LINEAR AND SPATIAL ORGANIZATION OF ACTIVE GENES

DIETMAR RIEDER

DOCTORAL THESIS

Graz University of Technology
Institute for Genomics and Bioinformatics
Petersgasse 14, 8010 Graz, Austria

Graz, April 2008

Abstract

There is increasing evidence that the cell nucleus is highly organized. A growing number of defined bodies and substructures with different functions has already been identified. Non random radial positioning of chromosome territories and single genes was observed. Moreover, close spatial association upon activation was reported for some selected functionally related genes. However, little is known about the mechanisms or factors that determine the nuclear positioning and it is unclear whether the spatial association of active genes is a general aspect for co-expression or if it applies only to special examples.

This study was designed to provide an unbiased and general insight into the linear and spatial organization of active genes. Expression profiles derived from a microarray study on human adipocyte differentiation were analyzed to assess the chromosomal arrangement of co-transcribed genes. The different analyses revealed a frequent grouping into tandems suggesting a non random linear organization of the human genome. Seven genes were then selected based on their similar expression profile and were simultaneously visualized by 3D combinatorial multiplex fluorescence in situ hybridization (M-FISH). The radial organization of these seven genes was found to be primarily determined by local gene density and did not correlate with their transcription levels. Furthermore, it could be shown that close spatial association of active genes is not restricted to some special examples but can also be observed for co-expressed genes of unrelated function. In addition, it could be demonstrated that the global spatial organization of active genes is non random and is preserved between different nuclei under different physiological conditions. This work provides further insight into the linear- and nuclear architecture of the human genome. It is a step towards a better understanding of complex organizational patterns involving multiple genes.

Keywords: nucleus, spatial organization, active genes, chromosome, transcription factory, fluorescence microscopy, 3D multiplex FISH, adipogenesis, hMADS

Publications

This thesis was based on the following publications and on unpublished work:

Papers

G. Stocker, D. Rieder, Z. Trajanoski. ClusterControl: a Web interface for distributing and monitoring bioinformatics applications on a Linux cluster. Bioinformatics. 20: 805-807 (2004)

Müller WG, Rieder D, Kreth G, Cremer C, Trajanoski Z, McNally JG. Generic features of tertiary chromatin structure as detected in natural chromosomes. Mol Cell Biol. 24: 9359-9370 (2004)

Kowalska A, Bozsaky E, Ramsauer T, Rieder D, Bindea G, Lorch T, Trajanoski Z, Ambros PF. A new platform linking chromosomal and sequence information. Chromosome Res. 15: 327-339 (2007)

Müller WG, Rieder D, Karpova TS, John S, Trajanoski Z, McNally JG. Organization of chromatin and histone modifications at a transcription site. J Cell Biol. 177: 957-967 (2007)

Stocker G, Fischer M, Rieder D, Bindea G, McNally J, Trajanoski Z. iLAP: a novel work flow oriented approach for microscopy data management and protocol development. *in preparation* (2008)

Rieder D, Scheideler M, Stocker G, Fischer M, Müller WG, McNally J, Trajanoski Z. Linear and Spatial Organization of Active Genes. *in preparation* (2008)

Contents

List of Figures

Introduction

Among many researchers the cell nucleus is often seen as a microscopic reaction tube containing chromatin, different nuclear factors and enzymes. In this simplified view, processes like replication and transcription are carried out in a uniformly distributed way at random locations in the nuclear space. But in fact, the nucleus is a highly organized cellular compartment. The DNA of eukaryotic cells is organized in form of chromosomes. In the interphase nucleus, these chromosomes build distinct entities, the so called chromosome territories, which occupy defined nuclear subvolumes and can be visualized by fluorescence microscopy [16, 53, 3, 17]. Besides chromosome territories several other nuclear substructures have been characterized including PML bodies, Cajal bodies, and splicing factor compartments (SFCs) also known as nuclear speckles (reviewed in [19, 70]). The most prominent nuclear substructure is the nucleolus where the rRNAs are produced by RNA-Polymerase I and biogenesis of ribosomal subunit takes place [2]. Thus, the nucleolus is one example for the link between nuclear organization and specific function. Also gene transcription and regulation happens in the context of an organized nucleus (reviewed in [34, 7, 22, 40, 64, 47]). For example, it has been demonstrated that the nuclear periphery can play a role in both gene silencing and gene activation (reviewed in [34, 46]). The ability to repress the transcriptional activity of certain genes was also reported for the nuclear periphery [21, 55], whereas the nuclear interior was often associated with transcriptional activity [41, 82, 81]. In addition, radial positioning of chromosomes or chromosomal domains could be linked to chromosomes size or gene density [74, 15, 3, 39].

Furthermore, not only the positioning of genes with respect to certain nuclear landmarks, but also the relative positioning to each other may reflect an important aspect of nuclear structure as indicated by the nucleolar organization of rDNA genes or the spatial aggregation of tRNA genes which could be observed in *S. cerevisiae* [76]. The transcription factory model [30, 31, 14, 49] describes a close relative positioning of multiple transcribed genes that spatially cluster at sites enriched in active RNA Polymerase II molecules and transcription factors. It provides also an explanation of the fact that only a few hundred to a few thousand (dependent on the cell type) discrete foci of active RNA Polymerase II can be detected by immuno staining in cells where tens of thousand molecules are actively transcribing. Experimental evidence for sharing of transcription factories by different genes was initially found in mouse erythroid cells where several coordinately expressed genes, which were separated by 25 or more megabase pairs on the same chromosome, co-localized in an unexpected high number of observed nuclei [50]. A recent study on human α- and β-*globin* both located on different chromosomes (16, 11), revealed a frequent spatial association of these genes. This observation was directly correlating with transcriptional activity [5]. In mouse B lymphocytes a juxtaposition of *Myc* and *Igh* was found. Upon activation, these two genes occupied the same transcription factory although they are located on different chromosomes 15 and 12 [51].

Besides the three-dimensional organization of the genome a linear arrangement of co-regulated genes could be uncovered in different organisms. The linear clustering of genes into so called ”operons” is a long known feature for gene regulation in prokaryotic organisms. However, it is not restricted to them. Co-regulated genes that were organized in tandems and grouped in chromosomal domains were also found in *S. cerevisiae* [11], *C. elegans* [60], mouse [35], and other higher eukaryotes [4, 79, 54, 45]. The non-random linear genome organization may facilitate coordinated gene regulation and influence their nuclear positioning (reviewed in [34]).

However, little is known about the underlying mechanisms that determine the nuclear positioning. It is unclear, whether the spatial association of co-transcribed genes can be observed in general, or if it applies only to a few selected examples [10]. This uncertainty is even bigger when considering co-regulated genes located on different chromosomes. Aside from the limited information on close encounters of active genes, there is also a lack of knowledge regarding their global nuclear organization relative to each other or relative to nuclear landmarks: Is it dependent on the transcription rate? Is the organization preserved between nuclei in different cells? Does it change

under different physiological conditions?

To the best of our knowledge, these questions have not been previously addressed by simultaneously assessing the three-dimensional positioning of multiple genes in single cells and under different physiological conditions. Therefore, a study was designed that allowed for an unbiased and general insight into the linear and spatial organization of active genes. Based on the expression profiles derived from microarray data the chromosomal arrangement was analyzed by different bioinformatic and statistical methods. Seven genes located on five different chromosomes and exhibiting similar expression profiles were selected for examining their spatial organization. These genes were simultaneously visualized by 3D combinatorial multiplex fluorescence in situ hybridization (M-FISH). The systematic and comprehensive analysis revealed that local gene density primarily accounts for radial positioning while the expression level does not correlate with the observed nuclear gene arrangement. Furthermore it could be shown that close spatial association of active genes is not restricted to some special examples but can be observed in general. In addition, it could be demonstrated that the global spatial organization of active genes is non random and is preserved between different nuclei.

Results

Chromosomal organization of co-regulated genes

Co-regulated genes are arranged in tandems along chromosomes. To examine the linear organization of co-regulated genes on human chromosomes data from microarray experiments was analyzed. This data was generated in a time course gene expression study on human adipocyte differentiation [62]. The 29,952 features present on the microarray were mapped to ∼14,000 genes and assigned to their linear basepair location along the corresponding chromosomes. The microarray experiments identified a set of 848 genes that were differentially expressed (2-fold) during adipogenesis, which is comparable to the number found in another mammalian system [25]. These genes are hereafter in a broader sense also referred to as "co-regulated" genes (see also [35]). They were clustered into 12 subgroups by the k-means (k=12) algorithm implemented in the Genesis [73] software package (Figure 1). To test whether they are linearly organized in groups, as suggested by differ-

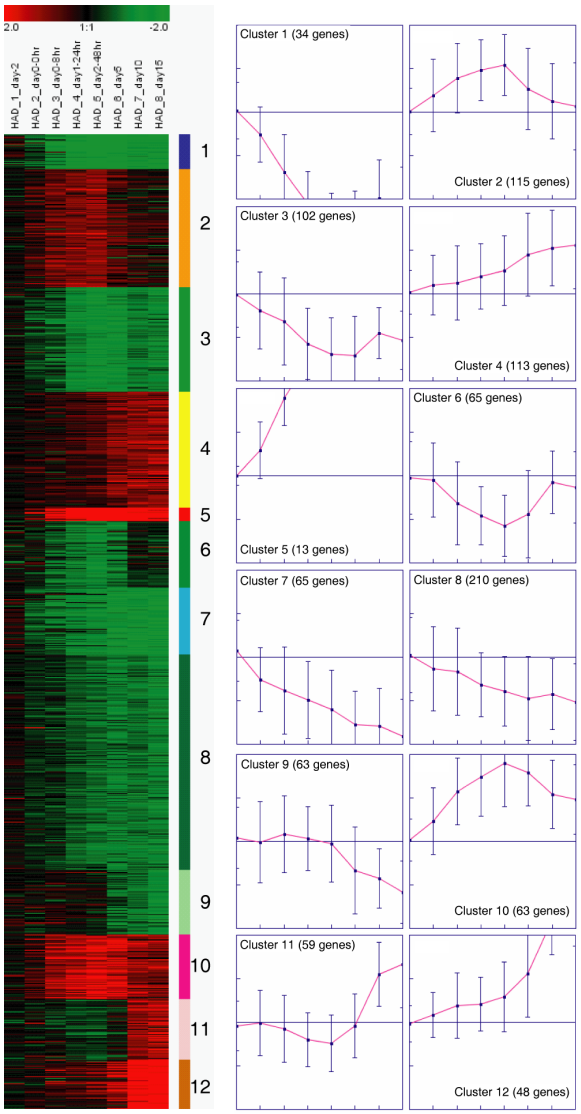


Figure 1: Human adipocyte differentiation regulated genes. A microarray study on adipogenesis in human multipotent adipose-derived stem (hMADS) cells revealed 848 differentially expressed genes which were clustered (k-means) into 12 groups based on their expression profile.

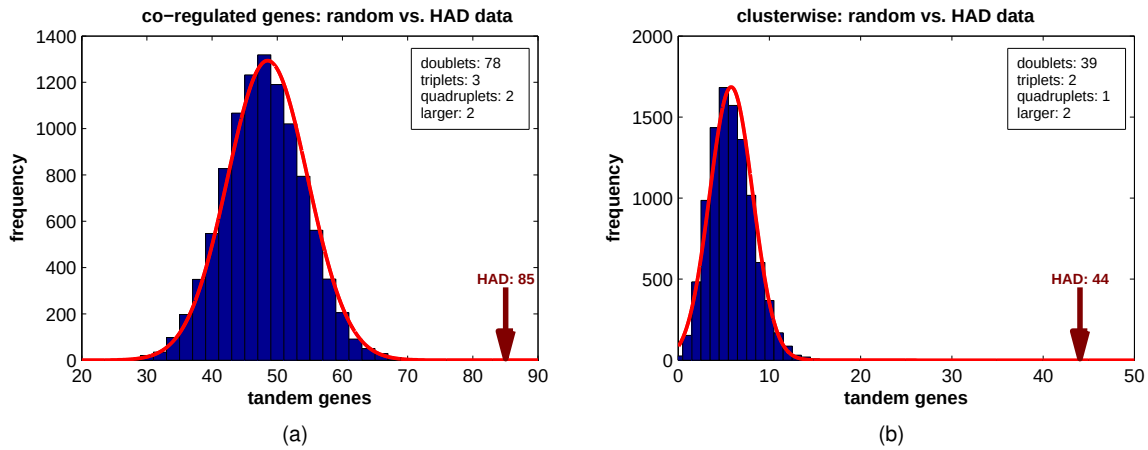


Figure 2: Co-regulated genes are organized in tandems along the chromosomes. The blue bars of the histograms illustrate the counts of gene tandems found in 10,000 iterations of the random permutation test. (a) shows the results for the entire set of 848 regulated genes found in human adipocyte differentiation. (b) displays the number of tandems found in an identical random permutation test that scanned for each of the 12 clusters of the co-regulated genes separately. The arrows indicate the numbers (85 and 44 respectively) of tandems derived from the real microarray dataset. The red line delineates a curve that fits the histogram with a Gaussian distribution.

ent models [11, 4, 79, 54, 45], the chromosomes were scanned for tandem genes. Tandems were defined as clusters of adjacent, differentially expressed genes which are not disrupted by an intercalating, "unregulated" gene on the microarray. The scan was performed for the whole set of the 848 regulated genes and for each of the 12 clusters separately. The analysis revealed 85 groups of tandem genes for the entire set and 44 for the individual clusters. In the majority of the cases, tandems were organized as doublets (90%) and to a lesser amount as triplets or larger groups, so that 18.4% (156/848) of the regulated genes can be found in pairs that are located in direct proximity on the same stretch of chromosome. The results were then compared to a random gene set in which the positions of the $\sim$14,000 genes present on the microarray were randomly shuffled and the scan was again run using this random assignment. This procedure was repeated 10,000 times. However, the random permutation test did not produce results which resembled the real data. A χ-square analysis demonstrated that the frequency of "real" tandems is significantly ($p < 10^{-5}$) different from random. Figure 2 summarizes the data from the permutation test and clearly shows that the observed number of gene tandems is very unlikely to be obtained by chance, neither when testing the entire set of 848 genes, nor when testing the 12 clusters separately. These findings and numbers are consistent with the results of a recent work focusing on hematopoiesis in mouse [35] and reflect a common tendency of co-regulated genes to be adjacent on the chromosomal level. This preference for tandem organization of genes regulated in human adipocyte differentiation resembles the observations made by studying cell differentiation in other species. For a cell the linear clustering of genes may facilitate their coordinated and efficient regulation. So called "expression hubs" may be established were regulatory proteins are concentrated and can bind to their target sequences. In addition, enhancers and other regulatory elements can change the chromosomal conformation at the tandem sites by spreading histone modifications [18, 34].

A sliding-window analysis was performed to determine whether genes that are co-regulated in human adipocyte differentiation exhibit an extended linear organization, exceeding doublets or triplets. This type of analysis helps to overcome the relative infrequency of adipogenesis related genes (848) compared to the number of genes present on the microarray ($\sim$14,000) [35]. Analogous to the analysis conducted by Kosak et al., a sliding window of 10 megabase pairs (Mbps) was moved

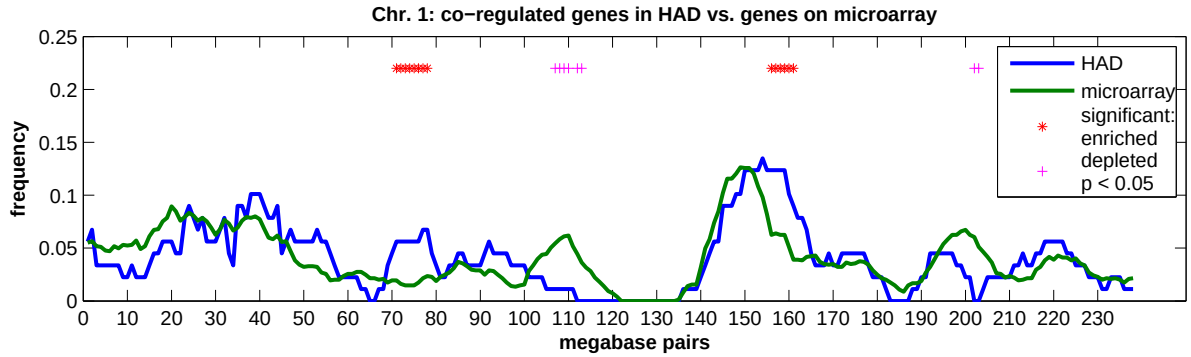


Figure 3: Sliding window analysis of human chromosome 1. A sliding window analysis reveals the gene density distribution along the chromosome. The green line represents the genes on the microarray and the blue line delineates the differentially expressed genes in adipocyte differentiation. The gene frequencies are denoted as ratios between the gene count in each window and the total number of genes on the chromosome or the number of regulated genes on the chromosome respectively. An exact binomial test uncovers domains that show gene frequencies that are significantly higher or lower in the adipogenesis related gene set than expected by the gene distribution on the microarray, these are marked with $*$ and $+$ respectively.

along each chromosome in steps of 1 Mbps and the frequency counts of genes located in the window were recorded. The proposed 10 Mbps window size was adopted due to its biological relevance. Kosak and his co-workers argued that the size of syntenic regions which are shared between murine and human chromosomes is in the range of $\sim$10-15 Mbps (see [13, 24]). This suggests a functional constrain on gene domain size. To confirm this reasoning, the syntenic blocks that are shared between human, mouse and rat, were extracted from the Ensembl database [29] and have been analyzed regarding their size. The results showed that 50% of the blocks larger than 1 Mbps are ranging between 3 and 15 Mbps in size. Figure 3 displays the data from the sliding window analysis on human chromosome 1. One can identify mountains and valleys in the curve representing the distribution of the microarray gene set. These mountains and valleys mark gene dense and gene poor regions. This observation was also made in each of the remaining 23 chromosomes (data not shown). When comparing the microarray curve with the curve that represents the co-regulated genes, a similarity between the two profiles can be observed. This is explainable by the fact that gene dense regions often correlate with "regions of increased gene expression" (RIDGEs) and gene poor regions in general are weakly transcribed [8]. In RIDGEs the transcription rate of single genes is increased seven fold compared to the genomic average. By having a closer look at the two graphs, differences in several domains were observed throughout the entire genome. In these domains the number of genes regulated during adipogenesis is significantly (exact binomial test, $p < 0.05$) higher or lower than expected by the gene count on the microarray. These discrepancies do not mean that those regions are per se different from the total gene distribution, but suggest that groups of co-regulated genes found in human adipocyte differentiation show a tendency for density, meaning that they are concentrated in larger domains. By counting the domains enclosing 11 or more genes in both - the random and the real gene set - this tendency is clearly indicated for the adipogenesis related genes, whereas the random genes do not exhibit such a preference (Fisher's exact test, $p < 0.05$).

These results resemble the findings from the mouse hematopoiesis study. It should be noted that the gene dense domains of linage-specific data, which significantly differ from the gene dense domains of random data, contain a smaller number of genes (> 5) compared to the values presented here (> 11). This might be due to the differences in the datasets, organisms, and studies: 11,000 versus 14,000 genes on the microarray, 539/594 versus 848 regulated genes, mouse versus human, adipogenesis versus hematopoiesis.

These data provide strong evidence that the genes co-regulated in human adipocyte differentiation are linearly organized at two levels. First, co-regulated genes are arranged in tandems, and second groups of such genes tend to be concentrated in larger domains.

Spatial organization of active genes

With this clear evidence for linear organization of co-regulated genes in human adipocyte differentiation in mind, several new questions arose: Do active genes also exhibit an organization in the three dimensional space of the cell nucleus? Furthermore, do co-regulated genes located on different chromosomes share expression hubs? Is there a change in the organization which is dependent on the expression status, and is the organization preserved between different cells?

To investigate the spatial organization of active genes, a subset of genes from cluster 10 (see Figure 1) was studied by 3D-DNA-FISH [67]. This subset was composed of seven genes, that are up-regulated (6-10 fold) in the early phase of adipocyte differentiation. All of them exhibited a peak in their expression level after 24 and 48 hours of induction. In order to be able to simultaneously visualize and discriminate the seven genes by using only 3 different labels, a combinatorial labeling [56, 71, 20, 80] approach was chosen and the probes were labeled according to the schema shown in Table 1.

Human multipotent adipose-derived stem cells (hMADS), which are known to show a normal karyotype [59], were grown on coverslips and nuclei were three-dimensionally preserved [27, 68, 67, 75], in preconfluent (∼80-90% confluence) and induced state. The nuclei were hybridized with the labeled probe, and over 400 - ∼200 preconfluent and ∼200 induced - were imaged as three-dimensional image stacks. The raw image stacks were processed by 3D-deconvolution [43] and the resulting images were corrected for color-shift errors and reconstructed in 3D. Figure 4 illustrates examples of raw and reconstructed image data (compare panel a,e with b,f). The image acquisition and processing produced over 89,000 single images that were stored and managed by the *i*LAP system [72]. A semi-automatic analysis routine for detecting nuclei and FISH-signals and assigning the corresponding gene names was realized in Imaris-XT and Matlab. Automatic data collection and measurement methods were implemented in the analysis routine, so that the following information could be recorded for each cell: 91 gene-gene distances, 14 gene-center distances, 14 gene-nucleus-surface-distances, 1 nucleus volume, 1 ellipsoid approximating the nucleus (see Figure 4 c,d,g,h).

	Label		
	DNP	Biotin	Digoxigenin
C4orf18	☑	☐	☐
GLUL	☐	☑	☐
SAMHD1	☐	☐	☑
AGTR1	☐	☑	☑
ANGPTL1	☑	☑	☐
FKBP5	☑	☐	☑
SPON2	☑	☑	☑
	AF-488	AF-647	Rhodamine

Table 1: 3D combinatorial M-FISH labeling schema. In order to discriminate between the seven different genes, a combinatorial labeling technique involving the listed hapten-fluorochrome-pair combinations was used.

Radial positioning of active genes does not reflect changes of their expression level. To assess the preferred nuclear positions of the seven genes, each nucleus was first approximated by its enclosing ellipsoid. The resulting volume was then subdivided into three volumes of equal size, delineated by three concentric ellipsoids of equal shape. By this method an inner, middle, and outer shell was created and each single gene was assigned to the shell containing it. Figure 5 illustrates and summarizes the results of this analysis performed with induced hMADS cells. In preconfluent cells, 52.38% of the investigated genes were found to be located in the nuclear interior confined by the inner shell. In the middle shell 39.17% and in the outer shell 8.44% of the genes were found. A similar distribution was found in the induced cells, where 55.80% of the gene locations were assigned to the nuclear interior, 36.50% and 7.71% were assigned to the middle- and outer shell respectively. These two shells together are

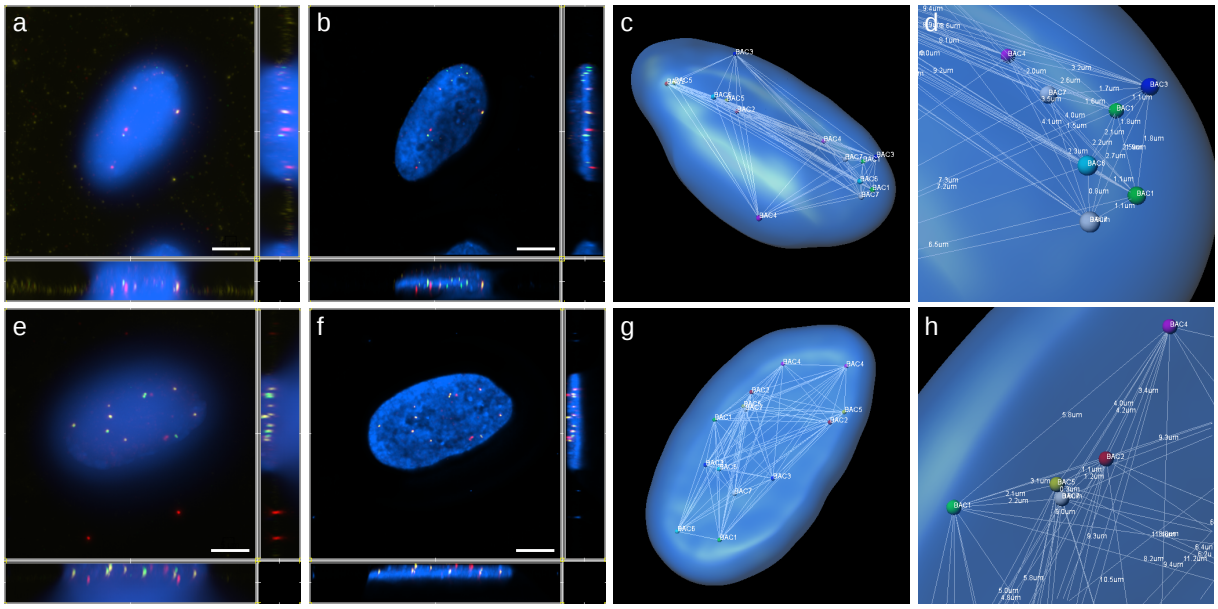


Figure 4: 3D combinatorial M-FISH. The spatial organization of seven active and co-regulated genes located on 5 different chromosomes was studied by 3D combinatorial M-FISH. The raw image stacks (a,e) of over 400 nuclei were 3D-deconvolved and corrected for color shift errors (b,f). The nucleus surface and FISH signals were detected (c,g) and various measurement data, like coordinates and distances, was collected (d,h). The seven genes were found in different locations and showed different distances to each other. However, there are some pairs that show a greater tendency of being located in spatial proximity than others (c,d,g,h). (white bar = 5μm)

further on referred to as nuclear periphery. Although the numbers are similar, the difference in the number of genes counted in the nuclear interior, in preconfluent and induced cells is significant (Fisher's exact test, $p < 0.02$). This observation suggests that the analyzed genes may change their position from the nuclear periphery to the interior, depending on their activity.

However, these results reflect just a summary over the positioning of all seven genes pooled together. Therefore it is likely that some of them do not change their radial position at all, some change toward the interior and others vice versa. Given these possibilities, each single gene was examined individually and members for each of these categories could be identified. Furthermore, a general preference for certain radial locations, independent of the change in the expression level, could be revealed. As Figure 6 illustrates, the genes that were preferentially located at the nuclear periphery are *C4orf18*, *GLUL*, *ANGPTL1*, and *AGTR1*. Interestingly, the first three of them exhibit a slight but significant change in the frequency with which they were found in the nuclear interior up on induction (Fisher's exact test, $p < 0.05$), while this was not observed for *AGTR1*. In contrast, *SAMHD1*, *FKBP5*, and *SPON2* were preferentially located at the nuclear interior. In this group only *SAMHD1* showed a significant alteration of its nuclear location. This gene was found to show a decreased frequency of being located at the nuclear interior after the cells were induced (Figure 6). These findings indicate that, for the herein studied genes, the radial positioning within the cell nucleus does not dramatically change when their expression level increases (6-10 fold).

Radial positioning does not correlate with transcription rates of genes or domains. Since the genes were selected based on a two-color microarray study, they must be actively transcribed in both states, preconfluent and induced. The fold change in the expression level measured by a microarray is a relative value and gives no information about the actual expression level in ei-

ther of the two compared conditions. Therefore, a gene that is weakly transcribed in preconfluent cells could still be less active after induction than a strongly transcribed one before induction.

If the positioning of a gene is determined by its activity level (high → interior, low → periphery) [36, 41, 63, 7], then this could explain why the genes tested here do not dramatically change their locations. To test whether the radial positioning of a gene is associated with its expression rate, a rough estimation of the expression level for each gene was derived from the intensities of the specific spots on the microarray. A linear regression analysis was then performed by using these estimated values together with the corresponding radial position counts. The theory that more active genes are preferentially located at the nuclear interior, could not be confirmed by the results of the regression analysis (R^2_{adj} = 0.15).

Given the possibility that chromatin positioning within the nuclear space is not affected by the activity of only a single gene, but by the overall activity of its chromosomal domain, the regression analysis was repeated by taking into account the neighborhood of the genes. Hence, the spot intensities of the genes 1 Mbps up- and downstream of each gene were averaged and the test was run again. By looking at this larger domains, a correlation of location and gene expression level could not be revealed (R^2_{adj} = -0.15). After having a closer look at the individual expression levels in the domains, a single gene (*LAMC1*) with a unproportionally high spot intensity in the "*GLUL* domain" could be identified and by excluding this gene from the regression analysis better fit (R^2_{adj} = 0.40) was obtained. However, since the correlation tends to vary based on the assumptions made for the input data, it is difficult to derive a general rule that can be applied on every gene or domain.

In summary, the present data suggests that there is no general correlation between gene expression level and radial gene positioning even not when considering larger chromosomal domains.

Local gene density accounts for radial positioning. Since, the activity dependent model for the radial positioning did not explain the observed gene positions in a robust and satisfying way, the question arose, whether the preference for the nuclear interior or periphery of a given gene could be linked to gene density or chromosome size as suggested by various other studies [74, 15, 3, 39]. Therein it was reported that small or gene dense chromosomes are located more toward the nuclear interior and large or gene poor chromosomes are preferentially found at the nuclear periphery. The same observations were made for single gene dense or gene poor chromosomal domains. Thus, a possible connection between chromosome sizes and radial positions or global gene density on the chromosomes and radial positions was tested first. Still, a linear regression analyses did not exhibit a significant correlation for neither of the two alternatives (chromosome size: R^2_{adj} = 0.17, global gene density: R^2_{adj} = -0.20).

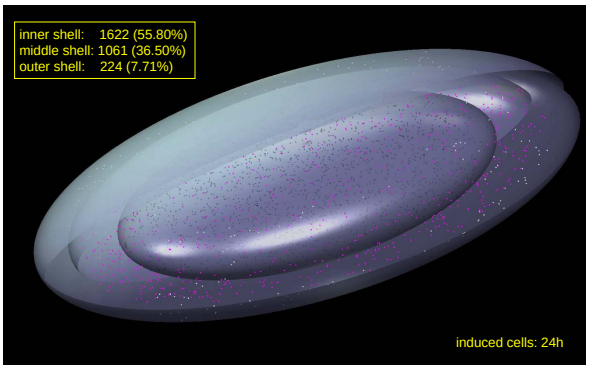


Figure 5: Nuclear gene positioning. Each nucleus was approximated by its enclosing ellipsoid, which was then subdivided into 3 shells of equal volume. Each of the seven genes was then assigned to the shell containing it. Here a standardized nucleus (median size) is shown. It summarizes the shell locations of all detected genes from ∼200 induced cells.

Next, the possibility of a linear relationship between chromosomal domain specific gene density and radial position was investigated. Therefore, gene density profiles along each chromosome were generated by a sliding window analysis, using a 10 Mbps window size and a 1 Mbps step size. The gene densities of the 10 Mbps domains to which any of the seven genes belonged to were then fed into a linear regression model together with the radial position frequencies. In contrast to the expression-, chromosome size-, and global gene density dependend models, a remarkable good correlation (R^2_{adj} = 0.67) of positioning and chromosomal domain specific gene density

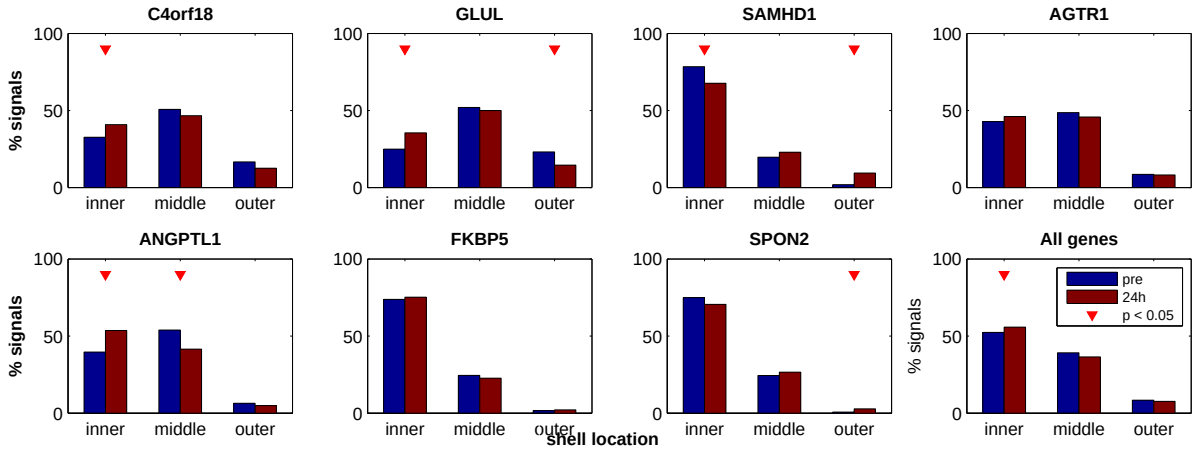


Figure 6: Nuclear gene positioning. The nuclear location counts of the individual genes are illustrated. The blue and red bars represent the percentages of gene signals found in a certain shell in preconfluent and induced cells respectively. The small red triangle indicates significant changes between the preconfluent and induced state (Fisher's exact test, $p \ll 0.05$).

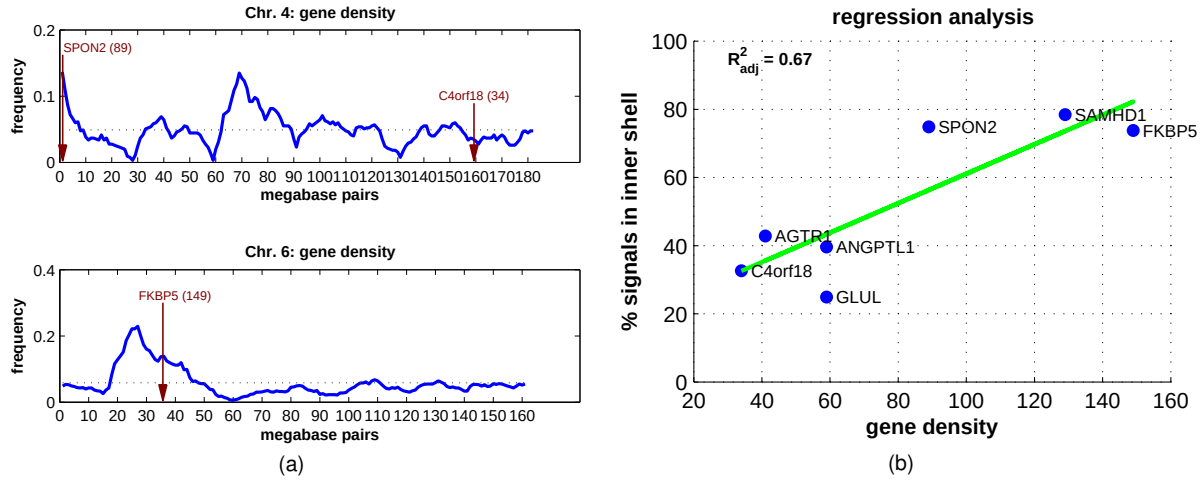


Figure 7: Chromosomal domain specific gene density and radial positioning. (a) A sliding window analysis: the blue line represents the gene density profile along the chromosome (4 upper- and 6 lower chart). The gene frequencies are denoted as ratios between the gene count in each window and the total number of genes present on the chromosome. The red arrows indicate the linear positions of *SPON2, C4orf18* and *FKBP5* with the number of genes in the corresponding 10 Mbps domains in brackets. (b) Regression analysis: the counts of genes positioned in the nuclear interior are plotted against the gene density of the chromosomal domain they belong to. A regression analysis shows that there is a high correlation between radial positioning and local gene density. The green line represents the regression line.

