

REVIEW

Form follows function: the genomic organization of cellular differentiation

Steven T. Kosak¹ and Mark Groudine^{1,2,3}

¹Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA; ²Department of Radiation Oncology, University of Washington School of Medicine, Seattle, Washington 98195, USA

The extent to which the nucleus is functionally organized has broad biological implications. Evidence supports the idea that basic nuclear functions, such as transcription, are structurally integrated within the nucleus. Moreover, recent studies indicate that the linear arrangement of genes within eukaryotic genomes is nonrandom. We suggest that determining the relationship between nuclear organization and the linear arrangement of genes will lead to a greater understanding of how transcriptomes, dedicated to a particular cellular function or fate, are coordinately regulated. Current network theories may provide a useful framework for modeling the inherent complexity the functional organization of the nucleus.

Louis Sullivan, whose early efforts helped pioneer the development of the skyscraper, is considered one of the most important architects of the last century. However, it is his dictum—"form ever follows function"—for which he is perhaps best known. Just as this imperative has influenced generations of architects, the idea that structure reflects function provides a useful perspective for a biologist's view of the cell. In many ways, the structure and function of cellular and subcellular organelles are inseparable; that is, disruptions in organelle function can lead to perturbations in its structure. Upon inhibition of rRNA transcription, for example, the nucleolus becomes disordered and ultimately disappears (Leung and Lamond 2003). This integration of structure and cellular function allows for conservation of resources and facilitates regulation at multiple levels.

Although a completely sequenced genome may represent a genetic blueprint, molecular biologists currently lack a key with which to fully grasp how this sequence is related to the development and subsequent maintenance of a given organism. Following Sullivan's example, a comprehensive understanding of genomic sequence may require considering its arrangement in the nucleus; the form DNA takes in the nucleus reveals not only its higher-order structure, but it may impart information

regarding its function. The current paradigm of gene regulation includes the binding of site-specific transcription factors, the recruitment of cofactors and general transcription factors, and the incorporation of multiple modifications to both the DNA and the histones that organize it (Felsenfeld and Groudine 2003). This description of transcription belies its enormous complexity, fueled by an ever-increasing catalog of proteins dedicated in one way or another to its regulation. Additionally, evidence supporting the role of nuclear localization in transcriptional regulation indicates that it is insufficient to know the components of transcription (Francastel et al. 2000). Rather, a thorough understanding of the process requires knowing its functional organization within the nucleus. In this sense, transcription should not be viewed simply as a process that turns on a specific gene, but as a process that governs within the genome an entire network of genes (a transcriptome) that gives rise to a particular cellular function or fate (such as cell division, differentiation, or apoptosis). Therefore, the challenge is to uncover the nuclear organization of gene activity and to determine whether genomes are specifically structured.

The form DNA takes in the nucleus is a result of at least three prevailing components, its organization into chromatin, the linear order of genes and repetitive elements along their respective chromosomes, and the spatial localization of genes and repeats within the nucleus. Current efforts with molecular, cell biological, and genomic approaches are attempting to elucidate the role each of these components of DNA plays in regulating nuclear processes. Clearly, the forms of DNA permissive for gene transcription and gene silencing are of particular importance. This review will survey what is currently known about the localization of genes spatially within the nucleus and linearly in the genome, focusing on how these organizational states may help facilitate the orchestrated gene expression that results in cellular differentiation. Finally, the review will explore how this coordinated expression may be modeled by current network theories.

Spatial organization of gene activity within the nucleus

Cellular differentiation is generally accompanied by coordinated changes in gene expression and alterations in

[*Keywords:* Nuclear structure; genomic organization; transcription; expression neighborhood; genetic networks]

³Corresponding author.

E-MAIL markg@fhcrc.org; FAX (206) 667-5894.

Article and publication are at <http://www.genesdev.org/cgi/doi/10.1101/gad.1209304>.

nuclear structure. For example, as genes are silenced, the extent of chromatin condensation often increases, and extended regions of DNA are packaged into heterochromatin (John 1988). The amount and distribution of condensed chromatin is similar in differentiated cells of the same lineage, but the pattern varies in the nuclei of different cell types (Leitch 2000). These types of observations have led to the idea that nuclear organization may be cell-type-specific; the topological organization of the interphase nucleus may reflect the differentiated state of the cell and may be involved in the establishment and propagation of tissue-specific patterns of gene expression (Francastel et al. 2000). Although we are just beginning to understand the link between the form and function of chromatin in the interphase nucleus, technological advances, such as fluorescence in situ hybridization (FISH) and multidimensional (3-D and 4-D) fluorescence microscopy, have already led to a greater understanding of the functional organization of the nucleus.

The form of transcriptional activity

Nuclear bodies

The nucleolus, whose function and components are known in detail, is the most well-characterized nuclear body and is exclusively reviewed elsewhere (e.g., Leung and Lamond 2003). The idea that other nuclear processes may occur in discrete sites within the nucleus owes an initial example to Hewson Swift (see Acknowledgments). Using electron microscopy, Swift described the nonhomogenous fine structure of the nucleus, identifying areas of low-electron density into which fibrils extend and associate with small, dense structures, now known as interchromatin granule clusters (IGCs; Swift 1959; Lamond and Spector 2003). Evidence that IGCs contain RNA and colocalize with transcribed genes led to the idea that they may represent sites of active transcription (Lamond and Spector 2003). However, immunofluorescence microscopy with antibodies against components of the splicing apparatus (such as snRNPs and SR proteins, pre-messenger RNA splicing factors with characteristic arginine-serine repeats) has revealed that IGCs most likely serve as a reservoir of proteins involved in mRNA processing (Huang and Spector 1996). Localization of active genes near IGCs may therefore facilitate transcription by providing concentrations of splicing components. For example, the association of SR proteins with IGCs is modulated by phosphorylation, and overexpression of an SR kinase results in dissipation of IGCs and a concomitant reduction in pre-mRNA splicing (Sacco-Bubulya and Spector 2002).

Similarly, the Cajal body, whose protein components include coilin and fibrillarin, colocalizes with a subset of active genes and is disrupted by perturbation of RNA polymerase II (Pol II) transcription or protein translation (Ogg and Lamond 2002). However, immunofluorescence analyses have revealed that the Cajal body contains snRNPs and snoRNPs (Gall 2000). Furthermore, Cajal bodies contain guide RNAs that facilitate the modification

of snRNAs by base pairing and aligning the modifying enzymes (Darzacq et al. 2002). Importantly, fibrillarin is structurally similar to methyltransferases, and mutation of its yeast ortholog results in unmethylated pre-rRNA and a loss of ribosomes in the cytoplasm (Tollervey et al. 1991; Wang et al. 2000). Given these characteristics, the Cajal body is likely a site of snoRNP and snRNP processing. Therefore, the IGC and the Cajal body reveal that the machinery involved in the processing of transcripts is organized within the nucleus. Although these nuclear structures do not actually participate in the mechanism of transcription, they provide evidence for the emerging idea that transcription, splicing, and further transcript modification are integral (Bentley 2002) and may therefore be spatially organized within the nucleus (Fig. 1).

PML nuclear bodies (PML-NBs) were first observed as a result of the t(15;17) translocation detected in nearly all cases of acute promyelocytic leukemia (APL). This translocation involves the fusion of two genes, *PML* and the *retinoic acid receptor α* (*RAR α*), and leads to the disruption of PML-NBs (Ruggero et al. 2000). Although the actual mechanistic function of the PML-NB has remained elusive, it has been implicated in many fundamental cellular processes, including transcription (Zhong et al. 2000). In support of a role in transcription, regulatory factors such as cyclic AMP response-binding protein (CBP) and retinoblastoma protein (pRB) have been shown to interact directly with PML (Ruggero et al. 2000). A recent localization analysis of gene-dense domains in the human genome revealed that transcriptionally robust loci are associated with PML-NBs (Wang et al. 2004). Although RNA FISH and RNA interference (RNAi) experiments did not reveal a strict correlation of transcriptional activity and localization to PML-NBs, the majority of RNA transcript foci analyzed were positioned at the bodies. Furthermore, treatment of an APL cell line with all-*trans* retinoic acid resulted in the generation of PML-NBs and the concomitant localization of an active locus to the emergent structures (Wang et al. 2004). It is therefore plausible that PML-NBs play a role in transcription, possibly as a reservoir of concentrated transcriptional regulators (Fig. 1). Given that they also contain proteasomes, PML-NBs may play a further regulatory role by degrading transcription factors in response to external stimuli.

