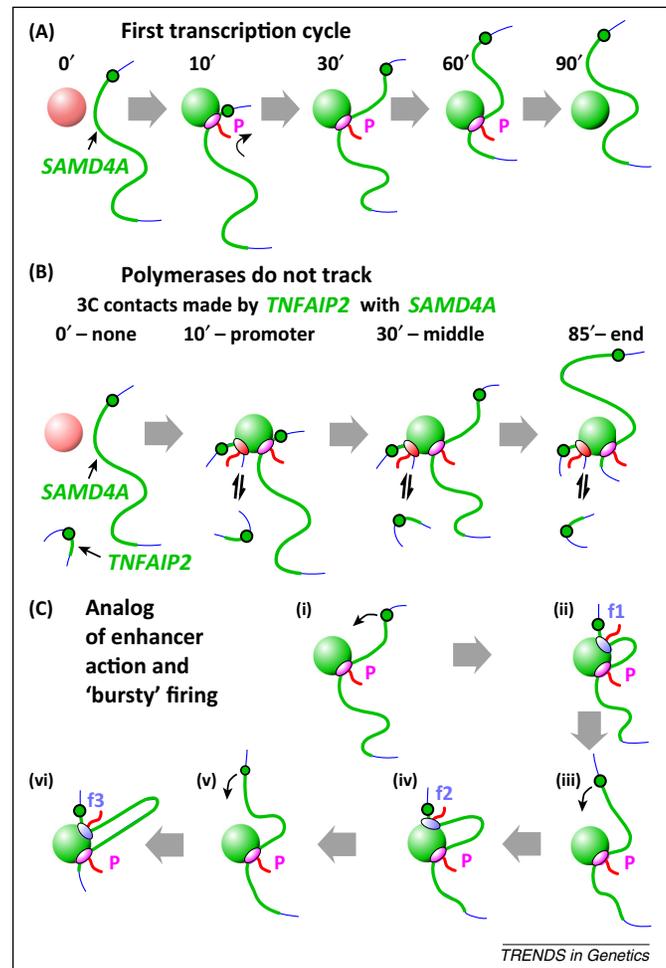
Some car factories make Fords, others Hondas; do factories also specialize in transcribing specific gene sets? They do [9]. The nucleolus provides the prototypic example – a place where many rRNA genes are cotranscribed by polymerase I. Active polymerases II and III are also found in their own nucleoplasmic factories, and many different

types of polymerase II factories have now been identified [9]; for example, 'ER $\alpha$ ' (estrogen receptor  $\alpha$ ) and 'KLF1' (Krueppel-like factor 1) factories transcribe genes involved in the estrogen response and globin production, respectively [20,21].

### An example – *SAMD4A*

We now illustrate how a gene might become active, using *SAMD4A* (sterile alpha motif domain containing 4A) as an example (Figure 2A). Although many other genes have been studied in great detail, this gene has various advantages in this context [22] (and we will use it as a concrete example throughout, but the general arguments apply to all genes). First, it is 221 kb in length, and this great length allows the two techniques used to assess proximity in nuclear space – 3C and RNA FISH (fluorescence *in situ* hybridization) coupled to super-resolution localization – to be applied with high precision [22–25]. Second, it can be 'switched on' rapidly in human umbilical vein endothelial cells (HUVECs) – cells studied by the ENCODE project [3]. Third, these cells are diploid and – in the cases discussed – synchronized in G0 phase, so there are no complicating effects of additional gene copies. Thus, *SAMD4A* is normally inactive as the relevant factor – nuclear factor  $\kappa$ B (NF- $\kappa$ B) – is sequestered in the cytoplasm. The promoter also contacts few other genes; it seems to diffuse through empty 'outer space' around factories, and – although it might occasionally visit one – it rarely initiates because the NF- $\kappa$ B concentration is low. When tumor necrosis factor  $\alpha$  (TNF $\alpha$ ) is added, NF- $\kappa$ B floods into nuclei and facilitates initiation during one of these visits. Consequently, a 'pioneering' polymerase begins transcribing *SAMD4A* after ~10 min, and the attached polymerase reels in the template as it extrudes the transcript. The gene is now in a crowd of other active units on the factory surface, and therefore makes many 3C contacts. This pioneer then continues to transcribe steadily (at ~3 kb/min) until it reaches the terminus after ~75 min, when it detaches. Subsequently, the factory develops into one that specializes in transcribing responsive genes (Box 2). Unlike *SAMD4A*, many genes that respond to TNF $\alpha$  are marked by a 'paused' polymerase [26]; we imagine these are pre-attached to a factory, and thus 'poised' ready to respond rapidly to the cytokine.

