Seminars in Cancer Biology 23 (2013) 65–71

Contents lists available at SciVerse ScienceDirect

Seminars in Cancer Biology

journal homepage: www.elsevier.com/locate/semcancer

Review

Transcription factories, chromatin loops, and the dysregulation of gene expression in malignancy

Binwei Deng, Svitlana Melnik[1], Peter R. Cook*

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

ARTICLE INFO

Keywords:
Chromatin loop
Enhancer
Factory
RNA polymerase
Silencer
Transcription

ABSTRACT

Pathologists recognize and classify cancers according to nuclear morphology, but there remains little scientific explanation of why malignant nuclei possess their characteristic features, or how those features are related to dysregulated function. This essay will discuss a basic structure–function axis that connects one central architectural motif in the nucleus–the chromatin loop–to the vital nuclear function of transcription. The loop is attached to a "transcription factory" through components of the transcription machinery (either polymerases or transcriptional activators/repressors), and the position of a gene within a loop determines how often that gene is transcribed. Then, dysregulated transcription is tightly coupled to alterations in structure, and vice versa. We also speculate on how the experimental approaches being used to analyze loops and factories might be applied to study the problems of tumour initiation and progression.

1. Introduction

Pathologists recognize and classify cancers according to nuclear morphology but there remains little scientific explanation of why malignant nuclei possess their characteristic features, or how those features are related to dysregulated function. This essay will discuss a basic structure-function axis that connects one central architectural motif in the nucleus–the chromatin loop–to the vital nuclear function of transcription. We will argue that there is a direct and immediate connection between the two: the transcription machinery defines the loop, and the position of a gene in the loop directly influences how often that gene is transcribed. We also suggest (like others) that there is an immediate connection between many malignancies and dysregulated transcription. First, just think of the roles that transcription factors like p53 [1], the Smads [2], and nuclear factor κB (NFκB; [3]) play in the progression of many tumours. Second, it is now forty years since the two-hit hypothesis explained the role of a recessive tumour-suppressor gene in one dominantly-inherited and rare cancer–retinoblastoma; complete inactivation of both gene copies was all that was required [4]. But with frequently-seen cancers, we now know that just a partial inactivation of relevant tumour-suppressor genes–often achieved through dysregulated transcription–is the critical driver [5]. We will outline the evidence for such a loop-transcription axis, before speculating on ways that changes in this structure-function axis might underpin the changes recognized by pathologists.

2. An alternative model for transcription

Our students are still taught that an RNA polymerase transcribes by diffusing to a gene wherever that gene might be in the nucleus, and then–once the polymerase initiates–it tracks down the template as it makes its transcript. In separate lectures, these students are told of the way the DNA fibre is organized in 3D nuclear space–for example, by a first wrapping around a nucleosome and then a final folding into a chromosomal territory. However, little is usually said about the inter-connection between structure and function at the intervening levels in the organization, other than a passing reference to the activity of "open" euchromatin and the inactivity of tightly-packed heterochromatin. We first discuss recent evidence supporting an alternative view of how transcription occurs–one that involves a direct and immediate connection between structure and function: the transcription machinery defines the structure, and the structure directly influences how often a gene is transcribed.

Abbreviations: 3C, chromosome conformation capture; CTCF, CCCTC-binding factor; DamID, DNA adenine methylation identification; DMR, differentially methylated domain; ES cells, embryonic-stem cells; HUVECs, human umbilical endothelial vein cells; KLF1, Kruppel-like factor 1; LAD, lamin-associated domain; LOCK, large organized chromatin lysine modification; NFκB, nuclear factor B; NOR, nucleolar organizing region; Smad, a protein that combines properties of the *Caenorhabditis elegans* protein; SMA, with those encoded by the "*mothers against decapentaplegic*" or *MAD* gene of the fly; TNFα, tumour necrosis factor; UBF, upstream binding factor.

* Corresponding author. Tel.: +44 1865 275528; fax: +44 1865 275515.
E-mail address: peter.cook@path.ox.ac.uk (P.R. Cook).
[1] Present address: Division of Molecular Biology of the Cell II, German Cancer Research Center, Heidelberg, Germany.

1044-579X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.semcancer.2012.01.003