The interchromatin compartment

The nuclear bodies described above indicate an inherent tendency toward organization of the transcription and transcript-processing machinery. Moreover, the perturbation of function associated with each body results in the loss of its nuclear form. The existence of such structures raises the question of where they (and the processes they support) are positioned in the nucleus relative to their substrate, chromatin. Euchromatin defines the gene-rich, transcriptionally active (or potentiated) portion of the genome. Furthermore, euchromatin has a characteristic histone modification pattern and is

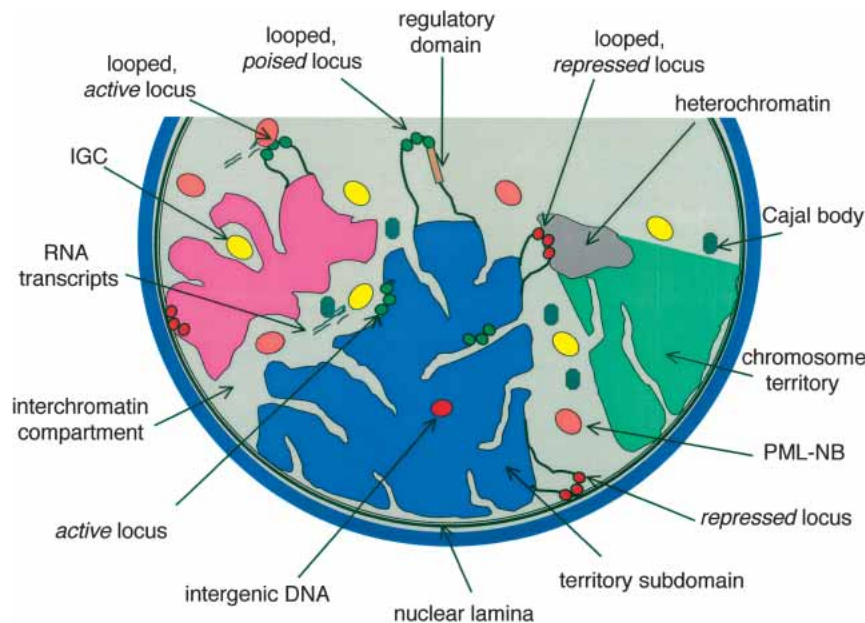


Figure 1. The nuclear organization of gene activity and gene silencing. This schematic representation indicates what is currently known of how active and potentiated genes are positioned within the nucleus and how silenced genes are compartmentalized in repressive subcompartments. See text for details.

further defined by nucleosomes with specific histone variants, such as H3.3 (Vermaak et al. 2003). FISH analysis with combined locus-specific probes and whole-chromosome paints has permitted the visualization of genes relative to their chromosome territories (CTs), the discrete structures that individual chromosomes form in the interphase nucleus (Fig. 1). Initial studies revealed that genes are preferentially positioned at territory surfaces, whereas intergenic DNA is found within the CT (Zirbel et al. 1993; Kurz et al. 1996). These observations led to the idea that an intervening compartment runs throughout the nucleus in the space between the discrete CTs, creating an interchromosome domain enriched for the nuclear bodies involved in transcription and splicing (Cremer et al. 1993). More recent studies have confirmed that transcription does occur at the surface of CTs, but that this surface runs throughout the invaginated contours of a territory, creating an interchromatin compartment (Verschure et al. 1999; Visser et al. 2000). Therefore, the interchromatin compartment (IC) model predicts that active genes are organized at the continuous surfaces of CTs to facilitate their regulation by bringing them into proximity with the nuclear bodies positioned in the IC (Cremer and Cremer 2001; Fig. 1).

In addition to being at the surface or the interior of a CT, a third type of territorial position has recently emerged, cell- and activity-dependent organization of multigene loci in a large loop (several megabase pairs) emanating from the CT. For example, it has been reported that active loci consisting of coordinately regulated genes, such as the major histocompatibility and epidermal differentiation complexes, are looped away from the central body of the CT during robust transcription (Volpi et al. 2000; Williams et al. 2002). In addition, gene-rich domains, with generally ubiquitous expression patterns, have a propensity to be looped away from their

CTs more often than gene-poor domains (Mahy et al. 2002). These results suggest that CTs, as determined by whole-chromosome paints, may represent the relatively more condensed domains of a chromosome. A gene or multigene locus in a state of open chromatin modification and structure may therefore be excluded from the CT when visualized by FISH. Regardless, the looping of a gene array from its CT may increase its association with the nuclear bodies that facilitate transcription and transcript processing.

Analysis of the endogenous wild-type and derivative mutant β -globin gene loci has helped to clarify the significance of looping of a locus from its CT (Ragoczy et al. 2003). In erythroid cells, the β -globin locus is looped away from its CT at a high frequency prior to transcriptional induction. Thus, looping is not a consequence of transcription per se, but may also represent a poised state prior to activation. However, in the absence of the locus control region (LCR), which is required for the high-level globin gene transcription induced upon terminal differentiation, the locus is positioned at the CT surface. Furthermore, if the β -globin LCR is replaced by sequences from the B-cell-specific *immunoglobulin heavy chain (IgH)* 3'C α LCR, an element that represses transcription of reporters in non-B cells (Madisen and Groudine 1994), looping is partially restored, but is now correlated with localization of the looped locus to pericentromeric heterochromatin (PCH) in another chromosome territory. Interestingly, the *IgH* locus is looped from its CT specifically in pro-B cells (where it is transcriptionally active), but is not positioned near heterochromatin. These results argue against a simple correlation of elevated transcriptional activity and looping away from CTs; rather, extrusion from the CT may play a significant role in cell-type-specific transcriptional activation or repression of a locus by localizing it to a particular position within the nucleoplasm.

The form of transcriptional repression

Heterochromatin

Silenced genes possess a distinct chromatin configuration and are specifically compartmentalized into transcriptionally repressive nuclear subcompartments (Felsenfeld and Groudine 2003). Staining of interphase nuclei with DNA-intercalating dyes reveals that the nucleus is organized into regions of weak and intense labeling, which correlate with euchromatic (active) and heterochromatic (inactive) chromatin domains, respectively. Heterochromatin is further classified as either constitutive heterochromatin (CH) or facultative heterochromatin (FH). Although the exact structure of CH has yet to be characterized, it demonstrates a regular nucleosomal spacing (as opposed to euchromatin and FH) and is refractory to DNase I and endonuclease enzymes, indicating a highly organized and condensed structure (Dillon and Festenstein 2002). Furthermore, CH is highly methylated, gene poor, late replicating, and transcriptionally repressive (Wallrath 1998; Bridger and Bickmore 1998). CH is comprised of arrays of tandem repeats (or satellites), whereas FH is a consequence of euchromatin being packaged into a condensed, transcriptionally repressive structure during cellular development (Dillon and Festenstein 2002). Therefore, CH represents a cell-type-independent organization of chromatin, whereas FH is lineage dependent and actively formed. Underscoring these elemental differences, a recent study has demonstrated that FH in erythrocytes can form in the absence of heterochromatin protein 1 (HP1), which is a requisite component of CH (Gilbert et al. 2003).

Although much remains to be determined about the structure and function of CH, its role in gene regulation has been well documented. *Pericentromeric heterochromatin* (PCH) describes the less-homogenous regions of satellite DNA adjacent to true centromeres, which may localize to the periphery of CH clusters (Lundgren et al. 2000). In *Drosophila*, the *bw^D* allele exerts its dominance over the wild-type locus by forcing its association with PCH (Csink and Henikoff 1996). The large insertion of satellite DNA in the *bw^D* allele causes it to organize with similar repeats found in heterochromatin, and somatic pairing serves to recruit the wild-type allele to this repressive domain. In murine-developing B cells, the lymphocyte-restricted transcriptional regulator Ikaros colocalizes with PCH through direct DNA binding (Brown et al. 1997; Cobb et al. 2000). T-cell-specific genes and developmentally regulated B-cell genes associate with heterochromatic Ikaros clusters specifically when they are inactive (Brown et al. 1997). Ikaros appears to associate with binding sites in a gene's regulatory element, and then recruits the gene to PCH through Ikaros-binding sites found in CH (Trinh et al. 2001). Furthermore, a study of the *immunoglobulin gene* (*Ig*) loci in immature B cells has demonstrated a nonrandom association with PCH, which may have implications in the allelic exclusion occurring at these loci (Skok et al. 2001). In developing T cells, which derive from a common progenitor as B cells, the specific expression pat-

terns of particular cytokines expressed during Th1 versus Th2 differentiation can be explained by association of the inactivated genes with PCH (Grogan et al. 2001). These studies suggest that organization of genes into heterochromatin can lead to transcriptional repression; however, they do not demonstrate the means by which this silencing is achieved (Fig. 1).

Analyses of the native β -globin locus and derivative transgenes in erythroid cells have provided a direct link between PCH association and gene activity. For example, when linked to a reporter gene integrated at PCH, an erythroid-specific enhancer (5'HS2) derived from the β -globin LCR confers localization of the transgene away from PCH and stable reporter expression (Francastel et al. 1999). Analysis of wild-type and mutant human β -globin loci have also shed light on the role of PCH association and gene activity. In erythroid cells, the wild-type locus is located away from PCH, and displays an active chromatin structure as assayed by nuclease sensitivity and histone H3 and H4 hyperacetylation (Schübeler et al. 2000). In contrast, a β -globin locus carrying a large naturally occurring deletion encompassing the LCR and 35 kb upstream ($\gamma\delta\beta^0$ thalassemia) colocalizes with PCH and adopts an inactive chromatin structure as revealed by nuclease insensitivity and histone H3 and H4 hypoacetylation, a state resembling the inactive wild-type β -globin locus in lymphocytes (Brown et al. 2001).