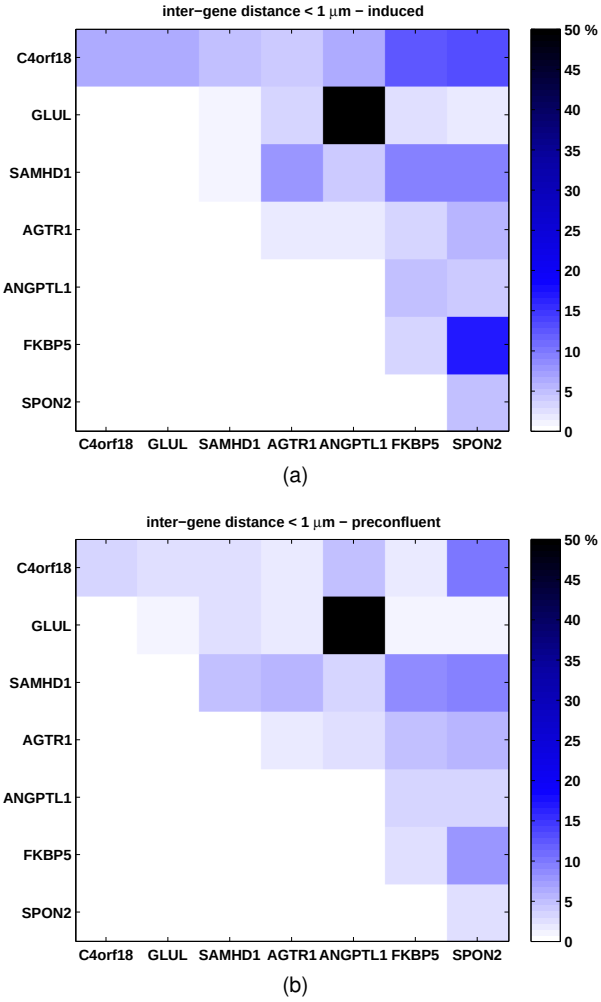


Figure 8: Inter-gene distances in induced (a) and preconfluent (b) cells. The gene-gene distances were measured in ∼200 nuclei. The heatmaps represent color-coded percentages of gene pairs found to be separated by a spatial distance of < 1 μm. The distances were measured from center to center of the two corresponding DNA-FISH signals.

was obtained. Figure 7a shows the gene density profiles along chromosome 4 and chromosome 6. *C4orf18* and *SPON2* are both located on chromosome 4, but they reside in chromosomal domains of different gene densities (∼3 fold). When looking at their radial positions, a clear preference for the nuclear interior can be noticed for the gene located in the domain with the higher gene density, whereas this is not the case for the gene residing in the domain bearing a low gene density. The same tendency was found for *FKBP5*, a gene in a gene dense domain on chromosome 6. Interestingly the two chromosomes, comparable in size, have a rather different global gene density. Nevertheless, the local gene density seems to be the determining factor for radial positioning, since neither the regression analysis on chromosome size, nor the one on global gene density revealed a high correlation. Furthermore, the regression graph in Figure 7b clearly demonstrates that these findings do not only reflect the situation for these three genes, but are true for each of the seven genes.

In summary, the results of the different analyses provide strong evidence for a radial positioning of genes which is primarily driven by local domain specific gene density. A link between positioning and chromosome size or global gene density could not be derived from the acquired data. To some degree the transcriptional activity of a gene or domain might have an influence on its radial positioning.

Active genes show close spatial association. According to the transcription factory model [30, 31, 14, 49], different genes can share discrete sites in the nucleus, where factors necessary for producing pre-mRNA, like RNA-Polymerase II, are concentrated. Genes associated with such sites can be transcribed simultaneously. Since the genes examined in the present study are co-regulated and co-expressed in human adipocyte differentiation, the question arose if they are also spatially associated and share a transcriptional environment. Therefore, the gene-gene distances in nuclei of induced cells were determined by measuring the distance between the centers of two corresponding DNA-FISH signals. Gene pairs that were separated by less than 1 μm were scored as spatially associated (see also [5]). In Figure 8 each square represents the color-coded percentage of nuclei that meet this criterion for the specific gene pair. Surprisingly, several gene pairs were found to be spatially associated in a remarkable high number of nuclei, namely *FKBP5* - *SPON2* in 17.2%, *C4orf18* - *SPON2* in 13%, *C4orf18* - *FKBP5* in 12.8%, and *GLUL* - *ANGPTL1* in 68%.

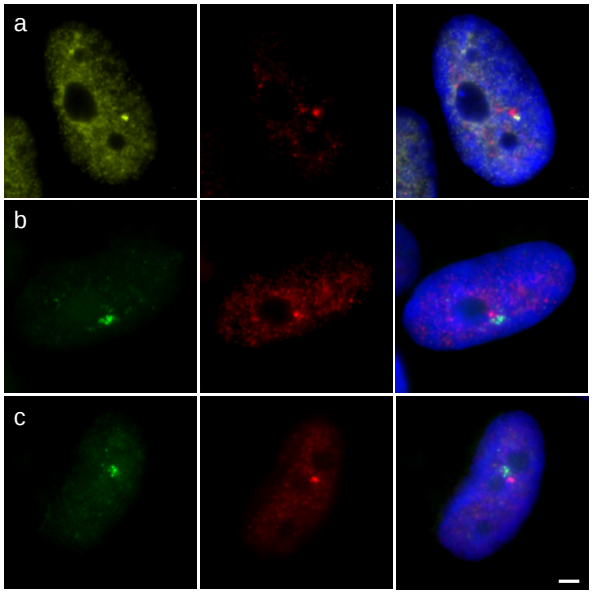


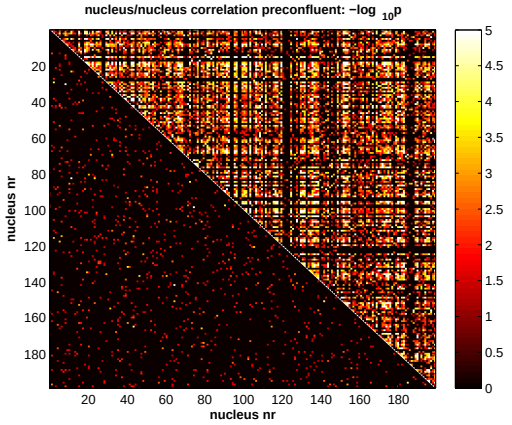
Figure 9: RNA-FISH of *FKBP5* - *SPON2* in induced cells. A dual color RNA-FISH experiment shows that *FKBP5* (green) and *SPON2* (red) are frequently localized in next proximity. Sometimes the signals overlap to some degree (b), sometimes they touch each other (a), and sometimes they are separated by $\ll 1$ μm (c). The white bar corresponds to a length of 2 μm.

As for the last pair, there is a simple explanation: both genes were located just 3.5 Mbps apart on the same chromosome. However, the genes from the remaining three pairs were all located on different chromosomes, yet they unexpectedly often appeared in close spatial proximity. The frequency of close encounters expected for a completely random nuclear organization is about 0.65%. This number was estimated based on the volumes of the nucleus and the DNA-FISH signals.

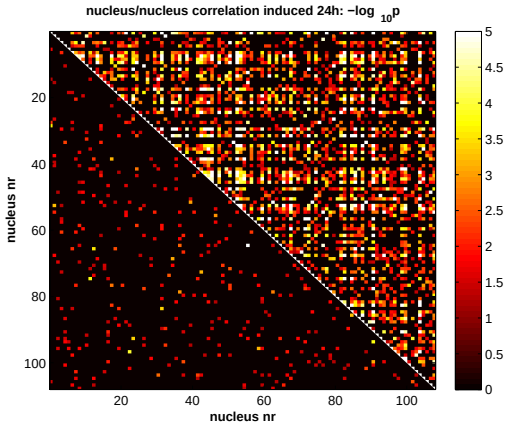
Next, the same analysis was done using preconfluent cells and the resulting data (Figure 8b) was compared with data obtained from induced cells (Figure 8a). By performing a Fisher's exact test, significant differences in the frequency of spatial association could be revealed for two gene pairs. For *C4orf18* - *FKBP5* an increase from 1.8% to 12.8% ($p < 10^{-5}$) and for *FKBP5* - *SPON2* an increase from 8.2% to 17,2% ($p < 6 \times 10^{-3}$) was detected. This increase goes along with a significant (6-10 fold) increase of the transcriptional activity of the three genes. Their spatial proximity could therefore imply that they associate in and with a nuclear environment that facilitates and/or enhances their transcription. Whether it also implies that they occupy one and the same RNA-Polymerase II transcription factory, was not directly tested, but it is likely that it frequently happens. As Osborne et al. [50] were showing, DNA-FISH signals of gene pairs that were associated with a RNA-Polymerase II focus had a median distance of $\sim$1 μm, and a notable subsection of them was associated with the very same Polymerase II focus. Since the distances presented here, were derived from center to center measurements of BAC probe ($\sim$70-200 kbps) DNA-FISH signals, the actual gene-gene distances might well be overestimated. As we reported previously [49], the DNA-FISH procedure might not be able to detect highly decondensed and therefore more fragile chromatin at higher denaturation temperatures, but exactly this loose consistency of chromatin is expected for actively transcribed genes. Thus it is likely, that the detected FISH-signals represent a more compact form of chromatin, from which the transcribed parts are looping out.

In order to test this possibility, a dual color RNA-FISH experiment was performed using intron probes for *FKBP5* and *SPON2*. This method detects primary transcripts which are expected to be localized in abundance at their transcription site. If the two co-expressed genes would indeed share a transcriptionally favored environment or a transcription factory, the signals from the RNA-FISH experiment should be located in next proximity [50]. Figure 9 displays examples for closely located primary transcripts of *FKBP5* and *SPON2*, found in induced hMADS cells. Consistent with the findings of Osborne, the FISH signals of these two co-expressed genes could be frequently observed in direct spatial vicinity. This provides support for the theory that co-expressed genes in general can exhibit interchromosomal interactions which further facilitate their transcription and that this mechanism is not restricted to just a few selected examples [10].

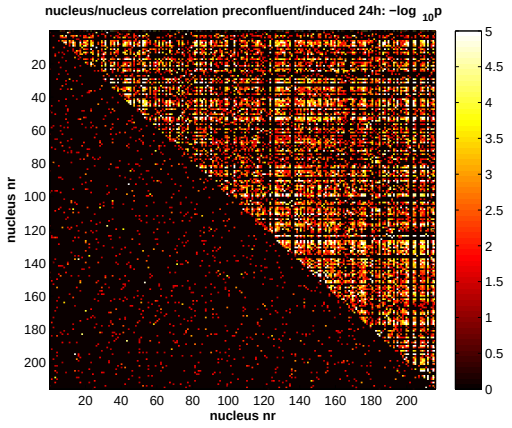
The global spatial organization of active genes is non random and preserved between different nuclei. After investigating the direct spatial association of active genes, the focus was set back on their global nuclear organization, try-



(a)



(b)



(c)

Figure 10: Correlation analysis of 3D-maps describing the spatial gene organization in each nucleus. The upper triangles of the heatmaps display significant correlations ($p < 0.05$) between each pair of nuclei in preconfluent cells (a), induced cells (b), and in a comparison of both (c). The lower triangles show the results of a random permutation test (see text for details).

ing to answer the following key questions: Is there a non random organization beyond the radial positioning? If so, is it preserved between the individual nuclei and maintained after the cells are induced for adipocyte differentiation?

To address these questions, the spatial organization of the seven genes in nuclei of preconfluent and induced cells, was described by relative 3D-maps. Such maps are defined by the distances of the map-coordinates and may therefore be reconstructed based on the knowledge of these distances. To define the nuclear organization, only the shorter distance between the allele coordinates of a given gene pair was considered. This resulted into 28 distances per nucleus which were then normalized by the individual nuclear volume, to allow for a comparison of the different nuclei. Vectors, each describing the spatial organization in a given nucleus, were created by assigning these normalized distances to their corresponding element positions. The pairwise Pearson correlation coefficients of these vectors were then calculated. The upper triangles of Figure 10 illustrate heatmaps of significant ($p < 0.05$) correlations. Each square represents the color-coded $-\log_{10}p$ value of the corresponding pairwise correlation. Surprisingly, there is a high number of nuclei that show a significant similarity in their spatial organization of the seven studied genes. This could be noticed in preconfluent cells (Figure 10a), as well as in induced ones (Figure 10b). Moreover it seems that this organization is preserved not only among nuclei of the same cell status, but also between nuclei of different status (Figure 10c).

To eliminate the possibility that the observed correlation could be obtained by chance, the measured intra-nuclear gene-gene distances were randomly permuted within each vector and the correlation coefficients as well as their p-values were recalculated. The lower triangles in Figure 10 demonstrate that a correlation, such as seen when analyzing vectors with the correct distance assignment, can not be created by an arbitrary assignment of the measured distances. These results give evidence for a non random spatial gene organization in hMADS cells. This organization is preserved between different nuclei and does not undergo dramatic changes up on induction of differentiation, although local changes can be observed.

Discussion

The here presented report provides a systematic, unbiased, and detailed analysis on the chromosomal and spatial organization of active and co-regulated genes. The results revealed a so far unreported non random chromosomal organization of adipogenesis related genes in a human cell system. Furthermore, this is the first study that investigated nuclear organization by simultaneously assessing the 3D-positions of multiple active genes in single cells under different physiological conditions. Thereby several important observations were made: First, it could be shown that the radial gene positioning is not dramatically changing upon induction of adipocyte differentiation. Second, it was observed that the radial positions of the active genes are not correlating with their expression level and it could be neither linked with the chromosome size nor with the global chromosome gene density. Third, it was found that the gene density in the chromosomal domains of the single genes accounts best for its radial positioning. In addition, it could be demonstrated that close encounters of active genes can also be observed for more general examples and not only for selected genes [10]. Using a novel and unique approach to describe the nuclear organization of multiple genes it was possible to compare the spatial gene arrangement of different nuclei. This revealed that the nuclear organization is non-random and preserved among different cells and even between cells under different physiological conditions.

Adipogenesis related genes are arranged non-randomly along human chromosomes. A microarray study on human adipocyte differentiation in hMADS cells identified 848 genes that were differentially expressed during the time course of adipogenesis [62]. By examining their chromosomal positions it could be demonstrated that a significant portion of the co-regulated genes was occupying adjacent locations. A random permutation test indicated that the observed number of gene tandems can not be obtained by chance (Figure 2). Furtheron, groups of co-regulated genes could be identified by extending the analysis of the chromosomal gene organization to domains (10 Mbps) (Figure 3). Such groups are expected for RIDGEs [8] that are normally represented by gene dense regions. Interestingly, grouping of genes was also found in regions of lower gene density, demonstrating that co-regulated genes, found by the microarray experiments, in general exhibit a tendency of being clustered in larger domains.

These results are consistent with the findings from recent studies in murine and human cell systems and tissues [35, 54], showing that the genome of mammalian organisms is organized in a linear fashion. Tandems and domains represent two levels of linear gene arrangement. It reflects a higher organization of the genome which may facilitate coordinated gene expression and/or repression [78]. If genes are grouped in linear proximity, then they may share regulatory environments in the nucleus where specific factors are concentrated. The placement of genes into linear neighborhood could also permit spreading histone modifications over larger chromosomal domains including multiple genes [18, 34].

Radial positioning of active genes is largely independent of their expression level. It was shown in previous studies that the activity of certain genes is associated with their positioning within the nucleus. Some authors have identified silencing of chromatin that is located at the nuclear periphery. Others have reported that the periphery is not necessarily a silencing compartment by showing examples of active genes and gene activation located at the nuclear periphery (reviewed in [34, 46, 40]). Transcriptional activity was also often linked with positioning of genes in the nuclear interior [41, 82, 81]. All these observations have been made on a few selected genes like β*-globin, IgK, IgH*, and *Mash1*. Recently published studies describe approaches for tethering chromosomal domains to the nuclear periphery [38, 55, 21]. Analyses regarding the activity of the hereby affected genes led to different conclusions. Finlan et al. [21] as well as Reddy et al. [55] found repression for active genes that were targeted to the nuclear periphery, whereas Kumaran [38] reports that the repositioned domain exhibited similar kinetics of induction compared to that observed in the nuclear interior.

In the here presented study the radial position-

ing seven genes on five different human chromosomes was investigated. These genes were selected in an unbiased fashion by considering their expression profiles and not by looking at their function or any other information. All of them showed a ∼6-10 fold increased transcriptional activity after 24 hours of induction for adipocyte differentiation. For some, a positional change with regard to the nuclear interior or periphery could be observed after induction. Interestingly, a general correlation of gene activity and positioning, as suggested by previous work [41, 82, 81], could not be deduced, even not when considering the transcriptional activity of entire domains that included the expression levels of flanking genes.

In a recent study on the positioning of 11 cancer-related genes Meaburn et al. came to a similar result [44]. This suggests that transcription rates of a gene or domain do not sufficiently explain its radial position and that changes of this rate do not necessarily lead to dramatic changes of its nuclear location. However, this does not rule out that nuclear positioning may play an important role in gene expression since a good correlation between activity and location was reported for certain genes that changed from a completely inactive to an active state or vice versa [41, 82, 81]. The results in this work suggest that positional changes depend sometimes on single genes in a given domain, but with the current data it is not possible to postulate a robust and general model based on the present data. This remains a challenging task for future investigations.

Local gene density accounts for radial positioning. The lack of correlation between expression level and nuclear positioning led to the question if the radial positions of the individual genes are related to gene density or chromosome size as it has been reported previously [74, 15, 3, 48, 39]. In contrast to those results, no correlation between chromosome size or global gene density and the nuclear positioning of the seven genes could be observed. However, a strong relationship could be seen between chromosomal domain specific gene density and radial positioning. This result supports the findings made in other cell systems [48, 39] and suggests that gene dense chromosomal domains, spanning several megabase pairs, are preferentially located at the interior of the nucleus in contrast to gene poor domains which tend to exhibit a preference for the periphery. Since RIDGEs [8] are located in gene dense regions, this observation may partly explain why active genes were often found to be occupying the nuclear interior.

General examples for active genes show close spatial association. Besides the positioning of individual genes with respect to nuclear landmarks, also their relative positioning to each other has lately experienced increasing interest from the research field. Close spatial association of transcribed genes is predicted by the transcription factory model [30, 31, 14, 49] and was recently reported for some selected genes in different cell systems [50, 5, 51]. One key question of the study was, if sharing of a transcriptional environment between genes is a general characteristic for expression of co-regulated genes [10].
Three spatially associated gene pairs could be identified by investigating the gene-gene distances of the seven unbiasedly chosen genes. They were juxtaposed in a remarkably high number of nuclei, ranging from 12.8% to 17.2%. The expected number for such close encounters in a completely random organization is below 1%. Two of the pairs indicated a higher number of encounters when their transcriptional activity was increased after induction: *C4orf18* - *FKBP5* changed from 1.8% to 12.8% ($p < 10^{-5}$) and *FKBP5* - *SPON2* from 8.2% to 17,2% ($p < 6 \times 10^{-3}$). These results imply that the spatial association of transcriptionally active genes is not only restricted to a few selected examples. A close relative positioning may be important for a coordinated and enhanced expression of the partners.

The global spatial organization of active genes is non random and preserved between different nuclei. The relative positioning of multiple genes - beyond close spatial association - was not studied so far. By developing and employing a 3D combinatorial M-FISH procedure it was possible to simultaneously visualize and identify seven different genes under two distinct physiological conditions. This technique allowed for a detailed analysis of their spatial relationship by using a novel approach: Relative 3D-maps, which were

defined by the individual gene-gene distances, were adopted to describe the nuclear organization of the genes in ∼400 cells. A correlation analysis of these 3D-maps revealed a strikingly high and non random correlation between different nuclei. The correlation applied not to nuclei under the same physiological condition, it was even observed when comparing nuclei under different conditions.

This observation indicates that the global spatial arrangement of the investigated genes is non random and largely preserved between different cells. Furthermore, a change of the physiological conditions does not imply a dramatic change of the global nuclear gene organization. If this is true for any gene in any organism under any condition can not be assured, but it is likely that one can get similar results by examining the nuclear organization in other systems.

Conclusion

It was demonstrated that the genes that were differentially expressed during adipocyte differentiation are non randomly organized on the chromosomal level as well as in the three dimensional space of the nucleus. The results indicate a strong relationship between local gene density and nuclear positioning, whereas the transcription rates of the studied genes do not correlate with their radial positions. Still, a possible link between activity and positioning of single genes or domains can not be ruled out and remains subject to future investigations.

It could be demonstrated that not only radial positioning but also a spatial close association of active genes is a generic property of nuclear architecture. Furthermore, a novel approach for analyzing the spatial organization of multiple genes could be developed and revealed that this organization is non-random and is preserved between different nuclei.

This study provides further insight into the linear- and nuclear architecture of the human genome. It is a step towards a better understanding of complex organizational patterns involving multiple genes.

Materials and Methods

Bioinformatics

Microarray- and chromosomal data analysis. Co-regulated genes that were discovered in a microarray study [62] on adipogenesis in human multipotent adipose-derived stem cells (hMADS) [58, 57], were investigated to ascertain their linear organization on the chromosomes.

All 29,952 sequences of the oligonucleotide microarray (ArrayExpress, accession numbers A-MARS-2, E-MARS-5, and E-MARS-6, [37]) were mapped to the NCBI human genome alignment (version 36) using BLAT [33]. The corresponding gene names from HGNC [6] were assigned to the resulting locations. Redundant locations or locations with no associated HGNC name were removed from the dataset. This resulted in ∼14,000 genes present on the array and 848 differentially expressed genes (regulated) in hMADS cells during adipogenesis.

Regulated tandem gene sets were identified by scanning the chromosomes for groups of regulated genes that have no intercalating, unregulated gene present on the microarray. For the random permutation test, the locations of the ∼14,000 genes on the microarray were randomly permuted and the scanning for tandems was repeated as described above. This process was iterated 10,000 times. The tandem gene analysis was carried out for the complete set of differentially expressed genes and for each of the 12 clusters of co-regulated genes, revealed by k-means clustering (k=12).

A sliding window approach was employed to study the linear organization beyond tandem genes. Thereby a 10 megabase pair (Mbps) window was moved in 1 Mbps steps along the chromosomal sequence. The frequency of the genes on microarray, represented as ratio between gene count in the window and total microarray genes on the chromosome, was recorded. The frequency of regulated genes was registered in the same way. They were represented as ratio between the number of genes in the window and the total number of regulated genes on the chromosome. This analysis was applied to all human chromosomes and significantly different regions were uncovered

by an exact binomial test. For the entire linear gene organization analysis Matlab® R2007b (MathWorks Inc.) and Perl were used as programming languages.

Gene selection by microarray data analysis. The seven genes which were studied in terms of their nuclear localization were determined by analyzing the gene expression profiles obtained from the microarray study (see above). Briefly: microarray expression data was grouped into sets of genes characterized by similar transcription profiles (Figure 1). This was achieved by applying the k-means clustering algorithm (k=12) available with the Genesis software package [73]. Cluster 10 which contained 63 genes that showed an upregulation in the early phase of adipogenesis was selected. Further subdivision by k-means clustering (k=7) revealed 7 genes showing a verify similar expression profile (*C4orf18, GLUL, SAMHD1, AGTR1, ANGPTL1, FKBP5, SPON2*). They were selected for the nuclear localization study. Their genomic locations were determined by mapping the sequence of the corresponding oligonucleotide probe from the microarray to the genome sequence: *C4orf18* - Chr 4 159.3 Mbps, *GLUL* - Chr 1 180.6 Mbps, *SAMHD1* - Chr 20 35.0 Mbps, *AGTR1* - Chr 3 149.9 Mbps, *ANGPTL1* - Chr 1 177.1 Mbps, *FKBP5* - Chr 6 35.7 Mbps, *SPON2* - Chr 4 1.1 Mbps

DNA-FISH probe selection. FISH-mapped BAC-clones [9] containing the genomic sequences for the seven genes of interest were identified by mining the BAC clone database of the BAC Resource consortium and ordered from ImaGenes GmbH (former German Resource Centre for Genome Research, RZPD): RP11-89C4 for *C4orf18*, RP1-223H12 for *GLUL*, RP1-132F21 *SAMHD1*, RP11-505J9 for *AGTR1*, RP4-595C2 for *ANGPTL1*, RP3-368C2 for *FKBP5* and RP11-20I20 for *SPON2*. Each BAC clone was verified by colony PCR using specific primers: *C4orf18* AGCCATAGCGCCTCTATCAA, GCCGGACTTGCTTAATGTGT; *GLUL* GACGGGAGTTTCATCGTGTT, ACTGATCCCCTGCAAAAATG; *SAMHD1* GGGACCAGCTGATATCTCCA, CAAACACAACCCTGCCTTCT; *AGTR1* GGACTGTTATCCAGGGCAGA, TTTGATCACCTGGGTCGAAT; *ANGPTL1* CCAAGGTGCTGTTGGAAAAT, TAGCGCACATGGTAGCAAAG; *FKBP5* TTGGCATGAATTAGGCAACA, TTGTTGTGGCTTTTGGTTGA; *SPON2* CAGCGGACTTCAGACTTTCC, ACCATAGCAACCTCGGACAC

RNA-FISH probe selection. Probes for detecting newly synthesized primary transcripts of *FKBP5* and *SPON2* were designed by the following method. The genomic sequences of the genes of interest were obtained from the Ensembl genome database (release 35) [28, 29]. Intron sequences were extracted and the repeats were masked using the RepeatMasker [65] software. Intron sequences which did not show repeats were then used for generating specific RNA-FISH probes by PCR. The primers for the PCR products with a size of $\sim$2000 base pairs and a melting temperature of 60°C were designed using Primer3 [61]. The isPCR program [32] was used to verify the primers in silico. To check for possible of cross-hybridizations of the primers, a BLAST search [1] against the human genome sequence database from Ensembl (release 35) using WU-BLAST [23] was performed. By this method the following primer sequences were obtained: *FKBP5* forward primer TTGGCATGAATTAGGCAACA, reverse primer TTGTTGTGGCTTTTGGTTGA; *SPON2* forward primer CAGCGGACTTCAGACTTTCC, reverse primer ACCATAGCAACCTCGGACAC.

Cell preparation

Cell culture. hMADS were cultured in low glucose (1 g/L) Dulbecco's modified Eagle's medium (DMEM) supplied with 2 mM glutamine, 10mM HEPES, 10 μg/ml penicillin/streptomycin, 0.1 mg/ml normocin, 10% fetal bovine serum (FBS) and 0.0025 μg/ml human fibroblast growth factor 2 (hFGF2). The cells were seeded on sterilized coverslips (Corning #1.5). This proliferation medium was exchanged every second day. Preconfluent cells were fixed after the culture reached 80-90% confluence. For obtaining induced cells, the cells were grown to confluence and maintained in proliferation medium for two additional days. Thereafter the medium was exchanged for an induction medium consisting of 50% low glucose (1g/L)

DMEM, 50% Ham's F-12, 5 mM HEPES, 1 mM glutamine, 10 μg/ml penicillin/streptomycin, 0.1 mg/ml normocin, 5 μg/ml human insulin, 10 μg/ml apo-Transferrin, 0.0002 μM T3 Triiodothyronin, 0.1 μM Rosiglitazone, 0.1 mM isobutyl-methylxanthine (IBMX), 0.001 mM Dexamethason. The cells were kept for 24 hours in the induction medium before fixing.

Cell fixation and permeabilization. Cells were fixed and permeabilized according to procedures that preserve three-dimensional structure and morphology of the nuclei [27, 68, 67, 75]. Coverslips were removed from the culturing medium and permeabilized for 1 minute with 0.5% Triton X-100 in CSK Buffer (100 mM NaCl, 200 mM sucrose, 10 mM PIPES pH 6.8, 3 mM $MgCl_2$) on ice. When the cells were prepared for RNA-FISH experiments, the CSK Buffer was prepared with DEPC treated H_2O and 2 mM VRC (Vanadyl Ribonucleoside Complexes) in order to inhibit RNases. Immediately after permeabilization the cells were fixed for 25 minutes in freshly prepared 3.5% PFA (paraformaldehyde) dissolved in PBS. After fixation the cells were washed 2×5 min. in PBS. When the cells were not used immediately for an experiment, the coverslips were stored in 70% EtOH at 4°C up to several days and rehydrated for 10 minutes in PBS before use.

3D combinatorial M-FISH

Probe generation and indirect labeling. DNA from the seven BAC-clones described above was isolated using QIAGEN® Large-Construct Kit following the procedure from the products manual. The DNA was then labeled separately with BIO-16-dUTP (biotin), DIG-11-dUTP (digoxigenin) or DNP-11-dUTP (dinitrophenyl). The Biotin or DIG Nick Translation Mix (Roche Applied Science) was used for applying Biotin and DIG labels. Probes tagged with DNP labels were obtained by using a plain Nick Translation Mix (Roche Applied Science) in combination with DNP-11-dUTPs (Perkin Elmer®). All labeling reactions were performed according to the protocol supplied by the manufacturer of the used kits. In order to obtain the label combinations, shown in Table 1, one to three separate nick-translations of the corresponding BAC DNA, with each of the three labels, was necessary.

Hybridization mix. A discrimination between 7 different targets by using just 3 different filtersets and corresponding fluorochromes was achieved by a combinatorial approach [56, 71, 20, 80] in which all possible boolean combinations (n=3 labels, 2^n-1 = 7) of the nick-translated probes, carrying 3 different labels, were used (see Table 1). The hybridization cocktail was composed of 100 ng of each of the contributing nick-translated probes (1.2 μg in total), 56 μg salmon sperm DNA and 40 μg human cot1 DNA (Invitrogen™). The DNA was ethanol precipitated and redissolved in 50% formamide/2×SSC 5% dextran sulfate. Prior to hybridization the DNA was heat denaturated for 10 minutes at 85°C and pre-hybridized for 20 minutes at 37°C.

Hybridization. Nuclei that were preserved in their 3D morphology, were prepared for hybridization by a 20 minute RNase A (25 μg/ml in PBS) treatment followed by washes in PBS for 3×5 minutes, 1×10 minutes in 2xSSC and 1×3 minutes 70% formamide/2xSSC. The chromatin was then denaturated by heat for 4.5 minutes at 85°C in 70% formamide/2×SSC. Afterwards the nuclei were dehydrated by a series of washes in 70%, 90%, and absolute EtOH on ice for 5 minutes each. Then the coverslips were air-dried for 5 minutes and the hybridization cocktail, containing the combinatorially mixed probes for simultaneous detection of all 7 genes, was applied, sealed with rubber cement, and hybridized over 2 nights at 37°C in a humidified chamber. After the hybridization a series of stringency washes was performed: 2×15 minutes 50% formamide/2×SSC at 45°C, 10 minutes 0.1×SSC, 10 minutes 2×SSC, and 5 minutes 4×SSC.

Detection of the probes. The biotinilated DNA probes were detected by a 45 minute incubation at 37°C with streptavidin Alexa Fluor® 647 (Invitrogen™) diluted 1:200 in 4×SSC/0.1% BSA/0.01% Tween 20. Washes in 4×SSC/0.1% TWEEN20 for 10 minutes, 4×SSC for 2×10 minutes, 2×SSC for 10 minutes, and PBS for 5 min-

utes succeeded the incubation step. The digoxigenin and dinitrophenyl labeled DNA probes were simultaneously detected by a 60 minute incubation with anti-digoxigenin-rhodamine (Roche Applied Science) and anti-dinitrophenyl-KLH Alexa Fluor® 488 antibodies, both antibodies were diluted 1:200 in 4%BSA/0.1% Tween 20/PBS. After washing 3×5 minutes with PBS, the nuclei were counterstained with 0.05μg/ml DAPI in PBS for 5 minutes. The coverslips were mounted on glass slides with the SlowFade® Gold (Invitrogen™) mounting medium. The cells were imaged immediately after mounting.

RNA-FISH

Probe generation. Unlabeled RNA-FISH probes were generated via PCR in 45 cycles of amplification using the specific intron primers described above. Isolated BAC-DNA form BAC clones containing the sequences of *SPON2* (clone RP11-20I20) and *FKBP5* (clone RP3-368C2) was used as templates. The PCR products were purified using the QiaQuick® PCR purification Kit (QIAGEN®) and labeled directly by generating amine-modified DNA which was then coupeled to a fluorescent dye. The amine-modified probes were produced by nick-translation using the Nick Translation Mix (Roche Applied Science). The 5× labeling mix was composed of dATP, dCTP, dGTP, dTTP and amino-allyl-dUTP in a mixture ratio of 1 : 1 : 1 : 0.25 : 1.5. After nick-translation the probes were purified once again by the QiaQuick® PCR purification Kit according to the products protocol, but with the difference that the wash and elution buffers were exchanged for phosphate buffers (5 mM $KP0_4$ pH 8.0 / 80% EtOH as wash buffer, 4 mM KPO_4, pH 8.5 as elution buffer). The amine-modified DNA was dried in a SpeedVac® and coupled to an amine-reactive fluorescent dye (Cy3 for *FKBP5*, Alexa Fluor® 488 for *SPON2*) according to the dye coupling procedure from the ARES™DNA Labeling Kit (Invitrogen™). Purification of the dye-labeled DNA was performed utilizing the QiaQuick® PCR purification Kit. After drying the fluorescently labeled DNA in a SpeedVac® it was redissolved in 20 μl ddH_2O.

Hybridization. Three-dimensionally preserved and permeabilized nuclei on a 1/4 of a coverslip were dehydrated by a series of washes in 70%, 90%, and absolute EtOH on ice for 5 minutes each. The cells were air-dried and hybridized over night in a humidified chamber at 37°C with 5μl of the hybridization-cocktail consisting of differently labeled DNA probes for *FKBP5* and *SPON2* (20 ng each), 2×SSC, 5% dextran sulfate, 1 mg/ml yeast t-RNA and 50% formamide.
After hybridization the coverslip was washed 3×5 minutes in 50% formamide/2×SSC at 37°C, 15 minutes 2×SSC and 5 minutes in PBS. The nuclei were counterstained with 0.05μg/ml DAPI in PBS for 5 minutes.
The coverslips were mounted on glass slides with the SlowFade® Gold (Invitrogen™) mounting medium and imaged immediately.

Microscopy and image analysis

Microscopic image acquisition. All microscopic images were recorded on a Zeiss AxioImager Z1 epifluorescence microscope equipped with an AxioCam MRm CCD camera, a motorized stage and a Heidenhain closed loop system for highly precise and reproducible z-stacks of equidistant step sizes. The 3D-image stacks were collected using a 100× Plan-Apochromat 1.4 NA oil immersion lens with a axial sampling distance of 200 nm. This resulted in a voxel size of 64.3×64.3×200 nm. The stacks for the 3D combinatorial DNA M-FISH (4 channels) and the 2-color RNA FISH (3 channels) were composed of 28 slices (512×512 pixels) per channel, resulting in 112 or 84 single images per stack. The raw image stacks were stored as zvi files in the *i*LAP system [72] before further analysis.

Image restoration. All recorded image stacks were processed by 3D-deconvolution [43, 12] in order to reduce out-of-focus light and improve their quality. This task was performed by running 100 iterations of the maximum likelihood estimation algorithm available with the Huygens Deconvolution Software (Scientific Volume Imaging - SVI http://www.svi.nl). The computational intensive process was carried out on a high perfor-

mance computing cluster and coordinated automatically by the *i*LAP system. The deconvolved image stacks were then corrected for color-shift errors which derive from chromatic aberrations and produce misalignments of the single channels. The shift vectors were determined empirically by recording z-stacks of 0.5 μm Ø fluorescent beads (TetraSpeck™, Invitrogen™) in the channels in question and then calculating the maximum correlation vectors between them. This calculation was done in Huygens Scripting (SVI) for 10 different beads and the modal values for xyz of the single shift vectors between two channels were then used as standard channel shift vector for these two channels.

3D image reconstruction. For the detection and measurement of the FISH-signals, the deconvolved and color-shift-corrected image stacks were reconstructed in 3D. The reconstruction method described here was implemented and realized with Imaris (Bitplane Inc.) using the ImarisXT interface and the Matlab® R2007b (MathWorks Inc.) programming language which also included the image processing and statistics toolboxes. The procedure involved of the following 5 steps:

1) *Removal of nuclei touching the image borders in yx*. This was achieved by z-projecting the DAPI-channel using a mean intensity projection (MIP). The resulting MIP-image was then thresholded by Otsu's method [52] followed by an erosion and a dilation [77, 26] using a ball structure element. Structures that were lighter than their surroundings and that were connected to the image border [66] were suppressed. From the resulting image an image mask was created (zeros and ones) and each slice of the DAPI channel in the stack multiplied with this mask.

2) *Generation of an isosurface in the masked DAPI channel that delineates the nucleus*. To generate a nucleus surface, the "*DetectIsoSurface*" function of ImarisXT was used with the lower threshold set to 10% of the intensity threshold derived by Otsu's method, the upper threshold set to the maximum intensity of the DAPI channel and a Gaussian filter width set to 0.53.

3) *Masking the FISH signal channels*. Any extranuclear background was removed by setting voxel intensities outside of the nuclear surface to zero.

4) *Detection of FISH signals*. For the DNA FISH signals the "*DetectSpots*" function of ImarisXT was used. It selects for local intensity maxima in the image. By automatically adjusting the intensity threshold the expected number of Spots (8 DNA FISH signals per channel) could be detected. The estimated maximum diameter of the spots was set to 0.3 μm.

5) *Identification of the single combinatorial labeled BACs*. The combinatorial labeling was resolved by k-means clustering [69] of all detected spot coordinates from each channel. The k-value was thereby set to the number of the hybridized BACs, usually 14 (2×7). The corresponding gene names were then assigned to the centroids of the resulting cluster according to the labeling schema from Table 1.

The following data of the reconstructed 3D-image was measured and recorded: 3D-coordinates of the detected BACs and the nucleus center, distance of each BAC to the nucleus center, all distances between the single BACs, shortest distance of each single BAC to the nucleus surface, volume of the nucleus and axes lengths including the rotation matrix of the ellipsoid that approximates and describes the nucleus surface.