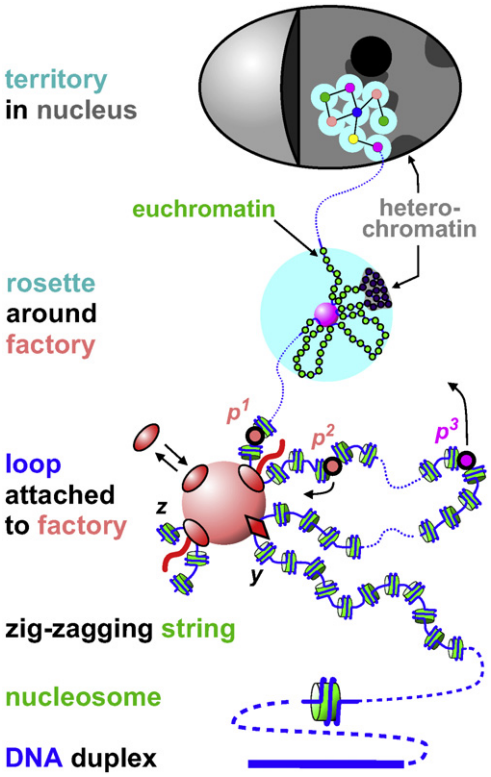

Fig. 1. A model for all genomes. In man, DNA is coiled into the nucleosome, and runs of nucleosomes form a zig-zagging string looped by attachment to factories through transcription factors (diamond) and engaged polymerases (ovals). In HeLa cells, the average contour length of a loop is 86 kbp (range 5–200 kbp), and the core of a nucleoplasmic factory has a diameter of ~90 nm and a mass ~10 MDa. The promoter, p^1, has just initiated, and a fixed polymerase in the factory is reeling in its template as it extrudes its transcript; this initiation has generated a new loop that extends to the right of the large pink factory. Components in a factory exchange continually with the soluble pool. About 16 such loops (only a few are shown) form a rosette around a factory; roughly half the attachments are mediated by active polymerases, half by transcription factors. Distal nucleosomes in long loops tend to be static and acquire a (heterochromatic) histone code that spreads down a fibre; they also aggregate on to the lamina, nucleoli, and chromocenters. A string of 30–180 successive rosettes forms a territory (the general path of DNA is shown). Different factories (circles of different colours) specialize in transcribing different sets of genes. Active transcription units that are near neighbors often form rosette-like structures, but the structure can be more complex; for example, *y* may be distant from *z* on the genetic map (which would generate a giant loop), and might perhaps be on a different chromosome. The transcription unit driven by p^1 "enhances" the activity of p^2 (by tethering it close to the pink factory) while "silencing" the activity of p^3 (by tethering it distant from the purple factory). This model is general in the sense that it can be applied to all genomes; for example, the genomes of yeast [69] and bacteria [70]) are now known to be tied into loops through components of the transcription machinery. Modified from [71].

Structures known as transcription factories [6–8] play a central role in this alternative view (Fig. 1). A "factory" is defined in The Oxford English Dictionary as '*a building or range of buildings with plant for the manufacture of goods.*' We define a transcription factory as a site containing at least two (usually more) polymerases and associated plant active on at least two (usually more) different templates. The raison d'être for all kinds of factories is the same–whether making jam, cars, or RNA: to enhance production by concentrating the relevant machines and raw materials in one place. For example, the nuclei of human cells contain a 1-μM pool of RNA polymerase II, but essentially all its transcripts are made in factories where the local concentration is ~1000-fold higher. A second important feature of this alternative view is the immobilization of engaged RNA polymerases; enzymes fixed to the surface of a factory reel in their templates as they extrude their transcripts and the transcribed DNA. These two features have prevented acceptance of this alternative view, simply because they run so counter to what we were taught.

A third feature of the model complicates analysis: the structure of each factory in a nucleus probably differs from that of all other factories, and it changes from moment to moment. Cell-to-cell variability is an inevitable consequence of the transcription of a typical gene being noisy, stochastic, and infrequent, with different alleles in different cells firing at different times [9]; then, loops inevitably appear/disappear as polymerases initiate/terminate, and transcription factors bind/dissociate. Even the core of one factory probably differs from all other cores. Thus, macromolecular structures the size of factories can be generated in two fundamentally different ways [10,11]. Some virus particles "self-assemble" to a fixed plan to attain a true thermodynamic equilibrium; the particles are stable and static, and can survive in the absence of a pool of unincorporated subunits once released from the host. On crystallization, every atom is found in the same place in the unit cell. But structures like the cytoskeleton are "self-organizing". They will never be crystallized simply because each one is different from all others, and each is intrinsically unstable, persisting only by exchanging subunits with others in their surroundings in an ATP-dependant manner; if those subunits (or the fuel) are depleted, they collapse and eventually disappear. Although sub-assemblies within factories like the polymerase may self-assemble, we suggest the larger factories will self-organize. Then, statements about their structure will necessarily be probabilistic rather than absolute.

3. Specialized transcription factories

Human nuclei contain three different kinds of RNA polymerase (i.e., polymerases I, II, and III [12]), and each is concentrated in a different kind of factory that transcribes a specific subset of genes to the exclusion of others. Thus, polymerase I is concentrated in nucleoli where it transcribes the repeated ribosomal cistrons, while the active forms of polymerases II (transcribing all protein-coding genes and some non-coding RNAs) and III (making non-coding RNAs like tRNA) are found in distinct nucleoplasmic factories [13]. Such specialization can be demonstrated using two mini-chromosomes carrying essentially identical transcription units driven by polymerase II; these are transcribed in the same set of factories. However, replacing one of the promoters with a polymerase I promoter (or a polymerase III promoter) now targets that mini-chromosome to another group of factories [14]. Other results indicate that distinct polymerase II factories further specialize in transcribing different genes [15]. For example, inserting an intron into an intron-less protein-coding gene now targets that gene to a different factory–presumably one that specializes in splicing [14]. In addition, factories transcribing genes encoding interleukins [16], various subunits of cytochrome c [17], and factors involved in globin production [18–20] have now been uncovered (amongst others).

4. Promoters go to a polymerase to be transcribed (not vice versa)

This alternative view requires that the promoter goes to a polymerase to be transcribed (rather than vice versa). But are genes sufficiently mobile, especially when packaged into the dense nucleus? Many live-cell experiments confirm they are [21]. For example, one study involved a Chinese hamster cell containing repeats of the lac operator integrated into one site in the genome [22]. [Integrants in other sites gave similar results.] The cell also expressed the lac repressor tagged with the green fluorescent protein so that tagged repressor bound to the operator could be followed in the living cell (Fig. 2). It appeared to diffuse throughout

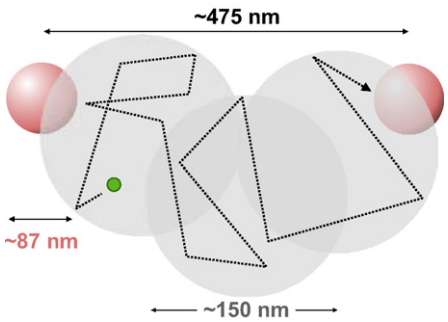

Fig. 2. Cartoon illustrating the way a promoter (the green spot) apparently diffuses through the nucleoplasm of a living cell, jumping from one ~150-nm "corral" (grey circle) to another, and so able to visit two different factories (pink spheres). Only the green spot is visible under the microscope. Data from [22].