The studies discussed above have revealed a correlation between the activity of a gene and its proximity to CH. In all of the cases, the genes that are ultimately sequestered at heterochromatin are significant in the differentiation of the involved cell type. That is, the genes localized to PCH are oftentimes the genes whose suppression is necessary in that cell type, or in that stage of cell development. It is likely, then, that PCH-association facilitates the formation of a repressive chromatin structure and that the examples described above are, in effect, facultative heterochromatin. The active recruitment to PCH may therefore be reserved for those genes that must be silenced for differentiation to occur, or those whose regulation must be modulated to ensure the precise developmental progression.

An analysis of transgenes comprised of copies of λ -5 (a gene involved in B-cell development) integrated into PCH, underscores the significance of CH association in cell development (Lundgren et al. 2000). Despite integration into PCH and localization to the outside of CH clusters, the transgenes demonstrate position effect variegation (PEV) in pre-B cells, indicating that proximity to CH does not preclude activity. However, loss of a potent HS site results in the internalization of the transgene into CH in fibroblasts (cells in which the gene is inactive) and in the reduction of expression in pre-B cells, although remaining at the surface of CH. The position of the Δ HS transgene in pre-B cells may reflect the availability of regulatory proteins that directly impact its activity, as a genetic background heterozygous for *EBF* (a gene required for early B-cell development) results in the internalization of the Δ HS transgene into CH and to a significant reduction in its activity. Therefore, the formation of

facultative heterochromatin at PCH may lead to the progressive silencing of genes that are obligately repressed for cellular differentiation, reflecting the need to localize silenced genes in a particular nuclear subcompartment to preclude their response to the fluctuating concentrations of regulatory proteins. Interestingly, the FH that forms during terminal erythroid differentiation coincides with a large-scale relocation of proteins associated with gene repression (e.g., MeCP2, HDAC1, and MafK) from CH to other nuclear subcompartments, reflecting the large-scale nuclear condensation that occurs at this developmental stage (Francastel et al. 2001).

Modifications to the N termini of histones can regulate the binding of proteins involved in chromatin organization and gene regulation (Felsenfeld and Groudine 2003). HP1 is perhaps the best-understood protein involved in the transcriptional repression by heterochromatin, having been shown to localize to CH clusters and to mediate gene silencing (Eissenberg and Elgin 2000). Studies of HP1 have shed light on a potential mechanism for the maintenance and spreading of repressive heterochromatin. Importantly, histone H3 methylated at Lys 9 specifically binds to HP1 (Bannister et al. 2001; Lachner et al. 2001). Furthermore, this association is dependent on the activity of histone methyltransferases (HMTs) that specifically modify histone H3 on Lys 9 (Rea et al. 2000). Because HP1 and the HMTs are colocalized in heterochromatin domains, these results suggest a means for transcriptionally repressive chromatin structures to be maintained as well as spread to adjacent *cis* sequences. In addition, it is also possible that this mechanism may function in *trans*, silencing genes brought to heterochromatin domains by heterochromatin-associating proteins, like Ikaros (Fig. 1).

Nuclear periphery

Despite early indications that transcription may be localized to the nuclear periphery (Hutchison and Weintraub 1985) and a recent demonstration that boundary activities (BAs, which protect the expression status of active domains) involve the tethering of active chromatin to the nuclear pore complex (NPC; Ishii et al. 2002), the nuclear periphery has primarily been demonstrated to represent a transcriptionally repressive nuclear compartment. The nuclear periphery's role in repression has been well established in budding yeast. A number of studies have collectively demonstrated that yeast telomeres form clusters at the nuclear periphery, which leads to an enrichment of the Sir proteins known to be involved in gene silencing (Cockell and Gasser 1999). The ability of this peripheral compartment to repress transcription was tested in a study in which a reporter gene was tethered to the nuclear envelope. Making use of a Gal4-DNA-binding domain/integral membrane protein fusion, a reporter flanked by Gal4-binding sites was inducibly repressed in a Sir-dependent manner (Andrulis et al. 1998).

An analysis of the murine *Ig* loci during lymphocyte development has shown the involvement of the nuclear

periphery in the regulation of these intricately regulated gene arrays (Kosak et al. 2002). In lymphoid progenitors (as well as embryonic stem cells), the inactive *IgH* and *Igκ* loci are sequestered at the nuclear periphery. During early B-cell development, both loci are relocated to the nuclear center, which may represent a transcriptionally permissive nuclear environment. Localization to the nuclear center is not necessarily a function of transcription, as the centrally positioned *Igκ* loci are not active. Interestingly, the *IgH* locus undergoes compaction (wherein distal ends of the 3-Mbp array colocalize, implying a looped structure) when it is centrally located in the nucleus and is poised to undergo long-range *V(D)J* recombination. A null mutation of the *interleukin-7 receptor α*, which results in a block early in B-cell development, abrogates relocation of the loci from the periphery and prevents the compaction of the *IgH* locus. A recent study has further delineated the steps involved in the compaction of the locus, indicating that the B-cell regulatory protein Pax-5 in conjunction with an unknown B-cell-specific factor may induce the close juxtaposition of the ends of the *IgH* array (Fuxa et al. 2004).

FISH analysis of whole chromosomes in human nuclei has revealed that a gene-poor chromosome (18) is preferentially localized to the nuclear periphery, whereas a gene-rich chromosome (19) is more centrally disposed in the nucleus (Croft et al. 1999). This preferential association is maintained, even in the context of a balanced translocation between the two chromosomes, with the translocated portions of 18 and 19 residing peripherally and centrally, respectively. Further analysis of gene-dense and gene-poor chromosomes has confirmed the tendency for gene-poor chromosomes to be positioned at the nuclear periphery (Boyle et al. 2001). Cross-species analysis has revealed that this behavior is not restricted to the human nucleus (Tanabe et al. 2002). As described below, gene-rich chromosomal domains are the most highly expressed regions of the human genome; therefore, the demonstrations of gene-rich chromosomes organized into the nuclear center may simply be a reflection of their overall level of activity.

The studies described above strongly suggest that the nuclear periphery may represent a transcriptionally repressive nuclear compartment distinct from CH, which often resides in perinuclear clusters. For example, the peripheral localization of silent *Ig* loci does not involve association with PCH (Kosak et al. 2002). In fact, this study indicates that the nuclear lamina itself may play a role in the sequestration and inactivity of perinuclear loci. The major components of the nuclear lamina are the lamins, type-V intermediate filament proteins that polymerize to form the lamin network that is juxtaposed to the inner nuclear membrane of the nuclear envelope. There are two classes of lamins, A type and B type. Expression of the A-type lamins is developmentally regulated, whereas the B-type lamins are ubiquitously expressed (Mounkes et al. 2003). The nuclear lamina, through lamin B, interacts directly with DNA and chromatin, as well as indirectly through lamin-binding pro-

teins (Gotzmann and Foisner 1999). In addition, proteins demonstrated to be involved in gene silencing have also been shown to associate with the lamina, including HP-1 (Kourmouli et al. 2000). Therefore, although much work needs to be done before a causative effect in gene silencing can be attributed to localization at the nuclear periphery, growing evidence supports the idea that it represents a transcriptionally repressive nuclear subcompartment (Fig. 1).

The relationship between structure and function in the nucleus is clearly evidenced by mutations in the A-type lamin gene (*LMNA*, with major splicing variants A and C), resulting in several human diseases, collectively termed “laminopathies.” These diseases include muscular dystrophy, cardiomyopathy, partial lipodystrophy, and progeria syndromes (Genschel and Schmidt 2000; Mounkes et al. 2003). Of particular interest is how mutation of a single gene that is broadly expressed in differentiated tissues could result in several tissue-specific disease phenotypes. One possible explanation is that lamin A/C, localized at the nuclear periphery as well as internal, perinucleolar foci, establishes a structure in differentiated cells on which transcriptional regulators and their respective target genes are organized. In support of this idea, lamin A/C has been found to interact with transcription factors, such as pRb and SREBP1, important in the differentiation of mesenchymal tissues, which are most affected by mutations in *LMNA* (Mancini et al. 1994; Lloyd et al. 2002).

Chromatin mobility

The organization of the transcriptional machinery and the compartmentalization of silenced genes suggest that genes must be mobile within the nucleus to be appropriately positioned. Currently, the study of chromatin mobility has yielded conflicting results in the comparison of human and yeast nuclei. In humans, small movements of 0.5 μm have been demonstrated, which allows a gene to sample a very small fraction of the total nuclear volume (Chubb et al. 2002). However, in light of the IC model, these small movements may be sufficient to localize a gene to a relevant nuclear body or repressive compartment. Also, use of the lac operator-repressor system (which allows visualization of chromatin through arrays of lac-binding sites) revealed that a late-replicating, heterochromatic domain undergoes large-scale movement from the nuclear periphery to the interior prior to replication (Li et al. 1998). In yeast (and *Drosophila*), 0.5 μm movements have also been detected, but given the significant difference in nuclear size, these movements permit a gene to travel upward of half the nuclear diameter (Gasser 2002). Furthermore, the movement of loci in yeast has been shown to be energy dependent, unlike the small-scale movements in humans. Despite the differences between human and yeast, the evidence for short, diffusional movements are compatible with the role of nuclear localization in gene regulation.