Data analysis

The distance-, localization- and statistical analysis was performed using the Matlab® R2007b (MathWorks Inc.) programming environment.

Nuclear localization. To assess the preferred nuclear localization of the individual genes, each nucleus was first approximated by an ellipsoid, which was then subdivided into three shells of equal volume delineated by concentric ellipsoids of equal shape. The frequency of gene localization in each shell was determined over all nuclei. A Fischer's exact test was applied to ascertain the statistical significance of differences in the preferred localizations.

Regression analysis. All regression analyses were conducted by running the "*regstats*" function of Matlab. The results were interpreted using adjusted R^2 (R^2_{adj}) statistics.

Expression level estimation. The expression level of the seven studied genes was roughly estimated by averaging the median microarray spot intensity over 3 biological replicates. This was eligible because the feature spots on the microarray were normalized to exhibit equal DNA concentrations.

Global and local gene density. The global gene density of a given chromosome was calculated as number of corresponding HGNC-genes (Ensembl release 48) per megabase pair. To assess the local domain-inherent gene density a sliding window analysis was performed by advancing a 10 Mbps window in 1 Mbps steps along the chromosomes and recording the number of HGNC-genes in each window. The frequencies are represented as ratio between genes in the window and total genes present on the chromosome.

Gene-Gene distances. The measurements of distances between single genes were made from center to center of the FISH-signal. Per cell, only the shortest distance of all possible distances between two genes (2+2 alleles) was taken into account. For the general comparison of spatial gene-gene relations, the distances were normalized for the volume of the nucleus in which they were measured [42].

Correlation of nuclear gene organization. To determine whether the different nuclei show a correlation in their spatial gene organization, the pairwise Pearson Correlation Coefficients between each nucleus, which was represented as vector of the volume-normalized gene-gene distances, was calculated. Only nuclei with no missing values were included in this analysis. Correlations were considered as significant if the p-value of the correlation coefficient was smaller then 0.05. To test for the possibility that the observed correlations between nuclei could be obtained by a random assignment of the measured gene-gene distances, the values of the inter-gene distance were randomly permuted within the vectors representing a nucleus. Using these permuted vectors, the correlation coefficients were calculated again.

References

[1] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, Oct 1990.

[2] François-Michel Boisvert, Silvana van Koningsbruggen, Joaquín Navascués, and Angus I Lamond. The multifunctional nucleolus. *Nat Rev Mol Cell Biol*, 8(7):574–585, Jul 2007.

[3] Andreas Bolzer, Gregor Kreth, Irina Solovei, Daniela Koehler, Kaan Saracoglu, Christine Fauth, Stefan Müller, Roland Eils, Christoph Cremer, Michael R Speicher, and Thomas Cremer. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol*, 3(5):e157, May 2005.

[4] Alexander M Boutanaev, Alla I Kalmykova, Yuri Y Shevelyov, and Dmitry I Nurminsky. Large clusters of co-expressed genes in the drosophila genome. *Nature*, 420(6916):666–669, Dec 2002.

[5] Jill M Brown, Joanne Leach, Joyce E Reittie, Ann Atzberger, Jane Lee-Prudhoe, William G Wood, Douglas R Higgs, Francisco J Iborra, and Veronica J Buckle. Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol*, 172(2):177–187, Jan 2006.

[6] Elspeth A Bruford, Michael J Lush, Mathew W Wright, Tam P Sneddon, Sue Povey, and Ewan Birney. The hgnc database in 2008: a resource for the human genome. *Nucleic Acids Res*, 36(Database issue):D445–D448, Jan 2008.

[7] Eva Bártová and Stanislav Kozubek. Nuclear architecture in the light of gene expression and cell differentiation studies. *Biol Cell*, 98(6):323–336, Jun 2006.

[8] H. Caron, B. van Schaik, M. van der Mee, F. Baas, G. Riggins, P. van Sluis, M. C. Hermus, R. van Asperen, K. Boon, P. A. Voûte,

S. Heisterkamp, A. van Kampen, and R. Versteeg. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*, 291(5507):1289–1292, Feb 2001.

[9] V. G. Cheung, N. Nowak, W. Jang, I. R. Kirsch, S. Zhao, X. N. Chen, T. S. Furey, U. J. Kim, W. L. Kuo, M. Olivier, J. Conroy, A. Kasprzyk, H. Massa, R. Yonescu, S. Sait, C. Thoreen, A. Snijders, E. Lemyre, J. A. Bailey, A. Bruzel, W. D. Burrill, S. M. Clegg, S. Collins, P. Dhami, C. Friedman, C. S. Han, S. Herrick, J. Lee, A. H. Ligon, S. Lowry, M. Morley, S. Narasimhan, K. Osoegawa, Z. Peng, I. Plajzer-Frick, B. J. Quade, D. Scott, K. Sirotkin, A. A. Thorpe, J. W. Gray, J. Hudson, D. Pinkel, T. Ried, L. Rowen, G. L. Shen-Ong, R. L. Strausberg, E. Birney, D. F. Callen, J. F. Cheng, D. R. Cox, N. A. Doggett, N. P. Carter, E. E. Eichler, D. Haussler, J. R. Korenberg, C. C. Morton, D. Albertson, G. Schuler, P. J. de Jong, B. J. Trask, and B. A. C. Resource Consortium. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*, 409(6822):953–958, Feb 2001.

[10] Chien-Hui Chuang and Andrew S Belmont. Close encounters between active genes in the nucleus. *Genome Biol*, 6(11):237, 2005.

[11] B. A. Cohen, R. D. Mitra, J. D. Hughes, and G. M. Church. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet*, 26(2):183–186, Oct 2000.

[12] José-Angel Conchello and Jeff W Lichtman. Optical sectioning microscopy. *Nat Methods*, 2(12):920–931, Dec 2005.

[13] Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915):520–562, Dec 2002.

[14] P. R. Cook. The organization of replication and transcription. *Science*, 284(5421):1790–1795, Jun 1999.

[15] Marion Cremer, Katrin Küpper, Babett Wagler, Leah Wizelman, Johann von Hase, Yanina Weiland, Ludwika Kreja, Joachim Diebold, Michael R Speicher, and Thomas Cremer. Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. *J Cell Biol*, 162(5):809–820, Sep 2003.

[16] T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2(4):292–301, Apr 2001.

[17] Thomas Cremer, Marion Cremer, Steffen Dietzel, Stefan Müller, Irina Solovei, and Stanislav Fakan. Chromosome territories–a functional nuclear landscape. *Curr Opin Cell Biol*, 18(3):307–316, Jun 2006.

[18] Wouter de Laat and Frank Grosveld. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*, 11(5):447–459, 2003.

[19] M. Dundr and T. Misteli. Functional architecture in the cell nucleus. *Biochem J*, 356(Pt 2):297–310, Jun 2001.

[20] C. Fauth and M. R. Speicher. Classifying by colors: Fish-based genome analysis. *Cytogenet Cell Genet*, 93(1-2):1–10, 2001.

[21] Lee E Finlan, Duncan Sproul, Inga Thomson, Shelagh Boyle, Elizabeth Kerr, Paul Perry, Bauke Ylstra, Jonathan R Chubb, and Wendy A Bickmore. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet*, 4(3):e1000039, Mar 2008.

[22] Peter Fraser and Wendy Bickmore. Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447(7143):413–417, May 2007.

[23] W. Gish. Wu-blast. *http://blast.wustl.edu*, 1996-2004.

[24] Simon G Gregory et al. A physical map of the mouse genome. *Nature*, 418(6899):743–750, Aug 2002.

[25] Hubert Hackl, Thomas Rainer Burkard, Alexander Sturn, Renee Rubio, Alexander

Schleiffer, Sun Tian, John Quackenbush, Frank Eisenhaber, and Zlatko Trajanoski. Molecular processes during fat cell development revealed by gene expression profiling and functional annotation. *Genome Biol*, 6(13):R108, 2005.

[26] R.M. Haralick and L. G. Shapiro. *Computer and Robot Vision*, volume 1, pages 158–205. Addison-Wesley, 1992.

[27] Claudia Hepperger, Simone Otten, Johann von Hase, and Steffen Dietzel. Preservation of large-scale chromatin structure in fish experiments. *Chromosoma*, 116(2):117–133, Apr 2007.

[28] T. Hubbard, D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen, T. Down, R. Durbin, E. Eyras, J. Gilbert, M. Hammond, L. Huminiecki, A. Kasprzyk, H. Lehvaslaiho, P. Lijnzaad, C. Melsopp, E. Mongin, R. Pettett, M. Pocock, S. Potter, A. Rust, E. Schmidt, S. Searle, G. Slater, J. Smith, W. Spooner, A. Stabenau, J. Stalker, E. Stupka, A. Ureta-Vidal, I. Vastrik, and M. Clamp. The ensembl genome database project. *Nucleic Acids Res*, 30(1):38–41, Jan 2002.

[29] T. J P Hubbard, B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney. Ensembl 2007. *Nucleic Acids Res*, 35(Database issue):D610–D617, Jan 2007.

[30] F. J. Iborra, A. Pombo, D. A. Jackson, and P. R. Cook. Active rna polymerases are localized within discrete transcription "factories' in human nuclei. *J Cell Sci*, 109 (Pt 6):1427–1436, Jun 1996.

[31] F. J. Iborra, A. Pombo, J. McManus, D. A. Jackson, and P. R. Cook. The topology of transcription by immobilized polymerases. *Exp Cell Res*, 229(2):167–173, Dec 1996.

[32] D. Karolchik, R. M. Kuhn, R. Baertsch, G. P. Barber, H. Clawson, M. Diekhans, B. Giardine, R. A. Harte, A. S. Hinrichs, F. Hsu, K. M. Kober, W. Miller, J. S. Pedersen, A. Pohl, B. J. Raney, B. Rhead, K. R. Rosenbloom, K. E. Smith, M. Stanke, A. Thakkapallayil, H. Trumbower, T. Wang, A. S. Zweig, D. Haussler, and W. J. Kent. The ucsc genome browser database: 2008 update. *Nucleic Acids Res*, 36(Database issue):D773–D779, Jan 2008.

[33] W. James Kent. Blat–the blast-like alignment tool. *Genome Res*, 12(4):656–664, Apr 2002.

[34] Steven T Kosak and Mark Groudine. Form follows function: The genomic organization of cellular differentiation. *Genes Dev*, 18(12):1371–1384, Jun 2004.

[35] Steven T Kosak, David Scalzo, Sam V Alworth, Fusheng Li, Stephanie Palmer, Tariq Enver, James S J Lee, and Mark Groudine. Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol*, 5(11):e309, Nov 2007.

[36] Stanislav Kozubek, Emilie Lukásová, Pavla Jirsová, Irena Koutná, Michal Kozubek, Alena Ganová, Eva Bártová, Martin Falk, and Renata Paseková. 3d structure of the human genome: order in randomness. *Chromosoma*, 111(5):321–331, Dec 2002.

[37] Birgit Kulterer, Gerald Friedl, Anita Jandrositz, Fatima Sanchez-Cabo, Andreas Prokesch, Christine Paar, Marcel Scheideler, Reinhard Windhager, Karl-Heinz Preisegger, and Zlatko Trajanoski. Gene expression profiling of human mesenchymal stem cells derived from bone marrow during expansion and osteoblast differentiation. *BMC Genomics*, 8:70, 2007.

[38] R. Ileng Kumaran and David L Spector. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J Cell Biol*, 180(1):51–65, Jan 2008.

[39] Katrin Küpper, Alexandra Kölbl, Dorothee Biener, Sandra Dittrich, Johann von Hase, Tobias Thormeyer, Heike Fiegler, Nigel P Carter, Michael R Speicher, Thomas Cremer, and Marion Cremer. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma*, 116(3):285–306, Jun 2007.

[40] Christian Lanctôt, Thierry Cheutin, Marion Cremer, Giacomo Cavalli, and Thomas Cremer. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet*, 8(2):104–115, Feb 2007.

[41] E. Lukásová, S. Kozubek, M. Kozubek, M. Falk, and J. Amrichová. The 3d structure of human chromosomes in cell nuclei. *Chromosome Res*, 10(7):535–548, 2002.

[42] Kirk J McManus, David A Stephens, Niall M Adams, Suhail A Islam, Paul S Freemont, and Michael J Hendzel. The transcriptional regulator cbp has defined spatial associations within interphase nuclei. *PLoS Comput Biol*, 2(10):e139, Oct 2006.

[43] J. G. McNally, T. Karpova, J. Cooper, and J. A. Conchello. Three-dimensional imaging by deconvolution microscopy. *Methods*, 19(3):373–385, Nov 1999.

[44] Karen J Meaburn and Tom Misteli. Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J Cell Biol*, 180(1):39–50, Jan 2008.

[45] Jason Mezey, Sergey Nuzhdin, Fangfei Ye, and Corbin Jones. Coordinated evolution of co-expressed gene clusters in the drosophila transcriptome. *BMC Evol Biol*, 8(1):2, Jan 2008.

[46] Tom Misteli. Spatial positioning; a new dimension in genome function. *Cell*, 119(2):153–156, Oct 2004.

[47] Tom Misteli. Beyond the sequence: cellular organization of genome function. *Cell*, 128(4):787–800, Feb 2007.

[48] Andrea E Murmann, Juntao Gao, Marissa Encinosa, Mathieu Gautier, Marcus E Peter, Roland Eils, Peter Lichter, and Janet D Rowley. Local gene density predicts the spatial position of genetic loci in the interphase nucleus. *Exp Cell Res*, 311(1):14–26, Nov 2005.

[49] Waltraud G Müller, Dietmar Rieder, Tatiana S Karpova, Sam John, Zlatko Trajanoski, and James G McNally. Organization of chromatin and histone modifications at a transcription site. *J Cell Biol*, 177(6):957–967, Jun 2007.

[50] Cameron S Osborne, Lyubomira Chakalova, Karen E Brown, David Carter, Alice Horton, Emmanuel Debrand, Beatriz Goyenechea, Jennifer A Mitchell, Susana Lopes, Wolf Reik, and Peter Fraser. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, 36(10):1065–1071, Oct 2004.

[51] Cameron S Osborne, Lyubomira Chakalova, Jennifer A Mitchell, Alice Horton, Andrew L Wood, Daniel J Bolland, Anne E Corcoran, and Peter Fraser. Myc dynamically and preferentially relocates to a transcription factory occupied by igh. *PLoS Biol*, 5(8):e192, Aug 2007.

[52] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[53] Luis A Parada, Philip G McQueen, Peter J Munson, and Tom Misteli. Conservation of relative chromosome positioning in normal and cancer cells. *Curr Biol*, 12(19):1692–1697, Oct 2002.

[54] Antje Purmann, Joern Toedling, Markus Schueler, Piero Carninci, Hans Lehrach, Yoshihide Hayashizaki, Wolfgang Huber, and Silke Sperling. Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality. *Genomics*, 89(5):580–587, May 2007.

[55] K. L. Reddy, J. M. Zullo, E. Bertolino, and H. Singh. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, 452(7184):243–247, Mar 2008.

[56] T. Ried, A. Baldini, T. C. Rand, and D. C. Ward. Simultaneous visualization of seven different dna probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy. *Proc Natl Acad Sci U S A*, 89(4):1388–1392, Feb 1992.

[57] A-M. Rodriguez, C. Elabd, Ez-Z. Amri, G. Ailhaud, and C. Dani. The human adipose tissue is a source of multipotent stem cells. *Biochimie*, 87(1):125–128, Jan 2005.

[58] Anne-Marie Rodriguez, Christian Elabd, Frédéric Delteil, Julien Astier, Cécile Vernochet, Perla Saint-Marc, Joëlle Guesnet, Anne Guezennec, Ez-Zoubir Amri, Christian Dani, and Gérard Ailhaud. Adipocyte differentiation of multipotent cells established from human adipose tissue. *Biochem Biophys Res Commun*, 315(2):255–263, Mar 2004.

[59] Anne-Marie Rodriguez, Didier Pisani, Claude A Dechesne, Claude Turc-Carel, Jean-Yves Kurzenne, Brigitte Wdziekonski, Albert Villageois, Claude Bagnis, Jean-Philippe Breittmayer, Hervé Groux, Gérard Ailhaud, and Christian Dani. Transplantation of a multipotent cell population from human adipose tissue induces dystrophin expression in the immunocompetent mdx mouse. *J Exp Med*, 201(9):1397–1405, May 2005.

[60] Peter J Roy, Joshua M Stuart, Jim Lund, and Stuart K Kim. Chromosomal clustering of muscle-expressed genes in caenorhabditis elegans. *Nature*, 418(6901):975–979, Aug 2002.

[61] S. Rozen and H. Skaletsky. Primer3 on the www for general users and for biologist programmers. *Methods Mol Biol*, 132:365–386, 2000.

[62] Marcel Scheideler, Christian Elabd, Laure-Emmanuelle Zaragosi, Chiara Chiellini, Sunaina Yadav, Kalina Duszka, Hubert Hackl, Fatima Sanchez-Cabo, Christine Paar, Andreas Prokesch, Reinhard Windhager, Gerard Ailhaud, Christian Dani, Ez-Zoubir Amri, and Zlatko Trajanoski. Comparative transcriptomics of human multipotent stem cells during adipogenesis and osteoblastogenesis. *submitted*, 2008.

[63] Markus O Scheuermann, Jian Tajbakhsh, Anette Kurz, Kaan Saracoglu, Roland Eils, and Peter Lichter. Topology of genes and nontranscribed sequences in human interphase nuclei. *Exp Cell Res*, 301(2):266–279, Dec 2004.

[64] Robert Schneider and Rudolf Grosschedl. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev*, 21(23):3027–3043, Dec 2007.

[65] AFA. Smit, R. Hubley, and R. Green. Repeatmasker open-3.0. *http://www.repeatmasker.org*, 1996-2004.

[66] P. Soille. *Morphological Image Analysis: Principles and Applications*, pages 164–165. Springer, 1999.

[67] I Solovei, J Walter, M Cremer, F Habermann, L Schermelleh, and TCremer. *FISH, a Practical Approach*, chapter FISH on three-dimensionally preserved nuclei, pages 93–118. Oxford University Press, 2002.

[68] Irina Solovei, Antonio Cavallo, Lothar Schermelleh, Françoise Jaunin, Catia Scasselati, Dusan Cmarko, Christoph Cremer, Stanislav Fakan, and Thomas Cremer. Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3d-fish). *Exp Cell Res*, 276(1):10–23, May 2002.

[69] Helmuth Spath. *The Cluster Dissection and Analysis Theory FORTRAN Programs Examples*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1985.

[70] David L Spector. The dynamics of chromosome organization and gene regulation. *Annu Rev Biochem*, 72:573–608, 2003.

[71] M. R. Speicher, S. Gwyn Ballard, and D. C. Ward. Karyotyping human chromosomes by combinatorial multi-fluor fish. *Nat Genet*, 12(4):368–375, Apr 1996.

[72] G. Stocker, M. Fischer, D. Rieder, G. Bindea, J.G. McNally, and Z. Trajanoski. ilap: a novel work flow oriented approach for microscopy data management and protocol development. *in preparation*, 2008.

[73] Alexander Sturn, Bernhard Mlecnik, Roland Pieler, Johannes Rainer, Thomas Truskaller, and Zlatko Trajanoski. Client-server environment for high-performance gene expression data analysis. *Bioinformatics*, 19(6):772–773, Apr 2003.

[74] H. B. Sun, J. Shen, and H. Yokota. Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys J*, 79(1):184–190, Jul 2000.

[75] R Tam, LS Shopland, CV Johnson, JA McNeil, and JB Lawrence. *FISH, a Practical Approach*, chapter Application of RNA FISH for visualizing gene expression and nuclear architecture, pages 93–118. Oxford University Press, 2002.

[76] Martin Thompson, Rebecca A Haeusler, Paul D Good, and David R Engelke. Nucleolar clustering of dispersed trna genes. *Science*, 302(5649):1399–1401, Nov 2003.

[77] Rein van den Boomgaard and Richard van Balen. Methods for fast morphological image transforms using bitmapped binary images. *CVGIP: Graph. Models Image Process.*, 54(3):252–258, 1992.

[78] Roel van Driel, Paul F Fransz, and Pernette J Verschure. The eukaryotic genome: a system regulated at different hierarchical levels. *J Cell Sci*, 116(Pt 20):4067–4075, Oct 2003.

[79] Jan H Vogel, Anja von Heydebreck, Antje Purmann, and Silke Sperling. Chromosomal clustering of a human transcriptome reveals regulatory background. *BMC Bioinformatics*, 6:230, 2005.

[80] J. Walter, B. Joffe, A. Bolzer, H. Albiez, P. A. Benedetti, S. Müller, M. R. Speicher, T. Cremer, M. Cremer, and I. Solovei. Towards many colors in fish on 3d-preserved interphase nuclei. *Cytogenet Genome Res*, 114(3-4):367–378, 2006.

[81] Ruth R E Williams, Véronique Azuara, Pascale Perry, Stephan Sauer, Maria Dvorkina, Helle Jørgensen, Jeffery Roix, Philip McQueen, Tom Misteli, Matthias Merkenschlager, and Amanda G Fisher. Neural induction promotes large-scale chromatin reorganisation of the mash1 locus. *J Cell Sci*, 119(Pt 1):132–140, Jan 2006.

[82] Daniele Zink, Margarida D Amaral, Andreas Englmann, Susanne Lang, Luka A Clarke, Carsten Rudolph, Felix Alt, Kathrin Luther, Carla Braz, Nicolas Sadoni, Joseph Rosenecker, and Dirk Schindelhauer. Transcription-dependent spatial arrangements of cftr and adjacent genes in human cell nuclei. *J Cell Biol*, 166(6):815–825, Sep 2004.

This work was supported by GENAU: BIN, Bioinformatics Integration Network. I would like to express my deepest gratitude to my mentor, Zlatko Trajanoski, for his encouragement, vision, and belief in me.
I thank those who have actively contributed to this work in alphabetical order: Maria Fischer, Hubert Hackl, Ingrid Hauser, Nicole Golob, Claudia Neuhold, Marcel Scheideler, Gernot Stocker.
Most special thanks go to James McNally, Waltraud Müller and colleagues at the "Laboratory of Receptor Biology and Gene Expression" (NCI, Bethesda) for giving a warm welcome and great support during the time that I could spend togther with them.

I am indebted to my family for their unfailing support, and to my wife, Lisi, for her love, encouragement and understanding.

Publications

G. Stocker, D. Rieder, Z. Trajanoski. ClusterControl: a Web interface for distributing and monitoring bioinformatics applications on a Linux cluster. Bioinformatics. 20: 805-807 (2004)

Müller WG, Rieder D, Kreth G, Cremer C, Trajanoski Z, McNally JG. Generic features of tertiary chromatin structure as detected in natural chromosomes. Mol Cell Biol. 24: 9359-9370 (2004)

Kowalska A, Bozsaky E, Ramsauer T, Rieder D, Bindea G, Lorch T, Trajanoski Z, Ambros PF. A new platform linking chromosomal and sequence information. Chromosome Res. 15: 327-339 (2007)

Müller WG, Rieder D, Karpova TS, John S, Trajanoski Z, McNally JG. Organization of chromatin and histone modifications at a transcription site. J Cell Biol. 177: 957-967 (2007)

BIOINFORMATICS APPLICATIONS NOTE Vol. 20 no. 5 2004, pages 805–807
DOI: 10.1093/bioinformatics/bth014

ClusterControl: a web interface for distributing and monitoring bioinformatics applications on a Linux cluster

*Gernot Stocker, Dietmar Rieder and Zlatko Trajanoski**

Institute of Biomedical Engineering and Christian-Doppler-Laboratory for Genomics and Bioinformatics, Graz University of Technology, Krenngasse 37, 8010 Graz, Austria

Received on June 2, 2003; accepted on August 19, 2003
Advance Access publication January 29, 2004

ABSTRACT

Summary: ClusterControl is a web interface to simplify distributing and monitoring bioinformatics applications on Linux cluster systems. We have developed a modular concept that enables integration of command line oriented program into the application framework of ClusterControl. The systems facilitate integration of different applications accessed through one interface and executed on a distributed cluster system. The package is based on freely available technologies like Apache as web server, PHP as server-side scripting language and OpenPBS as queuing system and is available free of charge for academic and non-profit institutions.

Availability: http://genome.tugraz.at/Software/ClusterControl

Contact: zlatko.trajanoski@tugraz.at

INTRODUCTION

With the introduction of high-throughput technologies, e.g. sequencing and microarrays for expression profiling the amount of data that has to be stored, managed and analyzed increased dramatically. Very soon, it became necessary to implement computer-intensive bioinformatics applications for analyzing the flood of data on multi-processor computing systems. As most of these applications are not parallelized, they have to be distributed and monitored using a queuing system. Several web interfaces (Jenuth, 2000; Ferlanti *et al*., 1999; Stoesser *et al*., 2003; Blanchet *et al*., 2000) and more generalized approaches (Letondal, 2001) for universal generation of web interfaces based on textual xml descriptions have been developed previously to enable delivery of calculations on a computing server. However, since these web interfaces were developed for execution on the same server on which the web server is installed, they are not able to use efficiently the power of computing cluster systems.

The first approach to combine distributed calculations using a low-cost PC computing cluster and a web server has been implemented by the BeoBLAST (Grant *et al*., 2001) project. However, this system is limited to the BLAST (Altschul *et al*., 1990) applications and there is currently no freely available tool, which enables integration of other command line oriented programs like HMMer (Durbin *et al*., 1998) or FASTA (Deshpande *et al*., 1995). Therefore, the objective of this work was to develop a platform-independent web interface for distributing and monitoring bioinformatics applications on PC-based cluster systems.

PROGRAM OVERVIEW

The modular framework of this system enables integration of every command line-driven tool (Fig. 1). By adding a web-form (that includes all required parameters of the program which should be integrated) and an additional PHP file (that assembles the appropriate command line), the framework can be extended by any user with moderate programming skills. The delivery to the queuing system and collection of the results are managed by the framework in the background. Hereby, single process oriented applications are supported as well as real parallel applications programmed with special communication libraries like MessagePassingInterface (MPICH). After obtaining the results, it is also possible to run a post-processing tool to improve the visualization of the result files. An example for an integrated application of this framework is the existing NCBI-Blast module that can be tested on http://genome.tugraz.at/Software/ClusterControl/Demo. Currently, the list of supported application modules contains NCBI-Blast, Fasta, WU-Blast and HMMer. Customizations of existing modules can be easily done by using simple but well-defined interface classes and procedures. Unsupported applications can be added immediately and can extend the functionality of ClusterControl by following the short instructions on how to write new modules.

For time-consuming applications, the user can log out from the web interface and can login later to monitor the status of the calculation. If the calculation is still running, an automatically refreshing web page will show the current status

*To whom correspondence should be addressed.

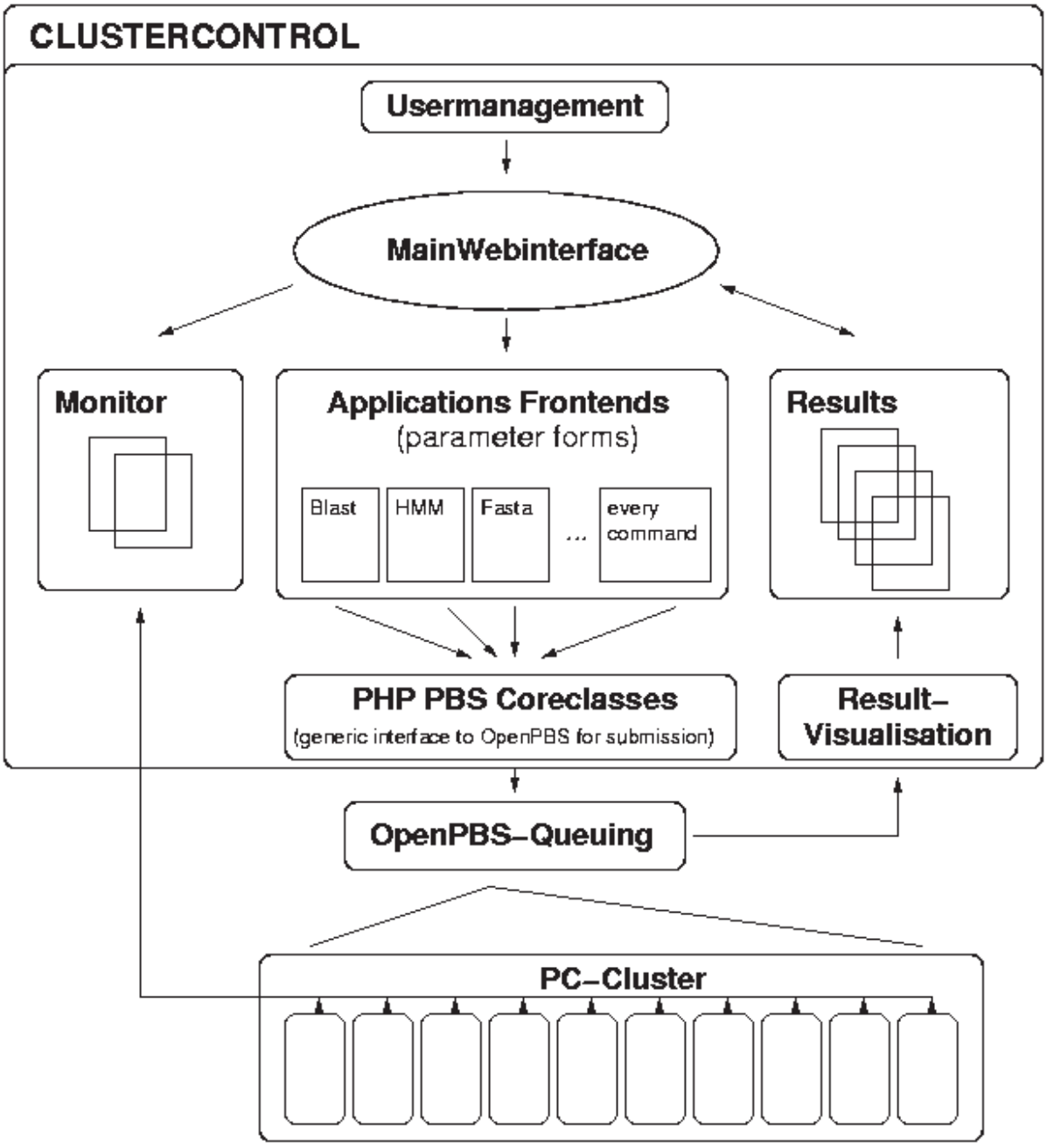


Fig. 1. Schematic overview of ClusterControl. The user can submit jobs through the applications front-ends and monitor the current status of the cluster. After submission, the job will be distributed through the backend PHP classes and the OpenPBS queuing system on the cluster. The job also appears in the Results section and shows its current status. After calculation, a visual improvement of the results can be performed and the final result will be shown in the Results section.

of this submitted job. Otherwise, the processed result will be presented immediately to the user.

An additional feature of this web interface is the ability to monitor all cluster nodes attached to this cluster system. To provide the real-time status and functionality of every calculation node, a lightweight program is installed on every calculation node within the cluster. For the transfer of this status information to the web interface a text protocol was defined, which allows request for information, such as system-, processor-, memory-load, etc.

For the management of the users two different systems are supported: a Lightweight Data Application Protocol (LDAP)-Server or an Unix-like password file that can be used as source of user-accounts. Thus, local cluster user and virtual user

stored in the password file can access the cluster system simultaneously without having direct access to the data, sequences, etc. of each other.

The object-oriented application programming interface (API) of ClusterControl can also be used in PHP scripts for farming out program execution outside the web server context. Similar to the Mollusc-API (Hokamp *et al*., 2003), which is programmed in Perl and wraps up distributed program execution, the backend library of ClusterControl can be used to submit calculation jobs from a PHP script run in a regular Unix shell.

INSTALLATION

For local installation, ClusterControl requires a running installation of the web server Apache (http://www.apache.org) with PHP (Version 4) and LDAP support, that comes with nearly every Linux distribution and a local installation of the queuing system OpenPortableBatchSystem (PBS), which can be downloaded from www.openpbs.org. If an LDAP-Server is already available for user management of the cluster, it can be easily integrated by modifying the configuration file. To activate the monitoring mechanism of ClusterControl, a status collecting server-process must be started on every cluster node. With the adaptation of one well-documented configuration file the installation process is completed. This web application is freely available and can be downloaded from http://genome.tugraz.at/Software/ClusterControl. It is licensed under the Gnu General Public License (GPL; http://www.gnu.org/licenses/gpl.html).

FUTURE DEVELOPMENT

Future development of ClusterControl will provide extensions of the already existing application set, improvements in integration of features of OpenPBS for job management and extension to other queuing systems. Beside already supported MPI implementations, some parallel bioinformatics applications are still using the parallelization software PVM. Therefore, support for this platform is also in preparation. Additional enhancements in the form of new modules by the scientific user community are encouraged and will be integrated in the open source distribution of ClusterControl.

ACKNOWLEDGEMENT

This work was supported by a grant Bioinformatics Integration Network from the Austrian GEN-AU Programme.

REFERENCES

Jenuth,J.P. (2000) The NCBI. Publicly available tools and resources on the Web. *Methods Mol. Biol.*, **132**, 301–312.

Ferlanti,E.S., Ryan,J.F., Makalowska,I. and Baxevanis,A.D. (1999) WebBLAST 2.0: an integrated solution for organizing and analyzing sequence data. *Bioinformatics*, **15**, 422–423.

Stoesser,G., Baker,W., van den Broek,A., Garcia-Pastor,M., Kanz,C., Kulikova,T., Leinonen,R., Lin,Q., Lombard,V., Lopez,R. *et al*. (2003) The EMBL Nucleotide Sequence Database: major new developments. *Nucleic Acids Res.*, **31**, 17–22.

Blanchet,C., Combet,C., Geourjon,C. and Deleage,G. (2000) MPSA: integrated system for multiple protein sequence analysis with client/server capabilities. *Bioinformatics*, **16**, 286–287.

Letondal,C. (2001) A Web interface generator for molecular biology programs in Unix. *Bioinformatics*, **17**, 73–82.

Grant,J.D., Dunbrack,R.L., Manion,F.J. and Ochs.M.F. (2001) BeoBLAST: distributed BLAST and PSI-BLAST on a Beowulf cluster. *Bioinformatics*, **18**, 765–766.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Durbin,R., Eddy,S., Krogh,A. and Mitchison,G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge.

Deshpande,A.S., Richards,D.S. and Pearson,W.R. (1995) Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith–Waterman and FASTA algorithms. *Protein Sci.*, **4**, 1145–1160.

Hokamp,K., Shields,D.C., Wolfe,K.H. and Caffrey,D.R. (2003). Wrapping up BLAST and other applications on Unix clusters. *Bioinformatics*, **19**, 441–442.

MOLECULAR AND CELLULAR BIOLOGY, Nov. 2004, p. 9359–9370 Vol. 24, No. 21
0270-7306/04/$08.00+0 DOI: 10.1128/MCB.24.21.9359–9370.2004

Generic Features of Tertiary Chromatin Structure as Detected in Natural Chromosomes

Waltraud G. Müller,[1]† Dietmar Rieder,[2]† Gregor Kreth,[3] Christoph Cremer,[3] Zlatko Trajanoski,[2] and James G. McNally[1]*

Fluorescence Imaging Group, Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, Bethesda, Maryland[1]; Institute for Genomics and Bioinformatics, Christian Doppler Laboratory for Genomics and Bioinformatics, Graz University of Technology, Graz, Austria[2]; and Kirchhoff Institute for Physics, University of Heidelberg, Heidelberg, Germany[3]

Received 12 May 2004/Returned for modification 21 June 2004/Accepted 13 August 2004

Knowledge of tertiary chromatin structure in mammalian interphase chromosomes is largely derived from artificial tandem arrays. In these model systems, light microscope images reveal fibers or beaded fibers after high-density targeting of transactivators to insertional domains spanning several megabases. These images of fibers have lent support to chromonema fiber models of tertiary structure. To assess the relevance of these studies to natural mammalian chromatin, we identified two different ~400-kb regions on human chromosomes 6 and 22 and then examined light microscope images of interphase tertiary chromatin structure when the regions were transcriptionally active and inactive. When transcriptionally active, these natural chromosomal regions elongated, yielding images characterized by a series of adjacent puncta or "beads", referred to hereafter as beaded images. These elongated structures required transcription for their maintenance. Thus, despite marked differences in the density and the mode of transactivation, the natural and artificial systems showed similarities, suggesting that beaded images are generic features of transcriptionally active tertiary chromatin. We show here, however, that these images do not necessarily favor chromonema fiber models but can also be explained by a radial-loop model or even a simple nucleosome affinity, random-chain model. Thus, light microscope images of tertiary structure cannot distinguish among competing models, although they do impose key constraints: chromatin must be clustered to yield beaded images and then packaged within each cluster to enable decondensation into adjacent clusters.