This model involves a transiently immobilized polymerase, in contrast to the traditional one where the polymerase tracks down the template. It turns out there is little evidence in favor of the traditional view, but considerable amounts supporting the alternative. Because this has already been reviewed [9], we summarize only two pieces of evidence here. First, consider what is often believed to be the strongest evidence for tracking polymerases – the iconic images of 'Christmas trees' illustrating genes in action. Such images are prepared by spreading the compact template such that polymerases are apparently frozen in the act of transcribing. However, these are static images that only indirectly tell us about relative movement. Thus, a tracking polymerase that transcribes a helical template would generate a transcript that is entwined about that template once for every 10 bp transcribed – and spreading would surely not un-entwine it so perfectly to give an



**Figure 2.** Transcription of *SAMD4A*. (A) Transcription cycle. Initially, *SAMD4A* is inactive; it visits the pink factory, but does not initiate (NF- $\kappa$ B is absent). Adding TNF $\alpha$  induces nuclear influx of NF- $\kappa$ B, and the promoter initiates on visiting the factory (now green). The 'pioneering' polymerase (p) begins transcribing; it reels in the template (arrow) as it extrudes its transcript. The pioneer now transcribes steadily to reach the terminus after ~85 min, when it detaches. (B) 3C evidence that active polymerases are transiently immobilized. Initially, two responsive genes (*SAMD4A*, *TNFAIP2*) are inactive and in 'outer space' (and so rarely in contact). Ten minutes after adding TNF $\alpha$ , both promoters have initiated; consequently, *TNFAIP2* now often contacts the *SAMD4A* promoter (but not the middle of the gene or the terminus). Because *TNFAIP2* is only 11 kb, it is soon transcribed and detaches; however, it often reattaches to be retranscribed (arrows). After 30 min the pioneer has transcribed 60 kb into *SAMD4A*, and reattached *TNFAIP2* now lies close to the middle of *SAMD4A* (but no longer the promoter) After 85 min, the pioneer on *SAMD4A* is about to terminate. When *TNFAIP2* reinitiates, it lies near the *SAMD4A* terminus (and no longer the middle). Such results are impossible to reconcile with a model involving tracking polymerases. (C). Analog of enhancer action. (i) The pioneer has transcribed ~60 kb; the promoter is tethered close to the factory and is therefore likely to reinitiate on revisiting it (arrow). (ii) A 'following' polymerase (f1) has initiated. (iii) For unknown reasons, the follower soon aborts (as the pioneer continues), but the promoter is still tethered close to the factory and is therefore likely to revisit it (arrow). (iv) The promoter has reinitiated, and thus another 'follower' (f2) transcribes *SAMD4A*. (v,vi) The follower soon aborts, the promoter detaches and reattaches as f3 initiates. Close tethering underlies enhanced firing and 'bursty' transcription.

extended 'branch' in the 'Christmas tree'. However, if transcription occurs as the template is pulled through a polymerase immobilized on the surface of a factory, ripping off the template plus the transcript would yield the iconic image without any entwinements. Paradoxically, then, these images provide strong support for polymerases being fixed when active.

**Box 2. How a factory might become specialized**

Before the addition of TNF $\alpha$  (Figure 1A), some genes with pink promoters are being transcribed in the pink factory; they detach on termination as other pink promoters initiate. Responsive promoters (green) are in 'outer' space; they may occasionally visit the pink factory but are unlikely to initiate (because NF- $\kappa$ B is absent). Stimulation with TNF $\alpha$  increases NF- $\kappa$ B levels, and *SAMD4A* will now initiate when it visits the factory (Figure 1B). Subsequently, gene A may initiate (Figure 1C). Now, some NF- $\kappa$ B is transiently bound in/around the factory; on dissociation, the local concentration of NF- $\kappa$ B increases (green halo), and this increases the chances that other responsive genes initiate. For example, in Figure 1D, C has initiated (binding of NF- $\kappa$ B to it further increases the local NF- $\kappa$ B concentration); the factory has evolved into one specializing in transcribing responsive genes (it is therefore now shown as green). Put another way, when NF- $\kappa$ B bound to a promoter dissociates, the local concentration in the soluble pool increases, and this triggers a virtuous cycle that increases it further: whenever another TNF $\alpha$ -responsive promoter (which also encodes NF- $\kappa$ B binding sites) collides with the factory, the marginal increase in local concentration makes it more likely to initiate, and – when it does – the concentration of closely-tethered NF- $\kappa$ B binding sites increases, any dissociated NF- $\kappa$ B is more likely to rebind, and this further increases local concentrations and thus initiation rates. Clustering of responsive genes/transcripts and active NF- $\kappa$ B is supported by 3C and microscopy [23], while computer modeling [68] and imaging [70] confirm that local concentrations can be maintained despite the homogenizing effects of diffusion.