Chromosome organization

Beyond the movement of individual genetic loci, there is evidence that chromosomes may themselves be mobile. Two recent studies utilized an H2B–GFP fusion protein and photobleaching to analyze the overall order of chromosomes through the cell cycle. In an analysis of HeLa cells, chromosomes were shown to maintain their localization in daughter cells in approximately half the nuclei studied (Walter et al. 2003). Furthermore, chromosomes were shown to be mobile during the early G₁ cell cycle (G₁). A similar analysis that modeled a random and nonrandom organization of chromosomes to be expected from the photobleaching experiment showed that the organization of chromosomes is significantly nonrandom, or maintained, during mitosis (Gerlich et al. 2003). Despite the discrepancies between these results, they both argue that an inherent chromosome organization may exist that is remembered upon cell division. In support of the suggestion of a defined chromosomal organization in the nucleus, studies of the chromosomes and gene loci involved in translocations that lead to leukemia have revealed a propensity for translocation partners to be spatially proximal (Parada et al. 2002; Roix et al. 2003). These results argue for a functional organization of the genome at the level of the chromosome. The exact nature of this organization, and whether the organization particular to a given cell type is altered as the cell responds to external stimuli or, in fact, differentiates, has yet to be determined.

Linear arrangement of gene activity within the genome

If the nuclear localization of a gene is involved in its regulation and the development of any cell type requires the regulation of hundreds and perhaps thousands of genes (its transcriptome), then cellular differentiation may be accompanied by large-scale nuclear reorganization. Regardless of whether this reorganization is due to a direct or indirect mechanism, the individual relocation of hundreds of spatially distant genes would be inefficient. Therefore, it is possible that an underlying linear order of genes along chromosomes exists to facilitate the coordinated regulation of the transcriptome. The idea that coregulated genes may share linear positions within a genome has two obvious precedents; namely, the operons of prokaryotic genomes (as well as that of *Caenorhabditis elegans*) and the gene arrays of eukaryotes, both of which have been instrumental in the understanding of transcriptional regulation. For example, the lineage-restricted *IgH* and *β -globin* loci are examples of gene arrays that share a common genomic position and are intricately regulated in specific cell types. Furthermore, as discussed above, it has also been shown that both of these loci have nuclear localization patterns that parallel their state of activity. Although these two gene arrays (and some others that have been characterized) are the result of duplication events, it is nevertheless likely that coregulated genes unrelated in sequence homology may be organized in linear clusters throughout the genome.

The advent of multiple genome-wide analytical techniques has provided the means to explore genomes for networks of unique genes that are clustered within the genome and involved in a common cellular function or in the differentiation of a particular lineage. If such linear gene clusters exist, they would support a model in which coregulated genes exhibit physical proximity along their chromosomes to facilitate their regulation. Evidence from all species so far examined has revealed that genomes are nonrandomly organized (Fig. 2).

Yeast

As discussed above, budding yeast has provided an excellent model for the study of nuclear localization affecting gene activity, specifically in the repressive nature of peripheral localization. Telomere position effect (TPE), which is caused by the tethering of telomeres at the periphery amid the localized enrichment of repressive SIR proteins, provides an important example of how a gene's linear position within the genome can affect its regulation (Hediger and Gasser 2002). An analysis of chromosome correlation maps of *Saccharomyces cerevisiae* has revealed that beyond TPE, there is an underlying order to the yeast genome. Correlation maps allow the expression patterns from various conditions or cell stages to be

plotted along the linear gene order of the chromosomes (Cohen et al. 2000). An analysis of these maps with expression data obtained from cell cycle phases, sporulation, and the pheromone response, revealed an inherent organization of the yeast genome; a highly significant percentage of nonduplicated, coexpressed genes are adjacent (and to a lesser degree form triplets) along the chromosomes. Furthermore, these adjacent genes also tend to be functionally related. To ascertain the nature of the coregulation of adjacent genes, their regulatory sequences were examined. Although adjacent genes do not necessarily have similar UASs, there are several examples in which one of the adjacent genes lacks a UAS (Cohen et al. 2000), indicating that adjacency may allow neighboring genes to share a single regulatory element.

Worm

Unlike other eukaryotes, which have not been shown to contain operons, as much as 25% of *C. elegans'* coding sequence may be organized into polycistronic operons of two to eight genes (Blumenthal 1998). Clearly, operons exemplify coregulation through proximal positioning within the genome. The multiple genes of an operon share a common regulatory domain, thereby ensuring the coexpression of genes typically involved in a com-

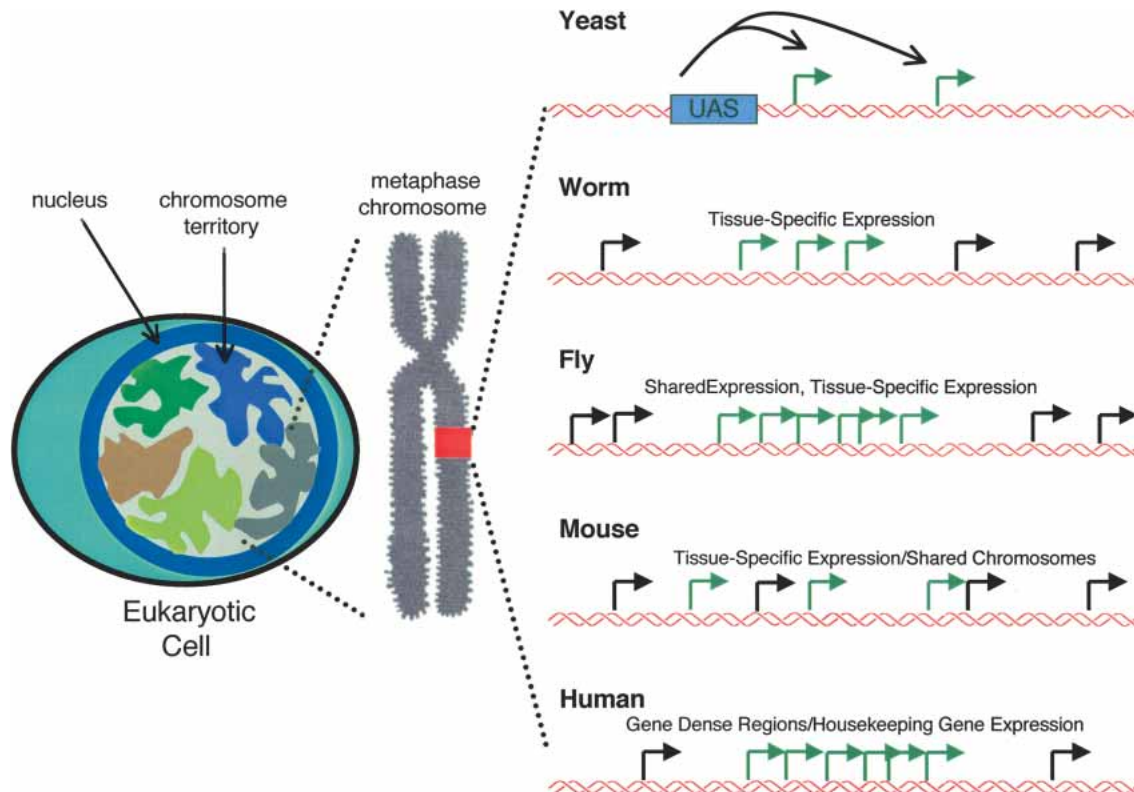


Figure 2. Eukaryotic genomes are nonrandomly organized. Evidence for a shared linear organization of similarly expressed genes has been uncovered in each of the species analyzed. Although the nature of the organization is not necessarily the same among the different species, the fact that all of the genomes do possess localized activity argues for its significance in gene regulation. See text for details on each organism.

Kosak and Groudine

mon function. Recently, however, an organization of individual, monocistronic genes has been uncovered within the worm genome (Roy et al. 2002). mRNA tagging [which makes use of an epitope-tagged poly(A)-binding protein] was developed for the isolation of tissue-specific transcripts from whole larvae. After excluding genes within an operon and tandem duplications, an analysis of muscle-specific genes revealed that they are clustered together in groups of two to five throughout the genome (Roy et al. 2002). Unlike yeast, however, these clustered genes do not necessarily share a common cellular function. Moreover, analysis of microarray data from germ-line cells showed that sperm, oocytes, and the germ line itself demonstrate an organization of tissue-specific expression similar to that of muscle (Roy et al. 2002). These data and those from yeast suggest that transcriptomes, dedicated to a cell state or to a particular cell type, exist in organized centers or neighborhoods, and that changes in expression patterns correspond with a shift in the genomic organization of the transcriptome.

Fly

Drosophila is widely known for its polytene chromosomes, an arrangement of the genome in salivary glands in which numerous rounds of replication without mitosis result in enormous polyploid chromosomes. Polytene chromosomes have been shown to exhibit puffs in regions of high levels of transcription, encompassing domains of presumably coregulated genes (Thummel 2002). Polytene puffs may therefore represent a physical manifestation of clustered genes with similar expression patterns. A microarray analysis of expression from *Drosophila* determined under 80 different experimental conditions, has revealed an organization of nonhomologous, coexpressed genes in groups of 10–30, covering between 20 and 200 kbp (Spellman and Rubin 2002). Although these genes demonstrate coregulation, shared functions for genes in each group was not established. Importantly, the grouped genes show highly correlated levels of expression, suggesting that the domain organization is a reflection of an active chromatin structure that stretches through the region. An analysis of expressed sequence tags (ESTs) databases has also demonstrated the clustering of genes within the *Drosophila* genome; however, this examination focused on the tissue-specific expression profiles from the testis, head-region, and embryo (Boutanaev et al. 2002). In each cell type, the coregulated genes were found to be significantly organized into clusters of three or more genes, with a trend toward large groupings. Therefore, the clustering of coregulated, lineage-restricted genes indicates a functional organization of transcriptomes that define a given cell type.