Chromatin structure can be classified hierarchically (44). Primary structure refers to the organization of DNA on nucleosomes, for example, the 10-nm fiber. Secondary structures arise from interactions between nucleosomes, for example, the 30-nm fiber. Tertiary structures are formed by interactions between secondary structures, and quaternary structures are formed by interactions between tertiary structures. Thus, tertiary and quaternary structures can involve the organized compaction of the chromatin fiber over domains of hundreds of kilobases (16). Here we use the term "tertiary structure " to refer generally to all higher-order chromatin structures beyond the level of the 30-nm fiber.

A knowledge of this tertiary structure and the molecules involved in regulating it will be critical for a complete understanding of the molecular processes involving DNA, including replication, repair, and transcription (26). Only the primary structure of chromatin packaging and, to some extent, the secondary structure are now known with certainty (5). However, these levels of packaging account for only a fraction of the compaction required to fit the cellular content of DNA into a nucleus (15). Thus, there is much to be learned about higher levels of chromatin packaging. A critical first step is to define the higher-order structures in a natural mammalian interphase chromosome. To date, most knowledge has come instead from several model systems that may or may not be representative of natural interphase chromatin.

Extracted metaphase chromosomes reveal a series of DNA loops attached to a protein scaffold (28). These observations have led to a radial-loop, protein scaffold model for metaphase chromosomes that has been extrapolated to interphase chromatin (14, 24, 27). Some meiotic chromosomes, namely, oocyte lampbrush chromosomes, also exhibit DNA loop structures (33). During interphase, the only direct structural evidence for DNA loops comes from the interpretation of electron microscope images of the Balbiani ring genes in polytene chromosome puffs of the fly *Chironomus* (1, 20). At present, it is not certain whether these model systems, namely, mitotic, meiotic, and polytene chromosomes, are representative of normal interphase chromatin and, if so, whether the structures observed after extraction and fixation are genuine.

To address these concerns, tandem array systems have been constructed and then visualized during interphase in living mammalian cells. The tandem arrays are based on green fluorescent protein (GFP)-tagged proteins that bind to a repeated series of target sites that can be transactivated (17, 23, 25, 30, 39, 40, 47). The resultant light microscope images of tertiary chromatin have been interpreted as fibers that elongate in response to transactivation. These interpretations support a chromonema fiber model in which progressive folding of nucleosomal DNA generates a series of thicker fibers: first, secondary structures, corresponding to fibers 30 nm thick; sec-

* Corresponding author. Mailing address: Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, Building 41, Room B516, 41 Library Dr., MSC 5055, Bethesda, MD 20892-5055. Phone: (301) 402-0209. Fax: (301) 496-4951. E-mail: mcnallyj@exchange.nih.gov.

† W.G.M. and D.R. contributed equally to this work.

ond, tertiary structures, corresponding to fibers 60 to 80 nm thick; and third, quaternary structures, corresponding to fibers 100 to 130 nm thick (4). Rather than a scaffold to provide structural support, it is thought that chromonema fibers are held together by fiber-fiber interactions, which could be regulated by different histone variants or tail modifications (16).

Tandem array systems, however, are also subject to questions about their general applicability to natural chromatin structure, since they are artificial. The arrays are composed of very densely packed head-to-tail repeats that recruit strong activating factors to very high densities along at least several megabases of chromatin. In contrast, natural chromosomes are often characterized by large intergenic regions and promoters that are typically much weaker than targeting the potent VP16 activation domain (32) used in *lac* operator arrays (40) or the binding of the glucocorticoid receptor to its cognate promoter, the mouse mammary tumor virus (MMTV) promoter, used in MMTV tandem arrays (23). Therefore, it is not known to what extent these systems are relevant to natural chromatin.

The different approaches to studying tertiary chromatin structure have yielded apparently conflicting models based on loops and scaffolds or on chromonema fibers. At present, it is not clear whether the results from these different systems can be reconciled at all or whether the differences arise because the systems studied are not representative or natural. Our goals in this study therefore were (i) to investigate tertiary chromatin structure in natural human chromosomes during interphase and (ii) to consider which models are consistent with structures observed in natural chromosomes.

The results obtained in the present study show that natural chromosomal domains of ~400-kb yield images in which a single puncta or "bead" decondenses into a series of adjacent beads dependent upon transcription. These features seen in images of tertiary chromatin are likely to be quite common, since they arise in both natural and artificial systems, despite significant functional differences among these various systems. We also show here how these images of tertiary structures can be explained by adapting a loop-scaffold model, a chromonema model, or even a random-chain model. Therefore, the conclusion that tandem array systems favor chromonema models (16, 23, 40), although potentially still correct, is premature.

MATERIALS AND METHODS

Cell culture. The MMTV array cell line (3617) was grown as previously described (23). MRC-5, Raji, and Jurkat cells were obtained from the American Type Culture Collection (Manassas, Va.) and grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 2 mM glutamine or RPMI medium supplemented with 10% FBS. For microscopy experiments, cells were grown on coverslips. These were coated with poly-L-lysine for Raji and Jurkat cells. For interferon induction of MRC-5 cells, cells were incubated with 200 U of gamma interferon/ml in fresh DMEM for 2 or 5 h.

DNA probes. Bacterial artificial chromosome (BAC) probes RP11-36N5, CTA-433F6, and CTA-526G4 for chromosomal region 22q11.21-22a and RP11-10A19 and P1-derived artificial chromosome (PAC) RP1-93N13 for the major histocompatibility complex (MHC) class II (MHCII) locus were obtained from Research Genetics, Huntsville, Ala. End sequences for each probe were obtained by primer extension sequencing with SP6 and T7 primers and then were compared to the entire chromosome 22q sequence or to the MHC sequence by using BLAST.

Plasmid probes 45.1DRα120 and 45.1DRβ008 (38) for the MHC locus were kind gifts from Eric Long (National Institute of Allergy and Infectious Diseases).

DNA FISH. Cells were fixed for 20 to 30 min with 3.5% paraformaldehyde in phosphate-buffered saline (PBS). DNA fluorescence in situ hybridization (FISH) for 3617 cells was performed essentially as described previously (23) but with the following modifications. (i) Dehydration through ethanol was done immediately after denaturation, omitting a 5-min incubation in 50% formamide–2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) on ice. (ii) Hybridization buffer contained 5% rather than 10% dextran sulfate. (iii) The DNA probe was prepared from the entire plasmid pM18, not a restriction fragment. (iv) The biotinylated probe was detected by immunofluorescence with a primary antibiotin antibody and a fluorescent secondary antibody. The DNA FISH procedure for MRC-5 cells was the same, except that digoxigenin was sometimes used to label the probe before detection with appropriate primary and secondary antibodies. The DNA FISH procedure for Raji and Jurkat cells differed from that for 3617 cells as follows. (i) Denaturation of cellular DNA was performed at either 90 or 95°C. (ii) Hybridization buffer contained 50% formamide instead of 25%, 1.4 mg of salmon sperm DNA/ml and 0.6 mg of human Cot-1 DNA/ml instead of tRNA, and 25 μg of DNA probe/ml instead of 5 to 10 μg/ml. (iii) The DNA probe was denatured at 85°C for 10 min and then incubated for 15 to 30 min at 37°C.

For colocalization of known MHC plasmid probes with the BAC and PAC probes for the MHC, the following procedure was used. Cells were fixed with 3.7% formaldehyde in PEM buffer [90 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES), 4.5 mM EGTA, 1.8 mM $MgCl_2$, 0.18% Triton X-100 (pH 6.8)]. The DNA FISH protocol was modified slightly from that used for Raji and Jurkat cells. Cellular DNA was denatured at 88°C in 70% formamide–2× SSC. Nick translation was used to tag the BAC and PAC probes with biotin and the plasmid probes with digoxigenin. For hybridization, the two probes were mixed at concentrations of 20 μg/ml for the BAC and PAC probes and 15 μg/ml for the plasmid probes. The tags were detected with fluorescein-conjugated antidigoxigenin antibody or rhodamine-conjugated avidin.

Bioinformatics. Levels of transcription from the q arm of human chromosome 22 (hereafter referred to as 22q) were extracted from published microarray data in which mRNA levels were compared between 69 different pairs of cell lines at all known and predicted exons on 22q (36). These data were originally used to confirm gene annotation on 22q by defining full-length transcripts based on coregulated exons. We reordered the exon expression data to match the known positioning of exons on 22q and then searched for domains along 22q with high levels of transcription. Crucial for our analysis was the availability of a finished sequence for 22q, enabling the unambiguous determination of gene order, and the complete coverage of 22q exons from microarray data, enabling an accurate measure of transcriptional output.

To identify active domains, we divided 22q into segments ~1 to 2 Mb wide. Segment boundaries were defined by a series of BAC clones that had been mapped by DNA FISH to cytogenetic locations on 22q (18). These BAC clones provided established markers for measuring by DNA FISH the degree of decondensation across a domain. Within each domain, relative transcriptional levels were estimated for each of the 69 pairs of cell lines. Exons which showed at least a twofold difference between cell lines were considered significant. The relative intensity difference for these exons was multiplied by the absolute microarray spot intensity to provide at each exon a measure of the magnitude of the difference in transcription between two cell lines. For each gene, the scores from all exons were averaged and then multiplied by the annotated length of the gene, including introns. This process yielded a measure of transcription proportional to the length of the transcribed region, a factor likely to be important for tertiary chromatin structure (23). Finally, the total score for a domain was the sum of scores for all genes contained in the domain.

Deconvolution microscopy. For high-resolution images, deconvolution microscopy was performed essentially as described previously (23), except that the focal plane spacing was 0.07 μm and the maximum-likelihood algorithm was run for 100 to 200 iterations for DNA FISH and 500 iterations for images of live or fixed cells containing the MMTV array.

Measurements of decondensation. Cells were imaged primarily on a Leica DMRA upright microscope with a Leica ×100/1.3 N.A. oil immersion objective. Images were obtained with a Sensys charge-coupled device camera (Roper, Trenton, N.J.). Length measurements of DNA FISH signals were performed with Metamorph software (Universal Imaging, Downingtown, Pa.). For nearly contiguous BAC or PAC pairs (the green and blue probes in Fig. 1b and 2c), two signals were typically observed in different focal planes of the same nucleus, consistent with the diploid karyotypes of the cells. For measurements, the larger of the two signals was always selected. This image was zoomed with smoothing to avoid pixilation. A curve was drawn to delineate the longest axis of the FISH signal, and then its length was recorded. For the "end-point" BACs on chromosome 22 (blue and red BACs in Fig. 2c), two pairs of signals were typically observed in both Raji and Jurkat cells. In most cells, it was easy to identify isolated signal pairs. In such unambiguous cells, the straight-line distance between one randomly selected pair of signals was measured.

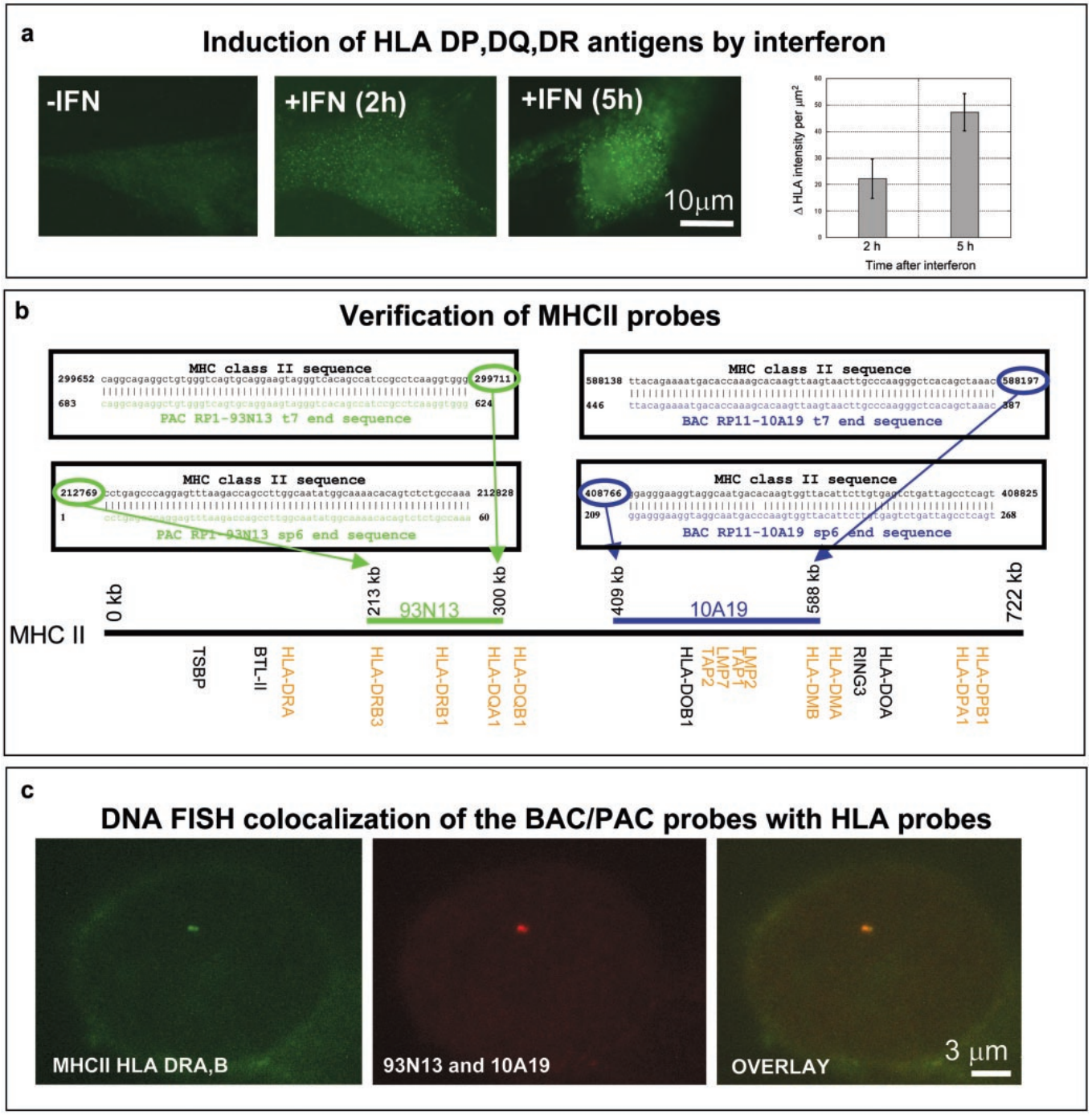


FIG. 1. Induction of MHCII by interferon and characterization of DNA probes for this region. (a) Interferon induction leads to expression of cell surface HLA antigens, as determined by immunofluorescence with antibodies against HLA-DP, HLA-DQ, and HLA-DR antigens. As determined with a *t* test, significant increases in intensity were observed after 2 and 5 h of interferon induction ($P < 0.01$). Standard errors are shown. (b) Two probes, PAC RP1-93N13 and BAC RP11-10A19, were used in DNA FISH to detect the MHCII region. These probes were end sequenced and found to align with the MHCII locus. Genes within the MHCII region are shown below the black line representing the locus; orange letters indicate genes induced by interferon. (c) Confirmation of probe specificity in DNA FISH by colocalization with known MHCII-specific probes.

Transcriptional inhibition. For MRC-5 cells, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) was used at 100 μg/ml for 0.5 h. For Raji and Jurkat cells, DRB was used for 1 h at 100 to 200 μg/ml. For 3617 cells, DRB was used at 100 μg/ml for 1.5 h. Actinomycin D was used at 10 μg/ml for 2 h for MRC-5 cells and 1.5 h for Jurkat cells.

Immunofluorescence. For anti-HLA staining of MRC-5 cells, cells were fixed for 20 min with 3.5% paraformaldehyde in PBS and then washed three times in PBS for a total of 15 min. Cells were incubated overnight at 4°C with an anti-HLA antibody (DakoCytomation, Carpenteria, Calif.) diluted 1:20 in PBS containing 4% bovine serum albumin and 0.1% Tween 20. After being washed in

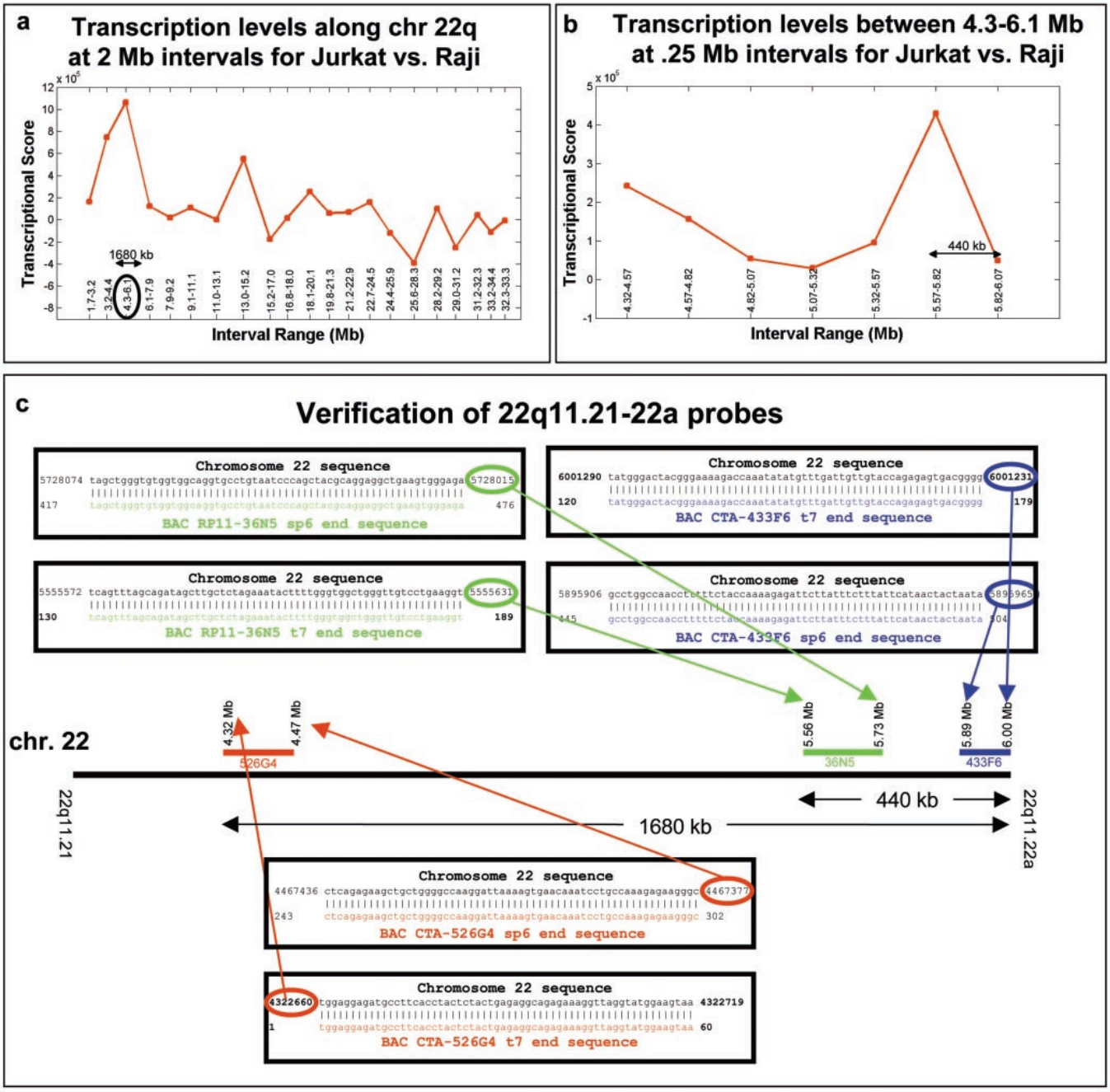


FIG. 2. Transcriptionally active domain on chromosome 22 and characterization of probes for this region. (a) Based on an analysis of microarray expression data (see Materials and Methods), a transcriptional score was calculated on the q arm of human chromosome 22 (chr 22q) for the expression of Jurkat cells relative to Raji cells. A peak was found in the domain from 4.3 to 6.1 Mb (encircled). (b) A finer-scale analysis of transcriptional levels within the 4.3- to 6.1-Mb domain revealed that transcriptional activity there peaked at between 5.3 and 5.8 Mb. (c) To probe tertiary chromatin structure in this domain, three probes were used. Their alignment with the 4.3- to 6.1-Mb domain was determined by end sequencing. Each probe was previously mapped by FISH (18), confirming its specificity. The red and blue BACs mark the end points of the 4.3- to 6.1-Mb domain and were used to measure the separation between these two locations. They span a 1,680-kb domain. The green and blue BACs were used to examine tertiary structure in the most transcriptionally active subdomain of the 4.3- to 6.1-Mb domain. These probes span a 440-kb domain.

PBS, cells were incubated with a Texas red-conjugated anti-mouse secondary antibody for 45 min at room temperature. Cells were washed once more in PBS and examined. Intensity measurements were made by computing the average intensity within a 3-μm-diameter spot positioned at the brightest location in the cell. Background intensity was determined from cells not exposed to interferon, and this intensity measurement was subtracted from the measurements at 2 and 5 h.

Nucleosome affinity, random-chain (NARC) model and simulated microscopy. To simulate a 400-kb chromatin domain, we used a biased, self-avoiding three-dimensional (3D) random walk in which nucleosomes exclusively occupy a certain position in space, whereas segments of linker DNA can occupy the same position. The chain was calculated at a resolution of 11.5 nm.

Each nucleosome in the chain accounted for 146 bp of DNA and occupied a single voxel element (11.5 nm^3). Nucleosomes were connected by linker DNA. Different average linker lengths were evaluated, namely, 50, 68, 85, 155, and 200

bp. This range was chosen in part based on measured linker lengths reported for a variety of organisms (~50 to 130 bp) (41) but also more generally to test the sensitivity of the simulation to this parameter. For each average linker length, 260 random chains were generated. For each chain, linker lengths varied around the average (L_{avg}) over the following ranges: 34 to 66 bp for an L_{avg} of 50 bp, 34 to 102 bp for an L_{avg} of 68 bp, 34 to 136 bp for an L_{avg} of 85 bp, 34 to 272 bp for an L_{avg} of 155 bp, and 34 to 378 bp for an L_{avg} of 200 bp. Linkers were presumed to be decondensed, double-helical DNA at 0.34 nm/bp. Given the resolution of the simulation (11.5 nm^3), linkers therefore occupied from 1 to 12 voxels, depending on their lengths. For linkers longer than one voxel, a random path was generated. Nucleosomes and linkers were added iteratively to the chain, with new nucleosome voxels also being selected at random locations with respect to the preceding voxel. Although overlapping linker DNA was permitted, neither linker DNA nor nucleosomes were allowed to overlap preexisting nucleosomes. When such an overlap occurred, a new random location was selected. This process was iterated until a 2,000-nucleosome chain corresponding to 400 kb was constructed.

To generate various degrees of decondensation of chains with a given average linker length, we allowed for different average distances between nucleosomes in the chain by setting a threshold distance between new elements added to a given chain and their nearest nucleosome. When the randomly selected location for either a new nucleosome or a linker segment exceeded the threshold distance from the nearest nucleosome, a new random location was selected until the distance to the nearest nucleosome was smaller than the threshold distance. By varying this threshold distance from one chain to the next over a range from 16 to 46 nm, we generated an assortment of random chains with different degrees of decondensation. This range of threshold distances was comparable to the range of measured nucleosome spacings in H1-depleted nucleosome chains (~15 to 60 nm) (46).

In some instances, this random-chain algorithm became "trapped, " with no new position available for the next element given the constraints of the algorithm. In practice, trapping occurred for high degrees of condensation, corresponding almost exclusively to single-bead structures in the simulations. To avoid trapping of the chain, we implemented a simple escape mechanism. If no free position was available, then additional linker DNA was added until a free position was found (typically a single additional voxel, corresponding to an additional 34 bp, was sufficient).

The average computation time to produce a 2,000-nucleosome chain was 10 min with an Intel XEON 2.6-GHz processor. Using a 50 CPU Linux cluster, we generated a total of 1,300 different chains at different degrees of decondensation to yield a representative assortment.

To simulate light microscope images from these chains, we assigned a uniform intensity of 255 to each nucleosome (corresponding to 146 bp). Linkers were assigned an intensity of 44 per voxel. When linker DNA overlapped with other linker DNA, the intensity of that voxel was increased accordingly. As another test of the robustness of the simulation, a separate set of 260 chains in which linkers were assigned a lower intensity of 22 per voxel was used, such that the intensities of nucleosomes relative to those of linkers were altered. In all instances, simulated images were generated by convolving the nucleosome chain data with a theoretically generated point spread function for a ×100/1.35 N.A. Olympus objective at a spatial resolution of 11.5 nm^3. XCOSM software was used to generate the point spread function, and the program Image3dm was used to perform the 3D convolution (a kind gift from Jose Conchello, Oklahoma Medical Research Foundation, Oklahoma City). The convolved data at a resolution of 11.5 nm^3 then were downsampled to a resolution of 70 nm^3 by using the program Reduction (a kind gift from Jose Conchello) in which the intensities from 6 by 6 adjacent pixels were averaged and the average was assigned to a single pixel. Finally, these blurred and downsampled images were deconvolved by using the maximum-likelihood algorithm in the Huygen's Essential software package. The point spread function parameters were identical to those used during deconvolution of the real images from the natural systems. The algorithm was run for 10 to 50 iterations with a background estimate set to zero.

RESULTS

Identification of natural systems with extended domains of transcriptional activation. We identified in natural chromosomes two different regions each containing extended domains (~400 kb) of transcriptional activation.

The first natural system was the classical MHCII locus on human chromosome 6, for which interferon induces 13 genes spanning ~700 kb. Earlier studies showed that after interferon induction, the MHCII region is more often external to the chromosome territory (42). However, the tertiary chromatin structure of MHCII and its dependence on transcription have not been investigated. We verified that under our conditions, interferon induced the expression of class II antigens (Fig. 1a), and then we identified a BAC and a PAC probe that together span 375 kb of the MHCII region (Fig. 1b). The specificity of these probes for the MHCII region was confirmed by DNA FISH colocalization with established probes for two loci in this region (Fig. 1c).

In a complementary approach, we used bioinformatics to identify another transcriptionally active domain in human cells. Several previous bioinformatic analyses demonstrated that coregulated genes tend to be clustered along the chromosome (8, 10, 31, 37), but in none of these studies has the corresponding tertiary chromatin structure been examined. To identify such a domain in cells amenable to light microscopy, we analyzed microarray expression data for a number of cultured cell lines (36). The microarray data represented relative expression levels in cell lines at all known and predicted exons on the q arm of human chromosome 22. We reordered the expression data to match the known positioning of exons on 22q and then searched for domains along 22q with high levels of transcription (see Materials and Methods). We identified a subdomain of 440 kb in which expression in one cell line (Jurkat) was markedly higher than that in another cell line (Raji) (Fig. 2b). Then we identified two BAC probes that span much of this domain (Fig. 2c).

Images of tertiary structure in natural chromosomes are beaded. We used the probes for the chromosome 6 and 22 domains to examine tertiary structure in these natural chromosomes by DNA FISH. In the transcriptionally active state, namely, in interferon-treated cells (chromosome 6) or in Jurkat cells (chromosome 22), the relevant probes for these domains typically yielded images characterized by a series of adjacent puncta or "beads" (hereafter referred to as "beaded images") (Fig. 3a and c). As detected by light microscopy, the number of adjacent beads in these images normally ranged from two to four, but occasionally as few as one bead or as many as six beads were observed. Bead diameters varied but typically ranged from 0.4 to 0.8 μm. In most images, beads were closely spaced, with dim intervening strands sometimes visible between them. Occasionally, larger spaces between beads were observed, and in such instances connecting regions were typically not visible. In contrast, before induction with interferon or in the transcriptionally less active Raji cell line, the same probes revealed single beads in most images (Fig. 3b and d).

As a control to test whether DNA FISH preserved tertiary structure at light microscopic resolution, we compared images of the MMTV tandem array after DNA FISH with those for live cells (Fig. 3e and f). Both types of images were similar, suggesting that DNA FISH preserves the basic features of tertiary structure, as detected by the light microscopy methods used.

Transcriptionally active states yield longer structures than inactive states. In tandem array systems, images of transcriptionally active states yield longer structures (23, 39, 40). To test the analog of this scenario in natural systems, we measured the

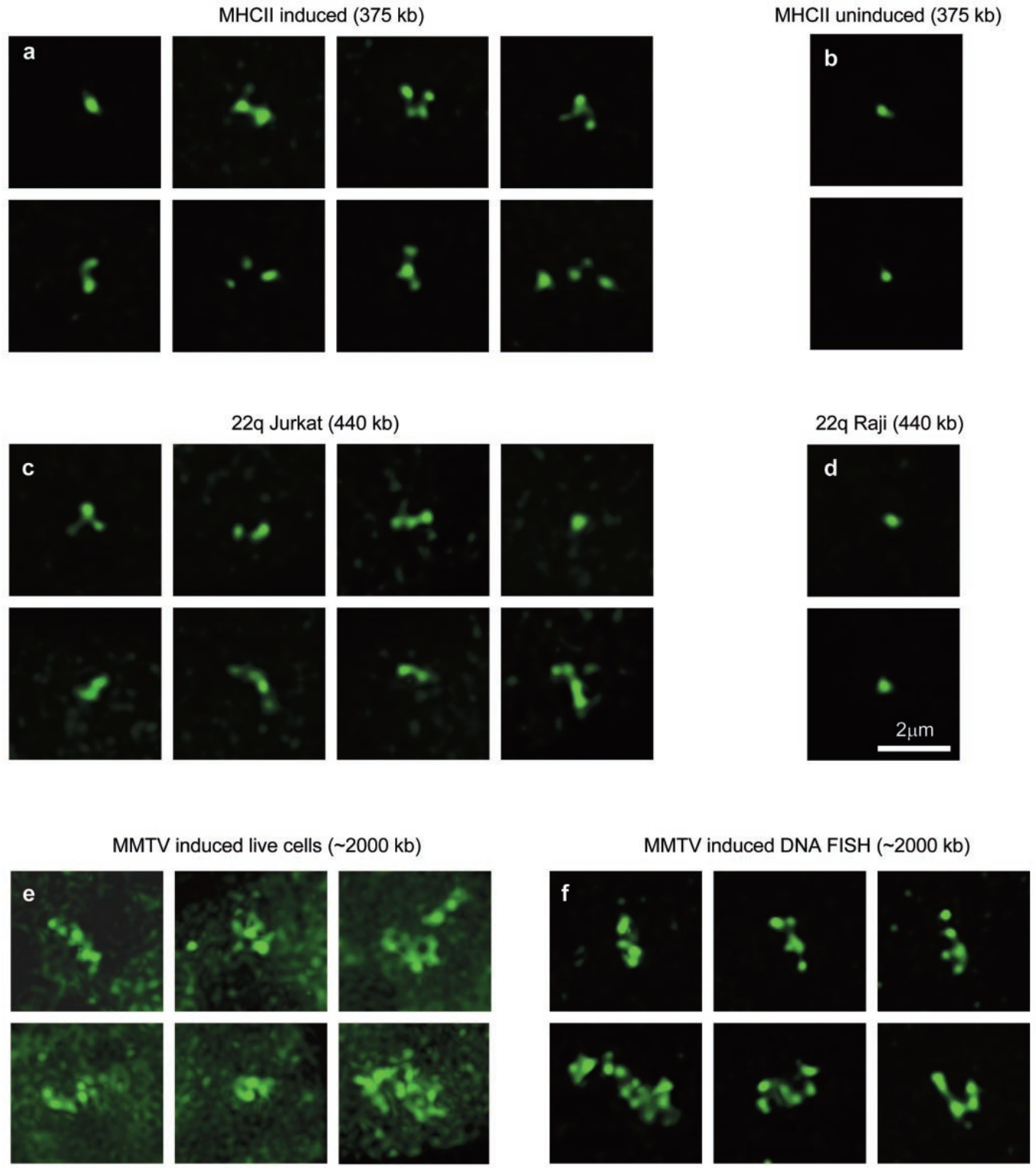


FIG. 3. Beaded images of tertiary structure arise in transcriptionally active regions in natural chromosomes. Projections of 3D deconvolved images of tertiary chromatin structure are shown. (a and c) Images of tertiary structure typically exhibit two to four beads in either the MHCII region on chromosome 6 after induction by interferon (a) or in a transcriptionally active domain on chromosome 22q in Jurkat cells (c). An assortment of decondensed states is shown, reflecting the natural variability in the cell population. Beads are always present; dimmer connecting strands sometimes are detected between them. In most images, the beads are nearly in contact, but wider separations sometimes occur. (b and d) In contrast, when these same natural domains are less transcriptionally active, images exhibit predominantly single beads. Single beads are detected when the MHCII region is examined before induction by interferon (b) or when chromosomal region 22q-11.21-22a is examined in Raji cells (d), which transcribe significantly less from this region than do Jurkat cells. (e and f) As a control for the preservation of structure by DNA FISH, we examined cells harboring the MMTV array after transcription was induced by hormone addition. (e) For live cells, images of the array containing up to ~15 beads are observed with a GFP-tagged glucocorticoid receptor. (f) Similar images are seen for cells examined after DNA FISH with a probe for MMTV DNA. (The higher background in panel e is due to binding of the GFP-tagged glucocorticoid receptor throughout the nucleoplasm, while in panel f a DNA probe specific for the tandem array was used.)

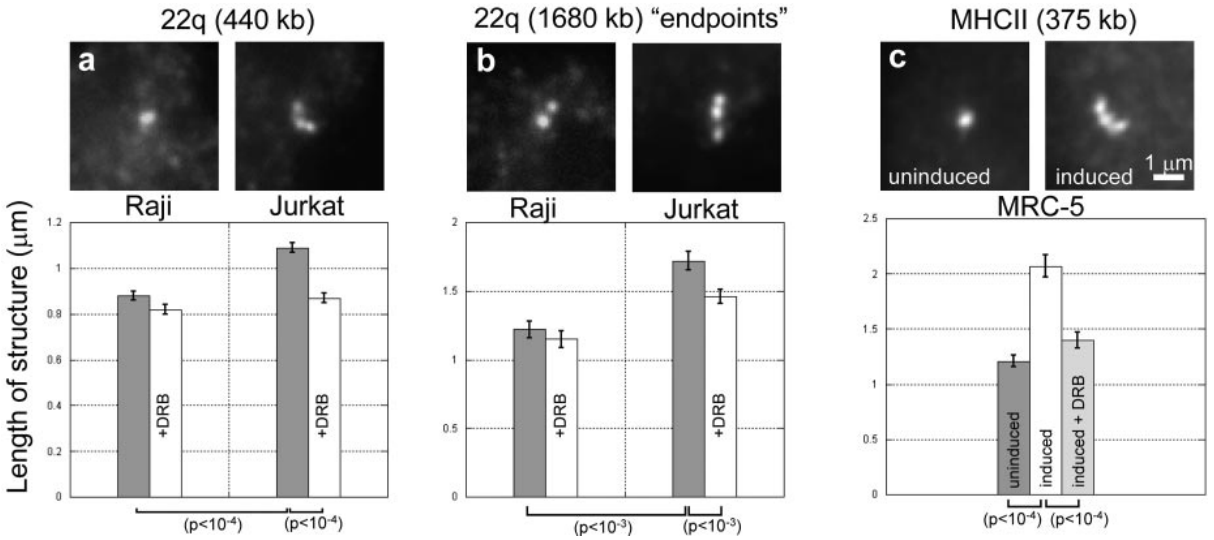


FIG. 4. Transcriptionally active states yield longer structures. Upper panels show examples of nondeconvolved images used for measurements of tertiary chromatin structures for the three different domains examined. Both the 440-kb domain on chromosome 22q and the 375-kb domain for MHCII were spanned by two nearly contiguous BAC or PAC probes. In contrast, the 1,680-kb domain on chromosome 22q was marked by a BAC at each end point, giving rise to images in which two separate structures could be discerned (b). Lower panels (bar charts) show the mean lengths of the structures measured in the different cell types in the presence (white bars) or absence (gray bars) of the transcriptional inhibitor DRB. The means are from 50 to 300 measurements, and each experiment was repeated at least three times with similar results; standard errors are shown. The brackets indicate that significant differences between means were found with a *t* test (*P* values are shown).

total lengths of the structures revealed by DNA FISH. In a comparison of Jurkat and Raji cells, we found that the same 440-kb domain yielded structures that were longer in Jurkat cells than in Raji cells (Fig. 4a).