a 250-nm "corral", and then–every minute or so–this spot "jumped" ~150 nm (over a period of ~1 s) into another "corral" [22]. This is the behavior expected of a tagged segment of a long polymer diffusing throughout a space bounded by other invisible and diffusing segments; as the confining fibres diffuse apart, the tagged segment escapes ("jumps") through the gap into a neighboring space that is–in turn–surrounded by additional confining fibres. Given the known dimensions of a factory (~87–130 nm diameter [23,24] and the known inter-factory spacing (calculated from the density of factories in a known nucleoplasmic volume [25]), a promoter will be able to visit several factories every few minutes. Then, the promoter will be most likely to initiate in the factory containing the appropriate transcription factors. Unfortunately, it has not yet been possible to resolve individual nucleoplasmic factories from each other in a living cell, as they are too small and numerous.

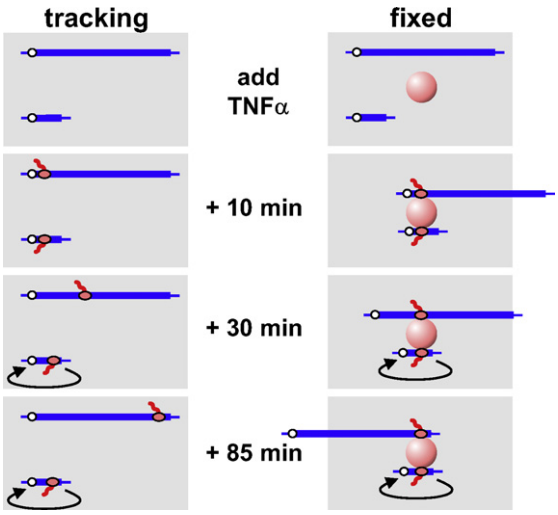

Fig. 3. Tracking v fixed polymerases. Left: before adding TNFα, both long and short genes are inactive. 10 min after adding the cytokine, RNA polymerases (ovals) initiate on both genes, and track down their templates. After 30 min, the one on the short gene has terminated and another has initiated (cycles of initiation and termination are indicated by the circular arrow); the pioneering polymerase is still tracking down the long gene. After 85 min, the pioneering polymerase on the long gene has reached the terminus, and yet another is transcribing the short gene. As the two genes lie on different chromosomes, no 3C product is seen at any time. Right: after adding the cytokine, the two genes diffuse to a factory specializing in transcribing TNFα-responsive genes, where they initiate; therefore, after 10 min the two promoters are now seen together. After 30 min, promoter:promoter contacts are lost, and the short gene now contacts one third of the way into the long gene. After 85 min, the short gene now contacts the terminus of the long gene. During the 85 min, only regions of the two genes that are being transcribed at that moment should yield a 3C product. This is the result obtained [27].

5. Active polymerases are immobile molecular machines

This alternative view also requires that the elongating polymerase is attached to a factory, so that it cannot track down the template. As with so many received ideas, there seems to be little (if any) evidence supporting such tracking in vivo. Indeed the idea seems to stem originally from a perception of relative size–we inevitably think that the smallest object (i.e., the polymerase) must move. But we now know that each mRNA-producing complex in a human cell contains (at least) the polymerase, nascent transcript plus associated proteins, and spliceosome–with diameters of at least 15, 14, and 25 nm, respectively [21]. Clearly, the active polymerizing complex is huge, and it is difficult to imagine how such a large complex could force its way along the template through the viscous nucleoplasm. We suggest the alternative is more likely: the template travels end on through the nucleoplasm as it is reeled in by a fixed polymerase–so the template with its small cross-section takes the path of least frictional resistance through the dense nucleoplasm. We also now know that fixed polymerases are powerful molecular motors able to reel in their templates in this way, with many single-molecule analyses relying on such enzyme immobilization [26]. [Of course, tracking and fixed polymerases use the same fuel–nucleotide triphosphates–to drive movement.] Recent evidence now confirms–indirectly, but nevertheless decisively–that RNA polymerases are indeed immobile when active.

The experiment involved two genes that can be switched on rapidly (Fig. 3; [27]). One short human gene of 10 kbp (*TNFAIP2*) was used as a reference point, while the other – 221-kbp *SAMD4A*–was long enough to provide the necessary spatial resolution. These two genes lie ~50 Mbp apart on the genetic map and so are not expected ever to lie near each other in 3D nuclear space. The powerful cytokine, tumour necrosis factor α (TNFα), was used as the gene switch; when applied to human umbilical endothelial vein cells (HUVECs), it signals through NFκB to activate and repress many genes, including these two which encode regulators of the signaling pathway. Both genes are turned on within 10 min, but *TNFAIP2* is so short that the pioneering polymerase soon terminates; subsequently, additional polymerases transcribe it repeatedly over the next few hours. However, *SAMD4A* is so long that the pioneering polymerase only reaches the terminus ~85 min after stimulation. Proximity between the two genes was monitored between 0 and 85 min by chromosome conformation capture (3C; [28]).