Mouse

Vertebrate genomes have many well-characterized loci that encompass gene arrays that demonstrate coregulation and a shared function, or even an ultimately recom-

bined protein in the case of the *Ig* loci. These arrays are found both for tissue-specific genes, as well as ubiquitously expressed genes. Therefore, there is a precedent for the proximal positioning of genes that share a common function, even though the example of arrays primarily indicate duplication events. Evidence also indicates that there may be a broader organization to tissue-restricted transcriptomes in the mouse. For example, an examination of ESTs from extraembryonic tissues from post-implantation mouse embryos (d7.5) revealed an organization of 155 cDNA clones localized in clusters on subregions of chromosomes 2, 7, 9, and 17 (Ko et al. 1998). Although the potential clustering of these genes was not tested, these data indicate that there is a non-random distribution of coregulated genes at the level of the chromosome. The *t*-complex itself, located on chromosome 17, represented 6.5% of all cDNAs. Similarly, an analysis of expression profiles from embryonic, neuronal, and hematopoietic stem cells revealed the *t*-complex to be enriched for shared stem-cell genes (Ramalho-Santos et al. 2002). Examination of the differentiation of a hematopoietic progenitor into erythroid and neutrophil cell types indicates an organization of transcriptomes into adjacent, coregulated genes that changes upon differentiation (S.T. Kosak, D. Scalzo, F. Li, S. Hall, T. Enver, and M. Groudine, in prep.).

Human

An integration of the human genomic sequence with SAGE (serial analysis of gene expression) data for genome-wide mRNA expression patterns from 12 tissue types, has provided a Human Transcriptome Map (Caron et al. 2001) that reveals the human genome is nonrandomly organized into regions of high and low levels of gene activity. The highly active regions, RIDGES (regions of increased gene expression), are separated by large regions of low activity (antiridges, not unlike valleys). Importantly, RIDGES and valleys coincide with gene-dense and gene-poor chromosomal domains, respectively. Therefore, gene activity is inherently compartmentalized along the chromosome, which is analogous to the further subdivision of lineage-restricted or coregulated genes being clustered in the genomes of the model organisms described above. RIDGES also demonstrate a high GC content, SINE density, and a low-intron length (Versteeg et al. 2003), implying a higher-order organization of the genome that may be a reflection of chromosomal structure and/or a strategy for gene regulation. Analysis of the linear organization of the mouse genome has revealed a nonhomogenous distribution similar to that of human, indicating that this type of genomic pattern has been conserved (Mural et al. 2002; S.T. Kosak, D. Scalzo, F. Li, S. Hall, T. Enver, and M. Groudine, in prep.). A recent analysis of SAGE data indicates that RIDGES may be a consequence of the population of these regions by the ubiquitously and highly expressed house-keeping genes and that tissue-restricted transcriptomes have a tendency to be clustered (Lercher et al. 2002). Also, paralleling the evidence from the worm, analysis of

a genomic transcript map of human skeletal muscle genes revealed that genes expressed in this lineage are concentrated on three chromosomes (17, 19, and X) in five chromosomal regions (Bortoluzzi et al. 1998). Therefore, in addition to the overall genomic organization of RIDGEs, there is a further level of organization of lineage-specific genes.

There have been many indications, such as operons and position effect, that genomes are not homogeneously organized. Now, from the genomic approaches described above, it appears that there is an elemental nonrandom organization of eukaryotic genomes. Coexpressed genes demonstrate a propensity to be adjacent or grouped along the genome. These gene clusters can be functionally related or involved in the transcriptome of a specific cell type. These latter features offer evidence that there may be evolutionary constraints upon the genomic organization of coexpressed genes. The profound effect on cellular differentiation of a single translocation giving rise to leukemia offers evidence of the importance of the regulatory consequence of a gene's chromosomal context (Rowley 1998). Therefore, the link between expression and position strongly indicate a role in gene regulation (although other processes, including splicing and replication, may also be related to the genome's overall organization). Further analysis will be necessary to determine whether an inherent organization of the nucleus exists that reflects the nonrandom linear arrangement of genes, and whether this nuclear organization is altered during differentiation.

The actual role clustering plays in gene regulation remains to be established. Nevertheless, the available data suggest several potential mechanisms for how expression neighborhoods may influence gene regulation (Oliver

et al. 2002). An obvious possibility is that proximal genes share enhancer elements. In the case of yeast and *C. elegans*, the sharing of a common regulatory element may in fact occur at adjacent genes (Cohen et al. 2000; Lercher et al. 2002). The evidence from other eukaryotes, however, suggests that a more general effect on transcriptional regulation may be at work. One possible explanation of this effect is that an increased local concentration of regulatory sequences, which are identical or involved in the regulation of related genes, create a hub of the proteins that, in turn, bind these sequences. Specifically, the grouping of genes may decrease the effective off-rate of regulatory proteins through the localization of binding sites. This is an attractive possibility, given that numerous FRAP studies have indicated a high-diffusion constant for both regulatory and structural nuclear proteins (Phair and Misteli 2000; Cheutin et al. 2003). The high mobility of regulatory proteins (such as transcription factors) is particularly significant, as a given binding site is found in many locations throughout the genome that are not germane to gene regulation (Bulyk 2003; Fig. 3A).

The study of lymphopoiesis has provided several important examples in support of the functional relevance of regulatory protein concentrations. Study of the Pax-5 transcription factor, which can act as a transcriptional repressor and activator, revealed that at lower nuclear concentrations, the protein maintains its positive regulatory role due to a 20-fold higher affinity of activator elements (Wallin et al. 1998). Additionally, as discussed above, localized concentrations of the EBF transcription factor can override the suppressive effects of PCH (Lundgren et al. 2000). Finally, the overall level of the transcription factor PU.1 in progenitor populations can affect the hematopoietic developmental pathway, with

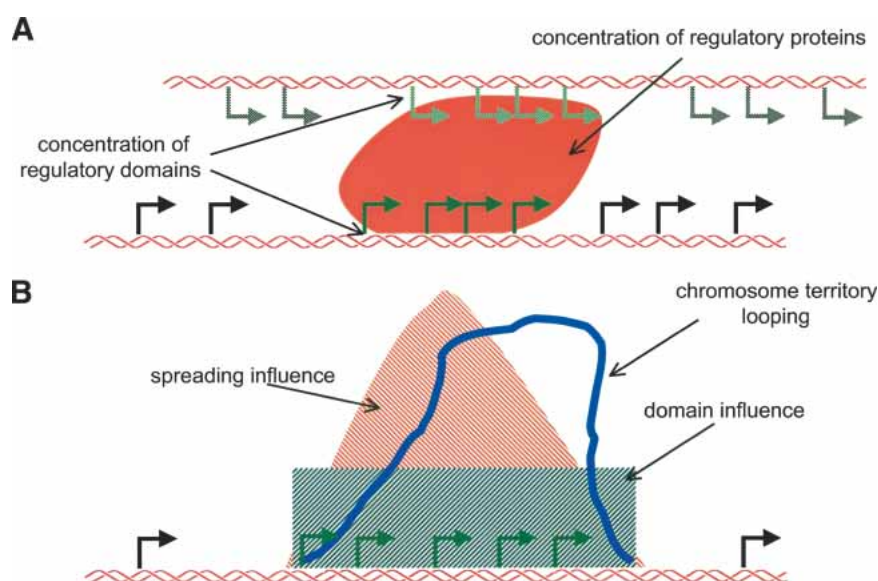


Figure 3. Potential roles for genetic clustering in gene regulation. (A) The clustering of genes may facilitate gene regulation by the formation of expression neighborhoods (or hubs) through a feedback of increased binding sites and a subsequent concentration of the regulatory proteins that bind them. Such a model would require the spatial association of these chromosome regions within the nucleus. The chromosome regions may include the alleles from homologous chromosomes (as depicted), or other regions that have clusters of similarly regulated genes. (B) As opposed to the active mechanism in A, the clustering of gene expression may be the result of a domain architecture of chromatin. In this model, a potent regulatory element or the additive spreading of a chromatin configuration may lead to the coexpression of adjacent genes. It is likely that boundary elements would be involved to prevent the spreading

of this effect, as well as to prevent the encroachment of silencing mechanisms. The degree of influence identified for these domains is similar to the extent of chromosome territory looping, and it is interesting to consider whether the domain architecture may reflect the spatial positioning of blocks of chromatin into a transcriptionally permissive nuclear subcompartment (see The Interchromatin Compartment).

high levels resulting in myeloid cells and low levels in B cells (DeKoter and Singh 2000). Therefore, creating an expression neighborhood through the localization of regulatory elements would be an efficient means to potentiate regional gene activity (Fig. 3A).