To test whether this behavior extended to larger domains, we used our bioinformatics analysis to identify a 1,680-kb domain encompassing the 440-kb subdomain. This larger domain was also more transcriptionally active in Jurkat cells than in Raji cells (Fig. 2a). To estimate the extension of this larger domain in the two cell lines, we measured the separation between two BACs that marked the end points of the domain (red and blue BACs in Fig. 2c). This distance was consistently greater in Jurkat cells than in Raji cells (Fig. 4b). Therefore, these data also show that the tertiary chromatin structure of a transcriptionally active domain is longer.

We then investigated whether this was also true for the MHCII region on chromosome 6 in MRC-5 fibroblasts. In this analysis, we used the two probes specific for this region and found that the MHCII region became longer in cells after transcription was induced by interferon (Fig. 4c). We conclude that transcriptional activation also induces changes in the tertiary structure of the MHCII region, resulting in its elongation.

Natural systems require transcription to maintain the elongated state. Since transcriptionally active regions were consistently longer, we examined whether they required transcription for their maintenance. In such analyses, contradictory results have been obtained with tandem array systems: inhibition of transcription had little effect on VP16 tandem arrays (40) but induced the recondensation of MMTV arrays (23). In classical systems, transcription is required for the decondensation of lampbrush chromosomes (22) and for normal puffing of genes in polytene chromosomes (2, 6).

To test the role of transcription in the chromosome 6 and 22 domains, cells were treated with the transcriptional inhibitor DRB (9), and then the lengths of the tertiary structures were measured. Results similar to those described below were also obtained with another transcriptional inhibitor, actinomycin D (data not shown). After transcriptional inhibition of Jurkat cells, the lengths of tertiary structures decreased significantly (Fig. 4a). Similarly, for the larger, 1,680-kb domain in Jurkat cells, transcriptional inhibition caused a decrease in the distance between the BACs marking the end points of this domain (Fig. 4b). Thus, transcription in Jurkat cells is necessary to maintain the extended chromatin conformation. In contrast, transcriptional inhibition of Raji cells yielded no significant reduction in length for either the 440-kb domain or the 1,680-kb domain (Fig. 4a and b), consistent with microarray data indicating that Raji cells showed relatively little transcriptional activity in these domains compared to Jurkat cells. To test the effects of transcriptional inhibition on the transcriptionally active MHCII region in MRC-5 cells, cells were pretreated with interferon and, 1.5 h later, transcription was inhibited for 0.5 h. As in Jurkat cells, this treatment resulted in a contraction of the MHCII region compared to the results obtained with the controls (Fig. 4c).

A NARC model yields beaded images. One goal of the present study was to determine which models are consistent with the principal features of tertiary chromatin in natural systems, namely, images that yield beads which elongate into a series of adjacent beads upon transcriptional activation. In the Discussion, we consider the applicability of chromonema and loop-scaffold models, but here we investigated a simpler possibility, namely, random folding of a chromatin fiber. Random walk models are widely used to describe the folding of polymers (13) and have been specifically applied in several studies to the modeling of secondary and tertiary chromatin structures

(3, 19, 29, 43). A characteristic feature of a random walk is its tendency to extensively explore a given locale and then jump to a new locale, where the exploration process is repeated (see Fig. 1.4 in reference 7). This local exploration process gives rise to local clusters that have the potential to explain the beaded images that we detected.

To explicitly test the applicability of a random walk model, we adapted and extended earlier models (21, 45) that introduced irregularity to account for the secondary structures of 30-nm fibers. These models incorporated randomness by varying the angle between successive nucleosomes. The resultant random chain then was blurred (21) based on the expected resolution of the atomic force microscope, and the predicted images were found to be quite similar to those obtained by atomic force microscopy of actual nucleosomal fibers.

We adopted a very similar approach here. We assume a randomly oriented linker DNA whose average length is 50 bp (linkers used in the earlier studies ranged from 44 to 64 bp). For simplicity, in our model, each nucleosome is 11.5 nm^3 and accounts for 146 bp of DNA, yielding ~200 bp per repeat element in the chain. Sufficient elements are then added to yield a 400-kb chain. Simulated light microscope images of this chain are produced by blurring the chain by using the optical characteristics of a high-numerical-aperture objective lens.

To account for the fact that transcriptional activation should lead to elongation of the chain, we made a simple, general assumption, namely, that the average spacing between nucleosomes in a random chain can increase upon transcriptional activation. This assumption is meant to incorporate in a simple way the many complex and unknown factors that could, in principle, regulate the decompaction of a nucleosome chain. In the model, this average spacing rule is enforced by setting a threshold limit for the maximal spacing between nucleosomes in a particular chain. If the randomly chosen location for a new nucleosome or linker is at a distance greater than the threshold, then that position is disallowed, and a new random location is sought and accepted only when its distance to the next closest nucleosome is less than the threshold. Our range of threshold values (16 to 46 nm) mirrors those measured by atomic force microscopy for H1-depleted nucleosome chains (~15 to 60 nm) (46). We refer to this feature of the model as nucleosome affinity, hence, the NARC model.

Random chains produced by this NARC model indeed produced the clustering characteristic of random walks (Fig. 5a and d). Simulated light microscope images of these clusters yielded a series of adjacent fluorescent spots (Fig. 5b, c, e, and f). By reducing the magnification of these simulated images to that used for real images, we found considerable similarity: both real and simulated images revealed beads with occasional dimmer connecting segments (compare Fig. 5g and Fig. 3). Very similar images were obtained when we changed some of the underlying assumptions in the model, such as the mean length of linker DNA (lengths of 68, 85, 155, and 200 bp were tested) or the relative fluorescence intensities assigned to nucleosomes versus linkers (data not shown). These results suggest that the model is robust and that its success is largely due to the tendency of random walks to cluster.

As a further test of the applicability of the NARC model, we performed a simple analysis to examine whether the structures it predicted could easily extend into longer structures. In both the natural and the MMTV tandem array systems, transcriptional activation yielded multiple beads from what were originally single beads in the transcriptionally inactive state (Fig. 3). For such a transition to occur, the underlying nucleosome chain would need to have been reasonably ordered to reduce entanglements that would arise upon separation into a series of adjacent local clusters of the chain.

To investigate this requirement, we used colors to mark positions along the nucleosome chain. We colored the first third of the chain blue, the second third green, and the last third red. Using this coloring scheme, we examined a number of random chains and found that colored domains remained largely distinct, with little overlap between them either for longer chains (256 kb) or for shorter segments (64 kb) within them (Fig. 5h). These results imply that on a macroscopic scale, the chain remains largely ordered and therefore could, in principle, split into separate domains without major entanglements.

DISCUSSION

Light microscopy images of tertiary chromatin structure are beaded. We identified transcriptionally active domains in natural human chromosomes and then examined their tertiary structure by light microscopy. The characteristic features of these images were a series of puncta which we have called beads. When transcriptionally inactive, a 400-kb region typically appeared as a single bead. When transcriptionally active, an image of the same chromosomal region typically exhibited two to six adjacent beads. Sometimes in these images we detected larger intervening regions between beads, and occasionally a strand connecting the beads could be discerned.

Beaded features can be recognized in images of several tandem array systems proposed previously as models for tertiary chromatin. Both the 2-Mb MMTV tandem array (23) and the 2-Mb *tet*-regulated *lac* operator arrays (see Fig. 6 in reference 39) exhibit such features. Since these are all live-cell systems, the beaded features cannot be artifacts of fixation or hybridization. To some extent, beads are also discernible in images of the first tandem array examined in live cells, the 90-Mb *lac* operator array activated by VP16 targeting. However, the latter array exhibited a much more pronounced ribbon-like fiber structure (see Fig. 2 in reference 40). The impact of this first study led to the prevailing view that chromonema fiber models best explain the images of tertiary chromatin seen in tandem arrays (16, 23, 40); therefore, despite the presence of beaded features in the other tandem array systems, they have not been emphasized.

Our examination of natural systems here shifts the emphasis from fibers to beads, which are in fact the common feature in images from all of these systems and the predominant feature in natural systems. This distinction is important. Describing these images of tertiary chromatin as beaded rather than as fibers not only is a more accurate description but also is a characterization that by its name favors no particular class of model. As we discuss below, beaded images can indeed be explained by a chromonema fiber model but also by a loop-scaffold model or by a self-avoiding random walk model; therefore, higher-resolution studies will be necessary to distinguish among these possibilities.

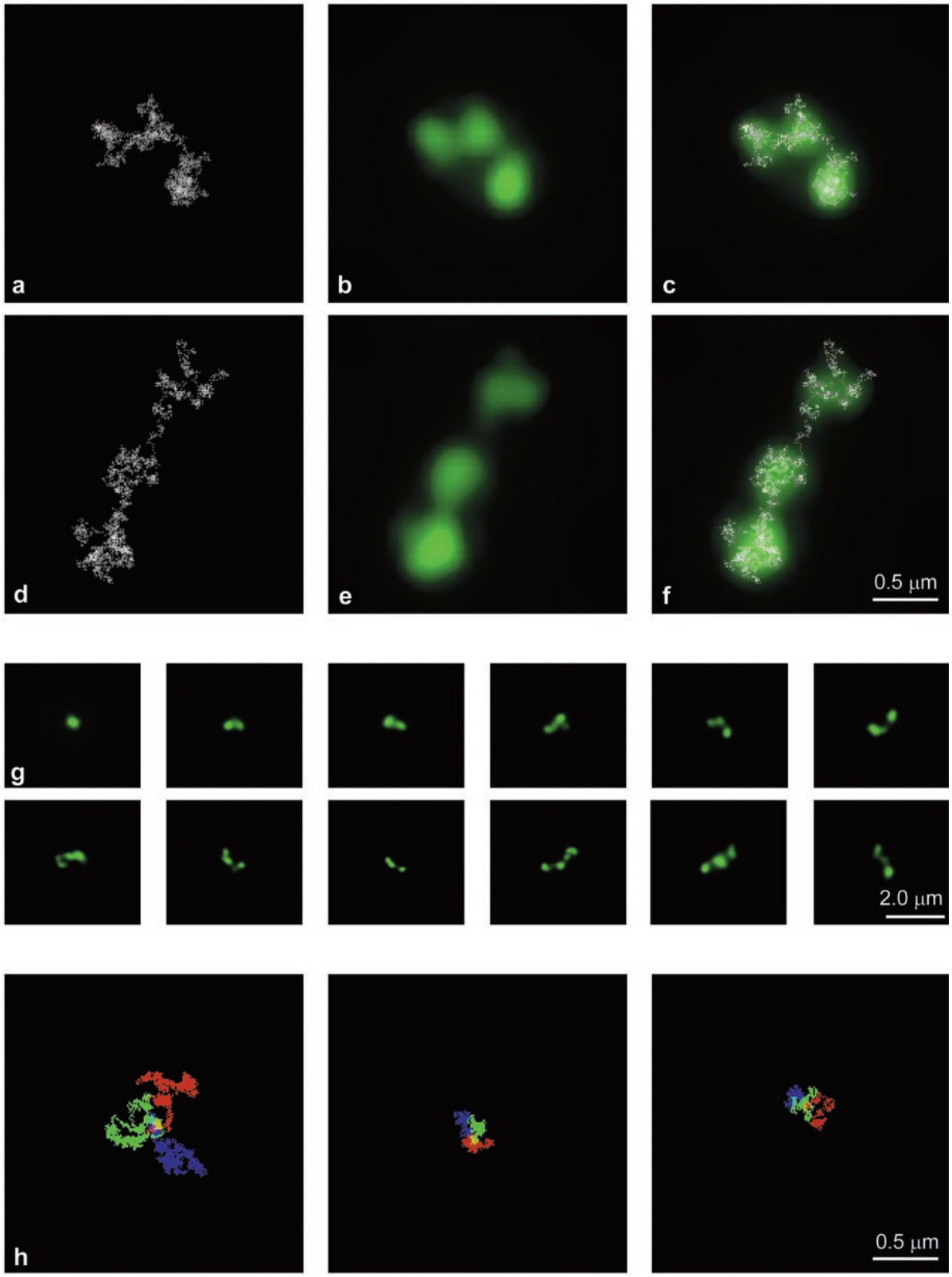


FIG. 5. A NARC model can explain beaded images. (a and d) Examples of 400-kb simulated nucleosome chains. (b and e) High-magnification views of the simulated images reveal the presence of fluorescence puncta; most details of the nucleosome chain are lost. (c and f) Overlays show that less densely packed regions of the chain may extend beyond the puncta and may be invisible. (g) Sampling of simulated projected images following deconvolution for 400-kb chains with different degrees of decondensation. Note that these images exhibit beads resembling those in the actual images (compare to Fig. 3). (h) When the first third of the chain is colored blue, the middle third is colored green, and the last third is colored red, distinct largely nonoverlapping regions are obtained either for long chains of 256 kb (left panel) or for shorter chains of 64 kb (middle and right panels). These data indicate that, on a macroscopic scale, neighbor-neighbor relations are preserved, and so the chain could unravel without significant entanglement.

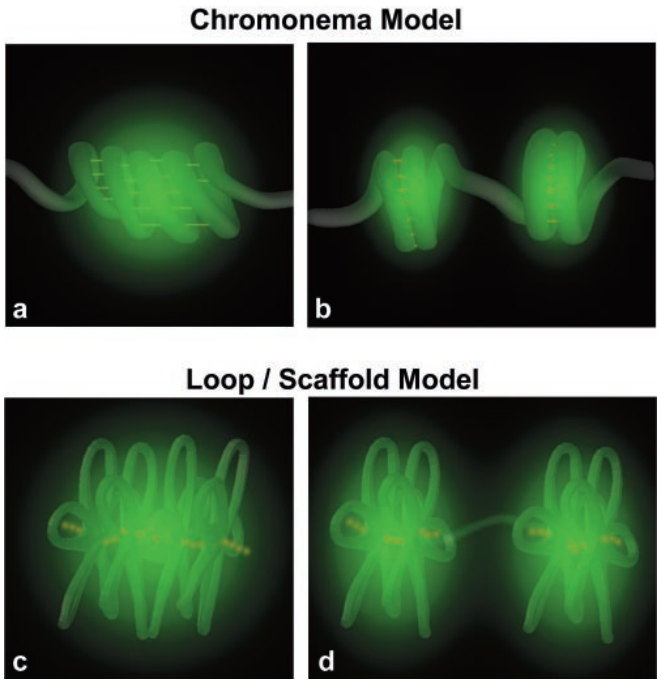


FIG. 6. Chromonema models and radial-loop, protein scaffold models can be adapted to account for beaded images. (a) Chromonema models propose a hierarchically folded chromatin fiber that yields a series of different fiber thicknesses (4, 35, 40). Here, a thinner fiber is helically wound to yield a thicker fiber that would appear as a bead (green haze) when viewed in the light microscope. Yellow struts represent fiber-fiber interactions, for example, between chromatin cross-linking proteins, that could stabilize the folded structure (16). The diameter of the bead detected by the light microscope would depend on the thickness of the underlying fiber and the size of the hollow core. If the underlying fiber were 30 nm in diameter, then the hollow helical core would need to be ~500 nm in diameter to match the measured bead diameter of 0.4 to 0.8 μm. However, if the underlying fiber were thicker, for example, 100 nm (4), then the hollow core would need to be only twice as large as the fiber diameter, namely, ~200 nm. (b) Local disruption of fiber-fiber interactions would lead to unraveling of the fibers and would yield a light microscope image of two adjacent beads. (c) Radial-loop models (14, 27) propose that the chromatin fiber forms loops attached to a protein scaffold (yellow spheres). A cluster of loops would be imaged as a bead by light microscopy. If a typical loop were 80 kb, then the radius of the bead would correspond to 40 kb. Assuming that loops are formed from 30-nm fibers which are ~40-fold compacted, the bead radius would be 40 kb × (1/40) × (0.34 nm/bp), or 0.34 μm, in reasonable agreement with the measured bead diameter of 0.4 to 0.8 μm. (d) The structure shown in panel c could decondense into two beads by local detachment of a loop from the scaffold, but in order to elongate into two beads, the scaffold would have to either fragment (as shown here) or stretch.

Transcription induces longer structures that require continued transcription for their maintenance. In both of the natural systems examined here, tertiary structures were longer after transcriptional activation. This finding demonstrates that elongation is a characteristic feature of transcriptionally active tertiary structures, as light microscope images of all of the artificial tandem arrays examined showed longer structures upon activation (17, 23, 39, 40).

The ability to elongate imposes constraints on the underlying folding of chromatin within tertiary structures. First, the fiber must be reasonably organized so that it can be easily unraveled to produce more extended structures. Second, if the fiber is attached to a structural scaffold of some form, then elongation requires that the scaffold either stretch or fragment. Explicit examples are discussed below and are shown in Fig. 6.

In order to maintain their elongated state, the two natural systems examined here required transcription. This is likely to be the norm, since it is also seen in the MMTV tandem array that uses a natural promoter (23) as well as with most genes examined in puffs or lampbrush chromosomes (2, 6). The one clear exception to this rule are *lac* tandem arrays, which do not require transcription for maintaining the decondensed state (40). These *lac* arrays do not use a natural promoter but rather target the potent VP16 activation domain to very high densities. This design may override the normal transcriptional control of the elongated state and therefore produce an unnatural result. Alternatively, it is possible that some natural systems not yet examined will also reveal no dependence of the elongated state on transcription.

Models of tertiary structure. To understand tertiary structure, one must ultimately determine the patterns of chromatin fiber folding. Visualization of folding patterns on the scale of a 30-nm fiber clearly requires a resolution higher than that afforded by conventional light microscopy. However, our light microscope images do provide an envelope into which higher-order folding patterns must fit.

The envelope described here is a bead. A beaded image will arise in the light microscope when chromatin is more densely packed in a set of relatively discrete clusters. Packing within a cluster need not be space filling, simply more dense than in surrounding regions, where fluorescence is found by light microscopy to be much dimmer or absent. The ability of clusters to split, yielding images with adjacent clusters, places an additional constraint on the patterns of chromatin folding within each cluster. We now consider below how these constraints can be realized with three explicit examples.

We found that a randomly folded fiber calculated according to the NARC model could explain our observations. Here, we were inspired by models previously developed to explain images of secondary chromatin structure, namely, atomic force microscopy images of 30-nm fibers. These images revealed irregular packing of nucleosomes into fibers whose diameters varied but were, on average, ~30 nm. These irregular secondary chromatin structures were successfully modeled by allowing for variable linker lengths, which in turn produced random variations in the angle between successive nucleosomes in the chain. Interestingly, we found that adaptation and extension of this concept of random angles in the form of our NARC model could also explain the tertiary structures that we observed by light microscopy in natural chromosomes. In addition, this NARC model could account for the ability of beads to easily separate into a series of adjacent beads. We found that within a bead, fiber folding was sufficiently organized so that local domains were preserved. This property would enable separation into two beads without major entanglements of the underlying fiber.

Although a simple NARC model can account for our observations, other models with more organized folding of the chromatin fiber can also be entertained. Indeed, the congruence between the predicted random-chain and real images does not ensure that the underlying chromatin structure is random. Diffusion, for example, can be modeled as a random walk, despite

the fact that individual particles undergo complex trajectories precisely defined by Newton's laws. Similarly, we cannot rule out a more complex organization of chromatin with a precise and varied sequence of bending angles. The average properties of such complex folding could, by analogy with the process of diffusion, yield macroscopic behavior at the resolution of conventional light microscopy that can be captured by a random-chain model.

While a macroscopic average over an assortment of complex, organized folding patterns could give rise to the images that we have observed, it is also possible to imagine more regular, repeated patterns of folding that could likewise explain our data. For example, chromonema fiber models can be adapted by assuming that beads correspond to locally thicker segments of the chromonema arising from higher-order folding of an underlying, thinner fiber (Fig. 6a). Although for the purposes of illustration we show here helical folding of the thinner fiber (35), any reasonably ordered fiber packing within the bead envelope would satisfy the requirement that major entanglements should be avoided upon separation into a series of adjacent beads. Chromonema models posit that tertiary structures are maintained by some form of interaction between adjacent fibers (16), shown in Fig. 6a and b as yellow cross-links. Breaking these cross-links at a specific location would lead to local unraveling of the fiber, yielding an image with two adjacent beads (Fig. 6b).

A radial loop-protein scaffold model could also be adapted to explain a beaded image, if a bead arises from a cluster of chromatin loops (Fig. 6c). Separation into a pair of adjacent beads would proceed by unraveling of one or several adjacent loops, generating two adjacent looped clusters (Fig. 6d). The radial looping of fibers in this model naturally preserves the sequence along the axis of the chromosome; therefore, this model easily accounts for the splitting of chromatin clusters without major entanglements. However, in this model, the observed elongation yielding a pair of adjacent beads requires that the protein scaffold either stretch or fragment at the site where the loops unravel. Therefore, in this model, elongation does not occur as simply as in a chromonema model, where unraveling requires only the loss of local fiber-fiber interactions.

In summary, a variety of models can be adapted to explain the characteristic features of tertiary structures seen by light microscopy. The chief requirement for any model is that chromatin be clustered into denser subdomains that yield the beaded images seen by light microscopy. Distinguishing among different models of bead substructure will require higher-resolution imaging of chromatin fiber folding to determine whether it is random, hierarchically folded, or looped. These questions may soon be addressed with new fluorescence microscopy techniques that provide improved resolution, such as 4Pi microscopy (34), spatially modulated illumination microscopy (12) and spectral precision distance microscopy (11).

Generic features of tertiary structures. Overall, our images of tertiary chromatin structures and the dependence of these structures on transcription largely support what has been found for artificial tandem arrays. This situation is remarkable, given the differences between tandem arrays and natural systems. The 440-kb chromosome 22 domain contains 6 genes and 60% coding sequence, while the 375-kb chromosome 6 domain contains 10 genes and 20% coding sequence. In contrast, some tandem arrays contain close to 100% coding sequence and/or recruit an abnormally high density of activators over large domains of at least 2 Mb. In addition, the methods of activation vary drastically across these systems, ranging from natural transcription factors recruited to their target promoters to chimeric transactivating factors recruited to viral promoters or simply tethered to chromatin in the absence of any promoter. The fact that these very disparate systems form similar light microscopy images suggests generic features of the underlying tertiary structures of activated chromatin, irrespective of the specific underlying sequences and functionality.

ACKNOWLEDGMENTS

We are indebted to both the National Cancer Institute and GEN-AU project BIN, Austrian Ministry for Education, Science, and Culture, for financial support.

We are grateful to the Graz University of Technology for computational support for the modeling studies. We thank Gordon Hager for 3617 cells, Eric Long for MHCII plasmid clones, and Jose Conchello for software used in the simulated microscopy. We are grateful to David Clark, Mike Kruhlak, Tatiana Karpova, Richard Leapman, and Andre Nussenzweig for comments on the manuscript. Microscopy was performed at the National Cancer Institute Fluorescence Imaging Facility, and we gratefully acknowledge Tatiana Karpova for expert assistance.

REFERENCES

1. **Andersson, K., B. Bjorkroth, and B. Daneholt.** 1980. The *in situ* structure of the active 75S RNA genes in Balbiani rings of *Chironomus tentans*. Exp. Cell Res. **130:**313–326.
2. **Andersson, K., B. Bjorkroth, and B. Daneholt.** 1984. Packing of a specific gene into higher order structures following repression of RNA synthesis. J. Cell Biol. **98:**1296–1303.
3. **Beard, D. A., and T. Schlick.** 2001. Computational modeling predicts the structure and dynamics of chromatin fiber. Structure (Cambridge) **9:**105–114.
4. **Belmont, A. S., and K. Bruce.** 1994. Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. J. Cell Biol. **127:**287–302.
5. **Belmont, A. S., S. Dietzel, A. C. Nye, Y. G. Strukov, and T. Tumbar.** 1999. Large-scale chromatin structure and function. Curr. Opin. Cell Biol. **11:**307–311.
6. **Berendes, H. D.** 1968. Factors involved in the expression of gene activity in polytene chromosomes. Chromosoma **24:**418–437.
7. **Berg, H. C.** 1983. Random walks in biology. Princeton University Press, Princeton, N.J.
8. **Caron, H., B. van Schaik, M. van der Mee, F. Baas, G. Riggins, P. van Sluis, M. C. Hermus, R. van Asperen, K. Boon, P. A. Voute, S. Heisterkamp, A. van Kampen, and R. Versteeg.** 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science **291:**1289–1292.
9. **Chodosh, L. A., A. Fire, M. Samuels, and P. A. Sharp.** 1989. 5,6-Dichloro-1-beta[scap]-d-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II *in vitro*. J. Biol. Chem. **264:**2250–2257.
10. **Cohen, B. A., R. D. Mitra, J. D. Hughes, and G. M. Church.** 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat. Genet. **26:**183–186.
11. **Esa, A., P. Edelmann, G. Kreth, L. Trakhtenbrot, N. Amariglio, G. Rechavi, M. Hausmann, and C. Cremer.** 2000. Three-dimensional spectral precision distance microscopy of chromatin nanostructures after triple-colour DNA labelling: a study of the BCR region on chromosome 22 and the Philadelphia chromosome. J. Microsc. **199:**96–105.
12. **Failla, A. V., U. Spoeri, B. Albrecht, A. Kroll, and C. Cremer.** 2002. Nanosizing of fluorescent objects by spatially modulated illumination microscopy. Appl. Opt. **41:**7275–7283.
13. **Flory, P. J.** 1953. Principles of polymer chemistry. Cornell University Press, Ithaca, N.Y.
14. **Gasser, S. M., and U. Laemmli.** 1987. A glimpse at chromosomal order. Trends Genet. **3:**16–22.
15. **Holmes, V. F., and N. R. Cozzarelli.** 2000. Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. Proc. Natl. Acad. Sci. USA **97:**1322–1324.
16. **Horn, P. J., and C. L. Peterson.** 2002. Chromatin higher order folding—wrapping up transcription. Science **297:**1824–1827.
17. **Janicki, S. M., T. Tsukamoto, S. E. Salghetti, W. P. Tansey, R. Sachidan-**

andam, K. V. Prasanth, T. Ried, Y. Shav-Tal, E. Bertrand, R. H. Singer, and D. L. Spector. 2004. From silencing to gene expression: real-time analysis in single cells. Cell **116:**683–698.

18. **Kirsch, I. R., E. D. Green, R. Yonescu, R. Strausberg, N. Carter, D. Bentley, M. A. Leversha, I. Dunham, V. V. Braden, E. Hilgenfeld, G. Schuler, A. E. Lash, G. L. Shen, M. Martelli, W. M. Kuehl, R. D. Klausner, and T. Ried.** 2000. A systematic, high-resolution linkage of the cytogenetic and physical maps of the human genome. Nat. Genet. **24:**339–340.
19. **Kreth, G., C. Munkel, J. Langowski, T. Cremer, and C. Cremer.** 1998. Chromatin structure and chromosome aberrations: modeling of damage induced by isotropic and localized irradiation. Mutat. Res. **404:**77–88.
20. **Lamb, M. M., and B. Daneholt.** 1979. Characterization of active transcription units in Balbiani rings of *Chironomus tentans*. Cell **17:**835–848.
21. **Leuba, S. H., G. Yang, C. Robert, B. Samori, K. van Holde, J. Zlatanova, and C. Bustamante.** 1994. Three-dimensional structure of extended chromatin fibers as revealed by tapping-mode scanning force microscopy. Proc. Natl. Acad. Sci. USA **91:**11621–11625.
22. **Mancino, G., I. Nardi, N. Corvaja, L. Fiume, and V. Marinozzi.** 1971. Effects of alpha-amanitin on Triturus lampbrush chromosomes. Exp. Cell Res. **64:** 237–239.
23. **Müller, W. G., D. Walker, G. L. Hager, and J. G. McNally.** 2001. Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. J. Cell Biol. **154:**33–48.
24. **Munkel, C., R. Eils, S. Dietzel, D. Zink, C. W. G. Mehring, T. Cremer, and J. Langowski.** 1999. Compartmentalization of interphase chromosomes observed in simulation and experiment. J. Mol. Biol. **285:**1053–1065.
25. **Nye, A. C., R. R. Rajendran, D. L. Stenoien, M. A. Mancini, B. S. Katzenellenbogen, and A. S. Belmont.** 2002. Alteration of large-scale chromatin structure by estrogen receptor. Mol. Cell. Biol. **22:**3437–3449.
26. **O'Brien, T. P., C. J. Bult, C. Cremer, M. Grunze, B. B. Knowles, J. Langowski, J. McNally, T. Pederson, J. C. Politz, A. Pombo, G. Schmahl, J. P. Spatz, and R. van Driel.** 2003. Genome function and nuclear architecture: from gene expression to nanoscience. Genome Res. **13:**1029–1041.
27. **Paul, A. L., and R. J. Ferl.** 1999. Higher-order chromatin structure: looping long molecules. Plant Mol. Biol. **41:**713–720.
28. **Paulson, J. R., and U. Laemmli.** 1977. The structure of histone depleted metaphase chromosomes. Cell **12:**817–828.
29. **Ponomarev, A. L., and R. K. Sachs.** 1999. Polymer chromosome models and Monte Carlo simulations of radiation breaking DNA. Bioinformatics **15:** 957–964.
30. **Robinett, C. C., A. Straight, G. Li, C. Willhelm, G. Sudlow, A. Murray, and A. S. Belmont.** 1996. *In vivo* localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J. Cell Biol. **135:**1685–1700.
31. **Roy, P. J., J. M. Stuart, J. Lund, and S. K. Kim.** 2002. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. Nature **418:** 975–979.
32. **Sadowski, I., J. Ma, S. Triezenberg, and M. Ptashne.** 1988. GAL4-VP16 is an unusually potent transcriptional activator. Nature **335:**563–564.
33. **Scheer, U., H. Spring, and M. F. Trendelenburg.** 1979. organizaton of transcriptionally active chromatin in lampbrush chromosome loops, p. 3–47. *In* H. Busch (ed.), The cell nucleus, vol. 7. Academic Press, Inc., New York, N.Y.
34. **Schrader, M., K. Bahlmann, G. Giese, and S. W. Hell.** 1998. 4Pi-confocal imaging in fixed biological specimens. Biophys. J. **75:**1659–1668.
35. **Sedat, J., and L. Manuelidis.** 1978. A direct approach to the structure of eukaryotic chromosomes. Cold Spring Harbor Symp. Quant. Biol. **42:**331–350.
36. **Shoemaker, D. D., E. E. Schadt, C. D. Armour, Y. D. He, P. Garrett-Engele, P. D. McDonagh, P. M. Loerch, A. Leonardson, P. Y. Lum, G. Cavet, L. F. Wu, S. J. Altschuler, S. Edwards, J. King, J. S. Tsang, G. Schimmack, J. M. Schelter, J. Koch, M. Ziman, M. J. Marton, B. Li, P. Cundiff, T. Ward, J. Castle, M. Krolewski, M. R. Meyer, M. Mao, J. Burchard, M. J. Kidd, H. Dai, J. W. Phillips, P. S. Linsley, R. Stoughton, S. Scherer, and M. S. Boguski.** 2001. Experimental annotation of the human genome using microarray technology. Nature **409:**922–927.
37. **Spellman, P. T., and G. M. Rubin.** 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. J. Biol. **1:**5.
38. **Tonnelle, C., R. DeMars, and E. O. Long.** 1985. DOβ: a new β chain gene in HLA-D with a distinct regulation of expression. EMBO J. **4:**2839–2847.
39. **Tsukamoto, T., N. Hashiguchi, S. M. Janicki, T. Tumbar, A. S. Belmont, and D. L. Spector.** 2000. Visualization of gene activity in living cells. Nat. Cell Biol. **2:**871–878.
40. **Tumbar, T., G. Sudlow, and A. S. Belmont.** 1999. Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. J. Cell Biol. **145:**1341–1354.
41. **Van Holde, K. E.** 1989. Chromatin. Springer-Verlag, New York, N.Y.
42. **Volpi, E. V., E. Chevret, T. Jones, R. Vatcheva, J. Williamson, S. Beck, R. D. Campbell, M. Goldsworthy, S. H. Powis, J. Ragoussis, J. Trowsdale, and D. Sheer.** 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. J. Cell Sci. **113:**1565–1576.
43. **Wedemann, G., and J. Langowski.** 2002. Computer simulation of the 30-nanometer chromatin fiber. Biophys. J. **82:**2847–2859.
44. **Woodcock, C. L., and S. Dimitrov.** 2001. Higher-order structure of chromatin and chromosomes. Curr. Opin. Genet. Dev. **11:**130–135.
45. **Woodcock, C. L., S. A. Grigoryev, R. A. Horowitz, and N. Whitaker.** 1993. A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. Proc. Natl. Acad. Sci. USA **90:**9021–9025.
46. **Yang, G., S. H. Leuba, C. Bustamante, J. Zlatanova, and K. van Holde.** 1994. Role of linker histones in extended chromatin fibre structure. Nat. Struct. Biol. **1:**761–763.
47. **Ye, Q., Y. F. Hu, H. Zhong, A. C. Nye, A. S. Belmont, and R. Li.** 2001. BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations. J. Cell Biol. **155:**911–921.

Chromosome Research (2007) 15:327–339
DOI: 10.1007/s10577-007-1129-y

A new platform linking chromosomal and sequence information

Agata Kowalska[1], Eva Bozsaky[1], Thomas Ramsauer[2], Dietmar Rieder[2], Gabriela Bindea[2], Thomas Lörch[3], Zlatko Trajanoski[2] & Peter F. Ambros[1]*
[1]*CCRI, Children's Cancer Research Institute, St. Anna Kinderkrebsforschung, 1090 Vienna, Austria; Tel: +43 1 40470 4110; Fax: +43 1 40470 7150; E-mail: peter.ambros@ccri.at;* [2]*Institute for Genomics and Bioinformatics, Graz University of Technology, Graz, Austria;* [3]*MetaSystems, Altlussheim, Germany*
* *Correspondence*

Received 27 November 2006. Received in revised form and accepted for publication by Adrian Sumner 24 January 2007

Key words: chromosome, cytoband, FISH, GGCC motif, genome sequence

Abstract

We have tested whether a direct correlation of sequence information and staining properties of chromosomes is possible and whether this combined information can be used to precisely map any position on the chromosome. Despite huge differences of compaction between the naked DNA and the DNA packed in chromosomes we found a striking correlation when visualizing the GGCC density on both levels. Software was developed that allows one to superimpose chromosomal fluorescence intensity profiles generated by chromolysin A3 (CMA3) staining with GGCC density extracted from the Ensembl database. Thus, any position along the chromosome can be defined in megabase pairs (Mb) besides the cytoband information, enabling direct alignment of chromosomal information with the sequence data. The mapping tool was validated using 13 different BAC clones, resulting in a mean difference from Ensembl data of 2 Mb (ranging from 0.79 to 3.57 Mb). Our results indicate that the sequence density information and information gained with sequence-specific fluorochromes are superimposable. Thus, the visualized GGCC motif density along the chromosome (sequence bands) provides a unique platform for comparing different types of genomic information.

Introduction

On the cytogenetic level, different chromosome banding techniques, including fluorescent and Giemsa staining methods, were developed in recent decades to visualize differential patterns along chromosomes. Previously published data have shown that staining intensities (i.e. chromosome bands) might be influenced by certain structural properties of the chromosome, among others, by staining behaviour of chromosomal proteins (Sumner 1974, 1984, Hliscs *et al.* 1997), DNA replication timing (Dutrillaux 1975a, b, Dutrillaux *et al.* 1976), density of genes (Craig & Bickmore 1993, Sumner *et al.* 1993) and base pair composition (Schweizer 1976, Ambros & Sumner 1987, Bernardi 1989, Saccone *et al.* 1993, Schweizer & Ambros 1994, Holmquist & Ashley 2006) as well as timing and degree of chromatin condensation (Drouin *et al.* 1991, Saitoh & Laemmli 1994, Craig *et al.* 1997).

The translation of the chromosome bands into a commonly agreed nomenclature is provided in the

Electronic Supplementary Material
The online version of this article (doi:10.1007/s10577-007-1129-3) contains supplementary material, which is available to authorized users.

ISCN (International System for Human Cytogenetic Nomenclature) (Mitelmann ed. 1995, Shaffer & Tommerup ed. 2005). The ISCN is an excellent uniform platform to describe every position along human chromosomes, as well as chromosome abnormalities, in a detailed fashion. The ISCN nomenclature is based on Giemsa (G)-banded ideograms (from Francke 1981, 1994), that correspond to a G-banding pattern of chromosomes. This staining pattern, obtained by trypsinization and subsequential Giemsa staining, reflects differences in the digestibility of proteins. Whether the G-banding pattern is due to differences in the GC content remains a matter of controversy (Sumner 1982, Korenberg & Rykowski 1988). Saccone *et al*. (2001) presented a correlation between GC density and chromosomal bands on the cytogenetic ideograms (from Francke 1994) of chromosomes 21 and 22. Interestingly, in contrast to the common belief that G bands are GC-poor, whilst R-bands are GC-rich (Holmquist *et al*. 1982), some Giemsa positive G bands of chromosome 22 were found to contain more G and C nucleotides than most of the Giemsa negative R-bands of chromosome 21 and, what was even more surprising, G bands 21q11.2 and 21q21.2 were characterized by a GC level higher than that of R bands 21q22.12 and 21q22.2. This indicated a lack of correspondence between ISCN ideograms and GC content distribution along the chromosomes.