If the conventional model for transcription applies, we would not expect the short gene to lie near the long gene at any time after stimulation (Fig. 3, left). But, if both responding genes are transcribed by polymerases that are transiently immobilized in the same "NFκB-factory" that specializes in transcribing TNFα-responding genes, the short gene–which repeatedly attaches to (and detaches from) the factory as it initiates (and terminates)–should always lie close to just the part of the long gene being transcribed at that particular moment. In other words, we would not expect to see any 3C products before stimulation (when both genes are inactive, and so unassociated with a factory). But after 10 min (when both initiate), the short gene should lie next to the promoter of the long gene (but no other part). And as the polymerase reels in the long gene, successive parts of that gene will be brought successively into the factory to lie transiently next to the short gene. After 85 min, when the pioneering polymerase on *SAMD4A* is about to terminate, only the terminus should lie next to the short gene. 3C products appear and disappear in this way, and the use of gene pairs on a different chromosome gives similar results [27]. Moreover, an independent method–RNA fluorescence in situ hybridization (FISH) used with probes targeting intronic sequences–confirms that the relevant nascent RNAs lie together at the appropriate times. Then,

"super-resolution" microscopy demonstrated that the pairs of nascent transcripts colocalize to the degree expected if they were randomly distributed within a 35-nm shell around an 87-nm factory. These experiments show that the DNA must move (not the polymerase), and that the two responsive genes congregate in the same specialized factory when transcribed.

6. Components of the transcription machinery tie the chromatin fibre in loops

Loops attached to factories through components of the transcription machinery–either polymerases or transcription factors (which might be activators or repressors)–constitute the central architectural feature of the model in Fig. 1. The widespread application of techniques like 3C, FISH, and DNA adenine methylation identification (DamID)–all powerful ways of detecting which sequences tend to lie next to each other in 3D nuclear space [29]–are confirming this organization. Thus, the one general feature that is emerging from the analysis of complete interactomes is that contacting DNA sequences either encode active transcription units, or binding sites for transcriptional activators/repressors like Mediator and the CCCTC-binding factor (CTCF; see, for example, [18,30–35]; reviewed in [36]). Therefore it is becoming increasingly accepted that active transcription units and their regulatory motifs are the sequences that stabilize loops. And as might be expected, contacts between heterochromatic regions, and between heterochromatin and the lamina, are superimposed on this organization [37,38].

7. Isolation of large fragments of factories

RNA polymerases can easily be purified from human cells using buffers containing unphysiological concentrations of salts, and the isolated polymerases can initiate–albeit poorly–on added templates [12]. But when cells are permeabilized in isotonic buffers, essentially all polymerases that are currently elongating remain tightly bound to the sub-structure [39,40]; as a result, they remain ill-characterized. [The large pool of soluble enzyme is easily extracted, but this pool is inactive.] Fortunately, it is now possible to partially purify large fragments of factories, and the associated active enzymes. The trick is to use caspases (which deconstruct nuclei during apoptosis) to release the fragments from the substructure (Fig. 4; [41]). [All subunits of the nuclear RNA polymerases lack sites recognized by the caspases used, except RPB9.] HeLa cells were permeabilized in a "physiological buffer", nuclei isolated and treated exhaustively with DNase I. After spinning to leave inactive chromatin in the supernatant, the resuspended pellet was treated with caspases, respun, and the supernatant retreated with DNase. Despite template truncation, ~50% of the nascent RNA and endogenous elongating capacity still remain. Electrophoresis in "blue native gels" then allowed resolution of three partially-overlapping complexes (named complex I, II, or III after the polymerases they contain); all ran slower than the largest (8 MDa) marker available. Finally, mass spectrometry showed that all complexes shared proteins like RNPs, while each possessed a characteristic set of others. For example, 83% proteins in complex I were also in the nucleolar proteome, while complex II contained 5 polymerase II subunits plus various transcription factors (e.g., AP-2, CEBPB, CTCF), epigenetic modifiers (e.g., histone-lysine N–methyl transferases EZH2, SUV39H1/2), components involved in adding 5′ caps and 3′ tails to mRNAs, and components of the proof-reading (i.e., nonsense-mediated decay) machinery. Each complex also contained the expected class of RNAs (e.g., complex I contained ~33-fold more nascent 45S rRNA, while complex II was richer in nascent mRNAs). Although these complexes disintegrated and aggregated when

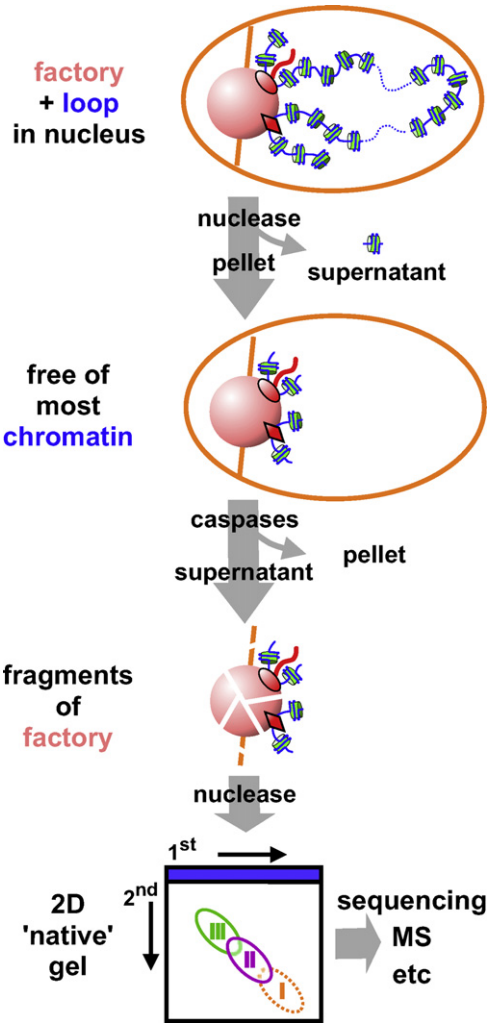

Fig. 4. A method for isolating large fragments of factories [41]. Cells are permeabilized in a "physiological" buffer; the cartoon on the top shows a chromatin loop with nucleosomes (green circles) tethered to a polymerizing complex (oval) in a factory (pink sphere), which is attached to the nuclear substructure (brown). Most chromatin is now detached with a nuclease (usually DNase I), chromatin-depleted nuclei washed, and polymerizing complexes released from the substructure with caspases; the resulting large fragments of factories are now retreated with a nuclease (usually DNase I), and partially resolved in a 2D gel (using "blue native" and "native" gels in the 1st and 2nd dimensions). Rough positions of >8-MDa complexes containing polymerases I, II, and III are shown. Finally, different regions can be excised, and their content analyzed (e.g., DNA and RNA by sequencing, proteins by mass spectrometry, MS). Use of a restriction enzyme instead of DNase I allows fragments to be isolated associated with more DNA; then, "native 3C"–which involves adding ligase to the excised fragments of the gel–permits analysis of which DNA sequences lie close together in the factories.