Another possibility for the role of linear arrangement in gene regulation, which is not mutually exclusive with protein concentrations, is that a potent regulatory element (or elements) may influence the expression status within a chromosomal region. This enhancer may directly activate the individual promoters of the adjacent genes, or it may simply lead to the spreading of histone modifications that, in turn, would affect the transcriptional status of surrounding genes (Fig. 3B). Evidence from the fly does not support a spreading effect, however, as there does not appear to be a gradual decline of influence the further from the center of the expression neighborhood (surrounding genes are instead either on or off; Spellman and Rubin 2002). In fact, the *Drosophila* data suggests the formation of a static domain, perhaps through the use of insulators, that delimits the local effect on expression. A recent genomic analysis of HP1 and Su(var)3-9 binding supports a domain architecture of gene expression, as developmentally regulated genes display uniform patterns of association with one or both of the proteins (Greil et al. 2003). Either by a spreading of modifications or the establishment of a domain, it is interesting that looping from CTs (as described above) appears to correlate with the range of effect seen at the genomic level (Ragoczy et al. 2003). It is possible that CT looping may be a physical manifestation of the potentiated or activated state of an expression neighborhood. Therefore, it will be very interesting to determine whether looped domains colocalize to repressive subcompartments (like PCH) or to regions permissive for transcription (like the nuclear center; Fig. 3B).

Cellular differentiation as a genetic network

The evidence reviewed above indicates that the nuclear localization of a particular gene can significantly affect its activity and that a shared characteristic of eukaryotic genomes is the nonrandom organization of genes related by shared expression patterns and possibly function. This spatial and linear organization of genes suggests a highly structured nucleus, in which form and function are intricately linked. However, as discussed above, an inherent organization of chromosomes for a particular progenitor cell type, which changes upon cellular differentiation, has yet to be demonstrated. Furthermore, a spatial analysis of all the active genes within a nucleus has not been performed. Therefore, if we are to approach the concept of a nonrandom, functionally coordinated genome, it is necessary to model the functional structure of the nucleus.

The concept of self-organization has been put forth to explain the behavior of the nucleus (in addition to other organelles; Misteli 2001a). As opposed to a self-assembly mechanism, in which constituent proteins form a static nuclear structure in a state of equilibrium, self-organiza-

tion describes a structure that forms from molecular interactions in a steady state. The idea that self-organization may describe the mechanism that forms the nuclear bodies and subcompartments (e.g., heterochromatin) is based largely on FRAP analysis (Misteli 2001b). These studies demonstrate the rapid diffusion of both regulatory and structural proteins in the nucleus. Because they are in constant flux, random interactions of proteins are thought to seed the formation of transient structures; in other words, a stable structure is achieved by the dynamic, continuous exchange of its components. Self-organization may therefore explain how the functional interaction of ribosomal proteins, rRNA transcription factors, and the rDNA template lead to the genesis of the nucleolus (Misteli 2001a). Interestingly, the introduction of rDNA into ectopic sites within the genome leads to the formation of micronucleoli around the integrated genes. As this example illustrates, the concept of self-organization holds promise in facilitating our understanding of the structural organization of cellular function. Self-organization, however, does not address the nonrandom nature of the components of a functional structure. Specifically, the underlying order of eukaryotic genomes argues against a random association of a gene and its requisite regulatory machinery. Form may indeed follow function, but the form of a self-organized structure may be predisposed by a nonrandom organization of its parts.

In recent years, graph (or network) theory, the mathematical field that explores how networks form, has made considerable progress in analyzing the nature of real-world networks (Barabási and Oltvai 2004). Network theory has, until recently, been dominated by the idea that networks are inherently random in their formation and consequent organization. The interconnections of a random network, defined by nodes (the entities) and links (the connections), follow a Poisson distribution, with the vast majority of nodes having a common, relatively small number of links and rare outlier nodes with many more or fewer links. The study of real-world networks, however, has revolutionized the field, revealing that random networks do not predominate in the natural world. Initial analysis of the World Wide Web and the Internet determined that these networks do not follow a Poisson distribution, but rather, they best fit a power-law degree distribution ($P(k) \sim k^{-\gamma}$). The diminishing tail of the power-law curve gave these networks their name, scale-free, which refers to the lack of a prevalent linkage number. Importantly, all scale-free networks so far examined demonstrate a γ (degree exponent) between 2 and 3, which is influenced by the very few nodes that have a tremendous number of links. Scale-free networks have two primary rules for their formation and maintenance, a scale-free network expands continuously with the addition of new nodes, and these nodes are added preferentially to sites that are already well connected (Barabási and Albert 1999). These principles explain the hubs (the highly connected nodes) seen in real-world networks (Fig. 4A). In essence then, scale-free networks describe a kind of self-organization

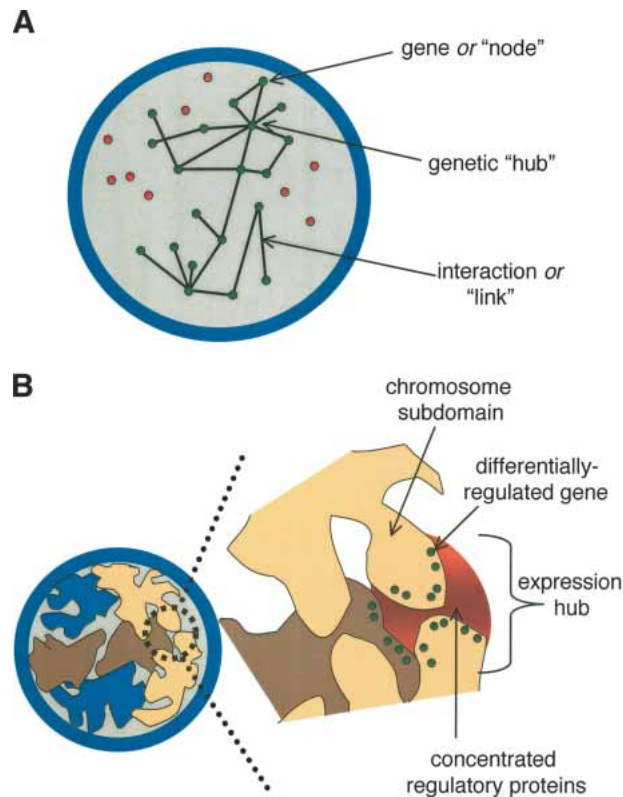


Figure 4. Network theory and genomic organization. (A) The recent description of numerous real-world networks has transformed network theory. “Scale-free” describes a special aspect of real-world networks, the organized centers, or hubs, of highly linked nodes. This behavior has been shown in metabolic as well as proteomic pathways, and is depicted here playing a role in genetic interactions. It is based upon the constant growth of real-world networks, and the preferential attachment of those added nodes. (B) The nonrandom organization of the linear and spatial genome suggests features of scale-free networks. The localized position of multiple coregulated genes would facilitate their intercommunication by helping to form and then utilize localized concentration of regulatory proteins. An expression hub, or neighborhood, of this kind may include any two alleles (chromosomes of the same color represent homologs), with their mirrored linear gene order, in addition to regions of other chromosomes with clustered genes undergoing similar regulation.

that is instructed by its inherent substructure. These qualities make network theory an engaging model with which to approach the genomic organization of the transcriptional regulation of differentiation. Scale-free models have, in fact, already been used to describe metabolic and proteomic networks in biological systems (Jeong et al. 2000, 2001; Giot et al. 2003).

In essence, a network is an exchange of information that obligates either a physical interaction or the sharing of regulatory information. The coordinate regulation of hundreds of genes therefore describes a network that guarantees the intercommunication of the integral, yet individually regulated genes that affect a cell response or cellular differentiation. Just as analysis of the yeast pro-

tein network has uncovered that those proteins with the greatest number of links are those most likely to be lethal (Jeong et al. 2001), an organization of genes within the genome and in the nucleus may indicate a mechanism by which proximity facilitates regulation. The proximal positioning of coregulated genes, described above, provide just the nodes that would permit modeling the genetic organization of cellular differentiation as a scale-free network. Although not elaborated in the description of expression neighborhoods, the majority of coregulated genes are not necessarily grouped within the genome. As demonstrated in the *Drosophila* genome, however, there are chromosome domains with a preponderance of coexpressed genes. These regions, in a scale-free model, would represent the hubs of genetic activity, as so many genes are contained in a confined genomic region (e.g., Featherstone and Broadie 2002). If the model is allowed to include the association of allelic regions of homologous chromosomes, then these hubs are even more pronounced in their local gene activity. Furthermore, if other highly expressed regions may also communicate in the coordinate activity of these domains, a significant regional neighborhood of expression would result (Fig. 4B).

We propose that there may be evolutionary constraints upon the disruption of localized gene expression, whether its origin is due to duplication events or otherwise, which have ensured that coregulated genes involved in a common function maintain a shared linear position within the genome. By spatially restricting the position of genes, their regulation can be coordinated through a concentration of regulatory proteins or by the spreading of chromatin modifications and activation mediated by enhancers. Further analysis will be required to verify that gene clustering truly facilitates coordinate gene regulation and determine whether this linear organization is reflected in the spatial organization of coregulated in the nucleus. In addition, whether the nonrandom order of genes on chromosomes necessitates a particular nuclear organization of chromosomes remains to be established.