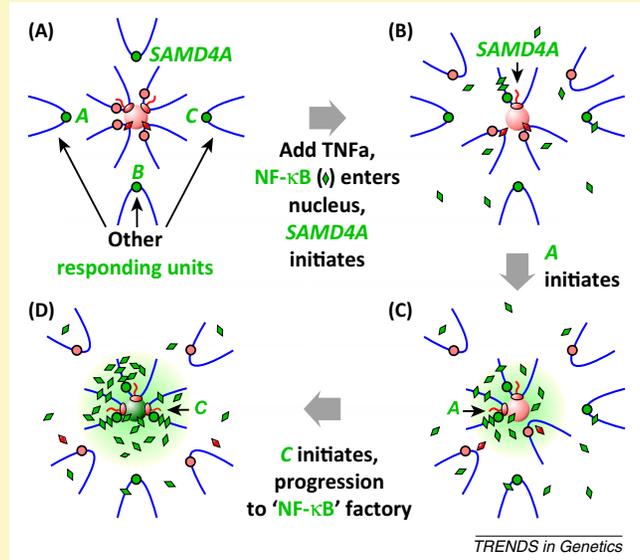


Figure 1. Development of an 'NF- $\kappa$ B' factory.

The second piece of evidence involves *SAMD4A* and another responsive gene – *TNFAIP2* (TNF $\alpha$ -induced protein 2) [22] (Figure 2B). Before adding TNF $\alpha$ , both genes rarely contact each other – and therefore are in 'outer space'. However, when the cytokine induces initiation, *TNFAIP2* often contacts the *SAMD4A* promoter – because both have attached to the same factory that specializes in transcribing TNF $\alpha$ -responsive genes. Since *TNFAIP2* is only 11 kb, its pioneering polymerase soon terminates, and *TNFAIP2* detaches. However, it may reattach to the same factory (owing to the high local concentrations of relevant factors), and recurring cycles of attachment/detachment underlie continued transcription. Then, it should often lie close to the middle of *SAMD4A* (but no longer the promoter) after 30 min, and to the terminus (but not the promoter or the middle) after 85 min. Moreover, intronic (nascent) RNAs copied from relevant segments of the two genes lie close enough together at the appropriate times to be on the surface of one factory (assessed using RNA FISH coupled to super-resolution localization). These results are impossible to reconcile with a model involving two tracking polymerases. If such a model applied, why should the two genes ever come together, and how could the particular pattern of contacts change in this highly specific way? In addition, if by chance the polymerases on the different genes happened to find themselves together, we would expect the two to become increasingly separated as the pioneer tracked down the great length of *SAMD4A* – but they remain together.

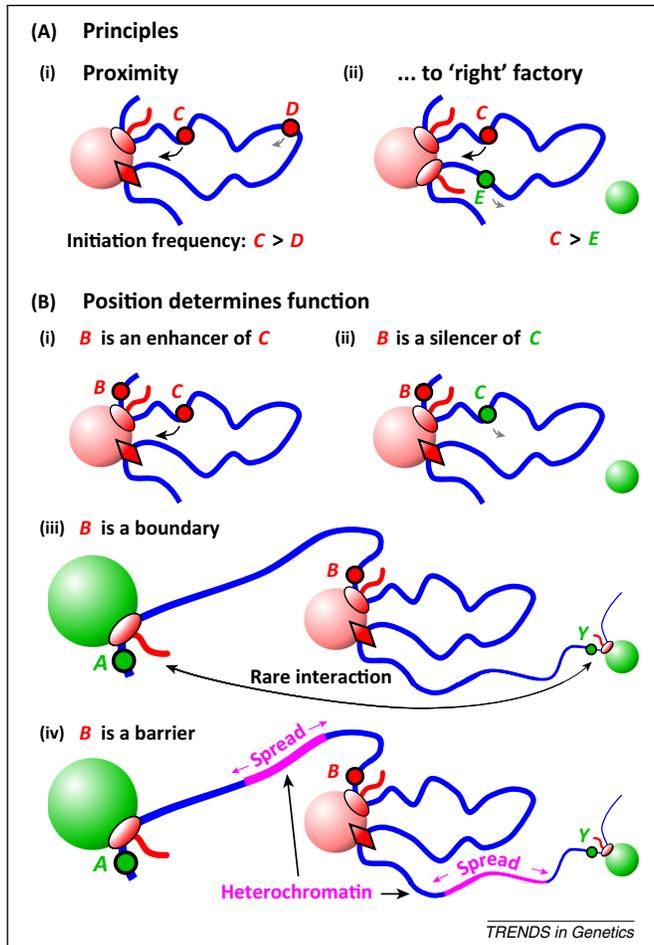
Studies on *SAMD4A* also provide insight into enhancer action [25]. After 30 min the pioneer has transcribed ~60 kb; consequently, the promoter is tethered close to the factory – which contains the appropriate factors (Figure 2Ci) – and therefore is likely to revisit it and reinitiate (Figure 2Cii). For unknown reasons, the 'following' polymerase now aborts (Figure 2Ciii), but continued