recovered from the 2D gels, this isolation procedure opens up the possibility of characterizing the factories–the active sites of transcription–in greater detail. If a restriction enzyme replaces DNase I during isolation, fragments associated with more DNA can be isolated, and this additionally allows analysis of transcribed sequences.

8. The general structure of transcription factories seen in fixed cells

Nucleolar factories are the best characterized. Human rDNA loci are carried on chromosomes 13, 14, 15, 21 and 22; each locus encodes many tens of tandem repeats of the 45S rRNA gene and these repeats form a secondary constriction–the nucleolar organizing region (NOR)–in the mitotic chromosome. Inactive RNA

polymerase I and its transcription factor UBF (upstream binding factor) are bound to some NORs, and when the cell exits from mitosis, these NORs fuse to form one or more nucleoli [42]. NORs lacking bound UBF and polymerase remain inactive and do not fuse [43]. In other words, the bound polymerase and UBF "bookmark" some NORs so the system can restart transcription on the same genes that were being transcribed during the previous cell cycle. Later during interphase, nascent rRNA is found in the "dense fibrillar component" on the surface of a protein-rich core containing polymerase I and UBF–the "fibrillar centre"; newly-completed transcripts are then processed in a surrounding "granular component" to emerge as mature ribosomal subunits into the nucleoplasm [44].

Both the number of nucleoli and the number of active units on the surface of each fibrillar centre (the factory core) vary greatly in different cell types, but a HeLa cell typically contains about 4 transcription units/core–each being transcribed by ~125 tightly-packed polymerases [44,45]. Then the core has ~500 active polymerizing machines on the surface, and we imagine that as each promoter is extruded from one machine, that promoter is immediately captured (to reinitiate) by a neighboring machine. The iconic image of a "Christmas tree" showing one 45S rRNA gene being transcribed by ~125 polymerases [46] is then obtained by stripping a unit off the core, and spreading it on the grid used for electron microscopy.

Although smaller, nucleoplasmic factories seem to follow the same general design principles. For example, nanoscale mapping of phosphorus and nitrogen in/around such factories in HeLa cells using a special electron microscope again reveals templates and transcripts on the surface of a protein-rich core; this core has a diameter of ~87 nm, a mass ~10 MDa, and a density about one-tenth that of the nucleosome–and so must be porous [23]. In fetal liver erythroblasts, factories are larger (diameter ~130 nm, mass ~26 MDa; [24]). Like their counterparts in nucleoli, nucleoplasmic factories are polymorphic and exhibit a roughly normal distribution of diameters about the mean.

There are only a few estimates of the total number of nucleoplasmic factories per cell, and fewer credible ones obtained using conditions where essentially all factories are detected. For example, several hundred foci containing the active form of RNA polymerase II can be seen in mouse erythroid cells by immunofluorescence [24,30]. However, this should be treated as a minimum estimate, as an unknown number of factories could lie below the level of detection. Higher estimates have been obtained using conditions where most nucleoplasmic factories are detected–between 3900 and 15,000 in diploid and totipotent mouse embryonic stem (ES) cells that differentiate along various developmental pathways, and ~10,000 in the large, rapidly-dividing (tetraploid) HeLa cell [13,25,45,47]. In HeLa, one fifth of these are polymerase III factories, while the remainder contain polymerase II [13]. To make these counts, engaged polymerases in permeabilized cells were allowed to extend their transcripts by a few nucleotides in BrUTP or biotin-CTP, the resulting nascent Br-/biotin-RNA immunolabeled with fluors (or gold particles), and fluorescent foci (or clusters of gold particles) marking the factories imaged in the light (or electron) microscope. If some factories lay below the level of detection, incorporation of twice as much BrUTP or biotin-CTP should lead to the detection of more factories (as some previously undetected ones acquire more label to rise above the background). However, conditions were chosen such that doubling the incorporation doubled the intensity of labeling of each focus/cluster without increasing factory number–a requirement if all factories are being detected. Marking sites with tagged precursors in this way has the advantage that each polymerizing complex contains one nascent transcript with many tagged residues; then, it is much easier to immunodetect the multiple tags compared to the single epitope in a polymerase. It also turns out that the approach targeting BrRNA is sensitive, as only one-twentieth the amount in a typical factory can be detected [13,47]. As <5% gold particles marking nascent BrRNA are also found outside the clusters, this puts an upper limit of 5% on the amount of extra-factory synthesis that might occur [13,47].

What happens to factory number and size as cells differentiate? Nucleolar factories grow and shrink in response to the demand for ribosomes. Thus, the ~234 fibrillar centres (equivalent to factory cores) in a fibroblast fall to ~156 on serum-starvation [48], while the ~9 in a peripheral blood lymphocyte rise to ~80 as it is stimulated to divide [49]. Then, increasing transcription correlates with an increase in surface area of the core as it splits, and–as that core is packed with polymerases–there will be an increase in number of polymerases accessible to promoters on the surface.