Acknowledgments

We acknowledge the many contributions of the late Hewson Swift to cell biology and to our understanding of the nucleus. In no small way, the substance of this review is indebted to him. We apologize to those researchers whose work we were not able to include due to size constraints. We also thank our colleagues in the Groudine lab for useful discussions, and Mike Bulger, Steve Henikoff, Dave Scalzo, Bill Schubach, Erica Smith, and Jon Soderholm for critical reading of this manuscript. S.K. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research. This work has been aided by a grant from The Jane Coffin Childs Memorial Fund for Medical Research and was supported by NIH grants (DK44746 and HL57620) to M.G.

References

- Andrulis, E.D., Neiman, A.M., Zappulla, D.C., and Sternglanz, R. 1998. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* **394**: 592–595.

Kosak and Groudine

- Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**: 120–124.
- Barabási, A.L. and Albert, R. 1999. Emergence of scaling in random networks. *Science* **286**: 509–512.
- Barabási, A.L. and Oltvai, Z.N. 2004. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**: 101–113.
- Bentley, D. 2002. The mRNA assembly line: Transcription and processing machines in the same factory. *Curr. Opin. Cell Biol.* **14**: 336–342.
- Blumenthal, T. 1998. Gene clusters and polycistronic transcription in eukaryotes. *Bioessays* **20**: 480–487.
- Bortoluzzi, S., Rampoldi, L., Simionati, B., Zimbello, R., Barbon, A., d'Alessi, F., Tiso, N., Pallavicini, A., Toppo, S., Cannata, N., et al. 1998. A comprehensive, high-resolution genomic transcript map of human skeletal muscle. *Genome Res.* **8**: 817–825.
- Boutanaev, A.M., Kalmykova, A.I., Shevelyov, Y.Y., and Nurminsky, D.I. 2002. Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* **420**: 666–669.
- Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., and Bickmore, W.A. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10**: 211–219.
- Bridger, J.M. and Bickmore, W.A. 1998. Putting the genome on the map. *Trends Genet.* **14**: 403–409.
- Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merckenschlager, M., and Fisher, A.G. 1997. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* **91**: 845–854.
- Brown, K.E., Amols, S., Horn, J.M., Buckle, V.J., Higgs, D.R., Merckenschlager, M., and Fisher, A.G. 2001. Expression of α - and β -globin genes occurs within different nuclear domains in haemopoietic cells. *Nat. Cell Biol.* **3**: 602–606.
- Bulyk, M.L. 2003. Computational prediction of transcription-factor binding site locations. *Genome Biol.* **5**: 201.
- Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A., et al. 2001. The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. *Science* **291**: 1289–1292.
- Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., and Misteli, T. 2003. Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science* **299**: 721–725.
- Chubb, J.R., Boyle, S., Perry, P., and Bickmore, W.A. 2002. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12**: 439–445.
- Cobb, B.S., Morales-Alcelay, S., Kleiger, G., Brown, K.E., Fisher, A.G., and Smale, S.T. 2000. Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. *Genes & Dev.* **14**: 2146–2160.
- Cockell, M. and Gasser, S.M. 1999. Nuclear compartments and gene regulation. *Curr. Opin. Genet. Dev.* **9**: 199–205.
- Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* **26**: 183–186.
- Cremer, T. and Cremer, C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**: 292–301.
- Cremer, T., Kurz, A., Zirbel, R., Dietzel, S., Rinke, B., Schrock, E., Speicher, M.R., Mathieu, U., Jauch, A., Emmerich, P., et al. 1993. Role of chromosome territories in the functional compartmentalization of the cell nucleus. *Cold Spring Harb. Symp. Quant. Biol.* **58**: 777–792.
- Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**: 1119–1131.
- Csink, A.K. and Henikoff, S. 1996. Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature* **381**: 529–531.
- Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., and Kiss, T. 2002. Cajal body-specific small nuclear RNAs: A novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J.* **21**: 2746–2756.
- DeKoter, R.P. and Singh, H. 2000. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science* **288**: 1439–1441.
- Dillon, N. and Festenstein, R. 2002. Unravelling heterochromatin: Competition between positive and negative factors regulates accessibility. *Trends Genet.* **18**: 252–258.
- Eissenberg, J.C. and Elgin, S.C. 2000. The HP1 protein family: Getting a grip on chromatin. *Curr. Opin. Genet. Dev.* **10**: 204–210.
- Featherstone, D.E. and Broadie, K. 2002. Wrestling with pleiotropy: Genomic and topological analysis of the yeast gene expression network. *Bioessays* **24**: 267–274.
- Felsenfeld, G. and Groudine, M. 2003. Controlling the double helix. *Nature* **421**: 448–453.
- Francastel, C., Walters, M.C., Groudine, M., and Martin, D.I. 1999. A functional enhancer suppresses silencing of a transgene and prevents its localization close to centromeric heterochromatin. *Cell* **99**: 259–269.
- Francastel, C., Schubeler, D., Martin, D.I., and Groudine, M. 2000. Nuclear compartmentalization and gene activity. *Nat. Rev. Cell Biol.* **1**: 137–143.
- Francastel, C., Magis, W., and Groudine, M. 2001. Nuclear relocation of a transactivator subunit precedes target gene activation. *Proc. Natl. Acad. Sci.* **98**: 12120–12125.
- Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E., and Busslinger, M. 2004. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes & Dev.* **18**: 411–422.
- Gall, J.G. 2000. Cajal bodies: The first 100 years. *Annu. Rev. Cell Dev. Biol.* **16**: 273–300.
- Gasser, S.M. 2002. Visualizing chromatin dynamics in interphase nuclei. *Science* **296**: 1412–1416.
- Genschel, J. and Schmidt, H.H. 2000. Mutations in the LMNA gene encoding lamin A/C. *Hum. Mutat.* **16**: 451–459.
- Gerlich, D., Beaudouin, J., Kalbfuss, B., Daigle, N., Eils, R., and Ellenberg, J. 2003. Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell* **112**: 751–764.
- Gilbert, N., Boyle, S., Sutherland, H., de Las Heras, J., Allan, J., Jenuwein, T., and Bickmore, W.A. 2003. Formation of facultative heterochromatin in the absence of HP1. *EMBO J.* **22**: 5540–5550.
- Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., et al. 2003. A protein interaction map of *Drosophila melanogaster*. *Science* **302**: 1727–1736.
- Gotzmann, J. and Foisner, R. 1999. Lamins and lamin-binding proteins in functional chromatin organization. *Crit. Rev. Eukaryot. Gene Expr.* **9**: 257–265.
- Greil, F., van der Kraan, I., Delrow, J., Smothers, J.F., de Wit, E., Bussemaker, H.J., van Driel, R., Henikoff, S., and van Steensel, B. 2003. Distinct HP1 and Su(var)3–9 complexes bind to sets of developmentally coexpressed genes depending