close tethering facilitates successive rounds of reinitiation/abortion (Figure 2Civ–viii). Put another way, the pioneer 'enhances' promoter firing in 'bursts'. Both phenomena (enhancer action and bursty firing) have been difficult to explain [27,28].

**A common mechanism for different regulators**

The various motifs regulating expression (enhancers, silencers, insulators, barriers, boundaries) are usually seen to be very different from one another – as their names indicate. However, they share one property: they encode active promoters [9]. Thus, (i) the canonical  $\beta$ -globin enhancer is a cluster of promoters, and active enhancers are now defined in genome-wide analyses simply as promoters that fire to yield 'eRNAs' (enhancer-encoded RNAs) [29–32]; (ii) the canonical silencer in yeast *HMR* (hidden mating-type locus right) is a tRNA promoter [33]; (iii) the most-famous insulator – CTCF (CCCTC-binding factor) – was originally defined as a transcription factor [34], other insulators are tRNA genes [35], and genome-wide studies now fail to distinguish insulators from promoters [36]; (iv) polymerases stalled at *Hox* (homeobox gene) promoters form chromosomal boundaries [37]; and (v) domain boundaries are both enriched in active promoters and carry the associated marks [e.g., bound polymerases, histone H3 acetylated at lysine 27 (H3K27Ac) [38]]. Significantly, most contacts seen using 3C-based techniques involve transcriptionally-active regulatory motifs, with many involving enhancers [18,19,28].

Now consider Figure 3A, where C is more likely to initiate than D or E because it is tethered closer to an appropriate factory – an intuition supported by computer modeling [39] and experiment [25]. Therefore, we suggest that any transcribed unit (whether non-coding or protein-coding) regulates the activity of nearby units, and we might call that transcribed unit an enhancer in one case, a



**Figure 3.** Position effects. Promoters are likely to initiate in factories of the same color (because they are rich in relevant factors). **(A)** Principles. (i) *C* is tethered closer to the pink factory than *D*, and therefore is more likely to initiate. (ii) *E* is tethered as close to the pink factory as *C*, but is unlikely to initiate there because it requires 'green' factors concentrated in the distant green factory. **(B)** How different regulatory motifs work. (i) *B* tethers *C* close to an appropriate factory, thus *B* is an enhancer of *C*. (ii) At a different stage in development, green factors suppressing initiation by *C* in the pink factory are present. *C* still visits the pink factory, but rarely initiates. Because the tether is too short to allow *C* to visit the green factory, *B* now silences *C*. (iii) Promoters *A* and *Y* often initiate in the green factories, but because they lie far apart in space, they rarely interact. Here, *B* is a boundary. (iv) Chromatin remote from factories is transcribed infrequently, and thus acquires inactive histone marks (purple). The activity of *B* prevents inactive marks spreading down the fiber, and *B* is therefore a barrier.

silencer in another, and so on. In other words, the context determines the way it acts, and this context is simply proximity to a factory containing the appropriate factors. Thus, in Figure 3B, imagine that *B* is often transcribed in the pink factory – and therefore is often attached to it. Then, *B* is an enhancer of *C*, because it often tethers *C* close to an appropriate factory (Figure 3Bi). If *B* is replaced by a cluster of closely spaced promoters that are often transcribed in the same pink factory (to increase the number of close-tethering points, and thus the fraction of time *C* lies near the factory), we would describe such a cluster as a 'super-enhancer' [40]. If new factors appear during development that reduce the likelihood that *C* can initiate in the pink factory while increasing its chances of initiating in the green one, *B* now silences *C* by tethering it far from the 'right' type of factory (Figure 3Bii). In the context of its surroundings, *B* can also act as a boundary or barrier

(Figure 3Biii,iv). Then, *B* is simultaneously an enhancer/silencer/insulator/etc. depending on the context and which gene is being considered.