In contrast, the clusters of gold particles marking nucleoplasmic factories remain roughly constant in diameter as mouse ES cells differentiate into larger, more active, cells [47]. Thus, as transcriptional activity increases, nucleoplasmic volume and factory number increase, as factory density remains constant and the contour lengths of loops decrease. Conversely, as activity falls, nucleoplasmic volume and factory number decrease, and loops become longer; as genes in these long loops are remote from factories, they become heterochromatic and so will pack together more tightly (with the result that factory density can still remain constant despite the presence of more DNA between each factory). This constancy in diameter and density is maintained in much larger newt nuclei containing an 11-fold larger genome and three-fold more active polymerases [47]. Again it seems that increasing activity is accompanied by an increase in factory number, rather than factory size, and–again–self-organization directly couples transcription to nuclear volume.

Do nucleoplasmic factories of different types have different sizes? Some do. We have seen that in fetal liver erythroblasts, factories have mean diameters and masses of ~130 nm and ~26 MDa, but the sub-set associated with the Kruppel-like transcription factor 1 (KLF1) are larger (~174 nm, ~36 MDa), and those with nascent globin transcripts larger still (~38 MDa; [24]).

How many different transcription units are found in one factory at any moment? We can calculate the answer after counting the number of factories, nascent transcripts and/or active polymerases, as well as individual polymerases engaged on each transcription unit. Nucleoli again provide the precedent. We have seen above that the larger the fibrillar centre (the core of the factory), the more dense fibrillar components (transcription units) tend to be associated with that factory on the surface. The similarly-sized nucleoplasmic factories in HeLa, mouse ES cells, and newt cells are all associated with ~8 active polymerases, each engaged on a different unit [13,25,45,47]. In contrast, the larger factories seen in mouse fetal liver seem to be associated with more active transcription units [24]–but this conclusion is based only on a guesstimate of the numbers of KLF1-responsive genes that might be active at any moment (rather than an accurate count). Given the normal distribution in diameter of nucleoplasmic factories [13,24,25], and the finding that most active polymerase II units are only being transcribed by one polymerase at any moment [9,45,50], the general consensus is that few "factories" will contain only one active transcription unit.

9. Gene activation and repression–enhancers, silencers, and bursting

Consider Fig. 1, where promoter p^2 is tethered closer to the factory than promoter p^3. Intuition then suggests that p^2 is more likely than p^3 to collide with the factory, and so initiate. Monte Carlo simulations of a model loop confirm this intuition [51]. Now consider the transcription unit driven by p^1. If p^2 requires

transcription factors that are found in this factory, p^1 "enhances" the activity of p^2 (by tethering p^2 close to a factory rich in the appropriate factors). But if p^3 requires "purple" factors concentrated in the purple factory above, p^1 will now "silence" p^3 (by tethering it away from the purple factory). So here p^1 acts either as an enhancer or silencer of different genes. This kind of argument, coupled to early evidence showing that the canonical motifs were transcribed, led to the suggestion that enhancers and silencers (and by extension other motifs like barriers and insulators) are all active transcription units that work by bringing their target promoters close to (or distant from) the appropriate factory [52]. Recent local and genome-wide studies now amply confirm that these regulatory motifs are indeed transcription units, that they loop back to contact their targets, and that they can often act as one or the other depending on the "context" [53–59].

Initiations do not occur completely randomly; rather, a gene might be silent for a while, before producing a brief "burst" of transcripts [9,60]. Such bursting is usually attributed to changes in binding of transcription factors, and/or chromatin remodeling or modification (e.g., by acetylation, phosphorylation, or methylation of nucleosomes). We suggest proximity to a factory will be another major factor. A gene might initially lie distant from a factory (and so silent), but–as genes around it are transcribed–it might be brought closer to a factory to increase its chances of initiating. And then, once it detaches after the first transcription cycle, it will still lie close to a factory that contained all the appropriate transcription factors–and this will make it more likely to re-initiate and so give a burst [9].

10. The development of specialized factories

We now speculate on how specialized factories might evolve from preexisting "naïve" ones, using "NFκB-factories" that transcribe TNFα-responding genes as an example. We imagine that responsive genes are "poised" prior to activation [61] at/near factories so they can respond rapidly (like promoter p^2 in Fig. 1). Before stimulation, the relevant promoters visit these factories every few minutes (Fig. 2), but few will initiate as the concentration of NFκB is low. But when stimulation induces activated NFκB to enter the nucleus, the transcription factor binds to responsive promoters, and this stabilizes attachment to one or other of these factories. Once productive transcription begins, responsive genes become stably tethered to the factory. As more responsive genes bind, the local concentration of NFκB will increase, so that the factory will evolve into one that predominantly–but not exclusively–transcribes TNFα-responsive genes.

11. Some speculations on cancer

We have seen how the transcription machinery defines the loop, and how position in a loop determines gene expression. Although we are a long way from uncovering how this structure-function axis is dysregulated during malignancy to yield the changes in nuclear morphology recognized by pathologists, it is nevertheless productive to try and answer the question: How might the model in Fig. 1 affect the experiments that cancer biologists do? We can suggest two general approaches that can be used to address the problems of tumour initiation and progression, respectively. For the first, we will use the *Myc* gene as an example. The *Myc* and *Igh* loci are the most frequent translocation partners in plasmacytoma and Burkitt lymphoma. Both loci are often found together in 3D nuclear space in the relevant cells [62], and 3C and FISH show they often share the same specialized factory [63]–presumably because they require the same transcription factors. Then, it is easy to imagine that tethering in close proximity should increase the frequency with which the critical initiating translocation occurs [64,65]. Therefore, it would seem sensible to analyze systematically whether the translocation partners involved in initiating other malignancies are similarly co-transcribed in specialized factories, and–if they are–characterize the genes, transcripts, and proteins associated with those factories using the approaches described above. The same argument applies to other genetic rearrangements occurring during malignancy like somatic copy number alterations [66,67].