- on chromosomal location. *Genes & Dev.* **17**: 2825–2838.
- Grogan, J.L., Mohrs, M., Harmon, B., Lacy, D.A., Sedat, J.W., and Locksley, R.M. 2001. Early transcription and silencing of cytokine genes underlie polarization of T helper cell subsets. *Immunity* **14**: 205–215.
- Hediger, F. and Gasser, S.M. 2002. Nuclear organization and silencing: Putting things in their place. *Nat. Cell. Biol.* **4**: E53–E55.
- Huang, S. and Spector, D.L. 1996. Dynamic organization of pre-mRNA splicing factors. *J. Cell Biochem.* **62**: 191–197.
- Hutchison, N. and Weintraub, H. 1985. Localization of DNAase I-sensitive sequences to specific regions of interphase nuclei. *Cell* **43**: 471–482.
- Ishii, K., Arib, G., Lin, C., Van Houwe, G., and Laemmli, U.K. 2002. Chromatin boundaries in budding yeast: The nuclear pore connection. *Cell* **109**: 551–562.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. 2000. The large-scale organization of metabolic networks. *Nature* **407**: 651–654.
- Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. 2001. Lethality and centrality in protein networks. *Nature* **411**: 41–42.
- John, B. 1988. *The biology of heterochromatin*. Cambridge University Press, Cambridge.
- Ko, M.S., Threat, T.A., Wang, X., Horton, J.H., Cui, Y., Pryor, E., Paris, J., Wells-Smith, J., Kitchen, J.R., Rowe, L.B., et al. 1998. Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the *t*-complex and under-representation on the X chromosome. *Hum. Mol. Genet.* **7**: 1967–1978.
- Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., and Singh, H. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**: 158–162.
- Kourmouli, N., Theodoropoulos, P.A., Dialynas, G., Bakou, A., Politou, A.S., Cowell, I.G., Singh, P.B., and Georgatos, S.D. 2000. Dynamic associations of heterochromatin protein 1 with the nuclear envelope. *EMBO J.* **19**: 6558–6568.
- Kurz, A., Lampel, S., Nickolenko, J.E., Bradl, J., Benner, A., Zirbel, R.M., Cremer, T., and Lichter, P. 1996. Active and inactive genes localize preferentially in the periphery of chromosome territories. *J. Cell. Biol.* **135**: 1195–1205.
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**: 116–120.
- Lamond, A.I. and Spector, D.L. 2003. Nuclear speckles: A model for nuclear organelles. *Nat. Rev. Mol. Cell. Biol.* **4**: 605–612.
- Leitch, A.R. 2000. Higher levels of organization in the interphase nucleus of cycling and differentiated cells. *Microbiol. Mol. Biol. Rev.* **64**: 138–152.
- Lercher, M.J., Urrutia, A.O., and Hurst, L.D. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.* **31**: 180–183.
- Leung, A.K. and Lamond, A.I. 2003. The dynamics of the nucleolus. *Crit. Rev. Eukaryot. Gene Expr.* **13**: 39–54.
- Li, G., Sudlow, G., and Belmont, A.S. 1998. Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: Precise choreography of condensation/decondensation and nuclear positioning. *J. Cell. Biol.* **140**: 975–989.
- Lloyd, D.J., Trembath, R.C., and Shackleton, S. 2002. A novel interaction between lamin A and SREBP1: Implications for partial lipodystrophy and other laminopathies. *Hum. Mol. Genet.* **11**: 769–777.
- Lundgren, M., Chow, C.M., Sabbattini, P., Georgiou, A., Minaae, S., and Dillon, N. 2000. Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* **103**: 733–743.
- Madisen, L. and Groudine, M. 1994. Identification of a locus control region in the immunoglobulin heavy-chain locus that deregulates *c-myc* expression in plasmacytoma and Burkitt's lymphoma cells. *Genes & Dev.* **8**: 2212–2226.
- Mahy, N.L., Perry, P.E., and Bickmore, W.A. 2002. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell. Biol.* **159**: 753–763.
- Mancini, M.A., Shan, B., Nickerson, J.A., Penman, S., and Lee, W.H. 1994. The retinoblastoma gene product is a cell cycle-dependent, nuclear matrix-associated protein. *Proc. Natl. Acad. Sci.* **91**: 418–422.
- Misteli, T. 2001a. The concept of self-organization in cellular architecture. *J. Cell. Biol.* **155**: 181–185.
- . 2001b. Protein dynamics: Implications for nuclear architecture and gene expression. *Science* **291**: 843–847.
- Mounkes, L., Kozlov, S., Burke, B., and Stewart, C.L. 2003. The laminopathies: Nuclear structure meets disease. *Curr. Opin. Genet. Dev.* **13**: 223–230.
- Mural, R.J., Adams, M.D., Myers, E.W., Smith, H.O., Miklos, G.L., Wides, R., Halpern, A., Li, P.W., Sutton, G.G., Nadeau, J., et al. 2002. A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* **296**: 1661–1671.
- Ogg, S.C. and Lamond, A.I. 2002. Cajal bodies and coilin—Moving towards function. *J. Cell. Biol.* **159**: 17–21.
- Oliver, B., Parisi, M., and Clark, D. 2002. Gene expression neighborhoods. *J. Biol.* **1**: 4.
- Parada, L.A., McQueen, P.G., Munson, P.J., and Misteli, T. 2002. Conservation of relative chromosome positioning in normal and cancer cells. *Curr. Biol.* **12**: 1692–1697.
- Phair, R.D. and Misteli, T. 2000. High mobility of proteins in the mammalian cell nucleus. *Nature* **404**: 604–609.
- Ragoczy, T., Telling, A., Sawado, T., Groudine, M., and Kosak, S.T. 2003. A genetic analysis of chromosome territory looping: Diverse roles for distal regulatory elements. *Chromosome Res.* **6**: 513–525.
- Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., and Melton, D.A. 2002. “Stemness”: Transcriptional profiling of embryonic and adult stem cells. *Science* **298**: 597–600.
- Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**: 593–599.
- Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A., and Misteli, T. 2003. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat. Genet.* **34**: 287–291.
- Rowley, J.D. 1998. The critical role of chromosome translocations in human leukemias. *Annu. Rev. Genet.* **32**: 495–519.
- Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. 2002. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418**: 975–979.
- Ruggero, D., Wang, Z.G., and Pandolfi, P.P. 2000. The puzzling multiple lives of PML and its role in the genesis of cancer. *Bioessays* **22**: 827–835.
- Sacco-Bubulya, P. and Spector, D.L. 2002. Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing. *J. Cell. Biol.* **156**: 425–436.
- Schubeler, D., Francastel, C., Cimbara, D.M., Reik, A., Martin, D.I., and Groudine, M. 2000. Nuclear localization and histone acetylation: A pathway for chromatin opening and transcriptional activation of the human β -globin locus. *Genes &*

Kosak and Groudine

- Dev.* **14**: 940–950.
- Skok, J.A., Brown, K.E., Azuara, V., Caparros, M.L., Baxter, J., Takacs, K., Dillon, N., Gray, D., Perry, R.P., Merken-schlager, M., et al. 2001. Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat. Immunol.* **2**: 848–854.
- Spellman, P.T. and Rubin, G.M. 2002. Evidence for large do-mains of similarly expressed genes in the *Drosophila* ge-nome. *J. Biol.* **1**: 5.
- Swift, H. 1959. Studies on nuclear fine structure. *Brookhaven Symp. Biol.* **12**: 134–152.
- Tanabe, H., Muller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M., Solovei, I., Cremer, C., and Cremer, T. 2002. Evolutionary conservation of chromosome territory arrange-ments in cell nuclei from higher primates. *Proc. Natl. Acad. Sci.* **99**: 4424–4429.
- Thummel, C.S. 2002. Ecdysone-regulated puff genes 2000. *In-sect Biochem. Mol. Biol.* **32**: 113–120.
- Tollervy, D., Lehtonen, H., Carmo-Fonseca, M., and Hurt, E.C. 1991. The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. *EMBO J.* **10**: 573–583.
- Trinh, L.A., Ferrini, R., Cobb, B.S., Weinmann, A.S., Hahm, K., Ernst, P., Garraway, I.P., Merken-schlager, M., and Smale, S.T. 2001. Down-regulation of TDT transcription in CD4⁺CD8⁺ thymocytes by Ikaros proteins in direct compe-tition with an Ets activator. *Genes & Dev.* **15**: 1817–1832.
- Vermaak, D., Ahmad, K., and Henikoff, S. 2003. Maintenance of chromatin states: An open-and-shut case. *Curr. Opin. Cell. Biol.* **15**: 266–274.
- Verschure, P.J., van Der Kraan, I., Manders, E.M., and van Driel, R. 1999. Spatial relationship between transcription sites and chromosome territories. *J. Cell. Biol.* **147**: 13–24.
- Versteeg, R., van Schaik, B.D., van Batenburg, M.F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H.J., and van Kam-pen, A.H. 2003. The human transcriptome map reveals ex-tremes in gene density, intron length, GC content, and re-peat pattern for domains of highly and weakly expressed genes. *Genome Res.* **13**: 1998–2004.
- Visser, A.E., Jaunin, F., Fakan, S., and Aten, J.A. 2000. High resolution analysis of interphase chromosome domains. *J. Cell. Sci.* **113**: 2585–2593.
- Volpi, E.V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ra-goussis, J., et al. 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in in-terphase nuclei. *J. Cell. Sci.* **113**: 1565–1576.
- Wallin, J.J., Gackstetter, E.R., and Koshland, M.E. 1998. Depen-dence of BSAP repressor and activator functions on BSAP concentration. *Science* **279**: 1961–1964.
- Wallrath, L.L. 1998. Unfolding the mysteries of heterochroma-tin. *Curr. Opin. Genet. Dev.* **8**: 147–153.
- Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., and Cremer, T. 2003. Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during sub-sequent interphase stages. *J. Cell. Biol.* **160**: 685–697.
- Wang, H., Boisvert, D., Kim, K.K., Kim, R., and Kim, S.H. 2000. Crystal structure of a fibrillarin homologue from *Metha-nococcus jannaschii*, a hyperthermophile, at 1.6 Å resolu-tion. *EMBO J.* **19**: 317–323.
- Wang, J., Shiels, C., Sasieni, P., Wu, P.J., Islam, S.A., Freemont, P.S., and Sheer, D. 2004. Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic re-gions. *J. Cell. Biol.* **164**: 515–526.
- Williams, R.R., Broad, S., Sheer, D., and Ragoussis, J. 2002. Sub-chromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp. Cell Res.* **272**: 163–175.
- Zhong, S., Salomoni, P., and Pandolfi, P.P. 2000. The transcrip-tional role of PML and the nuclear body. *Nat. Cell. Biol.* **2**: E85–E90.
- Zirbel, R.M., Mathieu, U.R., Kurz, A., Cremer, T., and Lichter, P. 1993. Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Res.* **1**: 93–106.



Form follows function: the genomic organization of cellular differentiation

Steven T. Kosak and Mark Groudine

Genes Dev. 2004, **18**:

Access the most recent version at doi:[10.1101/gad.1209304](https://doi.org/10.1101/gad.1209304)

References

This article cites 97 articles, 39 of which can be accessed free at:
<http://genesdev.cshlp.org/content/18/12/1371.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a dark background. On the left, it says "Dharmacon™ Reagents" with the tagline "Custom synthesis, RNAi, and CRISPR solutions" below it. In the center, the text "Infinite Reliability" is written in large white font. To the right of this text is a "More" button. On the far right, the "horizon" logo is displayed in white, with "a PerkinElmer company" written in smaller text underneath. The background features a colorful, abstract representation of DNA or protein structures in shades of purple, blue, and green.