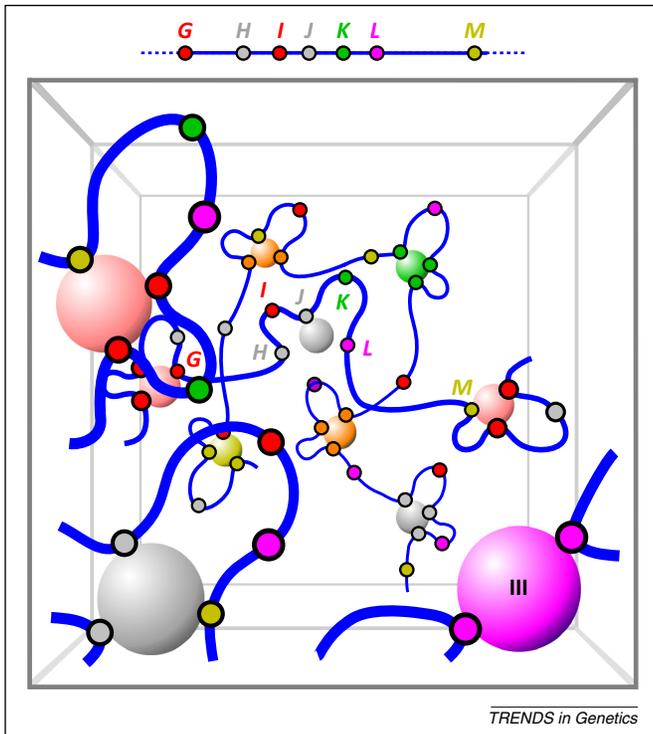
CTCF and cohesin play important roles in domain organization [19,41]; how do they fit in? We suggest in two inter-related ways: both are transcription factors [41], and both are enriched with active polymerases at domain boundaries [19,38]. Although 'knocking' down one/both influences domain organization, effects are modest [42–45]. Therefore, additional players must maintain domains, and we suggest these are polymerases and their factors such as ZNF143 (zinc finger protein 143) and YY1 (Yin Yang 1) that are co-enriched at boundaries [19,38].

### Regulators are widely distributed

Our thinking about how regulators work is rooted in classical studies of bacterial repressors. Although the situation in eukaryotes is more complicated, we assume the same basic mechanisms apply – but with more regulators binding to more targets near and far from promoters. The finding that co-expression of only four transcription factors (Oct3/4, Sox2, c-Myc, Klf4) can transform fibroblasts into pluripotent cells [46] encourages us to think we should be able to manipulate phenotypes at will if only we could identify the relevant 'master' regulators.

Nonetheless, a second set of evidence – rooted in quantitative genetics – leads to a different view. The copy-number of a transcript is a phenotypic trait, and genes influencing it can be identified in an unbiased way using genome-wide association studies (GWAS). We might then expect the regulatory loci detected to encode the master regulators and their binding sites; however, these are rarely seen. Thus, the first GWAS applied to transcript copy-number utilized microarrays to determine the levels of ~6000 mRNAs in the progeny of a cross between a lab strain of *Saccharomyces cerevisiae* and a wine strain [47,48]. Expression levels depended on parental origin, with 84% of the observed variation being heritable (defined as the fraction of phenotypic variance attributable to additive genetic effects). However, only 3% transcripts behaved as if their levels were controlled by one locus, with half being controlled (additively) by  $\geq 5$ . The majority of regulatory loci lay >10 kb away, sometimes on other chromosomes. Significantly, they rarely encoded transcription factors, and their targets rarely encoded binding sites [49].

As in yeast, the expression of most human genes is regulated by many loci around the genome (i.e., eQTLs) which do not usually encode transcription factors [50,51]. Put simply, transcript copy-number is a polygenic trait much like human height or susceptibility to type II diabetes – traits where hundreds of regulatory loci have been identified and where many more are expected to be found [52,53]. Because one third of eQTLs are symmetrically concentrated around promoters (which encode factor-binding sites) and another third around transcription end-sites (which do not) [54–56], the common characteristic of eQTLs is that they are transcription units – as a recent comprehensive analysis confirms [53]. Moreover, eQTLs are often in genes in the same functional pathway as the target, regulator and target are often in contact [57], and the target may be activated or repressed [50,56].