In 2002 Niimura & Gojobori further elucidated the relationship between ISCN Giemsa banding and the GC content of all human chromosomes by constructing an 'in-silico banding' based on the draft human genome sequence. The 'in-silico bands' were constructed computationally by plotting the GC content along every chromosome. Thus, the authors were able to develop an ideogram based on the GC distribution and to compare this ideogram with that obtained by Giemsa banding (according to Francke 1994). The authors showed that neither the positions nor the sizes of the GC-rich 'in-silico bands' are in concordance with Giemsa negative bands (the same holds true for GC-poor regions and Giemsa positive bands). The fact that 'in-silico' banding pattern matched only imprecisely with G bands would explain why, until now, no direct linkage between the ISCN G-banded ideograms and the sequence data has been possible.

In 2004 Gilbert *et al*. analysed the distribution of compact and open chromatin fibres across the human genome. The authors employed sucrose gradient fractionation to separate differentially packed chromatin fibre structures and hybridized them to metaphase chromosomes. Compact chromatin fibres hybridized mainly to heterochromatic regions (centromeres and juxtacentromeric regions – C bands). In contrast, regions of high gene density (R bands and T bands) were found to have an open chromatin fibre conformation. Interestingly, some satellite DNAs appeared to be in open chromatin fractions and, what was even more surprising, compact fibres, besides C bands, hybridized also to a subset of euchromatin sites – intensely staining G bands (e.g. 1p31, 1q31 and q41; 3p24 and q24; 5q34, 7p21 and q21; 9q31, 12q21, 16p12) that show a low gene density. These data implied that the relation between chromatin structure and G-banding pattern remains a contentious issue.

The sequencing of the human genome has provided a platform on which to assemble a wide range of diverse information. However, in order for cytogeneticists to fully take advantage of this resource, it became essential to find a direct way to link the chromosome phenotype with the human genome sequence data. So far a number of studies have been undertaken to translate cytogenetic information into the sequence map of the human genome (Korenberg *et al*. 1999, Kirsch *et al*. 2000, Cheung *et al*. 2001). These studies were supplemented by the development of databases which permit linkage of chromosomal features with the genome sequence maps and analysis of the clinical and research cytogenetic data on this common platform (Knutsen *et al*. 2005). Nevertheless, all these efforts did not solve the problem of the lack of a full and direct alignment of chromosome properties to the gene mapping data, which hampers detailed elucidation and interpretation of chromosomal mapping data. In 2003 Furey & Haussler developed a dynamic programming algorithm to predict G-band locations in the human genome sequence. They used FISH data from the BAC Resource Consortium (Cheung *et al*. 2001). However, the BAC clones used covered only approximately a quarter of the total genome. Therefore, we aimed to find a simple way to directly and fully link cytogenetic and DNA sequence data.

The human genome sequence is available worldwide in publicly accessible databases, e.g. Ensembl (http://www.ensembl.org). These databases provide almost complete information on the base pair composition of every individual chromosome. It is

well accepted, and also evidenced after sequencing of the entire genome, that the GC content is not distributed evenly along the genome, but is rather represented by clusters of GC-rich DNA followed by AT-rich domains (Macaya *et al*. 1976, Yunis 1976, Lander *et al*. 2001, Wright *et al*. 2001, Pavlicek *et al*. 2002, Gregory *et al*. 2006, http://www.genome.gov/). Differential distribution of the GC content along chromosomes can be demonstrated by the use of sequence-specific fluorochromes. First experiments of this kind, already undertaken in the 1970s and 1980s, clearly indicated that the base composition along chromosomes can be visualized with sequence-specific fluorochromes preferentially binding to GC- or AT-rich DNA in defined chromosomal regions (Schweizer 1976, 1981, Ambros & Sumner 1987, Schweizer & Ambros 1994). As already demonstrated in 1976, chromomycin A3 (CMA3) – a sequence-specific fluorescent dye, has a strong affinity to GC-rich DNA (Schweizer 1976). It binds preferentially to the minor groove in DNA helix via the GGCC binding motif (Hou *et al*. 2004). The staining results in higher fluorescence intensity of GC-rich chromosomal areas as compared to chromatin regions with higher AT content. Therefore, CMA3 gives a fluorescence pattern which directly visualizes the differences in GGCC base pair distribution along the chromosome. This fluorescence pattern resembles an R-banding type (Schweizer 1976) and accordingly highlights interchromomeric regions in meiotic chromosomes (Ambros & Sumner 1987).

Despite the fact that the chromosomal DNA is approximately compacted 10 000–20 000 fold compared to the underlying linear DNA, we hypothesized that the GGCC motif distribution found on the sequence level is mirrored on the chromosomal level. The purpose of this study was to test whether the differential GGCC motif distribution along the chromosomes corresponds to the GGCC densities of published DNA sequences (Ensembl sequence database) and whether this feature can be used for chromosome mapping. To verify our assumption we aligned the GGCC density profile calculated on the basis of the sequence information available in the Ensembl database with the fluorescence intensities of CMA3 specifically highlighting the GGCC sequences along the chromosomes. Here we present a computational procedure for adjusting and comparing GGCC density highlighted on chromosomes (sequence bands) with the GGCC density in the DNA sequence, which allows a precise mapping of chromosomal information.

Materials and methods

BAC clones

BAC clones RP11-275F13, RP11-205M9, RP11-480N14, RP11-328L16, RP11-351E10, RP11-99J9, RP11-667F14, RP11-536D21, RP11-300I6, RP11-700F9, PAC-dJ167K13, RP11-1152A10, RP13-650J16 (for details see Table 1) were kindly provided by Mariano Rocchi (Resources for Molecular Cytogenetics, http://www.biologia.uniba.it/rmc/). DNA preparations from cultured bacteria (containing required construct) were performed using the Large-Construction Kit (Qiagen) according to the protocol provided by the company. DNAs were labelled with biotin-16-dUTP or digoxigenin-11-dUTP by nick-translation (Roche, Biotin- or DIG-Nick Translation Mix, containing DNA polymerase I, DnaseI, 0.25 mM dATP, 0.25 mM dCTP, 0.25 mM dGTP, 0.17 mM dTTP and 0.08 mM biotin-16-dUTP or DIG-11-dUTP, respectively) according to the manufacturer's protocol. The fragment sizes of labelled probes, determined on 1.2% agarose gel, were 200–500 bp. Human Cot 1 DNA was added in the amount of total volume of labelled probe minus 10 µl. Unincorporated nucleotides were removed by ethanol precipitation.

Fluorescence in-situ hybridization

Metaphase chromosome preparations were made from normal human peripheral lymphocyte cultures. Cultures were harvested using standard cytogenetic methods.

Fluorescence *in-situ* hybridizations were performed according to usual laboratory procedures. Hybridization mixes containing 0.4–0.8 ng/µl of labelled DNA were placed on slides, which were then covered with coverslips and sealed by rubber cement. After denaturation (78°C, 6 min), the slides were incubated overnight at 37°C in a humidified chamber. After overnight hybridization the slides were washed for 5 min at room temperature in 2× SSC, then 15 min at 42°C in 50% formamide/2× SSC, and then at 42°C in 2× SSC (2×7 min). The probes labelled with biotin were detected by incubation for 30 min in 37°C with mouse anti-biotin antibody

Table 1. Validation of the sequence banding tool

Gene	Clone	Cytoband position in Ensembl	Cytoband position in NCBI	Mb position in Ensembl (gene)	Mean difference in Mb	Standard deviation in Mb	Range*	Median value†
AF1p	RP11-275F13	1p32.3	1p32	51.60	2.42	2.13	−7.58–8.64	1.09
TPM3	RP11-205M9	1q21.3	1q21.2	150.95	0.79	0.61	−1.79–2.10	0.26
MYCN	RP11-480N14	2p24.3	2p24.1	16.03	0.94	1.02	−2.26–5.10	−0.16
ALK	RP11-328L16	2p23.2-23.1	2p23	29.50	2.07	1.75	−3.67–6.39	0.22
—	RP11-351E10	2q37.3	2q37.3	242.50	2.11	2.12	−8.28–1.11	−0.59
HIP1	RP11-99J9	7q11.23	7q11.23	74.84	1.10	0.84	−2.81–2.30	−0.17
TES	RP11-667F14	7q31.3	7q31.2	115.45	2.42	1.68	−7.03–3.97	−1.15
FANCF	RP11-536D21	11p14.3	11p15	22.60	1.83	1.31	−4.79–4.25	−0.33
BCL1	RP11-300I6	11q13.3	11q13	69.17	3.40	2.00	−6.76–7.05	−1.23
DDX10	RP11-700F9	11q22.3	11q22-q23	108.20	3.57	2.04	−7.35–7.47	−0.90
MLL	PAC-dJ167K13	11q23.3	11q23	117.85	1.73	2.07	−9.51–2.66	0.14
RARA	RP11-1152A10	17q21.2	17q21	35.73	1.27	0.97	−1.28–4.11	1.04
ASPL	RP13-650J16	17q25.3	17q25	77.55	1.97	1.49	−6.12–3.39	−0.97

* 0 in range indicates the position in Ensembl.
† The value below which 50% of the cases fall.

(1:30, DAKO) in 2% bovine serum albumin (BSA)/PBS and rabbit anti-mouse TRITC (1:40, DAKO) in 2% BSA/PBS, digoxigenin-labelled probes were detected with anti-digoxigenin rhodamine (1:50, Roche). Subsequently, the slides were washed in 4× SSC/0,1% Tween 20 at 42°C (2×7 min), then in PBS at room temperature (2×5 min) and dehydrated.

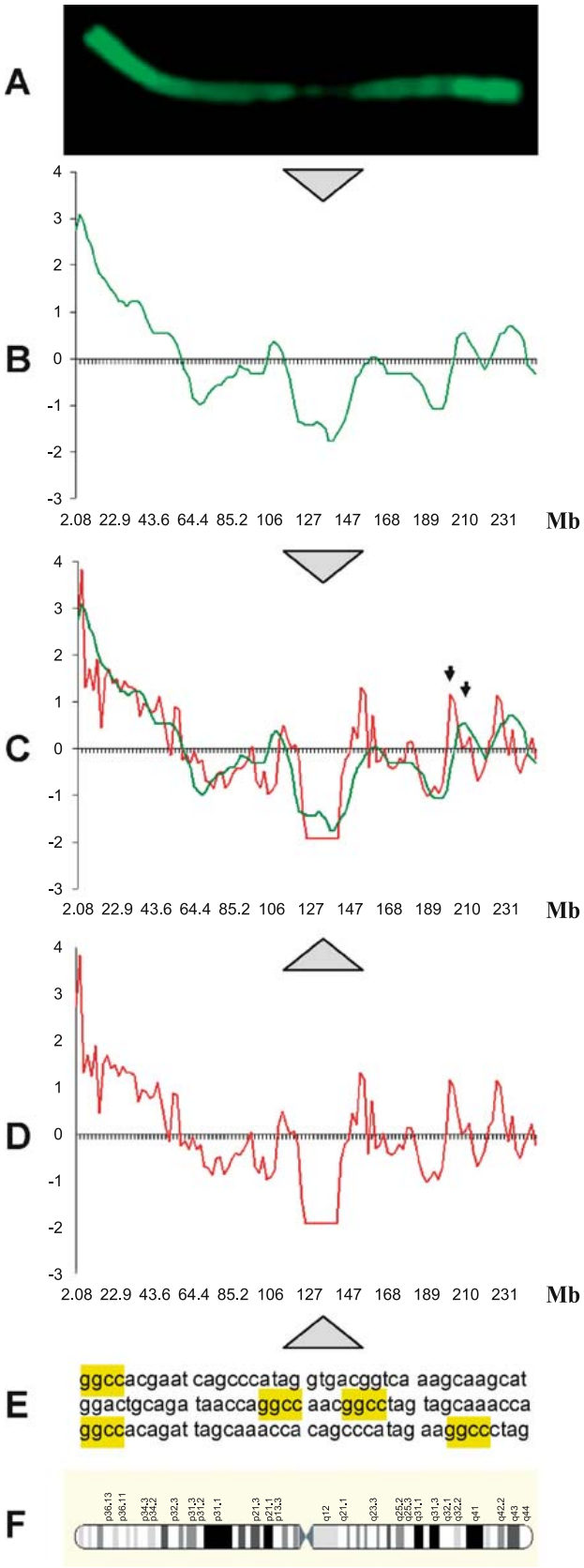
Chromomycin A3 staining

To simultaneously visualize fluorescence signals and R-banding pattern (GC content) on the chromosomes chromomycin A3 (CMA3), distamycin A (DA) and DAPI (CDD triple staining) were used as a counterstain. The CDD chromosome staining was performed on the slides according to the standard protocol (Schweizer 1976, Schweizer & Ambros 1994). CMA3 solution (0.5 mg/ml, in McIlvane's buffer (pH 7) diluted 1:1 with water, containing 0.005 volume 1 M $MgCl_2$) was placed on the slides, which were then covered with a plastic coverslip and incubated overnight at room temperature in the dark. After that the slides were rinsed in deionized water, blown dry and DA solution (1 mg/ml, in McIlvane's buffer (pH 7)) was applied for 10 min. This was followed by rinsing in deionized water and staining with DAPI solution (2 mg/ml in McIlvane's buffer (pH 7) diluted 1:1000 with PBS; 15 min). The slides were mounted in an antifade solution (87% Glycerol McIlvane's (pH 7) 1:1, plus 250 µl 1 M $MgCl_2$) mixed with a Vectashield mounting medium for fluorescence (Vector Laboratories) in 1:1 ratio and incubated at least 3 days at 37°C. The inclusion of Mg^{2+}, as well as the ageing of slides for at least 3 days, is crucial to obtain a stable CMA3 fluorescence.

Bioinformatics analyses

Regarding our first efforts to align sequence density information and CMA3 intensity information (Figure 1), for sequence analysis the assembled DNA chromosome files in FASTA format have been downloaded from the freely accessible Ensembl web server (ftp://ftp.ensembl.org/pub/). To obtain the absolute GGCC sequence motif count as well as the overall amount of G and C content in a specified window size, a computer script was developed using the Python programming language.

The program splits the DNA sequence into window slices of a previously defined length depending on the number of pixel counts obtained from the CMA3 fluorescence profile scan from the ISIS software package. Calculations such as the absolute GGCC sequence motif count in 5′ to 3′ direction and the appearance of the bases G and C relative to the length of the window slice were performed. The relative amount of the bases G and C in a DNA sequence slice, where gaps are present in the chromosome counting, were calculated separately. Gaps are normally displayed with the DNA letter code N, which stands for any of the four bases. After calculation the data information is stored in a TAB-



delimited spreadsheet file format which can be easily used for further processing. The output contains information of the chromosome position in bases and megabases, the absolute count of the DNA sequence motif defined by the user, the relative amount – as a percentage – of G and C, and finally all absolute values of the four bases A, C, G and T.

Regarding the Warp tool development, for sequence density analysis the 24 assembled DNA chromosome files in FASTA format have been downloaded from the freely accessible Ensembl web server (ftp://ftp.ensembl.org/pub/). The sequence of each chromosome was divided into intervals of 20 000 base pairs, and the absolute GGCC sequence motif count in each interval was saved to a single binary external profile set (XPS) file using an ISIS utility program (MetaSystems) for later use with the Warp tool.

The ISCN band numbers corresponding to sequence positions is not available in the FASTA files; therefore this information was directly extracted from the Ensembl web site (http://apr2006.archive.ensembl.org/Multi/newsview?rel=38#cat2), using java methods and ENSJ, one of Ensembl's Application Programming Interfaces (API). The connection to the database was established through an Ensembl driver. The driver provides adaptors that allow the retrieval of the required annotation information that will be stored in one ANSI text file per chromosome. The ISIS utility mentioned above will then read these 24 files and save them as a single ISIS external profile annotation (XPA) file.

ISIS external profiles function and Warp tool

The slides were imaged with a Zeiss Axioplan 2 microscope (Zeiss, Austria) with motorized reflector turret. Digital images were captured using an IMAC S30 integrating CCD camera and the ISIS software

Figure 1. Linkage of the banding properties and sequence information of chromosome 1. **A**: R-banding pattern obtained by staining of the chromosome with GGCC-specific fluorochrome, CMA3. **B**: CMA3 fluorescence intensity curve visualizing the distribution of the GGCC motif along the chromosome. **C**: Merge of **B** and **D** curves, arrows denote peak shifts. **D**: Graphical representation of the GGCC motif density along the chromosome generated on the basis of the Ensembl genome sequence database. **E**: Schematic visualization of the GGCC motif frequency. **F**: Ideogram giving cytoband information on every position along the chromosome. The horizontal axis in **B**, **C** and **D** uses an Mb scale.

(MetaSystems, Altlussheim, Germany). For each fluorochrome a specific fluorescence filter cube was used; switching between filter cubes was automatic under software control. FISH signals were captured using a TRITC filter cube. For each colour channel the optimal integration time was automatically determined by the ISIS software.

TRITC and CMA3 fluorescence intensity profiles were obtained using the ISIS software by interactively defining the longitudinal axis of individual chromosomes, followed by automatic measurement and normalization of fluorescence intensities along this line (see Figure 2A). The track of defined axis had to be set along one of the chromatids in order to achieve an appropriate and simultaneous measurement of the sequence banding and the FISH signal fluorescence intensities. Correlation with the density of the GGCC binding motif in the DNA sequence was done using the ISIS External Profile Function and the Warp tool developed by MetaSystems (Germany).

For this purpose the TRITC and CMA3 fluorescence intensity profiles are graphically displayed in the lower part of the screen (curves *a* and *b* in Figure 2B). The software currently supports up to nine fluorescence colour channels. After the user has selected the chromosome being analysed (circle *c* in Figure 2B), the GGCC sequence density information is read from the intermediate file mentioned above, normalized to the length of the fluorescence profiles, and displayed in the upper part of the screen (curve *d* in Figure 2B). By changing the smoothing power, the level of detail in the displayed sequence density profile can be adapted to that observed in the fluorescence intensity profiles, which makes the matching process easier.

To compensate for differential contraction artefacts in the chromosome fluorescence intensity profiles the user starts the 'Warp' mode (button *f* in Figure 2B) and first defines the telomeres and the centromere (primary landmarks) by simply clicking corresponding points in the upper (curve of GGCC sequence density created on the basis of Ensembl database) and lower part (fluorescence intensity curve of the CMA3-stained chromosome) of the screen (procedure illustrated in Electronic Supplementary Material S1). If additional characteristic peaks or valleys in the respective profiles can be matched, they can be marked in the same way (secondary landmarks), which increases the accuracy

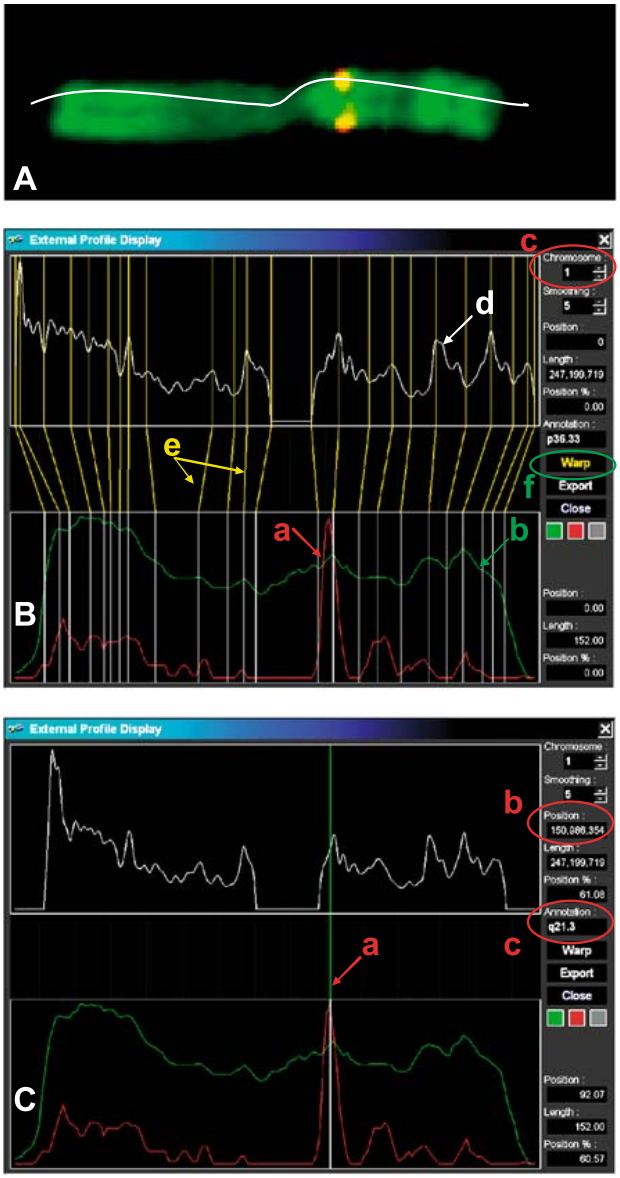


Figure 2. Assignment of TPM3 FISH signal to genome sequence information using Warp tool. **A**: Definition of longitudinal axis of normal human chromosome 1 after FISH of the TPM3 BAC clone and simultaneous staining with CMA3. The white line indicates the track of the defined axis along the chromatid. The FISH signal and the CMA3 fluorescence profile are visualized as intensity curves in **B**, red peak (a) represents the FISH signal; green curve (b) illustrates GGCC motif frequency based on the CMA3 fluorescence intensity; white one (d) reflects the GGCC density extracted from the Ensembl database; (e) denotes linkage of landmarks. Pter is defined as the first landmark at 0bp (on the left side), qter as the last position (on the right side) and the centromere as GC-poor domain (low fluorescence intensity in the middle). **C**: Assignment of the chromosomal fluorescence information to the sequence information and determination of the position of the BAC clone (a) in base pairs (b) and in cytobands (c).

of the position measurement. All landmarks are displayed as yellow/white lines connecting corresponding positions in the profiles (*e* in Figure 2B). When the 'Warp' mode is terminated these lines are no longer displayed, and the sequence density profile is linearly transformed (stretched/squeezed) in such a way that all landmark positions in both profile areas coincide (Figure 2C).

To perform a position measurement, the cursor – represented by a green/white vertical line – can be moved to a specific location by a mouse click (line *a* in Figure 2C). The system will then display the current cursor position in the warped sequence density profile in base pairs (circle *b* in Figure 2C) and the corresponding band number annotation (circle *c* in Figure 2C), as well as the total length of the sequence of the chromosome in base pairs, and the relative position of the cursor as a percentage (upper right in Figure 2C). In addition, the system will also display the cursor position in the fluorescence intensity profiles, i.e. the absolute position in pixels, the total length of the profile in pixels, and the relative position as a percentage (lower right of Figure 2C).

For the validation experiments the primary and secondary landmarks were interactively defined, and the cursor was set to the position of the BAC-FISH signal peak. The screen display then showed the absolute and relative sequence position of the FISH signal, and the corresponding band number annotation. By using the 'Export' button the fluorescence intensity profiles, and the GGCC density profiles as well as the position information, and the band number annotations were exported to a text file for further processing and analysis.

Three observers measured the signal positions in 30 metaphases per BAC clone. None of the observers had any information on the base pair and cytoband position of the analysed BAC. The difference between the BAC signal location measured with the Warp tool and the correct location of the gene according to Ensembl was calculated. Thus, a value of '0' (Figure 3) indicates the correct position. Descriptive statistical methods (median, minimum, maximum, quantiles) were used to describe the distribution of the deviation from the true value. In addition, exact 95% confidence intervals, based on the binomial distribution, for the median were calculated. A pitfall we found essential to be taken into consideration in order to fully take advantage of the mapping potential of the Warp tool is the proper alignment of all fluorescence channels of the captured metaphase. Even a slight merging discordance

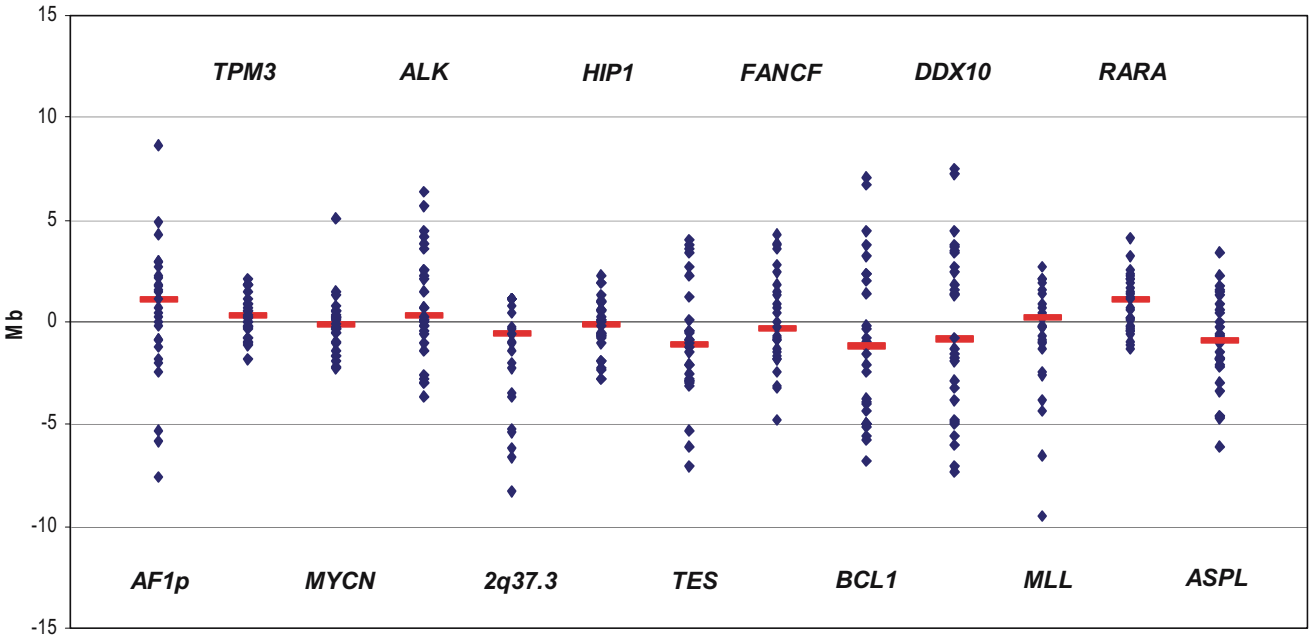


Figure 3. Differences between clone position measured with Warp tool and gene position given in Ensembl database. '0' Mb on Y axis indicates the gene position according to Ensembl. Each blue dot represents one single measurement. Red bars denote median values regarding measurements for each gene.

may cause an error in mapping results. Therefore, images composed of diverse fluorescence channels have to be analysed with great caution.

Results

Chromosomal bands highlighted with CMA3 staining correspond to DNA with high GGCC motif frequency

We tested whether the fluorescence pattern obtained by staining chromosomes with a sequence-specific dye can be correlated with the base pair density profile according to the sequence information. Therefore, we compared the fluorescence intensity of CMA3 along the chromosome (Figure 1A and B) with the GGCC frequency calculated on the basis of the sequence information available in a published sequence database (Figure 1D). The longitudinal axis of CMA3-stained chromosome was interactively defined, and the fluorescence profile was displayed on the screen. We correlated the fluorescence profiles with the graphically visualized GGCC densities extracted from the Ensembl database version 38 (http://www.ensembl.org) (Figure 1C). High similarities between the fluorescence intensities and the GGCC frequencies were found. However, longitudinal shifts between CMA3 fluorescence intensities and GGCC densities were also observed (indicated by arrows in Figure 1C).

GGCC content visualized on chromosomes can be precisely aligned to GGCC density in the genome sequence

In contrast to the DNA sequence information obtained from the sequence database, chromosomes frequently display stretched or squeezed regions, making it difficult to directly compare the two sets of information. Thus, the peaks of the chromosomal GGCC density profile visualized by staining with CMA3 are sometimes not in the same positions as GGCC density peaks on the sequence level. In order to perfectly match the CMA3 fluorescence intensity peaks with the sequence density peaks, the development of a tool to overcome stretching and/or squeezing artefacts was required. The External Profile Function and Warp tool, a program component of the ISIS software (MetaSystems), was developed to enable a direct alignment of both data sets. The definition of landmarks along the chromosomes, and the interpolation of the chromosomal regions in between, enable precise alignment of the chromosomal data to the sequence information (see Materials and Methods section for a detailed procedure description).

Validation of the sequence banding technique

To verify the accuracy of the sequence banding technique we hybridized 13 different BAC clones to CMA3-stained chromosomes. Their positions on the chromosomes were determined on the basis of the sequence banding information applying the Warp tool. After defining the chromosome axis the FISH signal and the CMA3 fluorescence profile (Figure 2A) were aligned with the GGCC density profile according to landmarks defined with the Warp tool (Figure 2B). This assignment allowed us to determine the positions of the BAC probes (peak of FISH signal) in cytobands and base pairs (Figure 2C). Differences between the obtained mapping positions and the expected ones (Ensembl database) are given in Table 1. The mean difference between our mapping data and the mapping data given in Ensembl was 2 Mb (range 0.79–3.57 Mb), the standard deviation between the measurements ranged from 0.61 (*TPM3*) to 2.13 Mb (*AF1p*). Figure 3 demonstrates differences (in Mb) between the BAC clone positions measured with the Warp tool and the corresponding gene locations obtained from the Ensembl genome database; it is shown for each BAC clone. The '0' Mb position on the *y* axis indicates the location of the gene according to Ensembl. Results of all measurements done by three different observers (30 metaphases were analysed per BAC) are included. Each blue dot represents one single measurement. Red horizontal bars denote the median values for position measurements of each clone. The median value for each clone was found to be within the mean difference between our mapping data and the gene positions reported in Ensembl (see Table 1). The calculated confidence intervals indicate the precision of the estimated median values and show the effect of sampling variation. If the value '0', representing the gene position in Ensembl, is not included in the confidence interval, this is an indication that the deviation between the measured and the correct location cannot be explained by imprecise measure-

ment or experimental variation alone. For two genes, RARA and TES, the 95% confidence interval with respect to the median value did not include the true value of '0', thus suggesting that the presumed 'correct' location indicated in Ensembl might be incorrect or that inter-individual variations (e.g. in GGCC density) may play a role and should be taken into consideration.

Discussion

Direct linkage of cytogenetic data and genome sequence applying the Warp tool

The possibility to chemically stain metaphase chromosomes has provided a technological tool that powered many important discoveries in the fields of cytogenetics, human and cancer genetics. Although more recently developed experimental techniques such as fluorescence *in-situ* hybridization (FISH) (Trask 1991), comparative genomic hybridization (CGH) (Kallioniemi *et al*. 1992, 1993), multicolor FISH (mFISH) (Speicher *et al*. 1996), spectral karyotyping (Schrock *et al*. 1996) and array-CGH (aCGH) (Pinkel *et al*. 1998) allow cytogeneticists to more precisely investigate chromosomal abnormalities, cytogenetic band analysis is still used extensively in research and medical diagnosis.

However, so far identification of positions on a chromosome has been hampered by the lack of a universally applicable and simple mapping system which enables one to align genome sequence data with chromosomal phenotypes (banding patterns). To overcome this limitation we used the information of the GGCC motif distribution, visualized on chromosomes with CMA3, and superimposed this information to the sequence data. Despite huge differences of compaction between the naked DNA and the DNA packed in chromosomes we found a striking correlation when visualizing the GGCC density profile on both levels. Prerequisite for this, however, was the development and testing of an efficient method which could be employed to correctly superimpose both sets of information. Chromosome staining with the base pair intercalating dye ethidium bromide (EtBr), which shows no base specificity, did not result in any banding pattern of preferentially highlighted euchromatic regions along chromosomes (data not shown).

In the genome sequence databases, GC content is described and elucidated directly on a linear DNA molecule. However, metaphase chromosomes dropped on slides are exposed to different forces, leading to a differential stretching or squeezing of chromosome regions. The stretching and/or squeezing can in consequence lead to shifts between fluorescence peaks and the corresponding GGCC density peaks from the Ensembl database. Such differences in the chromosome compaction can be clearly seen in Figure 1. The chromosome displayed in Figure 1A is visibly stretched in the centromeric region. Inspection of Figure 1C affirms that the centromeric region in the GGCC density profile on the sequence level (red curve) is relatively narrow in comparison to its width in the CMA3 fluorescence intensity profile (green curve). Due to this fact there is a noticeable systematic tendency in the shifts. CMA3 fluorescence intensity peaks are shifted towards pter and qter, which clearly explains the discrepancies between chromosome banding and the DNA sequence information (Figure 1C, see the distal part of the q arm, shifts are depicted with arrows). Additionally, the qter of the chromosome was not correctly assigned to the sequence density profile. To overcome these differential compactions the ISIS software was supplemented with a newly developed function, named a Warp tool. This tool allows direct alignment of chromosomal staining properties with the graphical visualization of the GGCC density profile, thus permitting a definition of any position along the chromosome not only in cytoband but also in base pairs (Figure 2). The tool requires one to manually define so-called landmarks, i.e. structures that can be detected on the chromosome as well as on the DNA sequence level. These are telomeres and centromeres (primary landmarks), and peaks or valleys of GGCC density (secondary landmarks). The user can identify, besides positive peaks, e.g. high CMA3 fluorescence intensity reflecting high density of GGCC motif in the DNA sequence, also negative peaks (valleys) indicating negative sites of low CMA3 fluorescence intensity corresponding to low GGCC density in the genome sequence. The landmarks denote fixed points which correspond to each other in both sets of information: fluorescence intensity of CMA3 according to preferential binding to the GGCC nucleotides and GGCC motif density from the Ensembl in the linear DNA sequence. When more secondary landmarks are defined the correla-

tion becomes more precise. Validation experiments, including hybridization of 13 different BAC clones to CMA3-stained chromosomes, revealed the high precision of this sequence-based mapping tool.

Cytoband annotations in genome sequence databases

The availability of genome sequence information prompted the development of publicly accessible databases, e.g. Ensembl (Birney *et al.* 2006, http://www.ensembl.org), NCBI (National Center for Biotechnology Information) database (Pruitt & Maglott 2001, Wheeler *et al.* 2005, http://www.ncbi.nlm.nih.gov/) or the University of California Santa Cruz (UCSC) Genome Browser Database (Kent *et al.* 2002, Karolchik *et al.* 2003, http://genome.ucsc.edu/).

Surprisingly, the cytoband information given in different databases is not identical. The ISCN ideograms based on the G-banded karyotype of Francke (1994) and Mitelmann ed. (1995) represent the basis of the NCBI (Entrez) Map Viewer, whilst the UCSC as well as the Ensembl genome browsers employ the ideograms based on Furey's and Haussler's cytogenetic band predictions (Furey & Haussler 2003). These ideograms were developed on the basis of G-band locations in the human genome sequence using a dynamic programming algorithm ('Bander' software). This approach employed data from chromosome FISH experiments made available by the BAC Resource Consortium (Cheung *et al.* 2001, http://www.ncbi.nlm.nih.gov/genome/cyto/hbrc.shtml). The 'Bander' software designed for this study allowed the authors to assign defined BAC to certain chromosomal bands; they were thus able to define the length (total number of bases) of each cytoband more precisely as compared to ISCN ideograms.

Comparison of chromosomal ideograms in terms of sequence and cytoband information

The 'Bander' predicted ideograms and the standard ISCN ideograms differ from each other in terms of the position and the length of certain cytobands (http://www.soe.ucsc.edu/research/compbio/cytobands/, Figure 4). This could be explained by the fact that the ISCN ideograms are based on visual estimates of the lengths of G bands (Francke 1994), and these may have been under- or overestimated regarding the actual length and content of nucleotides. Furey & Haussler (2003) showed that the lengths of the darkest G bands were consistently underestimated, while the opposite is true for the lighter bands. Thus, the 'Bander' ideograms contain more nucleotides in G positive bands and less in G negative ones in comparison to visually estimated sizes of the ISCN bands. As the ISCN ideograms are used in the NCBI Human Genome browser and the 'Bander' ideograms in Ensembl, the sequence information ascribed to cytobands and cytoband boundaries is not identical in these two databases. Thus, we could hypothesize that genes located in the distinct bands of 'Bander' ideograms (Ensembl) may possibly be found outside of these bands in the ISCN ideograms (NCBI). Therefore, an effort to assign a gene of interest to the chromosomal (cytoband) position may possibly result in discrepant annotations depending on the choice of the source database. Indeed, regarding database search for the *MYCN* gene, the Ensembl database displays its position in 2p24.3 (http://www.ensembl.org/Homo_sapiens/contigview?l=2:15998134-16004579), whilst NCBI reports the location in much more proximal 2p24.1 (http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?taxid=9606&chr=2&MAPS=ugHs,genes-r&cmd=focus&fill=40&query=uid(11094629)&QSTR=MYCN; Figure 4, red arrows indicate the *MYCN* gene position). It is noteworthy that, within the region of 2p, the two different ideograms show distinct discrepancies in terms of sizes and positions of cytobands. Another example of discrepant annotation is the *TPM3* gene, localized in 1p21.3 according to the Ensembl database and in 1p21.2 according to NCBI. Importantly, 1p21.3 is an R band, while 1p21.2 is a G band. Three of 13 genes used in our studies show different cytoband positions in the Ensembl and NCBI browsers. Only two of them displayed identical localizations. Positions of the other eight were reported less precisely in NCBI providing only the main cytoband annotation for these genes and no subband information as in Ensembl. Cytoband positions of all analysed genes according to both Ensembl and NCBI are given in Table 1.