A central problem faced when studying tumour progression is: how can we pick out the critical players (whether they be proteins, genes, or micro-RNAs) involved in the progression from the multitude of innocent bystanders? Here we will again use signaling through NFκB as an example. We have seen that on stimulation with the cytokine, responsive genes are transcribed in specialized "NFκB" factories. We suggest that an excellent way of achieving the required discrimination is to select a rapidly-responding gene like *SAMD4A*, and see which other sequences are co-transcribed with it in these "NFκB" factories. We anticipate some will act as enhancers and silencers (as in Fig. 1), and still others will encode key regulators of the pathway (whether they be proteins or micro-RNAs). It also becomes attractive to try and develop ways of separating these "NFκB" factories from the general pool, so that we can characterize their proteomes, and associated nucleic acids (DNA, coding RNAs, and non-coding RNAs). Isolation of the analogous factories from tumours would then allow is to home in on the critical players that are dysregulated in malignancy. We also imagine that groups of genes that tend to share the same factories will also tend to share the same epigenetic marks, and that these marks might extend over all the loops associated with a factory–and this would provide a physical basis for the Mbp-sized domains that are altered in malignancy [68], including differentially-methylated regions (DMRs), lamin-associated domains (LADs), and large organized chromatin lysine modifications (LOCKs).

Conflict of interest

None.

Acknowledgements

Work in PRC's lab described here was supported by the Biotechnology and Biological Sciences Research (through ERASysBio+/FP7 and ERASysBio+), the Medical Research Council, and the Wellcome Trust; PRC is the E.P. Abraham Professor of Cell Biology, and a fellow of Lincoln College.

References

[1] Junttila MR, Evan GI. p53–a Jack of all trades but master of none. Nat Rev Cancer 2009;9:821–9.
[2] Meulmeester E, Ten Dijke P. The dynamic roles of TGF-$ in cancer. J Pathol 2011;223:205–318.
[3] Staudt LM. Oncogenic activation of NF-kappaB. Cold Spring Harb Perspect Biol 2010;2:a000109.
[4] Knudson AG. Antioncogenes and human cancer. Proc Natl Acad Sci USA 1993;90:10914–21.
[5] Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. Nature 2011;476:163–9.
[6] Cook PR. A chromomeric model for nuclear and chromosome structure. J Cell Sci 1995;108:2927–35.
[7] Chakalova L, Fraser P. Organization of transcription. Cold Spring Harb Perspect Biol 2010;2:a000729.
[8] Cook PR. A model for all genomes; the role of transcription factories. J Mol Biol 2010;395:1–10.
[9] Finan K, Cook PR. Transcriptional initiation: frequency, bursting, and transcription factories. In: Rippe K, editor. Genome organization and function in the cell nucleus. KGaA: Wiley-VCH Verlag GmbH & Co.; 2011. p. 235–54.
[10] Misteli T. The concept of self-organization in cellular architecture. J Cell Biol 2001;155:181–5.
[11] Cook PR. Predicting three-dimensional genome structure from transcriptional activity. Nat Genetics 2002;32:347–52.