**Figure 4.** Factors affecting the position and activity of *K* and hypothetical non-genic units (*G*, *H*, *I*, *J*, *L*, *M*) in nuclear space. The genetic map is shown above the 3D structure. *K* is currently inactive, and other units are often (not exclusively) transcribed in factories of the same color that contain the appropriate factors. *K* is tethered close to an (inappropriate) gray and (appropriate) green factory. The frequency of initiation of *K* depends on the activity of neighbors. For example, if *J* is often transcribed in the gray factory, *J* silences *K* (*J* is simultaneously an enhancer of *H*); *K* is also silenced if *L* (perhaps a tRNA gene) becomes transcribed in the polymerase III factory (purple). During differentiation, changing concentrations of transcription factors and occupancy of binding sites should allow highly nuanced regulation of the 3D network and of the expression of individual transcription units.

These results show that gene regulation is widely distributed, often involving many other transcription units that can act positively or negatively. Once again, they sit comfortably with the model in Figure 4, where the activity of a unit is affected by that of many others (often not immediate neighbors), and one active unit might simultaneously function as an enhancer/silencer/boundary/etc. Moreover, regulators can be activators or repressors, and often contact their targets and share functional pathways (e.g., green units share green factories). Then, it is unsurprising that functionally related genes tend to be closely linked on chromosomes [58].

### Transcriptional activity underlies the activity of regulatory motifs

In our model transcriptional activity underlies the activity of regulatory motifs, but previous studies have concluded that inhibiting the elongating forms of RNA polymerase II with the inhibitor DRB (5,6-dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole) does not disrupt the contacts seen between the canonical globin enhancer and its target [59,60]. However, initiating forms of the polymerases and/or transcription factors can also maintain contacts, and these should be unaffected by the inhibitor.

Other evidence (which does not rely on the use of inhibitors) points to a direct relationship between transcription, looping, and enhanced transcription. For example, old

experiments show that, when transcription is progressively inhibited during development of chicken erythroblasts into completely inactive erythrocytes, loops are progressively lost until none remain [61]. Moreover, a recent comprehensive analysis of gene activation (involving 19 human and 14 mouse time courses as cells respond to growth-factors and/or differentiate) shows that enhancers fire shortly before their targets [32]. Then, enhancers would first attach to a factory, and fire. This creates a new long-lasting attachment and subdivides the loop, and the target gene thus becomes tethered closer to an appropriate factory to increase its chances of firing (it is now technically feasible to check this). In addition, varying the rate of production of a particular eRNA might sometimes affect local looping, and sometimes not (depending on how it affects competition for factories by nearby transcription units) – and this is again the case [28]. Finally, we would expect ‘pluripotent’ transcription factors such as Oct4 and Nanog to help to maintain genome organization in embryonic stem cells, and knock-outs to only slightly affect it; once again, this is the case [62].

### Towards the structural biology of the genome

One challenge posed by the sequence of the human genome is to understand how 1D sequence information specifies 3D architecture and function. Structure must be described in probabilistic terms as it depends on stochastic variation and past history (which effect the unique concentration of factors in each cell). Moreover, structures in two sisters are unlikely ever to be the same because they change from moment to moment – and, even within one cell, the structure of any two alleles will be different because their transcription is uncorrelated [63]. Nevertheless, we accept this challenge and present a detailed and comprehensive model. Importantly, we try to replace vague terms such as ‘transcriptionally-active compartment’ by specifying exactly what the most important players are (transcription factors, polymerases, promoters), how regulatory motifs

#### Box 3. Some major factors affecting the position of *SAMD4A* in nuclear space

*SAMD4A* lies in the middle of chromosome 14, and different parts of it will be tethered to various structures in the nucleus. During anaphase, the mitotic spindle segregates the chromosome to a local region in the prospective daughter; as the nucleus reforms, movement of decondensing chromosome 14 is restricted by neighboring ones. Nevertheless, individual segments within it diffuse locally, and – when those carrying (heterochromatic) histone marks nucleate lamin formation and/or visit the newly-forming lamina – they bind to lamin-associated proteins such as LBR (lamin B receptor [71]), BAF (barrier-to-autointegration factor [72]), and YY1 [73]. Consequently, heterochromatin (e.g., centromeric  $\alpha$ -satellite repeats, G bands, lamin-associated domains) tends to pack against the lamina. This is reinforced by the self-aggregation of heterochromatic nucleosomes, which also become concentrated around internal chromocenters. Such aggregation will be augmented by the depletion attraction, which drives thicker (heterochromatic) fibers to the periphery and together [74]. The nucleolus is another major anchor [9]. Promoters driving rRNA production bind UBF (upstream-binding factor), which ‘bookmarks’ them for future activity; then, on exit from mitosis, these promoters organize polymerase I factories (we suggest that analogous bookmarks [75] underlie assembly of nucleoplasmic factories). Consequently, *SAMD4A* might be distantly anchored to the nucleolus (because its chromosome encodes rDNA), and/or the periphery.