However, in order to integrate cytogenetics information with the sequence data, the Warp tool makes it possible to rely purely on the sequence information and sequence-specific banding. As a result we are able to omit any kind of G-band ideogram relation. As the chromosome mapping can be driven purely on the sequence information, independently of G-band ideogram information, the problem of divergent

Ideogram:	ISCN	'Bander'
Reference:	Francke 1994; ISCN 1995 (Mitelman ed.)	Furey and Haussler 2003
Basis:	G-banding (Seabright 1971); visual estimates	size/location of G-bands in the genome sequence predicted with 'Bander' software
Database:	NCBI http://www.ncbi.nih.gov/genome/guide/human/	Ensembl; UCSC http://www.ensembl.org/Homo_sapiens/index.html

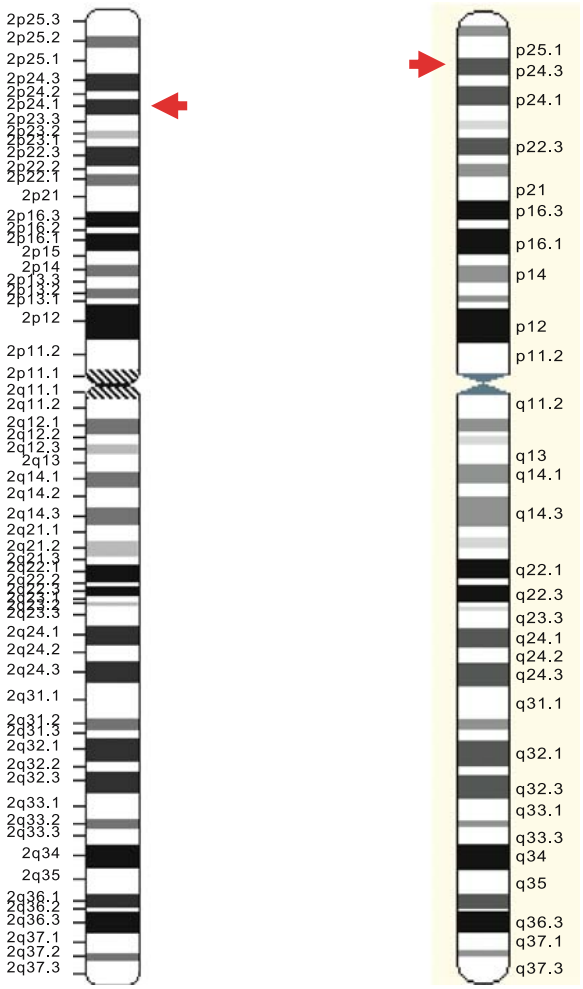


Figure 4. Comparison of chromosome 2 ideograms. The right ideogram is based on the 'Bander' predictions; the left one is the standard ISCN ideogram (www.soe.ucsc.edu/research/compbio/cytobands/). Red arrows indicate the *MYCN* gene position reported in the NCBI and the Ensembl browser.

ideograms existing in the different databases may be overcome.

Conclusion

Our results indicate that sequence density information and fluorescence intensity information gained with sequence-specific fluorochromes can be superimposed. Thus, sequence banding in combination with the External Profiles Function provides a new and promising approach to link the genomic with the chromosomal world. We are convinced that such a direct linkage of chromosomal with sequence data will facilitate a number of molecular cytogenetic applications, e.g. FISH, CGH and CESH (comparative expressed sequence hybridization) by bridging these two pieces of information, so far unconnected but in nature identical.

Acknowledgements

We thank Bettina Brunner for excellent technical assistance, Ulrike Pötschger for statistical analysis of the data, Dr Adrian T. Sumner and Dr Heinrich Kovar for valuable comments, and Marion Zavadil for proofreading. We are also very grateful to Prof. Dr. Mariano Rocchi for providing the BAC clones. This study was supported by the St Anna Kinderkrebsforschung, Vienna, Austria and the Austrian Federal Ministry for Science, Education, and Culture (GEN-AU Project Bioinformatics Integration Network).

References

Ambros PF, Sumner AT (1987) Correlation of pachytene chromomeres and metaphase bands of human chromosomes, and distinctive properties of telomeric regions. *Cytogenet Cell Genet* **44**: 223–228.

Bernardi G (1989) The isochore organization of the human genome. *Annu Rev Genet* **23**: 637–661.

Birney E, Andrews D, Caccamo M *et al*. (2006) Ensembl 2006. *Nucleic Acids Res* **34**: D556–D561.

Cheung VG, Nowak N, Jang W *et al*. (2001) Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**: 953–958.

Craig JM, Bickmore WA (1993) Chromosome bands–flavours to savour. *Bioessays* **15**: 349–354.

Craig JM, Boyle S, Perry P, Bickmore WA (1997) Scaffold attachments within the human genome. *J Cell Sci* **110**: 2673–2682.

Drouin R, Lemieux N, Richer CL (1991) Chromosome condensation from prophase to late metaphase: relationship to chromosome bands and their replication time. *Cytogenet Cell Genet* **57**: 91–99.

Dutrillaux (1975a) [Various banding simultaneously obtained on the same slides, after treatment by BrdU (author's transl)]. *Humangenetik* **30**: 297–306.

Dutrillaux B (1975b) [Discontinued treatment with BudR and staining with acridine orange: observation of R- or Q- or intermediary banding (author's transl)]. *Chromosoma* **52**: 261–273.

Dutrillaux B, Couturier J, Richer CL, Viegas-Pequignot E (1976) Sequence of DNA replication in 277 R- and Q-bands of human chromosomes using a BrdU treatment. *Chromosoma* **58**: 51–61.

Francke U (1981) High-resolution ideograms of trypsin–Giemsa banded human chromosomes. *Cytogenet Cell Genet* **31**: 24– 32.

Francke U (1994) Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet Cell Genet* **65**: 206–218.

Furey TS, Haussler D (2003) Integration of the cytogenetic map with the draft human genome sequence. *Hum Mol Genet* **12**: 1037–1044.

Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**: 555–566.

Gregory SG, Barlow KF, McLay KE *et al*. (2006) The DNA sequence and biological annotation of human chromosome 1. *Nature* **441**: 315–321.

Hliscs R, Muhlig P, Claussen U (1997) The nature of G-bands analyzed by chromosome stretching. *Cytogenet Cell Genet* **79**: 162–166.

Holmquist G, Ashley T (2006) Chromosome organization and chromatin modification: influence on genome function and evolution. *Cytogenet Genome Res* **114**: 96–125.

Holmquist G, Gray M, Porter T, Jordan J (1982) Characterization of Giemsa dark- and light-band DNA. *Cell* **31**: 121–129.

Hou MH, Robinson H, Gao YG, Wang AH (2004) Crystal structure of the [Mg2+-(chromomycin A3)2]-d(TTGGCCAA)2 complex reveals GGCC binding specificity of the drug dimer chelated by a metal ion. *Nucleic Acids Res* **32**: 2214–2222.

Kallioniemi A, Kallioniemi OP, Sudar D *et al*. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**: 818–821.

Kallioniemi OP, Kallioniemi A, Sudar D *et al*. (1993) Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors. *Semin Cancer Biol* **4**: 41–46.

Karolchik D, Baertsch R, Diekhans M *et al*. (2003) The UCSC Genome Browser Database. *Nucleic Acids Res* **31**: 51–54.

Kent WJ, Sugnet CW, Furey TS *et al*. (2002) The human genome browser at UCSC. *Genome Res* **12**: 996–1006.

Kirsch IR, Green ED, Yonescu R *et al*. (2000) A systematic, high-resolution linkage of the cytogenetic and physical maps of the human genome. *Nat Genet* **24**: 339–340.

Knutsen T, Gobu V, Knaus R *et al*. (2005) The interactive online SKY/M-FISH & CGH database and the Entrez cancer chromosomes search database: linkage of chromosomal aberrations with the genome sequence. *Genes Chromosomes Cancer* **44**: 52–64.

Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* **53**: 391–400.

Korenberg JR, Chen XN, Sun Z *et al*. (1999) Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and biomedicine. *Genome Res* **9**: 994–1001.

Lander ES, Linton LM, Birren B *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Macaya G, Thiery JP, Bernardi G (1976) An approach to the organization of eukaryotic genomes at a macromolecular level. *J Mol Biol* **108**: 237–254.

Mitelmann F (1995) *ISCN 1995: An International System for Human Cytogenetic Nomenclature*. Basel: Karger.

Niimura Y, Gojobori T (2002) In silico chromosome staining: reconstruction of Giemsa bands from the whole human genome sequence. *Proc Natl Acad Sci USA* **99**: 797–802.

Pavlicek A, Paces J, Clay O, Bernardi G (2002) A compact view of isochores in the draft human genome sequence. *FEBS Lett* **511**: 165–169.

Pinkel D, Segraves R, Sudar D *et al*. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* **20**: 207–211.

Pruitt KD, Maglott DR (2001) RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res* **29**: 137–140.

Saccone S, De Sario A, Wiegant J, Raap AK, Della Valle G, Bernardi G (1993) Correlations between isochores and chromosomal bands in the human genome. *Proc Natl Acad Sci USA* **90**: 11929–11933.

Saccone S, Pavlicek A, Federico C, Paces J, Bernard G (2001) Genes, isochores and bands in human chromosomes 21 and 22. *Chromosome Res* **9**: 533–539.

Saitoh Y, Laemmli UK (1994) Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell* **76**: 609–622.

Schrock E, du Manoir S, Veldman T *et al*. (1996) Multicolor spectral karyotyping of human chromosomes. *Science* **273**: 494–497.

Schweizer D (1976) Reverse fluorescent chromosome banding with chromomycin and DAPI. *Chromosoma* **58**: 307–324.

Schweizer D (1981) Counterstain-enhanced chromosome banding. *Hum Genet* **57**: 1–14.

Schweizer D, Ambros PF (1994) Chromosome banding. Stain combinations for specific regions. *Methods Mol Biol* **29**: 97– 112.

Shaffer LG, Tommerup N (2005) *ISCN 2005: An International System for Human Cytogenetic Nomenclature*. Basel: Karger.

Speicher MR, Gwyn Ballard S, Ward DC (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet* **12**: 368–375.

Sumner AT (1974) Involvement of protein disulphides and sulphydryls in chromosome banding. *Exp Cell Res* **83**: 438–442.

Sumner AT (1982) The nature and mechanisms of chromosome banding. *Cancer Genet Cytogenet* **6**: 59–87.

Sumner AT (1984) Distribution of protein sulphydryls and disulphides in fixed mammalian chromosomes, and their relationship to banding. *J Cell Sci* **70**: 177–188.

Sumner AT, de la Torre J, Stuppia L (1993) The distribution of genes on chromosomes: a cytological approach. *J Mol Evol* **37**: 117–122.

Trask BJ (1991) Fluorescence *in situ* hybridization: applications in cytogenetics and gene mapping. *Trends Genet* **7**: 149–154.

Wheeler DL, Barrett T, Benson DA *et al*. (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **33**: D39–D45.

Wright FA, Lemon WJ, Zhao WD *et al*. (2001) A draft annotation and overview of the human genome. *Genome Biol* **2**: RESEARCH0025.

Yunis JJ (1976) High resolution of human chromosomes. *Science* **191**: 1268–1270.

Organization of chromatin and histone modifications at a transcription site

Waltraud G. Müller,[1] Dietmar Rieder,[2] Tatiana S. Karpova,[1] Sam John,[1] Zlatko Trajanoski,[2] and James G. McNally[1]

[1]Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, Bethesda, MD 20892
[2]Christian Doppler Laboratory for Genomics and Bioinformatics, Institute for Genomics and Bioinformatics, Graz University of Technology, 8010 Graz, Austria

According to the transcription factory model, localized transcription sites composed of immobilized polymerase molecules transcribe chromatin by reeling it through the transcription site and extruding it to form a surrounding domain of recently transcribed decondensed chromatin. Although transcription sites have been identified in various cells, surrounding domains of recently transcribed decondensed chromatin have not. We report evidence that transcription sites associated with a tandem gene array in mouse cells are indeed surrounded by or adjacent to a domain of decondensed chromatin composed of sequences from the gene array. Formation of this decondensed domain requires transcription and topoisomerase IIα activity. The decondensed domain is enriched for the trimethyl H3K36 mark that is associated with recently transcribed chromatin in yeast and several mammalian systems. Consistent with this, chromatin immunoprecipitation demonstrates a comparable enrichment of this mark in transcribed sequences at the tandem gene array. These results provide new support for the pol II factory model, in which an immobilized polymerase molecule extrudes decondensed, transcribed sequences into its surroundings.

Introduction

There is increasing evidence for a spatial organization of transcription (Iborra and Cook, 2002; Chakalova et al., 2005). Pol II molecules form clusters within cells (Iborra et al., 1996), and nascent transcripts accumulate there, defining these clusters as transcription sites (Jackson et al., 1993; Wansink et al., 1996). These transcription sites can transcribe different genes from distant parts of the same chromosome or potentially even different chromosomes (Osborne et al., 2004).

According to the transcription factory model (Cook, 1999), transcription sites contain immobilized pol II molecules that spool the chromatin template in and out of the site. To date, however, evidence for movement of the chromatin template through a transcription site is largely theoretical. It has been argued that because transcripts appear within a restricted volume defining the transcription site, the polymerase cannot move very far, and so it is more likely that the chromatin template moves (Cook, 1999). This scenario also solves entanglement problems of the transcript and template (Cook, 1999).

If the chromatin template is reeled in and out of a transcription site, this site should be adjacent to or surrounded by decondensed, transcribed chromatin. In fact, transcription sites are surrounded by chromatin (Iborra et al., 1996), but, because most structures in the nucleus are found within chromatin, it has not been clear whether the chromatin seen around any one transcription site is associated with loci being transcribed by that site.

To investigate the spatial organization of chromatin at a transcription site, we have taken advantage of a mouse cell line harboring a tandem array. The array is composed of 200 directly repeated copies of a 9-kb element composed of the mouse mammary tumor virus (MMTV) promoter followed by reporter gene sequences (Walker et al., 1999).

Transcription from the array can be induced above basal levels by a hormone-stimulated GFP-tagged glucocorticoid receptor (GR) that also enables visualization of the array in live or fixed cells (McNally et al., 2000). Hormone induces a transcriptional response at the array comparable with that at single-copy MMTV promoters, including the recruitment of cofactors (Müller et al., 2001), specific nucleosome remodeling (Fragoso et al., 1998), and adaptation to prolonged hormone treatment (Fragoso et al., 1998). In addition, higher order chromatin structures at the array are indistinguishable from the structures observed in transcriptionally active domains of natural chromosomes (Müller et al., 2004).

Therefore, the array exhibits several features that are characteristic of normal transcription. Because of its size, however, the array is readily detected by light microscopy. Thus, it provides a useful model system for examining in a single cell the

Correspondence to James G. McNally: mcnallyj@exchange.nih.gov

Abbreviations used in this paper: BrUTP, bromo-UTP; ChIP, chromatin immunoprecipitation; GR, glucocorticoid receptor; MMTV, mouse mammary tumor virus.

The online version of this article contains supplemental material.

The Journal of Cell Biology, Vol. 177, No. 6, June 18, 2007 957–967
http://www.jcb.org/cgi/doi/10.1083/jcb.200703157

Supplemental Material can be found at:
http://www.jcb.org/cgi/content/full/jcb.200703157/DC1

spatial distribution of molecules associated with a transcriptionally active locus to construct a more unified picture of the nuclear organization of both these molecules and the chromatin at a transcription site. Using this approach, we report evidence for a previously undetected spatial organization at a transcription site, namely a domain of decondensed chromatin that borders or surrounds the transcription sites and appears likely to contain recently transcribed chromatin.

Results

Transcription sites and active pol II associate with the array and interdigitate between GFP-GR beads

We examined the location of transcription sites at the array by first using bromo-UTP (BrUTP) incorporation for detection of nascent transcripts. This consistently yielded a series of BrUTP puncta associated with the GFP-GR–tagged array (Fig. 1, a and b). As previously described, the array itself is composed of GFP-GR puncta or beads (Müller et al., 2004). To ascertain whether the BrUTP puncta overlaid the GFP-GR beads, we performed 3D deconvolution for improved resolution, including corrections for residual chromatic aberration along the optical axis. We consistently found that the BrUTP puncta did not directly colocalize with the GFP-GR beads but rather interdigitated between the beads, with some overlap at the edges of these two distributions (Fig. 1, c and d). These observations are consistent with earlier studies suggesting that transcription occurs predominantly at or near the surface of compact chromatin domains, namely in the interchromatin or perichromatin domains (Cmarko et al., 1999; Verschure et al., 1999).

Because the BrUTP incorporation procedure involves live cell permeabilization that might conceivably alter the relative distribution of transcription sites and GFP-GR beads, we used an alternate approach to address the same question. Fixed cells were probed with an antibody (H5) against the phosphorylated CTD domain of pol II to determine its association with GFP-GR at the array. This likewise yielded a punctate staining pattern for active pol II that was clearly enriched at the array (Fig. 1, e and f). This punctate pol II staining pattern is consistent with live cell images from a previous study that examined a GFP-tagged pol II in the array cell line (Becker et al., 2002). Again, using 3D deconvolution for improved resolution, we found that the active pol II, like the BrUTP incorporation sites, did not directly colocalize with the GFP-GR beads but interdigitated between them (Fig. 1, g and h).

Finally, we analyzed the degree of overlap between the active pol II and the BrUTP stains at the array and found a high degree of colocalization (Fig. 1, i–l). These observations identify the BrUTP puncta as transcription factories and are consistent with previous studies demonstrating transcription foci in various cell types (Jackson et al., 1993; Wansink et al., 1996). The pol II factories that we detected are larger than typical pol II factories but comparable in size with pol I factories (Hozak et al., 1994). This similarity in size may reflect the fact that the pol I factories also associate with a tandem array (in this case, of ribosomal genes). In summary, these results establish that the transcription sites at the array are located directly adjacent to the GFP-GR beads, with some overlap at the edges between these two distributions.

Decondensed chromatin from the MMTV array surrounds or is adjacent to the pol II transcription sites

It has been proposed that loops of highly decondensed chromatin extrude from transcription sites (Cook, 1995). If so, at the resolution afforded by light microscopy, each transcription site

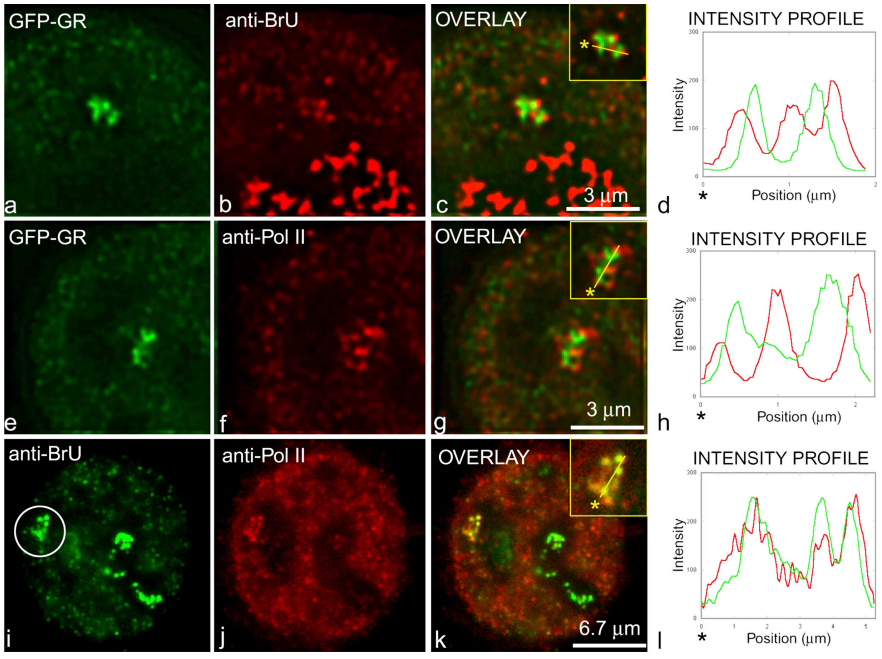


Figure 1. **Transcription sites interdigitate between GFP-GR beads.** (a–h) Transcription sites were visualized either by BrUTP incorporation followed by antibody detection of BrUTP (a–c) or by using an antibody against the active form of pol II (e–g). Images in a–c and e–g were deconvolved and corrected for chromatic aberration. Transcription sites at the array (b and f) are displaced from the GFP-GR beads, with occasional overlap at the edges of these distributions (c and g). Insets in c and g show the path starting at the asterisk over which red and green intensities were measured (d and h). (i and j) As a consistency check, triple-label analysis was performed using GFP-GR (not depicted) to identify the location of the array (circle in i), and immunofluorescence was performed with antibodies against BrUTP (i) and active pol II (j). (k and l) The antibody stains overlap considerably. The inset in k shows the path starting at the asterisk over which red and green intensities were measured (l). Note that as expected, pol I transcriptional activity in the nucleolus is marked by intense BrUTP incorporation (b and i). Because BrUTP levels in nucleoli are considerably higher than at the array, the nucleolar incorporation is saturated in these images to optimize visualization of the array signal.

at the array should be associated with a domain of decondensed chromatin. However, our previous DNA FISH experiments suggested that array chromatin exactly coincides with the GFP-GR beads (Fig. S1, available at http://www.jcb.org/cgi/content/full/jcb.200703157/DC1; Müller et al., 2001). This GFP-GR bead chromatin could, in principle, correspond to the predicted decondensed domain, but our previous estimates suggest it is considerably more condensed than expected for transcribed chromatin (Müller et al., 2001).

We reasoned that if additional, more decondensed chromatin was associated with transcription sites at the array, its fragility might make it difficult to preserve by our earlier procedure of DNA FISH with denaturation at 95°C (Müller et al., 2001). Thus, we performed DNA FISH at a lower denaturation temperature (70°C) and compared the results to DNA FISH with denaturation at 95°C. At 95°C, we once again detected beaded structures identical to those we had previously observed (Fig. 2 a). However, with denaturation at 70°C, we could also detect specific MMTV-labeled chromatin structures in every cell (Fig. 2 b). These structures contained some puncta that resembled the beads seen at 95°C, but the structures seen at 70°C also exhibited a haze interspersed between the puncta that was not as evident at 95°C. Furthermore, direct measurement of areas encompassed by the structures demonstrated that those detected at 70°C were significantly larger ($P < 10^{-6}$) than those detected at 95°C (Fig. 2 c). The 70°C structures were never detected in control experiments in which the specific DNA probe was omitted, although staining of the nuclear periphery and random spots within the nucleus was still apparent (Fig. S2 a, available at http://www.jcb.org/cgi/content/full/jcb.200703157/DC1). All of the specific structures detected by these two FISH protocols contain DNA, as an RNase treatment is always included in the DNA FISH procedures, and both the 70 and 95°C structures were eliminated by pretreating cells with DNase (unpublished data).

To determine whether there was any overlap between the DNA detected by the 70 and 95°C procedures, we devised a dual-temperature DNA FISH protocol that involved FISH at 70°C with a red-labeled probe followed by an additional fixation step to ensure preservation of the 70°C structure and FISH at 95°C with a green-labeled probe. This dual FISH procedure consistently enabled the preservation and detection of two distinct structures that showed virtually no overlap between the red (70°C) and green (95°C) labels in all cells (Fig. 2, d–i). This observation suggests that the 70 and 95°C structures are largely exclusive. Observation of many cells with the dual FISH procedure showed that the 70°C (red) structure typically surrounded the 95°C (green) structure (Fig. 2, d–f), although in a few cells, the 70°C (red) structure protruded largely from one side of the 95°C (green) structure (Fig. 2, g–i).

We also investigated whether we could reverse the preceding dual DNA FISH procedure; that is, we performed FISH first at 95°C and then at 70°C. In this reverse procedure, the decondensed domain was no longer detected (Fig. S2 b). This suggests that the chromatin within this domain is fragile and easily destroyed by 95°C treatment unless it is extensively prefixed.

The fragility of the 70°C structure suggests that it is more decondensed than the 95°C structure. Consistent with this, we found that the mean FISH intensity per unit area was approximately sevenfold higher in the 95 vs. 70°C structures, suggesting a considerable increase in DNA concentration within the GFP-GR beads relative to the decondensed domain. Note that

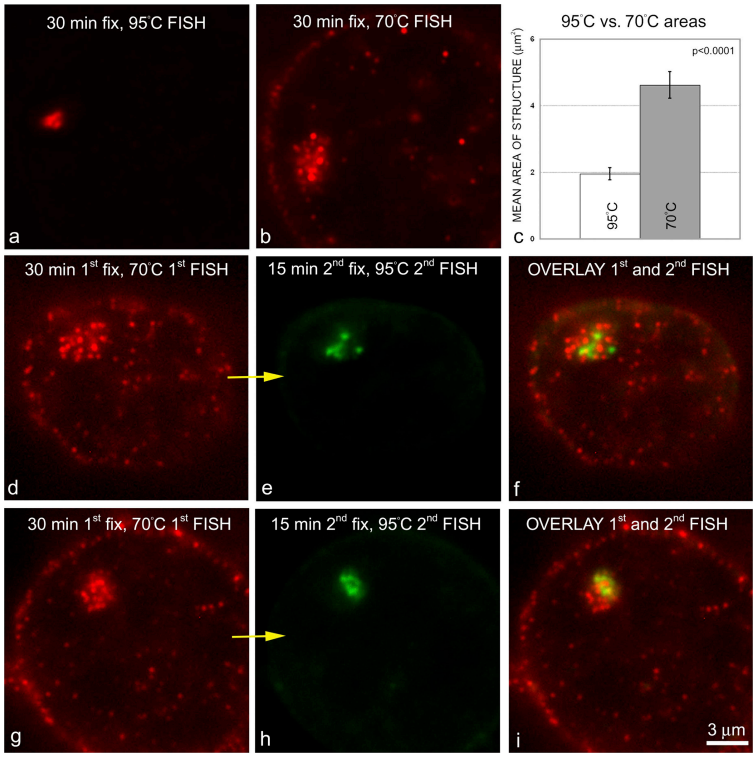


Figure 2. **Distinct array-specific structures are detected by DNA FISH at different denaturation temperatures.** (a) As previously reported (Müller et al., 2001, 2004), structures can be detected by DNA FISH with denaturation at 95°C that overlap the GFP-GR beads (Fig. S1, available at http://www.jcb.org/cgi/content/full/jcb.200703157/DC1). (b and c) However, with denaturation at 70°C (b), much larger structures are detected (c). (d–i) These 70 and 95°C structures are distinct and can be detected reproducibly in all cells using a double FISH protocol. Here, denaturation is first performed at 70°C, and the denatured DNA is labeled with a red probe (d and g). (e and h) This is followed by an additional fixation step (arrows) to preserve the more fragile 70°C structure, denaturation at 95°C, and the newly denatured DNA labeled with a green probe. (f and i) Double-label imaging of these cells demonstrates that in most cells, the 70°C structure surrounds the 95°C structure (f), although, occasionally, the 70°C structure extends largely to one side of the 95°C structure (i). Based on the fragility and estimated DNA density of the 70°C structure, we refer to it as the decondensed domain. Based on its overlap with the GFP-GR stain, we refer to the 95°C structure as either the GFP-GR beads or condensed domain. Error bars represent SEM.

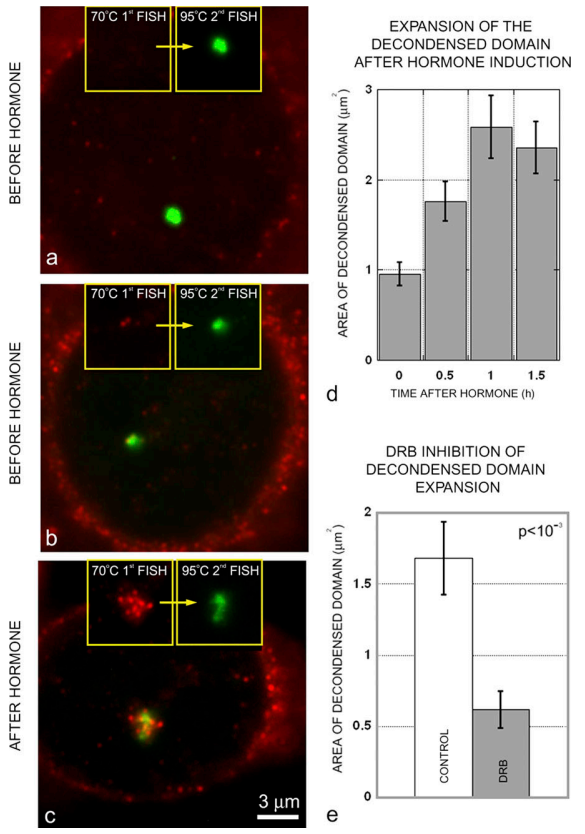


Figure 3. **The decondensed domain markedly expands upon transcriptional activation, and this expansion is inhibited by DRB.** (a and b) Consistent with low levels of basal expression from the MMTV promoter, decondensed domains were either absent (a) or very small (b) in cells before hormone addition. (c) Much larger decondensed domains were present 1.5 h after hormone addition. (a–c) The overlay images of the double FISH procedure are shown, with the insets showing separately the first (70°C; red) followed by (arrows) the second (95°C; green) steps of the double FISH. Areas of the decondensed domains were measured by thresholding the edge of the structure. (d and e) The mean area increased over time after hormone induction of transcription (d), but this increase could be significantly reduced ($P < 10^{-3}$) by the drug DRB, which prevents pol II elongation (e). Error bars represent SEM.

this is a rough approximation because the measurements were made in 2D instead of 3D and because of the possibility that the 70 and 95°C FISH protocols may have different efficiencies of DNA preservation or hybridization.

We conclude that there are two chromatin compartments at the array: a more condensed domain that corresponds directly to the GFP-GR beads surrounded by or adjacent to a more decondensed domain detectable only by lower temperature DNA FISH.

Formation of the decondensed domain requires transcription and topoisomerase IIα activity

We next investigated whether the decondensed domain arose as a result of transcription. To test this, we measured areas of the decondensed domain detected by DNA FISH at 70°C as a function of time before and after transcriptional activation by hormone induction. Before activation and consistent with the known low levels of basal transcription from the MMTV promoter (Toohey et al., 1990), small decondensed domains were visible in some cells, whereas in other cells, none could be detected (Fig. 3, a and b). In contrast, a single chromatin bead could always be detected by DNA FISH at 95°C, marking the site of the condensed array (Fig. 3, a and b). After activation, decondensed domains were present in every cell (Fig. 3 c), and their mean area increased substantially over time (Fig. 3 d).

As a second test of the decondensed domain's association with transcription, we induced transcription by the addition of hormone but simultaneously added a transcriptional inhibitor, DRB (5,6-dichloro-1-β-D-ribobenzimidazole; Chodosh et al., 1989). This significantly inhibited ($P < 10^{-3}$) formation of the decondensed domain (Fig. 3 e), also suggesting that formation of this domain is coupled to transcription.

As another test for the possible involvement of the decondensed domain in transcription, we investigated its association with a topoisomerase. Transcription generates positive supercoils in front of a polymerase and negative supercoils behind it (Liu and Wang, 1987). If not relieved by topoisomerase action, the resultant torsional strain may accumulate to levels that could stall transcription (Mondal et al., 2003).

We stained the array cell line with two different antibodies against topoisomerase IIα. For each antibody, we detected a similar association pattern with the array: a region of topoisomerase IIα staining extended around and beyond the GFP-GR beads (Fig. 4, a–c).

To determine the relationship of the topoisomerase IIα staining pattern with the decondensed domain, we performed immuno-FISH and found that the topoisomerase II stain and the decondensed domain consistently overlapped (Fig. 4, d–f). These results suggest that topoisomerase IIα associates with the decondensed domain and may perform some function there.

To test this, we inhibited topoisomerase IIα using the drug etoposide. We found that formation of the decondensed domain was impaired (Fig. 4 g) compared with controls in which cells were treated with vehicle only. As detected by RNA FISH, etoposide treatment also sharply reduced transcription from the array compared with the controls (Fig. 4 h). These results indicate that transcription sites at the array are associated with a surrounding region of topoisomerase IIα that is required both for transcription from the array and for formation of the decondensed domain around the array.

The decondensed domain is enriched in trimethyl H3K36, a histone mark characteristic of recent transcription

Several studies in both yeast and mammals have demonstrated that histones in recently transcribed chromatin are marked with a trimethyl H3K36 modification (Bernstein et al., 2005; Morris et al., 2005; Pokholok et al., 2005; Vakoc et al., 2006). We reasoned that if the decondensed domain contains recently transcribed chromatin extruded from the pol II factories, it should show increased levels of trimethyl H3K36.

To determine whether this mark was associated with transcribed chromatin from the MMTV array, we performed chromatin immunoprecipitation (ChIP) using an antibody specific for trimethyl H3K36 and compared the levels of this mark

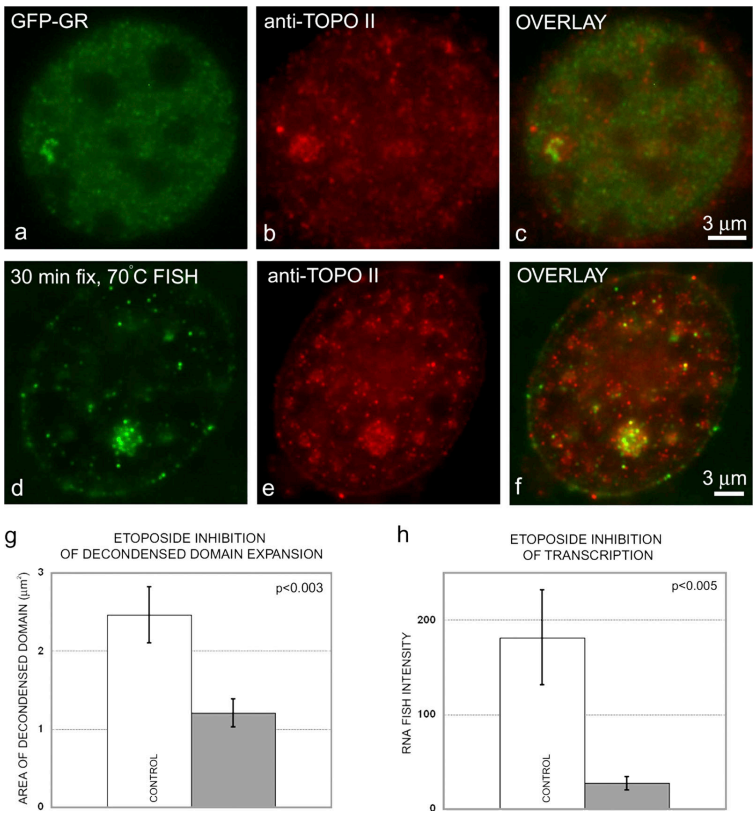


Figure 4. **Topoisomerase IIα associates with the decondensed domain and is required for expansion of the decondensed domain.** (a–c) Immunofluorescence reveals that topoisomerase IIα is distributed in a domain surrounding the GFP-GR–tagged array. (d–f) An immuno-FISH procedure demonstrates that the topoisomerase IIα stain colocalizes with the decondensed domain. (g) Expansion of the decondensed domain upon transcriptional activation is inhibited by etoposide. (h) As detected by RNA FISH intensity measurements, etoposide also dramatically reduces transcription from the array. Error bars represent SEM.

within the MMTV promoter to the downstream ras reporter gene. Consistent with previous studies of other genes (Bernstein et al., 2005; Morris et al., 2005; Pokholok et al., 2005; Vakoc et al., 2006), we found that compared with the MMTV promoter, the reporter gene sequence exhibited a substantial enrichment for the trimethyl H3K36 mark. This differential effect was enhanced upon the hormone induction of transcription but was still detected to a lesser degree without hormone (Fig. 5 a), which is consistent with basal transcription from the MMTV promoter (Toohey et al., 1990) and with our unpublished observations of RNA FISH accumulation at the array in the absence of hormone.