[12] Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, et al. Structure of eukaryotic RNA polymerases. Annu Rev Biophys 2008;37:337–52.
[13] Pombo A, Jackson DA, Hollinshead M, Wang Z, Roeder RG, Cook PR. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. EMBO J 1999;18:2241–53.
[14] Xu M, Cook PR. Similar active genes cluster in specialized transcription factories. J Cell Biol 2008;181:615–23.
[15] Bartlett J, Blagojevic J, Carter D, Eskiw C, Fromaget M, Job C, et al. Specialized transcription factories. Biochem Soc Symp 2006;73:67–75.
[16] Cai S, Lee CC, Kohwi-Shigematsu T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat Genet 2006;38:1278–88.
[17] Dhar SS, Wong-Riley MTT. Chromosome conformation capture of transcriptional interactions between cytochrome c oxidase genes and genes of glutamatergic synaptic transmission in neurons. J Neurochem 2010;115:676–83.
[18] Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet 2010;42:53–61.
[19] Soler E, Andrieu-Soler C, de Boer E, Bryne JC, Thongjuea S, Stadhouders R, et al. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes Dev 2010;24:277–89.
[20] Noordermeer D, de Wit E, Klous P, van de Werken H, Simonis M, Lopez-Jones M, et al. Variegated gene expression caused by cell-specific long-range DNA interactions. Nat Cell Biol 2011;13:944–51.
[21] Wachsmuth M, Caudron-Herger M, Rippe K. Genome organization: balancing stability and plasticity. Biochim Biophys Acta 2008;1783:2061–79.
[22] Levi V, Ruan Q, Plutz M, Belmont AS, Gratton E. Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. Biophys J 2005;89:4275–85.
[23] Eskiw CH, Rapp A, Carter DRF, Cook PR. RNA polymerase II activity is located on the surface of ~87 nm protein-rich transcription factories. J Cell Sci 2008;121:1999–2007.
[24] Eskiw CH, Fraser P. Ultrastructural study of transcription factories in mouse erythroblasts. J Cell Sci 2011;124:3676–83.
[25] Faro-Trindade I, Cook PR. Transcription factories: structures conserved during differentiation and evolution. Biochem Soc Trans 2006;34:1133–7.
[26] Herbert KM, Greenleaf WJ, Block SM. Single-molecule studies of RNA polymerase: motoring along. Annu Rev Biochem 2008;77:149–76.
[27] Papantonis A, Larkin JD, Wada Y, Ohta Y, Ihara S, Kodama T, et al. Active RNA polymerases: mobile or immobile molecular machines? PLoS Biol 2010;8:e1000419.
[28] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science 2002;295:1306–11.
[29] van Steensel B, Dekker J. Genomics tools for unraveling chromosome architecture. Nat Biotechnol 2010;28:1089–95.
[30] Osborne C, Chakalova L, Brown K, Carter D, Horton A, Debrand E, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 2004;36:1065–71.
[31] Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 2006;38:1348–54.
[32] Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 2009;462:58–64.
[33] Göndör A, Ohlsson R. Chromosome crosstalk in three dimensions. Nature 2009;461:212–7.
[34] Ohlsson R, Bartkuhn M, Renkawitz R. CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. Chromosome Res 2010;119:351–60.
[35] Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet 2011;43:1059–65.
[36] Göndör A, Woodbridge AF, Shi C, Aurell E, Imreh M, Ohlsson R. Window into the complexities of chromosome interactomes. Cold Spring Harb Symp Quant Biol 2010;75:493–500.
[37] Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 2008;453:948–51.
[38] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009;326:289–93.
[39] Jackson DA, Yuan J, Cook PR. A gentle method for preparing cyto- and nucleoskeletons and associated chromatin. J Cell Sci 1988;90:365–78.
[40] Jackson DA, Hassan AB, Errington RJ, Cook PR. Visualization of focal sites of transcription within human nuclei. EMBO J 1993;12:1059–65.
[41] Melnik S, Deng B, Papantonis A, Baboo S, Carr IM, Cook PR. The proteomes of transcription factories containing RNA polymerases I, II, or III. Nat Methods 2011;8:962–8.
[42] Dousset T, Wang C, Verheggen C, Chen D, Hernandez-Verdun D, Huang S. Initiation of nucleolar assembly is independent of RNA polymerase I transcription. Mol Biol Cell 2000;11:2705–17.
[43] Roussel P, André C, Comai L, Hernandez-Verdun D. The rDNA transcription machinery is assembled during mitosis in active NORs and absent in inactive NORs. J Cell Biol 1996;133:235–46.
[44] Hozák P, Cook P, Schöfer C, Mosgöller W, Wachtler F. Site of transcription of ribosomal RNA and intranucleolar structure in HeLa cells. J Cell Sci 1994;107:639–48.
[45] Jackson DA, Iborra FJ, Manders EMM, Cook PR. Numbers and organization of RNA polymerases, nascent transcripts and transcription units in HeLa nuclei. Mol Biol Cell 1998;9:1523–616.
[46] Miller OJ, Bakken A. Morphological studies of transcription. Acta Endocrinol Suppl (Copenh) 1972;168:155–77.
[47] Faro-Trindade I, Cook PR. A conserved organization of transcription during embryonic stem cell differentiation and in cells with high C value. Mol Biol Cell 2006;17:2910–20.
[48] Jordan EG, McGovern JH. The quantitative relationship of the fibrillar centres and other nucleolar components to changes in growth conditions, serum deprivation and low doses of actinomycin D in cultured diploid human fibroblasts (strain MRC-5). J Cell Sci 1981;52:373–89.
[49] Haaf T, Hayman DL, Schmid M. Quantitative determination of rDNA transcription units in vertebrate cells. Exp Cell Res 1991;193:78–86.
[50] Bon M, McGowan SJ, Cook PR. Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. FASEB J 2006;20:1721–3.
[51] Bon M, Marenduzzo D, Cook PR. Modeling a self-avoiding chromatin loop: relation to the packing problem, action-at-a-distance, and nuclear context. Structure 2006;14:197–204.
[52] Cook PR. Nongenic transcription, gene regulation and action at a distance. J Cell Sci 2003;116:4483–91.
[53] Geyer PK. The role of insulator elements in defining domains of gene expression. Curr Opin Genet Dev 1997;7:242–8.
[54] Core LJ, Lis JT. Paused pol II captures enhancer activity and acts as a potent insulator. Genes Dev 2009;23:1606–12.
[55] De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 2010;8:e1000384.
[56] Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature 2010;467:430–5.
[57] Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature 2010;465:182–7.
[58] Raab JR, Kamakaka RT. Insulators and promoters: closer than we think. Nat Rev Genet 2010;11:439–46.
[59] Gohl D, Aoki T, Blanton J, Shanower G, Kappes G, Schedl P. Mechanism of chromosomal boundary action: roadblock, sink, or loop. Genetics 2011;187:731–48.
[60] Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. Science 2011;332:472–4.
[61] Ferrai C, Xie SQ, Luraghi P, Munari D, Ramirez F, Branco MR, et al. Poised transcription factories prime silent uPA gene prior to activation. PLoS Biol 2010;8:e1000270.
[62] Roix JJ, McQueen PG, Munson PJ, Parada LA, Misteli T. Spatial proximity of translocation-prone gene loci in human lymphomas. Nat Genet 2003;34:287–91.
[63] Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, et al. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. PLoS Biol 2007;5:e192.
[64] Misteli T. Higher-order genome organization in human disease. Cold Spring Harb Perspect Biol 2010;2:a000794.
[65] Wijchers PJ, de Laat W. Genome organization influences partner selection for chromosomal rearrangements. Trends Genet 2011;27:63–71.
[66] De S, Michor F. DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. Nat Biotechnol 2011;29:1103–8.
[67] Fudenberg G, Getz G, Meyerson M, Mirny LA. High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. Nat Biotechnol 2011;29:1109–13.
[68] Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet 2011;43:768–75.
[69] Tanizawa H, Iwasaki O, Tanaka A, Capizzi JR, Wickramasinghe P, Lee M, et al. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic Acids Res 2010;38:8164–77.
[70] Wang W, Li GW, Chen C, Xie XS, Zhuang X. Chromosome organization by a nucleoid-associated protein in live bacteria. Science 2011;333:1445–9.
[71] Cook PR. Principles of nuclear structure and function. New York: J Wiley and Sons; 2001.