might act, and what forces drive the organization (many additional forces will be superimposed on the ones discussed, and some major ones are described in Box 3). Because our arguments are general, our model should apply to any complex genome. Indeed, position effects have recently been found in bacteria [64], where highly-active genes divide the genome into domains [65,66] and a 3C-based technique shows co-regulated operons cluster in 3D space [61]. Then, the words of Lewis were prescient: position determines ‘...the organization of the chromosomes as well as...the primary reactions of specific genes’. And the organization regulates transcription, while transcription specifies the organization. Fortunately, tagging with fluorescent proteins and ‘super-resolution’ techniques now allow us to test this (and other) models by monitoring the interplay between the ‘organization’ and the key players driving the ‘primary reactions’ (i.e., polymerases and associated factors).

### Acknowledgments

We thank The Medical Research Council for support.

### References

- Lewis, E.B. (1950) The phenomenon of position effect. *Adv. Genet.* 3, 73–115
- Akhtar, W. *et al.* (2013) Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* 154, 914–927
- Bernstein, B.E. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
- Nagano, T. *et al.* (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64
- Wachsmuth, M. *et al.* (2008) Genome organization: balancing stability and plasticity. *Biochim. Biophys. Acta* 783, 2061–2079
- Levi, V. *et al.* (2005) Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. *Biophys. J.* 89, 4275–4285
- Jonkers, I. and Lis, J.T. (2015) Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16, 167–176
- Eskiw, C.H. *et al.* (2010) Transcription factories and nuclear organization of the genome. *Cold Spring Harb. Symp. Quant. Biol.* 75, 501–506
- Papantonis, A. and Cook, P.R. (2013) Transcription factories; genome organization and gene regulation. *Chem. Rev.* 113, 8683–8705
- Jackson, D.A. *et al.* (1993) Visualization of focal sites of transcription within human nuclei. *EMBO J.* 12, 1059–1065
- Cisse, I.I. *et al.* (2013) Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* 341, 664–667
- Ghamari, A. *et al.* (2013) In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes Dev.* 27, 767–777
- Melnik, S. *et al.* (2011) The proteomes of transcription factories containing RNA polymerases I, II, or III. *Nat. Methods* 8, 962–968
- Caudron-Herger, M. *et al.* (2015) Dissecting the nascent human transcriptome by analyzing the RNA content of transcription factories. *Nucleic Acids Res.* Published online April 20, 2015. <http://dx.doi.org/10.1093/nar/gkv390>
- Simonis, M. *et al.* (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38, 1348–1354
- Yaffe, E. and Tanay, A. (2011) Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* 43, 1059–1065
- Zhang, Y. *et al.* (2013) Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* 503, 290–294
- Heidari, N. *et al.* (2014) Genome-wide map of regulatory interactions in the human genome. *Genome Res.* 24, 1905–1917
- Rao, S.S.P. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1–16
- Fullwood, M.J. *et al.* (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462, 58–64
- Schoenfelder, S. *et al.* (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* 42, 53–61
- Papantonis, A. *et al.* (2010) Active RNA polymerases: mobile or immobile molecular machines? *PLoS Biol.* 8, e1000419
- Papantonis, A. *et al.* (2012) TNF $\alpha$  signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J.* 31, 4404–4414
- Fanucchi, S. *et al.* (2013) Chromosomal contact permits transcription between coregulated genes. *Cell* 155, 606–620
- Larkin, J.D. *et al.* (2013) Space exploration by the promoter of a long human gene during one transcription cycle. *Nucleic Acids Res.* 41, 2216–2227
- Danko, C.G. *et al.* (2013) Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell* 50, 212–222
- Raj, A. and van Oudenaarden, A. (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216–226
- Plank, J.L. and Dean, A. (2014) Enhancer function: mechanistic and genome-wide insights come together. *Mol. Cell* 55, 5–14
- Andersson, R. (2015) Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays* 37, 314–323
- Andersson, R. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461
- Core, L.J. *et al.* (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* 46, 1311–1320
- Arner, E. *et al.* (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 347, 1010–1014
- Donze, D. and Kamakaka, R.T. (2001) RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in *Saccharomyces cerevisiae*. *EMBO J.* 20, 520–531
- Phillips-Cremins, J.E. and Corces, V.G. (2013) Chromatin insulators: linking genome organization to cellular function. *Mol. Cell* 50, 461–474
- Raab, J.R. *et al.* (2012) Human tRNA genes function as chromatin insulators. *EMBO J.* 31, 330–350
- Raab, J.R. and Kamakaka, R.T. (2010) Insulators and promoters: closer than we think. *Nat. Rev. Genet.* 11, 439–446
- Chopra, V.S. *et al.* (2009) Stalled Hox promoters as chromosomal boundaries. *Genes Dev.* 23, 1505–1509
- Dixon, J.R. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380
- Bon, M. *et al.* (2006) Modeling a self-avoiding chromatin loop: relation to the packing problem, action-at-a-distance, and nuclear context. *Structure* 14, 197–204
- Hnisz, D. *et al.* (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947
- Phillips, J.E. and Corces, V.G. (2009) CTCF: master weaver of the genome. *Cell* 137, 1194–1211
- Zuin, J. *et al.* (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 111, 996–1001
- Yan, J. *et al.* (2013) Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* 154, 801–813
- Seitan, V.C. *et al.* (2013) Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* 23, 2066–2077
- Sofueva, S. *et al.* (2013) Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 32, 3119–3129
- Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676
- Brem, R.B. *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755
- Rockman, M.V. and Kruglyak, L. (2006) Genetics of global gene expression. *Nat. Rev. Genet.* 7, 862–872
- Yvert, G. *et al.* (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* 35, 57–64