With this evidence for trimethyl H3K36 enhancement in the transcribed reporter gene sequence, we proceeded to examine the distribution of this mark at the MMTV array by confocal microscopy. Immunofluorescence with the same trimethyl H3K36 antibody used for ChIP revealed a staining pattern that surrounded the GFP-GR beads (Fig. 5, b–d). To follow up this observation, we also performed immunofluorescence with an antibody against the N terminus of the human huntingtin-interacting protein B (HYPB), which possesses H3K36 histone methyltransferase activity (Sun et al., 2005) and is an orthologue of the Set2 methyltransferase responsible for the H3K36 trimethylation mark in yeast (Strahl et al., 2002). This HYPB antibody also exhibited a staining pattern that surrounded the array (Fig. 5, e–g), suggesting that the trimethyl H3K36 mark itself as well as an enzyme potentially responsible for it were associated with the decondensed domain.

In contrast, strikingly different staining patterns were observed in confocal images of antibodies directed against histone marks typically associated with active promoters and 5′ regions (Liang et al., 2004; Schneider et al., 2004; Bernstein et al., 2005; Pokholok et al., 2005; Roh et al., 2005). Of the three antibodies tested (generically acetylated H4, trimethyl H3K4, and acetyl H3K9), all stained the condensed chromatin domain, yielding substantial colocalization with the GFP-GR beads, but showed little or no stain of the decondensed domain (Fig. 6).

The enhanced staining of the GFP-GR beads by antibodies specific for active promoters could reflect a preferential retention of 5′ sequences in the condensed domain compared with the decondensed domain or, alternatively, could reflect the fact that there is likely to be considerably more chromatin within the condensed domain compared with the decondensed domain (approximately seven times more based on our rough estimates; see Decondensed chromatin from the MMTV array…).

To distinguish between these possibilities, we performed DNA FISH at both 70 and 95°C with a probe for the MMTV promoter sequence. These FISH experiments revealed that this sequence was present in both the condensed (95°C FISH) and decondensed (70°C FISH) domains (Fig. 7, a, b, and e–g). Thus, promoter sequences do not appear to be preferentially retained within the condensed domain. Consequently, the enhancement of active promoter marks and GFP-GR staining in the condensed domain most likely reflects the increased chromatin concentration there.

We then repeated these probe-specific FISH experiments, but with a probe for the ras reporter gene sequence. Here, as for

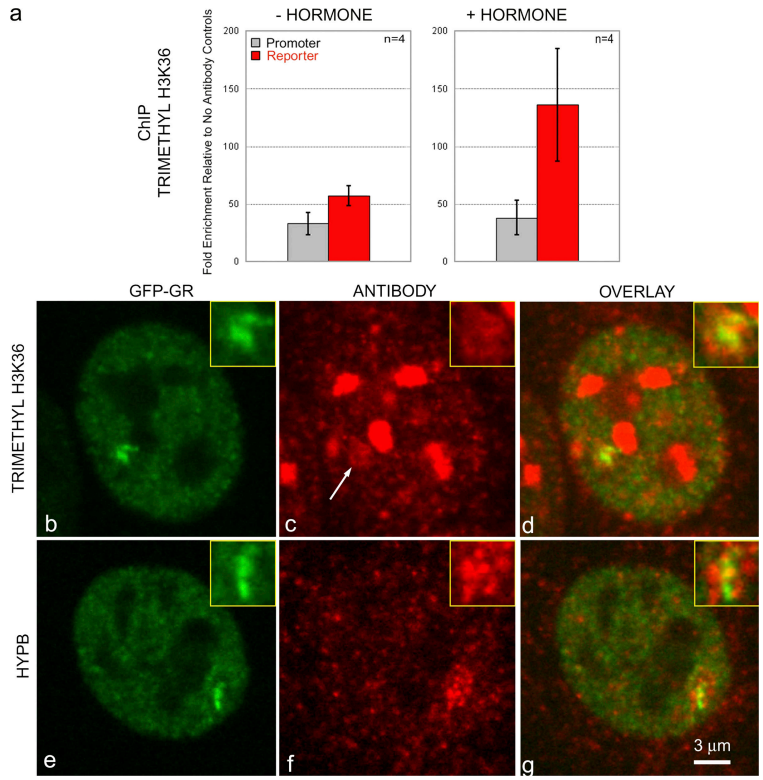


Figure 5. **A histone modification associated with transcribed regions is found in the decondensed domain.** Consistent with published studies of other genes in yeast and mammalian systems, ChIP reveals that the trimethyl H3K36 histone mark is enhanced in the transcribed region of the array relative to the promoter. (a) Transcribed region, red bars; promoter, gray bars. The mean fold enrichments with SEM (error bars) are shown from four separate chromatin isolations with and without hormone induction. Before hormone treatment (−hormone), some enhancement of the trimethyl H3K36 mark was reproducibly seen in the reporter relative to the promoter. After hormone treatment (+hormone), the enhancement of the trimethyl H3K36 mark in the reporter increased substantially in all experiments. (b–d) Immunofluorescence with the same trimethyl H3K36 antibody yields by confocal microscopy an enhanced stain (arrow in c) that surrounds the GFP-GR beads (b–d; with higher magnification views in the insets), suggesting that this is the domain containing recently transcribed sequences from the array. Note that a darker footprint corresponding to the location of the GFP-GR beads is present in the trimethyl H3K36 stain (compare insets in b and c), suggesting that this mark is largely excluded from the beads even though there is much more chromatin there. Also note that the trimethyl H3K36 antibody yields intense staining of nucleoli. This might conceivably reflect transcribed rDNA sequences, although there are no reports to date of whether transcribed pol I genes are also marked in this way. (e–g) An antibody against a histone methyltransferase thought to be responsible for the H3K36 methylation mark (HYPB, a Set2 orthologue) also stains a region surrounding the GFP-GR beads, but no comparable staining is seen in nucleoli, suggesting that another methyltransferase might be responsible for the H3K36 staining within nucleoli.

the promoter sequences, we could also detect reporter sequences in both the condensed and decondensed domains (Fig. 7, c, d, and h–j). Thus, despite both the increased chromatin concentration and the presence of reporter gene sequences in the condensed domain, staining for the trimethyl H3K36 mark is not enhanced in the condensed GFP-GR beads but rather only in the decondensed domain. Because the trimethyl H3K36 mark labels recently transcribed chromatin, this result argues (1) that the reporter gene sequences in the condensed domain have not yet been transcribed and (2) that as these sequences are transcribed, they appear in the decondensed domain.

Discussion

Overview

Structural analysis of transcriptionally active chromatin is challenging as a result of difficulties in identifying, preserving, and resolving the structures at such sites. We have overcome some of these limitations in this study by developing a new protocol for DNA FISH and applying it to a tandem gene array that is easily visualized by light microscopy. With these tools, we have now identified three different structures at the array that provide new insights into how transcription may occur there (Fig. 8). First, as we previously described (Müller et al., 2004), we find a series of adjacent puncta or beads of relatively condensed chromatin that can be identified by either conventional DNA FISH or in live or fixed cells by the accumulation of GFP-GR. Second, directly adjacent to this condensed domain, we find transcription sites identified by either BrUTP incorporation or by an antibody against the active form of pol II. Third, we find that these transcription sites are surrounded by and contained within a larger domain that is composed of more decondensed chromatin from the array. As explained below, our results suggest that this decondensed domain arises from the extrusion of transcribed sequences from an immobilized polymerase, providing new support to the pol II factory model of transcription.

Structural evidence for a decondensed domain

Some hints for chromatin-surrounding transcription sites had previously come from electron microscopy sections of HeLa cells in which transcription sites were detected by biotinylated RNA, and the presence of chromatin surrounding them was inferred by a uranyl-EDTA regressive staining technique (Iborra et al., 1996). Because this procedure detects all transcription sites and all chromatin, some amount of interpretation was required to imagine where the associated chromatin might begin and end for each transcription site or even whether the chromatin adjacent to a transcription site was composed of DNA associated with that site.

Our new evidence for a decondensed chromatin domain surrounding transcription sites at the array is more direct and substantial. Using a specially developed, gentler DNA FISH protocol, we were able to detect decondensed, array-specific chromatin extending to a clear boundary around only the transcription sites associated with the array. This demonstrates that a specific set of transcription sites is surrounded by a decondensed chromatin domain composed of sequences from the loci being transcribed. It seems likely that we and others have missed such decondensed domains before by DNA FISH

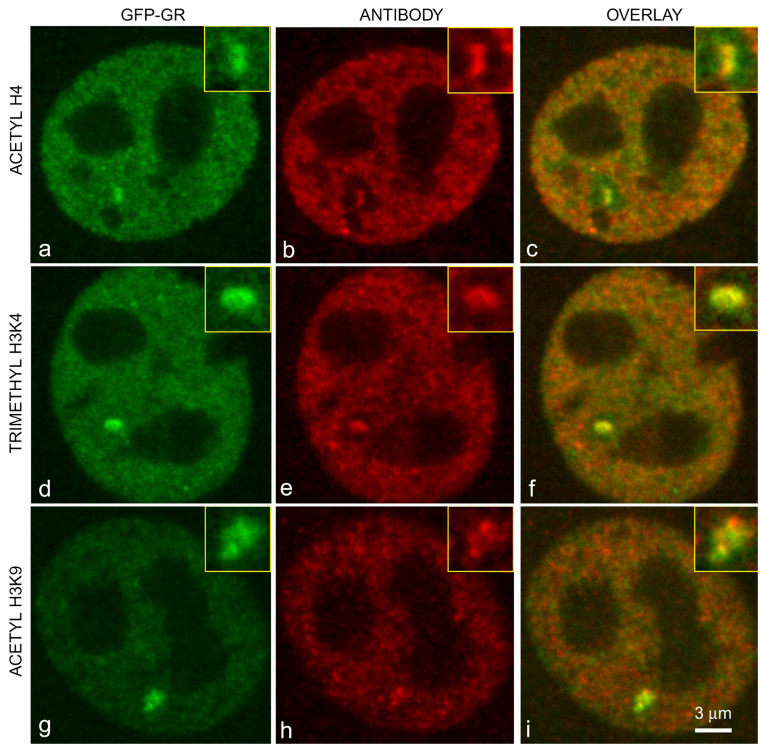


Figure 6. **The GFP-GR beads colocalize with histone modifications that are typically associated with promoters and the 5′ regions of transcribed genes.** Confocal microscope images show GFP-GR (green) and immunofluorescent images (red) with antibodies against generically acetylated H4 (a–c), trimethyl H3K4 (d–f), and acetyl H3K9 (g–i). Insets show higher magnification views of the array.

because they are difficult to preserve, are normally composed of a variety of different DNA sequences dependent on the genes being transcribed at the transcription site (Osborne et al., 2004), and are likely to be much smaller for a transcription site associated with single-copy genes of moderate transcriptional activity.

According to the simplest form of the pol II factory model, the decondensed domain surrounding a transcription site should be composed of loops of decondensed chromatin. Our light microscopy images cannot resolve such structures, but our DNA FISH detection procedure yields a punctate staining pattern in the decondensed domain that might reflect a more complex structural organization there. However, given the fragility of the chromatin within this domain, some alteration of fine structure might be expected after the fixation and denaturation procedures used to detect it. A live cell marker for the decondensed domain (analogous to GFP-GR for the beads) will be necessary to draw any firm conclusions with light microscopy about the substructure of this domain.

Evidence for involvement of the decondensed domain in transcription

We made several observations linking the array's decondensed domain with transcription. DRB treatment, which blocks transcriptional elongation (Chodosh et al., 1989), hinders formation of the decondensed domain, suggesting that transcriptional elongation is required for the formation of the decondensed domain. We also found that topoisomerase IIα associates with the decondensed domain and so is poised to remove supercoils that would arise on either side of a transcribing polymerase (Liu and Wang, 1987). Inhibition of topoisomerase II function by a brief (45 min) drug treatment impaired formation of the decondensed domain and dramatically reduced transcription. Both effects could arise if the drug treatment blocked the elongation of pol II either as a result of accumulated torsional strain or immobilized topoisomerase complexes, although effects of topoisomerase inhibition on promoter activation are also possible (Collins et al., 2001).

More direct molecular evidence for the role of the decondensed domain in elongation comes from the presence within the decondensed domain of a marker, trimethyl H3K36, which is characteristically found at multiple sites along transcribed genes (Bannister et al., 2005; Morris et al., 2005; Pokholok et al., 2005; Vakoc et al., 2006). Indeed, we found by ChIP that the trimethyl H3K36 mark is enhanced in the transcribed reporter sequences of the array compared with the promoter sequence. This indicates that when used in immunofluorescence, the same trimethyl H3K36 antibody should reveal the location and distribution of transcribed sequences at the array. This antibody stained a region that surrounded and extended well beyond the GFP-GR beads, suggesting that chromatin within the decondensed domain was recently transcribed. However, the BrUTP incorporation experiments demonstrate that transcription occurs only at the transcription sites directly adjacent to the GFP-GR beads (Fig. 1). Thus, it appears that transcribed sequences from the array do not remain at the transcription sites but instead are extruded into the surroundings, giving rise to the decondensed domain (Fig. 2).

The trimethyl H3K36 mark was unique among the histone modifications that we tested because it was the only one that showed enhanced staining that surrounded the GFP-GR beads. In contrast, active promoter marks instead showed enhancement

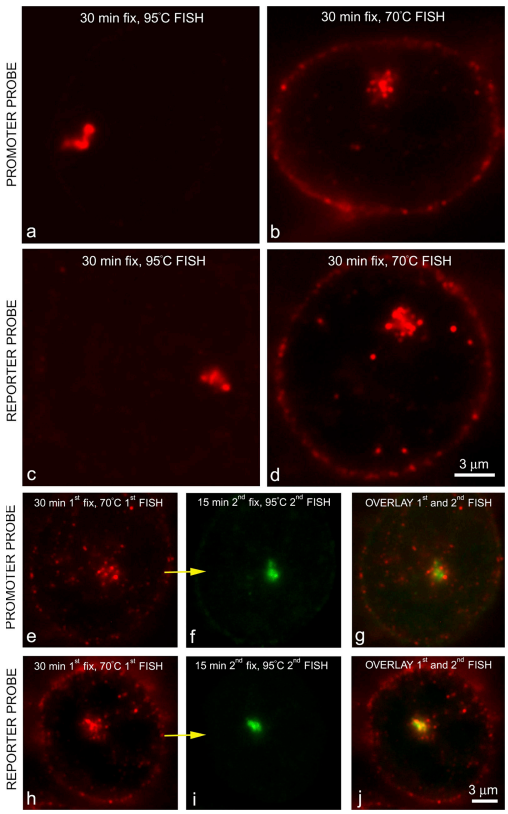


Figure 7. **Promoter and reporter sequences are present in both the condensed and decondensed domains.** Similar condensed (95°C FISH; a and c) and decondensed (70°C FISH; b and d) domain structures are detected with DNA probes specific for either the array's MMTV promoter (a and b) or ras reporter gene (c and d). (e–j) The same is true when the double FISH procedure is used, demonstrating that promoter and reporter sequences are present in both the condensed and decondensed domains. Arrows (e to f and h to i) indicate that the 70°C FISH is followed by the 95°C FISH.

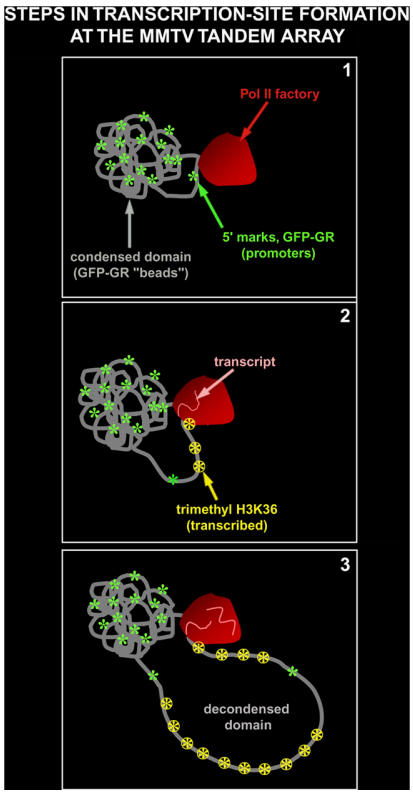


Figure 8. **A model for patterns of chromatin organization and histone modifications at the MMTV tandem array.** Upon hormone stimulation, GFP-GR binds to MMTV promoters within the condensed domain, thereby defining the GFP-GR beads visible in live cells. Some of these GFP-GR–bound promoters associate with transcription factories (step 1). A polymerase within the factory transcribes the downstream reporter sequence, extruding it into the surroundings with multiple trimethyl H3K36 marks attached (step 2). The tandem nature of the array promotes iteration of this process, thereby giving rise to large decondensed domains easily visible by light microscopy (step 3). Promoter marks (green) are enhanced within the condensed domain as a result of the density of chromatin there, but the trimethyl H3K36 mark is not detectable there (Fig. 5, b–d), suggesting that transcribed sequences are not found in the condensed domain. Rather, these transcribed marks are found only in the decondensed domain, which extends well beyond the sites of transcription, implying that the transcribed sequences are extruded from the transcription site.

within the GFP-GR beads. All of these marks for promoters, including GFP-GR, are probably higher within the GFP-GR beads because the chromatin concentration is substantially higher there than in the decondensed domain (approximately seven times more based on our rough estimate). We could find no evidence for the alternate possibility that promoter sequences are preferentially enriched within the beads because DNA FISH with a promoter-specific probe demonstrated that promoter sequences were present not only in the GFP-GR beads but also in the decondensed domain.

Despite the increased chromatin concentration within GFP-GR beads leading to the enhancement of 5′ marks there, no such enhancement was detected for the 3′ trimethyl H3K36 mark. However, DNA FISH with a reporter probe showed that reporter sequences were also present within both the GFP-GR beads and the surrounding decondensed domain. Thus, the lack of trimethyl H3K36 staining in the GFP-GR beads indicates that the reporter sequences within this condensed domain have not been transcribed. Because multiple trimethyl H3K36 marks are placed on each segment of transcribed chromatin, the substantial enhancement of this mark relative to any other histone modification in the decondensed domain is expected if this domain contains recently transcribed chromatin.

Together, these results suggest a model for transcription site formation at the MMTV array (Fig. 8). Promoter regions within the condensed domain are bound by GFP-GR, resulting in its visibility within live cells as the GFP-GR beads. Some of these GFP-GR–bound promoters then associate with pol II transcription factories. This leads to production within the pol II factory of nascent transcripts from the downstream reporter gene accompanied by deposition of the trimethyl H3K36 mark at multiple sites along the reporter gene. The transcribed sequences are extruded from the pol II factory, producing the decondensed domain and an enrichment of the trimethyl H3K36 mark in this region. The tandem nature of the gene array favors iteration of this process at consecutive promoters, thereby leading to a large decondensed domain visible by light microscopy.

Limitations of this study and future prospects

Our conclusions here are based on the premise that the array exploits the normal cellular transcription machinery, thereby yielding structural features at a transcription site that are amplified versions of those that occur at endogenous loci. Considerable biochemical evidence indicates that transcription occurs normally from the array, and further evidence suggests that the bead chromatin structure of the array also occurs in natural chromatin (see Introduction). However, the sequence composition at endogenous loci is less gene dense and more complex than the simple, repetitive nature of the array. Whether endogenous genes exhibit comparable decondensed domains can now be assayed using the new, gentler DNA FISH that we have developed here. Although we would predict that decondensed domains at endogenous loci should be considerably smaller, some may still be detectable by light microscopy if long transcripts are encoded.

Although our results provide new structural evidence for the immobilized pol II factory model, they do not provide definitive proof that transcripts actually move rather than the polymerase. This might be tested directly in the future if in vivo marks for the decondensed chromatin domain can be developed enabling time-lapse 3D imaging to assess whether chromatin within the decondensed domain moves in and out of the pol II transcription site.

In summary, we have identified a new decondensed chromatin domain surrounding transcription factories. This domain requires transcription for its formation and shows enrichment for a histone modification that is characteristic of recently transcribed chromatin. These observations provide new support for the immobilized pol II factory model and, in addition, suggest further tests of this model at both the tandem array and endogenous loci.

Materials and methods

Cell culture

The MMTV array cell line (3617) was grown as previously described (Müller et al., 2001). For microscopy experiments, cells were grown on #1.5 coverslips. To induce GR-mediated transcription from the MMTV array, 100 nM dexamethasone was added to cells for 0.5–1.5 h.

BrUTP incorporation

The protocol followed that in Elbi et al. (2002), with the following modifications. The permeabilization buffer contained 25 μg/ml instead of 5 μg/ml digitonin, 1 mM PMSF instead of 0.5 mM PMSF, and 100 nM dexamethasone. The transcription buffer contained 10 mM $MgCl_2$ instead of 5 mM, and the transcription reaction was run for 15 min at room temperature.

Immunofluorescence

Cells were fixed in either 3.5% PFA in PBS for 20 min followed by 0.5% Triton X-100 in PBS for 10 min or in 0.5% formaldehyde in PEM buffer (100 mM Pipes, 5 mM EGTA, 2 mM $MgCl_2$, pH 6.8, and 0.2% Triton X-100) for 5 min. The former fix tended to give more intense staining patterns for markers associated with the GFP-GR beads, whereas the latter fix tended to give more intense staining patterns for markers associated with the decondensed domain, although the pattern of staining itself was not dependent on the fixation protocol. Before antibody incubation, cells were washed three times for 10 min each in PBS.

The primary antibodies used were as follows: anti-BrdU mouse monoclonal (Caltag) or anti-BrdUTP rat monoclonal (Oxford Biotechnology); anti–topoisomerase IIα (Topogen); anti–pol II H5 (Covance); anti–human Set2 orthologue (HYPB) N terminus (Abgent); anti–trimethyl H3K36, anti–acetyl H4, and anti–acetyl H3K9 (Upstate Biotechnology); and anti–trimethyl H3K4 (Abcam).

The secondary antibodies used were as follows: Texas red–conjugated anti–mouse and anti–rabbit, rhodamine-conjugated anti–rat (Rockland), and Cy5-conjugated anti–mouse (Jackson ImmunoResearch Laboratories). Antibodies were diluted in PBS with 4% BSA and 0.1% Tween 20. Primary antibodies were incubated overnight at 4°C. After incubation, washes were performed three times with PBS for 20–30 min total. In some cases, the first wash also contained 0.1% Tween, and the incubation time was reduced to 5 min. Secondary antibodies were incubated from 40–60 min at room temperature and washed three times in PBS for a total of 20 min.

Array-specific DNA FISH

Decondensed domain–specific fixation and denaturation. Cells were fixed for 30 min by adding an equal volume of 7.0% PFA in PBS to the DME culture media. Improved staining was often achieved when this fix was preceded by a 5-min prefix in 0.5% formaldehyde in PEM buffer. Cells were then washed three times with PBS for 10 min each, permeabilized for 10 min with 0.5% Triton X-100 in PBS, and washed with PBS again. Then, cells were incubated in 50 μg/ml RNase for 30–60 min and washed three times in PBS for 10 min each. DNA was denatured by incubation at 70°C for 10 min in 70% formamide in 2× SSC followed by dehydration for 2–5 min each in 70, 90, and 100% ethanol kept on ice.

Condensed domain–specific fixation and denaturation. This was identical to the decondensed domain protocol described in the previous paragraph except that cells were fixed for 30 min with the 3.5% PFA fix described above, and denaturation was performed for 5 min at 95°C.

Probe preparation and hybridization. Three types of probe-specific DNA were used: array, promoter, and reporter. The array-specific probe was prepared as previously described (Müller et al., 2001), but with the following modifications: the biotin and digoxigenin nick translation mix was purchased from Roche, and the entire pM18 plasmid (Ostrowski et al., 1983) was used as a template. The promoter-specific probe was a 1.9-kb BstX1–BamH1 fragment of the pM18 plasmid. The reporter-specific probe was a 2-kb BamH1–Sal1 fragment of the pM18 plasmid. Hybridization was also performed essentially as previously described (Müller et al., 2001) except that the dextran sulfate concentration of the hybridization mix was reduced to 5%. The hybridized probe was detected as follows: for condensed domain FISH, probes were detected with streptavidin AlexaFluor488 (Invitrogen), whereas for decondensed domain FISH, probes were detected with an antidigoxigenin (sheep) primary antibody (Roche) followed by an AlexaFluor594-conjugated anti–sheep secondary antibody (Invitrogen).

Double FISH. The first stage of this protocol followed that for the decondensed domain–specific fixation and denaturation procedure, and the detection protocol for this first stage followed that for probe preparation and hybridization. Then, cells were fixed a second time for 15–20 min in 3.5% PFA in PBS, washed in PBS, permeabilized for 10 min in 0.5% Triton X-100, and washed with PBS again. DNA was then denatured for the second time according to the condensed domain–specific fixation and denaturation protocol. The second detection step followed that for aforementioned probe preparation and hybridization.

RNA FISH

RNA FISH was performed as previously described (Müller et al., 2001) except that cells were fixed for 30 min with 3.5% PFA in PBS, and the hybridized probe was detected with streptavidin AlexaFluor488 (Invitrogen).

Drug inhibition experiments

DRB. Transcription was induced with 100 nM dexamethasone, and DRB (Calbiochem) was added simultaneously at 100 μg/ml (from a 1-mg/ml stock solution in water dissolved by heating). After a 45-min incubation, the cells were prepared for decondensed domain–specific FISH.

Etoposide. Transcription was induced with 100 nM dexamethasone, and etoposide (Sigma-Aldrich) was added simultaneously at 250 μM (from a 500-mM stock solution in DMSO). After a 45-min incubation, the cells were prepared for decondensed domain–specific DNA FISH. The same protocol was used for RNA FISH measurements.

Microscopy

Images of BrUTP incorporation and active pol II were obtained by 3D deconvolution microscopy of PFA-fixed specimens mounted in PBS. Images were collected with a CCD camera (CoolSNAP HQ; Photometrics) mounted

on a microscope (IX70; Olympus) equipped with a 100× 1.35 NA oil immersion objective (Olympus). Voxel sizes were set at 0.07 × 0.07 × 0.07 μm with 16–32 focal planes. Images were collected and also preprocessed to correct for photobleaching using the softWoRx package (Applied Precision), and the maximum likelihood algorithm from the publicly available XCOSM software was run for 200 iterations.

For colocalization analysis, these deconvolved images were corrected for chromatic aberration. This was calibrated by imaging a 0.5-μm TetraSpeck multicolored fluorescent bead (Invitrogen). Using this bead, we found that xy shifts from the GFP channel to the Texas red channel were less than a pixel (0.07 μm) and so were not corrected, but a z shift of approximately two focal planes (0.14 μm) was present and subsequently corrected.

Immunofluorescence images of histone modification patterns were collected with a spinning disk confocal microscope (Ultraview LCI CSU10; PerkinElmer) mounted on a microscope (Axiovert 200; Carl Zeiss MicroImaging, Inc.) equipped with a 63× 1.4 NA objective.

DNA FISH images were acquired on an upright microscope (DMRA; Leica) with a 100× 1.3 NA oil-immersion objective (Leica). Images were obtained with a CCD camera (Sensys; Roper Scientific).

All specimens were mounted in PBS for imaging. Images were cropped within MetaMorph software (Molecular Devices). Overlay images were generated using Imaris (Bitplane AG). Figures were assembled in Photoshop (Adobe), in which contrast adjustments were also performed by setting minimum and maximum intensity values in each color and displaying a linear contrast range between these endpoints.

Image measurements

All image measurements were performed with MetaMorph software (Molecular Devices). RNA FISH intensities and mean areas of structures were determined as previously described (Müller et al., 2001).

ChIP and real-time PCR

Cells were treated with either vehicle or 100 nM dexamethasone for 60 min and were processed for ChIP using a ChIP assay kit (Upstate Biotechnology). In brief, cells were cross-linked for 10 min at 37°C in 0.5% formaldehyde followed by a quenching step for 10 min with 150 mM glycine. Soluble chromatin was immunoprecipitated with an antibody to a trimethyl H3K36 mark (Upstate Biotechnology) with 12 μl of antibody per reaction. DNA isolates from immunoprecipitates were used as templates for real-time quantitative PCR amplification.

Real-time assays were conducted on a real-time detection system (iCycler IQ; Bio-Rad Laboratories) using the intercalation dye SYBR green as the fluorescence agent (iQ SYBR Green Supermix; Bio-Rad Laboratories) and the manufacturer's recommended conditions. PCR was performed by denaturing at 95°C for 15 s and annealing/extending at 60°C for 60 s. Standard curves were created for each run using a plasmid (pM18) that contained the MMTV long terminal repeat and primers that spanned either the promoter or reporter regions of the MMTV array. 10-fold serial dilutions of pM18 (over three logs) were used to generate the standard curve. All PCR reactions were subjected to a melting curve to verify the integrity of the PCR product and to eliminate amplification of nonspecific products. The following primers were used for amplification: promoter primers, sense (5′-TTTCCATACCAAGGAGGGGACAGTG-3′) and antisense (5′-CTTACTTAAGCCTTGGGAACCGCAA-3′); reporter primers, sense (5′-CGTGAGATTCGGCAGCATAAA-3′) and antisense (5′-GACAGCACACACTTGCAGCTC-3′).

In each of the ChIPs, all Ct (threshold cycle) values were normalized to promoter primers with no antibody. The fold enrichment at the reporter or promoter was then calculated by dividing each immunoprecipitation value by the respective no-antibody value. Both reporter and promoter primers amplified input (genomic) DNA with similar efficiencies.

Online supplemental material

Fig. S1 shows immuno-FISH at 95°C with a GR antibody. Fig. S2 shows 70°C FISH controls (no probe DNA or 95°C pretreatment). Online supplemental material is available at http://www.jcb.org/cgi/content/full/jcb.200703157/DC1.

We are grateful to Dr. Valarie Barr for use of the Laboratory of Cellular and Molecular Biology microscopy core and thank Dr. Keji Zhao for comments on the manuscript.

This work was supported by the intramural program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research, the Human Frontier Sciences Program, the Genome Research in Austria project Bioinformatic Integration Network, and the Austrian Ministry for Education, Science, and Culture.

Submitted: 26 March 2007
Accepted: 17 May 2007

References

Bannister, A.J., R. Schneider, F.A. Myers, A.W. Thorne, C. Crane-Robinson, and T. Kouzarides. 2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* 280:17732–17736.

Becker, M., C. Baumann, S. John, D.A. Walker, M. Vigneron, J.G. McNally, and G.L. Hager. 2002. Dynamic behavior of transcription factors on a natural promoter in living cells. *EMBO Rep.* 3:1188–1194.

Bernstein, B.E., M. Kamal, K. Lindblad-Toh, S. Bekiranov, D.K. Bailey, D.J. Huebert, S. McMahon, E.K. Karlsson, E.J. Kulbokas III, T.R. Gingeras, et al. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 120:169–181.

Chakalova, L., E. Debrand, J.A. Mitchell, C.S. Osborne, and P. Fraser. 2005. Replication and transcription: shaping the landscape of the genome. *Nat. Rev. Genet.* 6:669–677.

Chodosh, L.A., A. Fire, M. Samuels, and P.A. Sharp. 1989. 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *J. Biol. Chem.* 264:2250–2257.

Cmarko, D., P.J. Verschure, T.E. Martin, M.E. Dahmus, S. Krause, X.D. Fu, R. van Driel, and S. Fakan. 1999. Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Mol. Biol. Cell*. 10:211–223.

Collins, I., A. Weber, and D. Levens. 2001. Transcriptional consequences of topoisomerase inhibition. *Mol. Cell. Biol.* 21:8437–8451.

Cook, P.R. 1995. A chromomeric model for nuclear and chromosome structure. *J. Cell Sci.* 108:2927–2935.

Cook, P.R. 1999. The organization of replication and transcription. *Science*. 284:1790–1795.

Elbi, C., T. Misteli, and G.L. Hager. 2002. Recruitment of dioxin receptor to active transcription sites. *Mol. Biol. Cell*. 13:2001–2015.

Fragoso, G., W.D. Pennie, S. John, and G.L. Hager. 1998. The position and length of the steroid-dependent hypersensitive region in the mouse mammary tumor virus long terminal repeat are invariant despite multiple nucleosome B frames. *Mol. Cell. Biol.* 18:3633–3644.

Hozak, P., P.R. Cook, C. Schofer, W. Mosgoller, and F. Wachtler. 1994. Site of transcription of ribosomal RNA and intranucleolar structure in HeLa cells. *J. Cell Sci.* 107:639–648.

Iborra, F.J., and P.R. Cook. 2002. The interdependence of nuclear structure and function. *Curr. Opin. Cell Biol.* 14:780–785.

Iborra, F.J., A. Pombo, D.A. Jackson, and P.R. Cook. 1996. Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J. Cell Sci.* 109:1427–1436.

Jackson, D.A., A.B. Hassan, R.J. Errington, and P.R. Cook. 1993. Visualization of focal sites of transcription within human nuclei. *EMBO J.* 12:1059–1065.

Liang, G., J.C. Lin, V. Wei, C. Yoo, J.C. Cheng, C.T. Nguyen, D.J. Weisenberger, G. Egger, D. Takai, F.A. Gonzales, and P.A. Jones. 2004. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc. Natl. Acad. Sci. USA*. 101:7357–7362.

Liu, L.F., and J.C. Wang. 1987. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA*. 84:7024–7027.

McNally, J.G., W.G. Müller, D. Walker, R. Wolford, and G.L. Hager. 2000. The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science*. 287:1262–1265.

Mondal, N., Y. Zhang, Z. Jonsson, S.K. Dhar, M. Kannapiran, and J.D. Parvin. 2003. Elongation by RNA polymerase II on chromatin templates requires topoisomerase activity. *Nucleic Acids Res.* 31:5016–5024.

Morris, S.A., Y. Shibata, K. Noma, Y. Tsukamoto, E. Warren, B. Temple, S.I. Grewal, and B.D. Strahl. 2005. Histone H3 K36 methylation is associated with transcription elongation in *Schizosaccharomyces pombe*. *Eukaryot. Cell*. 4:1446–1454.

Müller, W.G., D. Walker, G.L. Hager, and J.G. McNally. 2001. Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. *J. Cell Biol.* 154:33–48.

Müller, W.G., D. Rieder, G. Kreth, C. Cremer, Z. Trajanoski, and J.G. McNally. 2004. Generic features of tertiary chromatin structure as detected in natural chromosomes. *Mol. Cell. Biol.* 24:9359–9370.

Osborne, C.S., L. Chakalova, K.E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J.A. Mitchell, S. Lopes, W. Reik, and P. Fraser. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36:1065–1071.

Ostrowski, M.C., H. Richard-Foy, R.G. Wolford, D.S. Berard, and G.L. Hager. 1983. Glucocorticoid regulation of transcription at an amplified, episomal promoter. *Mol. Cell. Biol.* 3:2045–2057.

Pokholok, D.K., C.T. Harbison, S. Levine, M. Cole, N.M. Hannett, T.I. Lee, G.W. Bell, K. Walker, P.A. Rolfe, E. Herbolsheimer, et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*. 122:517–527.

Roh, T.Y., S. Cuddapah, and K. Zhao. 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* 19:542–552.

Schneider, R., A.J. Bannister, F.A. Myers, A.W. Thorne, C. Crane-Robinson, and T. Kouzarides. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat. Cell Biol.* 6:73–77.

Strahl, B.D., P.A. Grant, S.D. Briggs, Z.W. Sun, J.R. Bone, J.A. Caldwell, S. Mollah, R.G. Cook, J. Shabanowitz, D.F. Hunt, and C.D. Allis. 2002. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol. Cell. Biol.* 22:1298–1306.

Sun, X.J., J. Wei, X.Y. Wu, M. Hu, L. Wang, H.H. Wang, Q.H. Zhang, S.J. Chen, Q.H. Huang, and Z. Chen. 2005. Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J. Biol. Chem.* 280:35261–35271.

Toohey, M.G., J.W. Lee, M. Huang, and D.O. Peterson. 1990. Functional elements of the steroid hormone-responsive promoter of mouse mammary tumor virus. *J. Virol.* 64:4477–4488.

Vakoc, C.R., M.M. Sachdeva, H. Wang, and G.A. Blobel. 2006. Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell. Biol.* 26:9185–9195.

Verschure, P.J., van Der Kraan, E.M. Manders, and R. van Driel. 1999. Spatial relationship between transcription sites and chromosome territories. *J. Cell Biol.* 147:13–24.

Walker, D., H. Htun, and G.L. Hager. 1999. Using inducible vectors to study intracellular trafficking of GFP-tagged steroid/nuclear receptors in living cells. *Methods*. 19:386–393.

Wansink, D.G., O.C. Sibon, F.F. Cremers, R. van Driel, and L. de Jong. 1996. Ultrastructural localization of active genes in nuclei of A431 cells. *J. Cell. Biochem.* 62:10–18.