- 50 Cheung, V.G. *et al.* (2010) Polymorphic cis- and trans-regulation of human gene expression. *PLoS Biol.* 8, e1000480
- 51 Gaffney, D.J. (2013) Global properties and functional complexity of human gene regulatory variation. *PLoS Genet.* 9, e1003501
- 52 Wood, A.R. *et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46, 1173–1186
- 53 The GTEx Consortium (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660
- 54 Veyrieras, J.B. *et al.* (2008) High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* 4, e1000214
- 55 Stranger, B.E. *et al.* (2012) Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.* 8, e1002639
- 56 Hemani, G. *et al.* (2014) Detection and replication of epistasis influencing transcription in humans. *Nature* 508, 249–253
- 57 Duggal, G. *et al.* (2014) Higher-order chromatin domains link eQTLs with the expression of far-away genes. *Nucleic Acids Res.* 42, 87–96
- 58 Thévenin, A. *et al.* (2014) Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. *Nucleic Acids Res.* 42, 9854–9861
- 59 Palstra, R.J. *et al.* (2008) Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS ONE* 3, e1661
- 60 Mitchell, J.A. and Fraser, P. (2008) Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* 22, 20–25
- 61 Cook, P.R. and Brazell, I.A. (1976) Conformational constraints in nuclear DNA. *J. Cell Sci.* 22, 287–302
- 62 de Wit, E. *et al.* (2013) The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* 501, 227–231
- 63 Harper, C.V. *et al.* (2011) Dynamic analysis of stochastic transcription cycles. *PLoS Biol.* 9, e1000607
- 64 Bryant, J.A. *et al.* (2014) Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* 42, 11383–11392
- 65 Le, T.B. *et al.* (2013) High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* 342, 731–734
- 66 Xie, T. *et al.* (2015) Spatial features for *Escherichia coli* genome organization. *BMC Genomics* 16, 37
- 67 Rippe, K. (2001) Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* 26, 733–740
- 68 Brackley, C.A. *et al.* (2013) Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. U.S.A.* 110, E3605–E3611
- 69 Marenduzzo, D. *et al.* (2006) The depletion attraction: an underappreciated force driving cellular organization. *J. Cell Biol.* 175, 681–686
- 70 Kuhlman, T.E. and Cox, E.C. (2012) Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. *Mol. Syst. Biol.* 8, 610
- 71 Hirano, Y. *et al.* (2012) Lamin B receptor recognizes specific modifications of histone H4 in heterochromatin formation. *J. Biol. Chem.* 287, 42654–42663
- 72 Kind, J. and van Steensel, B. (2014) Stochastic genome–nuclear lamina interactions: modulating roles of Lamin A and BAF. *Nucleus* 5, 124–130
- 73 Harr, J.C. *et al.* (2015) Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J. Cell Biol.* 208, 33–52
- 74 Cook, P.R. and Marenduzzo, D. (2009) Entropic organization of interphase chromosomes. *J. Cell Biol.* 186, 825–834
- 75 Kadauke, S. and Blobel, G.A. (2013) Mitotic bookmarking by transcription factors. *Epigenet. Chromatin* 6, 6